2011

Comprehensive Gene-Based Association Study of a Chromosome 20 Linked Region Implicates Novel Risk Loci for Depressive Symptoms in Psychotic Illness

T. Bernard Bigdeli  
*Virginia Commonwealth University*, bigdelitb@vcu.edu

Brion S. Maher  
*Virginia Commonwealth University*, bsmaher@vcu.edu

Zhongming Zhao  
*Virginia Commonwealth University*, zzhao@vcu.edu

*See next page for additional authors*

Follow this and additional works at: [http://scholarscompass.vcu.edu/hgen_pubs](http://scholarscompass.vcu.edu/hgen_pubs)

Part of the [Medical Sciences Commons](http://scholarscompass.vcu.edu/hgen_pubs)

Copyright: © 2011 Bigdeli et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Downloaded from [http://scholarscompass.vcu.edu/hgen_pubs/8](http://scholarscompass.vcu.edu/hgen_pubs/8)

This Article is brought to you for free and open access by the Dept. of Human and Molecular Genetics at VCU Scholars Compass. It has been accepted for inclusion in Human and Molecular Genetics Publications by an authorized administrator of VCU Scholars Compass. For more information, please contact [libcompass@vcu.edu](mailto:libcompass@vcu.edu).
Comprehensive Gene-Based Association Study of a Chromosome 20 Linked Region Implicates Novel Risk Loci for Depressive Symptoms in Psychotic Illness

T. Bernard Bigdeli¹,², Brion S. Maher¹,²,³,⁴, Zhongming Zhao¹,²,³,⁵, Edwin J. C. G. van den Oord²,⁶, Dawn L. Thiselton²,³, Jingchun Sun⁵, Bradley T. Webb²,⁶, Richard L. Amdur⁷,¹¹, Brandon Wormley²,³, Francis A. O’Neill⁹, Dermot Walsh¹⁰, Brien P. Riley¹,²,³, Kenneth S. Kendler¹,²,³, Ayman H. Fanous²,³,⁷,⁸,¹¹

¹Department of Human and Molecular Genetics, Virginia Commonwealth University, Richmond, Virginia, United States of America, ²Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, Virginia, United States of America, ³Department of Psychiatry, Virginia Commonwealth University, Richmond, Virginia, United States of America, ⁴Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America, ⁵Departments of Psychiatry, Biomedical Informatics, and Cancer Biology, Vanderbilt University Medical Center, Vanderbilt, Tennessee, United States of America, ⁶Center for Biomarker Research and Personalized Medicine, Virginia Commonwealth University, Richmond, Virginia, United States of America, ⁷Department of Psychiatry, Keck School of Medicine of the University of Southern California, Los Angeles, California, United States of America, ⁸Department of Psychiatry, Queens University, Belfast, United Kingdom, ⁹The Health Research Board, Dublin, Ireland, ¹⁰Department of Psychiatry, Georgetown University School of Medicine, Washington, D. C., United States of America, ¹¹Department of Psychiatry, Virginia Commonwealth University, Richmond, Virginia, United States of America.

Abstract

Background: Prior genomewide scans of schizophrenia support evidence of linkage to regions of chromosome 20. However, association analyses have yet to provide support for any etiologically relevant variants.

Methods: We analyzed 2988 LD-tagging single nucleotide polymorphisms (SNPs) in 327 genes on chromosome 20, to test for association with schizophrenia in 270 Irish high-density families (ISHDSF, N = 270 families, 1408 subjects). These SNPs were genotyped using an Illumina iSelect genotyping array which employs the Infinium assay. Given a previous report of novel linkage with chromosome 20p using latent classes of psychotic illness in this sample, association analysis was also conducted for each of five factor-derived scores based on the Operational Criteria Checklist for Psychotic Illness (delusions, hallucinations, mania, depression, and negative symptoms). Tests of association were conducted using the PDTPHASE and QPDPHASE packages of UNPHASED. Empirical estimates of gene-wise significance were obtained by adaptive permutation of a) the smallest observed P-value and b) the threshold-truncated product of P-values for each locus.

Results: While no single variant was significant after LD-corrected Bonferroni-correction, our gene-dropping analyses identified loci which exceeded empirical significance criteria for both gene-based tests. Namely, R3HDL and C20orf39 are significantly associated with depressive symptoms of schizophrenia (P_{emp} < 2 \times 10^{-5}) based on the minimum P-value and truncated-product methods, respectively.

Conclusions: Using a gene-based approach to family-based association, R3HDL and C20orf39 were found to be significantly associated with clinical dimensions of schizophrenia. These findings demonstrate the efficacy of gene-based analysis and support previous evidence that chromosome 20 may harbor schizophrenia susceptibility or modifier loci.

