Published in final edited form as:

Psychiatr Genet. 2011 April ; 21(2): 57–68. doi:10.1097/YPG.0b013e328341a346.

Database of Genetic Studies of Bipolar Disorder

John E. Piletz¹, Xiaotong Zhang², Rajdeep Ranade¹, and Chunyu Liu²

¹Department of Psychiatry, Loyola University of Chicago, Stritch School of Medicine, Maywood, Illinois

²Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, Illinois

INTRODUCTION

Bipolar Disorder (BD) is a complex genetic trait that is known to be highly heritable. Up to 93% of the risk for type-1 BD has been attributed to heritable factors according to one large multi-centered US twin study (Kieseppa et al. 2004). These and many other factors have fueled an intense search for the genes that predispose for BD. Aberrant gene expression patterns – some involving neurotransmission pathways - have been well documented to exist in tissues from BD patients (Altar et al. 2009; Le-Niculescu et al. 2009). However, exactly which gene variants underlie the heritability of this disorder remains elusive.

Ever since the search for BD genes began in earnest some 20 years ago, there have been different schools of thought about the best approach for identifying genetic variants. At first, most investigators favored family-based linkage studies which employed chromosome recombination statistics (LOD scores). The concern that recombination resolution was not sufficient led other investigators to initiate genetic studies of unrelated populations (case-control association studies). Their view was that bipolar illness should be linked to common alleles because of its widespread and uniform dispersion at 1–3% of the world wide population (Weissman et al. 1996). Others developed a hybrid approach called family-based association tests (FBATs), where trios or quads of family members are studied. The HapMap project (Thorisson et al. 2005; Zheng and McPeck 2007) added more and more single nucleotide polymorphisms (SNPs) from many human populations (Kelsoe 2004). Of particular relevance were the haplotype blocks and tagged SNPs that demarcate them (e.g., tagged SNPs: tSNPs) (Haiman and Stram 2008; Nannya et al. 2007). Mapping tSNPs across virtually the entire human genome gave rise to a whole new genetic approach: the genome wide scanning approach. Overlaid on this were mathematical advances in the statistics for large SNP datasets to “impute” genotype data, and to assess gene-gene and gene-environment interactions (Askland et al. 2009; Baum et al. 2008b; Johnson et al. 2009; McMahon et al. 2010; Moskvina et al. 2009; Purcell et al. 2009).

Currently, the most favored approach is the Genome Wide Association (GWA) study - at least until technology advances and it becomes practical and affordable to perform whole genome sequencing on large cohorts of subjects. The appeal of GWA studies is their atheoretical nature combined with ability to scan large numbers of SNPs across the genome. To date, there have been 12 independently published GWA studies on the etiology of BD (Baum et al. 2008a; Ferreira et al. 2008; Hattori et al. 2009; Lee et al. 2010; McMahon et al. 2010; O'Donovan et al. 2008; Purcell et al. 2009; Scott et al. 2009; Sklar et al. 2008; Smith et al. 2009; Sullivan et al. 2009; Wellcome-Trust-Consortium 2007), plus a GWA study

solely on the pharmacogenomics of BD (Perlis et al. 2009). Traditional candidate association studies and linkage studies of BD have also continued to appear at a rapid rate. Many genetic variants have therefore been studied by all three major genetic approaches: linkage studies, candidate case-control association studies, and GWA studies.

The hope of the first wave of GWA studies had naturally been to unify the field around specific gene findings. Instead, the first GWA studies have yielded only marginally significant findings with little consensus. Some signs of consensus may be emerging now that the sample size is approaching over 10,000 – as discussed later in this article – but the findings appear to explain only a small fraction of the overall disease liability. To be fair, the disappointment in GWA studies is not limited to psychiatric phenotypes, but characterizes most complex genetic studies to date (Galvan et al. 2010; Niculescu and Le-Niculescu 2010a, b). But, the daunting news for BD comes from recalculating the statistical power to move forward. As one recent reviewer put it, “If real gene associations do exist, they must be so weak as to require sample sizes of 100,000 cases and 100,000 controls to detect them” (Moonesinghe et al. 2008). In hindsight, the problem may have been in placing too much expectation on the GWA technology when it first was developed. Rather than being an end point in themselves, we believe GWA studies should compliment other approaches. “Other” approaches that we refer to in this article include candidate gene association and linkage studies as well as studies focused on epigenetic, neurobiological, and environmental aspects of the disease. Herein, we describe a centralized database containing a variety of genetic study designs of BD.

The main motivation behind our database lies in the hope that if weak gene findings can be viewed together, they might guide future studies. Suppose, for instance, one holds to a theory about gene X in BD. At present, it would be extremely difficult without an organized database to compare all the previous positive findings of gene X with all the previous negative findings, and then design the right kind of study to move forward with. Likewise, there is difficulty for genetic researchers when ranking their new findings in context with all other contender gene findings for BD. These types of questions have been very difficult to answer even for those in the field, not to mention for researchers in other disciplines who may wish to use the accumulated findings in other ways.

The traditional approach for unifying diverse findings has of course been meta-analysis, a statistical strategy for pooling data from independent studies. To date, two meta-analyses have been published on GWA studies of BD (Liu et al. 2010; McMahon et al. 2010). However, meta-analysis cannot capture the entire literature on BD genes because it presumes that the data is of similar type and quality across studies. Important case-control candidate studies – not to mention linkage studies and FBAT studies – cannot be pooled with GWA data by meta-analysis. To deal with diverse genetic approaches and findings, our database displays semi-quantitative data (p values etc.) side-by-side with qualitative issues of study design, numbers of subjects, numbers of SNPs, diagnosis subtypes etc. Although we do not attempt an overall analysis in this article, the database has been designed with an eye towards future meta and mega analytical strategies similar to those described by Le-Niculescu et al (Le-Niculescu et al. 2009). We believe our database will eventually support new analytical methods that may yet be developed for complex datasets.

The database can be found online at <http://bioprogramming.bsd.uchicago.edu/BDStudies/>. Users can now compare the relative merits of *many types of* gene findings in both semi-quantitative and qualitative manners on a gene-by-gene and study-by study basis. Most importantly, the database looks forward to facilitating efficient meta and mega analyses. Herein, we describe the database as well as provide two lists of “top” candidate genes for BD to date. We welcome future collaborators who may wish to analyze the database further.

MATERIALS AND METHODS

Literature Search

To obtain primary information, PubMed (www.pubmed.gov) was searched repeatedly with two key words: “bipolar gene”. Extensive searches were conducted with these key words from May 1, 2006 – December 31, 2009. By the end date, 3,388 distinct citations had been looked at - at least their titles, abstracts and key words. We required abstracts to mention specific numbers of BD patients before the full articles were obtained and read. Two further cut-off criteria were then applied to each paper: (a) the papers had to contain both genotypic and phenotypic information, and (b) the data had to be genic (naming genes rather than just loci). By the first criteria, a large number of papers were excluded because they provided only gene expression data, with no genotypic data. By the second criteria, even more articles were excluded because they were linkage studies that failed to mention specific genes. To confirm and extend the list, another online search was then conducted using “manic depressive gene” as key words. This second PubMed search yielded 594 citations. This list underwent the same selection criteria as above. Most articles in the second search had already been found in the first search. Only 10 new papers met criteria. A master list of 574 articles was finally compiled as the basis for establishing the database.

Specific Data Selection

The 574 articles found by the PubMed searches were then subjected to specific rules governing data lifting. While these rules were our own creation, they were devised in order to make the database as bias-free as possible. The following *a priori* rules of data selection were thus employed for data lifting:

- A. Peer-reviewed data only: No private databases were obtained, consulted, or otherwise used for even supplementary information because it was not clear if these sources had been peer-reviewed. The lifting of data was almost completely restricted to published texts, tables, and figures of the articles themselves - except for a few rare occasions. Those rare occasions were when it became apparent that key information existed but could not be found in the main articles. In those rare instances, our search for data points went to alternative sources (i.e., appendices and/or associated websites the authors maintained publically) – but limited ourselves to filling out the columns already started in the database by the main articles. This rule was applied to our own published studies (Shi et al. 2007; Shi et al. 2008; Smith et al. 2009), as well.
- B. Regardless of polarity: Data would be lifted whether it was a positive or negative association finding.
- C. No “extra” filters applied to the data: There had to have been some filtering of the data prior to almost every publication, which we decided was enough. We simply lifted whatever data the authors provided.
- D. No data mining tools: That is, the data points were found by eye and manually lifted. This was done mostly by a single curator (JEP). To limit human error, data entry was subjected to consensus agreement between JEP and CYL. The database was also scrutinized on a line-by-line basis for over a month before it was placed on the internet.
- E. Every study counts: In recent years it has become common to publish “multiplex” articles which contain a parent study followed by one or more replication studies embedded in the same article. We decided each replication study should be considered on its own, assuming it was conducted independently from the parent study (which we found to be the normal situation). Therefore, each and every study

(even what some would call sub studies of one article) are listed as separate lines in the database. For concerned readers, the interconnectedness of the sub studies within multiplex articles has not been obscured in the database. Their interconnectedness should be obvious because they share identical authors, publication dates, and PubMed ID numbers in the database.

