

**Polygenic risk score for schizophrenia is not strongly associated with the expression of specific genes or gene sets**

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.

**Short title:** Polygenic risk score for schizophrenia

**Keywords**

Schizophrenia; polygenic risk score; RNA; expression.

## **Abstract**

The polygenic risk score (PRS) is derived from SNPs including both those which are genome-wide significant and also a large number of others more weakly associated with schizophrenia. Such variants are widely dispersed, though concentrated near genes expressed in the brain, and it has been proposed that these SNP associations result from impacts on cell regulatory networks which ultimately affect the expression or function of a modest number of "core" genes. A previous study demonstrated association of some GWAS-significant variants with expression of a number of genes, by examining pair-wise correlations of gene expression with SNP genotypes. The present study used data downloaded from the CommonMind Consortium site, consisting of SNP genotypes and RNAseq expression data from the dorsolateral prefrontal cortex, to examine whether the expression of individual genes or sets of genes correlated with PRS in 207 controls and 209 schizophrenia cases. Although the PRS was significantly associated with phenotype, the correlations with genes and genes sets followed distributions expected by chance. Thus, this analysis failed to demonstrate that the PRS captures a cumulative effect of multiple variants impacting the expression of a small number of genes and it failed to focus attention on a small number of genes of biological relevance. The multiple SNP associations observed in schizophrenia may result from other mechanisms, including effects mediated indirectly through environmental risk factors.

## **Introduction**

There is a substantial genetic contribution to the aetiology of schizophrenia which in part manifests as association with common variants, with over 100 loci meeting conventional standards for significance in a genome-wide association study (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). The same study showed that a polygenic risk score (PRS) was more strongly predictive of schizophrenia if it included not

only the variants which were statistically significant but also many other variants with only weak evidence for association. This clearly demonstrated that there are a very large number of common variants whose frequency differs between cases and non-cases. In a recent review of this field, the authors noted the prior evidence for the involvement of large numbers of variants which were very widely distributed across the genome and proposed that the term "omnigenic" could appropriately be used (Boyle et al., 2017). However in their own analyses they demonstrated that only variants near genes expressed in the brain contribute substantially to heritability of schizophrenia and that variants near genes expressed specifically in the brain made a larger contribution per variant than those near more widely expressed genes. They also noted that common variants show only a modest enrichment for genes by functional category and contrasted this with the findings from studies of rare variants and CNVs which implicate synaptic and neuronal genes (Fromer et al., 2014; Marshall et al., 2016; Purcell et al., 2014). Although they did not refer to it, an additional study of ultra-rare, gene disruptive variants also demonstrated enrichment in these categories (Genovese et al., 2016). To make sense of these observations, they proposed that a given phenotype might be directly affected by a modest number of "core" genes but that, because cell regulatory networks are highly interconnected, any expressed gene is likely to have some effect on the regulation or function of these core genes.

A previous study sought to investigate the relationship between common variants and gene expression by examining SNP genotypes and RNAseq results from the dorsolateral prefrontal cortex (DLPFC) (Fromer et al., 2016). This reported that about 20% of schizophrenia-associated loci contained variants which could contribute to altered gene expression or liability. Considering SNP-gene pairs within 1Mb of each other, there were 2,154,331 significant *cis*-eQTLs, and these were enriched for enhancer sequences in brain tissues, most strongly in DLPFC enhancers. There were also 45,453 significant *trans*-eQTLs, in which the SNP was more than 1Mb from the gene whose expression it correlated with. Only pair-wise comparisons of SNPs and genes were carried out and there was no

report of whether the expression of any genes correlated with the PRS, although it was reported that the PRS was higher in cases than controls.

If multiple common variants exert their effect indirectly through regulatory networks, it is plausible that the PRS might reflect a cumulative effect on the expression of one or more "core" genes. This cumulative effect on gene expression might be greater than any pairwise effect on expression due to a single SNP. Also, it might be more pronounced for the expression of the core genes themselves than for the genes which only exerted their effect indirectly and so could serve to focus attention on core genes. Hence, it seems logical to explore whether the schizophrenia PRS is associated with gene expression.

## **Materials and methods**

The dataset used in the previous gene expression study was downloaded from the CommonMind Consortium (CMC) Knowledge Portal (<https://www.synapse.org/#!Synapse:syn2759792/wiki/69613>) consisting of SNP genotypes and RNAseq results from DLPFC 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. 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 have been removed. The gene expression data had been obtained by RNA sequencing of tissue from the DLPFC followed by adjustment for ancestry and other appropriate covariates.

In order to obtain a polygenic risk score (PRS) for schizophrenia, the file called *scz2.prs.txt.gz*, containing ORs and p values for 102,636 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)). This training set was produced as part of the previously reported PGC2 schizophrenia GWAS (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). SNPs with p value  $< 0.05$  were selected and their  $\log(\text{OR})$  summed over sample genotypes using the `--score` function of *plink 1.09beta* in order to produce a PRS for each subject ([www.cog-genomics.org/plink/1.9/](http://www.cog-genomics.org/plink/1.9/)) (Chang et al., 2015; Purcell et al., 2007, 2009). The first 20 principal components for all genotyped SNPs were produced using the `--pca` and `--make-rel` functions of *plink*.

