Molecular Psychiatry (2009), 1–22

www.nature.com/mp

ORIGINAL ARTICLE

Identification of blood biomarkers for psychosis using convergent functional genomics

SM Kurian[1,6], H Le-Niculescu[2,6], SD Patel[2,6], D Bertram[2,3], J Davis[2,3], C Dike[2,3], N Yehyawi[2,3], P Lysaker[3], J Dustin[2], M Caligiuri[4], J Lohr[4], DK Lahiri[2], JI Nurnberger Jr[2], SV Faraone[5], MA Geyer[4], MT Tsuang[4], NJ Schork[1], DR Salomon[1] and AB Niculescu[2,3]

[1]*Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA, USA;* [2]*Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, USA;* [3]*Indianapolis VA Medical Center, Indianapolis, IN, USA and* [4]*Department of Psychiatry, UC San Diego, La Jolla, CA, USA;* [5]*Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY, USA*

There are to date no objective clinical laboratory blood tests for psychotic disease states. We provide proof of principle for a convergent functional genomics (CFG) approach to help identify and prioritize blood biomarkers for two key psychotic symptoms, one sensory (hallucinations) and one cognitive (delusions). We used gene expression profiling in whole blood samples from patients with schizophrenia and related disorders, with phenotypic information collected at the time of blood draw, then cross-matched the data with other human and animal model lines of evidence. Topping our list of candidate blood biomarkers for hallucinations, we have four genes decreased in expression in high hallucinations states (*Fn1*, *Rhobtb3*, *Aldh1l1*, *Mpp3*), and three genes increased in high hallucinations states (*Arhgef9*, *Phlda1*, *S100a6*). All of these genes have prior evidence of differential expression in schizophrenia patients. At the top of our list of candidate blood biomarkers for delusions, we have 15 genes decreased in expression in high delusions states (such as *Drd2*, *Apoe*, *Scamp1*, *Fn1*, *Idh1*, *Aldh1l1*), and 16 genes increased in high delusions states (such as *Nrg1*, *Egr1*, *Pvalb*, *Dctn1*, *Nmt1*, *Tob2*). Twenty-five of these genes have prior evidence of differential expression in schizophrenia patients. Predictive scores, based on panels of top candidate biomarkers, show good sensitivity and negative predictive value for detecting high psychosis states in the original cohort as well as in three additional cohorts. These results have implications for the development of objective laboratory tests to measure illness severity and response to treatment in devastating disorders such as schizophrenia.
Molecular Psychiatry advance online publication, 17 November 2009; doi:10.1038/mp.2009.117

Keywords: convergent functional genomics; blood; schizophrenia; hallucinations; delusions; biomarkers

Introduction

Our group has developed a powerful combined approach for extracting signal from noise in genetic and gene expression studies, termed convergent functional genomics (CFG). CFG translationally integrates multiple independent lines of evidence-genetic and functional genomic data, from human studies and animal models, as a Bayesian strategy for identifying and prioritizing findings, reducing the false-positives and false-negatives inherent in each individual approach. The CFG methodology has already been applied with some success to help identify and prioritize candidate genes, pathways and mechanisms for neuropsychiatric disorders such as bipolar disorder,[1,2] alcoholism[3] and schizophrenia,[4] as well as blood biomarker discovery for mood disorders.[5] We have now applied this approach (Figures 1 and 2) to blood biomarker discovery efforts for hallucinations and delusions, core symptoms of psychotic disorders. Objective blood biomarkers for illness state and treatment response would make a significant difference in our ability to asses and treat patients with psychotic disorders, eliminating subjectivity and reliance on patient's self-report of symptoms.

Correspondence: Professor AB Niculescu III, Indiana University School of Medicine, Staff Psychiatrist, Indianapolis VA Medical Center, Director, INBRAIN and Laboratory of Neurophenomics, Institute of Psychiatric Research, 791 Union Drive, Indianapolis, IN 46202-4887, USA.
E-mail: anicules@iupui.edu
[6]These authors contributed equally to the work.
Received 23 June 2009; revised 20 August 2009; accepted 18 September 2009

Materials and methods

Human subjects

We present data from four cohorts (Table 1). One cohort consisted of 31 subjects with psychotic

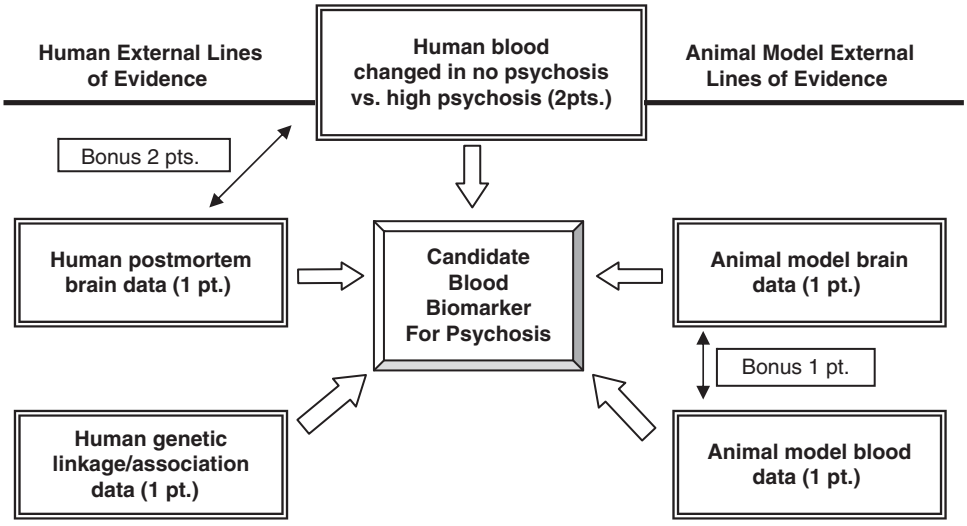

Figure 1 Convergent functional genomics approach for candidate biomarker prioritization. Scoring of independent lines of evidence (maximum score = 9 points).

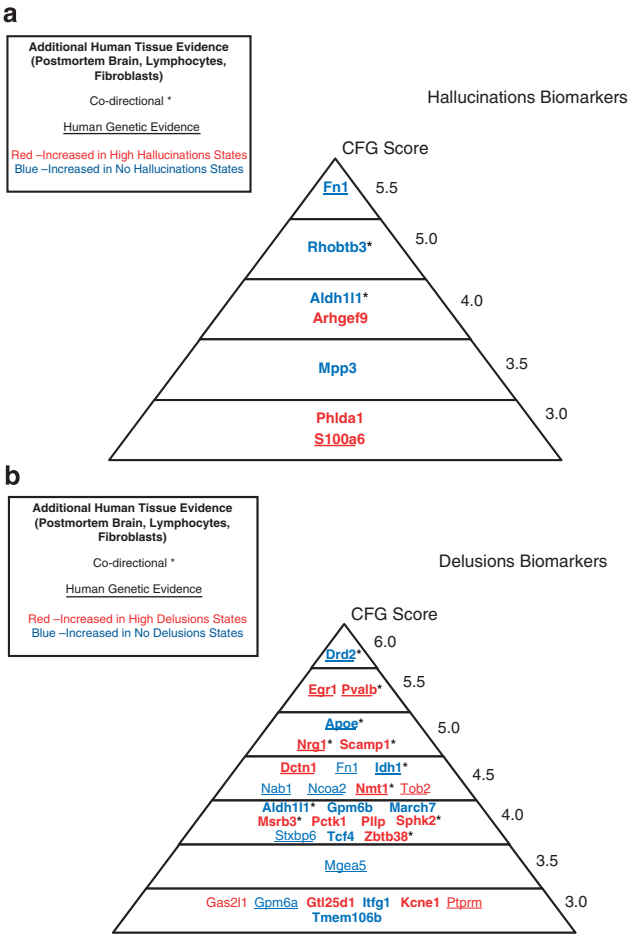

Figure 2 Top blood candidate biomarker genes for (**a**) hallucinations and (**b**) delusions. The CFG lines of evidence scoring are depicted on the right side of the pyramid.

disorders (schizophrenia, schizoaffective disorder and substance-induced psychosis), from which the primary biomarker data was derived, from testing done at their first visit (v1). A second cohort consisted of 17 subjects from the first cohort that had a change in psychotic symptom (hallucinations or delusions) Positive and Negative Symptom Scale (PANSS) scores at follow-up testing 3 months (v2) or 6 months (v3) later. A third cohort consisted of 10 new subjects with psychotic disorders, and the fourth cohort consisted of 9 subjects from the third cohort that had a change in symptom scores at follow-up testing 3 months (v2) later.

Subjects consisted primarily of men (and one woman) over 18 years of age. Subjects were recruited from the patient population at the Indianapolis VA Medical Center and the Indiana University School of Medicine. A demographic breakdown is shown in Table 1. We focused in our initial studies primarily on an age-matched male population, due to the demographics of our catchment area (primarily male in a VA Medical Center), and to minimize any potential gender-related state effects on gene expression, which would have decreased the discriminative power of our analysis given our relatively small sample size. The subjects were recruited largely through referrals from care providers, the use of brochures left in plain sight in public places and mental health clinics, and through word of mouth. Subjects were excluded if they had significant acute medical or neurological illnesses, or had evidence of active substance abuse or dependence. All subjects understood and signed informed consent forms detailing the research goals, procedure, caveats and safeguards. Subjects completed diagnostic assessments by an extensive structured clinical interview—Diagnostic Interview for Genetic Studies—at a baseline visit, followed by up to three testing visits, each three months apart. At each testing visit, they received a psychosis rating scale (PANSS), which includes items that score symptoms of hallucinations and delusions (see Table 2), and blood was drawn. Whole blood (10 ml)

Table 1 Demographics

Subject ID	*Diagnosis*	*Age*	*Gender*	*Ethnicity*	*Hallucination scores*	*Delusion scores*
(A) Individual demographic data						
*Cohort 1: Primary psychosis cohort (*n = *31)*						
phchp003v1	SZ	50	Male	African American	3	1
phchp004v1	SZA	55	Male	African American	1	3
phchp005v1	SZA	45	Male	Caucasian	1	1
phchp006v1	SZA	52	Male	African American	1	3
phchp008v1	SZ	47	Male	African American	4	1
phchp009v1	SZ	55	Male	African American	3	4
phchp010v1	SZA	45	Male	Caucasian	2	2
phchp012v1	SZA	55	Male	Caucasian	3	3
phchp013v1	SZA	53	Male	African American	3	4
phchp014v1	SubPD	55	Male	African American	3	2
phchp015v1	SubPD	48	Male	African American	1	1
phchp016v1	SZ	54	Male	African American	5	5
phchp018v1	SZA	54	Female	Caucasian	4	6
phchp019v1	SubPD	50	Male	African-American	2	3
phchp021v1	SZA	48	Male	Hispanic	5	5
phchp022v1	SZ	48	Male	Caucasian	1	2
phchp024v1	SZA	49	Male	African American	4	2
phchp025v1	SZ	42	Male	Caucasian	5	5
phchp026v1	SZA	49	Male	African American	4	4
phchp033v1	SZA	48	Male	Caucasian	5	4
phchp038v1	SZA	58	Male	African American	1	1
phchp040v1	SZA	50	Male	Caucasian	1	6
phchp041v1	SZ	62	Male	African-American	5	5
phchp042v1	SZA	43	Male	Caucasian	2	4
phchp046v1	SZA	45	Male	Caucasian	1	1
phchp047v1	SZA	57	Male	African American	5	4
phchp048v1	SZA	56	Male	African American	1	1
phchp049v1	SZA	46	Male	Caucasian	1	1
phchp057v1	SZA	47	Male	Caucasian	1	1
phchp061v1	SZ	49	Male	Caucasian	1	4
phchp062v1	SZ	56	Male	Caucasian	4	3
*Cohort 2: Primary psychosis cohort follow-up visit (*n = *17)*						
phchp003v2	SZ	50	Male	African American	3	4
phchp005v2	SZA	45	Male	Caucasian	2	2
phchp006v2	SZA	52	Male	African American	1	1
phchp010v3	SZA	45	Male	Caucasian	1	1
phchp012v2	SZA	55	Male	Caucasian	5	4
phchp013v3	SZA	54	Male	African American	5	4
phchp016v3	SZ	54	Male	African American	4	4
phchp021v3	SZA	49	Male	Hispanic	5	4
phchp022v2	SZ	48	Male	Caucasian	1	1
phchp026v3	SZA	49	Male	African American	1	1
phchp038v3	SZA	59	Male	African American	1	1
phchp040v2	SZA	50	Male	Caucasian	2	5
phchp042v2	SZA	43	Male	Caucasian	3	2
phchp046v2	SZA	45	Male	Caucasian	3	1
phchp047v2	SZA	57	Male	African American	5	5
phchp048v2	SZA	57	Male	African American	1	1
phchp062v2	SZ	56	Male	Caucasian	3	3
*Cohort 3: Second psychosis cohort (*n = *10)*						
phchp017v2	SZA	53	Male	African American	1	1
phchp058v1	SZ	56	Male	African American	1	1
phchp065v1	SZA	62	Male	Caucasian	2	5
phchp068v1	SZA	57	Male	African American	4	3
phchp069v1	SZ	47	Male	Caucasian	4	5
phchp072v1	SZA	60	Male	Caucasian	2	3
phchp073v1	SZA	50	Male	Caucasian	5	4