- F. Open to different interpretations: The literature contains occasional instances of “new” articles that actually contain no primary data. These articles usually describe a new analytic or statistical method applied to earlier samples, in essence shedding light on old findings. Sometimes, there can even be different interpretations. In these cases, the original results and interpretations from publication #1 are listed on one line of the database. A second line follows to describe the same samples given their new interpretation according to publication #2. The reader is left to compare the two interpretations on their own merits. We felt this better than choosing between the two articles.

Data Categorization

Two additional rules:

- A. Only a single concluding statement and one statistic per gene per study: A single study often describes multiple SNPs/gene and multiple types of statistical analyses/ SNP or per gene (e.g., single point analysis versus haplotype analysis). To minimize the complexity, it was decided that in these cases the database should present only one concluding statement and one p value for each gene in the study. But, in a few rare cases the author’s conclusions were actually different from the commonly accepted significance value ($p < 0.05$, uncorrected). In such cases, we wrote the Synopsis Statement in the database to be inline with the nominal significance level. This was done for the sake of uniformity.
- B. No information about subtypes given unless it was a central point of the paper: But, when a study was found to have its core focus on subtype(s) of BD cases, we used as narrow a phenotypic definition as possible when lifting p values.

Database Construction

The database was constructed as a Microsoft Excel spreadsheet. It was then uploaded for management as a MySQL database. A web-based user interface was implemented to enable readers to search the database based on gene names and aliases. Some of the gene names may be different than in their original articles because they were standardized to HUGO gene names (<http://www.genenames.org/>) for the database. To be accommodating, we have provided an option to query aliases of the gene names.

Synopsis Statements for Each Finding

Readers of the database should find the Synopsis Statement column informative. Information provided under this column typically begins with a direct quotation from the original article. This is meant to encapsulate the interpretation originally put on the finding. After this quotation, there will typically be a series of qualifying statements such as alternative p values – especially if those alternative p values tend to expose alternative conclusions. Caution is also revealed when the authors present “corrected” p values (i.e. for multiple measures) rather than uncorrected p values. The exact meaning behind each p value is also provided unless the usual two-tailed, uncorrected, lowest p value is listed. Under the Synopsis Statement one may also learn when “borrowing” subjects occurred. This happened in studies that pooled previously published samples with new ones. It is important to keep track of truly independent samples (new samples/ new data) for future meta and mega

analyses. In short, the Synopsis Statements provide not only the original author's statements but other critical information for understanding the finding.

RESULTS

Cautionary Statement

This Results section has purposely been limited to the broad parameters of the database itself. Those seeking specific genes - for instance lists of "top genes" - are referred to the Discussion section. The reason we have not listed specific genes under the Results section is it would be presumptuous and inappropriate to claim any truly original findings based on a database of other people's studies. Furthermore, the database is only a starting point for researchers to organize their thoughts. It is not capable of providing definitive statements about the genes that actually underlie BD, short of future sophisticated statistical analysis.

On its inception date (January 29, 2010), the database held 893 different genes (Fig. 1). These came from 574 unrelated research articles that had been found in PubMed. In actuality, the articles encompassed 638 different independent studies because many were multiplex articles (see Methods for definition of multiplex articles). Simply counting the studies in the database revealed that 32.6 % of the first findings had been called "positive" by the authors. Most, however, were later realized to have had the "winners curse" (Niculescu and Le-Niculescu 2010b). That is, they failed to be replicated subsequently. Most of the other findings were called "negatives", except for a few situations where findings were equivocal. This process of simply counting the positive and negative studies, however, does not in any way provide a definitive picture of the genes for BD. This is because scant few studies were performed gene-wide, that is interrogating all the haplotypic blocks in a single gene. Therefore, it remains for the reader to weigh each polymorphic finding on a study-by-study basis, rather than simply counting positive and negative studies.

The most common type of study found was the candidate gene association type. There were 587 candidate gene association studies on launching the database. From these, 544 different genes were gleaned for the database (Fig. 1). By comparison, only 12 GWA studies were enough to contribute 262 unique genes to the database (Fig. 1). The database also contained findings from 22 genic linkage studies (linkage studies focused on specific genes). Besides these, the database contained 4 rare allele studies (4 unique gene entries) and one twin study focused on a single gene (Kakiuchi et al. 2003). More than 70 different gene entries in the database were lifted from a mixture of different types of genetic studies in one or more article.

It was noted that 55.5 % of the genes in the database had been repeatedly interrogated by different researchers using different populations (i.e. replication studies). The remaining 44.5% of the genes had been studied just once in connection with BD. The more frequently studied genes were those thought to involve antidepressants. The three most studied genes in the database were: (a) the serotonin transporter gene (SLC6A4 gene) with 52 different genetic studies of BD etiology, of which 59.6% of the studies have yielded negative findings; (b) the Brain Derived Neurotrophic Factor gene (BDNF gene) with 35 different studies of BD etiology, of which 46% of the studies have yielded negative findings, and (c) the Catechol-O-Methyl Transferase gene (COMT gene) with 29 different studies of BD etiology, of which 59% of the studies have yielded negative findings.

DISCUSSION

Efforts to unravel the genetic architecture of BD have been ongoing for at least two decades. Because most reports have been subsequently unreplicated (winner's curse), there is now a

general assumption that if “real” positives do exist in the literature they almost surely have emerged amidst a sea of false positives too under-powered to be definitive (Hakon 2009). That being the said, we will highlight only the most significant and repeated findings from our database. Two lists of “top” genes are shown in Tables 1 and 2. The first list comes exclusively from candidate gene association literature (Table 1). The second list comes exclusively from GWA study literature (Table 2). Specific details are provided in other sections of this Discussion section. Suffice to say, for now, that there is no overlap between the top findings in these two lists (Tables 1 and 2). The reader is cautioned, however, that the p values in these tables are uncorrected for multiple measures and that only allelic-type p values are listed. We will discuss the individual genes in these lists under separate sections below and attempt to provide some rationale for the non-overlap of the two lists.

We wish to emphasize certain qualities of our database which tends to set it apart from other databases: (1) peer-reviewed summary-level data only, (2) inclusion of negative as well as positive findings, and (3) a comprehensive nature that exceeds most previous databases. The Schizophrenia Forum has a similar database for schizophrenia genes (MacDonald and Schulz 2009), but it barely touches on BD. The Psychiatric GWA Study Consortium also shares some genotype and phenotype data online (Cichon et al. 2009; O'Dushlaine et al. 2010), but it does not attempt to chronicle candidate gene association studies. A large database maintained jointly by the NIH and the US Center of Disease Control and Prevention (Lin et al. 2006) also provides additional summary-level findings, but not GWA findings. There are also online databases from independent laboratories covering a variety of psychiatric disorders that include BD (Johnson and O'Donnell 2009; Konneker et al. 2008; Lin et al. 2006). However, these databases contain mostly candidate gene association findings. The Sullivan Lab Evidence Project (SLEP) (Konneker et al. 2008) provides one of the best online databases of summary level gene findings. Although the SLEP offers a lot (linkage data, copy number variant (CNV) data, expression microarray data, meta-analysis data, and genome-wide data), it lacks candidate gene association data. It also does not include negative findings. Two other online databases are the Open Access Database of Genome Wide Association Results (Johnson and O'Donnell 2009) and a website of the National Human Genome Research Institute (<http://www.genome.gov/gwastudies>). The former offers raw data from 118 diverse GWA studies on different diseases (according to data sharing agreements), and the latter database offers summary-level findings from 559 diverse GWA studies on different diseases (both are growing as of the writing of this paper). Both, however, lack candidate gene association findings. All the URLs for these sites are given as links on our webpage. Our database therefore finds its niche amongst these other sites as the only database containing summary level findings of both negative and positive nature, from a comprehensive collection of diverse types of studies.

Since the entire genome is theoretically present on one platform in any large-scale GWA study, some have argued that GWA findings should be valued above all other findings (Pearson and Manolio 2008). We agree that the GWA researcher has many reasons for pride. But, a few limitations do exist. These limitations mostly pertain to the problems inherent in handling vast amounts of data. Several thousands of subjects are necessary in GWA studies to even approach sufficiency of statistical power. And, as mentioned in the Introduction section, the recently recalculated statistical power for these studies may actually require hundreds of thousands of subjects (Hakon 2009). GWA studies are also restricted to the common alleles, meaning only those with minor allele frequencies >1% (debatable >5%). More problematic, some frequently studied variants have not been included in any of the GWA platforms. Until the age of full genome sequencing arrives, we therefore caution against assuming that GWA findings automatically supersede all other findings made by other research designs.

