bioRxiv preprint doi: https://doi.org/10.1101/203521; this version posted November 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Weighted burden analysis of exome-sequenced case-control sample implicates synaptic genes in schizophrenia aetiology

[1,2]David Curtis*, [1]Leda Coelewij, [1]Shou-Hwa Liu, [1,3]Jack Humphrey, [1]Richard Mott.

[1]UCL Genetics Institute, University College London.

[2]Centre for Psychiatry, Barts and the London School of Medicine and Dentistry.

[3]Department of Neurodegenerative Disease, UCL Institute of Neurology, University College London.

*Correspondence: Darwin Building, Gower Street London WC1E 6BT, UK. d.curtis@ucl.ac.uk

Abstract

A previous study of exome-sequenced schizophrenia cases and controls reported an excess of singleton, gene-disruptive variants among cases, concentrated in particular gene sets. The dataset included a number of subjects with a substantial Finnish contribution to ancestry. We have reanalysed the same dataset after removal of these subjects and we have also included non-singleton variants of all types using a weighted burden test which assigns higher weights to variants predicted to have a greater effect on protein function. We investigated the same 31 gene sets as previously and also 1454 GO gene sets. The reduced dataset consisted of 4225 cases and 5834 controls. No individual variants or genes were significantly enriched in cases but 13 out of the 31 gene sets were significant after Bonferroni correction and the “FMRP targets” set produced a signed log p value (SLP) of 7.1. The gene within this set with the highest SLP, equal to 3.4, was *FYN*, which codes for a tyrosine kinase which phosphorylates glutamate metabotropic receptors and ionotropic NMDA receptors, thus modulating their trafficking, subcellular distribution and function. In the most recent GWAS of schizophrenia it was identified as a "prioritized candidate gene". Two of the subunits of the NMDA receptor which are substrates of *FYN* are coded for by *GRIN1* (SLP=1.7) and *GRIN2B* (SLP=2.1). Of note, for some sets there was a substantial enrichment of non-singleton variants. Of 1454 GO gene sets, 3 were significant after Bonferroni correction. Identifying specific genes and variants will depend on genotyping them in larger samples and/or demonstrating that they cosegregate with illness within pedigrees.

Introduction

Schizophrenia is a severe and disabling mental illness with onset typically in early adult life. It is associated with low fecundity but nevertheless remains fairly common with a lifetime prevalence of around 1% (Power et al., 2013). A variety of types of genetic variation contribute to risk. Many common variants demonstrate association with small effect sizes whereas extremely rare variants can have very large effect sizes. 108 SNPs have been reported to be genome-wide significant with odds ratio (OR) of 1.1-1.2 and it is likely that many other variants will achieve statistical significance when larger samples are genotyped (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Weak effects from common variants may arise from a number of mechanisms. The variant itself

bioRxiv preprint doi: https://doi.org/10.1101/203521; this version posted November 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

may exert a direct effect at some point in the pathogenic process, it may pick a up a more indirect effect through involvement in gene regulatory networks or it may be in linkage disequilibrium with other variants have a larger, direct effect (Boyle et al., 2017). A recent example of the last case is provided by SNPs in the HLA region which tag variant haplotypes of C4, the gene for complement component 4, the different haplotypes producing different levels of C4A expression associated with OR for schizophrenia risk of 1.3 (Sekar et al., 2016). Variants associated with small effect sizes will be subject to relatively little selection pressure and hence can remain common. By contrast, extremely rare variants such as some copy number variants (CNVs) or loss of function (LOF) variants of SETD1A may lead to a very high risk of developing schizophrenia (Deciphering Developmental Disorders Study, 2017; Rees et al., 2014; Singh et al., 2016). A proportion of cases of schizophrenia seem to be due to such variants with large effect size arising as de novo mutations (DNMs) (Fromer et al., 2014; Singh et al., 2017). Such variants are likely to be subject to strong selection pressure and may only persist for a small number of generations. Theoretically, variants acting recessively might persist in the population and still have reasonably large effect size but attempts to identify these have to date been unsuccessful (Curtis, 2015; Rees et al., 2015; Ruderfer et al., 2014).

In order to focus attention on only new or recent variants, the Swedish schizophrenia study of whole exome sequence data focussed on what were termed ultra-rare variants (URVs), that is variants which only occurred in a single subject and which were absent from ExAC. The effects of some of these variants on gene function were annotated as damaging or disruptive and these variants, termed dURVs, were found to be commoner in cases than controls across all genes, with the effect concentrated in particular sets of genes including FMRP targets, synaptically localised genes and genes which were LOF intolerant (Genovese et al., 2016). The present study seeks to analyse this dataset further in order to consider whether rare non-singleton sequence variants, as well as singleton variants, contribute to schizophrenia risk.

The dataset used in this study overlaps with a number of previously reported analyses. The full exome-sequenced dataset consists of 4968 cases with schizophrenia and 6245 controls. Although recruited in Sweden, it should be noted that some subjects have a substantial Finnish component to their ancestry (Genovese et al., 2016). The earlier phase of this dataset consisted of 2045 cases and 2045 controls and the primary analysis of these revealed subjects an excess among cases of very rare, disruptive mutations spread over a number of different genes though concentrated in particular gene sets (Purcell et al., 2014). This first phase of the dataset was also used for analyses which attempted to detect recessive effects and to identify Gene Ontology (GO) pathways with an excess of rare, functional variants among cases but which did not produce statistically significant results (Curtis, 2016, 2013). A subset of the full dataset with cases with Finnish ancestry removed was used to demonstrate a method for deriving an exome-wide risk score and to demonstrate an association of schizophrenia with variants in mir134 binding sites (Curtis, 2017; Curtis and Emmett, 2017). A genetically homogeneous subset of the full Swedish dataset was combined with a UK case-control association sample and nonsynonymous variants with Minor Allele Frequency (MAF)<0.001 which were present on the Illumina HumanExome and HumanOmniExpressExome arrays were analysed (Leonenko et al., 2017). This revealed an enrichment of these variant alleles in LOF intolerant genes and FMRP targets.

bioRxiv preprint doi: https://doi.org/10.1101/203521; this version posted November 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The present study uses a subset of the Swedish dataset after removal of subjects with a high Finnish ancestry component in order to avoid artefactual results produced by population stratification. It also utilises all rare (MAF<0.01) variants analysed using a weighted burden test to identify genes and sets of genes associated with schizophrenia risk.

Methods

The data analysed consisted of whole exome sequence variants downloaded from dbGaP from the Swedish schizophrenia association study containing 4968 cases and 6245 controls (Genovese et al., 2016). The dataset was managed and annotated using the GENEVARASSOC program which accompanies SCOREASSOC (https://github.com/davenomiddlenamecurtis/geneVarAssoc). Version hg19 of the reference human genome sequence and RefSeq genes were used to select variants on a gene-wise basis. Members of the protocadherin gamma gene cluster, whose transcripts overlap each other but which are entered separately in RefSeq, were treated as a single gene which was labelled PCDHG.

A number of QC processes were applied. Variants were excluded if they did not have a PASS in the Variant Call Format (VCF) information field and individual genotype calls were excluded if they had a quality score less than 30. Sites were also excluded if there were more than 10% of genotypes missing or of low quality in either cases or controls or if the heterozygote count was smaller than both homozygote counts in both cohorts. As previously reported (Curtis, 2017), preliminary gene-wise weighted burden tests revealed that several genes had an apparent excess of rare, protein-altering variants in cases but that these results were driven by variants which were reported in ExAC to be commoner in Finnish as opposed to non-Finnish Europeans (Lek et al., 2016). Accordingly, subjects with an excess of alleles more frequent in Finns were identified and removed from the dataset, comprising 743 cases and 411 controls. Once this had been done, leaving a sample of 4225 cases and 5834 controls, the gene-wise weighted burden test results conformed well to what would be expected under the null hypothesis with no evidence for inflation of the test statistic across the majority of genes not thought to be implicated in disease.

