

**Polygenic risk score for schizophrenia is more strongly associated with ancestry than with schizophrenia**

Running head: **Polygenic risk score associated with ancestry**

David Curtis  
*d.curtis@ucl.ac.uk*

UCL Genetics Institute, UCL, Darwin Building, Gower Street, London WC1E 6BT.  
Centre for Psychiatry, Barts and the London School of Medicine and Dentistry, Charterhouse Square, London EC1M 6BQ.

## **Abstract**

### **Background**

The polygenic risk score (PRS) for schizophrenia, derived from very large numbers of weakly associated genetic markers, has been repeatedly shown to be robustly associated with schizophrenia in independent samples and also with other diseases and traits.

### **Aims**

To explore the distribution of the schizophrenia PRS in subjects of different ancestry.

### **Method**

The schizophrenia PRS derived from the large genome-wide association study carried out by the Psychiatric Genetics Consortium was calculated using the downloaded genotypes of HapMap subjects from eleven different ancestral groups. It was also calculated using downloaded genotypes of European schizophrenia cases and controls from the CommonMind Consortium.

### **Results**

The PRS for schizophrenia varied significantly between ancestral groups ( $p < 2 \times 10^{-16}$ ) and was much higher in African than European HapMap subjects. The mean difference between these groups was ten times as high as the mean difference between European schizophrenia cases and controls. The distributions of scores for African and European subjects hardly overlapped.

### **Conclusions**

The PRS cannot be regarded as simply a measure of the polygenic contribution to risk of schizophrenia and clearly contains a strong ancestry component. It is possible that this could be controlled for to some extent by incorporating principal components as covariates but doubts remain as to how it should be interpreted. The PRS derived from European subjects cannot be applied to non-Europeans, limiting its potential usefulness in clinical settings and raising issues of inequity in health provision. Previous studies which have used the PRS should be re-examined in the light of these findings.

### **Declaration of interest**

The author declares he has no conflict of interest.

### **Keywords**

Schizophrenia; polygenic risk score; RNA; expression.

## Introduction

The polygenic risk score (PRS) is derived by estimating the effect size of large numbers of genetic variants from a case-control discovery sample and combining these across the genotypes observed in other subjects (Visscher *et al.*, 2017). Typically, many thousands of variants are used to contribute to the PRS with the expectation that only a proportion of them will in fact be truly associated with the trait in question. The first major application of this approach was to a genome wide association study (GWAS) of schizophrenia and was used to justify the claim that thousands of alleles contribute to the risk of schizophrenia and that there is shared genetic risk between schizophrenia and bipolar disorder (Purcell *et al.*, 2009). Since this landmark publication, the observation that the PRS derived from one sample is associated with schizophrenia in another sample has been replicated numerous times and is probably one of the most robust observations in psychiatric genetics. The effect size is substantial and in a larger GWAS the odds ratio for schizophrenia risk between subjects with PRS in the highest or lowest decile was around 10 (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). This study made summary statistics available for each marker and this would allow other researchers to use it as a discovery sample to calculate the PRS in other samples. However it should be noted that in order to carry out a GWAS cases and controls need to be matched for ancestry because marker allele frequencies vary between populations. Thus the summary statistics were produced from a meta-analysis of a number of matched case-control samples which were overwhelmingly of European ancestry. Of a total of 38,131 schizophrenia cases and 114,674 controls, all were of European ancestry except 1,866 cases and 3,418 controls who were of East Asian ancestry. Using the summary statistics and applying them to other traits allows one to test whether the variants associated with increased risk of schizophrenia are also associated with other disorders and a recent review identified 31 articles examining the association of the schizophrenia PRS with other psychiatric and non-psychiatric phenotypes (Mistry *et al.*, 2017).

