Network-Based Analyses of Pathological Gene Pathways in Neuropsychiatric Disorders

Vishesh Jain
Computational Biology, Sc. B.
Department of Computer Science
Brown University, 2014

Advisor: Dr. Ben J. Raphael
Reader: Dr. Sohini Ramachandran

Abstract

Complex disorders and diseases, such as cancer and heart disease, have multifactorial etiologies with genetic, physiological, and environmental factors. Understanding their interactions enables effective research, diagnosis, and treatment. Pathway analysis searches for sets of genes differentially expressed in certain phenotypes, and this genetic context provides a foothold for further work. This study extended the network-based algorithm HotNet2 to find gene clusters involved in neuropsychiatric diseases. Autism gene expression data from the Morrow Lab at Brown University revealed pathways involved in neurotransmitter release by soluble N-ethylmaleimide sensitive factor attachment protein receptors (SNARE) and regulation of the inhibitory neurotransmitter γ-aminobutyic acid (GABA). Analysis of a bipolar disorder genome-wide association study from the Wellcome Trust Case Control Consortium verified association of cell adhesion genes involved in synaptic development. The extension of HotNet to gene expression and single-nucleotide polymorphism data opens the door to a wide range of future computational studies and meta-analyses based on complementary biological experiments. This study's approach offers mechanistic understanding and logical treatment targets for complex disorders.

Introduction

Over the past few decades, studies across diverse disciplines have revealed the multifactorial nature of complex disorders like cardiovascular disease, Alzheimer's, and cancer [1,2]. At the genetic level, the complexity of biochemical interactions and the variability of driver abnormalities across patients make research, diagnosis, and treatment incredibly difficult [1,3]. Individual mutations in such diseases often have low penetrance and account for miniscule fractions of the observed phenotype. For example, model selection for diagnostic classification must account for the relative importance and interactions of various risk factors. Without knowing which of a host of potential causes is responsible for the disease of interest, targeted treatment remains elusive. In the worst case, only epiphenomena are treated.

The polygenic nature of complex disorders naturally leads to modelling them as diseases of pathways. Moreover, both developmental effects of genetic mutations and disease-stage abnormalities in gene expression may play a role, as is found in neurological diseases including autism and schizophrenia [4,5]. Rather than targeting single genes for analysis, it is crucial to produce a more comprehensive model aggregating important genes into biochemical pathways. For example, malfunction in a cell cycle pathway may be universally critical for tumor progression while the exact genes mutated within that pathway may vary from patient to patient. Identifying the pathway thus allows a diagnosis accounting for relative interactions and relevance of different factors.

Considerable effort in computational biology has been devoted to methods for extracting relevant pathways associated with a phenotype. Rather than identifying all genes associated with a disease (of which there may be many), these tools seek to identify clusters of genes which are related to each other and to the disease. This approach has several important advantages in the context of complex disorders. First, it reduces the number of genes that must be studied in subsequent *in vivo* and *in vitro* studies. Such methods may also account for the low penetrance and prevalence of mutations; while a single gene may not be mutated in all patients, a group of interacting genes may together be significantly associated with the phenotype of interest.

Most importantly, even for monogenic disorders, identification of a causal mutation is insufficient to diagnose and treat patients. Knowledge of the function of that segment of DNA in the normal and diseased states is required. The effect of a gene mutation on its protein product and interactions with other genes/proteins provide logical targets for treatment. Variation in gene-protein interactions may underlie the prevalence of patients non-responsive to therapies to which others respond well. Pathway-level analysis offers new insights into these subtleties relevant to personalized medicine.

Several algorithms already exist to identify clusters of relevant genes. For example, Dendrix [6] focuses on the observation that it is unlikely for two genes in the same pathway to be mutated in a tumor cell. This is potentially due to either a lack of selective advantage for the second mutation or synthetic lethality caused by severe dysfunction. However, this method ignores known biological information that could improve both selectivity and sensitivity in pathway prediction. In contrast, other algorithms such as HotNet [1] and MEMo [7] find subsets of a protein interaction network that are frequently mutated, though this dependence on prior biological knowledge can reduce the power of both methods to make new discoveries. Specifically, HotNet identifies subnetworks within an existing gene interaction network (based on whole genome and biochemical studies) that are enriched with somatic mutations, originally targeted toward cancer research [1,8]. However, enrichment with other types of abnormalities, such as altered expression of genes or common variants, may also be used to target key subnetworks and pathways in various diseases.

