A family-based study of common polygenic variation and risk of schizophrenia

Molecular Psychiatry (2011) **16,** 887–888; doi:10.1038/mp.2011.34; published online 12 April 2011

Recent studies have suggested that a large number of common variants may, in aggregate, underlie a substantial proportion of the heritability of complex traits such as schizophrenia, multiple sclerosis and height, even though the effect of each individual variant is typically very small.[1–5] A persistent concern for population-based genome-wide association studies, however, is that subtle population stratification could lead to bias. Although the abovementioned studies adhered to 'best-practice' for controlling bias in genome-wide association studies, some authors have speculated, with respect to the International Schizophrenia Consortium's schizophrenia study, that 'cryptic population stratification could substantially affect (the results)'.[6] In this study, we repeat the analysis described in ref. 1 using a family-based sample, to ask directly whether this is in fact a viable contention, as family-based designs, by definition, control for 'cryptic population stratification' completely.

We analyzed genome-wide association studies data on 694 parent-offspring schizophrenia trios from Bulgaria. Of these probands, 360 were cases in the original ISC manuscript. All samples were ascertained and genotyped on Affymetrix 6.0 arrays following the protocols described in ref. 1, with the exception that we further required single nucleotide polymorphisms (SNPs) to have greater than 99% genotyping and no more than one Mendel error. The original analysis in ref. 1 created 'scores' in case/control target samples, where the score per individual was a weighted sum of 'risk alleles', with the weights and the alleles determined by standard tests of association in an independent 'discovery' case/control sample. In this study, we designated the entire ISC sample (excluding all Bulgarian individuals) as the 'discovery' sample; our target comparison was between the transmitted and untransmitted alleles of the Bulgarian trios. In other words, we asked whether putative risk alleles from a case/control study tend to, on average, be overtransmitted to offspring with schizophrenia. This analysis does not constitute an independent replication of that in ref. 1 because of the overlapping cases; rather, our current purpose is solely to exclude bias because of cryptic population stratification as a possible source of inflated type I error.

Although within-family association statistics are free from bias because of population stratification, they are in fact susceptible to certain technical biases that do not impact population-based studies, arising from non-random genotyping error.[7] Particularly for low-frequency variants, heterozygotes are more likely to be misclassified as the common than the rare homozygote. In family-based studies, this leads to a bias in that the common allele is overtransmitted. This is because, if only one parent is heterozygous and transmits the minor allele, a miscall indicating the common homozygote in the parent will result in a Mendel error, whereas the same error in the offspring will result in an apparent transmission of the common allele. We observed this phenomenon in our total trio data set; using the transmission disequilibrium test,[8] the mean log of the odds ratio for the minor allele is –0.004, which is significantly different from 0 ($P < 2.2 \times 10^{-16}$). To overcome this bias, we removed 47 families with a relatively high number of Mendel errors (>300 out of a total of 525 571 SNPs). We also removed all SNPs with less than complete genotyping (because SNPs with lower call rates tend to have more miscalling errors), or a minor allele frequency less than 2%. The resulting data set showed no significant bias ($P = 0.34$).

We designated 'score alleles' in the discovery sample (2794 cases, and 2976 controls from ref. 1) for a subset of 45 544 SNPs selected to be in approximate linkage equilibrium, preferentially retaining SNPs with higher association in the discovery sample. In the target sample, we calculated the weighted sum of 'score alleles' at various discovery *P*-value thresholds ($P < 0.01$, 0.05, 0.1, 0.2, 0.3, 0.4 and 0.5) for the transmitted and untransmitted alleles.

Figure 1 shows the results of the primary analysis, in which we compared scores between transmitted and untransmitted chromosomes using logistic regression as in ref. 1. At every discovery *P*-value

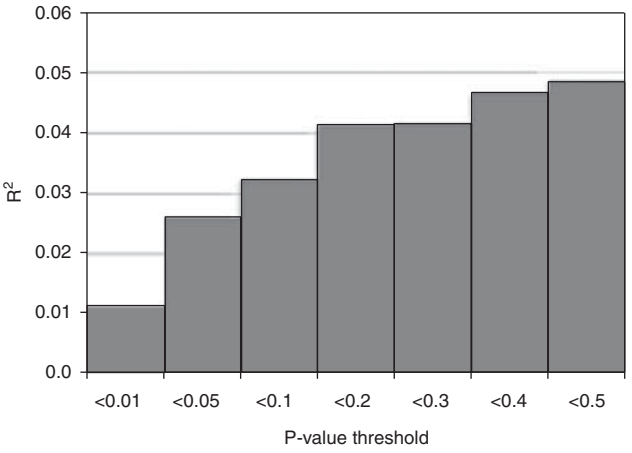

Figure 1 Variance explained by *P*-value threshold, corresponding significance at each threshold ($P = 0.0044$, 5.4×10^{-06}, 3.2×10^{-08}, 5.1×10^{-10}, 1.2×10^{-11}, 2.2×10^{-12} and 1.03×10^{-12}).

threshold, transmitted chromosomes had significantly higher rates of the score alleles, (ie, alleles that were more common in cases compared with controls in the independent discovery analysis). The estimate of variance explained (Nagelkerke's pseudo-R^2) by the observed score reached $\sim 5\%$, which is similar to values from ref. 1. We confirmed these results using a one-sided *t*-test on the difference in score between matched pairs of transmitted and untransmitted chromosomes (data not shown).

In summary, individuals with schizophrenia from distinct European populations show enrichment across a very large number of SNPs for the same sets of common alleles. As previously discussed,[1] this observation is consistent with a highly polygenic model of disease risk involving causal common variation. Further to the arguments already presented in ref. 1, we can reject cryptic population stratification as a viable alternative explanation.

Conflict of interest

The authors declare no conflict of interest.

DM Ruderfer[1,2,3], G Kirov[4], K Chambert[3], JL Moran[3], MJ Owen[4], MC O'Donovan[4], P Sklar[1,2,3] and SM Purcell[1,2,3]

[1]*Psychiatric and Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA;* [2]*Center for Human Genetics Research, Massachusetts General Hospital, Boston, MA, USA;* [3]*Stanley Center for Psychiatric Research, Broad Institute, 7 Cambridge Center, Cambridge, MA, USA and* [4]*MRC Centre for Neuropsychiatric Genetics & Genomics, School of Medicine, Cardiff University, Department of Psychological Medicine and Neurology, Heath Park, Cardiff, UK*
E-mail: shaun@pngu.mgh.harvard.edu

References

1 Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF *et al. Nature* 2009; **460**: 748–752.
2 Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR *et al. Nat Genet.* 2010; **42**: 565–569.
3 Bush WS, Sawcer SJ, de Jager PL, Oksenberg JR, McCauley JL, Pericak-Vance MA *et al. Am J Hum Genet.* 2010; **86**: 621–625.
4 Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F *et al. Nature.* 2010; **467**: 832–838.
5 Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU *et al. Nat Genet.* 2010; **42**: 937–948.
6 McClellan J, King MC. *JAMA.* 2010; **303**: 2523–2524.
7 Neale BM, Lasky-Su J, Anney R, Franke B, Zhou KX, Maller JB *et al. Am J Med Genet B* 2008; **147B**: 1337–1344.
8 Spielman RS, McGinnis RE, Ewens WJ. *Am J Hum Genet* 1993; **52**: 506–516.