Given the way that it is derived and how it is used, there is a natural tendency to assume that the PRS produced for a trait is a measure of polygenic susceptibility to that trait. However a recent study of autism showed in a supplementary figure that the PRS for schizophrenia stratified by ancestry (Weiner *et al.*, 2017) and in a separate study investigating whether the PRS predicted gene expression it was noted that the PRS for schizophrenia was strongly correlated with the first principal component of marker genotypes (a proxy for ancestry) in both cases and controls of the Common-Mind Consortium (CMC) dataset (Curtis, 2017). Other studies have also shown that the PRS for schizophrenia varies between populations (Martin *et al.*, 2017). A recent study demonstrated that the PRS for both type 2 diabetes and coronary heart disease varies between populations and pointed out that this would need to be taken into account if attempting to estimate an individual's risk (Reisberg *et al.*, 2017).

In order to investigate this further, the present study set out to examine the distribution of the PRS in cohorts of different ancestry genotyped for the HapMap project and to compare this with the distribution between schizophrenia cases and controls in the CMC dataset which had initially revealed the principal component correlation.

## Methods

The merged post-QC phase I+II and III HapMap (International HapMap 3 Consortium *et al.*, 2010) genotype files were downloaded from [ftp://ftp.ncbi.nlm.nih.gov/hapmap/genotypes/2010-08\\_phaseI+II/forward/](ftp://ftp.ncbi.nlm.nih.gov/hapmap/genotypes/2010-08_phaseI+II/forward/). These subjects are from eleven different ancestral groups and are assumed not to be affected with schizophrenia. Since the prevalence of schizophrenia is only 1% it seems reasonable to assume, even in the absence of a formal psychiatric assessment, that at least the vast majority are unaffected. In order to obtain a PRS for schizophrenia, the file called *scz2.prs.txt.gz*, containing ORs and p values for 102,636 LD-independent single nucleotide polymorphism markers (SNPs), was downloaded from the Psychiatric Genetics Consortium (PGC) website ([www.med.unc.edu/pgc/results-and-downloads](http://www.med.unc.edu/pgc/results-and-downloads)). The ORs had been obtained by carrying out meta-analysis of 49 cohorts of European ancestry along with 3 cohorts of East Asian ancestry and using principal components to control for population stratification. This training set was produced as part of the previously reported PGC2 schizophrenia GWAS (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). SNPs from this dataset were then selected only if they had also been genotyped in all 11 of the HapMap cohorts, yielding a reduced set of 32,588 SNPs. HapMap subjects with genotyping call rate < 0.90 were removed, leaving a sample of 1,397.

The dataset used in the previous gene expression studies was downloaded from the CMC Knowledge Portal (<https://www.synapse.org/#!/Synapse:syn2759792/wiki/69613>) consisting of SNP genotypes and RNAseq results from frontal cortex samples originating from tissue collections at Mount Sinai NIH Brain Bank and Tissue Repository (MSSM), University of Pennsylvania Brain Bank of Psychiatric illnesses and Alzheimer's Disease Core Center (Penn) and The University of Pittsburgh NIH NeuroBioBank Brain and Tissue Repository (Pitt), collectively referred to as the CMC MSSM-Pitt-Penn dataset (Fromer *et al.*, 2016). Genotypes and expression levels were available for 258 subjects with schizophrenia and 279 controls, though in the current analysis only the genotypes were used. The distributions of ethnicities were reported to be similar between subjects with schizophrenia and controls (Caucasian 80.7%, African-American 14.7%, Hispanic 7.7%, East Asian 0.6%). The methods for obtaining the genotypes and expression data have been described by the authors of the original study (Fromer *et al.*, 2016). Genotyping was performed on the Illumina Infinium HumanOmniExpressExome 8 v 1.1b chip (Catalog #: WG-351-2301) using the manufacturer's protocol. QC was performed using PLINK to remove markers with: zero alternate alleles, genotyping call rate < 0.98, Hardy-Weinberg p-value <  $5 \times 10^{-5}$ , and individuals with genotyping call rate < 0.90. Marker alleles were phased to the forward strand, and ambiguously stranded markers were removed.