Introduction

With a lifetime prevalence of 1 percent and an estimated annual cost of $62.7 billion in the United States [1], schizophrenia (Scz) is a debilitating neuropsychiatric disorder which poses a significant burden to public health. Whether schizophrenia represents a single or multiple disease processes is a source of persistent controversy, as patients vary considerably in onset, course and outcome of disease, and the particular combination of symptoms endorsed [2,3]. Models comprising continuous traits—often extracted in factor analysis of symptom profiles—have been adduced, typically distinguishing positive, negative, disorganization, and affective symptoms [4]. One explanation for this variability lies in the existence of more than one putative etiopathogenic mechanism, each imparting susceptibility to a more or less distinct disease subtype or influencing the character of...
illness dimensionally. Detection and subsequent replication of several putative risk variants, facilitated by genome-wide association studies (GWAS) [5–10], has seen renewed interest in this question among geneticists and diagnosticians alike [11–13].

Consistent with the observed variability in clinical presentation is the hypothesis that schizophrenia is likely genetically heterogeneous [6,8]. Linkage and candidate gene association studies have implicated a number of genes and genomic regions, with varying degrees of subsequent independent replication. Allelic heterogeneity has been demonstrated in meta-analyses of candidate genes such as DTNBP1 [14,15]. If the observed clinical heterogeneity of schizophrenia is in fact due to genetic heterogeneity, the use of more clinically homogenous phenotypes may increase the signal-to-noise ratio in gene-finding studies. A previous report by our group described detection of novel linkage to 20p using latent classes of psychotic illness [16]. Linkage analysis of Mania, Schizomania, Deficit Syndrome and Core Schizophrenia latent classes yielded several suggestively significant loci, in regions of chromosome 20 which had previously yielded very little evidence of linkage in our sample. Furthermore, the presence of susceptibility genes in chromosome 20 has been suggested by several previous linkage studies as well [17–23]. In addition to genes which increase susceptibility to more or less distinct clinical subtypes of illness, other genes may influence clinical features of disease in a dimensional fashion, without altering liability to the illness itself. These have previously been described as modifier loci [2]. Modifier loci may not be resolvable using traditional dichotomous phenotypes (simply “affected” or “unaffected”), but rather, by quantitative symptomatic measures. Several examples have been reported [24–30].

Recent GWAS of schizophrenia support a polygenic model in which potentially thousands of common variants individually impart small effects. Given the unprecedented multiple-comparison burden incurred in a genome-wide approach, hypothesis-based strategies remain viable alternatives for the study of complex disease. A gene-based approach is particularly convenient. In an analysis of bipolar and schizophrenia datasets, Moskvina and colleagues [31] observed significantly more SNPs within genes to specific patterns of linkage disequilibrium.

We describe a comprehensive, gene-based association survey of 327 genes in regions linked to chromosome 20 in our previous studies. In addition to testing for association with traditional diagnostic definitions of schizophrenia, we also sought to assess whether chromosome 20 harbors modifier loci. Association analysis was therefore also performed for five factor-derived scores, representing hallucinations, delusions, depressive symptoms, manic symptoms, and negative symptoms, in schizophrenia cases only. In addition to single-marker tests of allelic association, we employ two gene-based test-statistics, the minimum observed P-value per gene and the truncated product of P-values, to evaluate the efficacy of a gene-based approach as applied to a large, family-based study.

Results

Gene-wide Association Analyses

Following quality-control protocols, 2,988 single nucleotide polymorphisms in 327 genes were tested for association with a diagnosis of schizophrenia (Figure 1). Estimates of empiric significance (P_{emp}) were obtained via an adaptive permutation procedure employing the smallest observed P-value, as well as the truncated product of P-values (P_{trunc} ≤ 0.01) per gene. The number of genes carried forward in successive stages of this procedure, in both approaches, can be found in Table 1. Using the minimum gene-wide P-value approach, no genes were observed to be significantly associated with narrow (N = 1574), intermediate (N = 1749), or broad (N = 1808) diagnoses of schizophrenia. Next, we sought to identify those SNPs associated with clinical dimensions of schizophrenia in a subset of cases (N = 721) for which the OPCRIT was available. A previous report by Fanous and colleagues [16] supports linkage of latent classes derived from the OPCRIT to chromosome 20 in this sample. No genes were found to be significantly associated with the negative, manic, hallucinations or delusions factors. In the analysis of the clinical dimensions, R3HDML demonstrated significant evidence of association (P_{emp} < 2 × 10^{-5}) with the depressive factor using the minimum P-value approach. Using the truncated product of P-values, C20orf39 was also found to be significantly associated with the depressive factor (P_{trunc} < 2 × 10^{-5}). It is important to note that, for both C20orf39 and R3HDML, we observed fewer than ten simulated results more significant than the observed test-statistic after 100,000 permutations. Hence, our estimates of empirical significance may be conservative. However, extending our analyses to IAT permutations was not carried out as it was too computationally demanding.

Because validation of a truncated product approach in extended pedigrees relies on the permutation procedure faithfully conserving patterns of LD within each replicate dataset, we obtained a quantitative measure of how well haplotype-block structure was maintained for C20orf39 across actual and simulated datasets. In calculating an LD-corrected significance threshold, SNPSpD estimates the effective number of independent tests present in a set of markers. Using SNPSpD, 1,000 replicate datasets for C20orf39 were assessed for number of independent tests. When compared to the estimate based on the actual pattern of LD in C20orf39 (i.e., 26 independent tests), the distribution of these simulation-derived estimates demonstrates that the LD structure within each replicate does not differ significantly from the observed data (P = 0.409; 95% CI: [26,28]). This increases confidence in the truncated product finding for C20orf39. However, this may not hold for every gene and may be sensitive to specific patterns of linkage disequilibrium.

