Biological Psychiatry

Archival Report

No Evidence That Schizophrenia Candidate Genes Are More Associated With Schizophrenia Than Noncandidate Genes

Emma C. Johnson, Richard Border, Whitney E. Melroy-Greif, Christiaan A. de Leeuw, Marissa A. Ehringer, and Matthew C. Keller

ABSTRACT

BACKGROUND: A recent analysis of 25 historical candidate gene polymorphisms for schizophrenia in the largest genome-wide association study conducted to date suggested that these commonly studied variants were no more associated with the disorder than would be expected by chance. However, the same study identified other variants within those candidate genes that demonstrated genome-wide significant associations with schizophrenia. As such, it is possible that variants within historic schizophrenia candidate genes are associated with schizophrenia at levels above those expected by chance, even if the most-studied specific polymorphisms are not.

METHODS: The present study used association statistics from the largest schizophrenia genome-wide association study conducted to date as input to a gene set analysis to investigate whether variants within schizophrenia candidate genes are enriched for association with schizophrenia.

RESULTS: As a group, variants in the most-studied candidate genes were no more associated with schizophrenia than were variants in control sets of noncandidate genes. While a small subset of candidate genes did appear to be significantly associated with schizophrenia, these genes were not particularly noteworthy given the large number of more strongly associated noncandidate genes.

CONCLUSIONS: The history of schizophrenia research should serve as a cautionary tale to candidate gene investigators examining other phenotypes: our findings indicate that the most investigated candidate gene hypotheses of schizophrenia are not well supported by genome-wide association studies, and it is likely that this will be the case for other complex traits as well.

Keywords: Candidate genes, Complex traits, Gene set analysis, Genome-wide association study, GWAS, Schizophrenia, Single nucleotide polymorphism, SNP

http://dx.doi.org/10.1016/j.biopsych.2017.06.033

Schizophrenia is highly heritable (1), and since the 1960s candidate gene studies have played a major role in research dedicated to understanding the genetic etiology of schizophrenia (2). Most historical candidate genes were selected based on known drug treatment targets and corresponding neurobiological pathways (3). As family-based genetics studies began to reveal regions of the genome that appeared to be associated with psychiatric disorders, researchers began to consider additional candidate genes located in chromosomal regions suggested by linkage analyses [e.g., *NOTCH4* (4)]. Within genes chosen in this manner, candidate gene analyses typically focused on specific variants in regions of the genome thought likely to be functional.

The SZGene database (http://www.szgene.org) (2), a curated catalog of findings from genetic association studies for schizophrenia, comprising all studies published in a peer-reviewed English language journal from 1965 to 2012, lists over 1500 published studies for schizophrenia, the majority of which were candidate gene studies. However, few clear results have emerged from these studies, with many studies reporting contradictory results for the same candidate gene polymorphisms. Factors that may underlie this inconsistency include lack of statistical power, different genetic or environmental backgrounds across studies, incomplete coverage of relevant genetic variation within candidate genes, and false positives arising from, for example, publication bias (5,6). With the advent of genome-wide association studies (GWASs), investigators can now assess the vast majority of common genetic variation across the entire genome, enabling hypothesis-free exploration of the associations between common genetic variants and schizophrenia or other complex disorders. Owing to sample sizes that are two to three orders of magnitude larger than most candidate gene studies, adherence to analytic procedures shared in common across the field, and conservative significance thresholds, associations discovered by GWASs have proven to be more robust, replicable, and reflective of the true effect sizes of common genetic variants than those based on candidate gene

SEE COMMENTARY ON PAGE 696

 ISSN: 0006-3223

reports (7). In addition, the agnostic approach of GWASs mitigates incentives (i.e., findings are reported for all loci regardless of statistical significance) to selectively report results from just certain genes or polymorphisms of interest. Moreover, modern large-scale GWASs have ample statistical power to detect effect sizes typically reported in candidate gene studies. For example, the recent Psychiatric Genomics Consortium (PGC) GWAS (8) had >99% power to detect genome-wide significant ($\alpha = 5e^{-08}$) associations that explain a mere 0.04% of the variation in schizophrenia liability, an effect size much smaller than any discovery reported in candidate gene studies of schizophrenia. For these reasons, GWAS results can be used to determine the plausibility of previously reported findings on common candidate gene polymorphisms.

Two reports in the past 5 years have compared GWAS and candidate gene study results of schizophrenia. In 2012, Collins *et al.* (9) employed a pathway analysis approach to test for enrichment of lower *p* values for all ($n = 732$) schizophrenia candidate genes identified by the SZGene database in the International Schizophrenia Consortium (10) GWAS data ($N = 6909$). They found no evidence for *p* value enrichment in this set of 732 genes after correction for multiple testing. They also calculated a polygenic risk score based on the single nucleotide polymorphisms (SNPs) located in the 732 candidate genes they examined but did not see differences between cases and controls in an independent target sample [the Genetic Association Information Network study ($N = 2366$) included genotypes of 1230 schizophrenia cases and 1136 healthy controls of European ancestry (11)].

