Monoallelic Gene Expression In Human Neural Stem Cells

Perfect, Leo

Awarding institution:
King's College London

Monoallelic Gene Expression In Human Neural Stem Cells

Leo Perfect

Thesis submitted for the degree of

Doctor of Philosophy

at King's College London

2013

Department of Neuroscience

Institute of Psychiatry

King's College London

Abstract

Monoallelic expression is when only one of a gene's two copies is transcribed. It has long been recognised in the form of X-chromosome inactivation, genomic imprinting and random monoallelic expression (RME) of a small number of gene families. More recently, studies have reported evidence of widespread autosomal RME in human lymphoblastoid cells (Gimelbrant et al. 2007) and mouse neural stem cells (Wang et al. 2010; Li et al. 2012). However, the extent to which RME occurs in human neural tissue is unknown. Using a genome-wide analysis of allele-specific expression in human neural stem cells, I reveal that 1.6 to 2.2% of assayed autosomal genes display monoallelic expression and 0.5 to 1.1% show evidence of RME. This is largely retained after differentiation into neurons and glia. I also present evidence in support of RME altering gene transcript levels and exposing heterozygous functional variation. These results are a first demonstration that autosomal RME is widespread in human neural tissue and indicate that it is a potential source of phenotypic diversity between genetically identical neural stem cells *in vitro*. Furthermore, I find the novel RME genes reported in this thesis to be epigenetically distinct from biallelic genes in human foetal and adult brain in a manner consistent with monoallelic expression. Together, these findings support a model in which the human central nervous system is a mosaic of clones of cells, each with different gene expression potentials. Although the presence of widespread RME in the human central nervous system remains to be demonstrated *in vivo*, if present, this would be predicted to have significant implications for human neural development, function and disease. Finally, I map the genetic variants associated with the monoallelic expression of *PM20D1*, a gene located in the Parkinson's disease susceptibility locus *PARK16*.

Acknowledgements

I would like to take this opportunity to thank the many people who have helped me over the last three years. Firstly, I am grateful to Jack Price for his guidance, support and confidence in me since my appointment as a research assistant over four years ago. I would like to thank him for the freedom and financial backing he has given me to pursue my own ideas; I know this is a luxury that not all PhD students are afforded. Secondly, I wish to thank Sandrine Thuret for her advice and encouragement. The postdocs in our group, Aaron Jeffries, Dafe Uwanogho and Graham Cocks, have taught me everything I know in the lab. I thank them all for their patience, support and friendship. Nick Bray has been actively engaged in the development of my project. I thank him for his input and for the use of his precious samples. Equally, Nick's postdoc Matt Hill has always been ready to provide advice. I thank Jon Mill for allowing me to use his samples and his postdoc, Chloe Wong, for her time spent helping me run the microarrays. I thank all of the CCBB, and especially my fellow PhD students, Lotta Parviainen, Federica Buonocore and Doris Stangl, for making the last three years such an enjoyable experience. Last but not least, I would like to thank Alice and my parents. I have no doubt that, without their love and support, this work would not have been possible.

Table of Contents

Table of Figures

Table of Tables

Chapter 1. General Introduction

1.1 Gene expression

Life depends on gene expression. This fundamental process decodes genetic information into biological function. The sequential flow of information from DNA to RNA and then, typically, to protein is the central dogma of molecular biology (Crick 1970).

The initial step in gene expression is transcription. Here, general transcription factors bind to the core promoter upstream of the coding sequence. This complex enables RNA polymerase II to bind to the transcription start site and run along the template strand of the gene generating a complementary strand of RNA called precursor messenger RNA (pre-mRNA). For protein coding genes, the pre-mRNA is capped, polyadenylated and spliced to form mature mRNA. It is then exported to the cytoplasm where it engages with ribosomes to direct protein synthesis. A subset of genes directs the expression of non-coding RNAs (ncRNAs) that play an important role in regulating the expression of coding sequence.

The cells within each individual are genetically identical, yet they display enormous phenotypic diversity. It is the exquisite spatiotemporal regulation of gene expression that enables the differentiation of specialised cell types. Equally, the dynamic regulation of gene expression directed by signal transduction cascades in response to external cues is critical for maintaining homeostasis.

1.1.1 *Cis*- and *trans*-regulatory elements

Gene expression is controlled, in part, by a combination of *cis*- and *trans*-acting regulatory factors. *Cis*-regulatory elements are sequence motifs located on the same DNA molecule as the gene that they regulate. They are found upstream, downstream and within the target gene. In contrast, *trans*-regulatory factors are diffusible molecules, typically proteins or ncRNAs, which are transcribed from remotely located genes. These *trans*-factors bind to *cis*-elements to affect gene expression.

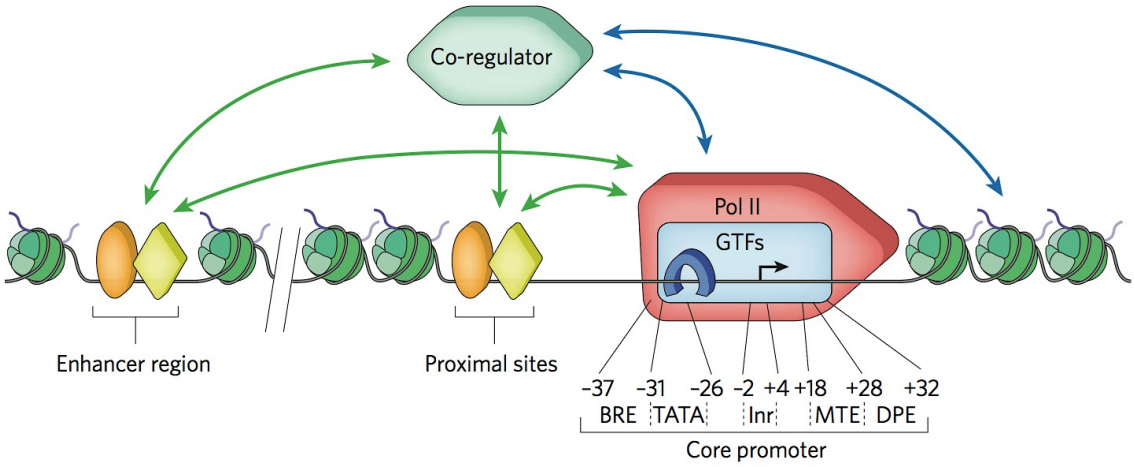


Figure 1.1 | Transcription regulatory interactions. See text below for description. Activator: orange oval; repressor: yellow diamond; Polymerase II (Pol II); general transcription factors (GTFs); B recognition element (BRE); TATA box (TATA); initiator (Inr); motif ten element (MTE); downstream promoter element (DPE) (Fuda et al. 2009).

The inter-relationship between *cis*- and *trans*-acting factors during transcription is illustrated in Figure 1.1. General transcription factors act in *trans*, binding to various *cis*-regulatory elements within the core promoter. They enable the binding of RNA polymerase II, forming the basal transcription complex. This formation is competent of driving basal, low level, transcription. Many genes have additional *cis*-regulatory

elements at proximal and distal sites to fine-tune gene expression. They are often bound by signalling molecules or tissue-specific transcription factors which can modulate the rate of transcription, acting as either activators or repressors. They do so either directly, by interacting with factors associated with the core promoter, or indirectly, via association with co-regulators. Combined, *cis*- and *trans*-regulatory elements allow for tight spatial, temporal and environment-aware regulation of gene expression.

1.1.2 Epigenetic regulation of gene expression

Epigenetic modification describes functional alteration to the genome that is maintained through cell division but, critically, involves no change to the nucleotide sequence (Berger et al. 2009). It regulates gene expression by altering the accessibility of transcription machinery to the DNA. This typically involves covalent alteration of the local chromatin (see Figure 1.2). Epigenetic processes play a key role in regulating gene expression during development, maintaining tissue-specific expression throughout life and enabling forms of allele-specific gene expression such as genomic imprinting and X-chromosome inactivation (Reik 2007; Goldmit & Bergman 2004). It is also thought to play a major role in the mediation of gene-environment interactions (Meaney 2010). Two major classes of epigenetic alteration include histone modification and DNA methylation, both of which are discussed in greater detail in the sections below.

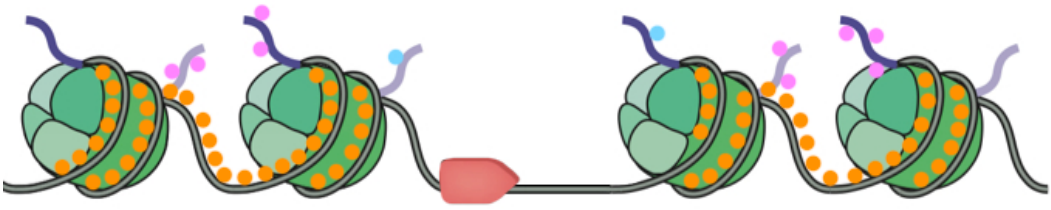


Figure 1.2 | Post-translational chromatin modification. Histone modification (blue and pink) and DNA methylation (orange) are the two main forms of epigenetic modification. They regulate the density of nucleosome packaging and, therefore, the accessibility of the DNA to the transcription machinery (red; adapted from Fuda et al. 2009).

1.1.2.1 Histone modification

DNA is tightly packaged into chromatin in eukaryote nuclei. Nucleosomes are the basic unit of chromatin and they consist of approximately 146 base pairs of DNA wrapped around two copies of each of the four histone proteins: H2A, H2B, H3 and H4. Chromatin can be broadly described as being in one of two states: euchromatic or heterochromatic. Euchromatin has more loosely coiled DNA and is transcriptionally active, while heterochromatin is tightly packed and transcriptionally inert. These states are regulated in large part by reversible post-translational modification of the histone proteins, predominantly at their N-terminals known as "histone tails". Types of histone modification include acetylation, phosphorylation and methylation.

Both histone acetylation and phosphorylation are thought to alter chromatin structure predominantly by reducing the positive charge of histones, disrupting electrostatic interactions between histones and DNA (Bannister & Kouzarides 2011). In general, this results in a less compact chromatin structure, allowing greater access to the DNA. There

are, however, exceptions to this rule as both acetylation at H4K16 and phosphorylation at H3S10 result in chromatin compaction (Shogren-Knaak et al. 2006; Wei et al. 1998).

Histone methylation occurs most frequently at lysine and arginine side chains. It does not alter the charge of histones; rather, it alters chromatin structure indirectly by regulating the binding of chromatin effectors. As such, histone methylation is site specific; it can either open or close chromatin depending on the effectors that it interacts with. For example, H3K4me3 is associated with active transcription (Barski et al. 2007) and it directly recruits euchromatin factors such as CHD1 (Sims et al. 2005), whereas H3K9me3 binds to HP1, a protein critical for gene repression and heterochromatin structure (Lachner et al. 2001).

1.1.2.2 DNA methylation

DNA methylation involves the transfer of a methyl group to a DNA cytosine residue. It occurs most frequently at CpG dinucleotides, although it is also observed at other cytosine residues (Lister et al. 2009). While CpGs are mostly dispersed throughout the genome at a low density, they also appear in densely packed regions known as "CpG islands" (CGI). DNA methylation is generally associated with the repression of gene expression (Bird 2002; Laurent et al. 2010).

DNA methylation plays a key role in regulating gene expression through development. Germline methylation is almost completely erased in the blastocyst before global *de novo* methylation is laid down at the stage of implantation (Meissner et al. 2008; Laurent et al. 2010). The *de novo* methylation is bimodal; lower density non-island CpGs are highly

methylated while CGIs escape methylation. This enables tissue-specific genes and endogenous viral sequences to be repressed while housekeeping genes with CGI promoters are expressed. Differentiation into mature cells is associated with a global reduction in methylation (Laurent et al. 2010) as the differentiating cells adopt a cell-type-specific methylation fingerprint (Han et al. 2011; Davies et al. 2012). This process permits cellular differentiation by presenting tissue specific genes to the transcription machinery only in the appropriate cells. In order for somatic cells to maintain their phenotype through cell division it is important that the methylation pattern is passed on to cellular progeny. This is achieved, at least in part, by the enzyme DNA methyl-transferase 1 (Li et al. 1992).

DNA methylation is thought to mediate gene silencing primarily by influencing chromatin structure. Early support for this notion came from the observation that unmethylated DNA transfected into mouse cells was packaged into euchromatin, while if the same sequence was pre-methylated, it remained DNase I resistant after transfection (Keshet et al. 1986). Consistent with this finding, it has been demonstrated that that DNA methylation directly induces histone H4 deacetylation as well as methylation of H3K9 and H3K4 in mice (Hashimshony et al. 2003). Together, these findings support an instructive role of DNA methylation in dictating histone modification and chromatin conformation.

1.2 Allele-specific gene expression

Autosomal genes are comprised of two alleles, one inherited from each parent. Both alleles are usually expressed at similar levels. Biallelic gene expression increases the genetic diversity of an organism, providing protection from recessive deleterious mutations. However, there are some well-characterised exceptions from this rule, where the risks associated with expressing a single allele are outweighed by the advantages. Classical examples of monoallelic gene expression include genomic imprinting, X-chromosome inactivation and autosomal random monoallelic gene expression (RME). Allelic imbalance, and monoallelic expression in extreme cases, can also occur through heterozygosity for genetic *cis*-regulatory variants. These examples are discussed in the following sections.

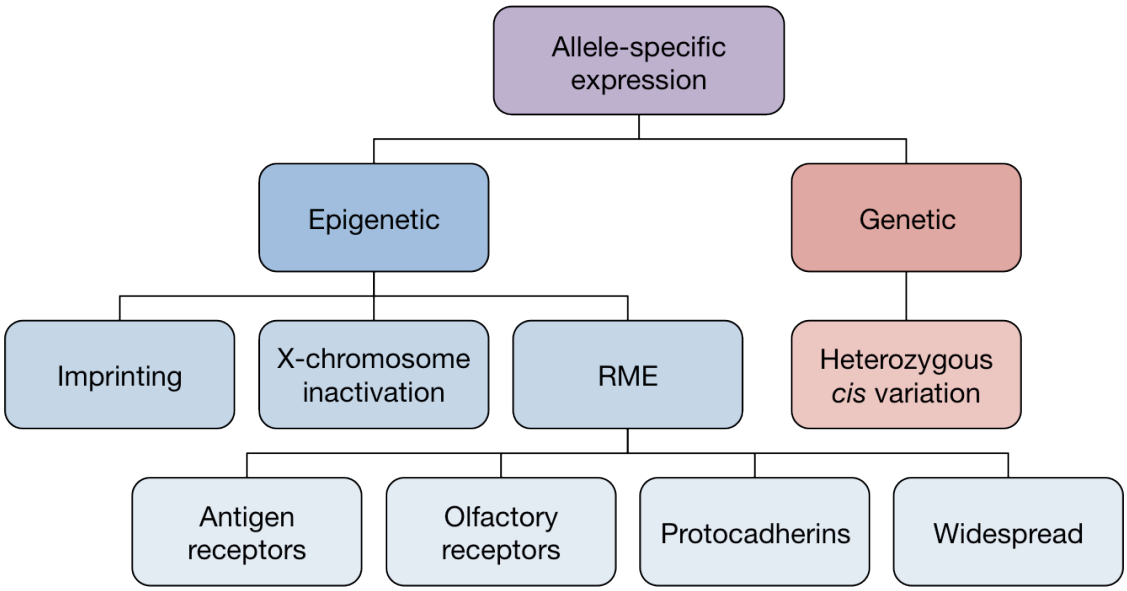


Figure 1.3 | Forms of allele-specific expression.

1.2.1 Genomic imprinting

Genomic imprinting is the most extensively studied form of monoallelic expression. It is an epigenetic process that causes select genes to be expressed from a single allele in a parent of origin-specific manner. It was first demonstrated in mouse (Surani et al. 1984; Barlow et al. 1991) and to date 118 imprinted genes have been validated in mouse and 80 have been validated in human (http://www.geneimprint.com). Imprinted genes usually appear in clusters under the control of a common imprinting control region (ICR), although this is not a universal rule. Most imprinted genes identified to date are regulators of embryonic growth, placental growth or adult metabolism (Morison et al. 2005), however, there is an increasing appreciation of its contribution to neural development, brain function and behaviour (Davies et al. 2005).

1.2.1.1 Mechanisms of genomic imprinting

Imprinted genes display differential methylation between the two parental chromosomes at ICRs (Ferguson-Smith & Sasaki 1993) which is critical for monoallelic expression (Sutcliffe & Nakao 1994; Wutz et al. 1997). The sex-specific epigenetic mark is acquired during gametogenesis and maintained in diploid cells after fertilisation. Exactly where the ICR is located in relation to the imprinted gene and how it acts to enforce monoallelic expression differs from locus to locus. The *KCNQ1* and *IGF2* clusters are two of the most studied and they are covered in greater detail below to illustrate the mechanisms typically involved in genomic imprinting.

For the maternally expressed *KCNQ1* cluster, the ICR lies within the promoter of *KCNQ1* overlapping transcript 1 (*KCNQ1OT1*), an antisense ncRNA (see Figure 1.4). The paternal ICR is unmethylated and as a result *KCNQ1OT1* is transcribed, leading to inactivation of the imprinted genes on the paternal chromosome (Mancini-Dinardo et al. 2006). *KCNQ1OT1* initiates the organisation of a nuclear silencing domain around the cluster by recruiting the histone-modifying proteins of polycomb repressive complex 2 (PRC2; Terranova et al. 2008; Redrup et al. 2009). In contrast, the maternal ICR is methylated, resulting in the silencing of *KCNQ1OT1* and expression of the genes within the cluster. Analogous ncRNA-dependent mechanisms have been described for the majority of other investigated maternally expressed imprinted genes, including the *Igf2r* (Sleutels et al. 2002) and *GNAS* (Plagge & Kelsey 2006) clusters.

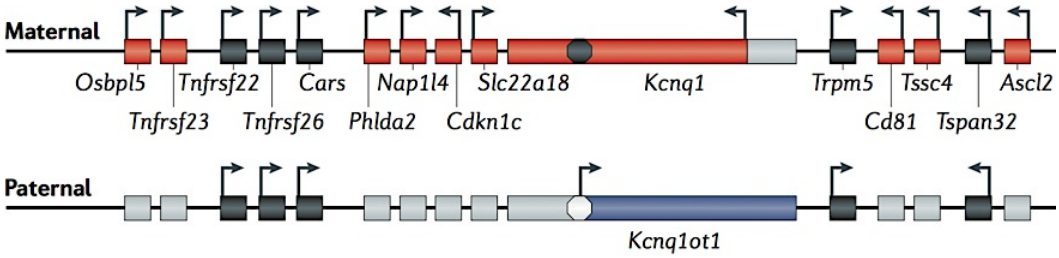


Figure 1.4 | The *KCNQ1* cluster. Red bar: maternally expressed gene; Blue bar: paternally expressed gene; Black bar: biallelically expressed gene; Grey bar: repressed gene; Black octagon: methylated ICR; Grey octagon: unmethylated ICR (Ferguson-Smith 2011).

The *IGF2-H19* cluster is regulated by an insulator dependent mechanism (see Figure 1.5). *IGF2* and *INS2* are exclusively expressed from the paternal chromosome. The differentially methylated ICR acts through its interaction with the zinc-finger protein CTCF, which only binds to the unmethylated maternal chromosome (Bell & Felsenfeld

2000). Here, CTCF blocks enhancer interaction with the maternal *IGF2* promoter, resulting in repression of *IGF2* and *INS2* on the maternal chromosome. Conversely, CTCF does not bind to the methylated paternal ICR and the paternal alleles are expressed. In contrast, the imprinted *H19* is expressed from the maternal chromosome. When CTCF is bound to the unmethylated maternal ICR, the chromatin topology is altered in a way that enables interaction between *H19* and the enhancer sequences which drive its expression (Murrell et al. 2004).

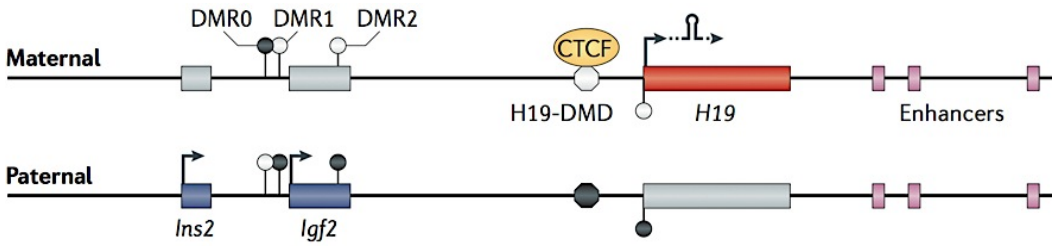


Figure 1.5 | The *IGF2R–H19* cluster. Red bar: maternally expressed gene; Blue bar: paternally expressed gene; Grey bar: repressed gene; Black octagon: methylated ICR; Grey octagon: unmethylated ICR; Black lollipop: methylated secondary DMR; Grey lollipop: unmethylated secondary DMR (Ferguson-Smith 2011).

1.2.1.2 Regional and temporal control of genomic imprinting

It is becoming clear that the monoallelic status of some imprinted genes is under strict spatiotemporal control. Of 82 mouse imprinted genes investigated, 23 show imprinted expression limited to a single tissue type (Prickett & Oakey 2012). Of this subset, 57% are exclusively imprinted in extra-embryonic tissues and 26% are solely imprinted in the brain. There is another subset of nine imprinted genes, which are monoallelic in all but one tissue type where they are expressed from both alleles. A recent study illustrated the potential importance of dynamic control of the monoallelic status of imprinted genes. It

has been demonstrated that the maternally imprinted NOTCH ligand *Dlk1* reverts to biallelic status in neural stem cells and niche astrocytes at postnatal stages, and that this switch is required for normal neurogenesis (Ferrón et al. 2011). The change to biallelic expression is associated with *de novo* DNA methylation at the ICR and results in an increase in *Dlk1* expression. The extent to which this process occurs in other imprinted genes is not yet clear and will require genome wide assessment of allele-specific expression in pure cell populations from different tissues. Nevertheless, this study illustrates the importance of allele-specific expression for controlling gene dosage during development.

1.2.1.3 Genomic imprinting and disease

The functional haploidity caused by imprinting means that inactivation of the expressed allele leads to a complete loss of function from that gene. The most frequently cited examples of genomic imprinting disorders are a result of *de novo* genetic variants, rather than any dysregulation of imprinting per se. For example, Angelman and Prader-Willi syndromes involve genomic deletions overlapping the imprinted *SNRPN* cluster (see Figure 1.6; Christian et al. 1995). Angelman syndrome presents when the deletion occurs on the maternal chromosome resulting in the loss of the paternally imprinted *UBE3A* (Kishino et al. 1997). Conversely, Prader-Willi syndrome is caused by deletion on the paternal chromosome resulting in a lack of expression of the maternally imprinted genes of this region (Ohta et al. 1999).

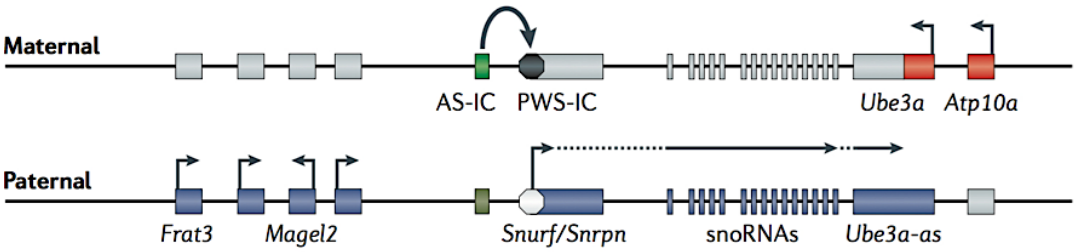


Figure 1.6 | The SNRPN cluster. Red bar: maternally expressed gene; Blue bar: paternally expressed gene; Grey bar: repressed gene; Black octagon: methylated ICR; Grey octagon: unmethylated ICR (Ferguson-Smith 2011).

Epigenetic dysregulation of genomic imprinting can also result in disease. Beckwith-Wiedemann syndrome (BWS) is a prime example; aberrant methylation of the imprinted genes at 11p15 is observed in more than 50% of cases (Cooper et al. 2005). BWS is a congenital overgrowth disorder with an incidence of one in 13,000. The most frequent aberration is hypomethylation of the *KCNQ1* cluster ICR. It results in biallelic expression of *KCNQ1OT1* and the subsequent downregulation of tumour suppresser *CDKN1C* (Lee et al. 1999; Diaz-Meyer et al. 2003). In approximately 5% of cases, BWS is caused by hypermethytlation of the *IGF2* ICR and the resulting biallelic expression of growth promoting *IGF2* (Reik et al. 1995). Loss of imprinting is the most abundant alteration in cancer (Feinberg et al. 2006). Epigenetic dysregulation of imprinted genes can promote cancer either by activation of a normally silenced copy of an oncogene such as *IGF2* (Ohlsson et al. 1999), or silencing of the normally active copy of a tumour-suppressor gene such as $p57^{KIP2}$ (Diaz-Meyer et al. 2003).

1.2.1.4 The imprinting paradox

The disorders described above highlight the evolutionary paradox of imprinting; its evolution and conservation indicates that there is a selective advantage that outweighs the vulnerability that monoallelic expression exposes. Genomic imprinting exists in plants and mammals; however, mechanistic differences indicate that they evolved independently (Scott & Spielman 2006). In mammals, most imprinted genes function to control prenatal growth, placenta formation and brain development (Morison et al. 2005). The parental conflict hypothesis is the dominant theory describing the evolution of genomic imprinting in mammals (Moore & Haig 1991). It states that imprinting is the result of a tug of war between parental chromosomes over access to maternal resources for developing offspring. This theory rationalises the imprinting of several paternally expressed growth promoting genes (e.g. *IGF2*, *KCNQ1OT1*, *Air*) and maternally expressed growth inhibitors (e.g. *Igf2r*, *CDKN1C*, *GRB10*). However, many imprinted genes are hard to reconcile with the parental conflict hypothesis (Haig 2004). The coadaptive theory has been proposed to explain the imprinting of genes associated with postnatal maternal nurturing, stating that it enabled rapid fixation of desirable coadaptive maternal-offspring traits (Curley et al. 2004; Wolf & Hager 2006). These theories are neither mutually exclusive nor all encompassing; different imprinted genes were likely under different selective pressures. Further research into the physiological roles of imprinted genes should provide a more satisfying resolution to the imprinting paradox.

1.2.2 X-chromosome inactivation

X-chromosome inactivation (XCI) was the first described form of monoallelic expression (Lyon 1961). It is the process where one of the female X-chromosomes is transcriptionally silenced to provide gene dosage compensation between males and females (Payer & Lee 2008). During development, mammalian females go through two kinds of XCI: imprinted and random.

1.2.2.1 Imprinted XCI

In a process referred to as imprinted XCI, the paternal X-chromosome is silenced in the female pre-implantation mouse embryo (Huynh & Lee 2003). Inactivation persists until the blastocyst stage, where the epigenetic marks are erased and the paternal X-chromosome is reactivated (Mak et al. 2004). The paternal and the maternal X-chromosome then have an equal chance of inactivation by random XCI (discussed below). As this process occurs before the embryonic stem cell stage, the majority of research has been carried out in mouse. The epigenetic state of human X-chromosomes in the pre-implantation embryo is not well understood, although a recent study indicates that genes from both female X-chromosomes are active at this stage (Okamoto et al. 2011).

1.2.2.2 Random XCI

Random XCI is initiated in all eutherian females at the stage of the late blastocyst. Unlike genomic imprinting or imprinted XCI, it occurs in a stochastic manner with respect to parent of origin (Gardner & Lyon 1971). First, the number of X-chromosomes in the cell

is counted and all but one are randomly selected for inactivation. This is mediated by physical interactions between the two X-chromosomes, where it is hypothesised that a mutually exclusive mark is laid down (Bacher et al. 2006; Xu et al. 2006; Masui et al. 2011). The ncRNA X inactive specific transcript (*XIST*), transcribed from the future inactivated X-chromosome (Xi), is both necessary and sufficient for silencing of the chromosome (Brown et al. 1991; Wutz & Jaenisch 2000; Chow et al. 2007). *XIST* coats the entire Xi in *cis*, leading to its inactivation through the recruitment of silencing factors such as PRC2 and the subsequent condensation of the X-chromatin (Clemson et al. 1996; Silva et al. 2003; Marks et al. 2009). *XIST* is tethered to the Xi by the transcription factor YY1, preventing it from operating in *trans* on the active X-chromosome (Jeon & Lee 2011). Once established, the Xi is irreversibly maintained through future cell divisions, resulting in somatic mosaicism in adult females (Wutz & Jaenisch 2000). Maintenance of XCI is managed by a diverse array of chromatin factors, including Orc2 and HP1a, which keep the Xi packaged tightly in heterochromatin (Chan et al. 2011). Nuclear organization is also thought to play a role in the maintenance of XCI through cell division; in both mouse and human, it has been observed that the Xi is sent to a distinct nuclear compartment, rich in the chromatin remodeling factor SNF2H, when it undergoes replication (Clemson et al. 2006; Zhang et al. 2007).

1.2.2.3 Skewed XCI

As stated above, adult females are mosaics of cells with different active X-chromosomes. This is advantageous as it reduces the effect of deleterious X-linked mutations. While most females are comprised of an even representation of the two X-chromosomes, approximately 30% have greater than 20% skew in XCI (Amos-landgraf et al. 2006). XCI

skewing is concordant between monozygotic twins (Wong et al. 2011). Extreme skewing of greater than 95% is very rare in the population but frequently observed in phenotypically normal carriers of X-linked mutations, where the X with the deleterious mutation is preferentially silenced (Plenge et al. 2002). This is thought to be a result of clonal selection during development; if cells expressing the complement of alleles from one X-chromosome have a growth advantage over cells expressing the alternate X-chromosome it would result in clonal outgrowth (Belmont 1996; Muers et al. 2007).

1.2.3 Autosomal random monoallelic expression

Random monoallelic expression (RME) describes the regulation of some autosomal genes that can be exclusively expressed from either allele, or in some cases both alleles, and the allelic choice is thought to be random. Once a cell has committed to an allele during development its cellular progeny maintains the choice. RME has classically been described in terms of a small number of gene families including antigen receptors, olfactory receptors and protocadherins. In these disparate gene families, RME plays a similar role: increasing cellular specificity and tissue diversity. Of particular interest to this thesis are the recent observations that indicate that RME is more widespread than had previously been appreciated (Gimelbrant et al. 2007; Wang et al. 2010; Li et al. 2012).

1.2.3.1 Antigen receptors

Monoallelic expression of antigen receptors, also known as allelic exclusion, was the first described form of autosomal RME (Pernis et al. 1965). The genes encoding antigen

receptors share a unique property in that they undergo DNA recombination before expression. DNA recombination along with monoallelic expression forms the central paradigm of allelic exclusion: “one lymphocyte - one antigen receptor” (Burnet 1957). It ensures that only a single receptor variant is present on each cell, enabling antigen specificity. For the sake of clarity I will focus on the mechanisms underpinning the allelic exclusion of immunoglobulin genes of B cells, but analogous mechanisms apply for T cell receptors (Brady et al. 2010).

B cell antigen receptors contain both heavy and light chain immunoglobulins, which are generated by the recombination of a limited number of V, D and J segments at the pro-B cell stage of development (Bassing et al. 2002). Exactly how the random allelic choice is coordinated is not yet clear. The probabilistic model states that both alleles are equally susceptible to recombination, and that the stochastic selection is born out of competition for a cooperative factor (Perry et al. 1980). However, it is becoming clear that recombination is preceded by epigenetic events that differentiate the two alleles. A key factor appears to be that antigen receptor loci start to replicate asynchronously after implantation (Mostoslavsky et al. 2001). Replication timing is determined randomly and once established, the pattern is maintained in the cell’s progeny. The early replicating allele is selected for recombination 80% of the time (Goldmit et al. 2004). Other asymmetric epigenetic alterations precede rearrangement, including chromatin modification, demethylation and locus contraction (Goldmit et al. 2004; Mostoslavsky et al. 1998; Sayegh et al. 2005). The instructive model states that the asymmetry of these alterations is determined by asynchronous replication, and they then influence which allele is selected for recombination (Mostoslavsky et al. 2001). As recombination is initiated, the homologous chromosomes pair and the uncleaved allele is marked and

subsequently repositioned to the pericentromeric heterochromatin where it is sheltered from the recombination machinery (Hewitt et al. 2009). Interestingly, the excluded allele can be recalled if the primary allele fails to make a functional receptor (Casellas et al. 2007). Once a functional receptor expressed, allelic exclusion is stably maintained by downregulation of recombination factors (Corcoran 2005), locus decontraction (Roldán et al. 2005) and histone deacetyltion (Stanton & Brodeur 2005).

1.2.3.2 Olfactory receptors

Olfactory receptors (ORs) are G-protein coupled receptors responsible for the detection of odorants. They are subject to both monoallelic and monogenic expression; therefore, only a single allele from a single OR gene is expressed per cell (Chess et al. 1994). This ensures specific detection of olfactory stimuli by olfactory sensory neurons. ORs also play an important role in the wiring of the olfactory system during brain development, ensuring that olfactory sensory neurons expressing the same OR converge on the same glomerus in the olfactory bulb (Imai et al. 2010). While it is understood that the expression of a functional OR allele initiates a feedback mechanism to prevent the expression of others (Serizawa et al. 2003; Lewcock & Reed 2004), the molecular mechanisms that regulate the initial OR selection have remained elusive. The observation that an enhancer region, the H element, associates with multiple OR gene promoters in *trans* lead to the suggestion that this element played a central role in the stochastic activation of a single OR allele (Lomvardas et al. 2006). However, subsequent work where the H element was genetically targeted refutes this hypothesis, as only the expression of ORs located near to the H element were disrupted (Fuss et al. 2007). A recent study in mice has demonstrated that all OR genes are packaged into compacted

heterochromatin prior to transcription (Magklara et al. 2011). The authors propose that, after the whole cluster has been silenced, a limited factor derepresses a randomly selected allele enabling its transcription. The expressed allele then initiates a feedback mechanism that prevents the selector from reactivating any other alleles. Further research is required to test this hypothesis and elucidate the control of RME in ORs. It is noteworthy that ORs also display asynchronous replication, a hallmark of monoallelic expression (Chess et al. 1994). While it is not clear what role it plays in regulating OR expression; one could hypothesise that it determines which chromosome is first exposed to the proposed reactivating factor. As with antigen receptor allelic exclusion, it has been demonstrated that if a non-functional OR is selected for expression the cell is able to reverse the decision and activate an alternate allele (Serizawa et al. 2003; Lewcock & Reed 2004).

1.2.3.3 Protocadherins

Protocadherins are cell surface proteins belonging to the cadherin superfamily of cell adhesion proteins. They are enriched at synapses and function during neural development (Kohmura et al. 1998; Wang et al. 2002). The three protocadherin sub families, Pcdhα, Pcdhβ and Pcdhγ, are all located in a single cluster (Wu & Maniatis 1999). Pcdhα and Pcdhγ gene structure is unusual in that they consist of twelve tandemly arrayed variable region (V) exons and three constant (C) exons. It has been demonstrated that Pcdhα and Pcdhγ genes are subject to monoallelic yet combinatorial expression of the variable exons (see Figure 1.7; Esumi et al. 2005; Kaneko et al. 2006).

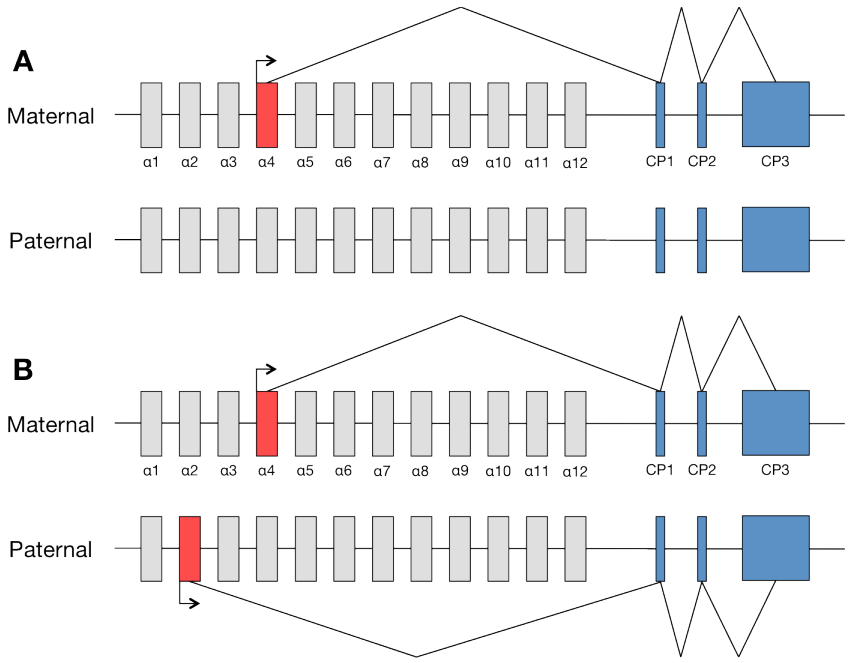


Figure 1.7 | Monoallelic and combinatorial expression of protocadherin genes. A variable exon is randomly selected (red) and spliced with the three constant exons (blue) to generate the mature mRNA. **A)** Monoallelic expression of a single isoform. **B)** Combinatorial monoallelic expression.

For each Pcdhα and Pcdhγ transcript, cells can express either the maternal, paternal, both or neither allele. This results in enormous diversity in protocadherin gene transcript sets between different cells. The expressed protocadherin proteins form heteromultimers exhibiting isoform-specific homophilic binding (Murata et al. 2004; Schreiner & Weiner 2010). Protocadherin complexes are hypothesised to provide a cell-specific fingerprint facilitating self-recognition, analogous to Dscam1 in invertebrates (Zipursky & Sanes 2010). Little is known about the mechanisms that regulate the monoallelic expression of protocadherins. A model has been proposed in which there is a limiting number of activators in each nucleus, and it is the competition for this limiting factor that gives rise to the stochastic monoallelic selection of each variable exon (Chess 2005). Further work

is required to unravel the regulation of protocadherin gene expression as well as their role in neural development.

1.2.3.4 Widespread random monoallelic expression

Recent evidence indicates that RME is not restricted to the established gene families reviewed above; instead, it is widespread throughout the human autosome. Using genotyping microarrays, Gimelbrant et al. (2007) carried out a genome-wide assessment of monoallelic expression in clonal human B-lymphoblastoid cell lines. They found evidence of RME in approximately 9% of the genes assayed. The genes were scattered throughout the genome and the allelic choice was made at the gene, rather than chromosome, level. The authors found these genes to be enriched for cell surface proteins and more than twice as likely to be located near noncoding sequences associated with human lineage-specific accelerated evolution. They also demonstrate monoallelic expression of a subset of these genes *in vivo*, in placental micropatches and fresh peripheral blood mononuclear cells. The study was replicated with mouse lymphoblastoid cells to assess the conservation between species (Zwemer et al. 2012). A similar frequency of RME was detected in mice as in human and, importantly, a significant overlap between genes subject to RME in mouse and human was observed. This, along with the enrichment for cell surface proteins, indicates that gene selection is not random; rather, some genes are predisposed for RME. Support for these findings comes from a recent next-generation sequencing study in clonal mouse neural stem cells (Li et al. 2012). The authors found evidence for RME in approximately 4% of autosomal genes and, notably, monoallelic expression was associated with a 30-35% reduction in transcript

levels. This demonstrates that widespread RME occurs in murine neural tissue *in vitro*, consistent with previous findings (Wang et al. 2007; Wang et al. 2010).

Together, these studies reveal an unprecedented source of epigenetic diversity between clonal cell lines. For each locus subject to RME, a cell can be in one of three cell states: biallelic, monoallelic for the paternal allele or monoallelic for the maternal allele. This could potentially drive functional variation between genetically identical cells by altering gene dosage, or by exposing functional heterozygosity. The extent to which RME occurs in human neural tissue, and the impact that it may have on brain function, development and disease remains to be determined.

1.2.4 Genetic *cis*-regulatory variants

The HapMap and 1000 Genomes projects have revealed that genetic variation is widespread on the human genome. Variants, such as single nucleotide polymorphisms (SNPs) and copy number variants (CNVs), can affect gene expression by altering regulatory elements or dosage of the protein coding sequence itself. While functional alteration of a *trans*-element would be predicted to alter both alleles equally, *cis*-regulatory variants only affect the expression of the allele on which they reside. Therefore, heterozygosity for genetic *cis*-regulatory variants can give rise to allelic expression imbalance (AEI), and in extreme cases, monoallelic expression.

Genetic *cis*-regulatory variation represents a significant source of AEI in the human genome. This was first demonstrated by a series of studies that assayed selected genes to estimate the prevalence of *cis*-acting variation in the transcriptome (Yan et al. 2002; Lo et

al. 2003; Bray et al. 2003). Recent genome-wide studies indicate that at least 20% of human genes display AEI caused by genetic *cis*-regulatory variation (Zhang et al. 2009; Ge et al. 2009; Heap et al. 2010). These variants are found upstream, downstream and within the gene they influence (Cheung et al. 2005). The presentation of AEI is dictated by the *cis*-regulatory element that is affected, as well as the manner in which it is altered. For example, one way in which *cis*-acting variants influence gene expression is by altering the affinity of a *trans*-factor to its binding site. It can do so by increasing, decreasing or abolishing binding. This can lead to tissue- or temporal-specific AEI when the variant is in the binding site of a *trans*-factor only present in certain cell types or developmental stages (Heinzen et al. 2008; Buonocore et al. 2010). Additional ways in which genetic *cis*-regulatory variants act to alter gene expression include disruption of RNA splicing and altering the stability of the message by modifying RNA structure or the polyadenylation site.

Recently, assays of allele-specific expression have been the preferred method for the detection of *cis*-acting variants (see section 1.2.5). This approach has an advantage over expression quantitative trait loci (eQTL) approaches in that each allele acts as an internal control for the other, eliminating confounding inter-sample variables.

1.2.5 Allele-specific expression assays

Assays of allele-specific expression typically utilise heterozygous SNPs as copy-specific tags to distinguish between a gene's two alleles. Individuals heterozygous for the tag SNP can then be assayed for allele-specific expression by measuring the relative abundance of each gene copy (see Figure 1.8).

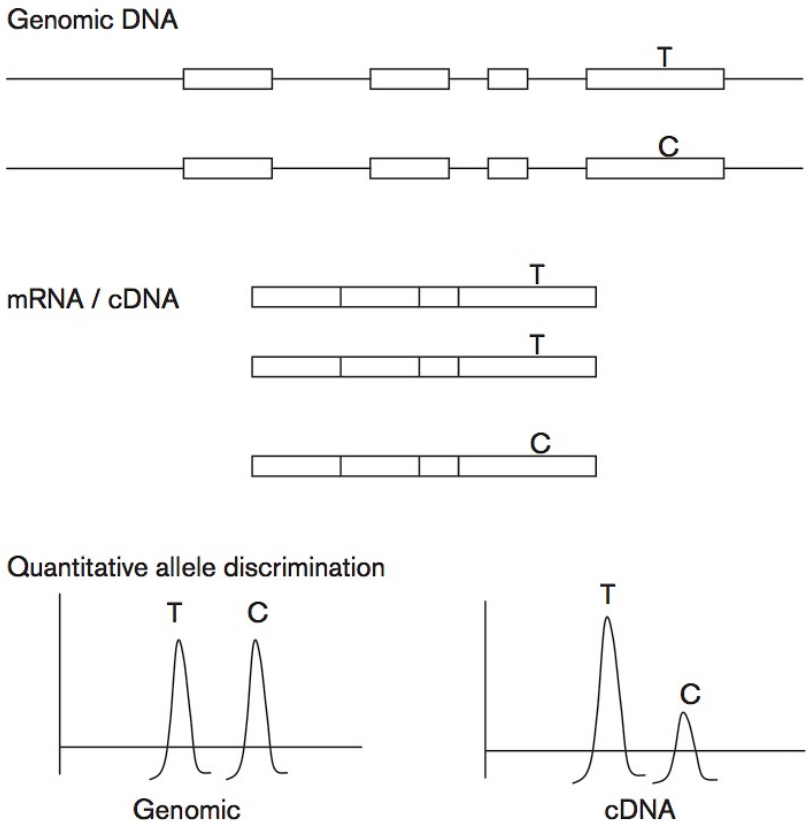


Figure 1.8 | Analysis of allele-specific expression. A heterozygous coding SNP (T/C in this example) is used to distinguish between the two gene copies. A quantitative method of allele discrimination is then used to measure the relative abundance mRNA / cDNA from each chromosome. A 1:1 allele ratio is observed in the genomic DNA, whereas a 2:1 imbalance is observed in the cDNA. Figure adapted from Bray and O'Donovan (2006).

Techniques used in the 1980s and early 1990s for the investigation of allele-specific expression were qualitative; they were only able to detect the presence or absence of a given allele (Nozari et al. 1986; Wu et al. 1989; Bourguin et al. 1990; Smrzka et al. 1995). In recent years, the need for quantitative measurement of the two alleles has been met by the development of a variety of techniques: allele-specific quantitative polymerase chain reaction (qPCR) (Zhu et al. 2004), polymerase colonies (Butz et al. 2004), sequencing (Sanger et al. 1977; Ge et al. 2005) and single base primer extension (Singer-

Sam et al. 1992; Bray & O'Donovan 2006). Modern single base primer extension assays utilising fluorescently tagged dideoxynucleotides and capillary gel electrophoresis are highly quantitative, enabling the detection of subtle allelic distortion (Norton et al. 2002; Bray & O'Donovan 2006).

Genome-wide assays have also been developed working on the same principle, enabling high throughput non-hypothesis driven assessment of allele-specific expression. Genotyping microarray-based approaches have been successfully implemented in the last five years for the detection of novel imprinted genes (Pollard et al. 2008), autosomal RME (Gimelbrant et al. 2007) and genetic *cis* variation (Ge et al. 2009). Recently, next generation sequencing (RNA-seq) has superseded microarrays in the investigation of allele-specific expression (Degner et al. 2009; Zhang et al. 2009; Heap et al. 2010; DeVeale et al. 2012). RNA-seq data is highly quantitative and its digital nature lends itself to the detection of allele-specific expression. Additionally, RNA-seq potentially enables the analysis of every heterozygous SNP in the transcriptome. The development of standardised RNA-seq allele-specific expression data analysis is expected to overcome technical biases observed in early studies (Rozowsky et al. 2011; Skelly et al. 2011).

1.3 Neural stem cells

Neural stem cells (NSCs) have been a focus of medical research in recent years due to their therapeutic potential as well as their application for modelling neural development and disease. Rather than a specific cell type, "neural stem cell" is an umbrella term covering a lineage of distinct cell populations that share the defining characteristics of self-renewal and the ability to generate neurons and glia. Typically, the later the NSC population emerges during development the more fate restricted it is. Various NSC populations can be isolated at different stages of development, or generated by *in vitro* differentiation of embryonic stem cells (ESCs; see Figure 1.9).

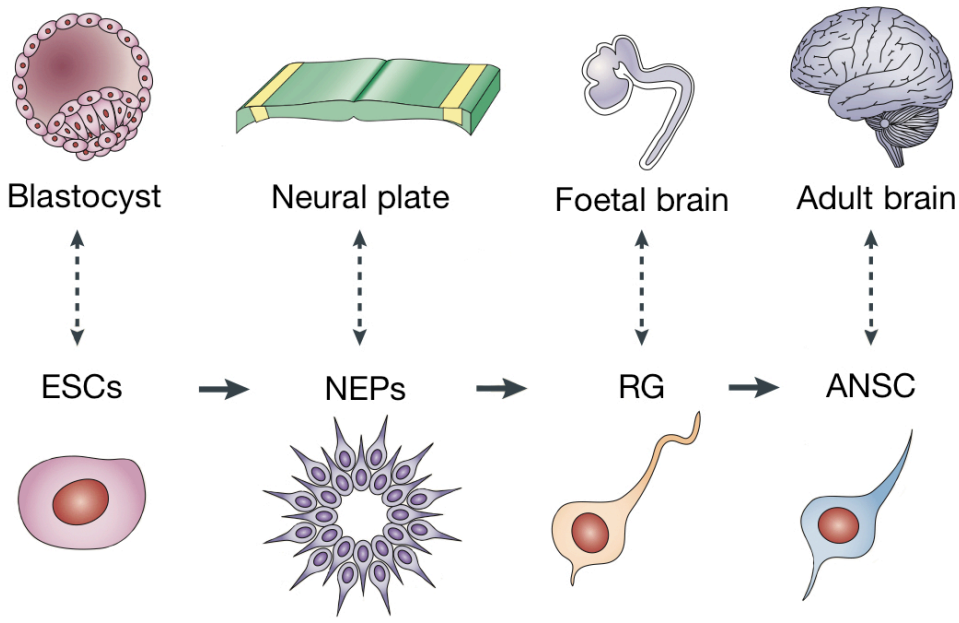


Figure 1.9 | NSC subtypes. NSCs can be isolated from the neuroepithelium, foetal brain or adult brain. Alternatively, less mature ESCs can be isolated and neuralised *in vitro* to form the NSC subtypes. NEP: neuroepithelial progenitor; RG: radial glia; ANSC: adult neural stem cell. Figure adapted from Conti and Cattaneo (2010).

