Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia

Anna C. Need[1,2,3], Joseph P. McEvoy[3], Massimo Gennarelli [4,5], Erin L. Heinzen[1,2], Dongliang Ge[1], Jessica M. Maia[1], Kevin V. Shianna[1,2], Min He[1], Elizabeth T. Cirulli [1], Curtis E. Gumbs[1], Qian Zhao[1], C. Ryan Campbell[1], Linda Hong[1], Peter Rosenquist[6], Anu Putkonen[7], Tero Hallikainen[7], Eila Repo-Tiihonen[7], Jari Tiihonen[7,8], Deborah L. Levy [9], Herbert Y. Meltzer [10], David B. Goldstein[1,11]

[1] Center for Human Genome Variation, Duke University School of Medicine, Durham, NC, 27708, USA

[2] Department of Medicine, Section of Medical Genetics, Duke University, Durham, NC, 27708, USA

[3] Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, 27710, USA.

[4] Genetic Unit, IRCCS "San Giovanni di Dio" - Fatebenefratelli, 25123, Brescia, Italy.

[5] Department of Biomedical Sciences and Biotechnologies, Biology and Genetic Division, University School of Medicine, 25121, Brescia, Italy.

[6] Department of Psychiatry and Behavioral Medicine, Wake Forest University, North Carolina Baptist Hospital, Winston-Salem, NC, 27157, USA.

[7] University of Eastern Finland, Department of Forensic Psychiatry, Niuvanniemi Hospital, FI-70240, Kuopio, Finland.

[8] Department of Clinical Neuroscience , Karolinska Institutet, SE-17177, Stockholm, Sweden

[9] Psychology Research Laboratory, McLean Hospital, Belmont, MA, 02478, USA

[10] Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, 60611, IL, USA

[11] Department of Molecular Genetics of Microbiology, Duke University School of Medicine, Durham, NC, 27708, USA

Corresponding authors:

Anna C. Need
Center for Human Genome Variation,
Duke University, 308 Research Drive,
Box 91009
LSRC B Wing, Room 330
Durham, NC 27708
(P) 919-684-4674 (F) 919-668-6787
anna.need@duke.edu

David B. Goldstein
Center for Human Genome Variation,
Duke University, 308 Research Drive,
Box 91009
LSRC B Wing, Room 330
Durham, NC 27708
(P) 919-684-0896 (F) 919-668-6787
d.goldstein@duke.edu

Abstract (248 words)

Schizophrenia is a severe psychiatric disorder with strong heritability and marked heterogeneity in symptoms, course and treatment response. There is strong interest in identifying genetic risk factors that can help to elucidate the pathophysiology and may result in the development of improved treatments. Linkage and genome-wide association studies (GWAS) suggest that the genetic basis of schizophrenia is heterogeneous. However, it remains unclear whether the underlying genetic variants are mostly moderately rare, and can be identified by genotyping variants observed in sequenced cases in large follow-up cohorts, or whether they will typically be much rarer and therefore more effectively identified by gene-based methods that seek to combine candidate variants.

Here we consider 166 persons with schizophrenia or schizoaffective disorder who have had either their genomes or exomes sequenced to high coverage. From these data, we selected 5155 variants that were further evaluated in an independent cohort of 2617 cases and 1800 controls. No single variant showed a study-wide significant association in the initial or follow up cohorts. However, a number of case-specific variants were identified, some of which may be real risk factors for schizophrenia, and these can be readily interrogated in other datasets. Our results indicate that schizophrenia risk is unlikely to be predominantly influenced by variants just outside the range detectable by GWAS. Rather, multiple rarer genetic variants must contribute substantially to the predisposition to schizophrenia, suggesting that both very large sample sizes and gene based association tests will be required to securely identify genetic risk factors.

Introduction

Schizophrenia [MIM 181500] is characterized by positive symptoms (e.g., delusions, hallucinations, disorganized thinking), negative symptoms (e.g., flat affect, loss of spontaneity, diminished initiative and capacity for pleasure, impaired volition), numerous cognitive dysfunctions of varying severity, mood disturbances and suicidality. Antipsychotic drugs are usually effective in treating positive symptoms, and to a much lesser extent, cognitive impairment. Clozapine, the prototypical atypical antipsychotic drug, has also been shown to uniquely reduce risk the risk for suicide in schizophrenia[1; 2]. Cognitive impairment is considered the major cause of deficits in functioning[3], but the other components of the illness contribute as well, collectively interfering substantially with quality of life and constituting significant burdens on the families of individuals with schizophrenia. Genetic studies that implicate variants in specific genes as risk factors for the syndrome may help to elucidate the pathophysiology of the syndrome and the identification of novel treatment targets.

Despite many years of study, the genetic basis of schizophrenia remains largely unknown. Many complex diseases, including neuropsychiatric disorders such as epilepsy and Alzheimer's disease [MIM 104300][4; 5], have been shown to have Mendelian forms; however, no single gene mutations of large effect have been conclusively identified in schizophrenia. Moreover, genome-wide association studies (GWAS) have identified only variants associated with extremely small effects on risk. For example, GWAS meta-analyses on 8000 cases and 19,000 controls identified several high frequency associations with very small odd ratios (1.1-1.3), effectively ruling out the possibility that risk for schizophrenia is determined primarily by a modest number of common variants (or even hundreds of common variants)[6-8]. On the other hand, studies of rare copy number variants (CNVs) have shown that a modest proportion of schizophrenia cases can be attributed to a heterogeneous collection of rare CNVs with high but incomplete penetrance, with estimated odds ratios ranging from 2.7 to 25 [9-12]. Because of their rarity, and the multiple genes involved in many of the CNV regions, they represent daunting targets for drug development unless they lead to more generalized downstream effects that affect a much higher percentage of people with schizophrenia .

With very little of the heritability of schizophrenia explained by non-genetic causes, and good evidence for a role for rare variants, there is intense interest in using next generation sequencing to identify additional rare variants associated with schizophrenia. These, in turn, might help to identify pathways that could inform and motivate novel drug development efforts. Two recent studies explored the role of highly penetrant individual sequence variants in schizophrenia by examining the number and function of *de novo* variants in apparent sporadic schizophrenia (in 14 and 53 cases, respectively)[13; 14]. The authors of both studies concluded that an excess of *de novo* variants was seen in the schizophrenia cases and, additionally, that more of these than expected were damaging, suggesting that at least some schizophrenia cases are caused by highly penetrant *de novo* variants. However, the high heritability of schizophrenia is not compatible with the hypothesis that most cases are the result of *de novo* mutations.

The genetic explanation of the majority of schizophrenia cases therefore remains unresolved. One possibility not excluded by current evidence, is that variants only slightly below the detection threshold for GWAS have appreciable effects on risk (for example variants with frequencies approaching 0.5% and relative risks of 2 or slightly more). If there were many such variants some could be readily detected by sequencing case genomes and genotyping identified variants in a large cohort of additional cases. This is the design we follow here. Another possibility is that most pathogenic mutations have frequencies well below the GWAS detection threshold. For such variants, the most efficient design will be to employ screens based on the combination of variants across particular genes or regions. There is a clear analogy here with CNVs. Although each individual mutation in a schizophrenia-associated CNV region appears to have arisen either *de novo* or very recently, and thus has extremely low frequency, collectively they reach frequencies that are significantly different in cases versus controls [9; 10; 15-19]. Whether this will be the requisite paradigm for identifying sequence variants remains to be resolved.