The tests previously carried out for an excess of dURVs among cases (Genovese et al., 2016) were performed on both the full and reduced datasets, with and without including covariates consisting of the total URV count and the first 20 principal components from the SNP and indel genotypes.

Weighted burden analysis of genes and gene sets as described below was carried out using SCOREASSOC, which analyses all variants simultaneously and can accord each variant a different weight according to its MAF and its predicted function (Curtis, 2016, 2012). Each variant was annotated using VEP, PolyPhen and SIFT (Adzhubei et al., 2013; Kumar et al., 2009; McLaren et al., 2016). GENEVARASSOC was used to generate the input files for SCOREASSOC and the default weights were used, for example consisting of 5 for a synonymous variant and 20 for a stop gained variant, except that 10 was added to the weight if the PolyPhen annotation was possibly or probably damaging and also if the SIFT annotation was deleterious. The full set of weights is shown in supplementary Table S1. SCOREASSOC also weights variants more highly than common ones but because it is well-

bioRxiv preprint doi: https://doi.org/10.1101/203521; this version posted November 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

established that no common variants have a large effect on the risk of schizophrenia we excluded variants with MAF>0.01 in the cases and in the controls, so in practice weighting by rarity had negligible effect. For each subject a gene-wise risk score was derived as the sum of the variant-wise weights, each multiplied by the number of alleles of the variant which the given subject possessed. These scores were then compared between cases and controls using a t test. To indicate the strength of evidence in favour of an excess of rare, functional variants in cases we took the logarithm base 10 of the p value from this t test and then gave it a positive sign if the average weighted sum was higher in cases and a negative sign if the average was higher in controls, to produce a signed log p (SLP).

In order to explore the contribution of singleton variants, for the analyses of gene sets three sets of variants were used: singleton variants which were only observed in a single subject and not in ExAC; non-singleton variants, observed in more than one subject (though still with MAF<0.01 in cases and/or controls); all variants, consisting of these singleton and non-singleton combined.

Weighted burden analysis within sets of genes was carried out using PATHWAYASSOC, which for each subject sums up the gene-wise scores to produce an overall score for the gene set. These set-wise scores can then be compared between cases and controls using a t-test. This approach has been demonstrated to produce appropriate p values through application to real data, supported by permutation testing (Curtis, 2016). This analysis was applied to the 31 gene sets used in the Swedish study separately using singleton, non-singleton and all variants. The analysis was also applied using all variants to the 1454 "all GO gene sets, gene symbols" pathways downloaded from the Molecular Signatures Database at http://www.broadinstitute.org/gsea/msigdb/collections.jsp (Subramanian, Tamayo et al. 2005).

Logistic regression analyses of dURVs were carried out using R (R Core Team, 2014). Weighted burden tests for genes and gene sets were carried out using SCOREASSOC and PATHWAYASSOC. Results from these programs are expressed as a Signed Log P (SLP) which is positive if there is an excess of variants among cases and negative if there is an excess among controls. Thus, a SLP of 3 would indicate that there was an excess of variants among cases with two-tailed significance $P<10^{-3}$.

Results

Preliminary analysis of the whole dataset, (i.e. all individuals before excluding those with Finnish ancestry), using a logistic regression analysis to test for an excess of dURVs among cases was significant ($P=8.7*10^{-10}$) when the total URV count and principal components were included as covariates. However without covariates this analysis was only marginally significant (P=0.031). Further investigation showed that subjects with a substantial Finnish component to their ancestry had a larger number of URVs than those who did not. Cases tended to have a larger number of dURVs than controls, but only relative to the total number of URVs, and more cases had a substantial Finnish ancestry component than controls. Thus, in the whole sample the relative excess of dURVs among cases was almost completely masked by the fact that more cases had Finnish ancestry and that these cases had a smaller absolute number of URVs, meaning that overall there was only a small excess in the absolute number of dURVs among cases. Including the total URV count or the principal components or both as covariates allowed the relative excess among cases to

bioRxiv preprint doi: https://doi.org/10.1101/203521; this version posted November 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

become apparent. The analysis was then repeated on the reduced dataset without those subjects with a substantial Finnish ancestry component. Once this had been done, there was a significant absolute excess of dURVs among cases ($P=2.7*10^{-5}$), without needing to include either total URV count or principal components as covariates.

The weighted burden tests evaluated 1,042,483 valid variants in 22,023 genes. As described in the *Methods* section, in preliminary analyses using the full dataset a number of genes yielded high SLPs. An example was *COMT*, with SLP=7.4. On inspection, it seemed that this gene-wise result was largely driven by SNP rs6267, which was heterozygous in 51/6242 controls and 94/4962 cases (OR=2.3, $p=8*10^{-7}$). However this variant is noted in ExAC to have MAF=0.002 in non-Finnish Europeans but MAF=0.05 in Finns. Hence, its increased frequency among cases appeared to be due to the excess of cases with Finnish ancestry. Once all subjects with a substantial Finnish ancestry component were excluded, the SLP for *COMT* fell to 1.7 and for rs6267 there were 36/5831 heterozygous controls s and 36/4221 cases (OR=1.4, p=0.2). A similar effect was observed for other genes with excessively high SLPs in the full dataset but not in the reduced dataset, suggesting that removing subjects with substantial Finnish ancestry seemed to produce a satisfactorily homogeneous dataset. QQ-plots for the gene-wise analyses using the reduced dataset are shown in Figure 1. All of the plots are symmetrical, indicating that the test is unbiased. When only singleton variants are used the gene-wise tests are somewhat underpowered and the gradient is less than 1. However for the tests using non-singleton variants or all variants the SLPs almost exactly follow the distribution expected under the null hypothesis. One outlier is apparent. This is caused by the gene *CDCA8* which produces an SLP of -5.49 with all variants. Further inspection showed that this result was mainly driven by 22 highly weighted variant alleles among controls but only 5 among cases. For a gene-wise test to be exome-wide significant with 22,023 genes the absolute value of the SLP would need to exceed 5.64, so this result is still within chance expectation.