Neuroepithelial progenitors (NEPs) are the most immature NSC subtype with the broadest differentiation potential. Appearing in the neural plate, they are responsible for the preliminary surge of neurogenesis via the generation of intermediate basal progenitors (Haubensak et al. 2004). NEPs are characterised by the expression of the markers PAX6, SOX1 and OTX2. They can be isolated directly from the neural plate (Elkabetz et al. 2008) or generated *in vitro* by SMAD inhibition of ESCs (Chambers et al. 2009). These cells are highly responsive to patterning signals that direct differentiation towards region-specific neuronal fates and they are capable of adopting both central and peripheral nervous system fates (Elkabetz et al. 2008). While NEPs represent an ideal population of NSCs for many *in vitro* applications, stable long-term expansion has not yet been achieved.

Radial glia (RG), spawned from NEPs in the neural tube, are the major source of neurons in the foetal brain and spinal cord (Malatesta et al. 2000, 2003; Anthony et al. 2004; Anthony & Heintz 2008). The majority loose their neuronal lineage towards the end of development where they terminally differentiate into astrocytes (Culican et al. 1990). RG are a heterogeneous cell population characterised by their bipolar morphology and astroglial marker expression. They are more lineage restricted than NEPs, retaining distinct spatiotemporal identities *in vitro* after isolation (Malatesta et al. 2003; Pinto et al. 2008). RG-like NSCs can also be generated *in vitro* by exposing less mature SOX1 positive neural progenitors to the mitogens fibroblast growth factor 2 (FGF2) and epidermal growth factor (EGF; Luciano Conti et al. 2005). These cells can be expanded as a homogenous population, which retain differentiation potential after extended *in vitro* expansion (Glaser et al. 2007).

Contrary to the long-held belief that the human brain does not contain proliferative cells, it is now understood that NSCs are preserved in the adult central nervous system (Kuhn et al. 1996; Eriksson et al. 1998). For the most part, adult NSCs reside in two neurogenic niches: the subgranular zone (SGZ) of the dentate gyrus (Kuhn et al. 1996; Eriksson et al. 1998) and the subventricular zone (SVZ) of the forebrain (Corotto et al. 1993; Barbaro et al. 2004). It has been demonstrated that the NSCs that populate the mouse SVZ are direct decedents of foetal RG (Merkle et al. 2004). In animal studies, adult neurogenesis has been shown to be activated on neuronal loss (Arvidsson et al. 2002; Tattersfield et al. 2004) and hippocampal neurogenesis to be associated with learning and memory (Deng et al. 2010) as well as affective disorders (Samuels & Hen 2011). The physiological role of adult NSCs in humans remains unknown, largely due to limitations of the available methodology. Human adult NSCs have successfully been isolated and expanded *in vitro*, demonstrating their self-renewal, multipotency and ability to generate physiologically active neurons (Johansson et al. 1999; Kukekov et al. 1999; Roy et al. 2000).

Induced pluripotent stem cells (iPSCs) represent an additional source of NSCs. iPSCs are generated *in vitro* by the forced expression of *OCT4*, *SOX2*, *KLF4* and *c-MYC* in somatic cells (Takahashi & Yamanaka 2006). They are ESC-like and can be differentiated *in vitro* to generate various NSC populations (Chambers et al. 2009). While iPSCs closely resemble ESCs in terms of pluripotency (Zhao et al. 2009) and gene expression profiles (Stadtfeld et al. 2010), questions remain whether their epigenetic reprogramming is complete. Epigenetic memory has been observed from the somatic cell of origin (Doi et al. 2009; Kim et al. 2010) as well as the accumulation of novel aberrant epigenetic states (Doi et al. 2009; Lister et al. 2011). Nevertheless, iPSCs represent an attractive source of NSCs for clinical purposes as well as for modelling development and disease.

1.3.1 NSC expansion *in vitro*

The expansion of homogeneous, phenotypically stable NSCs *in vitro* is critical for many applications, from models of neurogenesis to cell replacement therapies. There are two well established approaches to expand NSCs in culture: neurospheres and monolayers.

Neurosphere assays have been used in some of the pioneering studies that first sought to characterise NSCs *in vitro* (Reynolds & Weiss 1992; Uchida et al. 2000). The cells, acquired by microdissection of primary tissue or by ESCs neuralisation, are plated as single cell suspensions in N2 and/or B27 medium supplemented with EGF and FGF2. Differentiating cells are expected to die, while the NSCs proliferate and generate clonal aggregates. Neurospheres enable the expansion of RG-like NSCs that are capable of differentiating into neurons, astrocytes and oligodendrocytes (Reynolds & Weiss 1992, 1996). The neurosphere assay has two key limitations that have resulted in a decline in its popularity. Firstly, they lose neurogenic potential after extended passages (Fricker et al. 1999). Secondly, they generate a highly heterogeneous population consisting of only a small proportion of genuine neurosphere-forming NSCs (Suslov et al. 2002; Reynolds & Rietze 2005). Combined, these factors make neurospheres a suboptimal method to conduct informative assays of neurogenesis or generate well characterised, phenotypically stable cells for clinical applications.

Monolayer assays have become the method of choice for NSC culture in recent years. Cells are plated in adherent tissue culture vessels, often coated with extracellular matrix proteins such as laminin and fibronectin, in the presence of EGF and FGF2. Using this approach, it is possible to expand ESC-derived late-NEPs (Koch et al. 2009) and ESC-,

foetal brain- and adult brain-derived RG-like NSCs (Conti et al. 2005; Pollard et al. 2006). On exposure to differentiating conditions, these cells are capable of generating glia and electrically excitable neurons. A key advantage that this method offers over neurospheres is that it facilitates the generation of homogenous clonal cells lines that can be profiled for gene expression as well as growth and differentiation characteristics. Complete and proper characterisation is essential for transplantation applications as well as for models of disease and development. The monolayer system also generates a greater proportion of neurons and shows reduced karyotypic abnormalities after extended passaging than neurospheres (Goffredo et al. 2008).

1.3.2 Conditional immortalisation of NSCs

While progress has been made in the long-term expansion of NSCs *in vitro*, they eventually undergo growth arrest and senescence, limiting their use for large-scale applications. In addition, the inherent tumourigenic potential of NSCs is a hurdle in the way of cell transplantation therapies. A technique called conditional immortalisation tackles these problems, improving both the efficient long-term propagation of uniform clonal NSC lines as well as their safety after transplantation (Pollock et al. 2006). In this study, cells extracted from first trimester human foetal cortex were transduced with the conditional immortalising transgene c-MycER^{TAM}. The construct encodes the growth-promoting *c-Myc* fused to a hormone receptor that is driven by the synthetic 4-hydroxy-tamoxifen (4-OHT). These cells are multipotent and showed therapeutic benefit in a rat model of stroke (Pollock et al. 2006). The cell line, CTX0E03, is currently being used in the first clinical trial of NSC therapy for stoke patients (ClinicalTrials.gov identifier: NCT01151124). CTX0E03 and similar NSC lines are used in this thesis.

1.4 Study aims

In this introduction I have highlighted the critical role that different forms of monoallelic expression play in regulating gene expression through development and adulthood. With recent studies indicating that RME is widespread across the autosome (Gimelbrant et al. 2007; Wang et al. 2010; Li et al. 2012), the intriguing possibility arises that the human central nervous system is a mosaic of different clones of cells with different combinations of monoallelic and biallelic genes. Neighbouring cells could differ in terms of gene dosage as well as by exposing functional heterozygosity. This would be expected to have significant implications for brain development, function and disease.

The analysis of RME has been limited by the difficulty of detection in mixed cell populations *in vivo*. At the time of this study, no genome wide analysis of allele-specific expression has been carried out in human neural tissue with the power to detect RME. Therefore, the extent to which this process occurs in the developing and adult human central nervous system is unknown.

Using a genome-wide screen for monoallelic expression in clonal human NSCs I aim to assess the extent and nature of autosomal monoallelic expression in human neural tissue.

Specifically, this thesis aims to:

1) Determine the extent of autosomal monoallelic expression in clonal human NSC lines
2) Investigate autosomal monoallelic gene expression through NSC differentiation
3) Explore the functional implications of RME
4) Screen for novel imprinted genes

Chapter 2. General materials and methods

This section details methods relevant to multiple result chapters. Each results chapter contains an additional materials and methods section with methods specific to that chapter.

2.1 Cell Culture

2.1.1 Cell lines

The human NSC lines employed in this study were generated from three different foetal tissue types: spinal cord, cerebral cortex and striatum. Each tissue type was derived from a different donor and gave rise to three clonal cell lines (see Figure 2.1).

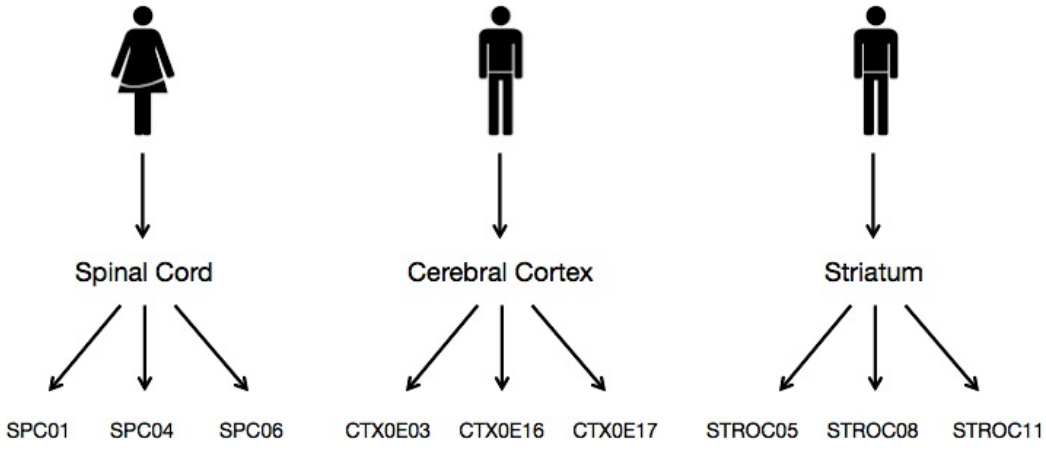


Figure 2.1 | The relationship between the human NSC lines. The spinal cord lines were generated from a female donor, while the cerebral cortex and striatum lines were derived from two different male donors.

The cortical and striatal lines were generated by ReNeuron Ltd. and the spinal cord lines were generated by my colleague Dr Graham Cocks (paper submitted). They were created by conditionally immortalising human somatic stem cells as described by (Pollock et al. 2006). In brief, primary cells were extracted from 12-week-old foetal central nervous system tissue and transduced with a retrovirus containing c-MycER^{TAM}. This transgene is a mutant oestrogen receptor (G525R) fused to the C-terminus of the mitogenic c-Myc (Littlewood et al. 1995). c-MycER^{TAM} is exclusively responsive to the synthetic drug 4-hydroxy-tamoxifen (4-OHT) and, in its presence, c-Myc expression drives cell proliferation. Clonal cell lines were generated from the transduced primary cells. The clonal cell lines accurately represent human NSCs; in their proliferative phase, they are self-replicative and under differentiating conditions they are multipotent and retain their positional specification in terms of gene expression and the specificity of the neurons they generate (Johansson et al. 2008; El-Akabawy et al. 2011). All cell lines used in this study are of a normal human karyotype.

2.1.2 Cell growth

All cells were grown in Binder CB150 incubators at 37°C, saturated humidity and 5% CO_2. They were cultured as monolayers in Nunclon™ Δ Surface tissue culture vessels (Thermo Scientific). Tissue culture vessels were coated with laminin (Sigma-Aldrich) at 20 μg/ml in Dilbecco's Modified Eagle's Media F-12 (DMEM:F12; Sigma-Aldrich) for three hours at 37°C. The flask was then rinsed with DMEM:F12 before the addition of the growth media. Spinal cord and cortical lines were cultured in Reduced Modified Media (RMM; see Table 2.1) and the striatal lines were cultured in Human Media (see Table 2.2).

Table 2.1 | Components of Reduced Modified Media

Component	Supplier	Concentration
DMEM:F12	Sigma-Aldrich	-
Human serum albumin	Baxter	0.03%
Human apo-transferin	Sigma-Aldrich	100 µg/ml
Putrescine dihydrochloride	Sigma-Aldrich	16.2 µg/ml
Human recombinant insulin	Sigma-Aldrich	5 µg/ml
Progesterone	Sigma-Aldrich	60 ng/ml
L-Glutamine	GIBCO	2 mM
Sodium selenite	Sigma-Aldrich	40 ng/ml
Human bFGF	Peprotech	10 ng/ml
Human EGF	Peprotech	20 ng/ml
4-OHT	Sigma-Aldrich	100 nM

Table 2.2 | Components of Human Media

Component	Supplier	Concentration
DMEM:F12	Sigma-Aldrich	-
Human serum albumin	Baxter	0.03%
Human apo-transferin	Sigma-Aldrich	100 µg/ml
Putrescine dihydrochloride	Sigma-Aldrich	16.2 µg/ml
Human recombinant insulin	Sigma-Aldrich	5 µg/ml
Progesterone	Sigma-Aldrich	60 ng/ml
L-Glutamine	GIBCO	2 mM
Sodium selenite	Sigma-Aldrich	40 ng/ml
T4 (L-Thyroxine)	Sigma-Aldrich	400 ng/ml
T3 (Tri-iodo-thyronine)	Sigma-Aldrich	337 ng/ml
Heparin sodium	Sigma-Aldrich	10 Units/ml
Corticosterone	Sigma-Aldrich	40 ng/ml
Human bFGF	Peprotech	10 ng/ml
Human EGF	Peprotech	20 ng/ml
4-OHT	Sigma-Aldrich	100 nM

Vials were thawed from liquid nitrogen at 37°C and revived in laminin-coated Nunclon™ Δ Surface 75 cm^2 tissue culture flasks (Thermo Scientific) with pre-warmed growth media. Cells were kept between 30% and 80% confluency, typically requiring passaging every 72 hours. For passaging, growth media was removed by aspiration and cells were washed with pre-warmed Hank's Balanced Salt Solution (HBSS) without Ca^{2+} or Mg^{2+} (Sigma-Aldrich). The HBSS was removed by aspiration and replaced with pre-warmed Accutase (Sigma-Aldrich) which was left for 3 minutes at 37°C. After the cells detached from the culture vessel, the cell suspension was transferred to a 15 ml centrifuge tube and centrifuged for 5 minutes at 400 rpm. The cell pellet was resuspended in 2 ml of growth media and the cells were added to a new laminin-coated Nunclon™ Δ Surface tissue culture vessel (Thermo Scientific) with pre-warmed growth media at a seeding density of 1.6 x 10^4 cells/cm^2. Cell counting was carried out using a standard Neubauer hemocytometer. Growth media was changed every 48 hours. The volumes of the reagents described above varied depending of the size of the culture vessel, the appropriate volumes are displayed in Table 2.3. Cells harvested during proliferations were collected one week or more after thawing. The differentiation protocol is detailed in section 4.2.1 on page 102.

Table 2.3 | Tissue culture reagent volumes for different vessels.

Vessel	Media	HBSS	Accutase
96 well plate	200 µl	100 µl	100 µl
6 well plate	3 ml	2 ml	1 ml
25 cm^2 flask	7 ml	5 ml	3 ml
75 cm^2 flask	15 ml	10 ml	5 ml
175 cm^2 flask	40 ml	20 ml	10 ml

2.2 Molecular Biology

2.2.1 DNA extraction from human NSCs

Cells from a Nunclon™ Δ Surface 75 cm^2 tissue culture flask (Thermo Scientific) were harvested and pelleted at 80% confluency as described in section 2.1.2. The cell pellet was resuspended and homogenised by pipetting in 1 ml of cell lysis buffer (Meulenbelt et al. 1995). The lysis buffer comprised of 100 mM NaCl, 10 mM Tris-HCl pH 8, 10 mM ethylene diaminetetraacetic acid (EDTA) pH 8 and 0.5% sodium dodecyl sulphate (SDS). 25 μg of Ribonuclease A (Fermentas) was added to the homogenate and incubated at 37°C for 30 minutes. 200 μg of Proteinase K (Fermentas) was then added and the sample was incubated at 50°C for 2 hours. Yeast reagent 3 (Autogen Bioclear Ltd.) was diluted with an equal volume of ethanol to make an organic deproteinisation reagent (ODPR) as described by Freeman et al. (2003). 200 μl of the ODPR was added to the sample which was then shaken vigorously by hand for 30 seconds and centrifuged for 5 minutes at 13,000 rpm to pellet the cell debris. The supernatant was recovered and transferred to a clean tube where a further 200 μl of ODPR was added. The sample was centrifuged for a further 5 minutes at 13,000 rpm after which the supernatant was recovered and transferred to a clean tube. An equal volume of isopropanol was added to precipitate the DNA. The tube was inverted 5 times and then centrifuged for 15 minutes at 13,000 rpm. The isopropanol was discarded and 3 ml of 70% ethanol added to the DNA pellet. The sample was centrifuged for a further 10 minutes at 46000 rpm. The ethanol was discarded and the DNA pellet air-dried before resuspension in TE buffer (Invitrogen) at 4°C for 16 hours. The yield and quality of the DNA was assessed as described in section 2.2.3.

2.2.2 Total RNA extraction from human NSCs

Cells from a Nunclon™ Δ Surface 75 cm^2 tissue culture flask (Thermo Scientific) were harvested and pelleted at 80% confluency as described in section 2.1.2. The cell pellet was lysed by resuspension in 1 ml of TRIzol® Reagent (Invitrogen) and pipetting until the solution turned clear. After complete cell lysis, 200 µl of chloroform (Sigma-Aldrich) was added, and the homogenate was mixed by 10 inversions. The sample was transferred to a Phase Lock Gel Heavy 2 ml tube (5PRIME) and incubated on ice for 3 minutes. The sample was then centrifuged at 12,000 g for 15 minutes at 4°C to separate the aqueous and the phenol-chloroform phase. The aqueous phase was recovered and transferred to a fresh 1.5 ml tube where the RNA was precipitated with 500 µl of isopropanol (Sigma-Aldrich). The sample was inverted 10 times and incubated at room temperature for 15 minutes. The precipitated RNA was then pelleted by centrifugation at 12,000 g for 30 minutes at 4°C. The isopropanol was removed and the RNA pellet was rinsed with 75% ethanol. The sample was centrifuged again at 12,000 g for 15 minutes at 4°C after which the supernatant was carefully removed and the RNA pellet was left to air dry. On turning transparent, the RNA pellet was resuspended in 40 µl of nuclease-free H_20 (Sigma-Aldrich). The yield and quality of the RNA was assessed as described in section 2.2.3.

2.2.3 Quantification of nucleic acid concentration and purity

The concentration and purity of extracted nucleic acids was measured using the NanoDrop™ 1000 spectrophotometer (Thermo Scientific). The 260/280 and 260/230 absorbance ratios were used to assess the purity of the nucleic acids. A 260/280 ratio of >1.8 was considered pure for DNA and a ratio of >2.0 was considered pure for RNA. A

260/230 ratio of >2.0 was considered pure for DNA and RNA. If a sample did not meet these standards, all samples from that study were re-precipitated as described in section 2.2.4. The integrity of extracted nucleic acids was assessed by agarose gel electrophoresis (see section 2.4).

2.2.4 Nucleic acid re-precipitation

Three volumes of molecular biology grade ethanol (Sigma-Aldrich) and 0.1 volume of 3 M sodium acetate (Fermentas) was added to the sample and mixed by 10 inversions. It was incubated for 16 hours at -80°C and then centrifuged at 14,000 rpm for 20 minutes. The supernatant was discarded and 1 ml of 80% molecular biology grade ethanol was added. The sample was centrifuged again at 14,000 rpm for 5 minutes, the supernatant discarded and a further 1 ml of 80% molecular biology grade ethanol was added to the nucleic acid pellet. The sample was centrifuged a final time for 5 minutes at 14,000 rpm before the supernatant was removed and the pellet air-dried. The nucleic acid was resuspended in the appropriate amount of nuclease-free H_20 to make approximately 1 µg/µl. Nucleic acid concentration and purity was calculated as described in section 2.2.3.

2.2.5 DNase treatment of RNA

To remove residual genomic DNA contamination, all RNA samples were treated with TURBO DNA-*free*™ (Ambion). 5 µg of extracted RNA was combined with 2 units of TURBO DNase, 1 x TURBO DNase Buffer and made up to 25 µl with nuclease-free H_20 (Sigma-Aldrich). The sample was incubated for 50 minutes at 37°C. The DNase was subsequently inactivated by incubation with 6 µl of DNase Inactivation Reagent at room

temperature for 5 minutes, vortexing every 30 seconds. The sample was then centrifuged at 13,000 rpm for 90 seconds and the supernatant recovered and transferred to a new microcentrifuge tube and stored at -80°C.

2.2.6 Genomic DNA contamination test

To verify complete removal of genomic DNA, 1 µl of DNase treated RNA was used as template for a quantitative polymerase chain reaction (qPCR; see section 2.6). The sequence of the *GAPDH* primer pair used is presented in Table 2.4 on page 62. The RNA sample was run alongside a no-template negative control as well ten serial diluted genomic DNA samples. Only samples without amplification were reverse transcribed.

2.2.7 Reverse transcription

Complimentary DNA (cDNA) was synthesised using SuperScript® III Reverse Transcriptase (Invitrogen). 1 µg of DNase-treated RNA was combined with 250 ng of random hexamers (Invitrogen) and 1 mM dNTP mix (New England BioLabs) made up to 10 µl with nuclease-free H_20 (Sigma-Aldrich). The mix was incubated for 5 minutes at 65°C on a heated block to denature RNA secondary structure. It was then promptly placed on ice and incubated for 1 minute. The mix was then made up to 20 µl with the following reagents: 1x First Strand Buffer (Invitrogen), 5 mM Dithithretiol (Invitrogen), 40 units RNaseOUT™ (Invitrogen), 200 units SuperScript® III Reverse Transcriptase (Invitrogen) and 3 µl of nuclease-free H_20 (Sigma-Aldrich). Samples were incubated at 25ºC for 2 minutes, 42ºC for 2 hours and then 70ºC for 15 minutes. For the microarray

studies, the cDNA was not diluted. For polymerase chain reaction (PCR) applications, the cDNA was diluted in 140 µl of nuclease-free H_20 (Sigma-Aldrich).

2.3 Polymerase Chain Reaction

2.3.1 Amplification primer design

PCR primer pairs were designed to amplify an amplicon of approximately 200 base pairs. DNA sequence was obtained from the University of California, Santa Cruz, genome browser (http://genome.ucsc.edu) and primers designed using Primer3Plus software (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi). They were designed to have minimal self-complementarity and no complementarity to the other primer. GC content was kept between 40% and 60% and primers were designed with an annealing temperature between 55ºC and 65ºC with only 1ºC between the pair. The nucleotide sequence was sent to Sigma-Aldrich who synthesised the oligonucleotides. They were resuspended in nuclease-free H_20 to a stock concentration of 100 µM and a working concentration of 2 µM.

2.3.2 PCR protocol

PCR reactions were carried out in a total reaction volume of 15 µl. The reaction comprised of template nucleic acid (48 ng of DNA or 6 µl of PCR concentration cDNA as described in section 2.2.7), 0.4 unit HOT FIREPol® DNA Polymerase (Solis BioDyne), 0.2 µM primer mix, 400 µM dNTP mix (New England BioLabs), 1x Buffer B2 (Solis

BioDyne), 1.5 mM $MgCl_2$ (Solis BioDyne) and nuclease-free H_20 (Sigma-Aldrich) to make a total reaction volume of 15 µl. All PCR reactions were carried out with a negative control for each primer pair, where the template was replaced with H_20. Thermal cycling was carried out on a G-Storm GS4 thermal cycler (Somerton Biotechnology Centre) using the following parameters:

Initial denaturation	95ºC	15 minutes	
Denaturation	95ºC	30 seconds	35 cycles
Annealing	50ºC - 60ºC	30 seconds	
Extension	72ºC	30 seconds	
Final extension	72ºC	10 minutes	

The annealing temperature was typically 5ºC cooler than the lowest melting temperature of the primer pair as predicted by the Primer3Plus software.

2.3.3 PCR optimisation

Where sub optimal-PCR results were observed, for example by low amplification or primer dimerisation, the optimal annealing temperature was determined using the temperature gradient function of the G-Storm GS4 thermal cycler (Somerton Biotechnology Centre). The temperature that yielded the brightest specific band after agarose gel electrophoresis analysis (see section 2.4) was used for subsequent PCRs.

2.4 Agarose gel electrophoresis

Agarose gel electrophoresis was used to test the integrity of extracted nucleic acids as well as to assess the level and specificity of PCR amplification. Agarose gels were made by melting agarose (Sigma-Aldrich) into 1x tris-acetate-EDTA (TAE) buffer. 0.8% and 1.2% gels were prepared for assessing the integrity of for extracted genomic DNA and RNA respectively. For PCR products, the concentration of agarose was dependent on the size of the amplicon. 3% agarose was used if the amplicon was smaller than 100 base pairs (bp), 2% if 100 to 200 bp and 1% if greater than 300 bp. After cooling, ethidium bromide (1% solution in water; MERK) was added at a concentration of 1 µl / 100 ml of TAE. The molten agarose was then poured into a mould and combs were inserted. After solidifying, the gel was placed into an electrophoresis tank with 1x TAE. Samples were mixed with 1x DNA Loading Dye (Fermentas) before being loaded into the wells. Samples were run alongside an appropriate sized DNA ladder (100 bp DNA ladder, Solis Biodyne; O'GeneRuler™ 1kb DNA Ladder, Fermentas). Electrophoresis was carried out at 4 V / cm between the electrodes for approximately 30 minutes until desired separation was achieved. Gels were analysed by ultra violet transillumination using the BioSpectrumAC Imaging System (UVP) and imaged using the Vision Works LS software (UVP).

2.5 SNaPshot single base primer extension assay

Single base primer extension with SNaPshot™ chemistry (Applied Biosystems) was carried out for the purpose of genotyping DNA samples and allelic expression analysis.

2.5.1 Target amplification

SNPs within exons present in all isoforms were targeted where possible for allelic expression analysis. When genotyping samples for tag SNPs, heterozygosity rates from the International HapMap database (www.hapmap.ncbi.nim.nih.gov) were taken into consideration. A ~200 bp region containing the target SNP was amplified by PCR as described in section 2.3. The intensity and specificity of the PCR product was then examined by agarose gel electrophoresis as described in section 2.4. For allele-specific expression assays, a minimum of four genomic DNA samples and a no-template negative control were included for each primer pair assayed. cDNA samples from the NSC lines were always run in biological triplicate.

2.5.2 PCR product cleanup

PCR products underwent an enzymatic cleanup step to eliminate unincorporated primers and dNTPs. 7 µl of PCR product was combined with 1 unit rAPid Alkaline Phosphatase (Roche), 4 units Exonuclease I (New England BioLabs) and made up to 10 µl with nuclease-free H_20 (Sigma-Aldrich). The reaction was incubated at 37ºC for 1 hour and then 85ºC for 15 minutes to inactivate the enzymes.

2.5.3 Primer extension

Single base primer extension was carried out using the SNaPshot™ Multiplex Kit (Applied Biosystems). The system involves the binding of an extension primer to the PCR product immediately upstream of the SNP of interest. The polymerase, AmpliTaq®,

extends the primer by one base incorporating a fluorescently labelled ddNTP complimentary to the nucleotide in the amplified sequence on its 3' end. Measuring the relative intensity of incorporated fluorescent ddNTPs enables quantitative genotyping (Norton et al. 2002).

2.5.3.1 Extension primer design

Extension primers were deigned using the online tool FP Primer (www.m034.pc.uwcm.ac.uk/FP_Primer.html) with all settings at default. 50 bp of sequence flanking either side of the SNP of interest was pasted into FP Primer from dbSNP and the highest ranked primer was selected. The sequence was synthesised by Sigma-Aldrich and rehydrated to 1 μM working concentration in nuclease-free H_20.

2.5.3.2 Single base primer extension protocol

2 μl of cleaned PCR product was combined with 1.25 μl SNaPshot™ Multiplex Kit mix (AmpliTaq® DNA polymerase, F-ddNTPs and buffer; Applied Biosystems), 0.1 μM extension primer and made up to a total volume of 10 μl with nuclease-free H_20 (Sigma-Aldrich). Temperature cycling was carried out on a G-Storm GS4 thermal cycler (Somerton Biotechnology Centre, UK) following the parameters below:

Initial denaturation	95ºC	2 minutes	
Denaturation	95ºC	10 seconds	30 cycles
Annealing	50ºC	5 seconds	
Extension	60ºC	10 seconds	

2.5.3.3 Electrophoresis

2 µl of single base primer extension product was combined with 8 µl of genetic analysis grade Hi-Di™ Formamide (Applied Biosystems). Electrophoresis was carried out on the 3130 Genetic Analyzer (Applied Biosystems) with a 36 cm capillary and POP-7™ polymer (Applied Biosystems). The default SNaPshot settings were used as recommended by the manufacturer.

2.5.4 Analysis of allele-specific expression data

Data files from the 3130 Genetic Analyzer (Applied Biosystems) were analysed using GeneMarker® 2.2.0 (SoftGenetics LLC®). The quantity of each allele detected is represented by the peak height. The average allele ratio derived from heterozygous genomic DNA was representative of the 1:1 ratio, correcting for any inherent primer or dye bias. The relative expression of the two alleles was determined by correcting cDNA allele ratios with those from the genomic DNA samples as illustrated below:

$$\text{Allele ratio} = \frac{A_C / B_C}{\left(\sum \frac{A_g}{B_g}\right) / n}$$

Where A_c and B_c are the cDNA peak heights for alleles A and B, A_g and B_g are the genomic DNA peak heights for alleles A and B, and n is the number of genomic DNA samples. This calculation was carried out for each replicate. Ratios were transformed into percentages for clearer graphical representation.

2.6 Quantitative real-time PCR

2.6.1 Primer design

Primers for quantitative PCR (qPCR) were designed as described in section 2.3.1 with some minor alterations. qPCR primers were designed to amplify ~100 bp amplicons and, where possible, they were also designed to span large exon boundaries to prevent the amplification of any residual genomic DNA. However, the *GAPDH* primer pair used to test for genomic DNA contamination (see section 2.2.6) did not span exon-boundaries.

2.6.2 Reference gene selection

The selection of appropriate reference, or housekeeping, genes is essential for accurate comparison of gene expression between biological samples. However, many commonly used reference genes show significant variability in expression between tissues, reducing their efficacy (Lee et al. 2002). The reference gene primers *ACTB*, *B2M*, *GAPD*, *HMBS*, *HPRT1*, *RPL13A*, *SDHA*, *UBC* and *YWHAZ* from Vandesompele et al. (2002) as well as *RPL27*, *RPL30*, *OAZ1*, *RPL22* and *RPS29* from de Jonge et al. (2007) were first tested for efficiency and specificity on my samples. *ACTB*, *GAPDH*, *HMBS*, *SDHA*, *UBC* and *OAZ1* were subsequently redesigned for optimal results as described in section 2.6.1. Primer sequences are presented in Table 2.4. These reference genes were then assayed across samples using the methods described in section 2.6.3. The programme geNorm^{PLUS} (Vandesompele et al. 2002) in qbase^{PLUS} version 3 (Biogazelle) was used to assess reference stability between samples and the optimal normalisation factor was determined to be the geometric mean of the reference targets *HPRT1* and *RPL13A* (see Figure 2.2).

Table 2.4 | Reference gene primer sequences for qPCR. Forward (F), reverse (R) and extension (E) sequences are presented. *GAPDH** is the primer used for testing genomic DNA contamination in RNA samples.

Gene	Primer sequence (5'-3')	
ACTB	F	TCGTGCGTGACATTAAGGAG
	R	AGGAAGGAAGGCTGGAAGAG
B2M	F	TGCTGTCTCCATGTTTGATGTAT
	R	TCTCTGCTCCCCACCTCTAAGT
GAPDH	F	CCTGACCTGCCGTCTAGAAA
	R	ATCCTGGTGCTCAGTGTAGCC
*GAPDH**	F	ATCATCAGCAATGCCTCCTGC
	R	ATGGCATGGACTGTGGTCATG
HMBS	F	CTCTGCGGAGACCAGGAGT
	R	ATTGCCGTTACCAGACATGG
HPRT1	F	TGACACTGGCAAAACAATGCA
	R	GGTCCTTTTCACCAGCAAGCT
OAZ1	F	GGTCTCCCTCCACTGCTGTA
	R	ACTATTCCCTCGCCCACCT
RPL13A	F	CCTGGAGGAGAAGAGGAAAGAGA
	R	TTGAGGACCTCTGTGTATTTGTC
RPL27	F	ATCGCCAAGAGATCAAAGATAA
	R	TCTGAAGACATCCTTATTGACG
RPL30	F	ACAGCATGCGGAAAATACTAC
	R	AAAGGAAAATTTTGCAGGTTT
RPS29	F	GCACTGCTGAGAGCAAGATG
	R	ATAGGCAGTGCCAAGGAAGA
SDHA	F	AGGAATCAATGCTGCTCTGG
	R	CTGCTCCGTCATGTAGTGGA
UBC	F	CCCCAGTATCAGCAGAAGGA
	R	TATAATCATCGGCGTTCACC
YWHAZ	F	ACTTTTGGTACATTGTGGCTTCA
	R	CCGCCAGGACAAACCAGTAT

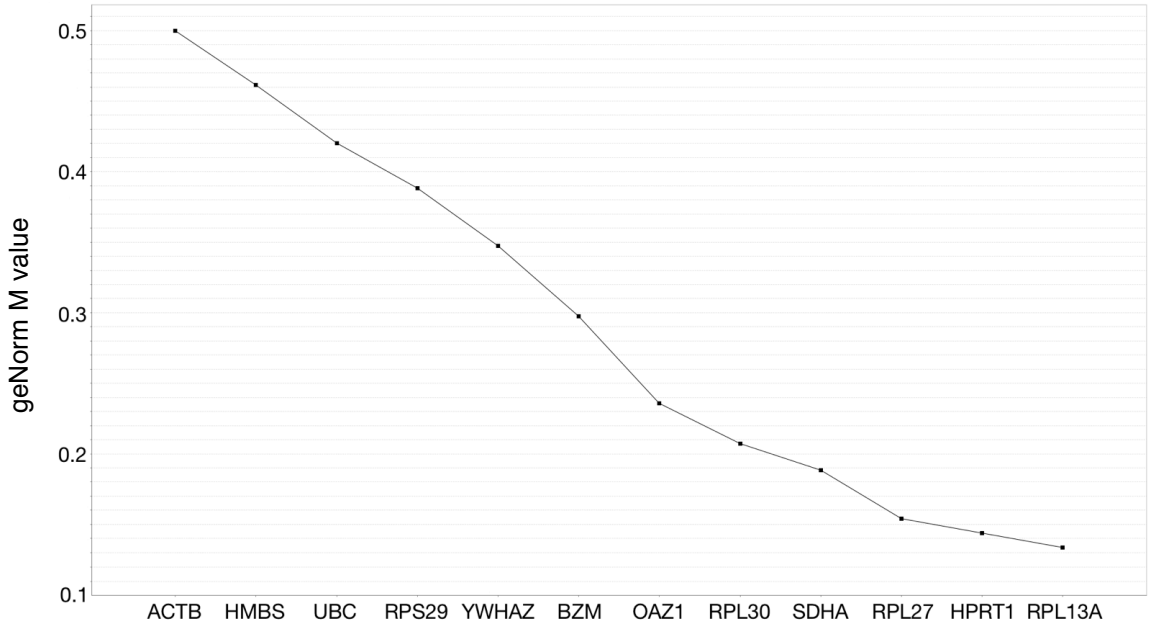


Figure 2.2 | Average expression stability of reference targets. geNorm M values are displayed for twelve reference target genes. A geNorm M value of < 0.2 is considered highly stable.

2.6.3 qPCR protocol

Experiments were carried out following the sample maximisation approach, where all samples assayed for the same target are included in the same plate (Hellemans et al. 2007). White welled qPCR plates (4titude FrameStar 96) were exposed to UV for 10 minutes in the UV Stratalinker 1800 (Stratagene) to crosslink DNA prior to plating. 4 µl of PCR concentration cDNA was combined with 0.2 µM primer mix and 1x HOT FIREPol® EvaGreen® qPCR Mix Plus ROX (Solis Biodyne). qPCR reactions were carried out on a Chromo4™ Real-Time PCR detector (Bio-Rad) using the following parameters:

Initial denaturation	95ºC	15 minutes	
Denaturation	95ºC	30 seconds	45 cycles
Annealing	60ºC	30 seconds	
Extension	72ºC	30 seconds	

Fluorescence was recorded at the end of each cycle. A melt curve was carried out from 60ºC-95ºC with a 10 second hold and a plate read and at every 1ºC increment. Reactions were carried out in biological triplicate and technical duplicate. Target genes were assayed alongside the reference genes selected by geNorm^{PLUS} (see section 2.6.2). Two negative controls were included for every primer pair, where 4 µl of H_20 replaced the template.

2.6.4 qPCR data analysis

Data files were opened in Opticon Monitor™ (Bio-Rad), where the threshold was selected and melting profiles were assessed. Cycle threshold (CT) values and efficiency values were exported and opened in Excel 2011 (Microsoft). The Pfaffl mathematical model for relative transcript quantification was used for data analysis (Pfaffl 2001). The formula is illustrated below:

$$\text{Ratio} = \frac{(E_{target})^{\Delta CT_{target}\ (control\ -\ sample)}}{(E_{ref})^{\Delta CT_{ref}\ (control\ -\ sample)}}$$

Where ΔCT_{target} = target gene cycle threshold, ΔCT_{ref} = mean of reference gene CTs, and E = PCR efficiency. This method enables relative quantification of a target gene in comparison to an endogenous standard, the reference genes. Technical duplicates were averaged and used to generate the biological replicate values that were used for statistical analysis.

2.7 Statistics

All statistical analysis in this thesis was carried out in Prism (GraphPad). *P* value summaries displayed on figures are as follows: $p \leq 0.05$ = *; $p \leq 0.01$ = **; $p \leq 0.001$ = ***. T-tests were used to compare the means of two samples. Unless stated otherwise, t-tests were unpaired and two-tailed with 95% confidence intervals. F-tests were carried out in parallel to compare variances. If unequal variance was found, Welch's correction was applied. Binomial exact tests, chi-square tests and Fisher's exacts tests were used to test categorical outcomes. The binomial exact test was used to compare the observed to the expected distribution for one nominal variable with only two outcomes. Fisher's exact test was used to test two nominal variables. Chi-squared test was used to test two nominal variables of large values (Prism defaults to chi-squared tests when values exceed its threshold), or to test more than two nominal variables. Linear regression and Pearson's coefficients were calculated to assess the correlation between two datasets.

Chapter 3. A genome-wide assessment of monoallelic expression in human neural stem cells

3.1 Introduction

The importance of monoallelic gene expression has long been appreciated in respect to genomic imprinting, XCI and the RME of a small number of autosomal gene families (see section 1.2). Genomic imprinting and XCI play critical roles in regulating gene dosage, while RME acts to enhance cell specificity and tissue diversity. The recent reports that RME is widespread on the autosome in human B-lymphoblastoid cells (Gimelbrant et al. 2007) and mouse NSCs (Wang et al. 2010; Li et al. 2012) raise important questions about the role this process may play in human neural development, function and disease. The analysis of RME in neural tissue has been limited by the difficulty of detection in mixed cell populations *in vivo*. Therefore, the extent to which this process occurs in the developing and adult human central nervous system is unknown.

In this chapter, I report the first genome-wide assessment of monoallelic gene expression in human neural tissue with the power to detect RME. I aimed to assess the extent and nature of RME in clonal human NSCs. Some of the studies presented in this chapter have been published in *Stem Cells* (Jeffries et al. 2012).

3.2 Methods

Cell culture and molecular biology was carried out as described in Chapter 2. Experimental details specific to this results chapter are included below.

3.2.1 Microarray-based assay of monoallelic expression

The Illumina Omni1-Quad BeadChip platform was used to assess allele-specific expression in the three spinal cord, cortical and striatal NSC lines. Genomic DNA from each of the three donors, and cDNA from each of the clonal lines, was used as template. The spinal cord line SPC01 was run in triplicate to test reproducibility.

3.2.1.1 BeadChip procedure

Samples were sent to the Wellcome Trust Centre for Human Genetics at the University of Oxford to be processed. The manufacturers protocol was followed with the only exception of using cDNA as the template for some samples. 5 µl of undiluted reverse transcription reaction product was used as template for cDNA samples and 750 ng of DNA was used as template for the genomic DNA samples. In brief, the template was denatured, neutralised and then underwent whole-genome amplification with MA1 reagent (Illumina) for 22 hours at 37ºC before being enzymatically fragmented with FMS reagent (Illumina) for 1 hour at 37ºC. The fragmented template was then precipitated with isopropanol, pelleted by centrifugation and resuspended in the hybridization buffer RA1 reagent (Illumina) for 1 hour at 48ºC. The samples were denatured at 95ºC for 20 minutes and then loaded onto the BeadChip where the template was hybridised to complementary

50-mers for 20 hours at 48ºC. The BeadChip was washed to remove unbound template before the application of TEM reagent (Illumina) to extend the template-bound oligos by a single labelled nucleotide. The hybridised template was removed with 95% formamide/1 mM EDTA and the BeadChip stained with the XStain HD reagents (Illumina). The fully processed BeadChip was then imaged with the Illumina iScan™ Reader. The scanner uses laser excitation to determine the single base extension product at each probe.

3.2.1.2 BeadChip data analysis

Data files were processed in GenomeStudio v2010.1 (Illumina) using Genotyping Module v1.6.3. Genotypes from the genomic DNA samples were called using Illumina's EGT cluster file with a stringent gencall score threshold of 0.25. Genotypes from the genomic DNA samples, and allelic intensity values (Xraw and Yraw) from all samples were exported as CSV files. Quantile normalisation of allelic intensity values between channels was carried out in R (www.r-project.org) using the Bioconductor package Limma (see Figure 3.1). This allows for cross-array analysis and corrects for any inherent dye bias. The genomic DNA samples were normalised separately from the cDNA samples to account for copy number differences.

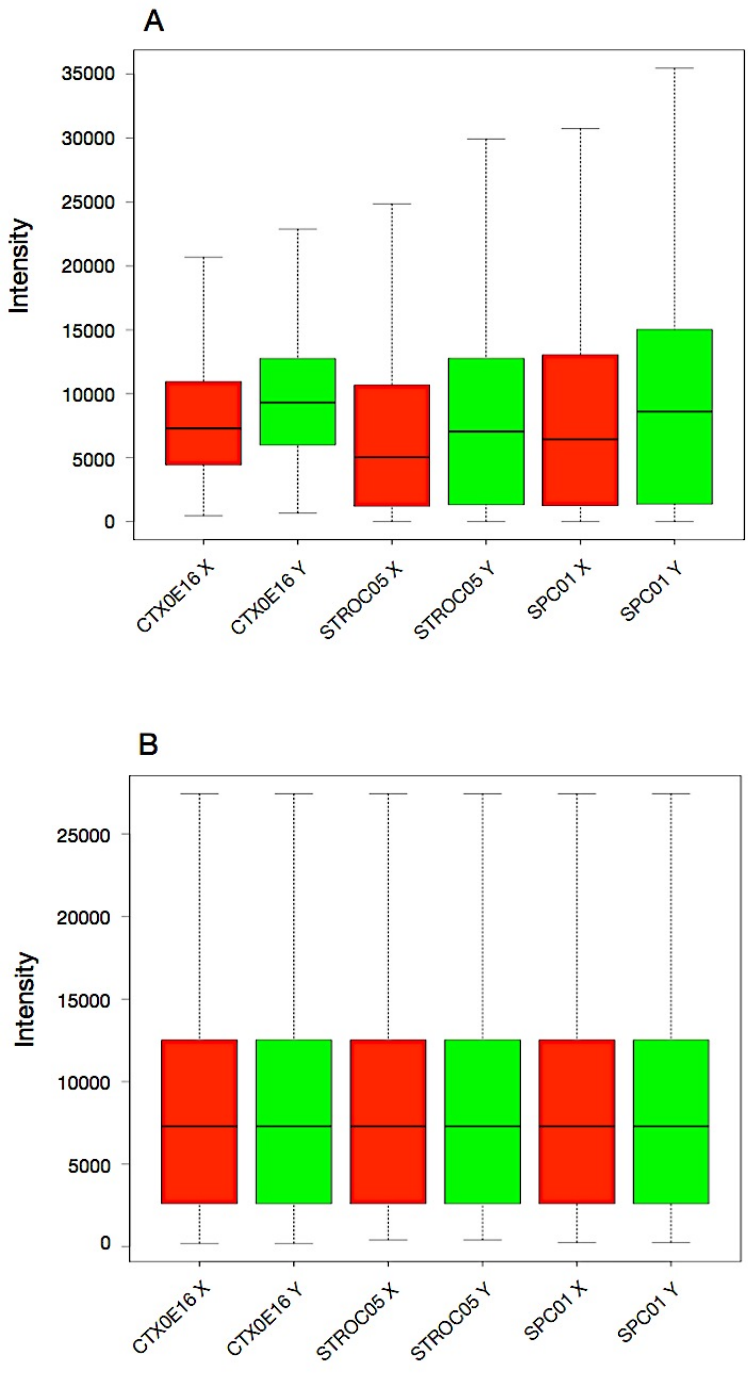


Figure 3.1 | Quantile normalisation of allelic intensity values. Allelic intensity values for the genomic DNA samples are displayed before (**A**) and after (**B**) quantile normalisation. cDNA values were normalised in the same way but separately from the genomic samples to account for copy number differences (see Appendix 9.1 on page 204 and Appendix 9.2 on page 205).

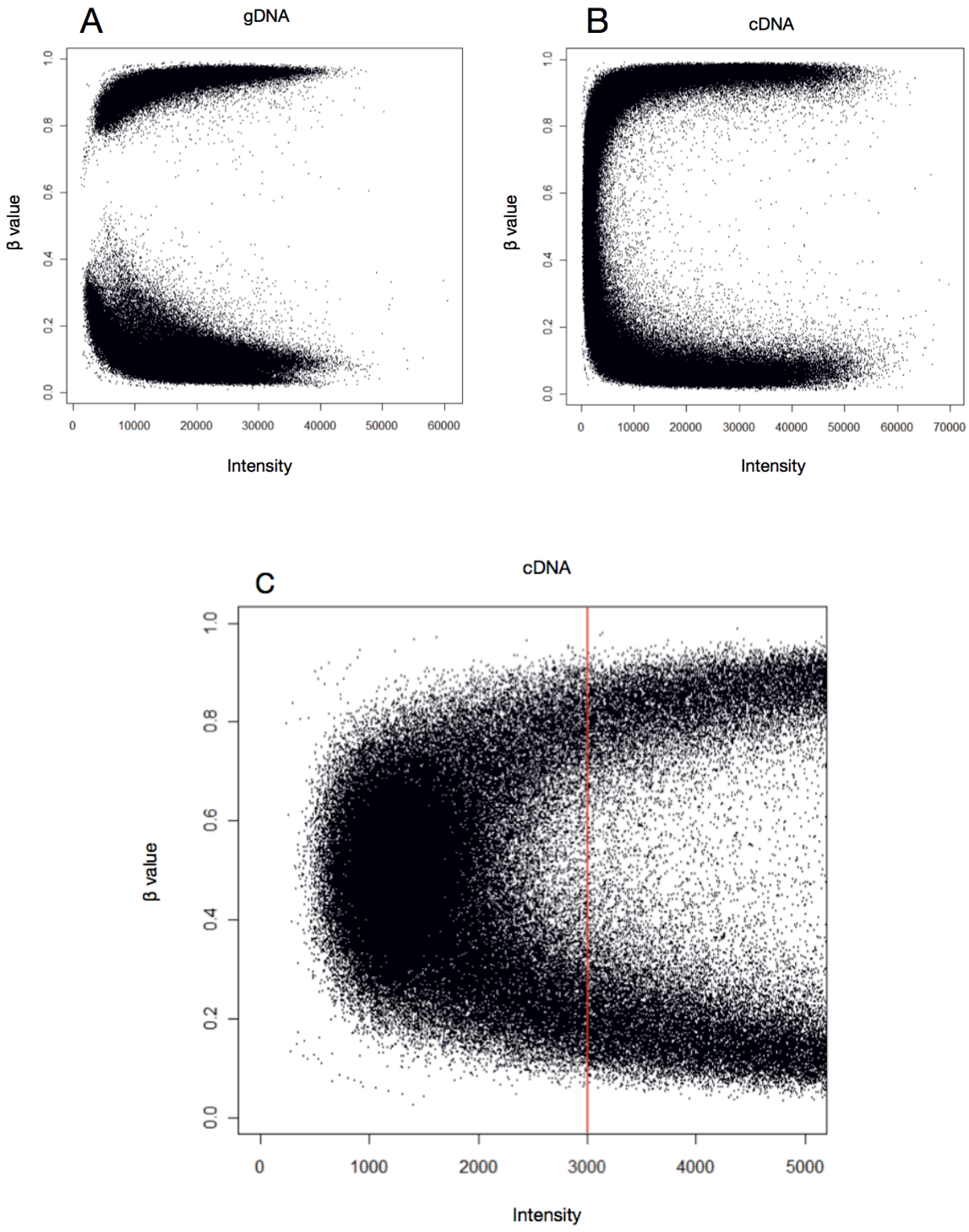


Figure 3.2 | Intensity threshold to remove background signal. A) β values of homozygous SNPs are plotted against total intensity. **B)** The same SNPs are plotted using cDNA from the same genotype. The cDNA β values deviate from the genomic distribution at low intensity. **C)** An intensity threshold of 3000 was set to remove the background signal probes.

