

Allele-Biased Expression in Differentiating Human Neurons: Implications for Neuropsychiatric Disorders

Mingyan Lin¹, Anastasia Hrabovsky², Erika Pedrosa², Tao Wang³, Deyou Zheng^{1,4,5}, Herbert M. Lachman^{1,2,4,6*}

1 Department of Genetics, Albert Einstein College of Medicine, Bronx, New York, United States of America, **2** Department of Psychiatry and Behavioral Sciences, Albert Einstein College of Medicine, Bronx, New York, United States of America, **3** Department of Epidemiology & Population Health, Albert Einstein College of Medicine, Bronx, New York, United States of America, **4** Dominick Purpura Department of Neuroscience, Albert Einstein College of Medicine, Bronx, New York, United States of America, **5** Department of Neurology, Albert Einstein College of Medicine, Bronx, New York, United States of America, **6** Department of Medicine, Albert Einstein College of Medicine, Bronx, New York, United States of America

Abstract

Stochastic processes and imprinting, along with genetic factors, lead to monoallelic or allele-biased gene expression. Stochastic monoallelic expression fine-tunes information processing in immune cells and the olfactory system, and imprinting plays an important role in development. Recent studies suggest that both stochastic events and imprinting may be more widespread than previously considered. We are interested in allele-biased gene expression occurring in the brain because parent-of-origin effects suggestive of imprinting appear to play a role in the transmission of schizophrenia (SZ) and autism spectrum disorders (ASD) in some families. In addition, allele-biased expression could help explain monozygotic (MZ) twin discordance and reduced penetrance. The ability to study allele-biased expression in human neurons has been transformed with the advent of induced pluripotent stem cell (iPSC) technology and next generation sequencing. Using transcriptome sequencing (RNA-Seq) we identified 801 genes in differentiating neurons that were expressed in an allele-biased manner. These included a number of putative SZ and ASD candidates, such as *A2BP1* (*RBFOX1*), *ERBB4*, *NLGN4X*, *NRG1*, *NRG3*, *NRXN1*, and *NLGN1*. Overall, there was a modest enrichment for SZ and ASD candidate genes among those that showed evidence for allele-biased expression (chi-square, $p = 0.02$). In addition to helping explain MZ twin discordance and reduced penetrance, the capacity to group many candidate genes affecting a variety of molecular and cellular pathways under a common regulatory process – allele-biased expression – could have therapeutic implications.

Citation: Lin M, Hrabovsky A, Pedrosa E, Wang T, Zheng D, et al. (2012) Allele-Biased Expression in Differentiating Human Neurons: Implications for Neuropsychiatric Disorders. PLoS ONE 7(8): e44017. doi:10.1371/journal.pone.0044017

Editor: Gabriel Alejandro de Erasquin, University of South Florida, United States of America

Received: March 9, 2012; **Accepted:** July 30, 2012; **Published:** August 30, 2012

Copyright: © 2012 Lin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the National Institutes of Health [MH087840 and MH073164 to HML]. This publication was also supported in part by the CTSA Grant UL1RR025750, KL2RR025749 and TL1RR025748 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH roadmap for Medical Research. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Herb.Lachman@einstein.yu.edu

Introduction

Stochastic and imprinted monoallelically expressed genes influence differentiation, development and cellular function. Imprinted genes are expressed in a parent-of-origin manner, whereas in stochastic monoallelic expression either the maternal or paternal allele is active in a given cell. Imprinted genes are marked during gametogenesis by differential methylation at CpG-rich islands and by chromatin modifications, and typically maintain their transcription-competent state after fertilization [1–3]. Some genes, however, acquire parent-of-origin imprints post-fertilization [4,5]. Approximately 100 imprinted genes have been identified [4]. Whole genome expression studies expanded the family of parent-of-origin gene expression in the brain, although a subsequent re-analysis has shown that many imprinted gene calls were likely false positive findings and needed independent confirmation [6–8]. Imprinting plays a key role in some

neuropsychiatric conditions, such as Prader-Willi Syndrome and Angelman Syndrome [9]. Parent-of-origin effects have also been observed in some families with schizophrenia (SZ), bipolar disorder (BD) and autism spectrum disorders (ASD) [10–20]. In addition, some investigators have suggested that genetic imprinting influences clinical phenotype, with an imbalance between the effects of paternally and maternally expressed genes in the developing brain resulting in an extreme paternal (ASD) or maternal (SZ) pattern of behavior [14,21,22].

Stochastic monoallelic expression is widespread in mammalian genomes. Examples include T-cell receptor and immunoglobulin genes, pheromone receptors, p120 catenin, odorant receptors, and the 5q31-linked *PCDH* family of protocadherins [23–26]. Recently, Gimelbrant et al. showed that random monoallelic expression affected nearly 10% of genes expressed in lymphoblasts, and similar to some imprinted genes, there was a degree of plasticity in that biallelic expression was observed in some clones [27].

Stochastic monoallelic expression in the brain could help explain some interesting epidemiological features of neuropsychiatric disorders, such as discordance in monozygotic (MZ) twins, where a range of ~30–90% has been found in SZ, ASD and BD (see discussion) [28–31].

Two experimental tools have emerged that provide the means to evaluate the role of allele-biased expression in neuronal differentiation and neuropsychiatric disorders; iPSC technology, and next generation sequencing (RNA-Seq). We, along with other groups, are using iPSCs for *in vitro* disease modeling in a variety of neuropsychiatric disorders [32–37]. In addition to their utility for disease modeling in terms of identifying patient vs control differences in gene expression, morphology and neuronal function, iPSCs can also be used to study human neurogenesis, which is particularly relevant to SZ and ASD considering that both have a neurodevelopmental basis [38–40].

RNA-Seq provides increased sensitivity and the capacity to detect novel transcripts [41–44]. It is also an ideal platform to assess allele-biased expression by quantifying differences in expression that might occur across heterozygous single nucleotide polymorphisms (SNPs) [45]. We have used this approach to screen for allele-biased expression in differentiating human neurons. The findings highlight the degree to which allele-biased expression occurs during human neurogenesis and suggest a plausible mechanism to explain incomplete penetrance and MZ twin discordance in SZ, ASD, and other neuropsychiatric disorders.

Materials and Methods

iPSC Development and Neuronal Differentiation

iPSCs were grown and induced to differentiate into neurons using two different protocols: A and B, which are described in detail in Methods S1. Briefly, in protocol A, iPSCs were maintained on irradiated mouse embryo fibroblasts supplemented with FGF2 (10 ng/ml). Colonies were subsequently detached and grown as embryoid bodies (EBs) on non-adherent plates in the absence of FGF2. After 4 days, EBs were plated on laminin, which resulted in the development of clusters of neurons [37]. These were manually dissected after 10 days. The neurons derived from this protocol are primarily glutamatergic (~90%) [37]. In protocol B, iPSCs were maintained on matrigel and mTeSR1® (StemCell Technologies, Vancouver, Canada). EBs were formed and neural rosettes were cultivated using standard techniques [34,36]. Neurons emerged after rosettes were isolated and grown on Poly-dL-Ornithine/laminin coated plates. Neurons (generally an equal mix of glutamatergic and GABAergic neurons) were harvested after 14 days.

RNA-Seq was carried out using line iPSC-1, which was derived from a control female. Validation by Sanger sequencing was carried out on iPSC-2 (a control male), and 3 lines derived from male subjects with SZ (SZ39, SZ97 and iPSC-15, the latter of which has a 22q11.2 deletion).

