An integrative functional genomics approach for discovering biomarkers in schizophrenia

Marquis P. Vawter, Firoza Mamdani and Fabio Macciardi
Advance Access publication date 7 December 2011

Abstract
Schizophrenia (SZ) is a complex disorder resulting from both genetic and environmental causes with a lifetime prevalence world-wide of 1%; however, there are no specific, sensitive and validated biomarkers for SZ. A general unifying hypothesis has been put forward that disease-associated single nucleotide polymorphisms (SNPs) from genome-wide association study (GWAS) are more likely to be associated with gene expression quantitative trait loci (eQTL). We will describe this hypothesis and review primary methodology with refinements for testing this paradigmatic approach in SZ. We will describe biomarker studies of SZ and testing enrichment of SNPs that are associated both with eQTLs and existing GWAS of SZ. SZ-associated SNPs that overlap with eQTLs can be placed into gene–gene expression, protein–protein and protein–DNA interaction networks. Further, those networks can be tested by reducing/silencing the gene expression levels of critical nodes. We present pilot data to support these methods of investigation such as the use of eQTLs to annotate GWASs of SZ, which could be applied to the field of biomarker discovery. Those networks that have association with SNP markers, especially cis-regulated expression, might lead to a more clear understanding of important candidate genes that predispose to disease and alter expression. This method has general application to many complex disorders.

Keywords: expression quantitative trait loci; cis-regulatory SNPs; GWAS; gene expression; lymphoblastoid cell lines

INTRODUCTION
Schizophrenia (SZ) is a mental disorder which contributes to a substantial chronic medical burden, and a reduced lifespan that is ~15 years shorter than the general population [1]. While SZ is believed to have a heritability of 80% based upon family studies [2], only a handful of genes in increasingly large sample sizes have been statistically associated despite the numerous results of SZ genome-wide association studies (GWASs). Overall, SZ etiology is composed of both genetic and environmental influences [2]. Current estimates of polygenic variation implicate hundreds (or thousands of genes), each with low risk effects, and moderately penetrant rarer copy number and structural variants. Based on studies to date, a single gene mutation causing schizophrenia is highly unlikely. The polygenic SZ-associated variation assessed through GWAS statistically accounts...
for 33–40% of genetic variation in SZ [3], but this finding is solely based on statistical evidence. Although this same statistical method accounts for a smaller proportion of polygenic variance component in a Japanese cohort [4] it does appear to capture some common risk elements between studies of SZ. The high percentage of associated single nucleotide polymorphisms (SNPs) are likely due to a slight excess of SNPs having statistically significant association to SZ beyond that expected by chance and not based on confirmatory validation through alternate genetic methodologies. Thus, it is expected that multiple transcripts might be involved in this overall polygenic risk model.

There is an unmet need for early diagnostic markers [5]. According to the World Federation of Societies of Biological Psychiatry Task Force on Biological Markers in schizophrenia ‘At present, not a single biological trait in schizophrenia is available which achieves sufficient specificity, selectivity and is based on causal pathology and predictive validity to be recommended as diagnostic marker’ [6]. Our group and others [5–21] have been involved in biomarker studies of SZ and other psychiatric disorders [22–27], and have shown that selected blood-derived transcripts have high correlation with brain ($r = 0.98$) and can make suitable biomarkers [28] and ultimately have very strong association with genetic variants [29].

In this review article, we report currently available information on the status of biomarkers in SZ and the utility of expression quantitative trait loci (eQTL) and their connection with disease-associated SNPs to predict disease susceptibility. This review will be presented in two main sections, gene expression studies, focusing on peripheral blood and possible overlap with post-mortem brain, and secondly expanding on eQTLs and use for annotation of GWAS-associated SNPs. Thus, we present two approaches for the identification of robust and plausible biomarkers for SZ susceptibility.

POST-MORTEM GENE EXPRESSION IN SCHIZOPHRENIA

Although, results stemming from post-mortem brain gene expression studies cannot become easily employable biomarkers, they can still provide some direction for peripheral studies and are, thus, briefly mentioned.