SNPs with p value < 0.05 in the PGC2 training set were selected and their log(OR) summed over sample genotypes using the `--score` function of *plink 1.09beta* in order to produce a PRS for each subject in both the HapMap and CMC datasets ([www.cog-genomics.org/plink/1.9/](http://www.cog-genomics.org/plink/1.9/)) (Purcell *et al.*, 2007, 2009; Chang *et al.*, 2015). The first 20 principal components for both datasets were produced using the `--pca` and `--make-rel` functions of *plink*.

As described previously (Curtis, 2017) ancestral outliers were removed from the CMC dataset by removing subjects with values for the first or second principal component exceeding -0.01. This left a fairly homogeneous sample of 264 subjects in which the PRS

was not correlated at  $p < 0.05$  with 19 of the first 20 principal components, although it was correlated with the 11th at  $p = 0.0019$ .

Statistical tests and data manipulation were carried out using R version 3.3.2 (R Core Team, 2014). The distributions of the PRS were compared between HapMap cohorts using anova and additionally a correlation analysis of PRS with the principal components was performed. For the CMC dataset, a t test was used to compare the PRS between subjects with schizophrenia and controls. In both samples, the residuals of the PRS after correction for the first 20 principal components were also compared.

## Results

The distribution of schizophrenia PRS between HapMap cohorts is shown in Table 1 and Figure 1. The SD of the PRS is fairly similar in all cohorts, varying from 1.71 to 2.59. However the means are very different. In CEU and TSI, the cohorts of European ancestry, the average PRS is -2.90 and -2.64. In ASW, LWK, MKK and YRI, the cohorts of African ancestry, the average PRS is 7.75, 9.48, 7.08 and 10.27. As can be seen from Figure 1, the differences between the cohorts are so marked that the scores for the European and African cohorts scarcely overlap. The anova testing for a difference between cohort means was formally statistically significant at  $p < 2 \times 10^{-16}$ . In the correlation analysis with the first 20 principal components, the PRS was highly significantly correlated with each of the first 6 principle components, at values of  $p < 2 \times 10^{-12}$  or lower. The residual PRS from this analysis did not significantly differ between cohorts.

In the CMC dataset, the PRS was significantly higher in subjects with schizophrenia (mean - 5.14) than controls (mean -6.08), difference = 0.94,  $p = 2.5 \times 10^{-7}$ . The residual PRS after correcting for the first 20 principal components in this dataset remained significantly higher in the subjects with schizophrenia,  $p = 7.4 \times 10^{-7}$ , though with a somewhat smaller difference between the means (0.41 versus -0.47, difference = 0.88).

## Discussion

There are striking differences in the schizophrenia PRS between cohorts with different ancestries. The differences between subjects of European and African ancestry are much larger, by a factor of around 10, than the differences between subjects with schizophrenia and controls of European ancestry.

The underlying mechanisms responsible for producing these results are not immediately obvious. Although differences in the frequencies of individual variants between groups of different ancestries would be not unexpected, it is harder to see why there should be a systematic effect such that alleles found more commonly in European schizophrenic subjects in the PGC should generally tend to be commoner in subjects with African ancestry. Two kinds of explanation for this effect suggest themselves.

The more benign explanation, from the point of view of the usefulness of the PRS, is that the PRS does indeed indicate genetic susceptibility to schizophrenia and that the contributing alleles are under weaker negative selection in African than non-African environments. This

could arise if SNPs used to generate the PRS did not themselves confer risk but were in LD with variants which did, and if these LD relationships were stronger in Europeans than in Africans, which might be expected given the bottleneck produced by the out of Africa migration. Then negative selection could operate on the causative variants and also pull down the frequencies of marker SNPs in LD with them. This would result in SNPs associated with schizophrenia being at lower frequency in Europeans than Africans, leading to a higher PRS in Africans.

