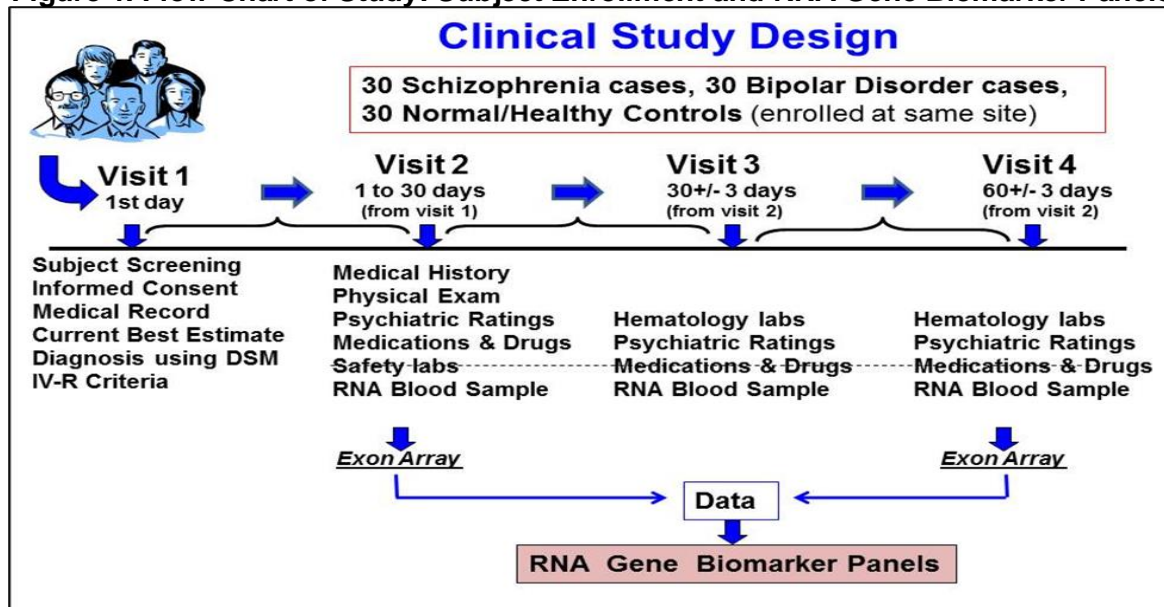


Published in Molecular Neuropsychiatry. Our clinical tests and initial clinical trial were based on original work and were published in Molecular Neuropsychiatry in 2018. – Vawter MP, Philibert R, Rollins B, Ruppel PL, Osborn TW. Exon Array Biomarkers for the Differential Diagnosis of Schizophrenia and Bipolar Disorder. Mol Neuropsychiatry. 2018;3(4):197-213. © 2018 S. Karger AG, Basel

Study Summary.

This study developed potential blood-based biomarker tests for diagnosing and differentiating schizophrenia (SZ), bipolar disorder type I (BD), and normal control (NC) subjects using mRNA gene expression signatures. Ninety subjects ($n = 30$ each for the three groups of subjects) provided blood samples at three visits. (At the time, this was the first gene expression study with multiple time points in these disorders.) See Study Design in **Figure 1**. The Affymetrix exon microarray was used to profile the expression of over 1.4 million probesets. We selected potential biomarker panels using the probesets' temporal stability and back-tested them at two visits for each subject. Using logistic regression modeling, the 18-gene biomarker panels correctly differentiated the three groups of subjects with high accuracy across the two different clinical visits (83–96% accuracy). The primary endpoint was the area under the curve (AUC) of the receiver operating characteristic (ROC), using transcript abundance from Affymetrix exon array analysis to classify SZ and BD cases against HC. The optimal cut point from the ROC analysis for each gene and gene composite marker was used to assess each marker's sensitivity, specificity, and odds ratio. Our 20 RNA gene panel diagnosed schizophrenia and bipolar disorder with high selectivity and high sensitivity (both over 95%), plus it effectively differentiated between SZ and BD patients. The results were consistent with the "leave-one-out" analyses, indicating that the models should be predictive when applied to independent data cohorts. Many of the SZ and BD subjects were taking antipsychotic and mood stabilizer medications at the time of the blood draw, raising the possibility that these drugs could have affected some of the differential transcription signatures (See Drug-Free Subjects Section.) We confirmed select transcripts using two different platforms: quantitative PCR and the Nanostring nCounter® System. The episodic nature of psychiatric disorders might lead to highly variable results depending on when blood is collected in relation to the severity of the disease/symptoms. We have found stable trait gene panel markers for lifelong psychiatric disorders that may have diagnostic utility in younger undiagnosed subjects where there is a critical unmet need. The study requires replication in subjects for ultimate proof of the utility of the differential diagnosis.

