ARCHIVAL REPORT

Common Genetic Variants and Gene-Expression Changes Associated with Bipolar Disorder Are Over-Represented in Brain Signaling Pathway Genes

Inti Pedroso, Anbarasu Lourdusamy, Marcella Rietschel, Markus M. Nöthen, Sven Cichon, Peter McGuffin, Ammar Al-Chalabi, Michael R. Barnes, and Gerome Breen

Background: Despite high heritability, the genetic variants influencing bipolar disorder (BD) susceptibility remain largely unknown. Low statistical power to detect the small effect-size alleles believed to underlie much of the genetic risk and possible heterogeneity between cohorts are an increasing concern. Integrative biology approaches might offer advantages over genetic analysis alone by combining different genomic datasets at the higher level of biological processes rather than the level of specific genetic variants or genes. We employed this strategy to identify biological processes involved in BD etiopathology.

Method: Three genome-wide association studies and a brain gene-expression study were combined with the Human Protein Reference Database protein–protein interaction network data. We used bioinformatic analysis to search for biological networks with evidence of association on the basis of enrichment among both genetic and differential-expression associations with BD.

Results: We identified association with gene networks involved in transmission of nerve impulse, Wnt, and Notch signaling. Three features stand out among these genes: 1) they localized to the human postsynaptic density, which is crucial for neuronal function; 2) their mouse knockouts present altered behavioral phenotypes; and 3) some are known targets of the pharmacological treatments for BD.

Conclusions: Genetic and gene-expression associations of BD cluster in discrete regions of the protein–protein interaction network. We found replicated evidence for association for networks involving several interlinked signaling pathways. These genes are promising candidates to generate animal models and pharmacological interventions. Our results demonstrate the potential advantage of integrative biology analyses of BD datasets.

Key Words: Bipolar disorder, depression, GWAS, mania, networks, pathways, postsynaptic density, systems biology

Bipolar disorder (BD) is a chronic and episodic psychiatric illness characterized by extremes of mood ranging from mania to severe depression. Despite a convincing and substantial genetic contribution to the etiology of the disorder (1), its genetic and molecular underpinnings remain largely unknown. Its diagnosis is based solely on observed clinical features. Individual genome-wide association studies (GWAS) and linkage studies have highlighted several genomic regions, and recently replicated evidence implicating specific loci have also been reported (2–6). The GWAS of common genetic variation have reinforced the notion that many low-risk genetic variants are involved in the etiology of BD. Therefore, an important challenge of genetic studies is to devise analytical strategies to extract biologically relevant associations from those under the genome-wide significance threshold needed for multiple testing correction, $p = 5 \times 10^{-8}$ (7). Currently, large meta-analyses of GWAS represent the major approach used to increase power to detect BD risk alleles (8). A potential limitation to these studies is allelic and locus heterogeneity (i.e., two or more polymorphisms within a gene being independently associated, and different sets of genes associated in different studies). This has been reported in different diseases (9–11). There is some evidence to suggest this might also be true for BD (12), although the extent to which this occurs in BD or other traits is not yet clear.

Prior knowledge can be used to boost signal-to-noise ratio and tackle heterogeneity in large-scale genomic experiments (13,14). Prior information can be used to filter out data, on the basis that they are unlikely to carry useful information, or to aggregate them into biologically relevant groups to allow their signal to stand out above the noise generated by multiple testing. Such an approach can be used with multiple data sources, and increasing evidence suggests that gene-expression studies can help prioritize GWAS results (15–17). For example, Zhong *et al.* (16) showed that gene-expression changes and disease-susceptibility alleles cluster in common biological pathways conferring risk for type 2 diabetes.

Analysis of the Wellcome Trust Case Control Consortium (WTCCC) BD GWAS and its meta-analysis with another GWAS provided evidence of association within biological processes involved in the modulation of transcription and cellular activity, including that of hormone action and adherens junctions (11,18,19).

Here we present an integrative biology analysis aimed to identify biological processes associated with BD susceptibility. Three GWAS of BD susceptibility were integrated via a gene-wide analysis followed by protein–protein interaction network (PPIN) analysis

From the Medical Research Council Social, Genetic and Developmental Psychiatry Centre (IP, AL, PM, MRB, GB); National Institute for Health Research Biomedical Research Centre for Mental Health at South London and Maudsley National Health Service Foundation Trust (IP, AL, PM, GB); Medical Research Council Centre for Neurodegeneration Research (AA-C), Department of Clinical Neuroscience, Institute of Psychiatry, King's College London, London; Computational Biology (MRB), GlaxoSmithKline Pharmaceuticals, Stevenage, Hertfordshire, United Kingdom; Central Institute of Mental Health (MMN, SC), Division of Genetic Epidemiology in Psychiatry, Mannheim; Institute of Human Genetics (MMN, SC), University of Bonn; Department of Genomics (MMN, SC), Life & Brain Center, University of Bonn; German Center for Neurodegenerative Diseases (DZNE) (MMN), Bonn; Institute of Neuroscience and Medicine (INM-1) (SC), Research Center Juelich, Juelich, Germany.

