A COMPUTATIONAL METHOD FOR DETECTING MULTI-LOCUS ASSOCIATIONS: AN APPLICATION TO BIPOLAR DISORDER

By

Michael A. Mooney

A DISSERTATION

Presented to the Department of Medical Informatics & Clinical Epidemiology
and the Oregon Health & Science University
School of Medicine
in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

May 2011

School of Medicine

Oregon Health & Science University

Certificate of Approval

This is to certify that the PhD Dissertation of

Michael A. Mooney

"A Computational Method for Detecting Multi-locus Associations: An Application to Bipolar Disorder"

Has been approved

_____

Dissertation Advisor – Shannon McWeeney, PhD

_____

Committee Member – Thomas Barrett, MD, PhD

_____

Committee Member – Kemal Sonmez, PhD

_____

Committee Member – Beth Wilmot, PhD

TABLE OF CONTENTS

List of Figures

List of Tables

Acknowledgements

I'd like to express my sincere gratitude to everyone who helped make this work possible. My dissertation committee, Shannon McWeeney, PhD; Thomas Barrett MD, PhD; Kemal Sonmez, PhD; and Beth Wilmot, PhD, were instrumental in helping me develop the ideas presented here. Shannon particularly has been a great source of inspiration, support and encouragement over the past four years, and my accomplishments are due, in large part, to her efforts as my advisor.

I'd like to thank the members of the Bipolar Genome Study, especially John Kelsoe, MD, for allowing me to work with their data and for encouraging me that the work I've done is of importance to the field.

Over the past few years I've had the opportunity to work with a number of wonderful people in the Department of Medical Informatics & Clinical Epidemiology, all of whom helped me navigate graduate school and helped to expand my perspective on scientific research. Although this is not a complete list, I'd like to thank the following people: Diane Doctor, Andrea Ilg, Lynne Schwabe, Bill Hersh, MD; Joan Ash, PhD; Karen Eden, PhD; Suzi Fei, PhD; and Ted Laderas.

And, of course, this work would not have been possible without support from the National Library of Medicine Training Fellowship in biomedical informatics.

Finally, I am especially thankful for my family. Without the love and support of my parents Margaret and Charles Mooney, and my sister Sara Mooney, my successes would not have been possible, nor as meaningful.

Abstract

Enormous data collection efforts and improvements in technology have made large-scale genome-wide association studies (GWAS) a promising approach to better understanding the genetics of common, complex diseases. However, the limited success of these studies so far suggests that genetic susceptibility may be due to a combined effect of multiple genetic variants (or interactions between variants), and that there may be a significant amount of genetic heterogeneity among those affected with complex diseases. It is clear that new data analysis methods are needed to address these hypotheses.

Using data from the NIMH-sponsored Bipolar Genome Study, this project attempted to discover groups of SNPs that are jointly associated with the disease, thereby explaining a greater portion of disease susceptibility than can be achieved by examining SNPs individually. A machine-learning technique, known as a genetic algorithm, was used to search for these multi-locus associations, and was guided by a variety of genomic information, such as protein-protein interactions, gene expression patterns, and gene functions.

A subset of the data was used to tune the algorithm's parameters and evaluate its performance. With the most appropriate parameters, and when allowed to run for a sufficient amount of time, the algorithm was consistently able to find multi-locus associations that met the conventional threshold for genome-wide significance. Also, it was able to discover these associations without performing a prohibitively high number of statistical tests.

When the algorithm was applied to the entire Bipolar Genome Study data set, three particularly interesting SNP pairs were discovered in the following genes: *CREB1* and *BMP7*; *GFRA1* and *ENPP2*; *NDUFV2* and *SLC35F1*. Each SNP pair was significantly associated with the disease (chi-square p-values: 7.48×10^{-10}, 1.17×10^{-7} and 3.26×10^{-7}), although none of the six SNPs had strong individual associations. Furthermore, a number of the genes implicated by these multi-locus associations have functions that are relevant to bipolar disorder and/or have been associated with the disease in previous studies.

The results presented here show that in the absence of individual SNP associations in a GWAS it is nevertheless possible to find strong multi-SNP associations. This approach offers promise not only for primary analyses, but also for secondary analyses in order to further mine the large GWAS data sets already available. However, significantly larger sample sizes will be needed to search for larger multi-SNP associations (more than two SNPs). And further modification of this approach, by incorporating specific genetic models (such as additive or conditional models) directly into the algorithm, should further improve the algorithm efficiency and the interpretation of results.

1 Introduction

1.1 Genome-wide Association Studies

With the sequencing of the human genome and our growing understanding of the genetic variation among the human population, new techniques have emerged for investigating the genetic component of common diseases. It is now a fairly standard procedure to genotype close to a million positions of an individual's genome in a single experiment. These variable positions in the genome are known as single-nucleotide polymorphisms, or SNPs (pronounced "snips"). By performing this large-scale genotyping on a sample of cases (patients with a specific disease thought to have a significant genetic component) and another sample of controls (healthy subjects) we are able to calculate correlations (or associations) between specific genotypes and the disease. This technique is referred to as a genome-wide association study (GWAS). When a specific genotype is seen significantly more frequently in cases than in controls, this suggests the genotype is in some way influencing disease susceptibility. For the most part, these studies are meant to generate hypotheses—they are a first step in the path to better understanding the biological mechanism of the disease.

Two important assumptions underlie GWAS. First, that susceptibility to the disease under study is influenced by common genetic variants (the DNA variant's frequency in the population is greater than one percent). Second, that the markers tested for association are either causal variants themselves, or are in linkage disequilibrium with causal variants. Hence, the design and interpretation of these studies depends heavily on our

knowledge of the genome structure within the study population. The validity of these assumptions has become a matter of serious debate [1].

In the past six years, over 800 large GWAS have been performed in an attempt to better understand the genetic component of complex diseases [2]. For some diseases, GWAS have proved successful and have provided significant insight into the nature of the disease. In the case of age-related macular degeneration, Klein et al. discovered an intronic SNP in the *CFH* gene with an odds ratio of 4.6 [3]. After sequencing the region, this SNP was found to be in linkage disequilibrium with a missense polymorphism in a region of the protein that binds C-reactive protein. This discovery has provided significant insight into the role of the inflammatory response in macular degeneration.

For other diseases, however, the results from GWAS have been frustratingly inconclusive, with very little agreement among numerous studies in different patient populations. For instance, in the Wellcome Trust Case-Control Consortium (WTCCC) study [4] no associations of genome-wide significance (generally accepted as a p-value smaller than 5×10^{-7}) were found for hypertension, despite the fact that there is evidence suggesting genetics play an important role in the disease. The sibling recurrence risk ratio for hypertension has been reported to be between 2.5 and 3.5. (The sibling recurrence risk ratio, also referred to as the sibling relative risk, is a standard way of estimating the genetic component of a disease, and is defined as the risk to siblings of an affected individual relative to the risk in the general population.)

There are a number of hypotheses about the nature of complex diseases that may explain the limited success of GWAS so far:

1. Susceptibility to complex diseases is due to many different genetic variants, each with a small effect. Our ability to detect these small effects may be hampered by a number of factors, such as insufficient sample size and noise caused by imprecise phenotype definition.

2. Susceptibility may be due to interactions between multiple genetic variants. For the most part, GWAS so far have looked for associations with individual SNPs only [5].

3. Some complex diseases may have a significant amount of genetic heterogeneity, meaning that the genetic variants that confer susceptibility in one family (or population) may be different from those in another.

4. Phenotypic variation among sample populations may be large, and may in fact represent multiple sub-types of the disease of interest (each with a different genetic mechanism). Therefore, a precise phenotypic definition is important for reducing noise in the data and increasing the power of the study.

5. Rare variants (rather than common SNPs) may play a significant role in causing common diseases [6, 7], though this is still a matter of debate. Dickson and colleagues [1] have recently shown that "synthetic" associations found in GWAS can actually be the result of common SNPs being linked to distant rare variants, and that rare variants may be the true causal agents in certain common genetic diseases.

6. While GWAS focus on the association between SNPs and the disease of interest, there are other types of genetic variation that may also play a role in disease susceptibility. Copy number variation (CNV) is one type of variation that has attracted particular interest in psychiatric disorders [8, 9].

7. Beyond genetic variation, there is also the possibility that epigenetic variation (variation without changes in the DNA sequence) plays a role in disease susceptibility. While this type of variation, particularly DNA methylation patterns, has been associated with some cancers, its role in common hereditary diseases is still an open question.

This project will focus on addressing the first two of these hypotheses, with the hope of discovering groups of SNPs associated with the disease of interest. A more detailed description of the projects goals are described in the Specific Aims below.

1.2 The Use Case: Bipolar Disorder

Bipolar disorder, also known as manic-depressive illness, is a mood (affective) disorder that affects about 1 % of the US population. Although the experiences of bipolar patients can vary significantly, typical symptoms include episodes of mania—elevated mood or euphoria, increased energy, low attention span, grandiosity, etc.—and, in many cases, episodes of depression—persistent feelings of sadness, fatigue, lack of interest and motivation, sleep disturbances, etc.—that interfere with normal functioning.

The official diagnostic criterion for bipolar I, according to the fourth edition of the American Psychiatric Association's (APA) Diagnostic and Statistical Manual (DSM-IV), is at least one episode of mania (which is often, but not necessarily, accompanied by episodes of major depression). A diagnosis of bipolar II requires one or more major depressive episodes accompanied by at least one hypomanic episode (having an elevated mood and increased energy, but still fully functional).

The clinical definition of bipolar disorder, particularly as it relates to highly recurrent forms of depression, is still a matter of some debate. Due to the large amount of variation

seen among patients with mood disorders, there is growing support for the concept of a bipolar (or manic-depressive) spectrum. Rather than attempting to place patients within pre-defined categories, patients would be characterized based on where they fall along a number of different dimensions (e.g., no mania to severe mania). While the spectrum model is a way to conceptualize the vast amount of variation among individuals (from normal temperament to severe affective disorder), it is also a way to reconcile the fact that many distinct disorders (e.g., unipolar depression, bipolar, schizophrenia) share common features and may indeed share common genetic causes [10].

Despite the ongoing struggle to conceptualize the wide variety of clinical presentations among patients with mood disorders, it is essential for genetic researchers to have a precise definition of the trait (phenotype) under study. This is especially true given the evidence that certain features of bipolar disorder—for example, early age-of-onset and rapid cycling—may be passed within families, indicating different genetic mechanisms than patients without these features [10].

The disorder appears to have a significant genetic component. Family and twin studies have found bipolar disorder to have a heritability between 0.70 and 0.80, with a sibling recurrence risk ratio of approximately 10 [10]. These values are comparable to other disease, such as age-related macular degeneration (heritability = 0.6 – 0.7) and Crohn's Disease (heritability = 0.5 – 0.6), for which GWAS have provided meaningful results [3, 11]. (Heritability is defined as the proportion of a population's phenotypic variation that can be explained by genetic variation among individuals.)

At least ten large-scale GWAS on bipolar disorder have been reported to date (see Appendix) [2]. The results from these studies vary widely. A study by Ferreira and

colleagues found 20 different SNPs associated with bipolar disorder (p-values < 10^{-5}), while three other studies found no associations at all [12]. Despite the apparent discrepant results across studies, Baum and colleagues have compared two bipolar GWAS and suggest that, when not focusing solely on the top hits, there may be significant points of agreement [13]. In other words, given the heterogeneity among data sets, it may not be required that a strong association found in one data set also have a strong association in a second data set. Even a modest association may lend support for the association of interest. In their study, Baum and colleagues simply ask the question: "does the set of SNPs implicated by Baum et al. [14] show more evidence of association in the WTCCC sample than would be expected by chance?"

Craddock and colleagues have modeled several possibilities for the mode of transmission of bipolar disease and found the most likely scenario is a multiplicative model by which disease susceptibility is due to the interaction of three or more genes [5]. However, up to this point, association studies have not looked at genetic interactions of this complexity.

In general terms, it is thought that bipolar disorder is due to the interaction of multiple defective or dysregulated susceptibility genes, which lead to a periodic imbalance in brain chemistry (loss of homeostasis). The majority of hypotheses about the underlying biological mechanism of bipolar disorder relate to the disruption of signaling pathways, an imbalance (either an excess or deficiency) of neurotransmitters, or the dysfunction of neurotransmitter receptors. The major neurotransmitter systems implicated in bipolar disorder are the noradrenergic, dopaminergic, serotonergic, cholinergic, GABAergic and

glutamatergic systems. Various other hormones and proteins that function in the brain have also been associated with the disease [10].

Still, the biochemical imbalances observed in patients with manic-depressive illness may be secondary effects, with the primary causes being defects in cellular signaling pathways [15]. Not only do signaling pathways regulate the various neurotransmitter systems linked to bipolar disorder, but they are also targeted by bipolar-associated hormones as well as some of the drugs used to effectively treat mood disorders [10].

1.3 The Polygenic Model of Complex Diseases

As discussed briefly above, in recent years, large genome-wide association studies have been successful in revealing a number of genes thought to play a role in complex diseases. Yet, in most cases, the genetic associations discovered have accounted for only a small portion of the "genetic component" of these diseases. It's clear that new hypotheses about the way genetic variation affects disease risk need to be examined. One such hypothesis is the polygenic model of complex disease—that susceptibility to the disease is due to the combined effect of multiple genetic variants.

A number of researchers have tested the polygenic hypothesis, with varying methods, and have discovered that, indeed, risk does increase when individuals carry multiple "risk alleles", each having a weak association with the disease. Baum and colleagues performed a genome-wide association study of bipolar disorder and examined the additive effects of ten of the most significantly associated SNPs in their data. They found that the proportion of cases rises with the number of risk alleles carried, and that individuals with 19 or more risk alleles were 3.8 times more likely to be classified as a

case than as a control. However, they do clearly note that the variants studied confer only a susceptibility to the disease, given that many controls also carried a significant number of risk alleles [14].

Purcell and colleagues examined the question of polygenic risk on a much larger scale. In the context of a large GWAS, they first defined large sets of SNPs (ranging from just under 1000 SNPs to over 38000 SNPs) that met various p-value thresholds. Next, they calculated scores for each individual in the GWAS based on the number of risk alleles carried (Purcell et al. are careful to use the term "score alleles" rather than risk alleles since many SNPs in the sets are sure to be false positives). They go on to show that these scores are significantly associated with schizophrenia, and also with bipolar disorder in two independent data sets [4, 16, 17].

While both of these studies suggest an additive polygenic model of susceptibility in bipolar disorder, discovery of an interactive or conditional effect among SNPs is a more complicated matter due to the problem of combinatorial complexity—the enormous number of possible SNP combinations. But despite the difficulty of the problem, a number of different approaches have been used to discover multi-locus associations in GWAS [18, 19, 20, 21]. For the most part, previous studies have dealt with the issue of combinatorial complexity by reducing the number of SNPs included in the analysis or by looking only for SNP pairs, ignoring the possibility of higher-order interactions.

Baum and colleagues (in the same study referenced earlier) looked for interactions among only those SNPs with the strongest individual associations. The most significant SNP interaction found was between SNPs in *SORCS2* and *DGKH* (joint p-value = 1.2×10^{-8}). However, the SNP in *DGKH* had a strong individual association (p-value =

1.5×10^{-8}, OR = 1.59) and the odds ratio for the SNP pair was not reported, so it is not clear whether this was a true interactive effect [14].

Emily and colleagues used protein-protein interaction data to define the population of SNP pairs to be tested for multi-locus association within the WTCCC data set. They discovered a significant interaction between SNPs in *PDGFRB* and *KITLG* (p-value = 8.32×10^{-8}) [22]. In this case, neither SNP was individually associated with bipolar disorder, which makes this an interesting example of an interactive effect. And because the joint contingency table was reported, we can see that there appears to be a conditional effect, whereby the SNP in *PDGFRB* becomes significantly associated when conditioned on the T/T genotype of the SNP in *KITLG*.

All of these studies support the idea of a polygenic model of susceptibility to bipolar disorder. Our aim was to build upon this work by developing a method that allows for a full-scale search for multi-locus associations, without necessarily reducing the number of SNPs in the analysis. In order to aid the biological interpretation of our results, we exploited the information gained from various other genomic datasets, such as protein-protein interaction data and previous linkage studies. We also chose to limit the size of the SNP sets examined to between two and four SNPs.

2 Research Objectives

2.1 Research Questions

The overall goal of the dissertation was to use a computational technique to search for multi-locus genetic associations among a large-scale GWAS data set, and to show the feasibility and the usefulness of this approach. Essentially, the intention was to answer the following questions: Is it possible to perform a full-scale search (i.e., without necessarily reducing the number of SNPs in the analysis) for multi-locus associations? And, if we approach the GWAS analysis with a polygenic model of disease susceptibility in mind from the outset, are meaningful results discovered?

To attempt to answer the above questions, the research was broken into two main parts:

1. Algorithm Evaluation and Refinement
2. Full-scale Analysis and Biological Interpretation of Top Hits

2.2 Specific Aims

Aim 1: Evaluate the ability of a network-guided genetic algorithm to search for and discover significant multi-locus associations in the context of a large-scale GWAS.

The evaluation will be done on a subset of the data, and will involve the measurement of various indicators of the efficiency and thoroughness of the search. In addition, the value of prioritizing genes in the network will be examined. The search algorithm will be guided by various types of information, such as previous linkage studies and gene

functions. An efficient and well-informed algorithm is essential given the extreme combinatorial complexity of the problem.

Ultimately, this evaluation will provide an opportunity for tuning the algorithm's parameters, which will give the best possible chance of producing meaningful results in the second half of the project. To determine the best parameter set and the most informative data types, the algorithm will be run multiple times with various inputs.

Aim 2: Compare the type and strength of associations (i.e., statistical significance and odds ratio) found when looking at groups of SNPs versus individual SNPs.

A search for groups of SNPs associated with bipolar disorder will be conducted on the full data set. The main question is whether or not new genetic associations can be discovered by looking at multiple variants simultaneously, rather than simply looking at individual SNPs. The following hypotheses will be tested:

Hypothesis 1: Disease susceptibility is due to multiple "hits" within a single biological pathway. The search will be limited to connected sub-networks (representing single pathways or functional modules) within the larger network.

Hypothesis 2: Disease susceptibility is due to a combination of genetic variants in multiple distinct pathways. The search will allow for groups of disconnected nodes in the network (representing SNPs in multiple separate pathways or cellular processes).

Aim 3: If any significant associations are discovered, provide a thorough review of the literature on the genes implicated, and present any supporting evidence for their role in the pathophysiology of bipolar disorder.

3 Data & Methods

3.1 The Genotype Data

The primary genotype data used in this study is from the Bipolar Genome Study (BiGS) [23], and is a combination of two independent datasets: 2035 individuals genotyped by the Genetic Association Information Network (GAIN) and 1865 individuals genotyped by the Translational Genomics Research Institute (TGen) [24]. Both datasets were genotyped using the Affymetrix Human Genome SNP 6.0 array, with over 900000 SNPs. Genotype calls were made with the CRLMM algorithm [25].

The following criteria were used to filter out low quality samples/SNPs: low genotyping call rates per sample (< 0.97), excessively high or low sample heterozygosity rates, low genotyping call rates per SNP (< 0.95), low minor allele frequency (< 0.01), and deviation from Hardy-Weinberg Equilibrium (p-value < 10^{-6}). After quality control measures were applied to the entire dataset, 916 cases and 1018 controls from the GAIN dataset, 1117 cases and 402 controls from the TGen dataset, and 636169 SNPs remained.

Phenotypic information for all cases was collected with the Diagnostic Interview for Genetic Studies [26], and only cases diagnosed as bipolar I disorder (BPI) were included in the analysis.

Genotype data from the WTCCC was used solely for the purposes of replicating significant associations found in the BiGS data. The WTCCC data consists of 1998 bipolar disorder cases and 1500 controls. Genotyping was done with the Affymetrix GeneChip Human Mapping 500K Array Set, with approximately 500000 SNPs. And

imputation was performed for another 2139483 SNPs in the HapMap database. Genotype calls were made with the CHIAMO algorithm [4].

3.2 The Gene Interaction Network

To constrain the search, and to facilitate the biological interpretation of results, SNPs were first mapped to genes within a gene interaction network. Interaction data was gathered from the STRING database (ver. 8.1) [27], while gene definitions were taken from Ensembl (ver. 54, NCBI build 36) [28]. A SNP was mapped to a gene if it fell within the gene itself or within a 3 kb window upstream or downstream of the gene. Of the 636169 SNPs that passed quality control, 217012 SNPs mapped to 12793 genes in the interaction network.

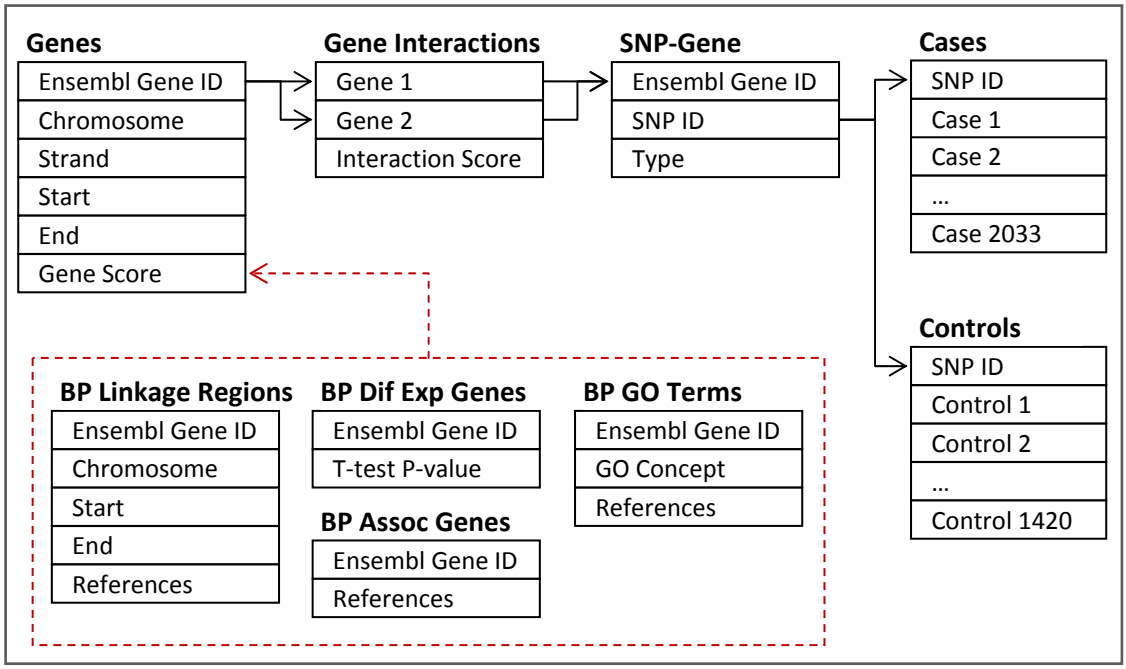

Figure 1. The relational database used by the algorithm to navigate throughout the gene interaction network.

In an attempt to focus the search even further, we compiled results from various studies about the pathophysiology and genetics of bipolar disorder. Evidence from previous linkage studies, gene association studies, and gene expression studies was used to assign scores to genes in the interaction network. In addition, gene function information (Gene Ontology terms [29]) and network connectivity were also used to score genes. The algorithm is able to use the scores to prioritize the genes in the network as it searches for multi-locus associations. The following paragraphs explain how each type of evidence was used to assign score to genes in the interaction network. All the information used to assign gene scores, and references for all the studies used as evidence are listed in the Appendix.

Linkage Regions and Gene Associations: Each linkage region and gene was categorized as having "weak" evidence, "moderate" evidence, or "strong" evidence based on the number of studies that reported each particular finding. If the finding had only one or two references it was considered weak evidence; if it had three, four or five references it was considered moderate evidence; and if it had more than five references it was considered strong evidence. Genes were assigned 1, 2 and 3 points for weak, moderate, and strong evidence of involvement in bipolar disorder, respectively. Points were assigned for each type of evidence separately, so if a gene was implicated in a linkage study and a gene association study it would be assigned points for both.

Functional Concepts: The Gene Ontology (GO) terms for each gene were searched for concepts related to bipolar disorder, and 1 point was assigned each time a gene's GO term matched a relevant concept. For example, if a gene's GO terms were "acetylcholine

metabolism" and "dopamine receptor", the gene would be assigned 2 points, since both GO terms contain a concept considered relevant to bipolar disorder.

Gene Expression Data: Two publicly available data sets from the Gene Expression Omnibus (GEO) were used to prioritize genes in the interaction network: GSE5392 and GSE7036 [30, 31] (GSE5392 contains two data sets from different brain regions, and GSE7036 is a dataset of discordant monozygotic twins). Genes in the network were assigned 1 point for each data set in which it was significantly differentially expressed (unadjusted p-value $< 5 \times 10^{-4}$).

Data from the GEO datasets are from Affymetrix arrays and were analyzed with the Affy package within the R statistical computing environment under the Bioconductor framework. All data from the Affymetrix arrays were Robust Multichip Average (RMA) background corrected, normalized and summarized at the probeset level with median polish, Perfect Match (PM) only. Two-sample t-statistics for tests of differences in means (case vs. control) were performed at the probeset level. A paired t-statistic was also performed for the twins dataset (GSE7036).

Network Connectivity: Because of evidence showing that genes with high connectivity can sometimes play an important role in regulating the behavior of biological networks [32, 33], genes considered hubs in the interaction network were assigned higher scores. Genes with more than 200 connections were assigned 1 point, while those with more than 500 connections were assigned 2 points.

Once genes were assigned points for all evidence types, a scaling factor was applied. In this study, a scaling factor of 50 was used. For instance, a gene with a linkage score of 1 would have a scaled linkage score of 50.

Genes were assigned points for each type of evidence individually (i.e., each gene had a linkage score, a function score, a differential expression score, etc.). The scaled scores for each type of evidence were then summed to create a cumulative score. Genes without any evidence of involvement in bipolar disorder were simply assigned one point, so that they still had a small chance of being selected by the algorithm.

3.3 The Algorithm

A standard genetic algorithm was adapted to the task of searching for groups of SNPs that are jointly associated with bipolar disorder, in the context of a standard case-control GWAS. A genetic algorithm (GA) is a general purpose search and optimization algorithm based on the process of natural selection, and has been used in a variety of applications, from engineering design to RNA structure prediction [34, 35, 36, 37]. After all, natural selection can be seen as an algorithm for optimizing the fitness of a population of organisms, whereby the population repeatedly goes through the steps of selection, reproduction, and mutation, increasing the population's fitness over the generations.

For the GA, the population of organisms is replaced by a population of candidate solutions to some computational problem. And these candidate solutions go through the same steps of selection (an evaluation of how well a solution solves the problem), reproduction, and mutation. As the solutions evolve over time, the population moves

closer and closer to an optimal solution to the problem at hand. Figure 2 demonstrates a standard GA.

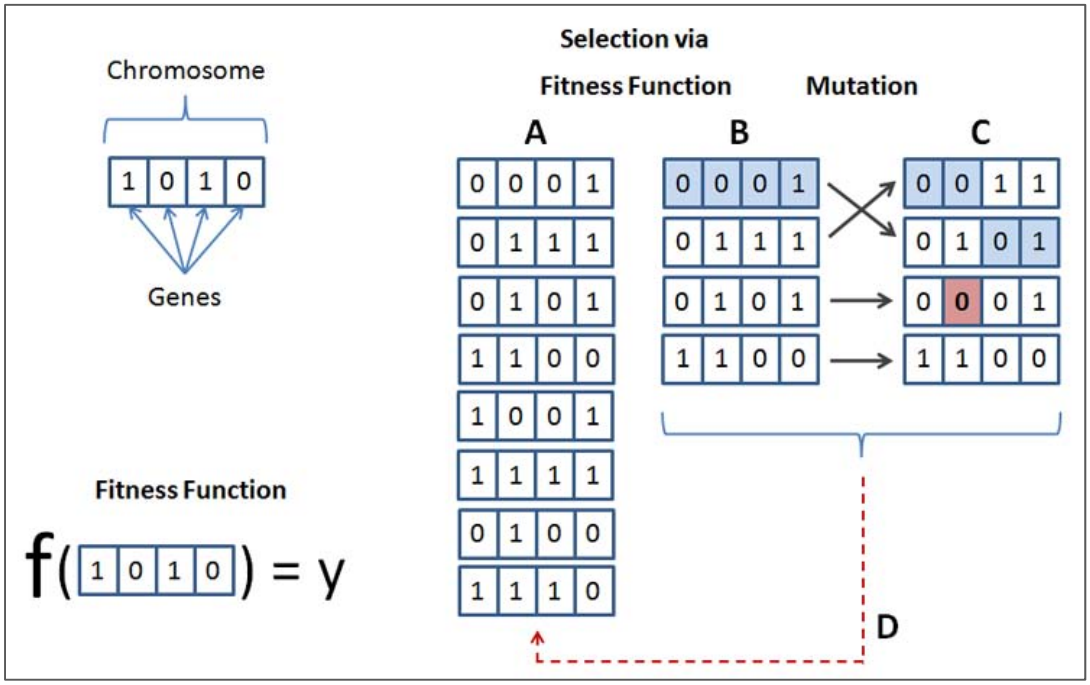

Figure 2. A standard genetic algorithm. The conventional data representation used in a genetic algorithm is a bit string (upper left), where each bit is analogous to a gene and the entire string is analogous to a chromosome. These bit strings represent candidate solutions for the problem of interest, and can be interpreted as either a binary number, or a list of features to be selected (as in a classification problem). The steps taken by the algorithm are as follows: A) a population of bit strings is randomly created. B) A fitness function (lower left) is applied to each candidate solution to evaluate how well it solves the problem, and based on these fitness values half of the population is selected. C) These selected candidate solutions are then altered slightly through crossover (the top two rows) or mutation (the third row). D) Finally, the selected solutions ("parents") and the newly

created altered solutions ("offspring") become the new population for the next iteration (or "generation").