The less benign explanation would be to say that the PRS is basically an indicator of African ancestry and that for some reason, perhaps through mechanisms such as social adversity, subjects in the PGC with schizophrenia have a higher African ancestry component than controls. However this does not seem a likely explanation because the PRS is associated with schizophrenia risk in a homogeneous sample after correction for principal components. Although there is some evidence that rates of schizophrenia are higher among subjects with African ancestry resident, for example, in the UK (Pinto, Ashworth and Jones, 2008), the prevalence appears to be fairly similar across the world and if anything lower in developing countries (Bhugra, 2005). It is certainly not plausible that the genetic risk associated in subjects with African ancestry could be as high as a naïve interpretation of the PRS would imply.

A number of other issues might be relevant. There might be different patterns of LD between different populations which impacted on the way variants were tagged and the extent to which marker SNPs tended to co-occur with each other. The SNPs contributing to the PRS are pruned so as not to be in LD with each other so if there were different LD relationships in a target population then this might be problematic. The p-value threshold chosen to select SNPs might be relevant in affecting how the PRS performs in different populations. More complex effects may be relevant and simulations in a coalescent-based framework have shown that genetic drift can produce unpredictable biases in any direction even in the absence of selection effects, so that in any particular scenario differences between populations may be essentially random (Martin *et al.*, 2017). Indeed, that study reported that the PRS for schizophrenia was actually lower in African than in European and Asian populations, although with smaller differences than found in the present study. These discrepancies may be relate to differences in the methods used to select SNPs and the use of a p value threshold of 0.01 rather than 0.05. They serve to illustrate that ancestry impacts on PRS in a complex way which will need to be systematically studied.

Whatever the explanation, these results have important implications for the interpretation of the PRS. As has been pointed out in the case of type 2 diabetes and coronary heart disease, the fact that the PRS varies according to ancestry would raise problems were one to attempt to use it to estimate an individual's risk in a clinical situation (Reisberg *et al.*, 2017). A recent study using the PRS for schizophrenia to distinguish cases with first episode psychosis from controls showed that it performed better in subjects with European rather than African ancestry (Vassos *et al.*, 2017). Any proposal to use the PRS to predict risk in individual subjects would need to explain very clearly how ancestry issues were to be dealt with. For certain applications, the PRS might be adjusted using principal components and other measures but given the very strong ancestry effects it would be reasonable to be concerned about whether such adjustments could ever be fully effective at the level of the individual. It may be worth noting that the PRS used for this study was obtained from analyses of mostly European samples in which principal components had already been incorporated but that

nevertheless the resulting measure still correlated very strongly with the principal components of an ancestrally heterogeneous sample. It might be that the PRS is a measure of relative polygenic risk in sufficiently homogeneous populations. However it would seem wise to exercise caution and to ensure that ancestry effects are appropriately dealt with whenever the PRS or similar methods are employed. If the PRS has some applicability, this would seem to be restricted to subjects having similar ancestry to that of the samples used to generate it. Given the results reported here one could not attempt to apply the PRS derived from European subjects to individual subjects with African ancestry. This highlights difficult questions about the fact that people of non-European ancestry may be disadvantaged if they cannot benefit from research which has been carried out largely or exclusively in European subjects. It is uncertain how much credence should be given to previous studies which may have used the PRS without appropriately accounting for ancestry effects. Although the current investigation has only dealt with the PRS for schizophrenia, it seems that similar considerations apply to the PRS for type 2 diabetes and coronary heart disease and this is likely to be the case for many other phenotypes (Reisberg *et al.*, 2017).

