

## Stochastic Choice of Allelic Expression in Human Neural Stem Cells

AARON R. JEFFRIES,<sup>a</sup> LEO W. PERFECT,<sup>a</sup> JULIA LEDDEROSE,<sup>a</sup> LEONARD C. SCHALKWYK,<sup>b</sup>  
NICHOLAS J. BRAY,<sup>a</sup> JONATHAN MILL,<sup>b</sup> JACK PRICE<sup>a</sup><sup>a</sup>King's College London, Institute of Psychiatry, Centre for the Cellular Basis of Behaviour, Department of Neuroscience and <sup>b</sup>King's College London, Institute of Psychiatry, MRC SGDP Research Centre, London, United Kingdom**Key Words.** Neural stem cells • Epigenomics • Allelic imbalance • Monoallelic expression • DNA methylation • Brain

## ABSTRACT

Monoallelic gene expression, such as genomic imprinting, is well described. Less well-characterized are genes undergoing stochastic monoallelic expression (MA), where specific clones of cells express just one allele at a given locus. We performed genome-wide allelic expression assessment of human clonal neural stem cells derived from cerebral cortex, striatum, and spinal cord, each with differing genotypes. We assayed three separate clonal lines from each donor, distinguishing stochastic MA from genotypic effects. Roughly 2% of genes showed evidence for autosomal MA, and in about half of these, allelic expression was

stochastic between different clones. Many of these loci were known neurodevelopmental genes, such as *OTX2* and *OLIG2*. Monoallelic genes also showed increased levels of DNA methylation compared to hypomethylated biallelic loci. Identified monoallelic gene loci showed altered chromatin signatures in fetal brain, suggesting an *in vivo* correlate of this phenomenon. We conclude that stochastic allelic expression is prevalent in neural stem cells, providing clonal diversity to developing tissues such as the human brain. *STEM CELLS* 2012;30:1938–1947

Disclosure of potential conflicts of interest is found at the end of this article.

## INTRODUCTION

Gene expression in diploid eukaryotic cells is generally biallelic, with transcription from both parental alleles. Classic exceptions are imprinted genes that show monoallelic expression (MA) in a parent-of-origin-specific manner [1], X-chromosome inactivation in mammalian females [2], and allelic imbalance through genetic variation in *cis* [3]. An additional form of MA occurs randomly, with individual cells expressing either one or both of the parental alleles. Like genomic imprinting, this stochastic choice of allelic expression occurs at the gene level rather than in a chromosome-specific manner and appears to be stably maintained in cellular progeny [4]. Stochastic choice of allelic expression allows the self-identity of individual cells and also yields potential functional variation within individual cells of a complex tissue. Stochastic MA has classically been described in a small number of gene families such as odorant receptors [5], immune receptors [6],

and (in the case of the developing nervous system) alpha and gamma protocadherins [7, 8].

In a genome-wide assessment of allelic expression in clonal human lymphoblastoid cells, Gimelbrant et al. found stochastic MA to be considerably more widespread than previously believed [9]. Many of these genes were cell surface molecules and loci characterized by lineage-specific accelerated evolution. If this finding extrapolates to somatic cell populations, such as neural stem cells, it could be of considerable functional importance. Imprinted genes are known to have important roles in the human brain [10], and recent studies suggest a large number of brain region-specific and cell-specific imprinted genes in the adult mouse brain [11, 12]. Random MA, if it exists in the developing nervous system, would add significant cell-cell variation within a system that commonly uses cues from neighboring cells for development. Clonal neural stem cell populations are an ideal resource to test this hypothesis. However, to our knowledge, no genome-wide allelic expression assessment of clonally derived human

Author contributions: A.R.J.: conception and design, provision of study material or patients, collection and/or assembly of data, data analysis and interpretation, manuscript writing, and final approval of manuscript; L.W.P.: provision of study material or patients, collection and/or assembly of data, data analysis and interpretation, and final approval of manuscript; J.L.: provision of study material or patients and final approval of manuscript; L.C.S.: data analysis and interpretation and final approval of manuscript; N.J.B.: conception and design, data analysis and interpretation, manuscript writing, and final approval of manuscript; J.M.: conception and design, financial support, collection and/or assembly of data, data analysis and interpretation, manuscript writing, and final approval of manuscript; J.P.: conception and design, financial support, data analysis and interpretation, manuscript writing, and final approval of manuscript.

Correspondence: Aaron R. Jeffries, Ph.D., King's College London, Institute of Psychiatry, Department of Neuroscience, CCBB, The James Black Centre, 125 Coldharbour Lane, London SE5 9NU, U.K. Telephone: +44(0)20-7848-0412; Fax: +44(0)20-7848-0986; e-mail: Aaron.Jeffries@kcl.ac.uk Received February 9, 2012; accepted for publication May 24, 2012; first published online in *STEM CELLS EXPRESS* June 19, 2012; available online without subscription through the open access option. © AlphaMed Press 1066-5099/2012/\$30.00/0 doi: 10.1002/stem.1155

neural stem cells, or indeed any adult stem cell population, has been made.

In this study, we assessed the frequency of stochastic choice of allelic expression in human neural cells by performing a global allelic expression analysis on a series of conditionally immortalized clones derived from human fetal brain. These neural cells are generated by the transduction of human fetal brain tissue with a c-mycER<sup>TAM</sup> construct encoded in a retroviral vector [13]. In many ways, the resultant clonal cell lines accurately represent human neural stem cells. In their proliferative phase, they are self-replicative, retain their positional specification (in terms of gene expression and the specificity of the neurons they generate), and are multipotential, generating human neurons and glia in differentiating conditions [14, 15]. Importantly, they retain a stable karyotype and phenotype over extended life in vitro.

## MATERIALS AND METHODS

### Samples

Samples consisted of three clonal lines derived from three independent donors (nine cDNAs total). A cell line from each donor was used as reference genomic DNA (gDNA). The cerebral cortex clones (CTX0E03, CTX0E16, and CTX0E17) and striatal clones (STR0C05, STR0C08, and STR0C11) were kindly provided by ReNeuron Ltd., Guildford, U.K., the spinal cord lines (SPC01, SPC04, and SPC06) were created from fetal cervical spinal cord. Cells were grown as previously described in Pollock et al. and El-Akabay et al. [13, 14]. In brief, derivation consisted of primary cell isolation from 12-week-old fetal brain. Cells were expanded on laminin-coated dishes to 60% confluence, transfected with virus containing c-mycER<sup>TAM</sup> for 12 hours, and followed by neomycin selection of transfected cell colonies. Growth/proliferation of cells was maintained by the presence of 4-hydroxytamoxifen in the media. Differentiation was achieved by removal of 4-hydroxytamoxifen and growth factors.

### Immunocytochemistry

Cells were fixed with 4% paraformaldehyde for 10 minutes, washed with Tris-buffered saline (TBS), and permeabilized using TBS with 0.1% Triton X-100 and 10% normal donkey serum before overnight 4°C incubation with primary antibody in TBS and 1% normal donkey serum. Antibodies for glial fibrillary acidic protein (Millipore, Billerica, MA, www.millipore.com), Tau (Dako, Cambridgeshire, U.K.), Nestin (Abcam, Cambridge, U.K., www.abcam.com), and O1 (Sigma-Aldrich, Dorset, U.K., www.sigmaaldrich.com) were used. Cells were then washed with TBS with 0.1% Triton X-100 before 1 hour incubation with secondary antibody (Alexa Fluor 488, Life Technologies, Paisley, U.K., www.lifetechnologies.com). Preparations were then washed with TBS, nuclear stained (Hoescht 33342, Sigma-Aldrich, Dorset, U.K., www.sigmaaldrich.com), and observed on a Leica TCS SP5 confocal microscope.