We would also emphasize the few advantages of the humble candidate gene association design. Its advantages are as follows: (1) Candidate gene studies hold inherent “face validity” because candidate genes are chosen precisely because of their known biological and/or pharmacological underpinnings. (2) Candidate gene studies stand to out-perform GWA studies when it comes to studying isolated populations or pedigrees. This is especially true if multiple rare mutations are involved (Walsh et al. 2008). This is because rare alleles fall entirely outside the boundaries of detection by GWA studies. (3) Candidate gene studies usually possess credible statistical power because of limited numbers of genotypes per study compared to GWA studies. This is not to say that candidate gene studies are perfect - or even optimal. Until recently, candidate gene studies fell well short of being able to interrogate even the known haplotype blocks of a single targeted gene. Many candidate gene studies were restricted to “alleles of interest”. This generally meant alleles encoding amino acid changes. While this seemed OK at the time, our awareness of regulatory elements in DNA has put to rest the notion that non-protein-coding variants should be waived as of no value. The problem of incomplete gene coverage in almost all early candidate gene association studies must be factored in.

Having a correct gene model also must be factored in. In the case of BD, the model of inheritance clearly is not classical (i.e., standard Mendelian single gene dominant or recessive). But, beyond this broad statement, little is certain. Most recent investigations have approached the study of BD genes from a “common alleles/ common disease” model, arguing that this model best explains the prevalence of the illness (Hemminki et al. 2008; Iyengar and Elston 2007; Tesli et al. 2009a). In fact, the common alleles/ common disease model lies at the heart of the GWA studies. Other investigators, however, have approached the puzzle of BD by looking for rare alleles (Walsh et al. 2008) and their often-related copy number variants (Saus et al. 2010). Still others have looked for interactions between allelic variation and environmental factors (Cornelis et al. 2010; Kieseppa et al. 2004; Liu and Piletz 2006). Others have focused on epigenetic variation (Crow 2007). There are also models of inheritance involving biochemical pathways (O'Dushlaine et al. 2010; Zhang et al. 2010). These models consider pleiotropy whereby genetic aberrations may cause a multiplicity of phenotypes through biochemical pathways. Depending on the starting model that one chooses, entirely different types of studies will emerge.

List of Top Candidate Gene Association Findings

Candidate gene association studies of BD began in the early 1990's. At that time, genotyping was confined to microsatellites, restriction fragment length polymorphisms (RFLPs) and/or copy number variants detected on gels. Today, almost all association studies use automated genotyping platforms. The development of the HapMap Project (Kelsoe 2004) made it possible to interrogate virtually all known heritable blocks (i.e., haplotypes) in each gene (Farrall and Morris 2005). There have been several good review articles written about SNPs in linkage disequilibrium and tagged SNPs (tSNPs) in genetic studies of BD (Badner and Gershon 2002; McInnis et al. 2003; McQueen et al. 2005; Segurado et al. 2003; Serretti and Mandelli 2008). There are also reproducible “hot” chromosome regions from these linkage studies, and all of these regions have generated candidate genes. Credible candidate gene association studies now required to enlist at least 200 cases and 200 controls for publication (being averaged-sized studies in recent years according to the database).

The top gene findings from candidate studies are listed in Table 1. To be selected for this list, a gene – not necessarily a particular SNP – had to have been studied at least 5 times in independent populations and come through with at least two more positive than negative studies. For simplicity sake, only the most robust studies are being shown for each gene in Table 1. It should also be realized that none of the genes listed in Table 1 can claim absolute consensus (i.e. without at least one negative report in their portfolios). There were invariably

failures to replicate somewhere in the literature. Keep in mind also that the p values listed in Table 1 are not bad when compared to those of the top findings from the GWA literature (Table 2) if one adjusts for starting statistical powers. In alphabetical order, the first gene in Table 1 is the **BDNF** gene. Shown is data for rs6265 of the Brain Derived Neurotrophic Factor (BDNF) gene on chromosome 11p13. It is followed by the **CD36** gene. Shown is data for rs2637777 of the Thrombospondin receptor gene (called CD36 gene) on chromosome location 7q11.2. Next is the **DAOA** gene. Shown is data for rs1935062 of the D-Amino acid Oxidase Activator (DAOA) gene on chromosome 13q34. Next is the **DISC1** gene. Shown is data for rs821616 and rs1411771 of the Disrupted Schizophrenia 1 (DISC1) gene on chromosome 1q42.1. Next is the **GRIN1** gene. Shown is data for rs35655437 of the Glutamate Receptor Ionotropic NMDA 1 (GRIN1) gene on chromosome 9q34.3. Next is the **NDUFV2** gene. Shown is data for rs1156044 of the NADH dehydrogenase Ubiquinone FlaVoprotein 2 (NDUFV2) gene on chromosome 18p11.31-p11.2. Next is the **TPH2** gene. Shown is data for rs4131348 of the Tryptophan Hydroxylase 2 (TPH2) gene on chromosome 12q21.1. Finally, there is the **TRPM2** gene. Shown is data for rs1618355 of the Transient Receptor Potential cation channel subfamily M2 (TRPM2) gene on chromosome 21q22.3. Again, each of these genes had been studied at least 5 times in independent populations, and yielded at least two more positive than negative associations with BD as of writing this report.

List of Top Genome Wide Association Findings

The top 30 gene findings from the first wave of GWA studies are listed in Table 2. This list was compiled by ranking p values from individual studies. One of the first things to be noticed is that the list in Table 2 seems at first glance to be completely unrelated to the list of Table 1. The reason for this is unknown. Some previous reviewers have dug deeply into the published GWA findings to here and there find some weak significance for the genes in Table 1 also coming forth in some GWA studies (Wellcome-Trust-Consortium 2007). However, the evidence is weak. We might speculate that because the designs of the studies in Tables 1 and 2 are vastly different (targeted association studies versus GWA studies), this may explain the results. For instance, the genes in Table 1 may simply pertain only to the inheritance of isolated BD populations, while the genes in Table 2 may pertain only to the inheritance of large general BD populations. However, the point we raise next is that some of the findings in Table 2 are beginning to achieve consensus from replication findings. Our database may be particularly well-suited for exposing the replicability of these findings.

In chronological order, the first gene to note in Table 2 is the Diacylglycerol Kinase eta gene (**DGKH**). Its association with BD was considered the main finding of the first GWA study (Baum et al. 2008a). These scientists found a triad of SNPs in DGKH to be statistically associated with BD in *both* their “parent” and “replication” studies. The grand finale analysis in that report yielded $p = 1.5 \times 10^{-8}$ for the association with BD (Baum et al. 2008a). This is still considered a strong finding, particularly because DGKH is known to encode a protein in the lithium-sensitive phosphatidyl inositol pathway. Two targeted gene association studies have also replicated the DGKH finding (Ollila et al. 2009; Squassina et al. 2009). But, another two targeted gene association studies have failed to replicate it (Tesli et al. 2009b; Yosifova et al. 2009). The jury is still out on DGKH.

The second major gene reported by Baum et al (2008a) was the **SORCS2** gene on chromosome 4p16.1. Finding a “hit” at chromosome 4p16.1 was of interest because this chromosome region had previously been associated with BD in a large consensus linkage study (Serretti and Mandelli 2008). The SORCS2 finding was also anchored by a triad of SNPs in *both* the parent study and the replication study of Baum and coworkers (2008a). Their grand finale analysis of the top SORCS2 SNP yielded significance at $p = 1.4 \times 10^{-5}$ for association with BD (Baum et al. 2008a). The SORCS2 finding has also enjoyed

multiple replications in the literature (Johnson et al. 2009; Ollila et al. 2009; Smith et al. 2009). It was replicated first by a convergence genome analysis using SNPs with p values < 0.05 within genes overlapping multiple bipolar samples (Monte Carlo simulation at $P < 0.00001$) (Johnson et al. 2009). It was later replicated by a gene-centric analysis (Baum et al. 2008a) of the Wellcome Trust Case Control Consortium (WTCCC (Wellcome-Trust-Consortium 2007)), and of Sklar et al (Sklar et al. 2008), and of a case-control comparison of substance abusers who may share genes predisposing for BD (Johnson et al. 2009). The SORCS2 finding was also replicated in a European GWA study of BD (Smith et al. 2009) and in a targeted gene association study of Finnish people by Ollila et al (Ollila et al. 2009). In fact, we could find only one study that has mentioned failing to associate the SORCS2 gene with BD (Yosifova et al. 2009) (though not mentioning it is worrisome in some studies). The SORCS2 gene is also of interest because it encodes a vacuolar protein sorting 10 (VPS10) protein of a domain-containing receptor and it seems to be strongly expressed in the central nervous system (CNS).