For diseases such as autism, gene expression studies can extract abnormal transcription and translation at the current disease stage, which may have effects beyond the genome itself. Genome-wide association studies (GWAS) have the advantage of wide, relatively unbiased coverage, but it can be difficult to interpret the mechanistic effect of a single-nucleotide polymorphism (SNP) in a gene without knowledge of the protein's structure and function. Furthermore, special care must be taken when converting this SNP data to gene level data to avoid introducing too much bias from gene length or gene-wise SNP counts. To this end, we seek not only to generalize HotNet for application to these data types but also to other disorders beyond cancer. Here, we focus on two neuropsychiatric disorders, autism and bipolar disorder. In autism, multiple mutations

can affect neurodevelopment to converge upon similar phenotypes, while none of them account for a large fraction of total incidence [9,10]. On the other hand, bipolar disorder is well known as a highly multifactorial disease, dependent on environmental factors such as early life stress as well as molecular risk factors [11,12]. Subtle differences between individuals with similar phenotypes makes them difficult to treat and suitable for network-based analysis.

Methods

HotNet2, the current version of HotNet, was used for all tests, and differs from the original algorithm in several ways [8,13]. It uses a random walk with restart to model diffusion of heat, or gene scores, through the network, and filters the graph by an automatic threshold selection. Rather than selecting connected components in an undirected influence graph, HotNet2 chooses identifies strongly connected components in a directed influence graph. The primary input to HotNet2 consists of a list of genes and gene scores, as well as a network, represented by an influence matrix describing interaction levels between genes.

Input Data and Influence Matrix

Two sets of input data were used for network-based analysis. For both sets, the used network was derived from iRefIndex, an index of protein interactions compiled from several primary interaction databases [14]. The resulting iRef influence matrix, represented as a sparse square matrix in MATLAB, represented a network of 12129 genes by defining an influence score for each gene pair based on iRefIndex's gene interactions and a heat diffusion process.

Autism gene expression data was provided by Dr. Eric Morrow, a collaborator at Brown University in the Department of Molecular Biology, Cell Biology, and Biochemistry. These data consisted of fold changes in expression from healthy control subjects to autism patients for 9935 genes. Corresponding p-values were based on the magnitude of these fold change and gene-specific variability of expression. Pre-processing to remove duplicates and unmeasured genes produced 8718 genes that were

assigned p-values for this study. Of these, 6605 genes were contained in the iRef network. For this data set, only the 1445 genes with p-values less than 0.05 were used for input.

Bipolar disorder SNP data was taken from a GWAS from the Wellcome Trust Case Control Consortium (WTCCC), which analyzed a diverse array of disorders in the United Kingdom [15]. This data was transformed via PLINK from .map and .ped files to asymptotic p-values for each SNP using an allele-based chi-squared association test [16]. In total, 490032 SNP markers were measured, from which gene scores for 13098 genes were generated (see below for details). Of these, 8484 genes with scores were contained in the iRef network, all of which were input into HotNet2.

Choice of Gene Scores

Since larger scores are assumed by HotNet2 to be more important, the p-values for both sets of data had to be inverted. Thus, the score used for analysis was $-\log_{10}(p)$, which produced an appropriate distribution of scores to differentiate genes without extreme variance. While the p-values from the autism gene expression data could be used without further processing, the SNP p-values had to be converted to gene-level p-values for further analysis.

The naïve method of using the minimum p-value of all SNPs in a gene (minSNP) would be the simplest approach, but it has several issues. Most importantly, this score would not take into account various differences between genes that must be controlled; for instance, a gene with 100 SNPs would have a much larger chance of seeming related to the disease than a gene with only 1 SNP in its coding region. To address this issue, we use a permutation test to derive empirical gene-level p-values, denoted permSNP. The disease phenotypes were permuted among all subjects randomly for 10,000 repetitions, and the empirical p-value was calculated by the fraction of permutations in which the minSNP p-value for a gene was less than its true minSNP p-value [17]. Note that as with any such permutation procedure, the minimum possible empirical p-value is limited by the number of repetitions. Furthermore, while this method controls somewhat for linearly dependent gene structure, gene length, and the number of SNPs, it neither resolves these issues entirely nor accounts for other potential issues, such as ancestry-dependent variation.

As shown in Figure 1, the distribution of minSNP p-values has an excessive concentration below 0.05, composing over 35% of the genes. However, this is somewhat fixed by the permSNP procedure, which reduces the number of p-values below 0.05 and flattens the curve, distributing more genes across higher p-values.