RNA-Seq

Paired-end RNA-Seq was carried out using an Illumina HiSeq 2000 instrument, as described in Methods S1 and in our previously published work (GEO accession number GSE32625) [36]. In the current study, we reprocessed the previously reported RNA-Seq data to identify allele-biased gene expression [36]. The read length was 104 bases for each of the paired-end reads. A summary of the data relevant to this study is shown in Table 1.

Genomic DNA was analyzed using the Affymetrix Genome-Wide Human 6.0 array and genotypes were scored using the Affymetrix Power Tools (APT) software based on the Birdseed calling algorithm. A total of 905,625 SNPs were called (call rate was 98.98%), and the average rate of heterozygosity and homozygosity was 28.04% and 70.94%, respectively (253,585 heterozygous SNPs were identified). Parental DNA was also genotyped using the Affymetrix Genome-Wide Human 6.0 array (paternal call rate, 98.44%; maternal call rate 98.34%). To reduce the known bias in mapping RNA-Seq reads towards reference alleles, we aligned the raw RNA-Seq reads to a modified reference human genome (hg19) by the software bowtie 2 in which bases at the identified heterozygous sites were replaced with “N” [46–48]. Reads mapped to heterozygous sites were then analyzed. To reduce false positives, we focused on sites that were covered by at least 10 high-quality reads and excluded bases with a Phred score lower than 13. We used samtools/bcftools (mpileup command) to record high-quality reads mapped across each heterozygous SNP site. We then called the consensus base(s) (by maximizing the posterior probability given the read distribution and average base quality), and calculated the phred-scaled probability of the base being called a homozygote or heterozygote [49]. Subsequently, the numbers of reads for the reference and alternative alleles were counted and used for a binomial test to determine if the ratio of the two numbers significantly deviated from 0.5, reflecting the null hypothesis that both alleles were equally expressed. The resulting p-values were further adjusted by multiple-testing correction using the B-H method [50,51]. Sites with an adjusted p-value (i.e., FDR) <0.05 and a phred-scaled probability ≥ 20 were considered to be expressed in an allele-biased manner. These analyses were carried out for iPSCs and differentiating neurons separately; the resulting allele-biased expressed SNP sites are listed in Table S1. Genes with allele-biased expression were those containing SNPs shown in Table S1 that mapped between transcription start and termination sites, having an expression value >1 FPKM.

RNA-Seq reads for some genes showed a preference for one allele in iPSCs, with a switch to the other in differentiating neurons. Heterozygous SNPs showing >4-fold changes in allele preference during differentiation were tabulated (Table 2).

We also assessed potential sequencing chemistry and mapping bias for allele-biased SNP calls. There was no preference towards any one of the four nucleotides (Figure S1).

Selected monoallelic SNPs were validated by capillary sequencing using RT-PCR amplified material (see Methods S1).

Results

RNA-Seq was carried out on iPSCs and differentiating neurons harvested from iPSC-1 using procedure A, which leads primarily to the production of glutamatergic neurons [37]. Of the 253,585 heterozygous SNPs subjected to analysis for allele-biased expression, a total of 1,909 satisfied the criteria described in the methods section (Table S1). These mapped to a total of 314 genes in iPSCs and 801 in differentiating neurons (Tables S2 and S3, respectively). These included 181 neuronal genes that contain two or more highly significant allele-biased SNPs, of which several are candidate genes for SZ, ASD, and intellectual disability (see below). Based on genotyping data obtained from the parents of iPSC-1, the ratio of paternal to maternal allele-biased SNPs is 1.2:1 in iPSCs and 1.1:1 in neurons, excluding the X-chromosome, indicating that there was no significant parent-of-origin bias.

Overall, 24.5% of all genes expressed in an allele-biased manner in iPSCs and 12.4% in neurons were X-linked. The abundance of

Table 1. Summary of RNA-Seq and SNP array data.

	iPSCs	NEURONS
N of reads (pairs)	269,672,486	222,127,542
N of paired-match reads	121,620,928	115,280,074
N of single-match reads	39,276,866	33,721,936
Ratio of mapped fragments	74.30%	82.30%
Total SNP calls	905,625	905,625
Heterozygotes	253,585	253,585
Total SNPs (>= 1 read)	51,161	63,489
Total SNPs (>= 10 high quality reads)	8,358	13,576
Binomial (FDR adjusted p-value <0.05) and monoallelic consensus by samtools	543	1542
Number of genes (FPKM>1)	314	801

doi:10.1371/journal.pone.0044017.t001

X-linked genes is consistent with the finding that human iPSCs often maintain a clonal X-chromosome active state [52]. However, in other studies the pattern of X-chromosome inactivation has been found to be heterogeneous [53,54]. All of the X-linked monoallelic SNPs were maternal, with the exception of *XIST*, which is expressed on the inactive X-chromosome (paternal in this sample), and *WASF4P* and *KIF4*, suggesting that the strict criteria used for our definition of allele-biased expressed SNPs resulted in few false positives.

Validation of X-linked and Imprinted Genes by Sanger Sequencing

Technical factors can inflate the false positive rate when RNA-Seq data are used to evaluate allele-biased expression; consequently, we validated several genes using Sanger dideoxy sequencing [47]. We also analyzed iPSCs and neurons that were maintained and induced to differentiate using a different protocol to determine whether culture conditions might influence allele-biased expression (see Methods S1). Finally, when SNPs were informative, we validated allele-biased expression in other iPSC lines.

First, we analyzed four X-linked genes (*GABRA3*, *TSPAN6*, *NXT2*, and *KALI*). Heterozygosity for each was confirmed by sequencing gDNA (Figure 1). However, only one peak was detected when cDNA was sequenced from template RNA derived from iPSCs and neurons derived from the line iPSC-1 grown using protocol A, which was used in the RNA-Seq experiment. An exception is *KALI* in neurons, which showed allele-bias rather than complete monoallelic expression.

Each of the X-linked validated SNPs met the criteria described in the methods section except for rs6627588 in *GABRA3*. The expression of only a single allele shows that our conservative method for determining allele-biased expression is leading to some degree of type II error.

Interestingly, a somewhat different pattern was observed using iPSCs and neurons grown under condition B. Here there is also clear evidence for monoallelic expression in *GABRA3* and *TSPAN6* in both iPSCs and neurons, and in iPSCs for *KALI*. However, biallelic expression is seen for *KALI* in neurons and for *NXT2* in iPSCs and neurons, suggesting that X-activation/inactivation is somewhat variable under different culture/differentiation conditions.

X-linked markers could not be verified in the other iPSC lines since they were each derived from male subjects.

We also analyzed *CTNNA3*, which has been shown to be imprinted in placenta, and *KCNQ1*, a known imprinted gene [3,55,56]. *KCNQ1* was not included in the list shown in Tables S2 and S3 because the minimum FPKM >1 threshold was not achieved for this gene. Nevertheless, Sanger sequencing clearly shows allele-biased expression for both genes (Figure 2). Thus, similar to the expression pattern for *GABRA3*, this further illustrates that our allele-biased algorithm is associated with some type II error.