Statistical tests and data manipulation were carried out using R version 3.3.2 (R Core Team, 2014).

Initial analyses showed that the PRS was higher in cases than controls ( $t = 3.5$ ,  $p = 0.00059$ ) but it was noticed that the association became much more highly significant when the SNP principal components were included as covariates ( $p = 2.5 \times 10^{-17}$ ). This seemed to be due the fact that the first principal component was very strongly negatively correlated with the PRS ( $r = -0.81$ ,  $p = 1.2 \times 10^{-126}$ ). This correlation was present in both the controls ( $r = -0.85$ ,  $p = 1.8 \times 10^{-78}$ ) and the cases ( $r = -0.80$ ,  $p = 2.5 \times 10^{-59}$ ). The plots of controls and cases of the first and second principal components are shown in Figures 1A and 1B. It can be seen that the distribution is similar in both groups. Most subjects have a very high value for the first principal component and a very low value for the second one but several subjects are clearly outliers from this main grouping and have low scores for the first principal component or, in a

few cases, high scores for the second. In order to filter out these outliers, the dataset was narrowed to include subjects with the first principal component value greater than 0.01 and second principal component value less than 0.04. Doing this excluded 51 controls and 72 cases, leaving 207 controls and 209 cases, and produced a more homogeneous sample in which the PRS was not significantly correlated with the first principal component but was very significantly higher in cases than controls ( $t = 8.3$ ,  $p = 1.3 \times 10^{-15}$ ).

Regression of the expression level of each of 16,423 genes was performed on the PRS including the 20 principal components as covariates. The minus log base 10 of the p value (MLP) was recorded for each gene, as was the signed log p value (SLP), which is the same as the MLP but given a positive sign if the regression coefficient is positive and a negative sign if the coefficient is negative.

As well as the analyses based on individual genes, gene set analyses were performed to see if it was possible to identify a set of genes whose expression tended to correlate with the PRS. Two lists of gene sets were used. The first consisted of the 32 gene sets reported to be enriched for ultra-rare, gene disruptive variants in a sample of exome-sequenced schizophrenia cases (Genovese et al., 2016). These included categories such as genes expressed in neurons, genes near GWAS hits and genes which are loss of function intolerant. The second group consisted of 1454 "all GO gene sets, gene symbols" pathways downloaded from the Molecular Signatures Database at <http://www.broadinstitute.org/gsea/msigdb/collections.jsp> (Subramanian et al., 2005).

It is challenging to detect whether the expression pattern of a set of genes correlates with the PRS. One cannot simply test whether genes in the set are more strongly correlated with PRS than those not in the set because it is expected that the expression patterns of genes within a set will correlate with each other so that their correlations with PRS are not expected to be independent. Nor can one test whether the mean or total expression of members of the set correlates with PRS because it is expected that if a biological effect is present then

expression of some genes will be increased and others decreased. Instead, the MLPs for each gene in the set were summed and the total compared with that which would be expected by chance by carrying out permutations of the PRS across subjects and then recalculating and totalling the gene-wise MLPs for each permuted dataset. The proportion of times the totalled MLPs for permuted datasets exceeded the total for the real dataset was used to obtain an empirical p value, and minus log base 10 of this empirical p value was used to obtain an MLP for the gene set.

## Results

The Q:Q plot for the SLPs for the correlation of PRS with the expression of each of 16,423 genes are shown in Figure 2. It can be seen that the results almost exactly follow what would be expected under the null hypothesis if the expression patterns of the genes were independent. The most highly significant correlation was for SULF1 with  $p=0.000005$  (MLP=5.3) but this would not reach conventional standards of significance if a Bonferroni correction were applied. SULF1 codes for an extracellular heparan sulfate endosulfatase and does not seem to be a likely candidate for involvement in the aetiology of schizophrenia. It could be argued that expression patterns of genes are likely to be correlated with each other and hence that a Bonferroni correction might be overly conservative. Table 1 lists the genes achieving  $MLP > 3$ .

Table 2 shows the permutation-derived MLPs for each of the gene sets previously used to test for enrichment for damaging ultra-rare variants. It can be seen that none of the sets produces highly significant evidence for correlation with PRS. For only two of the sets, *alid* and *mir137*, does the MLP exceed 2 and again these results would not be significant if a Bonferroni correction were applied. Again, it could be argued that this might be conservative because the sets overlap each other and hence the MLPs, although individually valid, are not independent.

Of the 1454 GO gene sets, none produced a result which would be significant after Bonferroni correction. However 11 did produce MLPs exceeding 3 and these are listed in Table 3. Again, these are not independent but include overlapping sets of genes. In particular, the sets relating to histone modification contain RBM14, which has a gene-wise MLP of 3.42, whereas the regulation of cell differentiation sets all include MAFB, which has a gene-wise MLP of 3.19. Table 4 lists all the genes which occur in at least one of these sets and which have a gene-wise  $MLP > 1.3$ , equivalent to  $p < 0.05$ .

