



# miRNA-mediated risk for schizophrenia in 22q11.2 deletion syndrome

Linda M. Brzustowicz<sup>1,2\*</sup> and Anne S. Bassett<sup>3,4,5</sup>

<sup>1</sup> Department of Genetics, Rutgers University, Piscataway, NJ, USA

<sup>2</sup> Human Genetics Institute of New Jersey, Rutgers University, Piscataway, NJ, USA

<sup>3</sup> Clinical Genetics Research Program, Centre for Addiction and Mental Health, Toronto, ON, Canada

<sup>4</sup> Department of Psychiatry, University of Toronto, Toronto, ON, Canada

<sup>5</sup> Department of Psychiatry and Division of Cardiology, Department of Medicine, University Health Network, Toronto, ON, Canada

## Edited by:

Peng Jin, Emory University School of Medicine, USA

## Reviewed by:

Timothy Bowen, Cardiff University

School of Medicine, UK

Ricardo Silva, Universidad Simón

Bolívar, Venezuela

## \*Correspondence:

Linda M. Brzustowicz, Department of Genetics, Rutgers University, 145 Bevier Road, Room 231, Piscataway, NJ 08854, USA.  
e-mail: brzustowicz@biology.rutgers.edu

In humans, the most common genomic disorder is a hemizygous deletion of a 1.5–3 Mb region of chromosome 22q11.2. The resultant 22q11.2 deletion syndrome (22q11.2DS) can affect multiple organ systems, and most notably includes cardiac, craniofacial, and neurodevelopmental defects. Individuals with 22q11.2DS have a 20–25-fold risk of developing schizophrenia compared to individuals from the general population, making 22q11.2DS the strongest known molecular genetic risk factor for schizophrenia. Although the deleted region includes *DGCR8*, a gene coding for a miRNA processing protein, the exact mechanism by which this deletion increases risk is unknown. Importantly, several lines of evidence suggest that miRNAs may modulate risk for schizophrenia in other, non-22q11.2DS populations. Here we present a theory which mechanistically explains the link between 22q11.2DS, miRNAs, and schizophrenia risk. We outline the testable predictions generated by this theory and present preliminary data in support of our model. Further experimental validation of this model could provide important insights into the etiology of both 22q11.2DS and more common forms of schizophrenia.

**Keywords:** schizophrenia, genetic, miRNA, *DGCR8*, copy number variation

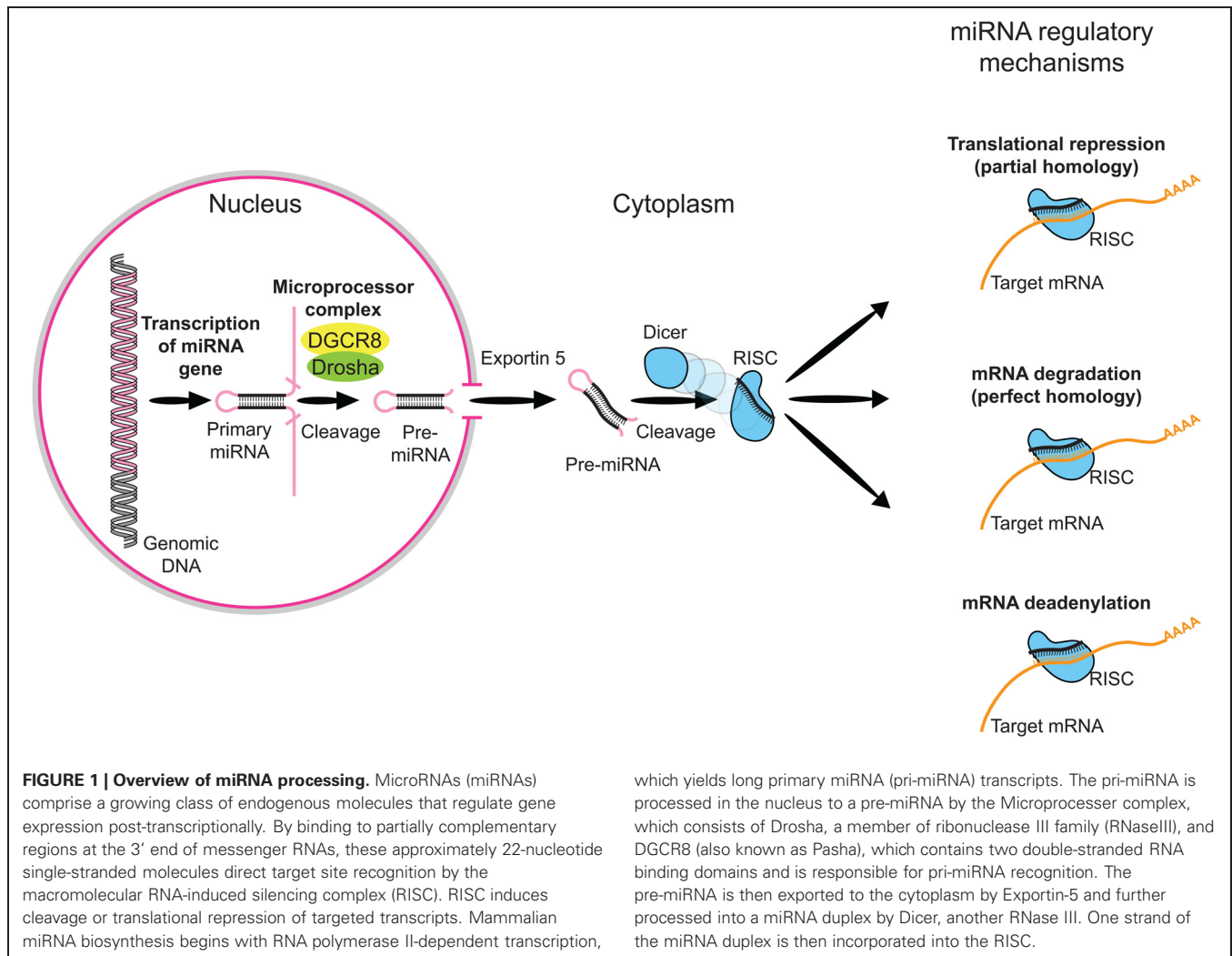
## BACKGROUND

Microdeletion of chromosome 22q11.2 or 22q11.2 deletion syndrome (22q11.2DS) (MIM #188400/#192430) is the most common human deletion syndrome with an estimated prevalence of 1 in 4000 live births (Goodship et al., 1998). Formerly known as DiGeorge or velocardiofacial syndromes, the expression is variable in severity and number of associated features (Bassett et al., 2011). While classically these include developmental features such as congenital cardiac and palatal anomalies, an often subtle facial phenotype and developmental delays and/or learning difficulties, later onset conditions are also commonly associated. These include hypocalcemia and thyroid abnormalities as well as psychiatric illnesses (Bassett et al., 2011). Over 60% of patients develop treatable psychiatric disorders by adulthood (Fung et al., 2010). In particular, about one in every four adults develops schizophrenia (Fung et al., 2010) and due to this high prevalence in 22q11.2DS patients, the 22q11.2 region is considered to be one of the main schizophrenia susceptibility loci in humans (Bassett and Chow, 2008; Insel, 2010).