Two informative SNPs were analyzed in *CTNNA3*, rs1925621 and rs1925622. As seen in Figure 2, a single allele at rs1925621 is expressed in undifferentiated iPSCs and neurons derived from iPSC-1 grown using protocol A. However, when the cells were cultivated under protocol B, expression was essentially biallelic for both SNPs in iPSC-1. Interestingly, when a biological replicate of iPSC-1 was analyzed (iPSC-1 (II)), the other parental allele was activated in neurons grown using protocol A. In addition, in the replicate sample, allele-bias was also seen in neurons grown using protocol B. Similarly, when these SNPs were analyzed in another iPSC line (iPSC-2, which was derived from another control subject), allele-biased expression was seen in iPSCs and neurons cultivated under both protocols.

The same results were obtained when rs1925622 was analyzed.

These data confirm allele-biased expression for *CTNNA3* under some culture and differentiation conditions, and suggest that the phenomenon is random in this system and not imprinted.

Validation of SZ and ASD Candidate Genes

Considering the parent-of-origin effects seen in some families and the possibility that allele-biased expression could help explain MZ twin discordance and reduced penetrance, we were interested in specifically validating SZ and ASD candidate genes that showed evidence for allelic imbalance in the RNA-Seq experiment. Overall, there were 48 SZ candidate genes and 26 ASD candidates expressed in an allele-biased manner in neurons, while in iPSCs, there were 17 and 9, respectively (downloaded from Allen Brain Atlas list of 883 SZ and 244 ASD candidate genes) (Table S4). A Chi-square test was used to determine whether a statistically significant enrichment of SZ/ASD genes is found in the allele-biased gene sets. More specifically, a 2x2 table was set up to

Table 2. Allele switching during transition from iPSCs to neurons.

chr	SNP ID	ref	alt	iPSC ref	iPSC alt	Neurons ref	Neurons alt	Genes
chr1	rs12123760	T	C	2	8	21	5	ST6GALNAC3
chr1	rs10492963	C	T	1	13	13	1	RERE
chr2	rs3811568	A	G	11	2	2	9	FMNL2
chr2	rs6757809	T	C	2	13	10	2	NOL10
chr2	rs4674015	A	G	1	15	10	2	SPATS2L
chr3	rs9860614	A	G	8	2	0	11	RUVBL1
chr3	rs9817055	G	A	16	1	0	16	CLASP2
chr3	rs1975760	T	C	2	16	13	3	ZNF148
chr3	rs3846072	C	T	1	9	11	1	MED12L/P2RY14
chr3	rs7644975	G	A	1	10	23	4	ERC2
chr4	rs7678728	T	G	1	9	10	2	ANK2
chr4	rs236985	G	A	2	8	9	2	AFF1
chr5	rs7712332	A	G	13	2	2	11	MSH3
chr5	rs2115436	T	A	2	10	13	2	SCAMP1
chr6	rs9373571	T	A	22	1	2	8	ASCC3
chr6	rs577372	A	G	2	9	11	2	MTO1
chr7	rs2691561	T	A	13	3	2	15	SNX13
chr7	rs258654	C	T	9	1	1	14	CACNA2D1
chr7	rs674462	A	T	2	9	9	2	MRPL32
chr8	rs6987331	T	G	12	3	1	9	SNTB1
chr8	rs2980683	G	C	1	10	10	1	AGPAT5
chr9	rs7019027	G	A	12	3	0	11	KDM4C
chr9	rs1411675	G	A	13	3	1	9	FXN
chr9	rs2297499	C	G	9	2	4	16	TLE4
chr9	rs7020390	A	G	2	9	10	2	UHRF2
chr10	rs1127047	G	A	9	1	2	8	PITRM1
chr10	rs1937971	A	C	17	0	3	16	NRG3
chr10	rs11817793	G	A	8	2	1	10	LARP4B
chr10	rs2418929	C	G	12	2	1	10	KIF20B
chr10	rs10795321	C	G	0	13	32	5	FAM188A
chr10	rs7084542	A	G	3	14	39	9	RP11-429G19.2
chr12	rs4130296	A	C	3	15	22	3	FBXW8
chr13	rs11619378	T	G	12	2	2	16	ABHD13
chr13	rs7983071	G	A	10	2	1	20	GPC6
chr13	rs9535499	G	C	3	34	21	4	DLEU7/DLEU7-AS1
chr14	rs12893288	C	T	16	4	3	13	intergenetic
chr14	rs10131048	T	C	3	14	15	3	ZFYVE26
chr16	rs17139246	T	C	10	2	0	14	RBFox1
chr17	rs2106663	G	A	2	10	9	1	BRIP1
chr17	rs2302235	G	A	4	21	9	2	FAM20A
chr17	rs9646411	C	T	0	10	10	0	AMAC1L3
chr18	rs7233697	T	C	8	2	1	17	MEP1B

Genes that showed >4 changes in allele distribution occurring during the transition from iPSCs to neurons. Abbreviations: chromosome number (chr); SNP identification number (SNP ID); reference allele (ref); alternative allele (alt); high quality reads in iPSCs (iPSC ref and iPSC alt); high quality reads in neurons (Neuron ref and Neurons alt).

doi:10.1371/journal.pone.0044017.t002

compare the allelic expression status (biased vs non-biased) of SZ/ASD candidates vs. all others; genes in both groups were selected for containing at least one SNP covered by ≥ 10 reads. A

significant enrichment was found for SZ/ASD candidate genes expressed in neurons ($p = 0.02$), while no significant enrichment was observed in iPSCs ($p = 0.14$).