Figure 1. Flow Chart of Study: Subject Enrollment and RNA Gene Biomarker Panels



Methods

Subject Enrollment

Subject enrollment occurred at a single clinical site at the University of Iowa. The University of Iowa and the Institutional Review Board approved the procedures for the study. The chronic SZ and BD type I outpatients aged 18–45 years provided consent for the study. All subjects (SZ, $n = 30$; BD, $n = 30$; and NC, $n = 30$) met the DSM-IV-R criteria and completed the study. Clinical assessments included the Scale for the Assessment of Positive Symptoms and Scale for the Assessment of Negative Symptoms (SAPS, SANS), medications, and drugs for the SZ and BD subjects, as well as the Young Mania Rating Scale (YMRS) and Hamilton Rating Scale for Depression (HAM-D or HRSD) for the BD subjects. These neuropsychiatric assessment data were analyzed separately for state-biomarker relationships. The mental state examination for the NC subjects consisted of the Mini-Mental State Examination. The outline of the study is shown in **Figure 1**. The demographics of the SZ, BD, and NC subjects are shown in online supplementary Table 1 (see www.karger.com/doi/10.1159/000485800) for the subjects' age, sex, duration of illness, and ethnicity.

Human Exon Array for Biomarker Profiling

There are advantages to using the Affymetrix exon arrays [65] compared to whole transcriptome shotgun sequencing (RNA-Seq). At the time of sample collection, the cost factor was favorable for future clinical biomarker trials that would require hundreds of arrays compared to the cost of RNA-Seq for the entire validation. The processing time and data storage requirements are more feasible for a study of this size using exon arrays. Whole blood samples were collected in Tempus Blood RNA tubes (Thermo Fisher Scientific) from the SZ, BD, and NC subjects at 4 visits spanning 3 months. For this report, the Tempus tubes from visits 2 and 4 were extracted, and RNA gene expression was measured using Affymetrix exon arrays for both visits on all 90 subjects. High-quality RNA was extracted from the Tempus tubes utilizing the manufacturer's protocol, and quality was assessed on an Agilent Bioanalyzer using the RNA integrity number. The exon arrays were run at the Functional Genomics Laboratory, University of California, using the manufacturer's protocol (Affymetrix, Santa Clara, CA, USA). The Functional Genomics Laboratory has run over 1,000 Affymetrix arrays with high-quality call rates.

Data Analysis

The Affymetrix exon array CEL files were imported into Partek Genomics using batch effect removal. The batch effect was based upon exon array scan dates, as usually 12 arrays were scanned in a single day. The mean intensity of probes was summarized at the probeset level. The resulting probesets were then median-centered within each exon array sample individually ($n = 180$). A two-factor ANOVA was run for each probe- set, using diagnosis, visit, and diagnosis \times visit interaction. Visit was a repeated measure to filter out genes that change significantly between visits. A false discovery rate (FDR) of 6×10^{-8} was established for diagnosis effect based upon 835,000 probe- sets. Three filters were used to select probesets from the ANOVA results that passed the FDR for diagnosis: (1) the most significant p values for BD compared to NC; (2) the most significant p values for BD compared to SZ; and (3) the most significant p values for SZ compared to NC. This resulted in a list of top probesets that was then reduced to probesets that mapped to known RefSeq genes. The top 100 RefSeq probesets for each of the three filters above were combined, and the resulting top 300 probesets were evaluated for biomarker signature.