Here we examined the first possibility – whether a substantial proportion of schizophrenia cases can be explained by individual, moderately rare variants with strong effects. We took a set of 5155 rare variants identified in 166 sequenced schizophrenia genomes and exomes, and genotyped them in an additional unrelated 2617 cases and 1800 controls to determine whether rare variants over-represented in unrelated cases relative to controls could be detected in samples of this size.

Methods

Study Participants

Discovery Cases were 166 individuals with a diagnosis of schizophrenia or schizoaffective disorder, enriched for treatment resistant cases and/or cases with a strong family history. The original sequenced cases comprised 47 Finnish individuals with schizophrenia defined as treatment-resistant (as indicated by qualifying for treatment with clozapine), 87 US individuals with treatment-resistant schizophrenia/ schizoaffective disorder (all whole exome sequenced), and 32 US individuals with a diagnosis of schizophrenia as well as a family history of schizophrenia or other severe neuropsychiatric disorders (whole genome sequenced). The sample was approximately 10% African ancestry, 89% European ancestry and 1% other (Native American, Hispanic). The institutional review boards (IRBs) of Duke University Medical Center and collaborating institutions approved all procedures. **Follow-up cases** (n=2756) were 544 US individuals with a family history of schizophrenia or other severe neuropsychiatric disorders (including 20 overlapping with whole genome sequenced for QC purposes), 364 US samples, including 168 treatment resistant (79 overlapping with exome sequenced), 360 Italian samples, and 1567 samples obtained from the Rutgers repository. Follow up cases were 45% African ancestry, 54% European ancestry and 1% other (Hispanic, Asian, Middle Eastern). Informed consent was obtained from all participants or their legal guardians.

Discovery Controls (n=307) and **follow-up controls** (n=1932, including 65 discovery controls for quality control purposes) were subjects not enriched for (but not specifically screened for) neuropsychiatric disorders enrolled in Duke IRB approved protocols who consented to future unrelated research, and from samples received from outside institutions under a Duke IRB exemption. Discovery controls were 6% African ancestry, 92% European ancestry and 2% other (Hispanic, Native American, Middle Eastern). Follow-up controls were 37% African ancestry, 59% European ancestry and 3% other (Hispanic, Asian, Middle Eastern).

Power Calculations were performed with Power for Genetic Association Analyses (PGA) [20].

Targeted capture and exome/ genome sequencing. For whole exome sequencing, the target regions were captured using the Agilent SureSelect Human All Exon 37Mb or 50Mb Kit (Agilent Technologies, Santa Clara, CA) following vendor provided protocols. Sequencing was performed in the CHGV Genomic Analysis Facility, using either Illumina GAII or HiSeq machines. Whole genome sequencing was performed as previously described[21]. Each read was then aligned to the reference genome (NCBI human genome assembly build 36; Ensembl core database release 50_361 1) using the Burrows-Wheeler Alignment (BWA) tool [22] and single nucleotide variants (SNVs) and small insertion-deletions (indels) were identified by using SAMtools[23]. PCR duplicates were removed using the Picard software (for URL, see Web Resources).

Variant Selection. We focused on variants annotated as functional, where functional was defined as non-synonymous, nonsense or located in the canonical splice sites. The analysis was restricted to variants with a MAF< 0.05 or, for the recessive model, MAF<0.3. For each variant tested, only individuals with a minimum coverage of 10x at that site were included (n= 152,511 SNVs for the allelic model and 172,886 for the recessive model).

We first tested for genetic association, using ATAV (for URL see Web Resources)to run Fisher's exact tests comparing exome-captured regions from cases and controls. We then removed variants with greater than 50% of the individuals missing due to low coverage, a Hardy-Weinberg equilibrium p value < 0.001, variants with a p<0.05 but whose frequency was higher in controls, and those variants that were clearly driven by differences between the Finnish or African American subjects and the other samples. All other associated (p<0.05) variants were put forward for scoring in the iSelect follow-up (n=316 allelic, 206 recessive).

Because of an expectation of high locus heterogeneity, we recognized that a gene may play a role in only a single case in our sequenced samples, and so we also selected variants that formed genotypes in cases but were not present in controls. From this set, we removed: variants that were present in only one individual ("unique variants") if that individual had greater than the mean + 1 standard deviation of unique variants (suggesting that these may have been poor quality calls) (n=485 allelic, 5 recessive), unique variants that were present only in the Finnish samples (since we had no Finnish controls) (n=293 allelic, 12 recessive) and unique variants in a low coverage sample (23.5-25x, n=4 samples; n=33 allelic, 1 recessive).

We then included for scoring 1) all variants that were in an essential splice site, destroyed a stop codon or introduced a new stop codon (n=1643 allelic, n=43 recessive), 2) all variants in a region associated through rare copy number variation with epilepsy, schizophrenia, autism or intellectual disability (n=863 allelic, 2 recessive), and 3) all non-synonymous variants that got a PolyPhen2[24] rating of 'probably damaging' (n=3736 allelic, n=51 recessive). These 6,860 variants were submitted to Illumina's online Assay Design Tool to predict the likelihood of a successful assay. All variants with a score below 0.6 were removed, leaving 5,788 variants.

iSelect QC. The raw data (idats) from the custom iSelect genotyping were brought into the Illumina GenomeStudio software. All variants were clustered using the GenomeStudio default parameters. The call rates were inspected and any samples with a call rate below 0.95 were removed from analysis. Each variant call was manually inspected for clustering accuracy and any obvious miscalls were corrected or deleted if correcting was not possible due to irregular clustering. Genders were checked based on X and Y chromosome variants, and mismatched samples were excluded. For a subset of samples, we also had Illumina genome-wide SNP array data, which we used to check the concordance for overlapping SNPs, and discordant samples were excluded from analysis. Of these 5788 variants, 5155 variants were successfully genotyped and passed all QC checks. We obtained genotypes for these variants in an additional 2617 cases and 1800 controls that passed all QC checks.

Individuals with epilepsy were also genotyped with the iselect array, and all putative schizophrenia-associated variants were investigated in cases with epilepsy to evaluate possible variable expressivity[25].).

Data Analysis. Because of allele frequency differences between subjects of African and European decent, we analyzed these two population groups separately and then, where appropriate, combined the p values using a Fisher's trend test. Fisher's allelic and recessive tests were performed using PLINK (see Web Resources). According to PLINK defaults, only females were included in the Fisher's exact test for X chromosome variants.

Comparison of Validation Rates. To determine if there were differences between the validation rates among different classes of functional variants, we took all autosomal variants that were seen in only a single sample in the initial sequencing cohort. These were considered to be the most vulnerable calls. We then selected from this group only the variants that were present in an individual that was genotyped with the iSelect array, and determined for each variant whether it had also been identified by the iSelect genotyping. Those identified were classed as validated, and we compared % validated among different functional groups.

Results

Association Testing. We performed whole exome (n=134) or whole genome sequencing (n=32) on 166 cases with schizophrenia and on 307 controls (n=256 exome and 51 genome) and followed this with Fisher's Exact test to look for association with schizophrenia using both an allelic and a recessive model after excluding common variants. This resulted in a total of 337,312 variants, for which Bonferroni correction for multiple testing required a $p< 1.5 \times 10^{-7}$. At this p value threshold in a dataset of this size (focusing exclusively on rare variants), we would have limited power to detect anything but variants with an extremely high relative risk (Figure 1). As expected, no variant achieved the required level of significance in the initial sequence data. This is not surprising because variants that would show significance in this initial sequence dataset would have to have had effect sizes that would have been expected to have been identified in linkage analyses.