A stringent post-normalisation total intensity (Xraw + Yraw) threshold of 3,000 was used for cDNA samples to eliminate spurious results arising from the analysis of non-expressed SNPs (see Figure 3.2). Normalised allelic intensity values that met the intensity threshold were used to calculate β values using the formula below:

$$\beta \text{ value} = \frac{\text{Xraw}}{\text{Xraw} + \text{Yraw}}$$

The β value is a score of allelic expression, where 0.5 represents equal expression of the two alleles and 1 or 0 represents exclusive expression of either allele. The β value of a heterozygous SNP in the genomic DNA provides a theoretical 1:1 allele ratio. Deviation from the genomic DNA β value in the cDNA samples indicates imbalance in allele-specific expression. Therefore, Δ_β values were calculated for all heterozygous SNPs to assess allelic-expression using the following formula:

$$\Delta_\beta \text{ value} = \beta_{gDNA} - \beta_{cDNA}$$

Assayed SNPs were associated with RefSeq accessioned transcripts using Galaxy (http://usegalaxy.org). The mean Δ_β value of all SNPs within a transcript was used to calculate the transcript Δ_β value. A penalty-based weighting score was applied to transcripts: SNPs with a $\Delta_\beta > 0.1$ scored +1 while values < 0.1 received a -2 penalty. Transcripts with a total score of one or below were rejected.

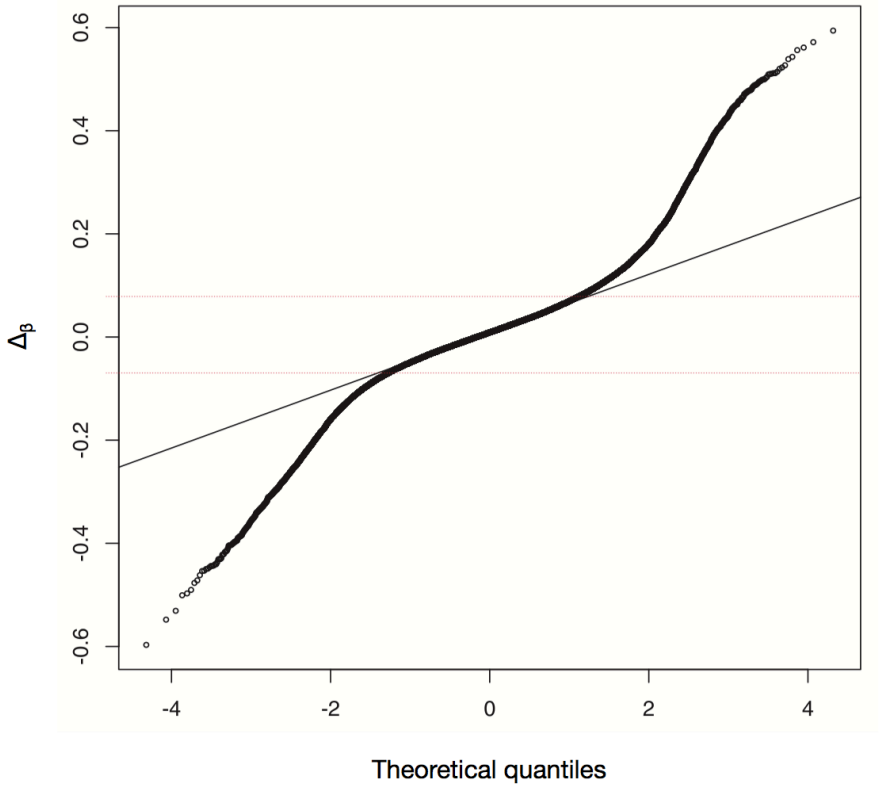


Figure 3.3 | Normal QQ plot of autosomal genic SNPs. Heavy tails are observed deviating from normal distribution at Δ_β values +0.07 and -0.07. This chart illustrates data from CTX0E03 and is representative of the other cDNA samples (data not shown).

A transcript Δ_β value of 0.1 (60:40 ratio) has been deemed to represent biologically significant allelic imbalance in other similar microarray studies (Pastinen et al. 2004; Serre et al. 2008; Lee et al. 2009). This threshold was validated in our data by generating normal QQ plots in R for each clonal line. The data deviated from normal distribution at Δ_β values of approximately +0.07 and -0.07 with a kurtosis of > 5 (see Figure 3.3). Monoallelic expression was defined as a transcript Δ_β value of > 0.2 (70:30 ratio). Biallelic expression was defined as a $\Delta_\beta < 0.05$.

3.2.2 BeadChip validation by SNaPshot primer extension

The SNaPshot single base primer extension assay, as described in section 2.5, was carried out for 23 genes in multiple cell lines to validate the microarray assay. The genes assayed and the primers used are presented in Table 3.1. Δ_β values were calculated in Excel 2011 (Microsoft) and the mean Δ_β from three biological replicates was correlated with Δ_β values derived from the BeadChips. The plot was drawn and Pearson's correlation calculated in Prism (GraphPad).

3.2.3 DNA methylation analysis

My collaborators, Dr Chloe Wong and Dr Ruth Pidsley, ran 750 ng of bisulfite-treated genomic DNA from each of the cortical and spinal cord NSCs on the Infinium HumanMethylation27 BeadChip (Illumina) following the manufactures recommended protocol. DNA methylation β values were calculated using the GenomeStudio Methylation module (v1.6.1). Differences between DNA methylation β values for monoallelic and biallelic genes were tested for statistical significance using a two-tailed t-test in Prism.

Table 3.1 | Primer sequences for single nucleotide primer extension. Forward (F), reverse (R) and extension (E) sequences are presented. Continued overleaf**.**

Gene	Tag SNP	Primer sequence (5'-3')	
C1QL3	rs4747277	F	TGGACACATAAGGGCTTTTTC
		R	CCTGGGAATGATTCAACTTCA
		E	ACATAAGGGCTTTTTCTAAATACCGTAC
CAT	rs769217	F	GCCTGGGACCCAATTATCTT
		R	TGTCCTGCATGCACATCG
		E	TCGAGTGGCCAACTACCAGCGTGA
CHL1	rs2117046	F	TTGGGTTTTGTTCTTGTATTGC
		R	TGGGGTAGATGTACATTCAACAAT
		E	TTTCAACAGTTTCAAAATAAAATATCATA
DNAJC15	rs1047775	F	TCTCCTTACGTAGCAGCCAAA
		R	TGGTGGAAGATAAGACTGTGG
		E	AGAATATTTTGCAGGGCTTTTTTTTTTT
ENG	rs3739817	F	GGCACACTTTGTCTGGATCA
		R	TAGGCTGCAGACCTCACC
		E	GGCTACAAGTGTCCTTGGGAGGAGT
GABRG3	rs10873636	F	CTACTGGTGACCCACCACCT
		R	CGGTGTCTGCAACATAGGAA
		E	ACACTGTCAGGATCTTGCTCATAG
GRIA1	rs707176	F	TCCTGCAGAAAGTCCTGGAT
		R	TGAGGCGTTCTGATTCACAG
		E	GTATCCCTCCTCTGTGGTTGTCAA
GRID1	rs1054979	F	CCGACCAGCAGAGCTTTTTA
		R	CCCCACTCCATTCTGTCATT
		E	CGAGTGTGTGTGGTTTGTGTTGTT
JAG2	rs1057744	F	TGAACGGGTACCAGTGTGTG
		R	GGGCAGTGGCAGTGGAAG
		E	CGGCATTGCGAGCTGGAACGAGAC
KIAA1324L	rs12535941	F	GGCTACATAGGGGAAGACTAAGC
		R	GCAGCAGAACAAGGCTCTTC
		E	ATGCATTTCTACTGAAGGATCACTAAAAA
NECAB2	rs2280027	F	TCCATCCTCCACAAGAAGGT
		R	GTGTGGTCAGTGTGGGTCAT
		E	CCCCTGCCTCCTGGTCCTGGCCTCTCCCC

Table 3.1 | Continued.

Gene	Tag SNP	Primer sequence (5'-3')	
PAX8	rs1478	F	AAGGAGAGAAGCCCCACAG
		R	TTTGTTTGGGTCAAGCTTCC
		E	CATTTCACTCAGAGGCCAAAGTCTGGGGG
PM20D1	rs1104899	F	ATCACAGAGTTCTTGTCGTCCA
		R	GAGGTGCCCCCATTCTCT
		E	AGAGTTCTTGTCGTCCAGTGTGCCCC
PMP2	rs2229015	F	TTTTCAGGTGTGGGGTTAGC
		R	TCTCTCTCAAGCAGCCCACT
		E	ACGAACTGAAAGTACCTTTAAAAATACAGA
RALYL	rs10094238	F	AAATGTGTTTTCCCCTTTGC
		R	CAGAATTGCAATATCACCTGACTT
		E	TGTGAAAGGCATTTATGAATGGTAAGGGA
SEMA5A	rs786843	F	CACCTTTCTGCACCTGATGT
		R	TAGCACTGGGAGCCACTTTC
		E	ATTTAACTTCTAAAATTATTTCTGAATACAGTTG
SFMBT2	rs11255084	F	GGGCACCTCCATGTTTTCTA
		R	TGCCCCATATTTTGCATTTT
		E	AAGGCAATCAGTGAACTTCCTC
SLC6A15	rs17183577	F	ATGTGTGATTTGCCCACAGA
		R	TCTGAAAGTGGGGGCTTAAA
		E	TCCCTCTCTCAAATTCTGTTACAG
THNSL2	rs4129190	F	TTCTCTGGCTATGCACCTGA
		R	GCTCAGAGCCAATGAAGAGG
		E	AAGAGCTCCCACAGTTGGACAGAG
TNFRSF10D	rs1133782	F	CGCAACGAGACCCTGAGTA
		R	GACGCTGTGGCTCCTCTG
		E	CAGAGCTAACAGGTGTGACTGTAGAGT
TTN	rs12463674	F	CTGCCCACCATCAATTCTAA
		R	TGAGGTTGAACTCCCAAGAA
		E	TCCTTATCTCCTCTTAACCACTCA
VAX2	rs3771389	F	CCTGCCCAATCCTGTAAGTG
		R	CCTTCTTCCCTGGATACTCAGA
		E	CATGGTCCCAATTCAGGCAGAG
ZFP3	rs12600437	F	CCTGCCACTGCTGACAGATA
		R	AGGGGCAAAACCAAGAAGAG
		E	AAAGCACTTAAGTTCTCATGACCTGGACC

3.2.4 Demethylation study

The spinal cord NSC lines were treated with the demethylating agent 5-Aza-2'-deoxycytodine (5-Aza). The cells were plated at normal seeding density in growth media as described in section 2.1. Two hours after plating, 5-Aza was added to culture vessels to make a final concentration of 1 μM. Control cells were cultured in parallel and exposed to the same concentration of vehicle (1 μl of DMSO per 1 ml of media). Cells were treated for four days with a full media change and fresh 5-Aza added each day. The study was carried out in biological triplicate, with cells separated by multiple passages. RNA was extracted and cDNA generated as described in sections 2.2.2 and 2.2.7 respectively for SNaPshot primer extension analysis.

The allelic expression of twelve RME genes was assessed before and after 5-Aza treatment by SNaPshot primer extension as described in section 2.5. The genes assayed were *C1QL3*, *CAT*, *DNAJC15*, *GRIA1*, *GRID1*, *JAG2*, *KIAA1324L*, *PMP2*, *SFMBT2*, *TNFRSF10D*, *TTN* and *VAX2*. Primer sequences are presented in Table 3.1. Difference in the percentage representation of the silenced allele before and after treatment was tested for statistical significance using a two-tailed t-test in Prism.

3.2.5 Gene ontology analysis

Autosomal genes subject to RME were interrogated for enrichment of specific gene classes and functional terms using the DAVID Gene Functional Classification Tool (http://david.abcc.ncifcrf.goc/). The RME gene lists from each donor were tested against a

reference gene set comprised of all expressed genes from that donor. The large protocadherin family were removed from the gene sets to avoid bias.

3.2.6 Foetal brain epigenetic analysis

The epigenetic status of the monoallelic and biallelic genes were compared in public data from human foetal brain. The analysis was carried for all novel monoallelic genes combined (same allelic choice, RME and unclassified) and RME were also tested independently. DNA methylation data was obtained from the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo; reference: GSM664920) and histone modification (reference: UW H-22510) and DNase sensitivity (reference: H-22510) data was retrieved from the NIH Epigenomics Atlas (http://www.genboree.org/epigenomeatlas/). Differences between the RME and biallelic gene lists derived from each genotype were tested for statistical significance using a two-tailed t-test in Prism.

3.3 Personal contribution

The microarrays were run at the Wellcome Trust Centre for Human Genetics at the University of Oxford. My colleague, Dr Aaron Jeffries, carried out the initial data processing to calculate gene Δ_β values. I validated the BeadChip data by single base primer extension. We worked in combination to interpret the data and I carried out further bioinformatic analyses. I also carried out the demethylation study. Collaborators Dr Chloe Wong and Dr Ruth Pidsley ran the methylation microarrays. I used the resulting data to assess the methylation status of the novel RME genes.

3.4 Results

3.4.1 Microarray assay validation

The reproducibility of the microarray assay was demonstrated with the spinal cord line SPC01, which was evaluated in triplicate. A Pearson's correlation of the Δ_β values was carried out between each pair and an R between 0.87 and 0.89 was observed (see Figure 3.4).

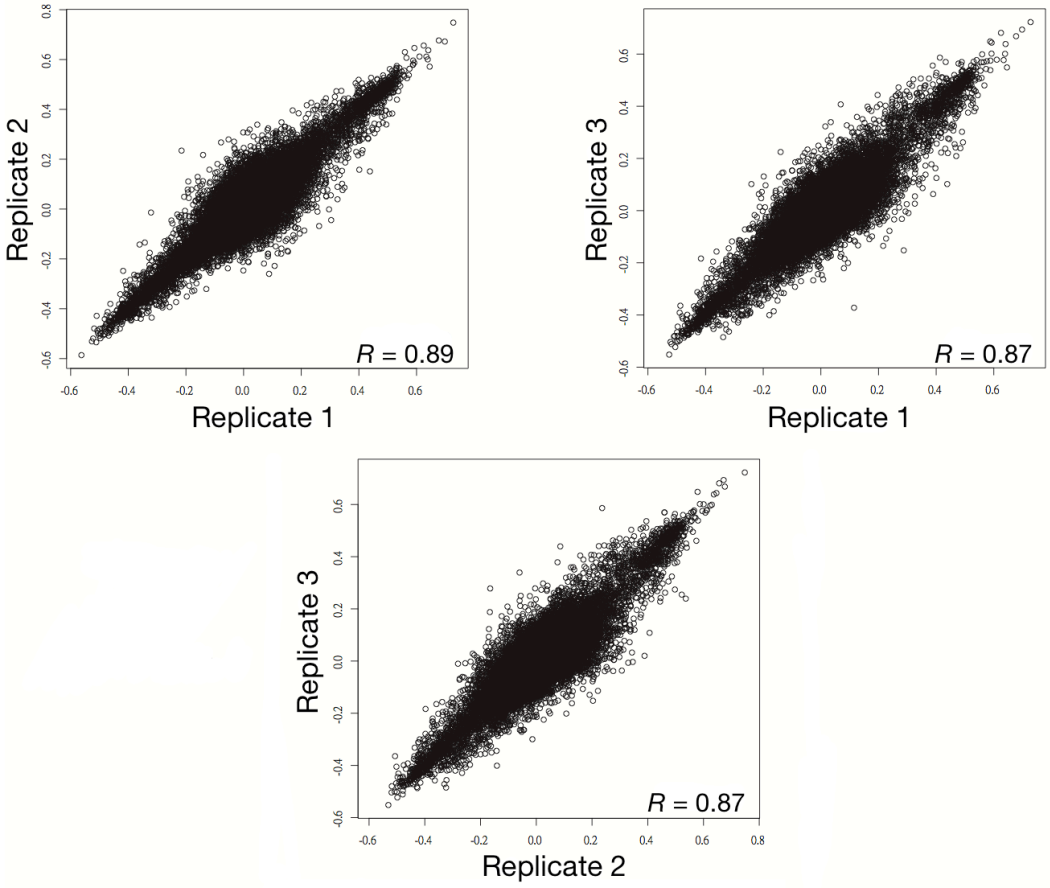


Figure 3.4 | Comparison of autosomal SNP probe Δ_β values between the three biological replicates of spinal cord line SPC01. Pearson's coefficient is presented.

In order to validate the use of the Illumina's BeadChip platform for the analysis of allele-specific expression, the detection of XCI in the female spinal cord lines was used as a proof of principle. 85% of assayed SNPs on the X chromosome, and 78% of genes, were called monoallelic (see Figure 3.5). This is consistent with previous estimates that approximately 15% of the genes on the human Xi escape silencing (Carrel & Willard 2005). Known imprinted genes were also called monoallelic in our dataset; of the 21 human imprinted genes that were assayed (listed on http://www.geneimprint.com), 16 were called monoallelic in all cell lines assayed, 20 were called monoallelic in at least one cell line, and only one did not meet the threshold for monoallelic expression in any cell line assayed.

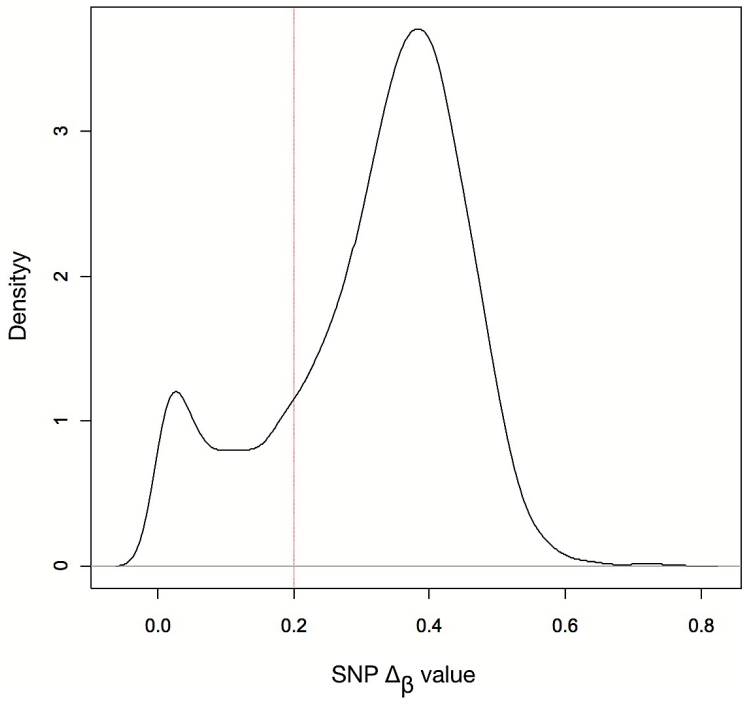


Figure 3.5 | Density plot of SNP Δ_β values across the X chromosome. A Δ_β value of 0.2, marked by the red line, is the threshold for monoallelic expression.

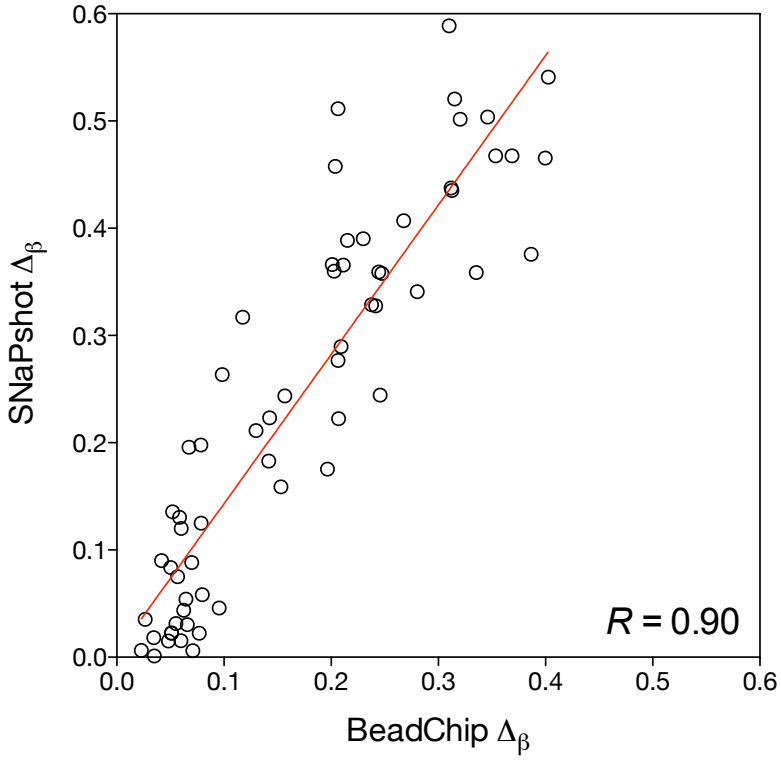


Figure 3.6 | Correlation between BeadChip and SNaPshot Δ_β values. SNaPshot Δ_β values are the mean of three biological replicates.

The accuracy of the BeadChip allelic expression measures was tested by plotting BeadChip Δ_β values against those from the highly quantitative SNaPshot primer extension assay. A significant positive correlation was observed between the two datasets ($p < 0.0001$; $R = 0.90$; see Figure 3.6), demonstrating that the microarray assay is capable of quantitative detection of allele-specific expression. Notably, the microarray data is compressed with a slope of 1.4. Therefore, the threshold Δ_β value for allelic imbalance of 0.1 actually represents 0.14 (an allele ratio of 64:36), and the threshold Δ_β value for monoallelic expression of 0.2 reflects an actual Δ_β of 0.28 (a ratio of 78:22). The high validation rates also demonstrate the stability of allele-specific expression in biological replicates separated by time.

3.4.2 Autosomal monoallelic expression is widespread

More than 9,000 autosomal genes were assayed in lines from each genotype, of which 1.6 to 2.2% showed evidence for monoallelic expression in at least one clone (see Table 3.2). Of the monoallelic genes detected, 7.4 to 10.5% are known imprinted genes. The novel monoallelic genes were dispersed throughout the genome (see Figure 3.7, Figure 3.8 and Figure 3.9). Between 34 and 57% of the novel monoallelic genes showed evidence for RME, where either monoallelic and biallelic clones are observed from the same genotype (see Figure 3.10A), or monoallelic clones with a different allelic choice are observed from the same genotype (see Figure 3.10B). Another set of novel monoallelic genes displayed the same allelic choice in all three clones from a single donor. Finally, genes were designated "unclassified monoallelic" when only one or two clones showed detectable monoallelic expression in the same direction.

Table 3.2 | Autosomal monoallelic expression in human NSCs summary.

	Cortical	Striatal	Spinal
Assayed genes	10,150	9,417	9,085
Monoallelic	185 (1.8%)	203 (2.2%)	143 (1.6%)
Known imprinted	15 (0.1%)	15 (0.1%)	15 (0.1%)
Random monoallelic	89 (0.9%)	107 (1.1%)	43 (0.5%)
Same allelic choice	19 (0.2%)	22 (0.2%)	25 (0.3%)
Unclassified monoallelic	63 (0.6%)	60 (0.6%)	61 (0.7%)

Figure 3.7 | Distribution of novel autosomal monoallelic genes (red) detected in the three cortical NSC cell lines. Known imprinted genes and protocadherins are excluded.

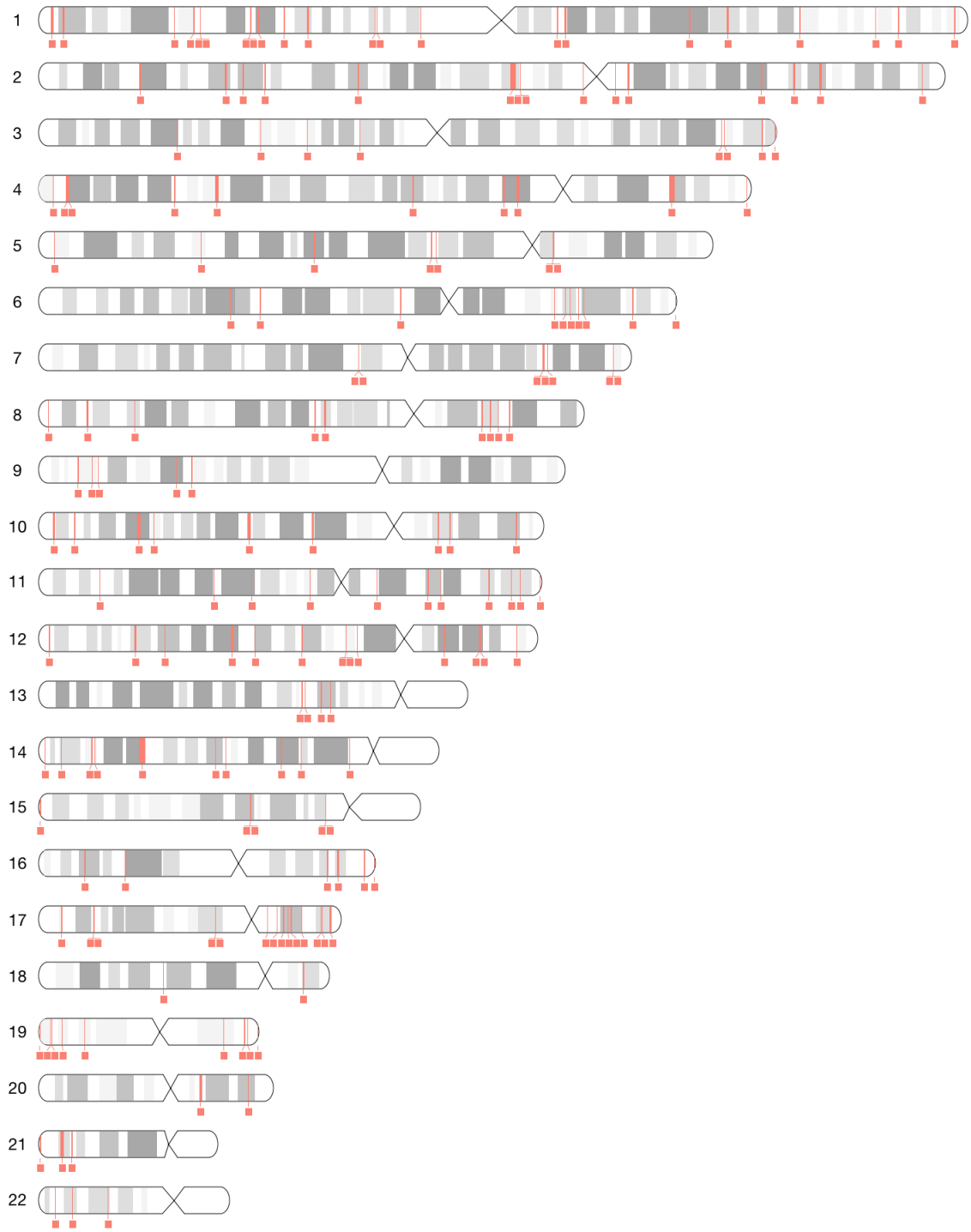


Figure 3.8 | Distribution of novel autosomal monoallelic genes (red) detected in the three striatal NSC cell lines. Known imprinted genes and protocadherins are excluded.

Figure 3.9 | Distribution of novel autosomal monoallelic genes (red) detected in the three spinal cord NSC cell lines. Known imprinted genes and protocadherins are excluded.

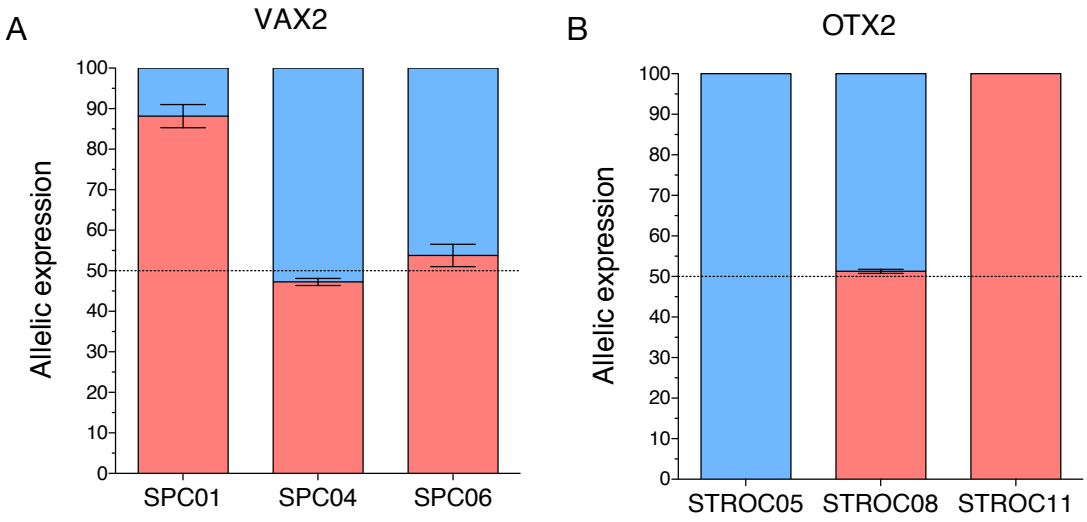


Figure 3.10 | Two presentations of RME. Bars illustrate the mean percentage representation of each allele (n=3) and error bars signify the SEM. Red represents the major frequency allele and blue represents the minor frequency allele of the tag SNP. Data is from SNaPshot primer extension allele-specific expression assay.

3.4.3 Common monoallelic genes between genotypes

A significant overlap of the novel monoallelic genes was observed between samples from different genotypes (least significant pairwise chi squared p value < 0.0001). No more than 2.4 genes are expected shared between two donors by chance and none are expected shared by all three. However, between 14 and 24 genes are observed shared between two donors, and two are shared between all three (see Figure 3.11). The two genes shared between all three donors are *TNFRSF10D* and *ACCS*. A statistically significant overlap between genotypes was also observed for the RME subset (least significant pairwise chi squared p value < 0.0001). Furthermore, a 4.3 fold enrichment of the novel RME genes detected in this study are observed in the monoallelic loci identified by Gimelbrant et al. (2007; $p < 0.0001$, Fisher's exact test).

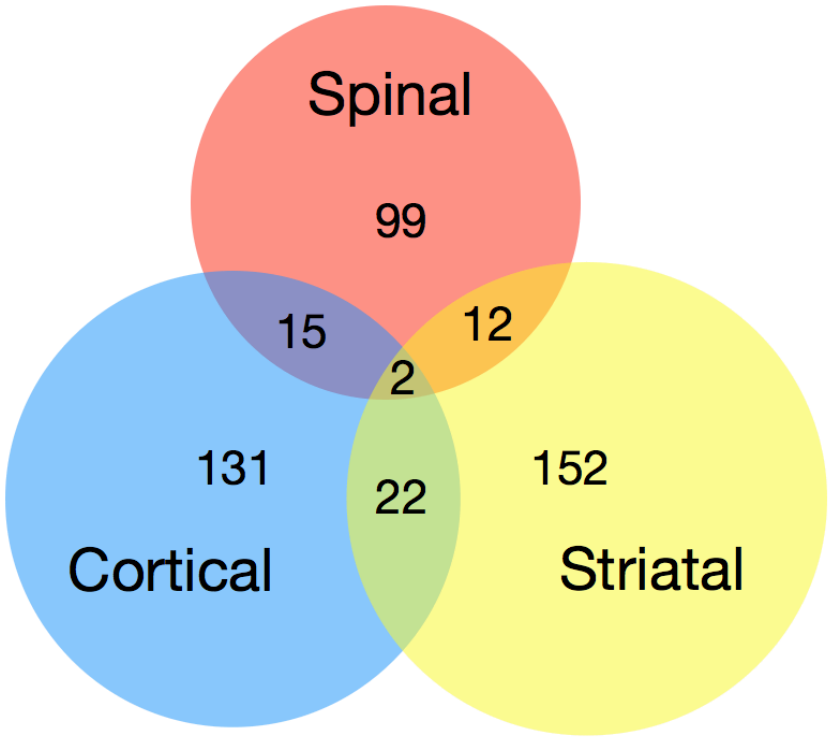


Figure 3.11 | Area-proportional Venn diagram illustrating the overlap of novel autosomal monoallelic genes between genotypes. Known imprinted genes were omitted from this analysis.

3.4.4 RME genes show bias in allelic choice

The randomness of allelic choice was assessed for the novel RME genes detected in this study. Of the RME set that show monoallelic expression in two clones and biallelic expression in the third, 71% (29 genes) are expressed from the same allele in both monoallelic clones, whereas only 29% (12 genes) are expressed from alternate alleles. This is a significant deviation from the 1:1 distribution expected if allelic choice was genuinely random ($p = 0.006$, binomial exact test).

3.4.5 RME gene ontology

Within the novel RME gene set, significant enrichment for the gene ontology term "signal" (1.8 to 2.3-fold) and the topological domains "extracellular" (1.8 to 2.8-fold) and "transmembrane" (1.4 to 1.7-fold) were observed in all three genotypes. More than 30% of all genes subject to RME detected in this study were transmembrane glycoproteins. Among the RME genes encoding transmembrane signalling proteins were the neurotransmitter receptor subunits GABA-A receptor gamma 3 (*GABRG3*), glutamate receptor ionotropic alpha 1 (*GRIA1*) and glutamate receptor ionotropic delta 1 (*GRID1*). Developmental terms were particularly enriched in the brain derived NSCs. For example, a 3.8 to 4.9-fold enrichment for the gene ontology term "cell morphogenesis involved in differentiation" was found. Developmental genes subject to RME included orthodenticle homeobox 2 (*OTX2*; see Figure 3.10B), oligodendrocyte lineage transcription factor 2 (*OLIG2*), ventral anterior homeobox transcription factor 1 (*VAX1*), neurotrophin-3 (*NTF3*) and neurexin 3 (*NRXN3*). Summary tables displaying count, percentage, fold enrichment and *p* values are included in Appendix 9.3 on page 206 and Appendix 9.4 on page 207. Full annotation can be retrieved at http://onlinelibrary.wiley.com-/doi/10.1002/stem.1155/full supporting information File S4.

3.4.6 RME genes show increased DNA methylation

Data from the Infinium HumanMethylation27 BeadChip (Illumina) was used to compare DNA methylation levels at RME and bialleleic loci. I show that the RME loci detected in this study show increased levels of DNA methylation compared to biallelic loci (see Figure 3.12). The increased level of methylation was statistically significant for RME genes detected in each of the six NSC lines assayed (least significant p = 0.0248). Statistical significance was also observed when comparing all novel monoallelic genes (RME, same allelic choice and unclassified combined) to biallelic genes (data not shown).

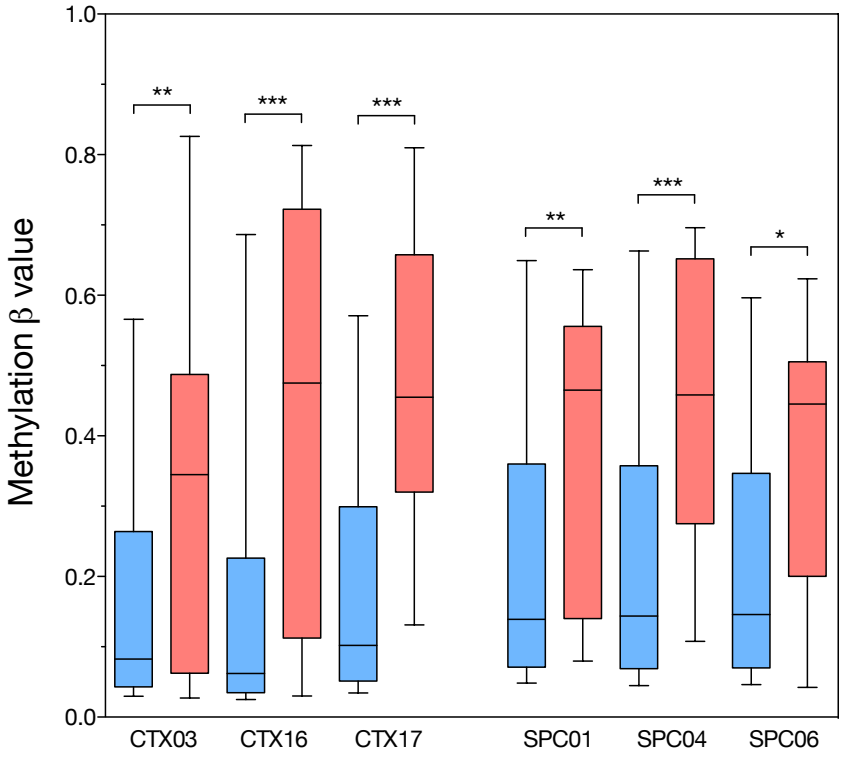


Figure 3.12 | Methylation state of novel RME genes in human NSCs. Methylation β values are presented for all biallelic (blue) and RME (red) genes detected in each of the cortical and spinal cord NSC lines. T-test p value summaries are presented. Whiskers represent the 10-90th percentile.

3.4.7 The impact of DNA demethylation on RME

In order to determine the extent that DNA methylation is deterministic in RME, the spinal cord NSC lines were treated with the demethylating agent 5-Aza. The allelic expression of twelve RME genes was assayed before and after treatment. All twelve genes showed a significant reduction in allelic expression imbalance (see Figure 3.13). The mean increase in representation of the silenced allele after treatment was 14% ($p < 0.0001$, paired two-tailed t-test).

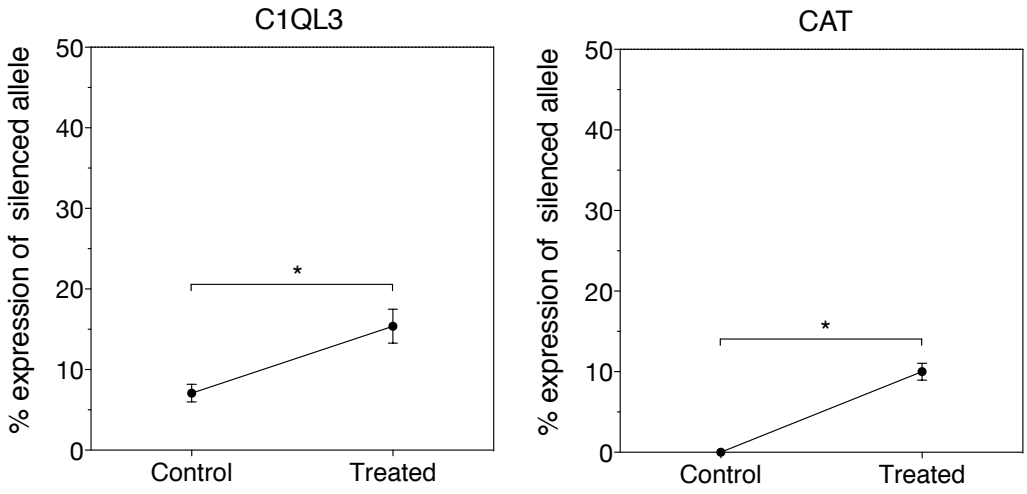


Figure 3.13 | Allele-specific expression of twelve RME genes in monoallelic clones before and after treatment with the demethylating agent 5-Aza. T-test *p* value summaries and SEM are presented. A summary box plot combining data from all genes assayed is included. Whiskers display minimum to maximum value and a paired t-test *p* value summary is presented (continued overleaf).

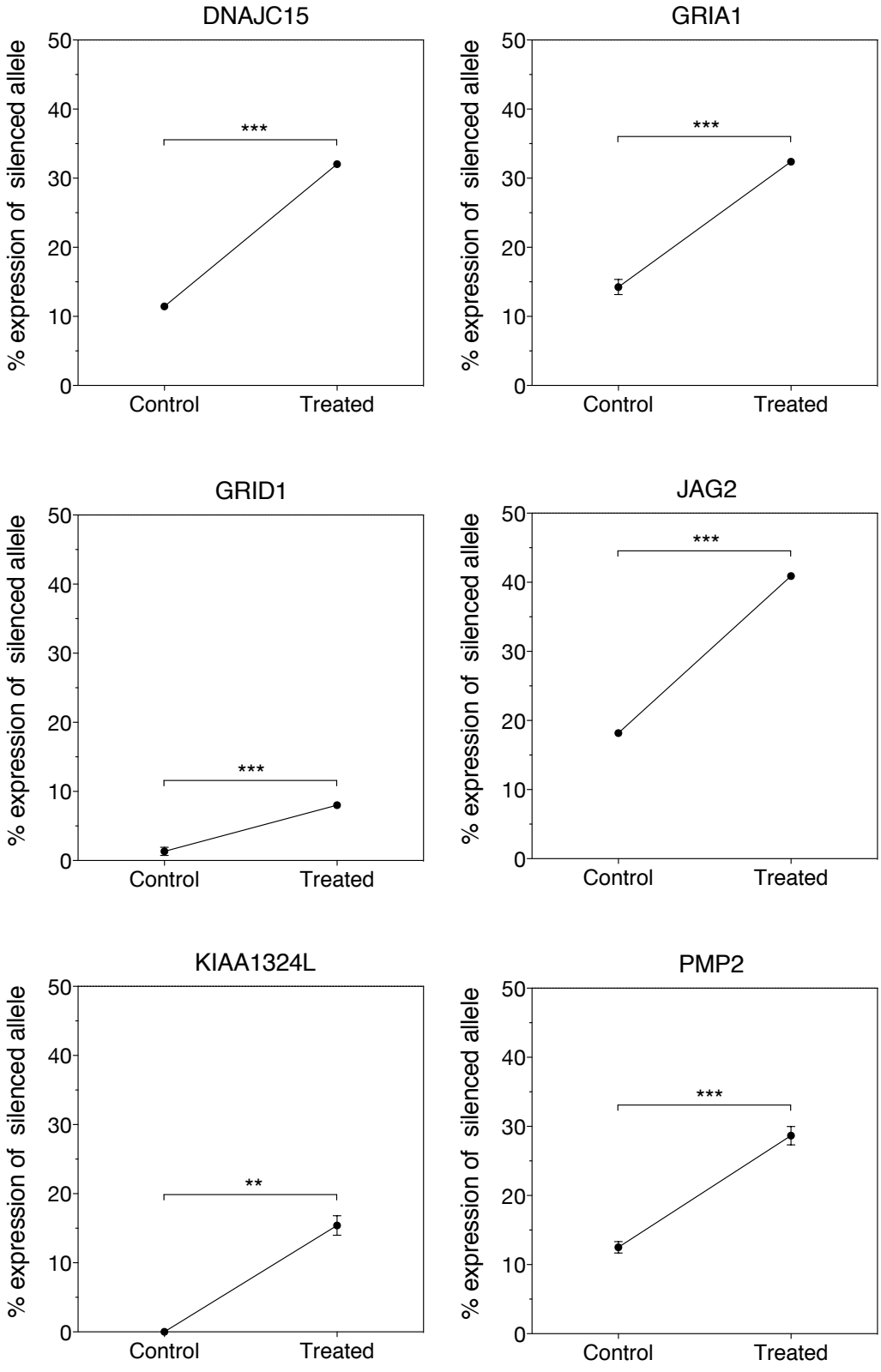


Figure 3.13 | Continued.

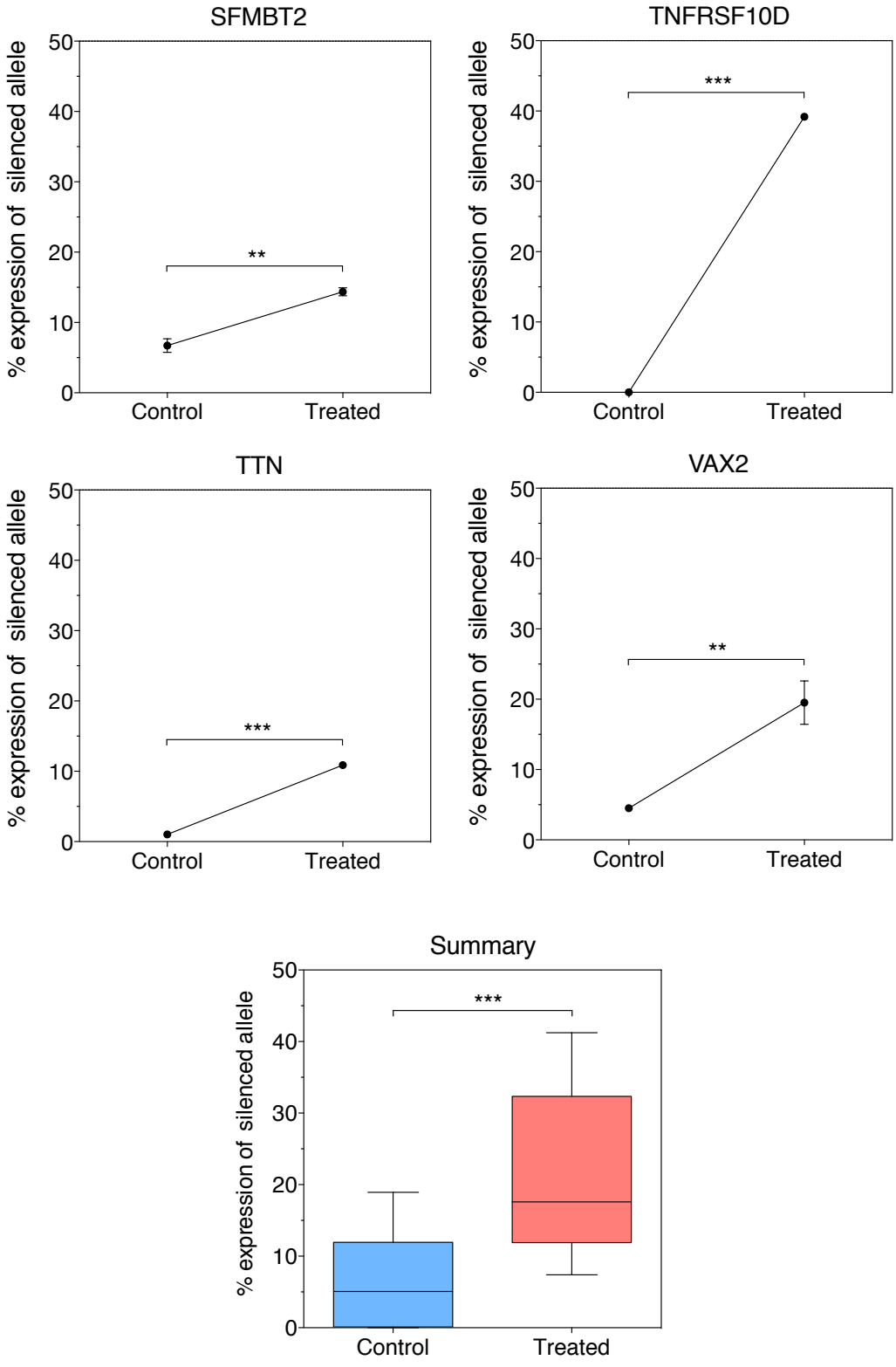


Figure 3.13 | Continued.