Address correspondence to Gerome Breen, Ph.D., King's College London, Institute of Psychiatry, Medical Research Council Social, Genetic and Developmental Psychiatry Centre, P080 SGDP, De Crespigny Park, London SE5 8AF, UK; E-mail: gerome.breen@kcl.ac.uk.

Received Apr 28, 2011; revised Dec 13, 2011; accepted Dec 15, 2011

0006-3223/$36.00
doi:10.1016/j.biopsych.2011.12.031

and comparison with a brain gene-expression study of BD patients and matched control subjects. Our integrative approach revealed convergent evidence for association of genes and biological processes with BD susceptibility.

Methods and Materials

Samples and Genotype Data

We reanalyzed GWAS of BD from the Wellcome Trust Case Control Consortium (20), Cichon *et al.* (4), and Sklar *et al.* (21) studies, which we refer to as WTCCC, Bonn, and Sklar, respectively. We used individual level genotype and phenotype data from the Bonn study and summary statistics from each of the other studies. Genotype data from the Bonn study were quality controlled by the sample and single nucleotide polymorphism (SNP) missing rate and Hardy-Weinberg. All BD samples met DSM-IV criteria to establish BD diagnoses. We also analyzed summary statistics from six GWAS in common nonpsychiatric disorders reported by the WTCCC (20). See Supplement 1 for additional method descriptions.

Gene-Based Association

We calculated gene-wide *p* values with the FORGE software suite (see Supplement 1 for a detailed description of the software). We included in our analyses approximately 21,000 protein-coding, long noncoding RNA and micro RNA genes annotated in Ensembl v59 (www.ensembl.org) and mapped them to SNPs if the SNP was within 20 kb of the annotated coordinates. The FORGE software combined the *m* association *p* values within genes with the fixed-effects *Z* score method

$$Z_{fix} = \left(\frac{\sum_{i=1}^{m} W_i Z_i}{\sum_{i=1}^{m} W_i} \right) \cdot \sqrt{V_{fix}}$$

where z_i are the *p* values transformed to *Z* scores with the standard normal distribution inverse cumulative distribution function (c.d.f.) and V_{fix} is the variance of Z_{fix}. With the approximation of the multivariate-normal distribution

$$V_{fix} = \sum_{i=1}^{m} \sum_{j=1}^{m} w_i w_j \rho_{ij}$$

where *w* are weights that we set to $1/m$ and ρ_{ij} is the correlation between the z_i and z_j that we approximate as the correlation between the SNP_i and SNP_j, because we only use summary statistics. We used the simulation-based strategy of Liu *et al.* (22) to estimate the significance of the Z_{fix} statistics, because it was shown to provide very good correlation with empirical estimates. Briefly, N gene-wide statistics for each gene were calculated with sets of *m* random Gaussian variables (*Z* scores) with correlations defined by the correlation matrix between the SNPs (see Liu *et al.* [22] for details). We set N to a maximum of 10^6 simulations or until the value of Z_{fix} was observed 10 times. The gene-wide association test significance is equal to $(R+1)/(N+1)$, where R is the number of times a statistic $\geq Z_{fix}$ was observed.

Application of these methods to GWAS has been reported elsewhere (23). Before calculating gene *p* values we applied genomic control to the SNP *p* values, if the study λ median was >1 (24).

Network Analyses

To identify subnetworks of interacting genes enriched with genetic associations, we used the greedy search introduced by Ideker *et al.* (25), which we implemented in a Perl script distributed together with FORGE. The algorithm starts subnetwork searches from each node (seed node) in the PPIN. A subnetwork is defined by sequentially adding the direct neighbors of the nodes of the subnetworks (initially only the seed node). We allowed searches to go to a maximum of five interactions from the seed node and generate subnetworks of 2 to 500 nodes in size. For each dataset, we calculate the aggregate *Z* score of the subnetwork (S_{Net}) with

$$S_{Net} = \left(\frac{\sum_{d=1}^{k} W_i g_i}{\sum_{d=1}^{k} W_i} \right) \cdot \sqrt{V_{SNet}}$$

where g_i is the *Z* score of the *i*th gene in the subnetwork, *k* is the number of genes in the subnetwork, and V_{genes} is the variance-covariance matrix of the statistic of the gene that we calculated with the method described by Luo *et al.* (23),

$$V_{SNet} \approx corr(g_i, g_j) = \frac{\sum_{u=1}^{ki} \sum_{v=1}^{kj} corr(z_{iu}, z_{jv})}{\sqrt{\left(\sum_{u=1}^{ki} \sum_{u=1}^{ki} \rho_{ij}\right) \cdot \left(\sum_{v=1}^{kj} \sum_{v=1}^{kj} \rho_{ij}\right)}}$$

To identify groups of highly overlapping networks, we constructed a gene-to-network membership matrix filled with values 0 or 1, depending on whether a gene was part of the network or not. This matrix was used to calculate correlations between networks and to perform hierarchical clustering with the Heatplus R package (http://www.bioconductor.org/packages/2.3/bioc/html/Heatplus.html).

