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.