Table 1 Continued

Subject ID	*Diagnosis*	*Age*	*Gender*	*Ethnicity*	*Hallucination scores*	*Delusion scores*
phchp075v1	SZA	57	Male	Caucasian	4	3
phchp083v1	SZ	50	Male	African American	1	1
phchp085v1	SZA	57	Male	Caucasian	1	4
Cohort 4: Second psychosis cohort follow-up visit (n = *9)*						
phchp058v2	SZ	56	Male	African American	4	3
phchp065v2	SZA	62	Male	Caucasian	1	4
phchp068v2	SZA	57	Male	African American	3	2
phchp069v2	SZ	47	Male	Caucasian	5	6
phchp072v2	SZA	60	Male	Caucasian	2	2
phchp073v2	SZA	50	Male	Caucasian	4	5
phchp075v2	SZA	58	Male	Caucasian	5	3
phchp083v2	SZ	50	Male	African American	1	1
phchp085v2	SZA	57	Male	Caucasian	1	1

	Primary psychosis cohort (n = *31)*			*Primary psychosis cohort follow-up visit* (n = *17)*		
	Schizo-affective	*Schizo-phrenia*	*Substance-induced psychotic disorder*	*Schizoaffective*	*Schizophrenia*	*Substance-induced psychotic disorder*
(B) Aggregate demographic data						
Number of subjects	19	9	3	13	4	0
Gender (males: females)	18:1	9:0	3:0	13:0	4:0	NA
Age, mean years (s.d.) range	50.3 (4.6) 43–58	51.4 (5.9) 48–55	51 (3.6) 48–55	50.8 (5.3) 43–59	52 (3.6) 48–56	NA
Duration of illness mean years (s.d.) range	27.7 (9.7) 5–42	30 (7.6) 42–62	25 (6.2) 20–32	28.7 (9.8) 5–42	27.5 (5.8) 21–35	NA
Ethnicity (Caucasian/other)	10/9	4/5	0/3	6/7	2/2	NA
	Second psychosis cohort (n=*10)*			*Second psychosis cohort follow-up visit* (n=*9)*		
Number of subjects	7	3	0	6	3	0
Gender (males: females)	7:0	3:0	NA	6:0	3:0	NA
Age, mean years (s.d.) range	56.6 (4.0) 53–62	51 (4.6) 47–56	NA	57.3 (4.1) 50– 62	51 (4.6) 47–56	NA
Duration of illness mean years (s.d.) range	35.1 (6.2) 23–43	28 (8) 20–36	NA	35.3 (6.8) 23–43	28 (8) 20–36	NA
Ethnicity (Caucasian/ Other)	5/2	1/2	NA	5/1	1/2	NA

Abbreviations: NA, not available; SubPD, substance-induced psychosis; SZ, schizophrenia; SZA, schizoaffective disorder. Diagnosis established by DIGS comprehensive structured clinical interview. PANSS hallucination and delusion scores at the time of blood draw, on a scale of 1 (no symptoms) to 7 (extremely severe symptoms).

was collected in two RNA-stabilizing PAXgene tubes, labeled with an anonymized ID number, and stored at −80 °C in a locked freezer until the time of future processing. Whole blood (predominantly lymphocyte) RNA was extracted for microarray gene expression studies from the PAXgene tubes blood, as detailed below.

Human blood gene expression experiments and analysis

RNA extraction. 2.5–5 ml of whole blood was collected into each PaxGene tube by routine venipuncture. PaxGene tubes contain proprietary reagents for the stabilization of RNA. The cells from

Table 2 Hallucinations and Delusions scoring as part of administration of the Positive and Negative Symptom Scale (PANSS)

(A) Hallucinatory behavior

Hallucinations score	*State*	*Definition*
1	Absent	Definition does not apply
2	Minimal	Questionable pathology; may be at the upper extreme of normal limits
3	Mild	One or two clearly formed but infrequent hallucinations, or else a number of vague abnormal perceptions which do not result in distortions of thinking or behavior
4	Moderate	Hallucinations occur frequently but not continuously, and the patient's thinking and behavior are affected only to a minor extent
5	Moderate severe	Hallucinations are frequent, may involve more than one sensory modality, and tend to distort thinking and/or disrupt behavior. Patient may have delusional interpretations of these experiences and respond to them emotionally and, on occasion, verbally as well
6	Severe	Hallucinations are present almost continuously, causing major disruption of thinking and behavior. Patient treats these as real perceptions, and functioning is impeded by frequent emotional and verbal responses to them
7	Extreme	Patient is almost totally preoccupied with hallucinations, which virtually dominate thinking and behavior. Hallucinations are provided a rigid delusional interpretation and provoke verbal and behavioral responses, including obedience to command hallucinations

(B) Delusions

Delusions score		
1	Absent	Definition does not apply
2	Minimal	Questionable pathology; may be at the upper extreme of normal limits
3	Mild	Presence of one or two delusions which are vague, uncrystallized, and not tenaciously held. Delusions do not interfere with thinking, social relations, or behavior
4	Moderate	Presence of either a kaleidoscopic array of poorly formed, unstable delusions or a few well-formed delusions that occasionally interfere with thinking, social relations, or behavior
5	Moderate severe	Presence of numerous well-formed delusions that are tenaciously held and occasionally interfere with thinking, social relations or behavior
6	Severe	Presence of a stable set of delusions which are crystallized, possibly systematized, tenaciously held, and clearly interfere with thinking, social relations and behavior
7	Extreme	Presence of a stable set of delusions which are highly systematized or very numerous, and which dominate major facets of the patient's life. This frequently results in inappropriate and irresponsible action, which may even jeopardize the safety of the patient or others

Hallucinatory behavior: Verbal report or behavior indicating perceptions which are not generated by external stimuli. These may occur in the auditory, visual, olfactory or somatic realms. Basis for rating: verbal report and physical manifestations during the course of interview.
Delusions: Beliefs which are unfounded, unrealistic, and idiosyncratic. Basis for rating: thought content expressed in the interview.

whole blood were concentrated by centrifugation, the pellet washed, resuspended and incubated in buffers containing Proteinase K for protein digestion. A second centrifugation step was done to remove residual cell debris. After the addition of ethanol for an optimal binding condition the lysate was applied to a silica-gel membrane/column. The RNA bound to the membrane as the column was centrifuged and contaminants were removed in three wash steps. The RNA was then eluted using diethylpyrocarbonate-treated water.

Globin reduction. To remove globin messenger RNA (mRNA), total RNA from whole blood was mixed with a biotinylated Capture Oligo Mix that is specific for human globin mRNA. The mixture was then incubated for 15 min to allow the biotinylated oligonucleotides to hybridize with the globin mRNA. Streptavidin magnetic beads were then added, and the mixture was incubated for 30 min. During this incubation, streptavidin binds the biotinylated oligonucleotides, thereby capturing the globin mRNA on the magnetic beads. The streptavidin magnetic beads were then pulled to the side of the tube with a magnet, and the RNA, depleted of the globin mRNA, was transferred to a fresh tube. The treated RNA was further purified using a rapid magnetic bead-based purification method. This consists of adding an RNA binding bead suspension to the samples, and using magnetic capture to wash and elute the GLOBINclear RNA.

Sample labeling. Sample labeling was performed using the Ambion MessageAmp II-Biotin*Enhanced* amplified RNA (aRNA) amplification (Ambion Inc., Austin, TX, USA). The procedure is briefly outlined below and involves the following steps:

1. *Reverse transcription to synthesize first strand complementary DNA (cDNA)* is primed with the T7 Oligo(dT) Primer to synthesize cDNA containing a T7 promoter sequence.

Table 3 High- and low-threshold analyses in the primary psychosis cohort ($n = 31$)

Thresholds	*Hallucination analyses (12 no hallucinations and 11 high hallucinations)*	*Delusion analyses (9 no delusions and 13 high delusions)*
High-threshold candidate biomarker genes (changed in $\geqslant$ 75% subjects; that is, at least 3-fold enrichment)	9/12 no hallucinations vs 9/11 high hallucinations A/P and P/A analysis	7/9 no delusions vs 10/13 high delusions A/P and P/A analysis
Low-threshold candidate biomarker genes (changed in $\geqslant$ 60% subjects; that is, at least 1.5-fold enrichment)	8/12 no hallucinations vs 7/11 high hallucinations A/P and P/A analysis	6/9 no delusions vs 8/13 high delusions A/P and P/A analysis

Genes are considered candidate biomarkers for high hallucinations or high delusions if they are called by the Affymetrix MAS5 software as absent (A) in the blood of no hallucination or, no delusion subjects, and detected as present (P) in the blood of high hallucination or high delusion subjects. Conversely, genes are considered candidate biomarkers for no hallucinations or no delusions if they are detected as present (P) in no hallucination or no delusion subjects and absent (A) in high hallucination or high delusion subjects.

2. *Second strand cDNA synthesis* converts the single-stranded cDNA into a double-stranded DNA template for transcription. The reaction employs DNA Polymerase and RNase H to simultaneously degrade the RNA and synthesize second strand cDNA.
3. *cDNA purification* removes RNA, primers, enzymes and salts that would inhibit *in vitro* transcription.
4. *In vitro transcription to synthesize aRNA* with Biotin-NTP Mix generates multiple copies of biotin-modified aRNA from the double-stranded cDNA templates; this is the amplification step.
5. *aRNA purification* removes unincorporated NTPs, salts, enzymes and inorganic phosphate to improve the stability of the biotin-modified aRNA.

Microarrays. Biotin-labeled aRNAs were hybridized to Affymetrix HG-U133 Plus 2.0 GeneChips (Affymetrix, Santa Clara, CA, USA; with over 40 000 genes and expressed sequence tags (ESTs)), according to the manufacturer's protocols. http://www.affymetrix.com/support/technical/manual/expression_manual.affx. Arrays were stained using standard Affymetrix protocols for antibody signal amplification and scanned on an Affymetrix GeneArray 2500 scanner with a target intensity set at 250. Present/absent calls were determined using GCOS software with thresholds set at default values. Quality control measures including 3′/5′ ratios for glyceraldehyde 3-phosphate dehydrogenase and β-actin, scale factors, background and *Q* values were within acceptable limits.

Analysis

We have used the subject's psychosis scores at time of blood collection, specifically the scores for hallucinations (from 1—no symptoms to 7—extreme symptoms) and the scores for delusions (1–7), obtained from a PANSS scale (Table 2). We looked only at all or nothing gene expression differences that are identified by Absent (A) vs Present (P) Calls in the Affymetrix MAS software. We classified genes whose expression is detected as Absent in the asymptomatic subjects (no hallucinations or no delusions, scores of 1) and detected as Present in the highly symptomatic subjects (high hallucinations or high delusions, scores of 4 and above), as being candidate biomarker genes for high hallucinations or high delusions states, respectively. Conversely, genes whose expression are detected as Present in the asymptomatic subjects and Absent in the highly symptomatic subjects are being classified as candidate biomarker genes for no hallucinations or no delusions states, respectively.

We employed two thresholds for analysis of gene expression differences between no symptoms and high symptoms (Table 3). First we used a high threshold, with at least 75% of subjects in the cohort showing a change in expression from Absent to Present between no symptoms and high symptoms (reflecting an at least threefold psychosis state related enrichment of the genes thus filtered). We also used a low threshold, with at least 60% of subjects in the cohort showing a change in expression from Absent to Present between no symptoms and high symptoms (reflecting an at least 1.5-fold psychosis state related enrichment of the genes thus filtered).

The higher threshold would identify candidate biomarker genes that are more common for all subjects, with a lower risk of false positives, whereas the lower threshold will identify genes that are present in more restricted subgroups of subjects, with a lower risk of false negatives. The high threshold candidate biomarker genes have, as an advantage, a higher degree of reliability, as they are present in at least 75% of all subjects with a certain hallucinations state (high symptoms or no symptoms) tested. They may reflect common aspects related to psychosis across a diverse subject population, but may also be a reflection of the effects of common medications used in the population tested, such as antipsychotics. The low threshold genes may have lower reliability, being present in at least 60% of the subject population

tested, but may capture more of the diversity of genes and biological mechanisms present in a genetically diverse human subject population.

Animal model gene expression studies

Our schizophrenia pharmacogenomic model consists of phencyclidine (PCP) and clozapine treatments in mice (see Le-Niculescu *et al*.[4] for experimental details and analysis/categorization of brain gene expression data).

For the current work, we repeated that series of experiments, to obtain blood gene expression data. All experiments were performed with male C57/BL6 mice, 8–12 weeks of age, obtained from Jackson Laboratories (Bar Harbor, ME, USA), and acclimated for at least 2 weeks in our animal facility prior to any experimental manipulation.

Mice were treated by intraperitoneal injection with single-dose saline, PCP ($7.5\,mg\,kg^{-1}$), clozapine ($2.5\,mg\,kg^{-1}$), or a combination of PCP and clozapine (7.5 and $2.5\,mg\,kg^{-1}$). Three independent *de novo* biological experiments were performed at different times. Each experiment consisted of three mice per treatment condition, for a total of nine mice per condition across the three experiments.

Mouse blood collection. Twenty-four hours after drug administration, following behavioral testing, the mice were decapitated to harvest blood. The headless mouse body was put over a glass funnel coated with heparin and approximately 1 ml of blood/mouse was collected into a PAXgene blood RNA collection tubes (Qiagen/BD Diagnostics, Valencia, CA, USA). Blood samples from three mice per treatment condition were pooled. The PAXgene blood vials were stored in $-4\,^{\circ}C$ overnight, and then at $-80\,^{\circ}C$ until future processing for RNA extraction.