The results for the 31 gene sets which had previously been used in the Swedish study are shown in Table 1. Using the weighted burden test many, though not all, of the sets show an excess of variants among cases. For *neurons*, *pLI09*, *fmrp* and *mir137* the non-singleton variants make a substantial contribution but for *psd*, *rbfox13* and *rbfox2* the bulk of the effect comes from only the singleton variants. Given that there are 31 sets, a simple Bonferroni correction would mean that a set could be declared statistically significant if the SLP using all variants exceeded -log(31/0.05)=2.8 although this threshold should be regarded as conservative because the sets overlap each other. For the 13 sets where SLP>2.8 using all variants, the genes with the highest gene-wise SLPs are shown in Table 2. As expected, there is some overlap between the sets with several genes making contributions to more than one set. The gene with the highest gene-wise SLP in the *fmrp* set is *FYN* (SLP=3.4) and it is also a member of 6 other sets. *FYN* codes for a tyrosine kinase which phosphorylates glutamate metabotropic receptors and ionotropic NMDA receptors, which modulates their trafficking, subcellular distribution and function (Mao and Wang, 2016a) In the most recent GWAS of schizophrenia *FYN* was identified as a "prioritized candidate gene" and an intronic marker, rs7757969, was significant at $p=4.8*10^{-8}$ (Li et al., 2017). The activity of *FYN* is regulated by dopamine DRD2 receptors (Mao and Wang, 2016b). *FYN* is involved in neuronal apoptosis, brain development and synaptic transmission and lower expression has been observed in the platelets of schizophrenic patients compared with controls (Ali and Salter, 2001; Du et al., 2012; Hattori et al., 2009). Two of the subunits of the NMDA receptor

bioRxiv preprint doi: https://doi.org/10.1101/203521; this version posted November 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

which are substrates of *FYN* are coded for by *GRIN1* (SLP=1.7) and *GRIN2B* (SLP=2.1). In all three of these genes, the signal seems to be produced from a number of highly weighted variants which are individually commoner in cases but all are very rare, with MAF<0.001 even among cases, so it is not possible to identify any obvious candidate variants.

Figure 2 shows the QQ plot for the set-wise analyses using the GO gene sets. Given that there is overlap of genes between sets, the SLPs are non-independent and it is expected that the gradient of the QQ plot will be less than 1. For those sets with a negative SLP this is indeed the case and these results are in accordance with the expectation under the null hypoethesis. However the gradient becomes steeper for sets with positive SLPs and this can be interpreted as showing that some sets have an excess of variants among cases above that which would be expected by chance. Given that 1454 GO gene sets were tested, a simple Bonferroni correction would mean that a test could be declared "exome-wide significant" if it achieved an SLP exceeding -log(1454/0.05)=4.5. Three sets did achieve this threshold. However, given the fact that the set-wise SLPs are not independent a Bonferroni correction might be viewed as conservative and Table 3 shows all sets achieving SLP>3. The full results are presented in supplementary Table S2. The most significant set, INTRACELLULAR_SIGNALING_CASCADE with SLP=5.4, contains *FYN* and two other genes with gene-wise SLP>3, *S1PR4* (SLP=3.7) and *RTKN* (SLP=3.2). S1PR4 codes for the type 4 receptor for sphingosine-1-phosphate and the mouse strain carrying the mutation with genotype S1pr4^{tm1Dgen}/S1pr4^{+} has decreased prepulse inhibition as a phenotype (http://www.informatics.jax.org/allele/genoview/MGI:3606610) (Blake et al., 2017; The Jackson Laboratory, n.d.). *RTKN* codes for rhotekin, a scaffold protein that interacts with GTP-bound Rho proteins. Again, inspecting results for individual variants within these genes did not reveal any obvious candidates. The full results for all genes and all gene sets can be downloaded at:
http:/www.davecurtis.net/downloads/SSS2WeightedBurdenAnalysisResults.zip.

Discussion

This analysis identifies a number of sets of genes that meet Bonferroni-corrected criteria for statistical significance. It differs from previous analyses in a number of ways.

In contrast to the original analysis of the Swedish dataset (Genovese et al., 2016) it uses non-singleton as well as singleton variants and it clearly demonstrates that there is a contribution to risk from these non-singleton variants. This is extremely important in terms of the prospects for identifying rare risk variants for schizophrenia. If only unique variants conferred risk, that is only variants which occur independently as *de novo* mutations and then disappear after a small number of generations, then it would not be possible to identify any single variant as definitively affecting risk. One could at best identify perhaps classes of variant occurring in particular genes. Without being able to conclude that any particular variant affected risk, one could not carry out functional studies in model systems with the confidence that one was indeed studying a true risk variant. Additionally, if only unique variants contributed to risk then strategies that might use linkage disequilibrium to implicate untyped variants could not succeed. If, on the other hand, there are risk variants which survive and spread in the population then potentially these could be tagged by haplotypes of common SNPs and imputed from GWAS data, in a way similar to that used to impute C4 risk

bioRxiv preprint doi: https://doi.org/10.1101/203521; this version posted November 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

variants (Sekar et al., 2016). Alternatively population sequencing may soon become cheap and accurate enough to identify these rare variants directly.

This study differs from both the Swedish study (Genovese et al., 2016) and the Swedish-UK study (Leonenko et al., 2017) in that it uses a homogeneous dataset. The original study did not exclude the subjects with a substantial Finnish ancestry component whereas the Swedish-UK study did use a homogeneous subset of the Swedish subjects but then combined them with a UK sample. This meant that both studies needed to incorporate principal components to control for population stratification and this to some extent complicates the interpretation of their results. For example, the highly significant enrichment for dURVs reported in the first study only becomes apparent when covariates are included. In the Swedish-UK study, the most highly significant variant ($p=3.4*10^{-7}$), which occurs in the *MCPH* gene, has MAF of 0.0046 in cases and of 0.0012 in controls, meaning that the unadjusted risk ratio is approximately 3.8. However after multivariate analysis including covariates the OR is reported as being only 1.2. By contrast, the reduced dataset we have used appears to be sufficiently homogeneous that the test statistic performs as expected without requiring any adjustment for population stratification. This allows for a simple, straightforward interpretation of the results obtained.

Another way our analysis differs is that it includes all variants in a single analysis. Variants are assigned different weights according to an arbitrary pre-specified set of weights designed to emphasise those variants more likely to affect gene function. This meant that we carried out only a single analysis for each gene or set of genes, reducing any correction for multiple-testing. Our analyses utilised 1,042,483 variants, compared with the 112,950 used in the Swedish-UK study. Using our method, 14 of the 32 candidate gene sets and 3 of the 1454 GO sets meet formal standards for statistical significance using a conservative Bonferroni correction.

As in the other studies, none of the results for individual genes reach formal standards for statistical significance, although the results obtained for *FYN* are possibly of interest. It seems likely that our results are detecting a real signal originating from rare variants concentrated within some of the genes that are members of the gene sets with high SLPs. These sets overlap each other to a considerable extent and it is difficult to tease out which ones best define a group of schizophrenia risk genes. An attempt to do this formally using exome-wide risk scores did not produce definitive results (Curtis, 2017). It should be noted that different sets might be implicated for different reasons. For example, it may be that the high SLP for targets of *miR-137* occurs because disruption of the regulation of these genes by *miR-137* can lead to increased risk of schizophrenia, as supported by the association of schizophrenia with markers for *miR-137* and with variants in its binding sites (Curtis and Emmett, 2017; Olde Loohuis et al., 2017). On the other hand, there is no reported association of *FMRP* itself with schizophrenia and the high SLP for its targets may simply reflect that this identifies a group of genes whose mRNA is localised to the synapse. In any event, it is clear that with samples currently available we are only able to identify very broad gene sets but not yet specific genes.

With increased sample sizes it will become possible to identify specific genes and variants which have a moderate or large effect on risk. However such variants, although not singletons, will still be very rare and serious attention should be focussed on complementary approaches to confirm them. One such approach would be to use exome sequence data

bioRxiv preprint doi: https://doi.org/10.1101/203521; this version posted November 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

from affected subjects to provide reference haplotypes for imputation into large GWAS datasets, analogously to the way C4 variants implicating risk were identified (Sekar et al., 2016). Another would be to search for affected relatives of subjects with candidate variants in order to see if the variants cosegregate with disease, a strategy which was successful in implicating *RBM12* in the aetiology of psychosis (Curtis, 2011; Steinberg et al., 2017). If and when specific variants are identified as having substantial effects on risk then they can be incorporated into model systems in order to gain insight into the mechanisms affecting the development of schizophrenia.