Many gene expression studies have been carried out using post-mortem brain tissue and analyses comparing individuals with schizophrenia and controls. A recent review [30] of this literature summarized 34 studies of post-mortem brain expression differences. Several studies of post-mortem brain implicated immune-related response pathways, those studies present a concerted upregulation of genes within immune and stress-related pathways [31–34]. Those pathways in brain might be relevant for potential peripheral biomarkers.

The likelihood of a finding a reliable and highly predictive gene expression biomarker can partly be rooted in its consistency and positive phenotypic association in several populations and concordance between several methodologies and/or technologies. Besides immune-related genes, there is evidence supporting abnormal GABAergic and glutamatergic neurotransmission as possible features of schizophrenia (reviewed in [30]). One major finding is the alteration in brain expression of glutamate decarboxylase 1 (GAD1), which is the key enzyme in the formation of gamma-aminobutyric acid (GABA) from glutamate. Interestingly, the direction of alterations in GAD1 gene expression is highly influenced by age in schizophrenia cohorts. The expression of GAD1 is high in brain and levels are generally decreased in post-mortem brain of SZ compared with controls [35–40]. The levels of GAD1 are ~16–32-fold higher compared to lymphoblastoid cell lines (LCLs), and there is a nominal decrease in GAD1 expression in LCLs from patients with SZ ($n = 42$) compared to controls ($n = 32$) by 7% ($P = 0.054$). Thus, the use of GAD1 as a biomarker alone would not have an effect size that parallels brain, likely reducing its use as a biomarker.

PERIPHERAL GENE EXPRESSION IN SCHIZOPHRENIA

Currently, gene expression studies have also moved to more accessible tissues such as peripheral cells and blood due to the relative ease of acquisition and the potential utility of such samples including the possibility to collect larger sample sizes with a minimally invasive procedure. The peripheral blood transcriptome shares >80% homology with genes expressed in the brain [28], heart, liver, spleen, colon, kidney, prostate and stomach [41]. The intensity of peripheral blood transcriptome gene expression for a number of biological processes is comparable to
that of the prefrontal cortex [42]. Recently, Rollins et al. [28] demonstrated considerable overlap between gene expression in brain and peripheral blood, from the same individual, using two independent populations and different high-throughput array platforms.

The use of blood cells to perform microarray studies has other advantages. The use of blood, from living subjects, to study gene expression avoids the influence of confounding variables associated with post-mortem brain studies, such as the post-mortem interval, low pH and other factors that decrease the integrity of mRNA and which must be accounted for in subsequent analyses [43]. A recent gene expression study among psychiatric patients demonstrated the possibility of discriminating between schizophrenia and bipolar disorder using a blood-based protocol [16]. This same group has also confirmed findings implicating the selenium-binding protein 1 gene in schizophrenia using both brain and blood samples [8].

Peripheral studies in SZ have been increasing in number as recently reviewed [30]. Highlights of blood-derived biomarker studies are briefly presented. Zhang and colleagues [44] explored the plausibility of neuregulin-1 (NRG-1) gene expression as a state and trait dependent biomarker. They showed that patients with SZ had lower levels of NRG-1 in lymphocytes in comparison to both sibling and non-sibling controls. Interestingly, a longitudinal investigation of NRG-1 levels following antipsychotic treatment induction demonstrates gradual increases in gene expression, which could be used as a treatment biomarker following validation in a larger sample (SZ, treated N = 31). Kuzman et al. [11] performed a similar study, looking at array-based gene expression in whole blood of first-episode SZ compared to controls. They identified 180 probe sets having significantly altered gene expression in SZ and validated four genes.