### Nucleic Acid Extraction

RNA was collected and extracted using Trizol (Life Technologies, Paisley, U.K.). Five micrograms of RNA was DNase treated (DNA Turbo free, Life Technologies, Paisley, U.K., www.lifetechnologies.com), and RNA quality was assessed with an Agilent Bioanalyser ensuring RNA integrity number > 9. Quantitative PCR was performed on 200 ng DNase-treated RNA to ensure no gDNA remained. cDNA synthesis was carried out on 1 µg of RNA with random hexamers and Superscript III (Life Technologies, Paisley, U.K., www.lifetechnologies.com) at 42°C for 2 hours.

gDNA was extracted by incubating the cell pellet in sodium chloride-Tris-EDTA buffer containing 0.5% SDS with RNaseA

(10 µg/ml) for 30 minutes at 37°C followed by Proteinase K (0.2 mg/ml) addition and further incubation at 50°C for 90 minutes. Phenol/chloroform extraction was then performed.

### Allelic Expression Analysis

gDNA (750 ng) for each donor and 1/5 cDNA reaction for each clonal line were assayed on the Illumina Omni1-Quad beadchip (Illumina, San Diego, CA, www.illumina.com). Scans were imported into Illumina GenomeStudio software (v2010.1) and genotypes called using Illumina's standard cluster file and a 0.25 gencall threshold. A loss of heterozygosity was noted on chromosome 7q of the striatal donor gDNA. gDNA genotypes together with raw allelic intensity (Illumina Xraw and Yraw fields) for all samples were exported as comma separated value files. Quantile normalization of the raw allele intensities between channels (Limma Bioconductor package—<http://www.bioconductor.org>) was then performed on cDNA and gDNA datasets before calculating  $\beta$  values ( $X/(X + Y)$ ) for each probe when the intensity ( $X + Y$ ) was above background level which, based on homozygous calls in gDNA appearing as incorrect heterozygous calls in cDNA, we determined to be 3,000. Unlike genotype calls,  $\beta$  provides a quantitative scale of allelic expression, where values around 0.5 represent equal representation of both alleles (heterozygotes), whereas those closer to 0 or 1 represent single alleles (homozygotes).  $\beta$  values for heterozygous single-nucleotide polymorphisms (SNPs) in the gDNA provide the ideal biallelic model, with significant deviations of  $\beta$  in the cDNA representing MA. Allelic expression measurement,  $\Delta\beta$  or  $\Delta_\beta$ , was therefore calculated using  $\beta_{\text{gDNA}} - \beta_{\text{cDNA}}$  when the gDNA SNP is heterozygous. Transcript-based  $\Delta_\beta$  estimates used the mean  $|\Delta_\beta|$  of expressed informative SNPs within RefSeq accessioned transcripts. A penalty-based weighting score was applied to transcripts, where SNP probes with  $\Delta_\beta > 0.1$  scored +1, whereas values below 0.1 received a -2 penalty. Genes with a total score of one or below were rejected. Additional filtering was also applied based on the expressed SNP probe density. Genes with 2–10 SNPs required at least 50% SNPs to be detectable above background. Eleven to nineteen SNPs required at least six detectable SNPs, 20–29 SNPs required seven, 30–49 SNPs required eight, 50–74 SNPs required nine, and 75+ SNPs required 10. Monoallelic expressed genes were then defined by a mean SNP  $\Delta_\beta > 0.2$ , whereas biallelic expression was defined as  $\Delta_\beta < 0.1$ . Gene expression estimates were made using the mean ( $\text{Intensity}_{\text{cDNA}}/\text{Intensity}_{\text{gDNA}}$ ) for all SNP probes within the transcript. Identification of intergenic or chains of SNPs was achieved using Boolean operators in Microsoft Excel 2010.

SNP probes were annotated using GALAXY (<http://main.g2.bx.psu.edu/>). Transcript-based analysis was performed in an isoform-specific manner. SNP probes lying in segmentally duplicated regions (data track at <http://genome.ucsc.edu/>) or defined duplicated regions [16] were removed from the transcript-based analysis. Known imprinted genes were identified from [genem印rint.com](http://www.genem印rint.com) and supplemented with loci identified by Morcos et al. [17]

### DNA Methylation Analysis

Seven hundred and fifty nanograms of bisulfite-treated gDNA from cortical and spinal cord clonal lines was assayed on the Illumina Infinum HumanMethylation27 BeadChip (Illumina, San Diego, CA, www.illumina.com) using the standard manufacturer's protocol, with DNA methylation  $\beta$ -values calculated using the GenomeStudio Methylation module (v1.6.1). DNA methylation  $\beta$  values were then mapped to the allelic expression results and statistics carried out in R (<http://www.r-project.org/>). Clonal bisulfite sequencing for a region upstream of TNFRSF10D was carried out on bisulfite-treated DNA (EpiTech, Qiagen, Crawley, U.K., www.qiagen.com) using the primers 5'-AGGATT TTGGGGTTTAGGAGTTAT-3' and 5'-AATACTAAAAA CCAACCTAAATACC-3'. Sixteen clonal sequence reads per sample were analyzed using BiQ software with the default quality control parameters [18].

## Bioinformatics

Gene ontology analysis was carried out using DAVID (<http://david.abcc.ncifcrf.gov/>) and Ingenuity (<http://www.ingenuity.com/>). To avoid bias, the large protocadherin family was removed from the gene lists. A reference gene set was defined for each donor based on genes with sufficient detectable number of SNPs to be scored (see Allelic Expression Analysis). EpiGraph (<http://epigraph.mpi-inf.mpg.de/>) was used for genetic and epigenetic (histone measures from lymphocytes) analysis at transcriptional start sites (1,000 bp upstream and 500 bp downstream) and the full length of the transcript. The same loci were also used to retrieve fetal brain epigenomics data from the NIH Epigenomics Atlas.  $dN/dS$  values for human/macaque comparison were obtained from Ensembl (<http://www.ensembl.org/>). DNA methylation fetal brain values were obtained from the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM664920>).

## Validation Experiments

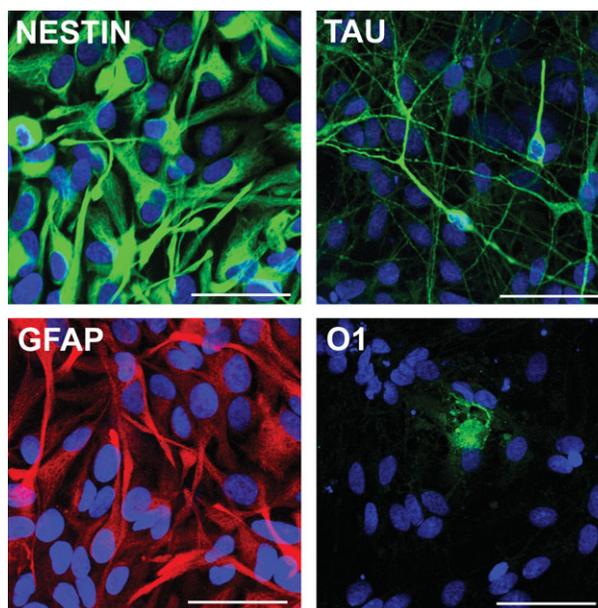
PCR amplicons between 100 and 300 bp were designed with Primer3plus (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>) to amplify the informative SNP of interest from cDNA and gDNA. Amplicons were treated with Exonuclease I (New England Biolabs, Hertfordshire, U.K., [www.neb.uk.com](http://www.neb.uk.com)) and Rapid Alkaline Phosphatase (Roche, Welwyn Garden City, U.K., [www.roche.co.uk](http://www.roche.co.uk)) prior to downstream analysis. Sanger sequencing was carried out with the PCR primers (Big-Dye v3.1 chemistry, Applied Biosystems, Warrington, U.K., [www.appliedbiosystems.com](http://www.appliedbiosystems.com)) or an additional primer for single primer nucleotide primer extension analysis (SNaPshot multiplex assay, Applied Biosystems, Warrington, U.K., [www.appliedbiosystems.com](http://www.appliedbiosystems.com)). Both were analyzed on an ABI 3130 genotyper/sequencer (Applied Biosystems, Warrington, U.K., [www.appliedbiosystems.com](http://www.appliedbiosystems.com)). Peak heights at the informative SNP were measured using PeakPicker [19] or GeneMarker (SoftGenetics, State College, PA, [www.softgenetics.com](http://www.softgenetics.com)).  $\Delta\beta$  values could then be calculated as previously described. Quantitative PCR was carried out using EVA-green mastermix (Solis Biodyne, Tartu, Estonia, [www.sbd.ee](http://www.sbd.ee)) on an MJ Research Chromo 4 thermal cycler (Bio-Rad, Hertfordshire, U.K., [www.bio-rad.com](http://www.bio-rad.com)). Relative gene expression was calculated as described in Pfaffl [20] using five reference genes.