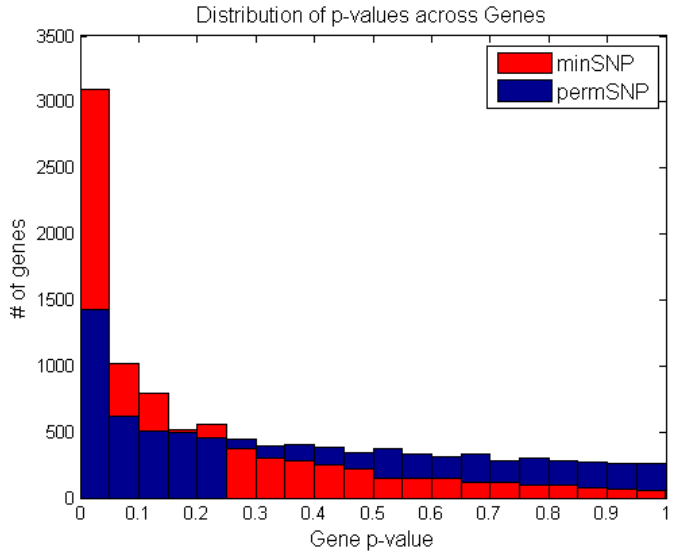


Figure 1: Histograms of p-values for the minSNP and permSNP methods. For both methods, p-values of 8484 genes were counted. The red and blue bars represent the minSNP and permSNP p-values respectively. The permSNP procedure generated less than half as many genes with p-values below 0.05 as did the minSNP procedure. Furthermore, these were redistributed in a fairly uniform manner across the higher bins, leaving all 20 bins with at least 2% of the total mass.

Although some genes contained hundreds of SNPs, over 90% of genes contained at most 34 SNPs. As shown in Figure 2, the high concentration of p-values below 0.05 in the minSNP method is largely due to genes with high numbers of SNPs. In contrast, the permSNP procedure yields a much more even distribution of p-values, reducing the dependency of the p-value on the number of SNPs.

Interestingly, there is a dramatic increase in variance in the p-values of genes with high SNP counts, which may be caused by a few different factors. Since many genes are largely unaffected by the change in procedure while others are greatly impacted, the variance may increase as a function of the greater variability in bias among genes with more SNPs. Alternatively, there may truly be more variation among genes with more SNPs, since each potential mutation may theoretically has a distinct structural or functional effect, even if many are silent with respect to phenotype.

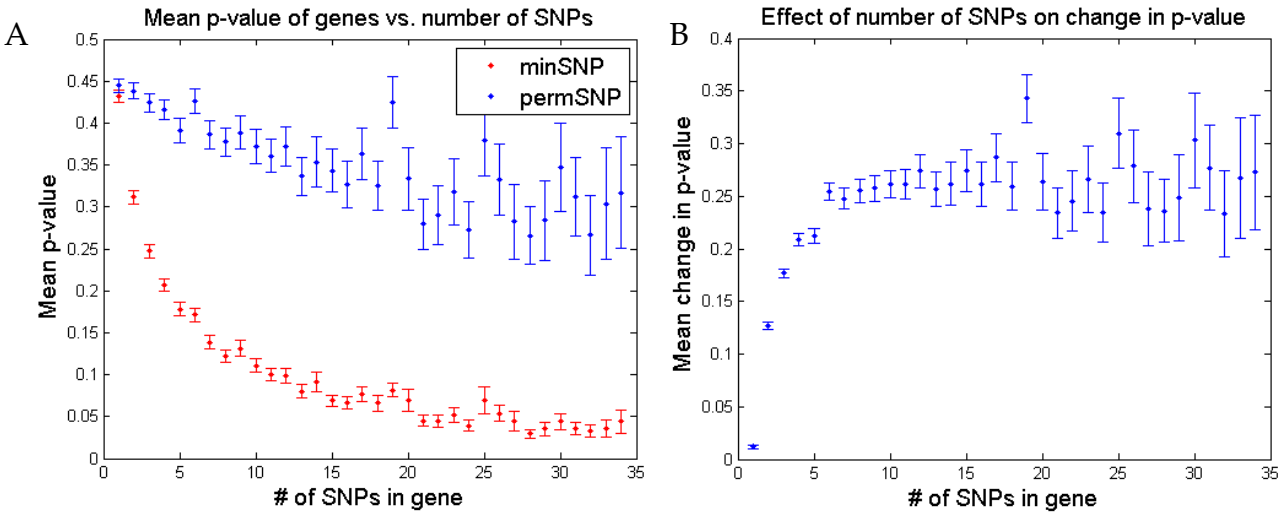


Figure 2: Dependence of p-values on the number of SNPs using the minSNP and permSNP procedures. (A) Red and blue points indicate the mean p-value for genes with a given number of SNPs using the minSNP and permSNP methods respectively. The tendency for high-SNP genes to have low p-values with the minSNP procedure is largely addressed by the permSNP method. (B) Points indicate the mean change in p-value from the minSNP to the permSNP method. In both of the plots, the error bars are drawn to indicate standard error of the mean.

Furthermore, as Figure 2B demonstrates, genes with more SNPs were adjusted more strongly, thus verifying that the permSNP procedure corrects the skewed genes preferentially. While genes containing a single SNP were barely adjusted, genes containing at least 10 SNPs showed an increase in p-value of approximately 0.25 on average. Thus, in the results below, the permSNP method was used to generate the gene-level p-values which were converted to gene scores for the WTCCC data.