Yao and colleagues [45] performed a study testing the applicability of utilizing peripheral gene expression and verifying if differences found in previous peripheral gene expression studies can translate to new samples. The authors were unable to replicate findings for all the genes tested. However, this should not discount the plausible utility of these genes as biomarkers since the sample size tested in was small. Bousman et al. [46] performed a study in a sample of 19 SZ and bipolar disorder (BD) patients with psychosis, to assess correlations between previously found SZ-associated ubiquitin gene peripheral expression with positive and negative symptoms. This study presents encouraging evidence for the utility of peripheral gene expression to assess specific aspects and more robust phenotypes within each disorder.

Takahashi and collaborators [47] took a slightly different approach although still starting from array-based peripheral gene expression data. Their bioinformatics-based approach identified a 14 probe set biomarker with high diagnostic accuracy using artificial neural networks analysis. However, when employing an unsupervised hierarchical clustering—with the 14 probe sets diagnostic separation was not attained.

Recently, there have been investigations into the use of peripheral blood microRNA expression as possible biomarkers for SZ. Lai and colleagues [48] used a TaqMan array composed of 365 human microRNAs, in a learning set of 30 cases and controls. Their significant findings, when comparing cases and controls, were subsequently tested in an independent case-control sample of 60 SZ and 30 control subjects. The authors were able to identify a disease signature, producing adequate sensitivity and specificity, with seven microRNAs. Beveridge et al. [49], previously found evidence for microRNA mediated regulation of expression in post-mortem SZ brain. As an independent follow-up in peripheral blood, Gardiner and colleagues [50] investigated microRNA expression in 112 SZ and 76 controls, and were able to identify 33 microRNAs passing false discovery rate correction. These studies demonstrate the utility of microRNAs in discriminating disease state.

These reports however show little overlap between the results of each study, most likely due to disease heterogeneity, medications, different preparations of blood cells, array platform differences small samples and other confounders [8, 16, 20, 22, 23]. Nevertheless the studies are encouraging in that they provide positive evidence that some peripheral markers are comparable to those from post-mortem studies, thus lending support to the use of peripheral blood samples as an advantageous alternative in the quest for the biological markers of brain-based disorders. However, there is room for additional improved methodologies, utilizing various levels of genomic information to identify biomarkers, one of which is investigating the genetics of gene expression—specifically eQTLs and associated SNPs.

Suggestions that we offer for selection of a predictive biomarker: (i) The gene should have
differential expression in SZ in LCL (or blood or peripheral blood mononuclear cell (PBMC)) in at least one or more prior studies, (ii) the gene ideally should be abundant in cell lines and whole blood, (iii) the gene should show differential expression within the SZ sample in brain [34] and (iv) the gene should not represent treatment effects (unless this is an aim of the study being undertaken), this criterion can be achieved through the collection of peripheral blood from first-episode SZ prior to the induction of treatment. As a working example of these selection criteria, the multidimensional scatter plot, in Figure 1, shows that the resulting gene list can be refined to using only four probe sets from the above criteria for complete classification between SZ and controls. These probe sets located in four different genes reliably predict subjects with SZ from controls in LCLs in our particular data sets, and predictions are based upon independent criteria applied to independent data sets.

**Figure 1:** There were four exons from different genes (BAT2L, DEK, GSR and DBC1) that significantly discriminated first-episode individuals with SZ (blue) and controls (red). The individual exons with bootstrap P-values <10^{-8} were plotted in multidimensional scaling, complete separation of SZ and C was observed, except for one subject with SZ which could not be classified.

EQTLS AND THE GENETICS OF GENE EXPRESSION

Current evidence suggests there is an important role of genetics on gene expression in SZ and possible relationship with phenotypic characteristics of SZ [51, 52]. The influence of genetics on gene expression can take a variety of forms, through an effect on gene regulation, thought to be particularly strong when the genetic variants are in *cis*-locations. For example, SNPs or insertion/deletions within transcription start sites, promoter and repressor elements, micro-RNA binding sites, transcription factor binding sites, protein splicing binding sites, and CpG islands, to name a few, can potentially regulate gene expression, change DNA structure, and alter expression. Therefore, in cases where there is high correlation between a *cis*-SNP and gene expression, the SNP could be used as a proxy for alterations in transcript levels. SNPs and other types of polymorphisms within regulatory elements are useful for annotation and serve to ‘functionate’ genomic variation, taken together these examples can be thought of broadly as the effect of genetics on gene expression. This broad concept of gene expression as eQTL and altered expression in schizophrenia as a biomarker is shown in Figure 2.