RNA extraction and microarray work. For the whole mouse blood RNA extraction, PAXgene blood RNA extraction kit (PreAnalytiX, a Qiagen/BD Company) was used, followed by GLOBINclear–Mouse/Rat (Ambion Inc.) to remove the globin mRNA. All the methods and procedures were carried out as per manufacturer's instructions. The quality of the total RNA was confirmed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). The quantity and quality of total RNA was also independently assessed by 260 nm ultraviolet absorption and by 260/280 ratios, respectively with a Nanodrop spectrophotometer (Thermo Scientific, Wilmington, DE, USA). Starting material of total RNA labeling reactions was kept consistent within each independent microarray experiment. Equal amounts of total RNA extracted from pooled blood samples was used for labeling and microarray assays. We used Mouse Genome 430 2.0 arrays (Affymetrix). The GeneChip Mouse Genome 430 2.0 Array contain over 45 000 probe sets that analyze the expression level of transcripts and variants from over 34 000 well-characterized mouse genes. Standard Affymetrix protocols were used to reverse transcribe the mRNA and generate biotinlylate cRNA (http://www.affymetrix.com/support/downloads/manuals/expression_s2_manual.pdf). The amount of cRNA used to prepare the hybridization cocktail was kept constant intra-experiment. Samples were hybridized at $45\,^{\circ}C$ for 17 h under constant rotation. Arrays were washed and stained using the Affymetrix Fluidics Station 400 and scanned using the Affymetrix Model 3000 Scanner controlled by GCOS software. All sample labeling, hybridization, staining and scanning procedures were carried out as per manufacturer's recommendations. All arrays were scaled to a target intensity of 1000 using Affymetrix MASv 5.0 array analysis software. Quality control measures including 3′/5′ ratios for glyceraldehyde 3-phosphate dehydrogenase and β-actin, scaling factors, background, and *Q* values were within acceptable limits.

Microarray data analysis. Data analysis was performed using Affymetrix Microarray Suite 5.0 software (MAS v5.0). Default settings were used to define transcripts as present (P), marginal (M) or absent (A). A comparison analysis was performed for each drug treatment, using its corresponding saline treatment as the baseline. 'Signal,' 'Detection,' 'Signal Log Ratio,' 'Change' and 'Change *P*-value' were obtained from this analysis. Only transcripts that were called Present in at least one of the two samples (saline or drug) intra-experiment, and that were reproducibly changed in the same direction in at least two out of three independent experiments, were analyzed further.

Cross-validation and integration: CFG

Gene identification. The identities of transcripts were established using NetAffx (Affymetrix), and confirmed by cross-checking the target mRNA sequences that had been used for probe design in the Mouse Genome 430 2.0 Array GeneChip or the Affymetrix Human Genome U133 Plus 2.0 GeneChip with the GenBank database. Where possible, identities of ESTs were established by BLAST searches of the nucleotide database. A National Center for Biotechnology Information (NCBI) (Bethesda, MD, USA) BLAST analysis of the accession number of each probe-set was done to identify each gene name. BLAST analysis identified the closest known gene existing in the database (the highest known gene at the top of the BLAST list of homologues) which then could be used to search the GeneCards database (Weizmann Institute, Rehovot, Israel). Probe sets that did not have a known gene were labeled 'EST' and their accession numbers kept as identifiers.

Human postmortem brain convergence. Information about our candidate genes was obtained using GeneCards, the Online Mendelian Inheritance of Man database (http://ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM), as well as database searches using

PubMed (http://ncbi.nlm.nih.gov/PubMed) and various combinations of keywords (gene name, psychosis, schizophrenia, schizoaffective, human, brain, postmortem). Postmortem convergence was deemed to occur for a gene if there were published reports of human postmortem data showing changes in expression of that gene in brains from patients with psychotic disorders (schizophrenia, schizoaffective d/o). In terms of concordance of direction of change in expression between published postmortem brain data and our human blood data, we made the assumption that schizophrenia postmortem brain data reflected a highly symptomatic phase of the illness. While this may arguably be the case, it is nevertheless an assumption, as no consistent objective data exists regarding the phase of the illness when the subjects deceased, which is one of the limitations of human postmortem brain data to date.

Human genetic data convergence. To designate convergence for a particular gene, the gene had to have published positive reports from candidate gene association studies, or map within 10 cM of a microsatellite marker for which at least one published study showed evidence for genetic linkage to psychotic disorders (schizophrenia or schizoaffective disorder). The University of Southampton's sequence-based integrated map of the human genome (The Genetic Epidemiological Group, Human Genetics Division, University of Southampton: http://cedar.genetics.soton.ac.uk/public_html/) was used to obtain cM locations for both genes and markers. The sex-averaged cM value was calculated and used to determine convergence to a particular marker. For markers that were not present in the Southampton database, the Marshfield database (Center for Medical Genetics, Marshfield, WI, USA: http://research.marshfieldclinic.org/genetics) was used with the NCBI Map Viewer web-site to evaluate linkage convergence.

CFG analysis scoring. Genes were given the maximum score of 2 points if changed in our human blood samples with high threshold analysis, and only 1 point if changed with low threshold (see Figure 1). They received 1 point for each external cross-validating line of evidence: other human tissue data, human genetic data (1 point for assoc., 0.5 point for linkage), animal model brain data, and animal model blood data. Genes received additional bonus points if changed in other human tissue and our blood data, as follows: for brain-2 points if changed in the same direction, 1 point if changed in opposite direction; for lymphoblastoid cell lines and fibroblasts, 1 point if changed in same direction, 0.5 point if changed in opposite directions. Genes also received additional bonus points if changed in brain and blood of the animal model, as follows: 1 point if changed in the same direction in the brain and blood, and 0.5 points if changed in opposite direction. Thus the total maximum CFG score that a candidate biomarker gene can have is 9 $(2+4+2+1)$. As we are interested in discovering blood biomarkers, and because of caveats discussed above, we weighted more heavily our human own live subject human blood data (if it made the high threshold cut) than literature-derived human postmortem brain data, human genetic data, or our own animal model data. We also weighted more heavily the human blood–brain concordance than the animal model blood–brain concordance. Other ways of weighing the scores of line of evidence may give slightly different results in terms of prioritization, if not in terms of the list of genes *per se*. Nevertheless, we feel that this empirical scoring system provides a good separation of genes based on our focus on identifying human blood candidate biomarkers.

Pathway analysis. Ingenuity Pathway Analysis 7.0 (Ingenuity Systems, Redwood City, CA, USA) was used to analyze the biological roles (molecular and cellular functions) categories of the top candidate genes resulting from our CFG analysis.

Results

Hallucinations biomarkers

Using our approach, out of over 40 000 genes and ESTs on the Affymetrix Human Genome U133 Plus 2.0 GeneChip, we have ended up with 50 candidate biomarker genes (Supplementary Table S1) which had a CFG score of 2 or above, meaning either maximal score from the A/P analysis *or* at least one other line of prior independent evidence for potential involvement in psychotic disorders. Of interest, one of our candidate biomarker genes (*Phlda1*[6]) had been previously reported to be changed in expression in the same direction, in lymphoblastoid cell lines from schizophrenia subjects. Another one, Adrbk2 (adrenergic receptor kinase, beta 2), also known as Grk3, has been previously reported by us to be decreased at a protein level in lymphoblastoid cell lines from bipolar patients.[1]

Selecting the top CFG scoring candidate biomarkers for hallucinations (CFG score of 3 and above, meaning, for example, a maximal score from the A/P analysis *and* at least one other line of prior independent evidence for potential involvement in psychotic disorders), we generated a panel of seven biomarkers for hallucinations (Table 4). To test the predictive value of our panel (to be called the BioM-7 hallucinations panel), we have looked in the cohort of 31 psychotic disorders subjects, containing the 23 subjects (12 no hallucinations, 11 high hallucinations) from which the candidate biomarker data was derived, as well as 8 additional subjects with hallucinations symptoms in the intermediate range (PANSS hallucination scores of 2 or 3). We derived a prediction score for each subject, based on the presence or absence of the biomarkers of the panel in their blood GeneChip data. Each of the biomarkers gets a score of 1 if it is detected as Present (P) in the

Table 4 Top candidate biomarker genes for hallucinations prioritized by CFG score for multiple independent lines of evidence

Affymetrix Probe Set ID/Entrez ID	*Gene symbol/name*	*Human blood hallucinations*	*Human tissue evidence (post-mortem brain, lymphocytes and fibroblasts)*	*Human tissue concordance/ co-directionality*	*Human genetic linkage/ association*	*Pharmacogenomic mouse model brain*[44]	*Pharmacogenomic mouse model blood*	*CFG score*
1558199_at/2335	Fn1, fibronectin 1	D (HT)	D (SZ fibroblasts)[25,26]	Yes/yes	2q35 SZ[45]	VT Cat-II (decreased)		5.5
216048_s_at/22836	Rhobtb3, Rho-related BTB domain containing 3	D	D (SZ suicide brain)[46]	Yes/yes	5q15	VT Cat-III (CLZ is increased)		5.0
205208_at/10840	Aldh1l1, aldehyde dehydrogenase 1 family, member L1	D	D (SZ suicide brain)[46]	Yes/yes	3q21.2			4.0
203264_s_at/23229	Arhgef9, Cdc42 guanine nucleotide exchange factor (GEF) 9	I	D (SZ brain)[47]	Yes/no	Xq11.2	PFC Cat-I (decreased) AMY Cat-III (CLZ is increased) NAC Cat- I (decreased)		4.0
206186_at/4356	Mpp3, membrane protein, palmitoylated 3 (MAGUK p55 subfamily member 3)	D	I (SZ lymphocytes)[48]	Yes/no	17q12-q21	VT Cat-III (decreased) HIP Cat- I (decreased)		3.5
225842_at/22822	Phlda1, pleckstrin homology-like domain, family A, member 1	I	I (SZ leukocytes)[6]	Yes/yes	12q21.2			3.0
228923_at/6277	S100a6, S100 calcium binding protein A6 (calcyclin)	I	D (SZ lymphocytes)[49]	Yes/no	1q21.3 SZ[50]			3.0

Abbreviations for human blood data: D, decreased in high hallucination states/increased in no hallucination states; HT, high threshold; I, increased in high hallucination states.
Abbreviations for post-mortem brain data: AMY, amygdala; CLZ, clozapine; CP, caudate putamen; D, decreased; HIP, hippocampus; I, increased; NAC, nucleus accumbens; PCP, phencyclidine; PFC, prefrontal cortex; SZ, schizophrenia; SZA, schizoaffective; VT, ventral tegmentum. Roman numerals in the multiple brain region data column represent the category of the gene.
Top seven candidate biomarker genes for hallucinations, with a CFG score of 3 and above, out of 50 with a CFG score of 2 and above (see Supplementary Data, Table 3S).

blood from that subject, 0.5 if it is detected as Marginally Present (M), and 0 if it is called Absent (A). The ratio of the average of the high hallucinations biomarker scores divided by the average of the no hallucinations biomarker scores is multiplied by 100, and provides a prediction score. If the ratio of high hallucinations biomarkers to no hallucinations biomarkers is 1, that is, the two sets of genes are equally represented, the prediction score is $1 \times 100 = 100$. The higher this score, the higher the predicted likelihood that the subject will have high hallucinations. We then compared the predictive score with actual PANSS hallucination scores. A prediction score of above 100 had an 80% sensitivity and a 70% specificity for predicting high hallucinations (Table 6).

Additionally, we have also conducted human blood gene expression analysis in three other cohorts, subsequently collected. Cohort 2 consisted of 17 subjects from the first cohort that had a change in psychotic symptom (hallucinations and/or delusions) scores at follow-up testing 3 months (v2) or 6 months (v3) later. Cohort 3 consisted of 10 new subjects with psychotic disorders, and Cohort 4 consisted of 9 subjects from Cohort 3 that had a change in symptom scores at follow-up testing 3 months (v2) later.

These cohorts were used as replication cohorts, to verify the predictive power of the hallucinations state biomarker panel identified by analysis of data from the primary psychosis cohort. Overall, the BioM-7 panel had good sensitivity and negative predictive value for high hallucinations state across the different cohorts (Figure 3 and Table 6). Detecting and not missing patients who have high symptom levels is arguably the critical clinical issue, as well as a potential practical application. As such, the sensitivity of the tests for high symptoms (high hallucinations), as well as its negative predictive value, is the most important parameter in that regard.

Delusions biomarkers

Using our approach, we have identified 107 candidate biomarker genes (Supplementary Table S2) which had a CFG score of 2 or above, meaning either maximal score from the A/P analysis or at least one other line of prior independent evidence for potential involvement in psychotic disorders.

Selecting the top CFG scoring candidate biomarkers for delusions (CFG score of 3 and above), we generated a panel of 31 biomarkers (Table 5). To test the predictive value of our panel (to be called the BioM-31 delusions panel), we have looked in the cohort of 31 psychotic disorders subjects, containing the 23 subjects (9 no delusions, 13 high delusions) from which the candidate biomarker data was derived, as well as 9 additional subjects with delusions symptoms in the intermediate range (PANSS delusions scores of 2 or 3). We derived a prediction score for each subject, based on the presence or absence of the biomarkers of the panel in their blood GeneChip data. As for hallucinations, each of the biomarkers gets a score of 1 if it is detected as Present (P) in the blood form that subject, 0.5 if it is detected as Marginally Present (M), and 0 if it is called Absent (A). The ratio of the average of the high delusions biomarker scores divided by the average of the no delusions biomarker scores is multiplied by 100, and provides a prediction score. If the ratio of high delusions biomarkers to no delusions biomarkers is 1, that is, the two sets of genes are equally represented, the prediction score is $1 \times 100 = 100$. The higher this score, the higher the predicted likelihood that the subject will have high delusions. We then compared the predictive score with actual PANSS delusions scores. A prediction score of above 100 had a 92.3% sensitivity and a 61.1% specificity for predicting high delusions (Figure 4 and Table 6).