Gene List Filtering

A key decision for the network permutation test performed by HotNet2 is the choice of a null hypothesis. A common distribution of p-values may be assumed across all genes in the network, all genes measured, or genes that were both measured and in the network. The last is the most conservative option, but it would be the most correct if genes in the network are more likely to have p-values below 0.05 than genes not in the network, within a given study. This is likely to be the case due to ascertainment bias in genetic research; often, genes will be added to interaction networks due to published findings, and genes of interest are more likely to be researched. Therefore, it is possible

that genes in such compiled networks are more likely to have a high mutation rate or expression fluctuation. Low p-values below 0.05 indicate significant fold changes for the autism study; in the bipolar study, they represent association between the disease and SNPs in the gene.

To address the issue of ascertainment bias, we examined the p-values from the autism gene expression data and minSNP p-values computed from the bipolar GWAS data. First, we plotted cumulative distribution functions of the p-values for genes inside and outside the iRef network for a simple visual indication of such an effect (Figure 3).

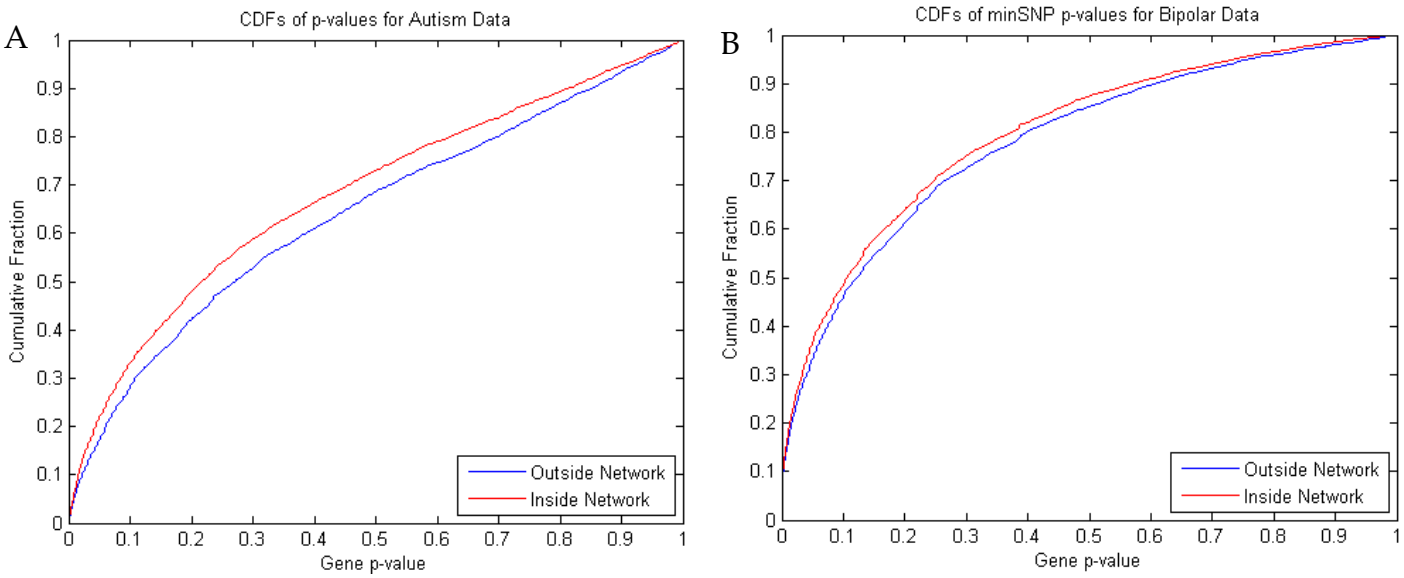


Figure 3: CDFs for p-values of genes inside and outside the iRef network. For the autism (A) and bipolar minSNP (B) p-values, genes inside the network seem to have lower p-values than those outside of them.

Next, we performed a 2-proportion z-test for each data set to quantify whether the number of significant genes for which $p(g) \leq 0.05$ differs between measured genes in and not in the network. We define the following variables:

Let H be the set of genes whose p-values were measured.
Let N be the set of genes in the iRef network, and let $O = H \setminus N$ and $I = H \cap N$ be the sets of measured genes outside and inside the network, respectively.
For any of these sets of genes, define $n_A = |A|$ as the size of the set A.
Further define $x_A = |\{g \in A \mid p(g) \leq 0.05\}|$ as the size of the set of significant genes in A.

