npg

SHORT REPORT

No evidence for rare recessive and compound heterozygous disruptive variants in schizophrenia

Douglas M Ruderfer[1,2,3], Elaine T Lim[4], Giulio Genovese[3], Jennifer L Moran[3], Christina M Hultman[5], Patrick F Sullivan[6], Steven A McCarroll[3,7], Peter Holmans[8], Pamela Sklar[1,9] and Shaun M Purcell*,[1,2,3,4]

Recessive inheritance of gene disrupting alleles, either through homozygosity at a specific site or compound heterozygosity, have been demonstrated to underlie many Mendelian diseases and some complex psychiatric disorders. On the basis of exome sequencing data, an increased burden of complete knockout (homozygous or compound heterozygous) variants has been identified in autism. In addition, using single-nucleotide polymorphism microarray data, an increased rate of homozygosity by descent, or autozygosity, has been linked to the risk of schizophrenia (SCZ). Here, in a large Swedish case-control SCZ sample (11 244 individuals, 5079 of whom have exome sequence data available), we survey the contribution of both autozygosity and complete knockouts to disease risk. We do not find evidence for association with SCZ, either genome wide or at specific loci. However, we note the possible impact of sample size and population genetic factors on the power to detect and quantify any burden that may exist.
European Journal of Human Genetics (2015) **23**, 555–557; doi:10.1038/ejhg.2014.228; published online 5 November 2014

INTRODUCTION

Schizophrenia (SCZ) is a highly heritable (h^2 ~ 0.6–0.8), debilitating disorder with an increasing number of implicated genetic loci[1–7] from genome-wide association studies (GWAS) of common single-nucleotide polymorphisms (SNPs) and rare copy number variants (CNVs). Recently, large-scale exome-sequencing studies have searched for rare variants that strongly increase risk,[8,9] although the relatively high baseline rate of rare neutral coding variation makes this challenging.[10] To focus on rare variants that are *a priori* more likely to be highly penetrant for disease, one approach is to study *de novo* mutations, as these are effectively uncensored by selection pressures.[11] A second strategy, adopted here, is to focus on the rare gene-disruptive variants, defined as nonsense, essential splice site or frameshift, where both copies of a gene are affected through homozygosity at a single site or compound heterozygosity (if an individual carries different disruptive variants on both the maternally and paternally derived copies of the gene). With respect to any one gene, we refer to these as 'complete knockout' variants. A recent study in autism reported complete knockout variants at a rate of 1 for every 20 healthy individuals (0.05 per person) with a twofold higher rate in cases.[12]

Whereas sequencing allows for direct discovery of rare complete knockout genotypes, rates of autozygosity (homozygosity by descent) can also be informative under recessive models. Previous studies have used SNP microarrays to identify multi-megabase runs of homozygosity (ROH), that likely reflect autozygosity arising from a degree of recent inbreeding.[13,14] A study of 9388, predominantly European, SCZ cases and 12 456 controls demonstrated a significant positive correlation between the extent of ROH per individual and the disease risk.[15] However, they did not identify any significant individual loci.

Here we apply both approaches in a large Swedish population-based sample[8,16] of SCZ cases and controls: we first test the recessive model indirectly by comparing the SNP-derived ROH burden in cases *versus* controls, and second, we use exome sequencing to assess directly the burden of complete knockout variants.

METHODS

Sample ascertainment and data generation

Samples were collected from national registers in Sweden, cases were identified through the Swedish Hospital Discharge Register and controls were randomly selected from Swedish population registers. All subjects were appropriately consented and the Institutional Human Subject Committees approved the research. Sequencing was performed at the Broad Institute using Illumina (San Diego, CA, USA) GAII or HiSeq2000 machines. Exome capture was done using either version 1 or version 2 of the Agilent (Santa Clara, CA, USA) SureSelect Human All Exon Kit. Additional detail on sample ascertainment and sequence data generation has been previously published.[8] BAM and VCF files are available in the dbGaP study phs000473.v1 (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000473.v1.p1).