If we use a cutoff of $p<0.05$ instead, however, we are powered to identify variants over a much broader range of effect sizes and allele frequencies (Figure 1), although the majority of these variants will be false positives. We therefore adopted a two-stage strategy in which 1) the discovery cohort was used to identify a set of variants that is likely to be enriched for schizophrenia-associated variants (those with $p<0.05$ selected as described in methods; n=428), and 2) a follow-up sample of 2617 cases and 1800 controls was used to look for corroborating evidence of association. The original sequencing data were combined with the iSelect data to produce a final dataset of 2785 cases and 2120 controls.

Because there were insufficient markers included to adjust for population stratification accurately, Fisher's exact tests (for both recessive and allelic models) were performed on individuals of African and European descent separately, and a logistic regression using self-described race (including only those of European or African ancestry) was used to look for variants that associated in both ancestry groups. Using a Bonferroni correction for all original 337,312 variants, $p< 1.5 \times 10^{-7}$ was required for study-wide significance. The lowest p values in the combined dataset were 0.0003 (allelic) and 0.01 (recessive). The most significantly associated variants in the African American only analysis had p values of 0.0006 (allelic) and 0.0005 (recessive), and the most significant variants in Europeans were $5.9\text{x}10^{-6}$ (allelic) and 0.01 (recessive). The most strongly associated variant was a non-synonymous variant in *AL589787.16* (HUGO:N/A, rs7098669) that was originally included because it showed association in the recessive model. This variant shows substantial variation in allele frequency across populations, and the association probably reflects population stratification. It is interesting to note, however, that this SNP lies within an open reading frame (*C10orf90)* found to contain two independent schizophrenia associations in the recent mega-GWAS, although the association did not hold up in the follow-up cohort of that study[26].

Evaluation of Genotypes Exclusive to Cases. Most very rare highly penetrant schizophrenia-associated genotypes would not be expected to show a significant association in a dataset of the size of our discovery cohort. Such genotypes would be found exclusively in cases, but only in very small numbers. We therefore also followed up all genotypes that were present in two cases and no controls (n= 861) and a subset of those present in one case and no controls (n=4498; see Methods for selection criteria), as this dataset is likely to be enriched for very rare variants with high penetrance. Table 1 summarizes the iSelect outcome for the different categories of variants. Of the 4028 genotypes originally present in only a single case and successfully genotyped in the follow up, 1588 (39%) were seen in an iSelect control, and 1989 (49%) were seen in only one or no further cases. To further explore the frequency of these variants in controls, we referred to the Exome Variant Server, NHLBI Exome Sequencing Project (ESP), Seattle, WA [accessed January 2012]. Supplementary tables 1 and 2 show all case-specific variants and case-specific homozygotes, respectively.

Some variants remained absent in all controls, and were found to be present in several additional cases (table 2). The variant present in the most cases (n=5) is a non-synonymous mutation (ENST00000380099 g1882C>T; Arg628Cys) in *KL* [MIM 604824], a gene that acts mainly in the renal and cardiovascular systems[27]. However, *KL* has also been implicated in the regulation of vitamin D metabolism. *Klotho* deficient mice, for example, show degeneration of mesencephalic dopamine neurons[28] and other aging phenotypes, such as hearing loss[29], which can be rescued with a vitamin D deficient diet. Although *KL* has not been previously associated with schizophrenia to our knowledge, substantial evidence implicates alterations in vitamin D levels, especially deficiency, as a risk factor for schizophrenia that may also help to explain certain epidemiological findings, such as season of birth and latitude gradient effects on geographic differences in incidence and prevalence [30-32]. In addition, prenatal vitamin D deficiency has been shown to have lasting changes in NMDA-mediated brain function in adult rats[33]. Other genes of interest with case-only variants include *EPB41L1* [MIM 602879]*, SLC1A2* [MIM 600300]*, STX4* [MIM 186591]*, HYDIN* [MIM 610812]*, PCLO* [MIM 604918] and *ZNF804B [MIM:NA]*. *EPB41L1* encodes the erythrocyte membrane protein band 4.1-like N[34], which co-localizes with AMPA receptors at excitatory synapses and is thought to mediate the interaction of the AMPA receptors with the cytoskeleton[35]. It has also been shown to be necessary for the formation of calcium waves in the mediation of neurite formation[36]. A recent study of synaptic protein sequencing reported a d*e novo* functional missense mutation in *EPB41L1* in nonsyndromic intellectual disability[37]. *SLC1A2* (aka (EAAT2, GLT1) encodes a glial high affinity glutamate transporter, and is the major transporter in the forebrain[38]. Mice lacking *Slc1a2* have spontaneous lethal seizures and are vulnerable to glutamate neurotoxicity after forebrain trauma[39]. *STX4* directs membrane fusion at excitatory glutamatergic synapses, and is essential for normal dendritic spine morphology, retrograde synaptic signaling and long-term potentiation (LTP) at hippocampal synapses[40]. A frameshift mutation of *Hydin* causes recessively transmitted hydrocephalus in the mouse[41]. Additionally, a paralog of *HYDIN* is included in 1q21.1 microdeletions and microduplications that have been associated with microcephaly, macrocephaly and neuropsychiatric disorders including schizophrenia[16]. *PCLO* encodes Picollo, which acts alongside Bassoon, Syntaxin, SNAP-25, and N-Cadherin, in the presynaptic active zone, a specialized region where synaptic vesicles dock and fuse[42]. Piccolo is upregulated in the nucleus accumbens in response to methamphetamine, and antisense suppression of Piccolo increases the behavioural response to amphetamine and causes synaptic accumulation of dopamine[43]. *ZNF804B* is a paralog of *ZNF804A* [MIM 612282], which has shown evidence in GWAS studies for an association with schizophrenia[44]. Some variants were present in both cases and controls but were homozygous only in cases (see supplementary table 2). Two variants were homozygous in 3 or more cases and absent in all controls, but they are both predicted to be benign.

Interestingly, it was recently shown that loss-of-function mutations have a lower validation rate than other classes of variants [45]. This is expected, because deleterious variants are likely to be less common than those with moderate or no function and thus, all else being equal, there is a lower prior that deleterious variants are real. We searched for a similar effect in our data, although we expected our validation rate in general to be higher because (i) we required high stringency for variants to be carried to follow up, and (ii) we looked almost exclusively at rare variants predicted to have strong functional effects, so the magnitude of the difference among them would be expected to be reduced. Nevertheless, we do see an attenuated effect as described by MacArthur *et al.* (2012), such that if we compare validation rates (as defined by a variant being genotyped on the iSelect array) among types of variants the rates for very rare nonsense (76%, n=388) and essential splice variants (73%, n=221) are lower than for non-synonymous variants (86%, n=2826).

Discussion

In this large-scale report of sequenced schizophrenia genomes, we have looked for evidence for association of individual, rare, highly penetrant sequence variants in a discovery cohort of 166 cases and 307 controls and in a follow-up cohort of 2756 cases and 1932 controls. We found no significantly associated variants after Bonferroni correction for multiple hypothesis testing.