## Discussion

This approach has failed to yield any genes or gene sets whose expression is strongly associated with the schizophrenia PRS. It would be possible to argue that the Bonferroni correction is overly conservative and that there may be some association but the main aim of the analysis, to provide a strong focus on a small set of "core" genes, has not been successful. The PRS represents a cumulation of genetic risk factors from multiple locations but it is not the case that the variants involved exert their effect on risk primarily by combining to effect the expression of a small number of genes. It is of course possible that there is some association between the PRS and the expression of some genes which would be possible to detect if a larger sample were available. In terms of effects on gene expression, the analyses described here do not add anything to the results previously reported for pair-wise analyses (Fromer et al., 2016).

As an aside, it is worth noting the very strong correlation of the schizophrenia PRS with the first principal component of the ethnically heterogeneous sample. It has recently been noted that the PRS for schizophrenia stratifies by ancestry (Weiner et al., 2017). Likewise, the PRS for type 2 diabetes and coronary heart disease vary between populations and this must be taken account of if attempting to estimate an individual's risk (Reisberg et al., 2017).

Although it may be expected that the mean PRS will vary somewhat according to ethnicity, it is perhaps a little surprising that the observed correlation with the first principal component is

quite so strong and this will be the subject of further investigation. However this phenomenon does not seem to be relevant to the findings reported here, which were obtained from the more homogeneous sub-sample.

This study does not support the proposal that multiple common variants exert a joint effect on schizophrenia risk through a cumulative effect on the expression of a small number of genes. It is not hard to conceive of different mechanisms whereby the "polygenic risk" might be observed. One kind of mechanism is that a variant has some effect on gene function or expression which has a small but direct effect on the pathophysiological processes which can lead to schizophrenia. However, one can also conceive of indirect effects which could be very distant from the "core biology". There are number of environmental factors which appear to influence risk of schizophrenia, including obstetric complications, cannabis use, maternal viral infection, maternal malnutrition and childhood adversity (Cannon et al., 2002; Gage et al., 2017; Khandaker et al., 2013; Stilo et al., 2013; Xu et al., 2009). Any genetic variant which affects exposure to such risk factors will be expected to be associated with schizophrenia and these mechanisms may work either within the proband or within their parents. To give one example, as pointed out previously, smoking during pregnancy is a risk factor for the subsequent development of schizophrenia and the observed genetic association between smoking behaviour and schizophrenia might be mediated via this mechanism, so that a genetic predisposition to smoke in the mother is associated with increased risk of schizophrenia in her offspring (Curtis, 2017; Hartz et al., 2017; Niemel et al., 2016). Likewise, genetic variants which influence pelvis size, foetal head size, susceptibility to viral infection, cannabis preference and susceptibility to be either a victim or perpetrator of childhood abuse will all be expected to be associated with schizophrenia risk. Of course, such indirect effects may be very small but GWAS sample sizes are such that very small effects can be detected. If this view is valid then a number of common variants may contribute to the "omnigenic" signal through indirect environmental mechanisms as well

as through the interactive effects in intra- and inter- cellular networks which have been proposed (Boyle et al., 2017).

### **Conflict of interest**

The author declares he has no conflict of interest.

### **Acknowledgments**

The author thanks the CommonMind Consortium for making this dataset 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).

### **References**

Bayés, A., van de Lagemaat, L.N., Collins, M.O., Croning, M.D.R., Whittle, I.R., Choudhary, J.S., Grant, S.G.N. (2011) Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat. Neurosci.* 14, 19–21.

Boyle, E.A., Li, Y.I., Pritchard, J.K., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Al., E., Consortium, Mig., Consortium, P., Study, L.C., Al., E., Consortium, G.L.G.G., ReproGen Consortium, Investigators, M., Al., E., Consortium, Mig., Consortium, P., Consortium, R., Consortium, G.L.G.G., Consortium, I.E., Al., E. (2017) An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 169, 1177–1186.

Cahoy, J.D., Emery, B., Kaushal, A., Foo, L.C., Zamanian, J.L., Christopherson, K.S., Xing, Y., Lubischer, J.L., Krieg, P.A., Krupenko, S.A., Thompson, W.J., Barres, B.A. (2008) A Transcriptome Database for Astrocytes, Neurons, and Oligodendrocytes: A New Resource for Understanding Brain Development and Function. *J. Neurosci.* 28, 264–278.

Cannon, M., Jones, P.B., Murray, R.M. (2002) Obstetric Complications and Schizophrenia: Historical and Meta-Analytic Review. *Am. J. Psychiatry* 159, 1080–1092.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7.

Cotton, A.M., Ge, B., Light, N., Adoue, V., Pastinen, T., Brown, C.J. (2013) Analysis of expressed SNPs identifies variable extents of expression from the human inactive X chromosome. *Genome Biol.* 14, R122.

Curtis, D. (2017) Association between smoking and psychosis may be mediated by maternal smoking during pregnancy. *Psychol. Med.* 1.