The 22q11.2 region is flanked by low copy repeats (LCRs) with greater than 97% DNA sequence identity. As for other genomic disorders (Kumar, 2008), the main mechanism giving rise to the hemizygous 22q11.2 deletion involves non-allelic homologous recombination that results from misalignment of LCRs during meiosis (Edelmann et al., 1999). Up to 90% of recombination events give rise to the common 3 Mb deletion (Shaikh et al., 2000), with approximately 8% of patients having a 1.5 Mb deletion, nested within the proximal region of the 3 Mb deletion.

Rare smaller or atypical deletions have been reported but there is no evidence for specific genotype–phenotype correlations (Bassett et al., 2011). Over 90% of cases are found to have a *de novo* 22q11.2 deletion with the remainder found to be inherited from a parent. The commonly deleted region in 22q11.2DS encompasses approximately 45 genes and the consequences of decreased gene dosage of multiple genes are believed to be involved in phenotypic expression (Meechan et al., 2011). The molecular substrates underlying the different clinical features of 22q11.2 deletion, however, are still debated. One gene of particular interest, in the proximal deletion region, is DiGeorge Critical Region Gene 8 (*DGCR8*), which plays a critical role in the biogenesis of miRNAs (Figure 1).

Multiple lines of evidence have implicated miRNAs in general aspects of neurodevelopment, neurodegeneration, and brain function and specifically in schizophrenia (reviewed in Beveridge and Cairns, 2012; Mellios and Sur, 2012; Saito and Saito, 2012). A very large genome-wide association study of over 50,000 individuals identified a SNP in miR-137 as strongly ( $p = 1.6 \times 10^{-11}$ ) associated with schizophrenia (Ripke et al., 2011). In addition, five genes associated with schizophrenia (*ZNF804A*, *CACNA1C*, *TCF4*, *CSMD1*, and *C10orf26*) have all been validated as targets of miR-137 (Kwon et al., 2011; Kim et al., 2012), suggesting that genes involved in schizophrenia susceptibility can be affected by either *cis*-acting mutations at the gene or via alterations in the miRNA regulatory system. *DICER1*, another critical gene for miRNA biogenesis, has also been potentially implicated in schizophrenia by a *de novo* CNV associated



which yields long primary miRNA (pri-miRNA) transcripts. The pri-miRNA is processed in the nucleus to a pre-miRNA by the Microprocessor complex, which consists of Drosha, a member of ribonuclease III family (RNaseIII), and DGCR8 (also known as Pasha), which contains two double-stranded RNA binding domains and is responsible for pri-miRNA recognition. The pre-miRNA is then exported to the cytoplasm by Exportin-5 and further processed into a miRNA duplex by Dicer, another RNase III. One strand of the miRNA duplex is then incorporated into the RISC.

with illness (Xu et al., 2008). Multiple suggestive level associations with schizophrenia have also been reported with SNPs in miRNA pathway genes (Zhang et al., 2012). Numerous post-mortem studies of miRNA expression in human brain have reported alterations in tissue from individuals with schizophrenia (reviewed in Beveridge and Cairns, 2012). Of particular interest, Moreau et al. (2011) reported a significant overlap between the miRNAs dysregulated in human postmortem tissue from individual with psychotic illness and the miRNAs previously reported dysregulated in brain tissue from a mouse model of 22q11.2DS (Stark et al., 2008). These findings support the involvement of miRNAs in schizophrenia via multiple mechanisms and suggest that there may be overlap in the specific miRNAs involved in 22q11.2DS and non-22q11.2DS mediated cases of schizophrenia.

## THE PUZZLE

Despite the strikingly increased risk for schizophrenia in 22q11.2DS, approximately 75% of individuals with a 22q11.2 deletion do not develop schizophrenia. Other high-risk features of this syndrome, e.g., severe congenital cardiac defects, show

comparable levels of reduced penetrance. A number of possible explanations for this variability exist, with these factors potentially acting alone or in combination.

## CHANCE VARIATION

Stochastic processes can result in individuals with identical genetic endowments and apparently identical environmental exposures developing along different pathways. One could imagine some key cellular factor with a threshold effect on development that is hovering right at the critical level in individuals with 22q11.2DS. Some individuals will be fortunate enough to have just enough of the factor to proceed down the usual developmental pathway, while others may fall just short due to random fluctuations in expression and proceed down an alternative pathway.

## ENVIRONMENTAL FACTORS

Triggering or protective environmental factors can interact with identical genomes to produce divergent outcomes in different individuals. Even the intrauterine environment of twins differs, as evidenced by frequently discordant *in utero* HIV transmission

(Biggar et al., 2003). The potential divergence of environmental exposures for an adult onset disorder such as schizophrenia would be expected to be even greater.

### VARIATIONS IN DELETION SIZE

The 22q11.2 region contains a set of low copy number repeats that can mediate a variety of specific deletion events. Deletions mediated by different specific LCRs will result in different sets of genes being deleted. Even deletions that appear grossly the same in size may exhibit some variability at the sites of the non-allelic homologous recombination events. These small differences near the deletion break-points may influence the expression of genes critical for determining some of the phenotypic variability of the syndrome. Although there is no evidence to date for this mechanism, careful sequencing studies of the deletion region may shed further light on this possibility.

### VARIATIONS IN THE HEMIZYGOUS REGION OF 22q11.2

Individuals with 22q11.2DS have only a single remaining copy of the genes in the deletion region. Any susceptibility variants present in these genes no longer have the chance of being compensated for by a non-susceptibility variant on the homolog. A variant in *PIK4CA* has been associated with risk of developing schizophrenia in individuals with 22q11.2DS, but alone is not sufficient to explain the variable penetrance of schizophrenia in these individuals (Vorstman et al., 2009). Sequencing studies of the deletion region will also help clarify this possibility. Interestingly, sequencing of a major congenital cardiac disease gene from the deletion region, *TBX1*, did not reveal such variants for expression of tetralogy of Fallot in 22q11.2DS (Guo et al., 2011).