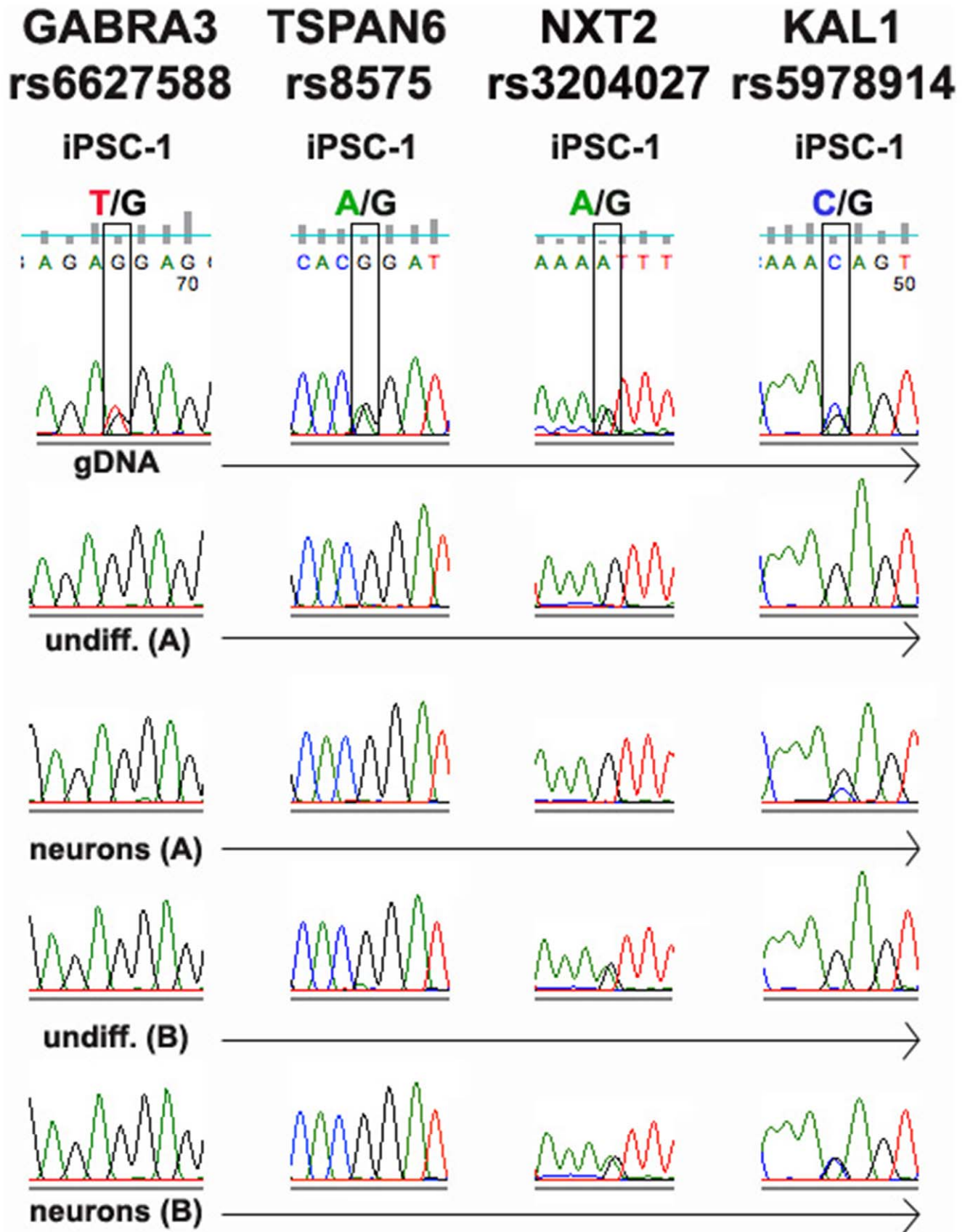


Figure 1. Sanger sequencing of selected X-linked genes. Top panel in each strip is sequence of genomic DNA (gDNA) for iPSC-1, a control subject, confirming heterozygosity of each SNP. Other strips are sequences of cDNA derived from undifferentiated (undiff) and neurons cultivated

under growth and differentiation conditions A and B (see Methods S1 for details). The RNA samples used to generate the Sanger sequencing data for iPSC-1 protocol A (undifferentiated iPSCs and neurons) were the same samples used for RNA-Seq.
doi:10.1371/journal.pone.0044017.g001

The SZ and ASD genes that showed the most robust evidence for allele-biased expression, based on containing more than one significant SNP, were *RBFOX1* (*A2BP1*) (15 SNPs), *NLGN4X* (14 SNPs), *NRG3* (10 SNPs), *NRG1* (9 SNPs), *CASK* (7 SNPs), and *ERBB4*, *CNTNAP2*, *NLGN1*, and *NPAS3*, each of which had 3 SNPs. Other genes of interest, such as *NRXN1* and *DISC1*, were only represented by one SNP.

We validated allele-biased expression in five SZ and ASP candidate genes, including *A2BP1* (Figure 3). RNA-Seq data showed that in iPSC-1, one parental allele was expressed at this locus in undifferentiated iPSCs (T allele), and the other (C allele) in neurons (Table S1; compare columns H, I, N and O for rs17139246). This was one of 42 genes that showed evidence for a switch from one parental allele to the other during the transition from iPSCs to neurons (Table 2). The switch in *A2BP1* alleles was confirmed by Sanger sequencing (Figure 3; undifferentiated iPSCs and neurons grown under protocol A, which was produced from the same sample used for RNA-Seq). However, when grown under protocol B, expression is essentially biallelic (there is a hint of allele-biased expression, but this should be viewed cautiously because capillary sequencing is only semi-quantitative).

Switching from one allele to the other could either reflect random activation of parental alleles occurring independently in iPSCs and neurons, or programmed expression of one parental allele in iPSCs and the other in neurons. Random expression is supported in the analysis of a biological replicate; iPSC-1 (II). As seen in Figure 3, the expression pattern was different compared to the original sample; some degree of allele-biased expression seen only for the “C” allele in iPSCs and neurons grown under protocol A suggesting that allele-biased expression in *A2BP1* during early neurogenesis is random. Whether other genes shown in Table 2 switch allelic preference during differentiation by the same random mechanism is under investigation.

Further validation of allele-biased expression in *A2BP1* is seen in three other iPSC lines, each of which is heterozygous at rs17139246. Allelic bias and a switch were clearly seen in SZ39 with protocol A. With protocol B however the predominate allele was C, as opposed to T in neurons grown under protocol A. These findings support the idea that allele-biased expression of *A2BP1* is a stochastic process.

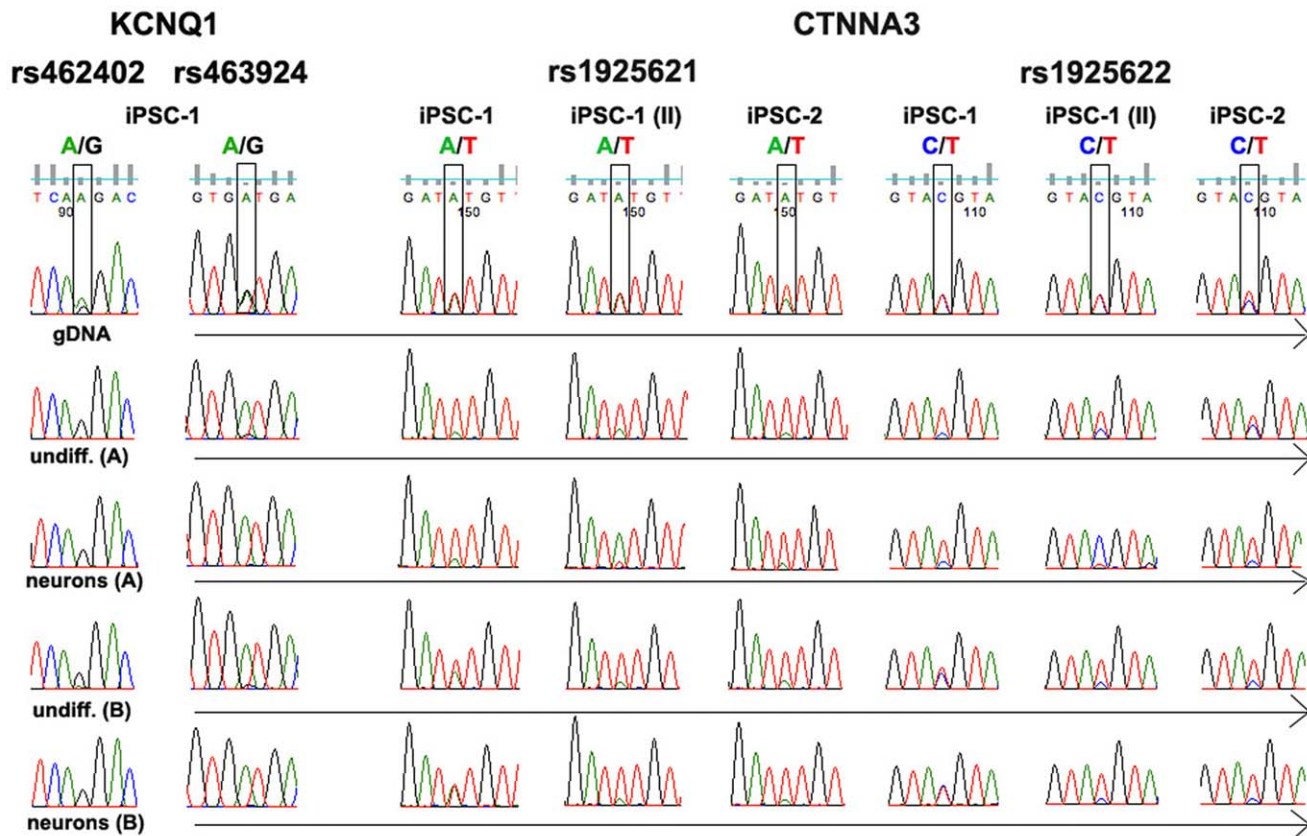


Figure 2. Imprinted genes. SNPs in *KCNQ1* and *CTNNA3* were validated by Sanger sequencing. See legend Figure 1 for details. iPSC-1 (II) is a biological replicate of iPSC-1.
doi:10.1371/journal.pone.0044017.g002