DNA Microarray Analysis

The gene-expression data of dorsolateral prefrontal cortex (DLPFC) tissue from 61 subjects and orbitofrontal cortex (OFC) tissue from 21 subjects reported by Ryan *et al.* (26) were downloaded from the ArrayExpress database (27) under accession number E-GEOD-5392. Raw intensity values of Affymetrix Human Genome U133A arrays were normalized with the Robust Multi-Average algorithm (28). Pre-filtering removed transcripts not detected (marked as "absent" with MAS5 detection call algorithm) in any sample and were not considered further. The generalized linear model with covariates was used to assess differential expression for each probe in each brain region. Covariates were used as in the original report: for the DLPFC samples we used the generalized linear model with disease status (control/BD) as the main effect while controlling for brain pH and fluphenazine equivalents; and for the OFC samples, we used fluphenazine equivalents as a covariate.

In Silico Characterization of Significant Networks

We interpreted the biology of significant subnetworks with MetaCore (GeneGo; http://www.genego.com). GeneGO provides gene ontologies as GeneGO Pathways Maps and Network Processes manually constructed from literature review. There are defined as: 1) GeneGO Pathways summaries of established, noncontradictory state-of-the-art knowledge on the major functional categories of human metabolism and cell signaling; and 2) Networks Processes descriptive of a biological function but containing more information than a Pathway Map and possibly having newer published results on them. Enrichment of subnetwork genes in these biological categories is calculated with hypergeometric distribution statistics as has been described elsewhere (29). For all analyses in GeneGO, we used the intersection between the PPIN and the three GWAS dataset genes as a background to account for biases in PPIN gene annotation. This reference list had 7924 genes, of which 3375 could be mapped to GeneGO Pathways and 7917 could be mapped to GeneGO Processes. In addition we compiled genes associated with mouse phenotypes by parsing the files provided by the Mammalian Phenotype Ontology database (30). Genes localized in the human postsynaptic density (hPSD) were obtained from the supplementary material of Bayes *et al.* (31). Enrichment of genes for membership of the Mammalian Phenotype Ontology, and the hPSD categories were calculated with

binomial statistics. Enrichment of BD gene expression results in subnetworks was performed with the parametric analysis of gene set enrichment (32).

Meta-Analyses

Gene and network level statistics from different studies were combined with the fixed-effects model previously described for gene-wide testing. We set study specific weights calculated as the square root of the sample size: WTCCC = 70.71, Sklar = 59.16, and Bonn = 44.72. When combining GWAS and gene-expression studies, equal weights were assigned. The Perl script used to perform the meta-analysis is distributed with the FORGE software.

Results

We combined results from three GWASs (4,20,21) and a gene-expression study of BD (26) with information of protein–protein interactions (33). Detailed descriptions of these datasets have been provided elsewhere (4,20,21). All studies used DSM-IV criteria to establish BD diagnosis, and quality control included sample and genetic marker filters to exclude low-quality data. The combined data were meta-analyzed at the gene and gene-set level to identify biological processes enriched with genetic variants and gene-expression changes associated with BD.

Gene-Based Meta-Analysis of BD Datasets

We calculated gene-wide association *p* values for each GWAS with the FORGE software and for the gene-expression studies with standard differential expression analysis (see Methods and Materials). First, the GWAS *p* values of each gene were meta-analyzed with a fixed-effects model, and this resulting *p* value (P_{GWAS}) was then meta-analyzed with the *p* value of that gene's test for differential gene-expression between cases and control subjects (P_{GE}). This was performed separately with the gene-expression results for two brain regions, DLPFC and OFC gene-expression, yielding two GWAS-gene-expression meta-analysis results we refer to as $P_{GWAS\text{-}DLPFC}$ and $P_{GWAS\text{-}OFC}$, respectively. One and eight genes showed significant association (false discovery rate <.05) in the GWAS-DLPFC and GWAS-OFC meta-analyses, respectively (Table 1). All genes, except *TRIM23* with $P_{GWAS} = .2$ and $P_{GE} = 2 \times 10^{-6}$, reached significance supported by both the gene-expression and GWAS data.

Network Analysis of GWAS and Gene-Expression Signals

To identify groups of interacting gene products enriched for BD association signals, we mapped gene-level results from the GWAS and gene-expression studies to the PPIN and tested the association of approximately 6100 subnetworks of between 2 and 200 nodes (products of genes) in size. We performed, as with our gene-level analyses, a combined analysis of the GWAS and gene-expression results at the PPIN subnetwork level. There were 11 subnetworks with false discovery rate < .05 in the GWAS-DLPFC meta-analysis (Table 2). Some of these showed significant overlap, and we selected four subnetworks as representatives (Figure 1). We used gene ontology analysis to characterize the biological function of the genes of the four subnetworks (Table 3). Two subnetworks, PPIN-5572 and PPIN-1576, were best characterized by similar ontologies, including transmission of nerve impulse, gonadotropin-releasing hormone signaling pathway, and Wnt and Notch signaling. The other two subnetworks were more distinctive; PPIN-6001 was enriched for neuropeptide signaling pathways genes, and PPIN-119 was enriched for genes involved in lipid metabolism pathways. We further characterized these networks by analyzing the phenotype of mouse knockdowns of their genes. We found that PPIN-5572 and PPIN-1576 were over-represented in mouse knockdowns showing alteration in phenotypes like behavior, growth, and size and nervous system phenotypes (Table 4). We did not find significant results after correction for multiple testing for PPIN-119 and PPIN-6001. Finally, we also found a significant overlap between these networks and genes reported to code for proteins localized in the human postsynaptic density (31). The PPIN-5572 and PPIN-1576 had the largest overlap, with a 2- and 2.4-fold enrichment, $p = 8 \times 10^{-4}$ and 7×10^{-3}, respectively (Table S5 in Supplement 1).