The second GWA study of BD was published by the WTCCC (Wellcome-Trust-Consortium 2007) (Table 2). The WTCCC studied BD in the context of six other genetic disorders in people from 12 British geographical regions. In total, the WTCCC genotyped 16,179 individuals: roughly 2,000 cases per disease ($n = 1,868$ BD) inclusive of 2,938 shared controls. Their strongest overall signal for BD was at SNP rs420259 under a recessive model, at chromosome position 16p12, which yielded a genotype-wise p value of 6.29×10^{-8} . This SNP lies close to the **PALB2** gene (Partner And Localizer of BRCA2) which is involved in stability of chromatin and the nuclear matrix. The PALB2 findings has support from one replication study (Yosifova et al. 2009) but not from another (Ollila et al. 2009).

The third GWA study of BD was reported by Sklar and colleagues (Sklar et al. 2008) (Table 2). This was a multiplex article with a parent GWA study (1,461 BD1 cases and 2,008 controls) followed by two replication studies of fewer subjects and pre-selected SNPs. Their first replication study was a FBAT and used NIMH-derived samples ($n = 409$ trios). Their second replication study was a case-control association design and used samples from the University of Edinburgh (UE: $n = 365$ cases, 351 controls). These investigators sought concordant findings across all three studies. This led them to highlight genes involving calcium. There was the **DFNB31** gene at 9q32-q34 (Deafness, autosomal recessive gene) and the **CACNA1C** gene at 12p13.3 (Calcium channel, voltage dependent, L-type, alpha 1C subunit). Our database notes that significant associations with BD for the DFNB31 gene were actually observed earlier in the GWA study of Baum and coworkers (2008a) and also in the WTCCC study (Wellcome-Trust-Consortium 2007). The DFNB31 gene encodes a calcium/calmodulin-dependent serine kinase-interacting protein involved in the formation of scaffolds that facilitates synaptic transmission in the CNS (Yap et al. 2003). The CACNA1C gene is discussed in more detail below.

Ferreira and colleagues have also reported (Ferreira et al. 2008) a GWA study of BD (Table 2 line 4). Their article amassed three large collections of cases and controls, totaling 4,387 BD cases and 6,209 healthy controls. Their top finding was with the gene for Ankyrin-3, **ANK3**. The top SNP in this gene, rs10994336, yielded a strong association statistic of $p = 9.1 \times 10^{-9}$. This was supported by many nearby SNPs in ANK3 (including two SNPs in the statistical range of $p < 5 \times 10^{-8}$). It was also supported by haplotype analysis. The database shows support for the ANK3 gene has also come from many outside studies. Seven independent GWA studies have mentioned some degree of positive findings for ANK3, though at different SNPs with BD (Lee et al. 2010; Purcell et al. 2009; Schulze et al. 2009). The Ankyrin-3 protein is also called Ankyrin-G. The protein regulates inactivation gating of the neuronal sodium channel, Nav1.6 (Shirahata et al. 2006). Therefore, the ANK3 gene seems to be achieving good consensus across multiple studies of BD.

The report by Ferreira and colleagues (2008) delved deeply into the **CACNA1C** gene which was found earlier by Sklar and colleagues (Sklar et al. 2008). They found SNP rs1006737 in the CACNA1C gene to achieve $p = 7.0 \times 10^{-8}$ in association with BD (Ferreira et al. 2008). Not long afterwards, this same SNP was reported (Green et al. 2009) to also confer risk for recurrent major depressive disorder (MDD) as well as for schizophrenia. Thus, SNP rs1006737 has raised attention in the psychiatric research community and may cross diagnostic boundaries with allelic odds ratios similar to those found for BD (~1.15) (Ferreira et al. 2008). It should also be noted that CACNA1C aligns with biochemical studies pointing to a primary defect in BD involving ionic dysregulation (Dubovsky et al. 1989; Dubovsky et al. 1991). For the record, the CACNA1C SNP rs1006737 associated with BD is not one of the better known channelopathy SNPs of the CACNA1C gene. Nonetheless, that SNP rs1006737 lies within the same gene as those often lethal SNPs, has not gone unnoticed.

Another recent GWA study was reported by McMahon et al (2010) (Table 2 line 9). These investigators reported a parent GWA study ($n = 645$ BD1 cases and $n = 1,310$ healthy controls) followed by a replication pooled genome wide analysis from the WTCCC (Wellcome-Trust-Consortium 2007), the NIMH (Smith et al. 2009), the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD) (Sklar et al. 2008), and some MDD cases from the Genetic Association Information Network MDD (GAIN-MDD) (Sullivan et al. 2009). These two studies were followed by three more replication studies focused on five targeted SNPs in two genes chosen from the parent study. Replication study-1 used a subset of BD cases and controls from a European ancestry GlaxoSK sample (Scott et al. 2009). Replication study-2 used a "Lausanne GlaxoSK" MDD sample ($n = 433$ MDD; $n = 916$ healthy controls). Replication study-3 used a "Munich GlaxoSK" MDD sample ($n = 866$ MDD; $n = 926$ healthy controls). McMahon and colleagues finished with a grand finale meta-analysis (excluding the MDD cases). The top finding was for SNP rs2251219 in the **PBRM1** gene found on chromosome 3p21.1. This SNP was modestly associated with BD in the parent study ($p = 2.0 \times 10^{-3}$) but also in most of the replication studies (p values from 10^{-4} - 10^{-2}), excepting three studies: the NIMH sample where the p value was 0.231, the targeted replication study-2 where the p value was 0.322, and the targeted replication study-3 where the p value was 0.123. Their grand finale meta-analysis included 6,683 BD cases (mixed BD1 and BD2 cases) and 9,068 healthy controls. Overall, their top SNP of PBRM1 achieved $p = 1.1 \times 10^{-8}$. PBRM1 encodes a protein involved in kinetochore localization during mitosis and may be a tumor suppressor found to be mutated in breast cancer. The WTCCC had earlier also reported a significant signal for PBRM1 at rs2878628 ($p = 0.0005$).

Finally, we would be remiss not to mention a series of GWA studies that began with schizophrenic patients by reporting associations for rs1344706 within the **ZNF804A** gene (Zinc-Finger protein 804A). The evidence for ZNF804A has been strongest when the schizophrenia phenotype is broadened to include bipolar disorder (Williams et al. 2010). Meta-analysis provides evidence for the association of rs1344706 that surpasses widely accepted benchmarks of significance by several orders of magnitude for both schizophrenia ($P = 2.5 \times 10^{-11}$, odds ratio 1.10, 95% confidence interval 1.07–1.14) and schizophrenia and bipolar disorder combined ($P = 4.1 \times 10^{-13}$, OR 1.11, 95% confidence interval 1.07–1.14) (Williams et al. 2010). The accepted cut-off statistic typically applied to GWA studies has been $p < 5 \times 10^{-8}$. Based on this established p value cut-off, the best genome wide findings for BD to date therefore appear to reside in the DGKH, ANK3, PBRM1, and ZNF804 genes.

There is debate about earlier simulated predictions using empirical HapMap data. In one early simulation performed by Nannya et al (2007), the statistical power needed to successfully identify genetic associations by the GWA study method had been estimated to be satisfactory with $n = 1,000$ cases and $n = 1,000$ controls. Other voices (Zaykin and

Zhivotovsky 2005), however, argued that positive findings would have just a 26% chance of falling into the top 1,000 findings in two independent GWA studies, even when the power to detect an effect was set at 85% (Zaykin and Zhivotovsky 2005). According to this simulation, very few replications would have been expected. The mere fact that we have a few replications so far (SORCS2, ANK3, DFNBS1, CACNA1C, PBRM1, and ZNF804) could therefore be viewed as highly encouraging. Nonetheless, the hype given to GWA studies when they first came forth may explain why most people today view our findings as discouraging.

Summary and Focus On the Future

Given so many sources of candidate genes, how many genes should underlie the inheritance of BD? One recent study that compared GWA studies of schizophrenia and BD (Purcell et al. 2009) estimated there might be hundreds or even thousands of different alleles that compose the polygenic inheritance of these illnesses. But, others point out that a polygenic model involving thousands of variants of weak effects is not likely to contribute to the heightened relative risks observed for BD in families (Nurnberger 2008). It has been argued that familial risks can only be explained by variants with large effects or by an oligogenic model involving at most two to three loci (Bodmer and Bonilla 2008; Hemminki et al. 2008).

