ORIGINAL ARTICLE

Identifying blood biomarkers for mood disorders using convergent functional genomics

H Le-Niculescu[1,2,3], SM Kurian[4], N Yehyawi[1,5], C Dike[1,5], SD Patel[1,3], HJ Edenberg[6], MT Tsuang[7], DR Salomon[4], JI Nurnberger Jr[3] and AB Niculescu[1,2,3,5]

[1]Laboratory of Neurophenomics, Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, USA; [2]INBRAIN, Department of Psychiatry, Indiana University School of Medicine, Indianapolis, IN, USA; [3]Institute of Psychiatric Research, Indiana University School of Medicine, Indianapolis, IN, USA; [4]Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA, USA; [5]Indianapolis VA Medical Center, Indianapolis, IN, USA; [6]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA and [7]Department of Psychiatry, University of California at San Diego, La Jolla, CA, USA

There are to date no objective clinical laboratory blood tests for mood disorders. The current reliance on patient self-report of symptom severity and on the clinicians' impression is a rate-limiting step in effective treatment and new drug development. We propose, and provide proof of principle for, an approach to help identify blood biomarkers for mood state. We measured whole-genome gene expression differences in blood samples from subjects with bipolar disorder that had low mood vs those that had high mood at the time of the blood draw, and separately, changes in gene expression in brain and blood of a mouse pharmacogenomic model. We then integrated our human blood gene expression data with animal model gene expression data, human genetic linkage/association data and human postmortem brain data, an approach called convergent functional genomics, as a Bayesian strategy for cross-validating and prioritizing findings. Topping our list of candidate blood biomarker genes we have five genes involved in myelination (*Mbp*, *Edg2*, *Mag*, *Pmp22* and *Ugt8*), and six genes involved in growth factor signaling (*Fgfr1*, *Fzd3*, *Erbb3*, *Igfbp4*, *Igfbp6* and *Ptprm*). All of these genes have prior evidence of differential expression in human postmortem brains from mood disorder subjects. A predictive score developed based on a panel of 10 top candidate biomarkers (five for high mood and five for low mood) shows sensitivity and specificity for high mood and low mood states, in two independent cohorts. Our studies suggest that blood biomarkers may offer an unexpectedly informative window into brain functioning and disease state.
Molecular Psychiatry advance online publication, 26 February 2008; doi:10.1038/mp.2008.11

Introduction

Research into the biological basis of mood disorders (bipolar disorders, depression) has, until recently, been advancing in human and animal studies more or less independently. The two avenues of research have complementary strengths and weaknesses. In human genetic studies, for example, in samples of patients with mood disorders and their family members, positional cloning methods such as linkage analysis, linkage disequilibrium mapping, candidate gene association analysis and, more recently, whole-genome association studies are narrowing the search for the chromosomal regions harboring risk genes for the illness and, in some cases, identifying plausible candidate genes and polymorphisms that will require further validation. Human postmortem brain gene expression studies have also been employed as a way of trying to identify candidate genes for mood and other neuropsychiatric disorders. In general, human studies suffer from issues of sensitivity—the signal is hard to detect due to noise generated by the genetic heterogeneity of individuals and the effects of diverse environmental exposures on gene expression and phenotypic penetrance. In animal studies, carried out in isogenic strains with controlled environmental exposure, the identification of putative neurobiological substrates of mood disorders is typically done by modeling human mood disorders through pharmacological or genetic manipulations. Animal model studies suffer from issues of specificity—questions regarding direct relevance to the human disorder modeled. Each independent line of investigation (that

Correspondence: Dr AB Niculescu, INBRAIN and Laboratory of Neurophenomics, Institute of Psychiatric Research, Indiana University School of Medicine, 791 Union Drive, Indianapolis, IN 46202-4887, USA.
E-mail: anicules@iupui.edu

2

is, human and animal studies) is contributing to the incremental gains in knowledge of mood disorders etiology witnessed in the past decade. Yet, it is now apparent that a lack of integration between these two lines of investigation may constitute a missed opportunity for accelerating the pace of discovery. Our group has developed an approach, termed convergent functional genomics (CFG), which translationally cross-matches animal model gene expression data with human genetic data and human tissue data (blood, postmortem brain), as a Bayesian strategy of cross-validating findings, reducing the false positives and false negatives inherent in each individual approach, and helping identify true candidate genes, pathways and mechanisms for neuropsychiatric disorders. The CFG approach has already been applied with some success to the study of bipolar disorders,[1,2] alcoholism[3] and schizophrenia.[4] In the work described in this paper, we extend this approach to peripheral biomarker discovery efforts.

Objective biomarkers of illness and treatment response would make a significant difference in our ability to diagnose and treat patients with psychiatric disorders, eliminating subjectivity and reliance of patient's self-report of symptoms. Blood gene expression profiling has emerged as a particularly interesting area of research in the search for peripheral biomarkers.[5–10] Most of the studies to date have focused on human lymphocytes gene expression profiling, comparison between illness groups and normal controls, and cross-matching with human postmortem brain gene expression data. They suffer from one of both of the following limitations: (1) the sample size used in most reports so far is small. Given the genetic heterogeneity in human samples and the effects of illness state and environmental history, including medications and street drugs, on gene expression, it is questionable if they have sufficient power to extract bona fide findings in and of themselves, despite the variety of sophisticated statistical methodologies used; (2) use of lymphoblastoid cell lines (LCLs). Passaged LCLs provide a self-renewable source of material, and are purported to avoid the effects of environmental exposure of cells from fresh blood. Fresh blood, however, with phenotypic state information gathered at the time of harvesting, may be more informative than immortalized lymphocytes, and avoid some of the caveats of Epstein–Barr virus immortalization and cell culture passaging.

The current state of our understanding of the genetic and neurobiological bases for bipolar mood disorder and depression in general, and of peripheral molecular biomarkers of illness in particular, is still inadequate. A rate-limiting step, which we propose to overcome, has been the lack of concerted integration across disciplines and methodologies. The use of such a multidisciplinary, integrative research framework should lead to a reduction in the high rate of inferential errors committed in studies of complex diseases like bipolar disorder and depression. To our knowledge, no one has reported to date a comprehensive investigation of human fresh blood gene expression studies tied to quantitative state measures of mood, and cross-validated that data with blood gene expression profiling in conjunction with brain gene expression studies in an pharmacogenomic animal model relevant to bipolar disorder, as well as integrated the findings in the context of the available human genetic linkage/association data, postmortem brain data and information on biological pathways. We present data showing that such an expanded CFG approach (Figure 2) may be fruitful for biomarker discovery, and overcome the caveats mentioned above. Moreover, we project, on the basis of our preliminary work, that panels of biomarkers rather than single biomarkers are going to emerge as clinically useful tools.

Materials and methods

Human subjects
We present data from three independent cohorts. One cohort consisted of 29 subjects with bipolar I disorder, from which the primary biomarker data were derived. A second (replication) cohort consisted of 19 subjects with bipolar I disorder, and a third cohort consisted of 30 subjects with psychotic disorders (schizophrenia, schizoaffective disorder and substance-induced psychosis). The diagnosis is established by a structured clinical interview— Diagnostic Interview for Genetic Studies (DIGS), which has details on the course of illness and phenomenology, and is the scale used by the Genetics Initiative Consortia for both Bipolar Disorder and Schizophrenia.

Subjects consisted of men and women over 18 years of age. Subjects were recruited from the patient population at the Indianapolis VA Medical Center, the Indiana University School of Medicine, as well as various facilities that serve people with mental illnesses in Indiana. 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 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 medical or neurological illness 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 (DIGS), and then a visual analog scale for mood (VAS Mood) at the time of blood draw. Whole blood (10 ml) was collected in two RNA-stabilizing PAXgene tubes, labeled with an anonymized ID number, and stored at $-80\,^{\circ}C$ in a locked freezer until the time of future processing.

Table 1 Demographics: (a) individual (b) aggregate

(a) Individual demographic data

Subject ID	Diagnosis	Age	Gender	Ethnicity	VAS mood (0–100)
Primary bipolar cohort					
174-1197-001	BP	37	Male	Caucasian	20
174-1055-001	BP	46	Male	Caucasian	20
phchp029v1	BP	56	Male	Caucasian	22
174-1126-001	BP	33	Male	Caucasian	24
174-1173-001	BP	56	Male	Caucasian	27
174-1161-001	BP	46	Male	Caucasian	29
174-1150-001	BP	52	Male	Caucasian	31
174-1042-001	BP	58	Male	Caucasian	37
174-1112-001	BP	24	Male	Caucasian	38
phchp027v1	BP	40	Male	Caucasian	38
174-1137-001	BP	48	Male	African American	39
phchp023v1	BP	52	Male	Caucasian	39
174-1115-001	BP	42	Male	American Indian	40
phchp020v1	BP	62	Male	Caucasian	42
phchp031v1	BP	51	Male	Caucasian	47
phchp028v1	BP	50	Female	Asian	52
phchp030v1	BP	49	Male	Caucasian	61
174-1107-001	BP	39	Male	Caucasian	63
174-1130-001	BP	21	Male	African American	65
174-5001-001	BP	23	Male	Caucasian	66
174-1132-001	BP	22	Male	African American	71
174-1160-001	BP	52	Male	African American	72
174-1171-001	BP	56	Female	Caucasian	72
174-1156-001	BP	57	Male	Caucasian	72
174-1037-001	BP	54	Male	Caucasian	73
174-5002-001	BP	26	Male	Caucasian	73
174-1119-001	BP	38	Male	Caucasian	80
phchp020v2	BP	62	Male	Caucasian	84
174-1193-001	BP	53	Male	African American	84
Psychosis cohort					
phchp022v2	SZ	48	Male	Caucasian	15
phchp005v2	SZA	45	Male	Caucasian	19
phchp025v1	SZ	42	Male	Caucasian	29
phchp021v2	SZA	49	Male	Hispanic	29
phchp006v2	SZA	52	Male	African American	33
phchp033v1	SZA	48	Male	Caucasian	35
phchp016v1	SZ	54	Male	African American	38
phchp021v1	SZA	48	Male	Hispanic	39
phchp019v1	SubPD	50	Male	African-American	41
phchp003v3	SZ	50	Male	African American	47
phchp010v1	SZA	45	Male	Caucasian	48
phchp024v1	SZA	49	Male	African American	49
phchp003v2	SZ	50	Male	African American	53
phchp009v1	SZ	55	Male	African American	54
phchp010v2	SZA	45	Male	Caucasian	55
phchp006v1	SZA	52	Male	African American	57
phchp026v1	SZA	49	Male	African-American	64
phchp022v1	SZ	48	Male	Caucasian	65
phchp010v3	SZA	45	Male	Caucasian	65
phchp014v1	SubPD	55	Male	African American	69
phchp004v1	SZA	55	Male	African American	69
phchp012v1	SZA	55	Male	Caucasian	70