Acknowledgements

The authors wish to thank Giulio Genovese for his assistance in providing supporting files and responding to queries. The datasets used for the analysis described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number *phs000473.v2.p2*. Samples used for data analysis were provided by the Swedish Cohort Collection supported by the NIMH grant R01MH077139, the Sylvan C. Herman Foundation, the Stanley Medical Research Institute and The Swedish Research Council (grants 2009-4959 and 2011-4659). Support for the exome sequencing was provided by the NIMH Grand Opportunity grant RCMH089905, the Sylvan C. Herman Foundation, a grant from the Stanley Medical Research Institute and multiple gifts to the Stanley Center for Psychiatric Research at the Broad Institute of MIT and Harvard. JH is supported by MRC studentship 516702.

Conflict of interest

The authors declare they have no conflict of interest.

bioRxiv preprint doi: https://doi.org/10.1101/203521; this version posted November 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1

Results showing SLPs obtained for the gene sets used in in the original analysis Swedish schizophrenia study (Genovese et al., 2016). The lists of genes were obtained directly from the first author. The symbol used is the same as that used for the name of the file containing the list.

Gene set	Symbol (Number of genes in set)	Singleton variants	Non-singleton variants	Both
OMIM intellectual disability (Hamosh et al., 2005)	*alid* (107)	0.2	1.0	1.1
Expression specific to brain (Fagerberg et al., 2014)	*brain* (2660)	4.0	1.3	3.1
Bound by CELF4 (Wagnon et al., 2012)	*celf4* (2675)	3.1	1.7	3.7
Missense-constrained (Samocha et al., 2014)	*constrained* (1005)	3.8	2.0	4.8
Involved in developmental disorder (Deciphering Developmental Disorders Study, 2017)	*dd* (93)	2.2	2.4	3.7
De novo variants in autism (Fromer et al., 2014)	*denovo.aut* (2927)	2.5	2.9	4.8
De novo variants in coronary heart disease (Fromer et al., 2014)	*denovo.chd* (249)	0.8	1.7	2.7
De novo variants in epilepsy (Fromer et al., 2014)	*denovo.epi* (322)	1.2	0.7	1.6
De novo duplications in ASD (Kirov et al., 2012)	*denovo.gain.asd* (1365)	0.9	1.2	1.8
De novo duplications in bipolar disorder (Kirov et al., 2012)	*denovo.gain.bd* (180)	0.8	0.5	1.2
De novo duplications in schizophrenia (Kirov et al., 2012)	*denovo.gain.scz* (200)	0.2	-0.1	0.1
De novo variants in intellectual disability (Fromer et al., 2014)	*denovo.id* (251)	0.5	1.8	2.8
De novo deletions in ASD (Kirov et al., 2012)	*denovo.loss.asd* (1179)	3.1	0.2	1.3
De novo deletions in bipolar disorder (Kirov et al., 2012)	*denovo.loss.bd* (130)	1.4	-0.3	0.2
De novo deletions in schizophrenia (Kirov et al., 2012)	*denovo.loss.scz* (246)	0.6	0.1	0.5
De novo variants in schizophrenia (Fromer et al., 2014)	*denovo.scz* (770)	1.7	1.3	2.3
Bound by FMRP (Darnell et al., 2011)	*fmrp* (1244)	7.0	3.3	7.2
Implicated by GWAS (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014)	*gwas* (91)	1.2	0.8	1.7
Targets of microRNA-137 (Robinson et al., 2015)	*mir137* (3260)	2.5	4.1	5.3

bioRxiv preprint doi: https://doi.org/10.1101/203521; this version posted November 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Expression specific to neurons (Cahoy et al., 2008)	*neurons* (4747)	3.4	4.3	6.9
NMDAR and ARC complexes (Kirov et al., 2012)	*nmdarc* (80)	1.8	-0.4	0.1
Loss-of-function intolerant (Lek et al., 2016)	*pLI09* (3488)	4.2	3.3	6.2
PSD-95 (Bayés et al., 2011)	*psd95* (120)	2.7	-0.2	0.5
Bound by RBFOX 1 or 3 (Weyn-Vanhentenryck et al., 2014)	*rbfox13* (3445)	5.7	1.3	4.2
Bound by RBFOX 2 (Weyn-Vanhentenryck et al., 2014)	*rbfox2* (3068)	6.4	1.0	4.1
Synaptic (Pirooznia et al., 2012)	*synaptome* (1887)	3.9	2.2	5.4
Escape X-inactivation (Cotton et al., 2013)	*x.escape* (213)	0.5	0.9	1.6
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)	*xlid.chicago* (77)	-0.1	1.8	1.4
X-linked intellectual disability, Greenwood Genetic Centre (Moeschler et al., 2006)	*xlid.gcc* (114)	-0.2	1.8	1.3
X-linked intellectual disability, OMIM (Hamosh et al., 2005)	*xlid.omim* (57)	-0.7	0.6	0.2
X-linked intellectual disability (combined)	*xlid* (122)	-0.3	1.8	1.2

Table 2

Gene-wise results for the genes with highest gene-wise SLPs in all sets with set-wise SLP>2.8. The top ten genes are shown, providing that the gene-wise SLP was at least 1.3, equivalent to p<0.05.

brain		*celf4*		*constrained*		*dd*		*denovo.aut*		*denovo.id*		*fmrp*	
DGKI	3.3	ADAMTSL1	4.3	KLHL11	3.7	GRIN2B	2.1	ADAMTSL1	4.3	ARFGEF2	2.5	FYN	3.4
SLC6A17	3.1	HPRT1	4.0	TMEM102	2.3	PACS1	2.0	TMC4	4.0	CDC42BPB	2.2	SLC6A17	3.1
AAK1	2.9	KLHL11	3.7	TIGD5	2.3	KCNQ3	1.8	OR10Z1	3.2	EPHB1	2.2	AAK1	2.9
EFNB3	2.8	PLK4	3.4	HERC1	2.3	ANKRD11	1.7	VAMP2	2.4	GRIN2B	2.1	AFF3	2.8
NDST3	2.7	DGKI	3.3	AGO3	2.2	KIF1A	1.6	FOCAD	2.4	TMPRSS12	1.8	PTK2	2.7
GLT6D1	2.6	GMCL1	3.3	DGKZ	2.2	KCNH1	1.5	C20orf96	2.3	KCNQ3	1.8	PREX2	2.5
TMEM174	2.5	CCDC112	3.1	SLIT1	2.2	DYNC1H1	1.3	HERC1	2.3	MBD5	1.7	ARFGEF2	2.5
HCRTR2	2.4	SLC6A17	3.1	DNMT3A	2.1	KAT6A	1.3	AGO3	2.2	TNK2	1.7	VAMP2	2.4
EPHA5	2.4	AAK1	2.9	KDM5C	2.1			RNF25	2.2	SETDB2	1.6	HERC1	2.3
PACSIN1	2.3	AFF3	2.8	TFAP2A	2.1			CDC42BPB	2.2	KCNH1	1.5	PACSIN1	2.3