So, how might the field move forward? A noteworthy attribute of the field is its willingness to pool data from different sources (Johnson and O'Donnell 2009). The single-most replicated finding to date, the ANK3 gene (Ferreira et al. 2008), could only have been achieved by pooling data from six different research teams: the WTCCC (Wellcome-Trust-Consortium 2007), STEP-BD (Nierenberg 2009), University College London (UCL), University of Edinburgh (ED), Trinity University of Dublin (DUB) and a second generation of STEP-BD called STEP2. And, another successful study that pooled data from different centers was able to highlight 69 “gene clusters” which were nominally yet positively associated with BD in at least three previous GWA studies (Johnson et al. 2009). In summary, pooling of samples and data is one way the field may move forward (Manolio et al. 2007).

Another way the field may move forward is through Convergent Functional Genomics (CFG) as described by Le Niculescu and colleagues (Le-Niculescu et al. 2009; Niculescu and Le-Niculescu 2010a, b). This approach, in addition to using p values from gene association studies, uses multiple independent lines of evidence related to the illness. These include mouse gene expression findings and human postmortem brain evidence. The CFG method takes account of several diversified datasets and prioritizes them within a cohort, similar to a Google PageRank algorithm. Genes prioritized by the CFG method may not have the highest p values from each GWA study, but are found to generalize across independent cohorts. Since the CFG approach is based on a gene-centric integration rather than a SNP-centric integration, it reduces heterogeneity. It can and has been used to mine GWA study datasets (Le-Niculescu et al. 2009), and to extract panels of top genes or biomarkers that reproduce well in independent cohorts. Studies that by themselves are relatively underpowered can be mined and made to yield results using CFG, by bringing to bear other large datasets and databases relevant to BD. In fact, we foresee the utility of similar approaches for a future analysis of our database.

Along the line of new analytical methods, a recent re-evaluation of the WTCCC dataset by “network analysis” has revealed the possibility of two-locus epistasis associated with BD (Emily et al. 2009). Epistasis is a phenomenon whereby one genetic variant requires another genetic variant to reveal the trait. Networks of genes might be explored more thoroughly as the field advances (O'Dushlaine et al. 2009; Peng et al. 2009; Walsh et al. 2008; Wang et al.

2007; Zamar et al. 2009). According to this approach a given variant might alone exert only small relative risk for the disease but may synergize through *co*-inheritance with other risk alleles in a biological pathway to produce a strong predisposition for the disease. If this model holds true, one of its logical corollaries might be that the likelihood of replicating any particular gene finding would depend not only on allele A in studies 1 and 2, but also on how closely the frequencies of key co-variants in genes B, C, and D match across studies.

There have also been papers highlighting the potential significance of interactomes in schizophrenia. At least two reports have described the DISC1 interactome (Camargo et al. 2007; Guo et al. 2009), as well as the DTNBD1 interactome (Guo et al. 2009). These studies are actually concerning schizophrenia, but are mentioned to address the likelihood of future interactome research in BD. The biological function of DISC1 remains poorly understood. Several reports on DISC1 have shown the protein interacts with two other gene products (the transcripts of NDEL1 and PDE4B genes), suggesting roles for DISC1 in the regulation of neurodevelopment and/or cAMP signaling pathways, respectively (Millar et al. 2005). Through functional and pathways analysis, there is evidence that DISC1 may intimately link to synapse function (Millar et al. 2003). The 'DISC1 Interactome' is envisioned as a network consisting of 127 proteins and 158 interactions. The other published interactome (Guo et al. 2009) involves the interaction between DISC1 and the dysbindin gene (DTNBD1). These observations suggest the involvement of the DTNBD1 gene which might affect dysbindin protein expression and thereby confer illness (Zinkstok et al. 2007). It is suggested that dysbindin and DISC1 proteins lie in a common biochemical pathway of relevance to Schizophrenia and BD.

There, of course, also remains the nagging likelihood that the diagnosis of BD might not itself be an entirely suitable phenotype to study. Type-1 of the disorder (BD1) is characterized by oscillating episodes of mania and depression, and type-2 (BD2) involves milder manic episodes (hypomania) followed by recurrent depressions. Both subtypes run together in pedigrees (Berrettini 2000). Moreover, major depressive disorder (MDD), schizoaffective disorder and schizophrenia are represented in most families of BD probands (Berrettini 2000). Therefore, BD may be best considered as a spectrum disorder. For future studies in this area we suggest that sub-symptoms be characterized according to the Bipolar Disorder Phenome Database (Potash et al. 2007) or the work of Deo and colleagues (Deo et al. 2010).

Finally, we will conclude with what findings stand out by their absence from our database. We refer to the paucity of positive genetic associations involving genes known to be involved with antidepressants. Proponents of the monoaminergic theory of depression would need to dig very deeply into our database to find any evidence of monoaminergic genes involved in the etiology of BD. This simple fact shows how open the field remains to new gene theories.

Acknowledgments

We thank Dr. Chindo Hicks of the Loyola University Medical Center and Dr. Elliott Gershon of the University of Chicago for useful comments on preliminary drafts of the manuscript. United States copyright protection for the database has been applied for.

References

- About Jamra R, Becker T, Georgi A, Feulner T, Schumacher J, Stromaier J, et al. Genetic variation of the FAT gene at 4q35 is associated with bipolar affective disorder. *Mol Psychiatry*. 2008; 13:277–284. [PubMed: 17938632]

- Altar CA, Vawter MP, Ginsberg SD. Target identification for CNS diseases by transcriptional profiling. *Neuropsychopharmacology*. 2009; 34:18–54. [PubMed: 18923405]
- Askland K, Read C, Moore J. Pathways-based analyses of whole-genome association study data in bipolar disorder reveal genes mediating ion channel activity and synaptic neurotransmission. *Hum Genet*. 2009; 125:63–79. [PubMed: 19052778]
- Badner J, Gershon E. Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Mol Psychiatry*. 2002; 7:405–411. [PubMed: 11986984]
- 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*. 2008a; 13:197–207. [PubMed: 17486107]
- Baum AE, Hamshere M, Green E, Cichon S, Rietschel M, Nothen MM, et al. Meta-analysis of two genome-wide association studies of bipolar disorder reveals important points of agreement. *Mol Psychiatry*. 2008b; 13:466–467. [PubMed: 18421293]
- Berrettini WH. Susceptibility loci for bipolar disorder: overlap with inherited vulnerability to schizophrenia. *Biol Psychiatry*. 2000; 47:245–251. [PubMed: 10682222]
- Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet*. 2008; 40:695–701. [PubMed: 18509313]
- Camargo LM, Collura V, Rain JC, Mizuguchi K, Hermjakob H, Kerrien S, et al. Disrupted in Schizophrenia 1 Interactome: evidence for the close connectivity of risk genes and a potential synaptic basis for schizophrenia. *Mol Psychiatry*. 2007; 12:74–86. [PubMed: 17043677]
- Chen Y, Akula N, Detera-Wadleigh S, Schulze T, Thomas J, Potash J, et al. Findings in an independent sample support an association between bipolar affective disorder and the G72/G30 locus on chromosome 13q33. *Mol Psychiatry*. 2004 Jan.9:87–92. image 85. [PubMed: 14699445]
- Cichon S, Craddock N, Daly M, Faraone SV, Gejman PV, Kelsoe J, et al. Genomewide association studies: history, rationale, and prospects for psychiatric disorders. *Am J Psychiatry*. 2009; 166:540–556. [PubMed: 19339359]
- Cornelis MC, Agrawal A, Cole JW, Hansel NN, Barnes KC, Beaty TH, et al. The Gene, Environment Association Studies consortium (GENEVA): maximizing the knowledge obtained from GWAS by collaboration across studies of multiple conditions. *Genet Epidemiol*. 2010; 34:364–372. [PubMed: 20091798]
- Crow TJ. How and why genetic linkage has not solved the problem of psychosis: review and hypothesis. *Am J Psychiatry*. 2007; 164:13–21. [PubMed: 17202538]
- Deo AJ, Costa R, DeLisi LE, DeSalle R, Haghighi F. A novel analytical framework for dissecting the genetic architecture of behavioral symptoms in neuropsychiatric disorders. *PLoS One*. 2010; 5:e9714. [PubMed: 20300526]
- Dubovsky SL, Christiano J, Daniell LC, Franks RD, Murphy J, Adler L, et al. Increased platelet intracellular calcium concentration in patients with bipolar affective disorders. *Arch Gen Psychiatry*. 1989; 46:632–638. [PubMed: 2735813]
- Dubovsky SL, Lee C, Christiano J, Murphy J. Elevated platelet intracellular calcium concentration in bipolar depression. *Biol Psychiatry*. 1991; 29:441–450. [PubMed: 2018817]
- Emily M, Mailund T, Hein J, Schauser L, Schierup MH. Using biological networks to search for interacting loci in genome-wide association studies. *Eur J Hum Genet*. 2009
- Farrall M, Morris AP. Gearing up for genome-wide gene-association studies. *Hum Mol Genet*. 2005; 14 Spec No. 2:R157–R162. [PubMed: 16244314]
- Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, et al. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet*. 2008; 40:1056–1058. [PubMed: 18711365]
- Galvan A, Ioannidis JP, Dragani TA. Beyond genome-wide association studies: genetic heterogeneity and individual predisposition to cancer. *Trends Genet*. 2010; 26:132–141. [PubMed: 20106545]
- Green EK, Grozeva D, Jones I, Jones L, Kirov G, Caesar S, et al. The bipolar disorder risk allele at CACNA1C also confers risk of recurrent major depression and of schizophrenia. *Mol Psychiatry*. 2009
- Guo AY, Sun J, Riley BP, Thiselton DL, Kendler KS, Zhao Z. The dystrobrevin-binding protein 1 gene: features and networks. *Mol Psychiatry*. 2009; 14:18–29. [PubMed: 18663367]