Table 1: Z-test statistics to quantify differences in genes inside and outside the iRef Network

	$O = H \setminus N$	$I = H \cap N$	2-proportion z-test
Autism	$\frac{x_O}{n_O} = \frac{360}{2113} = 17.0\%$	$\frac{x_I}{n_I} = \frac{1445}{6605} = 21.9\%$	$p = 8.79 * 10^{-7}$
Bipolar minSNP	$\frac{x_O}{n_O} = \frac{1533}{4614} = 33.2\%$	$\frac{x_I}{n_I} = \frac{3099}{8484} = 36.5\%$	$p = 7.96 * 10^{-5}$
Bipolar permSNP	$\frac{x_O}{n_O} = \frac{763}{4614} = 16.8\%$	$\frac{x_I}{n_I} = \frac{1424}{8484} = 16.5\%$	$p = 0.358$

For both the autism and bipolar minSNP p-values, genes in the network are statistically more likely to be significant, though the absolute difference in proportion is only about 4% (Table 1). Based on the above findings, we present results using the most conservative option, permuting within only genes that were both measured and inside the network. Interestingly, the permSNP procedure seems to correct this issue in addition to removing falsely significant p-values below 0.05. This correction may have been due to a higher proportion of scored genes with at least 10 SNPs in the iRef network (31.1) than outside it (26.5%); recall that genes with more SNPs were shown to have much lower p-values on average with the minSNP procedure compared to the permSNP procedure. Thus, for the autism data but not the bipolar data, a filter to genes with $p(g) \leq 0.05$ was used to both reduce noise and provide a further correcting step; since scores were then permuted among high scoring genes, high scores could not be spread farther throughout the network, which would have falsely increased the significance of results.

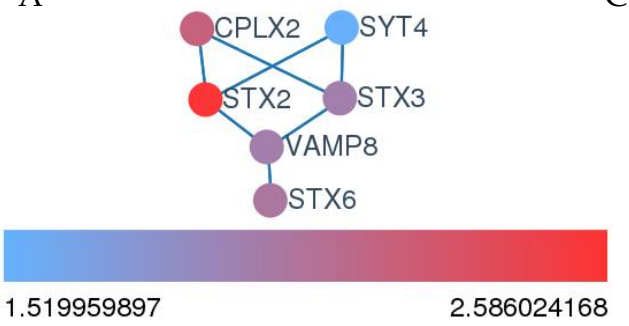
Results and Discussion

As discussed above, the following results were produced using HotNet2, using a gene score conversion of conversion of $-\log_{10}(p)$. Both the autism gene expression data set and the WTCCC bipolar data were analyzed with the iRef network. For both data sets, the resulting gene components output by HotNet2 were tested against the pathway database of the Kyoto Encyclopedia of Genes and Genomes (KEGG) [18,19]. Enrichment was calculated by modeling the presence of pathway genes in the components with a hypergeometric distribution.

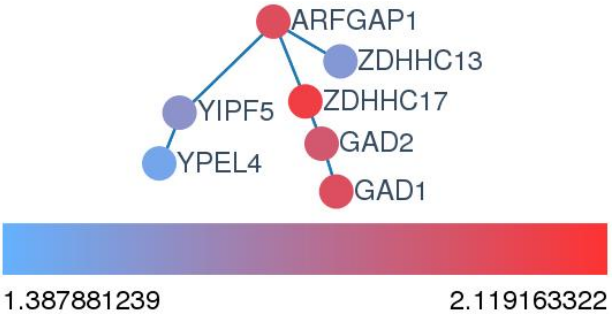
Autism Gene Expression Network Analysis

After analyzing the genes scores from the autism data in HotNet2, 2 clusters of at least 6 genes were found (Figure 4A,B). Based on HotNet2 tests permuting the gene scores across the tested genes, the probability of finding 2 clusters was 0.08 (Figure 4C). These two clusters were tested for enrichment in the KEGG pathway database [19]. Figure 4A contains genes involved in mediation of vesicle transport by soluble N-ethylmaleimide sensitive factor attachment protein receptors, or SNARE proteins (p = 1.26 * 10^{-6}). Figure 4B contains genes involved in taurine and hypotaurine metabolism (p = 2.18 * 10^{-3}). Enrichment p-values are based on a null hypergeometric distribution over genes in the HotNet2 components and KEGG pathways, and Bonferroni-corrected for the number of pathways and components tested. Each of these pathways will be examined in turn, based on research of existing literature regarding the genes and disorders. For cited articles regarding autism spectrum disorder (ASD) in which genes were shown to be up-regulated or down-regulated in disease states, fold changes (increase or decrease in expression) were in the same direction as found in this study.

A



B



C

Size	Expected	Actual	p-value
2	65.51	69	0.16
3	16.47	15	0.81
4	2.87	4	0.30
5	1.23	2	0.35
6	0.49	2	0.08
7	0.24	1	0.23
8	0.07	0	1.00

Figure 4: HotNet2 results for the autism gene expression data. (A) and (B) show clusters of genes in the SNARE and taurine pathways respectively. Color depicts gene scores. (C) shows p-values for the significance of clusters of a given size, based on expected numbers of clusters from permutation tests of gene scores.