Using the existing published results in the SZGene database, Farrell *et al.* (12) meta-analyzed the 25 most-studied schizophrenia candidate gene polymorphisms and found that none approached genome-wide significance ($p < 5e^{-08}$) in the PGC (8) schizophrenia GWAS study (34,241 cases and 45,604 controls). Moreover, the odds ratios of the significantly associated loci in the PGC study (~1.10) imply that almost all previous candidate gene studies examining genetic associations with schizophrenia diagnosis (the largest of which had a sample size some 16 times smaller than the PGC dataset) have been severely underpowered to detect any true association, much less potential associations at specific candidate polymorphisms. Though four of the most-studied candidate genes (*DRD2*, *GRM3*, *NOTCH4*, *TNF*) had genome-wide significant polymorphisms within 25 kb of their boundaries in the PGC study, only one of these associations (rs1800629 in *TNF*) was in linkage disequilibrium (LD) with the previously studied candidate polymorphism.

Still, the fact that four of the top 25 schizophrenia candidate genes contained significant GWAS signals raises the question of whether schizophrenia candidate genes themselves are supported by GWAS results, even if the specific candidate polymorphisms within them have not been. In other words, are SNPs within the most-studied schizophrenia candidate genes more strongly associated with schizophrenia than expected by chance? Previous studies have not addressed this question. Farrell *et al.* (12) focused solely on candidate polymorphisms rather than candidate genes and did not perform a gene set test for enrichment of lower GWAS *p* values for all variants within the candidate genes. Collins *et al.* (9) performed a gene set analysis for all 732 schizophrenia candidate genes identified in the SZGene database, but more than 75% of the genes currently listed in the SZGene database have been studied only once or twice, and most would not be considered "candidate genes" by researchers in the field. The current study used a gene set analysis approach and the latest PGC summary statistics (8) to determine whether polymorphisms within classic schizophrenia candidate genes are more related to schizophrenia risk than polymorphisms within other control sets of genes.

METHODS AND MATERIALS

Schizophrenia Candidate Genes

Our primary analysis focused on the same 25 top candidate genes examined by Farrell *et al.* (12) in their review (see Table 1). These 25 genes were either featured in previous reviews of schizophrenia research (13–16) or studied more than 20 times according to the SZGene database (2) and include what can be considered the "classic" candidate genes for schizophrenia (*COMT, DISC1, DRD3*, etc.). To ensure that no effects were missed, in a supplementary analysis we expanded this set to include all genes from SZGene that were 1) studied more than five times and 2) not originally motivated by GWASs. Eighty-six genes met both criteria (see Supplemental Table S1), approximately 23% of which were motivated by prior linkage results with the remaining motivated by involvement in promising biological pathways or pharmacological hypotheses. The distribution of the number of studies per gene is shown in Supplemental Figure S1.

Choosing Comparison Gene Sets

To compare the overall association of schizophrenia candidate genes to other sets of control genes, we selected genes containing polymorphisms significantly associated with one of two nonpsychiatric phenotypes genetically uncorrelated with schizophrenia according to LD Hub (17): type 2 diabetes and height. There were a total of 258 height-associated genes [keyword *height* in the database of reported associations from the GWAS Catalog (18)] and 70 type 2 diabetes-related genes (keyword *type 2 diabetes*) that did not overlap with the list of candidate genes. A list of 1028 unique genes related to pre- and postsynapse processes, chosen as a positive control, were downloaded from http://ctg.cncr.nl/software/genesets [originally curated by Ruano *et al.* (19) and Lips *et al.* (20)].

PGC GWAS Data

We downloaded the summary statistics (association *p* values for ~9.5 million imputed variants) from the PGC schizophrenia samples. Because the composition of the PGC schizophrenia sample is largely of European ancestry, we chose the 1000 Genomes Project (21) phase 1 European samples as a reference population to estimate LD between SNPs.

Statistical Analysis

We used the MAGMA software (22) to test whether the top 25 or top 86 schizophrenia candidate genes demonstrated enrichment of lower *p* values in the PGC schizophrenia GWAS data. We also used VEGAS2 software (23) to assess consistency of results across methods. Results were highly consistent (see Supplemental Methods and Supplemental Tables S5 and S6); for clarity, presented results are from MAGMA.