Darnell, J.C., Van Driesche, S.J., Zhang, C., Hung, K.Y.S., Mele, A., Fraser, C.E., Stone, E.F., Chen, C., Fak, J.J., Chi, S.W., Licatalosi, D.D., Richter, J.D., Darnell, R.B. (2011) FMRP Stalls Ribosomal Translocation on mRNAs Linked to Synaptic Function and Autism. *Cell* 146, 247–261.

Deciphering Developmental Disorders Study (2017) Prevalence and architecture of de novo mutations in developmental disorders. *Nature* 542, 433–438.

Fagerberg, L., Hallstrom, B.M., Oksvold, P., Kampf, C., Djureinovic, D., Odeberg, J., Habuka, M., Tahmasebpoor, S., Danielsson, A., Edlund, K., Asplund, A., Sjostedt, E., Lundberg, E., Szigyarto, C.A.-K., Skogs, M., Takanen, J.O., Berling, H., Tegel, H., Mulder, J., Nilsson, P., Schwenk, J.M., Lindskog, C., Danielsson, F., Mardinoglu, A., Sivertsson, A., von Feilitzen, K., Forsberg, M., Zwahlen, M., Olsson, I., Navani, S., Huss, M., Nielsen, J., Ponten, F., Uhlen, M. (2014) Analysis of the Human Tissue-specific Expression by Genome-wide Integration of Transcriptomics and Antibody-based Proteomics. *Mol. Cell. Proteomics* 13, 397–406.

Fromer, M., Pocklington, A.J., Kavanagh, D.H., Williams, H.J., Dwyer, S., Gormley, P., Georgieva, L., Rees, E., Palta, P., Ruderfer, D.M., Carrera, N., Humphreys, I., Johnson, J.S., Roussos, P., Barker, D.D., Banks, E., Milanova, V., Grant, S.G., Hannon, E., Rose, S.A., Chambert, K., Mahajan, M., Scolnick, E.M., Moran, J.L., Kirov, G., Palotie, A., McCarroll, S.A., Holmans, P., Sklar, P., Owen, M.J., Purcell, S.M., O'Donovan, M.C. (2014) De novo mutations in schizophrenia implicate synaptic networks. *Nature* 506, 179–184.

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., Toyoshiba, 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., Sklar, P. (2016) Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* 19, 1442–1453.

Gage, S.H., Jones, H.J., Burgess, S., Bowden, J., Davey Smith, G., Zammit, S., Munafò, M.R. (2017) Assessing causality in associations between cannabis use and schizophrenia

risk: a two-sample Mendelian randomization study. *Psychol. Med.* 47, 971–980.

Gécz, J., Shoubridge, C., Corbett, M. (2009) The genetic landscape of intellectual disability arising from chromosome X. *Trends Genet.* 25, 308–316.

Genovese, G., Fromer, M., Stahl, E.A., Ruderfer, D.M., Chambert, K., Landén, M., Moran, J.L., Purcell, S.M., Sklar, P., Sullivan, P.F., Hultman, C.M., McCarroll, S.A. (2016) Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* 19, 1433–1441.

Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A., McKusick, V.A. (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.* 33, D514-7.

Hartz, S.M., Horton, A.C., Hancock, D.B., Baker, T.B., Caporaso, N.E., Chen, L.-S., Hokanson, J.E., Lutz, S.M., Marazita, M.L., McNeil, D.W., Pato, C.N., Pato, M.T., Johnson, E.O., Bierut, L.J. (2017) Genetic correlation between smoking behaviors and schizophrenia. *Schizophr. Res.*

Khandaker, G.M., Zimbron, J., Lewis, G., Jones, P.B. (2013) Prenatal maternal infection, neurodevelopment and adult schizophrenia: a systematic review of population-based studies. *Psychol. Med.* 43, 239–257.

Kirov, G., Pocklington, A.J., Holmans, P., Ivanov, D., Ikeda, M., Ruderfer, D., Moran, J., Chambert, K., Toncheva, D., Georgieva, L., Grozeva, D., Fjodorova, M., Wollerton, R., Rees, E., Nikolov, I., van de Lagemaat, L.N., Bayés, À., Fernandez, E., Olason, P.I., Böttcher, Y., Komiyama, N.H., Collins, M.O., Choudhary, J., Stefansson, K., Stefansson, H., Grant, S.G.N., Purcell, S., Sklar, P., O'Donovan, M.C., Owen, M.J. (2012) De novo CNV analysis implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. *Mol. Psychiatry* 17, 142–153.