- Haiman CA, Stram DO. Utilizing HapMap and tagging SNPs. *Methods Mol Med*. 2008; 141:37–54. [PubMed: 18453083]
- Hakon H. Largest GWAS Analysis to Date Offers Only Two New Candidate Genes. *Schizophrenia Research Forum*. 2009 Available at: <http://www.schizophreniaforum.org/new/detailprint.asp?id=1532>.
- Hamshere ML, Green EK, Jones IR, Jones L, Moskvina V, Kirov G, et al. Genetic utility of broadly defined bipolar schizoaffective disorder as a diagnostic concept. *Br J Psychiatry*. 2009; 195:23–29. [PubMed: 19567891]
- Hattori E, Toyota T, Ishitsuka Y, Iwayama Y, Yamada K, Ujike H, et al. Preliminary genome-wide association study of bipolar disorder in the Japanese population. *Am J Med Genet B Neuropsychiatr Genet*. 2009; 150B:1110–1117. [PubMed: 19259986]
- Hemminki K, Forsti A, Bermejo JL. The 'common disease-common variant' hypothesis and familial risks. *PLoS One*. 2008; 3:e2504. [PubMed: 18560565]
- Iyengar SK, Elston RC. The genetic basis of complex traits: rare variants or "common gene, common disease"? *Methods Mol Biol*. 2007; 376:71–84. [PubMed: 17984539]
- Johnson AD, O'Donnell CJ. An open access database of genome-wide association results. *BMC Med Genet*. 2009; 10:6. [PubMed: 19161620]
- Johnson C, Drgon T, McMahon FJ, Uhl GR. Convergent genome wide association results for bipolar disorder and substance dependence. *Am J Med Genet B Neuropsychiatr Genet*. 2009; 150B:182–190. [PubMed: 19127564]
- Kakiuchi C, Iwamoto K, Ishiwata M, Bundo M, Kasahara T, Kusumi I, et al. Impaired feedback regulation of XBP1 as a genetic risk factor for bipolar disorder. *Nat Genet*. 2003; 35:171–175. [PubMed: 12949534]
- Kelsoe JR. Genomics and the Human Genome Project: implications for psychiatry. *Int Rev Psychiatry*. 2004; 16:294–300. [PubMed: 16194762]
- Kieseppa T, Partonen T, Haukka J, Kaprio J, Lonnqvist J. High concordance of bipolar I disorder in a nationwide sample of twins. *Am J Psychiatry*. 2004; 161:1814–1821. [PubMed: 15465978]
- Konneker T, Barnes T, Furberg H, Losh M, Bulik CM, Sullivan PF. A searchable database of genetic evidence for psychiatric disorders. *Am J Med Genet B Neuropsychiatr Genet*. 2008; 147B:671–675. [PubMed: 18548508]
- Le-Niculescu H, Patel SD, Bhat M, Kuczenski R, Faraone SV, Tsuang MT, et al. Convergent functional genomics of genome-wide association data for bipolar disorder: comprehensive identification of candidate genes, pathways and mechanisms. *Am J Med Genet B Neuropsychiatr Genet*. 2009; 150B:155–181. [PubMed: 19025758]
- Lee MT, Chen CH, Lee CS, Chen CC, Chong MY, Ouyang WC, et al. Genome-wide association study of bipolar I disorder in the Han Chinese population. *Mol Psychiatry*. 2010
- Lin BK, Clyne M, Walsh M, Gomez O, Yu W, Gwinn M, et al. Tracking the epidemiology of human genes in the literature: the HuGE Published Literature database. *Am J Epidemiol*. 2006; 164:1–4. [PubMed: 16641305]
- Liu, CY.; Piletz, J. Recent Developments in Nucleic Acids Research. Kerala, India: Research Signpost; 2006. Strategies and technologies for genetic study of complex diseases; p. 267-305.
- Liu Y, Blackwood DH, Caesar S, de Geus EJ, Farmer A, Ferreira MA, et al. Meta-analysis of genome-wide association data of bipolar disorder and major depressive disorder. *Mol Psychiatry*. 2010
- MacDonald AW, Schulz SC. What we know: findings that every theory of schizophrenia should explain. *Schizophr Bull*. 2009; 35:493–508. [PubMed: 19329559]
- Manolio TA, Rodriguez LL, Brooks L, Abecasis G, Ballinger D, Daly M, et al. New models of collaboration in genome-wide association studies: the Genetic Association Information Network. *Nat Genet*. 2007; 39:1045–1051. [PubMed: 17728769]
- McInnis MG, Lan TH, Willour VL, McMahon FJ, Simpson SG, Addington AM, et al. Genome-wide scan of bipolar disorder in 65 pedigrees: supportive evidence for linkage at 8q24, 18q22, 4q32, 2p12, and 13q12. *Mol Psychiatry*. 2003; 8:288–298. [PubMed: 12660801]
- McMahon FJ, Akula N, Schulze TG, Muglia P, Tozzi F, Detera-Wadleigh SD, et al. Meta-analysis of genome-wide association data identifies a risk locus for major mood disorders on 3p21.1. *Nat Genet*. 2010; 42:128–131. [PubMed: 20081856]

- 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. [PubMed: 16175504]
- Millar JK, Christie S, Porteous DJ. Yeast two-hybrid screens implicate DISC1 in brain development and function. *Biochem Biophys Res Commun.* 2003; 311:1019–1025. [PubMed: 14623284]
- Millar JK, Pickard BS, Mackie S, James R, Christie S, Buchanan SR, et al. DISC1 and PDE4B are interacting genetic factors in schizophrenia that regulate cAMP signaling. *Science.* 2005; 310:1187–1191. [PubMed: 16293762]
- Moonesinghe R, Khoury MJ, Liu T, Ioannidis JP. Required sample size and nonreplicability thresholds for heterogeneous genetic associations. *Proc Natl Acad Sci U S A.* 2008; 105:617–622. [PubMed: 18174335]
- Moskvina V, Craddock N, Holmans P, Nikolov I, Pahwa JS, Green E, et al. Gene-wide analyses of genome-wide association data sets: evidence for multiple common risk alleles for schizophrenia and bipolar disorder and for overlap in genetic risk. *Mol Psychiatry.* 2009; 14:252–260. [PubMed: 19065143]
- Mundo ETS, Neves-Pereira M, Dalton EJ, Macciardi F, Parikh SV, Bolonna A, Kerwin RW, Arranz MJ, Makoff AJ, Kennedy JL. Evidence that the N-methyl-D-aspartate subunit 1 receptor gene (GRIN1) confers susceptibility to bipolar disorder. *Mol Psychiatry.* 2003; 8:241–245. [PubMed: 12610658]
- Nannya Y, Taura K, Kurokawa M, Chiba S, Ogawa S. Evaluation of genome-wide power of genetic association studies based on empirical data from the HapMap project. *Hum Mol Genet.* 2007; 16:2494–2505. [PubMed: 17666406]
- Neves-Pereira M, Mundo E, Muglia P, King N, Macciardi F, Kennedy J. The brain-derived neurotrophic factor gene confers susceptibility to bipolar disorder: evidence from a family-based association study. *Am J Hum Genet.* 2002 Sep; 71:651–655. [PubMed: 12161822]
- Niculescu AB, Le-Niculescu H. Convergent Functional Genomics: what we have learned and can learn about genes, pathways, and mechanisms. *Neuropsychopharmacology.* 2010a; 35:355–356. [PubMed: 20010721]
- Niculescu AB, Le-Niculescu H. The P-value illusion: How to improve (psychiatric) genetic studies. *Am J Med Genet B Neuropsychiatr Genet.* 2010b
- Nierenberg AA. Lessons from STEP-BD for the treatment of bipolar depression. *Depress Anxiety.* 2009; 26:106–109. [PubMed: 19177467]
- Nurnberger JI Jr. A simulated genetic structure for bipolar illness. *Am J Med Genet B Neuropsychiatr Genet.* 2008; 147B:952–956. [PubMed: 18314863]
- O'Donovan MC, Craddock N, Norton N, Williams H, Peirce T, Moskvina V, et al. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet.* 2008; 40:1053–1055. [PubMed: 18677311]
- O'Dushlaine C, Kenny E, Heron E, Donohoe G, Gill M, Morris D, et al. Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. *Mol Psychiatry.* 2010
- O'Dushlaine C, Kenny E, Heron EA, Segurado R, Gill M, Morris DW, et al. The SNP ratio test: pathway analysis of genome-wide association datasets. *Bioinformatics.* 2009; 25:2762–2763. [PubMed: 19620097]
- Ollila HM, Soronen P, Silander K, Palo OM, Kieseppa T, Kaunisto MA, et al. Findings from bipolar disorder genome-wide association studies replicate in a Finnish bipolar family-cohort. *Mol Psychiatry.* 2009; 14:351–353. [PubMed: 19308021]
- Palo OM, Anttila M, Silander K, Hennah W, Kilpinen H, Soronen P, et al. Association of distinct allelic haplotypes of DISC1 with psychotic and bipolar spectrum disorders and with underlying cognitive impairments. *Hum Mol Genet.* 2007; 16:2517–2528. [PubMed: 17673452]
- Pearson TA, Manolio TA. How to interpret a genome-wide association study. *Jama.* 2008; 299:1335–1344. [PubMed: 18349094]
- Peng G, Luo L, Siu H, Zhu Y, Hu P, Hong S, et al. Gene and pathway-based second-wave analysis of genome-wide association studies. *Eur J Hum Genet.* 2009