Table 1. Descriptive Statistics of 25 Historical Candidate Genes

Gene	NCBI ID	Location	Size (kb)	Number of Studies	Average Number of Cases	Average Number of Controls	Most-Studied Polymorphism	Type (SNP, Repeat, etc.)	Association Statistic (*Z*) From MAGMA	Rank (Excluding MHC)
NOTCH4[a]	4855	6p21.3	29.2	24	347	792	rs367398	SNP	8.78	NA
DRD2	1813	11q23	65.7	67	197	266	rs1801028	SNP	5.92	129
KCNN3[b]	3782	1q21.3	172.8	23	159	154	1333T/C	SNP	5.03	301
GRM3	2913	7q21.1	221.0	15	590	681	rs2228595	SNP	4.6	435
TNF[a]	7124	6p21.3	2.8	21	164	214	rs1800629	SNP	4.28	NA
ZDHHC8	29801	22q11.21	16.2	9	308	401	rs175174	SNP	4.11	670
PPP3CC	5533	8p21.3	100.0	9	683	763	rs7837713	SNP	3.47	1214
BDNF[b]	627	11p13	67.2	40	236	292	270C/T	SNP	3.01	1774
DAO	1610	12q24	20.9	10	440	542	rs3918346	SNP	1.87	4291
SLC6A4	6532	17q11.2	39.6	32	173	207	5-HTTVNTR	VNTR	1.64	5101
MTHFR	4524	1p36.3	20.4	20	221	290	rs1801133	SNP	1.2	6926
COMT	1312	22q11.21	28.2	81	241	383	rs4680	SNP	0.85	8645
RGS4	5999	1q23.3	8.2	22	401	493	rs2661319	SNP	0.42	10,909
DRD3	1814	3q13.3	50.3	71	168	198	rs6280	SNP	0.11	12,652
AKT1	207	14q32.32	26.4	13	478	539	rs3730358	SNP	0.07	12,850
DRD4	1815	11p15.5	3.4	45	202	230	rs4646983	Indel	0.06	12,906
NRG1	3084	8p12	1125.7	41	384	487	rs62510682	SNP	0.03	13,040
PRODH	5625	22q11.21	23.8	10	235	320	rs383964	SNP	–0.16	14,059
DTNBP1	84062	6p22.3	140.3	32	400	444	rs3213207	SNP	–0.3	14,745
HTR2A	3356	13q14	63.7	57	215	224	rs6311	SNP	–0.32	14,860
CHRNA7	1139	15q14	139.7	12	315	316	rs28531779	SNP	–0.59	16,114
DISC1	27185	1q42.1	414.5	22	348	410	rs999710	SNP	–0.66	16,384
DAOA	267012	13q33.2	25.2	27	406	526	rs3916965	SNP	–0.83	17,012
SLC6A3	6531	5p15.3	52.6	22	176	234	rs28363170	VNTR	–0.91	17,295
APOE	348	19q13.2	3.6	32	143	211	ε2/3/4	triallelic	–1.3	18,167

The average numbers of cases and controls were calculated from the SZGene database, excluding genome-wide association studies and family-based studies. Gene rankings are based on the genes' *z* scores from MAGMA, which quantify each gene's association with schizophrenia, with rankings calculated across all genes in the genome excluding the major histocompatibility complex (MHC) region.

ID, identification number; NA, not applicable (located within the MHC region); NCBI, National Center for Biotechnology Information; SNP, single nucleotide polymorphism; VNTR, variable number tandem repeat.

[a]Denotes genes within the MHC region.

[b]The polymorphisms within these genes were mistakenly designated non-SNPs in Farrell *et al*. (12), but they have been correctly labeled as SNP markers here.

We calculated the overall strength of association for each gene, z_i, by summing the –log(*p*) for all SNP *p* values within each gene boundary. The distribution of this sum is unknown but is approximated by a scaled chi-square distribution with scaling and degrees of freedom a function of the squared SNP-SNP correlation matrix, which accounts for LD between SNPs within the gene. A gene-level *p* value was derived from this scaled chi-square distribution, which was then converted to a *z* score $z_i = \Phi^{-1}(1-p_i)$ where Φ^{-1} is the probit function. For an alternative set of analyses (primarily presented in Supplemental Table S2), we derived z_i using the minimum SNP *p* value per gene instead of the sum of –log(*p*).

After calculating the strength of association for each gene in the genome, we grouped the 25 or 86 schizophrenia candidate genes into gene sets and ran one of two gene set tests in MAGMA. In our primary analysis, we used MAGMA's "competitive" test to assess whether the candidate gene set was more associated with schizophrenia than were all other genes not in the gene set, controlling for potentially confounding gene characteristics. To understand whether our set of candidate genes showed stronger or weaker association with schizophrenia than control sets of genes (genes involved in type 2 diabetes, height, or synaptic processes) we used MAGMA's "relative" test (see Supplemental Methods for additional details of these tests). As recommended by the MAGMA authors (22), we report one-tailed *p* values for competitive tests but two-tailed *p* values for relative tests.