Lek, M., Karczewski, K.J., Minikel, E. V, Samocha, K.E., Banks, E., Fennell, T., O'Donnell-

Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., Tukiainen, T., Birnbaum, D.P., Kosmicki, J.A., Duncan, L.E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D.N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M.I., Moonshine, A.L., Natarajan, P., Orozco, L., Peloso, G.M., Poplin, R., Rivas, M.A., Ruano-Rubio, V., Rose, S.A., Ruderfer, D.M., Shakir, K., Stenson, P.D., Stevens, C., Thomas, B.P., Tiao, G., Tusie-Luna, M.T., Weisburd, B., Won, H.-H., Yu, D., Altshuler, D.M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J.C., Gabriel, S.B., Getz, G., Glatt, S.J., Hultman, C.M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Neale, B.M., Palotie, A., Purcell, S.M., Saleheen, D., Scharf, J.M., Sklar, P., Sullivan, P.F., Tuomilehto, J., Tsuang, M.T., Watkins, H.C., Wilson, J.G., Daly, M.J., MacArthur, D.G., Exome Aggregation Consortium, O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., Tukiainen, T., Birnbaum, D.P., Kosmicki, J.A., Duncan, L.E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D.N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M.I., Moonshine, A.L., Natarajan, P., Orozco, L., Peloso, G.M., Poplin, R., Rivas, M.A., Ruano-Rubio, V., Rose, S.A., Ruderfer, D.M., Shakir, K., Stenson, P.D., Stevens, C., Thomas, B.P., Tiao, G., Tusie-Luna, M.T., Weisburd, B., Won, H.-H., Yu, D., Altshuler, D.M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J.C., Gabriel, S.B., Getz, G., Glatt, S.J., Hultman, C.M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Neale, B.M., Palotie, A., Purcell, S.M., Saleheen, D., Scharf, J.M., Sklar, P., Sullivan, P.F., Tuomilehto, J., Tsuang, M.T., Watkins, H.C., Wilson, J.G., Daly, M.J., MacArthur, D.G., Exome Aggregation Consortium (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291.

Marshall, C.R., Howrigan, D.P., Merico, D., Thiruvahindrapuram, B., Wu, W., Greer, D.S., Antaki, D., Shetty, A., Holmans, P.A., Pinto, D., Gujral, M., Brandler, W.M., Malhotra, D., Wang, Z., Fajardo, K.V.F., Maile, M.S., Ripke, S., Agartz, I., Albus, M., Alexander, M.,

Amin, F., Atkins, J., Bacanu, S.A., Belliveau, R.A., Bergen, S.E., Bertalan, M., Bevilacqua, E., Bigdeli, T.B., Black, D.W., Bruggeman, R., Buccola, N.G., Buckner, R.L., Bulik-Sullivan, B., Byerley, W., Cahn, W., Cai, G., Cairns, M.J., Champion, D., Cantor, R.M., Carr, V.J., Carrera, N., Catts, S. V, Chambert, K.D., Cheng, W., Cloninger, C.R., Cohen, D., Cormican, P., Craddock, N., Crespo-Facorro, B., Crowley, J.J., Curtis, D., Davidson, M., Davis, K.L., Degenhardt, F., Del Favero, J., DeLisi, L.E., Dikeos, D., Dinan, T., Djurovic, S., Donohoe, G., Drapeau, E., Duan, J., Dudbridge, F., Eichhammer, P., Eriksson, J., Escott-Price, V., Essioux, L., Fanous, A.H., Farh, K.-H., Farrell, M.S., Frank, J., Franke, L., Freedman, R., Freimer, N.B., Friedman, J.I., Forstner, A.J., Fromer, M., Genovese, G., Georgieva, L., Gershon, E.S., Giegling, I., Giusti-Rodríguez, P., Godard, S., Goldstein, J.I., Gratten, J., de Haan, L., Hamshere, M.L., Hansen, M., Hansen, T., Haroutunian, V., Hartmann, A.M., Henskens, F.A., Herms, S., Hirschhorn, J.N., Hoffmann, P., Hofman, A., Huang, H., Ikeda, M., Joa, I., Kähler, A.K., Kahn, R.S., Kalaydjieva, L., Karjalainen, J., Kavanagh, D., Keller, M.C., Kelly, B.J., Kennedy, J.L., Kim, Y., Knowles, J.A., Konte, B., Laurent, C., Lee, P., Lee, S.H., Legge, S.E., Lerer, B., Levy, D.L., Liang, K.-Y., Lieberman, J., Lönngqvist, J., Loughland, C.M., Magnusson, P.K.E., Maher, B.S., Maier, W., Mallet, J., Mattheisen, M., Mattingsdal, M., McCarley, R.W., McDonald, C., McIntosh, A.M., Meier, S., Meijer, C.J., Melle, I., Meshulam-Gately, R.I., Metspalu, A., Michie, P.T., Milani, L., Milanova, V., Mokrab, Y., Morris, D.W., Müller-Myhsok, B., Murphy, K.C., Murray, R.M., Myin-Germeys, I., Nenadic, I., Nertney, D.A., Nestadt, G., Nicodemus, K.K., Nisenbaum, L., Nordin, A., O'Callaghan, E., O'Dushlaine, C., Oh, S.-Y., Olincy, A., Olsen, L., O'Neill, F.A., Van Os, J., Pantelis, C., Papadimitriou, G.N., Parkhomenko, E., Pato, M.T., Paunio, T., Perkins, D.O., Pers, T.H., Pietiläinen, O., Pimm, J., Pocklington, A.J., Powell, J., Price, A., Pulver, A.E., Purcell, S.M., Quedsted, D., Rasmussen, H.B., Reichenberg, A., Reimers, M.A., Richards, A.L., Roffman, J.L., Roussos, P., Ruderfer, D.M., Salomaa, V., Sanders, A.R., Savitz, A., Schall, U., Schulze, T.G., Schwab, S.G., Scolnick, E.M., Scott, R.J., Seidman, L.J., Shi, J., Silverman, J.M., Smoller, J.W., Söderman, E., Spencer, C.C.A., Stahl, E.A., Strengman, E., Strohmaier, J., Stroup, T.S., Suvisaari, J., Svrakic, D.M., Szatkiewicz, J.P., Thirumalai, S., Tooney, P.A.,