3.4.8 Epigenetic state of novel RME genes *in vivo*

In an attempt to test the relevance of these findings *in vivo*, the epigenetic status of the novel RME genes was assessed in publicly available datasets from human foetal brain. As observed in the NSCs, the RME genes detected from each donor have significantly increased levels of DNA methylation in human foetal brain when compared to biallelic loci (see Figure 3.14).

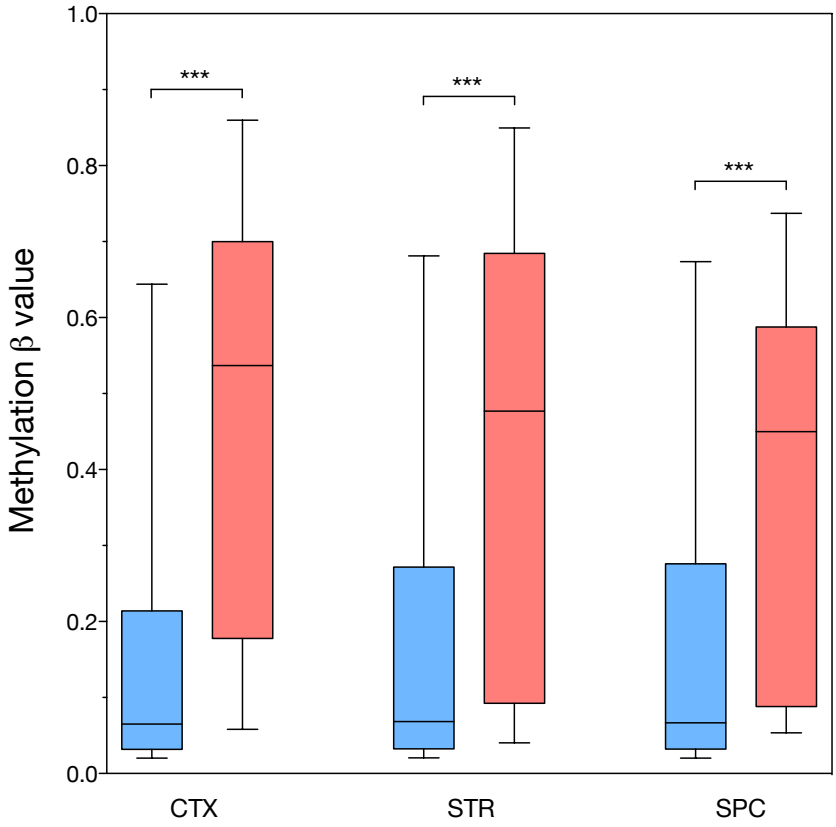


Figure 3.14 | Methylation state of novel RME genes in human foetal brain. Methylation β values derived from human foetal brain are presented for biallelic (blue) and RME (red) genes detected in the cortical (CTX), striatal (STR) and the spinal cord (SPC) NSC lines. T-test *p* value summaries are presented. Whiskers represent the 10-90th percentile.

In foetal brain, the novel RME genes reported in this study are also associated with an increase in repressive H3K27me3 and a reduction in signatures of open chromatin H3K4me3, H3K9ac and DNaseI hypersensitivity when compared to biallelic loci (see Figure 3.15). H3K9me3 is the only chromatin measure investigated that does not follow the expected trend; it is usually associated with repressive chromatin, but it is enriched in the biallelic gene set.

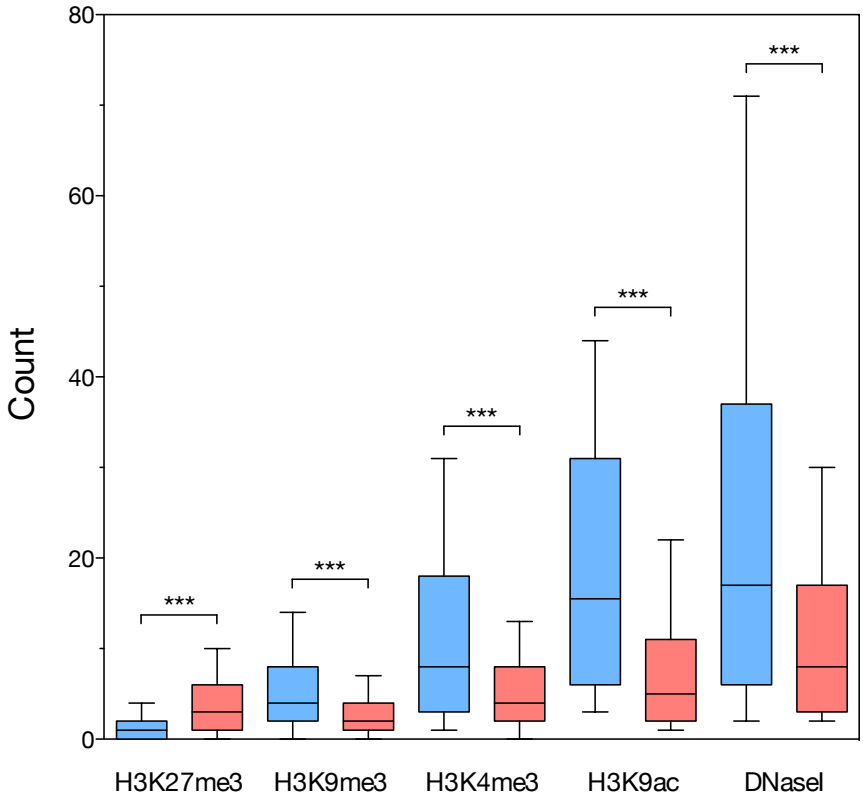


Figure 3.15 | Histone modification and DNaseI sensitivity of novel RME genes in human foetal brain. Chromatin data for biallelic (blue) and RME (red) genes detected in the cortical donor are presented. T-test *p* value summaries are displayed. Whiskers represent the 10-90th percentile. This data is representative of data from the other two donors (see Appendix 9.5 on page 208).

3.5 Discussion

This study demonstrates for the first time that autosomal monoallelic gene expression is widespread in human neural tissue. Using a genome-wide assay of allele-specific expression in NSCs, I find that 1.6 to 2.2% of assayed autosomal genes display monoallelic expression and 0.5 to 1.1% show evidence of RME.

The prevalence of autosomal monoallelic expression reported in this chapter is likely to be an underestimation for two reasons. Firstly, an additional 2,000 genes containing a single informative SNP were omitted from the final analysis. 5% of this group were predicted to be monoallelic and, despite many of these genes being independently validated, they were removed from the final analysis to minimise the inclusion of false positives. Secondly, this study had limited power to detect RME with only three clones per genotype; a significant proportion of genes subject to RME would be expected to be biallelic in the three clones by chance. Additionally, a significant proportion of genes subject to RME would be expected to be monoallelic for the same allele by chance. Therefore, the present study is expected to underestimate both the proportion of autosomal monoallelic genes as well as the fraction that display RME.

The number of autosomal monoallelic genes detected in this study is considerably lower than the 9.4% reported in human B-lymphoblastoid cells (Gimelbrant et al. 2007). This discrepancy could be explained by a major technical difference between our assays: the exclusion of genes with a single informative SNP in the present study. As stated above, this subset was removed to minimise the inclusion of false positives. However, Gimelbrant et al. (2007) included genes with a single informative SNP in their analysis.

The proportion of monoallelic genes they detected drops from 9.4 to 2.2% by excluding this subset. Furthermore, unlike the study reported by Gimelbrant et al., the present study implemented a gene expression threshold and rejected monoallelic genes that did not meet it. Another factor to consider is possible tissue-specific differences; indeed, it has been observed that autosomal monoallelic expression is considerably more widespread in lymphoblastoid cells than fibroblasts in mice (Zwemer et al. 2012). Recent studies in clonal mouse NSCs report the prevalence of autosomal RME to be between 2 and 4%, consistent with the data presented in this chapter (Wang et al. 2010; Li et al. 2012).

Unlike the vast majority of established monoallelic genes, those detected in this study are scattered across the autosome. XCI is coordinated at a chromosome-wide level, while genomic imprinting is classically associated with clusters of genes that are co-regulated under the control of common regulatory elements. Clustering is also observed for most established gene families subject to RME including antigen receptors, odorant receptors and protocadherins. The dispersed nature of the monoallelic genes reported in this chapter is consistent with other reports of widespread autosomal monoallelic expression in human lymphoblastoid cells (Gimelbrant et al. 2007) and mouse NSCs (Li et al. 2012) and indicates that these genes are independently regulated.

I assessed the randomness of allelic choice for the novel RME genes detected in this study. Of the RME set that show monoallelic expression in two clones and biallelic expression in the third, 71% are expressed from the same allele in both monoallelic clones, whereas only 29% are expressed from alternate alleles. This is a significant deviation from the 1:1 distribution expected if allelic choice was genuinely random and it is consistent with observations made in a genome-wide assessment monoallelic

expression in clonal mouse lymphoblasts (Zwemer et al. 2012). The bias towards one allele indicates that, at least for a proportion of this subset, the allelic choice is not genuinely random. Instead, it is suggestive of a genetic component where one allele is more susceptible to monoallelic expression than the other. Further studies with more clones per genotype would provide greater power to determine the randomness of allele selection for the novel RME genes detected.

Further support for the notion that RME has a genetic component comes from my observation that the novel RME genes detected show significant overlap between donors. Furthermore, I observed a highly significant 4.3 fold enrichment of the RME genes detected in this study in the monoallelic gene set reported by Gimelbrant et al. (2007) in B-lymphoblastoid cells, despite inherent tissue differences in expression.

These observations warranted further investigation to identify the genetic sequences that predispose genes to RME. It has previously been reported that long interspersed nuclear element (LINE)-1 sequences are associated autosomal monoallelic genes together with fewer CpG islands and short interspersed nuclear elements (SINEs; Allen et al. 2003). My colleague, Dr Aaron Jeffries, tested the novel monoallelic genes detected in this study for association with these motifs. No significant difference was observed for the prevalence of LINE-1 repeats between monoallelic and biallelic genes. However, a significant reduction in CpG density was observed at monoallelic gene transcriptional start sites ($p = 0.0008$, Wilcoxon rank-sum test) as well as a reduction in SINE repeats across the length of the transcript ($p < 0.0001$; Jeffries et al. 2012).

Taken together, the apparent non-randomness of allelic choice, the overlap between genotypes and the association with reduced CpG density and SINE repeats strongly argue for a genetic component predisposing genes for RME. As stated above, the dispersed nature of the monoallelic genes reported in this chapter indicates that these genes are regulated independently. Unravelling the mechanisms controlling RME will require further research at the gene-specific level.

Gene ontology analysis was carried out on the novel RME genes detected in this study with the aim of revealing the role they may play in brain development, function and disease. The RME genes detected in each of the three genotypes were significantly enriched for the gene ontology term "signal" and the topological domains "extracellular" and "transmembrane". This is consistent with observations made by Gimelbrant et al. (2007), who found a disproportionately large fraction of monoallelic cell surface proteins in human B-lymphoblastoid cells. Additionally, developmental terms were enriched in RME genes detected in the brain-derived NSC lines. These observations are revealing when considered in respect to the better-established gene families subject to RME; antigen receptors, odorant receptors and protocadherins are all transmembrane proteins involved in cell signalling, and the RME of these genes acts to increase cell specificity and tissue diversity. Furthermore, the RME of odorant receptors and protocadherins is known to be critical from proper wiring of the brain during development by generating a specific cell identity (Imai et al. 2010; Wang et al. 2002). The results presented in this chapter are consistent with widespread autosomal RME playing a similar role in human NSCs. Again, more detailed work on the gene-specific level, as well as *in vivo* validation, will be required to determine the biological significance of these findings.

The novel monoallelic loci identified in the spinal cord and cortical cell lines display significantly higher levels of DNA methylation at promoter regions than biallelic genes. This is true both when all novel monoallelic genes are combined (RME, same allelic choice and unclassified) and when RME genes are tested independently. Allele-specific DNA methylation is known to play a critical role in establishing and maintaining monoallelic expression in genomic imprinting, XCI and the RME of several gene families (Goldmit & Bergman 2004). Allelic skewing of DNA methylation has also been associated with genetic *cis* effects (Schalkwyk et al. 2010). While this assay was not allele-specific, the results are consistent with monoallelic silencing by DNA methylation.

While this data demonstrates that the novel RME genes are associated with increased levels of DNA methylation, the extent to which DNA methylation causes monoallelic expression remained to be demonstrated. After treatment with the demethylating agent 5-Aza, all twelve RME genes assayed showed significant reductions in allelic imbalance, with a mean reduction of 14%. This could either represent a partial reduction in allelic expression imbalance per cell, or complete reversion to biallelic status in a proportion of cells. Single cell analysis of the imprinted *PLAGL1* indicated that 5-Aza induced loss of imprinting is an all-or-none process (Diplas et al. 2009). Similar single cell analysis would be required to determine if this is true for the RME genes described in this chapter. Nonetheless, this data indicates that DNA methylation directly regulates the monoallelic expression of the RME genes assayed.

One possible criticism of the findings reported in this chapter is that they could represent an artefact of tissue culture, or the immortalisation process, where epigenetic regulation has gone awry. Indeed, concerns have been raised over the epigenetic stability of cell

lines after extended passages *in vitro* (Grafodatskaya et al. 2010). It is feasible that the enrichment of monoallelic loci observed between donors represents no more than genes prone to epigenetic dysregulation *in vitro*. In order to test the relevance of these findings *in vivo*, the epigenetic status of the novel RME genes was assessed in publically available data from human foetal brain. The RME genes discovered in this study were enriched for signatures of repressive chromatin (DNA methylation and H3K27me3), while the biallelic genes were associated with marks of open chromatin (H3K4me3, H3K9ac and DNaseI hypersensitivity). Notably, H3K9me3 does not follow the expected trend; it is usually associated with repressive chromatin, but it is enriched in the biallelic gene set. It has, however, previously been described in actively transcribed regions (Vakoc et al. 2005). While these observations do not directly prove that these genes are similarly expressed in foetal brain, it does demonstrate that the RME subset are distinct from the biallelic genes *in vivo* in a manner consistent with them being expressed from a single allele. This is evidence in support of biological relevance for the reported findings in the developing human brain. Further studies are required to determine whether these genes are also subject to RME *in vivo*. The non-clonal nature of brain tissue represents and obstacle in the way of investigating RME *in vivo*; neighbouring cells expressing the alternate allele, or both alleles, are expected to mask the effect. Single cell RNA-seq is now established (Tang et al. 2010), and this methodology represents the ideal platform to investigate whether RME is widespread in human neural tissue *in vivo*.

Finally, this study demonstrates that the Illumina BeadChip platform is suitable for genome-wide assays of allele-specific expression. High reproducibility between biological replicates was demonstrated as well as a strong correlation with data from the highly quantitative single base primer extension assay. This is consistent with the

observations made by other groups who have used Illumina's BeadChips for similar assays (Ge et al. 2009; Morcos et al. 2011). While RNA-seq is becoming the assay of choice, Illumina's BeadChip platform remains an efficient and cost-effective method to assay allele-specific expression on a genome-wide scale.

3.6 Conclusion

From the results presented in this chapter, I conclude that autosomal monoallelic expression is widespread in human NSCs *in vitro*. Approximately half of the novel monoallelic genes detected show evidence of RME. The overlap of RME genes between genotypes, as well as the bias in allelic choice, argues for a genetic component predisposing genes to RME. The novel RME genes are enriched for transmembrane signalling proteins, and those detected in the brain derived NSCs also show enrichment for developmental terms. Genes subject to RME are associated with increased levels of DNA methylation, and the results from 5-Aza treatment indicate that DNA methylation is directly regulating the monoallelic silencing of these genes. Finally, I demonstrate evidence in support of an *in vivo* relevance for these findings; the novel RME genes are epigenetically distinct from biallelic genes *in vivo* in a manner consistent with monoallelic expression. The presence of widespread RME *in vivo*, in the human central nervous system, remains to be demonstrated. The extent to which the novel monoallelic genes are retained through neural differentiation is investigated in Chapter 4 and potential functional implications are explored in Chapter 5. The monoallelic genes classified "same allele choice" in multiple genotypes are further investigated in Chapter 6.

Chapter 4. Monoallelic expression through neural differentiation

4.1 Introduction

Monoallelic gene expression is typically initiated early in development and stably maintained in adult tissues. This is true for XCI (Amos-landgraf et al. 2006) as well as the established autosomal gene families subject to RME described in section 1.2.3 (Pernis et al. 1965; Chess et al. 1994; Esumi et al. 2005). There are, however, examples of imprinted genes that show tissue and developmental stage specific monoallelic expression. For example, the imprinted *Dlk1* is exclusively expressed from the paternal allele during mouse embryogenesis before reverting to biallelic status in the neurogenic niche of the adult brain (Ferrón et al. 2011). The extent to which widespread autosomal RME is maintained through development has not been properly characterised. While a handful of select examples have been shown to retain monoallelic expression after the differentiation of mouse NSCs (Wang et al. 2010; Li et al. 2012), at the time of this study, no global assessment has been performed. Supposing that the novel monoallelic expression reported in Chapter 3 is occurring in human NSCs *in vivo*, the extent to which it is maintained through neural development, as well as the developmental stage at which it is initiated, will have significant implications for its role in neural development, function and disease. The NSC lines employed in this thesis offer the opportunity to address this question.

In this chapter I report a genome-wide screen for allele-specific expression in the three spinal cord NSC lines before and after differentiation. I aim to determine the extent to which non-imprinted autosomal monoallelic expression is retained after the NSCs are differentiated into neurons and glia. Additionally, I assess DNA methylation data derived from adult brain to test for *in vivo* relevance of this phenomenon and I perform a pilot study in an iPSC line generated from one of the assayed spinal cord NSC lines, assessing the allelic status of five RME genes after reprograming.

4.2 Methods

General molecular biology and cell culture methods are described in Chapter 2. Experimental details specific to this chapter are included below.

4.2.1 Cell differentiation

The three spinal cord lines were differentiated into neurons and glia. Cells differentiated for the purpose of nucleic acid extraction were grown in Nunclon™ Δ Surface 75 cm^2 tissue culture flasks (Thermo Scientific) and cells differentiated for immunocytochemistry were cultured in Nunclon™ Δ Surface 96 well optical bottom plates (Thermo Scientific). They were grown to 95% confluency in Reduced Modified Media (see section 2.1). At this point, the growth media was removed and replaced with media without 4-OHT, bFGF and EGF. The cells were treated with the γ-secretase inhibitor DAPT (10 μM; Sigma-Aldrich) for 48 hours. After the removal of DAPT, cells were cultured for 8 further days in the absence of 4-OHT, bFGF and EGF. The media was

changed every 72 hours. RNA extraction and cDNA synthesis was carried out as described in sections 2.2.2 and 2.2.7.

4.2.2 Immunocytochemistry

Media was removed from each well of the 96 well plate and replaced with 100 µl 4% paraformaldehyde (PFA; Alfa Aesar). After 15 minutes, the PFA was removed and replaced with 200 µl phosphate buffered saline (PBS; Sigma-Aldrich). Cells were then permeabalised with 100 µl of tris-buffered saline (TBS; Sigma-Aldrich) supplemented with 0.1% Triton X-100 (Sigma-Aldrich) and 10% normal donkey serum (NDS; Sigma-Aldrich) for 30 minutes at room temperature. Each well was rinsed with 100 µl TBS before adding the primary antibody (see Table 4.1) diluted in TBS with 1% NDS. Negative (isotype) controls were prepared using non-specific monoclonal IgGs (Abcam) in place of the primary antibody, matching host and concentration. The primary antibody was incubated for 16 hours at 4°C. Cells were then washed 4 times with TBS before incubation with the complimentary Alexa Flour® 488 secondary antibody (Invitrogen) diluted 1 in 500 in 75 µl of TBS for 1 hour at room temperature in the dark. The preparations were washed a further 4 times with TBS before incubation with 0.02 mg/ml Hoescht33342 (Sigma-Aldrich) for 5 minutes at room temperature and finally washed 4 more times with TBS.

Immunocytochemically stained cells were imaged by fluorescence microscopy. Both the TCS SP5 confocal microscope (Leica) in conjunction with the LAS AF software (Leica) and the 1X70 inverted microscope (Olympus) with the Axio Vision Digital Image Processing Software (Carl Zeiss Inc.) were used for image capture and processing.

All cell counting was carried out in biological triplicate, where the experiment was replicated with cells plated several passages apart. Each biological replicate consisted of three wells (technical replicates) for each condition. Each technical replicate consisted of three randomly placed non-overlapping images taken per well with the 40x objective. Images were imported into ImageJ and nuclei and target-positive cells were counted manually. The three well images were averaged to generate one technical replicate. The three technical replicates (wells) were averaged to generate one biological replicate (plate), which was then used for statistical analysis. All cell counting was carried out from images taken from the 1X70 inverted microscope (Olympus) and processed with the Axio Vision Digital Image Processing Software (Carl Zeiss Inc.). Exposure times were kept consistent for each target. Differences in the proportion of marker-positive cells between cell lines were tested for statistical significance using a one-way ANOVA for each marker (Prism).

Table 4.1 | Primary antibodies used for immunocytochemistry.

Target	Primary antibody			
	Supplier	**Code**	**Host**	**Conc.**
Ki67	Abcam	AB15580	Rabbit	1:500
NESTIN	Millipore	ABD69	Rabbit	1:500
SOX2	Millipore	AB5603	Rabbit	1:500
S100β	DAKO	Z0311	Rabbit	1:500
TAU	DAKO	A0024	Rabbit	1:2000
O1	Millipore	MAB344	Mouse	1:500

4.2.3 Microarray-based assay of monoallelic expression

A genome-wide assay of allele-specific expression was carried out for the three spinal cord NSC lines before and after differentiation. This study was carried out as described in section 3.2.1 with two key differences. Firstly, the OmniExpress BeadChips (Illumina) were used for this study rather than the Omni1-Quad BeadChips used previously. These microarrays were selected because the Omni1-Quad is no longer in production. The OmniExpress BeadChips have lower SNP probe density, covering ~700,000 loci compared to ~1,000,000 of the previous array. The second key difference in study design was the use of biological triplicates for each cDNA sample. Genomic DNA samples were assayed from each of the three spinal cord lines.

Unlike the microarrays used in Chapter 3, I processed and scanned these in-house. With the exception of using 5 µl of undiluted reverse transcription reaction product as template for the cDNA samples, the manufacturers recommended protocol was followed as summarised in section 3.2.1.1. The more recent Illumina HiScan™ Reader was used.

Data analysis for this study was carried out as described in section 3.2.1.2 with some small alterations. Due to differences in the array platform, the allelic intensity threshold for cDNA samples was altered to 1,500 using the same principle for selection (see Figure 4.1). Because this study was carried out with biological triplicate, the data analysis was modified slightly. Each replicate was analysed independently, as described in section 3.2.1.2 (see Appendix 9.6 on page 209 and Appendix 9.7 on page 210). For each clone, transcripts were assigned allelic status (biallelic, allele expression imbalance or monoallelic) when at least two of the three replicates agreed.

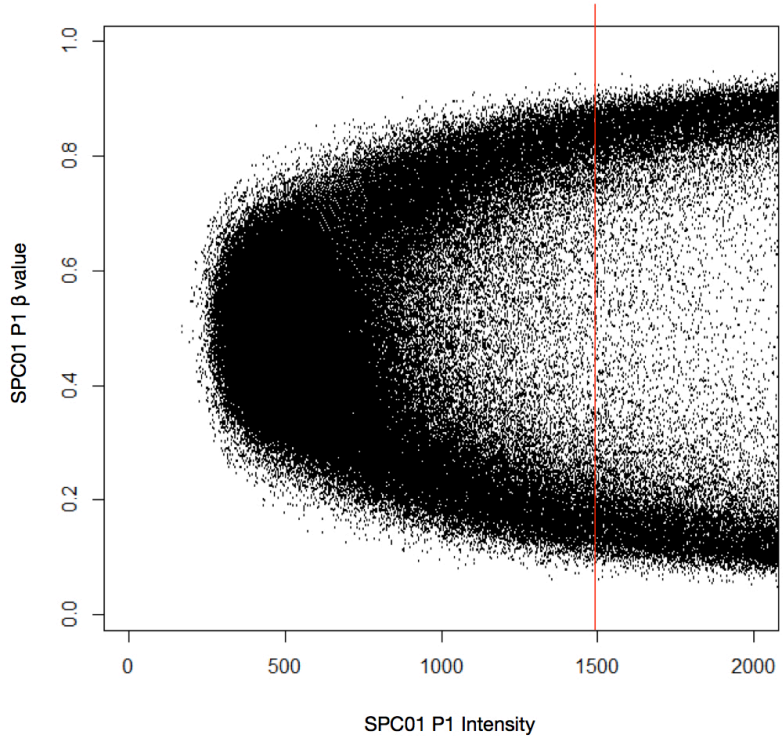


Figure 4.1 | SNP probe intensity threshold selection. β values of homozygous SNPs are plotted against total intensity for the SPC01 replicate 1 cDNA sample. As with the primary array study, a threshold was selected to remove background signal probes (observed as homozygous SNPs being called heterozygous in the cDNA). Due to different array platforms, this study had a lower threshold.

4.2.4 BeadChip validation by SNaPshot primer extension

The SNaPshot single nucleotide primer extension assay, as described in section 2.5, was carried out for 27 genes in multiple samples to validate the microarray assay. All samples were assayed in biological triplicate, separated by multiple passages. The genes assayed and the primers used are presented in Table 4.2. Δ_β values were calculated and correlated with those from the OmniExpress BeadChips. The plot was drawn and Pearson's correlation coefficient calculated in Prism (GraphPad).

Table 4.2 | Primer sequences for single nucleotide primer extension. Forward (F), reverse (R) and extension (E) sequences are presented. Continued overleaf.

Gene	Tag SNP	Primer sequence (5'-3')	
ADAMTS16	rs1863968	F	CCACTAGGCTGCTGGAAGTC
		R	TGGTCTCTGCCTACGAGGTT
		E	CTAGGCTGCTGGAAGTCCTCAGATCCA
C1QL3	rs4747277	F	TGGACACATAAGGGCTTTTTC
		R	CCTGGGAATGATTCAACTTCA
		E	ACATAAGGGCTTTTTCTAAATACCGTAC
C22ORF26	rs7510754	F	TTAGGGCTTGGACATCTTGG
		R	TGAGGAGGAGGAAGAGGACA
		E	TTCAGTCCCCTGGGGAGGTTCAGATGC
CAT	rs769217	F	GCCTGGGACCCAATTATCTT
		R	TGTCCTGCATGCACATCG
		E	TCGAGTGGCCAACTACCAGCGTGA
DDX11	rs10843881	F	GGTGTTTTCTGTCCCTGCTG
		R	TCAATGAAACCAAATGGGAAA
		E	GCAGGGGACTCATGTCTGTCCTACCCG
DSCAM	rs736977	F	AGACAGCCTTCTCTGGGACA
		R	TCAGGCCAATAGCATGTGAG
		E	CTAAAACTTCTGCTCGGATTTGCC
GABRG3	rs10873636	F	CTACTGGTGACCCACCACCT
		R	CGGTGTCTGCAACATAGGAA
		E	ACACTGTCAGGATCTTGCTCATAG
GRIA1	rs707176	F	TCCTGCAGAAAGTCCTGGAT
		R	TGAGGCGTTCTGATTCACAG
		E	GTATCCCTCCTCTGTGGTTGTCAA
GRID1	rs1054979	F	CCGACCAGCAGAGCTTTTTA
		R	CCCCACTCCATTCTGTCATT
		E	CGAGTGTGTGTGGTTTGTGTTGTT
HLA-DOA	rs3129304	F	GGGAGGTCCACCTAGACACA
		R	AATCTGTATTGCTCATTTTGGTC
		E	TAAATCACCATACCACATAGTTTATGTCA
HLA-DRA	rs14004	F	TCTTGTCTGTTCTGCCTCACTC
		R	TCCTAGCACAGGGACTCCAC
		E	CGAGCTCTACTGACTCCCAA

Table 4.2 | Continued.

Gene	Tag SNP	Primer sequence (5'-3')	
HSPA12B	rs3088007	F	ACTCCAGAGGGACAGGTGTG
		R	GCAAGTGGGGTCCTTCATAA
		E	TTCGAGACAAAACACCCGTCTGGGAAG
JAG2	rs1057744	F	TGAACGGGTACCAGTGTGTG
		R	GGGCAGTGGCAGTGGAAG
		E	CGGCATTGCGAGCTGGAACGAGAC
KIAA1324L	rs12535941	F	GGCTACATAGGGGAAGACTAAGC
		R	GCAGCAGAACAAGGCTCTTC
		E	ATGCATTTCTACTGAAGGATCACTAAAAA
NECAB2	rs2280027	F	TCCATCCTCCACAAGAAGGT
		R	GTGTGGTCAGTGTGGGTCAT
		E	CCCCTGCCTCCTGGTCCTGGCCTCTCCCC
PLXDC2	rs3817405	F	CTTCTGTAGGGAGGCTGGTG
		R	TTCTTCTCGAACCACCACAA
		E	TGAACTGGGTGGTTGTCGCTCCTA
PMP2	rs2229015	F	TTTTCAGGTGTGGGGTTAGC
		R	TCTCTCTCAAGCAGCCCACT
		E	ACGAACTGAAAGTACCTTTAAAAATACAGA
PSCA	rs2294008	F	AAGTCACCTGAGGCCCTCTC
		R	CAAGCCTGCCATCAACAG
		E	TCCACCACAGCCCACCAGTGACCA
RALYL	rs10094238	F	AAATGTGTTTTCCCCTTTGC
		R	CAGAATTGCAATATCACCTGACTT
		E	TGTGAAAGGCATTTATGAATGGTAAGGGA
RIC3	rs10839976	F	GCAGTTGTTTTCCCCTTTGA
		R	TTTTGAGCTATTTCTCTAGTGTTGC
		E	CTATACCTCCTTCAATTCTCAGCT
RIN3	rs3814830	F	AGAGGTGGGCAAAGGACAG
		R	CAGCCACCTCTTGGAAATTG
		E	GCACGCCCTTTGCCGCCCACCTCTGATGC
SFMBT2	rs11255084	F	GGGCACCTCCATGTTTTCTA
		R	TGCCCCATATTTTGCATTTT
		E	AAGGCAATCAGTGAACTTCCTC

Table 4.2 | Continued.

Gene	Tag SNP	Primer sequence (5'-3')	
SLC6A15	rs17183577	F	ATGTGTGATTTGCCCACAGA
		R	TCTGAAAGTGGGGGCTTAAA
		E	TCCCTCTCTCAAATTCTGTTACAG
THNSL2	rs4129190	F	TTCTCTGGCTATGCACCTGA
		R	GCTCAGAGCCAATGAAGAGG
		E	AAGAGCTCCCACAGTTGGACAGAG
TNFRSF10D	rs1133782	F	CGCAACGAGACCCTGAGTA
		R	GACGCTGTGGCTCCTCTG
		E	CAGAGCTAACAGGTGTGACTGTAGAGT
TTN	rs12463674	F	CTGCCCACCATCAATTCTAA
		R	TGAGGTTGAACTCCCAAGAA
		E	TCCTTATCTCCTCTTAACCACTCA
VAX2	rs3771389	F	CCTGCCCAATCCTGTAAGTG
		R	CCTTCTTCCCTGGATACTCAGA
		E	CATGGTCCCAATTCAGGCAGAG
ZFP3	rs12600437	F	CCTGCCACTGCTGACAGATA
		R	AGGGGCAAAACCAAGAAGAG
		E	AAAGCACTTAAGTTCTCATGACCTGGACC

4.2.5 Gene ontology

Novel autosomal monoallelic genes detected in the differentiated cell lines were interrogated for enrichment of specific gene classes and functional terms using the DAVID Gene Functional Classification Tool (http://david.abcc.ncifcrf.goc/). The novel monoallelic genes were tested against a background list comprised of all assayed genes in the differentiated cell lines.

4.2.6 Adult brain methylation analysis

The DNA methylation status of the RME genes detected in the differentiated cells, as well as all RME loci reported in this thesis combined, was compared to the equivalent biallelic gene set in publically available adult brain DNA methylation data (Ginsberg et al. 2012). The data was retrieved from the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo; reference: GSE38608) and comprised of nine control subjects. It was collected using the Infinium HumanMethylation27 BeadChip (Illumina). The raw data was converted into M values by the authors, an alternative scoring system to the β values described in Chapter 3. An M value of 0 represents an even proportion of methylated and unmethylated DNA. A positive M value means that more molecules are methylated than unmethylated, while a negative M value means the opposite. Statistical difference between the mean M value of RME versus biallelic genes sets was tested with a two-tailed t-test in Prism (GraphPad).

4.2.7 Assessment of RME genes after reprogramming

Five genes subject to RME in the spinal cord lines, and monoallelic in SPC01, were assayed using the SNaPshot primer extension assay in the iPSC line iPSC-SPC01. My colleague Dr Graham Cocks created the cell line by transducing SPC01 with a polycistronic excisable vector containing *C-MYC*, *KLF4*, *OCT4* and *SOX-2* as described elsewhere (Papapetrou et al. 2011). He carried out the cell culture and I extracted the RNA and synthesised cDNA as described in sections 2.2.2 and 2.2.7 respectively. The SNaPshot primer extension assay was carried out as described in section 2.5, for the following genes: *C1QL3*, *GRID1*, *SFMBT2*, *TNFRSF10D* and *VAX2*. Primer sequences are presented in Table 4.2. The difference between the mean percentage representation of the silenced allele before and after reprogramming was tested for statistical significance using a two-tailed t-test in Prism (GraphPad).

4.3 Results

4.3.1 Differentiation into neurons and glia

Immunocytochemical analysis was carried out to assess the multipotency of the spinal cord NSC lines. After ten days differentiation, the KI67, NESTIN and SOX2 positive NSCs successfully differentiate into approximately 75% S100β positive astrocytes, 12% TAU positive neurons and < 1% O1 positive oligodendrocytes (see Figure 4.2 and Figure 4.3). No significant difference was observed between cell lines for any of the markers (ANOVA). This is consistent with a more detailed characterisation carried out by my colleague Dr Graham Cocks, who demonstrated that these cell lines retain their ventral spinal cord identity and give rise to electrophysiologically active neurons (paper submitted).

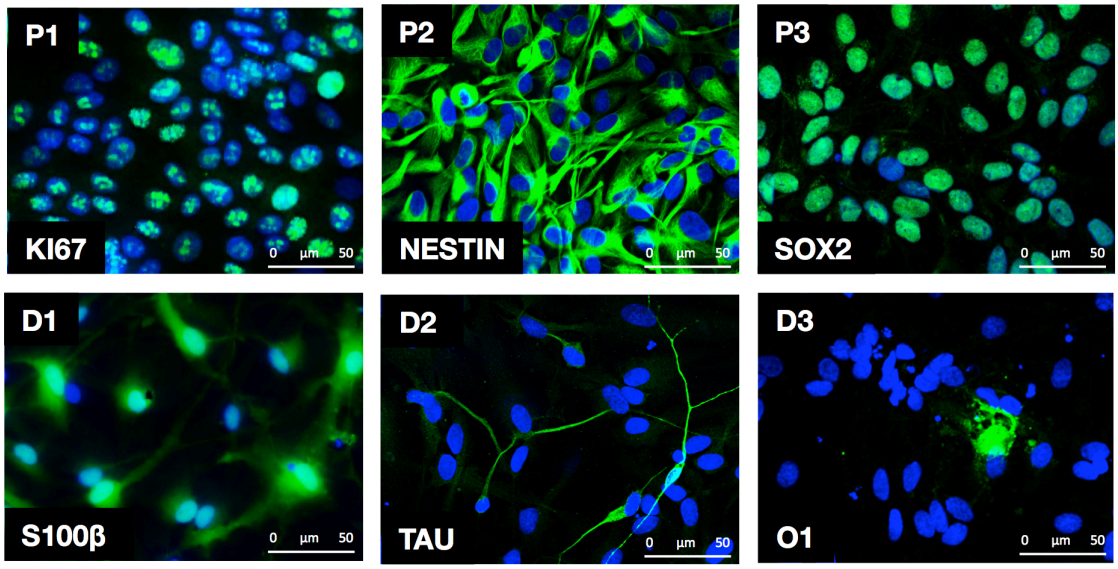


Figure 4.2 | Marker expression before and after differentiation of SPC01. Images **P1-3** show expression of KI67, NESTIN and SOX2 in the proliferating SPC01 NSC line. Images **D1-3** show expression of S100β, TAU and O1 after 10 days differentiation. Cells are nuclear stained with Hoechst 33342.

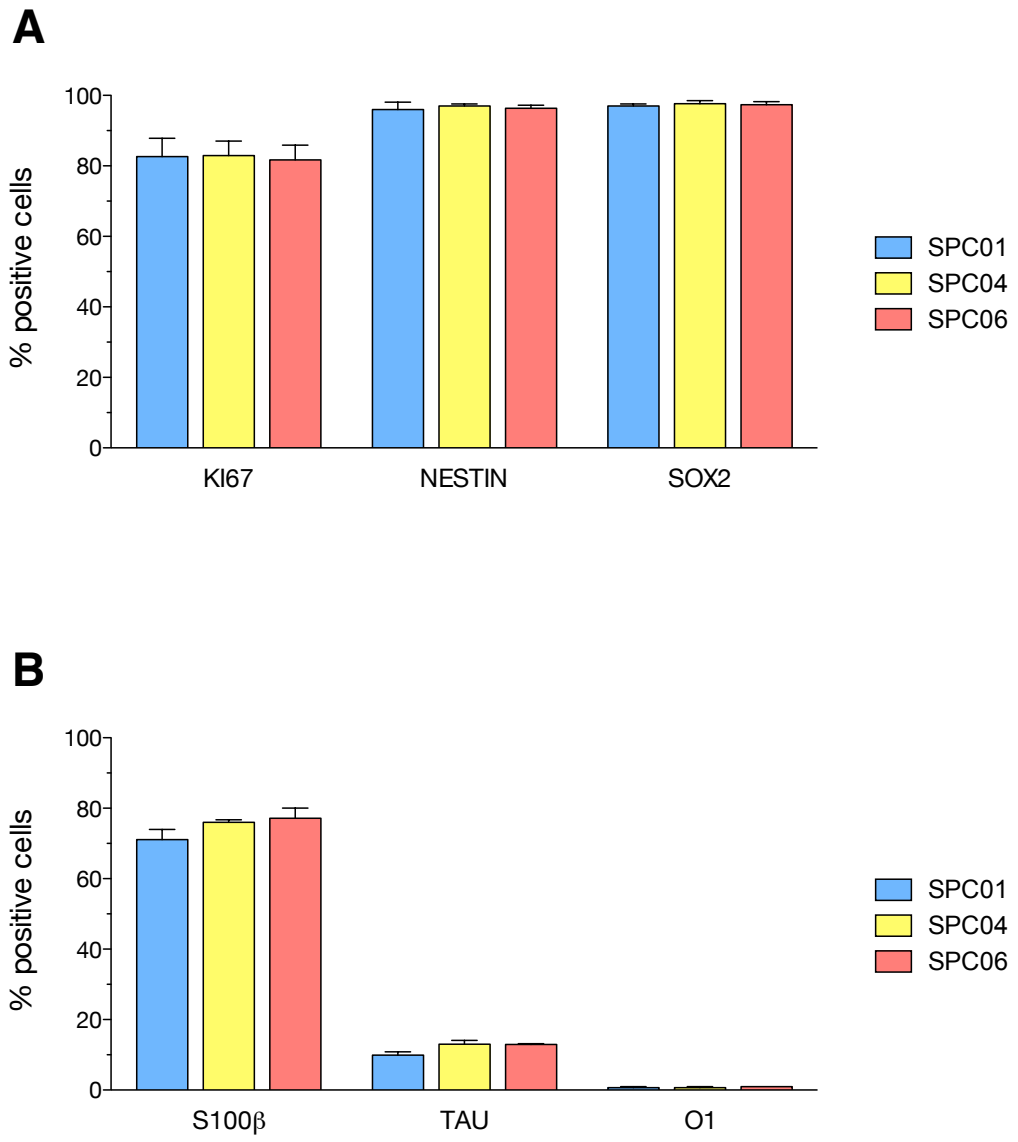


Figure 4.3 | Proportion of marker-positive cells in the three spinal cord NSC lines before (A) and after (B) differentiation. No significant difference between cell lines was observed for any of the markers (ANOVA).

4.3.2 Microarray assay validation

The OmniExpress BeadChips used in this study are different to those described in Chapter 3. To test the reproducibility of the allele-specific expression assay with these microarrays, a Pearson's correlation was carried out for assayed gene Δ_β values between replicates. *R* values ranged from 0.87 to 0.95 showing high reproducibility (see Figure 4.4).

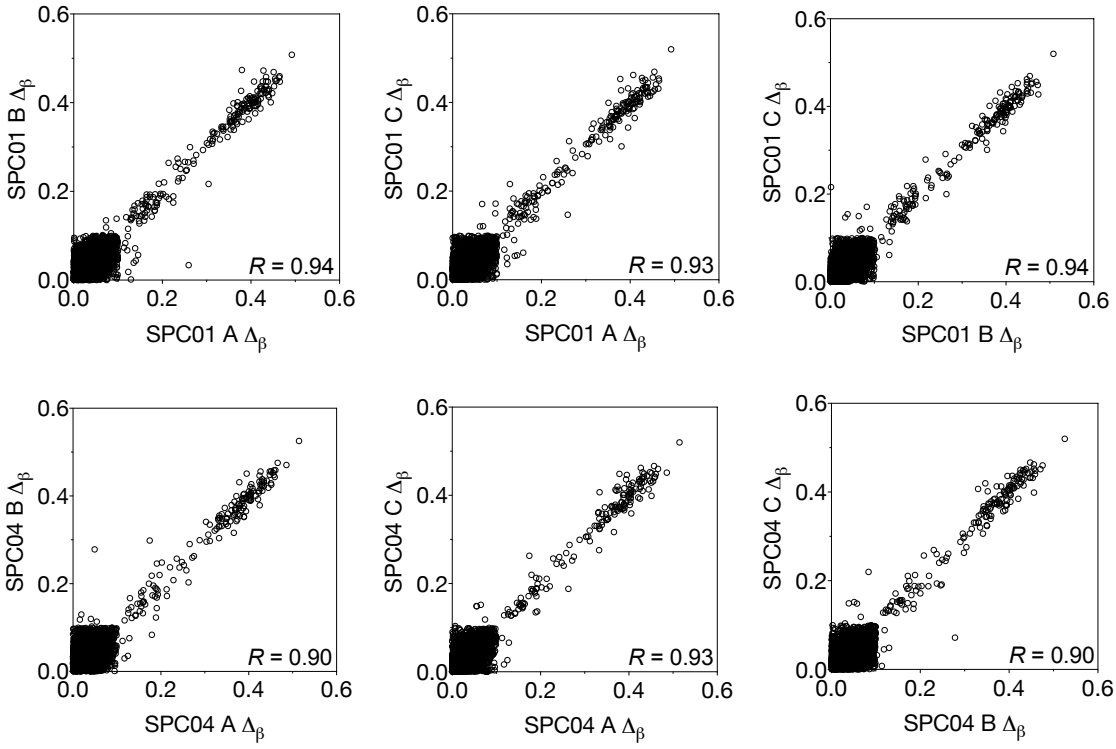


Figure 4.4 | Correlation of gene Δ_β values between biological replicates. Each pairwise comparison shows a strong, statistically significant, correlation demonstrating high reproducibility between replicates. Continued overleaf.

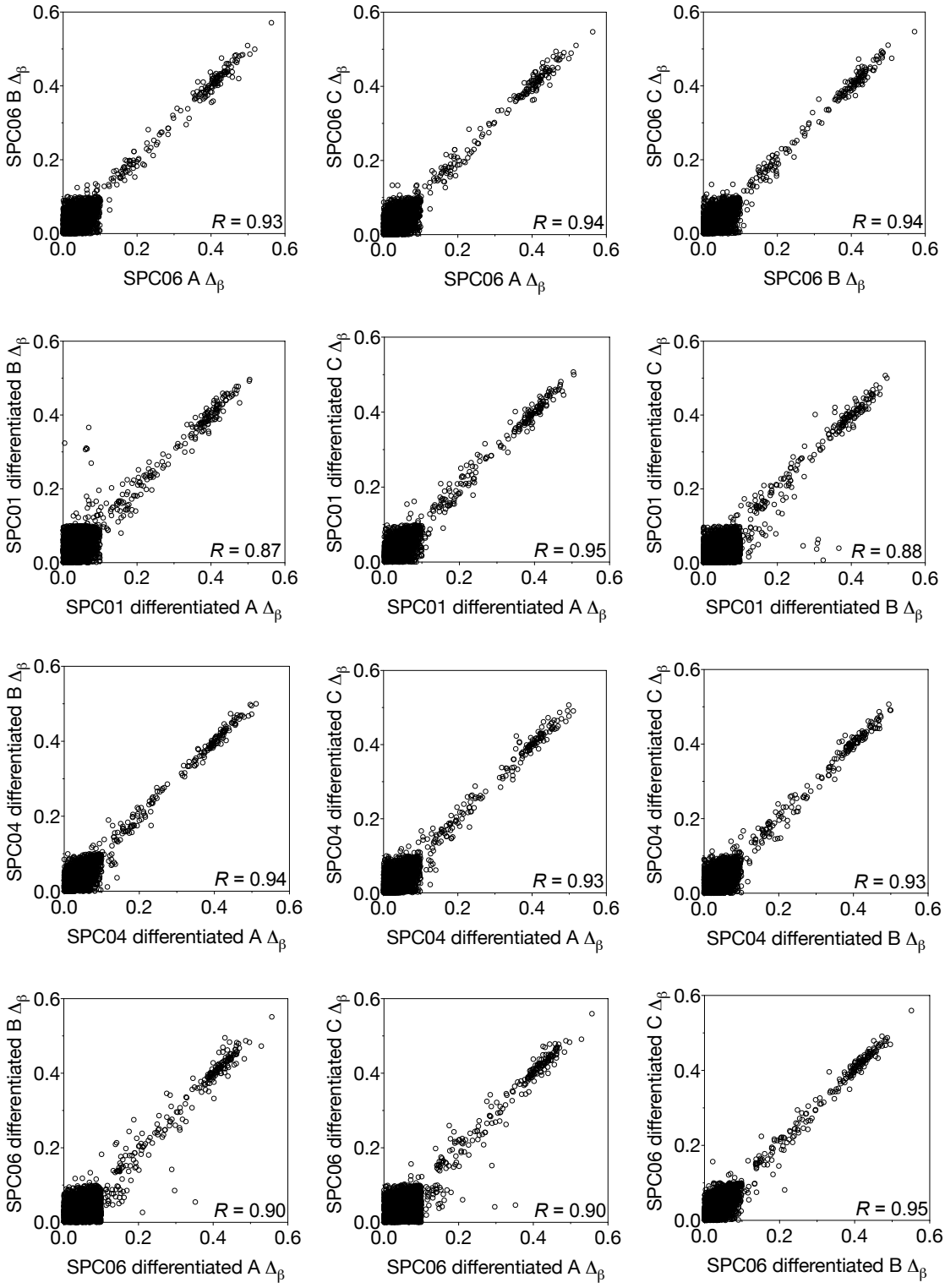


Figure 4.4 | Continued

After demonstrating biological and technical reproducibility, gene Δ_β values from the three replicates were averaged and correlated with those derived from the highly quantitative SNaPshot primer extension assay to assess the accuracy of BeadChip allele-specific expression measurements. A significant positive correlation was observed between the two datasets ($p < 0.0001$; $R = 0.91$; see Figure 4.5), demonstrating that Illumina's OmniExpress BeadChips are suitable for assays of allele-specific expression. The linear regression shows a slope of 1.3 indicating that, similar to the previous study, Δ_β values derived from the BeadChip underestimate the actual ratios of RNA derived from each allele.