Those biomarker genes identified by SNPs associated with eQTLs or with altered gene expression can be related into networks. One basic question is whether statistically associated SNPs within a complex disease such as SZ might also have a functional effect on expression. This approach can be partially exemplified by the genetic findings of ZNF804A [53–56], which is thought to have a ‘causative’ SNP in the intron of the gene, with an odds ratio of 1.1. While the ‘causative’ SNP in ZNF804A has been associated with both SZ and cognitive sparing in SZ, some research shows that expression of the ZNF804A gene [53, 57] is associated with the SZ risk allele in ZNF804A.

The effect of genetics on gene expression model can be expanded to predict that eQTL-associated SZ SNPs will also have a stronger association to disease in comparison to non-eQTL SNPs or gene expression alone. This increased disease association has been postulated by other groups [29, 58, 59] and has been recently shown for schizophrenia [4, 60]. However, we recognize some limitations of this approach, e.g. not all eQTLs will be associated with SZ, and of course, can be tissue-, developmental- and population specific. However, if further validated in SZ, this model will have high impact upon further SZ biomarker and drug discovery.

Convincing data demonstrating the robustness and increased power of eQTL-associated SNPs to
predict augmented SZ risk has recently been published [4, 52, 58]. Based upon these encouraging results, we and others [23, 27, 61] posit that eQTL SNPs will in many cases be the same SNPs associated to SZ in GWAS (or in strong linkage disequilibrium), implying that eQTLs, as well, will be strongly related to SZ. This approach has strong empirical support [62, 63] for use in a complex disorder like SZ with both genetic and environmental influences.

As an example, using LCL transcript expression, and exon array, we have found overlap in SZ versus control comparisons for biomarkers. Our early work in both brain and LCL for aspartylglucosaminidase (AGA) [64] has been replicated in a second cohort in our laboratory (Figure 3) and further independently confirmed by others for both our AGA cis-eQTL finding [65] and linkage to SZ [66]. Once eQTL SNPs with clear association signals, such as AGA, have been identified, it is possible to organize these SNPs into putatively functional gene regulatory networks (GRNs). This novel network composed of disease-associated biomarkers and strong eQTLs can be tested using robust centrality parameters, and compared between SZ and controls experimentally. This approach has multiple impacts, assigns functional meaning to SNPs via potential regulatory functions, can provide a network of multiple targets for drug screening, and ultimately therapeutics could be developed based on high-throughput screening of pharmacological treatment effects on network genes.
eQTLs in schizophrenia

Using exon array data, several groups including ours, have shown that eQTLs are heritable [29, 67] and can influence brain expression [25]. Expression of IRF5 has been shown to be one of the most heritable eQTL (10^{-52}), after removing any probes that contained SNPs, from the IRF5 eQTL analysis [29]. In our published work ([25] reproduced in Figure 4) IRF5 expression with the exon array platform was validated for an eQTL in different brain regions, although we and others had originally shown this to be an eQTL in LCL. Furthermore, IRF5 expression of probe set 3023264 was increased in LCLs of SZ subjects matched to unaffected family members by exon array (P-value 1.41 x 10^{-8}) [25]. The same directional effect of genotype × probe set was observed in three independent exon array studies [5, 34, 57]. The SNP–eQTL association of rs10954213, is due to a functional SNP, where the ‘A’ allele creates a functional polyadenylation site shortening its 3’-UTR [68].