Veijola, J., Visscher, P.M., Waddington, J., Walsh, D., Webb, B.T., Weiser, M., Wildenauer, D.B., Williams, N.M., Williams, S., Witt, S.H., Wolen, A.R., Wormley, B.K., Wray, N.R., Wu, J.Q., Zai, C.C., Adolfsson, R., Andreassen, O.A., Blackwood, D.H.R., Bramon, E., Buxbaum, J.D., Cichon, S., Collier, D.A., Corvin, A., Daly, M.J., Darvasi, A., Domenici, E., Esko, T., Gejman, P. V, Gill, M., Gurling, H., Hultman, C.M., Iwata, N., Jablensky, A. V, Jönsson, E.G., Kendler, K.S., Kirov, G., Knight, J., Levinson, D.F., Li, Q.S., McCarroll, S.A., McQuillin, A., Moran, J.L., Mowry, B.J., Nöthen, M.M., Ophoff, R.A., Owen, M.J., Palotie, A., Pato, C.N., Petryshen, T.L., Posthuma, D., Rietschel, M., Riley, B.P., Rujescu, D., Sklar, P., St Clair, D., Walters, J.T.R., Werge, T., Sullivan, P.F., O'Donovan, M.C., Scherer, S.W., Neale, B.M., Sebat, J., Sebat, J., CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium (2016) Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat. Genet.* 49, 27–35.

Moeschler, J.B. (2008) Genetic Evaluation of Intellectual Disabilities. *Semin. Pediatr. Neurol.* 15, 2–9.

Moeschler, J.B., Shevell, M., American Academy of Pediatrics Committee on Genetics (2006) Clinical Genetic Evaluation of the Child With Mental Retardation or Developmental Delays 117, 2304–2316.

Niemel, S., Sourander, A., Surcel, H.-M., Hinkka-Yli-Salomäki, S., McKeague, I.W., Cheslack-Postava, K., Brown, A.S. (2016) Prenatal Nicotine Exposure and Risk of Schizophrenia Among Offspring in a National Birth Cohort. *Am. J. Psychiatry* 173, 799–806.

Pirooznia, M., Wang, T., Avramopoulos, D., Valle, D., Thomas, G., Huganir, R.L., Goes, F.S., Potash, J.B., Zandi, P.P. (2012) SynaptomeDB: an ontology-based knowledgebase for synaptic genes. *Bioinformatics* 28, 897–9.

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., Sham, P.C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–

75.

Purcell, S.M., Moran, J.L., Fromer, M., Ruderfer, D., Solovieff, N., Roussos, P., O'Dushlaine, C., Chambert, K., Bergen, S.E., Kähler, A., Duncan, L., Stahl, E., Genovese, G., Fernández, E., Collins, M.O., Komiyama, N.H., Choudhary, J.S., Magnusson, P.K.E., Banks, E., Shakir, K., Garimella, K., Fennell, T., DePristo, M., Grant, S.G.N., Haggarty, S.J., Gabriel, S., Scolnick, E.M., Lander, E.S., Hultman, C.M., Sullivan, P.F., McCarroll, S.A., Sklar, P. (2014) A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* 506, 185–90.

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., Moran, J.L. (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 461, 8192–8192.

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

Rauch, A., Hoyer, J., Guth, S., Zweier, C., Kraus, C., Becker, C., Zenker, M., Hüffmeier, U., Thiel, C., Rüschenhoff, F., Nürnberg, P., Reis, A., Trautmann, U. (2006) Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. *Am. J. Med. Genet. Part A* 140A, 2063–2074.

Reisberg, S., Iljasenko, T., Läll, K., Fischer, K., Vilo, J. (2017) Comparing distributions of polygenic risk scores of type 2 diabetes and coronary heart disease within different populations. *PLoS One* 12, e0179238.

Robinson, E.B., Neale, B.M., Hyman, S.E. (2015) Genetic research in autism spectrum disorders. *Curr. Opin. Pediatr.* 27, 685–691.

Samocha, K.E., Robinson, E.B., Sanders, S.J., Stevens, C., Sabo, A., McGrath, L.M., Kosmicki, J.A., Rehnström, K., Mallick, S., Kirby, A., Wall, D.P., MacArthur, D.G., Gabriel, S.B., DePristo, M., Purcell, S.M., Palotie, A., Boerwinkle, E., Buxbaum, J.D., Cook, E.H., Gibbs, R.A., Schellenberg, G.D., Sutcliffe, J.S., Devlin, B., Roeder, K., Neale, B.M., Daly, M.J. (2014) A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* 46, 944–950.

