Convergent Functional Genomics of Psychiatric Disorders

Alexander B. Niculescu1,2*

1Department of Psychiatry, Indiana University School of Medicine, Indianapolis, Indiana
2Indianapolis VA Medical Center, Indianapolis, Indiana

Manuscript Received: 6 March 2013; Manuscript Accepted: 19 March 2013

Genetic and gene expression studies, in humans and animal models of psychiatric and other medical disorders, are becoming increasingly integrated. Particularly for genomics, the convergence and integration of data across species, experimental modalities and technical platforms is providing a fit-to-disease way of extracting reproducible and biologically important signal, in contrast to the fit-to-cohort effect and limited reproducibility of human genetic analyses alone. With the advent of whole-genome sequencing and the realization that a major portion of the non-coding genome may contain regulatory variants, Convergent Functional Genomics (CFG) approaches are going to be essential to identify disease-relevant signal from the tremendous polymorphic variation present in the general population. Such work in psychiatry can provide an example of how to address other genetically complex disorders, and in turn will benefit by incorporating concepts from other areas, such as cancer, cardiovascular diseases, and diabetes. © 2013 Wiley Periodicals, Inc.

Key words: convergent functional genomics; bipolar; animal models; GWAS; gene expression

INTRODUCTION

“Coming together is a beginning; keeping together is progress; working together is success”.

-Henry Ford

Psychiatric disorders are phenotypically and biologically complex, heterogeneous, overlapping, and interdependent [Niculescu, 2006; Niculescu et al., 2006; Niculescu and Le-Niculescu, 2010a]. Unraveling their genetic basis by human genetic studies has proven arduous. The combination of complex genetics with imprecise clinical nosology, relying on patient self-report rather than on objective laboratory tests, has made this one of the difficult challenges in science. Given that the rewards of a better understanding range from alleviating mental illness and suffering to improved brain performance and understanding how the mind works, the prize is commensurate with the degree of difficulty.
Animal Models

Animal models are developed and used for two main reasons: a better understanding of the disorder (including at a gene expression level), and the testing of new drugs. Animal models of bipolar disorder can broadly be classified into genetic and environmentally induced. We will confine our discussion to rodent models, which are much more experimentally tractable and widely used than those of other species (Table I). The genetic models arise from naturally occurring or inbred strains, or more often from transgenic manipulation (genetic engineering) of candidate genes hypothesized to be involved in bipolar disorder. For the environmentally induced models, pharmacological manipulation and different stress-related paradigms are used to mimic different aspects of bipolar disorder. Usually, the animal model recapitulates features of one or the other of the two antithetical phases of the illness—mania versus depression, with the sole exception to date of the DBP KO mouse model [Le-Niculescu et al., 2008]. It is important to note that while there is a nosological distinction between depression and bipolar disorder, the genetics, biology, and clinical symptomatology involved are likely part of a continuum–spectrum [Akiskal, 2007; Niculescu et al., 2010].

The most widespread pharmacological model to date involves use of stimulants (amphetamines, methamphetamine) to mimic the manic phase of bipolar disorder [Niculescu et al., 2000]. Withdrawal from the stimulant can also mimic the depressive phase of the disorder. Sometimes, an anxiolytic agent is added, on the premise that mitigating the anxiogenic side-effects of stimulants leads to modeling of euphoric mania [Kelly et al., 2009]. However, that approach is questionable, as human bipolar patients naturally often display co-morbid anxiety and/or irritability as part of their bipolar clinical picture.

A more systematic pharmacogenomic approach used a comparison of the gene expression effects of a disease-mimicking stimulant (methamphetamine) and a disease-treating mood stabilizing agent (valproate) [Ogden et al., 2004], as a way of prioritizing genes that are affected by both treatments, especially the genes that are changed in opposite directions by the disease agonist and the disease antagonist. Moreover, gene expression effects were mapped in key disease-relevant brain regions, not in the whole brain [Ogden et al., 2004]. That work was subsequently extended to look at the gene expression changes in blood from the animals on the different treatments, as a way of identifying brain–blood biomarkers [Le-Niculescu et al., 2009b].