Single Marker Association Analysis

Taking each SNP to represent an independent hypothesis but correcting for LD using SNPSpD, we found that no single marker met experiment-wide criteria for association (\(\text{SNP}_{\text{PD}} < 3.18 \times 10^{-5}\)) with either the three categorical diagnostic definitions used or our OPCRIT-derived factor scores (Tables 2, 3). The strongest evidence of association with a diagnosis of schizophrenia was in PLCB1 (rs1608205, P = 1.00 × 10^{-3}, intermediate Scz diagnosis). For the depressive factor, we observed the strongest associations experiment-wide at 20q13.12 (rs3761184, P = 3.31 × 10^{-5}) in R3HDML. This was very close to the LD-corrected significance threshold calculated using SNPSpD (P = 3.18 × 10^{-5}). Furthermore, rs11700002, in C20orf39 at 20p11.21 attained P = 1.01 × 10^{-4}.

Discussion

We have conducted a comprehensive gene-based association study of 327 genes on chromosome 20 in an Irish sample of 270 high-density schizophrenia families. This study sought to identify common variants conferring susceptibility to schizophrenia, following up reported linkage in this sample to clinical subtypes of psychotic illness [16], as well as previous studies reporting linkage to chromosome 20. Because those clinical subtypes were
derived from quantitative symptom dimensions, we also tested for association with these same dimensions. Although traditional single-marker tests failed to identify any SNPs meeting experiment-wide criteria for significance, application of gene-wide association metrics revealed two previously unimplicated loci, \textit{R3HDML} and \textit{C20orf39}, associated with depressive symptoms. Our findings support the power of gene-based association approaches. They also lend further support to previous evidence suggesting that genetic differences may underlie clinical heterogeneity in schizophrenia [2,3].

One of the aims of this study was to identify genomic loci predisposing to a particular form of illness or which modifies clinical presentation amongst affected individuals. Such genes have been described previously as “modifier” or “susceptibility-modifier” loci and are reviewed elsewhere [2]. Of the two loci showing the strongest associations, namely \textit{R3HDML} and \textit{C20orf39}, neither appears to affect the risk of the illness itself. That is, no single variant in either gene met even nominal significance criteria ($P < 0.05$) for association with narrow, intermediate, or broad diagnoses of schizophrenia. These two genes would therefore fulfill our definition of modifier genes [2]. However, the strength of evidence we observed for \textit{R3HDML} is greater than that observed for \textit{C20orf39}. \textit{R3HDML} was identified by application of the minimum $P$-value approach. Among affected individuals, those carrying the minor allele (G) of the corresponding SNP, rs3761184, had higher mean depression scores. On the other hand, for \textit{C20orf39}, empirical significance was attained using the truncated product of $P$-values. This makes it more difficult to identify a specific risk genotype. This is because the truncated product method only considers all variation within a gene jointly. In Figure 2, it is apparent that those markers contributing to the truncated product for \textit{C20orf39} comprise a block of LD distinct

Figure 1. Physical distribution of single-marker associations on chromosome 20, for both categorical diagnoses and clinical dimensions of Scz. Associations are displayed as log-transformed $P$-values ($-\log_{10}P$) at genomic positions in megabases (Mb). Where appropriate, a dotted line indicates the Bonferroni-corrected significance threshold, accounting for number of SNPs assayed experiment-wide. Similarly, a dashed line indicates the LD-corrected significance threshold, as estimated by SNPSpD.

doi:10.1371/journal.pone.0021440.g001

\textit{Novel Chr20 Associations with Depression in Scz}
from the surrounding region, with the majority showing association of the minor allele with higher depression scores. Whereas individually, none of the single-marker associations were significant to be considered significant at a particular stage of permutation, there may be no greater than 10 simulated null statistics which are more extreme than the observed. For each gene-based test, the number of permutations performed at each stage is displayed parenthetically (100, 1,000, 10,000, and 100,000). Dim codes “del”, “dep”, “hal”, “manic”, “neg” are delusions, depressive symptoms, hallucinations, mania, and negative symptoms, respectively, and described elsewhere in full.

For each diagnosis and symptom factor, the number of loci requiring additional permutations after each stage of our adaptive procedure, given for both the Min P and truncated-product methods. For an observed test-statistic to be considered significant at a particular stage of permutation, there may be no greater than 10 simulated null statistics which are more extreme than the observed. For each gene-based test, the number of permutations performed at each stage is displayed parenthetically (100, 1,000, 10,000, and 100,000). Dim codes “del”, “dep”, “hal”, “manic”, “neg” are delusions, depressive symptoms, hallucinations, mania, and negative symptoms, respectively, and described elsewhere in full.

1Min P finding for R3HDML.
2Truncated-product finding for C20orf39.

