Gene-Set Enrichment with Mathematical Biology (GEMB)

Amy L Cochran[1,*], Daniel B Forger[2,3], Sebastian Zöllner[4,5], and Melvin G McInnis[5]

[1]Department of Biostatistics and Medical Informatics, University of Wisconsin - Madison
[2]Department of Mathematics, University of Michigan
[3]Department of Computational Medicine and Bioinformatics, University of Michigan
[4]Department of Biostatistics, University of Michigan
[5]Department of Psychiatry, University of Michigan
[*]cochran4@wisc.edu

Abstract

Background: Multiple genes can contribute to disease risk. Common pathways can be identified among risk genes with gene-set analysis, which measures association between a *set* of pathway-related genes and the disorder of interest. Pathway-related genes, however, are often broadly defined having differential contributions to diverse pathway functions. Mathematical models of biology can predict the relative contribution of a gene to a specific function of a pathway. Our goal was to use model predictions to strengthen gene to disorder connections by identifying a specific function common among risk genes.

Results: We present a method (GEMB) to enrich gene-set analyses with models from mathematical biology. The method combines gene *weights* from model predictions and gene *ranks* from genome-wide association studies into a weighted gene-set test. We use the test to examine the hypothesis that intracellular calcium ion concentrations contribute to bipolar disorder, using publicly-available summary data from the Psychiatric Genetics Consortium (n=41,653; ∼9 million SNPs).

Conclusions: With our method, we can strengthen inferences from a P-value of 0.081 to 1.7×10^{-4} by moving from a general calcium signaling pathway to a specific model-predicted function. This dramatic difference demonstrates that incorporating math biology into gene-set analysis can strengthen gene to disease connections.

Keywords: GWAS; mathematical biology; gene ontology; genetic enrichment; gene-set analysis; bipolar disorder; calcium signaling.

1 Introduction

A single genetic variant confers little risk despite high heritability for many disorders, particularly in psychiatry [10, 67, 66]. Such disorders are considered to be *polygenic* [21], i.e. hundreds to thousands of genes of very small effect contribute to the disorder. This has led to the growing interest in measuring and reporting polygenic risk score in many disorders [56, 25, 2, 48, 30]. A challenge with polygenic disorders is that candidate genes are usually involved in disjunctive biological pathways

and hence have not enjoyed practical applications [68]. Computational approaches are emerging to better prioritize candidate genes [71, 23, 28, 15, 14, 3, 33, 27, 42, 6, 22, 39].

Gene-set analyses examine the association between a disorder and a *set* of genes rather than a single gene [57, 31, 64, 41, 5, 34]. When the set consists of genes from a biological pathway, strong associations suggest the pathway is involved in the disorder, helping to elucidate disease mechanisms. Such pathway-level enrichment could even strengthen associations when an abnormal biological pathway, and not single gene variants, is what ultimately contribute to a disorder's risk [19, 34]. Further, many statistical tests and software are available to perform gene-set analysis (cf. [51, 34]) to determine whether genes in a particular gene set are significantly associated with a phenotype (self-contained) or whether a phenotype is more strongly associated with genes in a set than genes not in the set (competitive) [41, 34, 70].

If we follow the argument that gene-set analysis strengthens gene-disorder associations, then function-level analyses will further strengthen these associations, since abnormal biological function, and not an abnormal pathway or single gene variant, might ultimately contribute to a disorder. For instance, *firing rate* (i.e. the number of action potentials in a time window divided by the duration of the time window) differs between induced pluripotent stem cells derived from bipolar individuals versus controls [13, 54]. If firing rate is important, gene-set analysis would prioritize genes that contribute to the firing rate. However, biological functions such as firing rate do not map one to one to genetic pathways. Instead, each function recruits subsets of genes from multiple biological pathways [63]. Hence, if perturbed biological functions drive disease risk, jointly testing genes in one pathway which includes genes of little impact and ignores genes in other pathways would result in a less powerful gene-set analysis. Rather, we should aim to define gene sets based on biological function.

Detailed mathematical models of biological pathways have been proposed and shown to be relevant to many medical fields including psychiatry (cf. [17, 18, 9, 8, 12]). These efforts have culminated in ModelDB (`https://senselab.med.yale.edu/modeldb/`) which hosts over 1000 publicly-available models [45]. Examples include models of the hypothalamic-pituitary-adrenal axis, monoamine systems, circadian rhythms, among others. By varying model parameters related to genes, these models offer a way to measure the relative contribution of genes to a specific biological function of interest (e.g., firing rate). Incorporating model predictions into gene-set tests might help to further strengthen the link between genes and disorders.

Here, we present a simple method (GEMB: Gene-set Enrichment with Math Biology) for measuring association between a disorder and model predictions of function. Our approach relies on (i) *ranking genes* in decreasing order of association strength to a disorder and (ii) *assigning weights* to a subset of genes to reflect an estimate of their relative contribution to a specific biological function. In particular, we illustrate how one could assign weights using neurobiological models to predict how various genes relate to specific biological functions. The ranks and weights are then combined into a single test statistic, which is then tested for significance. The test thus measures association between a biological function (as predicted by the neurobiological model) and the disorder of interest.

With GEMB and a model of intracellular calcium [4], we explore the hypothesis that intracellular calcium concentrations contribute to bipolar disorder. Bipolar disorder is a severe and chronic psychiatric disorder characterized by pathological and unstable swings of mood, ranging from mania to depression [7]. The heritability of this disorder is estimated to be 85% [46], and genetic association studies report several susceptibility loci [62], beginning with a voltage gated calcium gene [26], which remains among the strongest findings to date. Calcium signaling is complex and implicated in many human diseases [32], including bipolar disorder wherein calcium pathways are implicated in etiology [32] and management [44]. Further impetus is the combined findings from

replicated genetic association studies [26, 50] and replicated functional studies [13, 49]. These reports provided a sound basis for pursuing calcium signaling in bipolar disorder. In what follows, we demonstrate that our method can strengthen inferences from a P-value of 0.081 to 1.7×10^{-4} by moving from a general calcium signaling pathway to a specific function predicted by the model of Ashhad and Narayanan [4].

2 Method description

2.1 A weighted gene-set statistic

We assume a general set-up of a competitive gene-set test: individuals are phenotyped and analyzed for expression in n genes; each gene is measured for association to the phenotype; and a subset of m genes are determined to be of interest. From this set-up, we require only the rank of each gene in decreasing order of association strength to the phenotype, where genes that are most strongly-associated with the phenotype have the highest rank (i.e, closest to 1) and those that have most weakly-associated with the phenotype have the lowest rank (i.e., closest to n).