### VARIATIONS ELSEWHERE IN THE GENOME

A final potential source of phenotypic variability in 22q11.2DS is genetic variation elsewhere in the genome. Such variants may be considered genetic background effects. The importance of genetic background on variable expressivity and penetrance is well known from the mouse genetics literature, where the same mutation can manifest very differently in different strains (Linder, 2006). This background effect could be as simple as a single modifier locus that interacts with the deletion to determine penetrance of schizophrenia, or it could be similar to the complicated genetic risk proposed for non-22q11.2DS mediated schizophrenia, with the potential for multiple distributed liability loci acting in concert to cause illness. The profoundly increased risk of schizophrenia in 22q11.2DS does suggest that the non-22q11.2 genetic component for these cases of schizophrenia should be simpler than non-deletion mediated forms of the illness, making this an ideal subset for intensive study. Unfortunately, the relatively rarity of 22q11.2 deletions among individuals with schizophrenia (approximately 1% of cases of a disease that affects approximately 1% of the population) has hampered the assembly of the large cohorts necessary to conduct well-powered traditional genome-wide studies.

### A TWO-HIT MODEL

Canalization is a design principle wherein developmental pathways are stabilized to increase phenotypic reproducibility.

This helps the organism withstand genetic or environmental perturbations by providing a buffering system that confers robustness to developmental genetic programs. Hornstein and Shomron have proposed miRNAs as a mechanism for canalization (Hornstein and Shomron, 2006). In this model, miRNA regulation can compensate for variability in gene expression and promote constant levels of protein. Canalized traits have an increased capacity to absorb mutational variance. Specifically, there is a decreased penalty for mutations that alter gene regulation. Altered expression of transcription factors or sequence variation in *cis* binding sites have minimal phenotypic consequences as the miRNA regulatory system compensates for these changes, keeping gene expression at the required level. Because selective pressure is reduced, such mutations may silently accumulate, leading to a state where large numbers of cryptic variations may be distributed throughout the population. A sudden perturbation of the miRNA regulatory system, caused by haploinsufficiency of a key gene in miRNA biogenesis (*DGCR8*), could lead to multiple previously silenced mutations suddenly having an impact on the organism (Figure 2). The expected underlying population heterogeneity in these accumulated regulatory mutations could manifest as a high degree of variability in the specific phenotypic profile seen in individuals with *DGCR8* haploinsufficiency.

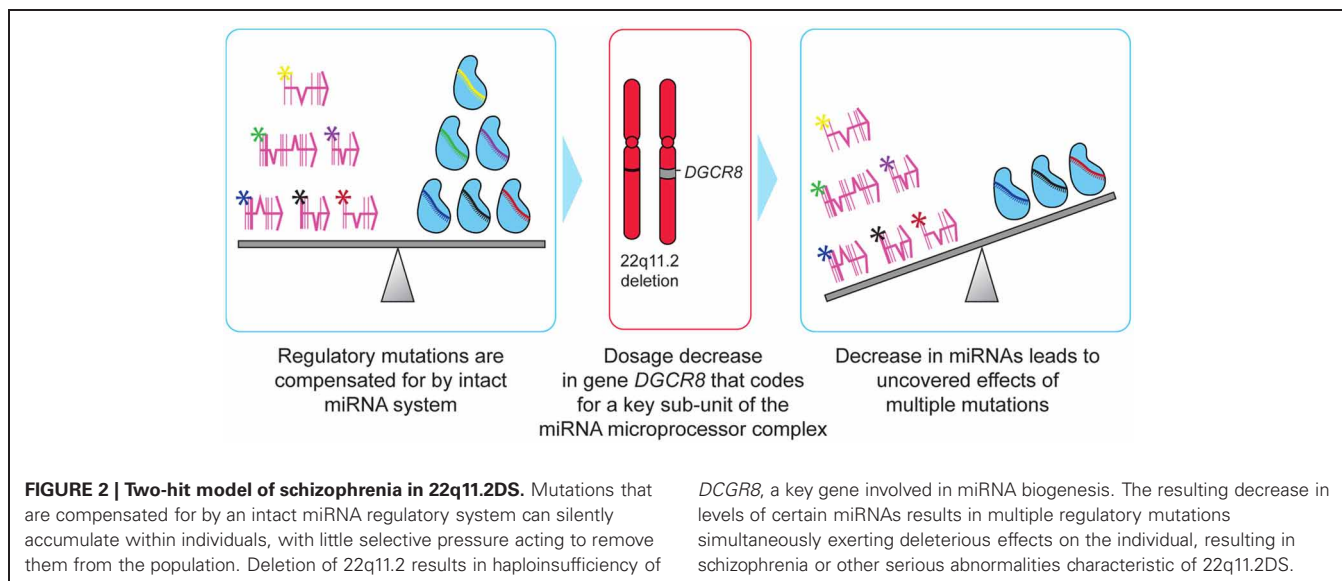
*DGCR8* is an essential component of the canonical miRNA biogenesis pathway. A non-canonical processing pathway also exists (Babiarz et al., 2008), with mouse knockout studies suggesting that the brain is particularly enriched for these non-canonical miRNAs as compared to other tissues (Babiarz et al., 2011). A mouse model of *DGCR8* haploinsufficiency has also demonstrated that not all miRNAs are equally affected by reduced *DGCR8* levels; despite reductions of *DGCR8* transcript abundance to approximately 60% of wild type levels, only 22% of mature miRNAs assayed exhibited significantly decreased levels in either prefrontal cortex or hippocampus in the haploinsufficient mice (Stark et al., 2008).

### A TESTABLE HYPOTHESIS

If loss of miRNA silencing of multiple regulatory mutations contributes to the risk of developing schizophrenia in 22q11.2DS, then one would expect these mutations to occur more frequently in genes regulated by the miRNAs most impacted by *DGCR8* haploinsufficiency. Even if tested in a relatively small (by genome-wide association study standards) sample, one would expect to see a pattern of enrichment of association findings to SNPs in or near genes regulated by the 22q11.2DS altered miRNAs. We sought to test this hypothesis using data from a set of 95 families with a 22q11.2DS proband.