Our network analysis strategy aimed to deal with potential locus heterogeneity, which has been described in gene-expression studies (e.g., Chuang *et al.* [13]) and diverse diseases studied by GWAS (10,11,19). We found, in common with these previous reports, that the genes driving the network associations (association *p* values < .05) were different in each dataset (Figure S2 and Table S6 in Sup-

Table 1. Significant Genes from Gene-Based Meta-Analysis Between GWAS and Gene-Expression

	GWAS				GWAS + Gene-Expression		
Hugo Symbol	WTCCC (4,30,21)	Sklar (4,30,21)	Bonn (4,30,21)	P_{GWAS}	P_{GE}	$P_{GWAS\text{-}GE}$	$FDR_{GWAS\text{-}GE}$
DLPFC							
LMBR1L	2E-03	8E-03	1E-02	6E-06	.02	4E-06	.03
OFC							
RAB7L1[a]	8E-03	.07	.1	1E-03	4E-04	4E-06	.04
TMEFF1	.07	.5	.03	.04	2E-05	2E-05	.04
ITGA10	3E-03	.02	.4	8E-04	5E-03	2E-05	.04
ITM2C	.3	.05	3E-03	6E-03	1E-03	3E-05	.04
LFNG[b]	.4	5E-03	.1	.01	3E-04	4E-05	.04
CSDA	2E-03	.4	.4	.01	4E-04	4E-05	.04
TRIM23	.4	.3	.2	.2	2E-06	4E-05	.04
HIST1H3F	.4	.02	.03	.01	3E-04	4E-05	.04

Reported are *p* values from individual genome-wide association studies (GWAS) and gene-expression studies and their meta-analysis results. The GWAS meta-analysis *p* value (P_{GWAS}) was combined with the *p* value of differential gene-expression (P_{GE}) to give the pooled GWAS and gene-expression estimate ($P_{GEAS\text{-}GE}$).

DLPFC, dorsolateral prefrontal cortex; $FDR_{GWAS\text{-}GE}$, false discovery rate estimate for $P_{GWAS\text{-}GE}$; OFC, orbitofrontal cortex.

[a]Loci with previous disease mapping results registered in the National Human Genome Research Institute GWAS Catalogue or Online Mendelian Inheritance in Man (OMIM): Parkinson's disease (52,53).

[b]Loci with previous disease mapping results registered in the National Human Genome Research Institute GWAS Catalogue or OMIM: autosomal recessive spondylocostal dysostis 3 (MIM 602576).

Table 2. Significant Gene-Sets in Combined GWAS and Gene-Expression Network Analysis

	GWAS				GWAS + Gene-Expression		
PPIN	WTCCC	Sklar	Bonn	P_{GWAS}	P_{DLPFC}	$P_{GWAS\text{-}DLPFC}$	$FDR_{GWAS\text{-}DLPFC}$
PPIN-5572	.1	4×10^{-3}	6×10^{-3}	4×10^{-4}	3×10^{-3}	1×10^{-5}	8×10^{-3}
PPIN-5042	.2	1×10^{-3}	.1	2×10^{-3}	2×10^{-3}	2×10^{-5}	8×10^{-3}
PPIN-7033	.04	.01	.2	2×10^{-3}	4×10^{-3}	5×10^{-5}	.01
PPIN-1576	.1	4×10^{-3}	.2	3×10^{-3}	1×10^{-3}	3×10^{-5}	.01
PPIN-3274	.4	3×10^{-3}	.02	5×10^{-3}	1×10^{-3}	5×10^{-5}	.01
PPIN-5348	.1	3×10^{-3}	6×10^{-3}	3×10^{-4}	.02	1×10^{-4}	.01
PPIN-7161	.2	.01	.02	3×10^{-3}	3×10^{-3}	6×10^{-5}	.01
PPIN-6001	.04	.1	.03	3×10^{-3}	3×10^{-3}	7×10^{-5}	.01
PPIN-6294	.1	.03	.04	4×10^{-3}	4×10^{-3}	1×10^{-4}	.03
PPIN-2675	.2	.01	.07	7×10^{-3}	2×10^{-3}	1×10^{-4}	.03
PPIN-119	5×10^{-3}	.4	.2	.02	1×10^{-3}	2×10^{-4}	.04

P_{GWAS}, $P_{GWAS\text{-}DLPFC}$, and $FDR_{GWAS\text{-}DLPFC}$ are *p* values obtained by pooling GWAS, GWAS + DLPFC, and FDR applied to the GWAS-DLPFC meta-analysis results, respectively. There were no significant results in the GWAS-OFC meta-analysis.

PPIN, protein–protein interaction network; WTCCC, Wellcome Trust Case Control Consortium; other abbreviations as in Table 1.

plement 1). This reinforces the notion that system level analyses are required to compare and integrate GWAS and gene-expression studies.