## RESULTS

### Illumina Beadchips Provide a Suitable Platform to Detect MA

We assessed genome-wide allelic expression in nine clonal neural stem cell lines derived from three different fetal donors. The three cortical lines (CTX0E03, CTX0E16, and CTX0E17) came from one donor; three striatal lines (STR0C05, STR0C08, and STR0C11) from a separate unrelated donor, and spinal cord lines (SPC01, SPC04, and SPC06) from a third donor. Initially, we assessed gene expression in proliferating undifferentiated cells. In this phase, all these cell lines express neural stem cell markers such as Nestin (Fig. 1), Musashi1, and Sox2 [13].

We used the Illumina Infinium Omni1-Quad beadchip to measure allelic representation at informative heterozygous SNPs for cDNA and gDNA. A quantitative scale termed  $\Delta\beta$  was used to measure the allelic ratio. This was calculated by comparing the amount of each expressed SNP allele in cDNA relative to the same SNP in donor gDNA (representing a 50:50 allelic ratio). Use of total RNA (rather than mRNA alone) enabled measurements from both exonic and intronic SNPs when expressed over the background signal. This resulted in approximately 100,000 autosomal informative intragenic SNPs in more than 9,000 genes. Biological reproducibility was demonstrated from three replicates for the pro-



**Figure 1.** Cell surface markers on clonal human neural stem cells showing multipotentiality. Clonal lines from the spinal cord donor show expression of the stem cell marker Nestin while they undergo proliferation. When growth factors are removed, the cells differentiate into neurons (stained with TAU), astrocytes (GFAP), and oligodendrocytes (O1). Nuclei are stained with 4',6-diamidino-2-phenylindole. Scale bar = 50  $\mu$ m. Abbreviation: GFAP, glial fibrillary acidic protein.

totypic spinal cord line SPC01 ( $\Delta\beta$  Pearson's correlation  $R$  between 0.86 and 0.89, Supporting Information Fig. S1), which is similar to a previous allelic expression study based on the Illumina platform [3]. We also looked for X-inactivation in female cell lines as proof of principle (Supporting Information Fig. S2), showing 85% of measurable X-chromosome SNPs or 78% (159 out of 202) of genes displaying MA/X inactivation. This agrees with previous estimates of 75%–85% of genes undergoing silencing and 15% escaping inactivation in human fibroblasts [21]. Known autosomal imprinted loci also demonstrated MA (Supporting Information File S1).

### Widespread MA Was Observed in Autosomal Genes, a Subset of Which Show Evidence for Stochastic Allelic Choice

Having detected monoallelic X inactivation, we sought to discover whether autosomal genes in clonal lines show similarly pronounced deviation in allelic expression. Allelic expression measurements ( $\Delta\beta$ ) for autosomal intragenic SNPs appear to follow a normal distribution but with "heavy tails" (kurtosis score  $>5$ ). Normal Q-Q plots of each individual clonal line showed the tails deviating away from a normal distribution at  $\Delta\beta$  values of approximately  $+0.07$  and  $-0.07$  (Supporting Information Fig. S2). Biological transcriptome noise is likely to be represented in a normal distribution, so these deviations at the distribution tails represent putative loci showing true allelic imbalances. Similar observations in other allelic expression studies have led to an accepted  $\Delta\beta$  threshold of 0.1, which represents a theoretical 40:60 allelic ratio, as a relevant allelic imbalance observation [22–24]. We therefore used this threshold and applied a weighted penalty scoring system to all transcripts that took into account the density of informative expressed SNPs within a transcript (see Materials and Methods). The weighting score provides a scale or rank of confidence for MA measurements. More than 9,000 genes

were assayable in each donor, and genes with an overall mean allelic expression  $\Delta_\beta > 0.2$  (representing an allelic ratio of 26:74 when directly measured in cDNA) were classified as showing MA (Table 1 and Supporting Information File S2). We observed that 1.82%, 2.16%, and 1.57% of the examined genes in cortical, striatal, and spinal cord donors, respectively, had at least one

**Table 1.** Results summary for CTX-, STR-, and SPC-derived clonal stem cells

	CTX	STR	SPC
Assayed genes	10,150	9,417	9,085
MA	185 (1.82%)	203 (2.16%)	143 (1.57%)
Known imprinted	16 (0.16%)	15 (0.16%)	15 (0.17%)
Stochastic allelic choice (St-MA)	89 (0.87%)	107 (1.14%)	43 (0.47%)
Same allelic choice	19 (0.19%)	22 (0.23%)	25 (0.28%)
Unclassified	65 (0.64%)	61 (0.64%)	62 (0.68%)

Assayed genes are shown (those containing informative expressed single-nucleotide polymorphisms [SNPs]) together with the number of genes with one or more clones showing MA, together with a breakdown of the types of allelic expression identified (stochastic allelic choice or St-MA, same allelic choice where all three clones show monoallelic expression in the same direction, and unclassified where only one or two clones showed detectable monoallelic gene expression in the same direction). Note: two identified imprinted genes in CTX and one in STR and SPC show one out of three clones with biallelic expression, meaning they are also counted in the St-MA group. Additionally, two genes in CTX and one gene in STR and SPC show the presence of one transcript isoform in the St-MA group and an additional alternate isoform of different SNP composition in the same allelic choice group.

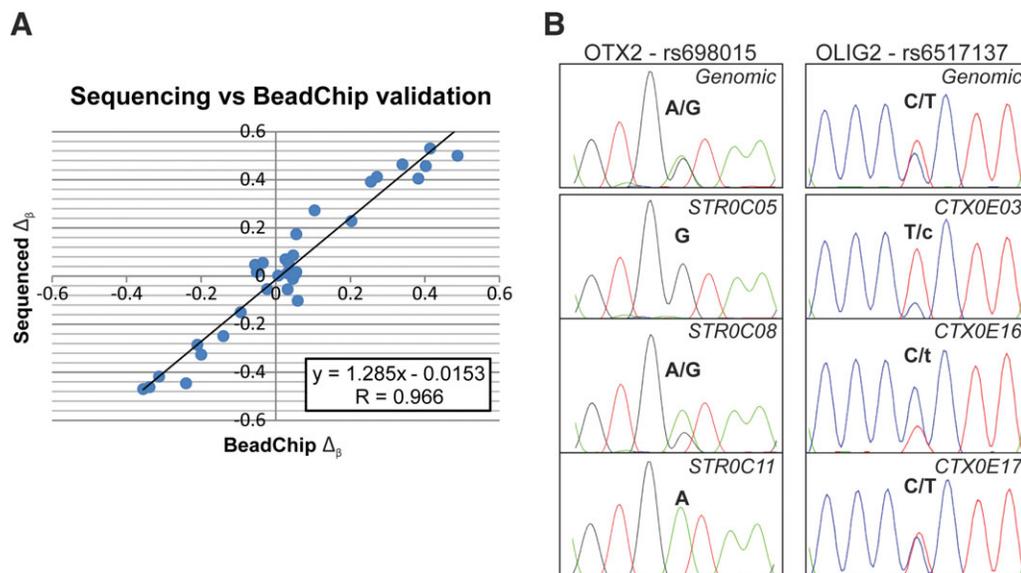
Abbreviations: CTX, cortical; MA, monoallelic expression; SPC, spinal cord; STR, striatal.