For the purposes of the current problem, the discovery of genetic associations, the algorithm is constrained by the structure of a gene interaction network, where each node in the network represents a gene and all SNPs that are located within or near that gene. Figure 3 gives a basic overview of the algorithm, and the algorithm's mutation procedure is illustrated in Figure 4. The fitness value used by the algorithm is the p-value from a standard chi-square test, but any statistical test of association could have been used. A chi-square test was chosen because it is easily and quickly calculated.

Two different hypotheses about the nature of the SNP groups were tested. Hypothesis 1 requires that all members of a SNP group be within a connected sub-network of genes (i.e., each gene must interact with at least one other gene in the group). Hypothesis 2 is the more general case and does not require that the genes be directly connected. Because Hypothesis 2 does not require that the network connections between genes be preserved, an additional genetic operator, the crossover function, can be used to introduce variation into the population of SNP groups. In this case, uniform crossover with a probability of 0.5 was used. This means that if two SNP groups were chosen for crossover (based on the specified crossover rate), the members of each group would be sorted and the two groups compared. Any non-matching members would then be swapped with a 50 % probability. Note that the size of the SNP groups was limited to between two and four SNPs.

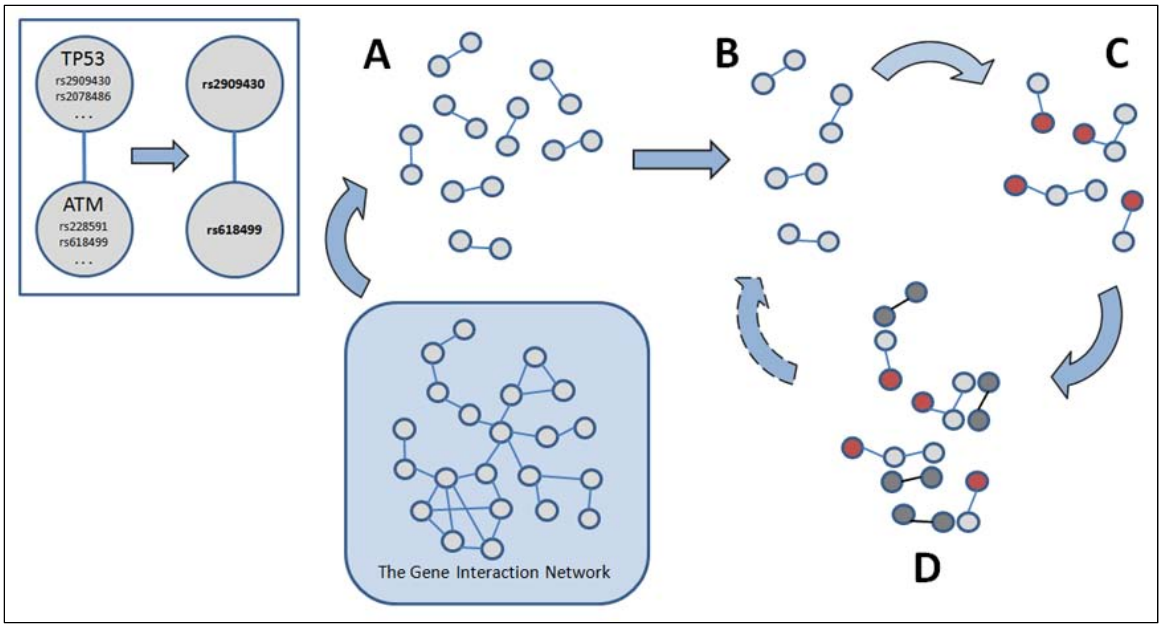

Figure 3. The genetic algorithm applied to a gene interaction network. The nodes in the gene interaction network represent genes and the SNPs that lie within or near each gene. The algorithm first selects a gene and then randomly chooses one of the gene's SNPs (top left). A) Using the structure of the interaction network, a population of SNP pairs is created randomly. B) The fitness of each SNP pair is calculated and, based on those fitness scores, half of the population is selected to become "parents" for the next generation. C) A specified proportion of parents are randomly altered, or mutated, to create "offspring". D) The parents plus the offspring become the new population and the process repeats, beginning with the calculation of fitness for each of the new offspring.

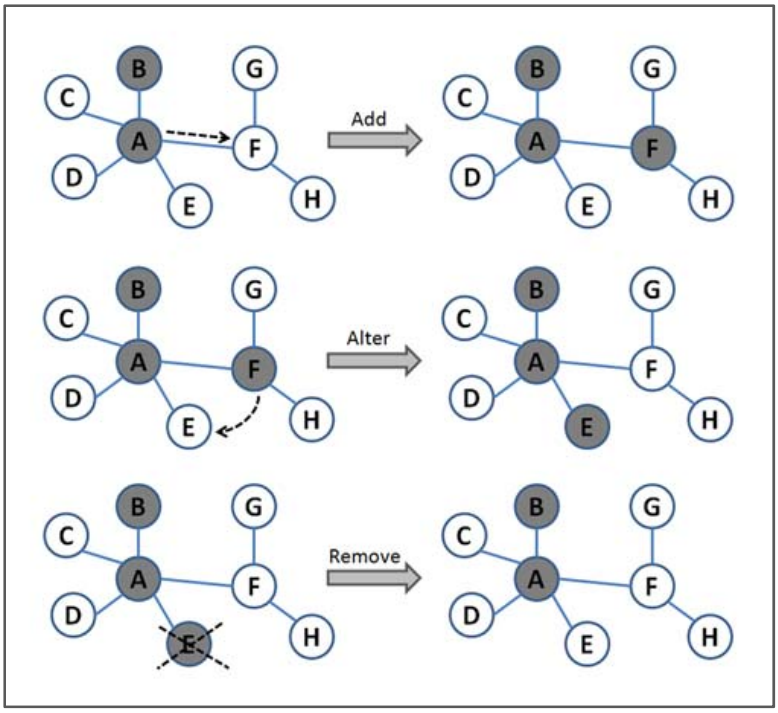

Figure 4. The "mutation" of a SNP group can occur in three ways, and is constrained by the structure of the gene interaction network. The nodes labeled A-F represent genes (each of which contain one or more SNPs, as shown in Figure 3), and the edges represent some evidence for interaction between genes (e.g., physical interaction, correlated expression, etc.).

As described above, because of the enormity of the search space, we compiled results from various studies about the genetics and pathophysiology of bipolar disorder in an attempt to force the algorithm to focus on certain areas of the network. Evidence from these studies, such as linkage regions or genes associated with the disease, was used to assign scores to genes in the interaction network. The algorithm is able to use the gene scores to preferentially select the most promising genes—those with prior evidence of involvement in bipolar disorder—as it searches the network for multi-locus associations.

These scores are used in two ways: 1) they influence the starting point of the search (i.e, the genes selected for the initial population of candidate solutions), and 2) they influence how the "mutation" procedure alters each SNP group. For instance, at the top of Figure 4, the algorithm is adding a gene to the group and must choose between genes C, D, E and F. If scores are applied to the gene interaction network, the algorithm will choose the gene—in a probabilistic manner—based on those scores. The algorithm won't automatically select the gene with the highest score, but rather it will choose randomly, with a higher probability given to those genes with higher scores.

Pseudocode for the algorithm, giving a more detailed description of the process, is included in the Appendix.

4 Results

4.1 The Algorithm Evaluation

Evaluation of the algorithm was done by performing multiple searches for multi-locus associations, and observing the effects of different parameter values. All analyses done in this stage of the project were done with only a subset of the data (TGen subjects only).The evaluation was done in four steps. First, to test the performance of the algorithm and to determine the most appropriate parameter values for the task at hand, a number of small-scale searchers were performed, each with different parameter combinations. While not meant to be a formal evaluation of all aspects of the algorithm, several features of the algorithm's behavior were demonstrated. (Note: I use the term "search" to mean a single execution of the algorithm, and the term "run" to mean a set of several searches all done with the same parameter values.)

Six different parameter combinations were tested for Hypothesis 1 and ten were tested for Hypothesis 2 (Table 1 and Table 2). For each parameter combination, five small-scale searches (200 generations) were performed and the results averaged. Various measures related to the effectiveness and efficiency of the search were recorded.

Run	Population Size	Generations	Crossover Rate	Mutation Rate	Selection	Migrants*
1	***50***	***200***	***NA***	***0.2***	***Truncate***	***0***
2	***100***	***200***	***NA***	***0.2***	***Truncate***	***0***
3	***50***	***200***	***NA***	***0.4***	***Roulette***	***0***
4	***100***	***200***	***NA***	***0.4***	***Roulette***	***5***
5	***50***	***200***	***NA***	***0.6***	***Hybrid***	***5***
6	***100***	***200***	***NA***	***0.8***	***Hybrid***	***10***

Table 1. Parameter combinations for the small-scale searches done under Hypothesis 1. *Migrants are the number of entirely new candidate solutions that are created randomly at the beginning of each generation (i.e., these new candidate solutions are not selected from the previous generation).

Run	Population Size	Generations	Crossover Rate	Mutation Rate	Selection	Migrants
7	***52***	***200***	***0.6***	***0.1***	***Truncate***	***0***
8	***100***	***200***	***0.4***	***0.1***	***Truncate***	***0***
9	***52***	***200***	***0.6***	***0.2***	***Roulette***	***0***
10	***100***	***200***	***0.4***	***0.2***	***Roulette***	***5***
11	***52***	***200***	***0.4***	***0.4***	***Hybrid***	***5***
12	***100***	***200***	***0.6***	***0.4***	***Hybrid***	***10***
13	***52***	***200***	***0.6***	***0.6***	***Hybrid***	***5***
14	***52***	***200***	***0.2***	***0.6***	***Hybrid***	***5***
15	***52***	***200***	***0.6***	***0.1***	***Hybrid***	***5***
16	***52***	***200***	***0.2***	***0.1***	***Hybrid***	***5***

Table 2. Parameter combinations for the small-scale searches done under Hypothesis 2.

Although the results vary significantly from one search to the next, certain trends were observed. For instance, it is clear that a higher mutation rate produces a broader search (more SNP groups are tested). This, of course, is expected since mutation is the method by which new candidate solutions (new SNP groups) are created. The same result can be achieved by increasing the GA population size (Figure 5). Furthermore, a larger

population size, when paired with an elitist selection method (one that necessarily selects the groups with the highest fitness), allows the algorithm to reach a good solution more quickly than a smaller population size (Figure 6). A larger population size increases the chance that good solutions (or potentially good solutions) are created in the early generations, and an elitist selection method ensures that those good candidates are carried on to the next generation. However, because of the tendency for the algorithm to become trapped in local minima, converging on a solution too quickly may be detrimental.

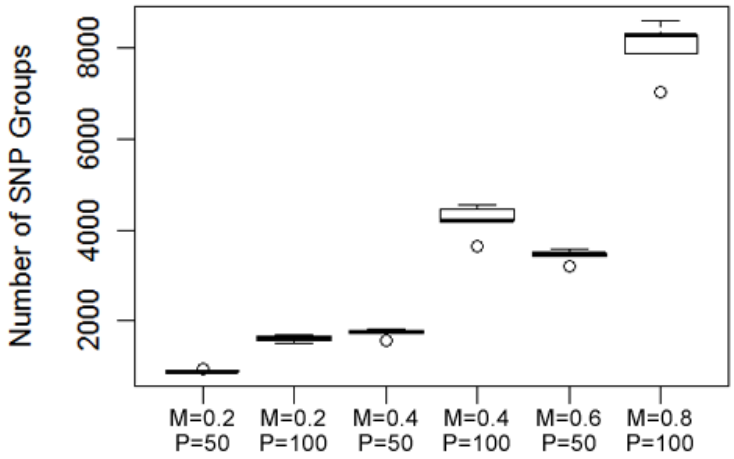

Figure 5. The number of SNP groups tested increases with GA population size and mutation rate. Each column shows the results of five searches done with the same parameters. M = mutation rate; P = population size. (Runs 1 through 6).

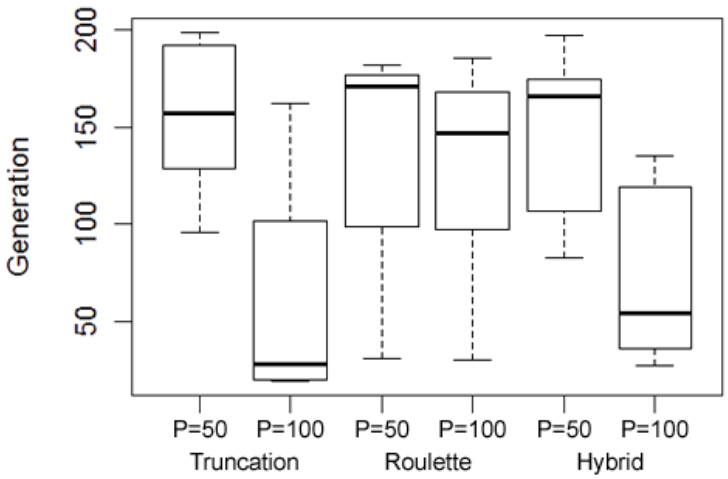

Figure. 6. Larger population size, when paired with an elitist selection method, allows the algorithm to converge on good solutions more quickly. Here we show the generation at which the best solution was found. P = population size. (Runs 1 through 6).

From the results of searches performed under Hypothesis 2, it appears the crossover rate had very little, if any, impact on performance. This suggests that exploring entirely new solutions created through mutation (and migration) is more important than recombining the current candidate solutions. But it is also clear that the impact of crossover depends on the selection method used and the rate at which new candidate solutions are introduced into the population (either through mutation or migration). However, a high crossover rate did not appear to have a negative effect, on solutions found or computation time (Figure 7), and it may be that the benefits of crossover were simply too subtle to be noticed in these small-scale searches. Crossover, as an additional method of creating diversity among candidate solutions, should not be discounted based on the results so far.

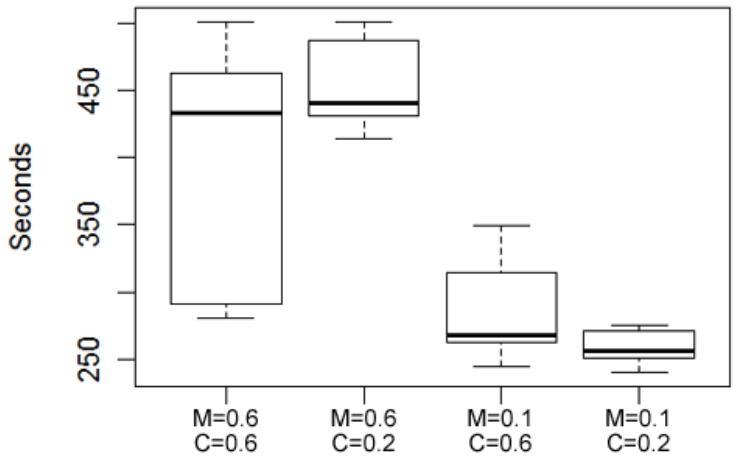

Figure 7. Algorithm run time was significantly impacted by the mutation rate, but NOT the crossover rate. M = mutation rate; C = crossover rate. (Runs 13 through 16).

Run	Best Fitness	-Log_{10}(Best)	Gen. of Best	Groups Tested	1st Gen. to 0.005	1st Gen. to 0.001	Time (sec.)
1	***1.16×10^{-3}***	***3.788***	***154.6***	***892.6***	***53.6***	***87.8***	***147.2***
2	***1.90×10^{-4}***	***3.857***	***66.2***	***1604.4***	***7.4***	***13***	***268***
3	***3.41×10^{-4}***	***3.857***	***132***	***1739.6***	***32.2***	***83.4***	***258.4***
4	***3.47×10^{-4}***	***3.657***	***125.6***	***4219***	***11***	***74.2***	***488.8***
5	***2.81×10^{-4}***	***4.728***	***145.6***	***3433.8***	***10***	***81.8***	***402.4***
6	***6.64×10^{-5}***	***4.225***	***74.2***	***8018.4***	***4.8***	***18.8***	***1029***

Table 3. Results of small-scale searches done under Hypothesis 1.

Run	Best Fitness	$-\text{Log}_{10}$(Best)	Gen. of Best	Groups Tested	1st Gen. to 0.005	1st Gen. to 0.001	Time (sec.)
7	***2.21×10^{-4}***	***3.847***	***81.4***	***551***	***5***	***68.2***	***171***
8	***1.14×10^{-4}***	***4.810***	***27.2***	***992.4***	***3***	***6.4***	***389.4***
9	***1.07×10^{-3}***	***3.363***	***83.6***	***984.4***	***9.2***	***106***	***232.4***
10	***1.59×10^{-4}***	***4.926***	***121.8***	***3674.8***	***2.6***	***19.2***	***450.2***
11	***5.90×10^{-5}***	***4.637***	***115.2***	***3393.6***	***6.4***	***25.6***	***357***
12	***1.37×10^{-5}***	***5.190***	***50.2***	***6717.8***	***2.8***	***7.4***	***604***
13	***1.72×10^{-5}***	***4.945***	***81***	***3986.8***	***6.4***	***15.8***	***393.6***
14	***1.84×10^{-5}***	***4.945***	***89.4***	***3836.2***	***6***	***15.8***	***454.6***
15	***3.52×10^{-5}***	***5.094***	***144.2***	***2867.4***	***3.2***	***32***	***287.4***
16	***2.12×10^{-5}***	***5.832***	***120.4***	***2104.2***	***12***	***17.4***	***258.6***

Table 4. Results of small-scale searches done under Hypothesis 2.

Results from the small-scale searches have revealed a number of important points about the behavior of the search algorithm. First, it's clear that creating and maintaining diversity among the population of candidate solutions is essential. A fairly high mutation rate and the creation of a few entirely new candidate solutions ("migrants") at each generation are important for broadening the search and discovering good solutions. Second, the Hybrid selection method appears to be the best choice. It ensures that a predefined number of "best" candidate solutions are passed on to the next generation, and it maintains diversity by allowing for probabilistic selection of the remaining candidate solutions. The effect of different selection methods can be seen in Figures 8, 9 and 10. Third, population size should be chosen carefully. A larger population size broadens the search and may allow for quicker discovery of good solutions, but also increases computation time. In addition, good solutions found early in a search may not always be the best possible solutions.

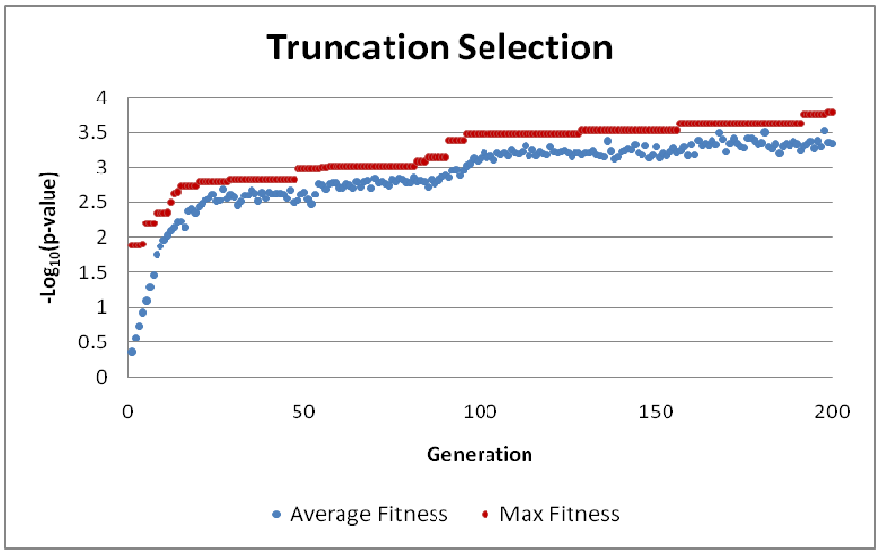

Figure 8. The truncation selection method (sometimes referred to as an elitist selection method) simply sorts the population of candidate solutions based on fitness values (p-values) and selects the top 50 % of the population. Since the best solutions are always chosen for the next generation, there is less diversity in the population. Notice there is little difference between the average fitness and the best fitness. (Run 1).

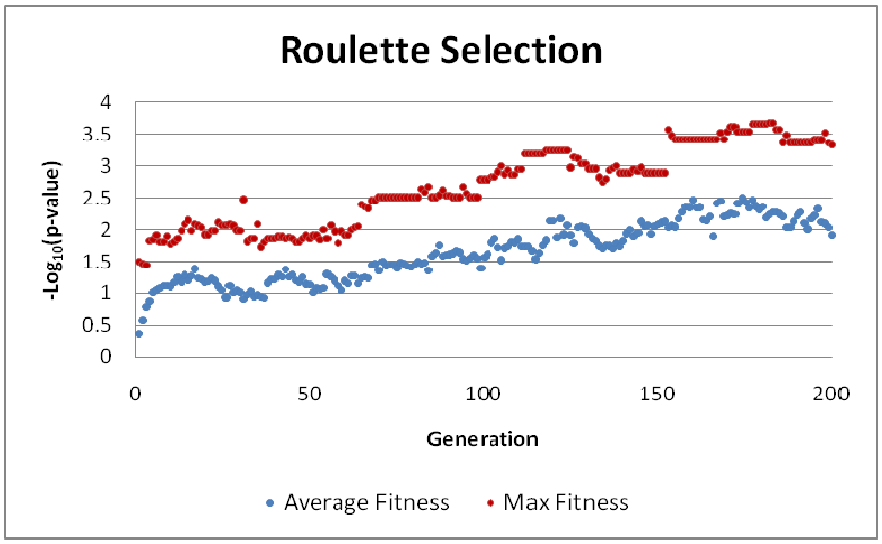

Figure 9. The roulette selection method randomly selects candidate solutions with a probability proportional to their fitness values (i.e., candidate solutions with small p-values are not automatically selected, but they have a higher probability of being

selected). A purely probabilistic selection method does not ensure that the best solutions are passed from one generation to the next. (Run 3).

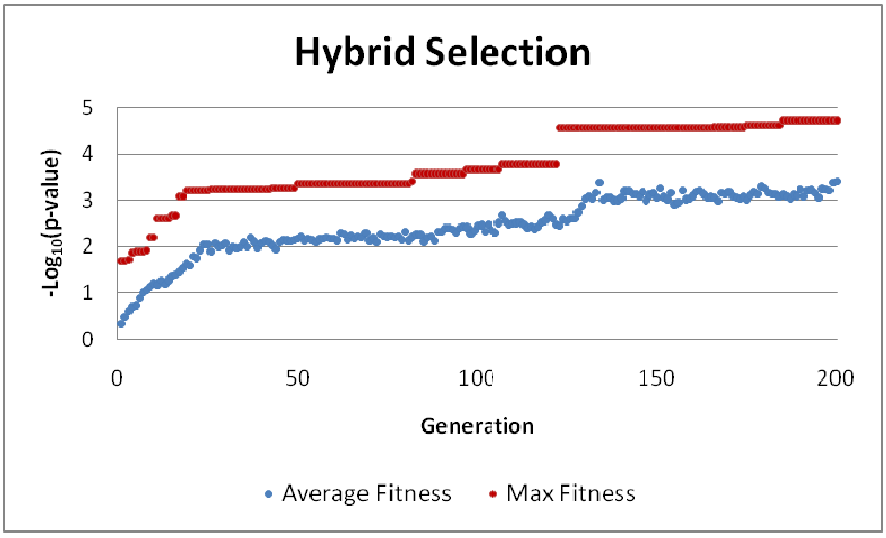

Figure 10. The hybrid selection method combines truncation selection with the roulette selection method. By combining an elitist selection method with a probabilistic method, the best solutions are passed on while still maintaining a good amount of diversity in the population. (Run 5).

In the second stage of algorithm evaluation, larger-scale searches were performed with the most appropriate parameter combinations learned from the earlier small-scale searches (Table 5 and Table 6). The goals here were to determine the number of generations needed to consistently find associations of genome-wide significance, to refine the algorithm parameters further, and to take a closer look at how thoroughly the algorithm is exploring the gene network.

Run	Population Size	Generations	Crossover Rate	Mutation Rate	Selection	Migrants
17	***100***	***1000***	***NA***	***0.6***	***Hybrid***	***10***
18	***100***	***2000***	***NA***	***0.6***	***Hybrid***	***10***
19	***150***	***1000***	***NA***	***0.6***	***Hybrid***	***15***
20	***150***	***1500***	***NA***	***0.8***	***Hybrid***	***15***
21	***50***	***5000***	***NA***	***0.8***	***Hybrid***	***5***

Table 5. Parameters for large-scale searches done under Hypothesis 1.

Run	Population Size	Generations	Crossover Rate	Mutation Rate	Selection	Migrants
22	***100***	***1000***	***0.6***	***0.4***	***Hybrid***	***10***
23	***100***	***2000***	***0.6***	***0.4***	***Hybrid***	***10***
24	***152***	***1000***	***0.6***	***0.6***	***Hybrid***	***15***
25	***152***	***1500***	***0.6***	***0.6***	***Hybrid***	***15***
26	***52***	***5000***	***0.6***	***0.6***	***Hybrid***	***5***
27	***100***	***2000***	***0.4***	***0.4***	***Hybrid***	***10***

Table 6. Parameters for large-scale searches done under Hypothesis 2.

The algorithm appears to have no problem exploring the vast majority of the gene interaction network. The large-scale searches visited between 85 % and 99 % of all genes in the network, and over 98 % of all chromosome cytobands. However, the majority of the search space complexity is due to the fact that multiple SNPs map to each gene (approximately 17 SNPs per gene on average). Despite nearly all genes in the network being visited during a typical large-scale search, fewer than 25 % of SNPs were ever included in a group and tested for association. So, while the search may be wide, its depth depends on how long the search is allowed to run. As seen so far, the algorithm may need many thousands of generations to adequately explore the search space.

Also, typical of GAs, the algorithm appears to be rather sensitive to starting conditions and therefore is somewhat inconsistent. While nearly all parameter

combinations tested in this stage were able to find statistically significant multi-locus associations, the best solutions varied significantly from one search to the next (Figure 11, and Tables 7 and 8).

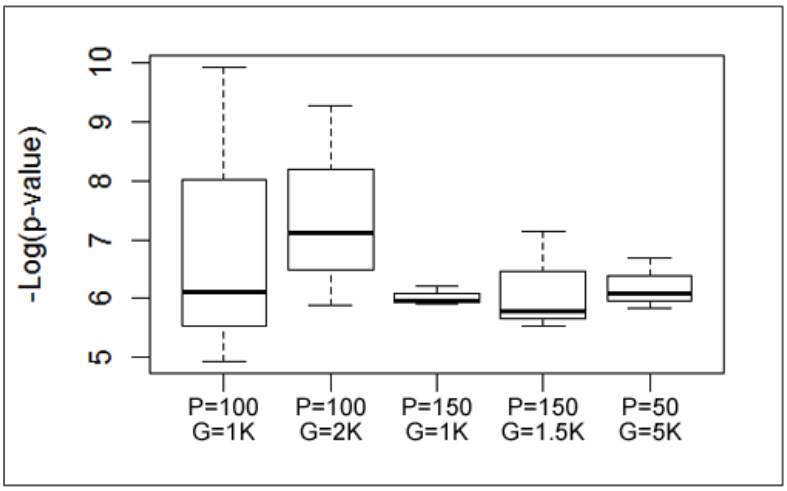

Figure 11. Results were inconsistent from one search to the next. Although differences in population size and run time didn't significantly impact the results across runs, the best solution found varied significantly within runs. All runs shown here were carried out under Hypothesis 1. Each column shows the results of three searches. P = population size; G = number of generations the GA was allowed to run. (Runs 17 through 21).

Run	Best Fitness	-Log_{10}(Best)	Gen. of Best	Groups Tested	1st Gen. to 0.005	1st Gen. to 0.001	Time (sec.)
17	***4.25×10^{-6}***	***6.9899***	***368.3***	***29210***	***4.3***	***15.3***	***3662***
18	***4.77×10^{-7}***	***7.4206***	***926.3***	***56630***	***14***	***22***	***10148***
19	***9.73×10^{-7}***	***6.0286***	***152***	***40922***	***3.3***	***12.3***	***6483***
20	***1.55×10^{-6}***	***6.1455***	***148.7***	***68213.3***	***2.3***	***7***	***12894***
21	***8.38×10^{-7}***	***6.1966***	***2803***	***76114***	***8.7***	***123.3***	***13753***

Table 7. Results of large-scale searches done under Hypothesis 1. Each row is the average of three searches.

Run	Best Fitness	$-\text{Log}_{10}$(Best)	Gen. of Best	Groups Tested	1st Gen. to 0.005	1st Gen. to 0.001	Time (sec.)
22	***1.29×10^{-5}***	***5.8454***	***567.7***	***30999***	***2.3***	***9***	***3257***
23	***1.17×10^{-5}***	***6.6555***	***318.3***	***58077***	***1.7***	***3***	***6952***
24	***2.84×10^{-7}***	***6.7865***	***905.7***	***47876***	***1***	***3.7***	***4978***
25	***1.24×10^{-7}***	***7.8837***	***696.7***	***72565***	***1.7***	***6***	***8723***
26	***1.43×10^{-6}***	***6.4416***	***1644***	***72883***	***3.7***	***10.3***	***8367***
27*	***4.82×10^{-10}***	***9.4461***	***297.7***	***24761***	***1.7***	***12***	***4746***

Table 8. Results of large-scale searches done under Hypothesis 2. Each row is the average of three searches. *Two of the searches in Run 27 were stopped early, when a p-value of 5×10^{-10} was reached.

These larger-scale searches also shed more light on the impact of GA population size on the algorithm's efficiency. And it appears that the impact may be different for Hypothesis 1 versus Hypothesis 2. When searching for connected sub-networks (Hypothesis 1) it seems that there is no benefit to increasing the population size beyond a certain point. Although searches with larger population sizes reached their best solutions more quickly, they were never able to reach solutions as good as some found with a moderate population size.