RESULTS

Importance of Candidate Genes as a Group

We found no evidence that the 25 candidate genes of interest showed enrichment for lower *p* values in the PGC GWAS compared with all other genes in the genome. This competitive test was nonsignificant, regardless of whether we controlled for gene size, SNP density, and minor allele count (β = .28, SE = .26, *p* = .14) or not (β = .34, SE = .26, *p* = .09). Because of the strong associations with schizophrenia previously shown to exist within the major histocompatibility complex (MHC) and the long-range LD, which makes it difficult to know which genes drive the multiple associations in this region, we also repeated these analyses after removing all MHC genes,

including *NOTCH4* and *TNF*; none of our conclusions changed. All results are presented in Table 2. Further, using MAGMA's relative test, the set of 25 schizophrenia candidate genes did not show a stronger association with schizophrenia than did genes associated with type 2 diabetes (β = .39, SE = .30, p = .19), genes associated with height (β = .23, SE = .27, p = .39), or genes involved in synaptic processes (β = .15, SE = .26, p = .57). No conclusions changed when we repeated the above analyses using strict gene boundaries (Supplemental Table S3) or the minimum SNP p value in a gene as the gene-level statistic (Supplemental Table S2). Using a resampling approach, we also confirmed that the discrepancy in gene set sizes in these three relative tests (25 schizophrenia candidate genes vs. 258 height, 70 type 2 diabetes, and 1028 synaptic genes, respectively) had no influence on our conclusions (see Supplement).

When we expanded our gene set to include the 86 candidate genes that had been studied more than five times according to the SZGene database (not including GWAS results), the gene set was more associated with schizophrenia compared with all other genes (β = .27, SE = .13, p = .01). This larger gene set was more strongly associated with schizophrenia than the set of genes associated with type 2 diabetes (β = .48, SE = .20, p = .02), but no more so than genes associated with height (β = .22, SE = .15, p = .14) or genes involved in synaptic processes (β = .13, SE = .13, p = .32; Table 2). Results were similar when using strict gene boundaries, except that the relative test comparing to type 2 diabetes–related genes was no longer significant (Supplemental Table S3). When the same analyses were performed using the minimum p value as the gene-level statistic, none of the gene set associations was significant except for the test relative to height-related genes when using strict gene boundaries (Supplemental Table S2). In addition, we repeated these analyses with genes within the MHC removed; none of our conclusions changed (Table 2). Finally, conclusions did not change after excluding the 20 candidate genes motivated by prior linkage studies (Supplemental Table S4).

Contrary to initial expectations, the expanded set of 86 candidate genes yielded more significant associations (13 significant results of 40 tests conducted) than the set consisting of the 25 most-studied candidate genes (0 of 40). This pattern of results might arise if less-studied candidate genes are more related to schizophrenia, or if the ability to correctly identify relevant candidate genes has increased over time (given that the most-studied candidate genes were typically first investigated longer ago). However, neither hypothesis was supported: there were no significant relationships between candidate genes' strengths of association with schizophrenia (genewise z values) and either the number of times each was studied (r = –.06, p = .61; Supplemental Figure S2) or the first year each was studied (r = .11, p = .34; Supplemental Figure S3).

A more likely reason we observed more significant associations with the expanded gene set is differences in statistical power between analyses. Whereas adding genes unassociated with a trait to a gene set is known to decrease power, when each potential member of a gene set contributes a small amount of heritability, on average, to a trait, adding additional genes to the set increases power in competitive tests due to increased variance explained by the set (24) (note that this is not the case with relative tests because gene set size is explicitly controlled in these tests). We confirmed this increased power with increased gene set size in our own data by permuting different set sizes of genes involved in synaptic processes (which, as a set, were significantly associated with schizophrenia; β = .152, SE = .04, p = 1.94e^{-05}), and finding a strong, negative relationship between gene set size and

Table 2. MAGMA Gene Set Analyses

Model	Target Gene Set	Comparison Gene Set	β (SE)	p Value
1	Historical 25 candidate genes	All other genes	.28 (.26)	.14
2	Historical 25 candidate genes	Height-associated genes	.23 (.27)	.39
3	Historical 25 candidate genes	Type 2 diabetes–associated genes	.39 (.30)	.19
4	Historical 25 candidate genes	Genes involved in synaptic processes	.15 (.26)	.57
5	86 Most-studied candidate genes	All other genes	.27 (.13)	.01[a]
6	86 Most-studied candidate genes	Height-associated genes	.22 (.15)	.14
7	86 Most-studied candidate genes	Type 2 diabetes–associated genes	.48 (.20)	.02[a]
8	86 Most-studied candidate genes	Genes involved in synaptic processes	.13 (.13)	.32
9	Historical 25 candidate genes minus MHC genes	All other genes	.18 (.27)	.24
10	Historical 25 candidate genes minus MHC genes	Height-associated genes	.12 (.28)	.66
11	Historical 25 candidate genes minus MHC genes	Type 2 diabetes–associated genes	.32 (.31)	.30
12	Historical 25 candidate genes minus MHC genes	Genes involved in synaptic processes	.05 (.19)	.88
13	86 Most-studied candidate genes minus MHC genes	All other genes	.25 (.13)	.02[a]
14	86 Most-studied candidate genes minus MHC genes	Height-associated genes	.19 (.15)	.20
15	86 Most-studied candidate genes minus MHC genes	Type 2 diabetes–associated genes	.50 (.21)	.02[a]
16	86 Most-studied candidate genes minus MHC genes	Genes involved in synaptic processes	.11 (.13)	.41

These analyses used the sum of the negative log of the p values as the gene-level test statistic, defined the genes with extended gene boundaries (including the ±25-kb regions upstream or downstream of gene start and endpoints), and controlled for gene size, single nucleotide polymorphism density, and minor allele count, as well as the log of each. None of the four significant tests would survive multiple testing corrections.