Having found no significantly associated SNVs, we focused on variants that were present in more than one case and absent in controls. At least some of these may ultimately prove to be statistically associated with schizophrenia, even though each was individually too rare to reach study-wide significance in this dataset. We found a small number of variants that remain good candidates as individual schizophrenia-associated variants and are high priority for follow up studies. However, because of their low frequency, large collaborative studies will be required to provide statistically significant evidence for their association with schizophrenia. As an illustrative example, we can consider the top hit, a missense mutation in *KL* that was present in 5/2780 cases and absent in 7417 controls. Assuming that the frequency in cases remains at 0.18%, and that we had equal numbers of cases and controls, we would need to see about 23 more schizophrenia cases with this variant, requiring approximately 13,000 schizophrenia cases and controls, to obtain study-wide significance for this individual variant ($p< 1.5 \times 10^{-7}$).

This study had 99% power to detect moderately rare (1-5%) variants with relative risks between 2 and 6. As some variants in this frequency range would have been poorly represented on GWAS chips[46], and effect sizes in part of this range would have been difficult to detect consistently with linkage studies[47], variants with these properties have not previously been systematically investigated on a genome-wide scale in relation to schizophrenia. On the basis of our findings and those of previous studies, some possible genetic architectures for schizophrenia risk appear increasingly unlikely, including (i) a small number of highly penetrant loci explaining the majority of cases (linkage studies), (ii) common variants with low relative risks underlying most cases (GWAS), and (iii) so called "goldilocks alleles", [48] or moderately rare variants with moderate RR explaining most cases (present study). This finding suggests that the majority of schizophrenia-associated variants will be of very low frequency and their association with schizophrenia will most readily be confirmed by their collective presence in genes associated with schizophrenia using collapsing methods and related approaches[49-52], rather than by variant-specific frequency differences between cases and controls.

These findings have a number of implications. First, a likely benefit of the 1000 genomes project will be novel GWAS arrays that extend to variants in the 1-5% frequency range. Our results suggest that the use of these arrays in relation to schizophrenia is unlikely to reveal significant associations for RRs above 2.

Second, our data strongly suggest that genetic risk factors for schizophrenia are outside the range of what is easily detected in sample sizes of the sort used here. For example, although our power to detect variants with frequencies below one percent is low, if there were many risk factors near this frequency cutoff with modest effect sizes, we should have detected evidence for some of them. Thus, at least in terms of the relatively rare and high impact variants contributing to schizophrenia risk, the genetic architecture is one characterized by high locus heterogeneity, or high allelic heterogeneity, or, more likely, both. Beyond locus and allelic heterogeneity, the results could also be consistent with oligogenic, polygenic or epistatic models where effect sizes for individual variants are very modest, and the genotypes are reasonably penetrant only in combination. We did not explicitly test for such models given the current sample sizes. Considering the number of tests required for open screens of interacting variants (even limited to interacting pairs and restricting analyses to common variants), we suspect that such interactions will need to be identified secondarily to the identification of a main effect, however small, for a single variant, or through testing of specific hypotheses analyzing small sets of variants in defined biological pathways.

It is valuable to consider these observations in light of the example of CNVs. Copy number variants were the first type of rare variants that could be detected on a genome-wide scale. Moderately rare CNVs with RR from 2-25 were immediately and definitively associated with schizophrenia risk across different cohorts and populations[9; 10; 17]. Because other types of rare genetic variation, such as SNVs and indels, were not available for comparison, it was not possible to determine whether these properties were specific to large structural variants or were representative of rare schizophrenia-associated variants in general. Here, we have been able to systematically search for associated SNVs with similar relative risks that explain as high a proportion of cases as some of the CNV regions, and have failed to identify any.

It is essential to appreciate that the schizophrenia-associated CNVs were detectable because recurrent mutation greatly increases the number of affected cases. Any given mutation event leading to a CNV appears to be responsible only for a few cases among closely related relatives[10]. If SNV mutations prove analogous, we would expect certain individual genes or gene pathways to associate with schizophrenia, although individual variants in those genes would not be frequent enough to reach statistical significance by themselves in the sample sizes used in our follow-up genotyping experiment. The clear implication is that collapsing methods that combine qualifying variants in individual genes, and/or much larger sample sizes will be required to find secure evidence of risk-conferring genes. Such a conclusion should not necessarily discourage the prospects for gene discovery in schizophrenia, rather it suggests that the clearest path to discovery should focus on both significantly expanded samples sizes and sophisticated methods to appropriately combine qualifying variants affecting the same genes[49-52]. Thus, we propose two possible strategies, one that sequences a sufficiently large discovery cohort to prioritize genes in terms of their load of qualifying mutations, followed by sequencing of those genes of interest in additional samples , and/or a second that generates complete sequence data on 10,000 or more case genomes. Significant findings from either approach will implicate particular genes as risk factors for schizophrenia, and these genes will hopefully converge on a limited number of pathways, thus providing valuable information about the genetic etiology of schizophrenia. Either way, only very rarely will an individual variant be determined to be causal in an individual schizophrenia case, thus limiting the scope for genetics in individual risk prediction.

Two conclusions emerge from this study. First, multiple genetic variants must underlie the predisposition to schizophrenia. Current sequence data cannot determine whether risk alleles will affect tens, hundreds, or thousands of different genes as our power to distinguish allelic and locus heterogeneity in this study is limited. If locus heterogeneity were low, however, we would expect linkage data to have been more consistent in implicating a small number of genomic regions. On balance, therefore, the data appear to point toward both high locus and allelic heterogeneity. Second, the risk variants for schizophrenia will be rare, although we cannot determine from a study of this size just how rare individual pathological variants will be.

The conclusions of this study must be tempered by a number of considerations. First, we analyzed only coding regions. It therefore remains possible that moderately rare regulatory variants with relatively strong effect sizes explain the majority of cases of schizophrenia. This would seem to be an unlikely model based on the prominent role of coding variants in Mendelian disorders, but will be tested in the near future as variants identified from whole-genome sequencing of individuals with schizophrenia are investigated in larger datasets. One example of a regulatory region that was not represented in this study was that around the microRNA MIR137 (miR-137), which was recently associated with schizophrenia through a "mega-GWAS" [26]. It is also likely that some of the variants that we included as likely to be functional are in fact errors of reference-based mapping, and that other functional variants (e.g., nonsense mutations) were excluded from the analysis due to mismapping [45].

Second, our initial analysis included cases from multiple populations, including Finns and African Americans. If rare schizophrenia-associated variants are population specific, power would be reduced by not focusing on single ethnic or racial groups. We know from previous studies that different racial and ethnic groups share rare CNVs [17], but this could be because they are in hypermutable regions and thus the same structural changes arise multiple independent times, which is much less likely to be the case for SNVs. Although we performed separate analyses for samples of self-reported European and African ancestry, we were not able to control for population structure within these groups. Stratification usually acts to inflate test statistics, but the effect on power was expected to be modest. Additionally, existing methods are not effective for the type of structure caused by rare variants[53]. We therefore took the approach to be maximally inclusive in this exploratory study, and that any variants that were supported in the second stage would be further investigated in cohorts that are robust to the effects of population stratification, such as family-based studies, as suggested in ref [53]. Finally, the discovery sample comprised mostly people with treatment resistant schizophrenia, who are a special population representing about 30% of people who meet DSM IV criteria for schizophrenia[54]. The replication sample had a much smaller percentage of treatment-resistant cases so there remains a possibility that real associated variants were not supported in the follow-up sample because they increase the risk for treatment-resistant schizophrenia specifically, but not for schizophrenia more generally.