We then diverge from many gene-set tests by requiring that non-negative weights are assigned to individual genes in the subset of interest. Formally, we require:

- genes labeled 1 to n;
- rank $r_i \in \{1, \ldots, n\}$ for each gene $i = 1, \ldots, n$;
- gene set $\mathcal{S} \subseteq \{1, \ldots, n\}$; and
- weights $w_i \geq 0$ for each gene $i \in \mathcal{S}$ with $\sum_{\mathcal{S}} w_i > 0$.

It will be without any loss of generality that we assume weights w_i sum to one. Then, we can define the following test statistic using a weighted-sum of ranks r_i ($i \in \mathcal{S}$):

$$v := \sum_{i \in \mathcal{S}} w_i r_i.$$

Each choice of weights encodes an *a priori* hypothesis about the relative contribution of weights to the phenotype. As a specific case, we can recover an unweighted gene-set test by setting $a_i := 1/m$. This choice of weights captures the *a priori* hypothesis that each gene in $\mathcal{S}$ contributes equally to the phenotype (or a lack of support for one gene over another). The statistic v would be the average rank of genes in $\mathcal{S}$. A value $v > \frac{n+1}{2}$ reflects that genes in $\mathcal{S}$ are ranked lower *on average* relative to genes not in $\mathcal{S}$; a value $v < \frac{n+1}{2}$ reflects that genes in $\mathcal{S}$ are ranked higher *on average* relative to genes not in $\mathcal{S}$; and $v = \frac{n+1}{2}$ means that genes in $\mathcal{S}$ are neither ranked higher nor lower on average relative to genes not in $\mathcal{S}$. In other words, small v suggests an association between the gene set and phenotype. We point out that genes in $\mathcal{S}$ do not need to be evenly distributed in rank to achieve $v \approx \frac{n+1}{2}$; they could be disproportionately ranked close to the average rank $\frac{n+1}{2}$ or ranked close to the extreme ranks 1 and n.

As another specific case, we could recover a single gene test by setting $a_j = 1$ for some $j \in \mathcal{S}$ and all other weights zero. This choice captures the *a priori* hypothesis that gene i specifically contributes to the phenotype. The statistic v would be the rank of gene i. More broadly, setting any weight to zero reflects the hypothesis that the corresponding gene does not contribute to the phenotype. The statistic v would be identical in value if we had simply removed the gene from s. Further, smaller weights means a smaller contribution to v.

With more general weights, the statistic v is interpreted similar to the unweighted version, but replacing an average of rank with a weighted average. Our interpretation is inherited from the fact that $v-(n+1)/2$ changes sign when genes are ranked in opposite order and increases when a gene in $\mathcal{S}$ is exchanged for a gene not in $\mathcal{S}$ with higher rank. So, small v can be thought of as genes in $\mathcal{S}$ have a higher (weighted) relative rank to genes not in $\mathcal{S}$.

2.2 A weighted gene-set test

To use v in a statistical test, we must specify a null distribution. For many gene-set analyses, a common null hypothesis is that the genes in $\mathcal{S}$ were chosen uniformly at random from the entire set of genes. Under this null hypothesis, we can construct a null distribution for v by drawing ranks for genes in our set, R_i for $i \in \mathcal{S}$, uniformly at random from

$$\{1, 2, \ldots, n\}$$

without replacement and then calculating

$$V := \sum_{i \in \mathcal{S}} w_i R_i.$$

The distribution of random variable V serves as the null distribution for v.

The alternative hypothesis is that genes in $\mathcal{S}$ were not chosen uniformly at random. In broad terms, they were chosen because of their relationship to the phenotype. Hence, we are interested in how often V with gene ranks chosen randomly suggests a stronger association between set $\mathcal{S}$ and a phenotype than the statistic v determined by the actual association to the phenotype. In other words, we use the probability, or p-value, associated with a one-sided test given by

$$\mathbb{P}\left(V \leq v\right)$$

to determine whether v is significant. Note, a two-sided test could also be defined by using

$$\mathbb{P}\left(\left|V - \frac{n+1}{2}\right| \geq \left|v - \frac{n+1}{2}\right|\right)$$

A simple way to estimate $\mathbb{P}\left(V \leq v\right)$ is to use Monte Carlo simulation, where V is repeatedly sampled from its distribution and we count how often a sample of V is less than or equal to v. This computation benefits from the fact that V is simple to calculate and can be sampled in parallel. The law of large numbers ensures a Monte Carlo estimate of $\mathbb{P}\left(V \leq v\right) = \mathbb{E}(1_{V \leq v})$ is unbiased and has variance $\Pr(V \leq v)/k$ where k is the number of Monte Carlo samples. Alternatively, we could estimate $\mathbb{P}(V \leq v)$ with

$$\Phi\left(\frac{v-\mu}{\sigma_w}\right)$$

where Φ is a standard normal distribution and $\mu = \frac{n+1}{2}$ and $\sigma_w^2 = \frac{n^2-1}{12}\sum_{i \in \mathcal{S}} w_i^2$. This approximation follows by making the simplifying assumption that ranks R_i are drawn uniformly at random from $\{1, \ldots, n\}$ *with* replacement (as opposed to without replacement) and then noting that the resulting V is a sum of independent random variables with respective means $w_i \frac{n+1}{2}$ and variances $w_i^2 \frac{n^2-1}{12}$ $(i \in \mathcal{S})$. Table 1 compares Monte Carlo estimates of one-sided P-values to estimates using a normal approximation.

			v			
l	n	n	$\mu - 4\sigma_w$	$\mu - 3\sigma_w$	$\mu - 2\sigma_1$	$\mu - \sigma_1$
0.5	10	1000	-3.0E-05	-5.7E-04	-1.2E-03	2.2E-03
0.5	10	10000	-2.9E-05	-5.3E-04	-8.0E-04	3.0E-03
0.5	100	1000	-2.1E-05	-5.5E-04	-4.7E-03	-1.1E-02
0.5	100	10000	-8.0E-06	-1.2E-04	-6.1E-04	-7.2E-04
1	10	1000	-3.1E-05	-7.4E-04	-1.5E-03	3.8E-03
1	10	10000	-3.1E-05	-7.0E-04	-1.1E-03	4.4E-03
1	100	1000	-1.8E-05	-5.0E-04	-4.0E-03	-8.7E-03
1	100	10000	-9.2E-06	-9.9E-05	-5.1E-04	-5.6E-04
2	10	1000	-3.2E-05	-1.1E-03	-2.4E-03	6.7E-03
2	10	10000	-3.2E-05	-1.1E-03	-2.1E-03	7.0E-03
2	100	1000	-1.7E-05	-4.4E-04	-3.1E-03	-6.2E-03
2	100	10000	-8.2E-06	-1.6E-04	-4.0E-04	-7.3E-05

Table 1: Difference between Monte Carlo estimates of a one-sided P-value $\mathbb{P}(V \leq v)$ for the weighted gene-set test and estimates using a normal approximation. Weights were defined as $w_i \propto i^l$ $(i = 1, \ldots, m)$ for various l, assuming $\mathcal{S} = \{1, \ldots, m\}$. A total of 10^7 Monte Carlo samples were used in each case.