### TABLE I. Animal Models for Bipolar Disorder: Recent and/or Key Studies

<table>
<thead>
<tr>
<th>Genetically engineered</th>
<th>Naturally occurring/inbred strains</th>
<th>Pharmacological models</th>
<th>Other environmental manipulations</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLOCK [Roybal et al., 2007; Mukherjee et al., 2010; Arey et al., 2013]</td>
<td>Flinders Sensitive Line (FSL) rats [Malkesman and Weller, 2009]</td>
<td>Methamphetamine/valproate [Ogden et al., 2004]</td>
<td>Isolation housing [Le-Niculescu et al., 2008, Niwa et al., 2013]</td>
</tr>
<tr>
<td>HINT1 [Barbier and Wang, 2009]</td>
<td></td>
<td>Other mood stabilizers: Lamotrigene [Li et al., 2010], Topiramate [Bourin et al., 2009]</td>
<td>Restraint stress [Johnson et al., 2009; Koo et al., 2010]</td>
</tr>
<tr>
<td>WFS1 [Kato et al., 2008]</td>
<td></td>
<td>GBR12909 [Young et al., 2010a]</td>
<td></td>
</tr>
<tr>
<td>DAT [van Enkhuizen et al., 2012; Young et al., 2010b]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERK1 [Engel et al., 2009]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRIK2 [Shaltiel et al., 2008; Malkesman et al., 2010]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DGK8 [Kakefuda et al., 2010]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP1A3 [Kirshenbaum et al., 2012]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FBXL3 [Keers et al., 2012]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BI-1 [Hunsberger et al., 2011]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CACNA1C [Dao et al., 2010]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANK3 [Leuissis et al., 2012]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Only one genetic model to date, the DBP (D-box binding protein) knock-out mouse, has been shown to mimic both phases of the illness, using clinically relevant environmental manipulations [Le-Niculescu et al., 2008]. DBP is a circadian clock gene candidate for bipolar disorder, that was identified in earlier gene expression studies [Niculescu et al., 2000] in pharmacogenomic models and maps to a locus implicated in bipolar disorder in humans. At baseline, the knock-out animals are depressed compared to wild-type controls. During exposure to chronic stress (isolation housing) and acute stress (experimental handling), the mice exhibit a switch in phenotype to a manic-like phase, characterized by increased activity and increased hedonic behavior. This two hit paradigm (genetic vulnerability, followed by environmental stressors) mimics very well the human condition. The fact that a single gene constitutive knock-out has such a broad phenotype exceeded apriori expectations. It may be due in part to the fact that the gene knocked-out is a transcription factor, responsible for setting in motion a cascade of other changes, and also due to the fact that it is a circadian clock gene, which are emerging as key molecular underpinnings of mood disorders. Comprehensive gene expression studies in brain and blood, with and without exposure to stress, were carried out in this animal model, generating additional candidate genes and blood biomarkers for bipolar disorder [Le-Niculescu et al., 2008]. Treatment studies in this model using omega-3 fatty acids led to a normalization of the phenotype [Le-Niculescu et al., 2011b].

Another genetic model, a knock-out of the circadian clock gene CLOCK, has been originally described to have a phenotype that mimics only the manic side of the illness [Roybal et al., 2007]. More recent work with it involving brain region-specific manipulation of gene expression has revealed a mixed mood phenotype [Mukherjee et al., 2010]. Other recently described genetically engineered models for manic-like behavior involve manipulation of the genes DAT (dopamine transporter) [Young et al., 2010b], GRIN2A (NMDA receptor subunit 2A) [Taniguchi et al., 2009], HINT1 (protein kinase C interacting protein) [Barbir and Wang, 2009], ERK1 (extracellular signal regulated kinase 1) [Engel et al., 2009], and GRK2 (metabotropic glutamate receptor 6) [Shaltiel et al., 2008; Malkesman et al., 2010]. Candidates emerging from genome-wide association studies of bipolar disorder (CACNA1C [Dai et al., 2010], ANK3 [Leussis et al., 2012]) were also validated in animal models.

An interesting model, supportive of a role for mitochondrial involvement in bipolar disorder, is that of POLG1 (mitochondrial DNA polymerase) transgenic mice, where mutant POLG1 is expressed in a neuron-specific manner. These mice exhibit periodic activity changes and altered circadian rhythm, similar to bipolar cycling [Kasahara et al., 2006]. Subsequent studies comparing gene expression changes in these mutant mice to human postmortem brain gene expression changes in bipolar subjects identified two overlapping genes [Kubota et al., 2010]. One of them, SFPQ (splicing factor proline/glutamine rich), is also a top candidate gene for bipolar disorder from the DBP KO mouse model described above, where it is increased in expression in the amygdala the activated (manic) phase. The second gene, PPFE, encodes cyclophilin D, a component of the mitochondrial permeability transition pore. A blood–brain barrier permeable cyclophilin D inhibitor improved the abnormal behavior of the POLG1 mice, suggesting a potential lead for new drug discovery efforts.

An area of emerging interest is that of small regulatory RNAs. It is possible that broad disease-relevant phenotypes may be obtained in mice in the future by manipulating microRNAs, which, similar to transcription factors like DBP, regulate many other genes [Miller et al., 2012].