## **Acknowledgments**

The author thanks Michael O'Donovan for helpful discussion and pointing out a mistake in an earlier version. He also thanks the CommonMind, HapMap and PGC Consortia for making these datasets available. The CommonMind Consortium data generation was supported by funding from Takeda Pharmaceuticals Company Limited, F. Hoffman-La Roche Ltd and NIH grants R01MH085542, R01MH093725, P50MH066392, P50MH080405, R01MH097276, RO1-MH-075916, P50M096891, P50MH084053S1, R37MH057881 and R37MH057881S1, HHSN271201300031C, AG02219, AG05138 and MH06692. Brain tissue for the study was obtained from the following brain bank collections: the Mount Sinai NIH Brain and Tissue Repository, the University of Pennsylvania Alzheimers Disease Core Center, the University of Pittsburgh NeuroBioBank and Brain and Tissue Repositories and the NIMH Human Brain Collection Core. CMC Leadership: Pamela Sklar, Joseph Buxbaum (Icahn School of Medicine at Mount Sinai), Bernie Devlin, David Lewis (University of Pittsburgh), Raquel Gur, Chang-Gyu Hahn (University of Pennsylvania), Keisuke Hirai, Hiroyoshi Toyoshiba (Takeda Pharmaceuticals Company Limited), Enrico Domenici, Laurent Essioux (F. Hoffman-La Roche Ltd), Lara Mangravite, Mette Peters (Sage Bionetworks), Thomas Lehner, Barbara Lipska (NIMH).

## **Ethics approval**

This study was approved by the UCL Research Ethics Committee.

## **References**

Bhugra, D. (2005) 'The global prevalence of schizophrenia.', *PLoS medicine*. Public Library of Science, 2(5), p. e151; quiz e175. doi: 10.1371/journal.pmed.0020151.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M. and Lee, J. J. (2015) 'Second-generation PLINK: rising to the challenge of larger and richer datasets', *GigaScience*. BioMed Central, 4(1), p. 7. doi: 10.1186/s13742-015-0047-8.

Curtis, D. (2017) 'Polygenic risk score for schizophrenia is not strongly associated with the expression of specific genes or gene sets', *bioRxiv*, p. 205518.

Fromer, M., Roussos, P., Sieberts, S. K., Johnson, J. S., Kavanagh, D. H., Perumal, T. M., Ruderfer, D. M., Oh, E. C., Topol, A., Shah, H. R., Klei, L. L., Kramer, R., Pinto, D., Gümüş, Z. H., Cicek, A. E., Dang, K. K., Browne, A., Lu, C., Xie, L., Readhead, B., Stahl, E. A., Xiao, J., Parvizi, M., Hamamsy, T., Fullard, J. F., Wang, Y.-C., Mahajan, M. C., Derry, J. M. J., Dudley, J. T., Hemby, S. E., Logsdon, B. A., Talbot, K., Raj, T., Bennett, D. A., De Jager, P. L., Zhu, J., Zhang, B., Sullivan, P. F., Chess, A., Purcell, S. M., Shinobu, L. A., Mangravite, L. M., Toyoshima, H., Gur, R. E., Hahn, C.-G., Lewis, D. A., Haroutunian, V., Peters, M. A., Lipska, B. K., Buxbaum, J. D., Schadt, E. E., Hirai, K., Roeder, K., Brennand, K. J., Katsanis, N., Domenici, E., Devlin, B. and Sklar, P. (2016) 'Gene expression elucidates functional impact of polygenic risk for schizophrenia', *Nature Neuroscience*, 19(11), pp. 1442–1453. doi: 10.1038/nn.4399.