A2BP1 rs17139246

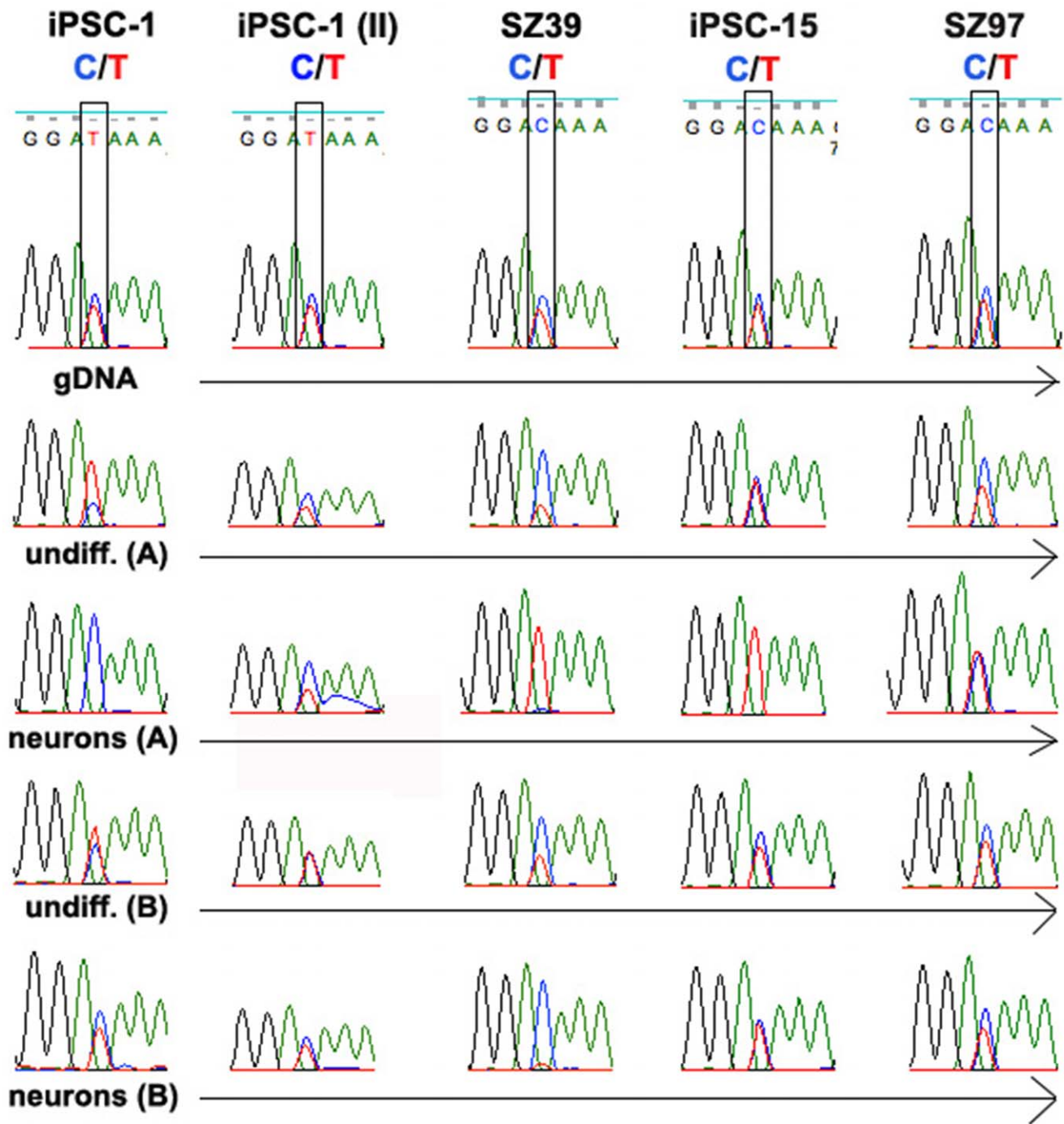


Figure 3. Validation of *A2BP1*. SZ39, SZ97 and iPSC-15 are iPSC lines developed using fibroblasts from patients with SZ.
doi:10.1371/journal.pone.0044017.g003

For the other two lines, iPSC-15 showed allele-biased expression protocol A neurons, but not in undifferentiated cells, while in SZ97, expression was essentially biallelic.

Validation of *NRG1* and *ERBB4* is shown in Figure 4. For *NRG1*, two SNPs were analyzed - rs4602844 and rs1481757, which map near the promoter of the *NRG1* isoform HRG- β 1C and an intron in the *GGF2*/*HRG*- β 1D isoforms, respectively.