- Perlis RH, Smoller JW, Ferreira MA, McQuillin A, Bass N, Lawrence J, et al. A genomewide association study of response to lithium for prevention of recurrence in bipolar disorder. *Am J Psychiatry*. 2009; 166:718–725. [PubMed: 19448189]
- Potash JB, Toolan J, Steele J, Miller EB, Pearl J, Zandi PP, et al. The bipolar disorder phenotype database: a resource for genetic studies. *Am J Psychiatry*. 2007; 164:1229–1237. [PubMed: 17671286]
- Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009; 460:748–752. [PubMed: 19571811]
- Saus E, Brunet A, Armengol L, Alonso P, Crespo JM, Fernandez-Aranda F, et al. Comprehensive copy number variant (CNV) analysis of neuronal pathways genes in psychiatric disorders identifies rare variants within patients. *J Psychiatr Res*. 2010
- Schulze TG, Detera-Wadleigh SD, Akula N, Gupta A, Kassem L, Steele J, et al. Two variants in Ankyrin 3 (ANK3) are independent genetic risk factors for bipolar disorder. *Mol Psychiatry*. 2009; 14:487–491. [PubMed: 19088739]
- Scott LJ, Muglia P, Kong XQ, Guan W, Flickinger M, Upmanyu R, et al. Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. *Proc Natl Acad Sci U S A*. 2009; 106:7501–7506. [PubMed: 19416921]
- Segurado R, Detera-Wadleigh SD, Levinson DF, Lewis CM, Gill M, Nurnberger JI Jr, et al. Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: Bipolar disorder. *Am J Hum Genet*. 2003; 73:49–62. [PubMed: 12802785]
- Serretti A, Mandelli L. The genetics of bipolar disorder: genome 'hot regions,' genes, new potential candidates and future directions. *Mol Psychiatry*. 2008; 13:742–771. [PubMed: 18332878]
- Shi J, Badner JA, Gershon ES, Liu C. Allelic association of G72/G30 with schizophrenia and bipolar disorder: A comprehensive meta-analysis. *Schizophr Res*. 2007
- Shi J, Badner JA, Hattori E, Potash JB, Willour VL, McMahon FJ, et al. Neurotransmission and bipolar disorder: a systematic family-based association study. *Am J Med Genet B Neuropsychiatr Genet*. 2008; 147B:1270–1277. [PubMed: 18444252]
- Shirahata E, Iwasaki H, Takagi M, Lin C, Bennett V, Okamura Y, et al. Ankyrin-G regulates inactivation gating of the neuronal sodium channel, Nav1.6. *J Neurophysiol*. 2006; 96:1347–1357. [PubMed: 16775201]
- Sklar P, Smoller JW, Fan J, Ferreira MA, Perlis RH, Chambert K, et al. Whole-genome association study of bipolar disorder. *Mol Psychiatry*. 2008; 13:558–569. [PubMed: 18317468]
- Smith EN, Bloss CS, Badner JA, Barrett T, Belmonte PL, Berrettini W, et al. Genome-wide association study of bipolar disorder in European American and African American individuals. *Mol Psychiatry*. 2009; 14:755–763. [PubMed: 19488044]
- Squassina A, Manchia M, Congiu D, Severino G, Chillotti C, Arda R, et al. The diacylglycerol kinase eta gene and bipolar disorder: a replication study in a Sardinian sample. *Mol Psychiatry*. 2009; 14:350–351. [PubMed: 19308020]
- Sullivan PF, de Geus EJ, Willemsen G, James MR, Smit JH, Zandbelt T, et al. Genome-wide association for major depressive disorder: a possible role for the presynaptic protein piccolo. *Mol Psychiatry*. 2009; 14:359–375. [PubMed: 19065144]
- Tesli M, Kahler AK, Andreassen BK, Werge T, Mors O, Mellerup E, et al. No association between DGKH and bipolar disorder in a Scandinavian case-control sample. *Psychiatr Genet*. 2009a; 19:269–272. [PubMed: 19478689]
- Tesli M, Kahler AK, Andreassen BK, Werge T, Mors O, Mellerup E, et al. No association between DGKH and bipolar disorder in a Scandinavian case-control sample. *Psychiatr Genet*. 2009b
- Thorisson GA, Smith AV, Krishnan L, Stein LD. The International HapMap Project Web site. *Genome Res*. 2005; 15:1592–1593. [PubMed: 16251469]
- Van Den Bogaert A, Slegers K, De Zutter S, Heyrman L, Norrback KF, Adolfsson R, et al. Association of brain-specific tryptophan hydroxylase, TPH2, with unipolar and bipolar disorder in a Northern Swedish, isolated population. *Arch Gen Psychiatry*. 2006; 63:1103–1110. [PubMed: 17015812]

- Walsh T, McClellan JM, McCarthy SE, Addington AM, Pierce SB, Cooper GM, et al. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*. 2008; 320:539–543. [PubMed: 18369103]
- Wang K, Li M, Bucan M. Pathway-Based Approaches for Analysis of Genomewide Association Studies. *Am J Hum Genet*. 2007; 81
- Weissman MM, Bland RC, Canino GJ, Faravelli C, Greenwald S, Hwu HG, et al. Cross-national epidemiology of major depression and bipolar disorder. *Jama*. 1996; 276:293–299. [PubMed: 8656541]
- Wellcome-Trust-Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; 447:661–678. [PubMed: 17554300]
- Williams HJ, Norton N, Dwyer S, Moskvina V, Nikolov I, Carroll L, et al. Fine mapping of ZNF804A and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder. *Mol Psychiatry*. 2010
- Xu C, Li PP, Kennedy JL, Green M, Hughes B, Cooke RG, et al. Further support for association of the mitochondrial complex I subunit gene NDUFV2 with bipolar disorder. *Bipolar Disord*. 2008; 10:105–110. [PubMed: 18199248]
- Xu C, Maciardi F, Li PP, Yoon IS, Cooke RG, Hughes B, et al. Association of the putative susceptibility gene, transient receptor potential protein melastatin type 2, with bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet*. 2006; 141B:36–43. [PubMed: 16252251]
- Yap CC, Liang F, Yamazaki Y, Muto Y, Kishida H, Hayashida T, et al. CIP98, a novel PDZ domain protein, is expressed in the central nervous system and interacts with calmodulin-dependent serine kinase. *J Neurochem*. 2003; 85:123–134. [PubMed: 12641734]
- Yosifova A, Mushiroda T, Stoianov D, Vazharova R, Dimova I, Karachanak S, et al. Case-control association study of 65 candidate genes revealed a possible association of a SNP of HTR5A to be a factor susceptible to bipolar disease in Bulgarian population. *J Affect Disord*. 2009; 117:87–97. [PubMed: 19328558]
- Zamar D, Tripp B, Ellis G, Daley D. Path: a tool to facilitate pathway-based genetic association analysis. *Bioinformatics*. 2009; 25:2444–2446. [PubMed: 19628508]
- Zaykin DV, Zhivotovsky LA. Ranks of genuine associations in whole-genome scans. *Genetics*. 2005; 171:813–823. [PubMed: 16020784]
- Zhang K, Cui S, Chang S, Zhang L, Wang J. i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. *Nucleic Acids Res*. 2010
- Zheng M, McPeck MS. Multipoint linkage-disequilibrium mapping with haplotype-block structure. *Am J Hum Genet*. 2007; 80:112–125. [PubMed: 17160899]
- Zinkstok JR, de Wilde O, van Amelsvoort TA, Tanck MW, Baas F, Linszen DH. Association between the DTNBP1 gene and intelligence: a case-control study in young patients with schizophrenia and related disorders and unaffected siblings. *Behav Brain Funct*. 2007; 3:19. [PubMed: 17445278]