Runs of homozygosity calling

We utilized previously published[16] SNP imputation data of the entire Swedish data set representing a superset of the individuals that have been sequenced. We selected ~ 55 k SNPs in approximate linkage equilibrium from high-quality genotypes called from imputed dosages in 5001 cases and 6243 controls. We sought to detect ROH using PLINK[17] with parameter settings ROH length > 1 Mb, at least one SNP per 100 kb and allowance of a single heterozygous event within a window of 50 SNPs, splitting calls if spanning 500 kb with no

marker. This method and parameter selection represents those suggested in the literature after a careful comparison[18] and is very similar to the approach used previously on a subset of these data.[15]

Compound heterozygous identification

To evaluate the role of complete knockout variants in SCZ we implemented a method to identify these variants within the framework of the statistical package Plink/Seq (http://research.mssm.edu/statgen/software.html). The method analyzes each transcript independently, ensuring that both the disruptive variants are affecting the same protein. As most pairs of variants will not be within the distance necessary to phase (length of sequencing read), we employed a previously used statistical phasing approach to identify compound heterozygous events.[12] We defined compound heterozygosity as an individual carrying a heterozygous genotype on each chromosome at different locations within the same gene.

Assessing significance using permutation

Significance of complete knockout variants was assessed by permuting individual alleles independently assuming Hardy–Weinberg equilibrium (HWE). Permutation was performed within 296 groups of individuals matched on IBS from GWAS data to control for subtle population substructure (average number of individuals per group = 17). We assessed the significance of gene sets by summing across multiple variants assuming independence.

RESULTS

Among 5001 cases and 6243 controls, we identified 4117 ROH in 2001 individuals; 16% of individuals carried at least one ROH. The average length of ROH was 8 Mb (median length of 6.8 Mb) and ranged from 1.6 to 55.2 Mb. We also quantified each individual's genome-wide degree of homozygosity using Wright's inbreeding coefficient (F). After including the first 10 multidimensional scaling (MDS) components to account for ancestry, homozygosity was similar in cases and controls as assessed by the number of ROH (0.38 cases, 0.32 controls, $P = 0.60$), total kilobases of ROH (3123 kb cases, 2544 kb controls, $P = 0.88$) or F (0.0026 cases, 0.0029 controls, $P = 0.09$, Table 1).

We next looked for complete knockout variants in a subset of 2477 cases and 2481 controls for whom exome sequence data were available.[8] We identified 462 complete knockout rare (minor allele frequency < 5%) genotypes at 193 genomic locations, a rate of 0.09 per person with no increased burden in cases (229 case; 233 control, Table 2). Of these, 405 were homozygous and 57 were compound heterozygous, with no significant enrichment in either class. Analysis of nonsynonymous and silent complete knockout variants yielded similar null results (data not shown).

Testing each site individually for association with increased risk (one-sided), 30 (16%) had $P < 0.05$ but none surpassed the Bonferroni correction for multiple testing (Supplementary Table S1). This excess of nominally significant tests is a product of testing only those sites carrying at least one complete knockout genotype: simulating null genotype data under HWE using the observed allele frequencies, we

found a similar proportion of nominally significant sites when following a similar analytic procedure (data not shown).

For the X chromosome, males and females were analyzed separately. A single hemizygous disruptive variant in males results in a complete loss of gene function as in the autosomal complete knockout scenario. Using a minor allele frequency cutoff of 0.25%, we observed 49 disruptive sites outside of the pseudoautosomal regions, a burden of 41 in cases and 41 in controls (one sided Fisher's exact test $P = 0.5$). No individual site achieved nominal significance (Supplementary Table S2).

Finally, we selected gene sets previously demonstrated to have a significant polygenic burden of rare disruptive alleles (primarily observed as a single heterozygous genotype) in the same set of individuals.[8] Across 2546 genes in seven sets, we identified 35 complete knockout or male X chromosome disruptive genotypes, 19 in cases and 16 in controls ($P = 0.38$; Supplementary Table S3). Although the set of genes carrying disruptive de novo alleles in SCZ displayed nominal significance (4:0, one-sided $P = 0.027$), this does not survive correction for the seven tests. We also considered genes with known syndromic autosomal recessive causes of autism or intellectual disability.[19] Three case genotypes and five control genotypes were found across 119 intellectual disability genes and one case genotype and zero control genotypes were found across 28 autism genes; neither test reached nominal significance (P intellectual disability = 0.72, P autism = 0.57).

DISCUSSION

Overall, we did not observe a significant enrichment of ROH in SCZ cases. In a larger sample (21 844), Keller et al.[15] reported a significantly higher burden of ROH in SCZ cases, suggesting that the current study may be under-powered. We note that the Keller et al. study included a small proportion of the Swedish individuals reported here, although the authors did not find evidence for increased ROH in the Swedish subset, consistent with our results. This suggests that population genetic and ancestry characteristics of a sample could be a further determinant of the relative role of rare recessive variants in the risk for SCZ.