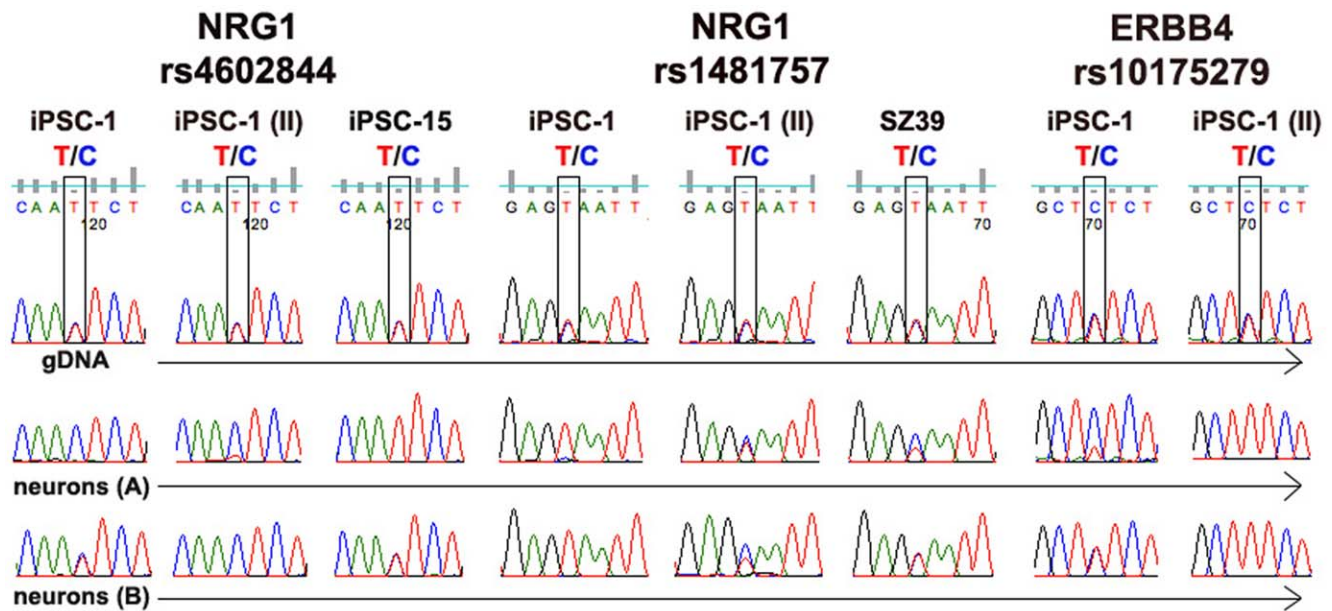


Figure 4. Validation of *NRG1* and *ERBB4*. The *NRG1* SNPs rs4602844 and rs1481757 map near the promoter of the *NRG1* isoform HRG- β 1C, and an intron in the GGF2/HRG- β 1D isoforms, respectively.
doi:10.1371/journal.pone.0044017.g004

Sanger sequencing confirmed the RNA-Seq allele-biased calls in iPSC-1 neurons for rs4602844 (expression was too low in undifferentiated cells for reliable sequencing). In addition, we were able to confirm allele-biased expression for this SNP in the replicate sample, iPSC-1 (II), and in iPSC-15. With protocol B, a different pattern emerged. Neurons derived from both iPSC-1 and iPSC-15 showed biallelic expression, but in iPSC-1 (II), allele bias was clearly seen. For rs1481757, allele-biased expression was seen in iPSC-1 but not in iPSC-1 (II), for both protocol A and B neurons. For SZ39, there was a suggestion of allele-biased expression only in protocol A.

These findings suggest that isoforms within the same gene may be subject to different regulatory signals with respect to allele-biased expression.

For rs10175279 (*ERBB4*) allele-biased expression of the C allele was seen in iPSC-1 in protocol A. However, in iPSC-1 (II), the other parental allele (T) is expressed. Furthermore, while biallelic expression was detected in iPSC-1 in protocol B, expression was biased towards the T allele in the replicate. These findings suggest that similar to *A2BP1*, allele-biased expression in *ERBB4* is random.

Finally, Sanger sequencing was also carried out for SNPs in *NRG3* and *AUTS2*. This analysis essentially showed a similar mixed pattern of allele-biased or biallelic expression for neurons that varied according to growth and differentiation protocol (Figure S2).

To summarize, the validation of X-linked, imprinted, and SZ and ASD candidate genes confirms the monoallelic SNP calls made in the RNA-Seq analysis and demonstrates that the phenomenon is not restricted to a single iPSC line. In addition, allele-biased expression is random for some genes, and can vary depending on culture conditions. These findings are similar to the change that occurs in *NANOG* mRNA (from a monoallelic to biallelic pattern of expression) in ES cells treated under different

culture conditions, as well observations made on allele-biased gene expression in lymphoblasts [27,57].

Discussion

Using RNA-Seq to identify allele-biased expression genome-wide has the capacity to add a novel dimension to our understanding of transcriptional regulation in different cell types. However, interpreting RNA-Seq data is complicated and can be confounded by false positive allele-biased calls [47]. One of the more important factors is *in silico* mapping bias when RNA-Seq reads are aligned to the reference genome, as described in the methods section. In this study, the reference allele was called with a 1.5:1 bias (Figure 5), despite precautions taken while aligning RNA-Seq reads (see methods). The cause of this persistent bias is under investigation, but does not appear to arise from sequencing bias (Figure S1). Nevertheless, the validation studies strongly support the RNA-Seq findings.

Allele-biased expression could be caused by epigenetic phenomena or a genetic process, such as copy loss or promoter variants. It has been suggested that most allele-biased expression is due to genetic factors [47,48]. This may well be the case for many of the genes in this study. However, several of the genes we validated, including *CTNNA3* and the SZ candidate genes *A2BP1* and *ERBB4*, showed random allele-biased expression suggesting an epigenetic phenomenon. Distinguishing between genetic and epigenetic processes (both random monoallelic expression and imprinting) will require a much more extensive analysis using several different lines and multiple replicates. This is an important consideration in SZ and ASD because an epigenetic process accounting for allele-biased expression could have an impact on determining the efficacy of epigenetics-based treatment. Confirmation in differentiating mouse and/or human neurons derived from embryonic neural progenitors is also needed.

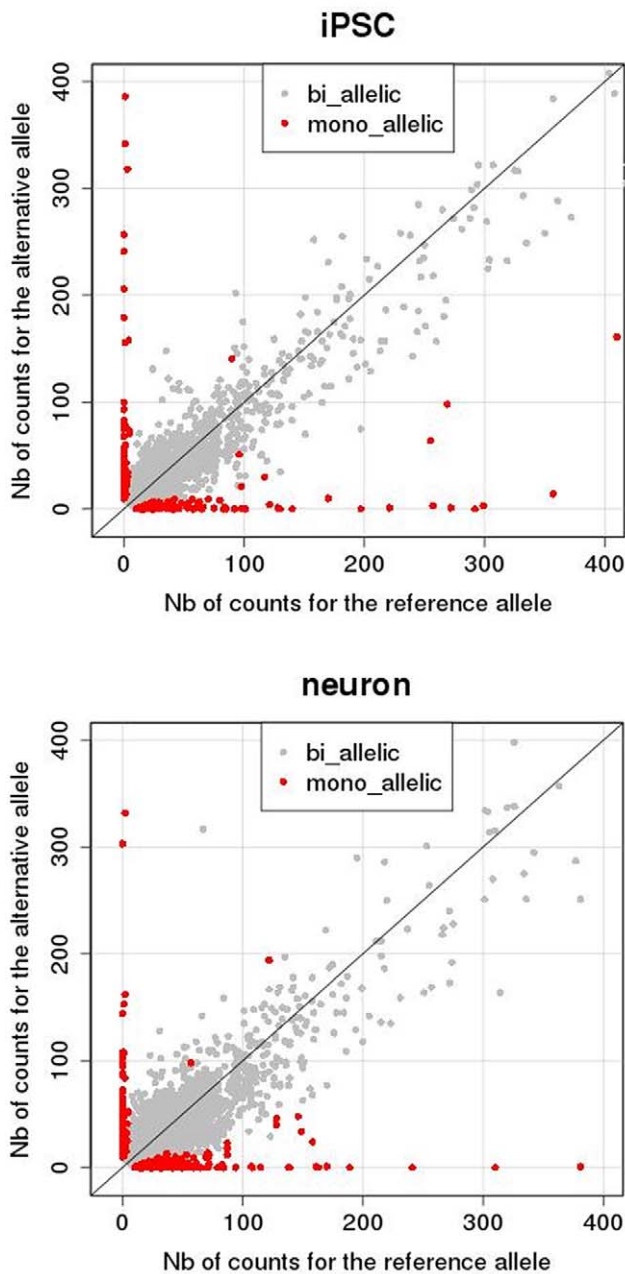


Figure 5. Reference vs alternative RNA-Seq reads at heterozygous SNPs. Y-axis, number (Nb) of reads with the alternative allele (211 in iPSCs, 619 in neurons); X-axis, number of reads with the reference allele (332 in iPSCs, 923 in neurons). Biallelic (grey); allele-biased expression (black).
doi:10.1371/journal.pone.0044017.g005

Finally, allele-biased expression, in particular a stochastic model, has the potential to explain the ~50% MZ twin discordance seen in SZ [28–31]. Here, a disease-causing mutant allele would have an equal probability of being expressed (causing

References

1. Delaval K, Feil R (2004) Epigenetic regulation of mammalian genomic imprinting. *Curr Opin Genet Dev* 14(2): 188–195. 10.1016/j.gde.2004.01.005.
2. Holmes R, Soloway PD (2006) Regulation of imprinted DNA methylation. *Cytogenet Genome Res* 113(1–4): 122–129. 10.1159/000090823.

disease), or suppressed (avoiding disease) if such a process occurred in a pathway or cortical segment that was clonally derived. A stochastic model is also consistent with the observation that the prevalence of SZ in the children of discordant and concordant MZ twins is similar [28].