For Hypothesis 2, larger population size did appear to have some benefit. A larger search space (i.e., groups are not limited to connected sub-networks) and the addition of the crossover function as a source of variation in the population, may allow the algorithm to take advantage of a larger population size under Hypothesis 2.

Table 9 shows the parameter values that were found to be most successful in the large-scale searches.

Hypothesis 1	
Population Size	***100***
Mutation Rate	***0.6 – 0.8***
Crossover Rate	***NA***
Selection Method	***Hybrid***
No. of Migrants	***10***
Hypothesis 2	
Population Size	***100 – 152[a]***
Mutation Rate	***0.4 – 0.6***
Crossover Rate	***0.4 – 0.6***
Selection Method	***Hybrid***
No. of Migrants	***10***

Table 9. After testing a variety of parameter values in multiple small- and large-scale searches, these values produced the best results. [a]For Hypothesis 2 the population size must be a multiple of 4.

Next, in an effort to deal with the problem of inconsistency and to help improve the depth of the searches, two strategies were implemented. The first strategy was to run the algorithm for a longer time (10000 generations), while periodically (every 1000 generations) re-initializing the GA population. In essence a new search, with a new starting point within the interaction network, is created every 1000 generations. The second strategy, similar to the first, was to re-initialize the population only when no improvement is seen in the previous 1000 generations. These searches were stopped if they reached a p-value $< 5 \times 10^{-10}$ (Table 10).

Run	Population Size	Generations	Crossover Rate	Mutation Rate	Selection	Migrants
28	***100***	***10000***	***NA***	***0.6***	***Hybrid***	***10***
29	***100***	***10000***	***NA***	***0.6***	***Hybrid***	***10***
30	***100***	***10000***	***0.4***	***0.4***	***Hybrid***	***10***
31	***100***	***10000***	***0.4***	***0.4***	***Hybrid***	***10***

Table 10. Parameters for large-scale searches with re-initialization. Runs 28 and 29 were done under Hypothesis 1, while runs 30 and 31 were done under Hypothesis 2.

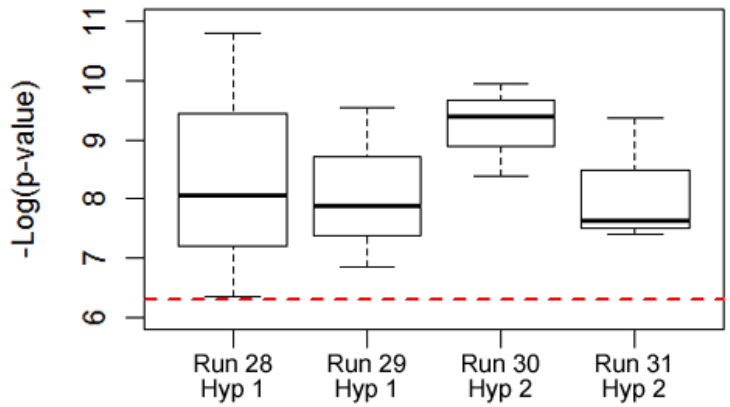

Figure 12. With a longer run time and periodic re-initialization of the GA population, the algorithm was able to consistently find solutions of genome-wide significance (p-values $< 5 \times 10^{-7}$; the dashed line). Runs 28 and 30 were re-initialized every 1000 generations; Runs 29 and 31 were re-initialized only when there was no improvement in the previous 1000 generations.

Run	Best Fitness	$-\text{Log}_{10}$(Best)	Gen. of Best	1st Gen. to $5\text{x}10^{-7}$	Groups Tested	Unique SNPs Visited
28	***1.48×10^{-7}***	***8.4116***	***6040***	***2419.3***	***226139***	***90249***
29	***5.14×10^{-8}***	***8.0987***	***7273***	***7223.3***	***262181***	***101586***
30	***1.58×10^{-9}***	***9.2314***	***4002***	***966***	***156835***	***64855***
31	***2.09×10^{-8}***	***8.1384***	***5053***	***1624.7***	***270091***	***94810***

Table 11. Results of large-scale searches with re-initialization. Runs 28 and 29 were done under Hypothesis 1, while runs 30 and 31 were done under Hypothesis 2. Runs 28 and 30 were re-initialized every 1000 generations; Runs 29 and 31 were re-initialized only when there was no improvement in the previous 1000 generations. Each row is the average of three searches.

Nearly all the larger-scale searches were able to find multi-locus associations with genome-wide statistical significance. And the strategy of increasing the search time while periodically re-initializing the GA population, improved the results further (Figure 12 and Table 11).

Not only was it possible to find multi-locus associations, it was possible to do so while performing far fewer statistical tests than a standard GWAS. For instance, Run 30 (three separate searches, with a total of 470467 SNP groups tested) was able to find 54 multi-locus associations that are statistically significant even after Bonferroni correction ($\alpha = 1.063 \times 10^{-7}$). Furthermore, nearly half of these associations (23 of 54) are interactions among three SNPs, which demonstrate the importance of looking for higher-order interactions. These results highlight an advantage of this approach over previous methods, which have mostly looked for interactions among pairs of SNPs only.

To examine whether the distribution of p-values found with the algorithm is different from what is expected by chance, large-scale searches were performed on two sets of

simulated data, one simulated using MERLIN [38] and another simulated with PLINK [39]. MERLIN is able to simulate genotype data while maintaining the patterns of missing data and linkage disequilibrium found in the actual bipolar disorder dataset (TGen). The distribution of p-values found when searching the simulated datasets was compared to the distribution of p-values found when searching the actual data (Figure 13). Clearly, the search performed on the actual bipolar disorder data discovered an excess of SNP groups with p-values smaller than 10^{-7}. Very similar results were seen with the data simulated in PLINK (data not shown).

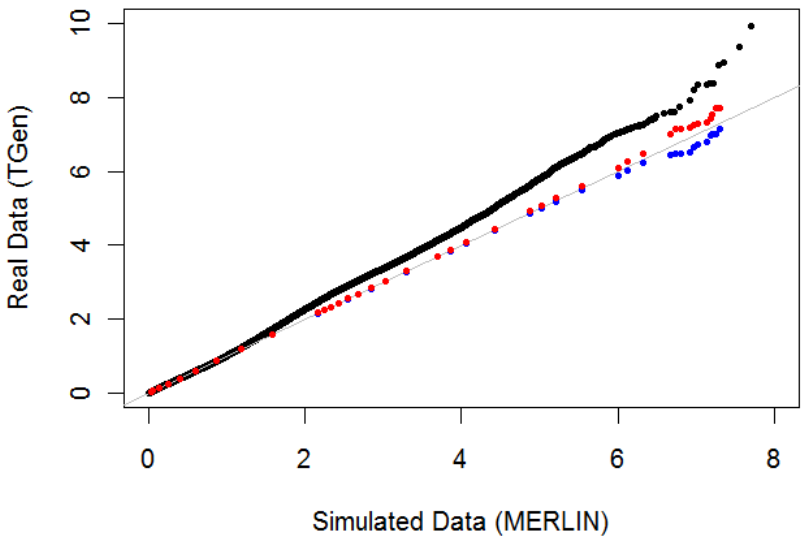

Figure 13. A quantile-quantile plot of the log scale p-values from a large-scale search for multi-locus associations among the actual bipolar disorder GWAS data versus data simulated with MERLIN. The simulated data was created so that no markers are associated with the disease. Also, patterns of missing data and linkage disequilibrium were simulated to match those of the real data. Red and blue dots represent the upper and lower bounds of a 95 % confidence interval for the simulated data quantiles.

The third step in the evaluation process was to perform a single-locus analysis, which revealed only two SNPs that had a significant association with the BPI group (p-values of 2.70×10^{-9} and 1.97×10^{-7}). Nevertheless, numerous multi-locus associations were found during the large-scale searches described in the previous section. To better understand the types of associations discovered, all statistically significant associations found during the most successful runs (Runs 28 – 31) were categorized in two ways. "Significant Interactions" were those where the joint p-value is smaller than all of the group members' individual p-values. And SNP groups were said to be "Absent of Main Effects" if none of the group members had an individual p-value that reached genome-wide significance (p-value $< 5 \times 10^{-7}$) (Table 12).

Run	Significant Groups	Significant Interactions	Absent of Main Effects
28	***37***	***26***	***16***
29	***73***	***9***	***8***
30	***169***	***88***	***84***
31	***82***	***59***	***57***

Table 12. The algorithm is able to detect multi-locus associations both with and without main effects. It appears that the algorithm is more successful under Hypothesis 2 (Runs 30 and 31). Each run represents the combined results of three searches.

The algorithm was able to find statistically significant multi-locus associations, both with main effects and without (pure interactions), under Hypothesis 1 and 2. However, it appears that finding pure interactions is more difficult under Hypothesis 1. This is not surprising, given that Hypothesis 1 requires that a SNP group be part of a connected sub-network, which limits the number of possible gene combinations.

Given the nature of the search algorithm there was some concern that the search may become focused on a small area of the gene network, and therefore a small area of the genome. To determine the distribution of these significant associations across the genome, the number of chromosomes, cytobands, genes, and SNPs present in the SNP groups were counted (Table 13). It's clear that these multi-locus associations are made up of genes spread widely across the genome. For instance, in Run 30, three searches detected 169 significant associations that contained 164 different genes (174 SNPs) from 23 chromosomes.

Run	Sig. Groups	Chrs.	Cytobands	Genes	SNPs
28	***37***	***19***	***37***	***39***	***51***
29	***73***	***19***	***37***	***40***	***76***
30	***169***	***23***	***119***	***164***	***174***
31	***82***	***22***	***77***	***95***	***94***

Table 13. The most significant multi-locus associations contain SNPs that are spread across the entire genome. However, it appears that many of the SNP groups found during a single search are dominated by one or two SNPs that appear in multiple groups (Figure 12).

However, it is sometimes the case that a few SNPs dominate the most significant SNP groups. For instance, consider the situation where a SNP with a very strong main effect is directly connected to ten other SNPs in the interaction network. It is possible that the algorithm will report all ten of these SNP pairs (the one strongly associated SNP paired with each of its neighbors) as significant multi-locus associations (Figure 14). While the algorithm presented here is a good first step in the search for multi-locus associations, it

is obvious that the SNP groups found will have to be scrutinized before being considered for any further study.

Following best practices for the analysis of GWAS data, we manually examined the clusters plots of SNPs in the most significant multi-locus associations. No anomalies were detected that would render the SNPs and corresponding associations as suspect.

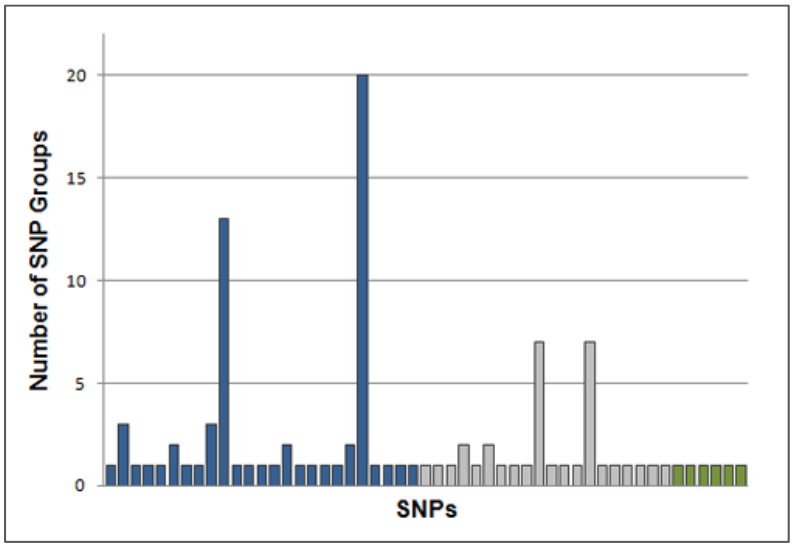

Figure 14. Some SNPs are found in multiple significant SNP groups. Here the columns represent each of the SNPs that were members of significant multi-locus associations found in the three searches of Run 28. In the first search (blue) there is one SNP that is involved in nearly all of the multi-locus associations (20 out of 23 SNP groups).

Although the results shown so far are promising, it's clear that the enormity of the search space poses a considerable challenge. One technique that may increase the chances of discovery is to assign scores to genes in the interaction network. Those genes with the highest scores (i.e., those thought to be involved in the disease of interest), and the SNPs located in or near them, will be preferentially chosen by the algorithm. While

this method may help to focus the search on the most promising genes, it does not reduce the search space outright.

All searches done so far in this study were done with a uniform distribution of gene scores (i.e., all genes were equally likely to be selected by the algorithm). In the next stage of the algorithm evaluation, various types of evidence (linkage studies, association studies, gene expression studies, gene functions, and network connectivity) were used to score the genes in the network. Points were assigned based on the strength of the evidence for involvement in the disease (e.g., the number of studies that reported a particular finding). Details about how scores were assigned are given in the Methods section above, and references for all the studies used as evidence are listed in the Appendix.

The distribution of cumulative gene scores (scores based on all of the evidence types combined) in the network follows an exponential distribution. Of the 12793 genes in the network, over half (6444 genes) are not implicated in bipolar disorder by any of the evidence compiled; 3643 genes have a score of 100 or greater (meaning they were implicated by a moderate amount of evidence); and only 45 genes have a score of 300 or greater (strong evidence for involvement in bipolar disorder).

The genetics of bipolar disorder have been studied extensively, so there is a great deal of information about the genes and cellular processes that are hypothesized to play a role in the disease. But it is very difficult to say that what works in the case of bipolar disorder will also work for other diseases (or even other datasets). Here I will simply point out some examples that show the potential value of integrating prior knowledge about the disease when searching for multi-locus associations.

Three searches (Table 14) were performed for each evidence type alone, and another three searches were done with all evidence cumulatively (i.e., the scores from all evidence types were added together). P-values from the most significant multi-locus associations detected during these searches are shown in Figure 15. It's difficult to make definitive conclusions from just a few searches, but it appears that linkage studies and gene association studies were helpful in guiding the search. The other evidence types were not helpful, and may actually have been detrimental. For instance, when the search was guided by network connectivity measures (focusing on hub genes) results appear worse than those achieved previously, without gene scores applied to the network.

Population Size	Generations	Crossover Rate	Mutation Rate	Selection	Migrants
100	***3000***	***0.4***	***0.4***	***Hybrid***	***10***

Table 14. Parameter combination for large-scale searches with gene prioritization. All searches were done under Hypothesis 2.

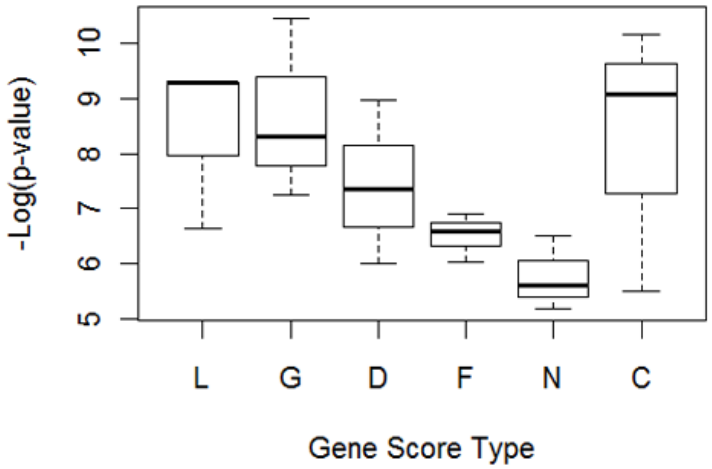

Figure 15. Not all evidence types were useful for guiding the search. Here we show the results from the searches done after gene prioritization. Each column represents the

results from three searches. L = linkage studies; G = gene association studies; D = differential expression; F = gene functions (GO terms); N = network connectivity (hubs); C = cumulative (all evidence types).

When linkage regions previously associated with bipolar disorder were used to assign gene scores, the algorithm consistently found associations of genome-wide significance within 650 generations. Furthermore, of the 258 SNPs that were part of all the significant multi-locus associations discovered, 155 (60 %) lie within the linkage regions previously associated with the disease.

When gene scores were based on previous implication in an association study, one search found a multi-locus association with the second smallest p-value (3.4×10^{-11}) found among all the searches performed so far. This association contained one SNP with a main effect and another that is directly connected in the network to a gene (*BDNF*) that has been previously implicated in two association studies of bipolar disorder (see Appendix).

To determine whether or not the gene scores were actually influencing the searches, gene scores from the top hits were compared between the searches done with the cumulative gene scores and Runs 30 and 31, which were done before gene prioritization. For SNPs from the top hits found by a search guided by cumulative gene scores, 86 % were in genes with a cumulative score of 100 or greater (moderate evidence of involvement in bipolar disorder). When gene scores were not used (Runs 30 and 31), only 39 % and 44 % of top-hit SNPs, respectively, were in genes with cumulative scores of 100 or greater. These results suggest that using a scoring method to prioritize the genes in

the interaction network may truly be able to improve the algorithm's efficiency by focusing the search on a set of particularly promising genes.

4.2 The GA vs. Random Search

In the final stage of the algorithm evaluation an attempt was made to show the computational benefit gained from using the genetic algorithm. This was done by comparing the results achieved when searching the network using the GA versus the results achieved when using a purely random search (i.e., at each generation SNP groups were chosen randomly, without any information from previous generations). Three searches using the GA, and three random searches were performed. All searches were allowed to run for 10,000 generations, but were stopped early if a p-value of 5×10^{-10} was reached.

It's clear that the GA was much more efficient in finding SNP groups with small p-values compared to the purely random search. Although the random search was able to find SNP groups with p-values less than 5×10^{-10}, the GA was able to find many more good solutions while performing far fewer statistical tests (Table 15). While fewer tests may not provide a statistical advantage for the GA, it definitely shows that the GA is able to find more good solutions given the same amount of computing time.

Search Type	Total SNP Groups Tested	Significant SNP Groups (p-value < 5×10^{-7})
Genetic Algorithm	***470467***	***168***
Random Search	***1988215***	***43***

Table 15. The GA vs. Random Search. The GA was able to detect four times as many significant multi-locus associations, while performing four times fewer tests, compared to a purely random search. The numbers shown are the combined results of three searches.

4.3 Multi-locus Associations Discovered

The algorithm evaluation described above provided insight into the search algorithms capabilities, and provided a set of parameter values likely to produce the best possible results. But in the evaluation phase, none of the associations discovered were examined more closely to determine biological relevance—the focus was on the statistical properties of the results. The next step in the project was to apply the algorithm to the full data set (TGen and GAIN samples combined) in an attempt to find multi-SNP associations that might shed some light on the biology of bipolar disorder.

Three separate searches were performed iteratively on the full data set. The first search was performed without the use of any prior knowledge about the genetics of bipolar disorder (i.e., information from previous linkage, association studies, etc. was not used to guide the search). The second search was identical to the first except that 14 SNPs, which were detected as part of statistically significant multi-locus associations by the first search, and later determined to have bad cluster plots, were removed from the analysis. The third search used information from previous linkage and gene association studies to focus on a set of genes thought to be involved in bipolar disorder. As

mentioned previously, this information was merely used to increase the probability that certain genes (and therefore the SNPs mapped to those genes) are selected by the algorithm, but it does not actually remove any genes from the algorithm's search space.

In total, across all three searches, 1632639 SNP groups were tested for association. Although the searches discovered hundreds of multi-locus associations with chi-square p-values $< 5 \times 10^{-7}$, only the 25 SNP groups with the smallest p-values from each search were examined closely. None of these statistically significant multi-locus associations contained SNPs with individual p-values smaller than 5×10^{-7} (i.e., no main effects).

This list of top hits was narrowed further by removing SNPs with bad cluster plots, and by discarding any SNP groups with very small numbers in their contingency table cells (most common in groups containing more than two SNPs). There was also a fair amount of overlap between SNP groups. For instance, when a SNP pair is strongly associated with the disease it is likely that a large number of triplets containing that pair will also be significantly associated. In most cases, the addition of the third SNP does not increase the strength of the association and so does not add any information. These uninformative triplets were discarded.

The reporting of odds ratios in the context of multi-locus associations is a complicated matter. In the case of single SNPs it makes sense to set one of the homozygous genotypes as the reference genotype and then calculate odds ratios for the other two genotypes relative to that reference. But when looking at the joint effects of two SNPs there are four different double homozygous genotype combinations. Emily and colleagues chose to report odds ratios relative to the most common double homozygous genotype combination [22]. However, we preferred to first examine all contingency

tables—the joint contingency table, as well as the contingency tables for each individual SNP in the multi-locus association. By looking at the SNPs individually we were able to determine, for those with at least a weak association, which allele was the "risk allele". Next we examined the joint contingency table to determine whether the multi-locus association fit an additive model (risk increases with the number of risk alleles) or a conditional model (the effect of one SNP depends on the genotype of a second SNP). Odds ratios were then reported in the context of these two specific models.

The three SNP pairs that survived the data cleaning process and that fit either an additive or conditional model are discussed in detail below:

The data suggests an additive or complementary effect between intronic SNPs in *BMP7* (rs6127985) and *CREB1* (rs10932201). Niether SNP was significantly associated with bipolar disorder in our data (chi-square p-values of 0.00282 and 0.000179, respectively). However, disease risk appears to increase with the cumulative number of risk alleles from the two SNPs (Table 20).

***BMP7* (rs6127985)**			
Genotype	Controls	Cases	OR
1	***959***	***1476***	***NS***
2	***412***	***497***	***NS***
3	***46***	***49***	***--***

Table 16. Genotype frequencies for rs6127985. 1 = G/G, 2 = A/G, 3 = A/A. NS = Not Significant. (Chi-square p-value = 2.82×10^{-3})

***CREB1* (rs10932201)**			
Genotype	Controls	Cases	OR
1	***369***	***658***	***1.28 (1.05, 1.56)***
2	***747***	***952***	***NS***
3	***304***	***423***	***--***

Table 17. Genotype frequencies for rs10932201. 1 = G/G, 2 = A/G, 3 = A/A. (Chi-square p-value = 1.79×10^{-4})

rs6127985 x rs10932201			
Genotype	Controls	Cases	OR
1,1	***233***	***490***	***3.41 (1.72, 6.95)***
1,2	***541***	***670***	***2.01 (1.03, 4.06)***
1,3	***185***	***316***	***2.77 (1.39, 5.68)***
2,1	***125***	***146***	***NS***
2,2	***179***	***261***	***2.36 (1.18, 4.87)***
2,3	***108***	***90***	***NS***
3,1	***10***	***18***	***NS***
3,2	***26***	***16***	***--***
3,3	***10***	***15***	***NS***

Table 18. Genotype combination frequencies for the SNP interaction. (Chi-square p-value = 7.48×10^{-10})

Additive Model			
Risk Alleles	Controls	Cases	OR
0	***10***	***15***	***--***
1	***134***	***106***	***NS***
2	***374***	***595***	***NS***
3	***666***	***816***	***NS***
4	***233***	***490***	***NS***

Table 19. The ratio of cases to controls tends to increase with the burden of risk alleles from the two SNPs. (Chi-square test of trend in proportions p-value = 1.97×10^{-5})

Collapsed Additive Model

Risk Alleles	Controls	Cases	OR
0 – 1	***144***	***121***	***--***
2 – 3	***1040***	***1411***	***1.61 (1.24, 2.10)***
4	***233***	***490***	***2.50 (1.86, 3.37)***

Table 20. When the groups are collapsed into low, moderate and high risk categories, the trend is much more apparent. (Chi-square test of trend in proportions p-value = 2.86×10^{-11})

An interactive effect between SNPs in *GFRA1* (rs3781538) and *ENPP2* (rs2305125) was found (Table 23), whereby the association of the SNP in *GFRA1* is strengthened when conditioned upon the G/G genotype of the SNP in *ENPP2*.

GFRA1 **(rs3781538)**

Genotype	Controls	Cases	OR
1	***524***	***650***	***--***
2	***697***	***1017***	***1.17 (1.01, 1.37)***
3	***198***	***366***	***1.48 (1.20, 1.85)***

Table 21. Genotype frequencies for rs6127985. 1 = C/C, 2 = C/T, 3 = T/T. (Chi-square p-value = 2.82×10^{-3})

ENPP2 **(rs2305125)**

Genotype	Controls	Cases	OR
3	***365***	***520***	***--***
2	***758***	***957***	***NS***
1	***294***	***554***	***1.32 (1.08, 1.61)***

Table 22. Genotype frequencies for rs10932201. 1 = G/G, 2 = A/G, 3 = A/A. (Chi-square p-value = 1.79×10^{-4})

rs3781538 x rs2305125			
Genotype	Controls	Cases	OR
1,1	***116***	***137***	***--***
1,2	***269***	***332***	***NS***
1,3	***138***	***181***	***NS***
2,1	***134***	***299***	***1.89 (1.35, 2.64)***
2,2	***385***	***464***	***NS***
2,3	***176***	***254***	***NS***
3,1	***44***	***118***	***2.27 (1.46, 3.57)***
3,2	***103***	***161***	***NS***
3,3	***51***	***85***	***NS***

Table 23. Genotype combination frequencies for the SNP interaction. (Chi-square p-value = 3.26×10^{-7})

Another interactive effect was detected between SNPs in *SLC35F1* (rs1334834) and *NDUFV2* (rs11661711), two genes involved in membrane transport (Table 26). In this case, the association with the SNP in *SLC35F1* is strengthened when conditioned upon the A/A genotype of the SNP in *NDUFV2*.

***SLC35F1* (rs1334834)**			
Genotype	Controls	Cases	OR
1	***225***	***432***	***1.51 (1.24, 1.83)***
2	***729***	***928***	***--***
3	***466***	***671***	***1.13 (0.97, 1.32)***

Table 24. Genotype frequencies for rs6127985. 1 = T/T, 2 = C/T, 3 = C/C. (Chi-square p-value = 9.70×10^{-5}).

***NDUFV2* (rs11661711)**			
Genotype	Controls	Cases	OR
1	***533***	***780***	***--***
2	***683***	***948***	***NS***
3	***203***	***301***	***NS***

Table 25. Genotype frequencies for rs10932201. 1 = A/A, 2 = A/G, 3 = G/G. (Chi-square p-value = 0.7).

rs1334834 x rs11661711			
Genotype	Controls	Cases	OR
1,1	***69***	***174***	***2.34 (1.68, 3.28)***
1,2	***119***	***211***	***1.64 (1.24, 2.19)***
1,3	***37***	***46***	***NS***
2,1	***295***	***318***	***--***
2,2	***324***	***466***	***1.33 (1.07, 1.66)***
2,3	***109***	***143***	***NS***
3,1	***169***	***287***	***1.57 (1.22, 2.03)***
3,2	***240***	***270***	***NS***
3,3	***57***	***113***	***1.84 (1.27, 2.68)***

Table 26. Genotype combination frequencies for the SNP interaction. (Chi-square p-value = 1.27×10^{-7}).

4.4 Replication Analysis

Genotype data was obtained from the WTCCC for the purposes of replicating the three significant associations. Because of the different genotyping array used in the WTCCC study, it was only possible to examine the association with the SNPs in *BMP7* (rs6127985) and *CREB1* (rs10932201) in this data set. Futhermore, the genotypes for SNP rs10932201 were arrived at through imputation, as this SNP was not directly

genotyped by the Affymetrix 500K Mapping array. Only genotypes with a call probability greater than 0.9 were included in the analysis.

Neither SNP had even a borderline association with the disease in the WTCCC data, as was seen in the BiGS data. Also, for SNP rs6127985 the A allele would be considered the "risk allele" (the allele more common in cases than controls) in the WTCCC data. The opposite allele (G) was considered the "risk allele" in the BiGS data set. However, this discrepancy does not automatically discount this SNP. Because these SNPs should be considered merely as markers for a causal variant, it is possible that different marker alleles have become correlated with the true causal allele in different populations.

Despite this difference, there does appear to be a trend towards higher risk of disease with a greater burden of "risk alleles" (Table 30). While this is far from strong support for this association, it does suggest that SNPs in *BMP7* and *CREB1* should remain candidates for susceptibility markers in bipolar disorder. And an additive effect between these genes should be examined further in other data sets.

***BMP7* (rs6127985) from WTCCC**

Genotype	Controls	Cases	OR
1	***1015***	***1305***	***--***
2	***383***	***476***	***NS***
3	***36***	***54***	***NS***

Table 27. Genotype frequencies for rs6127985. 1 = G/G, 2 = A/G, 3 = A/A. (Chi-square p-value = 0.69)

***CREB1* (rs10932201) from WTCCC**

Genotype	Controls	Cases	OR
1	***413***	***553***	***NS***
2	***643***	***911***	***1.23 (1.02, 1.49)***
3	***309***	***354***	***--***

Table 28. Genotype frequencies for rs10932201. 1 = G/G, 2 = A/G, 3 = A/A. (Chi-square p-value = 0.074)

Additive Model from WTCCC

Risk Alleles	Controls	Cases	OR
0	***206***	***220***	***--***
1	***524***	***709***	***1.27 (1.01, 1.59)***
2	***441***	***553***	***NS***
3	***125***	***171***	***NS***
4	***7***	***19***	***2.53 (1.00, 7.29)***

Table 29. The ratio of cases to controls tends to increase with the burden of risk alleles from the two SNPs, although the trend is not statistically significant. (Chi-square test of trend in proportions p-value = 0.13)

Collapsed Additive Model from WTCCC

Risk Alleles	Controls	Cases	OR
0	***206***	***220***	***--***
1 – 2	***965***	***1262***	***1.22 (0.99, 1.52)***
3 – 4	***132***	***190***	***1.34 (1.00, 1.83)***

Table 30. When the groups are collapsed into low, moderate and high risk categories, the trend is more apparent. Note: The categories are defined differently than in the BiGS data to avoid having small numbers in any of the cells. (Chi-square test of trend in proportions p-value = 0.036)

Unfortunately, it was not possible to test for the other two associations in the WTCCC data, because two of the SNPs (one from each SNP-pair) were not genotyped (or imputed) in the WTCCC study. An attempt was made to use adjacent SNPs as surrogates for the actual SNPs of interest. However, when these surrogate associations were tested, neither association was significant. Negative results from these tests do not necessarily mean a negative result for the actual SNPs of interest, and therefore they do not provide any useful information. Moreover, when these adjacent SNPs were examined more closely it was evident they were not significantly correlated with the SNPs of interest.