Our findings will aid in the interpretation of future small-scale sequencing studies. Here we have investigated almost all obvious functional variants in a set of 166 schizophrenia cases and have found no study-wide significant associations. Notably, the majority of non-private variants that were exclusive to cases in the discovery cohort were either seen in controls in the follow-up sample (60%), or were not seen in any further cases (23%). Thus, associations based on small-scale sequencing studies, however biologically enticing, must be interpreted with caution pending validation in larger follow-up studies.

If the majority of variants occur at very low frequency, there are a number of potential investigative strategies. The most obvious and immediate strategy would be to use gene-based NGS approaches to comprehensively investigate candidate genes (e.g., from CNV studies or GWAS) in large datasets, as described above. This approach would allow us to accurately gauge how much of the heritability of schizophrenia is explained by genes that have already been identified. We can also take advantage of the relatively small early sequencing datasets to identify a small number of promising candidate rare variants such as those presented here that continue to be absent in controls, for genotyping in larger datasets. Although the majority of these candidates will turn out to be false positives, potential insight into schizophrenia etiology is worth the cost if even one is statistically proven. Finally, rare variants can be investigated using family studies, using either segregation analyses in multiplex families or looking for *de novo* or recessive variants in sporadic cases. Although both of these approaches are likely to produce multiple candidate variants per family, with adequate control datasets and by comparing multiple families in collaborative efforts, it should be possible to identify either recurring variants or repeatedly disrupted genes that will ultimately prove to contribute to schizophrenia risk.

Supplemental Data Description

Supplementary Table 1. Single Nucleotide Variants Present in Cases Only.

Supplementary Table 2. Single Nucleotide Variants that are Homozygous or Hemizygous in Cases Only.

Acknowledgements

We thank the participants who made their DNA available for research. Thanks to Kristen Linney, Jordan Silver, Marlyne Silver, Yates Creech; doctors and staff at Central Regional Hospital; Robert Millet, Edward Leuth, and other doctors and staff at Carolina Behavioral Care for NC patient recruitment; and Latasha Little, Ken Cronin, Melora McCall, Josh Mauro and Alex McKenzie for lab support. For the Italian cohort, we acknowledge Catia Scassellati, Cristian Bonvicini for sample handling and Stefano Bignotti, Giuseppe Rossi and Alessandra Minelli for recruitment in Italy. We thank the Murdock study community registry and biorepository pro00011196, D. Daskalakis, R. Brown, A. Holden, E. Behr, W. Lowe, P. Lugar, J. Milner, K. Welsh-Bohmer, C. Hulette, J. Burke, D. Valle, J. Hoover-Fong, N. Sobriera, D. Marchuk, S. Palmer, E. Pras, D. Lancet and Z. Farfel for control DNA. This work was funded by NIMH ARRA grant 3RC2MH089915-01W1 and R01 grant MH071523, NINDS ARRA grant RC2NS070344, NIAID grant UO1AIO67854, , Italian Ministry of Health (Ricerca Corrente) and AFaR grants, Ellison Funding (O.C-F), NIA P30 grant AG028377 (Bryan ADRC), the Sidney R. Baer, Jr. Foundation, the Essel Foundation, the Carmela and Menachem Abraham Fund, and by NARSAD (now the Brain and Behavior Research Foundation). This research was made possible through collaboration with the International Serious Adverse Events Consortium, whose membership currently includes Abbott, Amgen, Astra-Zeneca, Cerner, Daiichi-Sankyo, GlaxoSmithKline, Merck, Novartis, Pfizer, Takeda, and the Wellcome Trust. The NIMH Cell Repository at Rutgers University provided schizophrenia follow-up samples. HYM is a shareholder in SureGene, LLC.

Web Resources

ATAV, http://www.duke.edu/~minhe/atav/
HUGO Gene Nomenclature Committee Home Page, http://www.genenames.org
NHLBI Exome Sequencing Project (ESP) Exome Variant Server http://snp.gs.washington.edu/EVS
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org
Picard, http://picard.sourceforge.net
Plink, http://pngu.mgh.harvard.edu/~purcell/plink/
Reactome, http://www.reactome.org
SequenceVariantAnalyzer (SVA), http://www.svaproject.org

References

1. Tiihonen, J., Lonnqvist, J., Wahlbeck, K., Klaukka, T., Niskanen, L., Tanskanen, A., and Haukka, J. (2009). 11-year follow-up of mortality in patients with schizophrenia: a population-based cohort study (FIN11 study). Lancet 374, 620-627.
2. Meltzer, H.Y., Alphs, L., Green, A.I., Altamura, A.C., Anand, R., Bertoldi, A., Bourgeois, M., Chouinard, G., Islam, M.Z., Kane, J., et al. (2003). Clozapine treatment for suicidality in schizophrenia: International Suicide Prevention Trial (InterSePT). Arch Gen Psychiatry 60, 82-91.
3. Green, M.F., Kern, R.S., Braff, D.L., and Mintz, J. (2000). Neurocognitive deficits and functional outcome in schizophrenia: are we measuring the "right stuff"? Schizophr Bull 26, 119-136.
4. Blennow, K., de Leon, M.J., and Zetterberg, H. (2006). Alzheimer's disease. Lancet 368, 387-403.
5. Baulac, S., and Baulac, M. (2010). Advances on the genetics of Mendelian idiopathic epilepsies. Clin Lab Med 30, 911-929.
6. Stefansson, H., Ophoff, R.A., Steinberg, S., Andreassen, O.A., Cichon, S., Rujescu, D., Werge, T., Pietilainen, O.P., Mors, O., Mortensen, P.B., et al. (2009). Common variants conferring risk of schizophrenia. Nature 460, 744-747.
7. Shi, J., Levinson, D.F., Duan, J., Sanders, A.R., Zheng, Y., Pe'er, I., Dudbridge, F., Holmans, P.A., Whittemore, A.S., Mowry, B.J., et al. (2009). Common variants on chromosome 6p22.1 are associated with schizophrenia. Nature 460, 753-757.
8. Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., and Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460, 748-752.
9. Stone, J.L., O'Donovan, M.C., Gurling, H., Kirov, G.K., Blackwood, D.H., Corvin, A., Craddock, N.J., Gill, M., Hultman, C.M., Lichtenstein, P., et al. (2008). Rare chromosomal deletions and duplications increase risk of schizophrenia. Nature.
10. Stefansson, H., Rujescu, D., Cichon, S., Pietilainen, O.P., Ingason, A., Steinberg, S., Fossdal, R., Sigurdsson, E., Sigmundsson, T., Buizer-Voskamp, J.E., et al. (2008). Large recurrent microdeletions associated with schizophrenia. Nature 455, 232-236.
11. Vacic, V., McCarthy, S., Malhotra, D., Murray, F., Chou, H.H., Peoples, A., Makarov, V., Yoon, S., Bhandari, A., Corominas, R., et al. (2011). Duplications of the neuropeptide receptor gene VIPR2 confer significant risk for schizophrenia. Nature 471, 499-503.
12. Bassett, A.S., and Chow, E.W. (2008). Schizophrenia and 22q11.2 deletion syndrome. Curr Psychiatry Rep 10, 148-157.