GWAS and CNV discovery have resulted in the identification of numerous candidate genes for neuropsychiatric disorders, a number that is bound to increase dramatically with rare variants detected using exome and whole genome sequencing. Translating the heterogeneous mix of susceptibility genes into novel medications could prove to be very challenging. This study presents the possibility of incorporating seemingly disparate candidate genes into a common molecular pathway – epigenetic regulation leading to allele-biased gene expression. Based on this idea, it is conceivable that one approach to treating a subgroup of patients would be at the epigenetic level (assuming a non-genetic cause of allele-biased expression) using drugs capable of activating a dormant normal allele as a means to buffer the effects of their monoallelically expressed abnormal partner. Proof of principle for this concept has recently been demonstrated in an animal model of Angelman Syndrome [58].

Supporting Information

Figure S1 Distribution of nucleotides for allele-biased SNPs.

(JPG)

Figure S2 validation of replicates and additional genes.

(JPG)

Table S1 Heterozygous SNPs showing allele-biased expression.

(XLS)

Table S2 allele-biased genes in undifferentiated iPSCs.

(XLS)

Table S3 allele-biased genes in neurons.

(XLS)

Table S4 SZ and ASD genes showing allele-biased expression in iPSCs and neurons (from Allen Brain Atlas list of 883 SZ and 244 ASD candidate genes).

(XLSX)

Methods S1 Detailed methods, reagents and PCR primers.

(DOC)

Acknowledgments

The authors thank participating families, Dr. Vladislav Sandler for providing two iPSC lines, Dr. Steven Cohen for doing skin biopsies, and Drs. Shahina Maqbool and John Grealley of the Einstein Center for Epigenomics for carrying out RNA-Seq.

Author Contributions

Conceived and designed the experiments: HL DZ. Performed the experiments: AH EP. Analyzed the data: ML DZ TW. Wrote the paper: HL DZ ML.