Table 1 Continued

(a) Individual demographic data

Subject ID	Diagnosis	Age	Gender	Ethnicity	VAS mood (0–100)
phchp012v2	SZA	55	Male	Caucasian	71
phchp018v1	SZA	54	Female	Caucasian	73
phchp015v1	SubPD	48	Male	African American	76
phchp008v1	SZ	47	Male	African American	76
phchp005v1	SZA	45	Male	Caucasian	81
phchp017v2	SZA	53	Male	African American	84
phchp013v1	SZA	53	Male	African American	89
phchp003v1	SZ	50	Male	African American	93
Secondary bipolar cohort					
phchp039v1	BP	52	Male	Caucasian	11
phchp023v2	BP	52	Male	Caucasian	20
174-1216-001	BP	60	Male	Caucasian	23
174-1278-001	BP	22	Male	Caucasian	24
174-1232-001	BP	45	Male	Caucasian	32
phchp045v1	BP	36	Male	Caucasian	36
174-1203-001	BP	39	Male	African American	49
174-1199-001	BP	41	Male	Caucasian	53
174-1237-001	BP	36	Male	Caucasian	57
174-5006-001	BP	60	Male	Caucasian	66
phchp053v1	BP	58	Male	Caucasian	68
174-1211-001	BP	27	Male	Caucasian	75
phchp031v2	BP	51	Male	Caucasian	79
174-1204-001	BP	52	Male	Caucasian	81
174-1255-001	BP	50	Male	Caucasian	81
174-1220-001	BP	68	Male	Caucasian	82
174-1096-001	BP	50	Male	Caucasian	83
phchp056v1	BP	36	Male	Caucasian	84
174-1258-001	BP	36	Male	Caucasian	90

(b) Aggregate demographic data

	Primary bipolar cohort			Psychosis cohort			Secondary bipolar cohort		
	Bipolar low mood	Bipolar high mood	Bipolar overall	Schizoaffective	Schizophrenia	Substance-induced psychotic disorder	Bipolar low mood	Bipolar high mood	Bipolar overall
Number of subjects	13	13	29	18	9	3	6	10	19
Gender (males/females)	13:0	12:1	27:2	17:1	9:0	3:0	6:0	10:0	19:0
Age mean years (s.d.) range	45.4 (10.0)	41.9 (15.6)	45.0 (12.5)	49.8 (3.9)	49.3 (3.8)	51.0 (3.6)	44.5 (13.6)	48.8 (12.5)	45.8 (12.0)
Duration of illness mean years (s.d.) range	24–58 22.8 (10.2)	21–62 20.4 (17.1)	21–62 21.6 (13.7)	45–55 31.2 (6.3)	42–55 26.7 (4.7)	48–55 25.0 (6.0)	22–60 25.8 (16.1)	27–68 27.2 (6.6)	22–68 25.8 (10.5)
Ethnicity (Caucasian/other)	5–40 11/2	2–49 10/3	2–49 23/6	17–42 9/9	20–26 3/6	20–32 0/3	7–53 6/0	19–38 10/0	7–53 18/1

Abbreviations: BP, bipolar; SubPD, substance-induced psychosis; SZ, schizophrenia; SZA, schizoaffective disorder.
VAS Mood score at time of blood draw, on a scale 0 (lowest mood) to 100 (highest mood).
Diagnosis established by DIGS comprehensive structured clinical interview.

Whole-blood (predominantly lymphocyte) RNA was extracted for microarray gene expression studies[5,11,12] from the PAXgene tubes blood, as detailed below.

Human blood gene expression experiments and analysis

RNA extraction. Whole blood (2.5–5 ml) was collected into each Paxgene tube by routine venipuncture. Paxgene tubes contain proprietary reagents for the stabilization of RNA. The cells from 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 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 bound 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. That consisted 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* aRNA amplification kit. The procedure is briefly outlined below and involves the following steps:

(1) Reverse transcription to synthesize first strand cDNA is primed with the T7 Oligo(dT) primer to synthesize cDNA containing a T7 promoter sequence.
(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 aRNA were hybridized to Affymetrix Human Genome U133 Plus 2.0 GeneChips according to manufacturer's protocols http://www.affymetrix.com/support/technical/manual/expression_manual.affx. All glyceraldehyde-3-phosphate dehydrogenase (GAPDH) 3′/5′ ratios should be less than 2.0 and backgrounds under 50. 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.

Analysis. We used the subject's mood scores at time of blood collection, obtained from a VAS Mood scale (Figure 1). We looked at only 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 was detected as Absent in the low mood subjects and detected as Present in the high mood subjects, as being candidate biomarker genes for elevated mood. Conversely, genes whose expression were detected as Present in the low mood subjects and Absent in the high mood subjects were classified as candidate biomarker genes for low mood.

We employed two thresholds for analysis of gene expression differences between low and high mood (Table 2). 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 low and high mood (reflecting an at least threefold mood 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 low and high mood (reflecting an at least 1.5-fold mood state-related enrichment of the genes thus filtered).
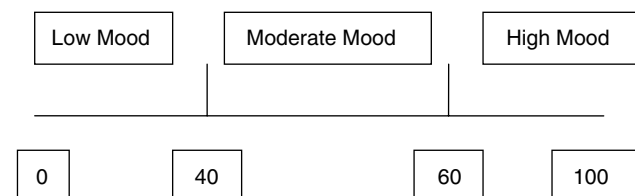

Figure 1 Visual analog mood scale (VAS) scoring. At the time of the blood draw, the subject draws a line on a 100 mm VAS to mark where subject's mood is at that moment in time, compared to worst subject has ever felt (0) and best subject has ever felt (100).

Table 2 High- and low-threshold analyses in primary bipolar cohort

Analysis	Bipolar subjects (n = 29) 13 low mood and 13 high mood
High-threshold candidate biomarker genes (changed in ⩾75% subjects; i.e. at least threefold enrichment)	10/13 low mood vs 10/13 high mood A/P and P/A analyses
Low-threshold candidate biomarker genes (changed in ⩾60% subjects; i.e. at least 1.5-fold enrichment)	8/13 low mood vs 8/13 high mood A/P and P/A analyses

Genes are considered candidate biomarkers for high mood if they are called by the Affymetrix MAS5 software as Absent (A) in the blood of low mood subjects and detected as Present (P) in the blood of high mood subjects. Conversely, genes are considered candidate biomarkers for low mood if they are detected as Present (P) in low mood subjects and Absent (A) in high mood subjects.

Animal model gene expression studies
Our bipolar pharmacogenomic model consisted of methamphetamine and valproate treatments in mice (see Ogden et al.[2] for experimental details and analysis/categorization of gene expression data).

For the current work, we repeated that series of experiments, to obtain blood gene expression data also. 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, methamphetamine (10 mg kg^{-1}), valproate (200 mg kg^{-1}) or a combination of methamphetamine and valproate (10 and 200 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 the 24 h time point behavioral test, 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 per mouse was collected into a PAXgene blood RNA collection tubes, BD Diagnostic (VWR.com). Blood samples from three mice per treatment condition were pooled. The PAXgene blood vials were stored at −4 °C overnight, and then at −80 °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–Human or GLOBINclear–Mouse/Rat (Ambion/Applied Biosystems Inc., Austin, TX, USA) 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 UV absorption and by 260/280 ratios, respectively (Nanodrop spectrophotometer). Starting material of total RNA labeling reactions was kept consistent within each independent microarray experiment.

Equal amount of total RNA extracted from pooled blood samples was used for labeling and microarray assays. We used Mouse Genome 430 2.0 arrays (Affymetrix, Santa Clara, CA, USA). The GeneChip Mouse Genome 430 2.0 array contains over 45 000 probe sets that analyze the expression level of over 39 000 transcripts and variants from over 34 000 well-characterized mouse genes. Standard Affymetrix protocols were used to reverse transcribe the messenger RNA and generate biotinylate 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 intraexperiment. Samples were hybridized at 45 °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 GAPDH 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. 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) intraexperiment, 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: convergent functional genomics

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) that 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 were 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, bipolar, depression, 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 mood disorders (bipolar disorder, depression). 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 bipolar postmortem brain data reflected a depressed phase of the illness. While this may arguably be the case, it is nevertheless an assumption, as no consistent objective data exist 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 mood disorders (bipolar disorder or depression). 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.

Convergent functional genomics 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. They received 1 point for each external cross-validating line of evidence (human postmortem brain data, human genetic data, animal model brain data and animal model blood data). Genes received additional bonus points if changed in human brain and blood, as follows: 2 points if changed in the same direction, 1 point if changed in opposite direction. 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 point if changed in opposite direction. Thus, the total maximum CFG score that a candidate biomarker gene could have was 9 $(2 + 4 + 2 + 1)$. As we are interested in discovering blood biomarkers, and because of caveats discussed above, we weighted more heavily our 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. It has not escaped our attention that 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 5.0 (Ingenuity Systems, Redwood City, CA, USA) was used to analyze the biological functions categories of the top candidate genes resulting from our CFG analysis (Table 5), as well as employed to identify genes in our data sets that are the target of existing drugs (Supplementary Table 2S).