NIH-PA Author Manuscript

Table 1

Top Findings of Candidate Gene Association Studies to Date

Candidate Gene Association Study	Source of Subjects	Numbers of Subjects	Gene Highlighted	Best Reported P Values*
(Neves-Pereira et al. 2002)	282 Canadian Caucasian families	282 BD probands; 564 parent CTLs (triads)	BDNF (rs6265)	3.9 E(-04) for 3-G allele haplotype TDT
(Mundo E 2003)	288 Canadian families (96.8% Caucasian)	276 BD probands; 576 parent CTLs (triads)	GRIN1 (rs35655437**)	7.0 E(-03) for haplotype TDT
(Chen et al. 2004)	North American Caucasian families	139 BD; 113 CTLs	DAOA (G72/G30) (rs1935062)	4.0 E(-03) for 3 allele haplotype
(Van Den Bogaert et al. 2006)	Northern Swedish Clinics	182 BD; 135 MDD; 364 CTLs	TPH2 (rs4131348 BD, rs11178997 MDD)	4.0 E(-03) for BD 1.0 E(-03) for MDD
(Xu et al. 2006)	Canadian (Caucasian) clinics	178 BD; 268 CTLs	TRPM2 (rs1618355)	2.3 E(-12) for seven marker haplotype
(Palo et al. 2007)	179 Finnish families	219 BD probands; 361 parent CTLs	DISC1 (rs821616 and rs1411771)	2.0 E(-04) for under-transmitted in haplotype TDT
(Abou Jamra et al. 2008)	German clinics	425 BD; 419 CTLs	CD36 (FAT) (rs2637777)	1.8 E(-05) for five marker haplotype
(Xu et al. 2008)	Canadian (Caucasian) clinics	178 BD; 120 CTLs	NDUFV2 (rs1156044)	2.0 E(-03)

* * Best p values were selected from amongst all reports on each particular gene, regardless of the same or different SNP rs numbers. They are given only in terms of allelic count association (two-tailed, prior to correction for multiple testing). The selection criteria for this table were: (a) only genes investigated in at least two peer-reviewed reports of bipolar disorder, totaling at least 5 independent case-control populations under study, and there had to be at least two more positive than negative reports; (b) the study which was chosen to represent the gene was the one with the best allelic p value regardless of rs numbers.

** The report on GRIN1 referred to variant 6608-A/G which may correspond to rs35655437 in intron 11.

Abbreviations: BD (bipolar disorder); CTLs (healthy controls); MDD (major depressive disorder); TDT (transmission disequilibrium test).

Table 2

Top Thirty Findings of Genome Wide Association Studies to Date

Genome Wide Study	Source of Subjects	Numbers of Subjects	Genes	Best Reported P Values*
(Baum et al. 2008a)	Cohort-1: NIMH-GI probands. Cohort-2: Patients from Universities of Bonn and Heidelberg	Cohort-1: 461 BD; 563 CTLs Cohort-2: 772 BD, 876 CTLs	1 DGKH (rs1170191, rs1012053, rs931585); 2 SORCS2 (rs4411993)	1 1.5 E(-08) 2 1.4 E(-05)
(Hamshire et al. 2009; Wellcome-Trust-Consortium 2007)	12 British regions. Controls were 50% from 1958 British Birth Cohort ('58C), and 50% selected blood donors	1,868 BD; 2,938 CTLs. Six other patients groups were also genotyped	1 PALB2, NDUFAB1, and DCTN5 in region 6p12 (rs420259). 2 KCNC2 (rs1526805) 3 B3GALT5 (rs4818065)	1 6.3 E(-08) for all three in cluster. 2 2.2 E(-07) 3 2.3 E(-07)
(Sklar et al. 2008)	Cohort-1: STEP-BD plus UCL for BD, NIMH-GI Rutgers Repository plus UCL for controls. Cohort-2: NIMH-GI parent-proband trios, plus University of Edinburgh based cases and controls	Cohort-1: 1,461 BD; 2,008 CTLs Cohort-2: 365 BD; 351 CTLs Cohort-3: WTCCC data	1 MYO5B (rs4939921); 2 TSPAN8 (rs1705236); 3 EGFR (rs17172438, rs729969)	1 2.0 E(-08) 2 7.6 E(-07) 3 8.4 E(-08)
(Ferreira et al. 2008)	New Cohort-1: ED-DUB-STEP2 Cohort-2: All above plus WTCCC, plus STEP-BD, plus UCL	New Cohort-1: 1,098 BD; 1,267 CTLs Cohort-2: 4,387 BD; 6,209 CTLs	1 ANK3 (rs10994336) 2 CACNA1C (rs1006737)	1 9.1 E(-09) 2 7.0 E(-08)
(Scott et al. 2009)	Cohort-1: NIMH-GI plus Prechter Cohort-2: GSK	Cohort-1: 1,177 BD; 772 CTLs Cohort-2: 899 BD; 904 CTLs Cohort-3: WTCCC data	1 MCTP1 (rs17418283), 2 GNL3 (rs2289247), 3 NEK4 (rs1029871), 4 ITIH3 (rs3617), 5 ITIH1 (rs7867133, rs1042779), 6 NFIX (rs472913),	1 9.7 E(-08) 2 1.0 E(-07) 3 1.8 E(-07) 4 1.0 E(-07) 5 1.8 E(-07) 6 2.0 E(-07)
(Perlis et al. 2009)	Treatment response in STEP-BD plus UCL	Cohort-1: 1,177 BD of which 458 were on lithium Cohort-2: 359 BD on lithium	1 SNP at 10p15 (rs10795189) 2 GRIA2 (rs9784453)	1 5.5 E(-07) 2 4.0 E(-04)
(Smith et al. 2009)	Cohort-1: European ancestry subjects collected in NIMH-GI five waves (EA) Cohort-2: African ancestry (AA)	Cohort-1: 1,001 BD; 1,933 CTLs. Cohort-2: 345 BD; 670 CTLs	1 SLITRK2 near Xq27.1 (rs5907577 in EA) 2 NCKAP5 (rs10093871 in EA) 3 DPY19L3 (rs2111505 in AA) 4 NTRK2 (rs2769605 in AA)	1 1.6 E(-06) 2 9.8 E(-06) 3 1.5 E(-06) 4 4.5 E(-05)
(Hattori et al. 2009)	Japanese	107 BD; 107 CTLs	KCNMB2 (rs9290656)	5.5 E(-04)
(McMahon et al. 2010)	New Cohort-1: German Cohort-2: German cohort plus WTCCC, NIMH-GI, STEP-BD, & GSK	New Cohort-1: 645 BD; 1,310 CTLs Cohort-2: 6,683 BD; 9,068 CTLs	1 PBRM1 (rs2251219) 2 GNL3 (rs2289247)	1 1.1 E(-08) 2 3.1 E(-08)

Genome Wide Study	Source of Subjects	Numbers of Subjects	Genes	Best Reported P Values*
(Lee et al. 2010)	Cohorts 1 and 2: Han Chinese	Cohort-1: 1,000 BD; 1,000 CTLs Cohort-2: 409 BD; 1,000 CTLs	1 SP8 (rs2709736)	1 4.9 E(-07)
			2 ST8SIA2 (rs8040009)	2 6.1 E(-06)
			3 KCTD12 (rs2073831)	3 9.7 E(-06)
			4 CACNB2 (rs11013860)	4 5.2 E(-05)
			5 ANK3 (rs1938526)	5 6.6 E(-05)

* Best p values were selected from amongst all reports on each particular gene, regardless of the same or different SNPs and rs numbers. They are given in terms of allelic associations (two-tailed, prior to correction for multiple testing).

Abbreviations: BD (bipolar disorder); CTLs (healthy controls); WTCCC (Wellcome Trust Case Control Consortium); NIMH-GI (NIMH Genetics Initiative); STEP-BD (Systematic Treatment Enhancement Program for Bipolar Disorder performed in the United States); UCL (University