Additionally, we also tested our BioM-31 delusions panel in the three other cohorts subsequently collected, used as replication cohorts, to verify the predictive power of the delusions state biomarker panel identified by analysis of data from the primary psychosis cohort. Overall, the BioM-31 panel had good sensitivity and negative predictive value for high delusions state, with the exception of one of the cohorts—Cohort 2 (Table 6). It may be that delusions are more private, diverse and ambiguous to assess by PANSS than hallucinations. If not asked specifically about a particular delusion, a subject may not endorse it. As some of our PANSS testing was done by testers who were not familiar clinically with the subject (that is, different testers had performed the Diagnostic Interview for Genetic Studies in those subjects), that could potentially have contributed to false negatives on the PANSS scoring for delusions, and as a consequence resulted in the apparent lower sensitivity of our test in Cohort 2. Regardless if that was the case or not, the reluctance of patients to report psychiatric symptoms underscores the necessity of developing objective tests such as the blood biomarker ones described in this paper, and the need to validate them in multiple cohorts.

Discussion

Strengths and limitations of our work

As a way of identifying biomarkers, we initially conducted gene expression profiling studies in peripheral whole blood from a primary cohort of 31 human subjects with psychotic disorders (30 males, 1 female) (see Table 1). We measured their psychological testing (PANSS) assessed hallucinations scores, respectively delusions scores (on a scale of 1 to 7) at the time of blood collection. We then looked at gene expression differences between the no symptoms of hallucinations, respectively delusions vs high symptoms of hallucinations, respectively delusions, groups. As in our previous work to identify mood biomarkers,[5] we have used an all or nothing Absent (A) vs. Present (P) Calls in the Affymetrix MAS software.

a Cohort 1: Primary pyschosis cohort (n=31)

Patient ID	Diagnosis	P3 Hallucinations Score (1-7)	No Hallucinations Biomarkers				High Hallucinations			BioM-7 Hallucination Prediction Score
			Aldh1l1	Fn1	Mpp3	Rhobtb3	Arhgef9	Phlda1	S100a6	
phchp005v1	SZA	1	A	M	P	A	A	A	A	0.00
phchp006v1	SZA	1	P	A	A	P	A	A	A	0.00
phchp009v1	SZ	3	A	P	A	A	A	A	A	0.00
phchp010v1	SZA	2	A	P	A	A	A	A	A	0.00
phchp013v1	SZA	3	A	A	P	A	A	A	A	0.00
phchp015v1	SubPD	1	M	P	A	A	A	A	A	0.00
phchp040v1	SZA	1	A	P	M	P	A	A	A	0.00
phchp049v1	SZA	1	M	A	P	M	A	A	A	0.00
phchp061v1	SZ	1	P	P	P	P	P	A	A	33.33
phchp048v1	SZA	1	M	P	P	P	P	A	A	38.10
phchp004v1	SZA	1	P	P	A	P	P	A	A	44.44
phchp057v1	SZA	1	P	P	P	A	A	A	P	44.44
phchp062v1	SZ	4	P	P	A	A	P	A	A	66.67
phchp038v1	SZA	1	P	P	M	P	A	P	P	76.19
phchp046v1	SZA	1	A	P	P	P	P	A	P	88.89
phchp025v1	SZ	5	A	A	P	P	A	M	P	100.00
phchp047v1	SZA	5	P	M	P	P	P	P	P	114.29
phchp021v1	SZA	5	A	A	P	P	P	A	P	133.33
phchp022v1	SZ	1	A	A	P	A	A	P	A	133.33
phchp014v1	SubPD	3	A	P	A	P	P	P	P	200.00
phchp024v1	SZA	4	P	A	P	A	P	P	P	200.00
phchp042v1	SZA	2	A	P	A	A	M	A	P	200.00
phchp003v1	SZ	3	M	A	A	P	M	P	P	222.22
phchp026v1	SZA	4	A	A	A	P	P	P	P	400.00
phchp033v1	SZA	5	M	A	A	A	P	A	M	400.00
phchp008v1	SZ	4	A	A	A	A	P	M	A	Infinity
phchp012v1	SZA	3	A	A	A	A	A	P	P	Infinity
phchp018v1	SZA	4	A	A	A	A	P	P	P	Infinity
phchp019v1	SubPD	2	A	A	A	A	A	M	P	Infinity
phchp041v1	SZ	5	A	A	A	A	P	P	P	Infinity
phchp016v1	SZ	5	A	A	A	A	A	A	A	ND

b Cohort 2: Primary pscyhosis cohort follow-up visit (n= 17)

Diagnosis	Gene Symbol	P3 Hallucinations Score (1-7)	No Hallucinations Biomarkers				High Hallucinations			BioM-7 Hallucination Prediction Score
			Aldh1l1	Fn1	Mpp3	Rhobtb3	Arhgef9	Phlda1	S100a6	
phchp038v3	SZA	1	A	P	P	M	A	A	A	0.00
phchp040v2	SZA	2	A	P	A	P	A	A	A	0.00
phchp042v2	SZA	3	P	A	A	A	A	A	A	0.00
phchp048v2	SZA	1	A	A	P	A	A	A	A	0.00
phchp062v2	SZ	3	P	A	A	A	A	A	A	0.00
phchp026v3	SZA	1	P	P	P	P	P	A	A	33.33
phchp006v2	SZA	1	A	P	P	P	A	A	P	44.44
phchp022v2	SZ	1	P	A	P	P	P	A	A	44.44
phchp046v2	SZA	3	A	A	P	A	A	M	A	66.67
phchp005v2	SZA	2	M	P	A	A	A	A	P	88.89
Phchp013v3	SZA	5	P	P	A	A	A	P	P	133.33
phchp021v3	SZA	5	A	A	P	P	P	P	A	133.33
phchp047v2	SZA	5	A	A	A	P	A	P	A	133.33
phchp010v3	SZA	1	P	A	A	A	A	M	P	200.00
phchp016v3	SZ	4	A	A	A	P	A	P	P	266.67
phchp012v2	SZA	5	A	A	A	A	P	P	P	Infinity
phchp003v2	SZ	3	A	A	A	A	A	A	A	ND

c Cohort 3: Second psychosis cohort (n= 10)

Diagnosis	Gene Symbol	P3 Hallucinations Score (1-7)	No Hallucinations Biomarkers				High Hallucinations			BioM-7 Hallucination Prediction Score
			Aldh1l1	Fn1	Mpp3	Rhobtb3	Arhgef9	Phlda1	S100a6	
phchp069v1	SZ	4	A	A	P	A	A	A	A	0.00
phchp017v2	SZA	1	A	A	P	P	P	A	A	66.67
phchp083v1	SZ	1	P	P	A	A	A	P	A	66.67
phchp068v1	SZA	4	A	P	P	M	A	P	P	106.67
phchp072v1	SZA	2	P	A	A	A	A	P	A	133.33
phchp075v1	SZA	4	A	P	A	A	A	P	A	133.33
phchp085v1	SZA	1	P	A	A	A	A	P	A	133.33
phchp058v1	SZ	1	A	A	A	A	A	P	A	Infinity
phchp065v1	SZA	2	A	A	A	A	A	P	P	Infinity
phchp073v1	SZA	5	A	A	A	A	A	A	A	ND

d Cohort 4: Second psychosis cohort follow-up visit (n=9)

Diagnosis	Gene Symbol	P3 Hallucinations Score (1-7)	No Hallucinations Biomarkers				High Hallucinations			BioM-7 Hallucination Prediction Score
			Aldh1l1	Fn1	Mpp3	Rhobtb3	Arhgef9	Phlda1	S100a6	
phchp068v2	SZA	3	P	A	P	A	A	A	A	0.00
phchp072v2	SZA	2	P	A	A	P	A	A	P	66.67
phchp075v2	SZA	5	A	A	M	P	A	A	P	88.89
phchp065v2	SZA	1	P	P	A	A	M	P	P	166.67
phchp058v2	SZ	4	A	A	A	P	A	P	P	266.67
phchp085v2	SZA	1	P	A	A	A	P	A	P	266.67
phchp069v2	SZ	5	A	A	A	A	A	A	P	Infinity
phchp073v2	SZA	4	A	A	A	A	A	P	A	Infinity
phchp083v2	SZ	1	A	A	A	A	A	A	A	ND

Figure 3 Comparison of BioM-7 hallucinations prediction scores and Positive and Negative Symptom Scale (PANSS) hallucinations scores. For hallucinations scores: blue—no hallucinations; red—high hallucinations; white—intermediate hallucinations. Hallucinations scores are based on PANSS scale administered at the time of blood draw. For biomarkers: A (blue)—called Absent by MAS5 analysis; P (red)—called Present by MAS5 analysis; M (yellow)—called Marginally Present by MAS5 analysis. A is scored as 0, M as 0.5 and P as 1. BioM Hallucinations Prediction Score is based on the ratio of the sum of the scores for high mood biomarkers and sum of scores for low mood biomarkers, multiplied by 100. We have used a cutoff score of above 100 for high hallucinations. Infinity—denominator is 0. ND—not determined.

Table 5 Top candidate biomarker genes for delusions prioritized by CFG score for multiple independent lines of evidence

Affymetrix Probeset ID/ Entrez ID	*Gene symbol/name*	*Human blood delusions*	*Human tissue evidence (post-mortem brain, lymphocytes and fibroblasts)*	*Human tissue concordance/ co-directionality*	*Human genetic linkage/ association*	*Pharmacogenomic mouse model brain*[44]	*Pharmaco-genomic mouse model blood*	*Mouse brain and blood concordance/ co-directionality*	*CFG score*
216938_x_at/1813	Drd2, dopamine receptor 2	D	D (SZ brain)[51–53] I (SZ lymphocytes)[54]	Yes/yes	11q23.2 SZ[15,55,56] (Assoc.)	AMY Cat-III (PCP is increased) PFC Cat-II (decreased)			6.0
201693_s_at/1958	Egr1, early growth response 1	I (HT)	D (SZ brain)[57] I (SZ leukocytes)[6]	Yes/no	5q31.2 SZ[58,59]	HIP Cat-II (increased)			5.5
205336_at/5816	Pvalb, parvalbumin	I	I (SZ brain)[60]	Yes/yes	22q12.3 SZ[61]	AMY Cat-II (increased)			5.5
212884_x_at/348	ApoE, apolipoprotein E	D	D (SZ suicide brain)[46]	Yes/yes	19q13.31 SZ[55,56,62] (Assoc.)				5.0
208241_at/3084	Nrg1, neuregulin 1	I	I (SZ brain)[63] (SZ leucocytes)[64] I (SZ lymphocytes)[6,65,66]	Yes/yes	8p12 SZ[47,55,67–72] (Assoc.)				5.0
1570210_x_at/9522	Scamp1, secretory carrier membrane protein 1	D	D (SZ brain)[73]	Yes/yes	5q14.1	AMY Cat-III (PCP is increased) VT Cat-III (CLZ is decreased)			5.0
211780_x_at/1639	Dctn1, dynactin 1 (p150, glued homolog, Drosophila)	I (HT)	D (SZ brain)[47]	Yes/no	2p13.1 SZ[59,74,75]				4.5
1558199_at/2335	Fn1, fibronectin 1	D	D (SZ fibroblasts)[25,26]	Yes/yes	2q35 SZ[45]	VT Cat- II (decreased)			4.5
242001_at/3417	Idh1, isocitrate dehydrogenase 1 (NADP+), soluble	D	D (SZ brain)[76]	Yes/yes	2q34 SZ[45]				4.5
208047_s_at/4664	Nab1, NGFI-A binding protein 1 (EGR1 binding protein 1)	D			2q32.2 SZ[77]	VT Cat-III (CLZ is increased)	Cat- III (CLZ- Increased)	Yes	4.5
205732_s_at/10499	Ncoa2, Nuclear receptor coactivator 2	D			8q13.3 SZ[61]	VT Cat-III (CLZ is increased)	Cat- III (PCP- Increased)	Yes	4.5
201159_s_at/4836	Nmt1, N-myristoyltransferase 1	I	I (SZ brain)[47]	Yes/yes	17q21.31 SZ[78]				4.5
221496_s_at/10766	Tob2, transducer of ERBB2, 2	I (HT)	I (SZ leukocytes)[6]	Yes/yes	22q13.2 SZ[61]				4.5
205208_at/10840	Aldh1l1, aldehyde dehydrogenase 1 family, member L1	D	D (SZ suicide Brain)[46]	Yes/yes	3q21.2				4.0
209168_at/2824	Gpm6b, Glycoprotein M6B	D	I (SZ brain)[79] D (SZ leukocytes)[6]	Yes/no	Xp22.2	AMY Cat-III (CLZ is increased)			4.0
1557704_a_at/64844	March7, membrane-associated ring finger (C3HC4) 7	D	I (SZ brain)[80]	Yes/no	2q24.2	VT Cat IV (PCP is decreased)			4.0
225790_at/253827	Msrb3, methionine sulfoxide reductase B3	I	I (SZ brain)[80]	Yes/yes	12q14.3				4.0
208823_s_at/5127	Pctk1, PCTAIRE-motif protein kinase 1	I	D(SZ brain)[47]	Yes/no	Xp11.3	VT Cat-III (CLZ is increased)			4.0
204519_s_at/51090	Pllp, plasma membrane proteolipid (plasmolipin)	I	D(SZ brain)[81,82]	Yes/no	16q13	AMY Cat-III (PCP is increased)			4.0
40273_at/56848	Sphk2, sphingosine kinase 2	I	I (SZ brain)[80]	Yes/yes	19q13.33				4.0