Table 1. Number of genes requiring additional simulations at each stage of adaptive permutation.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Min P (100)</th>
<th>Min P (1 K)</th>
<th>Min P (10 K)</th>
<th>Min P (100 K)</th>
<th>Trunc Prod P (100)</th>
<th>Trunc Prod P (1 K)</th>
<th>Trunc Prod P (10 K)</th>
<th>Trunc Prod P (100 K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic Category</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Narrow</td>
<td>19</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Int</td>
<td>12</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Broad</td>
<td>14</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Symptom Factor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>del</td>
<td>15</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>dep</td>
<td>10</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>9</td>
<td>4</td>
<td>2</td>
<td>1^1</td>
</tr>
<tr>
<td>hal</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>manic</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>neg</td>
<td>20</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

For each diagnosis and symptom factor, the number of loci requiring additional permutations after each stage of our adaptive procedure, given for both the Min P and truncated-product methods. For an observed test-statistic to be considered significant at a particular stage of permutation, there may be no greater than 10 simulated null statistics which are more extreme than the observed. For each gene-based test, the number of permutations performed at each stage is displayed parenthetically (100, 1,000, 10,000, and 100,000). Dim codes “del”, “dep”, “hal”, “manic”, “neg” are delusions, depressive symptoms, hallucinations, mania, and negative symptoms, respectively, and described elsewhere in full.

1Min P finding for R3HDML.
2Truncated-product finding for C20orf39.

doi:10.1371/journal.pone.0021440.t001

Table 2. Top ten Pedigree Disequilibrium Test results for categorical diagnoses of Schizophrenia.

<table>
<thead>
<tr>
<th>Chr/Mb</th>
<th>Gene</th>
<th>dbSNP</th>
<th>Nuc.(Minor)</th>
<th>Assoc.</th>
<th>Frq_assoc</th>
<th>Trios(Tr/NTr)</th>
<th>DSPs(Aff/Unaff)</th>
<th>Z-score</th>
<th>x^2</th>
<th>P</th>
<th>Dx</th>
</tr>
</thead>
<tbody>
<tr>
<td>20/0.92</td>
<td>R5PO4</td>
<td>rs6056462</td>
<td>A/G(G)</td>
<td>A</td>
<td>0.842</td>
<td>94/90</td>
<td>868/855</td>
<td>3.171</td>
<td>10.05</td>
<td>1.52 x 10^-3</td>
<td>Int</td>
</tr>
<tr>
<td>20/0.92</td>
<td>R5PO4</td>
<td>rs6056462</td>
<td>A/G(G)</td>
<td>A</td>
<td>0.843</td>
<td>97/93</td>
<td>911/897</td>
<td>3.170</td>
<td>10.05</td>
<td>1.53 x 10^-3</td>
<td>Broad</td>
</tr>
<tr>
<td>20/0.87</td>
<td>PLCB1</td>
<td>rs6108205</td>
<td>C/T(C)</td>
<td>C</td>
<td>0.491</td>
<td>67/65</td>
<td>536/467</td>
<td>3.289</td>
<td>10.82</td>
<td>1.00 x 10^-3</td>
<td>Int</td>
</tr>
<tr>
<td>20/0.87</td>
<td>PLCB1</td>
<td>rs6108205</td>
<td>C/T(C)</td>
<td>C</td>
<td>0.490</td>
<td>58/56</td>
<td>444/391</td>
<td>3.182</td>
<td>10.13</td>
<td>1.46 x 10^-3</td>
<td>Narrow</td>
</tr>
<tr>
<td>20/10.60</td>
<td>JAG1</td>
<td>rs6133987</td>
<td>C/T(T)</td>
<td>C</td>
<td>0.778</td>
<td>85/78</td>
<td>678/610</td>
<td>3.193</td>
<td>10.19</td>
<td>1.41 x 10^-3</td>
<td>Narrow</td>
</tr>
<tr>
<td>20/0.41</td>
<td>PPTRT</td>
<td>rs6072690</td>
<td>A/G(A)</td>
<td>A</td>
<td>0.442</td>
<td>67/59</td>
<td>586/502</td>
<td>3.224</td>
<td>10.40</td>
<td>1.26 x 10^-3</td>
<td>Broad</td>
</tr>
<tr>
<td>20/0.51</td>
<td>PPTRT</td>
<td>rs6130134</td>
<td>C/T(T)</td>
<td>C</td>
<td>0.792</td>
<td>94/93</td>
<td>880/826</td>
<td>3.152</td>
<td>9.940</td>
<td>1.62 x 10^-3</td>
<td>Int</td>
</tr>
<tr>
<td>20/4.27</td>
<td>TOX2</td>
<td>rs6103560</td>
<td>C/T(C)</td>
<td>C</td>
<td>0.705</td>
<td>111/95</td>
<td>855/822</td>
<td>3.200</td>
<td>10.24</td>
<td>1.37 x 10^-3</td>
<td>Broad</td>
</tr>
<tr>
<td>20/4.19</td>
<td>CD40</td>
<td>rs3765457</td>
<td>A/G(G)</td>
<td>G</td>
<td>0.175</td>
<td>24/24</td>
<td>235/170</td>
<td>3.154</td>
<td>9.950</td>
<td>1.33 x 10^-3</td>
<td>Broad</td>
</tr>
<tr>
<td>20/4.19</td>
<td>CD40</td>
<td>rs3765457</td>
<td>A/G(G)</td>
<td>G</td>
<td>0.176</td>
<td>23/23</td>
<td>223/160</td>
<td>3.210</td>
<td>10.30</td>
<td>1.61 x 10^-3</td>
<td>Int</td>
</tr>
</tbody>
</table>