MHC, major histocompatibility complex.

[a]Significant at $\alpha < .05$.

average p value in competitive tests (Supplemental Figure S4). Thus, the results from the 80 gene set analyses we performed are consistent with the hypothesis that schizophrenia candidate genes are weakly related to schizophrenia on average, and that tests involving larger sets were typically more significant only because the larger gene set contained more weakly related genes. This conclusion is supported by the similar β values (which estimate the predicted increase in average z value per gene for being in the set vs. not being in the set) for the competitive gene set tests for the top 25 (β_{25} = .28) and the top 86 (β_{86} = .27) candidate genes, with the difference in significance between the two tests being due to their different standard errors (SE_{25} = .26 vs. SE_{86} = .13).

Importance of Specific Candidate Genes

While we found little evidence to support the idea that the candidate genes as a group were more relevant to schizophrenia than were control sets of genes, especially when compared with genes involved in synaptic processes, several of the most-studied candidate genes were significantly related to schizophrenia, some of them highly so (Figure 1). In particular, nine genes in the set of 25 candidates were nominally ($p < .05$) associated with schizophrenia. To understand how surprising this result is for a highly polygenic trait such as schizophrenia, we permuted sets of 25 genes from the entire genome and observed nine or more nominally significant genes in 25.2% of permutations (Supplemental Figure S6), suggesting that nine significant genes of 25 is not unexpected for a highly polygenic trait such as schizophrenia. However, when we performed a relative test in MAGMA of the nine significant candidate genes versus all other genes significantly ($p < .05$) related to schizophrenia in the genome, we found evidence that the strength of the associations was greater among these nine genes than among all other significant genes (β = .789, SE = .28, p = .005). This result was largely driven by the five most significant candidate genes—*NOTCH4*, *DRD2*, *KCNN3*, *GRM3*, and *TNF*. Results were attenuated but remained significant when we dropped MHC genes from both sets (β = .738, SE = .32, p = .02) and when we compared the seven significant non-MHC candidate genes to all other significantly related non-MHC synaptic genes (β = .896, SE = .42, p = .03). Conclusions regarding the significantly related candidate genes among the broader set of less-studied 86 candidate genes were similar. Thus, there is evidence that some of the schizophrenia candidate genes are more strongly related to schizophrenia than expected by chance.

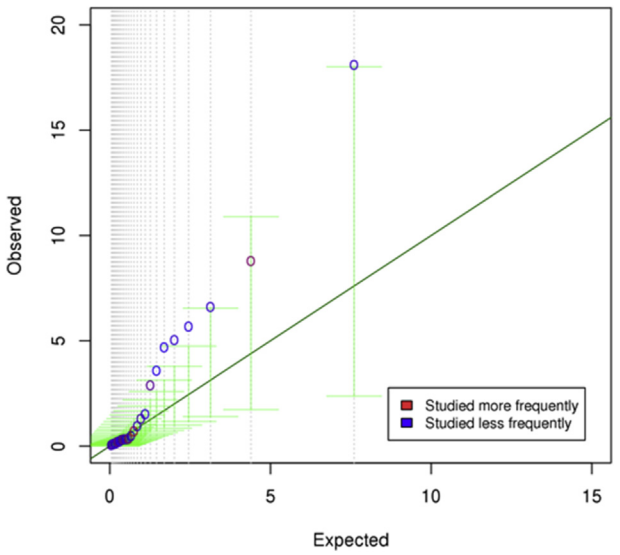

Figure 1. Quantile-quantile plot of the –log10 p values from the 25 most-studied candidate genes. Observed gene-level –log10 p values from MAGMA are plotted on the y axis, with expected –log10 p values plotted on the x axis. Points are heat map colored according to the number of times each gene has been studied, and the vertical green lines are bootstrapped 95% confidence intervals.