Additional replication of all associations will need to be done in other data sets.

4.5 Supporting Evidence for the Associations

The replication analysis done with the WTCCC data set failed to provide much support for the associations detected in the BiGS data. Still, there is a wealth of information in the literature about the genetics and pathophysiology of bipolar disorder that may be able lend additional evidence for the implicated genes, and shed some light on the possible roles they play in the disease. A thorough search of the literature revealed a number of studies that have previously associated the genes of interest with bipolar disorder, and others which provide an explanation for how these genes may be involved in the disease.

***BMP7* and *CREB1*:** There is evidence that both *BMP7* and *CREB1* play a role in neuron development, specifically axon guidance and axonogenesis [40, 41]. In addition, both genes have been shown to have significant or borderline differential expression in

post-mortem brain tissues of bipolar disorder patients [42, 30]. One study suggests a borderline association between *BMP7* and major depressive disorder [43], while other studies suggest *CREB1* is associated with mood disorders and lithium response [44, 45].

The complementary nature of *BMP7*—a member of the TGF-beta superfamily—and *CREB1* can possibly be explained by the fact that they are both downstream components of the signaling cascade activated by brain-derived neurotrophic factor (BDNF) (Figures 16 and 17) [46, 47, 48, 49, 50].

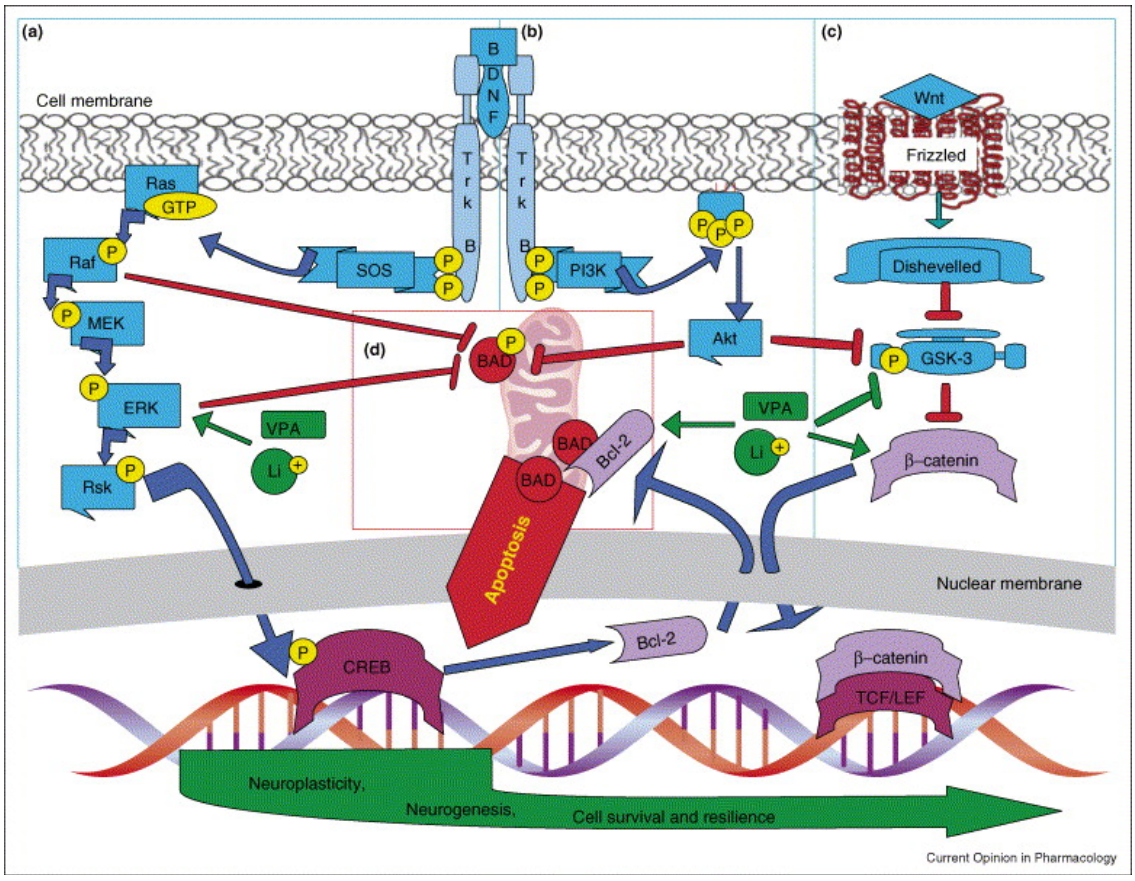

Figure 16. The BDNF signaling cascade. CREB is a downstream target of the signaling pathway, and is believed to activate the expression of genes involved in neurogenesis and cell survival. Figure taken from Shaltiel, et al. (2007) [50].

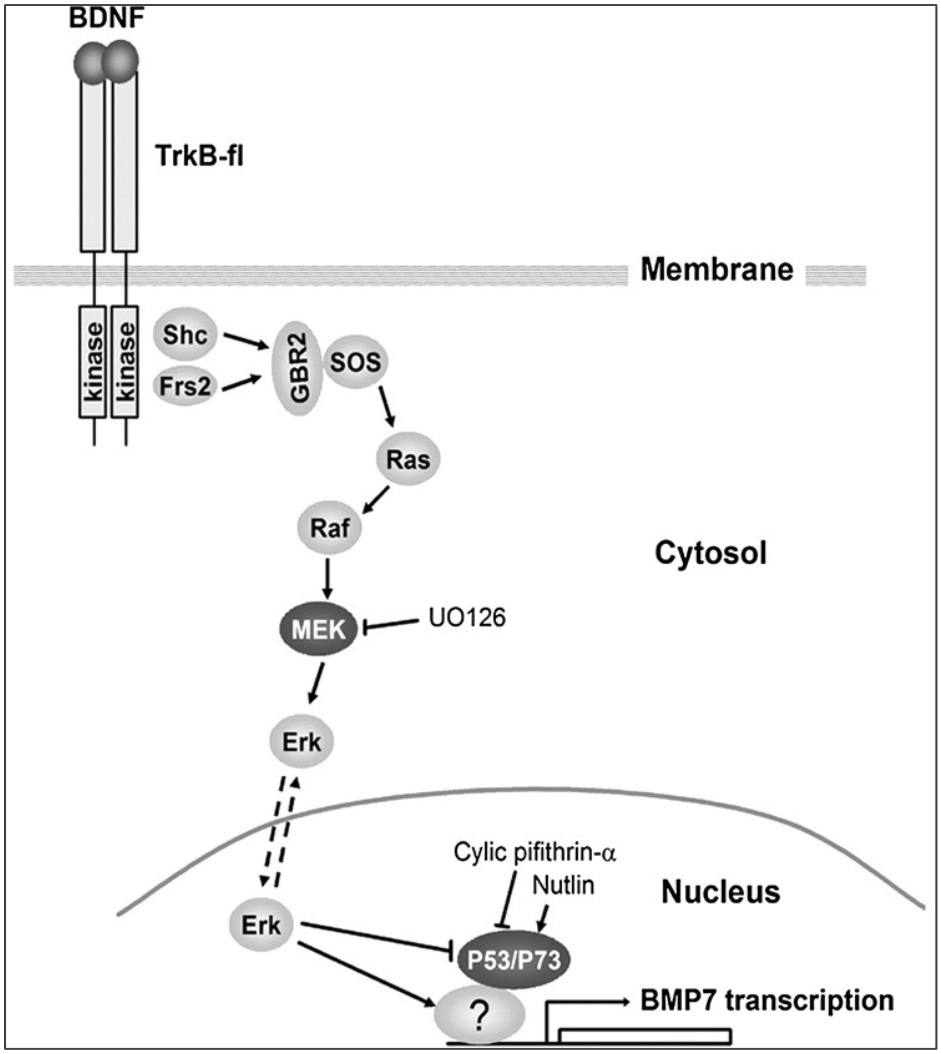

Figure 17. BDNF-induced expression of *BMP7*. Figure taken from Ortega, et al. (2010) [47].

***GFRA1* and *ENPP2*:** While no direct biochemical relationship between these genes has yet been described, both genes do appear to be involved in the same biological processes, namely neuroprotection.

ENPP2 encodes an enzyme responsible for the production of lysophosphatidic acid (LPA), which is involved in a number of processes possibly relevant to bipolar disorder, such as myelination and calcium transport [51, 52]. A number of studies have also shown

LPA to be released after CNS injury, and another has shown LPA to be present in higher concentrations in the CSF of Parkinson's disease patients [53, 54, 55]. Both findings suggest a role in the response to brain injury or nuerodegenerative disease. Additionally, one study has shown *ENPP2* to be differentially expressed in the temporal cortex of major depressive disorder patients [56]; it is also within a linkage region (8q24) that has been repeatedly implicated in bipolar disorder [57, 58, 59, 60, 61].

GFRA1 is a receptor for glial cell line-derived neurotrophic factor (GDNF), which has long been known to be involved in the protection of dopaminergic and motor neurons. A number of studies have shown that GDNF promotes survival and can even restore damaged or dying dopamine neurons [62, 63, 64, 65]. There is some evidence for lower expression and lower serum levels of GDNF in bipolar patients [66, 67].

The interactive effect of *GFRA1* and *ENPP2* could possibly be explained by their involvement in the activation of phosphatidylinositol 3-kinase (PI3K) (Figures 18 and 19) [68, 69, 70].

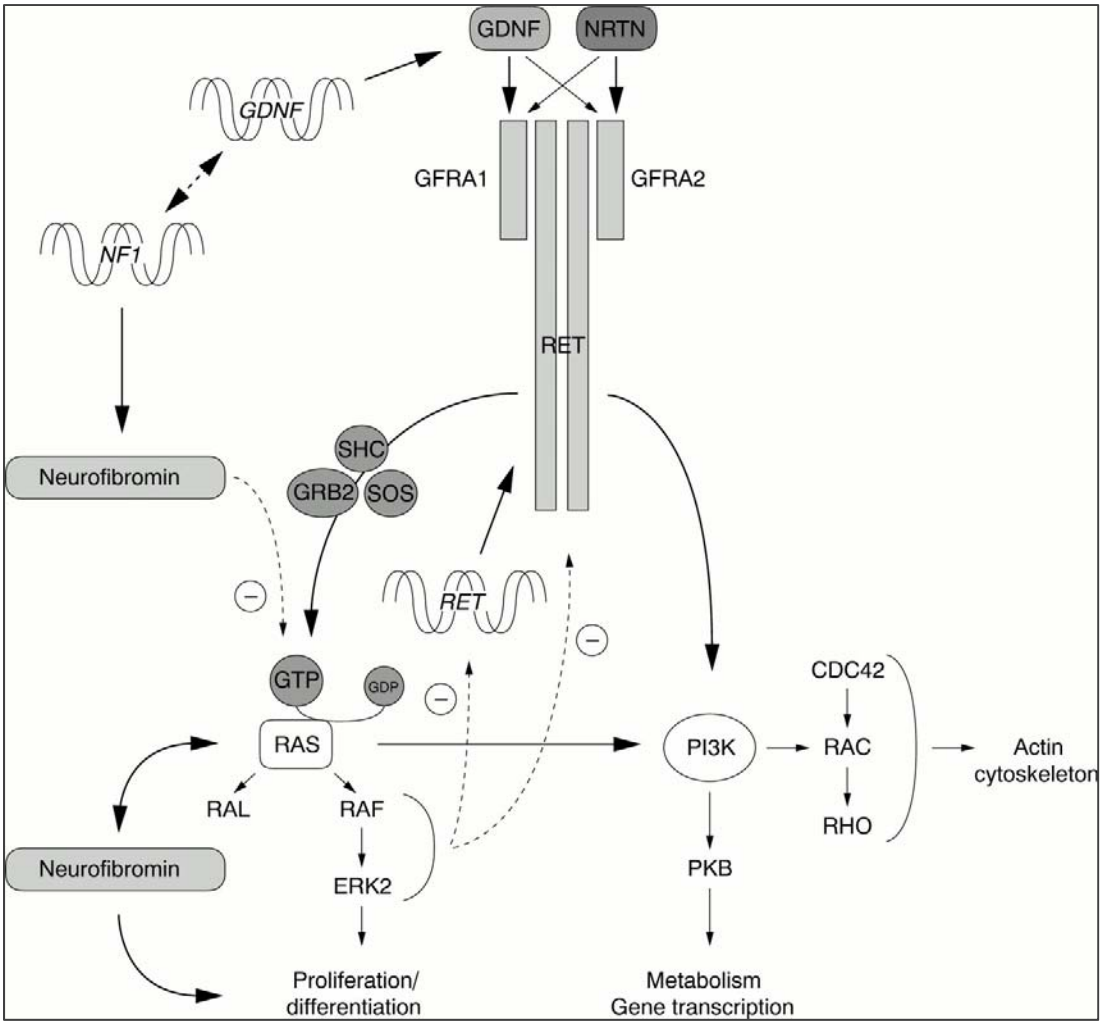

Figure 18. The GDNF signaling cascade. The signaling molecule PI3K is a downstream target of the pathway. Figure taken from Bahuau, et al. (2001) [69].

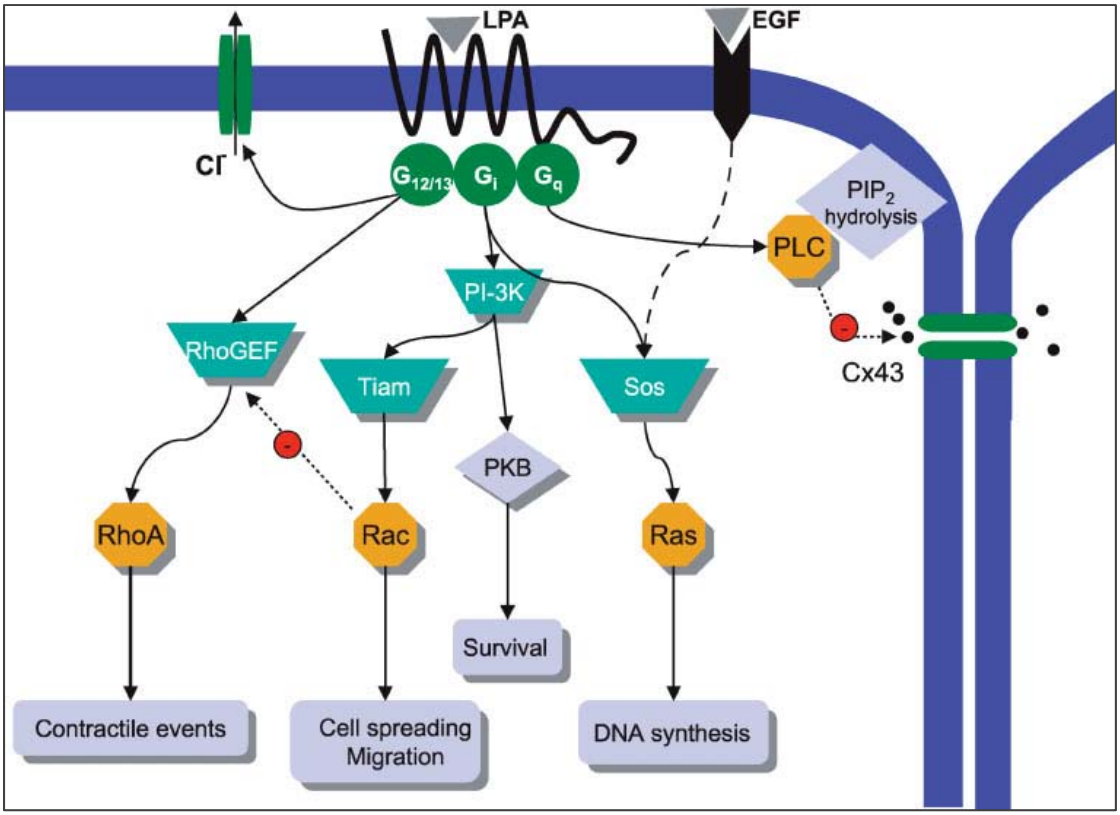

Figure 19. The LPA signaling cascade. The signaling molecule PI3K is shown to play an intermediary role in this pathway also. Figure taken from Moolenaar, et al. (2004) [70].

***SLC35F1* and *NDUFV2*:** Although no specific function has been identified for *SLC35F1*, it belongs to a family of nucleotide-sugar transporters, and has been found to be highly expressed in brain tissue from both adult and fetal samples [71]. *SLC35F1* lies within a linkage region (6q22) previously associated with bipolar disorder [58, 72, 73, 74], and was found to have at least borderline differential expression in the dorsolateral prefrontal cortex of bipolar patients [30].

NDUFV2 encodes a subunit of the mitochondrial complex 1. This gene has been implicated in bipolar disorder previously [75, 76, 77, 78], and resides in a linkage region (18p11.2) associated with bipolar disorder in multiple studies [79, 80, 81, 82]. There is some evidence that mutations in a number of the genes encoding subunits of complex 1

are associated with neurological disorders, including Parkinson's disease. And recently, Andreazza and colleagues found a significant decrease in complex 1 activity in the prefrontal cortex of bipolar patients [83].

5 Discussion & Conclusions

Detection of multi-locus associations is becoming an important way to extend the amount of information gained from a GWAS. While the method I have presented shows promise, this is an area of research that remains relatively unexplored, and there are a number of challenges that still need to be addressed.

Determining the threshold for statistical significance has always been a major issue in genomic studies. In 1995, Lander and Kruglyak provided an important summary of the statistical issues relevant to genetic studies of complex traits. And while their discussion was aimed specifically at genome-wide linkage analyses, their arguments are directly applicable to GWAS. Following Lander and Kruglyak's description of a linkage study, the process of a genome-wide association study can be demonstrated by three fairly simple steps: 1) perform a scan of a large number of SNPs across the entire genome, 2) calculate an appropriate measure of association at each SNP position, and 3) identify SNPs with association measures that deviate substantially from the null hypothesis of no association. However simple these three steps appear, the crucial question remains: since the measure of association "fluctuates substantially just by chance across an entire genome scan, what constitutes a 'significant' deviation?" As Lander and Kruglyak explain, it's clear that the true significance (or importance) of any particular measure of association, say a p-value of 5×10^{-7}, depends on how often such a p-value occurs by chance across the entire genome [84].

By convention, a false-positive rate of 5 % is considered acceptable, and is often used in biological and epidemiological research. When multiple statistical tests are performed

on the same data set, the probability of a false-positive test result occurring increases, and the threshold for statistical significance must be adjusted accordingly. For instance, in the case of a GWAS, where 10^6 SNPs are tested for association, a significance level of 5×10^{-8} will provide the same 5 % probability of a false positive occurring in the analysis (i.e., the 5×10^{-8} threshold provides a 95 % probability that no false-positive results will occur in the entire analysis) [85].

Because of the varying number of SNPs analyzed in GWAS, the conventional thresholds used have ranged between 10^{-5} and 10^{-8}, with the more stringent significance levels being more common. As Lander and Kruglyak pointed out in 1995, the selection of a standard threshold for statistical significance would have important consequences for a relatively young area of research. Too lax a threshold would flood the field with false-positive results, and would have a negative impact on the credibility of these types of studies. Too strict a threshold would cause initial studies to fail to produce meaningful results, and might inhibit progress in the field. However, thanks to improvements in technology over the past decade, GWAS are being conducted routinely, and hundreds, if not thousands, of genetic associations have been reported in the literature. We now have the opportunity to examine the overall outcome of the application of these theoretical thresholds for statistical significance. Has the application of the standard threshold produced a high proportion of replicable results, or has it led to the reporting of large numbers of false-positive associations?

The Human Genome Research Institute maintains a database of all published large-scale GWAS, which includes 42 diseases (or traits) that have been examined by five or more studies [2]. In an attempt to estimate the probability of replication for genetic

associations at various levels of statistical significance, 5 of the 42 well studied diseases/traits were chosen randomly (breast cancer, Crohn's disease, Parkinson's disease, LDL cholesterol, and smoking behavior). For each disease, the proportion of gene associations (at various significance levels) that were reported in at least two studies was recorded, and the results were averaged across the five diseases (Figure 20).

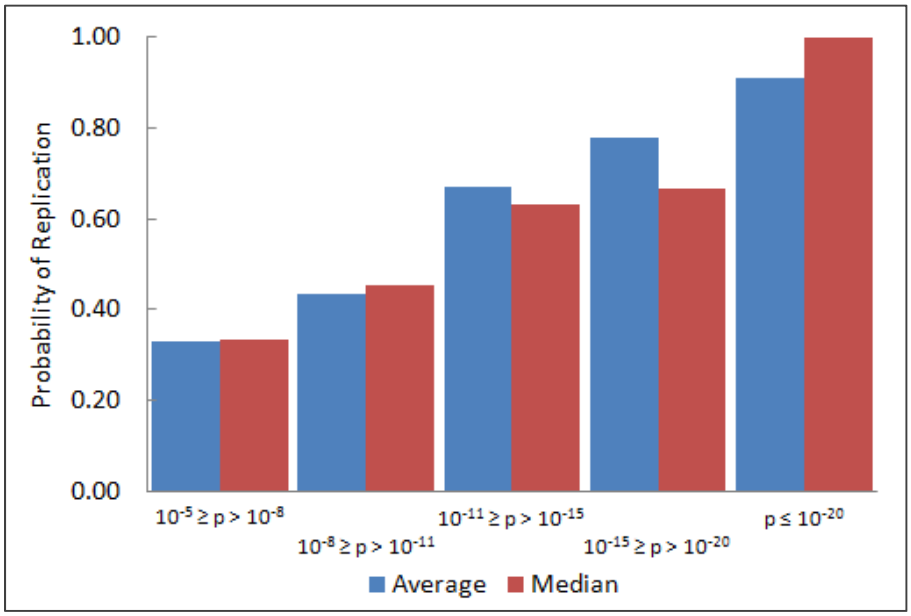

Figure 20. The probability of a genetic association being reported in multiple studies increases with the statistical significance of the association.

As expected, the likelihood of replication increases with smaller and smaller p-values. However, it is interesting to note that for associations with p-values near what is considered the standard threshold for statistical significance (between 10^{-5} and 10^{-11}), the probability of being reported in two or more studies is less than 50 %. In fact, it appears that a p-value smaller than 10^{-15} may be required for there to be a high level of confidence in an association.

The question remains whether or not the significance threshold used for single-SNP associations is appropriate for studies looking for multi-locus associations (SNP interactions). There are two things that make the search for multi-locus associations different from the standard GWAS: 1) the enormous amount of combinatorial complexity, especially when looking for higher-order SNP interactions (i.e., interactions among groups of more than two SNPs), means that the number of possible statistical tests is increased substantially, and 2) the high level of correlation among statistical tests due to that fact that many SNP groups will have individual members in common. The first observation suggests that the threshold for statistical significance may need to be more stringent for multi-locus associations than for single SNP associations, simply due to the increased number of statistical tests. However, because the tests are not independent, it may not be necessary to adjust the significance threshold for all possible SNP combinations. Emily and collegues calculated a "number of effective tests", which takes into account the correlation among SNP groups [22]. Although they tested approximately 3000000 SNP pairs for association in the WTCCC data set, they concluded that the effective number of tests ranged from 506173 to 600010 across the seven diseases in the data set, which correspond to significance thresholds very similar to those used in a standard GWAS (between 8.3×10^{-8} and 9.9×10^{-8}).

It is difficult to estimate the null distribution of p-values—those expected to occur by chance—when searching for multi-locus associations. I have presented one possible method, which involves using simulated genotype data, but there are other options as well. Some have suggested that certain computation-based tests, such as permutation tests, may be appropriate for addressing the huge multiple-testing problem of large-scale

GWAS [86]. The computational burden of permuting large data sets (i.e., randomly assigning case/control labels) and performing multi-locus association analyses for each permutation will be very heavy. However, this type of analysis may have benefits over the use of idealized simulated data. Cross-validation, or sample splitting methods, may also give us some insight into the distribution of p-values expected by chance, and therefore may help determine an appropriate threshold for statistical significance in multi-locus association studies [19].

Although not many studies have done full-scale, genome-wide searches for multi-locus associations, it appears that the significance thresholds used so far for reporting multi-locus associations have not been substantially different than those used for single-SNP associations (refer to examples in section 1.3). But interactions among SNPs are becoming an area of focus for researches exploring the genetic component of complex diseases, and it is only a matter of time before these types of studies become commonplace. The more data sets that are explored in this way, the better prepared we will be to interpret the significance of the results.

Beyond the question of statistical significance, the search for multi-locus associations also presents a considerable computational challenge, particularly when looking for higher-order interactions among SNPs. As seen from the results presented, one of the limitations of using of a genetic algorithm to search for multi-locus associations is the inconsistency of the results. While the algorithm may be able to reliably find good solutions (i.e., SNP groups with small p-values), the solutions may differ from one search to the next. Not being able to definitively say that all significant multi-locus associations

have been found can be somewhat troubling. However, as a hypothesis generating technique the GA is clearly successful.

Moreover, we have suggested a number of ways to increase the consistency and thoroughness of the algorithm. For example, allowing the algorithm to run for an extended period of time will undoubtedly make for a more thorough search. And assigning scores to the genes in the interaction network—highlighting those genes thought to be most promising—is able to focus the algorithm on a smaller search space.

Given the large number of genes and genetic variants that are now commonly examined in a single genetic study, gene prioritization has become an important area of research. Gene prioritization is a significant part of most genomic studies both *after* the primary analysis is done—such as narrowing down a list of dozens, possibly hundreds, of genes located within a linkage region—and as preparation *before* an analysis is done—such as testing for association among a limited set of candidate genes, rather than testing all genes across the genome.

There are two techniques that have been commonly used for gene prioritization, and are particularly relevant for multi-locus analyses. The first is the network-based approach, most commonly utilizing protein-protein interaction data, although other types of relationships between genes or proteins can be used (e.g., functional information). For the network-based approach, the hypothesis is that genes directly connected in the interaction network are more likely to have similar properties (e.g., functions, disease associations, etc.) than genes that are not connected in the network. Therefore, given a set of known genes, it is possible to use the relationships defined by the network to predict an additional set of candidate genes [87, 88].

The second technique is to perform a multi-stage analysis. In the first stage, all genes are analyzed individually, either in the same data set or a separate one. The results from this first stage are then used to produce a smaller set of candidate genes for the analysis in the second stage [14]. Alternatively, rather than simply using the results of the first-stage analysis to exclude genes from further study, the first-stage results could be used to guide a search for multi-locus associations in the second stage [89].

The algorithm presented in this dissertation is able to incorporate both of these gene prioritization techniques. And while it has shown promise as a hypothesis generating method, even without incorporating any disease-specific information, I believe it represents an ideal method for secondary analyses—mining available data sets for multi-locus associations after a group of candidate genes (or pathways, or cellular processes, etc.) have already been identified.

The reduced search space that comes with secondary analyses will address, to some extent, the computational and statistical issues discussed earlier. However, it is important to remember that any candidate genes fed into the algorithm can (and probably should) be chosen with a more permissive threshold than is used conventionally. This will ensure that the search capabilities of the algorithm are fully exploited, and it will also ensure that the possibility of discovering multi-locus associations without main effects is not ignored.

The main advantage of the method presented here is its flexibility. The algorithm is able to search for multi-locus associations within a full-scale GWAS dataset, and can search for interactions among any number of SNPs (of course the size of the SNP group is limited by the study sample size, since large SNP groups will result in very sparse contingency tables). And testing different genetic models or hypotheses is merely a

matter of changing the fitness function within the GA. For instance, rather than looking for interactions between SNPs, one could easily look at the additive risk of multiple alleles, similar to the analysis done by Baum and colleagues [14].

We have shown that our algorithm is able to discover significant multi-locus associations, even among SNPs that are not strongly associated with the disease. There has been some controversy about the idea of gene interactions without main effects, but our results and others [90] suggest the possibility of such interactions should not be ignored. This is true for the simple fact that any threshold used to define a significant main effect is arbitrary, and using such a threshold necessarily excludes any interactions that may be found among SNPs with a weak but real association with the disease.

There are numerous ways that the genetic algorithm described here can be adapted to fit the needs of researchers interested in similar problems, and I believe it has the potential to extract a great deal of information from the wealth of large-scale GWAS data already available.

Apart from evaluating the algorithm and showing the feasibility of performing a large-scale search for multi-locus associations, the second goal of this dissertation was to apply the method to a real-world dataset in the hope of finding biologically meaningful results that may provide insight into the processes involved in bipolar disorder.

From this real-world analysis a number of lessons were learned. The ability to discover significant multi-SNP associations, even in the absence of main effects, was confirmed. However, as the top hits were examined more closely, it became clear that interpreting how the effects of each individual SNP were combining to produce a joint

effect was far from straightforward in most cases. Many of the multi-SNP associations that achieved statistical significance appeared to be the result of small numbers in the contingency table cells. And very few of the associations demonstrated an obvious additive or conditional effect.

By incorporating more specific genetic models directly into the algorithm, it may be possible to remove much of the noise (i.e., false positives, or low confidence associations) from the results. This can be done fairly easily by improving the GA's fitness function. Rather than using a contingency table with all possible genotype combinations, the fitness function could include a test of trend (for an additive model), or specific conditional models (i.e., data stratification).

Finally, it was clear that sample size was a limitation when looking for associations with more than two SNPs. Larger sample sizes, or different analysis techniques, will be needed to avoid false positives due to small numbers in the contingency table cells (i.e., very rare genotype combinations). Although, it may be possible to alleviate this problem somewhat by restricting the analysis to only very common SNPs, such as those with a minor allele frequency greater than 10 %.