### EXPERIMENT I—GENOME-WIDE ASSOCIATION STUDY OF SCHIZOPHRENIA IN 22q11.2DS

The first experiment in testing our hypothesis was to conduct a GWAS of schizophrenia in 22q11.2DS. To protect against the effects of population stratification, we chose a family based association design. Adults with 22q11.2DS, followed at the Clinical Genetics Research Program, Centre for Addiction and Mental Health (CAMH) at the University of Toronto, Canada, and their families, were studied. Participants provided written informed



consent, and the study was approved by the Research Ethics Boards of the University of Toronto, CAMH, and University Health Network. Three hundred and seventeen subjects were included in the study. Comprehensive medical, psychiatric, and family history data were available for all subjects. Experienced psychiatrists provided lifetime DSM-IV psychiatric diagnoses using standard methods, as previously described (Bassett et al., 2007). These individuals comprised 95 families, each with an adult proband confirmed to have a chromosome 22q11.2 deletion by standard methods using metaphase chromosomes from peripheral blood lymphocytes and FISH techniques using a probe, most commonly TUPLE1 (Vysis) or N25 (ONCOR), from the commonly deleted 22q11.2 region (Driscoll et al., 1993). Forty-eight of the 22q11DS probands (26 female and 22 male) were diagnosed with schizophrenia or schizoaffective disorder, while psychotic illness was excluded in the remaining 47 22q11DS probands (28 female and 19 male). Of the 222 additional family members, five (four parents and a half-sibling) also had 22q11.2DS but no psychotic illness; four (two parents and two siblings) were also diagnosed with schizophrenia but did not have 22q11.2DS.

All subjects with available high-quality DNA samples ( $n = 244$ ) were genotyped on Affymetrix 250K Nsp SNP arrays at The Centre for Applied Genomics at the University of Toronto. Genotypes for all CEL files were processed as a single batch using Affymetrix APT software, version 1.10.0. All samples used for further analysis had  $<5\%$  missing data ( $n = 238$ ). All persons within any pedigree with a detected Mendel error were set to zero for the marker with the error. SNPs were further excluded from all subjects for  $>2\%$  missingness or deviations from Hardy–Weinberg equilibrium that were significant at  $p < 0.0001$ . After cleaning, 221,996 SNPs remained for analysis.

Association analysis was conducted using the software package KELVIN 2.4.0, which implements the posterior probability of linkage (PPL) class of models for measuring the strength of genetic evidence (Vieland et al., 2011b). The PPL was initially

developed as a linkage analysis method (PPL originally stood for posterior probability of linkage, but now is used to refer to the class of related measures), based upon an acting single-locus likelihood allowing for within-sample heterogeneity. All parameters of the trait model are integrated out of the PPL using simple model averaging, allowing for testing of dominant, recessive, and additive models without the inflationary effects of multiple testing or parameter maximization. The analysis framework is quite flexible, and many additional extensions have been implemented such as analysis of linkage disequilibrium, quantitative traits, liability classes, imprinting, sex specific recombination rates, and epistasis (Bartlett and Vieland, 2007; Huang et al., 2007; Huang and Vieland, 2010; Vieland et al., 2011a). Importantly, the PPL accumulates evidence against linkage or association as well as in favor of it. The PPL is on the probability scale, making its interpretation straightforward. As a measure of statistical evidence, and not a decision-making procedure, there are no associated “significance levels” and it is not interpreted in terms of associated error probabilities nor corrected for multiple testing (Royall, 1997; Vieland and Hodge, 1998; Vieland, 2006).

We chose this analysis approach for two important reasons. First, our study design is complex and we required an analysis approach that could appropriately incorporate all available data into the analysis. While our study fundamentally has a family based association design, it also has elements of a case-control design, with half of the 22q11.2DS probands affected with schizophrenia and half not. We also needed to be able to appropriately model two different clinical elements of interest, 22q11.2 deletion status and schizophrenia status, both of which are present in some of the first degree relatives. Our hypothesis is that mutations segregating within these pedigrees have a greater impact on causing disease in the individuals with 22q11.2 deletions. This led us to model schizophrenia as the primary categorical phenotype of interest for association analysis and model 22q11.2 deletion status as a liability class. In this type of model, individuals with the deletion would be expected to have higher



penetrance of schizophrenia with a given risk allele at a susceptibility locus when compared to relatives without the deletion but carrying the same risk allele. KELVIN's flexibility allowed our entire sample to be analyzed simultaneously, incorporating 22q11.2 deletion status as a liability class while testing for association to schizophrenia.

The second important reason for using the PPL framework for our analysis was the desire to use a method that would disambiguate strength of evidence for or against association from low power due to the relatively small size of our sample. The PPL provides a measure of strength of evidence for or against association. Beginning with a given prior probability of association, KELVIN calculates the posterior probability of association based on the provided sample data. If the sample is providing evidence against association, the posterior probability will be lower than the prior probability. If the sample is giving evidence supporting association, the posterior probability will be higher than the prior. The magnitude of the change, up or down, can further be used to determine the strength of the evidence and rank order findings in terms of likelihood of association.

The Affymetrix array data were first used to confirm that all subjects with 22q11.2DS had deletions that included *DGCR8*. Next, the 221,996 SNPs passing our quality control were tested for association. Of these, 9191 were uninformative. Of the remaining SNPs, 198,100 (93%) gave evidence against association, while only 14,705 (7%) provided any evidence for association. Given the small size of the sample (95 small pedigrees), we did not expect to detect compelling evidence for association. Starting from a prior probability of 0.04%, the maximum posterior probability of association was 1.7%, a large increase from the prior but still within the range of what might be seen by chance in a full GWAS. Despite the small magnitude of evidence for association in the best supported SNPs and the expectation that some of these findings were due to chance allele segregations, we reasoned that these top SNPs should be enriched for real association signals. We tested this possibility by a further application of our hypothesis of miRNA mediated risk for schizophrenia in 22q11.2DS.

## EXPERIMENT II—TEST FOR miRNA TARGET ENRICHMENT IN 22q11.2DS SCHIZOPHRENIA

The second experiment sought to correlate our GWAS results with miRNA data from two experimental sources. The work of Stark et al. had demonstrated that the levels of some, but not all, miRNAs are significantly decreased in the brains of mice haploinsufficient for *DGCR8* (Stark et al., 2008). Since a list of miRNAs decreased in the brains of humans with 22q11.2DS is not currently available, we began with the list from the mouse model as the best available data, recognizing that noise would undoubtedly be introduced into our analysis by this cross-species approach. The 65 mouse miRNAs reported as downregulated in either prefrontal cortex or hippocampus (Stark et al., 2008) were mapped to their human orthologs and used to query starBase v2.1 (Yang et al., 2011) to determine a list of human genes implicated by two or more CLIP-seq experiments to interact with the miRNAs of interest (Hafner et al., 2010). Overall, 2860 genes were identified as potentially interacting with one or more of the miRNA of interest. Of these, 1047 were predicted to interact with

**Table 1 | Distribution of CLIP-seq interactions with miRNAs of interest.**

Number of interacting miRNAs	Number of genes
1	1047
2	689
3	395
4	235
5	169
6–10	281
>10	44
Total	2860

a single miRNA with the remaining genes predicted to interact with between 2 and 34 of the miRNAs investigated (**Table 1**).