For each gene, Chr/Mb denotes chromosome and genomic position (Mega bases), dbSNP is the rs-identifier for the assayed SNP, and Nuc is the nucleotide substitution at a SNP. Frq_assoc and Z-scores are with respect to the associated allele. Allelic transmissions from parent to affected child is given by Trios, where Tr and NTr represent number of transmissions and non-transmissions. Allele-sharing between phenotypically-discordant sib-pairs is given by DSPs, with Aff/Unaff denoting the number of associated alleles in affected and unaffected siblings, respectively. Dx codes “Core”, “Int”, and “Broad” are Core, Intermediate, and Broad diagnoses of schizophrenia, respectively, and described elsewhere in full.

doi:10.1371/journal.pone.0021440.t002
Novel Chr20 Associations with Depression in Scz

Table 3. Top ten Quantitative Pedigree Disequilibrium Test results for clinical dimensions of Schizophrenia.

<table>
<thead>
<tr>
<th>Chr/Mb</th>
<th>Gene</th>
<th>dbsSNP</th>
<th>Nuc.(Minor)</th>
<th>Assoc.</th>
<th>Frq_assoc</th>
<th>Gametes (maj/min)</th>
<th>Z-score</th>
<th>x²</th>
<th>P</th>
<th>Dim</th>
</tr>
</thead>
<tbody>
<tr>
<td>20/0.83</td>
<td>ANGPT4</td>
<td>rs976166</td>
<td>C/G(C)</td>
<td>C</td>
<td>0.308</td>
<td>1285/571</td>
<td>3.333</td>
<td>11.11</td>
<td>8.59×10⁻⁴</td>
<td>neg</td>
</tr>
<tr>
<td>20/0.70</td>
<td>PAX7</td>
<td>rs2327225</td>
<td>A/C(C)</td>
<td>A</td>
<td>0.718</td>
<td>1333/523</td>
<td>3.426</td>
<td>11.74</td>
<td>6.13×10⁻⁴</td>
<td>manic</td>
</tr>
<tr>
<td>20/10.59</td>
<td>JAG1</td>
<td>rs1633986</td>
<td>A/G(A)</td>
<td>A</td>
<td>0.089</td>
<td>1691/165</td>
<td>3.523</td>
<td>12.41</td>
<td>4.27×10⁻⁴</td>
<td>manic</td>
</tr>
<tr>
<td>20/16.66</td>
<td>SNRPB2</td>
<td>rs6111262</td>
<td>C/T(T)</td>
<td>C</td>
<td>0.720</td>
<td>1337/519</td>
<td>3.281</td>
<td>10.76</td>
<td>1.04×10⁻³</td>
<td>del</td>
</tr>
<tr>
<td>20/17.58</td>
<td>RRPB1</td>
<td>rs3790310</td>
<td>A/T(A)</td>
<td>T</td>
<td>0.809</td>
<td>1502/354</td>
<td>3.451</td>
<td>11.91</td>
<td>5.58×10⁻⁴</td>
<td>neg</td>
</tr>
<tr>
<td>20/24.54</td>
<td>C20orf39</td>
<td>rs11700002</td>
<td>A/G(A)</td>
<td>A</td>
<td>0.187</td>
<td>1509/347</td>
<td>3.930</td>
<td>15.44</td>
<td>8.50×10⁻⁴</td>
<td>dep</td>
</tr>
<tr>
<td>20/24.56</td>
<td>C20orf39</td>
<td>rs4815292</td>
<td>G/T(G)</td>
<td>G</td>
<td>0.353</td>
<td>1201/655</td>
<td>3.362</td>
<td>11.30</td>
<td>7.73×10⁻⁴</td>
<td>dep</td>
</tr>
<tr>
<td>20/24.58</td>
<td>C20orf39</td>
<td>rs11696125</td>
<td>G/T(T)</td>
<td>T</td>
<td>0.230</td>
<td>1430/426</td>
<td>3.392</td>
<td>11.50</td>
<td>6.94×10⁻⁴</td>
<td>dep</td>
</tr>
<tr>
<td>20/24.58</td>
<td>C20orf39</td>
<td>rs11067473</td>
<td>A/G(A)</td>
<td>A</td>
<td>0.230</td>
<td>1430/426</td>
<td>3.392</td>
<td>11.50</td>
<td>6.94×10⁻⁴</td>
<td>dep</td>
</tr>
<tr>
<td>20/42.40</td>
<td>R3HDML</td>
<td>rs3761184</td>
<td>A/G(G)</td>
<td>G</td>
<td>0.169</td>
<td>1543/313</td>
<td>4.120</td>
<td>16.98</td>
<td>3.78×10⁻³</td>
<td>dep</td>
</tr>
</tbody>
</table>