DISCUSSION

The overarching goal of this study was to examine whether the results from a highly powered GWAS support the hypothesis that the most-studied schizophrenia candidate genes are particularly relevant to schizophrenia. The scientific community has invested enormous time, talent, and effort in candidate gene studies over the years. It has been estimated that at least $250 million has been invested in candidate gene studies in the 1990s and 2000s (7). These studies have contributed to an improved characterization and understanding of the biological functions of many of these candidate genes. However, we found little evidence that common SNPs within these genes are any more relevant to schizophrenia than SNPs within control sets of noncandidate genes. The set of top 25 candidate genes showed no evidence of being more associated with schizophrenia compared with all other genes, or relative to genes involved with type 2 diabetes and height, which are not expected to harbor a disproportionate number of risk alleles for schizophrenia. When we expanded the gene set to include all candidate genes that had been studied more than five times, there was marginal evidence that these 86 genes were more associated with schizophrenia when compared with all other genes and relative to genes associated with type 2 diabetes, but they were no more significantly associated than were genes involved in height or the ~1000 genes in a functional category (synaptic processes) we hypothesized a priori might be related to schizophrenia. Furthermore, only two of the 16 gene set results were significant when we measured strength of association using the minimum p value per gene.

Although our results suggest that, taken as a group, schizophrenia candidate genes are no more associated with schizophrenia than random sets of control genes are, they do not imply that this is true of all the candidate genes. Indeed, we present evidence that several of the most-studied candidate genes—particularly *NOTCH4*, *DRD2*, *KCNN3*, *GRM3*, and *TNF*—are more strongly related to schizophrenia than would be expected by chance. It is important to put this evidence in perspective. First, two of these five genes (*NOTCH4* and *TNF*) are in the MHC, and given the long-range and complex nature of LD in this region, it is unclear whether it is variants in these two genes or variants in some of the other ~240 genes in the MHC that are relevant, and indeed recent evidence suggests the signal driving the *NOTCH4* association comes not from *NOTCH4* but from the nearby complement component 4 (*C4A* and *C4B*) MHC genes (25). For this reason, we hereafter

restrict discussion of the most significantly associated candidate genes to those not in the MHC. Second, while variants in *DRD2*, *KCNN3*, and *GRM3* all appear related to schizophrenia above chance, the specific polymorphisms most studied in these genes were not particularly related to schizophrenia and were not close to being among the 108 genome-wide significant SNPs discovered in the PGC study (12) (p = .22, 3.3 × 10^{-5}, and .58 respectively). Thus, our results do not agree with most previous positive findings on these candidate genes. Finally, although there can be alternative motivations for further study of these candidate genes (e.g., the fact that they are well studied and well characterized in animal studies), our results suggest that the statistical rationale for further prioritization of these genes is weak. There were 128, 300, and 434 genes more related to schizophrenia than *DRD2*, *KCNN3*, and *GRM3*, respectively. In Supplemental Table S7, we present results from the 100 top-ranked non-MHC genes, any one of which is arguably a better target of future studies than any previously studied candidate gene.

There are some important limitations to acknowledge in this study. First, we tested for enrichment of lower GWAS p values from imputed, common SNPs. Thus we cannot rule out the possibility that these candidate genes contain rare variants important to the etiology of schizophrenia. Similarly, our study focused on variants ±25 kb of gene boundaries, and ignored effects of regulatory elements that can occur at great distances (26). However, for all of the top 25 candidate genes except *DISC1* (27), the focus in the literature has been on variants within candidate gene boundaries, and typically on a specific common polymorphism within those genes. Understanding the role that rare variants might play on schizophrenia risk in these candidate genes awaits sequence or more accurate imputation data, and understanding the role of transregulatory elements awaits better understanding of long-range gene regulation and the incorporation of patterns of chromatin binding (28) or gene expression data with GWASs (29). Nevertheless, our results do not support the original hypotheses involving the most-studied candidate genes, and thus provide no reason to believe that rare variants in these genes, or the elements that regulate them, will be particularly relevant to schizophrenia.

Additionally, the imputed data used by the PGC GWAS analysis did not adequately capture common polymorphisms within all the candidate genes. For example, the most frequently studied candidate polymorphisms for *DRD4* and *NOTCH4* were neither genotyped nor successfully imputed in the PGC GWAS. The most commonly studied polymorphism in *DRD4* is a 13-bp indel, which is not well captured in the kind of sequencing done by the 1000 Genomes Project (21), the reference panel used to impute in the PGC data. The most commonly studied polymorphism in *NOTCH4,* rs367398, was also missing from the 1000 Genomes Project phase 1 reference panel, and the PGC data included no SNPs in LD with that polymorphism at $R^2 > .3$. As already noted, this SNP is located in the MHC region, which has made genotyping and analyzing this section of the genome difficult, although a recent study suggests that a large proportion of the genetic risk to schizophrenia in the MHC is specific to a particular locus in the *C4A* gene (25).