clone with MA (Table 1). Approximately 0.16% of examined genes was known imprinted genes, leaving the occurrence of novel autosomal loci showing MA to be between 1.4% and 2.0%. While some of these genes show the same MA across all clonal lines from a donor, 0.87%, 1.14%, and 0.47% of cortical, striatal, and spinal cord assayed genes showed evidence of stochastic allelic choice (St-MA). By stochastic, we mean that one clone would show MA, while a second sister clone (from the same donor) may show biallelic expression or MA for the alternate allele. We also identified ~2,000 additional genes containing a single informative SNP, with 5% showing putative MA (Supporting Information File S3) but excluded these from further analysis. We refer to genes that showed biallelic expression in all three clonal lines as BA. Full Allelic expression results are hosted at <http://epigenetics.iop.kcl.ac.uk/nsc> for visual inspection in UCSC Genome Browser.

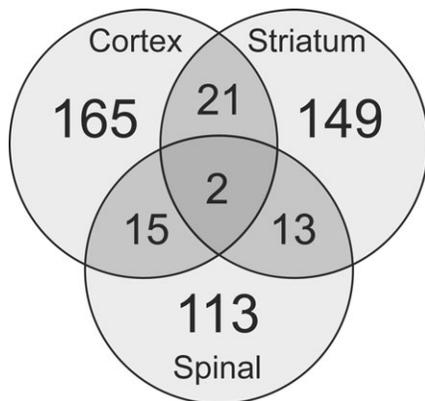
To independently verify allelic expression measurements, we sequenced RNA from 12 arbitrarily selected expressed genes across the clonal lines from the three donors (Fig. 2). Strong correlation was observed between the beadchip and direct sequencing (Pearson's correlation  $R = 0.966$ ), for both allelic discrimination together with the degree of allelic imbalance.

### Independent Samples Show Overlap for Autosomal Genes Susceptible to MA

We asked whether the same set of MA genes reoccurred in independent donors. A simulation showed that on average, four MA genes would be expected by chance to be detected in two donors, whereas typically zero MA genes would be expected between three or more donors. Our observed values of between thirteen to twenty one genes shared by two donors and also two genes common in all three donors (Fig. 3) indicate that MA gene expression occurs at a nonrandom/specific set of loci in these neural stem lines ( $p < .0001$ , chi-squared test). Thus, while the selection of allele might be stochastic, this is not true



**Figure 2.** Validation of allelic expression measurements. (A): Allelic expression  $\Delta_\beta$  measurements from the Illumina beadchip (x-axis) plotted against Sanger sequencing derived values (y-axis) for 12 genes showing measurable expression. A strong correlation is observed (Pearson's correlation  $R = 0.966$ ) highlighting the validity of the beadchip allelic analyses. The analysis also illustrates the underestimation of allelic expression from the beadchip. For example, a  $\Delta_\beta$  measurement of 0.20 actually represents a value of 0.24 (with an allelic ratio of 26:74 or a minimum of a 2.85-fold difference between the two alleles). (B): Examples of stochastic allelic choice. The transcription factor OTX2, which plays a pivotal role in forebrain specification and is thought to be important in modulating synaptic plasticity during the critical period of cortical development, shows monoallelic expression (MA) in clone STR0C05 and expression of the alternate allele in STR0C11, whereas STR0C08 shows biallelic expression, with a slight imbalance toward the A allele. A second example, OLIG2—a regulator of neural progenitor cell fate and oligodendrocyte development, also shows MA in two clones, albeit with weaker repression of the minor allele in CTX0E03 and CTX0E16.



**Figure 3.** Relationship of monoallelic genes identified in the three donors. The Venn diagram shows the number of genes identified with at least one clone from a donor showing monoallelic expression (known imprinted genes are excluded). The overlap of monoallelic genes between both two and three donors represents more than that would be expected by chance. The two nonimprinted autosomal genes common to all three donors are *ACCS* and *TNFRSF10D*.

for the selection of the genetic loci, instead suggesting a subset of loci preferentially susceptible to MA. This is supported by a 2.4–3.8-fold enrichment of monoallelic loci identified by Gimelbrant et al. [9] overlapping with ours, despite the expression differences inherent between lymphoblastoid and neural stem cells (Supporting Information Table S1).

### Stochastic Allelic Choice Genes Are Enriched for Neurodevelopmental Functions

We examined the functional classification of St-MA-expressed genes for each donor with the functional annotation tools DAVID [25] and ingenuity pathway analysis. We found that >30% of the St-MA genes in all three donors was transmembrane glycoproteins. Developmental terms were particularly enriched in St-MA genes identified in the clones derived from the brain (striatal and cortical clones) and included genes such as ventral anterior homeobox 1 (*VAX1*) transcription factor, neurotrophin-3 (*NTF3*), and neurexin 3 (*NRXN3*). The transcription factor binding sites LHX3 and CHX10 also ranked very highly in all three donors, being present in more than 50% of the genes in spinal and striatal, 70% of cortical genes, although this does only represent a 1.5-fold (spinal and striatal) to 1.9-fold (cortical) enrichment. Complete annotations are shown in Supporting Information File S4.

### Allelic Choice Is Largely Maintained After Differentiation

Our findings indicate that neural stem cell clones express a subset of genes with MA expression. The question arises as to whether the allelic expression pattern is retained when these progenitor cells differentiate into neurons and glia. To investigate the effect of differentiation on allelic expression, we allowed cells to differentiate in vitro for 1 week into neurons and glia, as visualized by their morphological differentiation and the expression of cell-specific markers (Fig. 1) [13–15]. We measured allelic expression of the genes *TMEM132D*, *GRID1*, *TNFRSF10D*, and *PMP2*. Quantitative PCR analysis showed that although *TMEM132D* expression remains unchanged after differentiation, *GRID1*, *TNFRSF10D*, and *PMP2* exhibit upregulation (Supporting Information Fig. S3). Despite gene expression changes, the allelic expression status of these genes is maintained in the differentiated progeny

(Fig. 4). Thus, any functional impact St-MA expression is likely to be maintained in the neurons and glia.

### MA Found Within Intergenic Regions of the Genome

As expected, intergenic SNPs existing outside of the classic transcript boundaries mostly showed low SNP probe intensities. Nevertheless, some regions show detectable, and sometimes MA, expression. We first validated this observation by successful PCR amplification and sequencing of five monoallelic expressed intergenic SNPs (Supporting Information Fig. S5). We then identified chains of multiple SNP probes sequentially positioned on the genome as representing monoallelically transcribed regions (Supporting Information Table S2). In CTX0E17, for example, we identified 164 chains of three or more monoallelically expressed SNPs, comprising 753 total SNPs. Unsurprisingly, many corresponded to already identified expressed transcripts. However, 35% of the SNPs was present within intergenic spaces. Of these, 44% shows overlap with expressed sequence tags (ESTs) and likely represent novel transcripts or expressed repetitive elements. Some may also be explained by their proximity to genes showing similar allelic expression status, representing unannotated alternative isoforms or antisense transcripts. For example, SNPs upstream of *OTX2* match its MA status in the striatal lines, an observation likely explained by the overlapping antisense RNA transcript *OTX2OS1* (Supporting Information Fig. S6) found at this locus. Nonetheless, approximately 56% of the intergenic SNPs identified does not appear to show any overlap with ESTs.