Table 5 Continued

Affymetrix Probeset ID/ Entrez ID	*Gene symbol/name*	*Human blood delusions*	*Human tissue evidence (post-mortem brain, lymphocytes and fibroblasts)*	*Human tissue concordance/ co-directionality*	*Human genetic linkage/ association*	*Pharmacogenomic mouse model brain*[44]	*Pharmaco-genomic mouse model blood*	*Mouse brain and blood concordance/ co-directionality*	*CFG score*
220995_at/29091	Stxbp6 syntaxin binding protein 6 (amisyn)	D			14q12 SZ[74]	NAC Cat-III (PCP is decreased) VT Cat-III (CLZ is increased)	Cat-III (PCP- Increased)	No	4.0
212385_at/6925	Tcf4, transcription factor 4	D	I (SZ brain)[80]	Yes/no	18q21.2	NAC Cat III (PCP is increased)			4.0
1558733_at/253461	Zbtb38, zinc-finger and BTB domain containing 38	I	I (SZ brain)[80]	Yes/yes	3q23				4.0
235868_at/10724	Mgea5, Meningioma expressed antigen 5 (hyaluronidase)	D (HT)			10q24.32 SZ[83]	VT Cat-III (CLZ is decreased)			3.5
209729_at/10634	Gas2l1, growth arrest-specific 2 like 1	I (HT)			22q12.2		Cat- I (Increased)		3.0
222644_s_at/79709	Glt25d1, glycosyltransferase 25 domain containing 1	I	D (SZ brain)[80]	Yes/no	19p13.11				3.0
209470_s_at/2823	Gpm6a, glycoprotein m6a	D (HT)			4q34.2 SZ[75,84] (Assoc.)				3.0
239044_at/81533	Itfg1, integrin alpha FG-GAP repeat containing 1	D	I (SZ brain)[80]	Yes/no	16q12.1				3.0
236407_at/3753	Kcne1, potassium voltage-gated channel, Isk-related family, member 1	I	D (SZ brain)[85]	Yes/no	21q22.12				3.0
203329_at/5797	Ptprm, protein tyrosine phosphatase, receptor type, M	I			18p11.23 SZ[86] (Assoc.)	VT Cat-III (CLZ is increased)			3.0
233666_at/54664	Tmem106b, transmembrane protein 106B	D	I (SZ brain)[47]	Yes/no	7p21.3				3.0

Abbreviations for human blood data: D, decreased in high delusion states/increased in no delusion states; HT, high threshold; I, increased in high delusion states.
Abbreviations for post-mortem brain data: AMY, amygdala; CLZ, clozapine; CP, caudate putamen; Down, decreased; HIP, hippocampus; NAC, nucleus accumbens; PCP, phencyclidine; PFC, prefrontal cortex; SZ, schizophrenia; SZA, schizoaffective; Up, increased; VT, ventral tegmentum; roman numerals in the multiple brain region data column represent the category of the gene.
Top 31 candidate biomarker genes for delusions, with a CFG score of 3 and above, out of 107 genes with a CFG score of 2 and above (see Supplementary Data, Table 3S).

Table 6 Psychosis biomarkers panels: sensitivity for predicting high hallucination and high delusion states

	Cohort 1, primary cohort (from which biomarkers were derived) N = *31*	*Cohort 2, primary cohort follow-up visit* N = *17*	*Cohort 3, second cohort (independent)* N = *10*	*Cohort 4, second cohort follow-up visit*, N = *9*
BioM-7 hallucinations				
Sensitivity	**80.00%**	**100.00%**	**66.60%**	**75.00%**
Specificity	70.00%	90.90%	33.30%	50.00%
Negative predictive value	**87.50%**	**100.00%**	**66.67%**	**66.67%**
Positive predictive value	57.14%	83.33%	33.33%	60.00%
BioM-31 delusions				
Sensitivity	**92.30%**	**42.80%**	**100.00%**	**100.00%**
Specificity	61.10%	60.00%	16.67%	16.67%
Negative predictive value	**91.70%**	**60.00%**	**100.00%**	**100.00%**
Positive predictive value	63.15%	42.85%	44.44%	37.50%

BioM-7 hallucinations is a seven biomarker panel for predicting hallucinations. BioM-31 delusions is a 31 biomarker panel for predicting delusions. Detecting and not missing patients who have high symptom levels is arguably the critical clinical issue, as well as potential practical application. As such, we have bolded in the table the sensitivity of the tests for high symptoms (high hallucinations, high delusions), as well as the negative predictive value, the most important parameters in that regard.

Given the genetic heterogeneity and variable environmental exposure, it is possible, indeed likely, that not all subjects will show changes in all the biomarker genes. Hence we have used two stringency thresholds: changes in 75% of subjects, and in 60% of subjects with no symptoms vs high symptoms. Moreover, our approach, as described above, is predicated on the existence of multiple cross-validators for each gene that is called a candidate biomarker (Figure 1): (1) is it changed in human blood, (2) is it changed in animal model brain, (3) is it changed in animal model blood, (4) is it changed in postmortem human brain, and (5) is there any human genetic data (linkage, association) implicating the gene in psychosis. All these lines of evidence are the result of independent experiments. The virtues of this networked Bayesian approach are that, if one or another of the nodes (lines of evidence) becomes questionable/non-functional upon further evidence in the field, the network is resilient and maintains functionality. The prioritization of candidates is similar conceptually to the Google PageRank algorithm[7]—the more links (lines of evidence) to a candidate, the higher it will come up on our priority list. As more evidence emerges in the field for some of these genes, they will move up in the prioritization scoring.[8] Using such an approach, we were able to identify and focus on a small number of genes as likely candidate biomarkers, out of the over 40 000 transcripts (about half of which are detected as Present in each subject) measured by the microarrays we used.

By cross-validating with other human datasets and with animal model data using CFG (Figure 1), we were able to extract a shorter list of genes for which there are external corroborating line of evidence (human genetic evidence, human postmortem brain data, animal model brain and blood data) linking them to psychotic disorders, thus reducing the risk of false positives. This cross-validation identifies candidate biomarkers that are more likely directly related to the relevant disease neuropathology, as opposed to being potential artifactual effects related to a particular cohort or indirect effects of lifestyle and environment. The power of our CFG approach is exemplified in the fact that our biomarker panels had good predictive power in independent cohorts, a key litmus test in our view, and one that needs to be applied more systematically in this nascent field.

All subjects recruited will be on priorly prescribed medications. We cannot exclude, and in fact would anticipate that medications may have an effect on biomarker expression levels. However, of note, the patients were on a very diverse list of antipsychotics, mood stabilizers, and other psychotropic medications (Supplementary Table S4).

While that makes pinpointing a particular medication effect not feasible with our current design (clinical trials with specific medications are a better setting for identifying such effects), it is re-assuring that we are obtaining with our CFG approach consistent findings that show predictive power in independent cohorts, despite this diversity of medications and of a variety of other environmental effects.

Clozapine, modeled in the pharmacogenomic animal model work, is a broad-spectrum drug, one of the current gold standards, and encompasses many of the actions of some of the other antipsychotics currently

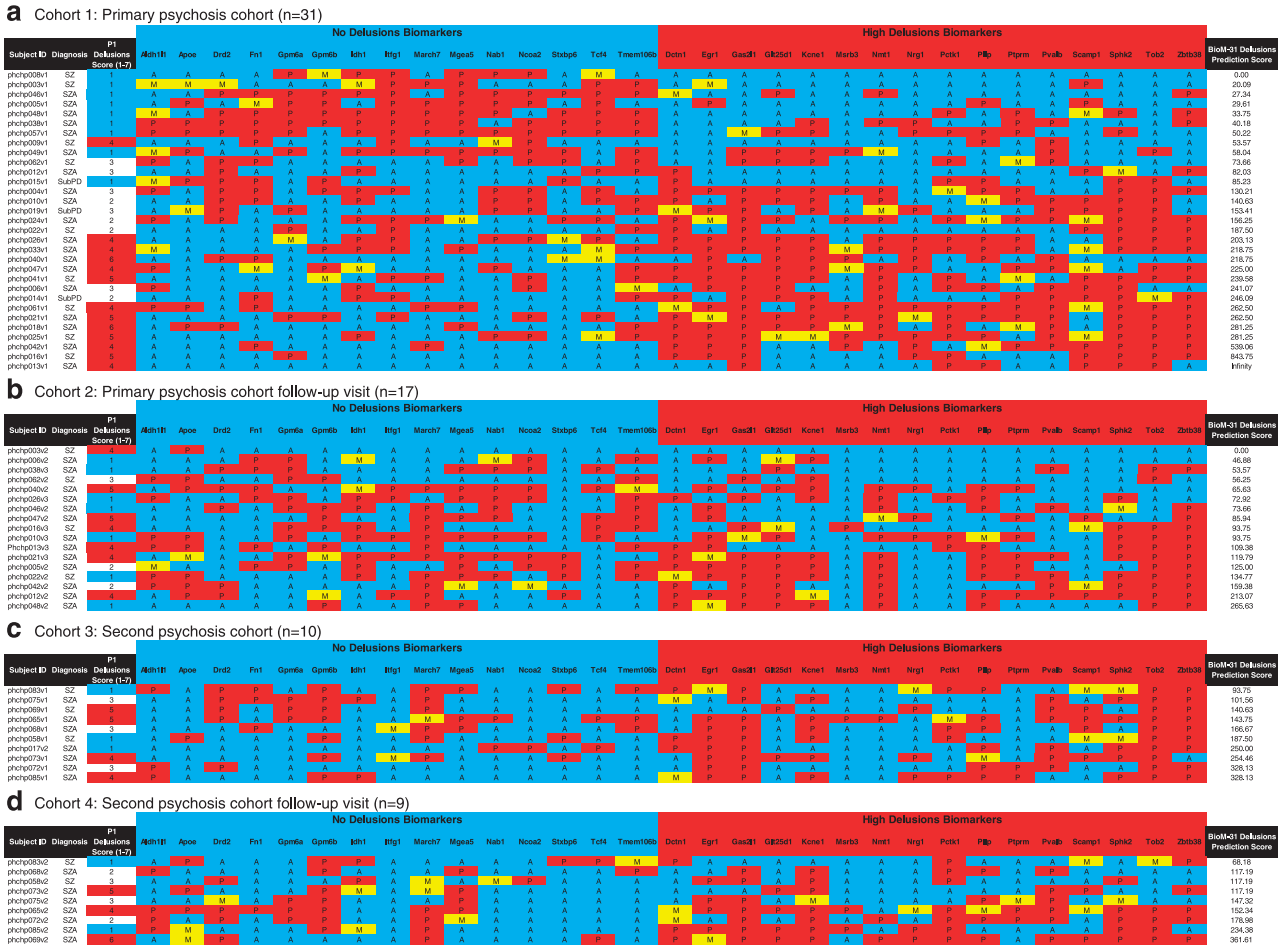

Figure 4 Comparison of BioM-31 delusions prediction scores and actual Positive and Negative Symptom Scale (PANSS) delusions scores. For delusion scores: blue—no delusion; red—high delusion; white—intermediate delusion. Delusion scores are based on PANSS scale administered at the time of blood draw. For biomarkers: A (blue)—called Absent by MAS5 analysis. P (red)—called Present by MAS5 analysis. M (yellow)—called Marginally Present by MAS5 analysis. A is scored as 0, M as 0.5 and P as 1. BioM delusions Prediction Score is based on the ratio of the sum of the scores for high mood biomarkers and sum of scores for low mood biomarkers, multiplied by 100. We have used a cutoff score of above 100 for high delusion. Infinity-denominator is 0. ND—not determined.

used in schizophrenia. The premise of using it, along with PCP, in a pharmacogenomic animal model of schizophrenia,[4] was that they may modulate the expression of genes involved in the pathogenesis of schizophrenia. The findings in that model, cross-validated with other independent approaches and lines of evidence, support its validity.[4] Comparisons with the non-medicated normal control group will in the future permit additional distinctions regarding medication effects, as will systematic large-scale within-subject comparisons of subjects whose medications remain constant but symptoms state and markers change from one visit to the next.

Moreover, psychosis state and blood gene expression changes may be influenced not only by the presence or absence of medications, but also of drugs of abuse. While we had access to the subject's medical records and clinical information as part of the informed consent for the study, medication non-compliance, on the one hand, and substance abuse, on the other hand, are not infrequent occurrences in psychiatric patients.

More extensive follow-up studies may benefit from the prospective inclusion of toxicology and medication levels testing. That medications and drugs of abuse may have effects on psychosis state and gene expression is in fact being partially modeled in the mouse pharmacogenomic model, with clozapine and PCP treatments respectively. In the end, it is the association of blood biomarkers with psychosis state that has been the primary goal of the work reported in this paper, regardless of the proximal causes, which could be diverse and will need to be the subject of subsequent hypothesis-driven studies beyond the scope of this initial work.

