

DOI : 10.31363/2313-7053-2019-4-1-8-11

Using polygenic risk score approaches to investigate the common-variant genetic architecture of schizophrenia

Escott-Price V., Professor in Biostatistics and Bioinformatics
Neuroscience and Mental health research institute, Dementia Research institute, Cardiff University, UK

Summary. In this paper we present a summary of our association analyses of schizophrenia polygenic risk score with a number of phenotypes in a large cohort of people from the UK population (N=442,192). We show that individuals with higher genetic loading to schizophrenia who have not been diagnosed with neurodevelopmental disorders are likely to have some cognitive deficits. Although these deficits may be subtle, they can result in significant effects on educational attainment and professional occupation. We also show that the relationship between schizophrenia liability and fecundity is consistent with sexual selection, with liability in unaffected people being associated with a net increase in fecundity, thereby supporting the persistence of schizophrenia risk alleles.

Keywords: polygenic risk score, schizophrenia, genome-wide association study.

Применение подходов с использованием полигенных шкал риска для изучения генетической архитектуры шизофрении

Эскотт-Прайс В., профессор биостатистики и биоинформатики
Исследовательский институт нейронаук и психического здоровья, Институт исследования деменции, Кардиффский университет, Великобритания

Резюме. В данной статье представлено резюме нашего ассоциативного исследования шкал полигенного риска шизофрении с различными фенотипами в крупной когорте населения Великобритании (N=442,192). Мы показали, что у людей с более высоким генетическим грузом по шизофрении без диагностированных нарушений развития нервной системы с большей вероятностью отмечались некоторые когнитивные нарушения. Хотя данные нарушения могли быть выражены в мягкой степени, они значительно влияли на уровень образования и профессиональные достижения. Мы также отметили, что взаимосвязь между подверженностью шизофрении и плодовитостью согласуется с половым отбором, при этом подверженность у не затронутых шизофренией индивидумов связана с чистым повышением плодовитости. Это в свою очередь ведет к сохранению аллелей риска шизофрении в популяции.

Ключевые слова: полигенные показатели риска, шизофрения, полногеномный поиск ассоциаций.

Introduction

Schizophrenia has a substantial heritability [2, 14] to which alleles across the frequency spectrum contribute [11, 17]. A substantial component of liability to schizophrenia is conferred by risk alleles that are common and which individually confer small effects on risk [8, 9]. Genome-wide association (GWA) studies have proved a powerful method to identify susceptibility alleles for complex diseases. GWA study datasets can be used to determine a polygenic contribution of common SNPs that show disease association but fail to meet the accepted P-value threshold for genome-wide significance ($p < 5 \times 10^{-8}$). The Polygenic risk score (PRS) approach encompasses more of the causal variance, as a genetic risk score is calculated based not solely on genome-wide significant polymorphisms, but on all nominally associated variants at a defined significance threshold (typically thousands of variants). This type of analysis has recently shown significant polygenic contribution in complex genetic diseases. The method can also be used to identify overlap in genetic determinants between related disorders, e.g. schizophrenia and bipolar disorder; depression and anxiety [6]. While the polygenic method undoubtedly introduces noise by including some variants that are not involved in

disease susceptibility (i.e. false positives), this is more than offset by the increased power to identify those at highest/lowest risk of disease. Trait differences between those with highest/lowest polygenic risk scores have also been identified. For example, in a study of the Lothian Birth Cohort, increased polygenic risk of schizophrenia was associated with lower cognitive ability at age 70 and greater relative decline in general cognitive ability between the ages of 11 and 70 [1].

Polygenic risk scores

Originally PRSs were designed to summarise genome-wide genotype data into a single variable that measures genetic liability to a disorder or a trait and have originally been described by the International Schizophrenia Consortium [8]. The PRS analysis requires two independent datasets. For the first, result data is sufficient as this dataset is used to select the SNPs, the risk score alleles and their genetic effects. The second dataset is used to test whether the polygenic risk scores differ in cases and controls and requires the genotypes for each individual. The PRS is then calculated from genome-wide association study summary statistics, summing the number of risk alleles carried by an individual, weighted by the effect size from the discovery GWAS.