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

Stilo, S.A., Di Forti, M., Mondelli, V., Falcone, A.M., Russo, M., O'Connor, J., Palmer, E., Paparelli, A., Kolliakou, A., Sirianni, M., Taylor, H., Handley, R., Dazzan, P., Pariante, C., Marques, T.R., Zoccali, R., David, A., Murray, R.M., Morgan, C. (2013) Social Disadvantage: Cause or Consequence of Impending Psychosis? *Schizophr. Bull.* 39, 1288–1295.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102, 15545–15550.

Wagon, J.L., Briese, M., Sun, W., Mahaffey, C.L., Curk, T., Rot, G., Ule, J., Frankel, W.N. (2012) CELF4 Regulates Translation and Local Abundance of a Vast Set of mRNAs, Including Genes Associated with Regulation of Synaptic Function. *PLoS Genet.* 8, e1003067.

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., Robinson, E.B. (2017) Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat. Genet.* 49, 978–985.

Weyn-Vanhentenryck, S.M., Mele, A., Yan, Q., Sun, S., Farny, N., Zhang, Z., Xue, C., Herre, M., Silver, P.A., Zhang, M.Q., Krainer, A.R., Darnell, R.B., Zhang, C. (2014) HITS-CLIP and Integrative Modeling Define the Rbfox Splicing-Regulatory Network Linked to Brain Development and Autism. *Cell Rep.* 6, 1139–1152.

Xu, M.-Q., Sun, W.-S., Liu, B.-X., Feng, G.-Y., Yu, L., Yang, L., He, G., Sham, P., Susser, E., St. Clair, D., He, L. (2009) Prenatal Malnutrition and Adult Schizophrenia: Further Evidence From the 1959-1961 Chinese Famine. *Schizophr. Bull.* 35, 568–576.

**Table 1.** List of genes for which the correlation of PRS with expression is significant at  $p < 0.001$  (MLP > 3).

<b>Gene</b>	<b>Beta</b>	<b>MLP</b>
SULF1	-0.037	5.30
JAM2	-0.025	4.95
PMEPA1	0.026	4.64
CHRD	-0.020	4.52
TPTE2P1	-0.050	4.38
FIGN	-0.021	4.08
HDAC9	0.016	4.03
PARP16	0.022	3.99
EPHB1	-0.026	3.88
CALB1	0.035	3.85
SLC38A11	-0.041	3.65
CNTNAP1	-0.015	3.62
RRP12	-0.019	3.61
GTF2A2	-0.022	3.53
LONP2	0.008	3.52
MSRB3	-0.014	3.52
ZNF436	-0.017	3.51
GLRA3	0.020	3.44
SLC6A5	-0.048	3.44
RBM14	0.018	3.42
DNAJC24	0.012	3.41
BCAT1	-0.016	3.34
TBC1D3F	-0.059	3.32

GPR135	0.016	3.27
ECH1	0.018	3.26
BOLA3	-0.020	3.25
ANKRD6	0.021	3.25
CYP26B1	-0.031	3.24
MAFB	-0.028	3.19
SLC6A11	-0.025	3.17
ENO4	-0.025	3.16
FLJ20021	-0.026	3.15
GADD45A	-0.037	3.12
AC107021.2	-0.037	3.11
SERTM1	0.029	3.11
RAB11FIP1	0.018	3.10
DOC2B	0.033	3.10
RNF123	-0.011	3.08
HTR2A	0.017	3.08
CCDC126	-0.012	3.05
KCNG1	0.041	3.05
FAM13B	-0.017	3.04
ZNF98	-0.046	3.03
AC084033.3	-0.015	3.03
MR1	0.022	3.02
RASGRF2	0.016	3.01
UTP14A	-0.020	3.00

**Table 2** Permutation-derived MLPs for each of the 32 gene sets previously used to test for enrichment for damaging ultra-rare variants in schizophrenia. The definition of each set is provided in the online methods section of the original publication (Genovese et al., 2016).