Our sample size for human subjects ($n = 31$ for the primary cohort; $n = 17$, $n = 10$, $n = 9$ for the other three cohorts) is relatively small, but comparable to the size of cohorts for human postmortem brain gene expression studies.[9,10,11] We have in essence studied live

donor blood samples instead of postmortem donor brains, with the advantage of better phenotypic characterization, more quantitative state information, and less technical variability. Our approach also permits repeated intra-subject measures when the subject is in different psychosis states, which is an area of future interest and work. In fact, two of our psychosis cohorts are composed of a subset of subjects from the primary and secondary psychosis cohorts, that displayed a different psychosis state (no symptoms vs. intermediate vs. high symptoms) when they were re-tested at a later time point, 3 or 6 months later. Overall, our design was geared towards validating state biomarkers for psychosis while minimizing the noise of genetic and environmental background differences. For trait biomarkers, larger population studies and comparisons with normal controls may be needed. Of note, we have studied almost exclusively male subjects, which means our results might be male-specific. Future studies looking at potential gender differences are warranted.

Overall, our approach of: (1) using individual phenes[12] reflecting internal subjective experiences (hallucinations or delusions), which are the hallmark of psychosis (as opposed to more complex and disease specific state/trait clinical instruments or DSM categorical diagnosis); (2) looking at extremes of state; combined with (3) robust differential expression based on A/P calls, and (4) CFG prioritization, seems to be able to identify state biomarkers for psychosis that may be, at least in part, generalizable to independent cohorts.

In the work reported here, similar to our previously published mood biomarker work,[5] we decided to focus on using CFG scoring as a cut-off to decide which biomarkers to include in panels, rather than find best panel size by fit-to-data and receiver operating characteristic curve. We reasoned that an objective CFG scoring cut-off would pick up signal relevant to illness and increase generalizability of our panels across independent cohorts, while a fit-to-data receiver operating characteristic approach, while it might achieve excellent results in the primary cohort, driven at least in part by the noise particular to that cohort, would have poorer results in independent cohorts. In fact, CFG prioritization has been shown to lead to generalizabilty across cohorts not only in our previous[5] and current biomarker work, but also when we applied it to genome-wide association studies data,[13] where *P*-value criteria are the equivalent of fit-to-data analyses.

While it appears that panels of biomarkers chosen by CFG scoring criteria are the way to go due to population heterogeneity and impact of environmental factors on gene expression, it remains an open empirical question for future work as to how large the panels should be, and whether it may be possible to identify particular single biomarkers that have almost as good a predictive power as that of a larger panel. Ongoing studies are also examining the issue of using incremental differential expression comparisons as opposed to all or nothing A/P calls to identify biomarkers, and are expected to yield an expanded repertoire of biomarkers.

Finally, some of the top candidate biomarker genes identified by our human blood work reported here have no previous evidence for involvement in psychotic disorders other than our mapping them to schizophrenia genetic linkage loci (Supplementary Tables S1 and S2), and thus constitute novel candidate genes for schizophrenia and related disorders. They merit further exploration in genetic candidate gene association studies, as well as comparison with emerging results from whole-genome association studies of schizophrenia and related disorders. Moreover, as more evidence accumulates in the field, all grist for the mill for our CFG approach, and as prospective studies are done, it is possible that the composition of top biomarker panels for hallucinations and for delusions will be refined or changed for different sub-populations. That being said, it is likely that a large number of the biomarkers that would be of use in different panels and permutations are already present in our lists of candidate biomarker genes ($n = 50$ for hallucinations—Supplementary Table S1; $n = 107$ for delusions—Supplementary Table S2).

Hallucinations and delusions: similarities and differences

There are more genes with high CFG scores for delusions than for hallucinations, reflecting the fact that more prior evidence exists for them in terms of involvement in schizophrenia and related disorders, and perhaps there is a higher degree of diversity in the genetic architecture of delusions, a more evolved cognitive phene, compared to that of hallucinations, a more primitive sensory phene. As a consequence, using the same CFG cut-off, the panel size for delusions was larger than that for hallucinations. Of note, there is co-directional overlap between the candidate biomarkers for delusions (Supplementary Table S2) and hallucinations (Supplementary Table S1) identified by us, which is reassuring in terms of the technical reliability of our assessments and findings, as these symptoms are often co-morbid clinically. More interestingly, there is some overlap between candidate biomarkers for hallucinations, delusions and mood state previously identified by us[5] (Supplementary Figure S1), with the mood markers being generally contra-directional to the psychosis markers. Taken together with the heterogeneity of biomarker expression seen in patients that have a similar psychiatric diagnosis (Figures 3 and 4), our work is consistent with an emerging Lego-like model of complexity, heterogeneity, overlap and interdependence of major psychiatric disorders.[4,14] Practical implications and predictions of this view would be the relative lack of specificity of single genes and biomarkers for a particular disorder, and the need to use profiling with panels of markers to achieve some disease specificity.

Table 7 Biological roles

(A) Top bio functions for hallucination biomarkers
Diseases and disorders

	P-value range	No. of molecules
Cancer	8.54E-05–4.98E-02	20
Hematological disease	2.94E-04–1.69E-02	5
Connective tissue disorders	3.46E-04–4.25E-02	6
Inflammatory response	2.35E-03–4.16E-02	3
Reproductive system disease	2.72E-03–4.71E-02	7

Molecular and cellular functions

	P-value range	No. of molecules
Cellular assembly and organizations	1.34E-05–4.38E-02	16
Cell-to-cell signaling and interaction	1.38E-05–4.44E-02	12
Cellular function and maintenance	2.63E-05–2.80E-02	9
Cell morphology	4.66E-05–4.71E-02	13
Cellular movement	7.39E-05–4.98E-02	8

Physiological system development and function

	P-value range	No. of molecules
Skeletal and muscular system development and function	1.38E-05–4.98E-02	5
Tissue development	1.38E-05–4.98E-02	8
Cardiovascular system development and function	2.92E-04–4.71E-02	5
Connective tissue development and function	9.23E-04–4.71E-02	8
Reproductive system development and function	2.25E-03–3.89E-02	3

Top canonical pathways

	P-value	Ratio
IL-8 signaling	1.31E-03	4/185 (0.022)
Chemokine signaling	1.32E-03	3/77 (0.039)
Thrombin signaling	2.05E-03	4/204 (0.02)
IL-15 production	2.43E-03	2/30 (0.067)
Semaphorin signaling in neurons	9.11E-03	2/52 (0.038)

(B) Top bio functions for delusions biomarkers
Diseases and disorders

	P-value range	No. of molecules
Cancer	3.73E-04–2.22E-02	25
Neurological disease	3.73E-04–1.84E-02	17
Reproductive system disease	5.77E-04–1.84E-02	17
Genetic disorder	1.32E-03–2.45E-02	19
Metabolic disease	1.32E-03–1.84E-02	6

Molecular and cellular functions

	P-value range	No. of molecules
Cell morphology	3.78E-05–1.92E-02	14
Cell-to-cell signaling and interaction	1.42E-04–2.34E-02	19
Cellular movement	2.05E-04–2.30E-02	16
Lipid metabolism	2.25E-04–2.45E-02	9
Small molecule biochemistry	2.25E-04–2.45E-02	21

Physiological system development and function

	P-value range	No. of molecules
Nervous system development and function	1.42E-04–1.84E-02	12
Reproductive system development and function	2.05E-04–1.84E-02	6
Skeletal and muscular system development and function	2.43E-04–2.30E-02	14
Cardiovascular system development and function	5.57E-04–1.84E-02	11
Tissue morphology	5.57E-04–1.84E-02	12

Top canonical pathways

	P-value	Ratio
IL-15 production	5.34E-04	3/30 (0.1)
LPS/IL-1-mediated inhibition of RXR function	3.06E-02	4/198(0.02)
Aryl hydrocarbon receptor signaling	5.87E-02	3/155 (0.019)
Agrin interactions at neuromuscular junction	7.31E-02	2/72 (0.0028)
LXR/RXR activation	7.49E-02	2/85 (0.024)

Abbreviations: IL, interleukin; LPS, lipopolysaccharide; LXR, *liver* X receptor; RXR, retinoid X receptor.
Ingenuity pathway analysis (IPA) of biological functions categories among our blood candidate biomarkers for hallucinations (A) and delusions (B). Genes from Tables 3S ($n = 50$) and 4S ($n = 107$).

From biomarkers to biology

Remarkably, among our candidate blood biomarker genes for delusions (Table 5) are key genes with extensive evidence in brain pathophysiology in psychotic disorders (*dopamine receptor 2—Drd2*,[15] *neuroregulin 1—Nrg1*[16,17]) and neurodegenerative disorders (*apolipoprotein E—ApoE*). A polymorphism in Drd2 was reported to be associated specifically with delusions and disorganization symptomatology in major psychoses.[18] Of interest, delusion symptoms were reported to be associated with ApoE epsilon4 allelic variant in late-onset Alzheimer's disease.[19] Moreover, plasma ApoE has been reported to be significantly decreased in treatment-free subjects with schizophrenia spectrum disorders and bipolar disorder,[20] consistent with our findings of ApoE being decreased in expression in high delusion states. Recently, variations in levels of expression of ApoE have also been tied by us to the risk and progression of Alzheimer's disease (AD) irrespective of ϵ4 status.[21] Overall, the ApoE connection warrants future empirical work as a possible molecular underpinning of the Kraepelinian view of schizophrenia as *dementia praecox*.

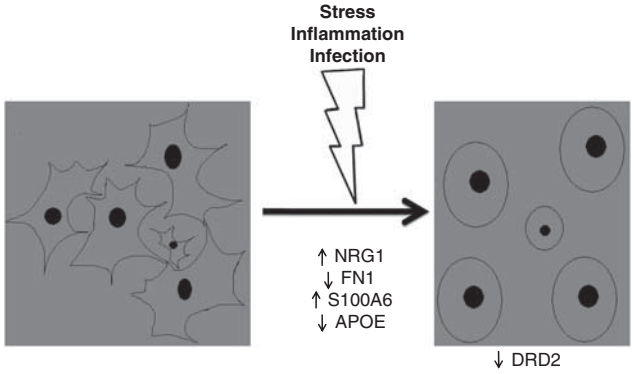

Figure 5 Psychosis: disconnection and de-differentiation.

At the top of our list of candidate biomarker genes for hallucinations (Table 4), we have four genes decreased in expression in high hallucinations states (*Rhobtb3*, *Aldh1l1*, *Mpp3*, *Fn1*), and three genes increased in high hallucinations states (*Arhgef9*, *Phlda1*, *S100a6*). Although all of these genes have prior evidence of differential expression in schizophrenia patients, they are less well known than the candidate biomarker genes for delusions discussed above. A non-obvious top candidate biomarker for hallucinations, increased in high hallucinations state, is Arhgef9 (Cdc42 guanine nucleotide exchange factor 9, also known as collybistin). Arhgef9 can regulate actin cytoskeletal dynamics and may also modulate GABAergic neurotransmission through binding of a scaffolding protein, gephyrin, at the synapse.[22] Interestingly, it has also been implicated in X-linked mental retardation with sensory hyperarousal.[23] Aldh1l1, another non-obvious candidate, is a folate metabolic enzyme with antiproliferative effects, expressed in astrocytes.[24]

Table 8 Connectivity map interrogation of drugs that have similar gene expression signatures to that of (A) high hallucinations and (B) high delusions

Rank	*Instance_id*	*Cmap name*	*Batch*	*Dose*	*Cell line*	*Score*	*Up*	*Down*
(A) Connectivity map detailed result for BioM-7 hallucinations panel genes								
1	5247	Cephaeline	726	6 μM	MCF7	1	0.714	−0.523
2	6817	Verteporfin	744	3 μM	MCF7	0.972	0.824	−0.378
3	5021	Suloctidil	707	12 μM	MCF7	0.949	0.91	−0.264
4	2801	Emetine	663	7 μM	MCF7	0.925	0.639	−0.506
5	3443	Monensin	670	6 μM	MCF7	0.913	0.807	−0.321
6096	7077	Trichostatin A	1073	1 μM	PC3	−0.937	−0.731	0.41
6097	1220	Vorinostat	603	10 μM	PC3	−0.957	−0.623	0.542
6098	7079	MG-262	1073	100 nM	PC3	−0.96	−0.835	0.335
6099	5106	Dropropizine	719	17 μM	PC3	−0.977	−0.621	0.569
6100	7068	MG-262	1069	100 nM	PC3	−1	−0.74	0.478
(B) Connectivity map detailed results for BioM-31 delusions panel genes								
1	4631	Josamycin	712	5 μM	PC3	1	0.33	−0.278
2	4457	Rosiglitazone	727	10 μM	PC3	0.995	0.265	−0.341
3	3258	7-Aminocephalosporanic acid	654	15 μM	MCF7	0.881	0.215	−0.321
4	1328	Pepstatin	631	6 μM	HL60	0.844	0.271	−0.242
5	2520	Tetrandrine	648	6 μM	HL60	0.837	0.318	−0.191
6096	6619	Tracazolate	709	12 μM	PC3	−0.815	−0.207	0.306
6097	5964	Fulvestrant	1012	1 μM	MCF7	−0.825	−0.201	0.318
6098	4527	Rifabutin	703	5 μM	PC3	−0.878	−0.278	0.275
6099	4184	Trichostatin A	692	100 nM	PC3	−0.893	−0.223	0.339
6100	494	**Fluphenazine**	69	10 μM	SKMEL5	−1	−0.251	0.379

A score of 1 indicates a maximal similarity with the gene expression effects of high hallucinations/delusions, and −1 indicates a maximal opposite effect. Bold indicates antipsychotic medication.

Fn1 (Fibronectin 1), one of our top scoring candidate biomarkers for hallucinations and for delusions (Figure 2), is decreased in high hallucinations states and high delusions states, was also previously reported to be decreased in fibroblasts from schizophrenia patients.[25,26] It has also been identified as a top candidate gene for alcoholism in previous work from our group.[3] This raises interesting issues about the psychosis-modulating properties of alcohol, specifically hallucinations and delusions symptoms in alcoholism, as well as the more general issue of clinical co-morbidity between schizophrenia and alcoholism.