In terms of the biological significance of the multi-SNP associations discovered, it is interesting that all three of the SNP interactions discussed above have some relation to Parkinson's disease and, in turn, neuroprotection. This suggests there may be disease processes in common between bipolar disorder and Parkinson's disease. Indeed, at least two studies have shown an increase risk for developing Parkinson's disease among patients with affective disorders [91, 92]. It is possible that defects in neuroprotective processes, and an inability to correct the damage or perturbation caused by certain

exposures (substance abuse, psychological trauma, etc.), may contribute to a susceptibility to bipolar disorder.

6 Future Work

Developing methods for revealing the genetic component of complex diseases is an exciting area of research. And the polygenic view has the potential to produce significant advances in our understanding. There are a number of areas that hold particular promise and may yield important results in the near future.

One of the most serious issues facing GWAS is the lack of reproducibility, which is thought to be due, in part, to heterogeneity—in terms of both genetic risk factors and phenotypes—among patient populations. Addressing this problem of heterogeneity will be important for understanding the genetics of all common complex diseases. This has been particularly apparent from the studies of bipolar disorder.

Making an effort to carefully and precisely define the phenotypic characteristics of cases in genetic studies will be a necessary step for improving agreement across studies. And collaboration across disciplines will be needed for understanding what phenotypic characteristics will be most important.

Exploring the question of genetic heterogeneity will require creative new analysis techniques. It will be crucial to move beyond the search for associations with single DNA variants, and towards associations with entire biochemical pathways or modules of genes involved in a common cellular process. This type of analysis has already begun to be developed and has been referred to as "network-based association studies" [93, 94]. The goal of these studies is to search a large gene interaction network to find sub-networks associated with the trait of interest. For example, to find sub-networks containing a

higher proportion of genes individually associatiated with the trait than would be expected by chance (i.e., clusters of associated genes).

However, as association studies progress from single SNPs to entire pathways and sub-networks, it will be important to make a special effort to focus on the interpretability of results. Making sure studies incorporate specific models of how the effects of multiple genes can combine to produce a disease state is a good way to increase the potential that result can be translated into information that is clinically relevant.

Finally, extending association studies even further will be possible by including different types of genetic variation such as CNVs, rare variants, insertions and deletions, etc. New sequencing technologies will undoubtedly allow for these types of advances to be made.

Many of these techniques are natural extensions for the work presented in this dissertation. And given the flexibility of the network-guided genetic algorithm, I hope to be able to incorporate many of these techniques—from performing analyses with alternative phenotype definitions, to utilizing new statistical methods adapted for network-based association studies—into my future research endeavors.

References

1. Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB. Rare variants create synthetic genome-wide associations. *PLoS Biol*. 2010 Jan 26, 8(1):e1000294. PMID: 20126254

2. Hindorff LA, Junkins HA, Mehta JP, and Manolio TA. A Catalog of Published Genome-Wide Association Studies. Available at: www.genome.gov/gwastudies. Accessed April, 2011.

3. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J. Complement factor H polymorphism in age-related macular degeneration. *Science*. 2005 Apr 15, 308(5720):385-9. PMID: 15761122

4. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007 Jun 7, 447(7145):661-78. PMID: 17554300

5. Craddock N, Khodel V, Van Eerdewegh P, Reich T. Mathematical limits of multilocus models: the genetic transmission of bipolar disorder. *Am J Hum Genet*. 1995 Sep, 57(3):690-702. PMID: 7668299

6. Kohn Y. Focus on rare genetic variants in bipolar disorder: how outliers help understand complex disorders. *International Journal of Neuropsychopharmacology*. 2005 Dec, 8(4):491-493. PMID: 16202180

7. Meyer J, Johannssen K, Freitag CM, Schraut K, Teuber I, Hahner A, Mainhardt C, Mössner R, Volz HP, Wienker TF, McKeane D, Stephan DA, Rouleau G, Reif A,

Lesch KP. Rare variants of the gene encoding the potassium chloride co-transporter 3 are associated with bipolar disorder. *International Journal of Neuropsychopharmacology*. 2005 Dec, 8(4):495-504. PMID: 16098236

8. Alaerts M, Del-Favero J. Searching genetic risk factors for schizophrenia and bipolar disorder: learn from the past and back to the future. *Hum Mutat*. 2009 Aug, 30(8):1139-52. PMID: 19626716
9. Tam GW, Redon R, Carter NP, Grant SG. The role of DNA copy number variation in schizophrenia. Biol Psychiatry. 2009 Dec 1, 66(11):1005-12. PMID: 19748074
10. Goodwin FK and Jamison KR. *Manic-Depressive Illness: Bipolar Disorders and Recurrent Depression*, 2nd edition. New York: Oxford University Press, 2007.
11. Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmada MM, Bitton A, Dassopoulos T, Datta LW, Green T, Griffiths AM, Kistner EO, Murtha MT, Regueiro MD, Rotter JI, Schumm LP, Steinhart AH, Targan SR, Xavier RJ; NIDDK IBD Genetics Consortium, Libioulle C, Sandor C, Lathrop M, Belaiche J, Dewit O, Gut I, Heath S, Laukens D, Mni M, Rutgeerts P, Van Gossum A, Zelenika D, Franchimont D, Hugot JP, de Vos M, Vermeire S, Louis E; Belgian-French IBD Consortium; Wellcome Trust Case Control Consortium, Cardon LR, Anderson CA, Drummond H, Nimmo E, Ahmad T, Prescott NJ, Onnie CM, Fisher SA, Marchini J, Ghori J, Bumpstead S, Gwilliam R, Tremelling M, Deloukas P, Mansfield J, Jewell D, Satsangi J, Mathew CG, Parkes M, Georges M, Daly MJ. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet*. 2008 Aug, 40(8):955-62. PMID: 18587394

12. Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, Fan J, Kirov G, Perlis RH, Green EK, Smoller JW, Grozeva D, Stone J, Nikolov I, Chambert K, Hamshere ML, Nimgaonkar VL, Moskvina V, Thase ME, Caesar S, Sachs GS, Franklin J, Gordon-Smith K, Ardlie KG, Gabriel SB, Fraser C, Blumenstiel B, Defelice M, Breen G, Gill M, Morris DW, Elkin A, Muir WJ, McGhee KA, Williamson R, MacIntyre DJ, MacLean AW, St CD, Robinson M, Van Beck M, Pereira AC, Kandaswamy R, McQuillin A, Collier DA, Bass NJ, Young AH, Lawrence J, Ferrier IN, Anjorin A, Farmer A, Curtis D, Scolnick EM, McGuffin P, Daly MJ, Corvin AP, Holmans PA, Blackwood DH, Gurling HM, Owen MJ, Purcell SM, Sklar P, Craddock N; Wellcome Trust Case Control Consortium. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet*. 2008 Sep, 40(9):1056-8. PMID: 18711365

13. Baum AE, Hamshere M, Green E, Cichon S, Rietschel M, Noethen MM, Craddock N, McMahon FJ. Meta-analysis of two genome-wide association studies of bipolar disorder reveals important points of agreement. *Mol Psychiatry*. 2008 May, 13(5):466-7. PMID: 18421293

14. Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B, Schulze TG, Cichon S, Rietschel M, Nöthen MM, Georgi A, Schumacher J, Schwarz M, Abou Jamra R, Höfels S, Propping P, Satagopan J, Detera-Wadleigh SD, Hardy J, McMahon FJ. A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Mol Psychiatry*. 2008 Feb, 13(2):197-207. PMID: 17486107

15. Manji HK, Lenox RH. Signaling: cellular insights into the pathophysiology of bipolar disorder. *Biol Psychiatry*. 2000 Sep 15, 48(6):518-30. PMID: 11018224

16. International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, Sklar P. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009 Aug 6, 460(7256):748-52. PMID: 19571811

17. Sklar P, Smoller JW, Fan J, Ferreira MA, Perlis RH, Chambert K, Nimgaonkar VL, McQueen MB, Faraone SV, Kirby A, de Bakker PI, Ogdie MN, Thase ME, Sachs GS, Todd-Brown K, Gabriel SB, Sougnez C, Gates C, Blumenstiel B, Defelice M, Ardlie KG, Franklin J, Muir WJ, McGhee KA, MacIntyre DJ, McLean A, VanBeck M, McQuillin A, Bass NJ, Robinson M, Lawrence J, Anjorin A, Curtis D, Scolnick EM, Daly MJ, Blackwood DH, Gurling HM, Purcell SM. Whole-genome association study of bipolar disorder. *Mol Psychiatry*. 2008 Jun, 13(6):558-69. PMID: 18317468

18. Nunkesser R, Bernholt T, Schwender H, Ickstadt K, Wegener I. Detecting high-order interactions of single nucleotide polymorphisms using genetic programming. *Bioinformatics*. 2007 Dec 15, 23(24):3280-8. PMID: 18006552.

19. Gayán J, González-Pérez A, Bermudo F, Sáez ME, Royo JL, Quintas A, Galan JJ, Morón FJ, Ramirez-Lorca R, Real LM, Ruiz A. A method for detecting epistasis in genome-wide studies using case-control multi-locus association analysis. *BMC Genomics*. 2008 Jul 31, 9:360. PMID: 18667089.

20. Zhang X, Huang S, Zou F, Wang W. TEAM: efficient two-locus epistasis tests in human genome-wide association study. *Bioinformatics*. 2010 Jun 15,26(12):i217-27. PMID: 20529910.

21. Moore JH, White BC. Genome-wide Genetic Analysis Using Genetic Programming: The Critical Need for Expert Knowledge. *Genetic Programming Theory and Practice IV*, R. Riolo, T. Soule, B. Worzel, eds., Springer, New York, pp. 11-28, 2007.

22. 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 Oct, 17(10):1231-40. PMID: 19277065

23. Smith EN, Bloss CS, Badner JA, Barrett T, Belmonte PL, Berrettini W, Byerley W, Coryell W, Craig D, Edenberg HJ, Eskin E, Foroud T, Gershon E, Greenwood TA, Hipolito M, Koller DL, Lawson WB, Liu C, Lohoff F, McInnis MG, McMahon FJ, Mirel DB, Murray SS, Nievergelt C, Nurnberger J, Nwulia EA, Paschall J, Potash JB, Rice J, Schulze TG, Scheftner W, Panganiban C, Zaitlen N, Zandi PP, Zöllner S, Schork NJ, Kelsoe JR. Genome-wide association study of bipolar disorder in European American and African American individuals. *Mol Psychiatry*. 2009 Aug, 14(8):755-763. PMID: 19488044

24. Smith EN, Craig DW, Schork NJ, Kelsoe JR, The Bipolar Genome Study. Genome-wide Association of Bipolar Disorder Suggests an Enrichment of True Associations in Regions Near Exons. (73), Presented at the 60th Annual Meeting of the American Society of Human Genetics, Nov. 2, 2010, Washington, D.C.

25. Carvalho B, Bengtsson H, Speed TP, Irizarry RA. Exploration, normalization, and genotype calls of high-density oligonucleotide SNP array data. *Biostatistics*. 2007 Apr, 8(2):485-99. PMID: 17189563.

26. Nurnberger JI Jr, Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J, Severe JB, Malaspina D, Reich T. Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. *Arch Gen Psychiatry*. 1994 Nov, 51(11):849-59. PMID: 7944874

27. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C. STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res*. 2009 Jan, 37(Database issue):D412-6. PMID: 18940858.

28. Hubbard TJ, Aken BL, Ayling S, Ballester B, Beal K, Bragin E, Brent S, Chen Y, Clapham P, Clarke L, Coates G, Fairley S, Fitzgerald S, Fernandez-Banet J, Gordon L, Graf S, Haider S, Hammond M, Holland R, Howe K, Jenkinson A, Johnson N, Kahari A, Keefe D, Keenan S, Kinsella R, Kokocinski F, Kulesha E, Lawson D, Longden I, Megy K, Meidl P, Overduin B, Parker A, Pritchard B, Rios D, Schuster M, Slater G, Smedley D, Spooner W, Spudich G, Trevanion S, Vilella A, Vogel J, White S, Wilder S, Zadissa A, Birney E, Cunningham F, Curwen V, Durbin R, Fernandez-Suarez XM, Herrero J, Kasprzyk A, Proctor G, Smith J, Searle S, Flicek P. Ensembl 2009. *Nucleic Acids Res*. 2009 Jan, 37(Database issue):D690-7. PMID: 19033362. Available at: www.ensembl.org. Accessed May, 2009.

29. The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nature Genet*. 2000 May, 25(1):25-29. Available at: www.geneontology.org. Accessed May, 2009.

30. Ryan MM, Lockstone HE, Huffaker SJ, Wayland MT, Webster MJ, Bahn S. Gene expression analysis of bipolar disorder reveals downregulation of the ubiquitin cycle and alterations in synaptic genes. *Mol Psychiatry*. 2006 Oct, 11(10):965-78.

31. Matigian N, Windus L, Smith H, Filippich C, Pantelis C, McGrath J, Mowry B, Hayward N. Expression profiling in monozygotic twins discordant for bipolar disorder reveals dysregulation of the WNT signalling pathway. *Mol Psychiatry*. 2007 Sep, 12(9):815-25. PMID: 17440432

32. Jeong H, Mason SP, Barabási AL, Oltvai ZN. Lethality and centrality in protein networks. *Nature*. 2001 May 3, 411(6833):41-2. PMID: 11333967

33. Horvath S, Zhang B, Carlson M, Lu KV, Zhu S, Felciano RM, Laurance MF, Zhao W, Qi S, Chen Z, Lee Y, Scheck AC, Liau LM, Wu H, Geschwind DH, Febbo PG, Kornblum HI, Cloughesy TF, Nelson SF, Mischel PS. Analysis of oncogenic signaling networks in glioblastoma identifies ASPM as a molecular target. *Proc Natl Acad Sci U S A*. 2006 Nov 14, 103(46):17402-7. PMID: 17090670

34. Holland JH, *Adaptation in Natural and Artificial Systems*, 2nd Edition. MIT Press, 1992.

35. Forrest S. Genetic algorithms: principles of natural selection applied to computation. *Science*. 1993 Aug 13, 261(5123):872-8. PMID: 8346439

36. Man KF, Tang KS, Kwong S. Genetic algorithms: concepts and applications [in engineering design]. *IEEE Transactions on Industrial Electronics*. 1996 Oct, 43(5):519-534.

37. van Batenburg FH, Gultyaev AP, Pleij CW. An APL-programmed genetic algorithm for the prediction of RNA secondary structure. *J Theor Biol*. 1995 Jun 7, 174(3):269-80. PMID: 7545258

38. Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet*. 2002 Jan, 30(1):97-101. PMID: 11731797

39. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007 Sep, 81(3):559-75. PMID: 17701901.

40. Cox LJ, Hengst U, Gurskaya NG, Lukyanov KA, Jaffrey SR. Intra-axonal translation and retrograde trafficking of CREB promotes neuronal survival. *Nat Cell Biol*. 2008 Feb;10(2):149-59. Erratum in: *Nat Cell Biol*. 2008 Mar, 10(3):370. PMID: 18193038

41. Wen Z, Han L, Bamburg JR, Shim S, Ming GL, Zheng JQ. BMP gradients steer nerve growth cones by a balancing act of LIM kinase and Slingshot phosphatase on ADF/cofilin. *J Cell Biol*. 2007 Jul 2, 178(1):107-19. PMID: 17606869

42. Yuan P, Zhou R, Wang Y, Li X, Li J, Chen G, Guitart X, Manji HK. Altered levels of extracellular signal-regulated kinase signaling proteins in postmortem frontal

cortex of individuals with mood disorders and schizophrenia. *J Affect Disord*. 2010 Jul, 124(1-2):164-9. PMID: 19913919

43. Garriock HA, Kraft JB, Shyn SI, Peters EJ, Yokoyama JS, Jenkins GD, Reinalda MS, Slager SL, McGrath PJ, Hamilton SP. A genomewide association study of citalopram response in major depressive disorder. *Biol Psychiatry*. 2010 Jan 15, 67(2):133-8. PMID: 19846067
44. Mamdani F, Alda M, Grof P, Young LT, Rouleau G, Turecki G. Lithium response and genetic variation in the CREB family of genes. *Am J Med Genet B Neuropsychiatr Genet*. 2008 Jun 5, 147B(4):500-4. PMID: 18189280
45. Dong C, Wong ML, Licinio J. Sequence variations of ABCB1, SLC6A2, SLC6A3, SLC6A4, CREB1, CRHR1 and NTRK2: association with major depression and antidepressant response in Mexican-Americans. *Mol Psychiatry*. 2009 Dec, 14(12):1105-18. PMID: 19844206
46. Juhasz G, Dunham JS, McKie S, Thomas E, Downey D, Chase D, Lloyd-Williams K, Toth ZG, Platt H, Mekli K, Payton A, Elliott R, Williams SR, Anderson IM, Deakin JF. The CREB1-BDNF-NTRK2 Pathway in Depression: Multiple Gene-Cognition-Environment Interactions. *Biol Psychiatry*. 2011 Jan 5. PMID: 21215389
47. Ortega JA, Alcántara S. BDNF/MAPK/ERK-induced BMP7 expression in the developing cerebral cortex induces premature radial glia differentiation and impairs neuronal migration. *Cereb Cortex*. 2010 Sep, 20(9):2132-44. PMID: 2003854
48. Cox S, Harvey BK, Sanchez JF, Wang JY, Wang Y. Mediation of BMP7 neuroprotection by MAPK and PKC IN rat primary cortical cultures. *Brain Res*. 2004 Jun 4, 1010(1-2):55-61. PMID: 15126117

49. Zhu Y, Yang GY, Ahlemeyer B, Pang L, Che XM, Culmsee C, Klumpp S, Krieglstein J. Transforming growth factor-beta 1 increases bad phosphorylation and protects neurons against damage. *J Neurosci*. 2002 May 15, 22(10):3898-909. PMID: 12019309

50. Shaltiel G, Chen G, Manji HK. Neurotrophic signaling cascades in the pathophysiology and treatment of bipolar disorder. *Curr Opin Pharmacol*. 2007 Feb, 7(1):22-6. PMID: 1705533

51. Fukushima N. LPA in neural cell development. *J Cell Biochem*. 2004 Aug 1, 92(5):993-1003. PMID: 15258920

52. Uemura T, Green M, Corson TW, Perova T, Li PP, Warsh JJ. Bcl-2 SNP rs956572 associates with disrupted intracellular calcium homeostasis in bipolar I disorder. *Bipolar Disord*. 2011 Feb, 13(1):41-51. PMID: 21320251

53. Goldshmit Y, Munro K, Leong SY, Pébay A, Turnley AM. LPA receptor expression in the central nervous system in health and following injury. *Cell Tissue Res*. 2010 Jul, 341(1):23-32. PMID: 20495828

54. Okudaira S, Yukiura H, Aoki J. Biological roles of lysophosphatidic acid signaling through its production by autotaxin. *Biochimie*. 2010 Jun, 92(6):698-706. PMID: 20417246

55. Guo J, Sun Z, Xiao S, Liu D, Jin G, Wang E, Zhou J, Zhou J. Proteomic analysis of the cerebrospinal fluid of Parkinson's disease patients. *Cell Res*. 2009 Dec, 19(12):1401-3. PMID: 19949427

56. Aston C, Jiang L, Sokolov BP. Transcriptional profiling reveals evidence for signaling and oligodendroglial abnormalities in the temporal cortex from patients

with major depressive disorder. *Mol Psychiatry*. 2005 Mar, 10(3):309-22. PMID: 15303102

57. McInnis MG, Lan TH, Willour VL, McMahon FJ, Simpson SG, Addington AM, MacKinnon DF, Potash JB, Mahoney AT, Chellis J, Huo Y, Swift-Scanlan T, Chen H, Koskela R, Stine OC, Jamison KR, Holmans P, Folstein SE, Ranade K, Friddle C, Botstein D, Marr T, Beaty TH, Zandi P, DePaulo JR. Genome-wide scan of bipolar disorder in 65 pedigrees: supportive evidence for linkage at 8q24, 18q22, 4q32, 2p12, and 13q12. *Mol Psychiatry*. 2003 Mar, 8(3):288-98. PMID: 12660801

58. McQueen MB, Devlin B, Faraone SV, Nimgaonkar VL, Sklar P, Smoller JW, Abou Jamra R, Albus M, Bacanu SA, Baron M, Barrett TB, Berrettini W, Blacker D, Byerley W, Cichon S, Coryell W, Craddock N, Daly MJ, Depaulo JR, Edenberg HJ, Foroud T, Gill M, Gilliam TC, Hamshere M, Jones I, Jones L, Juo SH, Kelsoe JR, Lambert D, Lange C, Lerer B, Liu J, Maier W, Mackinnon JD, McInnis MG, McMahon FJ, Murphy DL, Nothen MM, Nurnberger JI, Pato CN, Pato MT, Potash JB, Propping P, Pulver AE, Rice JP, Rietschel M, Scheftner W, Schumacher J, Segurado R, Van Steen K, Xie W, Zandi PP, Laird NM. 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 Oct, 77(4):582-95. PMID: 16175504

59. Cichon S, Schumacher J, Müller DJ, Hürter M, Windemuth C, Strauch K, Hemmer S, Schulze TG, Schmidt-Wolf G, Albus M, Borrmann-Hassenbach M, Franzek E, Lanczik M, Fritze J, Kreiner R, Reuner U, Weigelt B, Minges J, Lichtermann D, Lerer B, Kanyas K, Baur MP, Wienker TF, Maier W, Rietschel M, Propping P,

Nöthen MM. A genome screen for genes predisposing to bipolar affective disorder detects a new susceptibility locus on 8q. *Hum Mol Genet*. 2001 Dec 1, 10(25):2933-44. Erratum in: *Hum Mol Genet*. 2002 Jul 1, 11(14):1685. PMID: 11741836

60. Avramopoulos D, Willour VL, Zandi PP, Huo Y, MacKinnon DF, Potash JB, DePaulo JR Jr, McInnis MG. Linkage of bipolar affective disorder on chromosome 8q24: follow-up and parametric analysis. *Mol Psychiatry*. 2004 Feb, 9(2):191-6. PMID: 14966477

61. McInnis MG, Dick DM, Willour VL, Avramopoulos D, MacKinnon DF, Simpson SG, Potash JB, Edenberg HJ, Bowman ES, McMahon FJ, Smiley C, Chellis JL, Huo Y, Diggs T, Meyer ET, Miller M, Matteini AT, Rau NL, DePaulo JR, Gershon ES, Badner JA, Rice JP, Goate AM, Detera-Wadleigh SD, Nurnberger JI, Reich T, Zandi PP, Foroud TM. Genome-wide scan and conditional analysis in bipolar disorder: evidence for genomic interaction in the National Institute of Mental Health genetics initiative bipolar pedigrees. *Biol Psychiatry*. 2003 Dec 1, 54(11):1265-73. PMID: 14643094

62. Lin L-F, Doherty DH, Lile JD, Bektesh S, Collins F. GDNF: A glial cell line-derived neurotrophic factor for midbrain dopaminergic neurons. *Science*. 1993, 260:1130–1132. PMID: 8493557

63. Gill SS, Patel NK, Hotton GR, O'Sullivan K, McCarter R, Bunnage M, Brooks DJ, Svendsen CN, Heywood P. Direct brain infusion of glial cell line-derived neurotrophic factor in Parkinson disease. *Nature Medicine*. 2003, 9:589–595. PMID: 12669033

64. Slevin JT, Gerhardt GA, Smith CD, Gash DM, Kryscio R, Young B. Improvement of bilateral motor functions in patients with Parkinson disease through the unilateral intraputaminal infusion of glial cell line-derived neurotrophic factor. *J Neurosurg*. 2005, 102:216–222. PMID: 15739547

65. Lang AE, Gill S, Patel NK, Lozano A, Nutt JG, Penn R, Brooks DJ, Hotton G, Moro E, Heywood P, Brodsky MA, Burchiel K, Kelly P, Dalvi A, Scott B, Stacy M, Turner D, Wooten VG, Elias WJ, Laws ER, Dhawan V, Stoessl AJ, Matcham J, Coffey RJ, Traub M. Randomized controlled trial of intraputamenal glial cell line-derived neurotrophic factor infusion in Parkinson disease. *Ann Neurol*. 2006, 59:459–466. PMID: 16429411

66. Otsuki K, Uchida S, Watanuki T, Wakabayashi Y, Fujimoto M, Matsubara T, Funato H, Watanabe Y. Altered expression of neurotrophic factors in patients with major depression. J Psychiatr Res. 2008 Oct, 42(14):1145-53. PMID: 18313696

67. Zhang X, Zhang Z, Sha W, Xie C, Xi G, Zhou H, Zhang Y. Effect of treatment on serum glial cell line-derived neurotrophic factor in bipolar patients. *J Affect Disord*. 2010 Oct, 126(1-2):326-9. PMID: 20350767

68. Xing B, Xin T, Zhao L, Hunter RL, Chen Y, Bing G. Glial cell line-derived neurotrophic factor protects midbrain dopaminergic neurons against lipopolysaccharide neurotoxicity. *J Neuroimmunol*. 2010 Aug 25, 225(1-2):43-51. PMID: 20471698

69. Bahuau M, Pelet A, Vidaud D, Lamireau T, LeBail B, Munnich A, Vidaud M, Lyonnet S, Lacombe D. GDNF as a candidate modifier in a type 1

neurofibromatosis (NF1) enteric phenotype. *J Med Genet*. 2001 Sep, 38(9):638-43. PMID: 11565554

70. Moolenaar WH, van Meeteren LA, Giepmans BN. The ins and outs of lysophosphatidic acid signaling. *Bioessays*. 2004 Aug;26(8):870-81. PMID: 15273989

71. Nishimura M, Suzuki S, Satoh T, Naito S. Tissue-specific mRNA expression profiles of human solute carrier 35 transporters. *Drug Metab Pharmacokinet*. 2009, 24(1):91-9. PMID: 19252338

72. Dick DM, Foroud T, Flury L, Bowman ES, Miller MJ, Rau NL, Moe PR, Samavedy N, El-Mallakh R, Manji H, Glitz DA, Meyer ET, Smiley C, Hahn R, Widmark C, McKinney R, Sutton L, Ballas C, Grice D, Berrettini W, Byerley W, Coryell W, DePaulo R, MacKinnon DF, Gershon ES, Kelsoe JR, McMahon FJ, McInnis M, Murphy DL, Reich T, Scheftner W, Nurnberger JI Jr. Genomewide linkage analyses of bipolar disorder: a new sample of 250 pedigrees from the National Institute of Mental Health Genetics Initiative. *Am J Hum Genet*. 2003 Jul, 73(1):107-14. Erratum in: *Am J Hum Genet*. 2003 Oct, 73(4):979. PMID: 12772088

73. Schulze TG, Buervenich S, Badner JA, Steele CJ, Detera-Wadleigh SD, Dick D, Foroud T, Cox NJ, MacKinnon DF, Potash JB, Berrettini WH, Byerley W, Coryell W, DePaulo JR Jr, Gershon ES, Kelsoe JR, McInnis MG, Murphy DL, Reich T, Scheftner W, Nurnberger JI Jr, McMahon FJ. Loci on chromosomes 6q and 6p interact to increase susceptibility to bipolar affective disorder in the national institute of mental health genetics initiative pedigrees. *Biol Psychiatry*. 2004 Jul 1, 56(1):18-23. PMID: 15219468

74. Middleton FA, Pato MT, Gentile KL, Morley CP, Zhao X, Eisener AF, Brown A, Petryshen TL, Kirby AN, Medeiros H, Carvalho C, Macedo A, Dourado A, Coelho I, Valente J, Soares MJ, Ferreira CP, Lei M, Azevedo MH, Kennedy JL, Daly MJ, Sklar P, Pato CN. Genomewide linkage analysis of bipolar disorder by use of a high-density single-nucleotide-polymorphism (SNP) genotyping assay: a comparison with microsatellite marker assays and finding of significant linkage to chromosome 6q22. *Am J Hum Genet*. 2004 May, 74(5):886-97. PMID: 15060841

75. Nakatani N, Hattori E, Ohnishi T, Dean B, Iwayama Y, Matsumoto I, Kato T, Osumi N, Higuchi T, Niwa S, Yoshikawa T. Genome-wide expression analysis detects eight genes with robust alterations specific to bipolar I disorder: relevance to neuronal network perturbation. *Hum Mol Genet*. 2006 Jun 15, 15(12):1949-62. PMID: 16687443

76. Zhang J, Li X, Wang Y, Ji J, Yang F, Feng G, Wan P, Lindpaintner K, He L, He G. Association study on the mitochondrial gene NDUFV2 and bipolar disorder in the Chinese Han population. *J Neural Transm*. 2009 Mar, 116(3):357-61. PMID: 19194776

77. Xu C, Li PP, Kennedy JL, Green M, Hughes B, Cooke RG, Parikh SV, Warsh JJ. Further support for association of the mitochondrial complex I subunit gene NDUFV2 with bipolar disorder. *Bipolar Disord*. 2008 Feb, 10(1):105-10. PMID: 18199248

78. Washizuka S, Iwamoto K, Kazuno AA, Kakiuchi C, Mori K, Kametani M, Yamada K, Kunugi H, Tajima O, Akiyama T, Nanko S, Yoshikawa T, Kato T. Association of mitochondrial complex I subunit gene NDUFV2 at 18p11 with bipolar disorder

in Japanese and the National Institute of Mental Health pedigrees. *Biol Psychiatry*. 2004 Oct 1, 56(7):483-9. PMID: 15450783

79. Detera-Wadleigh SD, Badner JA, Berrettini WH, Yoshikawa T, Goldin LR, Turner G, Rollins DY, Moses T, Sanders AR, Karkera JD, Esterling LE, Zeng J, Ferraro TN, Guroff JJ, Kazuba D, Maxwell ME, Nurnberger JI Jr, Gershon ES. A high-density genome scan detects evidence for a bipolar-disorder susceptibility locus on 13q32 and other potential loci on 1q32 and 18p11.2. *Proc Natl Acad Sci USA*. 1999 May 11, 96(10):5604-9. PMID: 10318931