It is necessary to test whether there is an enrichment of disease and expression associated SNP overlap above that expected by chance [29, 58]. The null hypothesis (H0) is that the expected overlap between the eQTL-associated SNPs and disease-associated SNPs would be no more than expected for two randomly generated lists of HapMap SNPs of the same size. This approach has provided evidence that eQTLS found in LCLs derived from bipolar disorder, have strong association with disease compared to SNPs not found in eQTLS. This approach can be especially promising when using SNPs associated with exon expression [29, 58].

We next show an example of network construction using SZ GWAS SNPs and eQTLS. The NHGRI (www.genome.gov/gwastudies/), other GWASs from ISC [3] and GAIN, and other sources for SZ candidate SNPs (e.g. [27, 61, 69, 70]) can be used. Those disease-associated SNPs can be merged using the criterion of P < 10^{-6} as evidence of association. Evidence of those same SNPs associated with eQTLS can be found using SCAN [58] (http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/) using different thresholds. As a working example, to test the Ho that SZ-associated SNPs are associated with eQTLS we obtained the most replicated GWAS SNPs in studies of schizophrenia from genome.nih.gov, as of September 2010, annotated the 28 significant SNPs (minimum P < 10^{-6}), and then used this list to associate with any eQTLS (minimum P < 10^{-4}) derived from LCLs using SCAN [58]. This procedure gave a list of 62 genes within significantly associated eQTLS and GWAS SNPs for SZ, which became the input for a network analysis using IPA. Interestingly, the central network node was HNF4A (Figure 5). Increased interest for the HNF4A gene comes from our prior study of SZ brain gene expression where we found AGXT2L1, which is directly regulated by HNF4A [71], to be our strongest candidate gene for SZ [34]. Further we reported that HNF4A was the central node in a dysregulated network of LCL gene expression in SZ using matched family members that were affected for SZ and unaffected [25].

Thus, we would predict that alteration of HNF4A, a central node and transcription factor in a pathway altered in SZ, would lead to perturbation of many other key SZ genes such as AGXT2L1, and other network members shown in Figure 5.
Taking this example further, we and others found that AGXT2L1 expression was altered in SZ, BD and suicide dorsolateral prefrontal cortex (DLPFC) [32, 34, 72, 73]. As well, individuals with alterations in AGXT2L1 expression appear to be at a higher risk of developing a psychiatric disorder, based on a highly significant odds ratio [34]. Thus, we predict that through silencing of HNF4A in cell lines it would be plausible to determine whether AGXT2L1 and other key connections, such as serine racemase (SRR) a gene regulated by HNF4A [74], and having known association to SZ [75], are altered in this network.

**METHODOLOGICAL CONSIDERATIONS FOR eQTL ANALYSES**
Subject selection for eQTL analyses should be based upon homogeneous sets of matched sex and age matched cases and controls. These subjects, for example if using array-based genomic measures, can be further screened by principal components analysis (PCA) over all probe sets for both SZ and control subjects, to identify outliers and minimize population stratification in the eQTL analysis [29, 44, 58, 76, 77]. It is worthwhile to also use either PBMC or LCLs that are obtained under similar conditions, with total RNA being extracted in the same batch.

**Expression profiling and SNP array genotyping**
Exon arrays have been widely used for the study of genetic variation in coding regions [25, 29, 76, 78–85], correlate positively with RNA-Seq across most levels of transcript expression [86], and in some cases have less false positive detection than RNA-Seq [87–89]. However, with the continuous advancements being made in RNA-Seq analyses and reductions in cost, it will be an adequate alternative.
to hybridization strategies. We and others have shown the effects of potential SNPs on probe hybridization artifacts that can give erroneous eQTL information [76, 79, 84]. Thus, RNA-Seq will have more power to detect eQTLs than exon array as well as provide allele specific expression, when read coverage is at least 20 × [90]. However, for a fraction of the cost of RNA-Seq, a current Affymetrix version of the human transcriptome array [90] is being currently tested for allelic-specific expression, and alternative splicing.