2.3 Determining gene weights

Weights can be chosen to capture any *a priori* hypothesis justified by functional data, literature surveys, or experiments. Our goal, however, was more specific: we want weights to reflect the relative contribution of genes to a specific biological function. That way, our weighted gene-set test evaluates the hypothesis that a specific biological function (captured by the weights) is important to a phenotype.

We propose a general approach to accomplish our goal. We start with a neurobiological model that can return a scalar measure of the function of interest. Next, we identify model parameters that are controlled by genes. Then, we perform a global sensitivity analysis to measure the relative contribution of each parameter to a specific function of interest. Last, we use some function of these contributions for each parameter to assign weights to genes that are matched to that parameter.

For example, we might be interested in average calcium ion concentration in a cell during an action potential. We could use a Hodgkin-Huxley type model of action potentials that includes Calcium ions [24]. Conductance of the L-type Calcium ion channel is a common model parameter that modulates Calcium ion flux into the cell. Meanwhile, four genes (CACNA1[C-D,F,S]) are known to impact L-type Calcium ion channels. We might thus measure the relative change in average Calcium ion concentration if the conductance of L-type Calcium ion channel is varied by $\pm 5\%$ its baseline value. We could then weigh the L-type Calcium ion channel genes by the magnitude of this relative change.

While any global sensitivity analysis can be used, we opt for the popular and relatively simple sampling-based approach based on the partial rank correlation coefficient (PRCC) [43]. Parameters are sampled using Latin hypercube sampling; the model is simulated for the sampled parameters to measure the function of interest; and a PRCC for each parameter is estimated from the sampled parameters and functional measures.

We remark that the association between genes and parameters need not be one-to-one. On one hand, multiple genes may be associated with a single variable, since the model may not be

sufficiently detailed to capture individual contribution of a gene. As noted above, four genes are known to modulate formation of L-type Calcium ion channels, and most mathematical models with L-type ion channels are not sufficiently specific to model differential contributions of each gene. On the other hand, multiple parameters may be associated with a single gene, since the model may include different variables to represent the same gene. For example, models of neuronal action potential often distinguish between sodium currents and persistent sodium currents [55] even though both currents may be regulated by the same genes [38].

3 Method application

To illustrate our method, we explore the hypothesis that *intracellular calcium ion (Ca^{2+}) concentrations in excitable neurons contributes to bipolar disorder*. Our interest in this particular hypothesis stems from several considerations. First, we found initial support for this hypothesis by testing our method with a relatively small dataset (n=544) from the Prechter Bipolar Cohort [47] (details in Appendix). Thus, the results reported here represent a replication of our initial finding and validation of an *a priori* hypothesis with a much larger dataset. Second, calcium signaling is an ideal candidate because it has been both extensively modeled and implicated in bipolar disorder. Variation in the CACNAC1 gene, for example, has been repeatedly associated with bipolar disorder [26, 57, 58, 60]. See Dupont et al [24] for an overview on models of Calcium signaling and Casamassima et al [11] for an overview on the role of L-Type Calcium channels in bipolar disorder. Accompanying code is available at `https://github.com/cochran4/GEMB`.

3.1 Genes and their weights

We used a detailed model of an intracellular Ca^{2+} concentration in a hippocampus CA1 pyramidal cell developed by Ashhad and Narayanan in [4]. The model has a number of attractive features. First, it is publicly-available in modelDB (Model 150551) and written with free Neuron software. Second, it simulates important components that contribute to intracellular Ca^{2+} concentrations, including dynamics of ions (K^+, Na^+, and Ca^+) across the cell membrane; Ca^+ into and out of the sarcoplasmic endoplasmic reticulum; synaptic plasticity; and mediating receptors such as inositol triphosphate ($InsP_3$), ionotropic glutamate receptors, and metabotropic glutamate recptors (mGR). Third, the model also uses a morphologically realistic three-dimensional model of a hippocampus CA1 pyramidal cell accompanied by spatial dynamics giving rise to Ca^{2+} waves.

To identify genes of interest, we started with 182 genes making up the *Calcium signaling pathway* (Pathway ko04020) in the Kyoto Encyclopedia of Genes and Genomes (KEGG) [36, 37, 35]. Each gene was evaluated for whether it could modulate intracellular Ca^{2+} concentrations in the model. Gene function was determined using the Gene database from the National Center for Biotechnology Information (`https://www.ncbi.nlm.nih.gov/gene`). A total of 38 genes could modulate intracellular Ca^{2+} concentrations in the model, by way of ion channels, ion pumps, or receptors. We then found three ion channels (Na^+, A-type K^+, and delayed rectifying K^+) and two receptors (NMDA and AMPA) that could affect intracellular Ca^{2+} concentration in the model, but were not associated with genes in the KEGG Calcium signaling pathway. An additional 31 genes were found related to these channels or receptors. Of the 69 genes identified, 4 genes (ATP2B3,CACNA1F,GRIA3,KCND1) were excluded, because they were not associated with gene ranks (described below). A total of 65 genes were analyzed.

For each gene, we then identified a parameter that could modulate (up and down) the modeling component related to the gene. For ion channel genes, we used channel conductance. For pumps, we used rate of ion flux. For receptors, we used either initial density, formation rate, or permeability

depending on available parameters. Default parameter values were taken from the simulation in Figure 6 of [4]. The resulting genes, associated parameters, and default values are summarized in Table 2.

Genes	**Value**	**Parameter**
ATP2B[1-2,4]	0.008 μM ms^{-1}	Average rate γ_0 of Ca^{2+} flux density
ATP2A[1–3]	0.1 μM ms^{-1}	Amplitude V_{max} of SERCA pump uptake
CACNA1[C–D,S]	0.316 mS cm^{-1}	L-Type Ca^{2+} channel conductance g_{CaL}
CACNA1[G–I]	0.1 mS cm^{-1}	T-Type Ca^{2+} channel conductance g_{CaT}
GRM[1,5]	0.3e-3	Metabotropic glutamate receptor density $[mGR]_0$
GNA[Q,11,14–15]	100 ms^{-1}	G_α-bound activated PLC formation rate k_7
PLC[B1–B4,D1,D3–D4,E1,G1–G2,Z1]	0.83 ms^{-1}	PLC_α-bound PIP_2 formation rate k_9
ITPR[1–3]	1.85	IP3 receptor density $\bar{g}_{InsP_3R}$
GRIN[1,2A–2D,3A–3B]	1.938107025 nM s^{-1}	Maximum NMDA receptor permeability $\bar{P}_{NMDA}$
GRIA[1–2,4]	1.29207135 nM s^{-1}	Maximum AMPA receptor permeability $\bar{P}_{AMPA}$
KCN[A4,C3–C4,D2–D3]	22 mS cm^{-2}	A-type K^+ channel conductance g_{KA}
KCN[A1–A3,A6–A7,B1–B2,C1–C2]	3 mS cm^{-2}	Delayed rectifying K^+ channel conductance $\bar{g}_{KDR}$
SCN[1–5,8–11]A	90 mS cm^{-2}	Na^+ channel conductance $\bar{g}_{Na}$

Table 2: Calcium genes and associated model parameters. Calcium genes impact either ion channels, ion pumps, or receptors in the Ashhad and Narayanan model [4]. Baseline parameter values were taken from [4].