<b>Gene set</b>	<b>Symbol</b>	<b>MLP</b>
OMIM intellectual disability (Hamosh et al., 2005)	<i>alid</i>	2.22
Expression specific to brain (Fagerberg et al., 2014)	<i>brain</i>	0.00
Bound by CELF4 (Wagnon et al., 2012)	<i>celf4</i>	0.18
Missense-constrained (Samocha et al., 2014)	<i>constrained</i>	0.34
Involved in developmental disorder (Deciphering Developmental Disorders Study, 2017)	<i>dd</i>	0.34
De novo variants in autism (Fromer et al., 2014)	<i>denovo.aut</i>	1.05
De novo variants in coronary heart disease (Fromer et al., 2014)	<i>denovo.chd</i>	0.08
De novo variants in epilepsy (Fromer et al., 2014)	<i>denovo.epi</i>	0.34
De novo duplications in ASD (Kirov et al., 2012)	<i>denovo.gain.asd</i>	0.38
De novo duplications in bipolar disorder (Kirov et al., 2012)	<i>denovo.gain.bd</i>	0.11
De novo duplications in schizophrenia (Kirov et al., 2012)	<i>denovo.gain.scz</i>	0.65
De novo variants in intellectual disability (Fromer et al., 2014)	<i>denovo.id</i>	0.18
De novo deletions in ASD (Kirov et al., 2012)	<i>denovo.loss.asd</i>	1.04
De novo deletions in bipolar disorder (Kirov et al., 2012)	<i>denovo.loss.bd</i>	0.86
De novo deletions in schizophrenia (Kirov et al., 2012)	<i>denovo.loss.scz</i>	0.04
De novo variants in schizophrenia (Fromer et al., 2014)	<i>denovo.scz</i>	0.87
Bound by FMRP (Darnell et al., 2011)	<i>fmrp</i>	0.20
Implicated by GWAS (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014)	<i>gwas</i>	0.92
Targets of microRNA-137 (Robinson et al., 2015)	<i>mir137</i>	2.17

Expression specific to neurons (Cahoy et al., 2008)	<i>neurons</i>	0.71
NMDAR and ARC complexes (Kirov et al., 2012)	<i>nmdarc</i>	0.04
Loss-of-function intolerant (Lek et al., 2016)	<i>pLI09</i>	0.41
PSD-95 (Bayés et al., 2011)	<i>psd95</i>	0.18
Bound by RBFOX 1 or 3 (Weyn-Vanhentenryck et al., 2014)	<i>rbfox13</i>	0.04
Bound by RBFOX 2 (Weyn-Vanhentenryck et al., 2014)	<i>rbfox2</i>	0.26
Synaptic (Pirooznia et al., 2012)	<i>synaptome</i>	0.15
Escape X-inactivation (Cotton et al., 2013)	<i>x.escape</i>	0.04
X-linked intellectual disability, Genetic Services Laboratories of the University of Chicago (Gécz et al., 2009; Moeschler, 2008; Moeschler et al., 2006; Rauch et al., 2006)	<i>xlid.chicago</i>	0.32
X-linked intellectual disability, Greenwood Genetic Centre (Moeschler et al., 2006)	<i>xlid.gcc</i>	0.36
X-linked intellectual disability, OMIM (Hamosh et al., 2005)	<i>xlid.omim</i>	0.32
X-linked intellectual disability (combined)	<i>xlid</i>	0.04

**Table 3** GO gene sets achieving MLP>3 (out of 1454 sets in total).

Gene set	MLP
CHROMOSOME_ORGANIZATION_AND_BIOGENESIS	3.49
ESTABLISHMENT_AND_OR_MAINTENANCE_OF_CHROMATIN_ARCHITECTURE	3.48
HISTONE_MODIFICATION	3.44
NEGATIVE_REGULATION_OF_CELL_DIFFERENTIATION	3.42
REGULATION_OF_MYELOID_CELL_DIFFERENTIATION	3.41
TRANSCRIPTION_FACTOR_COMPLEX	3.36
COVALENT_CHROMATIN_MODIFICATION	3.35
CHROMATIN_MODIFICATION	3.31
SENSORY_ORGAN_DEVELOPMENT	3.18
TRANSCRIPTION_COACTIVATOR_ACTIVITY	3.12
NEGATIVE_REGULATION_OF_MYELOID_CELL_DIFFERENTIATION	3.07

**Table 4.** Genes from the GO sets with MLP>3 where the gene-wise MLP exceeds 1.3 (equivalent to  $p<0.05$ ).

Gene	beta	MLP
HDAC9	0.016	4.03
RBM14	0.018	3.42
MAFB	-0.028	3.19
ZMIZ2	-0.015	2.83
CHMP1A	0.011	2.64
JAG2	-0.014	2.35
FOXO3	-0.008	2.22
TAF11	-0.011	2.14
SNAPC4	-0.014	2.05
MAML2	-0.013	1.89
TGFB2	-0.015	1.81
SUPT4H1	-0.008	1.76
SMARCD3	0.010	1.70
NSD1	-0.005	1.67
DKC1	-0.009	1.65
HDAC6	0.006	1.59
HCFC1	-0.008	1.55
HUWE1	-0.006	1.53
SUPT16H	-0.007	1.51
ERCC4	-0.008	1.51
CARTPT	0.044	1.50
HDAC3	0.008	1.48

TERF2IP	-0.007	1.47
ACIN1	0.006	1.41
CENPE	0.017	1.40
GTF3C4	-0.006	1.37
PDS5B	-0.006	1.35
SIRT5	-0.007	1.32
RNF14	-0.006	1.32

**Figure 1.** Controls (a) and subjects with schizophrenia (b) plotted against first two principal components from SNP genotypes.

**Figure 2.** Q:Q plot of observed versus expected values for gene-wise SLPs assessing correlation between PRS and gene expression.