Biomarker Signature

The modeling proceeded in four steps to select the most predictive panel of probesets out of the top 300 in each step for discriminating between groups:

Step 1: NC versus BD + SZ

Step 2: NC versus SZ

Step 3: NC versus BD

Step 4: SZ versus BD

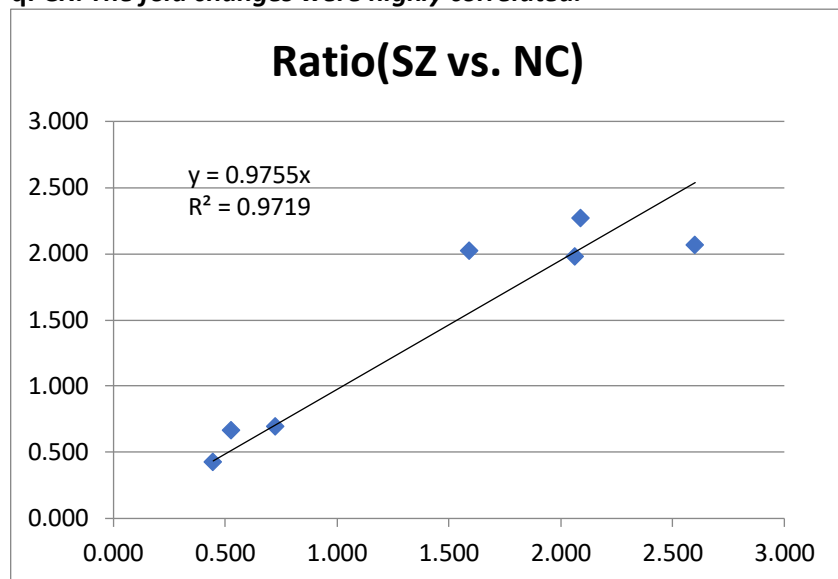
Multivariate logistic regression modeling with forward step-wise selection (SAS PROC LOGISTIC) was used on the combined visit 2 and 4 data from the groups included in the step to select the probesets that discriminated most strongly between the groups. We used forward step-wise regression to select probesets that differentiated two groups at a time (BD vs. SZ, BD vs. NC, and SZ vs. NC). A probeset was added to the model if the estimate was the most significant with $p < 0.001$, and the resulting ROC AUC also

retained statistical significance. Forward selection stopped when potential probesets were no longer statistically significant or did not improve the ROC AUC by more than 1%. Processing for each step resulted in a subset of the 300 probesets where each probeset contributed significantly to the model, and the panel represented the smallest number of probesets with a very high diagnostic utility based on the ROC AUC. Modeling for the diagnostic for each step was applied to the visit 2 data using the identified probesets. The optimal cut-point for discriminating between the groups based on the logistic model prediction was obtained by maximizing the Youden index J [67], where: $J = \text{True positive rate} - \text{FPR}$. The visit 2 prediction model was then applied to the visit 4 data to assess utility for the second set of data, which included stability over time. Further evaluation for each of the four panels included "leave-one-out" cross-validation where one subject was sequentially left out of the logistic model fit using the remaining subjects and then the model's predictability for the excluded subject was assessed. This tested whether outliers in the data set were driving the model.

Quantitative PCR

Transcripts were selected for quantitative PCR (qPCR) validation based upon significant differences using the ANOVA filter. We selected transcripts that represented a combination of the most significant ANOVA p values for SZ compared to NC and represented fold changes greater than 1.25. See **Figure 2**.

Figure 2. The fold change ratio (SZ compared to NC) for exon microarray results were compared to the fold change ratio (SZ compared to NC) of qPCR. The fold changes were highly correlated.