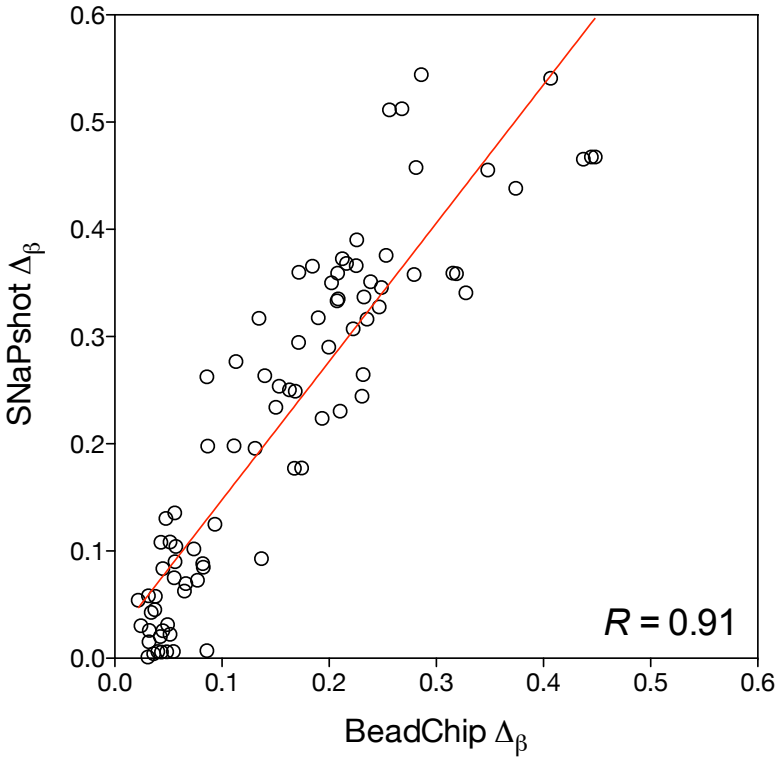


Figure 4.5 | Correlation between average BeadChip and SNaPshot Δ_β values. BeadChip Δ_β values are an average of the three biological replicates. A significant positive correlation between the two datasets is observed.

4.3.3 Monoallelic expression through neural differentiation

A similar proportion of novel autosomal monoallelic expression was detected before and after differentiation, 0.84% and 0.80% respectively. It is worth noting that this markedly is lower than the 1.4% reported in Chapter 3; this is thought to be due to technical differences between the BeadChips employed in each study (discussed in section 4.4). Average Δ_β values from genes called monoallelic in the NSCs were correlated with those assayed in the same line after differentiation (see Figure 4.6). A statistically significant positive correlation was observed ($p < 0.0001$, $R = 0.75$). Of these 42 genes, 32 (76%) retain monoallelic status. Of the 10 genes that lose monoallelic status after differentiation, nine retained biologically significant allele expression imbalance (Δ_β value > 0.1) and only one reverted to biallelic status. Of the novel monoallelic genes detected in the NSCs and assayed after differentiation, significantly more showed a reduction in Δ_β values after differentiation than would be expected by chance ($p = 0.033$, binomial exact test), indicating that, while monoallelic expression is largely maintained, there is a trend towards reduced Δ_β values.

Table 4.3 | Autosomal monoallelic expression before and after differentiation

	NSCs	Neurons and glia
Assayed genes	7,839	7,606
Monoallelic	77 (0.98%)	74 (0.97%)
Known imprinted	11 (0.14%)	13 (0.17%)
Random monoallelic	20 (0.26%)	15 (0.19%)
Same allelic choice	7 (0.09%)	10 (0.13%)
Unclassified monoallelic	39 (0.50%)	36 (0.47%)

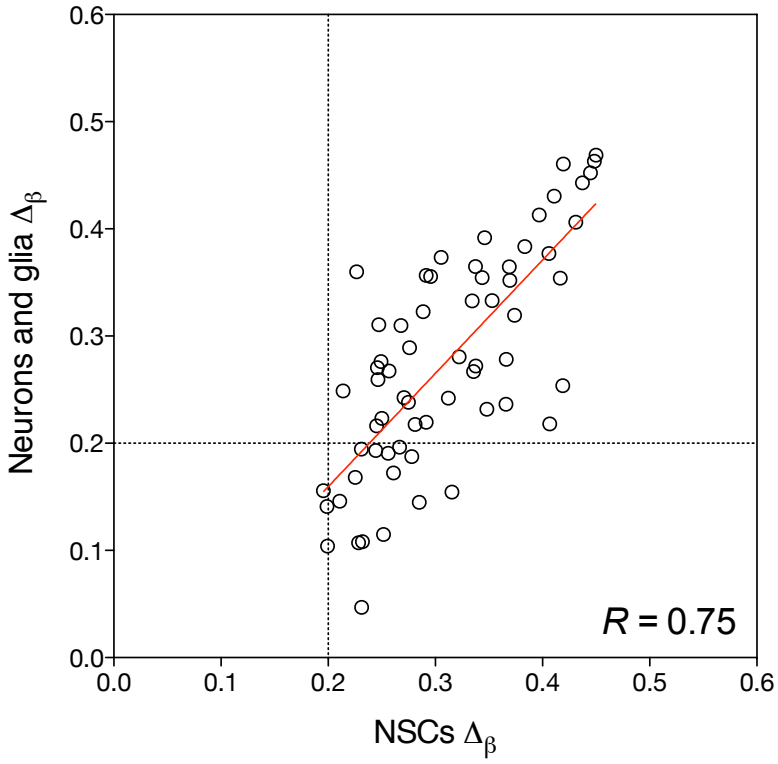


Figure 4.6 | Δ_β values from novel autosomal monoallelic genes detected in NSCs correlated before and after differentiation. Data from all three clones are presented.

Of the RME genes detected in the undifferentiated spinal cord NSCs that were also expressed after differentiation, 71% maintained monoallelic status. Only one of these genes reverts to biallelic status. A 27% reduction in the proportion of assayed genes subject to RME is observed after differentiation (0.19% compared to 0.26%). This is due to the fact that 30% of the RME genes detected in the NSCs before differentiation do not meet the expression threshold after differentiation. An additional 29 novel monoallelic genes emerge after differentiation, 18 of which were assayed before differentiation but were not classed as monoallelic. Eight of these genes are subject to RME.

4.3.4 DNA methylation of RME genes in adult brain

In Chapter 3 I describe that, when compared with biallelic loci, the novel RME genes discovered are associated with increased levels of methylation in the NSCs themselves, as well as in human foetal brain. Here, I have retrieved DNA methylation data from adult human occipital cortex to test if this association is preserved in the adult brain. I find the same relationship: RME loci show significantly higher levels of DNA methylation than biallelic loci (see Figure 4.7). This is true both when comparing all biallelic and RME loci reported in this thesis (p = 0.0003, t-test), as well as when comparing just the loci detected in the differentiated spinal cord cell lines (p = 0.0002, t-test).

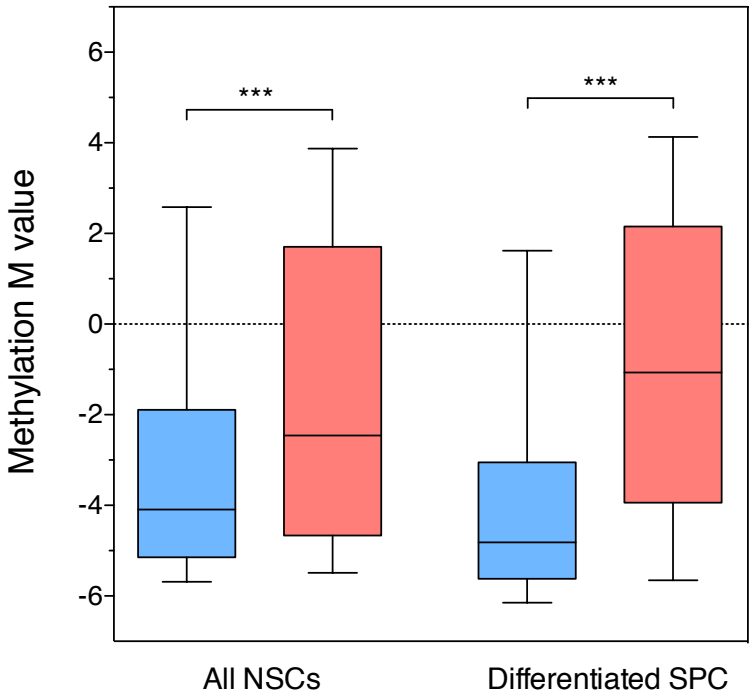


Figure 4.7 | DNA methylation status of biallelic and RME genes in adult brain. M values are grouped by biallelic (blue) and RME (red) genes. The pair on the left display data from all biallelic and novel RME genes reported in this thesis. The pair on the right display data from genes detected in the differentiated spinal cord NSC lines. Whiskers represent the 10-90th percentile. T-test p value summaries are presented.

4.3.5 Gene ontology analysis after differentiation

The autosomal monoallelic genes detected in the differentiated spinal cord NSC lines were interrogated for the enrichment of specific gene families and functional terms. Consistent with observations made in the proliferating NSCs reported in Chapter 3, I find a significant enrichment for the cellular component "extracellular region" (2.5 fold enrichment, $p = 0.009$) and keyword "signal" (2.2 fold enrichment, $p = 0.001$). Also of interest, the autosomal monoallelic genes in the differentiated spinal cord lines are enriched for the term "polymorphism" (1.3 fold enrichment, $p = 0.00042$). A summary table displaying count, percentage, fold enrichment and p value is included in the Appendix 9.8 on page 211. Note that, unlike in Chapter 3, this analysis was carried out for all novel monoallelic genes, rather than RME genes independently. This is because the number of RME genes detected after differentiation (15) is too low to carry out meaningful enrichment analysis.

4.3.6 RME after reprogramming

Using an iPSC line generated from the spinal cord NSC line SPC01, I have carried out a pilot study assessing the allelic status of five RME genes before and after reprogramming. All five genes lose monoallelic status, showing statistically significant reduction in the magnitude of allele expression imbalance (least significant $p = 0.002$, t-test; see Figure 4.8). Complete reversion to biallelic status was observed in four of the five genes while the fifth, *C1QL3*, showed a 26% shift towards biallelic expression.

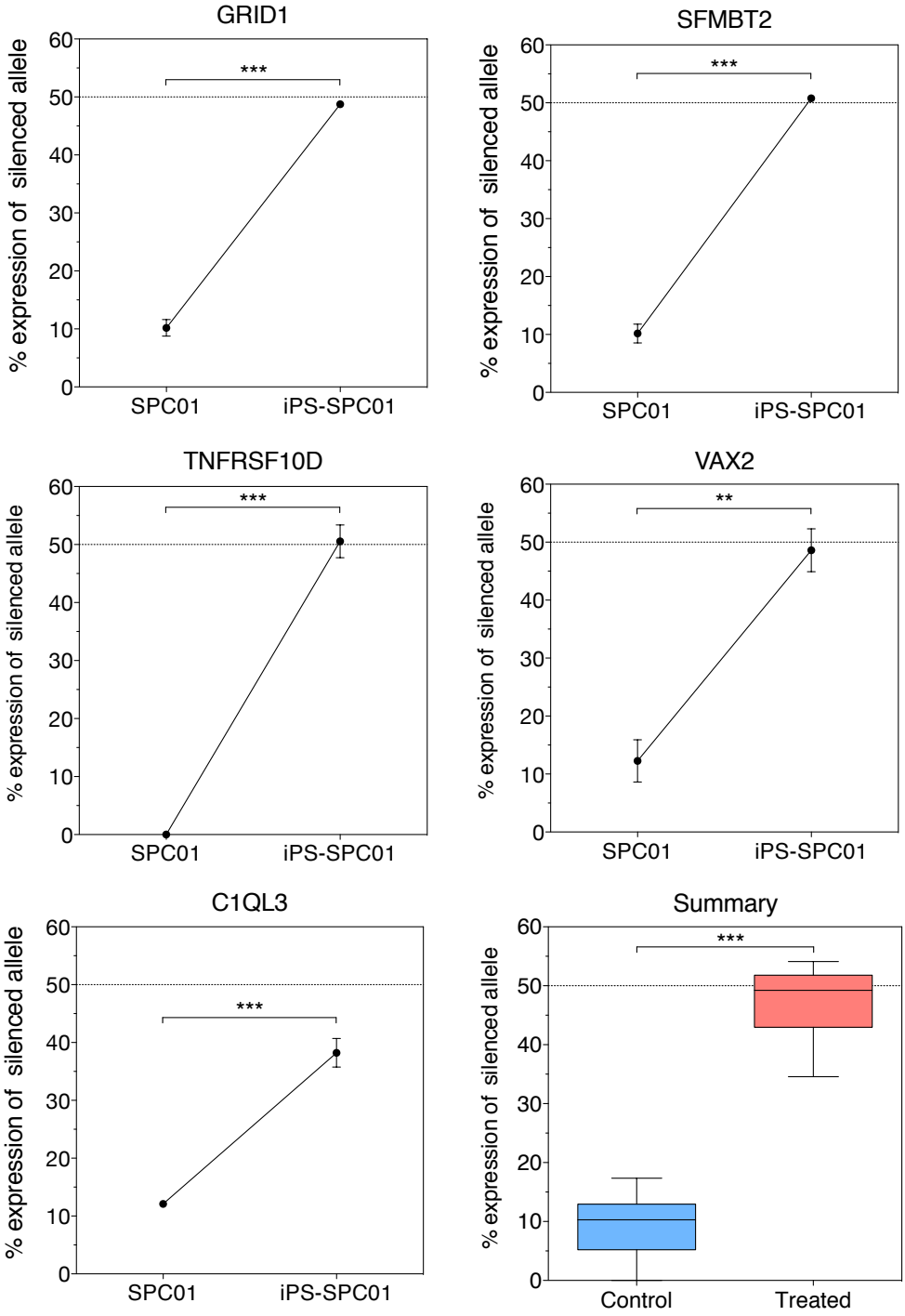


Figure 4.8 | Allelic expression of RME genes before and after reprogramming. Mean percentage expression of the silenced allele as detected by SNaPshot primer extension (n=3). T-test *p* value summaries and SEM is illustrated. The summary box plot box plot combines data from all genes assayed. Whiskers display minimum to maximum values and a paired t-test *p* value summary is presented.

4.4 Discussion

I first report immunocytochemical analysis of the spinal cord NSC lines before and after differentiation. The aim of this study was to verify that they had successfully differentiated into neurons and glia. I show that they express the generic NSC markers NESTIN and SOX2 in proliferative state, and after ten days of differentiation conditions they generate approximately 75% S100β positive astrocytes, 12% TAU positive neurons and < 1% O1 positive oligodendrocytes. No significant difference was observed between cell lines for any of the markers. This is consistent with a more in depth characterisation of these cell lines carried out by my colleague Dr Graham Cocks (paper submitted), who reports a similar fraction of neurons and astrocytes after differentiation (10% and 79% respectively). He finds that they retain their ventral specification; the neurons generated are predominantly NKX6, LHX3 and CHX10 positive V2a interneurons. He has also demonstrated that they give rise to electrophysiologically active neurons, with spontaneous Ca^{2+} oscillations and K^{+} induced depolarisation. These results indicate that the spinal cord NSC lines employed for this study show a good degree of functional maturation after differentiation.

In this chapter I report the first genome-wide assessment of allele-specific expression after human NSC differentiation. A near-equal proportion of novel autosomal monoallelic expression was detected before and after differentiation, 0.84% and 0.8% respectively. While a significant trend towards reduced Δ_{β} values was observed after differentiation, monoallelic expression is largely maintained. Of the monoallelic genes detected in the NSCs that were also assayed after differentiation, 76% remained monoallelic and 98% retained allele expression imbalance. RME genes show a similar relationship when

assessed independently; 71% assayed after differentiation retain monoallelic expression and 93% show allelic expression imbalance. However, a difference was observed in the proportion of RME genes detected before and after differentiation (0.26% compared to 0.19%). Additionally, a subset of novel monoallelic genes emerges after differentiation. Taken together, this data indicates that widespread autosomal monoallelic expression is largely maintained after NSC differentiation.

When assessing allele-specific expression in the differentiated cells, it is important to be mindful of the fact that a heterogeneous cell population is being assayed. This could result in an underestimation of the number of monoallelic genes in each cell. For example, if a gene were monoallelic in neurons and biallelic in astrocytes, it would not be detected in this study. Professor Judith Singer-Sam's research group report data for 19 non-imprinted autosomal monoallelic genes detected in mouse NSCs after differentiation into both pure neuron and astrocyte populations. Ten of these genes showed monoallelic expression in one cell type and biallelic expression in the other in at least one cell line (Wang et al. 2010; Li et al. 2012). This demonstrates that the allele-specific expression can differ dramatically between neural cell types and it may explain the trend towards reduced Δ_β values, as well as the reduced proportion of RME genes, observed after differentiation. Additionally, some monoallelic genes might only be expressed in one cell type resulting in them falling short of the gene expression threshold. A revealing follow up experiment could be conducted using fluorescence-activated cell sorting to generate pure cell populations. The genes that loose monoallelic status after differentiation could then be assayed in the pure populations. If evidence for a cell type-specific effect were observed, a genome-wide assay of allele-specific expression in these pure cell populations could be carried out.

Unexpectedly, the proportion of autosomal monoallelic genes detected in the undifferentiated spinal cord NSC lines is considerably lower than was reported in the same cell lines in the previous chapter (0.84% compared to 1.4%). This is unlikely to be a result of instability of the monoallelic genes as SNaPshot primer extension validation was carried out in biological triplicate with cells separated by multiple passages. Furthermore, correlating assayed gene Δ_β values from the two studies shows a strong correlation ($p < 0.0001$, $R = 0.72$; see Appendix 9.9 on page 212). Instead, it seems that monoallelic genes have been disproportionately omitted from the analysis; while 86% of the genes assayed using the Omni1-Quad BeadChips were assayed in this study, only 45% of the novel monoallelic genes were. Of the monoallelic genes reported in Chapter 3 that were not assayed in this study, 24% were omitted due to insufficient (less than two) SNPs and the remainder did not reach the expression threshold. The OmniExpress BeadChip has smaller hybridization beads in order to fit more samples on each chip and this may have resulted in reduced sensitivity. Indeed, the median post-normalisation cDNA SNP probe intensity value was ~800 on the Omni1-Quad BeadChips and ~300 on the OmniExpress BeadChips (see Appendix 9.2 on page 205 and Appendix 9.7 on page 210). Reduced gene dosage resulting from monoallelic expression may have biased the exclusion of monoallelic genes from this analysis. The relationship between monoallelic expression and total gene expression is explored in Chapter 5. It would have been preferable to carry out this study with the same genotyping arrays as the primary study, however, it was no longer in production. With cost rapidly reducing, RNA-seq is the ideal methodology for this analysis.

Taking these study limitations into account, it is difficult to accurately predict the proportion of autosomal genes subject to monoallelic expression in neurons or glia.

However, I can conclude that autosomal monoallelic expression remains widespread after NSC differentiation and, for the genes assayed, monoallelic expression is largely maintained. In addition, a subset of novel monoallelic genes emerge after differentiation, of which, a proportion shows evidence of RME. A more complete analysis of the state of autosomal monoallelic expression through neural differentiation could be carried out using RNA-seq, or genotyping microarrays closer matched to the Omni1-Quad BeadChips, with RNA from pure neuron and astrocyte populations.

The novel monoallelic genes detected in the differentiated cells were assessed for the enrichment of specific gene classes and functional terms. The number of RME genes detected after differentiation was too low to perform a meaningful analysis, so all novel monoallelic genes were assayed together. The results obtained were consistent with those reported in Chapter 3; a significant enrichment for the cellular component "extracellular region" and keyword "signal" was observed. This indicates that autosomal monoallelic expression is performing a similar function after differentiation and supports the hypothesis that RME is involved in cell-cell interaction.

While the results discussed above are of interest with respect to NSCs and their progeny *in vitro*, this study was also carried out with the ambition of providing clues as to the extent this process may occur *in vivo*, in the adult human CNS. NSC differentiation *in vitro* offers a powerful model that has revolutionised the understanding of neural development, however, the extrapolation of these observations to the adult CNS must be made cautiously. The detection of widespread autosomal monoallelic expression, including RME, in neurons and glia *in vitro* acts as a proof of principle.

In Chapter 3 I reported that the novel RME genes detected in this thesis are associated with increased levels of DNA methylation, both in the NSCs themselves, and in foetal brain. Using the demethylating agent 5-Aza, I also demonstrated that DNA directly regulates the monoallelic expression of all twelve RME genes assayed. In this chapter I assessed publically available DNA methylation data from adult human occipital cortex with the aim of determining whether or not this epigenetic distinction is retained into adulthood. I find the same relationship; the novel RME genes show significantly increased levels of DNA methylation in adult human cortex. While this does not demonstrate that these genes are similarly expressed in adult brain, it does demonstrate that they are epigenetically distinct from biallelic loci in a manner consistent with monoallelic expression.

Further research will be required to conclusively demonstrate that widespread monoallelic expression occurs in the developing and adult human CNS. One hurdle in the way of this study is the non-clonal nature of brain tissue; RME is expected to be masked by neighbouring cells with biallelic expression or monoallelic expression for the alternate allele. Additionally, the possibility of cell-type specific monoallelic expression would compound the issue, further reducing the power to detect RME in brain tissue. A recent publication reports a protocol for transcriptome-wide RNA-seq analysis of single cells (Tang et al. 2010). This technique should enable genome-wide assays of allele-specific expression from single cells and this would circumvent the obstacles highlighted above.

If widespread RME does occur *in vivo*, the stage at which it is initiated will dictate not only when this source of diversity is introduced, but also the size of clones that share the same repertoire of monoallelic genes. No previous study has addressed this issue. I

conducted a pilot study assessing the allele-specific expression of five RME genes in an iPSC line generated from the spinal cord NSC line SPC01. All five genes lost monoallelic status, and four of the five genes completely reverted to biallelic status. This data could be construed as evidence in support of RME being initiated between the ESC and NSC stage. However, there are two key factors that require consideration before drawing this conclusion.

Firstly, it is difficult to generalise about all RME genes when only five were assayed. In Chapter 3 I report that these genes are dispersed throughout the genome, and are therefore likely to be independently regulated by varying molecular mechanisms. This concern is being addressed; the study reported in this chapter is a pilot preceding a genome-wide analysis carried out by our laboratory. Preliminary results from the genome-wide study are consistent with this pilot; 83% of non-imprinted autosomal monoallelic genes detected in SPC01 lose their monoallelic status after reprogramming. This data taken together argues that RME is largely lost after reprogramming.

The second factor to be mindful of is how closely the iPSC line represents SPC01 cells when they were at the ESC stage. In other words, has the clock been turned back, or has a distinct cell population been generated? While iPSCs closely resemble ESCs in terms of pluripotency (Zhao et al. 2009) and gene expression profiles (Stadtfeld et al. 2010), questions remain whether their epigenetic reprograming is complete. Epigenetic memory has been observed from the somatic cell of origin (Doi et al. 2009; Kim et al. 2010) as well as the accumulation of novel aberrant epigenetic states (Doi et al. 2009; Lister et al. 2011). A global assessment of the epigenetic profile of the iPSC line used for this study has not been carried out and therefore, the extent to which it mirrors ESCs in this regard

is unknown. However, incomplete reprogramming would be expected to bias results in favour of the retention of monoallelic expression, rather than its loss. Ideally, a time course experiment would be conducted in ESCs, assessing allele-specific expression frequently throughout neural differentiation. Cells grown in parallel would be fixed for immunocytochemical analysis to identify the equivalent stage of neural development at which RME is initiated.

4.4.1 Conclusion

From the results presented in this chapter, I conclude that autosomal monoallelic expression is largely maintained after NSC differentiation, although a trend towards reduced Δ_β values is observed. A near-equal proportion of novel monoallelic genes are observed before and after differentiation, however, study limitations have precluded an accurate estimation of the proportion of non-imprinted autosomal monoallelic genes in neurons and glia. I propose follow up experiments in pure cell populations that will address these issues. The extent that widespread autosomal monoallelic expression occurs *in vivo*, in the adult human CNS, remains unknown. The observation that RME genes are epigenetically distinct from biallelic loci in the adult cortex in a manner consistent with monoallelic expression is evidence in support. Single cell analysis will be required to answer this important question. Finally, I find that RME is largely lost after reprogramming. This result should be interpreted cautiously but it indicates that RME is initiated between the ESC and NSC stage. Further work in ESCs will be required to determine the timing of RME initiation.

Chapter 5. The potential functional implications of RME

5.1 Introduction

There are two potential functional implications of monoallelic expression at the cellular level: altered transcript levels and the exposure of functional heterozygosity. If these hold true for the widespread RME reported in Chapters 3 and 4, it could provide a source of phenotypic diversity between genetically identical neural cells.

For both XCI and genomic imprinting, monoallelic expression is believed to have evolved as a means to regulate the expression of dosage-sensitive genes. XCI exists to match female X-linked gene dosage to that of males, while genomic imprinting has evolved, at least in part, as a result of a conflict between sexes over the expression of prenatal growth related genes. The importance of imprinting as a mechanism of regulating gene expression is illustrated by disorders that arise when it is disrupted; for example, biallelic expression of the growth promoting *IGF2* is sufficient to cause the congenital overgrowth disorder Beckwith Wiedemann syndrome (Reik et al. 1995). The dosage sensitivity of genes subject to RME is not well characterised. The RME of antigen receptors, odorant receptors and protocadherins exposes functional variation in order to increase cellular specificity and tissue diversity. The functional implications of widespread RME remain to be determined.

In this chapter I perform a global analysis of the relationship between allelic expression and total expression for the novel RME genes reported in Chapter 3. I also assess non-synonymous variation in RME genes with the aim of revealing the extent that widespread RME could impact phenotypic variation by exposing functional heterozygosity.

5.2 Methods

General molecular biology and cell culture methods used throughout this thesis are described in Chapter 2. Experimental details specific to this chapter are included below.

5.2.1 BeadChip gene expression estimates

Gene expression estimates were calculated from the Omni1-Quad BeadChip (Illumina) study reported in Chapter 3. The quantile normalised X and Y SNP probe intensity values from the cDNA samples were summed and normalised to that of the matching genomic DNA sample to control for any inherent SNP probe sensitivity bias (see the formula below). This method has previously been described by Wanger et al. (2010). The mean expression of all SNP probes (intronic and exonic) within a transcript was calculated for the gene expression estimate.

$$\text{Expression} = \frac{(\text{Xraw}_{\text{cDNA}} + \text{Yraw}_{\text{cDNA}})}{(\text{Xraw}_{\text{gDNA}} + \text{Yraw}_{\text{gDNA}})}$$

5.2.2 BeadChip expression estimate validation

The validity of this expression estimate to measure differential gene expression was tested by comparison with data obtained by qPCR. Sixteen genes were assayed in the three spinal cord NSC clones as described in section 2.6 with the two reference genes *HPRT1* and *RPL13A*. The genes assayed and primers used are presented in Table 5.1. Reference gene primer sequences can be found in section 2.6.2. Both BeadChip estimates and qPCR measures were used to calculate differential expression for each gene by normalising to one control clone. The data was presented on a log base 2 scale due to the non-normal distribution. The Pearson's correlation coefficient was calculated with the untransformed data in Prism (GraphPad).

5.2.3 Gene expression comparison

For the 201 RME genes reported in Chapter 3, which have both monoallelic and biallelic clones from the same genotype, BeadChip gene expression estimates were compared between biallelic and monoallelic clones from the same genotype. When a gene had two biallelic clones, or two monoallelic clones, the mean expression was used. A binomial exact test was used to test if more monoallelic clones showed reduced gene expression that would be expected by chance. A paired two-tailed t-test was used to assess the difference between the mean expression of monoallelic and biallelic clones.

5.2.4 qPCR analysis of 5-Aza treated cells

The transcript levels of twelve RME genes were measured before and after 5-Aza treatment in all three spinal cord NSC lines. RNA for this analysis was the same used in section 3.2.4 for allelic expression analysis after 5-Aza treatment. New cDNA was synthesised as described in section 2.2.7. qPCR was carried out as described in section 2.6 for the following genes: *C1QL3*, *CAT*, *DNAJC15*, *GRIA1*, *GRID1*, *JAG2*, *KIAA1324L*, *PMP2*, *SFMBT2*, *TNFRSF10D*, *TTN* and *VAX2*. Primer sequences are presented in Table 5.1. The two reference genes *HPRT1* and *RPL13A* were used. Reference gene primer sequences can be found in section 2.6.2.

5.2.5 Comparison of nsSNPs abundance

In order to compare the abundance of non-synonymous SNPs (nsSNPs) in biallelic and RME genes, the respective gene lists were uploaded to Galaxy (https://main.g2.bx.psu.edu; Goecks et al. 2010) along with chromosome coordinates. The RME gene list comprised of all 212 novel RME genes reported in Chapter 3. The biallelic gene list consisted of 975 genes called biallelic in all three genotypes. All nsSNPs (nonsense, missense and frameshift) were obtained from dbSNP build 135. Using the "join by genomic intervals" function in Galaxy, SNPs were assigned to genes. The number of nsSNP per coding base was also calculated to overcome transcript length bias. The mean nsSNP content per gene, and per coding base, was compared between biallelic and RME genes and tested for statistical significance using a two-tailed t-test in Prism (GraphPad). The same analysis was also carried out using common nsSNPs with a frequency in the population of >1%.

5.2.6 Prediction of functional effects of nsSNP

The online tool PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2; Adzhubei et al. 2010) was used to assess the impact amino acid substitution on protein function. The common nsSNPs located in biallelic and RME genes were uploaded to PolyPhen-2 and the default HumDiv classifier model was selected. The qualitative probabilistic classification of "neutral" or "deleterious" was extracted for all successfully annotated nsSNPs and a Fisher's exact test was carried out to test for association between either of the groups and outcomes using Prism (GraphPad).

Table 5.1 | Primer sequences for qPCR. Forward (F) and reverse (R) primers are presented.

Gene	Primer sequence (5'-3')	
C1QL3	F	ATCCCGGGCATCTACTTCTT
	R	GCTTTCCCGCCATCTAATTT
CAT	F	GCCTGGGACCCAATTATCTT
	R	TGTCCTGCATGCACATCG
CHL1	F	GCCTTCGGTCCCTTAATAGG
	R	GCGTAGGCACCAATAAATGA
DNAJC15	F	TCTCCTTACGTAGCAGCCAAA
	R	TGGTGGAAGATAAGACTGTGG
GABRA5	F	TCCGACAAAGCTGGAAAGAT
	R	GAAGTCCTCAAGCTGCATGG
GRIA1	F	TCCTGCAGAAAGTCCTGGAT
	R	TGAGGCGTTCTGATTCACAG
GRID1	F	CAGCCTCATGGATGAAGACA
	R	CTCTGCTCTGGCAGAAAGGT
JAG2	F	ATCGACGAGTGCCAGTCC
	R	GTGGGCAGCTACAGCGATAC
KIAA1324L	F	GGCTACATAGGGGAAGACTAAGC
	R	GCAGCAGAACAAGGCTCTTC
PAX8	F	GAGCAACAGGAGGACTCAGC
	R	TAACCACACAGGGAGTGTGC
PMP2	F	GGGGTTAGCCACCAGAAAAC
	R	CTGGCCTAGCTTGAAGGAGA
SEMA5A	F	CTTGGCCCATCTCTGGAATA
	R	CGGGTCCTCATATAGTGTCCA
SFMBT2	F	ATATTGACGGCCAAGCACTC
	R	GGCGTAGAAAGCCACTTTGA
TNFRSF10D	F	AAGTTCGTCGTCTTCATCGTC
	R	CTGGACACTCCTCCTCCTTG
TTN	F	CCACATCACAACCGTGAAAG
	R	GCATTGTCTGCTCCTTGACA
VAX2	F	AGCGGACACGTACATCCTTC
	R	CTGGGTCTCGGAGAGGTTC

5.3 Results

5.3.1 The impact of RME on transcript levels

In order to carry out a global analysis of the impact of RME on gene expression, the validity of using the BeadChip expression estimates to measure differential gene expression had to be demonstrated. A strong correlation was observed between qPCR and BeadChip differential gene expression measures ($p < 0.0001$, $R = 0.71$; see Figure 5.1).

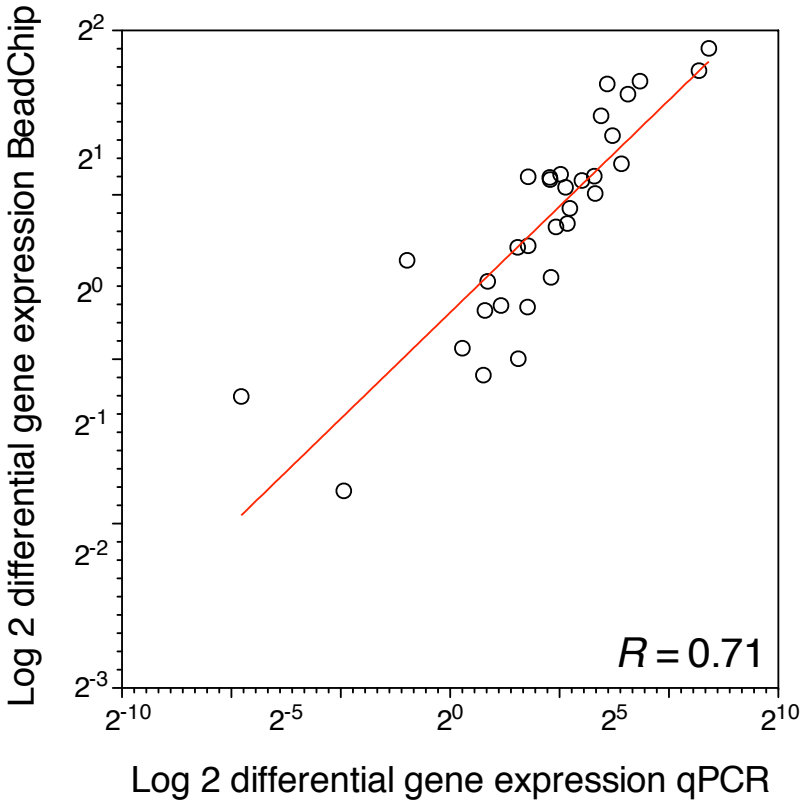


Figure 5.1 | Correlation of BeadChip and qPCR differential gene expression measures. Data is presented on a log2 scale. Pearson's correlation coefficient was calculated before log transformation. Trend line presented.

Using mean BeadChip expression estimates, the impact of RME on gene transcript levels was assessed for all 201 RME genes reported in Chapter 3, which have both monoallelic and biallelic clones of the same genotype. Figure 5.2 illustrates a trend towards reduced gene expression in the monoallelic clones. Of the 201 genes tested, 124 show reduced expression in monoallelic clones. Although a weak effect, this is significantly more than would be expected by chance (p = 0.0019, binomial exact test). In addition, the mean BeadChip expression estimate of RME genes in monoallelic clones was 15.4% lower than biallelic clones (p = 0.0002, paired two-tailed t-test). Two selected genes that illustrate this relationship are presented in Figure 5.3; both show reduced expression in monoallelic compared to biallelic clones.

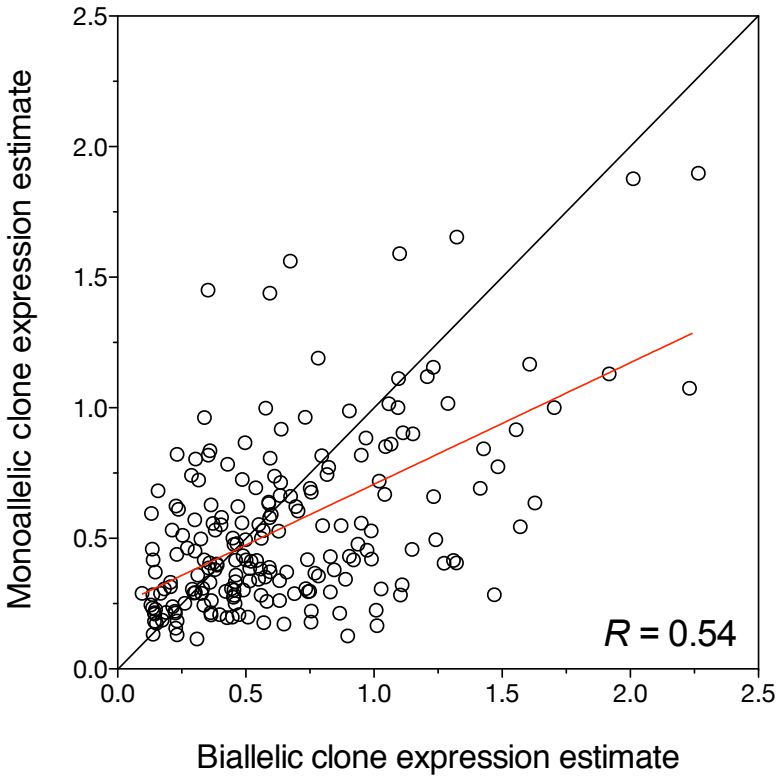


Figure 5.2 | Comparison of BeadChip gene expression estimates between monoallelic and biallelic clones of RME genes. Linear regression and Pearson's correlation coefficient presented.

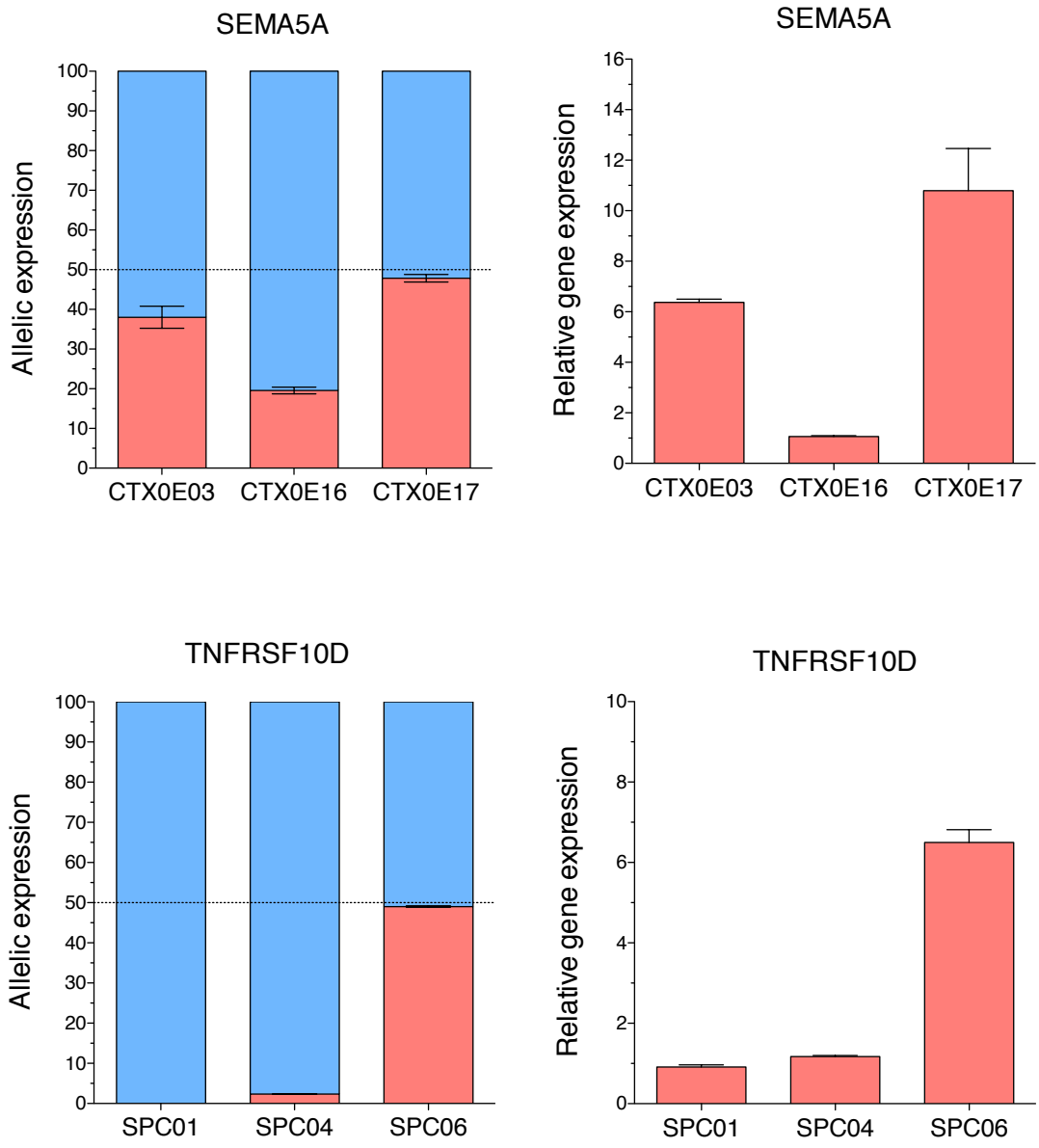


Figure 5.3 | Allele-specific expression (left) and total gene expression (right) of two selected RME genes. *SEMA5A* and *TNFRSF10D* illustrate the relationship between allelic expression and total transcript levels observed globally. Allele-specific expression data was collected using SNaPshot primer extension (see section 3.2.2) and total expression by qPCR (see section 5.2.2). Error bars represent the SEM.

In section 3.4.7 I demonstrate that all twelve RME genes assayed show a significant reduction in allele expression imbalance in monoallelic clones after 5-Aza treatment. In an attempt to further investigate the relationship between RME and total expression I carried out qPCR analysis of these genes before and after 5-Aza treatment. No correlation was found between the change in allelic expression and total expression (p = 0.18; R = 0.13; see Figure 5.4). Six of the twelve genes show a significant increase in expression after treatment and three show a significant decrease in monoallelic clones (see Figure 5.5). One, *C1QL3*, showed a significant increase in one monoallelic clone, and a significant decrease in another. Two showed no significant change. Importantly, ten of the twelve genes show a significant change in total expression in biallelic clones after treatment. Therefore, this is not isolating the impact of allelic expression on total expression.

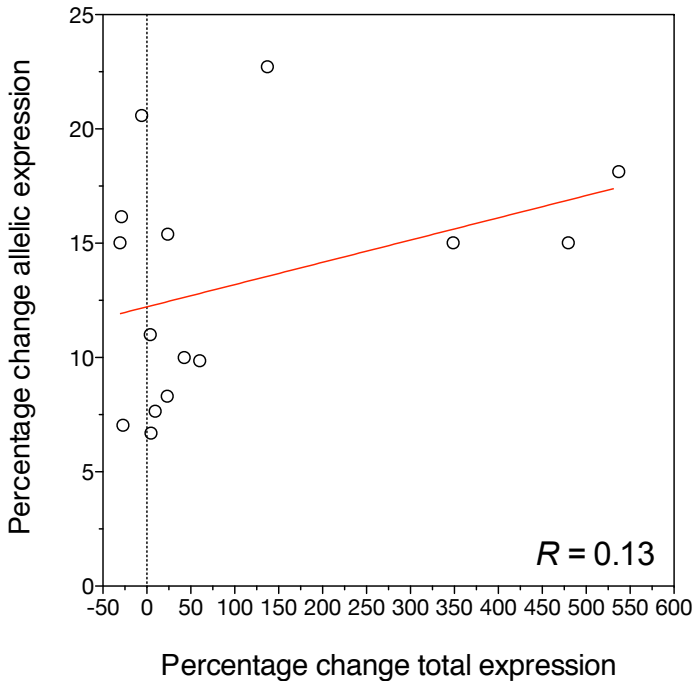


Figure 5.4 | Correlation of percentage change in allelic expression with percentage change in total expression of RME genes in monoallelic clones after 5-Aza treatment. Each data point is the mean of three biological replicates.

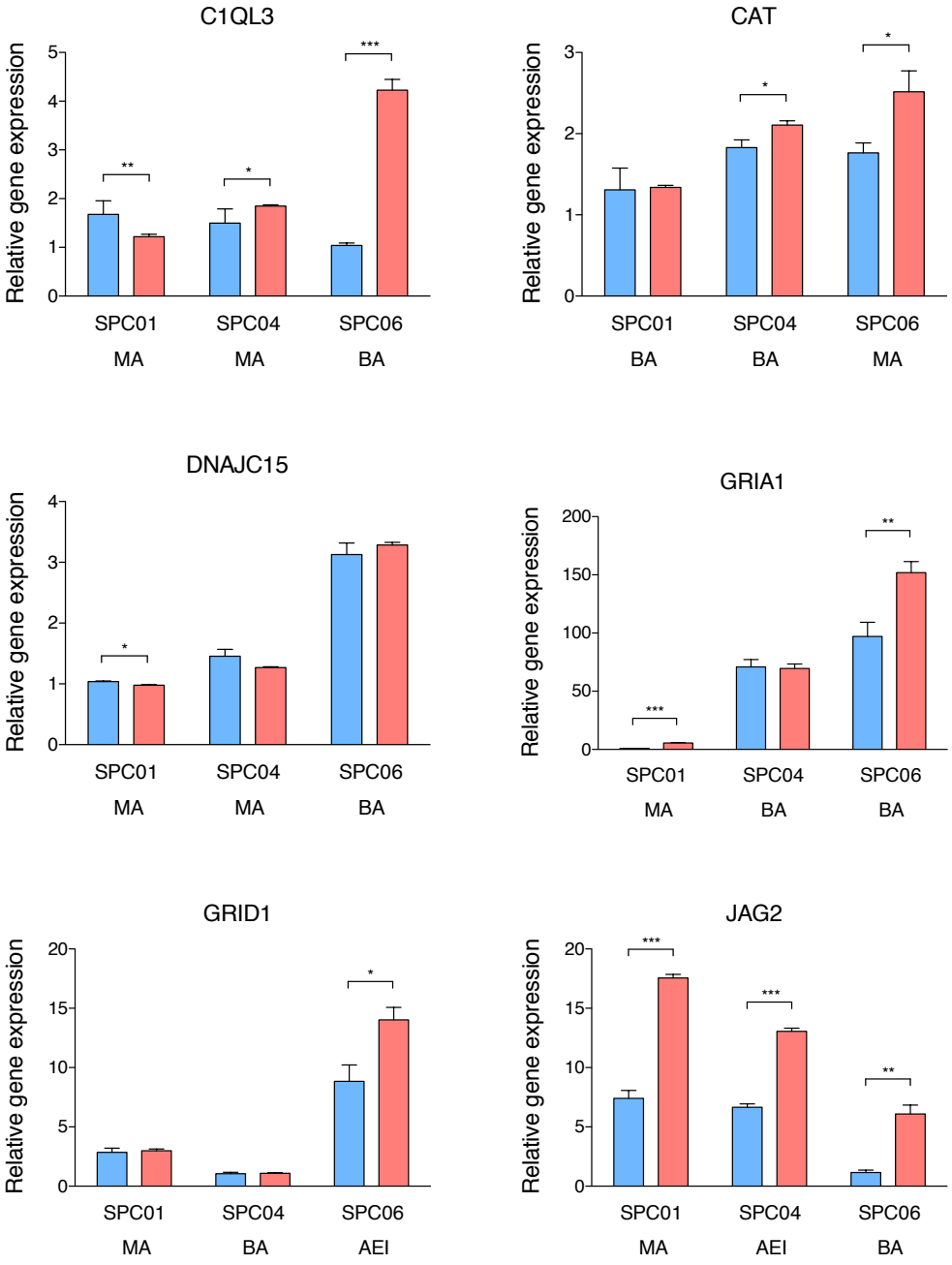


Figure 5.5 | Relative transcript levels of RME genes before (blue) and after (red) 5-Aza treatment in the three spinal cord NSC lines. The mean (n=3), SEM and t-test *p* value summary is presented. The allelic status of each clone is designated as MA (monoallelic), AEI (allelic expression imbalance) or BA (biallelic). Continued overleaf.