International HapMap 3 Consortium, D. M., Altshuler, D. M., Gibbs, R. A., Peltonen, L., Altshuler, D. M., Gibbs, R. A., Peltonen, L., Dermitzakis, E., Schaffner, S. F., Yu, F., Peltonen, L., Dermitzakis, E., Bonnen, P. E., Altshuler, D. M., Gibbs, R. A., de Bakker, P. I. W., Deloukas, P., Gabriel, S. B., Gwilliam, R., Hunt, S., Inouye, M., Jia, X., Palotie, A., Parkin, M., Whittaker, P., Yu, F., Chang, K., Hawes, A., Lewis, L. R., Ren, Y., Wheeler, D., Gibbs, R. A., Muzny, D. M., Barnes, C., Darvishi, K., Hurles, M., Korn, J. M., Kristiansson, K., Lee, C., McCarroll, S. A., Nemesh, J., Dermitzakis, E., Keinan, A., Montgomery, S. B., Pollack, S., Price, A. L., Soranzo, N., Bonnen, P. E., Gibbs, R. A., Gonzaga-Jauregui, C., Keinan, A., Price, A. L., Yu, F., Anttila, V., Brodeur, W., Daly, M. J., Leslie, S., McVean, G., Moutsianas, L., Nguyen, H., Schaffner, S. F., Zhang, Q., Ghorri, M. J. R., McGinnis, R., McLaren, W., Pollack, S., Price, A. L., Schaffner, S. F., Takeuchi, F., Grossman, S. R., Shlyakhter, I., Hostetter, E. B., Sabeti, P. C., Adebamowo, C. A., Foster, M. W., Gordon, D. R., Licinio, J., Manca, M. C., Marshall, P. A., Matsuda, I., Ngare, D., Wang, V. O., Reddy, D., Rotimi, C. N., Royal, C. D., Sharp, R. R., Zeng, C., Brooks, L. D. and McEwen, J. E. (2010) 'Integrating common and rare genetic variation in diverse human populations.', *Nature*, 467(7311), pp. 52–8. doi: 10.1038/nature09298.

Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., Daly, M. J., Bustamante, C. D. and Kenny, E. E. (2017) 'Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations.', *American journal of human genetics*. Elsevier, 100(4), pp. 635–649. doi: 10.1016/j.ajhg.2017.03.004.

Mistry, S., Harrison, J. R., Smith, D. J., Escott-Price, V. and Zammit, S. (2017) 'The use of polygenic risk scores to identify phenotypes associated with genetic risk of schizophrenia: Systematic review', *Schizophrenia Research*. Elsevier. doi: 10.1016/j.schres.2017.10.037.

Pinto, R., Ashworth, M. and Jones, R. (2008) 'Schizophrenia in black Caribbeans living in the UK: an exploration of underlying causes of the high incidence rate.', *The British journal of general practice : the journal of the Royal College of General Practitioners*. Royal College of General Practitioners, 58(551), pp. 429–34. doi: 10.3399/bjgp08X299254.

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., Sklar, P., Purcell Leader, S. M., Ruderfer, D. M., McQuillin, A., Morris, D. W., O'Dushlaine, C. T., Corvin, A., Holmans, P. a, Macgregor, S., Gurling, H., Blackwood, D. H. R., Craddock, N. J., Gill, M., Hultman, C. M., Kirov, G. K., Lichtenstein, P., Muir, W. J., Owen, M. J., Pato, C. N., Scolnick, E. M., St Clair, D., Sklar Leader, P., Williams, N. M., Georgieva, L., Nikolov, I., Norton, N., Williams, H., Toncheva, D., Milanova, V., Thelander, E. F., Sullivan, P. F.,



Kenny, E., Quinn, E. M., Choudhury, K., Datta, S., Pimm, J., Thirumalai, S., Puri, V., Krasucki, R., Lawrence, J., Quested, D., Bass, N., Crombie, C., Fraser, G., Leh Kuan, S., Walker, N., McGhee, K. a, Pickard, B., Malloy, P., Maclean, A. W., Van Beck, M., Pato, M. T., Medeiros, H., Middleton, F., Carvalho, C., Morley, C., Fanous, A., Conti, D., Knowles, J. a, Paz Ferreira, C., Macedo, A., Helena Azevedo, M., Kirby, A. N., Ferreira, M. a R., Daly, M. J., Chambert, K., Kuruville, F., Gabriel, S. B., Ardlie, K. and Moran, J. L. (2009) 'Common polygenic variation contributes to risk of schizophrenia and bipolar disorder.', *Nature*, 10(AuGuST), pp. 8192–8192. doi: 10.1038/nature08185.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J. and Sham, P. C. (2007) 'PLINK: a tool set for whole-genome association and population-based linkage analyses.', *American journal of human genetics*. Elsevier, 81(3), pp. 559–75. doi: 10.1086/519795.