Results

We have conducted gene expression profiling studies in peripheral whole blood (see 'Materials and methods' for methodological details) from a primary cohort of 29 human subjects with bipolar I disorder (27 men, 2 women; Table 1). In total, 13 had low self-reported mood scores (below 40) on the VAS Mood scale (Figure 1), and 13 had high self-reported mood scores (above 60). Three of them had intermediate mood scores (between 40 and 60). We have used their mood scores at the time of blood collection as a way of narrowing the field and identifying candidate biomarker genes for mood. We have 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 was detected as Absent in the low mood

8

subjects and detected as Present in the high mood subjects, as being candidate biomarker genes for elevated mood state (mania). Conversely, genes whose expression was detected as Present in the low mood subjects and Absent in the high mood subjects are being classified as candidate biomarker genes for low mood state (depression) (Tables 2 and 3). It has to be noted that it is possible that some of the genes associated with high or low mood state may not necessarily be involved in the induction of that state, but rather in its suppression as part of homeostatic regulatory networks or treatment–response mechanisms (similar conceptually to oncogenes and tumor-suppressor genes).

Second, we have employed two thresholds for analysis of gene expression differences between low and high mood (Table 2). First, we used a high threshold, with at least 75% of subjects in a cohort showing a change in expression from Absent to Present between low and high mood (reflecting an at least threefold mood state-related enrichment of the genes thus filtered). As psychiatric disorders are clinically and (likely) genetically heterogeneous, with different combinations of genes and biomarkers present in different subgroups, we also used a low threshold, with at least 60% of subjects in a cohort showing a change in expression from Absent to Present between low and high mood (reflecting an at least 1.5-fold mood state-related enrichment of the genes thus filtered). The high threshold will 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 mood state (high or low) tested. They may reflect common aspects related to mood disorders across a diverse subject population, but may also be a reflection of the effects of common medications used in the population tested, such as mood stabilizers. 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.

By cross-validating with animal model and other human data sets (Figure 2a) using CFG, we were able to extract a shorter list of genes for which there are external corroborating lines of evidence (human genetic evidence, human postmortem brain data, animal model brain and blood data) linking them to mood disorders (bipolar disorder, depression), thus reducing the risk of false positives. This cross-validation identifies candidate biomarkers that are more likely directly related to the relevant neuropathology, as opposed to being potential artifactual effects or indirect effects of lifestyle, environment and so on.

Using our approach for analyzing at first pass our human blood gene expression data, out of over 40 000 genes and ESTs on the Affymetrix Human Genome U133 Plus 2.0 GeneChip, by using the high threshold, we have ended up with 21 novel candidate biomarker genes (13 known genes and 7 ESTs; Table 3; Supplementary Table 3S), of which 8 had at least one line of prior independent evidence for potential involvement in mood disorders (that is, CFG score of 3 or above). In addition to the high threshold genes, by using the low threshold, we have a larger list totaling 661 genes (539 known genes and 122 ESTs; Table 3; Supplementary Table 3S), of which an additional 24 had at least two lines of prior independent evidence for potential involvement in mood disorders (that is, CFG score of 3 or above). Of interest, four of our low-threshold candidate biomarker genes (*Bclaf1* and *Rdx*,[8] *Gosr2* and *Wdr34*[42]) had been previously reported to be changed in expression in the same direction, in LCLs from bipolar subjects.

Making a combined list of all the high-value candidate biomarker genes identified as described above, consisting of the high-threshold genes with at least one other external line of evidence ($n = 8$) and of the additional low-threshold genes with at least two other external lines of evidence ($n = 24$) and the low-threshold genes with prior LCL evidence ($n = 4$), we end up with a list of 36 top candidate biomarker genes for mood, prioritized based on CFG score (Table 3).

Picking up the five top-scoring candidate biomarkers for high mood (*Mbp*, *Edg2*, *Fzd3*, *Atxn1* and *Ednrb*) and the five top-scoring candidate biomarkers for low mood (*Fgfr1*, *Mag*, *Pmp22*, *Ugt8* and *Erbb3*), we have established a panel of 10 biomarkers for mood disorder that may have diagnostic and predictive value.

To test the predictive value of our panel (to be called the BioM-10 Mood panel), we have looked in the cohort of 29 bipolar disorder subjects, containing the 26 subjects (13 low mood, 13 high mood) from which the candidate biomarker data were derived, as well as three additional subjects with mood in the intermediate range (self-reported mood scores between 40 and 60). We derived a prediction score for each subject, based on the presence or absence of the 10 biomarkers of the panel in their blood GeneChip data. Each of the 10 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 sum of the high mood biomarker scores divided by the sum of the low mood biomarker scores is multiplied by 100, and provides a prediction score. If the ratio of high biomarker genes to low mood biomarker genes is 1, that is, the two sets of genes are equally represented, the mood prediction score is $1 \times 100 = 100$. The higher the score, the higher the predicted likelihood that the subject will have high mood. We compared the predictive score with actual self-reported mood scores in the primary cohort of subjects with a diagnosis of

Table 3 Top candidate biomarker genes for mood prioritized by CFG score for multiple independent lines of evidence

Gene symbol/name	Entrez gene ID	Human blood data	Human postmortem brain, lymphocytes	Human brain and blood concordance/codirectionality	Human genetic linkage/association	Bipolar mouse model brain[2]	Bipolar mouse model blood	CFG score
Mbp (myelin basic protein)	4155	I	Up[13] (male) BP Down[13] (female) BP Down[14] Bipolar	Yes/Yes	18q23 BP[15–17]		Cat-IV Meth (I)	6
Edg2 (endothelial differentiation, lysophosphatidic acid G-protein-coupled receptor, 2)	1902	I	Down[18] MDD Down[19] (PFC) BP Up[19] (Parietal Cortex) BP	Yes/Yes	9q31.3 BP[20]			5
Fgfr1 (fibroblast growth factor receptor 1)	2260	D	Up[21] MDD	Yes/Yes	8p12 BP[22,23]			5
Fzd3 (frizzled homologue 3 (Drosophila))	7976	I	Down[24] BP	Yes/Yes	8p21.1 BP[22,23]			5
Mag (myelin-associated glycoprotein)	4099	D	Down[18] MDD	Yes/No	19q13.12 Depression[25]	CP Cat-IV Meth (I)		5
Pmp22 (peripheral myelin protein 22)	5376	D	Down[18] MDD	Yes/No	17p12 BP[22,26]	CP Cat-IV Meth (I)		5
Ugt8 (UDP glycosyltransferase 8 (UDP-galactose ceramide galactosyltransferase))	7368	D	Down[18] MDD	Yes/No	4q26 BP[27]	CP Cat-II (I)		5
Erbb3 (neuregulin receptor (v-erb-a erythroblastic leukemia viral oncogene homologue 4 (avian)))	2065	D	Down[28] MDD Down[14] BP	Yes/No	12q13.2 Depression[25]			4
Igfbp4 (insulin-like growth factor-binding protein 4)	3487	D	Down[29] BP	Yes/No	17q21.2 Depression[25]			4
Igfbp6 (insulin-like growth factor-binding protein 6)	3489	D	Down[29] BP	Yes/No	12q13 Depression[25]			4
Pde6d (phosphodiesterase 6D, cGMP-specific, rod, δ)	5147	D	Up[29] BP	Yes/Yes	2q37.1			4
Ptprm (protein tyrosine phosphatase, receptor type, M)	5797	D	Up[29] BP	Yes/Yes	18p11.23			4
Atp2c1 (ATPase, Ca^{2+} sequestering)	27032	D			3q21.3 BP[17,30]	CP Cat-IV Meth (D)		3
Atxn1 (ataxin 1)	6310	I			6p22.3 BP[31]			3
Btg1 (B-cell translocation gene 1, antiproliferative)	694	D			12q21.33 BP[17,32]		Cat-III Meth (D)	3
C6orf182 (chromosome 6 open reading frame 182)	285753	D			6q21 BP[22,27,30,33,34]			3
Dicer1 (Dicer1, Dcr-1 homologue (Drosophila))	23405	D	Down[18] MDD	Yes/No	14q32.13			3
Dnajc6 (DnaJ (Hsp40) homologue, subfamily C, member 6)	9829	D (HT)			1p31.3 BP[23,35] Depression[36]			3
Ednrb (endothelin receptor type B)	1910	I			13q22.3 BP[37]	CP Cat-III Meth (I)		3
Elovl5 (ELOVL family member 5, elongation of long-chain fatty acids (yeast))	60481	D			6p12.1 BP[27]		Cat-IV VPA (D)	3
Gnal (guanine nucleotide-binding protein, α-stimulating, olfactory type)	2774	D (HT)			18p11.21 BP[38]			3
Klf5 (Kruppel-like factor 5)	688	D (HT)			13q22.1 BP[37]			3
Lin7a (lin 7 homologue a (caenorhabditis elegans))	8825	D			12q21.31 BP[32]	Increased[1] (rat brain)		3
Manea (mannosidase, endo-α)	79694	D (HT)			6q16.1 BP[27,30,33] Depression[39]			3
Nupl1 (nucleoporin-like 1)	9818	D (HT)			13q12.13 BP[23]			3
Pde6b (phosphodiesterase 6B, cGMP-specific, rod, β (congenital stationary night blindness 3, autosomal dominant))	5158	D	Down[18] MDD	Yes/No	4p16.3			3
Slc25a23 (solute carrier family 25 (mitochondrial carrier; phosphate carrier), member 23)	79085	D (HT)			19p13.3	CP Cat-IV VPA (I)		3