Before we could compare this list of genes to our GWAS results, we needed to establish a method for identifying genes that might be regulated by an associated SNP. While *cis*-acting regulatory sequences can act on genes located over 1 Mb away, the sequences that act over the greatest distances tend to be found in gene-poor regions (reviewed in Kleinjan and Coutinho, 2009), suggesting that the physical extent of a candidate region surrounding an LD signal can be modulated by the gene density of the area. We therefore developed an algorithm to map SNPs to potential gene targets based on local gene density. The genomic locations of all RefSeq genes mapped to NCBI Build 36.1 of the human genome were downloaded from the UCSC genome browser (Lander et al., 2001; Kent et al., 2002; Fujita et al., 2011). We then used a four-tiered size approach for searching for potential gene targets. We began by identifying all genes within 25 Kb of a SNP of interest. If no genes were identified, the search window was expanded to 250 Kb on either side of the SNP. If no genes were identified at that step, the window was next expanded to 500 Kb and finally to 1 Mb on either side of the gene. This strategy allows for identification of distant genes from gene-poor regions while limiting the number of genes identified from gene-dense regions. It is important to note, however, that regardless of the interval used, genes are selected purely on the basis of distance from the SNP of interest and all genes within the given window will be identified as potentially regulated by this SNP. This is a very simplistic mapping procedure and even in cases where a SNP of interest actually does have a regulatory function, our algorithm is expected to identify the actual target of that regulatory SNP along with other nearby, but not necessarily regulated, genes. So only some subset of the genes on this list is expected to be of actual interest.

From the 14,705 SNPs producing some positive evidence of association in the GWAS, we arbitrarily selected the 0.5% (73 SNPs) giving the strongest evidence for further analysis. Using the algorithm described above, this set of SNPs identified a set of 72 genes as potential regulatory targets based on location. Of those 72 genes, 12 (17%) had been identified in starBase as potential targets of two or more of the miRNAs observed to be decreased in the *DGCR8* haploinsufficient mice. To determine if this represented enrichment over what would be expected by chance, we conducted a simulation study.

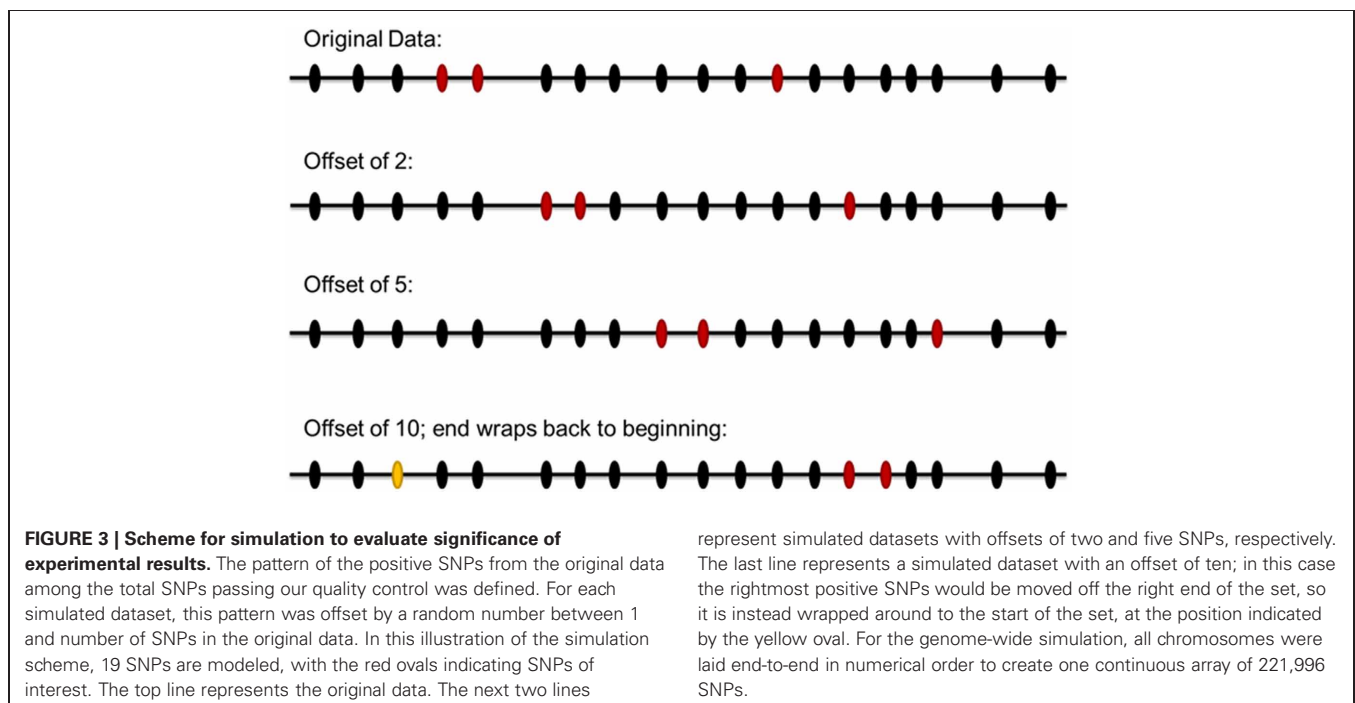
The distribution of the Affymetrix 250K SNPs across genes and the genome is neither perfectly uniform nor random. Similarly, the top association signals are not randomly distributed across the Affymetrix SNPs, with some grouping of positive signals among SNPs in strong linkage disequilibrium with each other. To try to best maintain any effects due to the actual location of the Affymetrix SNPs in the genome and the relative clustering of association signals within nearby SNPs, we devised the following simulation scheme. The pattern of the 73 positive SNPs from the analysis of our real data among the total 221,996 SNPs passing our quality control was defined. For each simulated dataset, this pattern was offset by a random number between 1 and 221,966 (Figure 3). This produced a simulated dataset with the same degree of clustering of SNPs as the real data, but with a different specific set of 73 Affymetrix 250K SNPs selected as targets. These SNPs would then be mapped to genes following the same algorithm that was used for the real data.