With parameters and genes identified, we used the Ashhad and Narayanan model to simulate intracellular Ca^{2+} concentrations during an established protocol for inducing synaptic plasticity at a synapse: 900 pulse stimulation at 10 Hz; see Figure 6 in [4]. We simulated intracellular Ca^{2+} concentrations for 1000 samples of a set of the 13 parameters. Parameters sets were chosen using latin-hypercube sampling from a normal distribution with mean given by the respective baseline parameter in [4] and in Table 1; standard deviation given by 5% of the respective baseline parameter; and zero correlation. For each parameter set, we simulated intracellular Ca^{2+} and measured average intracellular Ca^{2+} concentrations during initial transients induced in the first 3 seconds of the simulation.

We then estimated PRCC between each parameter and the measured concentrations controlling for the remaining parameters (Fig 1). We find, for example, a strong positive partial correlation between average intracellular Ca^{2+} concentrations and maximum permeability $\bar{P}_{NMDA}$ of NMDA receptors. We also find a strong negative partial correlation between average intracellular Ca^{2+} concentrations and the amplitude V_{max} of SERCA pump uptake.

Based on estimated PRCCs, we then defined weights for 65 genes as follows. If ρ_k was the PRCC for parameter k and there were N_k genes assigned to parameter k, we then defined weights $|\rho_k| / N_k$ to all genes assigned to parameter k and then re-normalized weights to sum to one. Note that we could use any function of ρ_k to assign weights to associated genes. We use only the magnitude of PRCC, since measured associations between genes and phenotypes and genes and parameters are not sufficiently specific to reflect the direction of association in addition to magnitude. We divide by the number of genes assigned to parameter k, so that a single component in the model is not weighted heavily simply because there are a large number of genes assigned to the component.

3.2 Gene ranks

Summary genetic data was obtained on bipolar disorder (n=20,129) and controls (n=21,524) from the Psychiatric Genomics Consortium (PGC) [58]. Association was measured between 8,958,989

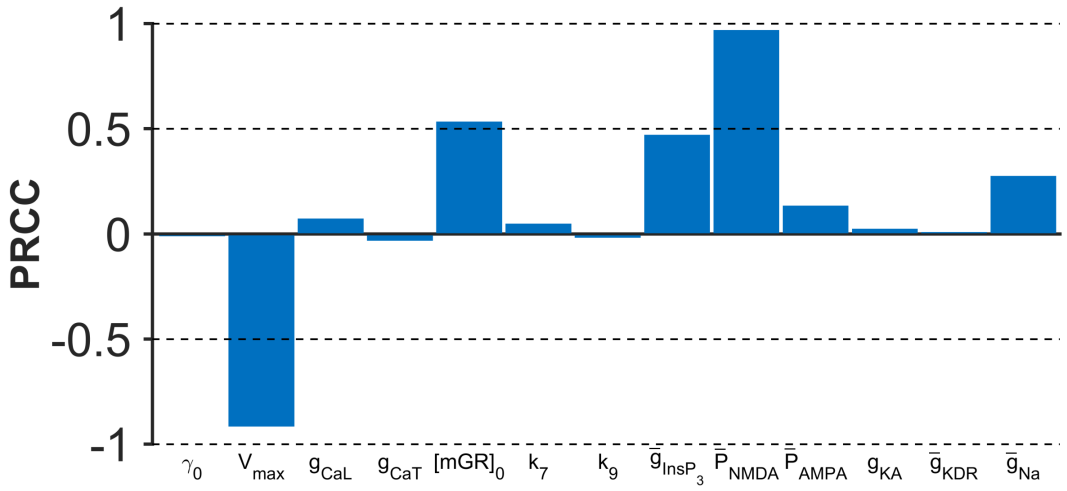

Figure 1: Estimated partial rank correlation coefficient (PRCC) estimated for 13 parameters that modulate intracellular Ca^{2+} concentrations in the Ashhad and Narayanan model [4]. PRCC measures partial correlation between a parameter and the functional measure of interest (average intracellular Ca^{2+} concentration) controlling for the contribution of other parameters.

SNPs and bipolar disorder, resulting in P values for each SNP. Data collection and analysis are detailed in Ruderfer et al [58]. Using SNP-level summary data, gene-level association with bipolar disorder was then measured using MAGMA software (`https://ctg.cncr.nl/software/magma`) [41]. Genes locations were defined using NCBI Build 37 (hg19). A total of 3,554,879 (39.68%) mapped to at least one gene, whereas at least one SNP mapped to 18,309 genes (out of 19,427 genes). Linkage disequilibrium between SNPs was estimated by MAGMA using reference data files created from Phase 3 of 1,000 Genomes [20]. A total of 18,309 genes were then ranked based on the measured association (P-value) with bipolar disorder, with smallest P-values ranked closest to 1.

3.3 Weighted gene-set test

We performed our weighted gene-set test, combining the gene ranks obtained from genetic analysis with gene weights obtained from the model of calcium signaling. We also performed our weighted gene-set test without the CACNA1C gene in order to investigate the degree to which our test is driven by the CACNA1C gene, since motivation for studying calcium was driven in part by prior PCG results that implicate CACNA1C. For comparison, we also performed an unweighted gene-set test using all 182 genes from the KEGG Calcium signaling pathway [36, 37, 35], which entailed performing our gene-set test with equal weights for all 182 genes. With these 182 genes, we also performed a typical over-representation analysis: genes were labeled as significant or not based on a significance level of 0.1 adjusted for false discovery rate [69] (a significance level 0.0044 for our problem), and then a one-sided Fisher's exact test was performed to test for over-representation of significant genes in the KEGG calcium signaling pathway compared to genes not in the KEGG calcium signaling pathway.