SNARE Interactions in Vesicular Transport

SNARE machinery is crucial for neurotransmitter release at the chemical synapses of the nervous system. Indeed, while axonopathy appears in various disorders, synaptic dysfunction is a quite common mechanism for neurological toxins and diseases. In figure 4A, the syntaxin genes STX2, STX3, and STX6 are shown. The proteins coded by these genes are associated with Munc, varieties of which bind SNAP-23 and SNAP-25 [20,21]. Synaptobrevin, coded by VAMP8, also interacts with the SNAPs to catalyze membrane fusion [19,22]. These complexes are critical for synaptic function; cleavage of syntaxins by botulinum toxin can halt neurotransmitter release, and dysfunction of this pathway has been associated with hyperactivity in ASD [23,24]. Excessive vesicle release may have varied effects based on the specific synapse. Abnormal regulation of post-synaptic receptor density as well as vesicle depletion are both potential results of such dysfunction.

Taurine and Metabolism and Ubiquitin-Mediated Proteolysis

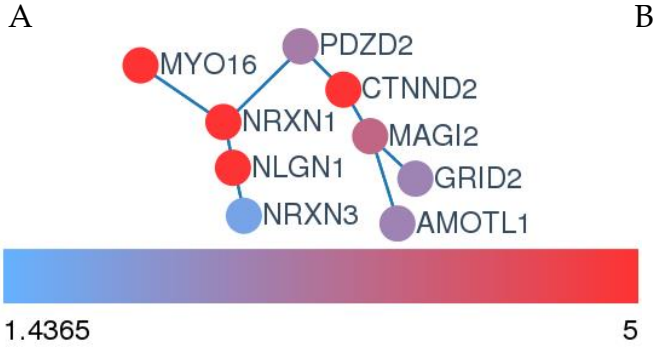
Evidence for involvement of these pathways in autism is less clear, but there is still reason to suggest that these may be worthwhile areas of inquiry. Limited evidence has been found for taurine deficiency in ASD; in Pangborn, 2002, 62% of autistic children were deficient in taurine by urine tests [25]. Furthermore, taurine has been shown to mediate activity of the most important inhibitory neurotransmitter, γ-aminobutyric acid (GABA) [26,27].

With respect to this study's results, GAD1, GAD2, and ARFGAP1 were all suppressed, with p-values within 0.01 ± 0.002. GAD1 and GAD2 are glutamate decarboxylases which produce GABA from glutamate; decreased levels of these enzymes could disturb GABAergic circuits, impact GABA immunoreactivity, and have been found in cerebellar dentate nuclei of autism patients [28,29]. Along with Sec24, ARFGAP1 may mediate placement of a GABA transporter, GAT1, on axon terminals, thus affecting axon-dendrite connections [30]. In addition, ZDHH17 is a zinc finger palmitoyltransferases, which can mediate targeting of proteins like SNAP-25 and GAD2 [31,32,33]. Disruption in the expression of these genes could thus cause alter the quantity and duration of GABA release in synapses.

Bipolar Disorder GWAS Network Analysis

P-values were generated via permSNP from the bipolar WTCCC GWAS data, and gene scores ($-\log_{10}(p)$) were used to run HotNet2 with the iRef network. One gene component of size 9 was found, and the probability of finding a cluster of at least that size was 0.07 based on permutation tests (Figure 5). This component was tested against the KEGG pathway database and was found to contain genes involved with cell adhesion (p = 7.49 * 10^{-3}) [19]. As before, the enrichment p-value is Bonferroni-corrected for the number of pathways and components tested.

A



B

Size	Expected	Actual	p-value
2	119.03	128	0.16
3	25.56	29	0.30
4	7.59	10	0.24
5	2.55	2	0.77
6	0.96	1	0.57
7	0.48	1	0.37
8	0.18	1	0.16
9	0.07	1	0.07
10	0.05	0	1.00

Figure 5: HotNet2 results for the bipolar GWAS data. (A) shows a gene cluster containing cell adhesion molecules. (B) shows p-values for the significance of clusters of a given size, based on expected numbers of clusters from gene score permutation tests.

Cell Adhesion Molecules

Cell adhesion molecules are important in the nervous system for development and structural stabilization of synapses. In particular, the associated genes NRXN1 and NRXN3, as well as NLGN1, have been associated with bipolar disorder as well as a number of other mental disorders, such as schizophrenia and autism [34,35]. Disruptions of these pathways can cause deleterious disassociation of synapses and abnormal neural communication. Another cell adhesion molecule that was associated in this data, MAGI2, is particularly involved in synaptic development along with MAGI1 and has been implicated in schizophrenia and bipolar disorder [36].

Myosin motor genes such as MYO16 are critical for neuronal growth, their depletion or dysfunction can result in neurite collapse. These motors help to drive neural dynamics and are essential for transportation along the actin cytoskeleton in the large cells [37]. Thus, SNPs in cell adhesion pathways are important targets for not only bipolar disorder but also related illnesses such as schizophrenia. These pathways demonstrate the need to jointly analyze disorders with common phenotypes, for disruption of the mechanisms in subtly different ways may result related diseases that we classify distinctly.