For each gene, Chr/Mb denotes chromosome and genomic position (Megabases), dbsSNP is the rs-identifier for the assayed SNP, and Nuc is the nucleotide substitution at a SNP. Gametes represents the number of major and minor alleles (maj/min) transmitted from parent to affected child or unshared between phenotypically-discordant sib-pairs. Frq_assoc and Z-scores are with respect to the allele corresponding to a higher trait mean. Dim codes “del”, “dep”, “hal”, “manic”, “neg” are delusions, depressive symptoms, hallucinations, mania, and negative symptoms, respectively, and described elsewhere in full.

doi:10.1371/journal.pone.0021440.t003

R3HDML, rs3761184, falls just upstream (<50 bp) of the second exon and 150 bp downstream of one such repeat-rich region. Roles in fertilization, spermatogenesis, and pathogen response have all been proposed for CRISP proteins, but these mechanisms are not immediately supportive of R3HDML as a schizophrenia candidate gene. However, recent implication of a number of HLA genes in large-scale GWAS suggest that genes involved in immune-related mechanisms, such as pathogen response, could be reasonable Scz candidates [8]. The presence of specific sequence features in the vicinity of the associated SNP may warrant more thorough bioinformatic inquiry. Additionally, R3HDML lies approximately 57 kb downstream of the GDAP1L1 locus, which appears to encode a gluthionine S-transferase (GST). Cell-culture studies have demonstrated a relationship between gluthionine deficiency and oxidative stress, mechanisms frequently purported to contribute to schizophrenia pathophysiology [33,34]. However, GDAP1L1 was not significantly associated.

Our empirically significant finding for C20orf39 presents additional challenges for interpretation, given its provisional status as an “open reading frame”. Provisionally known as TMEM90B, this locus encodes a predicted transmembrane protein. Of 33 SNPs assayed within C20orf39, the nine included in the truncated product bounded a region of LD corresponding to the coding region of C20orf39. The upstream, untranslated region of C20orf39, which itself corresponds to a distinct set of ESTs, yielded no SNPs meeting local significance criteria. Whether the markers driving this association simply lie in joint linkage disequilibrium with nearby causal variation, or actually demarcate an etiologically relevant genomic region, is unknown.

Depressive symptoms, especially suicidal ideation, comprise a considerable portion of morbidity and mortality in schizophrenia [35]. Therefore, follow up of these two genes could be important in the search for clues to more successful identification and treatment of this clinical dimension.

As demonstrated by Moskvina et al., polymorphisms mapping to functional elements are more likely to be associated with complex disease than intergenic variation [31]. Despite ongoing annotation and characterization of functional elements, however, our knowledge of genomic variation, functional or otherwise, remains incomplete. This is exemplified by C20orf39 and R3HDML, which are novel and unannotated.

A major benefit of gene-based approaches is that they are robust to allelic and haplotype heterogeneity across samples. This makes them particularly suited for use in replication and meta-analysis. In traditional replication of single-marker associations, the associated SNP in the discovery sample is usually assayed in all subsequent replication samples. This could inflate Type-II error in the presence of population differences in haplotype structure and allele frequencies [36]. Complex patterns of associations, whether spurious or due to genetic heterogeneity, have been more the rule rather than the exception in candidate gene studies of complex disease, as demonstrated by studies of DTNBP1 [14,15]. For discovery-based approaches, adoption of a gene-based strategy may be of even more immediate benefit, specifically by providing a straightforward means of multiple-test correction. Furthermore, traditional methods to correct for multiple-testing, such as Bonferroni correction or the less overtly conservative SNPSpD method, may be less robust in detecting small genetic effects. However, in spite of the advantages of gene-based association studies intergenic causative variants or variants in unrecognized genes might have been missed in this study.

Given the poor spatial resolution of linkage and intrinsic differences between these methodologies, we are currently unable to fully relate our association findings with the results of our previously published linkage study of latent classes. However, it is notable that R3HDML is located in a region which was linked to the “deficit syndrome” latent class, for which members were substantially more likely to fall below the median for depressive symptoms. Despite failing to demonstrate any evidence of association with a diagnosis of schizophrenia, R3HDML may be associated with a disease subtype characterized by low levels of depression. Because subtyping precludes use of our full sample for association analysis, statistical power is insufficient to test this hypothesis. Other methods aiming to identify more clinically homogenous subgroups have been applied to linkage analysis of schizophrenia. In a study of 168 affected sibling pairs, Hamshere and colleagues [37] demonstrated that inclusion of major depression as a covariate yielded suggestive evidence of linkage at 20q11.21, while schizophrenia as a whole did not. Taken together, these studies are compelling in their support of 20q11 harboring genes relevant to the affective component of schizophrenia. Emerging evidence supports a role for genetic variants...
conferring risk of both schizophrenia and bipolar disorder [8,38]. Furthermore, genome scans of both disorders have consistently implicated regions of chromosome 20 [39–44]. A recent study of 383 bipolar or schizoaffective relative pairs found suggestive linkage at 20q13.31 when conditioning on the presence of mood-incongruent psychosis, furthering the argument that chromosome 20 loci may have relevance to conditions containing admixtures of mood and psychotic symptoms [45].