**Human Genetic Studies**

Over the last few years, in concert with other fields, genetic studies for bipolar disorder have been dominated by genome-wide association studies (GWAS) [Wellcome Trust Case Control Consortium 2007; Baum et al., 2008; Sklar et al., 2008; Scott et al., 2009; Smith et al., 2009; Soronen et al., 2010], and to a lesser extent copy-number variants (CNV) studies [Lachman et al., 2007; Zhang et al., 2009], and more recently, whole-genome sequencing studies [Kiezun et al., 2012]. GWAS studies to date have identified few polymorphisms that meet the genome-wide statistical threshold for. Those few findings in turn are not reproduced as statistically significant in independent GWAS, although some show additional evidence in meta-analyses [Ferreira et al., 2008; Schulze et al., 2009; Liu et al., 2011; Williams et al., 2011; Green et al., 2012], especially when samples from different psychiatric disorders are combined, leading one to suspect those gene variants have to do with basic brain and body functions shared across disorders [Steinberg et al., 2012; Smoller et al., 2013]. Consistent with that view, the findings tend to be in obscure, housekeeping-type genes (ANK3, CACNA1C, ODZ4), in contrast to the more biologically interesting genes implicated by gene expression studies in animal models and in human postmortem brain from subjects with bipolar and related disorders. A discussion of the reasons for this limited success of GWAS has been ongoing in the field [Niculescu and Le-Niculescu, 2010b], but an emerging explanation is that genetic heterogeneity at the SNP level is a contributory factor. As such, gene-level analyses are much more likely to be reproducible [Ayalaw et al., 2012]. In addition, gene-level analyses permit cross-platforms, cross-methodologies, and cross-species integration [Le-Niculescu et al., 2009a; Patel et al., 2010], particularly with animal models and gene expression studies (Fig. 1), which can help with identification and prioritization of disease-relevant genes.

**Human Gene Expression Studies**

Gene expression data may be the Rosetta Stone helping to tie together and unravel epistasis (co-acting gene expression (CAGE) [Niculescu et al., 2000], “genes that change together work together”), as well as regulatory networks of non-coding SNPs [Dunham et al., 2012], epigenetic changes, chromatin modifications, non-coding RNAs, and transcription factors responsive to environmental stimuli. Human gene expression studies have been carried out in postmortem human brains tissue [Banigan et al., 2013], as well as in peripheral blood [Le-Niculescu et al., 2009b], fibroblasts [Yang et al., 2009], olfactory epithelium-derived neurons [Tajinda et al., 2010], and more recently in induced pluripotent stem cells (iPSC)-derived neurons [Lin et al., 2011;
Brennand et al., 2011]. Each particular approach has strengths and limitations.

Integration

The integration of animal model and human studies has occurred either as hypothesis-driven validation, or as discovery-driven convergent integration of datasets.

The first approach takes a finding from one line of work, and studies it in the other. For example, genetically engineered mice of human candidate genes for mood disorders have been generated (DBP, CLOCK, ANK3, CANA1C, and others), as described above, and are proving to be useful animal models for the disorder. Conversely, a gene expression finding from animal model studies is pursued in candidate gene association studies in human populations. One such example from our work is that of RORB (RAR-related orphan receptor beta), another circadian clock gene. RORB was identified as changed in expression in the brain of DBP KO mice [Le-Niculescu et al., 2008]. It was then tested and shown to have genetic association with bipolar disorder in a human pediatric bipolar population [McGrath et al., 2009]. The rationale for studying a pediatric bipolar population was that pediatric bipolar subjects exhibit more rapid cycling and changes in mood state (switching), which are likely underpinned at a molecular level by circadian clock genes.

The second approach, the discovery-based integration of animal model and human data, has had its most systematic embodiment to date through Convergent Functional Genomics (CFG) (Fig. 1). The approach is predicated on using large datasets as well as manually curated databases of the published literature to date [Bertsch et al., 2005; Niculescu and Le-Niculescu, 2010a]. Each individual line of work has strengths and limitations. Animal model data can provide sensitivity and ability to conduct experimental manipulations not feasible in humans. Human data provides more specificity and relevance to the human disease. Using a set of mouse experiments as a driving force [Ogden et al., 2004; Le-Niculescu et al., 2008], or using human blood biomarker [Le-Niculescu et al., 2009b] or GWAS data [Niculescu et al., 2009a; Patel et al., 2010] as a driving force, such studies have identified and prioritized candidate genes and biomarkers for bipolar disorder that show good reproducibility as well as predictive ability in independent cohorts [Le-Niculescu et al., 2009b; Kurian et al., 2011; Ayalew et al., 2012; Patel et al., 2010].