Table 3 Continued

Gene symbol/name	Entrez gene ID	Human blood data	Human postmortem brain, lymphocytes	Human brain and blood concordance/codirectionality	Human genetic linkage/association	Bipolar mouse model brain[2]	Bipolar mouse model blood	CFG score
Synpo (synaptopodin)	11346	D			5q33.1 BP[40]	PFC Cat-III Meth (D)		3
Tgm2 (transglutaminase 2, C polypeptide)	7052	D			20q11.23 BP[41]		Cat-III Meth (D)	3
Tjp3 (tight-junction protein 3 (zona occludens 3))	27134	D (HT)			19p13.3 BP[15]			3
Tpd52 (tumor protein D52)	7163	D (HT)			8q21.13 BP[26]			3
Trpc1 (transient receptor potential cation channel, subfamily C, member 1)	7220	D			3q23 BP[30]	CP Cat-IV VPA (I)		3
Bclaf1 (BCL2-associated transcription factor 1)	9774	D	Down[8] (lymphocytes) BP		6q23.3			2
Gosr2 (Golgi SNAP receptor complex member 2)	9570	D	Down[42] (lymphocytes) BP		17q21.32			2
Rdx (radixin)	5962	D[8]	Down[8] (lymphocytes) BP		11q22.3			2
Wdr34 (WD repeat domain 34)	89891	D	Down[42] (lymphocytes) BP		9q34.11			2

Abbreviations: HT, high threshold; MDD, major depressive disorder; METH, methamphetamine; VPA, valproate.
Top candidate biomarker genes for mood. For human blood data: I, increased in high mood (mania); D, decreased in high mood (mania)/increased in low mood (depression). For postmortem brain data: Up, increased; Down, decreased in expression. For mouse data: I, increased; D, decreased in expression.

bipolar mood disorder ($n = 29$). A prediction score of 100 and above had an 84.6% sensitivity and a 68.8% specificity for predicting high mood. A prediction score below 100 had a 76.9% sensitivity and an 81.3% specificity for predicting low mood (Table 4a; Figure 3a).

Additionally, we have conducted human blood gene expression analysis in an independent cohort consisting of 30 subjects with other psychotic disorders (schizophrenia, schizoaffective disorder, substance-induced psychosis), who also had mood state scores obtained at the time of the blood draw. The subjects in the psychosis cohort also had a distribution of low ($n = 9$), intermediate ($n = 7$) and high ($n = 14$) mood scores. This cohort was used as a way to verify the predictive power of the mood state biomarker panel, independent of a bipolar disorder diagnosis.

In the psychotic disorders cohort ($n = 30$), with various psychotic disorders diagnoses, a prediction score of 100 and above had a 71.4% sensitivity and a 62.5% specificity for predicting high mood. A prediction score below 100 had a 66.7% sensitivity and a 61.9% specificity for predicting low mood (Table 4b; Figure 3b).

Moreover, we have also conducted human blood gene expression analysis in a second independent bipolar disorder cohort, subsequently collected, consisting of 19 subjects. The subjects in the secondary bipolar cohort had a distribution of low ($n = 6$), intermediate ($n = 3$) and high ($n = 10$) mood scores. The second bipolar cohort was used as a replication cohort, to verify the predictive power of the mood state biomarker panel identified by analysis of data from the primary bipolar cohort.

In the second bipolar cohort ($n = 19$), a prediction score of 100 and above had a 70.0% sensitivity and a 66.7% specificity for predicting high mood. A prediction score below 100 had a 66.7% sensitivity and a 61.5% specificity for predicting low mood (Table 4c; Figure 3c).

The primary and secondary bipolar mood disorder cohorts are *a priori* more related and germane to mood state biomarkers identification, but may have blood gene expression changes, at least in part, due to the common pharmacological agents used to treat bipolar mood disorders. The psychotic disorders cohort may have blood gene expression changes related to mood state irrespective of the diagnosis and the different medication classes subjects with different diagnoses are on (Table 1; Figure 2b). The psychosis cohort was also notably different in terms of the ethnic distribution (Table 1b).

Lastly, we interrogated the MIT/Broad Institute Connectivity Map[43] with a signature query composed of the genes in our BioM-10 Mood panel of top biomarkers for low and high mood (Figure 4). We wanted to see which drugs in the Connectivity Map database have similar effects on gene expression as the effects of high mood on gene expression, and which drugs have the opposite effect to high mood. As such, as part of our signature query, the five biomarkers for high mood were considered as genes
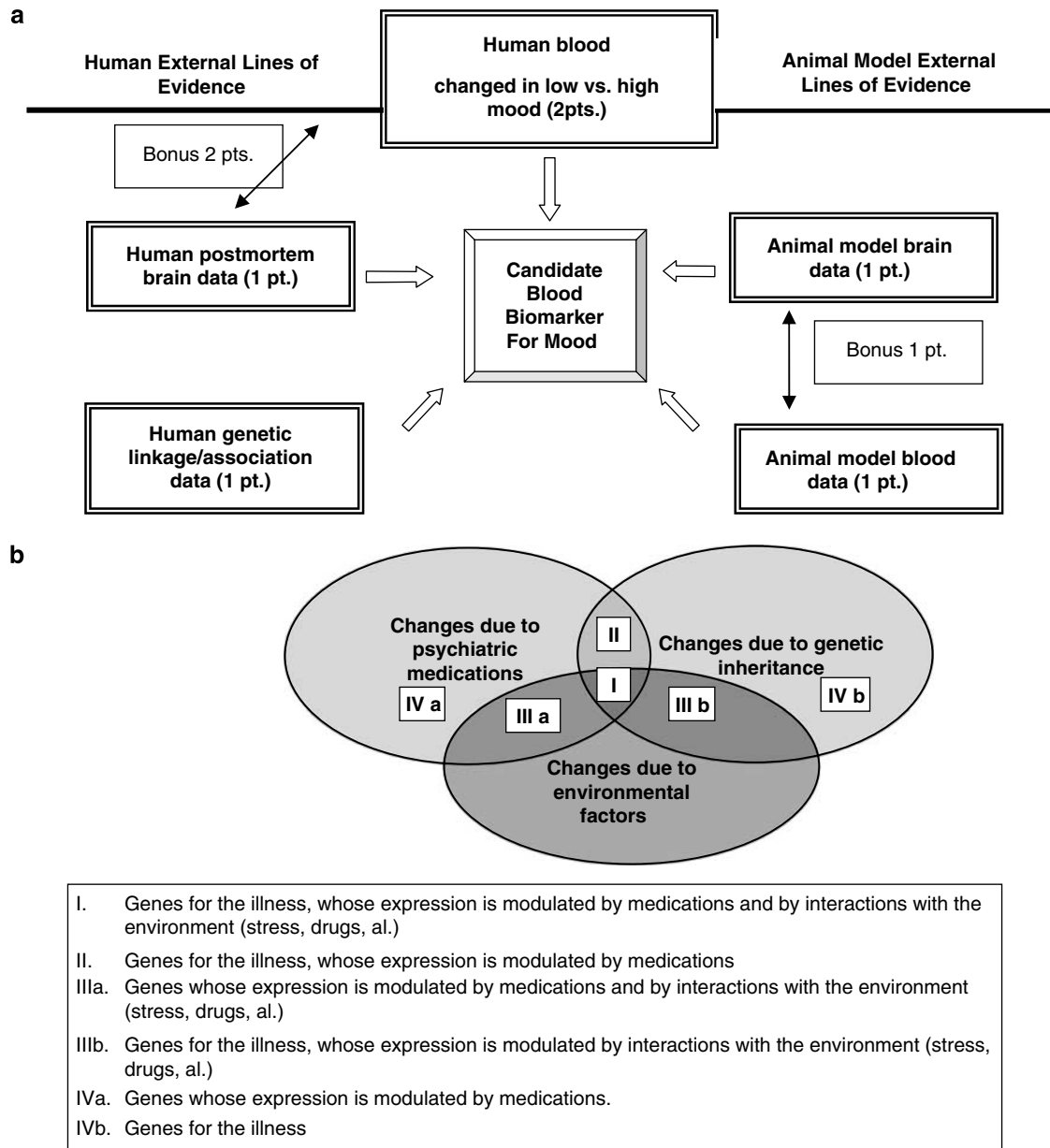
a

b

I.   Genes for the illness, whose expression is modulated by medications and by interactions with the environment (stress, drugs, al.)

II. Genes for the illness, whose expression is modulated by medications

IIIa. Genes whose expression is modulated by medications and by interactions with the environment (stress, drugs, al.)

IIIb. Genes for the illness, whose expression is modulated by interactions with the environment (stress, drugs, al.)

IVa. Genes whose expression is modulated by medications.

IVb. Genes for the illness

Figure 2 Prioritization (**a**) and conceptualization (**b**) of results. (**a**) Convergent functional genomics approach for candidate biomarker prioritization. Scoring of independent lines of evidence (maximum score = 9 points (pts.)). (**b**) Conceptualization of blood candidate biomarker genes.

'increased' by high mood, and the five biomarkers for low mood were genes 'decreased' by high mood. Our interrogation revealed that sodium phenylbutyrate exerts the most similar effects to high mood, and novobiocin the most similar effects to low mood.