Our gene-set test (GEMB) showed strong support for our hypothesis that intracellular Ca^{2+} concentration contributes to bipolar I disorder (P=1.7×10^{-4}; Fig 2). The CANA1C gene is alone significant, ranking 10th out of 18195 genes (P=10/18195=5.5×10^{-4}). However, our gene-set test

continues to provide strong support for this hypothesis even with the CACNA1C gene removed (P=1.9×10^{-4}). Thus, our result is only partly driven by the CACNA1C gene. Further, focusing on the entire KEGG Calcium signaling pathway without consideration of differential contributions to biological function provides little support for the hypothesis that Calcium signaling is important to bipolar (P=0.26 using our method GEMB with equal weights and P=0.08 using a one-sided Fisher's exact test). These discrepancies in P-values illustrate how incorporating weights to provide functional specificity could help illuminate biological factors that contribute to a psychiatric disorder.

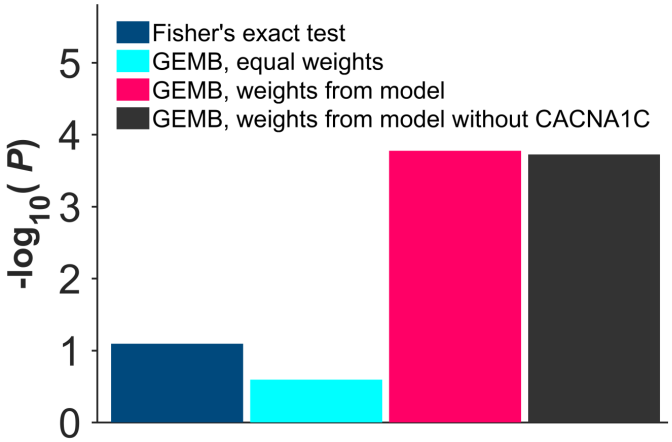

Figure 2: One-sided P-values estimated for gene-set tests. Four tests were performed: (1) over-representation test (Fisher's exact test) applied to the KEGG Calcium signaling pathway; (2) our gene-set test (GEMB) with equal weights applied to the entire KEGG Calcium signaling pathway, (3) our gene-set test (GEMB) with genes related to the Ashhad and Narayanan model [4] and weighted according to their relative contribution to our functional measure of interest (average intracellular Ca^{2+} concentration in an excitable cell); and (4) our gene-set test (GEMB) with weights and genes as determined for (3) but with the CACNA1C gene removed.

4 Discussion

We presented a method for examining associations between biological functions and psychiatric disorders which we call GEMB (Gene set Enrichment with Math Biology). Central to our method are gene weights that measure the relative contribution of a gene to a particular biological function, which we determine using a neurobiological model. We applied our approach to assess the hypothesis that intracellular Calcium ion concentrations are important to bipolar disorder. Gene weights were determined using a detailed model of calcium signaling from Ashhad and Narayanan [4] based on relative contribution of genes to intracellular calcium ion concentrations. Gene ranks were obtained using summary genetic data from the Psychiatric Genetics Consortium on bipolar disorder [59], consisting of 20,129 individuals with bipolar disorder and 21,524. Combining ranks and genes into our weighted gene-set test, we found strong support for our hypothesis that intracellular calcium concentrations contribute to bipolar I (P=1.7 × 10^{-4}) compared to little support for KEGG Calcium signaling (P= 0.10). This result suggests that biological pathways may be too

broad to capture biological *function* important to bipolar disorder. The model has the capacity to prioritize biological *function* for potential therapeutic targets. .

A benefit of our weighted gene-set test is that only gene ranks are needed from genetic data. Gene ranks can be shared across researchers more easily and requires fewer regulatory and computational resources to analyze compared to full genetic data. Sharing genetic resources and data has become the norm in genetic research, as the community moves towards large consortiums to achieve the sample sizes, level of evidence, and study consistency that are expected. The Psychiatric Genomics Consortium, for example, has ∼300 investigators and >75,000 subjects [65], and the National Institute of Mental Health (US) has made genetic data available to researchers. For similar reasons, many popular gene-set tests also need only measures of association (e.g., gene rank) rather than full genetic data [64, 70, 53]. Further, gene ranks can be recovered using MAGMA software from summary data (`https://ctg.cncr.nl/software/magma`) [41].

Once gene ranks are determined, our method then needs only gene weights from neurobiological models, which too has its benefits. Neurobiological models are numerous, experimentally-validated, and publicly-available in ModelDB (`https://senselab.med.yale.edu/modeldb/`). For example, we were able to quickly explore Calcium signaling in bipolar disorder, thanks to a detailed model developed by Ashhad and Narayanan [4] (Model 150551). Similar quick explorations could be used to examine other potentially important biological functions. In searching key words in ModelDB, we found 171 models that contain the concept of *Synaptic Plasticity*, 168 models that contain the concept of *Calcium dynamics*, 47 that contain the dopamine neurotransmitter, 9 models that contain the concept of *Circadian Rhythms*, to name a few. Together, these models could annotate genes based on model-predicted functional measures to add to current resources that annotate genes based on biological pathways, such as KEGG [36, 37, 35].

With GEMB, neurobiological models may inform genetic studies, but the reverse may also be true: genetic studies may inform neurobiological models. In psychiatry, for instance, there is growing emphasis on team science, affording many opportunities for researchers from the mathematical sciences to help tackle problems [1]. However, just as it is difficult to pin down genes to study in psychiatric disorders, it is also difficult to pin specific biological processes to study, since abnormal function is found for many neural systems in a psychiatric disorder [29]. Thus, GEMB could help identify, or ground, candidate neurobiological models for studying in psychiatry. The model of Ashhad and Narayanan [4] provides one such example.

Interest is high for ways to incorporate more functional information into gene set analysis. Network-based approaches, for example, try to incorporate measures of gene relevance based on where they lie in a network in which genes are nodes and gene interactions are edges [23, 28, 15, 14, 3, 33, 27]. Broadly, genes that are more relevant might be more connected or closer to other genes. Other examples include Bayesian approaches to account for overlap between gene sets [42, 6] or approaches based on gene expression levels [22, 39]. A benefit of our method is that it is sufficiently general that weights could thus also be determined from network analysis, experiments or meta-analysis. Weights need only be non-negative and sum to one.

The presented method GEMB was designed to be simple, which has certain limitations. First, our approach does not account for co-expression of genes unlike other genetic analyses [16, 72, 41]. Genes are known to interact in complex ways. Two gene variants, for instance, may lead to increased risk in a disorder that surpasses the additive risk of each variant alone. Second, we do not account for gene interactions in the neurobiological model. Sobol's method of global sensitivity analysis [61], for example, could measure relative contribution of parameters and their higher-order interactions. Our weighted gene-set test could be extended to incorporate these interactions. Third, neurobiological models are sure to be imperfect, meaning that gene weights are only *predicted* measures of biological function. This issue is, of course, common to all modeling. The question

then is not whether using a model leads to the correct answer, but rather whether using models to favor certain genes would strengthen inferences compared to treating the genes equally. This is an empirical question that only continued analyses and applications can answer.