R Core Team (2014) *R: A language and environment for statistical computing*. Vienna, Austria., Austria.: R Foundation for Statistical Computing.

Reisberg, S., Iljasenko, T., Läll, K., Fischer, K. and Vilo, J. (2017) 'Comparing distributions of polygenic risk scores of type 2 diabetes and coronary heart disease within different populations', *PLOS ONE*. Edited by G. Chaubey. Public Library of Science, 12(7), p. e0179238. doi: 10.1371/journal.pone.0179238.

Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) 'Biological insights from 108 schizophrenia-associated genetic loci', *Nature*, 511, pp. 421–427. doi: 10.1038/nature13595.

Vassos, E., Di Forti, M., Coleman, J., Iyegbe, C., Prata, D., Euesden, J., O'Reilly, P., Curtis, C., Koliakou, A., Patel, H., Newhouse, S., Traylor, M., Ajnakina, O., Mondelli, V., Marques, T. R., Gardner-Sood, P., Aitchison, K. J., Powell, J., Atakan, Z., Greenwood, K. E., Smith, S., Ismail, K., Pariante, C., Gaughran, F., Dazzan, P., Markus, H. S., David, A. S., Lewis, C. M., Murray, R. M. and Breen, G. (2017) 'An Examination of Polygenic Score Risk Prediction in Individuals With First-Episode Psychosis', *Biological Psychiatry*. Elsevier, 81(6), pp. 470–477. doi: 10.1016/j.biopsych.2016.06.028.

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A. and Yang, J. (2017) '10 Years of GWAS Discovery: Biology, Function, and Translation', *The American Journal of Human Genetics*, 101(1), pp. 5–22. doi: 10.1016/j.ajhg.2017.06.005.

Weiner, D. J., Wigdor, E. M., Ripke, S., Walters, R. K., Kosmicki, J. A., Grove, J., Samocha, K. E., Goldstein, J. I., Okbay, A., Bybjerg-Grauholm, J., Werge, T., Hougaard, D. M., Taylor, J., iPSYCH-Broad Autism Group, Psychiatric Genomics Consortium Autism Group, Skuse, D., Devlin, B., Anney, R., Sanders, S. J., Bishop, S., Mortensen, P. B., Børglum, A. D., Smith, G. D., Daly, M. J. and Robinson, E. B. (2017) 'Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders', *Nature Genetics*, 49(7), pp. 978–985. doi: 10.1038/ng.3863.

**Table 1.** Schizophrenia PRS distribution in HapMap cohorts.

Cohort	Description	Number of subjects	PRS mean (SD)
ASW	African ancestry in Southwest USA	87	7.75 (2.05)
CEU	Utah residents with Northern and Western European ancestry	165	-2.90 (2.09)
CHB	Han Chinese in Beijing, China	137	4.71 (1.78)
CHD	Chinese in Metropolitan Denver, Colorado	109	4.57 (1.71)
GIH	Gujarati Indians in Houston, Texas	101	1.91 (2.08)
JPT	Japanese in Tokyo, Japan	113	4.71 (1.59)
LWK	Luhya in Webuye, Kenya	110	9.48 (1.76)
MEX	Mexican ancestry in Los Angeles, California	86	1.50 (2.59)
MKK	Masai in Kinyawa, Kenya	184	7.08 (1.76)
TSI	Toscani in Italy	102	-2.64 (1.91)
YRI	Yoruba in Ibadan, Nigeria	203	10.27 (1.65)

**Figure 2.** Schizophrenia PRS for individual HapMap subjects. The boxes show the mean and SD for each cohort.