Also, we found no association with these four sub-networks in six nonpsychiatric complex diseases, except for PPIN-6001, which showed some association with two immune-system related disorders (Crohn's disease and rheumatoid arthritis) (Table S7 in Supplement 1).

Discussion

Our results show that genetic variants and gene-expression changes associated with BD are not randomly distributed across genes but instead cluster into discrete groups. These can be identified by integrating multiple layers of biological information, and here we used genetic association signals, evidence of direct or indirect interaction in the context of a PPIN, as well as differential expression in brains of BD patients compared with control subjects. The PPIN subnetworks associated were characterized by biological processes of relevance for a neuropsychiatric disorder (e.g., transmission of nerve impulse and Wnt signaling) (Table 3). Interestingly, we also found that these networks overlap significantly with the hPSD recently described by a large proteomic study (Table S5 in Supplement 1). The hPSD is enriched with genes affected by mutations causing neurological, central nervous system, and cognitive phenotypes, such as mental retardation or Alzheimer's disease (31). Therefore, not only rare mutations with large effect-size but also common variants and gene-expression changes seem to affect the hPSD as a mediator of disease risk. Furthermore, the data presented here and by others highlight the potential to shed light on the molecular basis of complex diseases via consideration of the statistical signals normally buried under strict GWAS significance thresholds alongside knowledge of protein function, (e.g., protein-protein interactions) (for examples see references [9–11,34] and gene-expression changes [15–17]). It is likely that these BD associated networks might also harbor risk variants for other psychiatric conditions, because the Wnt signaling and more broadly the hPSD have been linked to diverse conditions, including schizophrenia (31). Additional analyses in genetic and gene-expression samples ascertained for other disorders will be needed to explore the phenotypic specificity of the effects of genetic variants within these molecular networks. Our initial attempt showed no association with nonpsychiatric disorders, except for PPIN-6001 in two immune-system diseases. However, these results remain preliminary, because by analyzing one GWAS and no gene-expression data we anticipate that our statistical power is limited.

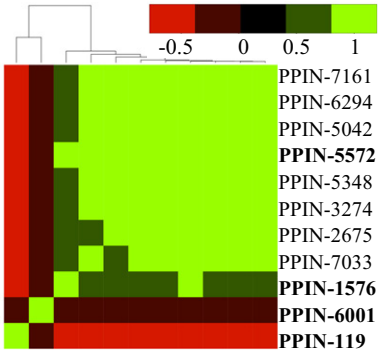

Figure 1. Overlap between significant networks. Hierarchical clustering of networks on the basis of gene overlap. Four networks were selected as representative (names in bold). Correlation between networks is color-coded as indicated in the figure key. PPIN, protein–protein interaction network.

Our results support Wnt signaling as a biological process of relevance to the understanding of BD etiology. There is evidence to suggest that lithium and valproic acid, two molecules used to treat the disorder, exert part of their pharmacological effects through changes in Wnt signaling (35). Previously, a genetic association study analyzed Wnt signaling genes and found some evidence of association with BD (36). Genetic variants within Wnt signaling genes have also been shown to affect gray matter volume in major depression patients (37). Wnt signaling is involved in the regulation of cell proliferation and tissue development across several major human organs, including brain (38). However, it remains challenging to propose a specific mechanism by which Wnt signaling genes influence BD risk, our results overall point to biological process of known relevance for brain function, supporting the notion that many subthreshold genetic associations hold valuable biological information to understand the etiology and propose new treatments. For example, drugs targeting neuropeptide signaling are thought to be promising alternatives in the treatment of mood disorders (39,40).

A major challenge in psychiatric genetics is the development of better animal models. The GWAS might provide a new avenue for their development by pointing to biological process that can be

Table 3. Subnetwork Characterization with Ontology Enrichment

	Biological Process	Enrichment *p* Value	Fraction of Genes
PPIN-5572	Neurophysiological process: transmission of nerve impulse	1×10^{-6}	17/121
	Signal transduction: Wnt signaling	3×10^{-6}	19/126
	Reproduction: GnRH signaling pathway	1×10^{-5}	15/83
PPIN-1576	Reproduction: GnRH signaling pathway	1×10^{-7}	15/83
	Signal transduction: Wnt signaling	4×10^{-7}	19/126
	Signal transduction: Notch signaling	8×10^{-7}	23/174
PPIN-6001	Signal transduction: neuropeptide signaling pathways	4×10^{-14}	23/94
PPIN-119	Regulation of lipid metabolism: regulation of fatty acid synthase activity in hepatocytes	3×10^{-4}	4/9
	Regulation of lipid metabolism: regulation of lipid metabolism via LXR, NF-Y and SREBP	4×10^{-4}	5/22

Enrichment of the subnetwork genes for biological pathways and processes. Each result is denoted by a broad category description followed by a specific category name and the uncorrected *p* value. All results had a false discovery rate <.05.