13. Girard, S.L., Gauthier, J., Noreau, A., Xiong, L., Zhou, S., Jouan, L., Dionne-Laporte, A., Spiegelman, D., Henrion, E., Diallo, O., et al. (2011). Increased exonic de novo mutation rate in individuals with schizophrenia. Nat Genet.

14. Xu, B., Roos, J.L., Dexheimer, P., Boone, B., Plummer, B., Levy, S., Gogos, J.A., and Karayiorgou, M. (2011). Exome sequencing supports a de novo mutational paradigm for schizophrenia. Nat Genet.

15. Rees, E., Moskvina, V., Owen, M.J., O'Donovan, M.C., and Kirov, G. (2011). De novo rates and selection of schizophrenia-associated copy number variants. Biol Psychiatry 70, 1109-1114.

16. Brunetti-Pierri, N., Berg, J.S., Scaglia, F., Belmont, J., Bacino, C.A., Sahoo, T., Lalani, S.R., Graham, B., Lee, B., Shinawi, M., et al. (2008). Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. Nat Genet 40, 1466-1471.

17. Need, A.C., Ge, D., Weale, M.E., Maia, J., Feng, S., Heinzen, E.L., Shianna, K.V., Yoon, W., Kasperaviciute, D., Gennarelli, M., et al. (2009). A genome-wide investigation of SNPs and CNVs in schizophrenia. PLoS genetics 5, e1000373.

18. Vrijenhoek, T., Buizer-Voskamp, J.E., van der Stelt, I., Strengman, E., Sabatti, C., Geurts van Kessel, A., Brunner, H.G., Ophoff, R.A., and Veltman, J.A. (2008). Recurrent CNVs disrupt three candidate genes in schizophrenia patients. Am J Hum Genet 83, 504-510.

19. Rujescu, D., Ingason, A., Cichon, S., Pietilainen, O.P., Barnes, M.R., Toulopoulou, T., Picchioni, M., Vassos, E., Ettinger, U., Bramon, E., et al. (2008). Disruption of the neurexin 1 gene is associated with schizophrenia. Hum Mol Genet.

20. Menashe, I., Rosenberg, P.S., and Chen, B.E. (2008). PGA: power calculator for case-control genetic association analyses. BMC Genet 9, 36.

21. Pelak, K., Shianna, K.V., Ge, D., Maia, J.M., Zhu, M., Smith, J.P., Cirulli, E.T., Fellay, J., Dickson, S.P., Gumbs, C.E., et al. (2010). The characterization of twenty sequenced human genomes. PLoS genetics 6.

22. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760.

23. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.

24. Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., and Sunyaev, S.R. (2010). A method and server for predicting damaging missense mutations. Nat Methods 7, 248-249.

25. Heinzen, E.L., Depondt, C., Cavalleri, G.L., Ruzzo, E.K., Walley, N.M., Need, A.C., Ge, D., He, M., Cirulli, E.T., Zhao, Q., et al. (2012). Exome sequencing followed by large scale genotyping fails to identify single rare variants of large effect in idiopathic generalized epilepsy. Am J Hum Genet this issue.

26. Ripke, S., Sanders, A.R., Kendler, K.S., Levinson, D.F., Sklar, P., Holmans, P.A., Lin, D.Y., Duan, J., Ophoff, R.A., Andreassen, O.A., et al. (2011). Genome-wide association study identifies five new schizophrenia loci. Nat Genet 43, 969-976.

27. Bernheim, J., and Benchetrit, S. (2011). The potential roles of FGF23 and Klotho in the prognosis of renal and cardiovascular diseases. Nephrol Dial Transplant 26, 2433-2438.

28. Kosakai, A., Ito, D., Nihei, Y., Yamashita, S., Okada, Y., Takahashi, K., and Suzuki, N. (2011). Degeneration of mesencephalic dopaminergic neurons in klotho mouse related to vitamin D exposure. Brain Res 1382, 109-117.

29. Carpinelli, M.R., Wise, A.K., and Burt, R.A. (2011). Vitamin D-deficient diet rescues hearing loss in Klotho mice. Hear Res 275, 105-109.

30. McGrath, J.J., Burne, T.H., Feron, F., Mackay-Sim, A., and Eyles, D.W. (2010). Developmental vitamin D deficiency and risk of schizophrenia: a 10-year update. Schizophr Bull 36, 1073-1078.

31. McGrath, J.J., Eyles, D.W., Pedersen, C.B., Anderson, C., Ko, P., Burne, T.H., Norgaard-Pedersen, B., Hougaard, D.M., and Mortensen, P.B. (2010). Neonatal vitamin D status and risk of schizophrenia: a population-based case-control study. Arch Gen Psychiatry 67, 889-894.
32. Amato, R., Pinelli, M., Monticelli, A., Miele, G., and Cocozza, S. (2010). Schizophrenia and vitamin D related genes could have been subject to latitude-driven adaptation. BMC evolutionary biology 10, 351.
33. Kesby, J.P., O'Loan, J.C., Alexander, S., Deng, C., Huang, X.F., McGrath, J.J., Eyles, D.W., and Burne, T.H. (2012). Developmental vitamin D deficiency alters MK-801-induced behaviours in adult offspring. Psychopharmacology (Berl) 220, 455-463.
34. Parra, M., Gee, S., Chan, N., Ryaboy, D., Dubchak, I., Mohandas, N., Gascard, P.D., and Conboy, J.G. (2004). Differential domain evolution and complex RNA processing in a family of paralogous EPB41 (protein 4.1) genes facilitate expression of diverse tissue-specific isoforms. Genomics 84, 637-646.
35. Shen, L., Liang, F., Walensky, L.D., and Huganir, R.L. (2000). Regulation of AMPA receptor GluR1 subunit surface expression by a 4. 1N-linked actin cytoskeletal association. J Neurosci 20, 7932-7940.
36. Fiedler, M.J., and Nathanson, M.H. (2011). The type I inositol 1,4,5-trisphosphate receptor interacts with protein 4.1N to mediate neurite formation through intracellular Ca waves. Neurosignals 19, 75-85.
37. Hamdan, F.F., Gauthier, J., Araki, Y., Lin, D.T., Yoshizawa, Y., Higashi, K., Park, A.R., Spiegelman, D., Dobrzeniecka, S., Piton, A., et al. (2011). Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. Am J Hum Genet 88, 306-316.
38. Benediktsson, A.M., Marrs, G.S., Tu, J.C., Worley, P.F., Rothstein, J.D., Bergles, D.E., and Dailey, M.E. (2012). Neuronal activity regulates glutamate transporter dynamics in developing astrocytes. Glia 60, 175-188.
39. Tanaka, K., Watase, K., Manabe, T., Yamada, K., Watanabe, M., Takahashi, K., Iwama, H., Nishikawa, T., Ichihara, N., Kikuchi, T., et al. (1997). Epilepsy and exacerbation of brain injury in mice lacking the glutamate transporter GLT-1. Science 276, 1699-1702.
40. Kennedy, M.J., Davison, I.G., Robinson, C.G., and Ehlers, M.D. (2010). Syntaxin-4 defines a domain for activity-dependent exocytosis in dendritic spines. Cell 141, 524-535.
41. Davy, B.E., and Robinson, M.L. (2003). Congenital hydrocephalus in hy3 mice is caused by a frameshift mutation in Hydin, a large novel gene. Hum Mol Genet 12, 1163-1170.
42. Zhai, R.G., Vardinon-Friedman, H., Cases-Langhoff, C., Becker, B., Gundelfinger, E.D., Ziv, N.E., and Garner, C.C. (2001). Assembling the presynaptic active zone: a characterization of an active one precursor vesicle. Neuron 29, 131-143.
43. Nitta, A., Hibi, Y., Miyamoto, Y., and Nabeshima, T. (2010). [Identification of Piccolo as a regulator of behavioral plasticity and dopamine transporter internalization]. Nihon Arukoru Yakubutsu Igakkai Zasshi 45, 525-529.
44. O'Donovan, M.C., Craddock, N., Norton, N., Williams, H., Peirce, T., Moskvina, V., Nikolov, I., Hamshere, M., Carroll, L., Georgieva, L., et al. (2008). Identification of loci associated with schizophrenia by genome-wide association and follow-up. Nat Genet 40, 1053-1055.
45. MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., et al. (2012). A systematic survey of loss-of-function variants in human protein-coding genes. Science 335, 823-828.
46. Barrett, J.C., and Cardon, L.R. (2006). Evaluating coverage of genome-wide association studies. Nat Genet 38, 659-662.
47. Botstein, D., and Risch, N. (2003). Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. NatGenet 33 Suppl, 228.
48. Antonarakis, S.E., Chakravarti, A., Cohen, J.C., and Hardy, J. (2010). Mendelian disorders and multifactorial traits: the big divide or one for all? Nat Rev Genet 11, 380-384.
49. Dering, C., Hemmelmann, C., Pugh, E., and Ziegler, A. (2011). Statistical analysis of rare sequence variants: an overview of collapsing methods. Genet Epidemiol 35 Suppl 1, S12-17.