In our laboratory we have good success, high between run reliability, running LCL RNA on the Affymetrix Human Gene Chip Exon 1.0 ST array [25, 28]. Gene expression measures derived with Robust Multichip Average (RMA) [91, 92] can be analyzed with diagnosis and probe sets as main effects and corresponding interactions. Probe sets which contained known SNPs [93] or which cross-hybridized are eliminated from the analysis.

SNP array genotyping
Subjects used for eQTL calculation can have genotypes from any platform, we routinely genotype subjects using the Affymetrix SNP 6.0 array. Using this platform, the genotypes are called using the birdseed v2 algorithm within Affymetrix Genotyping Console. Individual SNPs with call rates <95% were removed, while individuals with >95% call rate were retained for analysis. Furthermore, it is recommended to avoid batch effects through an appropriate randomization of cases with controls at each step of SNP array processing.

To minimize misclassification errors of eQTL status, it is recommended to use cis-eQTLs within 100 kb of the SNP (50 Kb on either side) following Frasier and Xie [29]. However, others have found that genomic segments 1–4 Mb away from a SNP still constitutes a cis-eQTL, since many regulatory elements are not directly adjacent to a gene, thus different sliding windows in this range can be reliably used. The identification of true positive trans-acting eQTL SNPs necessitates a much lower P-value threshold which we have previously estimated at being 10–15 through simulations (our unpublished data).

eQTL calculation and determination
The regression of genotype × expression × diagnosis can be implemented in PLINK [94] or MATLAB environment. A permutation test in which genotypes and gene expression are swapped randomly between each individual is used to empirically calculate false positive rates for eQTL regression. A permutation strategy can also control hidden biases that may be present at prior steps of the analysis. By determining eQTL in two independent samples, false positives will be minimized by looking for replication between samples, and calculation of the false negative rate using a balanced probability analysis in each sample [95].

Tests for over-representation of known and novel pathways of SNPs
Methods using ERMINEJ [96], DAVID [97] and Ingenuity Pathway are used by many researchers for analysis for gene over-representation in known pathways [34]. These are first-pass approaches to look for over-representation in well-described pathways of annotated genes. However, more careful analytical tools have been recently developed for testing SNP overlap in pathways. The SNP ratio test for enrichment of SNPs within KEGG pathways [98] in SZ GWAS is useful. The ALIGATOR algorithm [99], which corrects for linkage disequilibrium between SNPs, gene size differences and multiple testing of non-independent pathways with various levels of P-value criteria for defining which SNPs are significantly associated from eQTL results. There are limitations with the use of gene ontologic-al methods; however, these can be partially overcome through the investigation of networks based upon gene co-expression, gene neighborhoods and testing centrality parameters of networks. Using a variant of this approach for co-expression and imaging genetics QTL for brain activation levels, we predicted that miRNA 137 and related transcripts would be associated with SZ [27], which was independently replicated [100].

Validation of eQTLs
The high-confidence of observed eQTLs can be validated by qPCR, where confirmation is based on P-value and gene expression fold-change between SZ and control cell lines. As an example, in our recent study [25], 13 candidate genes were chosen for qPCR validation and all 13 were concordant with expression results in terms of the direction of fold changes, and the overall validation rate by qPCR was 100% (P < 0.1, two-tailed t-test). One can use programs such as GeNorm [101], or custom algorithms to select reference genes [102],
we routinely select several reference genes for validation of expression when studying eQTL [23, 28]. Using these methods, we have validated an eQTL for IRF5 (Figure 4), a gene also reported by others [29, 103] and AGA (Figure 3) as previously shown [23]. Of course, strong replication across independent studies is required for ultimate validation of an eQTL finding.

**Molecular signature analysis, cross validation and GRNs of eQTLs**

After finding eQTLs as described above, those SNP-expression pairs can be used as putative biomarkers in case–control cross validation within a study, and across studies. Further, biomarkers can be used that do not have an eQTL. The significant eQTL locus (both genotype and expression) can be entered into linear discriminant analysis to classify cases and controls, using one-half of the samples to analyze for the genotype × expression interaction, and the second half of the sample to test. Likewise, those genotype × expression pairs with high cross validation rates can be tested in different case–control studies of different tissues preferably brain. Strong tissue- and population-specific eQTLs might be identified, perhaps linked to disease-associated loci in each population.