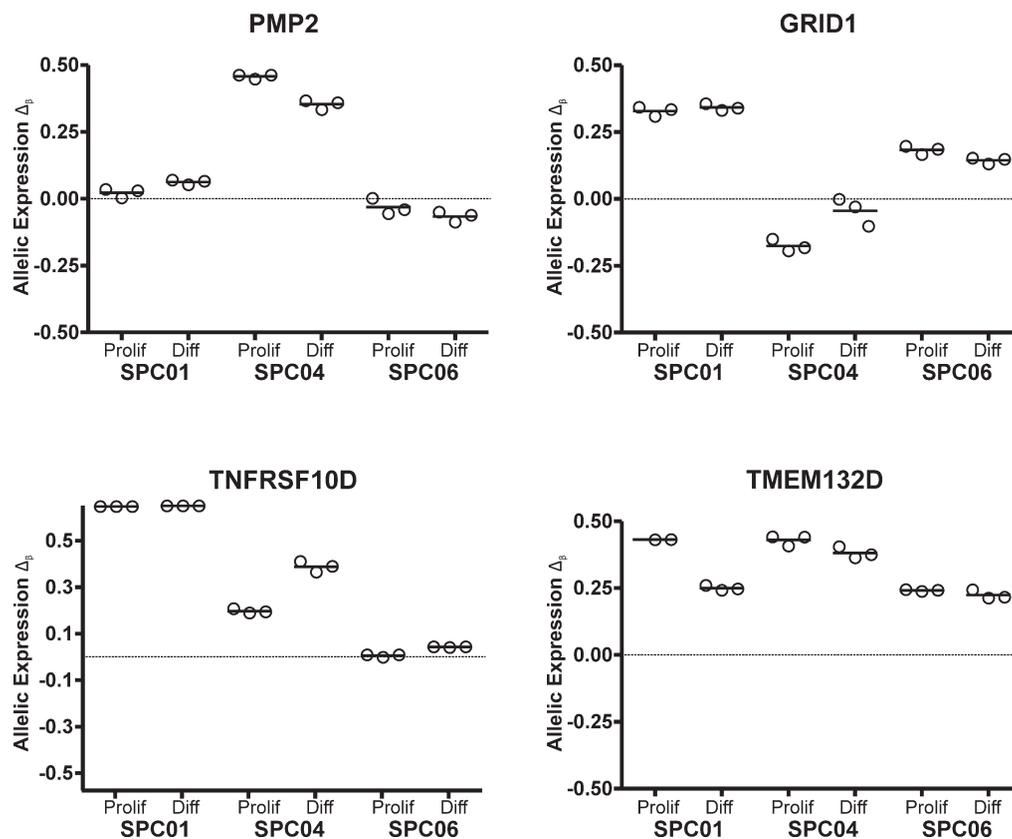
### MA Is Commonly Associated with Reduced Transcript Levels

One impact of MA might be to reduce expression levels: if one allele is silenced, then overall expression might be reduced. This was suggested in a clonal lymphoblastoid study [9]. We used SNP probe intensities to estimate gene expression that showed good accuracy when compared with quantitative PCR measurements (Pearson's correlation  $R = 0.871$ , Supporting Information Fig. S3). We therefore examined the impact of MA on transcript levels for all St-MA genes.

Plotting transcription levels against allelic expression shows a weak but significant ( $p$ -value = .01) negative correlation between the degree of allelic imbalance and total transcript levels (Pearson's correlation  $R = -0.159$ ), consistent with MA generally reducing the absolute expression of that gene (Supporting Information Fig. S4). This is also reflected in the absolute counts of monoallelic expressing clones showing either increased or decreased gene expression relative to their biallelic counterparts, the result of which indicates a twofold increased chance of reduced expression (88 upregulated vs. 171 downregulated in monoallelic clones and 50 upregulated vs. 62 downregulated in biallelic sister clones, chi-squared test  $p$ -value <.0001). Therefore, MA is more likely to result in reduced transcript levels, such as in the example shown in Figure 5A and 5B, although this is not a universal rule.

### DNA Sequence and DNA Methylation Differences Exist Between Monoallelic and Biallelic Gene Loci

A factor that may determine monoallelic and biallelic gene expression is the local DNA sequence acting in cis. The LINE-1/L1 transposon family has previously been associated with X-chromosome inactivation [26] and similarly associated with autosomal monoallelic genes together with fewer CpG islands and SINE repeats [27]. We used the tool EpiGraph [28] to look for any associated DNA sequence differences (Supporting



**Figure 4.** Allelic expression after differentiation. Allelic expression ratios for four genes were measured in the clonal spinal cord lines SPC01, SPC04, and SPC06 when in proliferative (Prolif) and differentiated (Diff) states. Measurements for three biological replicates (circles) are shown and the mean value (horizontal line) indicated.

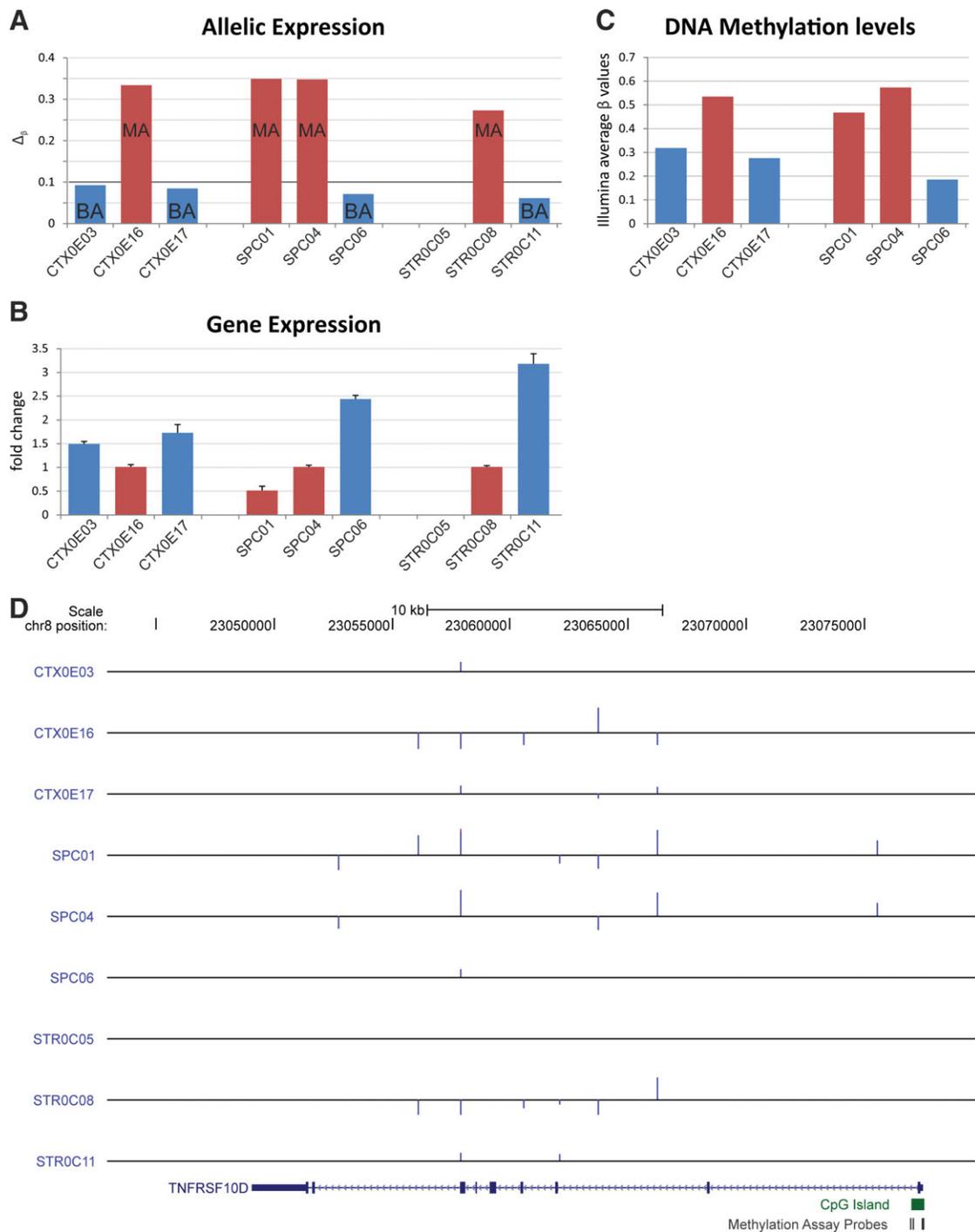
Information File S5). No significant difference was found for LINE-1 or overall LINE repeat occurrence between MA and BA expressed genes, although a marginal increased length of LINE-1 was noted with MA genes (Supporting Information Fig. S7). We find lower CpG density at transcriptional start sites of MA expressed genes (Wilcoxon rank sum  $p$ -value = .0008) together with depleted amounts of SINE repeats ( $p$ -value =  $1.4 \times 10^{-6}$ ) and increased long terminal repeats ( $p$ -value = .0001) across the length of the transcript.