In the field of genetics, candidate gene analyses have largely fallen out of favor owing to concerns about low power, false positives, low replication rates (30,31), insufficient biological knowledge to correctly identify plausible candidate genes, and the increasingly low cost of whole-genome array data. Yet candidate gene research continues in other fields, despite these issues. For example, a Google Scholar search performed on July 31, 2016, for *COMT* generated 3000 search results from 2016 alone, many of which are classic candidate gene association studies; the first five phenotypes from this search were circadian preferences (32), affective well-being across the lifespan (33), cognitive outcomes after electroconvulsive therapy (34), effect of opioid treatment on pain relief (35), and second-language learning in adults (36). It is of course possible that these and other traits are exceptions to what we now know about complex traits studied to date by GWASs, for which thousands of risk variants exist, each of which explains a very small amount of variation (37). Nevertheless, even if a given trait's genetic architecture is simple, our findings, as well as those of our colleagues (2,5,8,11), call into question the notion that scientists have been able to guess, a priori, which genes, much less which polymorphisms within those genes, will be relevant to any given trait. Given our inchoate understanding of the biological mechanisms underlying most complex traits, we suggest that future candidate gene studies should base gene choice not on historical precedent or proposed biological underpinnings, but on rigorous statistical evidence from the same or related traits, and that such studies should be sufficiently powered to detect effects on the order of those typically observed in modern GWASs.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was supported by National Institute on Alcohol Abuse and Alcoholism (NIAAA) Grant No. R01 AA017889 (to MAE), National Institute of Mental Health (NIMH) Grant No. R01 MH100141 (to MCK), Eunice Kennedy Shriver National Institute of Child Health and Human Development Grant No. T32 HD007289 (to ECJ), NIMH Grant No. T32 MH016880 (to RB), and NIAAA Grant No. T32 AA013525 (to WMG).

We thank the investigators who were part of the Psychiatric Genomics Consortium who put this data together, as well as Michelle Farris, Robin Hacker-Cary, Chris Klene, Valerie Slegesky, and Nicole Thatcher for helping with initial data collection and the many participants in these studies, all of whom helped make this study possible.

The authors report no biomedical financial interests or potential conflicts of interest.

ARTICLE INFORMATION

From the Department of Psychology and Neuroscience (ECJ, RB, MCK), Institute for Behavioral Genetics (ECJ, RB, MAE, MCK), and Department of Integrative Physiology (MAE), University of Colorado Boulder, Boulder, Colorado; Department of Neuroscience (WEM-G), The Scripps Research Institute, La Jolla, California; and the Department of Complex Trait Genetics (CAdL), Centre for Neurogenomics and Cognitive Research/VU University Amsterdam, Amsterdam; and Institute for Computing and Information Sciences (CAdL), Radboud University Nijmegen, Nijmegen, the Netherlands.

Address correspondence to Emma C. Johnson, Ph.D., Institute for Behavioral Genetics, 1480 30th St, Boulder, CO 80303; E-mail: emma.c.johnson@colorado.edu.

Received Dec 8, 2016; revised Jun 12, 2017; accepted Jun 28, 2017.

Supplementary material cited in this article is available online at http://dx.doi.org/10.1016/j.biopsych.2017.06.033.