Funding

This research is supported by the Heinz C Prechter Bipolar Research Fund, the Richard Tam Foundation, a Human Frontiers of Science Program Grant (RPG 24/2012), the National Science Foundation (US; DMS grant 1714094), and the National Institute of Mental Health (US; K01-MH112876).

References

[1] Abbott, A.: Us mental-health chief: psychiatry must get serious about mathematics. Nature News **539**(7627), 18 (2016)

[2] Agerbo, E., Sullivan, P.F., Vilhjalmsson, B.J., Pedersen, C.B., Mors, O., Børglum, A.D., Hougaard, D.M., Hollegaard, M.V., Meier, S., Mattheisen, M., et al.: Polygenic risk score, parental socioeconomic status, family history of psychiatric disorders, and the risk for schizophrenia: a danish population-based study and meta-analysis. JAMA psychiatry **72**(7), 635–641 (2015)

[3] Ala-Korpela, M., Kangas, A.J., Inouye, M.: Genome-wide association studies and systems biology: together at last. Trends in genetics **27**(12), 493–498 (2011)

[4] Ashhad, S., Narayanan, R.: Quantitative interactions between the a-type k+ current and inositol trisphosphate receptors regulate intraneuronal ca2+ waves and synaptic plasticity. The Journal of physiology **591**(7), 1645–1669 (2013)

[5] Askland, K., Read, C., Moore, J.: Pathways-based analyses of whole-genome association study data in bipolar disorder reveal genes mediating ion channel activity and synaptic neurotransmission. Human genetics **125**(1), 63 (2009)

[6] Bauer, S., Gagneur, J., Robinson, P.N.: Going bayesian: model-based gene set analysis of genome-scale data. Nucleic Acids Research **38**(11), 3523–3532 (2010). DOI 10.1093/nar/gkq045. URL `http://dx.doi.org/10.1093/nar/gkq045`

[7] Belmaker, R.: Bipolar disorder. New England Journal of Medicine **351**(5), 476–486 (2004)

[8] Best, J., Nijhout, H.F., Reed, M.: Serotonin synthesis, release and reuptake in terminals: a mathematical model. Theoretical Biology and Medical Modelling **7**(1), 34 (2010)

[9] Best, J.A., Nijhout, H.F., Reed, M.C.: Homeostatic mechanisms in dopamine synthesis and release: a mathematical model. Theoretical Biology and Medical Modelling **6**(1), 21 (2009)

[10] Cardno, A.G., Gottesman, I.I.: Twin studies of schizophrenia: from bow-and-arrow concordances to star wars mx and functional genomics. American journal of medical genetics **97**(1), 12–17 (2000)

[11] Casamassima, F., Hay, A.C., Benedetti, A., Lattanzi, L., Cassano, G.B., Perlis, R.H.: L-type calcium channels and psychiatric disorders: A brief review. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics **153**(8), 1373–1390 (2010)

[12] Chang, S.S., Chou, T.: A dynamical bifurcation model of bipolar disorder based on learned expectation and asymmetry in mood sensitivity. Computational Psychiatry pp. 1–18 (2018)

[13] Chen, H., DeLong, C., Bame, M., Rajapakse, I., Herron, T., McInnis, M., Oshea, K.: Transcripts involved in calcium signaling and telencephalic neuronal fate are altered in induced pluripotent stem cells from bipolar disorder patients. Translational psychiatry **4**(3), e375 (2014)

[14] Chen, J., Aronow, B.J., Jegga, A.G.: Disease candidate gene identification and prioritization using protein interaction networks. BMC bioinformatics **10**(1), 73 (2009)

[15] Chen, J., Bardes, E.E., Aronow, B.J., Jegga, A.G.: Toppgene suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Research **37**(suppl_2), W305–W311 (2009). DOI 10.1093/nar/gkp427. URL `http://dx.doi.org/10.1093/nar/gkp427`

[16] Choi, Y., Kendziorski, C.: Statistical methods for gene set co-expression analysis. Bioinformatics **25**(21), 2780–2786 (2009)

[17] Cochran, A., McInnis, M., Forger, D.: Data-driven classification of bipolar i disorder from longitudinal course of mood. Translational psychiatry **6**(10), e912 (2016)

[18] Cochran, A.L., Schultz, A., McInnis, M.G., Forger, D.B.: Testing frameworks for personalizing bipolar disorder. Translational psychiatry **8**(1), 36 (2018)

[19] Collins, A.L., Sullivan, P.F.: Genome-wide association studies in psychiatry: what have we learned? The British Journal of Psychiatry **202**(1), 1–4 (2013)

[20] Consortium, .G.P., et al.: A global reference for human genetic variation. Nature **526**(7571), 68 (2015)

[21] Consortium, I.S., et al.: Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature **460**(7256), 748 (2009)

[22] Davies, M.N., Meaburn, E.L., Schalkwyk, L.C.: Gene set enrichment; a problem of pathways. Briefings in functional genomics **9**(5-6), 385–390 (2010)

[23] Draghici, S., Khatri, P., Tarca, A.L., Amin, K., Done, A., Voichita, C., Georgescu, C., Romero, R.: A systems biology approach for pathway level analysis. Genome research **17**(10), 000–000 (2007)

[24] Dupont, G., Falcke, M., Kirk, V., Sneyd, J.: Models of calcium signalling, vol. 43. Springer (2016)

[25] Euesden, J., Lewis, C.M., Oreilly, P.F.: Prsice: polygenic risk score software. Bioinformatics **31**(9), 1466–1468 (2014)

[26] Ferreira, M.A., O'Donovan, M.C., Meng, Y.A., Jones, I.R., Ruderfer, D.M., Jones, L., Fan, J., Kirov, G., Perlis, R.H., Green, E.K., et al.: Collaborative genome-wide association analysis supports a role for ank3 and cacna1c in bipolar disorder. Nature genetics **40**(9), 1056 (2008)

[27] Fontanillo, C., Nogales-Cadenas, R., Pascual-Montano, A., De Las Rivas, J.: Functional analysis beyond enrichment: non-redundant reciprocal linkage of genes and biological terms. PloS one **6**(9), e24289 (2011)

[28] Glaab, E., Baudot, A., Krasnogor, N., Schneider, R., Valencia, A.: Enrichnet: network-based gene set enrichment analysis. Bioinformatics **28**(18), i451–i457 (2012)

[29] Goodwin, F.K., Jamison, K.R.: Manic-depressive illness: bipolar disorders and recurrent depression, vol. 1. Oxford University Press (2007)

[30] Hamshere, M.L., Langley, K., Martin, J., Agha, S.S., Stergiakouli, E., Anney, R.J., Buitelaar, J., Faraone, S.V., Lesch, K.P., Neale, B.M., et al.: High loading of polygenic risk for adhd in children with comorbid aggression. American journal of psychiatry **170**(8), 909–916 (2013)