Epigenetic factors can also be deterministic for allelic expression. The allele-specific expression of imprinted and X-inactivated loci is associated with allele-specific DNA methylation (ASM) across discrete differentially methylated regions. A recent study showed that ASM is common across the genome and tightly linked to allele-specific expression [29]. We investigated genome-wide patterns of DNA methylation in the spinal cord and cortical cell lines using the Illumina 27k Infinium methylation beadchip that targets CpG sites located in the proximal promoter regions of transcription start sites. The results shown in Figure 6A demonstrate that MA gene loci show increased levels of DNA methylation when compared with BA gene loci. Each pairwise combination shows high statistical significance (least significant  $t$  test  $p$ -value =  $1.12 \times 10^{-76}$ ), indicating a strong association between MA and intermediate DNA methylation levels, as seen in differentially methylated regions, at these gene loci. The St-MA choice gene *TNFRSF10D* provides a specific example of this association across all three donors (Fig. 5C and Supporting Information S8). A similar association of increased DNA methylation at our identified MA loci is also observed in fetal brain tissue (Fig. 6A).

### Repressive Chromatin Signatures Associate with Monoallelic Susceptible Gene Loci in Fetal Brain

Altered chromatin status is another epigenetic signal that may associate with monoallelic gene expression. We used the MA and BA gene loci identified in our neural stem cells to interrogate epigenomic data for a human fetal brain sample produced by the NIH Epigenomics Roadmap initiative [30]. Direct allelic expression measurement of a stochastic monoallelic expressing gene would typically show supposed “biallelic” expression due to the random nature of allelic choice in a nonclonal tissue. Nevertheless, any chromatin signatures associated with the subpopulation of monoallelic expressing cells should be detectable when compared with the more “consistent” chromatin status of a constitutive biallelic expressing gene. We therefore examined fetal brain chromatin measures at the transcriptional start sites for loci defined from our neural stem cell allelic expression data.

Using loci from the cortical donor as an example, increased chromatin accessibility was evident in BA genes as a 2.2-fold enrichment of DNase I hypersensitivity sites when compared with St-MA genes (Fig. 6B). St-MA expressed genes showed a 2.9-fold enrichment in H3K27me3 repressive marks, whereas BA genes exhibited 2.2-fold more H3K4me3 and 1.85-fold more H3K9ac, both measures of open/active chromatin. H3K9me3, a modification often linked with repression and heterochromatic silencing, showed a 1.85-fold enrichment in biallelic clones, although it has also been noted in actively transcribed regions [31, 32]. All comparisons showed high statistical significance (Wilcoxon rank test- $p$ -value  $< 2.2 \times 10^{-16}$ ). Similar results were also found for the loci identified in the other two independent donors (Supporting Information Fig. S9). A comparison of all MA identified

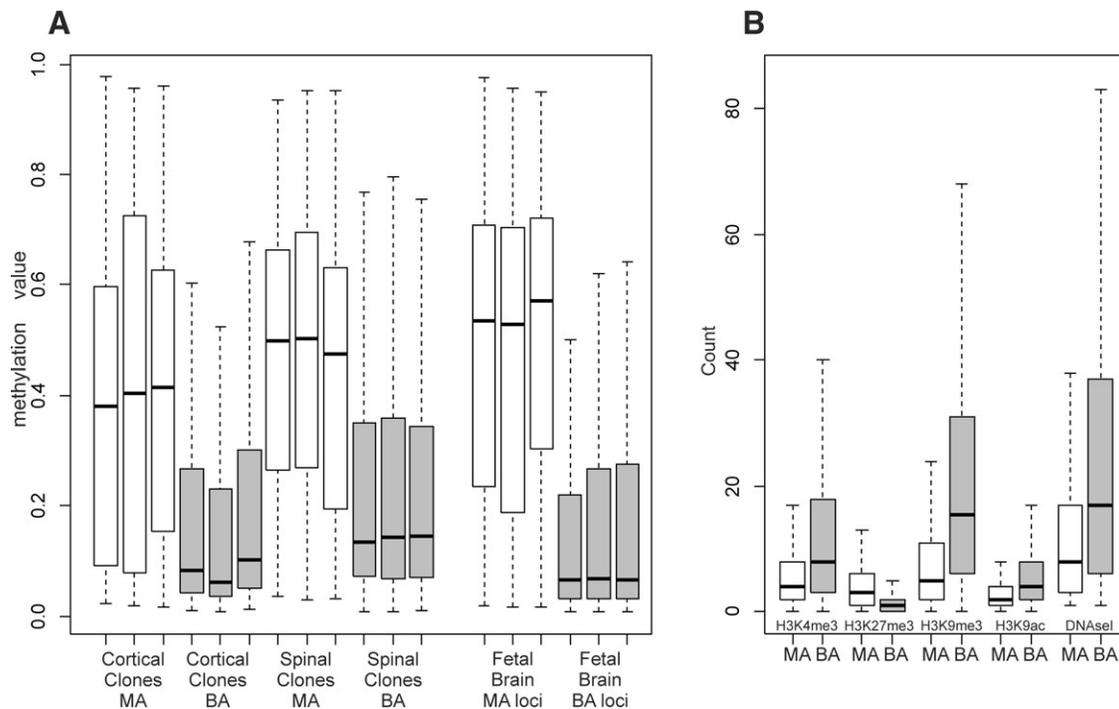


**Figure 5.** Measurements from the gene *TNFRSF10D* that shows stochastic allelic choice in all three donors. (A):  $\Delta\beta$  allelic expression measurements clearly show a number of monoallelic expressing clones (red, labeled MA) and biallelic expression (blue, labeled BA). (B): Gene expression from quantitative PCR showing increased *TNFRSF10D* transcript levels in biallelic expressing clones when compared with monoallelic sister lines from the same donor. (C): Methylation  $\beta$  levels deduced from five probes on the Illumina 27K methylation beadchip are shown for the cortical and spinal cord donors. (D): Genomic structure of *TNFRSF10D* shown, together with the location of a 5' CpG island, the position of the methylation probes, and also expressed SNPs showing MA expression. The SNPs are indicated as peaks representing  $\Delta\beta$  measurements. Abbreviations: BA, biallelic expression; MA, monoallelic expression.

loci compared to BA expressing genes also gave very similar results. Surprisingly, non-neural tissue also showed similar histone mark associations for H3K27me3 and H3K4me3 (Supporting Information Fig. S10), suggesting a common, relatively stable state of these measures across multiple tissues types. H3K9me3 enrichment with BA loci, conversely, appears to be unique to fetal tissue.