Discussion

General approach: strengths, limitations and caveats
Gene expression changes in specific brain regions and blood from a pharmacogenomic animal model developed in our group[2] were used as cross-validators to help with the identification of potential human blood biomarkers. Our pharmacogenomic mouse model of

relevance to bipolar disorder consisted of treatments with an agonist of the illness/bipolar disorder-mimicking drug (methamphetamine) and an antagonist of the illness/bipolar disorder-treating drug (valproate).[2] In essence, the pharmacogenomic approach is a tool for tagging genes that may have pathophysiological relevance. As an added advantage, some of these genes may be involved in potential medication effects present in human blood data (Figure 2). Nevertheless, this simple model, while useful as a gene-hunting tool,[44] does in all likelihood not cover the full spectrum of changes seen in human mood disorders. Other emerging rodent models of bipolar disorder[45–49] may be of interest for future

12

Table 4 BioM-10 mood panel sensitivity and specificity for predicting mood state

	Sensitivity (%)	Specificity (%)
A. Primary bipolar cohort		
High mood	84.6	68.8
Low mood	76.9	81.3
B. Other psychotic disorders cohort		
High mood	71.4	62.5
Low mood	66.7	61.9
C. Secondary bipolar cohort		
High mood	70	66.7
Low mood	66.7	61.5

cross-comparisons and mining of our data sets in a convergent fashion.

Human whole-blood gene expression studies were initially carried out in a primary cohort of bipolar subjects. We have used whole blood as a way of minimizing potential artifacts related to sample handling and separation of individual cell types, and also as a way of having a streamlined approach that lends itself well to scalability and future large-scale studies in the field. Genes differentially expressed in low vs high mood subjects were compared with (1) the results of our animal model brain and blood data, as well as (2) published human genetic linkage/association data and (3) human postmortem brain data, as a way of cross-validating our findings, prioritizing them and coming up with a short list of high-probability candidate biomarker genes (Figures 2a and 5).

We have used a focused approach looking at a discrete quantitative phenotypic item (phene), in our case a VAS Mood. This approach avoids the issue of corrections for multiple comparisons that would arise if we were to look in a discovery fashion at multiple phenes in a comprehensive phenotypic battery (Pheno-Chipping)[50] changed in relationship with all genes on a GeneChip microarray. Larger sample cohorts would be needed for the latter approach.

A panel of top candidate biomarker genes for mood state identified by our approach was then used to generate a prediction score for mood state (low vs high mood). This prediction score was compared to the actual self-reported mood scores from bipolar subjects in the primary cohort (Figure 3a). We also examined this panel of mood biomarkers and prediction score in a separate independent cohort of psychotic disorders patients for which we have gene expression data and mood state data (Figure 3b), as well as in a second independent bipolar cohort (Figure 3c).

Our sample size for human subjects ($n = 29$ for the primary bipolar cohort, $n = 30$ for the psychotic disorders cohort and $n = 19$ for the secondary bipolar cohort) is relatively small, but comparable to the size

of cohorts for human postmortem brain gene expression studies.[51–53] 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 intrasubject measures when the subject is in different mood states, which is an area of future interest and work.

It is to be noted that our experimental approach for detecting gene expression changes relies on a single methodology, Affymetrix GeneChip oligonucleotide microarrays. It is possible that some of the gene expression changes detected from a single biological experiment, with a one-time assay with this technology, are biological or technical artifacts. With that in mind, we have designed our analyses to minimize the likelihood of having false positives, even at the expense of potentially having false negatives, due to the high cost in time and resources of pursuing false leads. For the animal model work, using isogenic mouse strain affords us an ideal control baseline of saline-injected animals for our drug-injected animals. We performed three independent *de novo* biological experiments, at different times, with different batches of mice. This overall design is geared to factor out both biological variability and technical variability. It is to be noted that the concordance between reproducible microarray experiments using the latest generations of oligonucleotide microarrays and other methodologies such as quantitative PCR, with their own attendant technical limitations, is estimated to be over 90%.[54] For the human blood samples differential gene expression analyses, which are the results of single biological experiments, it has to be noted that our approach used a very restrictive and technically robust, all or nothing induction of gene expression (change from Absent Call to Present Call). It is possible that not all biomarker genes for mood will show this complete induction related to state, but rather some will show modulation in gene expression levels, and are thus missed by our filtering. Moreover, 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 low vs high mood. 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 2a): (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) does it map to a human genetic linkage locus. 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/nonfunctional upon further evidence in the field, the network is resilient and maintains functionality. The prioritization

a

Subject ID	Diagnosis	Mood Score	MBP	EDG2	FZD3	ATXN1	EDNRB	FGFR1	MAG	PMP22	UGT8	ERBB3	Mood BioM Prediction Score
					High Mood Biomarkers				Low Mood Biomarkers				
174-1173	BP	27	A	A	A	A	A	P	P	P	P	P	0.0
174-1150	BP	31	A	P	A	A	A	A	P	A	A	A	12.5
174-1126	BP	24	A	A	M	A	A	P	P	A	A	P	16.7
174-1055	BP	20	A	A	P	A	A	P	A	P	P	P	25.0
Phchp023v1	BP	39	P	A	A	M	A	P	P	P	P	M	33.3
Phchp027v1	BP	38	A	A	A	A	P	P	A	M	P	A	40.0
174-1112	BP	38	P	P	P	P	A	A	A	P	P	A	50.0
174-1115	BP	40	A	A	A	P	A	P	A	A	A	P	50.0
Phchp028v1	BP	52	P	M	P	A	A	P	P	P	P	P	50.0
174-1171	BP	72	P	A	A	A	M	P	P	A	A	A	50.0
174-1197	BP	20	A	A	A	P	P	P	P	P	A	A	66.7
174-1042	BP	37	A	P	A	A	P	P	A	P	M	M	66.7
174-5002	BP	73	M	A	M	M	M	P	A	A	A	P	75.0
Phchp029v1	BP	22	P	P	P	P	P	P	P	P	P	P	100.0
174-1161	BP	29	A	A	M	A	A	P	P	P	A	P	100.0
Phchp020v1	BP	42	P	P	A	P	A	P	P	P	A	A	100.0
Phchp031v1	BP	47	M	P	P	P	M	P	P	P	A	P	100.0
174-1119	BP	73	A	P	P	A	P	A	P	A	P	A	100.0
174-1107	BP	63	A	P	P	P	P	A	P	A	P	A	150.0
174-1156	BP	72	A	P	A	P	A	A	P	A	P	A	150.0
174-1132	BP	71	M	P	P	P	P	A	A	A	A	A	175.0
174-1137	BP	39	A	P	A	A	A	P	M	A	M	A	200.0
Phchp030v1	BP	61	P	P	P	A	A	A	A	P	A	P	200.0
174-1037	BP	72	A	P	A	P	A	A	A	P	A	A	200.0
174-1130	BP	65	A	P	P	A	P	A	A	M	A	A	450.0
174-5001	BP	66	P	P	P	P	M	P	A	A	A	A	600.0
174-1193	BP	84	P	P	P	P	P	M	A	A	A	A	1000.0
174-1160	BP	72	M	P	A	P	P	P	P	A	A	A	Infinity
Phchp020v2	BP	80	P	P	P	A	P	A	A	A	A	A	Infinity

Figure 3 (**a**) Comparison of BioM-10 mood prediction score and actual mood scores in the primary cohort of bipolar subjects (n = 29). BP, bipolar. For mood scores: blue, low mood; red, high mood; white, intermediate mood. Mood scores are based on subject self-report on mood visual analog scale (VAS) administered at 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-10 mood 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 100 and above for high mood. Infinity denominator is 0. (**b**) Comparison of BioM-10 mood prediction score and actual mood scores in an independent cohort of psychotic disorders subjects (n = 30). SZ, schizophrenia; SZA, schizoaffective disorder; SubPD, substance-induced psychosis. (**c**) Comparison of BioM-10 mood prediction score and actual mood scores in a secondary independent cohort of bipolar subjects (n=19).

of candidates is similar conceptually to the Google PageRank algorithm[55]—the more the 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.[44] 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.

A validation of our approach is the fact that our biomarker panel showed sensitivity and specificity, of a comparable nature, in both independent replication cohorts (psychotic disorder cohort and secondary bipolar cohort). Thus, our approach of using a VAS phene[50] reflecting an internal subjective experience of well-being or distress (as opposed to more complex and disease-specific state/trait clinical instruments), and looking at extremes of state combined with robust differential expression based on A/P calls and CFG prioritization, seems to be able to identify state biomarkers for mood, our intended goal, which are, at least in part, independent of specific diagnoses or medications. Nevertheless, a comparison with existing clinical rating scales (Supplementary Figure 1S), actimetry[56] and functional neuroimaging,[57] as well as analysis of biomarker data using such instruments may be of high interest for future work, as a way of delineating state vs trait issues, diagnostic boundaries or lack thereof and informing the design of clinical trials that may incorporate clinical and biomarker measures of response to treatment.