Conclusion

Many of the illnesses which the medical and scientific communities are struggling to combat in this decade are complex disorders, influenced by a multitude of genetic, physiological, and environmental factors. This class of diseases includes various types of cancer and cardiovascular disease as well as mental illnesses such as Alzheimer's, schizophrenia, and depression. Single-gene analysis or isolated experimentation are typically insufficient to resolve the mechanisms of these diseases, and therefore pathway-level network-based analysis is crucial for effective research, diagnosis, and treatment [1]. Focusing on two neuropsychiatric disorders, autism and bipolar disorder, this study sought to extend the network-based algorithm HotNet2 to novel diseases and data types.

Analysis of autism gene expression data from the Morrow Lab at Brown University revealed association of autism with pathways involved in SNARE-mediated vesicle release and GABAergic regulation [24,28]. These results suggest that the cellular mechanisms underlying the autism phenotype are fundamentally synaptic in nature, and the regulation of neurotransmitter exocytosis and uptake may be disrupted in several ways across different patients. Furthermore, autism is well-known to be comorbid to several other neurological disorders, including fragile X, epilepsy. Therefore, analysis of relevant pathways may produce insights into the common mechanisms of other disorders in a synergistic manner.

One example of such shared mechanisms is the common cell adhesion molecules involved in bipolar disorder and schizophrenia, two other diseases comorbid to autism [34,36]. This paper's results verified involvement of SNPs in cell adhesion genes with bipolar disorder, such as mutations in NRXN1 and NLGN1. In sum, we have demonstrated the utility of pathway-level analysis based on gene interaction networks for neuropsychiatric diseases and complex disorders in general and extended HotNet to analyze broader sets of diseases and genetic data.

Acknowledgments

I would like to thank Professors Ben J. Raphael and Sohini Ramachandran for their guidance, comments, and encouragement, as well as Jonathan V. Eldridge, Max Leiserson, and Priyanka Nakka of the Raphael and Ramachandran lab groups. I would also like to thank Dr. Eric Morrow and the Wellcome Trust Case Control Consortium for allowing us to use the gene expression and genome-wide association data for this research.

References

1. F. Vandin et al. "Algorithms for detecting significantly mutated pathways in cancer." *Journal of Computational Biology*, 2011. doi:10.1089/cmb.2010.0265
2. R. Padwal et al. "Cardiovascular risk factors and their effects on the decision to treat hypertension." *British Medical Journal*, 2001. doi:10.1136/bmj.322.7292.977
3. B.J. Maron et al. "Impact of laboratory molecular diagnosis on contemporary diagnostic criteria for genetically transmitted cardiovascular diseases." *Circulation*, 1998. doi:10.1161/01.CIR.98.14.1460
4. S. Gulsuner et al. "Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network." *Cell*, 2013. doi:10.1016/j.cell.2013.06.049
5. C.A. Walsh et al. "Autism and brain development." *Cell*, 2008. doi:10.1016/j.cell.2008.10.015
6. F. Vandin et al. "De novo discovery of mutated driver pathways in cancer." *Proceedings of the 15th Annual International Conference on Research in Computational Molecular Biology* (RECOMB 2011).
7. G. Ciriello et al. "Mutual exclusivity analysis identifies oncogenic network modules." *Genome Research*, 2012. doi:10.1101/gr.125567.111