50. Li, L., Zheng, W., Lee, J.S., Zhang, X., Ferguson, J., Yan, X., and Zhao, H. (2011). Collapsing-based and kernel-based single-gene analyses applied to Genetic Analysis Workshop 17 mini-exome data. BMC proceedings 5 Suppl 9, S117.
51. Li, B., and Leal, S.M. (2008). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet 83, 311-321.
52. Sun, Y.V., Sung, Y.J., Tintle, N., and Ziegler, A. (2011). Identification of genetic association of multiple rare variants using collapsing methods. Genet Epidemiol 35 Suppl 1, S101-106.
53. Mathieson, I., and McVean, G. (2012). Differential confounding of rare and common variants in spatially structured populations. Nat Genet 44, 243-246.
54. Meltzer, H.Y. (1997). Treatment-resistant schizophrenia--the role of clozapine. Curr Med Res Opin 14, 1-20.

Figure 1. Range of relative risks and minor allele frequencies that are detectable with 99% power at $p<0.05$ and $p<1.5 \times 10^{-7}$ (corrected for all included variants) in a cohort of 166 cases and 307 controls.

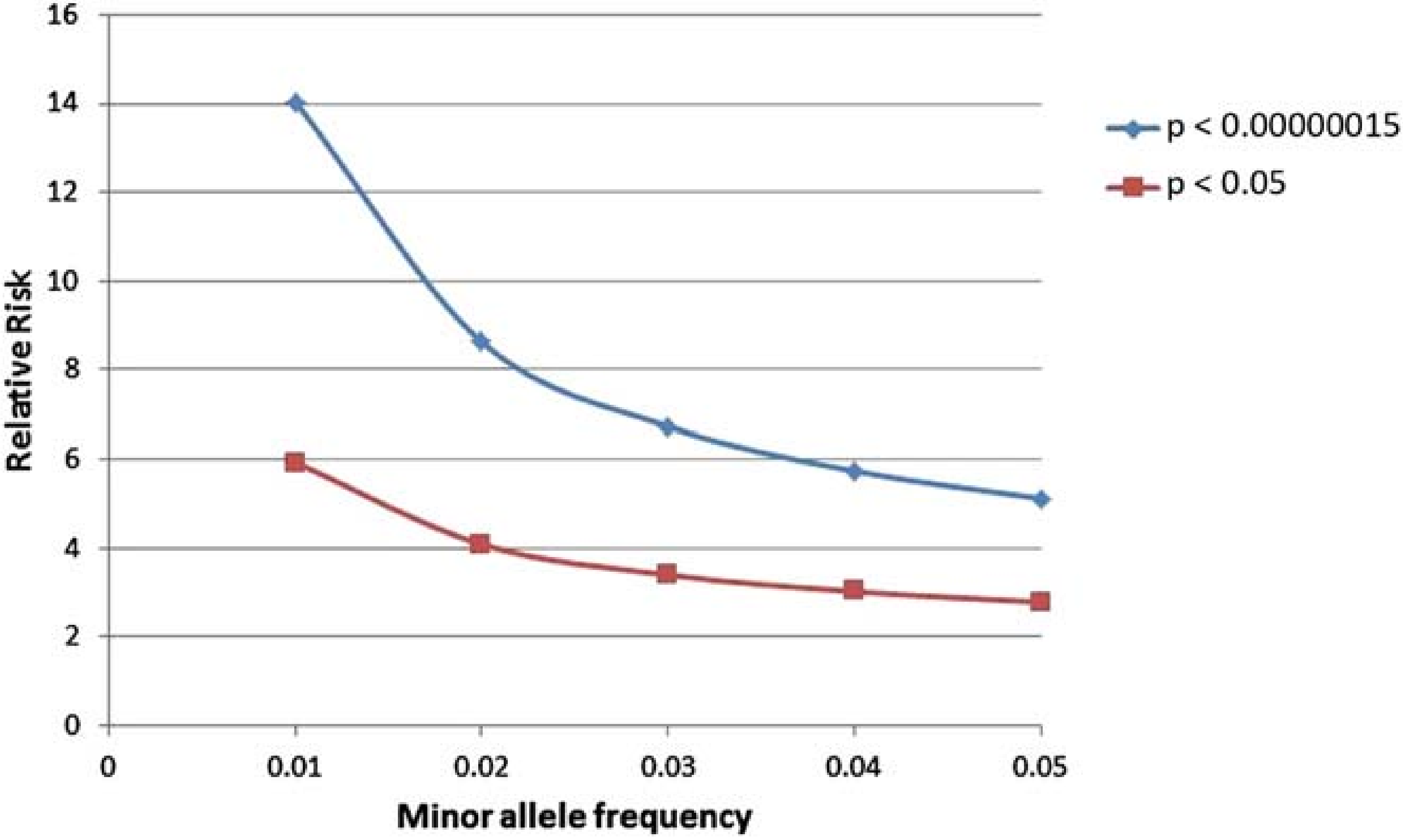

Table 1. Summary of Outcomes for Variants by Original Inclusion Criteria

Reason for original inclusion (n=5788)		Failed or excluded	Present in controls, p< 0.05				Absent in controls		
			In African-Americans only	In white only	In both ancestral groups	In neither ancestral group	In 0 additional cases	In 1 additional case	In >1 additional case
p<0.05 (n=428)	Allelic (n=242)	30	3	28	2	173	4	2	0
	recessive (n=186) [a]	37	5	11	1	131	1	0	0
in >1 case and 0 controls (n=861)	allelic (n=843)	94	9	15	0	476	192	34	23
	recessive (n=18) [a]	2	0	1	0	12	2	1	0
in 1 case and 0 controls (n=4498)	allelic (n=4280)	434	28	9	0	1443	1929	301	136
	recessive (n=218) [a]	36	2	3	0	103	60	7	7

[a] p-values given for recessive tests; counts are for homozygous genotypes.