b

	Diagnosis	Mood Score	High Mood biomarkers					Low Mood Biomarkers					Mood BioM Prediction Score
			MBP	EDG2	FZD3	ATXN1	EDNRB	FGFR1	MAG	PMP22	UGT8	ERBB3	
phchp015v1	SubPD	78	A	A	A	A	A	P	M	A	A	A	0.0
phchp033v1	SZA	38	A	A	A	A	P	P	A	P	P	P	25.0
phchp010v1	SZA	50	A	P	A	A	A	A	P	P	P	A	33.3
phchp025v1	SZ	30	A	P	P	A	A	P	P	P	P	A	50.0
phchp003v3	SZ	50	P	P	A	A	A	P	A	P	P	M	57.1
phchp024v1	SZA	50	P	A	P	P	A	P	P	P	P	P	60.0
phchp026v1	SZA	65	P	P	P	A	A	P	P	P	P	P	60.0
phchp006v2	SZA	33	M	P	P	A	A	P	P	P	P	A	62.5
phchp010v3	SZA	66	A	P	P	A	A	P	A	P	P	A	66.7
phchp021v1	SZA	39	P	P	M	A	P	P	P	P	P	P	70.0
phchp006v1	SZA	59	A	P	P	P	A	P	P	A	P	P	75.0
phchp019v1	SubPD	36	A	P	P	P	A	P	A	P	P	P	75.0
phchp022v2	SZ	15	P	P	A	A	P	P	P	P	A	P	75.0
phchp017v2	SZA	86	A	P	A	A	P	M	A	P	A	P	80.0
phchp004v1	SZA	70	P	P	P	P	A	P	P	P	P	A	100.0
phchp008v1	SZ	78	A	P	A	A	P	P	P	A	A	A	100.0
phchp009v1	SZ	55	A	P	P	P	A	P	P	A	P	A	100.0
phchp012v1	SZA	71	A	A	A	A	P	A	A	A	P	A	100.0
phchp013v1	SZA	90	A	A	A	A	P	A	A	A	P	A	100.0
phchp022v1	SZ	67	A	A	A	A	P	P	A	A	A	A	100.0
phchp014v1	SubPD	71	P	P	P	P	P	P	P	P	P	A	125.0
phchp005v1	SZA	81	A	P	A	P	A	A	A	P	A	M	133.3
phchp003v1	SZ	95	A	P	P	A	P	A	A	A	P	P	150.0
phchp010v2	SZA	56	A	P	A	P	P	P	A	A	P	A	150.0
phchp012v2	SZA	73	P	P	P	A	A	P	A	P	A	A	150.0
phchp018v1	SZA	76	P	A	P	P	P	P	P	A	A	A	200.0
phchp021v2	SZA	29	P	P	P	M	P	P	A	A	M	M	225.0
phchp005v2	SZA	18	P	P	P	P	P	A	A	P	P	A	250.0
phchp003v2	SZ	54	A	A	P	A	A	A	A	A	A	A	Infinity
phchp016v1	SZ	38	A	A	A	P	A	A	A	A	A	A	Infinity

c

Subject ID	Diagnosis	Mood Score	High Mood Biomarkers					Low Mood Biomarkers					Mood BioM Prediction Score
			MBP	EDG2	FZD3	ATXN1	EDNRB	FGFR1	MAG	PMP22	UGT8	ERBB3	
174-1232	BP	32	M	A	A	P	A	P	P	P	P	P	30.0
pchp045v1	BP	36	A	A	P	A	A	P	A	P	P	A	33.3
174-1278	BP	24	P	A	A	P	A	P	P	P	P	A	50.0
174-1237	BP	57	A	A	P	P	A	P	A	P	P	M	57.1
174-1216	BP	23	P	A	A	A	P	A	P	P	P	A	66.7
phchp031v2	BP	79	A	P	A	A	P	P	P	P	A	A	66.7
phchp056v1	BP	84	P	P	A	A	A	P	P	P	A	A	66.7
174-1199	BP	53	A	P	P	P	M	P	P	P	P	M	77.8
174-1096	BP	83	A	A	P	P	P	A	M	P	P	P	85.7
174-1203	BP	49	A	A	P	A	A	A	A	A	A	P	100.0
phchp053v1	BP	68	A	P	P	A	P	P	A	P	A	P	100.0
174-1211	BP	75	A	P	P	P	P	P	P	P	P	A	100.0
174-1258	BP	90	P	P	P	A	A	P	P	P	A	A	100.0
phchp039v1	BP	11	P	A	P	P	A	P	A	P	A	P	133.3
174-1204	BP	81	P	A	A	A	P	A	P	M	A	A	133.3
174-1220	BP	82	P	P	A	P	P	P	A	P	P	A	133.3
phchp023v2	BP	20	P	P	P	P	P	P	A	P	A	P	166.7
174-5006	BP	66	P	P	P	A	M	P	P	A	A	A	175.0
174-1255	BP	81	P	P	A	A	P	A	A	P	A	A	300.0

Figure 3 Continued.

rank	instance id	cmap name	batch	dose	cell line	score	up	down
1	341	sodium phenylbutyrate	31	100 ÂμM	MCF7	1	0.512	-0.553
453	499	novobiocin	70	100 ÂμM	ssMCF7	-1	-0.5	0.651

Figure 4 Connectivity Map interrogation of drugs that have similar gene expression signatures to that of high mood. A score of 1 indicates a maximal similarity with the gene expression effects of high mood, and −1 indicates a maximal opposite effects to high mood.
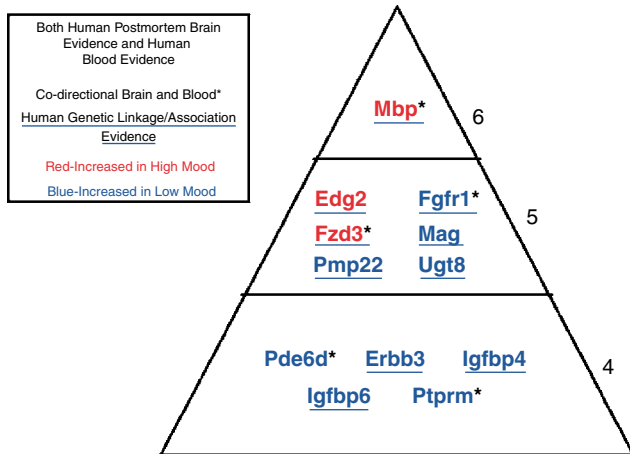
Figure 5 Top blood candidate biomarker genes for mood state. The lines of evidence scoring are depicted on the right side of the pyramid.

Human blood gene expression changes may be influenced by the presence or absence of both medications and drugs of abuse. While we have access to the subject's medical records and clinical information as part of the informed consent for the study, medication noncompliance, on the one hand, and substance abuse, on the other hand, are not infrequent occurrences in psychiatric patients. That medications and drugs of abuse may have effects on mood state and gene expression is in fact being partially modeled in our mouse pharmacogenomic model, with valproate and methamphetamine treatments, respectively. In the end, it is the association of blood biomarkers with mood state that is the primary goal of this study, regardless of the proximal causes, which could be diverse (Figure 2b), and will need to be the subject of subsequent hypothesis-driven studies beyond the scope of this initial report. Future studies are also needed to look at these candidate biomarkers at a protein level, in larger cohorts of both genders, in different age groups and in theragnostic settings measuring responses to specific treatments/medications.

Findings
Topping our list of candidate biomarker genes we have five genes involved in myelination (*Mbp*, *Edg2*, *Mag*, *Pmp22* and *Ugt8*), six genes involved in growth factor signaling (*Fgfr1*, *Fzd3*, *Erbb3*, *Igfbp4*, *Igfbp6* and *Ptprm*) and one gene involved in light transduction (*Pde6d*). These genes were selected as having a

line of evidence (CFG) score of 4 or higher (Table 3). That means, in addition to our human blood data, these genes have at least two other independent lines of evidence implicating them in mood disorders and/or concordance of expression in human brain and blood. Using this cutoff score, we ended up with 12 top genes (Figure 5), all of which have prior evidence of differential expression in human postmortem brains from mood disorder patients.

It is intriguing that genes that have a well-established role in brain functioning should show changes in blood in relationship to psychiatric symptoms state (Figure 5, Table 3; Supplementary Table 3S), and moreover that the direction of change should be concordant with that reported in human postmortem brain studies. It is possible that trait-promoter sequence mutations or epigenetic modifications influence expression in both tissues (brain and blood), and that state-dependent transcription factor changes that modulate the expression of these genes may be contributory as well. This clearly is an area that deserves attention and mechanistic elucidation by future hypothesis-driven research, and may point to the need for focusing research efforts not only on coding regions of genes of interest in psychiatric genetics, but also on their promoter regions as well.

Our top findings suggest that genes involved in brain infrastructure changes (myelin, growth factors) are prominent players in mood disorders, and are reflected in the blood profile, consistent with previous work in the field implicating neurotrophic mechanisms in mood disorders.[58–60] Myelin abnormalities have emerged as a common, if perhaps nonspecific, denominator across a variety of neuropsychiatric disorders.[2,14,44,61–66] Of note, *Mbp*, our top-scoring candidate biomarker (Figure 5), associated with high mood state, was also identified as a top candidate gene for bipolar disorder by a recent whole-genome association study.[15] Our findings regarding insulin growth factor signaling changes may provide an underpinning for the comorbidity with diabetes and metabolic syndrome often encountered in mood disorder patients. Whether these changes are etiopathogenic, compensatory mechanisms, side effects of medications or results of illness-induced lifestyle changes (Figure 2b) is an intriguing area for future research.

The fact that most of the top genes identified are associated with a low mood state (depression) as opposed to high mood state (Figure 5; Table 3) suggests that depression may have more of an impact on blood gene expression changes, perhaps

through a neuroendocrine–immunological axis,[67] as part of a whole-body reaction to a perceived hostile environment.

Of note, some of the other top candidate biomarker genes identified by our human blood work have no previous evidence for involvement in mood disorders other than our mapping them to bipolar genetic linkage loci (Table 3), and thus constitute novel candidate genes for bipolar disorder and depression. They merit further exploration in genetic candidate gene association studies, as well as comparison with emerging results from whole-genome association studies of bipolar disorder and depression. 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 mood will be refined or changed for different subpopulations. 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 the complete list of candidate biomarker genes ($n = 661$) identified by us using the low-threshold analysis (Supplementary Table 2S).