GnRH, gonadotropin-releasing hormone; LXR, liver X receptor; NF-Y, nuclear factor-Y; PPIN, protein–protein interaction network; SREBP, sterol regulatory element binding protein.

systematically studied to generate new animal models (e.g., with mouse knockdowns). These genetic models can be systematically phenotyped to establish phenotypic commonalities that potentially will translate into underlying processes of relevance for the understanding of human pathology. Our results suggest that this strategy might be fruitful, because the biological process associated in human also associates with phenotypic alteration in the mouse (Table 4). Although laborious, such a systematic screening will be streamlined by new gene targeting technologies, for example as demonstrated in Premsrirut *et al.* (41). This strategy might provide genetically driven models in which to test new drugs and explore pathology at a molecular and physiological level (42,43). Some progress along these lines has been made in BD (44–46). For example, Ogden *et al.* (45) used a convergent functional genomics approach combining gene expression from animal models and human linkage information from human studies to identify new candidate genes for mood disorders. In a follow-up study, Le-Niculescu *et al.* (44) characterized the mouse knockout of one of these candidate genes, which showed phenotypes of potential interest to understand mood disorders.

Several of the biological functions over-represented in the significant PPIN networks are in line with those reported in previous gene set analyses of BD GWAS (18,47). For example, Holmans *et al.* (18) found significant association with GO0005179 ($p < .0001$ in Table 4 of Holmans *et al.* [18]), which is a parent in the Gene Ontology classification of gonadotropin hormone-releasing activity (GO0005183), where we also found association (Table 3). Our gene-based meta-analysis highlighted several new potential loci but did not support previous associations with BD (Table 1). Not much is known about the biological function of our most significant gene, *LMBR1L*, but it has been shown to be a lipocalin receptor (48) and is expressed in many tissues, including the central nervous system (49). Our results did not provide compelling evidence for loci previously associated with BD in large GWASs (2), namely *ANK3* and *CACNA1C*. These genes were not part of the significantly associated networks, but they had nominal levels of association in the GWAS meta-analysis—*CACNA1C* gene $p = .01$, and *ANK3* gene $p = .25$—and gene-expression data; differential expression *p* value in the OFC for *CACNA1C* = .047 and *ANK3* = .047 and in the DLPFC for *CACNA1C* = .054 and *ANK3* = .27. Additional analyses in larger GWAS (e.g., those of the Psychiatric GWAS Consortium) (8) will provide a better synthesis of the association evidence at these loci.

We found, in line with previous reports (10,11,19), that system-level genetic associations in complex traits are heterogeneous (Figure S2 and Table S6 in Supplement 1). However, this heterogeneity is approachable, and GWAS can be mined for the biological processes underlying complex traits, as recently reviewed in Wang *et al.* (50). Although overcoming genetic heterogeneity was a major motivation of our study, we do acknowledge the merit of increasing sample size to improve statistical power; this has been an unmistakably successful approach for some complex traits (e.g., Lango *et al.* [9]). However, in some situations increasing sample size is not an option—for example, in the case of low population frequency diseases. Many nonreplicated signals might represent true disease-susceptibility alleles whose effect is not expressed, due to cryptic differences in genetic background between populations, their exposure to environmental factors, or ascertainment differences between studies, among other possibilities. Despite this, by clustering these weak signals with prior knowledge of biological pathways and networks, it is possible to identify molecular systems underlying complex traits.

A major limitation of our approach and those based on predefined gene-sets (e.g., [10,11,18,19,51]) was the relatively low or biased coverage of the protein interactome, which only included approximately 8000 genes of the approximately 21,000 genes covered by the GWAS analyzed. However, this will improve as more

Table 4. Representative Mouse Phenotypes Enriched for Significant Subnetworks

		PPIN-1576		PPIN-5572	
Phenotype Ontology	Ontology Description	Fold Enrichment	*p*	Fold Enrichment	*p*
MP:0005386	Behavior and neurological phenotype	2.36	1E-11	2.56	2E-09
MP:0005378	Growth and size phenotype	2.01	7E-10	2.05	8E-06
MP:0005376	Homeostasis and metabolism phenotype	1.91	2E-08	1.97	1E-05
MP:0002873	Nervous system phenotype	2.40	2E-07	1.88	8E-05
MP:0003631	Nervous system phenotype	1.93	2E-07	2.39	1E-04

Mammalian Phenotype Ontology (MPO) identifiers are provided for ontology category. Fold enrichment correspond to the enrichment of the networks compared with all genes included in the protein–protein interaction analysis. Reported *p* values are corrected for 120 tests (4 networks × 30 ontology categories) via a Bonferroni correction. There were no significant results for protein–protein interaction network (PPIN)-119 and PPIN-6001.

experiments are annotated in the public databases. In summary, we find a significant and replicable association with PPIN subnetworks involved in transmission of nerve impulse and signaling pathway as neuropeptides, Wnt and Notch with BD. We suggest that these approaches are highly complementary to large meta-analytical studies based on single SNP analyses. Our results strongly suggest that BD genetic research can benefit by integrating GWAS and gene-expression studies. However, our results do require replication in additional genetic and gene-expression samples, and functional studies will be needed to explore the molecular mechanisms mediating the effects of genetic variants within these networks.