### Monoallelic Genes Are Associated with Accelerated Evolutionary Noncoding Sequences

Gimelbrant et al. found that monoallelic genes in clonal lymphoblastoid cell lines were more than twice as likely as biallelic genes to be close to presumed regulatory conserved noncoding sequences believed to be characterized by human-



**Figure 6.** Epigenomic measures from fetal brain at cortical donor defined monoallelic and biallelic loci. **(A):** DNA methylation measures using the Illumina Methylation 27k beadchip show increased levels of methylation at monoallelic genes when compared with biallelic loci. Average  $\beta$  values, indicating level of methylation, are shown for monoallelic (white boxes labeled MA) and biallelic expression (gray boxes labeled BA) in cerebral cortex and spinal cord donors. The methylation status of loci defined in these loci is also shown for a fetal brain sample. **(B):** Chromatin measures from a fetal brain sample (human epigenome atlas) at transcriptional start sites of stochastic monoallelic genes (white boxes) and biallelic genes (gray boxes) identified in the cerebral cortex donor for Histone H3 lysine 4, H3 lysine 27, H3 lysine 9 trimethylation, H3 lysine 9 acetylation, and DNase I hypersensitivity sites (Supporting Information Fig. S9 for the other two donors). Abbreviations: BA, biallelic expression; MA, monoallelic expression.

lineage-specific accelerated evolution [9, 33]. Using three separate datasets that define sequences [33–35], we confirm that MA genes are significantly more likely to be located close to accelerated evolutionary noncoding sequences (Supporting Information Fig. S11). We also looked at positive selection pressure on the coding gene regions by calculating the nonsynonymous versus synonymous ( $dN/dS$ ) ratio from a comparison of Human and Macaque genome. No significant difference was found between MA and BA genes.

## DISCUSSION

A number of studies have reported allelic expression imbalances in non-neuronal cells and highlighted tissue-specific cis effects [3, 23, 24, 36–38]. Allelic imbalance is also commonly observed in human brain tissue, although any stochastic choice of allele within individual cells is likely to be missed by such analyses [39, 40]. We used a series of clonal human neural stem cell lines to carry out a global allelic expression survey, allowing detection of stochastic events in allelic choice. We have shown that between 1.4% and 2.0% of autosomal genes from clonal lines derived from three different tissues are subjected to MA expression. A subset of these (between 0.47% and 1.14% of assayed genes) show evidence for St-MA expression, many of which encoded proteins important in neurodevelopment. This frequency is similar number to the 1% estimate in a clonal murine neural stem cell study [41] but lower than that found in mouse and human

lymphoblastoid cell studies [9, 42]. We believe this study may underestimate the prevalence for the following reasons. First, we excluded single SNP genes or those with low expression due to the higher risk of false positives even though we could validate a number by sequencing. Second, examples can be found where small transcripts do not contain sufficient informative heterozygous SNPs to be classified yet are flanked by monoallelic expressed intergenic SNPs. In such cases, the transcript would presumably show similar allelic expression to that of the flanking SNPs due to the shared local heterochromatin. Finally, our use of three clones from each donor had limited power due to the limited amount of allelic heterozygosity.

Our data suggest that while allelic expression may often be stochastic, the actual loci undergoing such transcriptional control is not random, as indicated by the significant overlap of MA genes between different donors (genotypes) beyond that expected by chance. Furthermore, there is a significant overlap between our set of MA genes and those identified in lymphoblastoid cells by Gimelbrant et al. [9]. Thus, a specific subset of genes appears to be susceptible to MA regulation and may also be conserved across multiple cell or tissue types. A subset of the identified MA genes showed the same monoallelic choice in all clonal lines within a donor. Genetic cis-regulatory variants can lead to such an observation in genetically related cell lines, and the use of only three donors cannot totally rule out some of our observations falling into this category. Nonetheless, a proportion of these identified genes and intergenic regions may represent novel imprinted loci, although confirmation of this status remains to be demonstrated.

How might the allelic expression we observe in neural stem cells affect brain development? There are three reasons for believing that the allelic status of a gene might have a significant functional impact. First, we show that the allelic status is maintained, at least for a subset of genes, as the neural progenitor cells differentiate. This could contribute to clonal phenotypes in the resulting neurons and glia. Second, the St-MA genes are enriched for a number of genes important in neurodevelopment, functioning, and disease. Finally, epigenetic associations observed in human fetal brain infer possible *in vivo* occurrence of St-MA gene expression. We have examined cultured neural stem cells, since it is only in culture that clonal progenitors can effectively be examined biochemically. Nonetheless, our observations suggest a subset of genetic loci have a particular propensity for allelic regulation. We observe that the promoter regions of MA genes show intermediate levels of DNA methylation in comparison to BA genes, which are generally hypomethylated, a finding that is also mirrored in fetal brain. While the methylation assay used in this study was not allele-specific, these data concur with the notion that the unexpressed allele is silenced by DNA methylation. Also, when compared with BA gene loci, MA genes have a distinct combination of chromatin modifications in fetal brain tissue, suggesting that they are less "open." These findings do not directly prove that these genes are monoallelic *in situ* in the developing brain, but indicate that they are distinct from the BA genes in a manner consistent with what might be expected if one allele were more permissive to transcription.

Our data support a model in which the developing brain is a mosaic composed of distinct clones of cells, each with a unique combination of monoallelic and biallelic expressed genes. Differing combinations would then allow clonal diversity from either gene dosage variations or functional differences due to polymorphic variation between the two alleles. Alternatively, the impact of MA may be the unmasking of recessive alleles or repression of semidominant alleles [43], which, for example, would be limited to specific clones of cells due to their "salt-and-pepper" distribution in the brain. Moreover, the same clonal diversity patterns would not be precisely reproduced, even between individuals with the same genotype. This mechanism could explain some of the discordance observed between monozygotic twins for neurobiological phenotypes and psychiatric disease [44].

Even if this phenomenon is restricted to cells *in vitro*, it could have significance in that it might explain some of the

clonal diversity that is invariably observed in stem cell (and other) clonal lines. Like many others, we have observed that sister lines, which are demonstrably multipotential and expressing the requisite combination of markers to be deemed neural stem cells, nonetheless differ markedly in aspects of their phenotype. CTX0E16 and CTX0E03, for example, both have stable phenotype over multiple passages. Nonetheless, the former reproducibly generates more neurons when differentiated than its sister line. No doubt there are many causes for this diversity, but the stochastic allelic expression choice we have observed is likely to be one of them.

## CONCLUSIONS

To conclude, we have identified widespread MA in clonal human neural stem cell lines. Stochastic choice of allelic expression appears evident for a subset of these genes, with roles associated in nervous system development and functioning. The process has potential to allow significant cellular diversity within complex tissues such as the brain, although it still remains to be seen if stochastic allelic choice is widespread *in vivo*. Nevertheless, supportive epigenomic data from fetal brain suggest it may be a possibility.

## ACKNOWLEDGMENTS

We thank Chloe Wong and Ruth Pidsley for help with the methylation experiments. We thank Ioannis Ragoussis and Ghazala Mirza at the Wellcome Trust Centre for Human Genetics unit for their assistance with the Illumina beadchip studies. This work was supported by the Charles Wolfson Charitable Trust. J.M. is supported by NIH Grant AG036039.

## DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

JP is a consultant for ReNeuron Ltd., a U.K. stem cell biotech company.

## REFERENCES

- Reik W, Walter J. Genomic imprinting: Parental influence on the genome. *Nat Rev Genet* 2001;2:21–32.
- Lyon MF. X chromosomes and dosage compensation. *Nature* 1986; 320:313.
- Ge B, Pokholok DK, Kwan T et al. Global patterns of cis variation in human cells revealed by high-density allelic expression analysis. *Nat Genet* 2009;41:1216–1222.
- Gimelbrant AA, Ensminger AW, Qi P et al. Monoallelic expression and asynchronous replication of p120 catenin in mouse and human cells. *J Biol Chem* 2005;280:1354–1359.
- Chess A, Simon I, Cedar H et al. Allelic inactivation regulates olfactory receptor gene expression. *Cell* 1994;78:823–834.
- Mostoslavsky R, Singh N, Tenzen T et al. Asynchronous replication and allelic exclusion in the immune system. *Nature* 2001;414: 221–225.
- Esumi S, Kakazu N, Taguchi Y et al. Monoallelic yet combinatorial expression of variable exons of the protocadherin-alpha gene cluster in single neurons. *Nat Genet* 2005;37:171–176.
- Kaneko R, Kato H, Kawamura Y et al. Allelic gene regulation of Pcdh-alpha and Pcdh-gamma clusters involving both monoallelic and biallelic expression in single Purkinje cells. *J Biol Chem* 2006;281: 30551–30560.
- Gimelbrant A, Hutchinson JN, Thompson BR et al. Widespread monoallelic expression on human autosomes. *Science* 2007;318: 1136–1140.
- Davies W, Isles AR, Wilkinson LS. Imprinted gene expression in the brain. *Neurosci Biobehav Rev* 2005;29:421–430.
- Gregg C, Zhang J, Butler JE et al. Sex-specific parent-of-origin allelic expression in the mouse brain. *Science* 2010;329:682–685.
- Gregg C, Zhang J, Weissbourd B et al. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* 2010; 329:643–648.
- Pollock K, Stroemer P, Patel S et al. A conditionally immortal clonal stem cell line from human cortical neuroepithelium for the treatment of ischemic stroke. *Exp Neurol* 2006;199:143–155.
- El-Akabay G, Medina LM, Jeffries A et al. Purmorphamine increases DARPP-32 differentiation in human striatal neural stem cells through the hedgehog pathway. *Stem Cells Dev* 2011;20:1873–1887.
- Johansson S, Price J, Modo M. Effect of inflammatory cytokines on major histocompatibility complex expression and differentiation of human neural stem/progenitor cells. *Stem Cells* 2008;26:2444–2454.
- Musumeci L, Arthur JW, Cheung FS et al. Single nucleotide differences (SNDs) in the dbSNP database may lead to errors in genotyping and haplotyping studies. *Hum Mutat* 2010;31:67–73.

- 17 Morcos L, Ge B, Koka V et al. Genome-wide assessment of imprinted expression in human cells. *Genome Biol* 2011;12:R25.
- 18 Bock C, Reither S, Mikeska T et al. BiQ Analyzer: Visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics* 2005;21:4067–4068.
- 19 Ge B, Gurd S, Gaudin T et al. Survey of allelic expression using EST mining. *Genome Res* 2005;15:1584–1591.
- 20 Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 2001;29:e45.
- 21 Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* 2005;434:400–404.
- 22 Pastinen T, Sladek R, Gurd S et al. A survey of genetic and epigenetic variation affecting human gene expression. *Physiol Genomics* 2004;16:184–193.
- 23 Serre D, Gurd S, Ge B et al. Differential allelic expression in the human genome: A robust approach to identify genetic and epigenetic cis-acting mechanisms regulating gene expression. *PLoS Genet* 2008;4:e1000006.
- 24 Lee JH, Park IH, Gao Y et al. A robust approach to identifying tissue-specific gene expression regulatory variants using personalized human induced pluripotent stem cells. *PLoS Genet* 2009;5:e1000718.
- 25 Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4:44–57.
- 26 Lyon MF. X-chromosome inactivation: A repeat hypothesis. *Cytogenet Cell Genet* 1998;80:133–137.
- 27 Allen E, Horvath S, Tong F et al. High concentrations of long interspersed nuclear element sequence distinguish monoallelically expressed genes. *Proc Natl Acad Sci USA* 2003;100:9940–9945.
- 28 Bock C, Halachev K, Buch J et al. EpiGRAPH: User-friendly software for statistical analysis and prediction of (epi)genomic data. *Genome Biol* 2009;10:R14.
- 29 Schalkwyk LC, Meaburn EL, Smith R et al. Allelic skewing of DNA methylation is widespread across the genome. *Am J Hum Genet* 2010;86:196–212.
- 30 Bernstein BE, Stamatoyannopoulos JA, Costello JF et al. The NIH roadmap epigenomics mapping consortium. *Nat Biotechnol* 2010;28:1045–1048.
- 31 Vakoc CR, Mandat SA, Olenchok BA et al. Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. *Mol Cell* 2005;19:381–391.
- 32 Brink MC, van der Velden Y, de Leeuw W et al. Truncated HP1 lacking a functional chromodomain induces heterochromatinization upon in vivo targeting. *Histochem Cell Biol* 2006;125:53–61.
- 33 Prabhakar S, Noonan JP, Paabo S et al. Accelerated evolution of conserved noncoding sequences in humans. *Science* 2006;314:786.
- 34 Bird CP, Stranger BE, Liu M et al. Fast-evolving noncoding sequences in the human genome. *Genome Biol* 2007;8:R118.
- 35 Pollard KS, Salama SR, King B et al. Forces shaping the fastest evolving regions in the human genome. *PLoS Genet* 2006;2:e168.
- 36 Heap GA, Yang JH, Downes K et al. Genome-wide analysis of allelic expression imbalance in human primary cells by high-throughput transcriptome resequencing. *Hum Mol Genet* 2010;19:122–134.
- 37 Zhang K, Li JB, Gao Y et al. Digital RNA allelotyping reveals tissue-specific and allele-specific gene expression in human. *Nat Methods* 2009;6:613–618.
- 38 Palacios R, Gazave E, Goni J et al. Allele-specific gene expression is widespread across the genome and biological processes. *PLoS One* 2009;4:e4150.
- 39 Bray NJ, Buckland PR, Owen MJ et al. Cis-acting variation in the expression of a high proportion of genes in human brain. *Hum Genet* 2003;113:149–153.
- 40 Buonocore F, Hill MJ, Campbell CD et al. Effects of cis-regulatory variation differ across regions of the adult human brain. *Hum Mol Genet* 2010;19:4490–4496.
- 41 Wang J, Valo Z, Bowers CW et al. Dual DNA methylation patterns in the CNS reveal developmentally poised chromatin and monoallelic expression of critical genes. *PLoS One* 2010;5:e13843.
- 42 Zwemer LM, Zak A, Thompson BR et al. Autosomal monoallelic expression in the mouse. *Genome Biol* 2012;13:R10.
- 43 Schimenti J. Monoallelic gene expression in mice: Who? When?? How?? Why?? *Genome Res* 2001;11:1799–1800.
- 44 Dempster EL, Pidsley R, Schalkwyk LC et al. Disease-associated epigenetic changes in monozygotic twins discordant for schizophrenia and bipolar disorder. *Hum Mol Genet* 2011;20:4786–4796.



See [www.StemCells.com](http://www.StemCells.com) for supporting information available online.