Table 2. SNVs Seen in Two or More Follow-up Schizophrenia Cases that were Absent in All Study Controls and Also Invariant in the NHLBI Exome Sequenced Cohort. A full list of variants, with further details, including those that were absent in our study but either seen in NHLBI controls or not covered in the NHLBI Exome Sequenced Cohort, can be seen in supplementary table 1.

VARIANT (chr_hg18 position_variant allele)	Gene	MIM Number	Transcript (Ensembl 50_36l)	RefSeq mRNA	Annotated function;position of sequence and amino acid change	Total cases	European American case counts (hom/het/ref)	European American control counts (hom/het/ref)	African American case counts (hom/het/ref)	African American Control counts[a] (hom/het/ref)	Other case counts (hom/het/ref)	Other controlcounts (hom/het/ref)	Total samples (NHLBI cohort)	European Amerian samples (NHLBI cohort)	African Amerian samples (NHLBI cohort)
13_32533098_T	*KL*	604824	ENST00000380099	NM_004795.3	NS; c.1882C>T; p.Arg628Cys	5	0/1/1564	0/0/1359	0/4/1186	0/0/679	0/0/30	0/0/82	5379	3510	1869
19_2785443_A	*ZNF554*	NA	ENST00000317243	NM_001102651.1	NS; c.1210G>A; p.Gly404Arg	4	0/3/1561	0/0/1359	0/1/1189	0/0/679	0/0/30	0/0/82	5351	3508	1843
20_34245584_G	*EPB41L1*	602879	ENST00000344237	-	NS; c.59A>G; p.His20Arg	4	0/4/1560	0/0/1359	0/0/1187	0/0/679	0/0/30	0/0/82	5379	3510	1869
2_238150673_A	*RAB17*	602206	ENST00000264601	NM_022449.3	NS; c.401C>T; p.Thr134Met	4	0/0/1565	0/0/1359	0/4/1186	0/0/679	0/0/30	0/0/82	5379	3510	1869
21_44639800_T	*TRPM2*	603749	ENST00000397928	NM_003307.3	NS; c.1870G>T; p.Asp624Tyr	4	0/3/1561	0/0/1359	0/1/1189	0/0/679	0/0/30	0/0/82	5376	3508	1868
1_55290595_T	*PCSK9*	607786	ENST00000302118	NM_174936.3	NS; c.580C>T; p.Arg194Trp	3	0/0/1565	0/0/1358	0/3/1187	0/0/679	0/0/30	0/0/82	5379	3510	1869
11_111555378_T	*BCDO2*	611740	ENST00000393032	NM_001037290.2	X; c.154C>T; p.Arg52*	3	0/3/1562	0/0/1359	0/0/1190	0/0/679	0/0/30	0/0/82	5379	3510	1869
11_35290538_T	*SLC1A2*	600300	ENST00000395750	NM_001195728.2	NS; c.317G>A; p.Arg106His	3	0/2/1563	0/0/1359	0/1/1189	0/0/679	0/0/30	0/0/82	5379	3510	1869
12_12952740_A	*GPRC5A*	604138	ENST00000014914	NM_003979.3	NS; c.290G>A; p.Arg97His	3	0/0/1565	0/0/1359	0/3/1187	0/0/679	0/0/30	0/0/82	5379	3510	1869
15_41875227_T	*SERINC4*	614550	ENST00000299969	-	Splice; c.918-1C>T;	3	0/3/1562	0/0/1359	0/0/1190	0/0/679	0/0/30	0/0/82	5379	3510	1869
16_1083724_C	*C1QTNF8*	614147	ENST00000328449	NM_207419.3	X; c.537C>G; p.Tyr179*	3	0/3/1557	0/0/1359	0/0/1190	0/0/679	0/0/30	0/0/82	5368	3500	1868
16_30958421_A	*STX4*	186591	ENST00000313843	NM_004604.3	NS; c.761G>A; p.Arg254His	3	0/1/1564	0/0/1359	0/2/1188	0/0/679	0/0/30	0/0/82	5379	3510	1869
16_69543869_T	*HYDIN*	610812	ENST00000316490	-	NS; c.6340G>A; p.Gly2114Arg	3	0/3/1562	0/0/1359	0/0/1190	0/0/679	0/0/30	0/0/82	4650	3215	1435
17_7769756_A	*KCNAB3*	604111	ENST00000303790	NM_004732.3	X; c.508C>T; p.Arg170*	3	0/3/1562	0/0/1359	0/0/1190	0/0/679	0/0/30	0/0/82	5379	3510	1869
19_19506937_A	*AC011448.5*	NA	ENST00000397179	NM_198537.3	NS; c.413G>A; p.Arg138Gln	3	0/3/1562	0/0/1359	0/0/1190	0/0/679	0/0/30	0/0/82	5167	3422	1745
19_60041100_T	*KIR3DL3*	610095	ENST00000391729	-	X; c.328C>T; p.Gln110*	3	0/0/1559	0/0/1355	0/3/1186	0/0/679	0/0/30	0/0/82	5264	3418	1846
2_32578230_C	*BIRC6*	605638	ENST00000261359	-	NS; c.8497G>C; p.Glu2833Gln	3	0/0/1542	0/0/1337	0/3/1187	0/0/679	0/0/29	0/0/75	5379	3510	1869
3_114481460_A	*BOC*	608708	ENST00000355385	NM_033254.2	NS; c.2120G>A; p.Arg707His	3	0/2/1563	0/0/1359	0/0/1190	0/0/679	0/1/29	0/0/82	5379	3510	1869
3_42714733_A	*HHATL*	614071	ENST00000341477	-	NS; c.325C>T; p.Arg109Cys	3	0/1/1564	0/0/1359	0/0/1190	0/0/679	0/2/28	0/0/82	5379	3510	1869
6_117745049_A	*ROS1*	165020	ENST00000368507	-	NS; c.6067A>T; p.Met2023Leu	3	0/0/1565	0/0/1359	0/3/1187	0/0/679	0/0/30	0/0/82	5379	3510	1869
7_65077123_A	*GUSB*	611499	ENST00000345660	-	X; c.916C>T; Arg305*	3	0/3/1562	0/0/1359	0/0/1190	0/0/679	0/0/30	0/0/82	5379	3510	1869
7_82419591_T	*PCLO*	604918	ENST00000333891	-	NS; c.8407G>A; p.Val2803Ile	3	0/0/1512	0/0/1312	0/3/1186	0/0/679	0/0/29	0/0/80	5032	3389	1643
7_86891201_A	*ABCB4*	171060	ENST00000394680	-	NS; c.2168G>T; p.Gly723Val	3	0/0/1565	0/0/1359	0/3/1187	0/0/679	0/0/30	0/0/82	5379	3510	1869
7_88803894_A	*ZNF804B*	NA	ENST00000333190	NM_181646.2	NS; c.3662C>A; p.Ala1221Asp	3	0/3/1562	0/0/1359	0/0/1190	0/0/679	0/0/30	0/0/82	5379	3510	1869

[a.] **No variant had an rs number; NS= non-synonymous, X=Nonsense**