The interrogation of the MIT/Broad Institute Connectivity Map[43] with a signature query composed of the genes in our BioM-10 Mood panel of top biomarkers for low and high mood revealed that sodium phenylbutyrate exerts the most similar effects to high mood, and novobiocin the most similar effects to low mood (Figure 4). Sodium phenylbutyrate is a medication used to treat hyperammonemia that also has histone deacetylase (HDAC) properties, cell survival and antiapoptotic effects.[68] The mood-stabilizer drug valproate, also an HDAC inhibitor, as well as sodium phenylbutyrate and another HDAC inhibitor, trichostatin A, were shown to induce α-synuclein in neurons through inhibition of HDAC and that this α-synuclein induction was critically involved in neuroprotection against glutamate excitotoxicity.[69] Human postmortem brain studies,[52,70] as well as animal model[71] and clinical studies[72] have implicated glutamate abnormalities and HDAC modulation as therapeutic targets in mood disorders. Novobiocin is an antibiotic drug that also has antitumor activity and apoptosis-inducing properties, through Hsp90 inhibition of Akt kinase,[73,74] an effect opposite to that of the valproate, trichostatin A and sodium phenylbutyrate.[75]

This connectivity map analysis with our BioM-10 Mood panel genes provides an interesting external biological cross-validation for the internal consistency of our biomarker approach, as well as illustrates the utility of the Connectivity Map for nonhypothesis-driven identification of novel drug treatments and interventions.

Overall, these results, taken together with our top candidate biomarker genes results discussed above, and with the results of a biological pathway analysis (Table 5), are consistent with a trophicity model[76,77] for genes involved in mood regulation: cell survival and proliferation associated with high mood, and cell shrinkage and death associated with low mood. This perspective is of speculative evolutionary interest and pragmatic clinical importance. Speculatively, nature may have selected primitive cellular mechanisms for analogous higher organism level functions: survival and expansion in favorable, mood-elevating environments, withdrawal and death (apoptosis) in unfavorable, depressogenic environments. In this view, suicide is the organismal equivalent of cellular apoptosis (programmed cell death). Pragmatically, our results point to a previously underappreciated potential molecular and therapeutic overlap between two broad areas of medicine: mood disorders and cancer. This overlap has implications for the clinical comorbidity of mood disorders and cancer, as well as for empirical studies to evaluate the use of mood-regulating drugs in cancer, and of cancer drugs in mood disorders.

Lastly, in clinical practice there is a high degree of overlap and comorbidity between mood disorders, psychosis and substance abuse. Our results in bipolar and psychotic disorder cohorts point to the issue of heterogeneity, overlap and interdependence of major psychiatric syndromes as currently defined by Diagnostic and Statistical Manual of Mental Disorders,[78] and the need for a move toward comprehensive empirical profiling and away from categorical diagnostic classifications.[50]

Conclusions

There are to date no clinical laboratory blood tests for mood disorders. We propose, and provide proof of principle for, a translational convergent approach to help identify and prioritize blood biomarkers of mood state. Our preliminary studies suggest that blood biomarkers have the potential to offer an unexpectedly informative window into brain functioning and

Table 5 Biological roles

Cell death	1.43E-07–4.54E-02
Nervous system development and function	8.31E-07–4.63E-02
Cell morphology	2.25E-05–4.63E-02
Cellular assembly and organization	4.48E-05–4.56E-02
Neurological disease	7.46E-05–4.63E-02
Cellular growth and proliferation	1.11E-04–4.89E-02
Skeletal and muscular system development and function	1.11E-04–3.83E-02
Tissue morphology	1.12E-04–4.09E-02
Behavior	2.08E-04–4.63E-02
Digestive system development and function	2.08E-04–4.63E-02
Cellular development	2.86E-04–4.63E-02
Cancer	5.50E-04–4.89E-02

Ingenuity pathway analysis (IPA) of biological functions categories among our top blood candidate biomarker genes for mood. Genes from Table 3 ($n = 36$). Top categories, overrepresented with a significance of $P < 0.005$, are shown.

disease state. Panels of such biomarkers may serve as a basis for objective clinical laboratory tests, a long-standing Holy Grail for psychiatry. Biomarker-based tests may help with early intervention and prevention efforts, as well as monitoring response to various treatments. In conjunction with other clinical information, such tests will play an important part of personalizing treatment to increase effectiveness and avoid adverse reactions—personalized medicine in psychiatry. Moreover, they may be of scientific use in combination with imaging studies (imaging genomics[79]), and will be of use to pharmaceutical companies engaged in new neuropsychiatric drug development efforts, at both a preclinical and clinical (phases I, II and III) stages of the process.

Acknowledgments

References

1 Niculescu III AB, Segal DS, Kuczenski R, Barrett T, Hauger RL, Kelsoe JR. 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; **144**: 129–158.

5 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.

6 Tsuang MT, Nossova N, Yager T, Tsuang MM, Guo SC, Shyu KG *et al.* Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report. *Am J Med Genet B Neuropsychiatr Genet* 2005; **133**: 1–5.

7 Segman RH, Shefi N, Goltser-Dubner T, Friedman N, Kaminski N, Shalev AY. Peripheral blood mononuclear cell gene expression profiles identify emergent post-traumatic stress disorder among trauma survivors. *Mol Psychiatry* 2005; **10**: 500–513,425.

8 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.

9 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.

10 Sullivan PF, Fan C, Perou CM. Evaluating the comparability of gene expression in blood and brain. *Am J Med Genet B Neuropsychiatr Genet* 2006; **141**: 261–268.

11 Gladkevich A, Kauffman HF, Korf J. Lymphocytes as a neural probe: potential for studying psychiatric disorders. *Prog Neuropsychopharmacol Biol Psychiatry* 2004; **28**: 559–576.

12 Emamian ES, Hall D, Birnbaum MJ, Karayiorgou M, Gogos JA. Convergent evidence for impaired AKT1-GSK3beta signaling in schizophrenia. *Nat Genet* 2004; **36**: 131–137.

13 Chambers JS, Perrone-Bizzozero NI. Altered myelination of the hippocampal formation in subjects with schizophrenia and bipolar disorder. *Neurochem Res* 2004; **29**: 2293–2302.

14 Tkachev D, Mimmack ML, Ryan MM, Wayland M, Freeman T, Jones PB *et al.* Oligodendrocyte dysfunction in schizophrenia and bipolar disorder. *Lancet* 2003; **362**: 798–805.

15 Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B *et al.* A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Mol Psychiatry* 2007; **13**: 197–207.

16 Schulze TG, Chen YS, Badner JA, McInnis MG, DePaulo Jr JR, McMahon FJ. Additional, physically ordered markers increase linkage signal for bipolar disorder on chromosome 18q22. *Biol Psychiatry* 2003; **53**: 239–243.

17 Maziade M, Roy MA, Chagnon YC, Cliche D, Fournier JP, Montgrain N *et al.* Shared and specific susceptibility loci for schizophrenia and bipolar disorder: a dense genome scan in Eastern Quebec families. *Mol Psychiatry* 2005; **10**: 486–499.

18 Aston C, Jiang L, Sokolov BP. Transcriptional profiling reveals evidence for signaling and oligodendroglial abnormalities in the temporal cortex from patients with major depressive disorder. *Mol Psychiatry* 2005; **10**: 309–322.

19 Jurata LW, Bukhman YV, Charles V, Capriglione F, Bullard J, Lemire AL *et al.* Comparison of microarray-based mRNA profiling technologies for identification of psychiatric disease and drug signatures. *J Neurosci Methods* 2004; **138**: 173–188.

20 Badenhop RF, Moses MJ, Scimone A, Mitchell PB, Ewen-White KR, Rosso A *et al.* A genome screen of 13 bipolar affective disorder pedigrees provides evidence for susceptibility loci on chromosome 3 as well as chromosomes 9, 13 and 19. *Mol Psychiatry* 2002; **7**: 851–859.

21 Gaughran F, Payne J, Sedgwick PM, Cotter D, Berry M. Hippocampal FGF-2 and FGFR1 mRNA expression in major depression, schizophrenia and bipolar disorder. *Brain Res Bull* 2006; **70**: 221–227.

22 Park N, Juo SH, Cheng R, Liu J, Loth JE, Lilliston B *et al.* Linkage analysis of psychosis in bipolar pedigrees suggests novel putative loci for bipolar disorder and shared susceptibility with schizophrenia. *Mol Psychiatry* 2004; **9**: 1091–1099.

23 Cichon S, Schumacher J, Muller DJ, Hurter M, Windemuth C, Strauch K *et al.* A genome screen for genes predisposing to bipolar affective disorder detects a new susceptibility locus on 8q. *Hum Mol Genet* 2001; **10**: 2933–2944.

24 Ryan MM, Lockstone HE, Huffaker SJ, Wayland MT, Webster MJ, Bahn S. Gene expression analysis of bipolar disorder reveals downregulation of the ubiquitin cycle and alterations in synaptic genes. *Mol Psychiatry* 2006; **11**: 965–978.

25 Zubenko GS, Hughes HB, Stiffler JS, Zubenko WN, Kaplan BB. Genome survey for susceptibility loci for recurrent, early-onset

18

major depression: results at 10 cM resolution. *Am J Med Genet* 2002; **114**: 413–422.

26 Liu J, Juo SH, Dewan A, Grunn A, Tong X, Brito M et al. Evidence for a putative bipolar disorder locus on 2p13–16 and other potential loci on 4q31, 7q34, 8q13, 9q31, 10q21–24, 13q32, 14q21 and 17q11–12. *Mol Psychiatry* 2003; **8**: 333–342.

27 Lambert D, Middle F, Hamshere ML, Segurado R, Raybould R, Corvin A et al. Stage 2 of the Wellcome Trust UK–Irish bipolar affective disorder sibling-pair genome screen: evidence for linkage on chromosomes 6q16-q21, 4q12-q21, 9p21, 10p14-p12 and 18q22. *Mol Psychiatry* 2005; **10**: 831–841.

28 Aston C, Jiang L, Sokolov BP. Microarray analysis of postmortem temporal cortex from patients with schizophrenia. *J Neurosci Res* 2004; **77**: 858–866.

29 Nakatani N, Hattori E, Ohnishi T, Dean B, Iwayama Y, Matsumoto I et al. Genome-wide expression analysis detects eight genes with robust alterations specific to bipolar I disorder: relevance to neuronal network perturbation. *Hum Mol Genet* 2006; **15**: 1949–1962.