This work was supported by the European Community Health Seventh Framework Programme (FP7/2007-2013) (259867 to AAC). AAC thanks the ALS Therapy Alliance, the Angel Fund, the ALS Association and the Motor Neurone Disease Association of Great Britain and Northern Ireland. IP, GB, and AL thank funding from the National Institute for Health Research Biomedical Research Centre for Mental Health at the South London and Maudsley National Health Service Foundation Trust and Institute of Psychiatry, Kings College London. GB also thanks the Medical Research Council (G0701420), the Wellcome Trust, National Alliance for Research on Schizophrenia and Depression, and Biotechnology and Biological Sciences Research Council for funding. The Bonn genome-wide association study was supported by the German Federal Ministry of Education and Research, within the context of the National Genome Research Network 2, the National Genome Research Network plus, and the Integrated Genome Research Network MooDS (Grant 01GS08144 to SC and MMN, Grant 01GS08147 to MR). AAC receives royalties for The Brain (Oneworld Publications) and The Genetics of Complex Human Diseases (Cold Spring Harbor Laboratory Press). MB is a full-time employee of and holds stock in GlaxoSmithKline. All other authors report no biomedical financial interests or potential conflicts of interest.

Supplementary material cited in this article is available online.

1. McGuffin P, Rijsdijk F, Andrew M, Sham P, Katz R, Cardno A (2003): The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. *Arch Gen Psychiatry* 60:497–502.
2. Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, *et al.* (2008): Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet* 40:1056–1058.
3. Lewis CM, Levinson DF, Wise LH, DeLisi LE, Straub RE, Hovatta I, *et al.* (2003): Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet* 73:34–48.
4. Cichon S, Muhleisen TW, Degenhardt FA, Mattheisen M, Miro X, Strohmaier J, *et al.* (2011): Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *Am J Hum Genet* 88:372–381.
5. Breen G, Lewis CM, Vassos E, Pergadia ML, Blackwood DH, Boomsma DI, *et al.* (2011): Replication of association of 3p21.1 with susceptibility to bipolar disorder but not major depression. *Nat Genet* 43:3–5.
6. O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V, *et al.* (2008): Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet* 40:1053–1055.
7. Dudbridge F, Gusnanto A (2008): Estimation of significance thresholds for genomewide association scans. *Genet Epidemiol* 32:227–234.
8. Psychiatric GWAS Consortium (2009): A framework for interpreting genome-wide association studies of psychiatric disorders. *Mol Psychiatry* 14:10–17.
9. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, *et al.* (2010): Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* 467:832–838.
10. Wang K, Zhang H, Kugathasan S, Annese V, Bradfield JP, Russell RK, *et al.* (2009): Diverse genome-wide association studies associate the IL12/IL23 pathway with Crohn Disease. *Am J Hum Genet* 84:399–405.
11. Baranzini SE, Galwey NW, Wang J, Khankhanian P, Lindberg R, Pelletier D, *et al.* (2009): Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. *Hum Mol Genet* 18:2078–2090.
12. Schulze TG, Detera-Wadleigh SD, Akula N, Gupta A, Kassem L, Steele J, *et al.* (2009): Two variants in Ankyrin 3 (ANK3) are independent genetic risk factors for bipolar disorder. *Mol Psychiatry* 14:487–491.
13. Chuang HY, Lee E, Liu YT, Lee D, Ideker T (2007): Network-based classification of breast cancer metastasis. *Mol Syst Biol* 3:140.
14. Ideker T, Dutkowski J, Hood L (2011): Boosting signal-to-noise in complex biology: Prior knowledge is power. *Cell* 144:860–863.
15. Hsu YH, Zillikens MC, Wilson SG, Farber CR, Demissie S, Soranzo N, *et al.* (2010): An integration of genome-wide association study and gene expression profiling to prioritize the discovery of novel susceptibility loci for osteoporosis-related traits. *PLoS Genet* 6:e1000977.
16. Zhong H, Beaulaurier J, Lum PY, Molony C, Yang X, Macneil DJ, *et al.* (2010): Liver and adipose expression associated SNPs are enriched for association to type 2 diabetes. *PLoS Genet* 6:e1000932.
17. Zhong H, Yang X, Kaplan LM, Molony C, Schadt EE (2010): Integrating pathway analysis and genetics of gene expression for genome-wide association studies. *Am J Hum Genet* 86:581–591.
18. Holmans P, Green EK, Pahwa JS, Ferreira MA, Purcell SM, Sklar P, *et al.* (2009): Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. *Am J Hum Genet* 85:13–24.
19. O'Dushlaine C, Kenny E, Heron E, Donohoe G, Gill M, Morris D, *et al.* (2011): Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. *Mol Psychiatry* 16:286–292.
20. Wellcome Trust Case Control Consortium (2007): Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–678.
21. Sklar P, Smoller JW, Fan J, Ferreira MA, Perlis RH, Chambert K, *et al.* (2008): Whole-genome association study of bipolar disorder. *Mol Psychiatry* 13:558–569.
22. Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, *et al.* (2010): A versatile gene-based test for genome-wide association studies. *Am J Hum Genet* 87:139–145.
23. Luo L, Peng G, Zhu Y, Dong H, Amos CI, Xiong M (2010): Genome-wide gene and pathway analysis. *Eur J Hum Genet* 18:1045–1053.
24. Devlin B, Roeder K (1999): Genomic control for association studies. *Biometrics* 55:997–1004.
25. Ideker T, Ozier O, Schwikowski B, Siegel AF (2002): Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* 18(suppl 1):S233–S240.
26. Ryan MM, Lockstone HE, Huffaker SJ, Wayland MT, Webster MJ, Bahn S (2006): Gene expression analysis of bipolar disorder reveals downregulation of the ubiquitin cycle and alterations in synaptic genes. *Mol Psychiatry* 11:965–978.
27. Parkinson H, Kapushesky M, Kolesnikov N, Rustici G, Shojatalab M, Abeygunawardena N, *et al.* (2009): ArrayExpress update—from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Res* 37:D868–D872.
28. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP (2003): Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 31:e15.
29. Ekins S, Bugrim A, Brovold L, Kirillov E, Nikolsky Y, Rakhmatulin E, *et al.* (2006): Algorithms for network analysis in systems-ADME/Tox using the MetaCore and MetaDrug platforms. *Xenobiotica* 36:877–901.
30. Smith CL, Goldsmith CA, Eppig JT (2005): The Mammalian Phenotype Ontology as a tool for annotating, analyzing and comparing phenotypic information. *Genome Biol* 6:R7.
31. Bayes A, van de Lagemaat LN, Collins MO, Croning MD, Whittle IR, Choudhary JS, *et al.* (2011): Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat Neurosci* 14:19–21.
32. Kim SY, Volsky DJ (2005): PAGE: Parametric analysis of gene set enrichment. *BMC Bioinformatics* 6:144.
33. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, *et al.* (2009): Human Protein Reference Database—2009 update. *Nucleic Acids Res* 37:D767–D772.
34. Ruano D, Abecasis GR, Glaser B, Lips ES, Cornelisse LN, de Jong AP, *et al.* (2010): Functional gene group analysis reveals a role of synaptic heterotrimeric G proteins in cognitive ability. *Am J Hum Genet* 86:113–125.