PGC and CLOZUK

To define risk alleles, we used the largest available schizophrenia GWAS comprising a meta-analysis of two large studies which included 40,675 schizophrenia cases and 64,643 controls [7]. The first study by the Schizophrenia Working Group of the Psychiatric Genomics Consortium (PGC2) [17] reported the results of a multi-stage schizophrenia genome-wide association study where the authors identified 128 independent associations spanning 108 conservatively defined loci that meet genome-wide significance. The second study (CLOZUK), the largest single cohort genome-wide association study of schizophrenia (11,260 cases and 24,542 controls), where schizophrenia cases blood samples were collected from those with treatment-resistant schizophrenia (TRS) in the UK through the mandatory clozapine. The meta-analysis of these two studies provided 50 novel genetic loci for schizophrenia [7].

UKBB

The UK Biobank [13] is a large prospective cohort of more than half a million residents of the UK for which genetic data and seasonality of birth data are available. We restricted the sample to those who self-reported as being of white UK or Irish ancestry as the schizophrenia GWAS which we used to define risk alleles was of primarily European Ancestry (N=442,192 individuals).

We implemented PRS analysis of the UK Biobank (UKBB) participants to investigate the polygenic architecture of schizophrenia using the powerful PGC+CLOZUK dataset to define risk alleles. For the analyses described here we selected markers, based upon the schizophrenia association significance threshold $p < 0.05$ in the PGC2+CLOZUK GWAS [7], to construct a polygenic score in the UK Biobank data. $P < 0.05$ is the threshold that maximally captures polygenic risk in current studies of schizophrenia [17]. Using the UK Biobank genotypes, we performed clumped pruning using parameters $r^2 = 0.2$, a physical distance threshold for clumping SNPs of 1Mb. The clumped pruning process retains SNPs that are the most significantly associated with schizophrenia while excluding SNPs at which the genotypes are correlated with the selected SNPs. The PRS was calculated from the effect size-weighted sum of associated alleles within each subject and were adjusted for the array batch effect and the population based principal components and then standardised by subtracting the population mean for PRS and dividing by the standard deviation.

Association of PRS to SZ

First, we assessed how well the SZ PRS predicted those people with SZ (N cases=660) and psychosis (N cases=276) in the UK Biobank sample. To minimise a potential overlap between CLOZUK and UK Biobank participants, we excluded 13 individuals who declared that they take clozapine. As expected the SZ PRS significantly predicted the people with schizophrenia ($B = 0.64$, $95\%CI = [0.56-0.71]$, $p = 7.2 \times 10^{-60}$,

$R^2 = 0.028$) and psychosis ($B = 0.49$, $95\%CI = [0.38-0.61]$, $p = 2.1 \times 10^{-16}$, $R^2 = 0.015$). Both effect sizes were positive, indicating that people with SZ and psychosis have higher genetic loading for SZ, although the effect size when predicting psychosis was slightly lower than for SZ. The prediction accuracy for SZ cases, as measured using the Area Under the receiver operating characteristic Curve (AUC), was $AUC = 67\%$.

Further analyses aim to determine how the increased genetic risk conferred by common genetic variation affects individuals in the general population, who do not have the disorder. Therefore, for further analyses we excluded additionally people with self-reported schizophrenia (N=660), depression (N=27,761), bipolar disorder (N=1,489) and people who have not developed the full diagnoses but had significant psychopathology (N=276).

Association of PRS to reproductive fitness

We compared the burden of schizophrenia risk alleles, as indexed by a polygenic risk score (PRS), carried by UK Biobank participants who did not have schizophrenia, with the number of offspring of those individuals. We found that higher schizophrenia liability was weakly but significantly associated with having more children ($B = 0.004$, $95\%CI = [0.0015, 0.0059]$, $p = 0.001$). The relationship was dependent on gender, with a positive correlation between number of children and liability in females ($B = 0.007$, $95\%CI = (0.004, 0.010)$, $p = 3.5 \times 10^{-6}$) whereas in males, higher liability was associated with being childless ($OR = 0.968$, $95\%CI = (0.957, 0.978)$, $p = 2.7 \times 10^{-9}$) [18,19].