[31] Holmans, P., Green, E.K., Pahwa, J.S., Ferreira, M.A., Purcell, S.M., Sklar, P., Owen, M.J., O'Donovan, M.C., Craddock, N., Consortium, W.T.C.C., et al.: Gene ontology analysis of gwa study data sets provides insights into the biology of bipolar disorder. The American Journal of Human Genetics **85**(1), 13–24 (2009)

[32] Hörtenhuber, M., Toledo, E.M., Smedler, E., Arenas, E., Malmersjö, S., Louhivuori, L., Uhlén, P.: Mapping genes for calcium signaling and their associated human genetic disorders. Bioinformatics **33**(16), 2547–2554 (2017)

[33] Hu, T., Sinnott-Armstrong, N.A., Kiralis, J.W., Andrew, A.S., Karagas, M.R., Moore, J.H.: Characterizing genetic interactions in human disease association studies using statistical epistasis networks. BMC bioinformatics **12**(1), 364 (2011)

[34] Huang, D.W., Sherman, B.T., Lempicki, R.A.: Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic acids research **37**(1), 1–13 (2008)

[35] Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., Morishima, K.: Kegg: new perspectives on genomes, pathways, diseases and drugs. Nucleic acids research **45**(D1), D353–D361 (2016)

[36] Kanehisa, M., Goto, S.: Kegg: kyoto encyclopedia of genes and genomes. Nucleic acids research **28**(1), 27–30 (2000)

[37] Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M.: Kegg as a reference resource for gene and protein annotation. Nucleic acids research **44**(D1), D457–D462 (2015)

[38] Kiss, T.: Persistent na-channels: Origin and function: A review János salánki memory lecture. Acta Biologica Hungarica **59**(Supplement 2), 1–12 (2008)

[39] Le-Niculescu, H., Patel, S., Bhat, M., Kuczenski, R., Faraone, S., Tsuang, M., McMahon, F., Schork, N., Nurnberger Jr, J., Niculescu Iii, A.: Convergent functional genomics of genome-wide association data for bipolar disorder: Comprehensive identification of candidate genes, pathways and mechanisms. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics **150**(2), 155–181 (2009)

[40] Lee, S.H., Ripke, S., Neale, B.M., Faraone, S.V., Purcell, S.M., Perlis, R.H., Mowry, B.J., Thapar, A., Goddard, M.E., Witte, J.S., et al.: Genetic relationship between five psychiatric disorders estimated from genome-wide snps. Nature genetics **45**(9), 984 (2013)

[41] de Leeuw, C.A., Mooij, J.M., Heskes, T., Posthuma, D.: Magma: generalized gene-set analysis of gwas data. PLoS computational biology **11**(4), e1004219 (2015)

[42] Lu, Y., Rosenfeld, R., Simon, I., Nau, G.J., Bar-Joseph, Z.: A probabilistic generative model for go enrichment analysis. Nucleic Acids Research **36**(17), e109 (2008). DOI 10.1093/nar/gkn434. URL `http://dx.doi.org/10.1093/nar/gkn434`

[43] Marino, S., Hogue, I.B., Ray, C.J., Kirschner, D.E.: A methodology for performing global uncertainty and sensitivity analysis in systems biology. Journal of theoretical biology **254**(1), 178–196 (2008)

[44] McCarthy, M.J., Le Roux, M.J., Wei, H., Beesley, S., Kelsoe, J.R., Welsh, D.K.: Calcium channel genes associated with bipolar disorder modulate lithium's amplification of circadian rhythms. Neuropharmacology **101**, 439–448 (2016)

[45] McDougal, R.A., Morse, T.M., Carnevale, T., Marenco, L., Wang, R., Migliore, M., Miller, P.L., Shepherd, G.M., Hines, M.L.: Twenty years of modeldb and beyond: building essential modeling tools for the future of neuroscience. Journal of computational neuroscience **42**(1), 1–10 (2017)

[46] McGuffin, P., Rijsdijk, F., Andrew, M., Sham, P., Katz, R., Cardno, A.: The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. Archives of general psychiatry **60**(5), 497–502 (2003)

[47] McInnis, M.G., Assari, S., Kamali, M., Ryan, K., Langenecker, S.A., Saunders, E.F., Versha, K., Evans, S., OShea, K.S., Provost, E.M., et al.: Cohort profile: the heinz c. prechter longitudinal study of bipolar disorder. International journal of epidemiology **47**(1), 28 (2018)

[48] McIntosh, A.M., Gow, A., Luciano, M., Davies, G., Liewald, D.C., Harris, S.E., Corley, J., Hall, J., Starr, J.M., Porteous, D.J., et al.: Polygenic risk for schizophrenia is associated with cognitive change between childhood and old age. Biological psychiatry **73**(10), 938–943 (2013)

[49] Mertens, J., Wang, Q.W., Kim, Y., Diana, X.Y., Pham, S., Yang, B., Zheng, Y., Diffenderfer, K.E., Zhang, J., Soltani, S., et al.: Differential responses to lithium in hyperexcitable neurons from patients with bipolar disorder. Nature **527**(7576), 95 (2015)

[50] Moon, A.L., Haan, N., Wilkinson, L.S., Thomas, K.L., Hall, J.: Cacna1c: Association with psychiatric disorders, behavior, and neurogenesis. Schizophrenia bulletin **44**(5), 958–965 (2018)

[51] Mooney, M.A., Wilmot, B.: Gene set analysis: A step-by-step guide. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics **168**(7), 517–527 (2015)

[52] Nurnberger, J.I., Blehar, M.C., Kaufmann, C.A., York-Cooler, C., Simpson, S.G., Harkavy-Friedman, J., Severe, J.B., Malaspina, D., Reich, T.: Diagnostic interview for genetic studies: rationale, unique features, and training. Archives of general psychiatry **51**(11), 849–859 (1994)

[53] O'dushlaine, C., Kenny, E., Heron, E.A., Segurado, R., Gill, M., Morris, D.W., Corvin, A.: The snp ratio test: pathway analysis of genome-wide association datasets. Bioinformatics **25**(20), 2762–2763 (2009)

[54] O'shea, K.S., McInnis, M.G.: Induced pluripotent stem cell (ipsc) models of bipolar disorder. Neuropsychopharmacology **40**(1), 248 (2015)

[55] Paul, J.R., DeWoskin, D., McMeekin, L.J., Cowell, R.M., Forger, D.B., Gamble, K.L.: Regulation of persistent sodium currents by glycogen synthase kinase 3 encodes daily rhythms of neuronal excitability. Nature communications **7**, 13470 (2016)