Initial networks of gene–gene expression can be built from peripheral biomarkers and test nodes centered on GWAS and cis-regulatory eQTL SNPs. These functional networks can be tested by silencing key network nodes in vitro, and measuring those impacts in known genetic variants. False positive and possible incidental findings can be minimized by combining multiple genome-scale data sets from different laboratories, tissues and populations. These networks could also improve our understanding of the biological basis for psychoses, and could be used for risk assessment in early onset psychosis patients.

Networks can be derived through the implementation of recently developed algorithms, QCUT [104, 105] and CentiScaPe [106] to compute specific centrality parameters describing the network topology and the most significant nodes in a complex network [106]. These and other Cytoscape [107] plug-ins (Network Analyzer [108], Network Compare[108]) are routinely used to define and alter network topology. Differences between networks (e.g. SZ versus controls) can be evaluated by random permutation of network connections. Following the identification of significant differences in a community of genes between SZ and controls, indirect validation by in silico miswiring of the network, using substitution of gene expression values and rerunning the network centrality algorithm, e.g. CentiScaPe [106] can be tested. Further, disease related co-expression networks with another pioneering method—WGCNA (weighted correlation network analysis) [109, 110]—to find disease-related networks (modules) and disease related hub genes from biomarker expression data can be run.

After developing in silico models with SNP eQTL-trait association data, there remains a need to test the coherency and robustness of these models, based upon in vitro miswiring. Specific network criteria can be used to select genes for RNA interference (RNAi) validation. One criterion is for central nodes to have a high degree of connectedness, which show significant differential expression between SZ and controls or based upon highly significant eQTL/trait-associated genes. RNAi is a rapid, effective method to knock down gene expression and study the effects of loss of gene function. Many approaches have been developed for the introduction of RNAi into cells including short-interfering (siRNA) and short hairpin RNA (shRNA) [111–114].

**CONCLUSION**

The search for biomarkers in SZ is advancing at a rapid pace and the use of combinatorial and complementary analyses allows for better modeling of this complex disorder. The studies presented in this review attempt to refine biomarkers that are both useful for prediction, and reflecting changes occurring in the brain. Unfortunately, there are several caveats concerning the criteria for comparison of brain, peripheral blood and LCL expression profiles. Multiple factors affecting expression profiling, when using a microarray platform, have been described extensively [43, 115–117]. Nevertheless there has been definite success in determining and replication of the impact of SNPs upon gene expression [25, 64, 118]. Encouraging findings from peripheral studies carried out thus far have shown some overlap with brain expression differences and some studies show eQTL consistency across brain and periphery for a subset of eQTLs. It is noteworthy that eQTLs have particularly large genetic effects thus finding the same loci associated with disease provides, at minimum, compelling annotation of GWAS results.
These proposed methods for biomarker and eQTL analyses, and similar implementations, can then be the plausible basis for a clinically based study designed to be representative of a general patient population commonly ascertained at a clinical center, for biomarker validation and applicability.

**Key Points**
- An eQTL can be broadly defined as a gene expression trait found to significantly correlate with a genetic variant in a cis or trans location.
- EQTLs can be used to enhance annotation of GWASs.
- Integrative methodologies enable better modeling and robust investigation of complex psychiatric disorders.

**FUNDING**
The author acknowledges support from the Pritzker Family Philanthropic Fund (MPV), NIMH R01MH085801 (MPV), and the William Lion Penzner Foundation (UCI, Dept of Psychiatry).

**References**

Integrative approaches to biomarker discovery


30. Sequeira PA, Martin MV, Vawter MP. The first decade and beyond of transcriptional profiling in schizophrenia. *Neurobiol Dis* 2011 [Epub ahead of print; doi:10.1016/j.nbd.2011.03.001].