Overall, the top candidate biomarker genes results discussed above and the results of a biological functions analyses (Tables 6 and 7) suggest that genes involved in cancer, plasticity and connectivity (cell morphology, cell-to-cell signaling and interaction) are prominent players in psychotic disorders, and are reflected in the blood profile, consistent with previous work in the field implicating developmental and connectivity mechanisms in schizophrenia.[4,27,28] Unlike for our mood biomarker work,[29] we did not find myelin genes prominently represented among our top psychosis biomarkers. Interestingly, the top canonical pathways for both hallucinations and delusions had to do with interleukin signaling, consistent with previous work in the field implicating immune and inflammatory mechanisms in schizophrenia pathophysiology.[30] For example, IL-8 signaling, which was identified as the top canonical pathway in hallucinations, has been previously implicated as a maternal risk factor for schizophrenia in the offsprings,[31] and IL-8 levels have been reported to be elevated in neuroleptic-free schizophrenia patients compared to normal controls.[32]

The model that is emergent is that of increased plasticity and decreased connectivity[4] in high psychosis states compared to no psychosis states. This perspective is of speculative evolutionary interest and pragmatic clinical importance. Speculatively, nature may have selected primitive cellular mechanisms involved in the response to damage, insults and stressors for analogous higher organism level-functions (Figure 5). In this view, psychosis is the higher organismal/brain equivalent of cellular de-differentiation and disconnection such as occurs in early stages of inflammation[33], tissue re-modeling[34] and cancer metastasis.[35] Specifically, the decrease in FN1 expression and increase in NRG1 expression in high delusions states, as well as decrease in fibronectin expression and increase in calcyclin (S100A6) in high hallucination states, are consistent with increased metastatic potential, though not necessarily increased tumorigenesis/cellular proliferation. Indeed, there seems to be a decrease incidence of respiratory cancers in schizophrenia patients, despite the high incidence of smoking in that population. Pragmatically, the psychotic episodes may be correlated with metastasis in cancers.[36] Typical antipsychotic medications may have protective effects against cancer,[37] consistent also with our connectivity map results identifying fluphenazine as having an opposite gene expression profile to that of high delusions (Table 8). Lastly, the involvement of interleukin signaling canonical pathways suggests that anti-inflammatory and immune-modulating medications medications should to be more systematically evaluated for prevention and early intervention in psychotic disorders, consistent with some emerging clinical data.[38,39] In particular, omega-3 fatty acids may have a favorable effects to side-effects ratio and multiple whole-body health benefits in this patient population.[40]

Conclusions

We propose, and provide proof of principle for, a translational convergent approach to help identify and prioritize blood biomarkers for psychosis states, specifically for hallucinations and for delusions. A validation of our approach is the fact that our primary cohort-derived biomarker panels showed not only good sensitivity and specificity in the primary cohort, but also predictive ability in three other cohorts. Finally, a data-derived model for whole-body biological mechanisms associated with psychosis is proposed.

Biomarker-based tests may help with early detection, intervention and prevention efforts in schizophrenia[41,42] and related disorders,[43] as well as monitoring response to various treatments. In conjunction with other clinical information, such tests may come to play an important part in personalizing treatment to increase precision, effectiveness and avoid adverse reactions. Last but not least, new drug development efforts would particularly benefit from the use of such markers.

Conflict of interest

ABN and DRS are founders and hold an equity interest in Mindscape Diagnostics, Inc. MAG holds an equity interest in San Diego Instruments, Inc.

Acknowledgments

This work was supported by funds from INGEN (Indiana Genomics Initiative of Indiana University), INBRAIN (Indiana Center for Biomarker Research In Neuropsychiatry) and NARSAD Young Investigator Award to ABN, as well as NIMH R01 MH071912–01 to MTT and ABN. ABN would like to thank Howard Edenberg for excellent help and advice with animal model microarray data, as well as Sudharani Mamidipalli, Griffin Fitzgerald and Jesse Townes for their precise work with database maintenance and data analysis. Most importantly, we would like to thank the subjects who participated in these studies, their families and their caregivers. Without their generous participation, such work to advance the understanding of mental illness would not be possible.

References

1 Niculescu A, Segal D, Kuczenski R, Barrett T, Hauger R, Kelsoe J. Identifying a series of candidate genes for mania and psychosis: a convergent functional genomics approach. *Physiol Genomics* 2000; **4**: 83–91.

2 Ogden CA, Rich ME, Schork NJ, Paulus MP, Geyer MA, Lohr JB *et al.* Candidate genes, pathways and mechanisms for bipolar (manic-depressive) and related disorders: an expanded convergent functional genomics approach. *Mol Psychiatry* 2004; **9**: 1007–1029.

3 Rodd ZA, Bertsch BA, Strother WN, Le-Niculescu H, Balaraman Y, Hayden E *et al.* Candidate genes, pathways and mechanisms for alcoholism: an expanded convergent functional genomics approach. *Pharmacogenomics J* 2007; **7**: 222–256.

4 Le-Niculescu H, Balaraman Y, Patel S, Tan J, Sidhu K, Jerome RE *et al.* Towards understanding the schizophrenia code: an expanded convergent functional genomics approach. *Am J Med Genet B Neuropsychiatr Genet* 2007; **144B**: 129–158.

5 Le-Niculescu H, Kurian SM, Yehyawi N, Dike C, Patel SD, Edenberg HJ *et al.* Identifying blood biomarkers for mood disorders using convergent functional genomics. *Mol Psychiatry* 2009; **14**: 156–174.

6 Middleton FA, Pato CN, Gentile KL, McGann L, Brown AM, Trauzzi M *et al.* Gene expression analysis of peripheral blood leukocytes from discordant sib-pairs with schizophrenia and bipolar disorder reveals points of convergence between genetic and functional genomic approaches. *Am J Med Genet B Neuropsychiatr Genet* 2005; **136**: 12–25.

7 Morrison JL, Breitling R, Higham DJ, Gilbert DR. GeneRank: using search engine technology for the analysis of microarray experiments. *BMC Bioinformatics* 2005; **6**: 233.

8 Le-Niculescu H, McFarland MJ, Mamidipalli S, Ogden CA, Kuczenski R, Kurian SM *et al.* Convergent functional genomics of bipolar disorder: from animal model pharmacogenomics to human genetics and biomarkers. *Neurosci Biobehav Rev* 2007; **31**: 897–903.

9 Vawter MP, Crook JM, Hyde TM, Kleinman JE, Weinberger DR, Becker KG *et al.* Microarray analysis of gene expression in the prefrontal cortex in schizophrenia: a preliminary study. *Schizophr Res* 2002; **58**: 11–20.

10 Choudary PV, Molnar M, Evans SJ, Tomita H, Li JZ, Vawter MP *et al.* Altered cortical glutamatergic and GABAergic signal transmission with glial involvement in depression. *Proc Natl Acad Sci USA* 2005; **102**: 15653–15658.

11 Vawter MP, Tomita H, Meng F, Bolstad B, Li J, Evans S *et al.* Mitochondrial-related gene expression changes are sensitive to agonal-pH state: implications for brain disorders. *Mol Psychiatry* 2006; **11**, 615, 663–679.

12 Niculescu AB, Lulow LL, Ogden CA, Le-Niculescu H, Salomon DR, Schork NJ *et al.* PhenoChipping of psychotic disorders: a novel approach for deconstructing and quantitating psychiatric phenotypes. *Am J Med Genet B Neuropsychiatr Genet* 2006; **141**: 653–662.

13 Le-Niculescu H, Patel SD, Bhat M, Kuczenski R, Faraone SV, Tsuang MT *et al.* Convergent functional genomics of genome-wide association data for bipolar disorder: comprehensive identification of candidate genes, pathways and mechanisms. *Am J Med Genet B Neuropsychiatr Genet* 2009; **150B**: 155–181.

14 Niculescu 3rd AB. Polypharmacy in oligopopulations: what psychiatric genetics can teach biological psychiatry. *Psychiatr Genet* 2006; **16**: 241–244.

15 Glatt SJ, Faraone SV, Lasky-Su JA, Kanazawa T, Hwu HG, Tsuang MT. Family-based association testing strongly implicates DRD2 as a risk gene for schizophrenia in Han Chinese from Taiwan. *Mol Psychiatry* 2008; **14**: 885–893.

16 Georgieva L, Dimitrova A, Ivanov D, Nikolov I, Williams NM, Grozeva D *et al.* Support for neuregulin 1 as a susceptibility gene for bipolar disorder and schizophrenia. *Biol Psychiatry* 2008; **64**: 419–427.

17 Goes FS, Willour VL, Zandi PP, Belmonte PL, Mackinnon DF, Mondimore FM *et al.* Family-based association study of Neuregulin 1 with psychotic bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet* 2009; **150B**: 693–702.

18 Serretti A, Lattuada E, Lorenzi C, Lilli R, Smeraldi E. Dopamine receptor D2 Ser/Cys 311 variant is associated with delusion and disorganization symptomatology in major psychoses. *Mol Psychiatry* 2000; **5**: 270–274.

19 Spalletta G, Bernardini S, Bellincampi L, Federici G, Trequattrini A, Caltagirone C. Delusion symptoms are associated with ApoE epsilon4 allelic variant at the early stage of Alzheimer's disease with late onset. *Eur J Neurol* 2006; **13**: 176–182.

20 Dean B, Digney A, Sundram S, Thomas E, Scarr E. Plasma apolipoprotein E is decreased in schizophrenia spectrum and bipolar disorder. *Psychiatry Res* 2008; **158**: 75–78.

21 Maloney B, Ge YW, Petersen RC, Hardy J, Rogers JT, Perez-Tur J *et al.* Functional characterization of three single-nucleotide polymorphisms present in the human APOE promoter sequence: differential effects in neuronal cells and on DNA-protein interactions. *Am J Med Genet B Neuropsychiatr Genet*, advance online publication, 5 June 2009; e-pub ahead of print.

22 Papadopoulos T, Korte M, Eulenburg V, Kubota H, Retiounskaia M, Harvey RJ *et al.* Impaired GABAergic transmission and altered hippocampal synaptic plasticity in collybistin-deficient mice. *Embo J* 2007; **26**: 3888–3899.

23 Marco EJ, Abidi FE, Bristow J, Dean WB, Cotter P, Jeremy RJ *et al.* ARHGEF9 disruption in a female patient is associated with X linked mental retardation and sensory hyperarousal. *J Med Genet* 2008; **45**: 100–105.

24 Anthony TE, Heintz N. The folate metabolic enzyme ALDH1L1 is restricted to the midline of the early CNS, suggesting a role in human neural tube defects. *J Comp Neurol* 2007; **500**: 368–383.

25 Mahadik SP, Mukherjee S, Wakade CG, Laev H, Reddy RR, Schnur DB. Decreased adhesiveness and altered cellular distribution of fibronectin in fibroblasts from schizophrenic patients. *Psychiatry Res* 1994; **53**: 87–97.

26 Miyamae Y, Nakamura Y, Kashiwagi Y, Tanaka T, Kudo T, Takeda M. Altered adhesion efficiency and fibronectin content in fibroblasts from schizophrenic patients. *Psychiatry Clin Neurosci* 1998; **52**: 345–352.

27 Bassett DS, Bullmore E, Verchinski BA, Mattay VS, Weinberger DR, Meyer-Lindenberg A. Hierarchical organization of human cortical networks in health and schizophrenia. *J Neurosci* 2008; **28**: 9239–9248.

28 Sun D, Stuart GW, Jenkinson M, Wood SJ, McGorry PD, Velakoulis D *et al.* Brain surface contraction mapped in first-episode schizophrenia: a longitudinal magnetic resonance imaging study. *Mol Psychiatry* 2008; **14**: 976–986.

29 Le-Niculescu H, Kurian SM, Yehyawi N, Dike C, Patel SD, Edenberg HJ *et al.* Identifying blood biomarkers for mood disorders using convergent functional genomics. *Mol Psychiatry* 2008; **14**: 156–174.

30 Shirts BH, Wood J, Yolken RH, Nimgaonkar VL. Comprehensive evaluation of positional candidates in the IL-18 pathway reveals suggestive associations with schizophrenia and herpes virus seropositivity. *Am J Med Genet B Neuropsychiatr Genet* 2008; **147**: 343–350.

31 Brown AS, Hooton J, Schaefer CA, Zhang H, Petkova E, Babulas V *et al.* Elevated maternal interleukin-8 levels and risk of schizophrenia in adult offspring. *Am J Psychiatry* 2004; **161**: 889–895.

32 Zhang XY, Zhou DF, Zhang PY, Wu GY, Cao LY, Shen YC. Elevated interleukin-2, interleukin-6 and interleukin-8 serum levels in neuroleptic-free schizophrenia: association with psychopathology. *Schizophrenia Res* 2002; **57**: 247–258.

33 Garcia-Bueno B, Caso JR, Leza JC. Stress as a neuroinflammatory condition in brain: damaging and protective mechanisms. *Neurosci Biobehav Rev* 2008; **32**: 1136–1151.

34 Lau CG, Zukin RS. NMDA receptor trafficking in synaptic plasticity and neuropsychiatric disorders. *Nat Rev* 2007; **8**: 413–426.

35 Kanakry CG, Li Z, Nakai Y, Sei Y, Weinberger DR. Neuregulin-1 regulates cell adhesion via an ErbB2/phosphoinositide-3 kinase/Akt-dependent pathway: potential implications for schizophrenia and cancer. *PLoS ONE* 2007; **2**: e1369.