4. Radford EJ, Ferron SR, Ferguson-Smith AC (2011) Genomic imprinting as an adaptive model of developmental plasticity. *FEBS Lett* 585(13): 2059–2066. 10.1016/j.febslet.2011.05.063.
5. Edwards CA, Rens W, Clarke O, Mungall AJ, Hore T, et al. (2007) The evolution of imprinting: Chromosomal mapping of orthologous of mammalian imprinted domains in monotreme and marsupial mammals. *BMC Evol Biol* 7: 157. 10.1186/1471-2148-7-157.
6. Gregg C, Zhang J, Butler JE, Haig D, Dulac C (2010) Sex-specific parent-of-origin allelic expression in the mouse brain. *Science* 329(5992): 682–685. 10.1126/science.1190831.
7. Gregg C, Zhang J, Weissbourd B, Luo S, Schroth GP, et al. (2010) High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* 329(5992): 643–648. 10.1126/science.1190830.
8. Deveale B, van der Kooy D, Babak T (2012) Critical evaluation of imprinted gene expression by RNA-seq: A new perspective. *PLoS Genet* 8(3): e1002600. 10.1371/journal.pgen.1002600.
9. Cassidy SB, Schwartz S, Miller JL, Driscoll DJ (2011) Prader-will syndrome. *Genet Med*. 10.1097/GIM.0b013e31822bead0.
10. Fradin D, Cheslack-Postava K, Ladd-Acosta C, Newschaffer C, Chakravarti A, et al. (2010) Parent-of-origin effects in autism identified through genome-wide linkage analysis of 16,000 SNPs. *PLoS One* 5(9): e12513. 10.1371/journal.pone.0012513.
11. Francks C, Tozzi F, Farmer A, Vincent JB, Rujescu D, et al. (2008) Population-based linkage analysis of schizophrenia and bipolar case-control cohorts identifies a potential susceptibility locus on 19q13. *Mol Psychiatry*. 10.1038/mp.2008.100.
12. Berrettini WH, Ferraro TN, Goldin LR, Weeks DE, Detera-Wadleigh S, et al. (1994) Chromosome 18 DNA markers and manic-depressive illness: Evidence for a susceptibility gene. *Proc Natl Acad Sci U S A* 91(13): 5918–5921.
13. Ingason A, Kirov G, Giegling I, Hansen T, Isles AR, et al. (2011) Maternally derived microduplications at 15q11-q13: Implication of imprinted genes in psychotic illness. *Am J Psychiatry* 168(4): 408–417. 10.1176/appi.ajp.2010.09111660.
14. Crespi B, Badcock C (2008) Psychosis and autism as diametrical disorders of the social brain. *Behav Brain Sci* 31(3): 241–61; discussion 261–320. 10.1017/S0140525X08004214.
15. Badcock C, Crespi B (2006) Imbalanced genomic imprinting in brain development: An evolutionary basis for the aetiology of autism. *J Evol Biol* 19(4): 1007–1032. 10.1111/j.1420-9101.2006.01091.x.
16. Kopsida E, Mikaelsson MA, Davies W (2011) The role of imprinted genes in mediating susceptibility to neuropsychiatric disorders. *Horm Behav* 59(3): 375–382. 10.1016/j.yhbeh.2010.04.005.
17. Pun FW, Zhao C, Lo WS, Ng SK, Tsang SY, et al. (2011) Imprinting in the schizophrenia candidate gene GABRB2 encoding GABA(A) receptor beta(2) subunit. *Mol Psychiatry* 16(5): 557–568. 10.1038/mp.2010.47.
18. Kitiou-Tzeli S, Tzetzis M, Sofocleous C, Vrettou C, Xaidara A, et al. (2010) De novo interstitial duplication of the 15q11.2-q14 PWS/AS region of maternal origin: Clinical description, array CGH analysis, and review of the literature. *Am J Med Genet A* 152A(8): 1925–1932. 10.1002/ajmg.a.33447.
19. Kwasnicka-Crawford DA, Roberts W, Scherer SW (2007) Characterization of an autism-associated segmental maternal heterodisomy of the chromosome 15q11-13 region. *J Autism Dev Disord* 37(4): 694–702. 10.1007/s10803-006-0225-8.
20. Simic M, Turk J (2004) Autistic spectrum disorder associated with partial duplication of chromosome 15; three case reports. *Eur Child Adolesc Psychiatry* 13(6): 389–393. 10.1007/s00787-004-0414-y.
21. Wilkins JF (2010) Antagonistic coevolution of two imprinted loci with pleiotropic effects. *Evolution* 64(1): 142–151. 10.1111/j.1558-5646.2009.00826.x.
22. Smit H (2010) A conceptual contribution to battles in the brain. *Biol Philos* 25(5): 803–821. 10.1007/s10539-010-9195-5.
23. Roppolo D, Vollery S, Kan CD, Luscher C, Broillet MC, et al. (2007) Gene cluster lock after pheromone receptor gene choice. *EMBO J* 26(14): 3423–3430. 10.1038/sj.emboj.7601782.
24. Krueger C, Morison IM (2008) Random monoallelic expression: Making a choice. *Trends Genet* 24(6): 257–259. 10.1016/j.tig.2008.03.005.
25. Magklara A, Yen A, Colquitt BM, Clowney EJ, Allen W, et al. (2011) An epigenetic signature for monoallelic olfactory receptor expression. *Cell* 145(4): 555–570. 10.1016/j.cell.2011.03.040.
26. Yagi T (2008) Clustered protocadherin family. *Dev Growth Differ* 50 Suppl 1: S131–40. 10.1111/j.1440-169X.2008.00991.x.
27. Gimelbrant A, Hutchinson JN, Thompson BR, Chess A (2007) Widespread monoallelic expression on human autosomes. *Science* 318(5853): 1136–1140. 10.1126/science.1148910.
28. Gottesman II, Bertelsen A (1989) Confirming unexpressed genotypes for schizophrenia. risks in the offspring of Fischer's Danish identical and fraternal discordant twins. *Arch Gen Psychiatry* 46(10): 867–872.
29. Folstein S, Rutter M (1977) Infantile autism: A genetic study of 21 twin pairs. *J Child Psychol Psychiatry* 18(4): 297–321.
30. Allen MG (1976) Twin studies of affective illness. *Arch Gen Psychiatry* 33(12): 1476–1478.
31. Tsuang M (2000) Schizophrenia: Genes and environment. *Biol Psychiatry* 47(3): 210–220.
32. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, et al. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131(5): 861–872. 10.1016/j.cell.2007.11.019.
33. Brennand KJ, Simone A, Jou J, Gelboin-Burkhardt C, Tran N, et al. (2011) Modelling schizophrenia using human induced pluripotent stem cells. *Nature*. 10.1038/nature09915.
34. Marchetto MC, Carroneu C, Acab A, Yu D, Yeo GW, et al. (2010) A model for neural development and treatment of rett syndrome using human induced pluripotent stem cells. *Cell* 143(4): 527–539. 10.1016/j.cell.2010.10.016.
35. Chiang CH, Su Y, Wen Z, Yoritomo N, Ross CA, et al. (2011) Integration-free induced pluripotent stem cells derived from schizophrenia patients with a DISC1 mutation. *Mol Psychiatry*. 10.1038/mp.2011.13.
36. Lin M, Pedrosa E, Shah A, Hrabovsky A, Maqbool S, et al. (2011) RNA-seq of human neurons derived from iPSCs reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. *PLoS One* 6(9): e23356. 10.1371/journal.pone.0023356.
37. Pedrosa E, Sandler V, Shah A, Carroll R, Chang C, et al. (2011) Development of patient-specific neurons in schizophrenia using induced pluripotent stem cells. *J Neurogenet*. 10.3109/01677063.2011.597908.
38. Ginsberg SD, Mirnic K (2006) Functional genomic methodologies. *Prog Brain Res* 158: 15–40. 10.1016/S0079-6123(06)58002-1.
39. Deep-Soboslay A, Iglesias B, Hyde TM, Bigelow LB, Imamovic V, et al. (2008) Evaluation of tissue collection for postmortem studies of bipolar disorder. *Bipolar Disord* 10(7): 822–828. 10.1111/j.1399-5618.2008.00607.x.
40. Hill C, Keks N, Roberts S, Opekin K, Dean B, et al. (1996) Problem of diagnosis in postmortem brain studies of schizophrenia. *Am J Psychiatry* 153(4): 533–537.
41. Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, et al. (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 321(5891): 956–960. 10.1126/science.1160342.
42. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18(9): 1509–1517. 10.1101/gr.079558.108.
43. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat Methods* 5(7): 621–628. 10.1038/nmeth.1226.
44. Feng J, Li W, Jiang T (2011) Inference of isoforms from short sequence reads. *J Comput Biol* 18(3): 305–321. 10.1089/cmb.2010.0243.
45. Li SM, Valo Z, Wang J, Gao H, Bowers CW, et al. (2012) Transcriptome-wide survey of mouse CNS-derived cells reveals monoallelic expression within novel gene families. *PLoS One* 7(2): e31751. 10.1371/journal.pone.0031751.
46. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3): R25. 10.1186/gb-2009-10-3-r25.
47. Heap GA, Yang JH, Downes K, Healy BC, Hunt KA, et al. (2010) Genome-wide analysis of allelic expression imbalance in human primary cells by high-throughput transcriptome resequencing. *Hum Mol Genet* 19(1): 122–134. 10.1093/hmg/ddp473.
48. Degner JF, Marioni JC, Pai AA, Pickrell JK, Nkadori E, et al. (2009) Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics* 25(24): 3207–3212. 10.1093/bioinformatics/btp579.
49. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16): 2078–2079. 10.1093/bioinformatics/btp352.
50. Benjamini Y, Heller R, Yekutieli D (2009) Selective inference in complex research. *Philos Transact A Math Phys Eng Sci* 367(1906): 4255–4271. 10.1098/rsta.2009.0127.
51. Hochberg Y, Benjamini Y. (1990) More powerful procedures for multiple significance testing. *Stat Med* 9(7): 811–818.
52. Tchiew J, Kuoy E, Chin MH, Trinh H, Patterson M, et al. (2010) Female human iPSCs retain an inactive X chromosome. *Cell Stem Cell* 7(3): 329–342. 10.1016/j.stem.2010.06.024.
53. Bruck T, Benvenisty N (2011) Meta-analysis of the heterogeneity of X chromosome inactivation in human pluripotent stem cells. *Stem Cell Res* 6(2): 187–193. 10.1016/j.scr.2010.12.001.
54. Kim DH, Jeon Y, Anguera MC, Lee JT (2011) X-chromosome epigenetic reprogramming in pluripotent stem cells via noncoding genes. *Semin Cell Dev Biol*. 10.1016/j.semcdb.2011.02.025.
55. Kim KP, Thurston A, Mummery C, Ward-van Oostwaard D, Priddle H, et al. (2007) Gene-specific vulnerability to imprinting variability in human embryonic stem cell lines. *Genome Res* 17(12): 1731–1742. 10.1101/gr.6609207.
56. Oudejans CB, Mulders J, Lachmeijer AM, van Dijk M, Konst AA, et al. (2004) The parent-of-origin effect of 10q22 in pre-eclamptic females coincides with two regions clustered for genes with down-regulated expression in androgenetic placentas. *Mol Hum Reprod* 10(8): 589–598. 10.1093/molehr/gah080.
57. Miyanari Y, Torres-Padilla ME (2012) Control of ground-state pluripotency by allelic regulation of nanog. *Nature*. 10.1038/nature10807; 10.1038/nature10807.
58. Huang HS, Allen JA, Mabb AM, King IF, Miriyala J, et al. (2011) Topoisomerase inhibitors unsilence the dormant allele of Ube3a in neurons. *Nature* 21;481(7380): 185–189. Source 10.1038/nature10726.