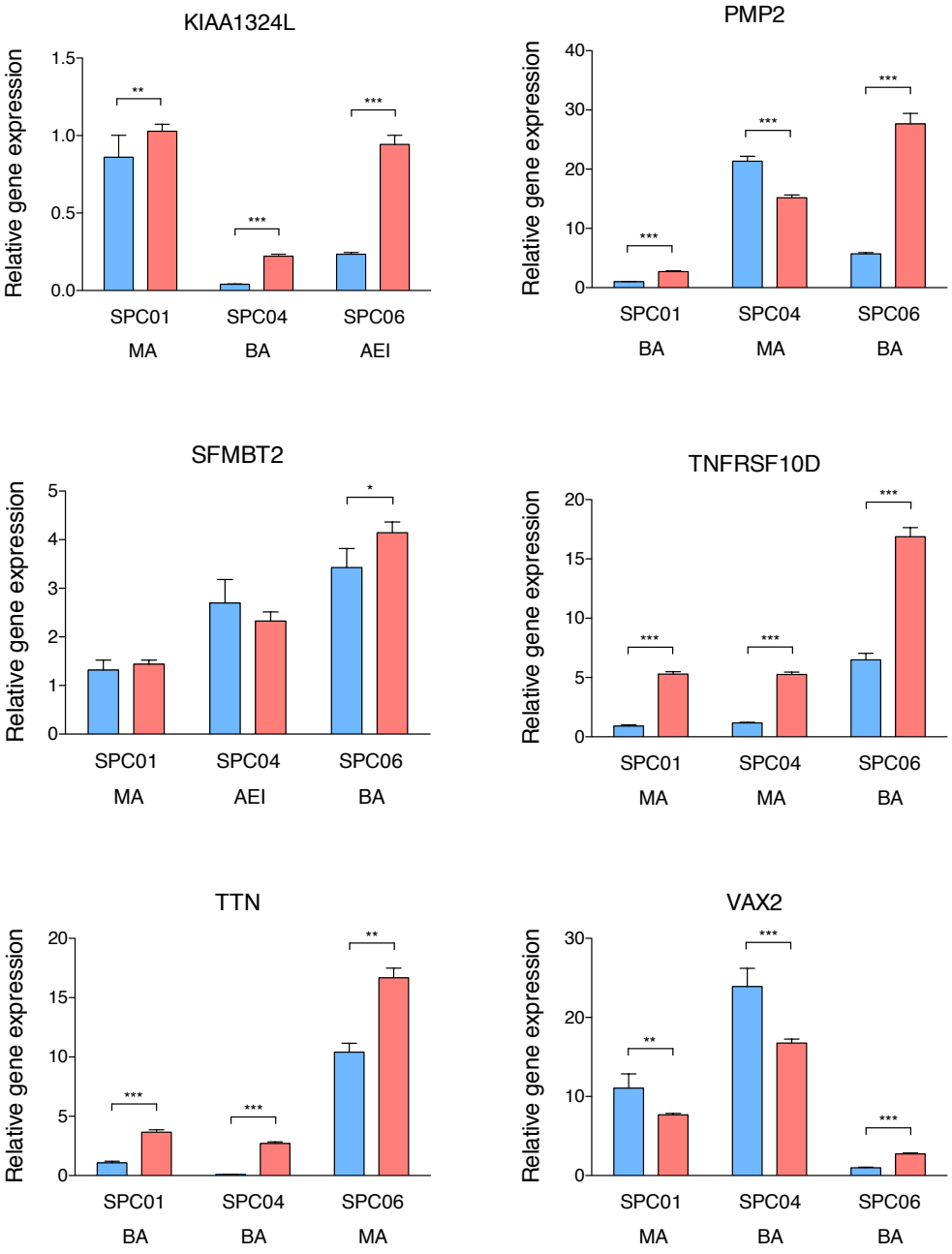


Figure 5.5 | Continued.

5.3.2 RME genes are enriched for nsSNPs

The novel RME genes reported in this study are enriched for nsSNPs when compared to biallelic genes. A mean of 3.9 common nsSNPs are found in RME genes compared to 2.3 in biallelic genes (p = 0.0014, t-test). One factor potentially biasing this comparison is coding sequence length. Indeed, the mean coding sequence of RME genes was 570bp (18%) longer than biallelic genes. To correct for this, the number of nsSNPs per coding base was calculated. Biallelic genes had a mean of 0.00091 common nsSNPs per coding base while the RME genes had 0.0014 (a 1.58 fold enrichment; see Figure 5.6). The difference between the means is highly significant ($p = 3.16 \times 10^{-9}$, t-test).

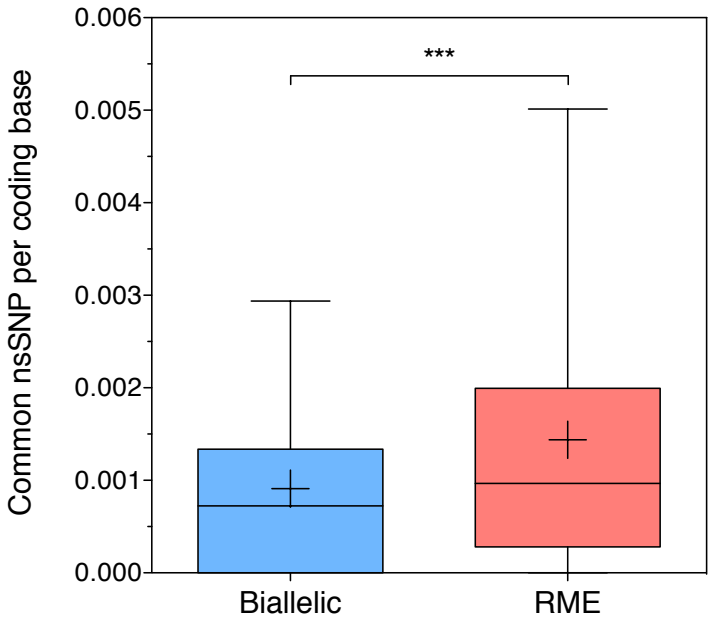


Figure 5.6 | Comparison of the number of common nsSNPs per coding base for biallelic and RME genes. Mean displayed as a plus, whiskers represent the 5-95th percentile. *P* value summary from a two-tailed t-test is presented.

A reduced, but still highly significant, 1.22 fold increase in the mean number nsSNPs per coding base was observed in RME genes when all nsSNPs in dbSNP were tested (as opposed to common nsSNPs; $p = 6.78 \times 10^{-8}$, t-test). The common nsSNPs located in biallelic and RME genes were uploaded to PolyPhen to assess the predicted impact of the amino acid substitution on protein function. Of the common nsSNPs located in RME genes, 35% were assigned "deleterious" compared to 30% of the common nsSNPs in biallelic genes. A Fisher's exact test revealed no significant association between groups and outcomes ($p = 0.45$).

5.4 Discussion

In this chapter I report a trend towards reduced transcript levels in monoallelic clones compared to biallelic clones for RME genes. Significantly more RME genes show reduced expression in monoallelic clones compared to biallelic clones than would be expected by chance, and the mean expression estimate is 15.4% lower in monoallelic than biallelic clones. This observation is consistent with previous studies of widespread RME, where individual genes were shown to be dosage sensitive (Gimelbrant 2007; Wang et al. 2010). Since the study reported in this chapter was completed, a transcriptome-wide assessment of RME in mouse NSCs using RNA-seq has been published (Li et al. 2012). The authors report a mean 35.8% reduction of expression levels in monoallelic compared to biallelic clones. Taken together, these studies indicate that widespread RME can increase cellular heterogeneity by altering gene expression levels. The extent that this relationship translates to the protein level remains to be determined.

While a trend towards reduced transcript levels in monoallelic clones is observed, the data presented in this chapter, and that reported by Li et al. (2012), indicate that this is not a general rule applying to all RME genes. 38% of genes tested do not show the expected reduction of transcript levels in monoallelic clones. This could be a result of dosage compensation; altered gene dosage will only alter the transcript level of dosage-sensitive genes. Studies of human aneuploidies have revealed that many genes have feedback mechanisms to compensate for altered gene dosage (FitzPatrick et al. 2002; Yahya-Graison et al. 2007). Therefore, it may be that a proportion of the RME genes that do not show reduced gene expression in monoallelic clones have compensatory mechanisms in place to maintain the desired transcript levels.

This study is limited in that it is not able to isolate the impact of monoallelic expression on total transcript levels; there are likely to be other, *trans*-acting, factors contributing to transcript level variation between clonal cell lines (Huang 2009). This variation is visible in Figure 5.5 where *VAX2* exemplifies the issue; one biallelic clone expresses *VAX2* at a higher level that the monoallelic clone, while the second biallelic clone shows lower expression than the monoallelic clone. Therefore, it is possible that the effect of RME on transcript levels of some genes is being masked. A more robust analysis of this relationship would require more cell lines per genotype in order to overcome interclonal variation in gene expression. Alternatively, candidate genes could be assessed independently for dosage sensitivity by genetic copy number manipulation.

I attempted to overcome the confounding factor of interclonal gene expression variation by assessing transcript levels before and after demethylation within the same cell line. It was hypothesised that the monoallelic clones would show an increase in gene expression

after 5-Aza treatment due to increased expression from the silenced allele. However, 5-Aza treatment affected transcript levels of biallelic clones as much a monoallelic clones, resulting in the same shortcoming: the impact of allele-specific expression on gene transcript levels was not being isolated. 5-Aza is a drug of broad affect, demethylating the genome indiscriminately. Therefore, it is not only acting on the target genes in *cis* to remove the allele-specific DNA methylation that enforces RME, but also in *trans* by altering the expression of upstream genes. The rational for this experiment came from a study of loss of imprinting (Diplas et al. 2009). Here, the authors report 5-Aza induced loss of imprinting of *PLAGL1* is associated with an increase in *PLAGL1* transcript levels. My results suggest that the increase in *PLAGL1* expression that they observe could be attributable in part to non-specific *trans*-acting factors rather than a result specific to the loss of imprinting. As mentioned above, further work will be required to fully characterise the dosage sensitivity of the RME genes reported in this thesis. Nonetheless, the results presented in this chapter, along with the other studies referenced, indicate that at least a proportion of RME genes are dosage sensitive.

In addition to altering gene expression levels, RME also has the potential to impact cellular phenotype by exposing functional heterozygosity. For this to occur, non-synonymous variants would have to be common in genes subject to RME. I observe that the novel RME genes reported in this thesis are significantly enriched for nsSNPs in comparison to biallelic genes; the mean number of common nsSNPs per coding base is 1.58 fold greater in RME genes. This is consistent with the enrichment of the gene ontology term "polymorphism" (1.3 fold enrichment, $p = 0.00042$) observed in novel monoallelic genes detected after differentiation of the spinal cord NSC lines in Chapter 4. The bioinformatics tool PolyPhen-2 predicted that 35% of the common nsSNPs would

alter protein function. This evidence for increased diversity of RME compared to biallelic genes supports a model of widespread RME increasing clonal diversity by exposing functional variation in a manner similar to the better established RME gene families. At present, it is not clear whether exposing functional variation is an adaptive advantage of RME. While this explanation is appealing, it is also possible that the RME genes are less critical than the biallelic genes and the enrichment of nsSNPs merely represents a reduced purifying pressure against amino acid variation.

The observation that nsSNPs are enriched in RME genes is also consistent with the hypothesis that RME could increase a gene's evolvability (Chess 2012). Chess suggests that RME may facilitate the selection of beneficial recessive alleles that would otherwise only increase the fitness of the organism when it is homozygous. As new advantageous alleles will initially be present predominantly in the heterozygous state, Chess suggests that RME could provide a mechanism to expose the new allele, increasing the fitness of the organism and reinforcing the selection of the new allele. This is especially interesting in light of the observation that the RME genes detected in human lymphoblastoid cells are more than twice as likely to be located near noncoding sequences associated with human lineage-specific accelerated evolution (Gimelbrant et al. 2007). My colleague Dr Aaron Jeffries replicated this finding in our dataset (Jeffries et al. 2012). The comparison of biallelic and RME gene ka/ks ratios could provide more information on the evolutionary rate of these genes.

5.5 Conclusion

As with all forms of monoallelic expression, functional hemizygosity will expose recessive deleterious mutations. With this apparent fitness cost, it would be predicted that RME provides a considerable benefit that has enabled its conservation. An adaptive advantage of widespread RME remains to be demonstrated. However, the data presented in this chapter supports a model in which widespread RME increases clonal diversity by reducing transcript levels and exposing non-synonymous variation. This diversification of NSCs would be predicted to have considerable implications for the developing central nervous system if present *in vivo*. The observation that RME genes are significantly enriched for common nsSNPs also lends support to the hypothesis that RME can increase a gene's evolvability.

Chapter 6. Identification of common genetic *cis* variants associated with monoallelic expression of *PM20D1*

6.1 Introduction

The genome wide assay of allele-specific expression reported in Chapter 3 revealed a subset of monoallelic genes that showed the same allele choice in all clones in more than one genotype. This pattern of allelic expression is consistent with both genomic imprinting and common *cis*-acting genetic regulatory variation. Genomic imprinting is an epigenetic process that causes genes to be expressed from a single allele in a parent of origin-specific manner (see section 1.2.1). To date, 80 human imprinted genes have been validated and many more are predicted (http://www.geneimprint.com). Microarray-based assays of allele-specific expression have been used successfully for the detection of novel imprinted genes (Pollard et al. 2008; Morcos et al. 2011). The complete characterisation of human imprinting is an important task as it promises to be informative for human disease, development and evolution. Heterozygosity for genetic *cis*-regulatory variation could also describe this pattern of monoallelic expression. While common *cis* variants typically have a subtle impact on allelic expression, extreme allelic expression imbalance, or monoallelic expression, is also observed (Yan et al. 2002). This can result from allele-specific disruption of a crucial regulatory element. Microarray based approaches have been used previously for the detection of genetic *cis* variation (Ge et al. 2009) and screens for monoallelic expression have successfully identified disease-associated genetic variants (Ben-David et al. 2011).

The present study initially sought to screen for novel imprinted genes in post-mortem adult cortex. While no novel imprinted genes were validated, I was able to map the genetic variants associated with the monoallelic expression of the *PM20D1,* a gene located in the Parkinson's disease (PD) susceptibility locus *PARK16*.

6.2 Methods

General molecular biology and cell culture methods used throughout this thesis are described in Chapter 2. Experimental details specific to this chapter are included below.

6.2.1 Samples

Post-mortem brain samples were of unrelated Caucasian European subjects obtained from the MRC London Neurodegenerative Diseases Brain Bank. Tissue was from the dorsolateral prefrontal cortex (DLPFC; Brodmann's area BA9). All individuals were free from psychiatric or neurological diagnosis at the time of death. Samples were selected based on heterozygosity for allele-specific expression tag SNPs. Blood samples were from the Twins Early Development Study (TEDS), a longitudinal study of behavioural development in twins. Twelve parent-child trios were selected from 32 when the genotypes of the allele-specific expression tag SNPs enabled the parent of origin of each allele to be determined. The children were healthy 12-years olds.

6.2.2 Total RNA extraction from *post-mortem* human brain

Frozen brain samples were thawed at room temperature and suspended in 10 µl of TRI Reagent® (Ambion) per 1 mg of tissue. The tissue was homogenised using FastPrep®-24 Instrument (MP Biomedicals) for 45 seconds at a speed setting of five. The sample was incubated for five minutes at room temperature and transferred to a new 1.5 ml tube. 0.1 volume of 1-Bromo-3-chloropropane (Sigma-Aldrich) was added to the homogenate, which was then vortexed for five seconds and incubated at room temperature for ten minutes. The sample was then transferred to a Phase Lock Gel Heavy 2 ml tube (5PRIME) and centrifuged at 12,000 g for 15 minutes at 4°C to separate the aqueous and the phenol-chloroform phase. RNA was precipitated, washed and resuspended as described in section 2.2.2. The yield and quality of the RNA was checked as described in section 2.2.3. RNA was DNase treated and cDNA was synthesised as described in sections 2.2.5 and 2.2.7 respectively.

6.2.3 Total RNA extraction from whole blood

Blood samples were collected in PAXgene Blood RNA Tubes (PreAnalytiX) as part of the Twins Early Development Study (TEDS) using standard protocols. RNA was extracted using the PAXgene Blood RNA Kit (PreAnalytiX). The manufacturer's recommended protocol was followed. In brief, the blood was pelleted, washed, and then resuspended with a lysis buffer and proteinase K to digest cellular proteins. The sample was then passed through the PAXgene Shredder spin column to homogenise the lysate and filter out cell debris. The supernatant of the flow-through was then passed through the PAXgene RNA spin column where the silica membrane selectively binds to the RNA.

After several wash steps the RNA is eluted and heat-denatured. The yield and quality of the RNA was assessed as described in section 2.2.3. RNA was DNase treated and cDNA was synthesised as described in sections 2.2.5 and 2.2.7 respectively.

6.2.4 Genotyping

Post-mortem brain and whole blood parent-child trio samples were genotyped by single base primer extension. In brief, PCR primers were designed to amplify approximately 200 bp of genomic DNA containing the target SNP as described in section 2.3.1. PCR amplification, enzymatic cleanup, primer extension and electrophoresis were carried out for each sample as described in section 2.5. Primer sequences for the purpose of allelic expression analysis are presented in Table 6.1 and primers designed to map the *PM20D1* regulatory variant are presented in Table 6.2. Genotypes were called by GeneMarker® 2.2.0 (SoftGenetics LLC®). Allele frequencies were compared to the HapMaP CEU database as a quality control. Genomic DNA had previously been extracted from the brain and blood donors by Dr Nicholas Bray and Dr Rebecca Smith respectively using standard phenol chloroform techniques.

6.2.5 SNaPshot allele-specific expression assay

Post-mortem brain and whole blood parent-child trio samples were assayed for allele-specific expression by single base primer extension using SNaPshot chemistry (Applied Biosystems). This assay was carried out as described in section 2.5 with a minor alteration. The nature of these samples precluded the use of biological replicates. Therefore, each sample was assayed with four technical replicates from two separate

reverse transcription reactions. Genes assayed, tag SNPs and primer sequences are presented in Table 6.1.

6.2.6 Bioinformatics

6.2.6.1 eQTL analysis

The influence of common SNPs on transcript levels of *THNSL2* and *PM20D1* in human brain was investigated using the SNPExpress eQTL database (Heinzen et al. 2008). The two genes were tested for *cis*-acting effects, where all common SNPs within 100 kb up- and down-stream of the target are tested for association with transcript levels. The database consists of genotype and expression data from 93 human brain samples. All settings were left as default.

6.2.6.2 Linkage disequilibrium analysis

Two bioinformatics tools were used to assess the linkage disequilibrium of SNPs of interest. CEU population genotype data from the International HapMap Project was analysed with HaploView version 4.2 (Barrett et al. 2005) and the Tagger function was used to aid tag SNP selection with an r^2 threshold of one (de Bakker et al. 2005). The web-based tool SNAP (Johnson et al. 2008) was used to determine whether rs708727 is in linkage disequilibrium with any SNPs from the 1000 Genomes Project.

Table 6.1 | Primer sequences for allele-specific expression analysis. Forward (F), reverse (R) and extension (E) sequences are presented.

Gene	Tag SNP	Primer sequence (5'-3')	
CD163L1	rs7306824	F	GGGCATAATGCAAGGTGAGT
		R	CTCTTCATGCTGCCAATGTG
		E	AGAAAGAGATATGGCATCTCCACA
CLUL1	rs8093432	F	GAAAAAGTGGTTCACAATCACC
		R	TTTTCTAGGAGGAGTCATGAAGAA
		E	CAAAATAAATGAAAGCTGCCCTCTCCT
FTCD	rs1980983	F	CTCTGGGAACCAAGCTTCTG
		R	CCCTGAGGGAACTACCTGCT
		E	CTGCTCCTCCACCAAGCCTGATGC
OGFR	rs1048802	F	CTCCTGGCCTGGCTGTGT
		R	AAGTCAAATGAATTTATTCAGAAAAGG
		E	TTATTCAGAAAAGGCCTTGCTTGG
PKDREJ	rs4508712	F	TTCCATTCATGCTGACCAAA
		R	TGACAATTTTGAAGACCCTCAG
		E	CCATCCAGGCTGCCCTCCC
PM20D1	rs1104899	F	ATCACAGAGTTCTTGTCGTCCA
		R	GAGGTGCCCCCATTCTCT
		E	AGAGTTCTTGTCGTCCAGTGTGCCCC
RPL22L1	rs13063927	F	GAGCAGGCTGGGTGTTTTTA
		R	GCAATACCTCAGTGCAGCAA
		E	TGGGAACAATGGGGCAGAG
SELP	rs6131	F	AGTGTCAGCACCTGGAAGC
		R	CAAGCCCCTCACTCTGTAGC
		E	TGAACACAGTCCATGGTTCCTTCA
THNSL2	rs4129190	F	TTCTCTGGCTATGCACCTGA
		R	GCTCAGAGCCAATGAAGAGG
		E	AAGAGCTCCCACAGTTGGACAGAG
TXLNB	rs10499208	F	GGCAGAAAACGTGTTGGTTT
		R	CCCAGCCCTCATCTCAATTA
		E	CTGATATTGAGACCCAGTGTAGATTT

Table 6.2 | Primer sequences for genotyping candidate *PM20D1* *cis*-regulatory variants. Forward (F), reverse (R) and extension (E) sequences are presented.

SNP	Primer sequence (5'-3')	
rs823114	F	TGATAGGCTCCACCTTCACC
	R	TGATCAAGGCAGACGAGATG
	E	CTAAGCAGCGACCATTTTTGTTTTGCGGC
rs708723	F	CCATGTGCATTTCTGCATCT
	R	AAACGCAGGTGCTGATTTCT
	E	GGCACTAACTGGCACTAATGTGAG
rs947211	F	GGTTGTCACATTTGCCTCCT
	R	GGGAACCACACAACCAGTTC
	E	TGAAAACTAAAGAGAAAATTTTGCTTCA
rs1772159	F	CCTACATCCAAATCCCCAGA
	R	CCCCATTGCTCAACCAGTAT
	E	TGGTAAGGAGAGGTGGAGGGAAGGGAAT
rs823154	F	GCCCTAGGAAGGGAATGGTA
	R	GCCTGAATTGGAAAGGATGA
	E	CCTTTACAGGGGATCAAGTGTCA
rs708727	F	GGACTCCAAGAAGCCACACT
	R	CCCACTTGCTCACCAGGTAT
	E	GCAGCAATGGGTGTGGCCAC
rs11240572	F	AAAAACCTTGGAGCCATCCT
	R	AGGTAGTGCATGTGAAGGTCTG
	E	ACATACCACATCCTATTGAACCCA
rs960603	F	GGAAGAACAGAAGGGGAAGC
	R	TCTTCTGACTCCAGCCCTTG
	E	CTCTTTATTTACACCTGTTGTATTAATCA
rs11240574	F	GTGACTTGCCCAGGGTCTAA
	R	GCAGTGGGTCCATCACCTAT
	E	TTTTCAGTTCTTTTTCCATCACACGTATG

6.2.6.3 Haplotype prediction

In order to phase the *PM20D1* tag SNP rs1104899 and rs708727, 92 individuals were genotyped for these two SNPs and eight more spanning approximately 100 kb (the SNPs assayed are presented in Table 6.2). The subjects were obtained from the MRC London Neurodegenerative Diseases Brain Bank and included those assayed for allelic expression of *PM20D1*. They were all free from psychiatric or neurological diagnosis at the time of death. Dr Nicholas Bray extracted the genomic DNA from these samples using standard phenol chloroform techniques. Genotyping was carried out as described in section 6.2.4. The primers sequences are presented in Table 6.2. Haplotypes were reconstructed from the genotype data using the program PHASE 2.02 (Stephens et al. 2001; Stephens & Donnelly 2003). All settings were set as default. The phase probability of the two SNPs was calculated from the x.out_pairs output file.

6.2.6.4 Transcription factor binding site prediction

Five SNPs (rs708727, rs823074, rs823075, rs9438393 and rs823080) were investigated for allele-specific modulation of transcription factor binding using three web-based predictive bioinformatics packages TFsearch (http://www.cbrc.jp/research/db/-TFSEARCH.html), MatInspector (Quandt et al. 1995) and AliBaba 2.1 (http://www.gene-regulation.com/pub/programs.html). All settings were left as default.

6.3 Results

6.3.1 Screen for novel imprinted genes

The genome-wide assay of allele-specific expression described in Chapter 3 detected 22 genes with a pattern of monoallelic expression consistent with genomic imprinting: when all clonal lines from a donor showed the same monoallelic choice, and this was observed in at least two different donors. Of these 22 genes, twelve were known imprinted genes. The remaining ten were assayed in adult post-mortem brain. Sixteen individuals were first genotyped for the expressed SNPs. Heterozygotes, for which tissue was available, were then assayed for allelic expression. Monoallelic expression in a manner consistent with genomic imprinting was detected in two genes: *PM20D1* and *THNSL2* (see Figure 6.1). Monoallelic expression of *PKDREJ* was detected in one individual and *RPL22L1* showed allelic imbalance in all three individuals assayed.

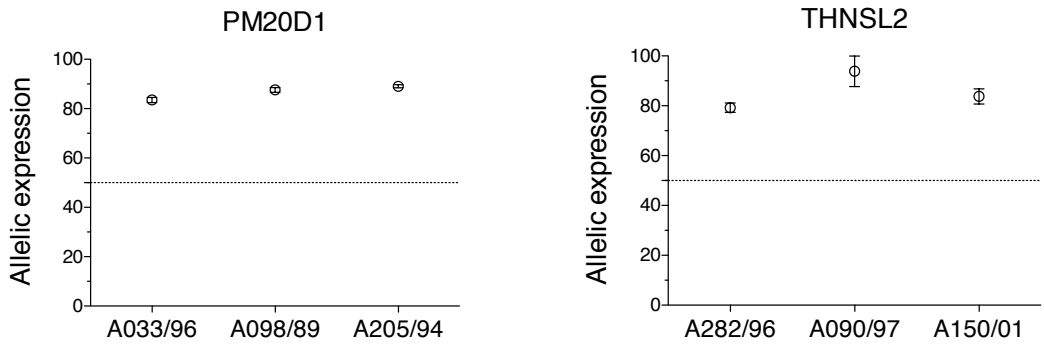


Figure 6.1 | Allelic expression of ten candidate imprinted genes in adult human cortex. Data points represent the mean percentage representation of the major allele (n=4) for each individual and error bars illustrate the SEM. Continued overleaf.

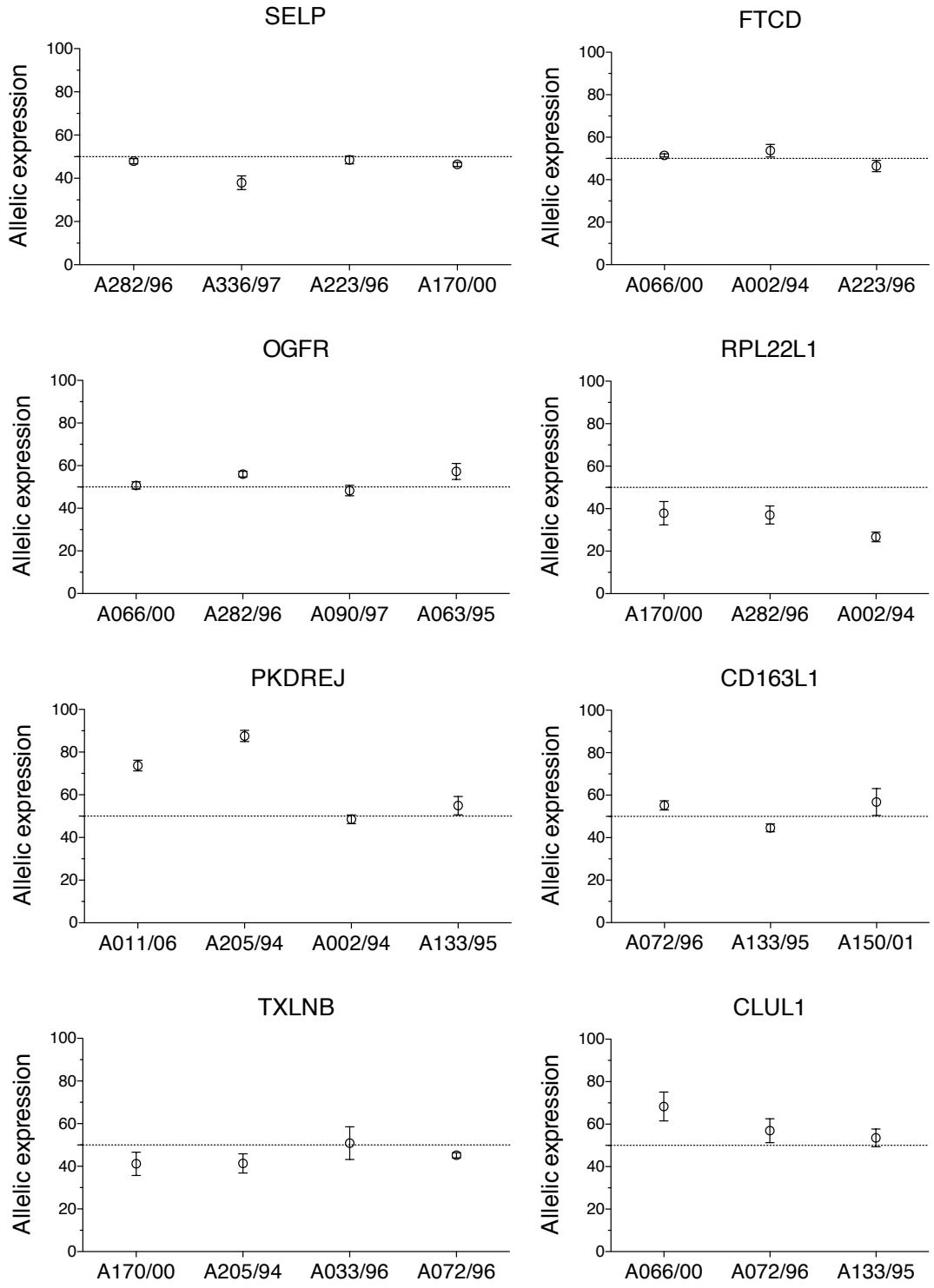


Figure 6.1 | Continued.

Genomic imprinting is parent of origin dependent. In order to determine whether or not *PM20D1* and *THNSL2* are classically imprinted, further work was required to trace the parent of origin of the silenced alleles. This information was not available for the brain samples so the two genes were assayed in whole blood parent-child trios. 32 trios were genotyped and seven were informative for each gene. *PM20D1* was not monoallelic in any of the individuals assayed in whole blood and therefore it was not possible to determine whether or not it is imprinted (see Figure 6.2). Monoallelic expression of *THNSL2* was observed in all individuals assayed; however, no parent of origin effect was observed (see Figure 6.3). The SNPExpress database was interrogated for *cis*-acting genetic variants to describe the monoallelic expression of *THNSL2*. Sixteen SNPs were significantly associated with altered transcript levels in human brain (see Table 6.3). rs4359651 and rs13019346 were called heterozygous by the genotyping arrays in both monoallelic NSC genotypes, making them good candidates for further study.

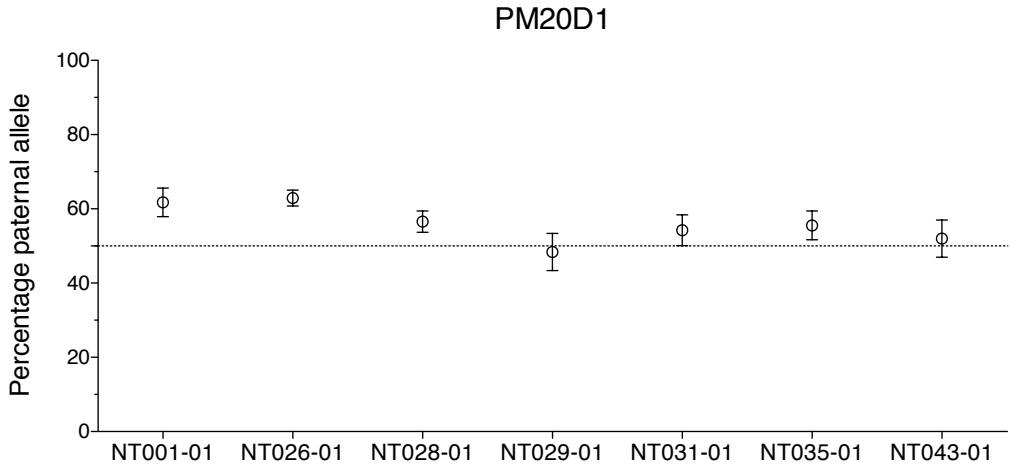


Figure 6.2 | Allelic expression of *PM20D1* in whole blood. Data points represent the mean percentage representation of the paternal allele (n=4) for each individual and error bars illustrate the SEM.

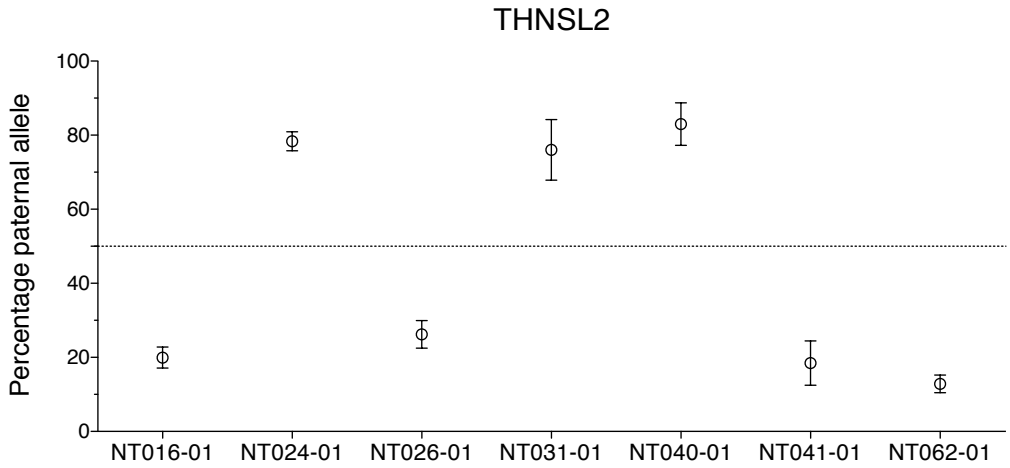


Figure 6.3 | Allelic expression of *THNSL2* in whole blood. Data points represent the mean percentage representation of the paternal allele (n=4) for each individual and error bars illustrate the SEM.

Table 6.3 | SNPs influencing the expression of *THNSL2* in human brain. P values from SNPExpress and genotypes of the monoallelic spinal cord (SPC) and striatal (STR) NSC donors are included.

SNP ID	P value	SPC	STR
rs988731	0.005926	BB	BB
rs7581571	0.01628	-	-
rs4386315	0.001242	-	-
rs10185660	0.009162	AA	AA
rs4359651	4.22×10^{-10}	AB	AB
rs4246599	1.99×10^{-5}	BB	AB
rs4129190	5.59×10^{-5}	AA	AB
rs2139100	1.19×10^{-4}	-	-
rs2176569	2.07×10^{-4}	AB	AA
rs3791357	0.02674	-	-
rs2292869	0.01712	BB	BB
rs9973903	2.07×10^{-5}	-	-
rs13019346	0.001996	AB	AB
rs2363731	0.007068	AB	BB
rs6724281	0.003975	-	-
rs4355112	0.001243	BB	BB

6.3.2 Further investigation of *PM20D1*

PM20D1 is located in the PD susceptibility locus *PARK16* and a SNP within *PM20D1*, rs11240572, has previously shown genome-wide association with PD (Satake et al. 2009). Therefore, determining the cause of monoallelic expression was of particular interest. A further twenty individuals were assayed in post-mortem cortex. A varied pattern of allelic expression was observed; eleven subjects display monoallelic expression and twelve show equal representation of the two alleles (see Figure 6.4). This was considered consistent with a causative *cis*-acting genetic variant. To test this hypothesis, the SNPExpress database was interrogated for SNPs that influence the expression of *PM20D1*. Seven showed a significant association with *PM20D1* transcript levels in human brain (see Table 6.4), and one of these, rs960603, was heterozygous in both monoallelic NSC genotypes.

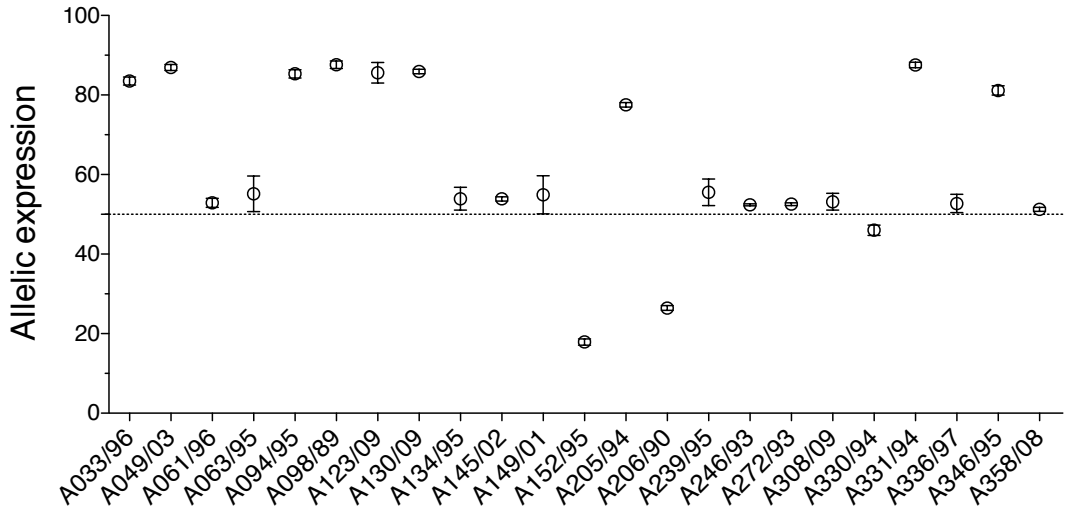


Figure 6.4 | Allelic expression of *PM20D1* in adult human cortex. Data points represent the mean percentage representation of the major allele (n=4) for each individual and error bars illustrate the SEM.

The 23 individuals assayed for *PM20D1* allelic expression were genotyped for rs960603 as well as two *PARK16* SNPs that showed genome-wide association with PD in the discovery genome-wide association study (Satake et al. 2009). rs11240572 is located in an intron in *PM20D1* and rs947211, located in *NUCKS1*, showed the strongest association with PD from the *PARK16* locus. None of the three SNPs defined the monoallelic expression of *PM20D1* (see Figure 6.5).

The NSC genotype microarray data was then sorted for all SNPs that were heterozygous in both monoallelic NSC donors within 100 kb of *PM20D1*. Six additional SNPs met these criteria. The Tagger function in HaploView was used to assess the linkage disequilibrium of these SNPs in the CEU population HapMap data. None met the r^2 threshold of one; so all six were genotyped in the 23 individuals assayed for allelic expression. Genotypes were grouped by homozygotes and heterozygotes for each SNP (see Figure 6.6). The SNP rs708727 showed a strong association with all homozygotes biallelic, and all heterozygotes monoallelic ($p < 0.0001$, Fisher's exact test). The SNPExpress eQTL database reports a trend between rs708727 genotype and *PM20D1* transcript levels ($p = 0.07978$).

Two monoallelic individuals (A152/95 and A206/90) preferentially express the alternate allele of the expressed *PM20D1* SNP rs1104899. To test whether this was a result of a phase inversion between the functional and expressed SNPs, the genotype data from ten SNPs across this region (including rs1104899 and rs708727) for 92 subjects (including those assayed for *PM20D1* allelic expression) was processed with the software package PHASE 2.02. The individuals A152/95 and A206/90 were the only two monoallelic subjects predicted to have a phase inversion (see Table 6.5).

Table 6.4 | SNPs influencing the expression of *PM20D1* in human brain. P values from SNPExpress and genotypes of the monoallelic cortical (CTX) and striatal (STR) donors are included.

SNP ID	P value	CTX	STR
rs823128	0.02148	AA	AA
rs823122	0.03389	AA	AB
rs823066	0.004208	BB	BB
rs708730	0.0304	AA	AB
rs823088	0.006034	AB	AA
rs11240572	0.04442	-	-
rs960603	0.03218	AB	AB

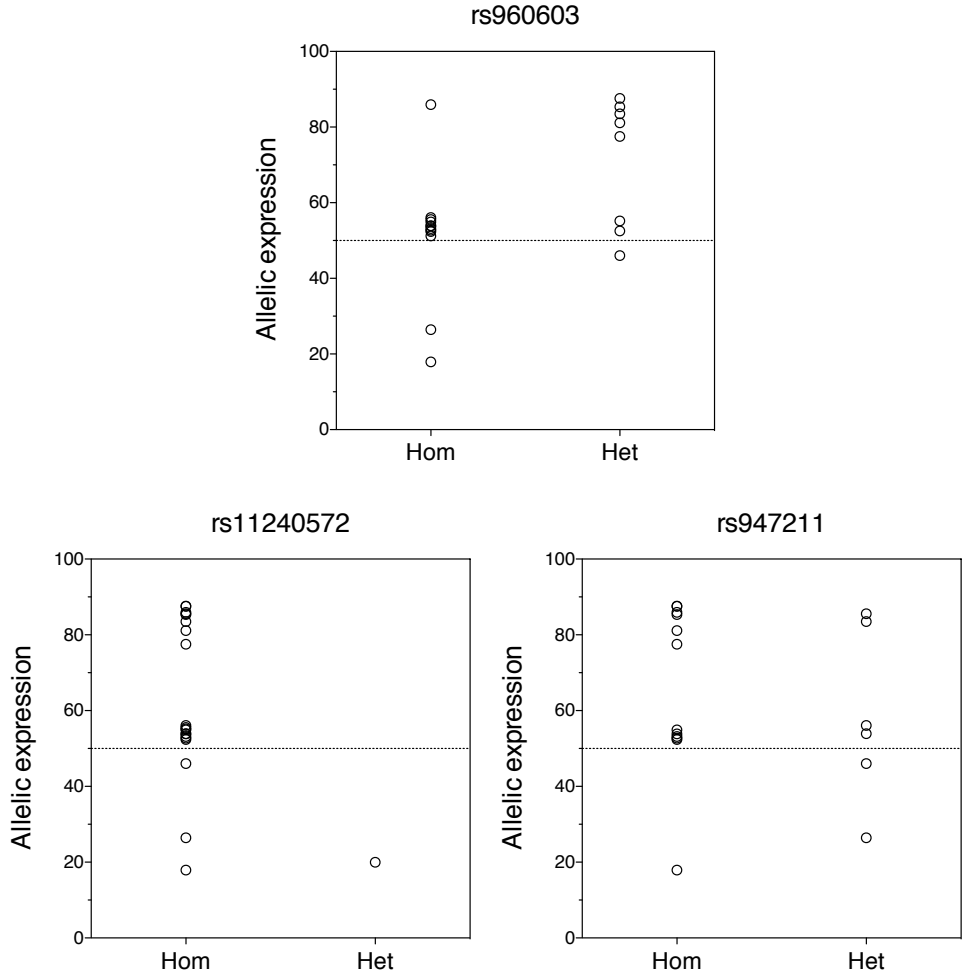


Figure 6.5 | Allelic expression of *PM20D1* grouped by genotype of candidate regulatory variants. Data points represent the mean percentage representation of the major allele (n=4) for each individual. Individuals are grouped by homozygotes (Hom) and heterozygotes (Het).

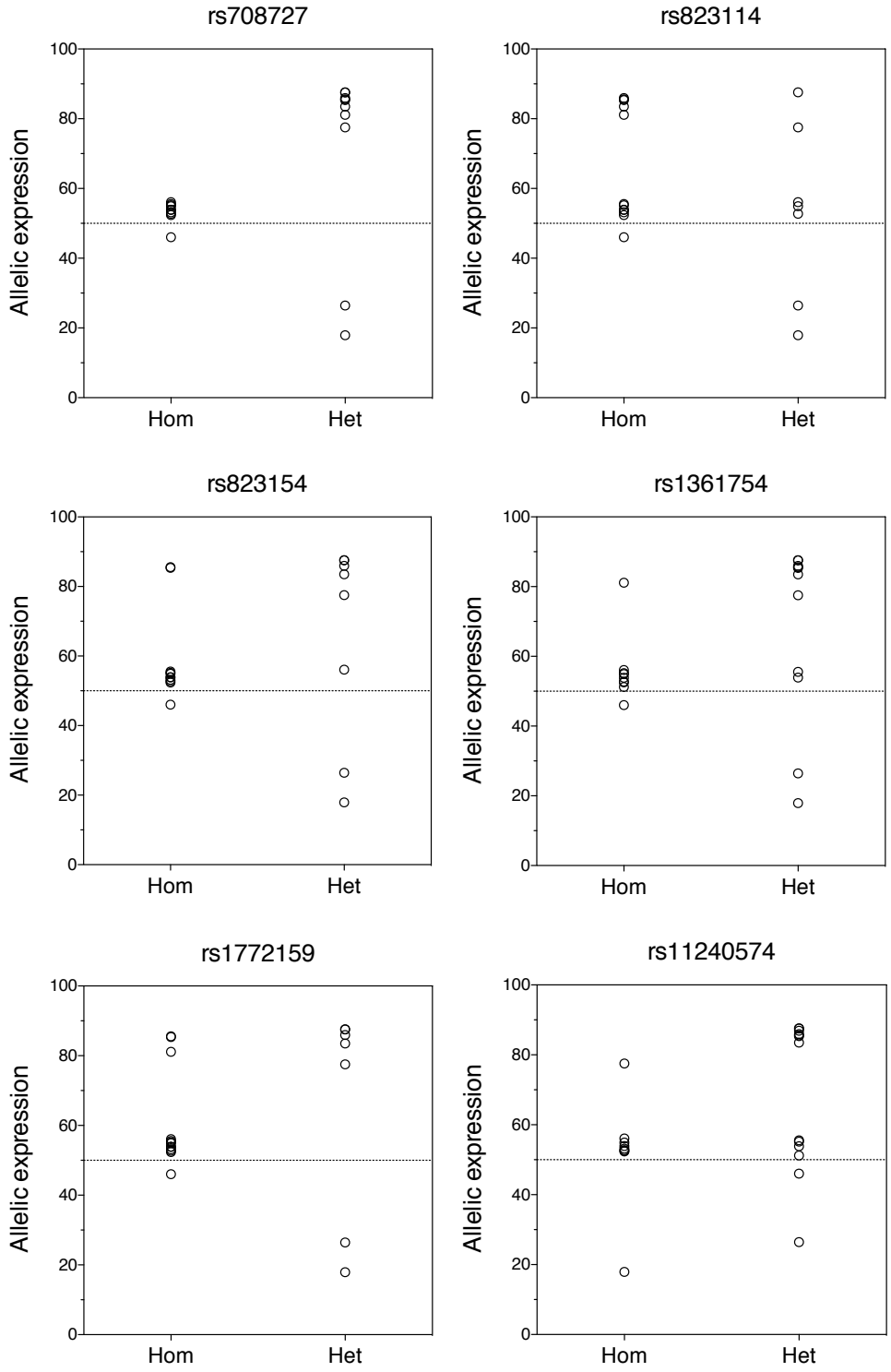


Figure 6.6 | Allelic expression of *PM20D1* grouped by genotype of candidate regulatory variants. Data points represent the mean percentage representation of the major allele (n=4) for each individual. Individuals are grouped by homozygotes (Hom) and heterozygotes (Het).

Table 6.5 | Phase predictions of rs708727 and the allelic expression tag SNP for monoallelic individuals. Sample A049/03 dropped out of phase analysis due to incomplete genotype data.

Individual	Strand	rs708727	rs1104899	Probability
A033/96	1a	G	A	0.943
	1b	A	G	
A094/95	1a	G	A	0.914
	1b	A	G	
A098/89	1a	G	A	0.921
	1b	A	G	
A123/09	1a	G	A	0.808
	1b	A	G	
A130/09	1a	G	A	0.628
	1b	A	G	
A152/95	1a	G	G	0.734
	1b	A	A	
A205/94	1a	G	A	0.853
	1b	A	G	
A206/90	1a	G	G	0.569
	1b	A	A	
A331/94	1a	G	A	0.961
	1b	A	G	
A346/95	1a	G	A	0.628
	1b	A	G	

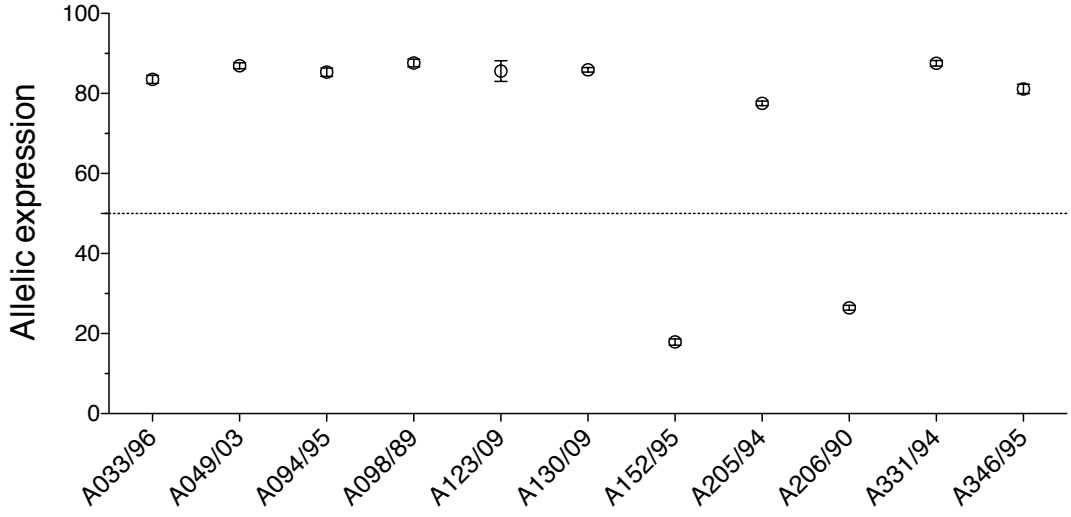


Figure 6.7 | Allelic expression of *PM20D1* in monoallelic individuals. Data points represent the mean percentage representation of the major allele (n=4) for each monoallelic individual and error bars illustrate the SEM.