The amount of recent inbreeding will contribute to the levels of homozygosity present in a population. Our sample was ascertained from Sweden, but includes a proportion of individuals of Finnish descent. Utilizing populations with recent inbreeding will increase the frequency of these events and the power to detect enrichment, an approach that has been used successfully for CNVs.[20]

In conclusion, this sample does not lend support for a major role of disruptive recessive or compound heterozygous variants in the risk of SCZ. There was no overall burden of ROH genome wide or of complete knockout variants exome wide, no specific genomic site was significantly associated and no gene achieved significance after correction. These results largely mirror those of de novo SCZ trio studies (Fromer et al.[11]), in which there is no significantly increased

Table 1 Summary of ROH and inbreeding coefficient (F) in full Swedish GWAS sample, P-values are from a logistic regression with 10 MDS components and wave as covariates

	Cases	Controls	P-value
N	5001	6243	
ROH rate	0.38	0.32	0.60
ROH (kb)	3123	2544	0.88
F	0.0026	0.0029	0.09

Table 2 Summary of two-hit disruptive genotypes

	Case	Control
N	2477	2481
Homozygous	200	205
Compound heterozygous	29	28
All two-hit	229	233
Rate	0.09	0.09

burden overall and very few individual genes are unambiguously implicated, although for *de novo* mutations, broader effects can be seen across subsets of functionally-related genes. Rare recessive and compound heterozygous sites may still have a role in disease risk although larger samples or different analytic approaches may be required.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1 Bassett AS, Chow EW, AbdelMalik P, Gheorghiu M, Husted J, Weksberg R: The schizophrenia phenotype in 22q11 deletion syndrome. *Am J Psychiatry* 2003; **160**: 1580–1586.
2 International Schizophrenia Consortium: Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* 2008; **455**: 237–241.
3 Ripke S, Sanders AR, Kendler KS *et al*: Genome-wide association study identifies five new schizophrenia loci. *Nature genetics* 2011; **43**: 969–976.
4 O'Donovan MC, Craddock N, Norton N *et al*: Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet* 2008; **40**: 1053–1055.
5 Walsh T, McClellan JM, McCarthy SE *et al*: Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 2008; **320**: 539–543.
6 McCarthy SE, Makarov V, Kirov G *et al*: Microduplications of 16p11.2 are associated with schizophrenia. *Nat Genet* 2009; **41**: 1223–1227.
7 Lichtenstein P, Yip BH, Bjork C *et al*: Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* 2009; **373**: 234–239.
8 Purcell SM, Moran JL, Fromer M *et al*: A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 2014; **506**: 185–190.
9 Kiezun A, Garimella K, Do R *et al*: Exome sequencing and the genetic basis of complex traits. *Nat Genet* 2012; **44**: 623–630.
10 Zuk O, Schaffner SF, Samocha K *et al*: Searching for missing heritability: Designing rare variant association studies. *Proc Natl Acad Sci USA* 2014; **111**: E455–E464.
11 Fromer M, Pocklington AJ, Kavanagh DH *et al*: *De novo* mutations in schizophrenia implicate synaptic networks. *Nature* 2014; **506**: 179–184.
12 Lim ET, Raychaudhuri S, Sanders SJ *et al*: Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron* 2013; **77**: 235–242.
13 Lencz T, Lambert C, DeRosse P *et al*: Runs of homozygosity reveal highly penetrant recessive loci in schizophrenia. *Proc Natl Acad Sci USA* 2007; **104**: 19942–19947.
14 Power RA, Keller MC, Ripke S *et al*: A recessive genetic model and runs of homozygosity in major depressive disorder. *Am J Med Genet B Neuropsychiatr Genet* 2014; **165**: 157–166.
15 Keller MC, Simonson MA, Ripke S *et al*: Runs of homozygosity implicate autozygosity as a schizophrenia risk factor. *PLoS Genet* 2012; **8**: e1002656.
16 Ripke S, O'Dushlaine C, Chambert K *et al*: Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet* 2013; **45**: 1150–1159.
17 Purcell S, Neale B, Todd-Brown K *et al*: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
18 Howrigan DP, Simonson MA, Keller MC: Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics* 2011; **12**: 460.
19 Betancur C: Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res* 2011; **1380**: 42–77.
20 Morrow EM, Yoo SY, Flavell SW *et al*: Identifying autism loci and genes by tracing recent shared ancestry. *Science* 2008; **321**: 218–223.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)