We initially selected qPCR to validate the exon array findings and later, after completing the entire biomarker panel analysis, used NanoString (see below) for validation. Standard SYBR Green qPCR methods previously described by the Functional Genomics Laboratory (University of California, Irvine) were used to confirm gene expression values derived from the exon array data set [31]. Briefly, in developing SYBR Green assays, we use exon junction-crossing primers to eliminate any genomic DNA from amplification. We assess primers for amplification consistency by single dissociation peaks to represent a single region of cDNA amplification, and minimal primer-dimer formation that could interfere with the amplification signal. We require the primers to amplify genes in our samples at fewer

than 35 cycles to be usable. We also run samples in triplicate, and use two housekeeping genes (*SDHA* and *HPRT1*) that have Ct within similar ranges to those of the genes being assayed.

Results

The top 300 probesets from the Affymetrix exon microarray based upon ANOVA significance for differentiating BD, SZ, and NC subjects. The resulting biomarker signature was composed of 23 probesets that condensed into 18 known Ref-Seq genes (biomarker panel; **Table 1**). The diagnostic logistic model was built in four steps, using the visit 2 transcripts shown in **Table 1**.

Table 1. Probesets that were found to reliably discriminate BD, SZ, and NC were assigned to known RefSeq transcripts

Biomarker panel (comparisons)	Affymetrix exon microarray transcript ID	Gene
BD-NC	2661992	OXTR
BD-NC	3195034	PTGDS
BD-NC	3333247	FADS2
BD-NC	4048241	HLA-DRB5
BD-SZ	2661992	OXTR
BD-SZ	3333247	FADS2
BD-SZ	3554818	CRIP2
BD-SZ-NC	2418570	SLC44A5
BD-SZ-NC	2545092	HADHA
BD-SZ-NC	2647109	CPA3
BD-SZ-NC	2661992	OXTR
BD-SZ-NC	2739160	CCDC109B
BD-SZ-NC	2906720	TREML4
BD-SZ-NC	3063536	TRIM4
BD-SZ-NC	3195034	PTGDS
BD-SZ-NC	3667890	HPR
BD-SZ-NC	3846538	EEF2
BD-SZ-NC	3908149	ZMYND8
SZ-NC	2401347	TCEA3
SZ-NC	2418570	SLC44A5
SZ-NC	2660617	IL5RA
SZ-NC	3329099	GYLTL1B
SZ-NC	3766893	DDX5

BD, bipolar disorder; SZ, schizophrenia; NC, normal control.

The resulting logistic predictive model based on visit 2 was then applied to the visit 4 data. The summary of individual steps in the construction of the biomarker gene panels is shown in **Table 2**. The diagnostic algorithm uses a four-step decision model: step 1, BD and SZ versus NC; step 2, SZ versus NC; step 3, BD versus NC; and step 4, SZ versus BD.

Table 2. Prediction model trained on combined visit 2 and 4 data and tested on each visit individually

Actual	Step 1		Step 2		Step 3		Step 4		Final call		
	NC	SZ + BD	NC	SZ	NC	BD	SZ	BD	NC	SZ	BD
Training visit 2											

30 NC	30	0	24	6	26	4		21	5	4	
30 SZ	3	27	1	29			29	1	3	26	1
30 BD	1	29			1	29	1	29	1	1	28
Testing visit 4											
30 NC	28	2	23	7	28	2			22	6	2
30 SZ	1	29	0	30			29	1	1	28	1
30 BD	1	29			0	30	0	30	1	0	29

The confusion matrices show the numbers of subjects correctly and incorrectly classified for each iteration of the classification. Step 1: 11 genes diagnostically differentiating BD + SZ from NC. Step 2: 5 genes diagnostically differentiate SZ from NC. Step 3: 4 genes diagnostically differentiate BD from NC. Step 4: 3 genes diagnostically differentiate SZ from BD. Final call: summary of steps 1, 2, 3, and 4 classifications. BD, bipolar disorder; SZ, schizophrenia; NC, normal control.

The 18-gene biomarker panels, using logistic regression modeling, correctly differentiated the three groups of subjects (SZ, BD type I, and NC) with high accuracy at visit 2 and visit 4. The visit 2 cut-point probabilities for the SZ-NC comparison were significantly correlated with the visit 4 cut-point probabilities ($p < 0.0001$) with $r = 0.74$ (95% CI 0.59–0.83) showing temporal stability (**Table 3**).