REFERENCES

1. Sullivan PF, Kendler KS, Neale MC (2003): Schizophrenia as a complex trait. Arch Gen Psychiatry 60:1187–1192.

2. Allen NC, Bagade S, McQueen MB, Ioannidis JP, Kavvoura FK, Khoury MJ, *et al.* (2008): Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: The SzGene database. Nat Genet 40:827–834.
3. Cichon S, Craddock N, Daly M, Faraone SV, Gejman PV, Kelsoe J, *et al.* (2009): Genomewide association studies: History, rationale, and prospects for psychiatric disorders. Am J Psychiatry 166:540–556.
4. Wei J, Hemmings GP (2000): The NOTCH4 locus is associated with susceptibility to schizophrenia. Nat Genet 25:376–377.
5. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG (2001): Replication validity of genetic association studies. Nat Genet 29:306–309.
6. Mutsuddi M, Morris DW, Waggoner SG, Daly MJ, Scolnick EM, Sklar P (2006): Analysis of high-resolution HapMap of DTNBP1 (Dysbindin) suggests no consistency between reported common variant associations and schizophrenia. Am J Hum Genet 79:903–909.
7. Visscher PM, Brown MA, McCarthy MI, Yang J (2012): Five years of GWAS discovery. Am J Hum Genet 90:7–24.
8. Ripke S, Neale BM, Corvin A, Walters JTR, Farh K-H, Holmans PA, *et al.* (2014): Biological insights from 108 schizophrenia-associated genetic loci. Nature 511:421–427.
9. Collins AL, Kim Y, Sklar P, O'Donovan MC, Sullivan PF (2012): Hypothesis-driven candidate genes for schizophrenia compared with genome-wide association results. Psychol Med 42:607–616.
10. International Schizophrenia Consortium, Purcell SSM, Wray NRN, Stone JJL, Visscher PM, O'Donovan MC, Sullivan PF, *et al.* (2009): Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460:748–752.
11. Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I, *et al.* (2009): Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature 460:753–757.
12. Farrell MS, Werge T, Sklar P, Owen MJ, Ophoff RA, O'Donovan MC, *et al.* (2015): Evaluating historical candidate genes for schizophrenia. Mol Psychiatry 20:555–562.
13. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN (2003): Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 33:177–182.
14. Owen MJ, Craddock N, O'Donovan MC (2005): Schizophrenia: Genes at last? Trends Genet 21:518–525.
15. Sullivan PF (2005): The genetics of schizophrenia. PLoS Med 2:e212.
16. Harrison PJ, Weinberger DR (2005): Schizophrenia genes, gene expression, and neuropathology: On the matter of their convergence. Mol Psychiatry 10:40–68.
17. Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, *et al.* (2017): LD Hub: A centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. Bioinformatics 33:272–279.
18. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, *et al.* (2014): The NHGRI GWAS catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 42:D1001–D1006.
19. Ruano D, Abecasis GR, Glaser B, Lips ES, Cornelisse LN, de Jong APH, *et al.* (2010): Functional gene group analysis reveals a role of synaptic heterotrimeric G proteins in cognitive ability. Am J Hum Genet 86:113–125.
20. Lips ES, Cornelisse LN, Toonen RF, Min JL, Hultman CM, Holmans PA, *et al.* (2012): Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. Mol Psychiatry 17:996–1006.
21. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, *et al.* (2012): An integrated map of genetic variation from 1,092 human genomes. Nature 491:56–65.
22. De Leeuw CA, Mooij JM, Heskes T, Posthuma D (2015): MAGMA: Generalized gene-set analysis of GWAS data. PLoS Comput Biol 11:e1004219.
23. Mishra A, Macgregor S, Amos CI, Wang L-E, Lee JE, Gershenwald JE, *et al.* (2015): VEGAS2: Software for more flexible gene-based testing. Twin Res Hum Genet 18:86–91.
24. De Leeuw CA, Neale BM, Heskes T, Posthuma D (2016): The statistical properties of gene-set analysis. Nat Rev Genet 17:353–364.
25. Sekar A, Bialas AR, de Rivera H, Davis A, Hammond TR, Kamitaki N, *et al.* (2016): Schizophrenia risk from complex variation of complement component 4. Nature 530:177–183.
26. Ardlie KG, Deluca DS, Segrè AV, Sullivan TJ, Young TR, Gelfand ET, *et al.* (2015): The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science 348:648–660.
27. Millar JK, Wilson-Annan JC, Anderson S, Christie S, Taylor MS, Semple CA, *et al.* (2000): Disruption of two novel genes by a translocation co-segregating with schizophrenia. Hum Mol Genet 9: 1415–1423.
28. Won H, de la Torre-Ubieta L, Stein JL, Parikshak NN, Huang J, Opland CK, *et al.* (2016): Chromosome conformation elucidates regulatory relationships in developing human brain. Nature 538:523–527.
29. Gusev A, Mancuso N, Finucane HK, Reshef Y, Song L, Safi A, *et al.* (2016): Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights [published online ahead of print Aug 2]. bioRxiv.
30. Ioannidis JPA, Tarone R, McLaughlin JK (2011): The false-positive to false-negative ratio in epidemiologic studies. Epidemiology 22: 450–456.
31. Koenen KC, Duncan LE, Liberzon I, Ressler KJ (2013): From candidate genes to genome-wide association: The challenges and promise of posttraumatic stress disorder genetic studies. Biol Psychiatry 74: 634–636.
32. Jawinski P, Tegelkamp S, Sander C, Häntzsch M, Huang J, Mauche N, *et al.* (2016): Time to wake up: No impact of COMT Val158Met gene variation on circadian preferences, arousal regulation and sleep. Chronobiol Int 33:893–905.
33. Turan B, Sims T, Best SE, Carstensen LL (2016): Older age may offset genetic influence on affect: The COMT polymorphism and affective well-being across the life span. Psychol Aging 31:287–294.
34. Bennett DM, Currie J, Fernie G, Perrin JS, Reid IC (2016): Differences in cognitive outcomes after ECT depending on BDNF and COMT polymorphisms. J ECT 32:243–250.
35. Elens L, Elisabeth N, Maja M, Rane A, Fellman V, van Schaik RH (2016): Genetic predisposition to poor opioid response in preterm infants: Impact of KCNJ6 and COMT polymorphisms on pain relief after endotracheal intubation. Ther Drug Monit 38:525–533.
36. Mamiya PC, Richards TL, Coe BP, Eichler EE, Kuhl PK, Geschwind DH, *et al.* (2016): Brain white matter structure and COMT gene are linked to second-language learning in adults. Proc Natl Acad Sci U S A 113:7249–7254.
37. Sullivan PF, Daly MJ, O'Donovan M (2012): Genetic architectures of psychiatric disorders: The emerging picture and its implications. Nat Rev Genet 13:537–551.