80. Berrettini WH, Ferraro TN, Goldin LR, Weeks DE, Detera-Wadleigh S, Nurnberger JI Jr, Gershon ES. Chromosome 18 DNA markers and manic-depressive illness: evidence for a susceptibility gene. *Proc Natl Acad Sci USA*. 1994 Jun 21, 91(13):5918-21. PMID: 8016089

81. Stine OC, Xu J, Koskela R, McMahon FJ, Gschwend M, Friddle C, Clark CD, McInnis MG, Simpson SG, Breschel TS, Vishio E, Riskin K, Feilotter H, Chen E, Shen S, Folstein S, Meyers DA, Botstein D, Marr TG, DePaulo JR. Evidence for linkage of bipolar disorder to chromosome 18 with a parent-of-origin effect. *Am J Hum Genet*. 1995 Dec, 57(6):1384-94. PMID: 8533768

82. Segurado R, Detera-Wadleigh SD, Levinson DF, Lewis CM, Gill M, Nurnberger JI Jr, Craddock N, DePaulo JR, Baron M, Gershon ES, Ekholm J, Cichon S, Turecki G, Claes S, Kelsoe JR, Schofield PR, Badenhop RF, Morissette J, Coon H, Blackwood D, McInnes LA, Foroud T, Edenberg HJ, Reich T, Rice JP, Goate A, McInnis MG, McMahon FJ, Badner JA, Goldin LR, Bennett P, Willour VL, Zandi PP, Liu J, Gilliam C, Juo SH, Berrettini WH, Yoshikawa T, Peltonen L, Lönnqvist

J, Nöthen MM, Schumacher J, Windemuth C, Rietschel M, Propping P, Maier W, Alda M, Grof P, Rouleau GA, Del-Favero J, Van Broeckhoven C, Mendlewicz J, Adolfsson R, Spence MA, Luebbert H, Adams LJ, Donald JA, Mitchell PB, Barden N, Shink E, Byerley W, Muir W, Visscher PM, Macgregor S, Gurling H, Kalsi G, McQuillin A, Escamilla MA, Reus VI, Leon P, Freimer NB, Ewald H, Kruse TA, Mors O, Radhakrishna U, Blouin JL, Antonarakis SE, Akarsu N. Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: Bipolar disorder. *Am J Hum Genet*. 2003 Jul, 73(1):49-62. PMID: 12802785

83. Andreazza AC, Shao L, Wang JF, Young LT. Mitochondrial complex I activity and oxidative damage to mitochondrial proteins in the prefrontal cortex of patients with bipolar disorder. *Arch Gen Psychiatry*. 2010 Apr;67(4):360-8. Erratum in: *Arch Gen Psychiatry*. 2010 Dec, 67(12):1254. PMID: 20368511

84. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet*. 1995 Nov, 11(3):241-7. PMID: 7581446

85. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science*. 1996 Sep 13, 273(5281):1516-7. PMID: 8801636

86. Hoh J, Ott J. Mathematical multi-locus approaches to localizing complex human trait genes. *Nat Rev Genet*. 2003 Sep, 4(9):701-9. PMID: 12951571

87. Jiang JQ, Dress AW, Chen M. Towards prediction and prioritization of disease genes by the modularity of human phenome-genome assembled network. *J Integr Bioinform*. 2010 Nov 22, 7(2). PMID: 21098881

88. Köhler S, Bauer S, Horn D, Robinson PN. Walking the interactome for prioritization of candidate disease genes. *Am J Hum Genet*. 2008 Apr, 82(4):949-58. PMID: 18371930

89. Ideker T, Ozier O, Schwikowski B, Siegel AF. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics*. 2002, 18 Suppl 1:S233-40. PMID: 12169552

90. Musani SK, Shriner D, Liu N, Feng R, Coffey CS, Yi N, Tiwari HK, Allison DB. Detection of gene x gene interactions in genome-wide association studies of human population data. *Hum Hered*. 2007, 63(2):67-84. PMID: 17283436

91. Nilsson FM, Kessing LV, Bolwig TG. Increased risk of developing Parkinson's disease for patients with major affective disorder: a register study. *Acta Psychiatr Scand*. 2001 Nov, 104(5):380-6. PMID: 11722320

92. Shiba M, Bower JH, Maraganore DM, McDonnell SK, Peterson BJ, Ahlskog JE, Schaid DJ, Rocca WA. Anxiety disorders and depressive disorders preceding Parkinson's disease: a case-control study. *Mov Disord*. 2000 Jul, 15(4):669-77. PMID: 10928577

93. Baranzini SE, Galwey NW, Wang J, Khankhanian P, Lindberg R, Pelletier D, Wu W, Uitdehaag BM, Kappos L; GeneMSA Consortium, Polman CH, Matthews PM, Hauser SL, Gibson RA, Oksenberg JR, Barnes MR. Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. *Hum Mol Genet*. 2009 Jun 1, 18(11):2078-90. PMID: 19286671

94. Califano A, Butte A, Friend S, Ideker T, and Schadt E. Integrative Network-based Association Studies: Leveraging cell regulatory models in the post-GWAS era.

Available from Nature Precedings (2011)

<http://dx.doi.org/10.1038/npre.2011.5732.1>

Appendix

Algorithm Pseudocode

*Main_GA (***M, N, P***) {*
    Population = *Randomly create* **M** *groups of SNPs // Create_population(***M***)*
    Do the following for **N** *generations {*
        Calculate the fitness of each of the **M** *group of SNPs // Fitness(***M***)*
        Parents = *Select* **M/2** *of the SNP groups based on their fitness*
        Offspring = *Mutate each parent with probability* **P**
        New Population = **Parents** + **Offspring**
    }
    Calculate the fitness for each group of SNPs in the final generation
}

*Create_population (***M***) {*
    Get all genes from the database and create a roulette wheel based on gene scores
    Do the following **M** *times {*
        Gene1 = *Spin the roulette wheel to select a gene*
        Get all SNPs that mapped to Gene1
        Create another roulette wheel based on the location of the SNP (intragenic, 5', 3')
        SNP1 = *Spin the roulette wheel and select a SNP*

        Get all genes that interact with Gene1
        Create a roulette wheel based on interaction score
        Gene2 = *Spin the roulette wheel to select a gene*
        Get all SNPs that mapped to Gene2 (and are not SNP1)
        Create another roulette wheel based on the location of the SNP (intragenic, 5', 3')
        SNP2 = *Spin the roulette wheel and select a SNP*

        Add this Gene pair/SNP pair to the population
    }
}

*Fitness (***M***) {*
    Do the following for each of the **M** *SNP groups in the population {*
        Get all SNPs in the group
        Create a contingency table with all genotype combinations for the SNP group
        Fill the contingency table with counts of all genotype combos for cases and controls
        Calculate a Chi-Square p-value for the contingency table
    }
}

*Select (***M, T, H, K***) {*
    *Migrants = Create_population(***K***) // Create* **K** *new SNP groups to maintain diversity*

    *If (***T** *= "Truncate") {*
        Get all SNP groups and their fitness
        Parents *= the (***M***/2 –* **K***) SNP groups with the smallest p-values*
    }

    *If (***T** *= "Roulette") {*
        Get all SNP groups and their fitness
        Create a roulette wheel based on –log$_{10}$*(fitness)*
        Parents *= Spin the roulette wheel (***M***/2 –* **K***) times*
    }

    *If (***T** *= "Hybrid") {*
        Get all SNP groups and their fitness
        Best Groups *= the* **H** *SNP groups with the lowest p-values*
        Create a roulette wheel based on –log$_{10}$*(fitness) for the remaining SNP groups*
        Parents *= Spin the roulette wheel (***M***/2 –* **K** *–* **H***) times*
        Parents = Parents + Best Groups
    }

    Add **Migrants** *and* **Parents** *to the next generation*
}

```
Mutate (M, P) { // For Hypothesis 1
    Get all M/2 SNP groups that were selected from the previous generation
    Do the following for all M/2 SNP groups {
        Mutate with probability P {
            Get the size of the SNP group
            Determine which actions are possible (Remove, Add, Alter)
            Randomly select which action to take
            If (Remove = TRUE) {
                Get all genes/SNPs in the group
                Randomly select a SNP to remove
                If two or more interacting genes remain, they are added to the population
                Else, a new group is created
            }
            If (Add = TRUE) {
                Get all genes/SNPs in the group
                Get all genes that interact with the group
                Create a roulette wheel based on gene scores and interaction scores
                Spin the roulette wheel to choose a new gene to be added to the group
                Get all SNPs in this gene and create a roulette wheel based on SNP location
                Spin the roulette wheel to choose a SNP
            }
            If (Alter = TRUE) {
                Get all genes/SNPs in the group
                Randomly select a gene/SNP to remove
                If two or more interacting genes remain {
                    Old Group = the remaining interactions from the original group
                    Get all genes that interact with these remaining genes
                    Create a roulette wheel based on gene scores and interaction scores
                    New Gene = Spin the roulette wheel to choose a gene to add the group
                    Get all SNPs in gene and create a roulette wheel based on SNP location
                    New SNP = Spin the roulette wheel to choose a SNP
                    New Group = Old Group + New SNP
                }
                Else {
                    Get all genes that interact with the remaining genes
                    Create a roulette wheel based on gene scores and interaction scores
                    New Gene = Spin the roulette wheel to choose a gene to add the group
                    Get all SNPs in gene and create a roulette wheel based on SNP location
                    New SNP = Spin the roulette wheel to choose a SNP
                    New Group = New SNP + SNP from the original group
                }
            }
        }
    }
}
```

Compiled Research on Bipolar Disorder

Linkage Regions Associated with Bipolar Disorder:

Cytoband	Genomic Coordinates	References
4p16	chr4:1-10900000	1, 2, 3, 4, 5
4q32	chr4:155100001-170400000	5, 6, 7
4q35	chr4:182600001-191273063	6, 8, 9
6q16-22	chr6:92100001-130400000	10, 11, 12, 13
8q24	chr8:117700001-146274826	6, 10, 14, 15, 16
12q24	chr12:107500001-132349534	5, 6, 17, 18, 19, 20
13q31-33	chr13:77800001-109100000	21, 22, 23, 24, 25, 26
16p12-13	chr16:1-27600000	4, 7, 16, 20, 27, 28, 29, 30, 31, 32, 170
18p11.2	chr18:7200001-15400000	22, 33, 34, 35
18p11.3	chr18:1-7200000	28, 35
18q12	chr18:23300001-41800000	31, 35, 36
18q21-23	chr18:41800001-76117153	6, 28, 31, 34, 35, 37, 38, 39, 40, 41, 42
21q22	chr21:30500001-46944323	43, 44, 45, 46
22q11-13	chr22:11800001-49691432	21, 22, 25, 26
Xq24-27	chrX:116800001-146900000	47, 48, 49, 50

Genes Associated with Bipolar Disorder:

Ensembl Gene ID	Gene Symbol	Other Gene Names	References
ENSG00000179869	ABCA13		51
ENSG00000159640	ACE		52, 53
ENSG00000135744	AGT		54
ENSG00000142208	AKT1		55, 56, 57
ENSG00000151150	ANK3		58, 59, 60, 61
ENSG00000176697	BDNF		62, 63
ENSG00000151067	CACNA1C		59, 64
ENSG00000093010	COMT		65
ENSG00000132437	DDC		66, 67
ENSG00000162946	DISC1		68
ENSG00000184845	DRD1		69, 70, 71, 72, 73
ENSG00000149295	DRD2		74, 75, 76
ENSG00000069696	DRD4		65, 77
ENSG00000182346	DAOA	G72, G30	81, 82
ENSG00000022355	GABRA1		83
ENSG00000011677	GABRA3		84
ENSG00000186297	GABRA5		85, 86, 87
ENSG00000183454	GRIN2A		88, 89
ENSG00000150086	GRIN2B		89, 90
ENSG00000149403	GRIK4		91, 92, 93, 94

ENSG00000100077	ADRBK2	GRK3	95, 96
ENSG00000082701	GSK3B		97, 98, 99
ENSG00000102468	HTR2A		100, 101, 102
ENSG00000147246	HTR2C		102, 103, 104
ENSG00000164270	HTR4		105
ENSG00000158748	HTR6		106
ENSG00000141401	IMPA2		107, 108, 109, 110, 111
ENSG00000166086	JAM3		112
ENSG00000189221	MAOA		113, 114, 115, 116, 117, 118
ENSG00000149294	NCAM1		119, 120, 121, 122, 123
ENSG00000178127	NDUFV2		124, 125, 126, 127, 128, 129, 130
ENSG00000151322	NPAS3		131, 132
ENSG00000157168	NRG1		59, 133, 134
ENSG00000170890	PLA2G1B	PLA2A	135, 136, 137
ENSG00000124181	PLCG1		138, 139
ENSG00000112033	PPARD		140
ENSG00000141873	SLC39A3	ZIP3	112
ENSG00000142319	SLC6A3	DAT1	141, 142
ENSG00000108576	SLC6A4	SERT, HTT, 5HTT	143
ENSG00000162009	SSTR5		144, 145
ENSG00000180176	TH		146, 147, 148, 149, 150
ENSG00000204490*	TNF	TNFA	151, 152, 153
ENSG00000129167	TPH1	TPH	154, 155, 156
ENSG00000139287	TPH2		157, 158, 159
ENSG00000109501	WFS1		160, 161, 162
ENSG00000100219	XBP1		163, 164, 165
ENSG00000128245	YWHAH		166

Functional Concepts Relevant to Bipolar Disorder [167]:

acetylcholine
adrenergic
calcitonin
cAMP
catecholamine
cholecystokinin
cholinergic
corticotropin
dexamethasone
dopamine
GABA
glucocorticoid
glutamate

glycogen synthase
g-protein
lithium
neurotransmitter
neurotensin
noradrenergic
norepinephrine
opiate
opioid
oxytocin
phosphoinositide
protein kinase C
serotonin
somatostatin
substance P
thyrotropin
tryptophan
valproate
vasoactive intestinal polypeptide
vasopressin

Previous Genome-wide Association Studies on Bipolar Disorder:

First Author	Genomic Region	Gene	SNP-Risk Allele	Odds Ratio	P-value
Smith [61]	19q13.11	DPY19L3	rs2111504-T	1.74	2.00E-6
Scott [171]	2q11.2	Intergenic	rs6733011-G	1.17	0.000003
Scott	3p21.1	ITIH1, NEK4	rs1042779-A	1.19	2.00E-07
Scott	4q12	KIT	rs2537859-T	1.16	0.000004
Scott	5q15	MCTP1	rs17418283-C	1.21	1.00E-07
Scott	3q26.1	NR	rs7427021-G	1.16	0.000005
Scott	2p12	CTNNA2	rs13409348-G	1.2	0.000003
Scott	9q34.13	NR	rs2905072-A	1.21	0.000006
Scott	17q21.33	NR	rs1035050-T	1.17	0.000009
Scott	15q23	NR	rs6494849-A	1.23	0.000007
Scott	19p13.3	NR	rs7250872-T	1.21	0.000002
Scott	11q13.1	NR	rs2242663-T	1.2	0.000001
Scott	8p12	NR	rs6990255-T	1.33	0.000006
Scott	1p32.1	NF1A	rs472913-C	1.18	2.00E-07
Hattori [172]	NR	NR	NR	NR	NS
Ferreira [59]	10q21.2	ANK3	rs10994336-T	1.45	9.00E-09
Ferreira	9q33.3	NR	rs4130590-?	1.16	0.000003
Ferreira	11q14.1	NR	rs12290811-A	1.2	4E-06
Ferreira	15q14	C15orf53,	rs12899449-?	1.2	4.00E-07

		RASGRP1			
Ferreira	2q11.2	Intergenic	rs2314398-?	1.17	0.000003
Ferreira	6q25.2	NR	rs17082664-G	1.22	4E-06
Ferreira	3p22.3	NR	rs4380451-?	1.18	4E-06
Ferreira	12q13.12	CACNA1C	rs1006737-A	1.18	7.00E-08
Ferreira	10q22.3	NR	rs703965-?	1.15	8E-06
Ferreira	3p26.2	NR	rs1601875-?	1.14	7E-06
Ferreira	18p11.32	NR	rs7226677-G	1.24	7E-06
Ferreira	15q25.1	NR	rs2278702-?	1.21	0.000006
Ferreira	11q24.2	NR	rs544368-T	1.22	0.000006
Ferreira	9q31.3	NR	rs7042161-?	1.15	0.000006
Ferreira	1p21.2	NR	rs1948368-?	1.15	0.000006
Ferreira	3p24.3	NR	rs11720452-?	1.15	0.000005
Ferreira	14q11.2	NR	rs12436436-C	1.3	0.000005
Ferreira	14q13.1	NR	rs8015959-T	1.59	0.000005
Ferreira	3p24.3	NR	rs3821396-A	1.23	0.000005
Ferreira	9p13.3	NR	rs216345-T	1.15	4E-06
Ferreira	15q14	NR	rs16966460-G	1.26	4E-06
Sklar [173]	NR	NR	NR	NR	NS
WTCCC [174]	16p12.1	PALB2, NDUFAB1, DCTN5	rs420259-A	2.08	6.00E-08
WTCCC	6p21	NR	rs6458307-?	1.19	4E-06
WTCCC	3q27	NR	rs683395-G	1.47	0.000005
WTCCC	14q32.33	NR	rs11622475-C	1.13	8E-06
WTCCC	14q23.1	NR	rs10134944-T	1.45	7E-06
WTCCC	2q37	NR	rs2953145-C	1.84	7E-06
WTCCC	20p13	NR	rs3761218-C	1.03	7E-06
Baum [175]	13q14.11	DGKH	rs1012053-A	1.59	2.00E-08
Cichon [176]	19q13.11	NCAN	rs1064395-A	1.17	2.00E-9
Djurovic [177]	NS	NS	NS	NS	NS
Lee [178]	7p21.1	SP8	rs2709736-G	1.44	5.00E-7
Lee	15q26.1	ST8SIA2, C15orf32	rs8040009-C	1.4	6.00E-6

*Only associations with p-value < 10^{-5} are shown; NR = Not Reported; NS = Not Significant. Data from *http://genome.gov/gwastudies*;

References for the Compiled Research on Bipolar Disorder

1. Blackwood DH, He L, Morris SW, McLean A, Whitton C, Thomson M, Walker MT, Woodburn K, Sharp CM, Wright AF, Shibasaki Y, St Clair DM, Porteous DJ, Muir WJ. A locus for bipolar affective disorder on chromosome 4p. *Nat Genet*. 1996 Apr, 12(4):427-30. PMID: 8630499
2. Ewald H, Degn B, Mors O, Kruse TA. Support for the possible locus on chromosome 4p16 for bipolar affective disorder. *Mol Psychiatry*. 1998 Sep, 3(5):442-8. PMID: 9774780
3. Ginns EI, St Jean P, Philibert RA, Galdzicka M, Damschroder-Williams P, Thiel B, Long RT, Ingraham LJ, Dalwaldi H, Murray MA, Ehlert M, Paul S, Remortel BG, Patel AP, Anderson MC, Shaio C, Lau E, Dymarskaia I, Martin BM, Stubblefield B, Falls KM, Carulli JP, Keith TP, Fann CS, Lacy LG, Allen CR, Hostetter AM, Elston RC, Schork NJ, Egeland JA, Paul SM. A genome-wide search for chromosomal loci linked to mental health wellness in relatives at high risk for bipolar affective disorder among the Old Order Amish. *Proc Natl Acad Sci U S A*. 1998 Dec 22, 95(26):15531-6. PMID: 9861003
4. Ewald H, Flint T, Kruse TA, Mors O. A genome-wide scan shows significant linkage between bipolar disorder and chromosome 12q24.3 and suggestive linkage to chromosomes 1p22-21, 4p16, 6q14-22, 10q26 and 16p13.3. *Mol Psychiatry*. 2002, 7(7):734-44. PMID: 12192618
5. Fullerton JM, Donald JA, Mitchell PB, Schofield PR. Two-dimensional genome scan identifies multiple genetic interactions in bipolar affective disorder. *Biol Psychiatry*. 2010 Mar 1, 67(5):478-86. PMID: 20022591

6. McInnis MG, Lan TH, Willour VL, McMahon FJ, Simpson SG, Addington AM, MacKinnon DF, Potash JB, Mahoney AT, Chellis J, Huo Y, Swift-Scanlan T, Chen H, Koskela R, Stine OC, Jamison KR, Holmans P, Folstein SE, Ranade K, Friddle C, Botstein D, Marr T, Beaty TH, Zandi P, DePaulo JR. Genome-wide scan of bipolar disorder in 65 pedigrees: supportive evidence for linkage at 8q24, 18q22, 4q32, 2p12, and 13q12. *Mol Psychiatry*. 2003 Mar, 8(3):288-98. PMID: 12660801

7. Ekholm JM, Kieseppä T, Hiekkalinna T, Partonen T, Paunio T, Perola M, Ekelund J, Lönnqvist J, Pekkarinen-Ijäs P, Peltonen L. Evidence of susceptibility loci on 4q32 and 16p12 for bipolar disorder. *Hum Mol Genet*. 2003 Aug 1, 12(15):1907-15. PMID: 12874110

8. Adams LJ, Mitchell PB, Fielder SL, Rosso A, Donald JA, Schofield PR. A susceptibility locus for bipolar affective disorder on chromosome 4q35. *Am J Hum Genet*. 1998 May, 62(5):1084-91. PMID: 9545396

9. Willour VL, Zandi PP, Huo Y, Diggs TL, Chellis JL, MacKinnon DF, Simpson SG, McMahon FJ, Potash JB, Gershon ES, Reich T, Foroud T, Nurnberger JI Jr, DePaulo JR Jr, McInnis MG. Genome scan of the fifty-six bipolar pedigrees from the NIMH genetics initiative replication sample: chromosomes 4, 7, 9, 18, 19, 20, and 21. *Am J Med Genet B Neuropsychiatr Genet*. 2003 Aug 15, 121B(1):21-7. PMID: 12898570

10. McQueen MB, Devlin B, Faraone SV, Nimgaonkar VL, Sklar P, Smoller JW, Abou Jamra R, Albus M, Bacanu SA, Baron M, Barrett TB, Berrettini W, Blacker D, Byerley W, Cichon S, Coryell W, Craddock N, Daly MJ, Depaulo JR, Edenberg HJ, Foroud T, Gill M, Gilliam TC, Hamshere M, Jones I, Jones L, Juo SH, Kelsoe JR,

Lambert D, Lange C, Lerer B, Liu J, Maier W, Mackinnon JD, McInnis MG, McMahon FJ, Murphy DL, Nothen MM, Nurnberger JI, Pato CN, Pato MT, Potash JB, Propping P, Pulver AE, Rice JP, Rietschel M, Scheftner W, Schumacher J, Segurado R, Van Steen K, Xie W, Zandi PP, Laird NM. 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 Oct, 77(4):582-95. PMID: 16175504

11. Dick DM, Foroud T, Flury L, Bowman ES, Miller MJ, Rau NL, Moe PR, Samavedy N, El-Mallakh R, Manji H, Glitz DA, Meyer ET, Smiley C, Hahn R, Widmark C, McKinney R, Sutton L, Ballas C, Grice D, Berrettini W, Byerley W, Coryell W, DePaulo R, MacKinnon DF, Gershon ES, Kelsoe JR, McMahon FJ, McInnis M, Murphy DL, Reich T, Scheftner W, Nurnberger JI Jr. Genomewide linkage analyses of bipolar disorder: a new sample of 250 pedigrees from the National Institute of Mental Health Genetics Initiative. *Am J Hum Genet*. 2003 Jul, 73(1):107-14. Erratum in: *Am J Hum Genet*. 2003 Oct, 73(4):979. PMID: 12772088

12. Schulze TG, Buervenich S, Badner JA, Steele CJ, Detera-Wadleigh SD, Dick D, Foroud T, Cox NJ, MacKinnon DF, Potash JB, Berrettini WH, Byerley W, Coryell W, DePaulo JR Jr, Gershon ES, Kelsoe JR, McInnis MG, Murphy DL, Reich T, Scheftner W, Nurnberger JI Jr, McMahon FJ. Loci on chromosomes 6q and 6p interact to increase susceptibility to bipolar affective disorder in the national institute of mental health genetics initiative pedigrees. *Biol Psychiatry*. 2004 Jul 1, 56(1):18-23. PMID: 15219468

13. Middleton FA, Pato MT, Gentile KL, Morley CP, Zhao X, Eisener AF, Brown A, Petryshen TL, Kirby AN, Medeiros H, Carvalho C, Macedo A, Dourado A, Coelho I, Valente J, Soares MJ, Ferreira CP, Lei M, Azevedo MH, Kennedy JL, Daly MJ, Sklar P, Pato CN. Genomewide linkage analysis of bipolar disorder by use of a high-density single-nucleotide-polymorphism (SNP) genotyping assay: a comparison with microsatellite marker assays and finding of significant linkage to chromosome 6q22. *Am J Hum Genet*. 2004 May, 74(5):886-97. PMID: 15060841

14. Cichon S, Schumacher J, Müller DJ, Hürter M, Windemuth C, Strauch K, Hemmer S, Schulze TG, Schmidt-Wolf G, Albus M, Borrmann-Hassenbach M, Franzek E, Lanczik M, Fritze J, Kreiner R, Reuner U, Weigelt B, Minges J, Lichtermann D, Lerer B, Kanyas K, Baur MP, Wienker TF, Maier W, Rietschel M, Propping P, Nöthen MM. A genome screen for genes predisposing to bipolar affective disorder detects a new susceptibility locus on 8q. *Hum Mol Genet*. 2001 Dec 1, 10(25):2933-44. Erratum in: *Hum Mol Genet*. 2002 Jul 1, 11(14):1685. PMID: 11741836

15. Avramopoulos D, Willour VL, Zandi PP, Huo Y, MacKinnon DF, Potash JB, DePaulo JR Jr, McInnis MG. Linkage of bipolar affective disorder on chromosome 8q24: follow-up and parametric analysis. *Mol Psychiatry*. 2004 Feb, 9(2):191-6. PMID: 14966477

16. McInnis MG, Dick DM, Willour VL, Avramopoulos D, MacKinnon DF, Simpson SG, Potash JB, Edenberg HJ, Bowman ES, McMahon FJ, Smiley C, Chellis JL, Huo Y, Diggs T, Meyer ET, Miller M, Matteini AT, Rau NL, DePaulo JR, Gershon ES, Badner JA, Rice JP, Goate AM, Detera-Wadleigh SD, Nurnberger JI, Reich T, Zandi PP, Foroud TM. Genome-wide scan and conditional analysis in bipolar

disorder: evidence for genomic interaction in the National Institute of Mental Health genetics initiative bipolar pedigrees. *Biol Psychiatry*. 2003 Dec 1, 54(11):1265-73. PMID: 14643094

17. Craddock N, McGuffin P, Owen M. Darier's disease cosegregating with affective disorder. Br J Psychiatry. 1994 Aug, 165(2):272. PMID: 7832868
18. Ewald H, Degn B, Mors O, Kruse TA. Significant linkage between bipolar affective disorder and chromosome 12q24. *Psychiatr Genet*. 1998 Autumn, 8(3):131-40. PMID: 9800214
19. Curtis D, Kalsi G, Brynjolfsson J, McInnis M, O'Neill J, Smyth C, Moloney E, Murphy P, McQuillin A, Petursson H, Gurling H. Genome scan of pedigrees multiply affected with bipolar disorder provides further support for the presence of a susceptibility locus on chromosome 12q23-q24, and suggests the presence of additional loci on 1p and 1q. *Psychiatr Genet*. 2003 Jun, 13(2):77-84. PMID: 12782963
20. Cassidy F, Zhao C, Badger J, Claffey E, Dobrin S, Roche S, McKeon P. Genome-wide scan of bipolar disorder and investigation of population stratification effects on linkage: support for susceptibility loci at 4q21, 7q36, 9p21, 12q24, 14q24, and 16p13. *Am J Med Genet B Neuropsychiatr Genet*. 2007 Sep 5, 144B(6):791-801. PMID: 17455214
21. Badner JA, Gershon ES. Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Mol Psychiatry*. 2002, 7(4):405-11. PMID: 11986984
22. Detera-Wadleigh SD, Badner JA, Berrettini WH, Yoshikawa T, Goldin LR, Turner G, Rollins DY, Moses T, Sanders AR, Karkera JD, Esterling LE, Zeng J, Ferraro

TN, Guroff JJ, Kazuba D, Maxwell ME, Nurnberger JI Jr, Gershon ES. A high-density genome scan detects evidence for a bipolar-disorder susceptibility locus on 13q32 and other potential loci on 1q32 and 18p11.2. *Proc Natl Acad Sci USA*. 1999 May 11, 96(10):5604-9. PMID: 10318931

23. Liu C, Badner JA, Christian SL, Guroff JJ, Detera-Wadleigh SD, Gershon ES. Fine mapping supports previous linkage evidence for a bipolar disorder susceptibility locus on 13q32. *Am J Med Genet*. 2001 May 8, 105(4):375-80. PMID: 11378853
24. Shaw SH, Mroczkowski-Parker Z, Shekhtman T, Alexander M, Remick RA, Sadovnick AD, McElroy SL, Keck PE Jr, Kelsoe JR. Linkage of a bipolar disorder susceptibility locus to human chromosome 13q32 in a new pedigree series. *Mol Psychiatry*. 2003 May, 8(5):558-64. PMID: 12808435
25. Kelsoe JR, Spence MA, Loetscher E, Foguet M, Sadovnick AD, Remick RA, Flodman P, Khristich J, Mroczkowski-Parker Z, Brown JL, Masser D, Ungerleider S, Rapaport MH, Wishart WL, Luebbert H. A genome survey indicates a possible susceptibility locus for bipolar disorder on chromosome 22. *Proc Natl Acad Sci USA*. 2001 Jan 16, 98(2):585-90. PMID: 11149935
26. Potash JB, Zandi PP, Willour VL, Lan TH, Huo Y, Avramopoulos D, Shugart YY, MacKinnon DF, Simpson SG, McMahon FJ, DePaulo JR Jr, McInnis MG. Suggestive linkage to chromosomal regions 13q31 and 22q12 in families with psychotic bipolar disorder. *Am J Psychiatry*. 2003 Apr, 160(4):680-6. PMID: 12668356