Table 3. The overall results show a stable clinical biomarker signature of mRNA expression across a 90-day test-retesting period with an accuracy of 88% on the retesting data

Predicted	Visit 2 actual			Visit 4 actual			Both actual		
	BD	SZ	NC	BD	SZ	NC	BD	SZ	NC
BD	28	1	4	29	1	2	57	2	6
SZ	1	26	5	0	28	6	1	54	11
NC	1	3	21	1	1	22	2	4	43
Accuracy	83%			88%			86%		
Sensitivity (SZ)	87%			93%			90%		
Sensitivity (BP)	93%			97%			95%		
Specificity (NC)	70%			73%			72%		

BD, bipolar disorder; SZ, schizophrenia; NC, normal control.

The initial model was developed to select stable probesets across visits, and all subjects and visits were incorporated to select the most informative probesets. To test that no single subject was overly influential in determining the model, the initial probesets were evaluated by a “leave-one-out” method, whereby a new model is fit to the remaining subjects, and the left-out subject is identified. “Leave-one-out” cross-validation is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction to estimate how accurately a predictive model will perform in practice. This cross-validation was applied to the visit 2 data from each of the four probeset panels (**Table 4**). The results are very consistent with the actual data, and the “leave-one-out” analyses indicate that the models should be predictive when applied to independent data cohorts.

Table 4. The overall results fitting a “leave-one-out” validation model to the remaining subjects show a stable clinical biomarker signature of mRNA expression across a 90-day test-retesting period with an accuracy greater than 87% for the retesting data at each of the model classification steps

	BD + SZ vs. NC (11- gene panel)	SZ vs. NC (5-gene panel)	BD vs. NC (4-gene panel)	SZ vs. BD (3-gene panel)
<i>Actual visit 2 data</i>				
Sensitivity (SZ)	93%	97%	–	97%
Sensitivity (BP)		–	97%	97%
Specificity (NC)	100%	80%	87%	–
Accuracy	96%	88%	92%	97%
<i>“Leave-one-out” cross-validation</i>				
Sensitivity (SZ)	90%	93%	–	90%
Sensitivity (BP)		–	93%	93%
Specificity (NC)	83%	80%	80%	–
Accuracy	88%	87%	87%	92%
BD, bipolar disorder; SZ, schizophrenia; NC, normal control.				

The AUC for each step was greater than 0.95, indicating the high combined sensitivity and specificity of the classification into three groups (**Table 5**). The data were not normalized to blood counts (CBC measurements) for our main data analysis reported in this paper.

Table 5. The diagnostic algorithm uses four individual steps, shown in each column.

BD + SZ vs. NC (11-gene panel)	SZ vs. NC (5-gene panel)	BD vs. NC (4-gene panel)	SZ vs. BD (3-gene panel)
AUC = 0.994 ($p < 0.0001$)	AUC = 0.954 ($p < 0.0001$)	AUC = 0.974 ($p < 0.0001$)	AUC = 0.998 ($p < 0.0001$)
BD, bipolar disorder; SZ, schizophrenia; NC, normal control.			

Discussion.

To determine stable temporal biomarkers, this study evaluated whole blood gene expression at two different time points using the same subjects (SZ, BD, and NC) for differential diagnosis. The diagnostic algorithm used logistic regression modeling and a total of 18 uniquely expressed exons within known mRNA transcripts. The model discriminated SZ and BD from each other, as well as both from healthy controls in four steps. The upper limit of accuracy achieved in this biomarker study was 88%, using an independent visit of the same patients. When using the “leave-one-out” evaluation algorithm, the results were very consistent with the actual data; thus, the “leave-one-out” analyses indicated that the models were not driven by outliers and that they should be predictive when applied to independent data cohorts.

It is expected that the application of these panels to first-episode or prodromal subjects may improve prediction for those subjects that ultimately convert to either illness, as well as for the millions of patients worldwide that have not received any clear diagnosis of their ongoing disorder. This will require an additional validation study of the biomarker signatures with a larger cohort in a follow-on project.