8. F. Vandin et al. “Discovery of mutated subnetworks associated with clinical data in cancer.” *Pacific Symposium on Biocomputing*, 2012. doi:10.1142/9789814366496_0006
9. J. Veenstra-Vanderweele et al. “Autism as a paradigmatic complex genetic disorder.” *Annual Review of Genomics and Human Genetics*, 2004. doi:10.1146/annurev.genom.5.061903.180050
10. G. S. Fisch. “Is autism a polygenic disorder?” *American Journal of Medical Genetics Part A*, 2008. doi:10.1002/ajmg.a.32438
11. E. Brietzke et al. “Towards a multifactorial approach for prediction of bipolar disorder in at risk populations.” *Journal of Affective Disorders*, 2012. doi:10.1016/j.jad.2012.02.016
12. K.N. Fountoulakis. “The contemporary face of bipolar illness: complex diagnostic and therapeutic challenges.” *CNS Spectrums*, 2008.
13. K.L. Kanchi et al. “Integrated analysis of germline and somatic variants in ovarian cancer.” *Nature Communications*, 2014. doi:10.1038/ncomms4156
14. S. Razick et al. “iRefIndex: a consolidated protein interaction database with provenance. *BMC Bioinformatics*, 2008. doi:10.1186/1471-2105-9-405
15. N. Craddock. Bipolar Disorder, WTCCC1. Wellcome Trust Case Control Consortium, 2005. http://www.wtccc.org.uk/
16. S. Purcell et al. “PLINK: a toolset for whole-genome association and population-based linkage analysis.” *American Journal of Human Genetics*, 2007. doi:10.1086/519795 Version 1.07. http://pngu.mgh.harvard.edu/purcell/plink/
17. D.H. Ballard et al. “Comparisons of multi-marker association methods to detect association between a candidate region and disease.” *Genetic Epidemiology*, 2010. doi:10.1002/gepi.20448
18. M. Kanehisa et al. “Data, information, knowledge and principle: back to metabolism in KEGG.” *Nucleic Acids Research*, 2014. doi:10.1093/nar/gkt1076
19. M. Kanehisa and S. Goto. “KEGG: Kyoto Encyclopedia of Genes and Genomes.” *Nucleic Acids Research*, 2000. http://www.genome.ad.jp/kegg/
20. B. Martín-Martín et al. “Involvement of SNAP-23 and syntaxin 6 in human neutrophil exocytosis.” *Blood*, 2000.
21. Y. Hata and T.C. Sudhof. “A novel ubiquitous form of Munc-18 interacts with multiple syntaxins: use of the yeast two-hybrid system to study interactions between proteins involved in membrane traffic.” *Journal of Biological Chemistry*, 1995.
22. C.G. Schuette et al. “Determinants of liposome fusion mediated by synaptic SNARE proteins.” *Proceedings of the National Academy of Sciences of the USA*, 2004. doi:10.1073/pnas.0400044101
23. V. O’Connor et al. “Disruption of syntaxin-mediated protein interactions blocks neurotransmitter secretion.” *Proceedings of the National Academy of Sciences of the USA*, 1997.

24. F.R. Guerini et al. “SNAP-25 single nucleotide polymorphisms are associated with hyperactivity in autism spectrum disorders.” *Pharmacological Research*, 2011. doi:10.1016/j.phrs.2011.03.015
25. J. Pangborn. “Autism: metabolic differentiation, role of DPPIV/CD26, some pertinent lab tests.” *Defeat Autism Now!*, Spring 2002 Conference. Autism Research Institute.
26. S.M. Molchanova et al. “Effect of taurine on the concentrations of glutamate, GABA, glutamine and alanine in the rat striatum and hippocampus.” *Proceedings of the Western Pharmacological Society*, 2007.
27. F. Jia et al. “Taurine is a potent activator of extrasynaptic $GABA_A$ receptors in the thalamus.” *Journal of Neuroscience*, 2008. doi:10.1523/JNEUROSCI.3996-07.2008
28. H. Chao et al. “Dysfunction in GABA signaling mediates autism-like stereotypies and Rett syndrome phenotypes.” *Nature*, 2010. doi:10.1038/nature09582
29. J. Yip et al. “Decreased GAD65 mRNA levels in select subpopulations of neurons in the cerebellar dentate nuclei in autism: an in situ hybridization study.” *Autism Research*, 2009. doi:10.1002/aur.62
30. V. Reiterer et al. “Sec24- and ARFGAP1-dependent trafficking of GABA Transporter-1 is a prerequisite for correct axonal targeting.” *Journal of Neuroscience*, 2008. doi:10.1523/JNEUROSCI.3451-08.2008
31. D.B. Rush et al. “Palmitoylation and trafficking of GAD65 are impaired in a cellular model of Huntington’s disease.” *Biochemical Journal*, 2012. doi:10.1042/BJ20110679
32. J. Greaves and L.H. Chamberlain. “Differential palmitoylation regulates intracellular patterning of SNAP25.” *Journal of Cell Science*, 2011. doi:10.1242/jcs.079095
33. J. Greaves et al. “Palmitoylation of the SNAP25 protein family: specificity and regulation by DHHC palmitoyl transferases.” *Journal of Biological Chemistry*, 2010. doi:10.1074/jbc.M110.119289
34. A.P. Corvin. “Neuronal cell adhesion genes: key players in risk for schizophrenia, bipolar disorder, and other neurodevelopmental disorders?” *Cell Adhesion and Migration*, 2010. doi:10.4161/cam.4.4.12460
35. J.T. Glessner et al. “Autism genome-wide copy number variation reveals ubiquitin and neuronal genes.” *Nature*, 2009. doi:10.1038/nature07953
36. R. Karlsson et al. “MAGI1 copy number variation in bipolar affective disorder and schizophrenia.” *Biological Psychiatry*, 2012. doi:10.1016/j.biopsych.2012.01.020
37. S.R. Wylie and P.D. Chantler. “Myosin IIC: a third molecular motor driving neuronal dynamics.” *Molecular Biology of the Cell*, 2008. doi:10.1091/mbc.E07-08-0744