27. Ewald H, Mors O, Flint T, Koed K, Eiberg H, Kruse TA. A possible locus for manic depressive illness on chromosome 16p13. *Psychiatr Genet*. 1995 Summer, 5(2):71-81. PMID: 7551966

28. McInnes LA, Escamilla MA, Service SK, Reus VI, Leon P, Silva S, Rojas E, Spesny M, Baharloo S, Blankenship K, Peterson A, Tyler D, Shimayoshi N, Tobey C, Batki S, Vinogradov S, Meza L, Gallegos A, Fournier E, Smith LB, Barondes SH, Sandkuijl LA, Freimer NB. A complete genome screen for genes predisposing to severe bipolar disorder in two Costa Rican pedigrees. *Proc Natl Acad Sci USA*. 1996 Nov 12, 93(23):13060-5. PMID: 8917544

29. Dick DM, Foroud T, Edenberg HJ, Miller M, Bowman E, Rau NL, DePaulo JR, McInnis M, Gershon E, McMahon F, Rice JP, Bierut LJ, Reich T, Nurnberger J Jr. Apparent replication of suggestive linkage on chromosome 16 in the NIMH genetics initiative bipolar pedigrees. *Am J Med Genet*. 2002 May 8, 114(4):407-12. PMID: 11992562

30. Edenberg HJ, Foroud T, Conneally PM, Sorbel JJ, Carr K, Crose C, Willig C, Zhao J, Miller M, Bowman E, Mayeda A, Rau NL, Smiley C, Rice JP, Goate A, Reich T, Stine OC, McMahon F, DePaulo JR, Meyers D, Detera-Wadleigh SD, Goldin LR, Gershon ES, Blehar MC, Nurnberger JI Jr. Initial genomic scan of the NIMH genetics initiative bipolar pedigrees: chromosomes 3, 5, 15, 16, 17, and 22. *Am J Med Genet*. 1997 May 31, 74(3):238-46. PMID: 9184305

31. Maziade M, Roy MA, Chagnon YC, Cliche D, Fournier JP, Montgrain N, Dion C, Lavallée JC, Garneau Y, Gingras N, Nicole L, Pirès A, Ponton AM, Potvin A, Wallot H, Mérette C. Shared and specific susceptibility loci for schizophrenia and

bipolar disorder: a dense genome scan in Eastern Quebec families. *Mol Psychiatry*. 2005 May, 10(5):486-99. PMID: 15534619

32. Mérette C, Roy MA, Bureau A, Fournier A, Emond C, Cliche D, Jomphe V, Chagnon YC, Maziade M. Replication of linkage with bipolar disorder on chromosome 16p in the Eastern Quebec population. *Am J Med Genet B Neuropsychiatr Genet*. 2008 Sep 5, 147B(6):737-44. PMID: 18165973

33. Berrettini WH, Ferraro TN, Goldin LR, Weeks DE, Detera-Wadleigh S, Nurnberger JI Jr, Gershon ES. Chromosome 18 DNA markers and manic-depressive illness: evidence for a susceptibility gene. *Proc Natl Acad Sci USA*. 1994 Jun 21, 91(13):5918-21. PMID: 8016089

34. Stine OC, Xu J, Koskela R, McMahon FJ, Gschwend M, Friddle C, Clark CD, McInnis MG, Simpson SG, Breschel TS, Vishio E, Riskin K, Feilotter H, Chen E, Shen S, Folstein S, Meyers DA, Botstein D, Marr TG, DePaulo JR. Evidence for linkage of bipolar disorder to chromosome 18 with a parent-of-origin effect. *Am J Hum Genet*. 1995 Dec, 57(6):1384-94. PMID: 8533768

35. Segurado R, Detera-Wadleigh SD, Levinson DF, Lewis CM, Gill M, Nurnberger JI Jr, Craddock N, DePaulo JR, Baron M, Gershon ES, Ekholm J, Cichon S, Turecki G, Claes S, Kelsoe JR, Schofield PR, Badenhop RF, Morissette J, Coon H, Blackwood D, McInnes LA, Foroud T, Edenberg HJ, Reich T, Rice JP, Goate A, McInnis MG, McMahon FJ, Badner JA, Goldin LR, Bennett P, Willour VL, Zandi PP, Liu J, Gilliam C, Juo SH, Berrettini WH, Yoshikawa T, Peltonen L, Lönnqvist J, Nöthen MM, Schumacher J, Windemuth C, Rietschel M, Propping P, Maier W, Alda M, Grof P, Rouleau GA, Del-Favero J, Van Broeckhoven C, Mendlewicz J,

Adolfsson R, Spence MA, Luebbert H, Adams LJ, Donald JA, Mitchell PB, Barden N, Shink E, Byerley W, Muir W, Visscher PM, Macgregor S, Gurling H, Kalsi G, McQuillin A, Escamilla MA, Reus VI, Leon P, Freimer NB, Ewald H, Kruse TA, Mors O, Radhakrishna U, Blouin JL, Antonarakis SE, Akarsu N. Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: Bipolar disorder. *Am J Hum Genet*. 2003 Jul, 73(1):49-62. PMID: 12802785

36. Maziade M, Roy MA, Rouillard E, Bissonnette L, Fournier JP, Roy A, Garneau Y, Montgrain N, Potvin A, Cliche D, Dion C, Wallot H, Fournier A, Nicole L, Lavallée JC, Mérette C. A search for specific and common susceptibility loci for schizophrenia and bipolar disorder: a linkage study in 13 target chromosomes. *Mol Psychiatry*. 2001 Nov, 6(6):684-93. PMID: 11673797

37. Freimer NB, Reus VI, Escamilla MA, McInnes LA, Spesny M, Leon P, Service SK, Smith LB, Silva S, Rojas E, Gallegos A, Meza L, Fournier E, Baharloo S, Blankenship K, Tyler DJ, Batki S, Vinogradov S, Weissenbach J, Barondes SH, Sandkuijl LA. Genetic mapping using haplotype, association and linkage methods suggests a locus for severe bipolar disorder (BPI) at 18q22-q23. *Nat Genet*. 1996 Apr, 12(4):436-41. PMID: 8630501

38. McMahon FJ, Simpson SG, McInnis MG, Badner JA, MacKinnon DF, DePaulo JR. Linkage of bipolar disorder to chromosome 18q and the validity of bipolar II disorder. *Arch Gen Psychiatry*. 2001 Nov, 58(11):1025-31. PMID: 11695948

39. McMahon FJ, Hopkins PJ, Xu J, McInnis MG, Shaw S, Cardon L, Simpson SG, MacKinnon DF, Stine OC, Sherrington R, Meyers DA, DePaulo JR. Linkage of

bipolar affective disorder to chromosome 18 markers in a new pedigree series. *Am J Hum Genet*. 1997 Dec, 61(6):1397-404. PMID: 9399888

40. De bruyn A, Souery D, Mendelbaum K, Mendlewicz J, Van Broeckhoven C. Linkage analysis of families with bipolar illness and chromosome 18 markers. *Biol Psychiatry*. 1996 Apr 15, 39(8):679-88. PMID: 8731454

41. Coon H, Hoff M, Holik J, Hadley D, Fang N, Reimherr F, Wender P, Byerley W. Analysis of chromosome 18 DNA markers in multiplex pedigrees with manic depression. *Biol Psychiatry*. 1996 Apr 15, 39(8):689-96. PMID: 8731455

42. Nöthen MM, Cichon S, Rohleder H, Hemmer S, Franzek E, Fritze J, Albus M, Borrmann-Hassenbach M, Kreiner R, Weigelt B, Minges J, Lichtermann D, Maier W, Craddock N, Fimmers R, Höller T, Baur MP, Rietschel M, Propping P. Evaluation of linkage of bipolar affective disorder to chromosome 18 in a sample of 57 German families. *Mol Psychiatry*. 1999 Jan, 4(1):76-84. PMID: 10089014

43. Straub RE, Lehner T, Luo Y, Loth JE, Shao W, Sharpe L, Alexander JR, Das K, Simon R, Fieve RR, et al. A possible vulnerability locus for bipolar affective disorder on chromosome 21q22.3. *Nat Genet*. 1994 Nov, 8(3):291-6. PMID: 7874172

44. Liu J, Juo SH, Terwilliger JD, Grunn A, Tong X, Brito M, Loth JE, Kanyas K, Lerer B, Endicott J, Penchaszadeh G, Gilliam TC, Baron M. A follow-up linkage study supports evidence for a bipolar affective disorder locus on chromosome 21q22. *Am J Med Genet*. 2001 Mar 8, 105(2):189-94. PMID: 11304836

45. Detera-Wadleigh SD, Badner JA, Goldin LR, Berrettini WH, Sanders AR, Rollins DY, Turner G, Moses T, Haerian H, Muniec D, Nurnberger JI Jr, Gershon ES.

Affected-sib-pair analyses reveal support of prior evidence for a susceptibility locus for bipolar disorder, on 21q. *Am J Hum Genet*. 1996 Jun, 58(6):1279-85. PMID: 8651306

46. Detera-Wadleigh SD, Badner JA, Yoshikawa T, Sanders AR, Goldin LR, Turner G, Rollins DY, Moses T, Guroff JJ, Kazuba D, Maxwell ME, Edenberg HJ, Foroud T, Lahiri D, Nurnberger JI Jr, Stine OC, McMahon F, Meyers DA, MacKinnon D, Simpson S, McInnis M, DePaulo JR, Rice J, Goate A, Gershon ES, et al. Initial genome scan of the NIMH genetics initiative bipolar pedigrees: chromosomes 4, 7, 9, 18, 19, 20, and 21q. *Am J Med Genet*. 1997 May 31, 74(3):254-62. PMID: 9184307

47. Mendlewicz J, Simon P, Sevy S, Charon F, Brocas H, Legros S, Vassart G. Polymorphic DNA marker on X chromosome and manic depression. *Lancet*. 1987 May 30, 1(8544):1230-2. PMID: 2884369

48. Lucotte G, Landoulsi A, Berriche S, David F, Babron MC. Manic depressive illness is linked to factor IX in a French pedigree. *Ann Genet*. 1992, 35(2):93-5. PMID: 1524415

49. Ekholm JM, Pekkarinen P, Pajukanta P, Kieseppä T, Partonen T, Paunio T, Varilo T, Perola M, Lönnqvist J, Peltonen L. Bipolar disorder susceptibility region on Xq24-q27.1 in Finnish families. *Mol Psychiatry*. 2002, 7(5):453-9. PMID: 12082562

50. Pekkarinen P, Terwilliger J, Bredbacka PE, Lönnqvist J, Peltonen L. Evidence of a predisposing locus to bipolar disorder on Xq24-q27.1 in an extended Finnish pedigree. *Genome Res*. 1995 Sep, 5(2):105-15. PMID: 9132265

51. Knight HM, Pickard BS, Maclean A, Malloy MP, Soares DC, McRae AF, Condie A, White A, Hawkins W, McGhee K, van Beck M, MacIntyre DJ, Starr JM, Deary IJ, Visscher PM, Porteous DJ, Cannon RE, St Clair D, Muir WJ, Blackwood DH. A cytogenetic abnormality and rare coding variants identify ABCA13 as a candidate gene in schizophrenia, bipolar disorder, and depression. *Am J Hum Genet*. 2009 Dec, 85(6):833-46. PMID: 19944402

52. Kucukali CI, Aydin M, Ozkok E, Bilge E, Zengin A, Cakir U, Kara I. Angiotensin-converting enzyme polymorphism in schizophrenia, bipolar disorders, and their first-degree relatives. *Psychiatr Genet*. 2010 Feb, 20(1):14-9. PMID: 20010451

53. Arinami T, Li L, Mitsushio H, Itokawa M, Hamaguchi H, Toru M. An insertion/deletion polymorphism in the angiotensin converting enzyme gene is associated with both brain substance P contents and affective disorders. *Biol Psychiatry*. 1996 Dec 1, 40(11):1122-7. PMID: 8931914

54. Meira-Lima IV, Pereira AC, Mota GF, Krieger JE, Vallada H. Angiotensinogen and angiotensin converting enzyme gene polymorphisms and the risk of bipolar affective disorder in humans. *Neurosci Lett*. 2000 Oct 27, 293(2):103-6. PMID: 11027844

55. Karege F, Perroud N, Schürhoff F, Méary A, Marillier G, Burkhardt S, Ballmann E, Fernandez R, Jamain S, Leboyer M, La Harpe R, Malafosse A. Association of AKT1 gene variants and protein expression in both schizophrenia and bipolar disorder. *Genes Brain Behav*. 2010 Jul, 9(5):503-11. PMID: 2021468

56. Magno LA, Miranda DM, Neves FS, Pimenta GJ, Mello MP, De Marco LA, Correa H, Romano-Silva MA. Association between AKT1 but not AKTIP genetic variants

and increased risk for suicidal behavior in bipolar patients. *Genes Brain Behav*. 2010 Jun 1, 9(4):411-8. PMID: 20132317

57. Toyota T, Yamada K, Detera-Wadleigh SD, Yoshikawa T. Analysis of a cluster of polymorphisms in AKT1 gene in bipolar pedigrees: a family-based association study. *Neurosci Lett*. 2003 Mar 13, 339(1):5-8. PMID: 12618287

58. Schulze TG, Detera-Wadleigh SD, Akula N, Gupta A, Kassem L, Steele J, Pearl J, Strohmaier J, Breuer R, Schwarz M, Propping P, Nöthen MM, Cichon S, Schumacher J; NIMH Genetics Initiative Bipolar Disorder Consortium, Rietschel M, McMahon FJ. Two variants in Ankyrin 3 (ANK3) are independent genetic risk factors for bipolar disorder. *Mol Psychiatry*. 2009 May, 14(5):487-91. PMID: 19088739

59. Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, Fan J, Kirov G, Perlis RH, Green EK, Smoller JW, Grozeva D, Stone J, Nikolov I, Chambert K, Hamshere ML, Nimgaonkar VL, Moskvina V, Thase ME, Caesar S, Sachs GS, Franklin J, Gordon-Smith K, Ardlie KG, Gabriel SB, Fraser C, Blumenstiel B, Defelice M, Breen G, Gill M, Morris DW, Elkin A, Muir WJ, McGhee KA, Williamson R, MacIntyre DJ, MacLean AW, St CD, Robinson M, Van Beck M, Pereira AC, Kandaswamy R, McQuillin A, Collier DA, Bass NJ, Young AH, Lawrence J, Ferrier IN, Anjorin A, Farmer A, Curtis D, Scolnick EM, McGuffin P, Daly MJ, Corvin AP, Holmans PA, Blackwood DH, Gurling HM, Owen MJ, Purcell SM, Sklar P, Craddock N; Wellcome Trust Case Control Consortium. Collaborative genome-wide association analysis supports a role for

ANK3 and CACNA1C in bipolar disorder. *Nat Genet*. 2008 Sep, 40(9):1056-8. PMID: 18711365

60. Lee MT, Chen CH, Lee CS, Chen CC, Chong MY, Ouyang WC, Chiu NY, Chuo LJ, Chen CY, Tan HK, Lane HY, Chang TJ, Lin CH, Jou SH, Hou YM, Feng J, Lai TJ, Tung CL, Chen TJ, Chang CJ, Lung FW, Chen CK, Shiah IS, Liu CY, Teng PR, Chen KH, Shen LJ, Cheng CS, Chang TP, Li CF, Chou CH, Chen CY, Wang KH, Fann CS, Wu JY, Chen YT, Cheng AT. Genome-wide association study of bipolar I disorder in the Han Chinese population. *Mol Psychiatry*. 2010 Apr 13. PMID: 20386566

61. Smith EN, Bloss CS, Badner JA, Barrett T, Belmonte PL, Berrettini W, Byerley W, Coryell W, Craig D, Edenberg HJ, Eskin E, Foroud T, Gershon E, Greenwood TA, Hipolito M, Koller DL, Lawson WB, Liu C, Lohoff F, McInnis MG, McMahon FJ, Mirel DB, Murray SS, Nievergelt C, Nurnberger J, Nwulia EA, Paschall J, Potash JB, Rice J, Schulze TG, Scheftner W, Panganiban C, Zaitlen N, Zandi PP, Zöllner S, Schork NJ, Kelsoe JR. Genome-wide association study of bipolar disorder in European American and African American individuals. *Mol Psychiatry*. 2009 Aug, 14(8):755-63. PMID: 19488044

62. Skibinska M, Hauser J, Czerski PM, Leszczynska-Rodziewicz A, Kosmowska M, Kapelski P, Slopien A, Zakrzewska M, Rybakowski JK. Association analysis of brain-derived neurotrophic factor (BDNF) gene Val66Met polymorphism in schizophrenia and bipolar affective disorder. *World J Biol Psychiatry*. 2004 Oct, 5(4):215-20. PMID: 15543516

63. Fan J, Sklar P. Genetics of bipolar disorder: focus on BDNF Val66Met polymorphism. Novartis Found Symp. 2008, 289:60-72; discussion 72-3, 87-93. PMID: 18497095

64. Luykx JJ, Boks MP, Terwindt AP, Bakker S, Kahn RS, Ophoff RA. The involvement of GSK3beta in bipolar disorder: integrating evidence from multiple types of genetic studies. *Eur Neuropsychopharmacol*. 2010 Jun, 20(6):357-68. PMID: 20226637

65. Lee SY, Chen SL, Chen SH, Huang SY, Tzeng NS, Chang YH, Wang CL, Lee IH, Yeh TL, Yang YK, Lu RB. The COMT and DRD3 genes interacted in bipolar I but not bipolar II disorder. *World J Biol Psychiatry*. 2010 Aug 10. PMID: 20698735

66. Jahnes E, Müller DJ, Schulze TG, Windemuth C, Cichon S, Ohlraun S, Fangerau H, Held T, Maier W, Propping P, Nöthen MM, Rietschel M. Association study between two variants in the DOPA decarboxylase gene in bipolar and unipolar affective disorder. *Am J Med Genet*. 2002 Jul 8, 114(5):519-22. PMID: 12116187

67. Børglum AD, Bruun TG, Kjeldsen TE, Ewald H, Mors O, Kirov G, Russ C, Freeman B, Collier DA, Kruse TA. Two novel variants in the DOPA decarboxylase gene: association with bipolar affective disorder. *Mol Psychiatry*. 1999 Nov, 4(6):545-51. PMID: 10578236

68. Schosser A, Gaysina D, Cohen-Woods S, Chow PC, Martucci L, Craddock N, Farmer A, Korszun A, Gunasinghe C, Gray J, Jones L, Tozzi F, Perry J, Muglia P, Owen MJ, Craig IW, McGuffin P. Association of DISC1 and TSNAX genes and affective disorders in the depression case-control (DeCC) and bipolar affective

case-control (BACCS) studies. *Mol Psychiatry*. 2010 Aug, 15(8):844-9. PMID: 19255581

69. Ni X, Trakalo JM, Mundo E, Macciardi FM, Parikh S, Lee L, Kennedy JL. Linkage disequilibrium between dopamine D1 receptor gene (DRD1) and bipolar disorder. *Biol Psychiatry*. 2002 Dec 15, 52(12):1144-50. PMID: 1248805

70. Del Zompo M, De Luca V, Severino G, Ni X, Mulas S, Congiu D, Piccardi MP, Kennedy JL. Haplotype association study between DRD1 gene and bipolar type I affective disorder in two samples from Canada and Sardinia. *Am J Med Genet B Neuropsychiatr Genet*. 2007 Mar 5, 144B(2):237-41. PMID: 17066478

71. Severino G, Congiu D, Serreli C, De Lisa R, Chillotti C, Del Zompo M, Piccardi MP. A48G polymorphism in the D1 receptor genes associated with bipolar I disorder. *Am J Med Genet B Neuropsychiatr Genet*. 2005 Apr 5, 134B(1):37-8. PMID: 15704231

72. Dmitrzak-Weglarz M, et al. Dopamine receptor D1 gene -48A/G polymorphism is associated with bipolar illness but not with schizophrenia in a Polish population. *Neuropsychobiology*. 2006, 53(1):46-50. PMID: 16397404

73. Rybakowski JK, Dmitrzak-Weglarz M, Suwalska A, Leszczynska-Rodziewicz A, Hauser J. Dopamine D1 receptor gene polymorphism is associated with prophylactic lithium response in bipolar disorder. *Pharmacopsychiatry*. 2009 Jan, 42(1):20-2. PMID: 19153942

74. Pérez de Castro I, Santos J, Torres P, Visedo G, Saiz-Ruiz J, Llinares C, Fernández-Piqueras J. A weak association between TH and DRD2 genes and bipolar affective disorder in a Spanish sample. *J Med Genet*. 1995 Feb, 32(2):131-4. PMID: 7760324

75. Massat I, Souery D, Del-Favero J, Van Gestel S, Serretti A, Macciardi F, Smeraldi E, Kaneva R, Adolfsson R, Nylander PO, Blackwood D, Muir W, Papadimitriou GN, Dikeos D, Oruc L, Segman RH, Ivezic S, Aschauer H, Ackenheil M, Fuchshuber S, Dam H, Jakovljevic M, Peltonen L, Hilger C, Hentges F, Staner L, Milanova V, Jazin E, Lerer B, Van Broeckhoven C, Mendlewicz J. Positive association of dopamine D2 receptor polymorphism with bipolar affective disorder in a European Multicenter Association Study of affective disorders. *Am J Med Genet*. 2002 Mar 8, 114(2):177-85. PMID: 11857579

76. Li T, Liu X, Sham PC, Aitchison KJ, Cai G, Arranz MJ, Deng H, Liu J, Kirov G, Murray RM, Collier DA. Association analysis between dopamine receptor genes and bipolar affective disorder. *Psychiatry Res*. 1999 Jun 30, 86(3):193-201. PMID: 10482338

77. Parsian A, Chakraverty S, Todd RD. Possible association between the dopamine D3 receptor gene and bipolar affective disorder. *Am J Med Genet*. 1995 Jun 19, 60(3):234-7. PMID: 7573178

78. López León S, Croes EA, Sayed-Tabatabaei FA, Claes S, Van Broeckhoven C, van Duijn CM. The dopamine D4 receptor gene 48-base-pair-repeat polymorphism and mood disorders: a meta-analysis. *Biol Psychiatry*. 2005 May 1, 57(9):999-1003. PMID: 15860340

79. Muglia P, Petronis A, Mundo E, Lander S, Cate T, Kennedy JL. Dopamine D4 receptor and tyrosine hydroxylase genes in bipolar disorder: evidence for a role of DRD4. *Mol Psychiatry*. 2002, 7(8):860-6. PMID: 1223277

80. Manki H, Kanba S, Muramatsu T, Higuchi S, Suzuki E, Matsushita S, Ono Y, Chiba H, Shintani F, Nakamura M, Yagi G, Asai M. Dopamine D2, D3 and D4 receptor and transporter gene polymorphisms and mood disorders. *J Affect Disord*. 1996 Sep 9, 40(1-2):7-13. PMID: 8882909

81. Schulze TG, Ohlraun S, Czerski PM, Schumacher J, Kassem L, Deschner M, Gross M, Tullius M, Heidmann V, Kovalenko S, Jamra RA, Becker T, Leszczynska-Rodziewicz A, Hauser J, Illig T, Klopp N, Wellek S, Cichon S, Henn FA, McMahon FJ, Maier W, Propping P, Nöthen MM, Rietschel M. Genotype-phenotype studies in bipolar disorder showing association between the DAOA/G30 locus and persecutory delusions: a first step toward a molecular genetic classification of psychiatric phenotypes. *Am J Psychiatry*. 2005 Nov, 162(11):2101-8. PMID: 16263850

82. Bass NJ, Datta SR, McQuillin A, Puri V, Choudhury K, Thirumalai S, Lawrence J, Quested D, Pimm J, Curtis D, Gurling HM. Evidence for the association of the DAOA (G72) gene with schizophrenia and bipolar disorder but not for the association of the DAO gene with schizophrenia. *Behav Brain Funct*. 2009 Jul 8, 5:28. PMID: 19586533

83. Horiuchi Y, Nakayama J, Ishiguro H, Ohtsuki T, Detera-Wadleigh SD, Toyota T, Yamada K, Nankai M, Shibuya H, Yoshikawa T, Arinami T. Possible association between a haplotype of the GABA-A receptor alpha 1 subunit gene (GABRA1) and mood disorders. *Biol Psychiatry*. 2004 Jan 1, 55(1):40-5. PMID: 14706423

84. Massat I, Souery D, Del-Favero J, Oruc L, Noethen MM, Blackwood D, Thomson M, Muir W, Papadimitriou GN, Dikeos DG, Kaneva R, Serretti A, Lilli R, Smeraldi

E, Jakovljevic M, Folnegovic V, Rietschel M, Milanova V, Valente F, Van Broeckhoven C, Mendlewicz J. Excess of allele1 for alpha3 subunit GABA receptor gene (GABRA3) in bipolar patients: a multicentric association study. *Mol Psychiatry*. 2002, 7(2):201-7. PMID: 11840313

85. Otani K, Ujike H, Tanaka Y, Morita Y, Katsu T, Nomura A, Uchida N, Hamamura T, Fujiwara Y, Kuroda S. The GABA type A receptor alpha5 subunit gene is associated with bipolar I disorder. *Neurosci Lett*. 2005 Jun 10-17, 381(1-2):108-13. PMID: 15882799

86. Papadimitriou GN, Dikeos DG, Karadima G, Avramopoulos D, Daskalopoulou EG, Vassilopoulos D, Stefanis CN. Association between the GABA(A) receptor alpha5 subunit gene locus (GABRA5) and bipolar affecttive disorder. *Am J Med Genet*. 1998 Feb 7, 81(1):73-80. PMID: 9514592

87. Oruc L, Verheyen GR, Furac I, Ivezić S, Jakovljević M, Raeymaekers P, Van Broeckhoven C. Positive association between the GABRA5 gene and unipolar recurrent major depression. *Neuropsychobiology*. 1997, 36(2):62-4. PMID: 9267853

88. Itokawa M, Yamada K, Iwayama-Shigeno Y, Ishitsuka Y, Detera-Wadleigh S, Yoshikawa T. Genetic analysis of a functional GRIN2A promoter (GT)n repeat in bipolar disorder pedigrees in humans. *Neurosci Lett*. 2003 Jul 10, 345(1):53-6. PMID: 12809987

89. Cherlyn SY, Woon PS, Liu JJ, Ong WY, Tsai GC, Sim K. Genetic association studies of glutamate, GABA and related genes in schizophrenia and bipolar

disorder: a decade of advance. *Neurosci Biobehav Rev*. 2010 May, 34(6):958-77. PMID: 20060416

90. Martucci L, Wong AH, De Luca V, Likhodi O, Wong GW, King N, Kennedy JL. N-methyl-D-aspartate receptor NR2B subunit gene GRIN2B in schizophrenia and bipolar disorder: Polymorphisms and mRNA levels. *Schizophr Res*. 2006 Jun, 84(2-3):214-21. PMID: 16549338
91. Whalley HC, Pickard BS, McIntosh AM, Zuliani R, Johnstone EC, Blackwood DH, Lawrie SM, Muir WJ, Hall J. A GRIK4 variant conferring protection against bipolar disorder modulates hippocampal function. *Mol Psychiatry*. 2009 May, 14(5):467-8. PMID: 19384319
92. Pickard BS, Knight HM, Hamilton RS, Soares DC, Walker R, Boyd JK, Machell J, Maclean A, McGhee KA, Condie A, Porteous DJ, St Clair D, Davis I, Blackwood DH, Muir WJ. A common variant in the 3'UTR of the GRIK4 glutamate receptor gene affects transcript abundance and protects against bipolar disorder. *Proc Natl Acad Sci USA*. 2008 Sep 30, 105(39):14940-5. PMID: 18824690
93. Pickard BS, Malloy MP, Christoforou A, Thomson PA, Evans KL, Morris SW, Hampson M, Porteous DJ, Blackwood DH, Muir WJ. Cytogenetic and genetic evidence supports a role for the kainate-type glutamate receptor gene, GRIK4, in schizophrenia and bipolar disorder. *Mol Psychiatry*. 2006 Sep, 11(9):847-57. PMID: 16819533
94. Blackwood DH, Pickard BJ, Thomson PA, Evans KL, Porteous DJ, Muir WJ. Are some genetic risk factors common to schizophrenia, bipolar disorder and

depression? Evidence from DISC1, GRIK4 and NRG1. *Neurotox Res*. 2007 Jan, 11(1):73-83. PMID: 17449450

95. Barrett TB, Hauger RL, Kennedy JL, Sadovnick AD, Remick RA, Keck PE, McElroy SL, Alexander M, Shaw SH, Kelsoe JR. Evidence that a single nucleotide polymorphism in the promoter of the G protein receptor kinase 3 gene is associated with bipolar disorder. *Mol Psychiatry*. 2003 May, 8(5):546-57. PMID: 12808434

96. Barrett TB, Emberton JE, Nievergelt CM, Liang SG, Hauger RL, Eskin E, Schork NJ, Kelsoe JR. Further evidence for association of GRK3 to bipolar disorder suggests a second disease mutation. *Psychiatr Genet*. 2007 Dec, 17(6):315-22. PMID: 18075471

97. Benedetti F, Bernasconi A, Lorenzi C, Pontiggia A, Serretti A, Colombo C, Smeraldi E. A single nucleotide polymorphism in glycogen synthase kinase 3-beta promoter gene influences onset of illness in patients affected by bipolar disorder. *Neurosci Lett*. 2004 Jan 23, 355(1-2):37-40. PMID: 14729229

98. Benedetti F, Serretti A, Colombo C, Lorenzi C, Tubazio V, Smeraldi E. A glycogen synthase kinase 3-beta promoter gene single nucleotide polymorphism is associated with age at onset and response to total sleep deprivation in bipolar depression. *Neurosci Lett*. 2004 Sep 23, 368(2):123-6. PMID: 15351432