Ten thousand simulated datasets were generated by this procedure. The mean number of genes identified by our mapping algorithm in the simulated datasets was 73.0 (s.d. = 11.8), compared to 72 in our real data, suggesting that the simulation procedure successfully generated comparably sized sets of genes. However, among the simulated datasets, only an average of 9.3% of the genes were identified in starBase as potential targets of two or more of the miRNAs observed to be decreased in the *DGCR8* haploinsufficient mice, compared to the 17% observed in the real data set. Based on the distribution of results seen in our simulated data sets, the results observed in the real data set represent a substantial enrichment of genes predicted to be regulated by the miRNAs most disrupted by *DGCR8* haploinsufficiency, significant at  $p = 0.03$ . This is the result that was predicted by our hypothesis.

## LIMITATIONS OF THIS STUDY

The results of these preliminary experiments support our model that the increased risk of schizophrenia in individuals with 22q11.2DS is at least partially due to a loss of miRNA compensation for SNP variants that alter regulation of susceptibility genes. As predicted by our model, our experimental results produced statistically significant enrichment in genes likely to be regulated by the miRNAs decreased by *DCGR8* haploinsufficiency over what would be expected by chance. This result is particularly striking given the number of significant limitations of this study.

First, we used data from *DGCR8* haploinsufficient mice (Stark et al., 2008) as the source of the list of the most disrupted miRNAs. It is possible that the levels of the orthologous miRNAs will not all be similarly altered in humans with 22q11.2DS. Furthermore, the effects of 22q11.2DS on primate and human specific miRNAs will not be informed by the murine model. Therefore, our starting list of miRNAs of interest is incomplete at best but may also contain miRNAs that are not significantly decreased in humans with 22q11.2DS. Ideally, we would start with a list of altered miRNAs obtained through study of human brain tissue from subjects with 22q11.2DS. Given the relative rarity of this syndrome and the general challenges in obtaining human post-mortem brain tissue suitable for gene expression studies, a sizable collection of such tissues will likely be difficult to obtain. Perhaps more logistically feasible would be to use neural tissue derived from induced pluripotent stem cells (iPSCs) made from individuals with 22q11.2DS, although such a derived cell population might include experimentally induced genetic or epigenetic changes. Even profiling of blood samples from individuals with 22q11.2DS might provide some valuable insights into alterations of primate and human specific miRNAs. Our studies clearly need to



be repeated with a list of miRNAs experimentally derived from human source tissue.

Second, we chose to use CLIP-seq data, rather than bioinformatic prediction, to identify potential targets of the miRNAs of interest. CLIP-seq data have the advantage of identifying miRNA targets with better specificity than bioinformatic approaches (Chi et al., 2009; Zisoulis et al., 2010). For example, querying the miRWalk database (Dweep et al., 2011) for predicted targets of the miRNAs of interest using the default setting (target identification by miRanda, miRDB, miRWalk, RNA22, and TargetScan) identifies 10,037 genes predicted to be targets of one or more of the miRNAs of interest, 3.5 times as many as were identified by the available CLIP-seq data. This larger dataset appears to have a worse signal-to-noise ratio than the CLIP-seq data, as repeating our experiments using this very large set of potential target genes failed to produce significant results. More stringent bioinformatic prediction procedures might well perform better. However, the use of CLIP-seq data also has limitations. As an experimentally based approach, the discovered interactions will be limited by the gene expression profiles of the cells that are interrogated. The currently available human CLIP-seq data are still somewhat limited, but as more tissue types are investigated the comprehensiveness of the interactions catalogued will increase and be of even greater value.

Third, we used a sample for this study that is too small to be likely to be able to produce compelling evidence in support of association in the context of a GWAS. While the PPL analysis framework has been demonstrated to have a superior signal-to-noise ratio when compared to other common analysis methods (Huang and Vieland, 2001; Vieland et al., 2001; Logue et al., 2005), the low score threshold for the SNPs investigated in this study clearly overlaps the range of scores that could be seen by chance in a GWAS. Therefore, we expect that some of the SNPs of interest represent false positives. As a group, however, we expect the SNPs of interest to be enriched for true association signals when compared to comparable random sets of SNPs. While subjects with 22q11.2DS are relatively rare, the importance of this syndrome as a model for studying the neurodevelopmental trajectory of schizophrenia has been recognized (Insel, 2010), leading to a new international focus on recruiting such individuals for genetic studies. Larger cohorts will have great power to separate real signals from background noise.

Fourth, we implemented a very simple SNP-to-gene mapping algorithm that identified multiple genes surrounding each associated SNP based solely on physical position. While our procedure used variable size intervals to limit the number of genes identified from the more gene-dense regions of the genome, our algorithm will typically return sets of genes for each SNP of interest, with only a fraction of the identified genes expected to be the target of a SNP with regulatory function. Additional layers of bioinformatic analysis could be added to attempt to further predict the potential regulatory function of the SNPs of interest (or SNPs in strong linkage disequilibrium with them). Functional validation studies of putative regulatory SNPs could also be of utility. Another resource for

potentially identifying the regulatory targets of SNPs of interest would be gene expression profiles from individuals with 22q11.2DS. Our hypothesis is that deleterious regulatory mutations are uncovered by damage to the miRNA regulatory system. Therefore, we would expect to see increased expression of the gene targets of the decreased miRNAs. Expression data could be used to further refine the list of putative targets of SNPs of interest.

Fifth, we conducted these experiments using relatively low-density SNP data from Affymetrix 250K NSP arrays. It is quite likely that important signals were missed due to the low SNP density. While this initial study has produced intriguing results, additional studies using a higher density SNP array will be necessary to more comprehensively search the genome for association signals.

### ADDITIONAL CONSIDERATIONS

While we have focused this study on the presence or absence of schizophrenia in individuals with 22q11.2DS, this model could also explain the variability in multiple symptoms of this syndrome. Under the canalization model of Hornstein and Shomron (2006), multiple regulatory mutations can accumulate in the population, free from selective pressure, in the presence of an intact miRNA regulatory system. One would expect the specific distribution of compensated mutations to differ from individual to individual, due to chance segregation of the various deleterious alleles in the population. The striking phenotypic variability of individuals with 22q11.2DS could be due, in part, to the underlying heterogeneity of regulatory mutations that are suddenly uncovered by *DGCR8* haploinsufficiency. One could repeat our experiments with additional phenotypes, such as cardiac malformations, to further test this hypothesis.