(continued)

mir137		*neurons*		*pLI09*		*rbfox13*		*rbfox2*		*synaptome*	
ADAMTSL1	4.3	ADAMTSL1	4.3	HPRT1	4.0	HPRT1	4.0	HPRT1	4.0	GLIPR2	3.6
HPRT1	4.0	HPRT1	4.0	FYN	3.4	KLHL11	3.7	KLHL11	3.7	FYN	3.4
GLIPR2	3.6	GMCL1	3.3	PITPNA	3.2	FYN	3.4	FYN	3.4	PITPNA	3.2
FYN	3.4	PITPNA	3.2	SLC6A17	3.1	DGKI	3.3	DGKI	3.3	SLC6A17	3.1
DGKI	3.3	CCDC112	3.1	CSRNP2	2.9	GMCL1	3.3	PITPNA	3.2	AAK1	2.9
PITPNA	3.2	SLC6A17	3.1	AAK1	2.9	PITPNA	3.2	SLC6A17	3.1	PTK2	2.7
CD302	2.9	AAK1	2.9	AFF3	2.8	SLC6A17	3.1	RABEP2	3.0	HCK	2.7
ALG6	2.9	AFF3	2.8	PSME3	2.8	AAK1	2.9	AAK1	2.9	RPS15A	2.5
AFF3	2.8	PSME3	2.8	PTK2	2.7	AFF3	2.8	AFF3	2.8	BLNK	2.5
CXCL11	2.8	EFNB3	2.8	BCL6	2.7	PSME3	2.8	PSME3	2.8	ARFGEF2	2.5

bioRxiv preprint doi: https://doi.org/10.1101/203521; this version posted November 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 3

Table showing all of the 1454 GO gene sets which produced set-wise SLP>3.

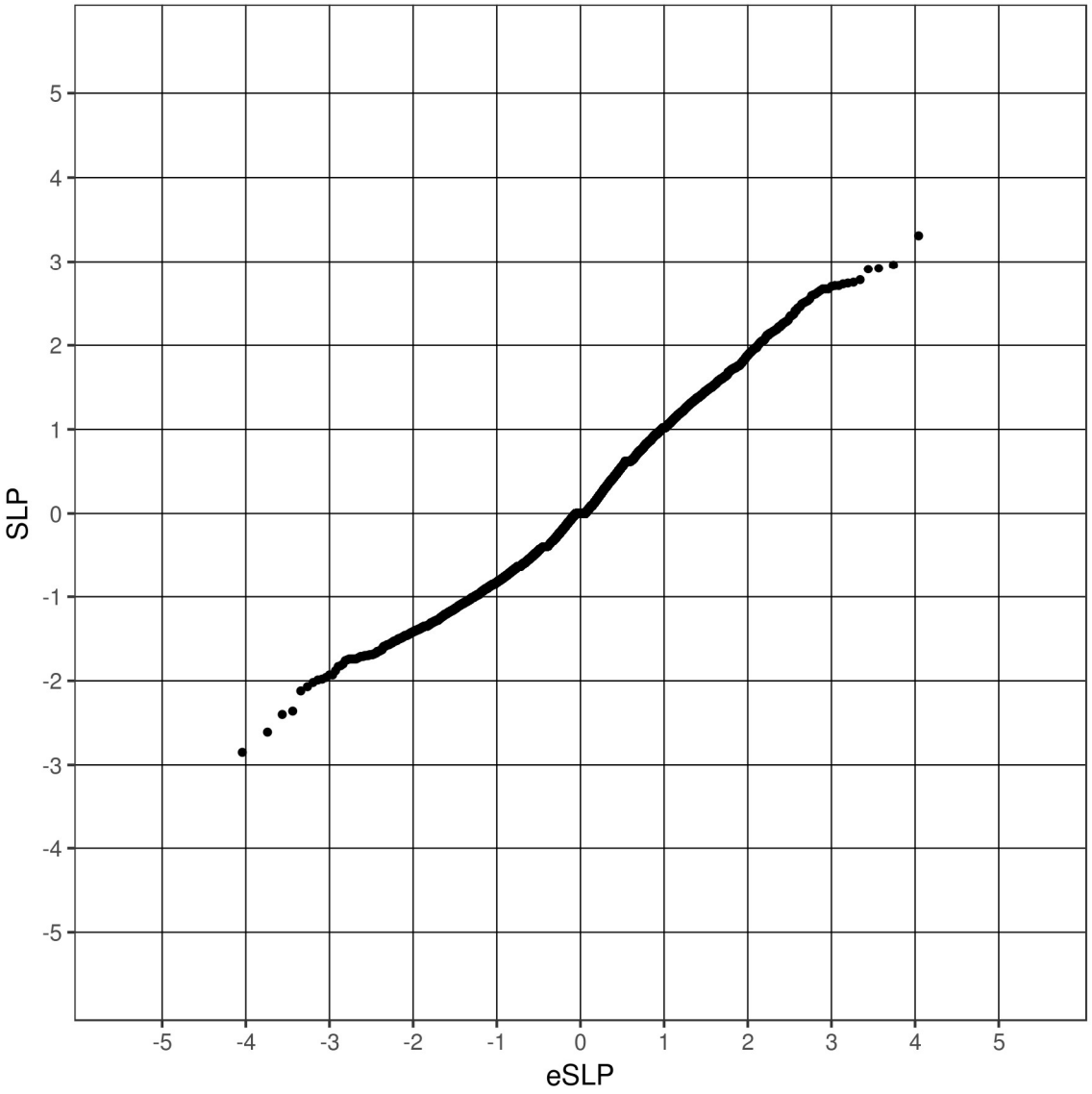
GO gene set	SLP
INTRACELLULAR_SIGNALING_CASCADE	5.39
CHROMOSOME_ORGANIZATION_AND_BIOGENESIS	4.70
ORGAN_DEVELOPMENT	4.64
SIGNAL_TRANSDUCTION	4.37
ION_BINDING	4.34
POSITIVE_REGULATION_OF_CELLULAR_PROCESS	4.22
REGULATION_OF_CELLULAR_METABOLIC_PROCESS	4.11
RHO_GUANYL_NUCLEOTIDE_EXCHANGE_FACTOR_ACTIVITY	4.11
CELL_DEVELOPMENT	4.09
CYTOPLASM	4.00
REGULATION_OF_METABOLIC_PROCESS	3.99
POSITIVE_REGULATION_OF_BIOLOGICAL_PROCESS	3.86
STRUCTURE_SPECIFIC_DNA_BINDING	3.69
PROTEIN_METABOLIC_PROCESS	3.38
TRANSMEMBRANE_RECEPTOR_ACTIVITY	3.37
SEXUAL_REPRODUCTION	3.34
FEEDING_BEHAVIOR	3.22
REGULATION_OF_PROTEIN_AMINO_ACID_PHOSPHORYLATION	3.22
NEGATIVE_REGULATION_OF_BIOLOGICAL_PROCESS	3.22
CELL_ACTIVATION	3.15
INTEGRAL_TO_MEMBRANE	3.15
REGULATION_OF_PHOSPHORYLATION	3.13
INTRINSIC_TO_MEMBRANE	3.10
GAMETE_GENERATION	3.10
REGULATION_OF_DEVELOPMENTAL_PROCESS	3.09
ESTABLISHMENT_AND_OR_MAINTENANCE_OF_CHROMATIN_ARCHITECTURE	3.09
MEMBRANE	3.06
BIOPOLYMER_METABOLIC_PROCESS	3.02
NEGATIVE_REGULATION_OF_CELLULAR_PROCESS	3.01

bioRxiv preprint doi: https://doi.org/10.1101/203521; this version posted November 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1