Association of SZ PRS to educational attainment

Educational achievement in UK Biobank was coded by integer values 1 to 5, indicating the highest achieved level of education in descending academic order (1- College/University degree, 2- A/AS levels or equivalent (advanced academic qualifications taken post-compulsory education age), 3- O levels/GCSE levels or equivalent (largely academic qualifications taken at the end of compulsory education, 4- CSEs or equivalent (certification of at the end of compulsory education, less stringent than O levels), 5- NVQ/HND/HNC (vocational qualifications) 6- none of the above. We excluded individuals with "other professional qualifications" and those who did not answer.

Higher genetic loading to schizophrenia was weakly associated with lower attainment when analysed using ordinal regression model ($B = 0.010$, $SE = 0.003$, $p\text{-value} = 1.95 \times 10^{-4}$; given the numbering system, the positive B-coefficient indicates higher SZ PRS is correlated lower educational attainment). To compare our findings with previously published work, we compared those with a college or university level degree versus the rest (but found no significant evidence for association ($B = 0.0058$; $SE = 0.0033$; $p\text{-value} = 0.082$; here 1, is a college or university level degree, 0- others). In contrast, dichotomizing at the largely academic

qualifications (university though to GCSEs) versus the rest resulted in a highly significant association; higher schizophrenia polygenic risk score being associated with an absence of academic qualification ($B=-0.026$; $SE=0.0034$; $p=1.4 \times 10^{-14}$) [20].

Association of SZ PRS to cognitive measures

UK Biobank participants completed a range of cognitive tests. Different numbers of participants were asked by the Biobank to complete the various tests [3]. To enable direct comparison between association of cognitive scores with common genetic variation, we analysed Z-scored data as reported in [3]. Association analyses of SZ PRS with cognitive performance show that pairs matching, reaction time, fluid intelligence and symbol digit substitution tests were highly significantly associated with SZ PRS (p-values ranging from 6.3×10^{-7} to 1.4×10^{-46}), indicating that higher genetic liability to SZ is associated with poorer performance in some cognitive tests. The strongest association effect size was for fluid intelligence test ($B=-0.067$, $95\%CI=(-0.077, -0.058)$). Pairs matching, reaction time, fluid intelligence and symbol digit substitution tests remained significant after Bonferroni correction for the numbers of statistical tests. Poorer performance in digit span and trail making tests were also significantly associated with higher SZ PRS.

Summary

Patients with SZ have reduced cognitive performance, occupational and educational achievements [10] and tend to have on average over 10 years shorter lives [4, 5]. Previous reports by us and others have shown that even in unaffected individuals, (higher) trait liability as defined by SZ PRS [15,16] or CNV carrier status [3,12] is associated with reduced performance for several measures of cognitive function. We have shown that in the UK Biobank sample, adults with higher genetic liability to schizophrenia had reduced cognitive performance and educational and occupational attainment, with all comparisons being highly significant [3, 20]. Our findings also suggest that at a population level, higher common variant genetic loading to schizophrenia is associated with a small reproductive advantage in female. The effect is not seen in males and there is evidence that those with the highest risk scores have fewer offspring. When males and females are considered together there is a net reproductive advantage associated with high PRS [19]. However, the effects of PRS are not sufficient to explain the persistence and allele frequencies of schizophrenia risk alleles in contemporary populations as they are small relative to the effects of schizophrenia itself on fecundity, which at a population level make a stronger impact.