The associated SNP, rs708727, is a synonymous coding variant located in exon six of the neighbouring gene *SLC41A1*. In the Caucasian population, rs708727 is in linkage disequilibrium ($r^2 = 1$) with four other SNPs from the 1000 Genomes Project: rs823074, rs823075, rs9438393 and rs823080. Two of these are located in an intron of *SLC41A1* and two are in the intergenic space between *SLC41A1* and *PM20D1* (see Figure 6.8). These five SNPs were subjected to bioinformatic analysis of transcription factor binding sites using three prediction packages. Each SNP is predicted to reside within multiple transcription factor binding sites and each is predicted to cause allele-specific transcription factor binding. There is no consistency of binding site prediction between packages (see Table 6.6).

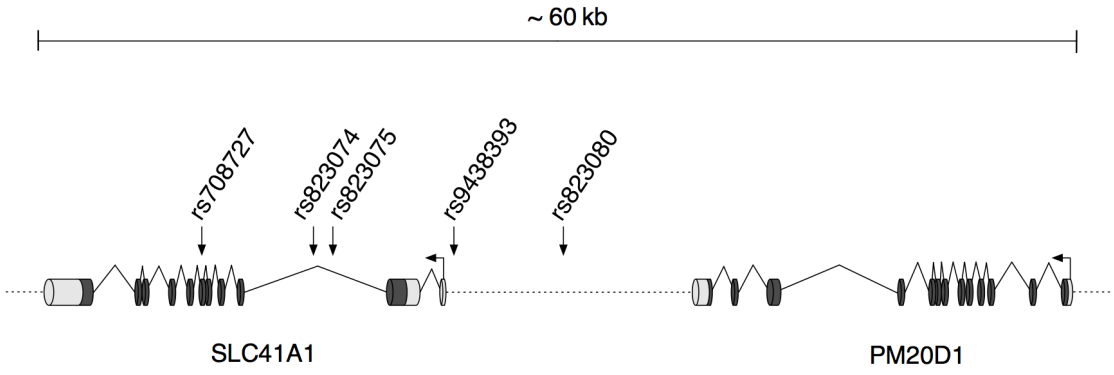


Figure 6.8 | Graphic representation of *PM20D1* and neighbouring *SLC41A1*. rs708727 and the four SNPs in linkage disequilibrium are illustrated.

Professor Huw Morris and Dr Nigel Williams tested rs708727 for association with PD (personal communication) using a meta-analysis dataset from five PD genome-wide association studies in Caucasian populations from Europe and the USA (Nalls et al. 2011). No association was found ($p = 0.7191$).

Table 6.6 | Predicted transcription factor binding motifs.

SNP	Software	Transcription factors	Binding allele
rs708727	TFSearch	-	-
	MatInspector	Doublesex and mab-3 related 1	A + G
		SMAD3	A + G
		X-box binding protein 1	G
		Myc-Max	G
	AliBaba	Repressor activator protein 1	G
rs823074	TFSearch	-	-
	MatInspector	Zinc finger protein 148	C + T
		Zinc finger protein 219	C + T
		GRE	C
	AliBaba	Specificity protein 1	C
		CCAAT/enhancer binding protein alpha	T
rs823075	TFSearch	-	-
	MatInspector	Homeobox protein H6 2	C + T
		Proximal sequence element	C + T
		General transcription factor IIIC	C
		NK3 homeobox 1	T
	AliBaba	-	-
rs9438393	TFSearch	-	-
	MatInspector	T-cell leukemia homeobox 1	A + G
		Grainyhead-like 1	A + G
		E2F-4/DP-2 heterodimeric complex	A
		Hmx2/Nkx5-2 homeodomain	A + G
		Olfactory neuron-specific factor	A + G
		Winged helix protein	G
	AliBaba	-	-
rs823080	TFSearch	Ikaros 2	A
	MatInspector	NF-kappaB	A + G
		Ikaros 3	A + G
		Glial cells missing homolog 1	A
		Mesoderm posterior 1 and 2	A + G
		Transcription factor yin yang 2	A + G
		Neurogenin 1 and 3	A + G
		SWI/SNF related, matrix associated	A
		Myc associated zinc finger protein	G
	AliBaba	Specificity protein 1	A + G
		CACCC-bi	A
		Early growth response 1	G
		CCAAT/enhancer binding proteins	G

6.4 Discussion

The study reported in this chapter initially sought to screen for novel imprinted genes. The genome-wide assay of allele-specific expression reported in Chapter 3 detected 22 genes that showed monoallelic expression of the same allele in all clonal NSC lines for at least two genotypes. Twelve of these genes were known imprinted genes and therefore, the remaining ten were deemed putatively imprinted. These candidates were assayed in post-mortem adult cortex where two genes, *PM20D1* and *THNSL2*, displayed a pattern of allelic expression consistent with genomic imprinting (i.e. monoallelic expression in all three initial samples). After further investigation it was determined that genomic imprinting does not cause the monoallelic expression of these two genes. Therefore, this study did not validate any novel imprinted genes.

With the exception of *PM20D1* and *THNSL2,* which appear to be explained by genetic effects (discussed below), these genes remain candidate novel imprinted genes. Tissue-specific monoallelic expression is common for imprinted genes (Prickett & Oakey 2012); therefore, the observation that they are not monoallelic in the adult cortex does not disqualify them from being imprinted in NSCs and other tissues. Further research in multiple NSC lines, or foetal brain, with parental genotype data will be required to answer this question. An alternative explanation for the allelic expression pattern of these genes is that the NSC donors are heterozygous for a *cis*-acting genetic regulatory variant while the subjects assayed in adult cortex are homozygotes. The screen in post-mortem brain tissue was in a small sample size with limited power to detect such effects. The observation that *PKDREJ* shows distortion in two of four individuals suggests that genetic effects could describe its allelic expression pattern. It has previously been

reported that a large proportion of *cis*-acting genetic variants have a tissue-specific effect on gene expression (Zhang et al. 2009; Dimas et al. 2009) and this could also explain the different allelic expression patterns observed in NSCs and adult cortex. Finally, the genome-wide assay was carried out in clonal NSCs and it cannot be ruled out that the putative imprinted genes are subject to RME in the NSCs with the same allele silenced in each clone by chance. RME is expected to be masked in non-clonal brain tissue by neighbouring cells that express the alternate, or both, alleles and this could explain the observed biallelelic expression of these genes in adult cortex. The remaining candidate imprinted genes require further investigation in other tissues to determine the cause of monoallelic expression observed in the NSCs.

The genome-wide assay of allele-specific expression reported in Chapter 3 was not designed with the intention of screening for novel imprinted genes and, therefore, it has some limitations when used for this application. Firstly, the small sample size of three genotypes reduces the power to discriminate between genomic imprinting and common genetic effects. Secondly, the validation of genomic imprinting requires the demonstration of parent of origin specific allele silencing and this information was not available for the NSC lines assayed. Together, these limitations meant that candidates were followed up in different tissues, exposing the confounding factor of tissue-specific regulation. Finally, screening for genomic imprinting in clonal tissue is not ideal as RME genes can be mistaken for imprinted genes when the same allele is silenced in all clones by chance. The complete definition of human genomic imprinting will be laborious task, requiring transcriptome-wide analysis of allele-specific expression in all human tissues from different developmental stages. These samples also need to be accompanied by

parental genotype data. This is an important job as it promises to be informative for both human disease and evolution.

The gene *PM20D1* was found to be monoallelic in all of the cortical and striatal NSC lines as well as eleven of the 23 subjects assayed in adult cortex. I found the monoallelic expression of *PM20D1* to be strongly associated with rs708727, a SNP in the neighbouring gene *SLC41A1*; all homozygotes were biallelic and all heterozygotes were monoallelic. rs708727 has a minor allele frequency of 0.46 in the CEU HapMap cohort and is rare in Asian and African HapMap populations. The SNPExpress eQTL database reports a trend towards altered transcript levels of *PM20D1* in brain associated with rs708727 genotype, although it did not meet the threshold for statistical significance (p = 0.07978). The association between rs708727 and the monoallelic expression of *PM20D1* does not prove causality; it could be tagging the causative variant. The location of rs708727, in an exon of a neighbouring gene, makes this a more likely explanation. It was found to be in linkage disequilibrium with four other SNPs from the 1000 Genomes Project (rs823074, rs823075, rs9438393 and rs823080), of which two are also in *SLC41A1* and two are in the intergenic space between *SLC41A1* and *PM20D1*. Bioinformatic analysis of these SNPs was carried out with the aim of revealing which is most likely to define the monoallelic expression of *PM20D1*. All were predicted to alter transcription factor binding, but there was no consistency between predictions from the three programs. The observation that *PM20D1* is expressed biallelically in all seven individuals assayed in blood indicates that the causative variant is in a tissue-specific regulatory region. However, it should be noted that the allelic status of *PM20D1* in the brain of those individuals is unknown. Further molecular biology will be required to identify the causative variant. One approach for further study is site directed mutagenesis;

rs708727 and the four other SNPs in linkage disequilibrium could each be altered to homozygous state in the monoallelic NSC lines. Allelic expression analysis would then reveal if any of these variants were causing the monoallelic expression of *PM20D1*.

Little is known about the functional role of PM20D1 (peptidase M20 domain containing 1) other than a link with peptidase activity and metal ion binding inferred from electronic annotation (http://www.ncbi.nlm.nih.gov/gene/). The observation that *PM20D1* is commonly monoallelic in NSCs and adult brain was of particular interest because it is located in the PD susceptibility locus *PARK16*. A SNP in *PM20D1* intron ten, rs11240572, has shown genome-wide significant association with PD in the Japanese population (Satake et al. 2009). While the strongest *PARK16* association signal came from a SNP in *NUCKS1,* linkage disequilibrium analysis indicated that rs11240572 represented an independent association signal. The association between rs11240572 and PD has twice been replicated in Chinese cohorts (Tan et al. 2010; Yan et al. 2011). The signal from rs11240572 was less robust in a Caucasian genome-wide association study ($p = 1.3 \times 10^{-4}$) and did not surpass Bonferroni-corrected significance (Simón-Sánchez et al. 2009). However, the significance was improved when the initial study was combined with a stage two replication study ($p = 6.11 \times 10^{-7}$) and the investigators concluded that it is likely to represent a true association. The allele frequency of the implicated SNP is low in the tested population (2 to 4%) and this could account for the weaker association detected in the Caucasian population. Professor Huw Morris and Dr Nigel Williams tested the SNP associated with monoallelic expression of *PM20D1*, rs708727, for association with PD in a meta-analysis dataset from five PD genome-wide association studies in European and North American populations (personal communication). No association was found ($p = 0.7191$), indicating that the monoallelic expression of *PM20D1* is not influencing PD

susceptibility in these populations. In the future it will be interesting to test for a compound heterozygote effect of rs708727 and rs11240572 on PD risk.

THNSL2 also showed an allelic expression pattern consistent with genomic imprinting after the initial screen in post-mortem cortex. Further investigation in parent-child trios revealed that *THNSL2* was also monoallelic in all seven individuals in whole blood, but monoallelic expression was not dependent on parent of origin. This is inconsistent with genomic imprinting and indicative of a *cis*-acting genetic variant. Interrogation of the SNPExpress eQTL database revealed sixteen SNPs associated with altered transcript levels in human brain. Of these, two were heterozygous in the two monoallelic NSC genotypes, making them candidates for further study. An additional six were not covered on the genotyping array and therefore they also remain candidates. The fact that all individuals assayed are monoallelic indicates that the allelic expression tag SNP is in linkage disequilibrium with, or is itself, the causative variant. The same approach used for the investigation of *PM20D1* could be carried out to map the causative variant. It cannot be ruled out that a parent of origin independent epigenetic effect causes the monoallelic expression of *THNSL2*.

THNSL2 (threonine synthase-like 2) is a conserved gene remnant encoding a protein with close homology to threonine synthase. It is incapable of L-threonine synthesis (metazoa have lost the L-threonine biosynthetic pathway) but can act as a catabolic phospholyase (Donini et al. 2006). The *THNSL2* mRNA splice variant B, named *SOFAT*, has been shown to encode a protein with cytokine activity capable of inducing IL6 production by oseteoblasts (Rifas & Weitzmann 2009). Considering the current understanding of THNSL2 function, the biological significance of its monoallelic expression is unclear.

6.5 Conclusion

In conclusion, the present study sought to test ten putative imprinted genes predicted from a genome-wide assay of monoallelic expression in human NSCs. No further evidence in support of imprinting was gained from screening in adult cortex samples. However, I find the monoallelic expression of *PM20D1* to be strongly associated with rs708727. This SNP shows no association with PD in the European and Northern American populations tested. Evidence in support of a genetic variant underlying the monoallelic expression of *THNSL2* is also presented. This study was not ideally suited for the identification of novel imprinted genes and the nature of the monoallelic expression of eight of the putative imprinted genes in the NSCs remains to be determined. However, this study validates the use of genome-wide screens for monoallelic expression as a method of identifying large-effect genetic *cis* variants. This approach could prove informative in screening for rare variants in disease cohorts.

Chapter 7. General discussion

The overall aim of the series of experiments presented in this thesis was to explore autosomal monoallelic gene expression in human neural tissue. Monoallelic expression has long been recognised in the form of XCI, genomic imprinting and RME of a small number of gene families. More recently, studies have reported evidence of widespread autosomal RME in human lymphoblastoid cells (Gimelbrant et al. 2007) and mouse NSCs (Wang et al. 2010; Li et al. 2012). If this were true for human NSCs, and by extension the developing human brain, RME would be predicted to have significant implications for neural development, function and disease. Therefore, I sought to assess the prevalence, nature and potential functional implications of RME in human NSCs and their differentiated progeny.

The initial objective of this thesis was to determine whether or not RME occurs in human NSCs, and if so, to what extent. In Chapter 3 I report the first genome-wide assessment of allele-specific expression in human neural tissue with the power to detect RME. A genotyping microarray array approach was used based on the method described by Gimelbrant et al. (2007). I find that 1.6 to 2.2% of assayed autosomal genes display monoallelic expression and 0.5 to 1.1% show evidence of RME. Therefore, I demonstrate for the first time that, as in human lymphoblastoid cells and mouse NSCs, autosomal RME is widespread in human NSCs. The prevalence of RME reported in this study is consistent with studies of monoallelic expression in mouse NSCs (Wang et al. 2010; Li et al. 2012) but lower than the 9% reported in human lymphoblastoid cells (Gimelbrant et al. 2007). This discrepancy is likely explained by technical differences between our studies;

unlike Gimelbrant et al., the present study excluded lowly expressed genes as well as those with a single informative SNP.

The novel RME genes reported in Chapter 3 are dispersed throughout the autosome and were found enriched for the functional term "signal" and the topological domains "extracellular" and "transmembrane". This is consistent with observations made by Gimelbrant et al. who found a disproportionate fraction of monoallelic genes encoding cell surface proteins (Gimelbrant et al. 2007). Additionally, the RME genes detected in the brain-derived NSCs were enriched for developmental terms. I find evidence in support of a genetic component to RME, predisposing alleles for RME: significantly more RME genes overlap between genotypes than would be expected by chance. Furthermore, I find that RME genes are significantly biased towards silencing the same rather than the alternate allele. This demonstrates that, at least for a proportion of the genes detected, the choice of which allele to silence is not truly random. Further investigation of the mechanisms that regulate this form of monoallelic expression is required. I find the novel RME genes to be associated with increased DNA methylation at promoter regions when compared to biallelic loci. Using the demethylating agent 5-Aza, I demonstrate that DNA methylation directly regulates the monoallelic expression of all twelve novel RME genes assayed in the spinal cord NSC lines.

Given the observation that autosomal RME is widespread in human NSCs, I assessed whether it is maintained after differentiation into the cell types that populate the adult central nervous system: neurons and glia. A microarray-based genome wide analysis of allele-specific expression was carried out using the spinal cord NSC lines, which have previously been shown to retain their positional specification and generate functional

neurons (paper submitted). I find that non-imprinted autosomal monoallelic expression was largely retained after neural differentiation; 71% of RME genes assayed after differentiation maintained monoallelic status. The genes that lost monoallelic status predominantly adopted an intermediate state of allele expression imbalance. It is unclear whether this reflects an intermediate allele ratio for each cell, or whether a proportion of cells remain monoallelic while others revert to biallelic status. Indeed, this differentiation protocol results in a heterogeneous cell population and neural cell type-specific RME has previously been described in mouse (Wang et al. 2010; Li et al. 2012). A revealing follow up study could be carried out assaying cell populations purified for specific cell types.

The studies described above demonstrate that autosomal RME is widespread in human NSCs *in vitro* and indicate that it is largely preserved after differentiation. Whether or not NSCs and their progeny are subject to the same regulation *in vivo* is of particular interest. Direct allele-specific expression analysis of brain tissue is complicated by the fact that RME is expected to be masked by the heterogeneous non-clonal nature of this tissue. Therefore, I interrogated publically available epigenetic data from human foetal and adult brain in an attempt to test the *in vivo* relevance of these findings. In both foetal and adult brain, RME genes were associated with significantly increased DNA methylation when compared to biallelic loci. Additional epigenetic measures were assessed in foetal brain; RME genes were found enriched for repressive H3K27me3, while biallelic genes were enriched for signatures of open chromatin (H3K4me3, H3K9ac and DNaseI hypersensitivity). Therefore, I find that the novel RME loci detected in this thesis are epigenetically distinct from biallelic loci in a manner consistent with monoallelic expression *in vivo*. Whether or not they are also subject to RME remains to be determined. An essential follow up study to advance our understanding of this process is

ex vivo single cell analysis from human brain. Single cell RNA-seq technology is becoming established (Tang et al. 2010) and provides the ideal platform to assess RME in heterogeneous non-clonal brain tissue.

If widespread RME is occurring *in vivo*, then the point at which it is initiated will dictate the size and distribution of clones that share the same pattern of monoallelic and biallelic genes. In Chapter 4 I report a pilot study investigating the allelic status of five novel RME genes after reprogramming to iPSCs. I find that all five genes loose monoallelic status and four completely revert to biallelic status. Preliminary results from the succeeding genome-wide analysis are consistent with this data, with 83% of non-imprinted autosomal monoallelic genes detected in SPC01 losing monoallelic status after reprogramming. These results indicate that RME is initiated between the ESC and NSC stage. Following iPSCs or ESCs through neural differentiation could enable the precise timing of RME initiation to be elucidated.

After exploring the extent and nature of RME in Chapters 3 and 4, the next aim was to investigate the possible functional implications that this process could have at the cellular level. Other forms of monoallelic expression are known to affect cells by either altering transcript levels or exposing functional diversity. In Chapter 5 I assess the transcript levels of RME genes in biallelic versus monoallelic clones using gene expression estimates from the genotyping microarrays. I find that significantly more monoallelic clones have lower transcript levels than sister biallelic clones than would be expected by chance. Furthermore, the mean expression of monoallelic clones was significantly lower than that of biallelic clones. These data indicate that RME is capable of influencing cell phenotype by altering gene transcript levels. However, it should be noted that this was not

a general rule, as nearly 40% of genes did not show the expected relationship. It is not clear whether this is accounted for by feedback mechanisms to correct for altered gene dosage, or whether interclonal variation in gene expression is masking the effect. A recent study on mouse NSCs using RNA-seq, which is better suited to this analysis, lends support to my findings; they report a mean 35.8% reduction of expression levels in monoallelic compared to biallelic clones (Li et al. 2012). Further investigation of the dosage sensitivity of the RME genes could be carried out by genetic manipulation of gene copy numbers. Whether or not this relationship translates to the protein level remains to be determined.

In addition to altering gene transcript levels, monoallelic expression can alter cellular phenotype by exposing functional variation. I tested the extent that this is likely to occur for the novel RME genes reported in this thesis by assessing the number of nsSNPs per coding base in RME and biallelic genes. Intriguingly, I find a significant 1.58 fold enrichment of nsSNPs per coding base in RME genes when compared to biallelic genes. This indicates that the widespread autosomal RME reported in this thesis has the potential to influence cellular phenotype by exposing functional variation. The reason for the enrichment of nsSNPs is, as yet, unclear. It could be that the generation of tissue diversity itself is an adaptive advantage in a manner similar to odorant receptors, antigen receptors and protocadherins. The enrichment of genes encoding transmembrane signalling proteins is evidence in support in support of this hypothesis, although further work at the gene level will be required to determine if this is the case. The observation that RME genes are enriched for nsSNPs is also consistent with a theory of RME enhancing a gene's evolvability (Chess 2012). This concept is appealing as it has previously been reported that RME are more than twice as likely to be located near noncoding sequences

associated with human lineage-specific accelerated evolution (Gimelbrant et al. 2007), and my colleague Dr Aaron Jeffries replicated this finding with our dataset (Jeffries et al. 2012). Investigating the evolutionary rate of these genes by comparison of RME and biallelic gene ka/ks ratios could be informative. It should also be noted that the enrichment of nsSNPs in RME genes could merely represent these genes being less critical for the cell and subsequently under less intensive purifying selective pressure.

Taken together, the findings discussed above demonstrate that autosomal RME is widespread in human NSCs *in vitro* and that it is largely maintained after differentiation into neurons and glia. Here, this process has the potential to drive phenotypic diversity between otherwise identical clones of cells by altering gene transcript levels and by exposing functional variation. This could go some way towards explaining the well-documented non-genetic heterogeneity observed between clonal cell populations (Huang 2009) and it highlights the importance of full and proper epigenetic characterisation of NSC lines for clinical and research applications. Furthermore, the data presented in this thesis supports a model in which the human central nervous system is a mosaic of clones of cells, each with different combinations of monoallelic and biallelic genes. RME can give rise to three cell states: biallelic, monoallelic for the paternal allele or monoallelic for the maternal allele (see Figure 7.1). This would be predicted to have significant implications for human neural development, function and disease. Widespread RME would have the potential to drive phenotypic diversity, not just between cells but also between individuals. Take the example of monozygotic twins; while they share identical genetic backgrounds, the stochastic nature of RME would result in distinct gene expression potentials for each twin in discrete brain regions. This could help explain the discordance between monozygotic twins for psychiatric disorders (Dempster et al. 2011).

The penetrance of a disease-causing allele subject to RME would be altered in heterozygotes based on the location, or connectivity, of cells exclusively expressing the mutant allele. In addition, widespread RME in somatic tissues could be relevant to human disease by way of exposing recessive deleterious mutations, although clones exclusively expressing the mutant allele could be selected against during development in manner similar to XCI skewing in females carrying deleterious X-linked mutations (Plenge et al. 2002). While these potential implications are intriguing, it should be stressed that widespread RME in human neural tissue remains to be demonstrated *in vivo*.

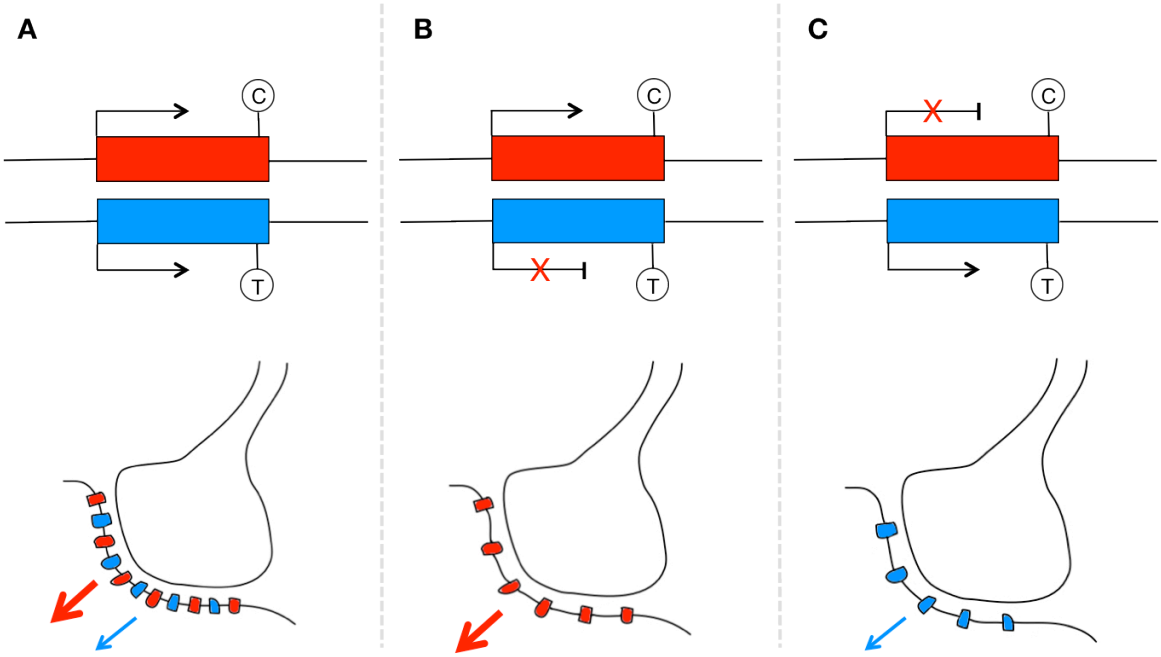


Figure 7.1 | RME can give rise to three cell states. A hypothetical scenario where a neurotransmitter receptor gene comprised of functionally distinct alleles is subject to RME. **A)** Biallelic expression: both receptor variants are expressed and present at the synapse resulting in maximal sensitivity. **B)** Monoallelic expression of the high sensitivity C variant: here, reduced receptor density decreases the excitability of the synapse. **C)** Monoallelic expression of the low sensitivity T variant: this scenario results in even further reduction of post-synaptic cell excitability.

In Chapter 6 I report a screen for novel imprinted genes using the genome-wide allele-specific expression data reported in Chapter 3. Ten putative novel imprinted were assayed in post-mortem adult cortex where two genes, *PM20D1* and *THNSL2*, had an allelic expression pattern consistent with genomic imprinting. After further investigation I was able to map genetic variants associated with the monoallelic expression of *PM20D1*. This gene was of particular interest as it is located in the PD susceptibility locus *PARK16* and a SNP within *PM20D1* has shown association with PD (Satake et al. 2009; Tan et al. 2010; Yan et al. 2011). The SNP associated with *PM20D1* monoallelic expression, rs708727, was not found to be associated with PD in the meta-analysis dataset tested, indicating that the causative variant does not influence PD susceptibility in the Caucasian population. However, a compound heterozygote effect cannot yet be ruled out. *THNSL2*, which also showed an allelic expression pattern consistent with genomic imprinting after the initial screen in post-mortem cortex, was also monoallelic in blood but showed no parent of origin dependence and is therefore inconsistent with genomic imprinting. Interrogation of the SNPExpress eQTL database revealed sixteen candidate SNPs that influence *THNSL2* expression in human brain. Two of these are heterozygous in both monoallelic NSC genotypes and therefore good candidates for further study. The cause of monoallelic expression of the remaining eight putative novel imprinted genes remains unknown. These genes remain candidate novel imprinted genes, although common genetic effects could also explain the observed allelic expression pattern. The work reported in Chapter 6 highlights the need for large sample sizes with parental genotype information when screening for genomic imprinting. It also validates the use of genome-wide assays of monoallelic expression for the detection of large-effect genetic *cis* variants. This approach could prove valuable in screening for rare variants in disease cohorts.

Conclusion

In this thesis I report evidence of widespread autosomal RME expression in human NSCs and their differentiated progeny *in vitro* for the first time. My data indicates that this process has the potential to drive phenotypic diversity between genetically identical cells. These findings highlight the importance of full and proper epigenetic characterisation of NSC lines for clinical and research applications. Furthermore, they support a model of widespread RME in the human central nervous system driving phenotypic diversity both between clones of cells and individuals. If this process does occur *in vivo*, it would be predicted to have significant implications for brain development, function and disease. These observations warrant a follow up investigation in *ex vivo* human foetal and adult brain. Single cell RNA-seq represents an ideal platform to progress our knowledge of this phenomenon. Finally, I have mapped the genetic variants associated with the monoallelic expression of *PM20D1*. While this variant does not appear to have any influence on PD susceptibility, the finding validates the use of screens for monoallelic expression in identifying *cis* acting genetic variants.

References

Adzhubei, I. a, Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S., et al. (2010). A method and server for predicting damaging missense mutations. *Nature methods*, *7*(4), 248–9.

Allen, E., Horvath, S., Tong, F., Kraft, P., Spiteri, E., Riggs, A. D., & Marahrens, Y. (2003). High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(17), 9940–5.

Amos-landgraf, J. M., Cottle, A., Plenge, R. M., Friez, M., Schwartz, C. E., Longshore, J., & Willard, H. F. (2006). X Chromosome Inactivation Patterns of 1,005 Phenotypically Unaffected Females, *79*(September), 493–499.

Anthony, T. E., & Heintz, N. (2008). Genetic lineage tracing defines distinct neurogenic and gliogenic stages of ventral telencephalic radial glial development. *Neural development*, *3*(30), 1–12.

Anthony, T. E., Klein, C., Fishell, G., & Heintz, N. (2004). Radial glia serve as neuronal progenitors in all regions of the central nervous system. *Neuron*, *41*(6), 881–90.

Arvidsson, A., Collin, T., Kirik, D., Kokaia, Z., & Lindvall, O. (2002). Neuronal replacement from endogenous precursors in the adult brain after stroke. *Nature medicine*, *8*(9), 963–70.

Aït Yahya-Graison, E., Aubert, J., Dauphinot, L., Rivals, I., Prieur, M., Golfier, G., Rossier, J., et al. (2007). Classification of human chromosome 21 gene-expression variations in Down syndrome: impact on disease phenotypes. *American journal of human genetics*, *81*(3), 475–91.

Bacher, C. P., Guggiari, M., Brors, B., Augui, S., Clerc, P., Avner, P., Eils, R., et al. (2006). Transient colocalization of X-inactivation centres accompanies the initiation of X inactivation. *Nature cell biology*, *8*(3), 293–9.

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell research*, *21*(3), 381–95.

Barbaro, N. M., Gupta, N., & Kunwar, S. (2004). Unique astrocyte ribbon in adult human brain contains neural stem cells but lacks chain migration. *Nature, 427*(2), 740–744.

Barlow, D., Stöger, R., & Herrmann, B. (1991). The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature*, *349*(6304), 84-87.

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, *21*(2), 263–5.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., Wei, G., et al. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, *129*(4), 823–37.

Bassing, C., Swat, W., & Alt, F. (2002). The mechanism and regulation of chromosomal V(D)J recombination. *Cell*, *109*(D), 45–55.

Bell, a C., & Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, *405*(6785), 482–5.

Belmont, J. W. (1996). Genetic control of X inactivation and processes leading to X-inactivation skewing. *American journal of human genetics, 58*(6), 1101.

Ben-David, E., Granot-Hershkovitz, E., Monderer-Rothkoff, G., Lerer, E., Levi, S., Yaari, M., Ebstein, R. P., et al. (2011). Identification of a functional rare variant in autism using genome-wide screen for monoallelic expression. *Human molecular genetics*, *20*(18), 3632–41.

Berger, S. L., Kouzarides, T., Shiekhattar, R., & Shilatifard, A. (2009). An operational definition of epigenetics. *Genes & development*, *23*(7), 781–3.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes & development*, *16*(1), 6–21.

Bourguin, a, Tung, R., Galili, N., & Sklar, J. (1990). Rapid, nonradioactive detection of clonal T-cell receptor gene rearrangements in lymphoid neoplasms. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(21), 8536–40.

Brady, B. L., Steinel, N. C., & Bassing, C. H. (2010). Antigen receptor allelic exclusion: an update and reappraisal. *Journal of immunology*, *185*(7), 3801–8.

Bray, N. J., Buckland, P. R., Owen, M. J., & O'Donovan, M. C. (2003). Cis-acting variation in the expression of a high proportion of genes in human brain. *Human genetics, 113*(2), 149-153.

Bray, N. J., & O'Donovan, M. C. (2006). Investigating cis-acting regulatory variation using assays of relative allelic expression. *Psychiatric genetics*, *16*(4), 173–7.

Brown, C., Ballabio, A., Rupert, J., Lafreniere, R., Grompe, M., Tonlorenzi, R., & Willard, H. (1991). A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*, *349*, 38–44.

Buonocore, F., Hill, M. J., Campbell, C. D., Oladimeji, P. B., Jeffries, A. R., Troakes, C., Hortobagyi, T., et al. (2010). Effects of cis-regulatory variation differ across regions of the adult human brain. *Human molecular genetics*, *19*(22), 4490–6.

Burnet, F. M. (1957). A modification of Jerne's theory of antibody production using the concept of clonal selection. *Australian journal of science,* (20), 67-69.

Butz, J. a, Yan, H., Mikkilineni, V., & Edwards, J. S. (2004). Detection of allelic variations of human gene expression by polymerase colonies. *BMC genetics*, *5*(3), 1-5.

Carrel, L., & Willard, H. F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature*, *434*(7031), 400–4.

Casellas, R., Zhang, Q., Zheng, N.-Y., Mathias, M. D., Smith, K., & Wilson, P. C. (2007). Igkappa allelic inclusion is a consequence of receptor editing. *The Journal of experimental medicine*, *204*(1), 153–60.

Chambers, S. M., Fasano, C. a, Papapetrou, E. P., Tomishima, M., Sadelain, M., & Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature biotechnology*, *27*(3), 275–80.

Chan, K. M., Zhang, H., Malureanu, L., Van Deursen, J., & Zhang, Z. (2011). Diverse factors are involved in maintaining X chromosome inactivation. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(40), 16699–704.

Chess, A., Cedar, H., & Axel, R. (1994). Allelic inactivation regulates olfactory receptor gene expression. *Cell*, *78*(5), 823–34.

Chess, A. (2005). Monoallelic expression of protocadherin genes. *Nature genetics*, *37*(2), 120–1.

Chess, A. (2012). Mechanisms and consequences of widespread random monoallelic expression. *Nature reviews. Genetics*, *13*(6), 421–8.

Cheung, V. G., Spielman, R. S., Ewens, K. G., Weber, T. M., Morley, M., & Burdick, J. T. (2005). Mapping determinants of human gene expression by regional and genome-wide association. *Nature*, *437*(7063), 1365–9.

Chow, J. C., Hall, L. L., Baldry, S. E. L., Thorogood, N. P., Lawrence, J. B., & Brown, C. J. (2007). Inducible XIST-dependent X-chromosome inactivation in human somatic cells is reversible. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(24), 10104–9.

Christian, S. L., Robinson, W. P., Huang, B., Mutirangura, a, Line, M. R., Nakao, M., Surti, U., et al. (1995). Molecular characterization of two proximal deletion breakpoint regions in both Prader-Willi and Angelman syndrome patients. *American journal of human genetics*, *57*(1), 40–8.

Clemson, C M, McNeil, J. a, Willard, H. F., & Lawrence, J. B. (1996). XIST RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *The Journal of cell biology*, *132*(3), 259–75.

Clemson, Christine Moulton, Hall, L. L., Byron, M., McNeil, J., & Lawrence, J. B. (2006). The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(20), 7688–93.

Conti, L., & Cattaneo, E. (2010). Neural stem cell systems: physiological players or in vitro entities? *Nature reviews. Neuroscience*, *11*(3), 176–87.

Conti, L., Pollard, S. M., Gorba, T., Reitano, E., Toselli, M., Biella, G., Sun, Y., et al. (2005). Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS biology*, *3*(9), e283.

Cooper, W. N., Luharia, A., Evans, G. a, Raza, H., Haire, A. C., Grundy, R., Bowdin, S. C., et al. (2005). Molecular subtypes and phenotypic expression of Beckwith-Wiedemann syndrome. *European journal of human genetics : EJHG*, *13*(9), 1025–32.

Corcoran, A. E. (2005). Immunoglobulin locus silencing and allelic exclusion. *Seminars in immunology*, *17*(2), 141–54.

Corotto, F. S., Henegar, J. a, & Maruniak, J. a. (1993). Neurogenesis persists in the subependymal layer of the adult mouse brain. *Neuroscience letters, 149*(2), 111–4.

Crick, F. (1970). Central dogma of molecular biology. *Nature, 227*(5258), 561-563.

Culican, S. M., Baumrind, N. L., Yamamoto, M., & Pearlman, a L. (1990). Cortical radial glia: identification in tissue culture and evidence for their transformation to astrocytes. *The Journal of neuroscience, 10*(2), 684–92.

Curley, J. P., Barton, S., Surani, A., & Keverne, E. B. (2004). Coadaptation in mother and infant regulated by a paternally expressed imprinted gene. *Proceedings of The Royal Society of London*, *271*(1545), 1303–9.

Davies, M. N., Volta, M., Pidsley, R., Lunnon, K., Dixit, A., Lovestone, S., Coarfa, C., et al. (2012). Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome biology*, *13*(6), R43.

Davies, W., Isles, A. R., & Wilkinson, L. S. (2005). Imprinted gene expression in the brain. *Neuroscience and biobehavioral reviews*, *29*(3), 421–30.

De Bakker, P. I. W., Yelensky, R., Pe'er, I., Gabriel, S. B., Daly, M. J., & Altshuler, D. (2005). Efficiency and power in genetic association studies. *Nature genetics*, *37*(11), 1217–23.

De Jonge, H. J. M., Fehrmann, R. S. N., De Bont, E. S. J. M., Hofstra, R. M. W., Gerbens, F., Kamps, W. a, De Vries, E. G. E., et al. (2007). Evidence based selection of housekeeping genes. *PloS one*, *2*(9), e898.

Degner, J. F., Marioni, J. C., Pai, A. a, Pickrell, J. K., Nkadori, E., Gilad, Y., & Pritchard, J. K. (2009). Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics*, *25*(24), 3207–12.

Dempster, E. L., Pidsley, R., Schalkwyk, L. C., Owens, S., Georgiades, A., Kane, F., Kalidindi, S., et al. (2011). Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. *Human molecular genetics*, *20*(24), 4786–96.

Deng, W., Aimone, J. B., & Gage, F. H. (2010). New neurons and new memories: how does adult hippocampal neurogenesis affect learning and memory? *Nature reviews. Neuroscience*, *11*(5), 339–50.

DeVeale, B., Van der Kooy, D., & Babak, T. (2012). Critical Evaluation of Imprinted Gene Expression by RNA–Seq: A New Perspective. *PLoS Genetics*, *8*(3), e1002600.

Diaz-Meyer, N., Day, C., Khatod, K., Maher, E., Cooper, W., Reik, W., Junien, C., et al. (2003). Silencing of CDKN1C (p57(KIP2)) is associated with hypomethylation at KvDMR1 in Beckwith – Wiedemann syndrome. *Journal of medical genetics,* 797–802.

Dimas, A. S., Deutsch, S., Stranger, B. E., Montgomery, S. B., Borel, C., Attar-Cohen, H., Ingle, C., et al. (2009). Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science*, *325*(5945), 1246–50.

Diplas, A. I., Hu, J., Lee, M.-J., Ma, Y. Y., Lee, Y. L., Lambertini, L., Chen, J., et al. (2009). Demonstration of all-or-none loss of imprinting in mRNA expression in single cells. *Nucleic acids research*, *37*(21), 7039–46.

Doi, A., Park, I.-H., Wen, B., Murakami, P., Aryee, M. J., Irizarry, R., Herb, B., et al. (2009). Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nature genetics*, *41*(12), 1350–3.

Donini, S., Percudani, R., Credali, A., Montanini, B., Sartori, A., & Peracchi, A. (2006). A threonine synthase homolog from a mammalian genome. *Biochemical and biophysical research communications*, *350*(4), 922–8.

El-Akabawy, G., Medina, L. M., Jeffries, A., Price, J., & Modo, M. (2011). Purmorphamine increases DARPP-32 differentiation in human striatal neural stem cells through the Hedgehog pathway. *Stem cells and development*, *20*(11), 1873–87.

Elkabetz, Y., Panagiotakos, G., Al Shamy, G., Socci, N. D., Tabar, V., & Studer, L. (2008). Human ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell stage. *Genes & development*, *22*(2), 152–65.

Eriksson, P. S., Perfilieva, E., Björk-Eriksson, T., Alborn, a M., Nordborg, C., Peterson, D. a, & Gage, F. H. (1998). Neurogenesis in the adult human hippocampus. *Nature medicine*, *4*(11), 1313–7.

Esumi, S., Kakazu, N., Taguchi, Y., Hirayama, T., Sasaki, A., Hirabayashi, T., Koide, T., et al. (2005). Monoallelic yet combinatorial expression of variable exons of the protocadherin-alpha gene cluster in single neurons. *Nature genetics*, *37*(2), 171–6.

Feinberg, A. P., Ohlsson, R., & Henikoff, S. (2006). The epigenetic progenitor origin of human cancer. *Nature reviews. Genetics*, *7*(1), 21–33.

Ferguson-Smith, A. C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. *Nature reviews. Genetics*, *12*(8), 565–75.

Ferguson-Smith, A., & Sasaki, H. (1993). Parental-origin-specific epigenetic modification of the mouse H19 gene. *Nature, 362*, 751-5.

Ferrón, S. R., Charalambous, M., Radford, E., McEwen, K., Wildner, H., Hind, E., Morante-Redolat, J. M., et al. (2011). Postnatal loss of Dlk1 imprinting in stem cells and niche astrocytes regulates neurogenesis. *Nature*, *475*(7356), 381–5.

FitzPatrick, D. R., Ramsay, J., McGill, N. I., Shade, M., Carothers, A. D., & Hastie, N. D. (2002). Transcriptome analysis of human autosomal trisomy. *Human molecular genetics*, *11*(26), 3249–56.

Freeman, B., Smith, N., Curtis, C., Huckett, L., Mill, J., & Craig, I. W. (2003). DNA from buccal swabs recruited by mail: evaluation of storage effects on long-term stability and suitability for multiplex polymerase chain reaction genotyping. *Behavior genetics*, *33*(1), 67–72.

Fricker, R. a, Carpenter, M. K., Winkler, C., Greco, C., Gates, M. a, & Björklund, a. (1999). Site-specific migration and neuronal differentiation of human neural progenitor cells after transplantation in the adult rat brain. *The Journal of neuroscience*, *19*(14), 5990–6005.

Fuda, N. J., Ardehali, M. B., & Lis, J. T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature*, *461*(7261), 186–92.

Fuss, S. H., Omura, M., & Mombaerts, P. (2007). Local and cis effects of the H element on expression of odorant receptor genes in mouse. *Cell*, *130*(2), 373–84.

Gardner, R., & Lyon, M. (1971). X chromosome inactivation studied by injection of a single cell into the mouse blastocyst. *Nature*, *231*, 358–386.

Ge, B., Gurd, S., Gaudin, T., Dore, C., Lepage, P., Harmsen, E., Hudson, T. J., et al. (2005). Survey of allelic expression using EST mining. *Genome research*, *15*(11), 1584–91.

Ge, B., Pokholok, D. K., Kwan, T., Grundberg, E., Morcos, L., Verlaan, D. J., Le, J., et al. (2009). Global patterns of cis variation in human cells revealed by high-density allelic expression analysis. *Nature genetics*, *41*(11), 1216–22.

Reynolds, B, & Weiss, S. (1992) Generation of neurons and astrocytes from isolated cells of the adult mammalian central nervous system. *Science, 255*(5052), 1707-10.

Gimelbrant, A., Hutchinson, J. N., Thompson, B. R., & Chess, A. (2007). Widespread monoallelic expression on human autosomes. *Science*, *318*(5853), 1136–40.

Ginsberg, M. R., Rubin, R. a, Falcone, T., Ting, A. H., & Natowicz, M. R. (2012). Brain transcriptional and epigenetic associations with autism. *PloS one*, *7*(9), e44736.

Glaser, T., Pollard, S. M., Smith, A., & Brüstle, O. (2007). Tripotential differentiation of adherently expandable neural stem (NS) cells. *PloS one*, *2*(3), e298.

Goecks, J., Nekrutenko, A., & Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome biology*, *11*(8), R86.

Goffredo, D., Conti, L., Febo, F. Di, Biella, G., Tosoni, A., Vago, G., Biunno, I., et al. (2008). Setting the conditions for efficient, robust and reproducible generation of functionally active neurons from adult subventricular zone-derived neural stem cells. *Cell death & differentiation, 15*(12), 1847–1856.

Goldmit, M, & Bergman, Y. (2004). Monoallelic gene expression: a repertoire of recurrent themes. *Immunological reviews*, *200*, 197–214.

Goldmit, Maya, Ji, Y., Skok, J., Roldan, E., Jung, S., Cedar, H., & Bergman, Y. (2004). Epigenetic ontogeny of the Igk locus during B cell development. *Nature Immunology*, *6*(2), 198–203.

Grafodatskaya, D., Choufani, S., Ferreira, J. C., Butcher, D. T., Lou, Y., Zhao, C., Scherer, S. W., et al. (2010). EBV transformation and cell culturing destabilizes DNA methylation in human lymphoblastoid cell lines. *Genomics*, *95*(2), 73–83.

Haig, D. (2004). Genomic imprinting and kinship: how good is the evidence? *Annual review of genetics*, *38*, 553–85.

Han, H., Cortez, C. C., Yang, X., Nichols, P. W., Jones, P. a, & Liang, G. (2011). DNA methylation directly silences genes with non-CpG island promoters and establishes a nucleosome occupied promoter. *Human molecular genetics*, *20*(22), 4299–310.

Hashimshony, T., Zhang, J., Keshet, I., Bustin, M., & Cedar, H. (2003). The role of DNA methylation in setting up chromatin structure during development. *Nature genetics*, *34*(2), 187–92.

Haubensak, W., Attardo, A., Denk, W., & Huttner, W. B. (2004). Neurons arise in the basal neuroepithelium of the early mammalian telencephalon: a major site of neurogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(9), 3196–201.

Heap, G. a, Yang, J. H. M., Downes, K., Healy, B. C., Hunt, K. a, Bockett, N., Franke, L., et al. (2010). Genome-wide analysis of allelic expression imbalance in human primary cells by high-throughput transcriptome resequencing. *Human molecular genetics*, *19*(1), 122–34.

Heinzen, E. L., Ge, D., Cronin, K. D., Maia, J. M., Shianna, K. V, Gabriel, W. N., Welsh-Bohmer, K. a, et al. (2008). Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS biology*, *6*(12), e1.

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., & Vandesompele, J. (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome biology*, *8*(2), R19.

Hewitt, S. L., Yin, B., Ji, Y., Chaumeil, J., Marszalek, K., Tenthorey, J., Salvagiotto, G., et al. (2009). RAG-1 and ATM coordinate monoallelic recombination and nuclear positioning of immunoglobulin loci. *Nature immunology*, *10*(6), 655–64.

Huang, S. (2009). Non-genetic heterogeneity of cells in development: more than just noise. *Development*, *136*(23), 3853–62.

Huynh, K. D., & Lee, J. T. (2003). Inheritance of a pre-inactivated paternal X chromosome in early mouse embryos. *Nature, 426*(6968), 857–862.

Imai, T., Sakano, H., & Vosshall, L. B. (2010). Topographic mapping - the olfactory system. *Cold Spring Harbor perspectives in biology*, *2*(8), a001776.

Meulenbelt, I., Droog, S., Trommelen, G. J., Boomsma, D. I., & Slagboom, P. E. (1995). High-yield noninvasive human genomic DNA isolation method for genetic studies in geographically dispersed families and populations. *American journal of human genetics*, *57*(5), 1252-4.

Jeffries, A. R., Perfect, L. W., Ledderose, J., Schalkwyk, L. C., Bray, N. J., Mill, J., & Price, J. (2012). Stochastic Choice of Allelic Expression in Human Neural Stem Cells. *Stem cells*, *30*(9), 1938-47.

Jeon, Y., & Lee, J. T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. *Cell*, *146*(1), 119–33.

Johansson, C., Svensson, M., Wallstedt, L., Janson, A. M., & Frisen, J. (1999). Neural stem cells in the adult human brain. *Experimental cell research, 253(2)*, 733–736.

Johansson, S., Price, J., & Modo, M. (2008). Effect of inflammatory cytokines on major histocompatibility complex expression and differentiation of human neural stem/progenitor cells. *Stem cells*, *26*(9), 2444–54.

Johnson, A. D., Handsaker, R. E., Pulit, S. L., Nizzari, M. M., O'Donnell, C. J., & De Bakker, P. I. W. (2008). SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*, *24*(24), 2938–9.

Kaneko, R., Kato, H., Kawamura, Y., Esumi, S., Hirayama, T., Hirabayashi, T., & Yagi, T. (2006). Allelic gene regulation of Pcdh-alpha and Pcdh-gamma clusters involving both monoallelic and biallelic expression in single Purkinje cells. *The Journal of biological chemistry*, *281*(41), 30551–60.

Keshet, I., Lieman-Hurwitz, J., & Cedar, H. (1986). DNA methylation affects the formation of active chromatin. *Cell*, *44*(4), 535–43.

Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., & Cahan, P. (2010). Epigenetic memory in induced pluripo- tent stem cells. *Nature*, *467*(7313), 285–290.