36 Schonfeldt-Lecuona C, Freudenmann RW, Tumani H, Kassubek J, Connemann BJ. Acute psychosis with a mediastinal carcinoma metastasis. *Med Sci Monit* 2005; **11**: CS6–CS8.

37 Wei Z, Qi J, Dai Y, Bowen WD, Mousseau DD. Haloperidol disrupts Akt signalling to reveal a phosphorylation-dependent regulation

of pro-apoptotic Bcl-XS function. *Cell Signalling* 2009; **21**: 161–168.
38 Riedel M, Strassnig M, Schwarz MJ, Muller N. COX-2 inhibitors as adjunctive therapy in schizophrenia: rationale for use and evidence to date. *CNS Drugs* 2005; **19**: 805–819.
39 Laan W, Smeets H, de Wit NJ, Kahn RS, Grobbee DE, Burger H. Glucocorticosteroids associated with a decreased risk of psychosis. *J Clin Psychopharmacol* 2009; **29**: 288–290.
40 Peet M. Omega-3 polyunsaturated fatty acids in the treatment of schizophrenia. *Israel J Psychiatry Relat Sci* 2008; **45**: 19–25.
41 Huang JT, Wang L, Prabakaran S, Wengenroth M, Lockstone HE, Koethe D *et al.* Independent protein-profiling studies show a decrease in apolipoprotein A1 levels in schizophrenia CSF, brain and peripheral tissues. *Mol Psychiatry* 2008; **13**: 1118–1128.
42 Sawa A, Cascella NG. Peripheral olfactory system for clinical and basic psychiatry: a promising entry point to the mystery of brain mechanism and biomarker identification in schizophrenia. *Am J Psychiatry* 2009; **166**: 137–139.
43 Kato T, Iwayama Y, Kakiuchi C, Iwamoto K, Yamada K, Minabe Y *et al.* Gene expression and association analyses of LIM (PDLIM5) in bipolar disorder and schizophrenia. *Mol Psychiatry* 2005; **10**: 1045–1055.
44 Le-Niculescu H, Balaraman Y, Patel S, Tan J, Sidhu K, Jerome RE *et al.* Towards understanding the schizophrenia code: an expanded convergent functional genomics approach. *Am J Med Genet B Neuropsychiatr Genet* 2007; **144**: 129–158.
45 Paunio T, Tuulio-Henriksson A, Hiekkalinna T, Perola M, Varilo T, Partonen T *et al.* Search for cognitive trait components of schizophrenia reveals a locus for verbal learning and memory on 4q and for visual working memory on 2q. *Hum Mol Genet* 2004; **13**: 1693–1702.
46 Kim S, Choi KH, Baykiz AF, Gershenfeld HK. Suicide candidate genes associated with bipolar disorder and schizophrenia: an exploratory gene expression profiling analysis of post-mortem prefrontal cortex. *BMC Genomics* 2007; **8**: 413.
47 Glatt SJ, Everall IP, Kremen WS, Corbeil J, Sasik R, Khanlou N *et al.* Comparative gene expression analysis of blood and brain provides concurrent validation of SELENBP1 up-regulation in schizophrenia. *Proc Natl Acad Sci USA* 2005; **102**: 15533–15538.
48 Vawter MP, Ferran E, Galke B, Cooper K, Bunney WE, Byerley W. Microarray screening of lymphocyte gene expression differences in a multiplex schizophrenia pedigree. *Schizophr Res* 2004; **67**: 41–52.
49 Bowden NA, Weidenhofer J, Scott RJ, Schall U, Todd J, Michie PT *et al.* Preliminary investigation of gene expression profiles in peripheral blood lymphocytes in schizophrenia. *Schizophrenia Res* 2006; **82**: 175–183.
50 Brzustowicz LM, Hodgkinson KA, Chow EW, Honer WG, Bassett AS. Location of a major susceptibility locus for familial schizophrenia on chromosome 1q21-q22. *Science* 2000; **288**: 678–682.
51 Dean B, Pavey G, Scarr E, Goeringer K, Copolov DL. Measurement of dopamine D2-like receptors in postmortem CNS and pituitary: differential regional changes in schizophrenia. *Life Sci* 2004; **74**: 3115–3131.
52 Seeman P, Guan HC, Nobrega J, Jiwa D, Markstein R, Balk JH *et al.* Dopamine D2-like sites in schizophrenia, but not in Alzheimer's, Huntington's, or control brains, for [3 H]benzquinoline. *Synapse* 1997; **25**: 137–146.
53 Torrey EF, Barci BM, Webster MJ, Bartko JJ, Meador-Woodruff JH, Knable MB. Neurochemical markers for schizophrenia, bipolar disorder, and major depression in postmortem brains. *Biol Psychiatry* 2005; **57**: 252–260.
54 Zvara A, Szekeres G, Janka Z, Kelemen JZ, Cimmer C, Santha M *et al.* Over-expression of dopamine D2 receptor and inwardly rectifying potassium channel genes in drug-naive schizophrenic peripheral blood lymphocytes as potential diagnostic markers. *Dis Markers* 2005; **21**: 61–69.
55 Sun J, Kuo PH, Riley BP, Kendler KS, Zhao Z. Candidate genes for schizophrenia: a survey of association studies and gene ranking. *Am J Med Genet B Neuropsychiatr Genet* 2008; **147B**: 1173–1181.
56 Allen NC, Bagade S, McQueen MB, Ioannidis JP, Kavvoura FK, Khoury MJ *et al.* Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet* 2008; **40**: 827–834.
57 Yamada K, Gerber DJ, Iwayama Y, Ohnishi T, Ohba H, Toyota T *et al.* Genetic analysis of the calcineurin pathway identifies members of the EGR gene family, specifically EGR3, as potential susceptibility candidates in schizophrenia. *Proc Natl Acad Sci USA* 2007; **104**: 2815–2820.
58 Straub RE, MacLean CJ, O'Neill FA, Walsh D, Kendler KS. Support for a possible schizophrenia vulnerability locus in region 5q22-31 in Irish families. *Mol Psychiatry* 1997; **2**: 148–155.
59 Devlin B, Bacanu SA, Roeder K, Reimherr F, Wender P, Galke B *et al.* Genome-wide multipoint linkage analyses of multiplex schizophrenia pedigrees from the oceanic nation of Palau. *Mol Psychiatry* 2002; **7**: 689–694.
60 Smalla KH, Mikhaylova M, Sahin J, Bernstein HG, Bogerts B, Schmitt A *et al.* A comparison of the synaptic proteome in human chronic schizophrenia and rat ketamine psychosis suggest that prohibitin is involved in the synaptic pathology of schizophrenia. *Mol Psychiatry* 2008; **13**: 878–896.
61 Faraone SV, Lasky-Su J, Glatt SJ, Van Eerdewegh P, Tsuang MT. Early onset bipolar disorder: possible linkage to chromosome 9q34. *Bipolar Disord* 2006; **8**: 144–151.
62 Kampman O, Anttila S, Illi A, Mattila KM, Rontu R, Leinonen E *et al.* Apolipoprotein E polymorphism is associated with age of onset in schizophrenia. *J Hum Genet* 2004; **49**: 355–359.
63 Hahn CG, Wang HY, Cho DS, Talbot K, Gur RE, Berrettini WH *et al.* Altered neuregulin 1-erbB4 signaling contributes to NMDA receptor hypofunction in schizophrenia. *Nat Med* 2006; **12**: 824–828.
64 Petryshen TL, Middleton FA, Kirby A, Aldinger KA, Purcell S, Tahl AR *et al.* Support for involvement of neuregulin 1 in schizophrenia pathophysiology. *Mol Psychiatry* 2005; **10**: 366–374, 328.
65 Chagnon YC, Roy MA, Bureau A, Merette C, Maziade M. Differential RNA expression between schizophrenic patients and controls of the dystrobrevin binding protein 1 and neuregulin 1 genes in immortalized lymphocytes. *Schizophrenia Res* 2008; **100**: 281–290.
66 Zhang HX, Zhao JP, Lv LX, Li WQ, Xu L, Ouyang X *et al.* Explorative study on the expression of neuregulin-1 gene in peripheral blood of schizophrenia. *Neurosci Lett* 2008; **438**: 1–5.
67 Blouin JL, Dombroski BA, Nath SK, Lasseter VK, Wolyniec PS, Nestadt G *et al.* Schizophrenia susceptibility loci on chromosomes 13q32 and 8p21. *Nat Genet* 1998; **20**: 70–73.
68 Chiu YF, McGrath JA, Thornquist MH, Wolyniec PS, Nestadt G, Swartz KL *et al.* Genetic heterogeneity in schizophrenia II: conditional analyses of affected schizophrenia sibling pairs provide evidence for an interaction between markers on chromosome 8p and 14q. *Mol Psychiatry* 2002; **7**: 658–664.
69 Gurling HM, Kalsi G, Brynjolfson J, Sigmundsson T, Sherrington R, Mankoo BS *et al.* Genomewide genetic linkage analysis confirms the presence of susceptibility loci for schizophrenia, on chromosomes 1q32.2, 5q33.2, and 8p21-22 and provides support for linkage to schizophrenia, on chromosomes 11q23.3-24 and 20q12.1-11.23. *Am J Hum Genet* 2001; **68**: 661–673.
70 Pulver AE, Mulle J, Nestadt G, Swartz KL, Blouin JL, Dombroski B *et al.* Genetic heterogeneity in schizophrenia: stratification of genome scan data using co-segregating related phenotypes. *Mol Psychiatry* 2000; **5**: 650–653.
71 Suarez BK, Duan J, Sanders AR, Hinrichs AL, Jin CH, Hou C *et al.* Genomewide linkage scan of 409 European-Ancestry and African American families with schizophrenia: suggestive evidence of linkage at 8p23.3-p21.2 and 11p13.1-q14.1 in the combined sample. *Am J Hum Genet* 2006; **78**: 315–333.
72 Kaufmann CA, Suarez B, Malaspina D, Pepple J, Svrakic D, Markel PD *et al.* NIMH Genetics Initiative Millenium Schizophrenia Consortium: linkage analysis of African-American pedigrees. *Am J Med Genet* 1998; **81**: 282–289.
73 Arion D, Unger T, Lewis DA, Levitt P, Mirnics K. Molecular evidence for increased expression of genes related to immune and chaperone function in the prefrontal cortex in schizophrenia. *Biol Psychiatry* 2007; **62**: 711–721.
74 DeLisi LE, Shaw SH, Crow TJ, Shields G, Smith AB, Larach VW *et al.* A genome-wide scan for linkage to chromosomal regions in 382 sibling pairs with schizophrenia or schizoaffective disorder. *Am J Psychiatry* 2002; **159**: 803–812.

75 Straub RE, MacLean CJ, Ma Y, Webb BT, Myakishev MV, Harris-Kerr C *et al.* Genome-wide scans of three independent sets of 90 Irish multiplex schizophrenia families and follow-up of selected regions in all families provides evidence for multiple susceptibility genes. *Mol Psychiatry* 2002; **7**: 542–559.
76 Clark D, Dedova I, Cordwell S, Matsumoto I. A proteome analysis of the anterior cingulate cortex gray matter in schizophrenia. *Mol Psychiatry* 2006; **11**: 459–470, 423.
77 Takahashi S, Faraone SV, Lasky-Su J, Tsuang MT. Genome-wide scan of homogeneous subtypes of NIMH genetics initiative schizophrenia families. *Psychiatry Res* 2005; **133**: 111–122.
78 Cardno AG, Holmans PA, Rees MI, Jones LA, McCarthy GM, Hamshere ML *et al.* A genomewide linkage study of age at onset in schizophrenia. *Am J Med Genet* 2001; **105**: 439–445.
79 Vawter MP, Barrett T, Cheadle C, Sokolov BP, Wood 3rd WH, Donovan DM *et al.* Application of cDNA microarrays to examine gene expression differences in schizophrenia. *Brain Res Bull* 2001; **55**: 641–650.
80 Mudge J, Miller NA, Khrebtukova I, Lindquist IE, May GD, Huntley JJ *et al.* Genomic convergence analysis of schizophrenia: mRNA sequencing reveals altered synaptic vesicular transport in post-mortem cerebellum. *PLoS ONE* 2008; **3**: e3625.
81 Aston C, Jiang L, Sokolov BP. Microarray analysis of postmortem temporal cortex from patients with schizophrenia. *J Neurosci Res* 2004; **77**: 858–866.
82 McInnes LA, Lauriat TL. RNA metabolism and dysmyelination in schizophrenia. *Neurosci Biobehav Rev* 2006; **30**: 551–561.
83 Fallin MD, Lasseter VK, Wolyniec PS, McGrath JA, Nestadt G, Valle D *et al.* Genomewide linkage scan for schizophrenia susceptibility loci among Ashkenazi Jewish families shows evidence of linkage on chromosome 10q22. *Am J Hum Genet* 2003; **73**: 601–611.
84 Sullivan PF, Lin D, Tzeng JY, van den Oord E, Perkins D, Stroup TS *et al.* Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol Psychiatry* 2008; **13**: 570–584.
85 Benes FM, Lim B, Matzilevich D, Subburaju S, Walsh JP. Circuitry-based gene expression profiles in GABA cells of the trisynaptic pathway in schizophrenics versus bipolars. *Proc Natl Acad Sci USA* 2008; **105**: 20935–20940.
86 Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM *et al.* Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science (New York, NY)* 2008; **320**: 539–543.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)