[56] Power, R.A., Steinberg, S., Bjornsdottir, G., Rietveld, C.A., Abdellaoui, A., Nivard, M.M., Johannesson, M., Galesloot, T.E., Hottenga, J.J., Willemsen, G., et al.: Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. Nature neuroscience **18**(7), 953 (2015)

[57] of the Psychiatric Genomics Consortium, C.D.G., et al.: Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. The Lancet **381**(9875), 1371–1379 (2013)

[58] Ruderfer, D.M., Fanous, A.H., Ripke, S., McQuillin, A., Amdur, R.L., Gejman, P.V., O'Donovan, M.C., Andreassen, O.A., Djurovic, S., Hultman, C.M., et al.: Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. Molecular psychiatry **19**(9), 1017 (2014)

[59] Ruderfer, D.M., Ripke, S., McQuillin, A., Boocock, J., Stahl, E.A., Pavlides, J.M.W., Mullins, N., Charney, A.W., Ori, A.P., Loohuis, L.M.O., et al.: Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. Cell **173**(7), 1705–1715 (2018)

[60] Sklar, P., Ripke, S., Scott, L.J., Andreassen, O.A., Cichon, S., Craddock, N., Edenberg, H.J., Nurnberger Jr, J.I., Rietschel, M., Blackwood, D., et al.: Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near odz4. Nature genetics **43**(10), 977 (2011)

[61] Sobol, I.M.: Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. Mathematics and computers in simulation **55**(1-3), 271–280 (2001)

[62] Stahl, E., Breen, G., Forstner, A., McQuillin, A., Ripke, S., Cichon, S., Scott, L., Ophoff, R., Andreassen, O.A., Kelsoe, J., et al.: Genomewide association study identifies 30 loci associated with bipolar disorder. biorxiv p. 173062 (2018)

[63] Stoney, R., Robertson, D.L., Nenadic, G., Schwartz, J.M.: Mapping biological process relationships and disease perturbations within a pathway network. NPJ systems biology and applications **4**(1), 22 (2018)

[64] Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al.: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences **102**(43), 15545–15550 (2005)

[65] Sullivan, P.F., Agrawal, A., Bulik, C.M., Andreassen, O.A., Børglum, A.D., Breen, G., Cichon, S., Edenberg, H.J., Faraone, S.V., Gelernter, J., et al.: Psychiatric genomics: an update and an agenda. American Journal of Psychiatry **175**(1), 15–27 (2017)

[66] Sullivan, P.F., Daly, M.J., O'donovan, M.: Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nature Reviews Genetics **13**(8), 537 (2012)

[67] Sullivan, P.F., Kendler, K.S., Neale, M.C.: Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Archives of general psychiatry **60**(12), 1187–1192 (2003)

[68] Tabor, H.K., Risch, N.J., Myers, R.M.: Candidate-gene approaches for studying complex genetic traits: practical considerations. Nature Reviews Genetics **3**(5), 391 (2002)

[69] Tarca, A.L., Bhatti, G., Romero, R.: A comparison of gene set analysis methods in terms of sensitivity, prioritization and specificity. PloS one **8**(11), e79217 (2013)

[70] Wang, L., Jia, P., Wolfinger, R.D., Chen, X., Zhao, Z.: Gene set analysis of genome-wide association studies: methodological issues and perspectives. Genomics **98**(1), 1–8 (2011)

[71] Zampieri, G., Van Tran, D., Donini, M., Navarin, N., Aiolli, F., Sperduti, A., Valle, G.: Scuba: scalable kernel-based gene prioritization. BMC bioinformatics **19**(1), 23 (2018)

[72] Zhang, X., Zhang, C., Wu, Z., Wang, Z., Peng, D., Chen, J., Hong, W., Yuan, C., Li, Z., Yu, S., et al.: Association of genetic variation in cacna1c with bipolar disorder in han chinese. Journal of affective disorders **150**(2), 261–265 (2013)

Appendix

Prior to applying our method to the PCG dataset discussed in the main text, we applied our method to genetic data obtained from the Prechter Bipolar Cohort, a longitudinal cohort of 1,111 individuals [47]. We focused on individuals with bipolar I disorder. Diagnoses of psychiatric illness (e.g. bipolar disorder type I) or lack of psychiatric illness (i.e. control) were determined using the Diagnostic Instrument for Genetic Studies, commonly used in psychiatric research [52]. Diagnoses obtained from the DIGS adhered to DSM-IV diagnostic criteria and were confirmed and re-confirmed annually through a consensus of three clinicians resulting in best estimate diagnoses. Participants provided whole blood samples at study intake for genetic testing of specific single nucleotide polymorphisms (SNPs). Methods pertaining to genetic testing are described in detail elsewhere [40]. About 0.5 million SNPs were analyzed initially, which were then used to impute alleles for other SNPs resulting in over 9.8 million SNPs in total.

For the application of our method, we used the same set of genes and the same gene weights obtained from simulation of the Ashhad and Narayanan model [4]. Gene ranks were obtained starting with 428 individuals with BPI and 193 controls without a psychiatric diagnoses. Genetic variation was first analyzed using PLINK software (http://zzz.bwh.harvard.edu/plink/) to account for population stratification and outliers. We performed principal component analysis on SNP data and visualized the participant loadings associated with the first two principal components. We removed any individuals who could be separated from the main cluster in this two-dimensional space either visually or with k-means clustering. This analysis was repeated until there were no participants that could be separated, leaving a total of 377 participants with BPI and 167 controls. Gene-level association to BPI was measured using MAGMA software (https://ctg.cncr.nl/software/magma) [41]. The 10 leading principal components obtained from the final principal component analysis were included as covariates. Genes locations were defined using NCBI Build 38. A total of 18,300 genes were ranked based on the measured association (P-value) with BPI, with smallest P-values ranked closest to 1.

With gene ranks and weights, we performed our weighted gene-set test (GEMB). We again compare our results to an unweighted gene-set test (applying our gene-set test with equal weights) using all 182 genes from the KEGG Calcium signaling pathway [36, 37, 35]. We also performed a typical over-representation analysis: genes were labeled as significant or not and then a one-sided Fisher's exact test was applied to test for over-representation of significant genes in the KEGG Calcium signaling pathway compared to genes not in the KEGG Calcium signaling pathway.

However, since the significance level of 0.1 adjusted for false discovery rate yielded no significant genes, we labeled the top 1% of genes as significant [69].

Our gene-set test (GEMB) showed moderate support for our hypothesis that intracellular Ca^{2+} concentration contributes to bipolar I disorder (P=0.04). By contrast, focusing on the entire KEGG Calcium signaling pathway provided little support for the hypothesis that Calcium signaling is important to bipolar I (P=0.63 using our method GEMB with equal weights and P=0.24 using a one-sided Fisher's exact test). These results provided the impetus to study intracellular calcium concentrations in the larger PCG dataset.