Kishino, T., Lalande, M., & Wagstaff, J. (1997). UBE3A/E6-AP mutations cause Angelman syndrome. *Nature genetics, 15*(1), 70-73.

Koch, P., Opitz, T., Steinbeck, J. a, Ladewig, J., & Brüstle, O. (2009). A rosette-type, self-renewing human ES cell-derived neural stem cell with potential for in vitro instruction and synaptic integration. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(9), 3225–30.

Kohmura, N., Senzaki, K., Hamada, S., Kai, N., Yasuda, R., Watanabe, M., Ishii, H., et al. (1998). Diversity revealed by a novel family of cadherins expressed in neurons at a synaptic complex. *Neuron*, *20*(6), 1137–51.

Kuhn, H., Dickinson-Anson, H., & Gage, F. (1996). Neurogenesis in the Dentate Gyrus of the Adult Rat: Age-Related Decrease of Neuronal Progenitor Proliferation. *The Journal of neuroscience*, *76*(6), 2027–2033.

Kukekov, V. G., Laywell, E. D., Suslov, O., Davies, K., Scheffler, B., Thomas, L. B., O'Brien, T. F., et al. (1999). Multipotent stem/progenitor cells with similar properties arise from two neurogenic regions of adult human brain. *Experimental neurology*, *156*(2), 333–44.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., & Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature*, *410*(6824), 116-120.

Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C. T., Low, H. M., et al. (2010). Dynamic changes in the human methylome during differentiation. *Genome research*, *20*(3), 320–31.

Lee, J.-H., Park, I.-H., Gao, Y., Li, J. B., Li, Z., Daley, G. Q., Zhang, K., et al. (2009). A robust approach to identifying tissue-specific gene expression regulatory variants using personalized human induced pluripotent stem cells. *PLoS genetics*, *5*(11), e1000718.

Lee, M., DeBaun, M., Mitsuya, K., Galonek, J., Brandenburg, S., Oshimura, M., & Feinberg, A. (1999). Paternally expressed transcript, with antisense orientation to KVLQT1, occurs frequently in Beckwith–Wiedemann syndrome and is independent of insulin-like growth factor II imprinting. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(9), 5203–5208.

Lee, P. D., Sladek, R., Greenwood, C. M. T., & Hudson, T. J. (2002). Control genes and variability: absence of ubiquitous reference transcripts in diverse mammalian expression studies. *Genome research*, *12*(2), 292–7.

Lewcock, J. W., & Reed, R. R. (2004). A feedback mechanism regulates monoallelic odorant receptor expression. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(4), 1069–74.

Li, E., Bestor, T. H., & Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, *69*(6), 915–26.

Li, S. M., Valo, Z., Wang, J., Gao, H., Bowers, C. W., & Singer-Sam, J. (2012). Transcriptome-wide survey of mouse CNS-derived cells reveals monoallelic expression within novel gene families. *PloS one*, *7*(2), e31751.

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, *462*(7271), 315–22.

Lister, R., Pelizzola, M., Kida, Y. S., Hawkins, R. D., Nery, J. R., Hon, G., Antosiewicz-Bourget, J., et al. (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature*, *471*(7336), 68–73.

Littlewood, T. D., Hancock, D. C., Danielian, P. S., Parker, M. G., & Evan, G. I. (1995). A modified oestrogen receptor ligand-binding domain as an improved switch for the regulation of heterologous proteins. *Nucleic acids research*, *23*(10), 1686–90.

Lo, H. S., Wang, Z., Hu, Y., Yang, H. H., Gere, S., Buetow, K. H., & Lee, M. P. (2003). Allelic variation in gene expression is common in the human genome. *Genome research, 13*(8), 1855-1862.

Lomvardas, S., Barnea, G., Pisapia, D. J., Mendelsohn, M., Kirkland, J., & Axel, R. (2006). Interchromosomal interactions and olfactory receptor choice. *Cell*, *126*(2), 403–13.

Lyon, M. (1961). Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature*, *190*, 372–373.

Magklara, A., Yen, A., Colquitt, B. M., Clowney, E. J., Allen, W., Markenscoff-Papadimitriou, E., Evans, Z. a, et al. (2011). An epigenetic signature for monoallelic olfactory receptor expression. *Cell*, *145*(4), 555–70.

Main, B. J., Bickel, R. D., McIntyre, L. M., Graze, R. M., Calabrese, P. P., & Nuzhdin, S. V. (2009). Allele-specific expression assays using Solexa. *BMC genomics*, *10*, 422.

Mak, W., Nesterova, T. B., De Napoles, M., Appanah, R., Yamanaka, S., Otte, A. P., & Brockdorff, N. (2004). Reactivation of the paternal X chromosome in early mouse embryos. *Science, 303*(5658), 666–9.

Malatesta, P, Hartfuss, E., & Götz, M. (2000). Isolation of radial glial cells by fluorescent-activated cell sorting reveals a neuronal lineage. *Development*, *127*(24), 5253–63.

Malatesta, Paolo, Hack, M. a, Hartfuss, E., Kettenmann, H., Klinkert, W., Kirchhoff, F., & Götz, M. (2003). Neuronal or glial progeny: regional differences in radial glia fate. *Neuron*, *37*(5), 751–64.

Mancini-Dinardo, D., Steele, S. J. S., Levorse, J. M., Ingram, R. S., & Tilghman, S. M. (2006). Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. *Genes & development*, *20*(10), 1268–82.

Marks, H., Chow, J. C., Denissov, S., Françoijs, K.-J., Brockdorff, N., Heard, E., & Stunnenberg, H. G. (2009). High-resolution analysis of epigenetic changes associated with X inactivation. *Genome research*, *19*(8), 1361–73.

Masui, O., Bonnet, I., Le Baccon, P., Brito, I., Pollex, T., Murphy, N., Hupé, P., et al. (2011). Live-cell chromosome dynamics and outcome of X chromosome pairing events during ES cell differentiation. *Cell*, *145*(3), 447–58.

Meaney, M. J. (2010). Epigenetics and the biological definition of gene x environment interactions. *Child development*, *81*(1), 41–79.

Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, *454*(7205), 766–70.

Merkle, F. T., Tramontin, A. D., García-Verdugo, J. M., & Alvarez-Buylla, A. (2004). Radial glia give rise to adult neural stem cells in the subventricular zone. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(50), 17528–32.

Muers, M. R., Sharpe, J. A., Garrick, D., Sloane-Stanley, J., Nolan, P. M., Hacker, T., ... & Gibbons, R. J. (2007). Defining the cause of skewed X-chromosome inactivation in X-linked mental retardation by use of a mouse model. *The American journal of human genetics, 80*(6), 1138-1149.

Moore, T., & Haig, D. (1991). Genomic imprinting in mammalian development: a parental tug-of-war. *Trends in genetics*, *7*(2), 45–9.

Morcos, L., Ge, B., Koka, V., Lam, K. C. L., Pokholok, D. K., Gunderson, K. L., Montpetit, A., et al. (2011). Genome-wide assessment of imprinted expression in human cells. *Genome biology*, *12*(3), R25.

Morison, I. M., Ramsay, J. P., & Spencer, H. G. (2005). A census of mammalian imprinting. *Trends in genetics*, *21*(8), 457–65.

Mostoslavsky, R, Singh, N., & Tenzen, T. (2001). Asynchronous replication and allelic exclusion in the immune system. *Nature*, *414*, 221–225.

Mostoslavsky, Raul, Singh, N., Kirillov, A., Pelanda, R., Cedar, H., Chess, A., & Bergman, Y. (1998). k chain monoallelic demethylation and the establishment of allelic exclusion. *Genes & development, 12*(12), 1801–1811.

Murata, Y., Hamada, S., Morishita, H., Mutoh, T., & Yagi, T. (2004). Interaction with protocadherin-gamma regulates the cell surface expression of protocadherin-alpha. *The Journal of biological chemistry*, *279*(47), 49508–16.

Murrell, A., Heeson, S., & Reik, W. (2004). Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nature genetics*, *36*(8), 889–93.

Nalls, M. a, Plagnol, V., Hernandez, D. G., Sharma, M., Sheerin, U.-M., Saad, M., Simón-Sánchez, J., et al. (2011). Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet*, *377*(9766), 641–9.

Norton, N., Williams, N. M., Williams, H. J., Spurlock, G., Kirov, G., Morris, D. W., Hoogendoorn, B., et al. (2002). Universal, robust, highly quantitative SNP allele frequency measurement in DNA pools. *Human genetics*, *110*(5), 471–8.

Nozari, G., Rahbar, S., & Wallace, R. (1986). Discrimination among the transcripts of the allelic human beta-globin genes beta A, beta S and beta C using oligodeoxynucleotide hybridization probes. *Gene*, *43*, 23–28.

Ohlsson, R., Cui, H., He, L., Tumorigenesis, W., & Heterogeneity, E. (1999). Mosaic allelic insulin-like growth factor 2 expression patterns reveal a link between Wilms' tumorigenesis and epigenetic heterogeneity. *Cancer research, 59*(16), 3889–3892.

Ohta, T., Gray, T. a, Rogan, P. K., Buiting, K., Gabriel, J. M., Saitoh, S., Muralidhar, B., et al. (1999). Imprinting-mutation mechanisms in Prader-Willi syndrome. *American journal of human genetics*, *64*(2), 397–413.

Okamoto, I., Patrat, C., Thépot, D., Peynot, N., Fauque, P., Daniel, N., Diabangouaya, P., et al. (2011). Eutherian mammals use diverse strategies to initiate X-chromosome inactivation during development. *Nature*, *472*(7343), 370–4.

Pant, P. V. K., Tao, H., Beilharz, E. J., Ballinger, D. G., Cox, D. R., & Frazer, K. a. (2006). Analysis of allelic differential expression in human white blood cells. *Genome research*, *16*(3), 331–9.

Papapetrou, E. P., Lee, G., Malani, N., Setty, M., Riviere, I., Tirunagari, L. M. S., Kadota, K., et al. (2011). Genomic safe harbors permit high β-globin transgene expression in thalassemia induced pluripotent stem cells. *Nature biotechnology*, *29*(1), 73–8.

Pastinen, T., Sladek, R., Gurd, S., Sammak, A., Ge, B., Lepage, P., Lavergne, K., et al. (2004). A survey of genetic and epigenetic variation affecting human gene expression. *Physiological genomics*, *16*(2), 184–93.

Payer, B., & Lee, J. T. (2008). X chromosome dosage compensation: how mammals keep the balance. *Annual review of genetics*, *42*, 733–72.

Pernis, B., & Chiappino, G. (1965). Cellular localization of immunoglobulins with different allotypic specificities in rabbit lymphoid tissues. *The Journal of experimental medicine, 122*(5), 853-876.

Perry, R. P., Kelley, D. E., Coleclough, C., Seidman, J. G., Leder, P., Tonegawa, S., Matthyssens, G., et al. (1980). Transcription of mouse K chain genes: Implications for allelic exclusion. *Proceedings of the National Academy of Sciences of the United States of America*, *77*(4), 1937–41.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic acids research*, *29*(9), e45.

Pinto, L., Mader, M. T., Irmler, M., Gentilini, M., Santoni, F., Drechsel, D., Blum, R., et al. (2008). Prospective isolation of functionally distinct radial glial subtypes--lineage and transcriptome analysis. *Molecular and cellular neurosciences*, *38*(1), 15–42.

Plagge, a, & Kelsey, G. (2006). Imprinting the Gnas locus. *Cytogenetic and genome research*, *113*(1-4), 178–87.

Plenge, R. M., Stevenson, R. A., Lubs, H. A., Schwartz, C. E., & Willard, H. F. (2002). Report Skewed X-Chromosome Inactivation Is a Common Feature of X-Linked Mental Retardation Disorders, 168–173.

Pollard, K. S., Serre, D., Wang, X., Tao, H., Grundberg, E., Hudson, T. J., Clark, A. G., et al. (2008). A genome-wide approach to identifying novel-imprinted genes. *Human genetics*, *122*(6), 625–34.

Pollard, S. M., Conti, L., Sun, Y., Goffredo, D., & Smith, A. (2006). Adherent neural stem (NS) cells from fetal and adult forebrain. *Cerebral cortex*, *16(1)* 112–20.

Pollock, K., Stroemer, P., Patel, S., Stevanato, L., Hope, A., Miljan, E., Dong, Z., et al. (2006). A conditionally immortal clonal stem cell line from human cortical neuroepithelium for the treatment of ischemic stroke. *Experimental neurology*, *199*(1), 143–55.

Prickett, A. R., & Oakey, R. J. (2012). A survey of tissue-specific genomic imprinting in mammals. *Molecular genetics and genomics : MGG*, *287*(8), 621–30.

Quandt, K., Frech, K., Karas, H., Wingender, E., & Werner, T. (1995). MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic acids research*, *23*(23), 4878–84.

Redrup, L., Branco, M. R., Perdeaux, E. R., Krueger, C., Lewis, A., Santos, F., Nagano, T., et al. (2009). The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing. *Development*, *136*(4), 525–30.

Reik, W, Brown, K., & Schneid, H. (1995). Imprinting mutations in the Beckwith-Wiedemann syndrome suggested by an altered imprinting pattern in the IGF2-H19 domain. *Human molecular genetics, 4*(12), 2379-2385.

Reik, Wolf. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature*, *447*(7143), 425–32.

Reynolds, B. a, & Rietze, R. L. (2005). Neural stem cells and neurospheres--re-evaluating the relationship. *Nature methods*, *2*(5), 333–6.

Reynolds, B., & Weiss, S. (1996). Clonal and population analyses demonstrate that an EGF-responsive mammalian embryonic CNS precursor is a stem cell. *Developmental biology*, *13*, 1–13.

Rifas, L., & Weitzmann, M. N. (2009). A novel T cell cytokine, secreted osteoclastogenic factor of activated T cells, induces osteoclast formation in a RANKL-independent manner. *Arthritis and rheumatism*, *60*(11), 3324–35.

Roldán, E., Fuxa, M., Chong, W., Martinez, D., Novatchkova, M., Busslinger, M., & Skok, J. a. (2005). Locus “decontraction” and centromeric recruitment contribute to allelic exclusion of the immunoglobulin heavy-chain gene. *Nature immunology*, *6*(1), 31–41.

Roy, N. S., Wang, S., Jiang, L., Kang, J., Benraiss, a, Harrison-Restelli, C., Fraser, R. a, et al. (2000). In vitro neurogenesis by progenitor cells isolated from the adult human hippocampus. *Nature medicine*, *6*(3), 271–7.

Rozowsky, J., Abyzov, A., Wang, J., Alves, P., Raha, D., Harmanci, A., Leng, J., et al. (2011). AlleleSeq: analysis of allele-specific expression and binding in a network framework. *Molecular systems biology*, *7*(522), 522.

Samuels, B. a, & Hen, R. (2011). Neurogenesis and affective disorders. *The European journal of neuroscience*, *33*(6), 1152–9.

Sanger, F., Nicklen, S., & Coulson, A. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America, 74*(12), 5463–5467.

Satake, W., Nakabayashi, Y., Mizuta, I., Hirota, Y., Ito, C., Kubo, M., Kawaguchi, T., et al. (2009). Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nature genetics*, *41*(12), 1303–7.

Sayegh, C. E., Sayegh, C., Jhunjhunwala, S., Riblet, R., & Murre, C. (2005). Visualization of looping involving the immunoglobulin heavy-chain locus in developing B cells. *Genes & development*, *19*(3), 322–7.

Schalkwyk, L. C., Meaburn, E. L., Smith, R., Dempster, E. L., Jeffries, A. R., Davies, M. N., Plomin, R., et al. (2010). Allelic Skewing of DNA Methylation Is Widespread across the Genome. *The American Journal of Human Genetics*, 196–212.

Schreiner, D., & Weiner, J. a. (2010). Combinatorial homophilic interaction between gamma-protocadherin multimers greatly expands the molecular diversity of cell adhesion. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(33), 14893–8.

Scott, R. J., & Spielman, M. (2006). Genomic imprinting in plants and mammals: how life history constrains convergence. *Cytogenetic and genome research*, *113*(1-4), 53–67.

Serizawa, S., Miyamichi, K., Nakatani, H., Suzuki, M., Saito, M., Yoshihara, Y., & Sakano, H. (2003). Negative feedback regulation ensures the one receptor-one olfactory neuron rule in mouse. *Science*, *302*(5653), 2088–94.

Serre, D., Gurd, S., Ge, B., Sladek, R., Sinnett, D., Harmsen, E., Bibikova, M., et al. (2008). Differential allelic expression in the human genome: a robust approach to identify genetic and epigenetic cis-acting mechanisms regulating gene expression. *PLoS genetics*, *4*(2), e1000006.

Shogren-Knaak, M., Ishii, H., Sun, J.-M., Pazin, M. J., Davie, J. R., & Peterson, C. L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science*, *311*(5762), 844–7.

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T. B., Webster, Z., Peters, A. H. F. M., et al. (2003). Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Developmental cell*, *4*(4), 481–95.

Sims, R., & Chen, C. (2005). Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. *Journal of Biological Chemistry*, *280*(51), 41789–41792.

Simón-Sánchez, J., Schulte, C., Bras, J. M., Sharma, M., Gibbs, J. R., Berg, D., Paisan-Ruiz, C., et al. (2009). Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nature genetics*, *41*(12), 1308–12.

Singer-Sam, J., LeBon, J. M., Dai, a, & Riggs, a D. (1992). A sensitive, quantitative assay for measurement of allele-specific transcripts differing by a single nucleotide. *Genome Research*, *1*(3), 160–163.

Skelly, D. A., Johansson, M., Madeoy, J., Wakefield, J., & Akey, J. M. (2011). A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome research, 21*(10), 1728–1737.

Sleutels, F., Zwart, R., & Barlow, D. P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature*, *415*(6873), 810–3.

Smrzka, O. W., Faé, I., Stöger, R., Kurzbauer, R., Fischer, G. F., Henn, T., Weith, a, et al. (1995). Conservation of a maternal-specific methylation signal at the human IGF2R locus. *Human molecular genetics*, *4*(10), 1945–52.

Stadtfeld, M., Apostolou, E., Akutsu, H., Fukuda, A., Follett, P., Natesan, S., Kono, T., et al. (2010). Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature*, *465*(7295), 175–81.

Stanton, M., & Brodeur, P. (2005). Stat5 Mediates the IL-7-Induced Accessibility of a Representative D-Distal V H Gene. *The Journal of Immunology, 174*(6), 3164-3168.

Stephens, M., & Donnelly, P. (2003). A comparison of bayesian methods for haplotype reconstruction from population genotype data. *American journal of human genetics*, *73*(5), 1162–9.

Stephens, M., Smith, N. J., & Donnelly, P. (2001). A New Statistical Method for Haplotype Reconstruction from Population Data. *Society*, 978–989.

Surani, M., Barton, S., & Norris, M. (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature, 308*(5959), 548-550.

Suslov, O. N., Kukekov, V. G., Ignatova, T. N., & Steindler, D. A. (2002). Neural stem cell heterogeneity demonstrated by molecular phenotyping of clonal neurospheres. *Proceedings of the National Academy of Sciences of the United States of America, 99*(22), 14506-14511.

Sutcliffe, J., & Nakao, M. (1994). Deletions of a differentially methylated CpG island at the SNRPN gene define a putative imprinting control region. *Nature genetics,8*(1), 52-58.

Takahashi, K., & Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, *126*(4), 663–676.

Tan, E.-K., Kwok, H.-H., Kwok, H.-K., Tan, L. C., Zhao, W.-T., Prakash, K. M., Au, W.-L., et al. (2010). Analysis of GWAS-linked loci in Parkinson disease reaffirms PARK16 as a susceptibility locus. *Neurology*, *75*(6), 508–12.

Tang, F., Barbacioru, C., Nordman, E., Li, B., Xu, N., Bashkirov, V. I., Lao, K., et al. (2010). RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nature protocols*, *5*(3), 516–35.

Tattersfield, a S., Croon, R. J., Liu, Y. W., Kells, a P., Faull, R. L. M., & Connor, B. (2004). Neurogenesis in the striatum of the quinolinic acid lesion model of Huntington's disease. *Neuroscience*, *127*(2), 319–32.

Terranova, R., Yokobayashi, S., Stadler, M. B., Otte, A. P., Van Lohuizen, M., Orkin, S. H., & Peters, A. H. F. M. (2008). Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Developmental cell*, *15*(5), 668–79.

Uchida, N., Buck, D. W., He, D., Reitsma, M. J., Masek, M., Phan, T. V, Tsukamoto, a S., et al. (2000). Direct isolation of human central nervous system stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(26), 14720–5.

Vakoc, C. R., Mandat, S. a, Olenchock, B. a, & Blobel, G. a. (2005). Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. *Molecular cell*, *19*(3), 381–91.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., & Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome biology*, *3*(7), R0034.

Wagner, J. R., Ge, B., Pokholok, D., Gunderson, K. L., Pastinen, T., & Blanchette, M. (2010). Computational analysis of whole-genome differential allelic expression data in human. *PLoS computational biology*, *6*(7), e1000849.

Wang, J., Valo, Z., Bowers, C. W., Smith, D. D., Liu, Z., & Singer-Sam, J. (2010). Dual DNA methylation patterns in the CNS reveal developmentally poised chromatin and monoallelic expression of critical genes. *PloS one*, *5*(11), e13843.

Wang, J., Valo, Z., Smith, D., & Singer-Sam, J. (2007). Monoallelic expression of multiple genes in the CNS. *PloS one*, *2*(12), e1293.

Wang, X., Weiner, J. a, Levi, S., Craig, A. M., Bradley, A., & Sanes, J. R. (2002). Gamma protocadherins are required for survival of spinal interneurons. *Neuron*, *36*(5), 843–54.

Wei, Y., Mizzen, C. a, Cook, R. G., Gorovsky, M. a, & Allis, C. D. (1998). Phosphorylation of histone H3 at serine 10 is correlated with chromosome condensation during mitosis and meiosis in Tetrahymena. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(13), 7480–4.

Wolf, J. B., & Hager, R. (2006). A maternal-offspring coadaptation theory for the evolution of genomic imprinting. *PLoS biology*, *4*(12), e380.

Wong, C. C. Y., Caspi, A., Williams, B., Houts, R., Craig, I. W., & Mill, J. (2011). A longitudinal twin study of skewed X chromosome-inactivation. *PloS one*, *6*(3), e17873.

Wu, D. Y., Ugozzoli, L., Pal, B. K., & Wallace, R. B. (1989). Allele-specific enzymatic amplification of beta-globin genomic DNA for diagnosis of sickle cell anemia. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(8), 2757–60.

Wu, Q., & Maniatis, T. (1999). A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell*, *97*(6), 779–90.

Wutz, a, & Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Molecular cell*, *5*(4), 695–705.

Wutz, a, Smrzka, O. W., Schweifer, N., Schellander, K., Wagner, E. F., & Barlow, D. P. (1997). Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature*, *389*(6652), 745–9.

Xu, N., Tsai, C.-L., & Lee, J. T. (2006). Transient homologous chromosome pairing marks the onset of X inactivation. *Science*, *311*(5764), 1149–52.

Yan, H., Yuan, W., Velculescu, V. E., Vogelstein, B., & Kinzler, K. W. (2002). Allelic variation in human gene expression. *Science*, *297*(5584), 1143.

Yan, Y.-P., Mo, X.-Y., Tian, J., Zhao, G.-H., Yin, X.-Z., Jin, F.-Y., & Zhang, B.-R. (2011). An association between the PARK16 locus and Parkinson's disease in a cohort from eastern China. *Parkinsonism & related disorders*, *17*(10), 737–9.

Zhang, K., Li, J. B., Gao, Y., Egli, D., Xie, B., Deng, J., Li, Z., et al. (2009). Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human. *Nature methods*, *6*(8), 613–8.

Zhang, L.-F., Huynh, K. D., & Lee, J. T. (2007). Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. *Cell*, *129*(4), 693–706.

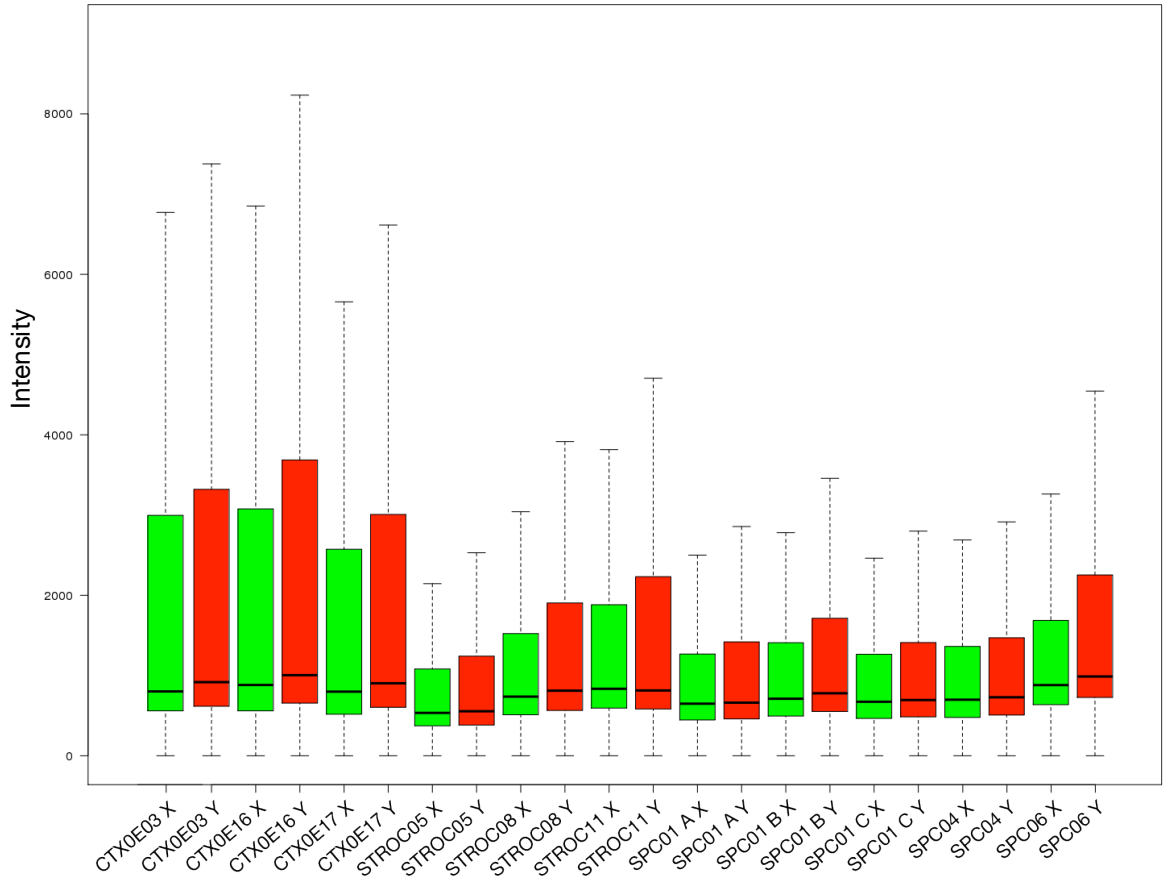
Zhao, X., Li, W., Lv, Z., Liu, L., Tong, M., Hai, T., Hao, J., et al. (2009). iPS cells produce viable mice through tetraploid complementation. *Nature*, *461*(7260), 86–90.

Zhu, G., Lipsky, R. H., Xu, K., Ali, S., Hyde, T., Kleinman, J., Akhtar, L. a, et al. (2004). Differential expression of human COMT alleles in brain and lymphoblasts detected by RT-coupled 5' nuclease assay. *Psychopharmacology*, *177*(1-2), 178–84.

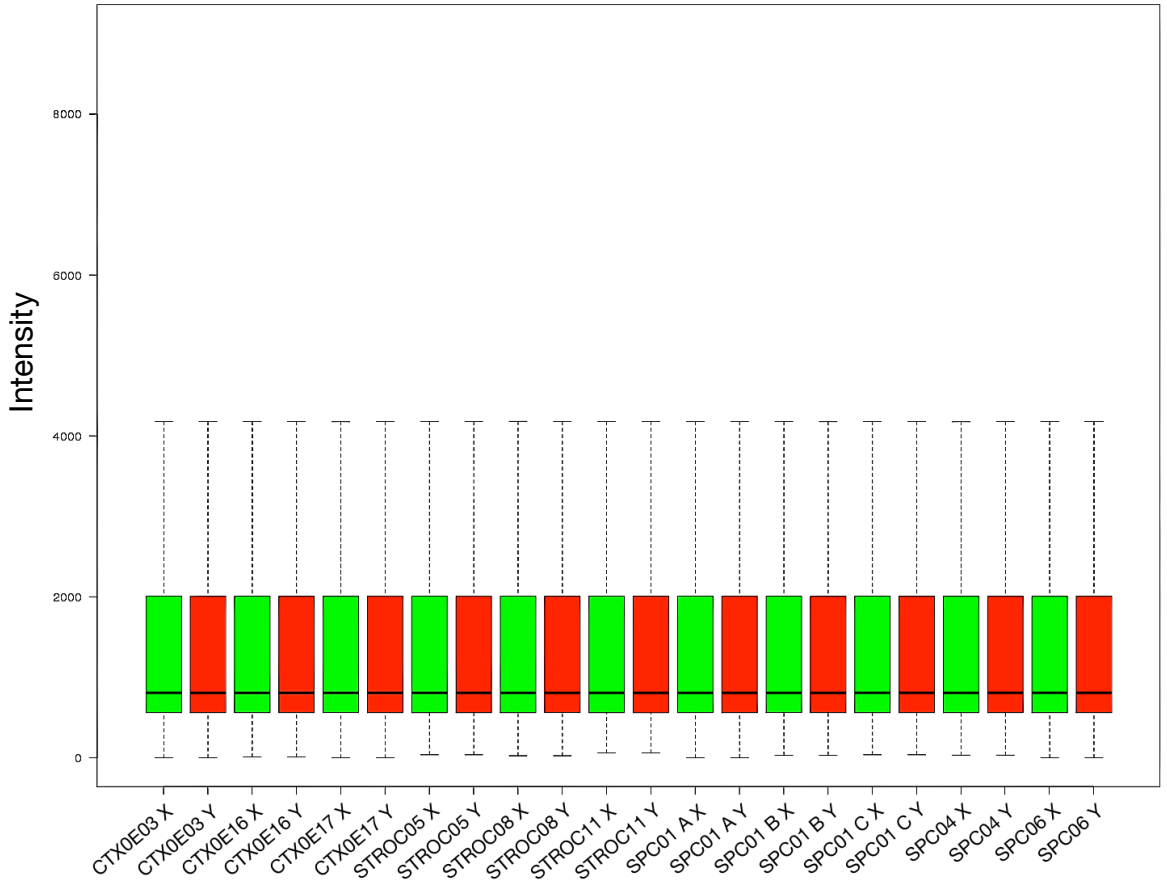
Zipursky, S. L., & Sanes, J. R. (2010). Chemoaffinity revisited: dscams, protocadherins, and neural circuit assembly. *Cell*, *143*(3), 343–53.

Zwemer, L. M., Zak, A., Thompson, B., Kirby, A., Daly, M. J., Chess, A., & Gimelbrant, A. a. (2012). Autosomal monoallelic expression in the mouse. *Genome biology*, *13*(2), R10.

Appendices



Appendix 9.1 | Raw SNP probe intensity values of cDNA samples from the microarray study described in Chapter 3.



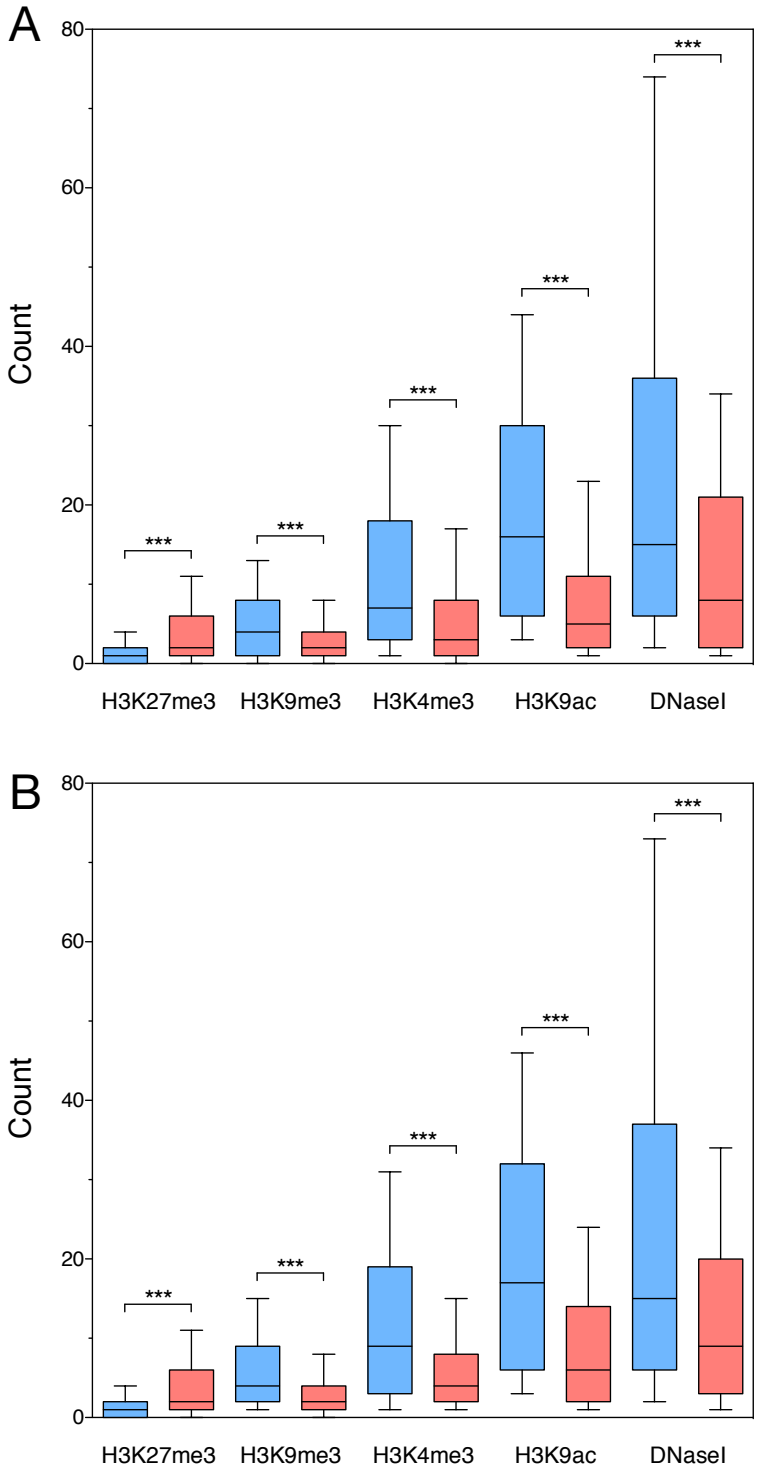
Appendix 9.2 | SNP probe intensity values of cDNA samples from the microarray study described in Chapter 3 after quantile normalisation.

Appendix 9.3 | Gene ontology terms enriched in the RME genes lists from all three donors.

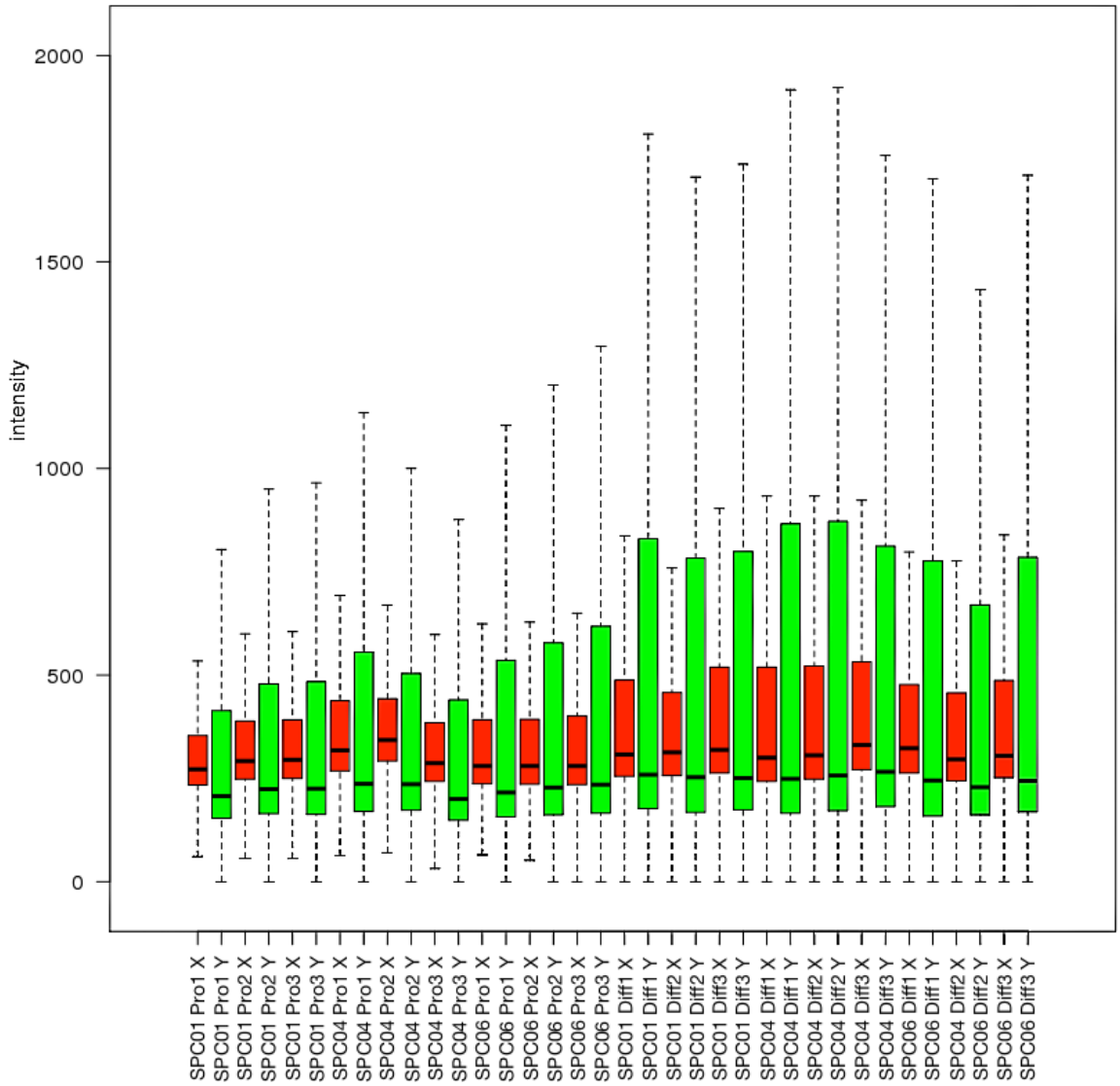
GO Term	Cortical donor				Striatal donor				Spinal cord donor			
	Count	%	Enrichment	*p* value	Count	%	Enrichment	*p* value	Count	%	Enrichment	*p* value
Transmembrane	26	30.95	1.41	0.048	37	35.92	1.62	0.0015	14	33.33	1.67	0.046
Extracellular	16	19.05	1.88	0.018	19	18.45	1.83	0.013	11	26.19	2.81	0.0032
Cytoplasmic	22	26.19	1.88	0.0038	27	26.21	1.88	0.0012	13	30.95	2.43	0.0034
Glycoprotein	33	39.29	2.2	5.58E-06	34	33.01	1.88	0.00018	13	30.95	1.96	0.019
Signal	23	27.38	2.07	0.00084	24	23.3	1.84	0.0037	11	26.19	2.28	0.014
Glycosylation site:N-linked	33	39.29	2.31	1.83E-06	34	33.01	1.96	0.000082	13	30.95	2.04	0.014

Appendix 9.4 | Gene ontology terms enriched in the RME genes lists from the two brain derived NSC lines.

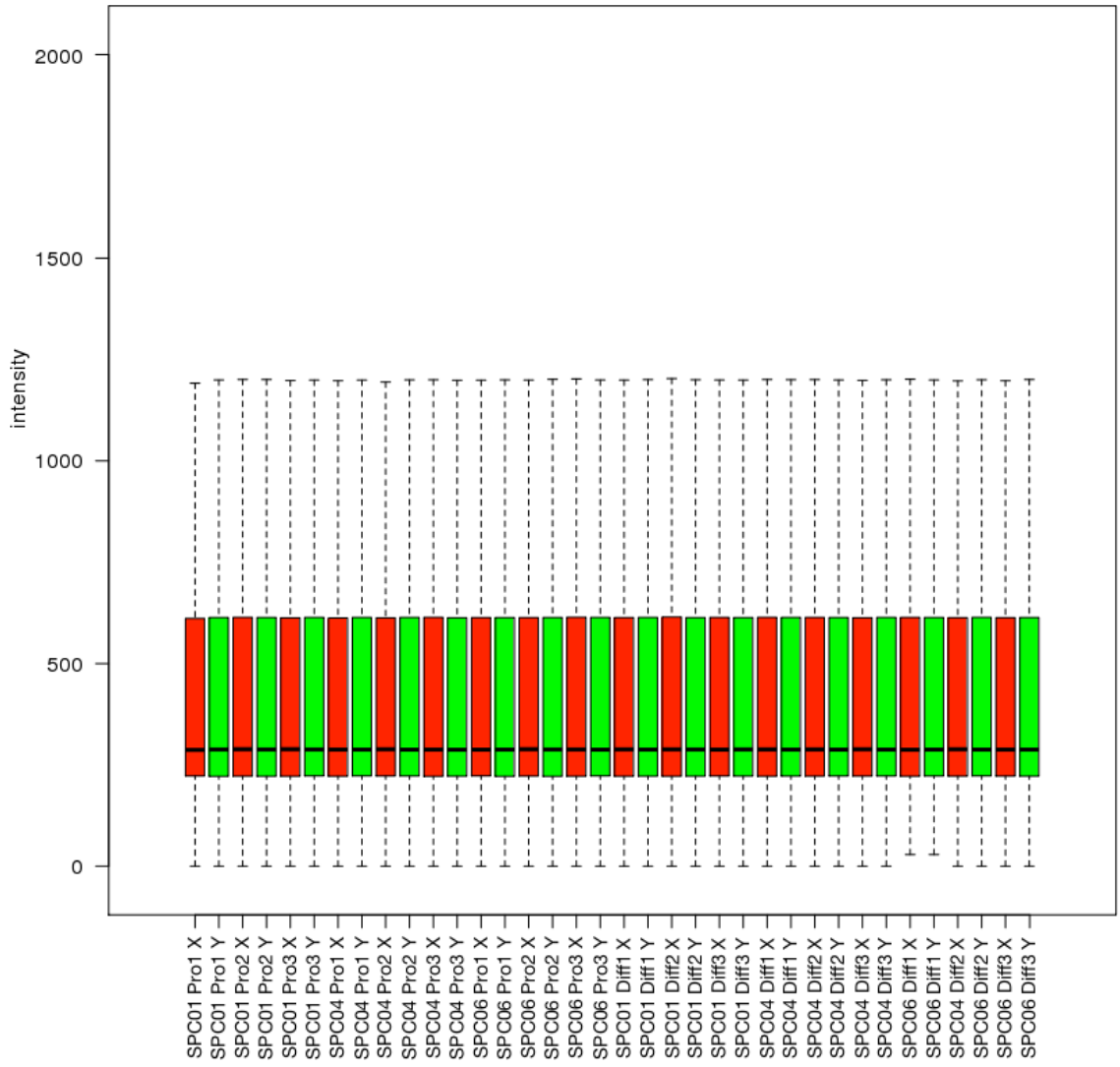
GO Term	Cortical donor				Striatal donor			
	Count	%	Enrichment	*p* value	Count	%	Enrichment	*p* value
Anatomical structure development	21	25	1.91	0.0031	20	19.42	1.62	0.027
Cellular component morphogenesis	8	9.52	3.77	0.0045	8	7.77	3.42	0.0079
System development	19	22.62	1.9	0.0059	19	18.45	1.7	0.021
Anatomical structure morphogenesis	13	15.48	2.36	0.0064	12	11.65	1.95	0.037
Cell morphogenesis	6	7.14	3.11	0.04	8	7.77	3.73	0.0049
Cell morphogenesis involved in differentiation	5	5.95	3.75	0.042	7	6.8	4.86	0.0028
Membrane	45	53.57	1.54	0.000097	50	48.54	1.34	0.005
Intrinsic to membrane	28	33.33	1.4	0.036	41	39.81	1.58	0.00069
Transmembrane	26	30.95	1.41	0.048	37	35.92	1.62	0.0015
Extracellular	16	19.05	1.88	0.018	19	18.45	1.83	0.013
Cytoplasmic	22	26.19	1.88	0.0038	27	26.21	1.88	0.0012
Glycoprotein	33	39.29	2.2	5.58E-06	34	33.01	1.88	0.00018
Signal	23	27.38	2.07	0.00084	24	23.3	1.84	0.0037
Glycosylation site:N-linked	33	39.29	2.31	1.8E-06	34	33.01	1.96	0.000082
Cell adhesion	8	9.52	4.61	0.0016	6	5.83	3.04	0.046
Disulfide bond	18	21.43	2.05	0.0047	20	19.42	1.93	0.0057



Appendix 9.5 | Histone modification and DNaseI sensitivity of novel RME genes in human foetal brain. Chromatin data for biallelic (blue) and RME (red) genes detected in the spinal cord (**A**) and striatal (**B**) cell lines. Whiskers represent the 10-90th percentile. T-test *P* value summaries are presented.



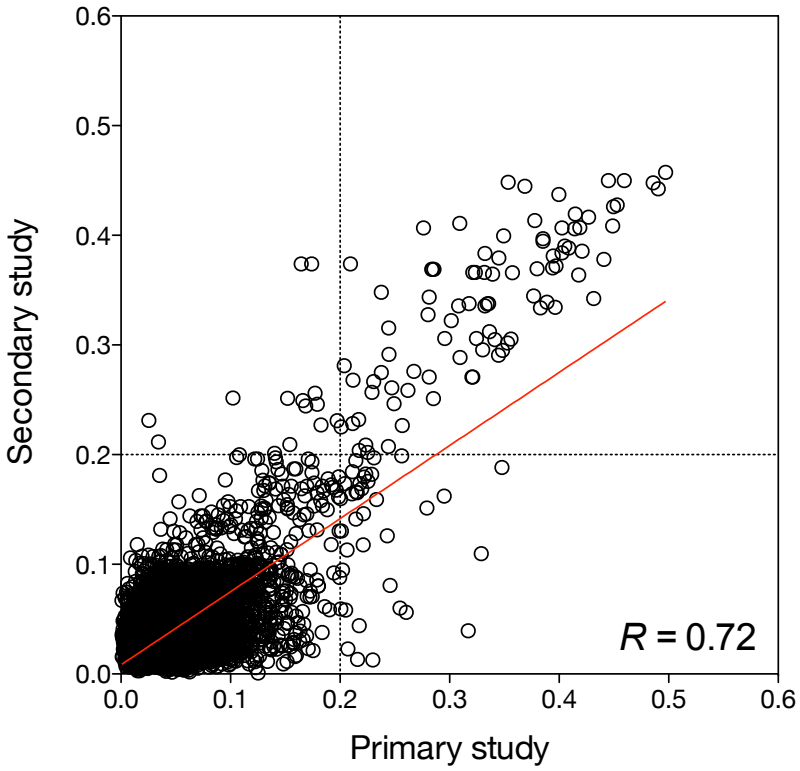
Appendix 9.6 | Raw SNP probe intensity values of cDNA samples from the microarray study described in Chapter 4.



Appendix 9.7 | SNP probe intensity values of cDNA samples from the microarray study described in Chapter 4 after quantile normalisation.

Appendix 9.8 | Gene ontology terms enriched in the monoallelic genes detected in the differentiated spinal cord lines.

GO Term	Differentiated spinal cord lines			
	Count	%	Enrichment	*p* value
Extracellular region	11	16.18	2.46	0.009
Glycoprotein	22	32.35	1.85	0.0035
Signal	19	27.94	2.20	0.0012
Disulfide bond	18	26.47	2.77	0.00012
Polymorphism	53	77.94	1.32	0.00042
Sequence variant	54	79.41	1.30	0.00046
Glycosylation site:N-linked	22	32.35	1.94	0.002
Cleavage on pair of basic residues	6	8.23	8.00	0.00081



Appendix 9.9 | Correlation of Δ_β values from all autosomal genes assayed in the three spinal cord NSC lines in the primary microarray study reported in Chapter 3 and the secondary study reported in Chapter 4. A good correlation is observed between the two datasets ($p < 0.0001$, $R = 0.72$).