The findings presented here provide additional support to published findings suggesting that schizophrenia modifier loci may exist on chromosome 20 and, more generally, that genetic differences underlie clinical heterogeneity in schizophrenia [46]. We await replication of the observed associations between these loci and either categorically defined illness or more or less distinct subtypes or clinical dimensions. There are two main limitations relevant to this study. First, the truncated product of \( P \)-values is particularly sensitive to patterns of LD (unpublished results), since markers could be significant only due to their LD with other significant markers. Applied to family-based analysis of extended pedigrees, the validity of gene-based testing relies on the permutation method realistically maintaining LD across simulated datasets. As discussed, for \( C2orf39 \), the LD structure for a random sample of simulated datasets did not differ significantly from the actual data \((P>0.05)\). Second, our analysis of multiple symptom dimensions may increase the Type-I error rate due to multiple testing. However, as we have previously shown, these dimensions are correlated [47], making Bonferroni correction overly conservative. It remains unclear whether the failure of traditional approaches to detect experiment-wide significant loci reflects the spurious nature of these findings or simply the limited power of this sample. Ultimately, the genotype-phenotype correlations reported herein require confirmation in independent samples for which comparable symptom measures are available. We are unaware of other family-based schizophrenia samples in which OPCRIT data are readily available. However, this is likely to be attempted in case-control samples by the Psychiatric GWAS Consortium Cross-Disorders Group [13].

**Methods**

**Ethics Statement**

This research was approved by the Institutional Review Boards of Virginia Commonwealth University School of Medicine and the

---

**Figure 2. Association of \( C2orf39 \) SNPs with depressive symptoms of Scz.** Magnitudes and directions of associations are displayed in the upper panel, with upwards-oriented triangles indicating a positive correlation with symptom factor score. A dashed line is provided at the inclusion threshold for the truncated product of \( P \)-values. Connecting lines relate the physical positions of associations to SNP labels in the corresponding LD-map \( (r^2) \). Plot generated using snp.plotter for R [66].

doi:10.1371/journal.pone.0021440.g002
Sample

Fieldwork for the Irish Study of High Density Schizophrenia Families (ISHDSF) was conducted between April 1987 and November 1992, with probands ascertained from public psychiatric hospitals in Ireland and Northern Ireland [46]. Selection criteria were two or more first-degree relatives meeting DSM-III-R criteria for schizophrenia or poor-outcome schizoaffective disorder (PO-SAD). Diagnoses were based on the Structured Interview for DSM-III-R Diagnosis (SCID) [49]. Independent review of all pertinent diagnostic information was made blind to pedigree assignment and marker genotypes by KSK and DW, with each diagnostician making up to three best-estimate DSM-III-R diagnoses. The Operational Criteria Checklist for Psychotic Illness (OPCRIT) [50] was completed by KSK for all subjects with probable lifetime histories of hallucinations or delusions (N = 755; N = 722 genotyped). Our diagnostic schema contains 4 concentric definitions of affection: narrow (D2) (schizophrenia, PO-SAD, and simple schizophrenia) (N = 577), intermediate (D5) which adds to D2 schizotypal personality disorder, schizophreniaiform and delusional disorders, atypical psychosis and good-outcome SAD (N = 700), broad (D8) (all disorders which significantly aggregated in relatives of probands) (N = 754) and very broad (D9), including any psychiatric illness (N = 961). Exploratory and confirmatory factor analysis of the OPCRIT was conducted previously by Fanous et al. [47]. This yielded a five-factor solution, comprising depressive, manic, and negative symptoms, delusions and hallucinations. Factor-derived scores were obtained by summing the scores of all items belonging to each factor.

Bioninformatics and SNP-selection

Using WebGestalt [51], a total of 378 genes were initially identified as mapping to the region of chromosome 20 corresponding to the peak SNP and to the positions corresponding to a SNP of at least 1 on either side, based on the Illumina version 4.0 linkage SNP map used for genotyping in a multicenter linkage study funded by R01-MH068881 [52]. While there was very little evidence of linkage in our published microsatellite-based scan [53], we did observe modest evidence using the map in the Holmans et al. study [52], which included our study sample (results available on request). We included predicted genes and open reading frames (ORFs) from the p-terminal to 45.85 Mb (20q13.13). Physical map positions for 362 genes were obtained from the UCSC Genome Browser [hg17/NCBI Build 35] [54]. Tagging SNPs were selected for each identified genomic region (excluding upstream and downstream regions of genes) using Tagger ($r^2 \geq 0.8$, minor allele frequency (MAF) $\geq 0.1$) [55], as applied to the HapMap CEPH dataset [56]. Of these, 31 genes were excluded on the basis of frequency (MAF) ($< 0.1$) [55], as applied to the HapMap CEPH dataset [56]. Empirical significance was calculated from the proportion of simulated gene-wise test statistics more significant than the actual genotypes, with individual phenotypes and pedigree structure also preserved within each simulated dataset. For markers in linkage disequilibrium ($r^2 \geq 0.1$), alleles were simulated using the haplotype frequencies for the marker clusters. To reduce computation time, those pedigrees of complexity greater than 70 bits were omitted from calculation of allele and haplotype frequencies. Each simulated dataset was analyzed as described above in two ways: retaining the minimum P-value per gene, as well as calculating the threshold-truncated product of P-values ($\prod_{k=1}^{n_{perm}} P_{k} \leq 0.01$) per gene. For the set of single-SNP hypotheses corresponding to a gene, the truncated product method considers the product of only those $P$-values falling below a specified threshold, evaluating the probability of observing as significant a product by chance. Whereas Fisher’s Combined Test assesses the overall evidence for departure from the null, the truncated product approach can be used to assess whether suggestive or significant findings are truly significant [64]. Previous reports support the use of a truncated product approach in conjunction with the PDT [65]. Empirical significance was calculated from the proportion of simulated gene-wise test statistics more significant than the actual results ($\prod_{k=1}^{n_{perm}} P_{k} \leq 0.01$). We used an adaptive permutation procedure, by which empirical $P$-values were obtained for 100, 1,000, 10,000, and 100,000 simulations. Only those observed associations for which there were not at least ten more significant simulated results were carried forward to each successive stage of permutation analysis.