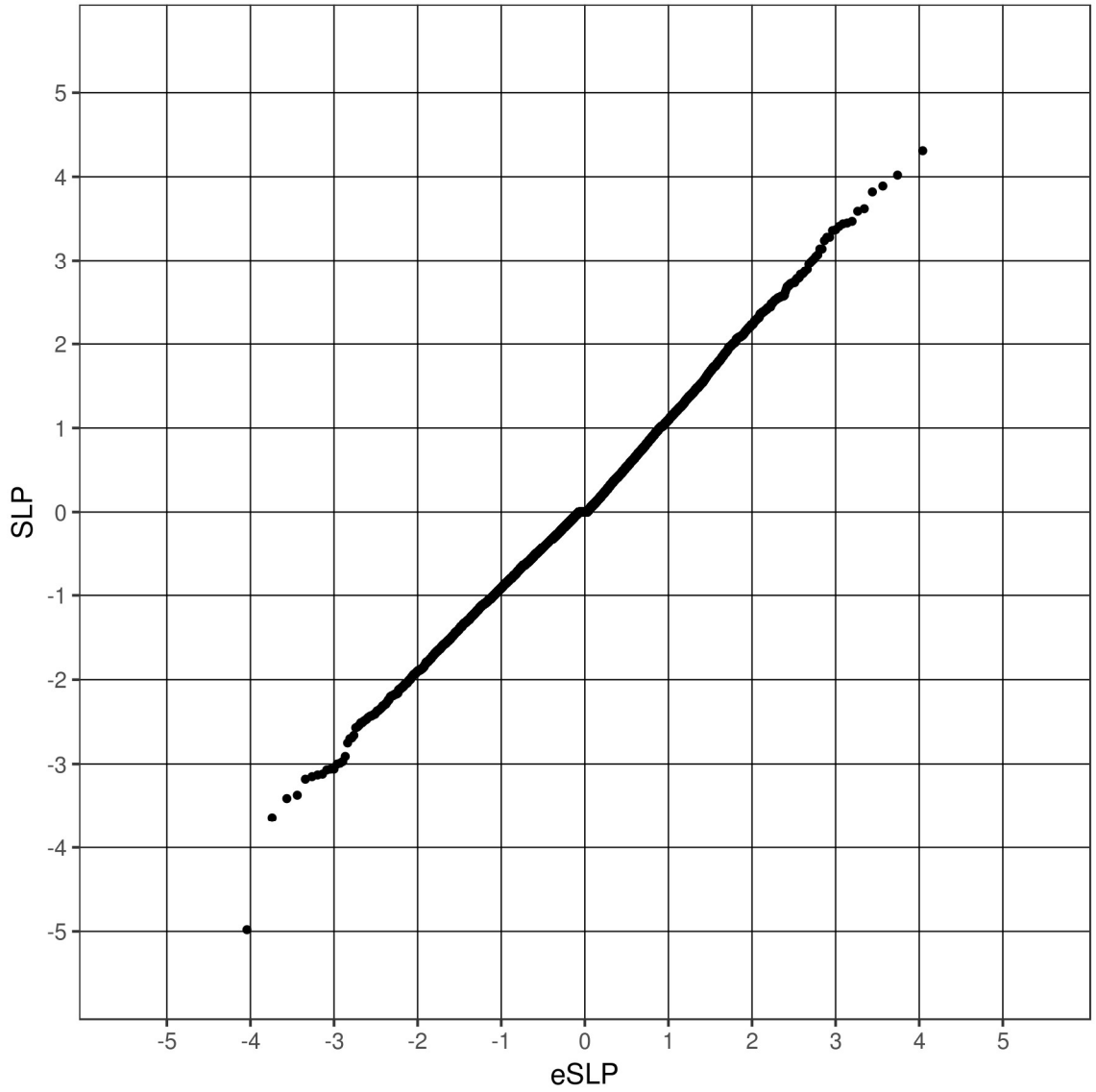
QQ plots of observed versus expected gene-wise SLP using (a) only singleton variants, (b) non-singleton variants and (c) both.

(a)



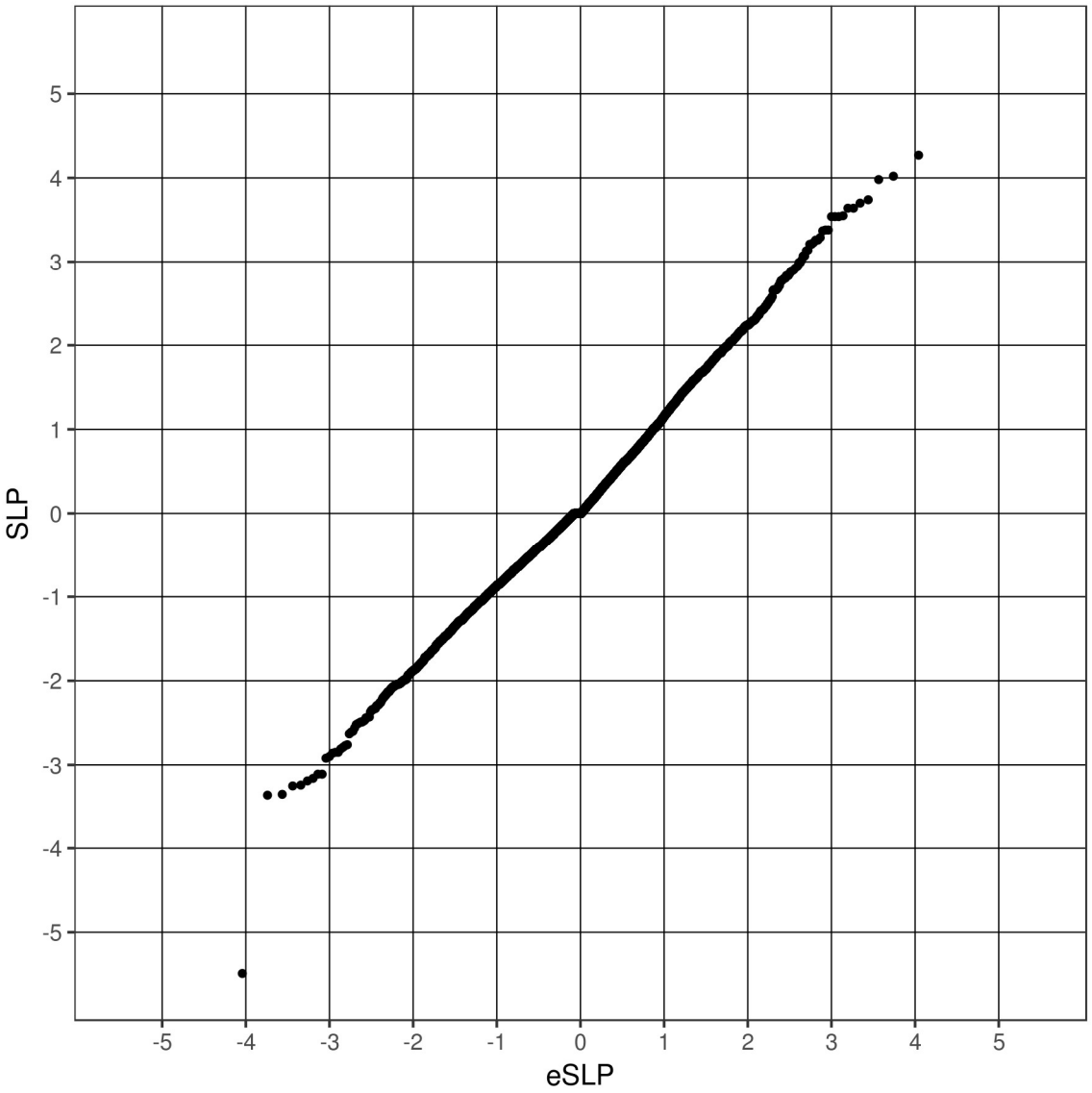
bioRxiv preprint doi: https://doi.org/10.1101/203521; this version posted November 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(b)



bioRxiv preprint doi: https://doi.org/10.1101/203521; this version posted November 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