The mining of GWAS data for bipolar disorder with a CFG approach was particularly successful [Le-Niculescu et al., 2009a; Patel et al., 2010], and holds generalizable lessons. The integration of GWAS data had as a first step the selection of SNPs. A nominal P-value threshold, not a genome-wide significance threshold, was used to select the positive SNPs from each GWAS, as it was assumed that most SNPs make only a small contribution to the disorder at a population level, and the work relied on the subsequent integration with other lines of evidence to identify and prioritize true positives. The second step is the conversion of SNPs into genes. From then on, all lines of evidence are tabulated at a gene level. The more lines of evidence, that is, the more times a gene shows up as positive finding across independent studies, platforms, methodologies and species, the higher its CFG score (Fig. 1). This is very similar conceptually to a Google PageRank algorithm, in which the more links to a page, the higher it comes up on the search prioritization list. Human and animal model, genetic and gene expression, datasets were integrated and tabulated. The top candidate genes were then assembled in a panel composed of their component SNPs, and tested in independent cohorts. Each subject in an independent cohort has a genetic risk prediction score (GRPS) based on how many of the SNPs in the panel it was positive for. Using such an approach, a polygenic panel of 56 top candidate genes for bipolar disorder, mined from GWAS using CFG, showed good predictive ability to differentiate, in independent cohorts, between bipolar and controls, as well as between less severe and more severe forms of bipolar disorder [Patel et al., 2010]. As an added feature, this approach identified top candidate genes that have a lot of prior biological evidence and disease relevance, as opposed to the mundane top findings from GWAS alone. For example, at the very top of the candidate gene list for bipolar disorder generated by this mega-analysis is ARNTL [Patel et al., 2010], another circadian clock gene also recently implicated in diabetes [Marcheva et al., 2010]. The top candidate genes for bipolar were then analyzed in terms of distribution in biological pathways and mechanisms, levels of analysis where there is less heterogeneity and a clearer picture emerges. The analysis resulted in the first comprehensive empirically derived model of bipolar disorder pathophysiology to date [Le-Niculescu et al., 2009a]. This led to a proposed understanding of mood as related to cellular and organismal energy, activity, and trophicity, as an adaptive clock-gene mediated synchronization to a favorable or hostile environment [Le-Niculescu et al., 2009a; Niculescu et al., 2010]. Excessive, discordant, or variable reactivity to environmental stimuli leads to clinical illness (depression in a favorable environment, mania in a hostile environment, cycling and switching from one mood state to the other that is not warranted by adaptation to the environment).

We have also used such a strategy for schizophrenia, with similar success [Kurian et al., 2011; Ayalew et al., 2012], showing
reproducibility in four independent cohorts. The polygenic approach exemplified in our earlier work [Le-Niculescu et al., 2009b; Kurian et al., 2011; Ayalew et al., 2012; Patel et al., 2010] has now become widely used, with varying success [Derks et al., 2012]. It seems clear, as our work has also shown, that good results depend on the strength of the (endo) phenotype [Fanous et al., 2012; Whalley et al., 2013] and the ability to prioritize genes.

Future Directions

The advances surveyed here occurred over the last decade as a result of collaborations between investigators with different backgrounds, using different approaches. They have opened the door to a better understanding of the genetics, biology, diagnosis and ultimately treatment of bipolar and related mood disorders, paving the way in the near future for individualized/personalized medicine. It is clear that such convergent strategies should continue to be employed and refined for bipolar disorder, in other psychiatric disorders, and in complex medical disorders in general. Psychiatric disorders share similarity at a genetic level with cancer and diabetes in terms of complex genetics, and even in terms of some of the molecular pathways involved [Le-Niculescu et al., 2009b; Marcheva et al., 2010; Niculescu et al., 2010]. Paradigms from cancer could be borrowed in psychiatric research, particularly the classification of genetic variants into risk genes (similar to oncogenes) and protective genes (similar to tumor-suppressor genes). An early proposal for such a classification used the terms psychogenes and psychosis-suppressor genes [Niculescu et al., 2000]. The complexity of these broad groups of disorders however is such that simple binary classifications may be insufficient, and only the complete understanding of the contextual cumulative combinatorics of common gene variants, development and environment may yield the ultimate answer [Patel et al., 2010]. Pathway analyses and mechanistic identification can lead to comprehensive, testable models that move the field beyond current nosological classifications [Niculescu et al., 2010], and lead to better diagnostics and treatments. Personalized medicine, that is pro-active in building resilience, preventive for avoiding environmental risk factors, predictive of whom is at risk and who will respond to treatment, and participatory for the patient and the family, is the ultimate outcome of our work.

ACKNOWLEDGMENTS

We would like to thank members of the Niculescu laboratory for their input, Ming Tsuang for mentorship and organizing wonderful collaborations over the years, and our colleagues in the field for the comprehensive work they are doing. This work was supported by an NIH Directors’ New Innovator Award (1DP2OD007363) and a VA Merit Award (1I01CX000139-01) to A.B.N.

NOTE ADDED IN PROOF

This article was published online on 31 May 2013. Subsequently, it was determined that the final version had not been published, and this was corrected on 3 June 2013. The primary change is found in References, where citations were expanded to the first ten authors where applicable.

REFERENCES