Identification and validation of the specific gene targets leading to schizophrenia in the setting of altered miRNAs in 22q11.2DS could provide valuable new insights into this disorder. These genes could represent novel targets for pharmacologic intervention, and it is possible that targeting one or a small number of them could be therapeutically effective. Alternatively, it might be possible to pharmacologically manipulate the miRNA regulatory system more globally. For example, the pri-miRNA processing activity of *DGCR8* has been linked to levels of ferric heme, leading to the suggestion that manipulation of the intracellular environment with heme derivatives might have utility in increasing *DGCR8* activity (Barr et al., 2012). Such a therapy would directly address the molecular lesion caused by *DGCR8* haploinsufficiency.

While 22q11.2DS contributes to a minority of all cases of schizophrenia, there may well be mechanistic overlap with non-22q11.2DS cases of illness. Post-mortem analysis of human brain tissue has indicated a significant overlap in the miRNAs dysregulated in non-22q11.2DS individuals with psychotic illness and those depleted in a mouse model of 22q11.2DS (Stark et al., 2008; Moreau et al., 2011), suggesting the potential for shared mechanisms between etiologically different forms of schizophrenia. Further understanding of 22q11.2DS may well inform our global understanding of the causes, and potential treatment, of all schizophrenia.

## ACKNOWLEDGMENTS

This work was supported by grant R01 MH80429 from the National Institutes of Mental Health (Linda M. Brzustowicz), and Canadian Institutes of Health Research grants MOP-97800 and MOP-89066 and a Canada Research Chair in

Schizophrenia Genetics and Genomic Disorders (Anne S. Bassett). Thanks to Heidi Chen and Anthony Marcketta for assistance with mouse to human miRNA comparisons and to Sean Bekeschus for assistance with figure preparation.

## REFERENCES

- Babiarz, J. E., Hsu, R., Melton, C., Thomas, M., Ullian, E. M., and Belloch, R. (2011). A role for noncanonical microRNAs in the mammalian brain revealed by phenotypic differences in Dgcr8 versus Dicer1 knockouts and small RNA sequencing. *RNA* 17, 1489–1501.
- Babiarz, J. E., Ruby, J. G., Wang, Y., Bartel, D. P., and Belloch, R. (2008). Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. *Genes Dev.* 22, 2773–2785.
- Barr, I., Smith, A. T., Chen, Y., Senturia, R., Burstyn, J. N., and Guo, F. (2012). Ferric, not ferrous, heme activates RNA-binding protein DGCR8 for primary microRNA processing. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1919–1924.
- Bartlett, C. W., and Vieland, V. J. (2007). Accumulating quantitative trait linkage evidence across multiple datasets using the posterior probability of linkage. *Genet. Epidemiol.* 31, 91–102.
- Bassett, A. S., Caluseriu, O., Weksberg, R., Young, D. A., and Chow, E. W. (2007). Catechol-O-methyl transferase and expression of schizophrenia in 73 adults with 22q11 deletion syndrome. *Biol. Psychiatry* 61, 1135–1140.
- Bassett, A. S., and Chow, E. W. (2008). Schizophrenia and 22q11.2 deletion syndrome. *Curr. Psychiatry Rep.* 10, 148–157.
- Bassett, A. S., McDonald-McGinn, D. M., Devriendt, K., Digilio, M. C., Goldenberg, P., Habel, A., et al. (2011). Practical guidelines for managing patients with 22q11.2 deletion syndrome. *J. Pediatr.* 159, 332–339 e331.
- Beveridge, N. J., and Cairns, M. J. (2012). MicroRNA dysregulation in schizophrenia. *Neurobiol. Dis.* 46, 263–271.
- Biggar, R. J., Cassol, S., Kumwenda, N., Lema, V., Janes, M., Pilon, R., et al. (2003). The risk of human immunodeficiency virus-1 infection in twin pairs born to infected mothers in Africa. *J. Infect. Dis.* 188, 850–855.
- Chi, S. W., Zang, J. B., Mele, A., and Darnell, R. B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* 460, 479–486.
- Driscoll, D. A., Salvin, J., Sellinger, B., Budarf, M. L., McDonald-McGinn, D. M., Zackai, E. H., et al. (1993). Prevalence of 22q11 microdeletions in DiGeorge and velocardiofacial syndromes: implications for genetic counselling and prenatal diagnosis. *J. Med. Genet.* 30, 813–817.
- Dweep, H., Sticht, C., Pandey, P., and Gretz, N. (2011). miRWalk-database: prediction of possible miRNA binding sites by “walking” the genes of three genomes. *J. Biomed. Inform.* 44, 839–847.
- Edelmann, L., Pandita, R. K., and Morrow, B. E. (1999). Low-copy repeats mediate the common 3-Mb deletion in patients with velocardio-facial syndrome. *Am. J. Hum. Genet.* 64, 1076–1086.
- Fujita, P. A., Rhead, B., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Cline, M. S., et al. (2011). The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.* 39, D876–D882.
- Fung, W. L., McEvilly, R., Fong, J., Silversides, C., Chow, E., and Bassett, A. (2010). Elevated prevalence of generalized anxiety disorder in adults with 22q11.2 deletion syndrome. *Am. J. Psychiatry* 167, 998.
- Goodship, J., Cross, I., Liling, J., and Wren, C. (1998). A population study of chromosome 22q11 deletions in infancy. *Arch. Dis. Child.* 79, 348–351.
- Guo, T., McDonald-McGinn, D., Blonska, A., Shanske, A., Bassett, A. S., Chow, E., et al. (2011). Genotype and cardiovascular phenotype correlations with TBX1 in 1, 022 velo-cardiofacial/DiGeorge/22q11.2 deletion syndrome patients. *Hum. Mutat.* 32, 1278–1289.
- Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129–141.
- Hornstein, E., and Shomron, N. (2006). Canalization of development by microRNAs. *Nat. Genet.* 38(Suppl.), S20–S24.
- Huang, J., and Vieland, V. J. (2001). Comparison of ‘model-free’ and ‘model-based’ linkage statistics in the presence of locus heterogeneity: single data set and multiple data set applications. *Hum. Hered.* 51, 217–225.
- Huang, Y., Bartlett, C. W., Segre, A. M., O’Connell, J. R., Mangin, L., and Vieland, V. J. (2007). Exploiting gene x gene interaction in linkage analysis. *BMC Proc.* 1(Suppl. 1):S64.
- Huang, Y., and Vieland, V. J. (2010). Association statistics under the PPL framework. *Genet. Epidemiol.* 34, 835–845.
- Insel, T. R. (2010). Rethinking schizophrenia. *Nature* 468, 187–193.
- Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., et al. (2002). The human genome browser at UCSC. *Genome Res.* 12, 996–1006.
- Kim, A. H., Parker, E. K., Williamson, V., McMichael, G. O., Fanous, A. H., and Vladimirov, V. I. (2012). Experimental validation of candidate schizophrenia gene ZNF804A as target for hsa-miR-137. *Schizophr. Res.* 141, 60–64.
- Kleinjan, D.-J., and Coutinho, P. (2009). Cis-rupture mechanisms: disruption of cis-regulatory control as a cause of human genetic disease. *Brief. Funct. Genomic. Proteomic.* 8, 317–332.
- Kumar, D. (2008). Disorders of the genome architecture: a review. *Genomic Med.* 2, 69–76.
- Kwon, E., Wang, W., and Tsai, L. H. (2011). Validation of schizophrenia-associated genes CSMD1, C10orf26, CACNA1C and TCF4 as miR-137 targets. *Mol. Psychiatry*. doi: 10.1038/mp.2011.170. [Epub ahead of print].
- Lander, E. S., Linton, L. M., Birren, B., Nussbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.
- Linder, C. C. (2006). Genetic variables that influence phenotype. *ILAR J.* 47, 132–140.
- Logue, M. W., George, A. W., Spence, M. A., and Vieland, V. J. (2005). Performance comparison of two-point linkage methods using microsatellite markers flanking known disease locations. *BMC Genet.* 6(Suppl. 1):S141. doi: 10.1186/1471-2156-6-S1-S141
- Meechan, D. W., Maynard, T. M., Tucker, E. S., and Lamantia, A. S. (2011). Three phases of DiGeorge/22q11 deletion syndrome pathogenesis during brain development: patterning, proliferation, and mitochondrial functions of 22q11 genes. *Int. J. Dev. Neurosci.* 29, 283–294.
- Mellios, N., and Sur, M. (2012). The emerging role of microRNAs in schizophrenia and autism spectrum disorders. *Front Psychiatry* 3:39. doi: 10.3389/fpsy.2012.00039
- Moreau, M. P., Bruse, S. E., David-Rus, R., Buyske, S., and Brzustowicz, L. M. (2011). Altered microRNA expression profiles in postmortem brain samples from individuals with schizophrenia and bipolar disorder. *Biol. Psychiatry* 69, 188–193.
- Ripke, S., Sanders, A. R., Kendler, K. S., Levinson, D. F., Sklar, P., Holmans, P. A., et al. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nat. Genet.* 43, 969–976.
- Royall, R. (1997). *Statistical Evidence: A Likelihood Paradigm*. London: Chapman and Hall.
- Saito, Y., and Saito, H. (2012). MicroRNAs in cancers and neurodegenerative disorders. *Front. Gene.* 3:194. doi: 10.3389/fgene.2012.00194
- Shaikh, T. H., Kurahashi, H., Saitta, S. C., O’Hare, A. M., Hu, P., Roe, B. A., et al. (2000). Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis. *Hum. Mol. Genet.* 9, 489–501.
- Stark, K. L., Xu, B., Bagchi, A., Lai, W.-S., Liu, H., Hsu, R., et al. (2008). Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. *Nat. Genet.* 40, 751–760.
- Vieland, V. J. (2006). Thermometers: something for statistical geneticists to think about. *Hum. Hered.* 61, 144–156.
- Vieland, V. J., Hallmayer, J., Huang, Y., Pagnamenta, A. T., Pinto, D., Khan, H., et al. (2011a). Novel method for combined linkage and genome-wide association analysis