Association Analysis

We performed association analysis for categorical diagnoses of schizophrenia using PDTPHASE (UNPHASED v. 2.404), an implementation of the pedigree disequilibrium test (PDT) with extensions to deal with uncertain haplotypes and missing data [58,59]. The PDT is an extension of the transmission disequilibrium test (TDT) to examine general pedigree structures and is similarly a test of association in the presence of linkage. Association with quantitative measures of disease was assessed using QPDTPHASE (UNPHASED v. 2.404), an implementation of the quantitative trait PDT with extensions to deal with uncertain haplotypes and missing data [59,60]. An LD-corrected significance threshold was obtained using the SNPSpD package for R [61,62]. For 2,988 SNPs, SNPSpD calculated an estimated 1,569 independent tests, with a corresponding significance threshold of $\alpha_{SNPSpD} = 3.16 \times 10^{-5}$, maintaining the type I error rate at 5%.

Gene-wide Tests of Empirical Significance

Estimates of empirical significance for association results were obtained by adaptive permutation of gene-dropping simulations created with MERLIN [63]. Simulated genotypes were of identical frequency, marker spacing, and pattern of missing data as the actual genotypes, with individual phenotypes and pedigree structure also preserved within each simulated dataset. For markers in linkage disequilibrium ($r^2 \geq 0.1$), alleles were simulated using the haplotype frequencies for the marker clusters. To reduce computation time, those pedigrees of complexity greater than 70 bits were omitted from calculation of allele and haplotype frequencies. Each simulated dataset was analyzed as described above in two ways: retaining the minimum $P$-value per gene, as well as calculating the threshold-truncated product of $P$-values ($\prod_{k=1}^{n_{perm}} P_{k} \leq 0.01$) per gene. For the set of single-SNP hypotheses corresponding to a gene, the truncated product method considers the product of only those $P$-values falling below a specified threshold, evaluating the probability of observing as significant a product by chance. Whereas Fisher’s Combined Test assesses the overall evidence for departure from the null, the truncated product approach can be used to assess whether suggestive or significant findings are truly significant [64]. Previous reports support the use of a truncated product approach in conjunction with the PDT [65]. Empirical significance was calculated from the proportion of simulated gene-wise test statistics more significant than the actual results ($\prod_{k=1}^{n_{perm}} P_{k} \leq 0.01$). We used an adaptive permutation procedure, by which empirical $P$-values were obtained for 100, 1,000, 10,000, and 100,000 simulations. Only those observed associations for which there were not at least ten more significant simulated results were carried forward to each successive stage of permutation analysis.

Supporting Information

Table S1 Chromosome 20 genes assayed, with corresponding boundary SNPs.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chromosome 20</th>
<th>Boundary SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC2</td>
<td>20</td>
<td>100000000</td>
</tr>
<tr>
<td>ABO</td>
<td>20</td>
<td>30000000</td>
</tr>
<tr>
<td>ACTA2</td>
<td>20</td>
<td>40000000</td>
</tr>
</tbody>
</table>

Table S1 Chromosome 20 genes assayed, with corresponding boundary SNPs. For each gene assayed, the corresponding number of SNPs, position of the first SNP, and its dbSNP identifier, and the position of the last SNP and dbSNP identifier are given.

Supporting Information

Table S1 Chromosome 20 genes assayed, with corresponding boundary SNPs.

Table S1 Chromosome 20 genes assayed, with corresponding boundary SNPs. For each gene assayed, the corresponding number of SNPs, position of the first SNP, and its dbSNP identifier, and the position of the last SNP and dbSNP identifier are given.

Supporting Information
Acknowledgments

We are grateful to the patients and their families for their generous participation in these studies.

Author Contributions

Conceived and designed the experiments: ZZ JS RL BW TN. Performed the experiments: TBB AHF BPR BW. Analyzed the data: TBB AHF BPR BW. Wrote the paper: TBB.