35. Gould TD, Manji HK (2002): The Wnt signaling pathway in bipolar disorder. *Neuroscientist* 8:497–511.
36. Zandi PP, Belmonte PL, Willour VL, Goes FS, Badner JA, Simpson SG, *et al.* (2008): Association study of Wnt signaling pathway genes in bipolar disorder. *Arch Gen Psychiatry* 65:785–793.
37. Inkster B, Nichols TE, Saemann PG, Auer DP, Holsboer F, Muglia P, *et al.* (2010): Pathway-based approaches to imaging genetics association studies: Wnt signaling, GSK3beta substrates and major depression. *Neuroimage* 53:908–917.
38. Logan CY, Nusse R (2004): The Wnt signaling pathway in development and disease. *Annu Rev Cell Dev Biol* 20:781–810.
39. Muller MB, Wurst W (2004): Getting closer to affective disorders: The role of CRH receptor systems. *Trends Mol Med* 10:409–415.
40. Refojo D, Holsboer F (2009): CRH signaling. Molecular specificity for drug targeting in the CNS. *Ann N Y Acad Sci* 1179:106–119.
41. Premsrirut PK, Dow LE, Kim SY, Camiolo M, Malone CD, Miething C, *et al.* (2011): A rapid and scalable system for studying gene function in mice using conditional RNA interference. *Cell* 145:145–158.
42. Geschwind DH, Konopka G (2009): Neuroscience in the era of functional genomics and systems biology. *Nature* 461:908–915.
43. Konopka G, Geschwind DH (2010): Human brain evolution: Harnessing the genomics (r)evolution to link genes, cognition, and behavior. *Neuron* 68:231–244.
44. Le-Niculescu H, McFarland MJ, Ogden CA, Balaraman Y, Patel S, Tan J, *et al.* (2008): 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* 147B: 134–166.
45. Ogden CA, Rich ME, Schork NJ, Paulus MP, Geyer MA, Lohr JB, *et al.* (2004): Candidate genes, pathways and mechanisms for bipolar (manic–depressive) and related disorders: An expanded convergent functional genomics approach. *Mol Psychiatry* 9:1007–1029.
46. Patel SD, Le-Niculescu H, Koller DL, Green SD, Lahiri DK, McMahon FJ, *et al.* (2010): Coming to grips with complex disorders: Genetic risk prediction in bipolar disorder using panels of genes identified through convergent functional genomics. *Am J Med Genet B Neuropsychiatr Genet* 153B:850–877.
47. Heron EA, O'dushlaine C, Segurado R, Gallagher L, Gill M (2011): Exploration of empirical Bayes hierarchical modeling for the analysis of genome-wide association study data. *Biostatistics* 12:445–461.
48. Wojnar P, Lechner M, Merschak P, Redl B (2001): Molecular cloning of a novel lipocalin-1 interacting human cell membrane receptor using phage display. *J Biol Chem* 276:20206–20212.
49. Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, *et al.* (2009): BioGPS: An extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol* 10:R130.
50. Wang K, Li M, Hakonarson H (2010): Analysing biological pathways in genome-wide association studies. *Nat Rev Genet* 11:843–854.
51. Wang K, Li M, Bucan M (2007): Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet* 81:1278–1283.
52. Satake W, Nakabayashi Y, Mizuta I, Hirota Y, Ito C, Kubo M, *et al.* (2009): Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat Genet* 41:1303–1307.
53. Simon-Sanchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, Berg D, *et al.* (2009): Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat Genet* 41:1308–1312.