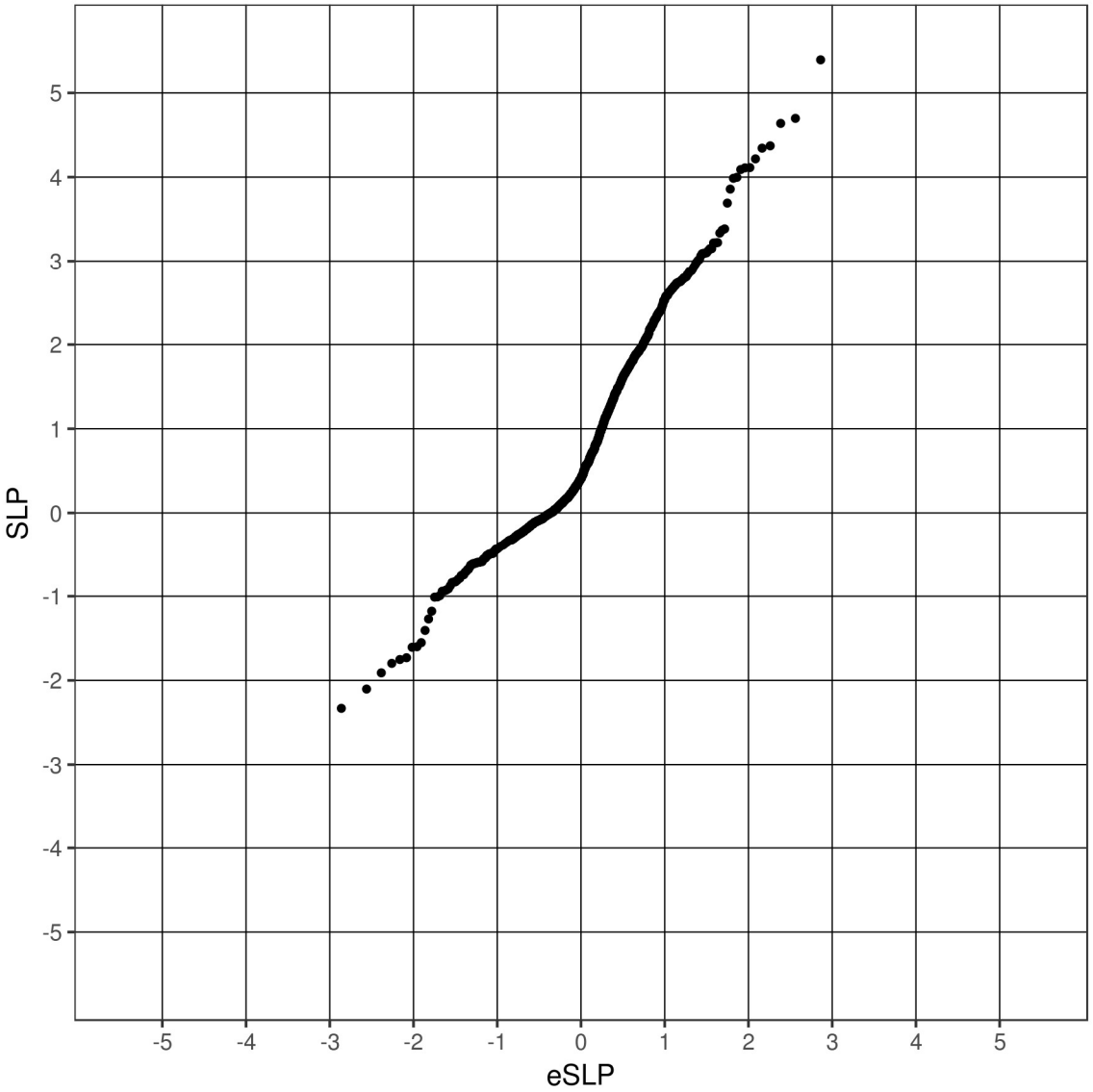
(c)



bioRxiv preprint doi: https://doi.org/10.1101/203521; this version posted November 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 2

QQ plot for set-wise SLPs for GO sets against the expected SLP if all sets were non-overlapping and independent.



bioRxiv preprint doi: https://doi.org/10.1101/203521; this version posted November 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Compliance with ethical standards

JH is supported by MRC studentship 516702. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

References

Adzhubei, I., Jordan, D.M., Sunyaev, S.R. (2013) Predicting functional effect of human missense mutations using PolyPhen-2. Curr. Protoc. Hum. Genet. 7 Unit7.20.

Ali, D.W., Salter, M.W. (2001) NMDA receptor regulation by Src kinase signalling in excitatory synaptic transmission and plasticity. Curr. Opin. Neurobiol. 11, 336–42.

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.

Blake, J.A., Eppig, J.T., Kadin, J.A., Richardson, J.E., Smith, C.L., Bult, C.J., the Mouse Genome Database Group (2017) Mouse Genome Database (MGD)-2017: community knowledge resource for the laboratory mouse. Nucleic Acids Res. 45, D723–D729.

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.

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. (2011) Assessing the Contribution Family Data Can Make to Case-Control Studies of Rare Variants. Ann Hum Genet 75, 630–638.

Curtis, D. (2012) A rapid method for combined analysis of common and rare variants at the level of a region, gene, or pathway. Adv Appl Bioinform Chem 5, 1–9.

Curtis, D. (2013) Approaches to the detection of recessive effects using next generation sequencing data from outbred populations. Adv Appl Bioinform Chem 6, 29.

Curtis, D. (2015) Investigation of Recessive Effects in Schizophrenia Using Next-Generation Exome Sequence Data. Ann Hum Genet.

Curtis, D. (2016) Pathway analysis of whole exome sequence data provides further support for the involvement of histone modification in the aetiology of schizophrenia. Psychiatr.

bioRxiv preprint doi: https://doi.org/10.1101/203521; this version posted November 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Genet. 26, 223–7.

Curtis, D. (2017) Construction of an exome-wide risk score for schizophrenia based on a weighted burden test. Ann. Hum. Genet.

Curtis, D., Emmett, W. (2017) Association study of schizophrenia with variants in miR-137 binding sites. bioRxiv.

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.

Du, C.-P., Tan, R., Hou, X.-Y. (2012) Fyn Kinases Play a Critical Role in Neuronal Apoptosis Induced by Oxygen and Glucose Deprivation or Amyloid-β Peptide Treatment. CNS Neurosci. Ther. 18, 754–761.

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.

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.

Hattori, K., Fukuzako, H., Hashiguchi, T., Hamada, S., Murata, Y., Isosaka, T., Yuasa, S., Yagi, T. (2009) Decreased expression of Fyn protein and disbalanced alternative splicing patterns in platelets from patients with schizophrenia. Psychiatry Res. 168, 119–28.

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

bioRxiv preprint doi: https://doi.org/10.1101/203521; this version posted November 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

implicates specific abnormalities of postsynaptic signalling complexes in the pathogenesis of schizophrenia. Mol. Psychiatry 17, 142–153.

Kumar, P., Henikoff, S., Ng, P.C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 4, 1073–1081.