- finds evidence of distinct genetic architecture for two subtypes of autism. *J. Neurodev. Disord.* 3, 113–123.
- Vieland, V. J., and Hodge, S. E. (1998). Review of statistical evidence: a likelihood paradigm. *Am. J. Hum. Genet.* 63, 283–289.
- Vieland, V. J., Huang, Y., Seok, S. C., Burian, J., Catalyurek, U., O'Connell, J., et al. (2011b). KELVIN: a software package for rigorous measurement of statistical evidence in human genetics. *Hum. Hered.* 72, 276–288.
- Vieland, V. J., Wang, K., and Huang, J. (2001). Power to detect linkage based on multiple sets of data in the presence of locus heterogeneity: comparative evaluation of model-based linkage methods for affected sib pair data. *Hum. Hered.* 51, 199–208.
- Vorstman, J. A., Chow, E. W., Ophoff, R. A., Van Engeland, H., Beemer, F. A., Kahn, R. S., et al. (2009). Association of the PIK4CA schizophrenia-susceptibility gene in adults with the 22q11.2 deletion syndrome. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 150B, 430–433.
- Xu, B., Roos, J. L., Levy, S., Van Rensburg, E. J., Gogos, J. A., and Karayiorgou, M. (2008). Strong association of *de novo* copy number mutations with sporadic schizophrenia. *Nat. Genet.* 40, 880–885.
- Yang, J. H., Li, J. H., Shao, P., Zhou, H., Chen, Y. Q., and Qu, L. H. (2011). starBase: a database for exploring microRNA-mRNA interaction maps from Argonaute CLIP-Seq and Degradome-Seq data. *Nucleic Acids Res.* 39, D202–D209.
- Zhang, F., Chen, Y., Liu, C., Lu, T., Yan, H., Ruan, Y., et al. (2012). Systematic association analysis of microRNA machinery genes with schizophrenia informs further study. *Neurosci. Lett.* 520, 47–50.
- Zisoulis, D. G., Lovci, M. T., Wilbert, M. L., Hutt, K. R., Liang, T. Y., Pasquinelli, A. E., et al. (2010). Comprehensive discovery of endogenous Argonaute binding sites in *Caenorhabditis elegans*. *Nat. Struct. Mol. Biol.* 17, 173–179.
- that could be construed as a potential conflict of interest.

Received: 13 September 2012; paper pending published: 10 October 2012; accepted: 26 November 2012; published online: 13 December 2012.

Citation: Brzustowicz LM and Bassett AS (2012) miRNA-mediated risk for schizophrenia in 22q11.2 deletion syndrome. *Front. Gene.* 3:291. doi: 10.3389/fgene.2012.00291

This article was submitted to *Frontiers in Non-Coding RNA*, a specialty of *Frontiers in Genetics*.

Copyright © 2012 Brzustowicz and Bassett. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.

**Conflict of Interest Statement:** Dr. Brzustowicz serves as a consultant for the Janssen Pharmaceutical Companies of Johnson & Johnson. Dr. Bassett declares that the research was conducted in the absence of any commercial or financial relationships