The differences in expression of 3 genes (PTGDS, FADS2, and HADHA) related to polyunsaturated fatty acid (PUFA) and prostaglandin biosynthesis were used for the final biomarker panels to differentiate between SZ, BD, and NC. Previously, these genes have been associated with psychiatric disorders such as BD, major affective disorder, SZ, and anxiety. PTGDS is involved in the synthesis of PGD2 from PGH2, the cyclooxygenase-mediated product of arachidonic acid which is a PUFA [4]. PTGDS is a top anxiety gene modulated by changes in PUFA (omega-3 fatty acid docosahexaenoic acid) [73] on the convergent functional genomics scale. Increased expression of FADS2 has been found in SZ and BD brains post mortem [74, 75]. FADS2 activity is increased in BD and is associated with suicidal behavior [76].

In the present study, we found an increased expression of FADS2 in BD, in agreement with the FADS2 findings reported. The increased activity of FADS2 could reduce PUFA levels of both arachidonic acid and eicosapentaenoic acid by promoting conversion to longer-chain fatty acids, shown in both the *n*-3 and the *n*-6 pathway (Fig. 5 - See complete article). Thus, PUFA supplementation with *n*-3 fatty acids in mood disorder was effective in reducing mood symptoms in 4 out of 7 well-controlled studies [76]. The expression data for FADS2, while interesting, could be subject to dietary influence, such as amounts and types of daily dietary intake of fatty acids, the timing of intake, and also medication effects on these genes. Further, genetics plays a significant role, especially in modulating levels of fatty acids and FADS2 expression. Another limitation to the assessment of these genes as representing actual pathophysiological markers is that, potentially, stress could modulate the biomarker panel genes. Many patients with BD and SZ experience higher levels of stress than controls, which might explain differences in immune cell activation and prostaglandin synthesis. We examined our biomarker panel of 18 genes in our unpublished stress data set (M. Martin and MPV) using the same exon array approach and Tempus tube approach on healthy volunteers who underwent sleep deprivation and 9 repeated blood draws over 54 h, i.e., every 6 h. We checked our results for the healthy controls and found that 4 transcripts that passed Bonferroni corrections were affected by time of day and, potentially, stress induction as well (DDX5, EEF2, HADHA, and CCDC109B). However, these 4 genes did not vary in the present study, even though in the stress data set these genes varied by time of day. Taken together, although these 4 genes were dysregulated as a consequence of time of day and sleep deprivation, these genes in the present study were stable across 8 weeks of time, and would have been expected to show some fluctuation with stress levels or time of day.

Over 100,000 adolescent Americans suffer from symptoms of psychosis each year, as well as millions of patients worldwide that have not received any clear differential diagnosis of their ongoing disorder; yet, currently, there are no biomarker tests that are FDA approved to classify SZ or BD. There is a serious need for “objective” clinical laboratory tests for an early diagnosis of these mental disorders, since today these disorders may typically take months or even years to reach a diagnosis and for patients to receive effective treatment. The lag in treatment is associated with an increase in suicide rates and recurrent episodes of psychosis and mood dysregulation. There is a large increase in deaths reported among first-episode psychotic subjects due to lack of treatment after the first year of illness [1]. Thus, it is important to have objective biomarkers to help implement treatment at an early stage. One estimate of the direct and indirect annual costs in the USA for SZ is USD 174 billion [88], with an additional cost of USD 151 billion for BD [89]. Biomarker signatures could lead to faster and more accurate diagnoses, reducing the duration of untreated psychosis, suicidality, and cognitive decline and adding to an understanding of the shared and unique pathophysiologies of each disorder. The blood test results that are described in this paper, if further validated in a larger number of subjects, will offer molecular diagnostic support for psychiatrists’ clinical evaluation with rapid clinical laboratory test results.

Disclosure Statement.

A SBIR Phase I project “Gene expression exon array biomarkers to diagnose schizophrenia” (R43 MH090806) was awarded (TWO) for developing a blood-based commercial biomarker test for psychiatric disorders. The authors (M.P.V. and T.W.O.) are officers of Laguna Diagnostic, LLC, and co-inventors of a US patent to commercialize the results presented.

References. (See Complete Article)