99. Benedetti F, Serretti A, Pontiggia A, Bernasconi A, Lorenzi C, Colombo C, Smeraldi E. Long-term response to lithium salts in bipolar illness is influenced by the glycogen synthase kinase 3-beta -50 T/C SNP. *Neurosci Lett*. 2005 Mar 7, 376(1):51-5. PMID: 15694273

100. McAuley EZ, Fullerton JM, Blair IP, Donald JA, Mitchell PB, Schofield PR. Association between the serotonin 2A receptor gene and bipolar affective disorder in an Australian cohort. *Psychiatr Genet*. 2009 Oct, 19(5):244-52. PMID: 19584773

101. Ranade SS, Mansour H, Wood J, Chowdari KV, Brar LK, Kupfer DJ, Nimgaonkar VL. Linkage and association between serotonin 2A receptor gene polymorphisms and bipolar I disorder. *Am J Med Genet B Neuropsychiatr Genet*. 2003 Aug 15, 121B(1):28-34. PMID: 12898571

102. Manchia M, Zai CC, Squassina A, Vincent JB, De Luca V, Kennedy JL. Mixture regression analysis on age at onset in bipolar disorder patients: investigation of the role of serotonergic genes. *Eur Neuropsychopharmacol*. 2010 Sep, 20(9):663-70. PMID: 20452754

103. Massat I, Lerer B, Souery D, Blackwood D, Muir W, Kaneva R, Nöthen MM, Oruc L, Papadimitriou GN, Dikeos D, Serretti A, Bellivier F, Golmard JL, Milanova V, Del-Favero J, Van Broeckhoven C, Mendlewicz J. HTR2C (cys23ser) polymorphism influences early onset in bipolar patients in a large European multicenter association study. *Mol Psychiatry*. 2007 Sep, 12(9):797-8. PMID: 1776714

104. Lerer B, Macciardi F, Segman RH, Adolfsson R, Blackwood D, Blairy S, Del Favero J, Dikeos DG, Kaneva R, Lilli R, Massat I, Milanova V, Muir W, Noethen M, Oruc L, Petrova T, Papadimitriou GN, Rietschel M, Serretti A, Souery D, Van Gestel S, Van Broeckhoven C, Mendlewicz J. Variability of 5-HT2C receptor cys23ser polymorphism among European populations and vulnerability to affective disorder. *Mol Psychiatry*. 2001 Sep, 6(5):579-85. PMID: 11526472

105. Ohtsuki T, Ishiguro H, Detera-Wadleigh SD, Toyota T, Shimizu H, Yamada K, Yoshitsugu K, Hattori E, Yoshikawa T, Arinami T. Association between serotonin 4 receptor gene polymorphisms and bipolar disorder in Japanese case-control samples and the NIMH Genetics Initiative Bipolar Pedigrees. *Mol Psychiatry*. 2002, 7(9):954-61. PMID: 12399948

106. Vogt IR, Shimron-Abarbanell D, Neidt H, Erdmann J, Cichon S, Schulze TG, Müller DJ, Maier W, Albus M, Borrmann-Hassenbach M, Knapp M, Rietschel M, Propping P, Nöthen MM. Investigation of the human serotonin 6 [5-HT6] receptor gene in bipolar affective disorder and schizophrenia. *Am J Med Genet*. 2000 Apr 3, 96(2):217-21. PMID: 10893499

107. Bloch PJ, Weller AE, Doyle GA, Ferraro TN, Berrettini WH, Hodge R, Lohoff FW. Association analysis between polymorphisms in the myo-inositol monophosphatase 2 (IMPA2) gene and bipolar disorder. *Prog Neuropsychopharmacol Biol Psychiatry*. 2010 Aug 25. PMID: 20800640

108. Ohnishi T, Yamada K, Ohba H, Iwayama Y, Toyota T, Hattori E, Inada T, Kunugi H, Tatsumi M, Ozaki N, Iwata N, Sakamoto K, Iijima Y, Iwata Y, Tsuchiya KJ, Sugihara G, Nanko S, Osumi N, Detera-Wadleigh SD, Kato T, Yoshikawa T. A promoter haplotype of the inositol monophosphatase 2 gene (IMPA2) at 18p11.2 confers a possible risk for bipolar disorder by enhancing transcription. *Neuropsychopharmacology*. 2007 Aug, 32(8):1727-37. PMID: 17251911

109. Dimitrova A, Milanova V, Krastev S, Nikolov I, Toncheva D, Owen MJ, Kirov G. Association study of myo-inositol monophosphatase 2 (IMPA2) polymorphisms

with bipolar affective disorder and response to lithium treatment. *Pharmacogenomics J*. 2005, 5(1):35-41. PMID: 15505643

110. Sjøholt G, Ebstein RP, Lie RT, Berle JØ, Mallet J, Deleuze JF, Levinson DF, Laurent C, Mujahed M, Bannoura I, Murad I, Molven A, Steen VM. Examination of IMPA1 and IMPA2 genes in manic-depressive patients: association between IMPA2 promoter polymorphisms and bipolar disorder. *Mol Psychiatry*. 2004 Jun, 9(6):621-9. PMID: 14699425
111. Yoon IS, Li PP, Siu KP, Kennedy JL, Cooke RG, Parikh SV, Warsh JJ. Altered IMPA2 gene expression and calcium homeostasis in bipolar disorder. *Mol Psychiatry*. 2001 Nov, 6(6):678-83. PMID: 1167379
112. Baum AE, Hamshere M, Green E, Cichon S, Rietschel M, Noethen MM, Craddock N, McMahon FJ. Meta-analysis of two genome-wide association studies of bipolar disorder reveals important points of agreement. *Mol Psychiatry*. 2008 May, 13(5):466-7. PMID: 18421293
113. Furlong RA, Ho L, Rubinsztein JS, Walsh C, Paykel ES, Rubinsztein DC. Analysis of the monoamine oxidase A (MAOA) gene in bipolar affective disorder by association studies, meta-analyses, and sequencing of the promoter. *Am J Med Genet*. 1999 Aug 20, 88(4):398-406. PMID: 10402508
114. Fan M, Liu B, Jiang T, Jiang X, Zhao H, Zhang J. Meta-analysis of the association between the monoamine oxidase-A gene and mood disorders. *Psychiatr Genet*. 2010 Feb, 20(1):1-7. PMID: 20010318

115. Lin YM, Davamani F, Yang WC, Lai TJ, Sun HS. Association analysis of monoamine oxidase A gene and bipolar affective disorder in Han Chinese. *Behav Brain Funct*. 2008 May 24, 4:21. PMID: 18501009

116. Müller DJ, Serretti A, Sicard T, Tharmalingam S, King N, Artioli P, Mandelli L, Lorenzi C, Kennedy JL. Further evidence of MAO-A gene variants associated with bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet*. 2007 Jan 5, 144B(1):37-40. PMID: 16958037

117. Preisig M, Ferrero F, Malafosse A. Monoamine oxidase a and tryptophan hydroxylase gene polymorphisms: are they associated with bipolar disorder? *Am J Pharmacogenomics*. 2005, 5(1):45-52. PMID: 15727488

118. Gutiérrez B, Arias B, Gastó C, Catalán R, Papiol S, Pintor L, Fañanás L. Association analysis between a functional polymorphism in the monoamine oxidase A gene promoter and severe mood disorders. *Psychiatr Genet*. 2004 Dec, 14(4):203-8. PMID: 15564894

119. Wakabayashi Y, Uchida S, Funato H, Matsubara T, Watanuki T, Otsuki K, Fujimoto M, Nishida A, Watanabe Y. State-dependent changes in the expression levels of NCAM-140 and L1 in the peripheral blood cells of bipolar disorders, but not in the major depressive disorders. *Prog Neuropsychopharmacol Biol Psychiatry*. 2008 Jul 1, 32(5):1199-205. PMID: 18430502

120. Atz ME, Rollins B, Vawter MP. NCAM1 association study of bipolar disorder and schizophrenia: polymorphisms and alternatively spliced isoforms lead to similarities and differences. *Psychiatr Genet*. 2007 Apr, 17(2):55-67. PMID: 17413444

121. Arai M, Itokawa M, Yamada K, Toyota T, Arai M, Haga S, Ujike H, Sora I, Ikeda K, Yoshikawa T. Association of neural cell adhesion molecule 1 gene polymorphisms with bipolar affective disorder in Japanese individuals. *Biol Psychiatry*. 2004 Apr 15, 55(8):804-10. PMID: 15050861

122. Vawter MP, Howard AL, Hyde TM, Kleinman JE, Freed WJ. Alterations of hippocampal secreted N-CAM in bipolar disorder and synaptophysin in schizophrenia. *Mol Psychiatry*. 1999 Sep, 4(5):467-75. PMID: 10523820

123. Vawter MP, Hemperly JJ, Hyde TM, Bachus SE, VanderPutten DM, Howard AL, Cannon-Spoor HE, McCoy MT, Webster MJ, Kleinman JE, Freed WJ. VASE-containing N-CAM isoforms are increased in the hippocampus in bipolar disorder but not schizophrenia. *Exp Neurol*. 1998 Nov, 154(1):1-11. PMID: 9875262

124. Zhang J, Li X, Wang Y, Ji J, Yang F, Feng G, Wan P, Lindpaintner K, He L, He G. Association study on the mitochondrial gene NDUFV2 and bipolar disorder in the Chinese Han population. *J Neural Transm*. 2009 Mar, 116(3):357-61. PMID: 19194776

125. Xu C, Li PP, Kennedy JL, Green M, Hughes B, Cooke RG, Parikh SV, Warsh JJ. Further support for association of the mitochondrial complex I subunit gene NDUFV2 with bipolar disorder. *Bipolar Disord*. 2008 Feb, 10(1):105-10. PMID: 18199248

126. Washizuka S, Iwamoto K, Kakiuchi C, Bundo M, Kato T. Expression of mitochondrial complex I subunit gene NDUFV2 in the lymphoblastoid cells derived from patients with bipolar disorder and schizophrenia. *Neurosci Res*. 2009 Mar, 63(3):199-204. PMID: 19135101

127. Ben-Shachar D, Karry R. Neuroanatomical pattern of mitochondrial complex I pathology varies between schizophrenia, bipolar disorder and major depression. *PLoS One*. 2008, 3(11):e3676. PMID: 18989376

128. Washizuka S, Iwamoto K, Kazuno AA, Kakiuchi C, Mori K, Kametani M, Yamada K, Kunugi H, Tajima O, Akiyama T, Nanko S, Yoshikawa T, Kato T. Association of mitochondrial complex I subunit gene NDUFV2 at 18p11 with bipolar disorder in Japanese and the National Institute of Mental Health pedigrees. *Biol Psychiatry*. 2004 Oct 1, 56(7):483-9. PMID: 15450783

129. Washizuka S, Kametani M, Sasaki T, Tochigi M, Umekage T, Kohda K, Kato T. Association of mitochondrial complex I subunit gene NDUFV2 at 18p11 with bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet*. 2003 Jul 1, 120B(1):72-8. PMID: 12815743

130. Nakatani N, Hattori E, Ohnishi T, Dean B, Iwayama Y, Matsumoto I, Kato T, Osumi N, Higuchi T, Niwa S, Yoshikawa T. Genome-wide expression analysis detects eight genes with robust alterations specific to bipolar I disorder: relevance to neuronal network perturbation. *Hum Mol Genet*. 2006 Jun 15, 15(12):1949-62. PMID: 16687443

131. Pickard BS, Christoforou A, Thomson PA, Fawkes A, Evans KL, Morris SW, Porteous DJ, Blackwood DH, Muir WJ. Interacting haplotypes at the NPAS3 locus alter risk of schizophrenia and bipolar disorder. *Mol Psychiatry*. 2009 Sep, 14(9):874-84. PMID: 18317462

132. Huang J, Perlis RH, Lee PH, Rush AJ, Fava M, Sachs GS, Lieberman J, Hamilton SP, Sullivan P, Sklar P, Purcell S, Smoller JW. Cross-Disorder Genomewide

Analysis of Schizophrenia, Bipolar Disorder, and Depression. *Am J Psychiatry*. 2010 Aug 16. PMID: 20713499

133. Prata DP, Breen G, Osborne S, Munro J, St Clair D, Collier DA. An association study of the neuregulin 1 gene, bipolar affective disorder and psychosis. *Psychiatr Genet*. 2009 Jun, 19(3):113-6. PMID: 19339916

134. Thomson PA, Christoforou A, Morris SW, Adie E, Pickard BS, Porteous DJ, Muir WJ, Blackwood DH, Evans KL. Association of Neuregulin 1 with schizophrenia and bipolar disorder in a second cohort from the Scottish population. *Mol Psychiatry*. 2007 Jan, 12(1):94-104. PMID: 16940976

135. Dawson E, Gill M, Curtis D, Castle D, Hunt N, Murray R, Powell J. Genetic association between alleles of pancreatic phospholipase A2 gene and bipolar affective disorder. *Psychiatr Genet*. 1995 Winter, 5(4):177-80. PMID: 8750360

136. Jacobsen N, Daniels J, Moorhead S, Harrison D, Feldman E, McGuffin P, Owen MJ, Craddock N. Association study of bipolar disorder at the phospholipase A2 gene (PLA2A) in the Darier's disease (DAR) region of chromosome 12q23-q24.1. *Psychiatr Genet*. 1996 Winter, 6(4):195-9. PMID: 9149325

137. Jacobsen NJ, Franks EK, Owen MJ, Craddock NJ. Mutational analysis of phospholipase A2A: a positional candidate susceptibility gene for bipolar disorder. *Mol Psychiatry*. 1999 May, 4(3):274-9. PMID: 10395219

138. Turecki G, Grof P, Cavazzoni P, Duffy A, Grof E, Ahrens B, Berghöfer A, Müller-Oerlinghausen B, Dvoráková M, Libigerová E, Vojtechovský M, Zvolský P, Joober R, Nilsson A, Prochazka H, Licht RW, Rasmussen NA, Schou M, Vestergaard P, Holzinger A, Schumann C, Thau K, Rouleau GA, Alda M. Evidence for a role of

phospholipase C-gamma1 in the pathogenesis of bipolar disorder. *Mol Psychiatry*. 1998 Nov, 3(6):534-8. PMID: 9857980

139. Løvlie R, Berle JO, Stordal E, Steen VM. The phospholipase C-gamma1 gene (PLCG1) and lithium-responsive bipolar disorder: re-examination of an intronic dinucleotide repeat polymorphism. *Psychiatr Genet*. 2001 Mar, 11(1):41-3. PMID: 11409699
140. Zandi PP, Belmonte PL, Willour VL, Goes FS, Badner JA, Simpson SG, Gershon ES, McMahon FJ, DePaulo JR Jr, Potash JB; Bipolar Disorder Phenome Group; National Institute of Mental Health Genetics Initiative Bipolar Disorder Consortium. Association study of Wnt signaling pathway genes in bipolar disorder. *Arch Gen Psychiatry*. 2008 Jul, 65(7):785-93. PMID: 18606951
141. Mick E, Kim JW, Biederman J, Wozniak J, Wilens T, Spencer T, Smoller JW, Faraone SV. Family based association study of pediatric bipolar disorder and the dopamine transporter gene (SLC6A3). *Am J Med Genet B Neuropsychiatr Genet*. 2008 Oct 5, 147B(7):1182-5. PMID: 18361424
142. Ohadi M, Keikhaee MR, Javanbakht A, Sargolzaee MR, Robabeh M, Najmabadi H. Gender dimorphism in the DAT1 -67 T-allele homozygosity and predisposition to bipolar disorder. *Brain Res*. 2007 May 4, 1144:142-5. PMID: 17316573
143. Ewald H, Flint T, Degn B, Mors O, Kruse TA. A functional variant of the serotonin transporter gene in families with bipolar affective disorder. *J Affect Disord*. 1998 Mar;48(2-3):135-44. PMID: 9543202
144. Nyegaard M, Børglum AD, Bruun TG, Collier DA, Russ C, Mors O, Ewald H, Kruse TA. Novel polymorphisms in the somatostatin receptor 5 (SSTR5) gene

associated with bipolar affective disorder. *Mol Psychiatry*. 2002, 7(7):745-54. PMID: 12192619

145. Saus E, Brunet A, Armengol L, Alonso P, Crespo JM, Fernández-Aranda F, Guitart M, Martín-Santos R, Menchón JM, Navinés R, Soria V, Torrens M, Urretavizcaya M, Vallès V, Gratacòs M, Estivill X. Comprehensive copy number variant (CNV) analysis of neuronal pathways genes in psychiatric disorders identifies rare variants within patients. *J Psychiatr Res*. 2010 Apr 14. PMID: 20398908
146. Meloni R, Biguet NF, Mallet J. Post-genomic era and gene discovery for psychiatric diseases: there is a new art of the trade? The example of the HUMTH01 microsatellite in the Tyrosine Hydroxylase gene. *Mol Neurobiol*. 2002 Oct-Dec, 26(2-3):389-403. PMID: 12428766
147. Serretti A, Macciardi F, Verga M, Cusin C, Pedrini S, Smeraldi E. Tyrosine hydroxylase gene associated with depressive symptomatology in mood disorder. *Am J Med Genet*. 1998 Mar 28, 81(2):127-30. PMID: 9613850
148. Serretti A, Macciardi F, Cusin C, Verga M, Pedrini S, Smeraldi E. Tyrosine hydroxylase gene in linkage disequilibrium with mood disorders. *Mol Psychiatry*. 1998 Mar, 3(2):169-74. PMID: 9577842
149. Lobos EA, Todd RD. Cladistic analysis of disease association with tyrosine hydroxylase: application to manic-depressive disease and alcoholism. *Am J Med Genet*. 1997 May 31, 74(3):289-95. PMID: 9184313
150. Meloni R, Leboyer M, Bellivier F, Barbe B, Samolyk D, Allilaire JF, Mallet J. Association of manic-depressive illness with tyrosine hydroxylase microsatellite marker. *Lancet*. 1995 Apr 8, 345(8954):932. PMID: 7707843

151. Barbosa IG, Huguet RB, Mendonça VA, Sousa LP, Neves FS, Bauer ME, Teixeira AL. Increased plasma levels of soluble TNF receptor I in patients with bipolar disorder. *Eur Arch Psychiatry Clin Neurosci*. 2010 May 6. PMID: 20446090

152. Kauer-Sant'Anna M, Kapczinski F, Andreazza AC, Bond DJ, Lam RW, Young LT, Yatham LN. Brain-derived neurotrophic factor and inflammatory markers in patients with early- vs. late-stage bipolar disorder. *Int J Neuropsychopharmacol*. 2009 May, 12(4):447-58. PMID: 18771602

153. Czerski PM, Rybakowski F, Kapelski P, Rybakowski JK, Dmitrzak-Weglarz M, Leszczyńska-Rodziewicz A, Słopień A, Skibińska M, Kaczmarkiewicz-Fass M, Hauser J. Association of tumor necrosis factor -308G/A promoter polymorphism with schizophrenia and bipolar affective disorder in a Polish population. *Neuropsychobiology*. 2008, 57(1-2):88-94. PMID: 18515978

154. Efimov RV, Mavliudova ASh, Golimbet VE. [Association between the tryptophan hydroxylase (TPH) gene polymorphic markers and endogenous psychoses.] *Genetika*. 2009 Dec, 45(12):1668-73. Russian. PMID: 20198979

155. Chen C, Glatt SJ, Tsuang MT. The tryptophan hydroxylase gene influences risk for bipolar disorder but not major depressive disorder: results of meta-analyses. *Bipolar Disord*. 2008 Nov, 10(7):816-21. PMID: 19032713

156. Bellivier F, Leboyer M, Courtet P, Buresi C, Beaufils B, Samolyk D, Allilaire JF, Feingold J, Mallet J, Malafosse A. Association between the tryptophan hydroxylase gene and manic-depressive illness. *Arch Gen Psychiatry*. 1998 Jan, 55(1):33-7. PMID: 9435758

157. Lin YM, Chao SC, Chen TM, Lai TJ, Chen JS, Sun HS. Association of functional polymorphisms of the human tryptophan hydroxylase 2 gene with risk for bipolar disorder in Han Chinese. *Arch Gen Psychiatry*. 2007 Sep, 64(9):1015-24. PMID: 17768266

158. Van Den Bogaert A, Sleegers K, De Zutter S, Heyrman L, Norrback KF, Adolfsson R, Van Broeckhoven C, Del-Favero J. Association of brain-specific tryptophan hydroxylase, TPH2, with unipolar and bipolar disorder in a Northern Swedish, isolated population. *Arch Gen Psychiatry*. 2006 Oct, 63(10):1103-10. PMID: 17015812

159. Harvey M, Gagné B, Labbé M, Barden N. Polymorphisms in the neuronal isoform of tryptophan hydroxylase 2 are associated with bipolar disorder in French Canadian pedigrees. *Psychiatr Genet*. 2007 Feb, 17(1):17-22. PMID: 17167340

160. Kakiuchi C, Ishigaki S, Oslowski CM, Fonseca SG, Kato T, Urano F. Valproate, a mood stabilizer, induces WFS1 expression and modulates its interaction with ER stress protein GRP94. *PLoS One*. 2009, 4(1):e4134. PMID: 19125190

161. Koido K, Kõks S, Nikopensius T, Maron E, Altmäe S, Heinaste E, Vabrit K, Tammekivi V, Hallast P, Kurg A, Shlik J, Vasar V, Metspalu A, Vasar E. Polymorphisms in wolframin (WFS1) gene are possibly related to increased risk for mood disorders. *Int J Neuropsychopharmacol*. 2005 Jun, 8(2):235-44. PMID: 15473915

162. Kawamoto T, Horikawa Y, Tanaka T, Kabe N, Takeda J, Mikuni M. Genetic variations in the WFS1 gene in Japanese with type 2 diabetes and bipolar disorder. *Mol Genet Metab*. 2004 Jul, 82(3):238-45. PMID: 15234338

163. Kim B, Kim CY, Lee MJ, Joo YH. Preliminary evidence on the association between XBP1-116C/G polymorphism and response to prophylactic treatment with valproate in bipolar disorders. *Psychiatry Res*. 2009 Aug 15, 168(3):209-12. PMID: 19564049

164. Masui T, Hashimoto R, Kusumi I, Suzuki K, Tanaka T, Nakagawa S, Kunugi H, Koyama T. A possible association between the -116C/G single nucleotide polymorphism of the XBP1 gene and lithium prophylaxis in bipolar disorder. *Int J Neuropsychopharmacol*. 2006 Feb, 9(1):83-8. PMID: 15927087

165. Kakiuchi C, et al. Impaired feedback regulation of XBP1 as a genetic risk factor for bipolar disorder. *Nat Genet*. 2003 Oct, 35(2):171-5. PMID: 12949534

166. Grover D, Verma R, Goes FS, Mahon PL, Gershon ES, McMahon FJ, Potash JB; NIMH Genetics Initiative Bipolar Disorder Collaborative, Bipolar Disorder Phenome Group, Gershon ES, McMahon FJ, Potash JB. Family-based association of YWHAH in psychotic bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet*. 2009 Oct 5, 150B(7):977-83. PMID: 19160447

167. Frederick K. Goodwin, Kay R. Jamison. *Manic-depressive Illness: bipolar disorders and recurrent depression*, 2nd edition. Oxford University Press, 2007.

168. Ryan MM, Lockstone HE, Huffaker SJ, Wayland MT, Webster MJ, Bahn S. Gene expression analysis of bipolar disorder reveals downregulation of the ubiquitin cycle and alterations in synaptic genes. *Mol Psychiatry*. 2006 Oct, 11(10):965-78. PMID: 16894394

169. Matigian N, Windus L, Smith H, Filippich C, Pantelis C, McGrath J, Mowry B, Hayward N. Expression profiling in monozygotic twins discordant for bipolar

disorder reveals dysregulation of the WNT signalling pathway. *Mol Psychiatry*. 2007 Sep, 12(9):815-25. PMID: 17440432

170. Ross J, Berrettini W, Coryell W, Gershon ES, Badner JA, Kelsoe JR, McInnis MG, McMahon FJ, Murphy DL, Nurnberger JI Jr, Foroud T, Rice JP, Scheftner WB, Zandi P, Edenberg H, Byerley W. Genome-wide parametric linkage analyses of 644 bipolar pedigrees suggest susceptibility loci at chromosomes 16 and 20. *Psychiatr Genet*. 2008 Aug, 18(4):191-8. PMID: 18628681

171. Scott LJ, Muglia P, Kong XQ, Guan W, Flickinger M, Upmanyu R, Tozzi F, Li JZ, Burmeister M, Absher D, Thompson RC, Francks C, Meng F, Antoniades A, Southwick AM, Schatzberg AF, Bunney WE, Barchas JD, Jones EG, Day R, Matthews K, McGuffin P, Strauss JS, Kennedy JL, Middleton L, Roses AD, Watson SJ, Vincent JB, Myers RM, Farmer AE, Akil H, Burns DK, Boehnke M. Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. *Proc Natl Acad Sci USA*. 2009 May 5, 106(18):7501-6. PMID: 19416921

172. Hattori E, Toyota T, Ishitsuka Y, Iwayama Y, Yamada K, Ujike H, Morita Y, Kodama M, Nakata K, Minabe Y, Nakamura K, Iwata Y, Takei N, Mori N, Naitoh H, Yamanouchi Y, Iwata N, Ozaki N, Kato T, Nishikawa T, Kashiwa A, Suzuki M, Shioe K, Shinohara M, Hirano M, Nanko S, Akahane A, Ueno M, Kaneko N, Watanabe Y, Someya T, Hashimoto K, Iyo M, Itokawa M, Arai M, Nankai M, Inada T, Yoshida S, Kunugi H, Nakamura M, Iijima Y, Okazaki Y, Higuchi T, Yoshikawa T. Preliminary genome-wide association study of bipolar disorder in the

Japanese population. *Am J Med Genet B Neuropsychiatr Genet*. 2009 Dec 5, 150B(8):1110-7. PMID: 19259986

173. Sklar P, Smoller JW, Fan J, Ferreira MA, Perlis RH, Chambert K, Nimgaonkar VL, McQueen MB, Faraone SV, Kirby A, de Bakker PI, Ogdie MN, Thase ME, Sachs GS, Todd-Brown K, Gabriel SB, Sougnez C, Gates C, Blumenstiel B, Defelice M, Ardlie KG, Franklin J, Muir WJ, McGhee KA, MacIntyre DJ, McLean A, VanBeck M, McQuillin A, Bass NJ, Robinson M, Lawrence J, Anjorin A, Curtis D, Scolnick EM, Daly MJ, Blackwood DH, Gurling HM, Purcell SM. Whole-genome association study of bipolar disorder. *Mol Psychiatry*. 2008 Jun, 13(6):558-69. PMID: 18317468

174. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007 Jun 7, 447(7145):661-78. PMID: 17554300

175. Baum AE, Akula N, Cabanero M, Cardona I, Corona W, Klemens B, Schulze TG, Cichon S, Rietschel M, Nöthen MM, Georgi A, Schumacher J, Schwarz M, Abou Jamra R, Höfels S, Propping P, Satagopan J, Detera-Wadleigh SD, Hardy J, McMahon FJ. A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Mol Psychiatry*. 2008 Feb, 13(2):197-207. PMID: 17486107

176. Cichon S, Mühleisen TW, Degenhardt FA, Mattheisen M, Miró X, Strohmaier J, Steffens M, Meesters C, Herms S, Weingarten M, Priebe L, Haenisch B, Alexander M, Vollmer J, Breuer R, Schmäl C, Tessmann P, Moebus S, Wichmann HE, Schreiber S, Müller-Myhsok B, Lucae S, Jamain S, Leboyer M, Bellivier F, Etain

B, Henry C, Kahn JP, Heath S; Bipolar Disorder Genome Study (BiGS) Consortium, Hamshere M, O'Donovan MC, Owen MJ, Craddock N, Schwarz M, Vedder H, Kammerer-Ciernioch J, Reif A, Sasse J, Bauer M, Hautzinger M, Wright A, Mitchell PB, Schofield PR, Montgomery GW, Medland SE, Gordon SD, Martin NG, Gustafsson O, Andreassen O, Djurovic S, Sigurdsson E, Steinberg S, Stefansson H, Stefansson K, Kapur-Pojskic L, Oruc L, Rivas F, Mayoral F, Chuchalin A, Babadjanova G, Tiganov AS, Pantelejeva G, Abramova LI, Grigoroiu-Serbanescu M, Diaconu CC, Czerski PM, Hauser J, Zimmer A, Lathrop M, Schulze TG, Wienker TF, Schumacher J, Maier W, Propping P, Rietschel M, Nöthen MM. Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *Am J Hum Genet*. 2011 Mar 11, 88(3):372-81. PMID: 21353194

177. Djurovic S, Gustafsson O, Mattingsdal M, Athanasiu L, Bjella T, Tesli M, Agartz I, Lorentzen S, Melle I, Morken G, Andreassen OA. A genome-wide association study of bipolar disorder in Norwegian individuals, followed by replication in Icelandic sample. *J Affect Disord*. 2010 Oct, 126(1-2):312-6. PMID: 20451256

178. Lee MT, Chen CH, Lee CS, Chen CC, Chong MY, Ouyang WC, Chiu NY, Chuo LJ, Chen CY, Tan HK, Lane HY, Chang TJ, Lin CH, Jou SH, Hou YM, Feng J, Lai TJ, Tung CL, Chen TJ, Chang CJ, Lung FW, Chen CK, Shiah IS, Liu CY, Teng PR, Chen KH, Shen LJ, Cheng CS, Chang TP, Li CF, Chou CH, Chen CY, Wang KH, Fann CS, Wu JY, Chen YT, Cheng AT. Genome-wide association study of bipolar I disorder in the Han Chinese population. *Mol Psychiatry*. 2011 May, 16(5):548-56. PMID: 20386566