References

- Demirkan A, Penninx BW, Hek K, Wray NR, Amin N, Aulchenko YS, et al. Genetic risk profiles for depression and anxiety in adult and elderly cohorts. *Molecular psychiatry*. 2011;16(7):773-83.
- Cardno AG, Gottesman, II. Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. *Am J Med Genet*. 2000;97(1):12-17.
- Kendall KM, Rees E, Escott-Price V, Eion M, Thomas R, Hewitt J, O'Donovan MC, Owen MJ, Walters JTR, Kirov G. Cognitive performance among carriers of pathogenic copy number variants: analysis of 152,000 UK Biobank subjects. *Biological Psychiatry*. 2017;82(2):103-110.
- Lawrence D, Hancock KJ, Kisely S. The gap in life expectancy from preventable physical illness in psychiatric patients in Western Australia: retrospective analysis of population based registers. *BMJ*. 2013;346:f2539.
- McGrath J, Saha S, Chant D, Welham J. Schizophrenia: a concise overview of incidence, prevalence, and mortality. *Epidemiol Rev*. 2008;30:67-76.
- McIntosh AM, Gow A, Luciano M, Davies G, Liewald DC, Harris SE, et al. Polygenic risk for schizophrenia is associated with cognitive change between childhood and old age. *Biological psychiatry*. 2013;73(10):938-43.
- Pardinas AF, Holmans, P, Pocklington, A.J., Escott-Price, V., et al Common schizophrenia alleles are enriched in mutation-intolerant genes and maintained by background selection. *Nature Genetics*. 2018;50(3):381-389.
- Purcell SM, Wray NR, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460(7256):748-752.
- Purcell SM, Moran JL, Fromer M, et al. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature*. 2014;506(7487):185-190.
- Rajji TK, Voineskos AN, Butters MA, Miranda D, Arenovich T, Menon M, et al. Cognitive performance of individuals with schizophrenia across seven decades: a study using the MATRICS consensus cognitive battery. *Am J Geriatr Psychiatry*. 2013;21:108-118.
- Rees E, Walters JT, Georgieva L, et al. Analysis of copy number variations at 15 schizophrenia-associated loci. *Br J Psychiatry*. 2014;204(2):108-114.
- Stefansson H, Meyer-Lindenberg A, Steinberg S, Magnúsdóttir B, Morgen K, Arnarsdóttir S, et al. CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature*. 2014;505:361-366.
- Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M4, Liu B, Matthews P, Ong G, Pell J, Silman A, Young A, Sprosen T, Peakman T, Collins R. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;2(3):e1001779.
- Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-

- analysis of twin studies. *Arch Gen Psychiatry*. 2003;60(12):1187-1192.
15. Whalley HC, Adams MJ, Hall LS, Clarke TK, Fernandez-Pujals AM, Gibson J, et al. Dissection of major depressive disorder using polygenic risk scores for schizophrenia in two independent cohorts. *Transl Psychiatry*. 2016;6:e938.
 16. Hagenaars SP, Harris SE, Davies G, Hill WD, Liewald DC, Ritchie SJ, et al. Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N=112 151) and 24 GWAS consortia. *Molecular psychiatry*. 2016;21:1624-1632.
 17. Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511(7510):421-427.
 18. Escott-Price V, Pardiñas AF, Santiago E, Walters J, Kirov G, Owen MJ, O'Donovan MC. The Relationship Between Common Variant Schizophrenia Liability and Number of Offspring in the UK Biobank. *Am J Psychiatry*. 2019;176(8):661-666.
 19. Escott-Price V, Pardiñas AF, Santiago E, Walters G, Kirov G, Owen MJ, O'Donovan MC. The Relationship Between Common Variant Schizophrenia Liability and Number of Offspring in the UK Biobank: Response to Lawn et al. *Am J Psychiatry*. 2019;176(7):574-575.
 20. Escott-Price V, Bracher-Smith M, Menzies G, Walters J, Kirov G, Owen M, O'Donovan M. Genetic liability to schizophrenia is negatively associated with educational attainment in UK Biobank. *Molecular Psychiatry*. 2018; doi: 10.1038/s41380-018-0328-6.

Сведения об авторах

Valentina Escott-Price—Professor in Biostatistics and Bioinformatics, Neuroscience and Mental health research institute, Dementia Research institute, Cardiff University, UK.