30 Dick DM, Foroud T, Flury L, Bowman ES, Miller MJ, Rau NL et al. Genomewide linkage analyses of bipolar disorder: a new sample of 250 pedigrees from the National Institute of Mental Health Genetics Initiative. *Am J Hum Genet* 2003; **73**: 107–114.

31 Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; **447**: 661–678.

32 Morissette J, Villeneuve A, Bordeleau L, Rochette D, Laberge C, Gagne B et al. Genome-wide search for linkage of bipolar affective disorders in a very large pedigree derived from a homogeneous population in Quebec points to a locus of major effect on chromosome 12q23–q24. *Am J Med Genet* 1999; **88**: 567–587.

33 Schulze TG, Buervenich S, Badner JA, Steele CJ, Detera-Wadleigh SD, Dick D et al. Loci on chromosomes 6q and 6p interact to increase susceptibility to bipolar affective disorder in the National Institute of Mental Health Genetics Initiative pedigrees. *Biol Psychiatry* 2004; **56**: 18–23.

34 McQueen MB, Devlin B, Faraone SV, Nimgaonkar VL, Sklar P, Smoller JW et al. Combined analysis from eleven linkage studies of bipolar disorder provides strong evidence of susceptibility loci on chromosomes 6q and 8q. *Am J Hum Genet* 2005; **77**: 582–595.

35 Rice JP, Goate A, Williams JT, Bierut L, Dorr D, Wu W et al. Initial genome scan of the NIMH genetics initiative bipolar pedigrees: chromosomes 1, 6, 8, 10, and 12. *Am J Med Genet* 1997; **74**: 247–253.

36 Nurnberger Jr JI, Foroud T, Flury L, Su J, Meyer ET, Hu K et al. Evidence for a locus on chromosome 1 that influences vulnerability to alcoholism and affective disorder. *Am J Psychiatry* 2001; **158**: 718–724.

37 Potash JB, Zandi PP, Willour VL, Lan TH, Huo Y, Avramopoulos D et al. Suggestive linkage to chromosomal regions 13q31 and 22q12 in families with psychotic bipolar disorder. *Am J Psychiatry* 2003; **160**: 680–686.

38 Segurado R, Detera-Wadleigh SD, Levinson DF, Lewis CM, Gill M, Nurnberger Jr JI et al. Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: bipolar disorder. *Am J Hum Genet* 2003; **73**: 49–62.

39 Camp NJ, Lowry MR, Richards RL, Plenk AM, Carter C, Hensel CH et al. Genome-wide linkage analyses of extended Utah pedigrees identifies loci that influence recurrent, early-onset major depression and anxiety disorders. *Am J Med Genet B Neuropsychiatr Genet* 2005; **135**: 85–93.

40 Etain B, Mathieu F, Rietschel M, Maier W, Albus M, McKeon P et al. Genome-wide scan for genes involved in bipolar affective disorder in 70 European families ascertained through a bipolar type I early-onset proband: supportive evidence for linkage at 3p14. *Mol Psychiatry* 2006; **11**: 685–694.

41 Radhakrishna U, Senol S, Herken H, Gucuyener K, Gehrig C, Blouin JL et al. An apparently dominant bipolar affective disorder (BPAD) locus on chromosome 20p11.2–q11.2 in a large Turkish pedigree. *Eur J Hum Genet* 2001; **9**: 39–44.

42 Matigian N, Windus L, Smith H, Filippich C, Pantelis C, McGrath J et al. Expression profiling in monozygotic twins discordant for bipolar disorder reveals dysregulation of the WNT signalling pathway. *Mol Psychiatry* 2007; **12**: 815–825.

43 Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ et al. The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006; **313**: 1929–1935.

44 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.

45 Kasahara T, Kubota M, Miyauchi T, Noda Y, Mouri A, Nabeshima T et al. Mice with neuron-specific accumulation of mitochondrial DNA mutations show mood disorder-like phenotypes. *Mol Psychiatry* 2006; **11**: 577–593,523.

46 Prickaerts J, Moechars D, Cryns K, Lenaerts I, van Craenendonck H, Goris I et al. Transgenic mice overexpressing glycogen synthase kinase 3beta: a putative model of hyperactivity and mania. *J Neurosci* 2006; **26**: 9022–9029.

47 Roybal K, Theobold D, Graham A, Dinieri JA, Russo SJ, Krishnan V et al. From the cover: mania-like behavior induced by disruption of CLOCK. *Proc Natl Acad Sci USA* 2007; **104**: 6406–6411.

48 Kato T, Kubota M, Kasahara T. Animal models of bipolar disorder. *Neurosci Biobehav Rev* 2007; **31**: 832–842.

49 Le-Niculescu H, McFarland MJ, Ogden CA, Balaraman Y, Patel S, Tan J et al. Phenomic, Convergent Functional Genomic, and biomarker studies in a stress-reactive genetic animal model of bipolar disorder and co-morbid alcoholism. *Am J Med Genet B Neuropsychiatr Genet* 2008; e-pub ahead of print 4 February 2008.

50 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.

51 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.

52 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.

53 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**: 615,663–79.

54 Quackenbush J. Genomics. Microarrays—guilt by association. *Science* 2003; **302**: 240–241.

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

56 Benedetti F, Radaelli D, Bernasconi A, Dallaspezia S, Falini A, Scotti G et al. Clock genes beyond the clock: CLOCK genotype biases neural correlates of moral valence decision in depressed patients. *Genes Brain Behav* 2008; **7**: 20–25.

57 Hariri AR, Weinberger DR. Imaging genomics. *Br Med Bull* 2003; **65**: 259–270.

58 Duman RS, Monteggia LM. A neurotrophic model for stress-related mood disorders. *Biol Psychiatry* 2006; **59**: 1116–1127.

59 Agid Y, Buzsaki G, Diamond DM, Frackowiak R, Giedd J, Girault JA et al. How can drug discovery for psychiatric disorders be improved? *Nat Rev Drug Discov* 2007; **6**: 189–201.

60 Shaltiel G, Chen G, Manji HK. Neurotrophic signaling cascades in the pathophysiology and treatment of bipolar disorder. *Curr Opin Pharmacol* 2007; **7**: 22–26.

61 Davis KL, Stewart DG, Friedman JI, Buchsbaum M, Harvey PD, Hof PR et al. White matter changes in schizophrenia: evidence for myelin-related dysfunction. *Arch Gen Psychiatry* 2003; **60**: 443–456.

62 Lewohl JM, Wixey J, Harper CG, Dodd PR. Expression of MBP, PLP, MAG, CNP, and GFAP in the human alcoholic brain. *Alcohol Clin Exp Res* 2005; **29**: 1698–1705.

63 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.

64 Haroutunian V, Katsel P, Dracheva S, Stewart DG, Davis KL. Variations in oligodendrocyte-related gene expression across multiple cortical regions: implications for the pathophysiology of schizophrenia. *Int J Neuropsychopharmacol* 2007; 1–9.

65 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.

66 Sokolov BP. Oligodendroglial abnormalities in schizophrenia, mood disorders and substance abuse. Comorbidity, shared traits, or molecular phenocopies? *Int J Neuropsychopharmacol* 2007; **10**: 547–555.

67 Kling MA, Alesci S, Csako G, Costello R, Luckenbaugh DA, Bonne O *et al.* Sustained low-grade pro-inflammatory state in unmedicated, remitted women with major depressive disorder as evidenced by elevated serum levels of the acute phase proteins C-reactive protein and serum amyloid A. *Biol Psychiatry* 2007; **62**: 309–313.

68 Qi X, Hosoi T, Okuma Y, Kaneko M, Nomura Y. Sodium 4-phenylbutyrate protects against cerebral ischemic injury. *Mol Pharmacol* 2004; **66**: 899–908.

69 Leng Y, Chuang DM. Endogenous alpha-synuclein is induced by valproic acid through histone deacetylase inhibition and participates in neuroprotection against glutamate-induced excitotoxicity. *J Neurosci* 2006; **26**: 7502–7512.

70 McCullumsmith RE, Kristiansen LV, Beneyto M, Scarr E, Dean B, Meador-Woodruff JH. Decreased NR1, NR2A, and SAP102 transcript expression in the hippocampus in bipolar disorder. *Brain Res* 2007; **1127**: 108–118.

71 Einat H, Manji HK. Cellular plasticity cascades: genes-to-behavior pathways in animal models of bipolar disorder. *Biol Psychiatry* 2006; **59**: 1160–1171.

72 Zarate Jr CA, Singh J, Manji HK. Cellular plasticity cascades: targets for the development of novel therapeutics for bipolar disorder. *Biol Psychiatry* 2006; **59**: 1006–1020.

73 Haendeler J, Hoffmann J, Rahman S, Zeiher AM, Dimmeler S. Regulation of telomerase activity and anti-apoptotic function by protein–protein interaction and phosphorylation. *FEBS Lett* 2003; **536**: 180–186.

74 Neckers L. Using natural product inhibitors to validate Hsp90 as a molecular target in cancer. *Curr Top Med Chem* 2006; **6**: 1163–1171.

75 De Sarno P, Li X, Jope RS. Regulation of Akt and glycogen synthase kinase-3 beta phosphorylation by sodium valproate and lithium. *Neuropharmacology* 2002; **43**: 1158–1164.

76 Niculescu AB. Genomic studies of mood disorders—the brain as a muscle? *Genome Biol* 2005; **6**: 215.

77 Evans SJ, Choudary PV, Neal CR, Li JZ, Vawter MP, Tomita H *et al.* Dysregulation of the fibroblast growth factor system in major depression. *Proc Natl Acad Sci USA* 2004; **101**: 15506–15511.

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

79 Goldberg TE, Weinberger DR. Genes and the parsing of cognitive processes. *Trends Cogn Sci* 2004; **8**: 325–335.

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