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.

Leonenko, G., Richards, A.L., Walters, J.T., Pocklington, A., Chambert, K., Al Eissa, M.M., Sharp, S.I., O'Brien, N.L., Curtis, D., Bass, N.J., McQuillin, A., Hultman, C., Moran, J.L., McCarroll, S.A., Sklar, P., Neale, B.M., Holmans, P.A., Owen, M.J., Sullivan, P.F., O'Donovan, M.C. (2017) Mutation intolerant genes and targets of FMRP are enriched for nonsynonymous alleles in schizophrenia. Am. J. Med. Genet. Part B Neuropsychiatr. Genet.

Li, Z., Chen, J., Yu, H., He, L., Xu, Y., Zhang, D., Yi, Q., Li, C., Li, X., Shen, J., Song, Z., Ji, W., Wang, M., Zhou, J., Chen, B., Liu, Y., Wang, J., Wang, P., Yang, P., Wang, Q., Feng, G., Liu, B., Sun, W., Li, B., He, G., Li, W., Wan, C., Xu, Q., Li, W., Wen, Z., Liu, K., Huang, F., Ji, J., Ripke, S., Yue, W., Sullivan, P.F., O'Donovan, M.C., Shi, Y. (2017) Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. Nat. Genet. 49.

Mao, L.-M., Wang, J.Q. (2016a) Tyrosine phosphorylation of glutamate receptors by non-receptor tyrosine kinases: roles in depression-like behavior. Neurotransmitter (Houston, Tex.) 3.

Mao, L.-M., Wang, J.Q. (2016b) Dopamine D2 receptors are involved in the regulation of fyn and metabotropic glutamate receptor 5 phosphorylation in the rat striatum in vivo. J. Neurosci. Res. 94, 329–338.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., Cunningham, F. (2016) The Ensembl Variant Effect Predictor. Genome Biol. 17, 122.

Moeschler, J.B. (2008) Genetic Evaluation of Intellectual Disabilities. Semin. Pediatr. Neurol.

bioRxiv preprint doi: https://doi.org/10.1101/203521; this version posted November 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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.

Olde Loohuis, N.F.M., Nadif Kasri, N., Glennon, J.C., van Bokhoven, H., Hébert, S.S., Kaplan, B.B., Martens, G.J.M., Aschrafi, A. (2017) The schizophrenia risk gene MIR137 acts as a hippocampal gene network node orchestrating the expression of genes relevant to nervous system development and function. Prog. Neuro-Psychopharmacology Biol. Psychiatry 73, 109–118.

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.

Power, R.A., Kyaga, S., Uher, R., MacCabe, J.H., Långström, N., Landen, M., McGuffin, P., Lewis, C.M., Lichtenstein, P., Svensson, A.C. (2013) Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. JAMA Psychiatry 70, 22–30.

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.

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., Rüschendorf, 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.

Rees, E., Kirov, G., Walters, J.T., Richards, A.L., Howrigan, D., Kavanagh, D.H., Pocklington, A.J., Fromer, M., Ruderfer, D.M., Georgieva, L., Carrera, N., Gormley, P., Palta, P., Williams, H., Dwyer, S., Johnson, J.S., Roussos, P., Barker, D.D., Banks, E., Milanova, V., Rose, S.A., Chambert, K., Mahajan, M., Scolnick, E.M., Moran, J.L., Tsuang, M.T., Glatt, S.J., Chen, W.J., Hwu, H.-G., Taiwanese Trios Exome Sequencing Consortium, Neale, B.M., Palotie, A., Sklar, P., Purcell, S.M., McCarroll, S.A., Holmans, P., Owen, M.J., O'Donovan, M.C. (2015) Analysis of exome sequence in 604 trios for recessive genotypes in schizophrenia. Transl. Psychiatry 5, e607.

Rees, E., Walters, J.T.R., Georgieva, L., Isles, A.R., Chambert, K.D., Richards, A.L., Mahoney-Davies, G., Legge, S.E., Moran, J.L., McCarroll, S.A., O'Donovan, M.C., Owen, M.J., Kirov, G. (2014) Analysis of copy number variations at 15 schizophrenia-associated loci. Br. J. Psychiatry 204, 108–114.

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

Ruderfer, D.M., Lim, E.T., Genovese, G., Moran, J.L., Hultman, C.M., Sullivan, P.F., McCarroll, S.A., Holmans, P., Sklar, P., Purcell, S.M. (2014) No evidence for rare recessive and compound heterozygous disruptive variants in schizophrenia. Eur J Hum Genet.

bioRxiv preprint doi: https://doi.org/10.1101/203521; this version posted November 24, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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.

Sekar, A., Bialas, A.R., de Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N., Tooley, K., Presumey, J., Baum, M., Van Doren, V., Genovese, G., Rose, S.A., Handsaker, R.E., Daly, M.J., Carroll, M.C., Stevens, B., McCarroll, S.A. (2016) Schizophrenia risk from complex variation of complement component 4. Nature.

Singh, T., Kurki, M.I., Curtis, D., Purcell, S.M., Crooks, L., McRae, J., Suvisaari, J., Chheda, H., Blackwood, D., Breen, G., Pietiläinen, O., Gerety, S.S., Ayub, M., Blyth, M., Cole, T., Collier, D., Coomber, E.L., Craddock, N., Daly, M.J., Danesh, J., DiForti, M., Foster, A., Freimer, N.B., Geschwind, D., Johnstone, M., Joss, S., Kirov, G., Körkkö, J., Kuismin, O., Holmans, P., Hultman, C.M., Iyegbe, C., Lönnqvist, J., Männikkö, M., McCarroll, S.A., McGuffin, P., McIntosh, A.M., McQuillin, A., Moilanen, J.S., Moore, C., Murray, R.M., Newbury-Ecob, R., Ouwehand, W., Paunio, T., Prigmore, E., Rees, E., Roberts, D., Sambrook, J., Sklar, P., Clair, D.S., Veijola, J., Walters, J.T.R., Williams, H., Sullivan, P.F., Hurles, M.E., O'Donovan, M.C., Palotie, A., Owen, M.J., Barrett, J.C. (2016) Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. Nat. Neurosci. 19, 571–577.

Singh, T., Walters, J.T.R., Johnstone, M., Curtis, D., Suvisaari, J., Torniainen, M., Rees, E., Iyegbe, C., Blackwood, D., McIntosh, A.M., Kirov, G., Geschwind, D., Murray, R.M., Di Forti, M., Bramon, E., Gandal, M., Hultman, C.M., Sklar, P., Palotie, A., Sullivan, P.F., O'Donovan, M.C., Owen, M.J., Barrett, J.C., Owen, M.J., Barrett, J.C. (2017) The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. Nat. Genet.

Steinberg, S., Gudmundsdottir, S., Sveinbjornsson, G., Suvisaari, J., Paunio, T., Torniainen-Holm, M., Frigge, M.L., Jonsdottir, G.A., Huttenlocher, J., Arnarsdottir, S., Ingimarsson, O., Haraldsson, M., Tyrfingsson, T., Thorgeirsson, T.E., Kong, A., Norddahl, G.L., Gudbjartsson, D.F., Sigurdsson, E., Stefansson, H., Stefansson, K. (2017) Truncating mutations in RBM12 are associated with psychosis. Nat. Genet.

The Jackson Laboratory (n.d.) Mouse Genome Database (MGD) at the Mouse Genome Informatics website [WWW Document]. URL http://www.informatics.jax.org

Wagnon, 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.

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.