Table of Contents

[Consistency of Primary GWAS Results in an Independent Dataset 2](#_Toc47697861)

[Polygenicity versus other sources of Inflation 3](#_Toc47697862)

[Sex based analyses 4](#_Toc47697863)

[Autosomal 4](#_Toc47697864)

[Schizophrenia XWAS and Dosage Compensation 4](#_Toc47697865)

[Heterogeneity in SNP effects 4](#_Toc47697866)

[Meta-analysis 5](#_Toc47697867)

[X-chromosome gene inactivation status 5](#_Toc47697868)

[Results X chromosome analyses 5](#_Toc47697869)

[Conclusion 6](#_Toc47697870)

[Outcome of 128 independent associations from PGC2 SCZ study (Nature, 2014) 6](#_Toc47697871)

[Heritability, SNP-based heritability, variance explained in out of sample prediction, and variance explained by genome-wide significant SNPs. 7](#_Toc47697872)

[Conditional Analysis 11](#_Toc47697873)

[FINEMAP 12](#_Toc47697874)

[Mutation intolerance metrics in fine-mapped genes 13](#_Toc47697875)

[Summary Based Mendelian Randomisation 14](#_Toc47697876)

[Prioritising SMR Genes 16](#_Toc47697877)

[Combining eQTL and GWAS Fine-mapping 16](#_Toc47697878)

[Combining SMR, FINEMAP and chromatin conformation analysis 16](#_Toc47697879)

[Mendelian randomisation analyses of causal effects across schizophrenia and complex traits and diseases 17](#_Toc47697880)

[Potential Heterogeneity due to Sample Ascertainment 18](#_Toc47697881)

# **Consistency of Primary GWAS Results in an Independent Dataset**

In the context of the primary GWAS sample, the sample from deCODE (**Methods**) is too small to reliably test for replication at individual loci, and instead we use it to show that our SNP associations *en masse* replicate in an independent sample. We conducted two tests of replication. The first used counts of the number of independent, index SNPs whose directions of effect were the same between the discovery and replication samples as a test statistic. Under the null hypothesis of randomly oriented effects, we use the exact binomial test to obtain a one-sided p-value.

Algorithm in R:

* binom.test(NSUM - NPOS, NSUM, alternative="less")$p.value
* NSUM = number of total observations
* NPOS = number of same direction observations

For our second test, we calculated the expected number of same direction effects taking into account the discovery magnitude of effect, and the replication effect-estimate precision. For each variant we replicate, we calculate the following probability of matching effect direction:

Algorithm in R:

* pnorm(0, abs(log(OR\_disc)),sd=SE\_repl,lower.tail=F)
* OR\_disc = odds ratio in the discovery summary statistics
* SE\_repl = standard error in the combined replication summary statistics

Assuming independent associations, we sum these probabilities over the GWS autosomal index SNPs, and derive an expectation that 90.4% of SNPs should have the same direction of effect if all the discovery associations are true positives. For details please refer to **Supplementary Table 22**.

Of the index SNPs that were genome wide significant in our primary GWAS, and available from deCODE (N=289), 87% had the same direction of effect in the deCODE cohort, a level greater than expected by chance (Binomial P=6.3x10-41). This is not significantly different (chi square 1df, P=0.15) from expectation if all the discovery associations are true positives, an expectation which is inflated as it does not allow for the winner’s curse, which is the phenomenon where discovery effect sizes are overestimated (and therefore so is power to replicate) in imperfectly powered studies. At a more relaxed significance threshold (P=1x10-6 - 5x10-8) of 289 index SNPs available in the deCODE cohort, 216 showed congruent effects. Again, this is significantly different from the null (Binomial P=7.0x10-18), indicating the presence of large numbers of true associations that just fail to meet the GWS threshold.

# **Polygenicity versus other sources of Inflation**

In well-powered GWAS of a polygenic disorder where a very large number of true genetic effects exist, the test statistics are expected to be inflated over the null distribution. However, test statistic inflation can also occur as a result of confounding factors such as such as cryptic relatedness between study participants, and population stratification. Linkage Disequilibrium Score Regression (LDSR) has been widely employed to distinguish between these two broad sources of inflation1, with deviation of the intercept from 1 being indicative of residual confounding. However, elevation in the LDSR intercept can also deviate from 1 as a function of greater sample size, particularly for traits with high heritability and consequently, in large sample sizes of highly heritable traits like schizophrenia, the unadjusted LDSR intercept is not a reliable guide to confounding2,3. It has recently been shown that the attenuation ratio statistic, calculated as the value of the (LDSC(intercept) − 1)/(mean of association chi-square statistics − 1) provides a better measure of the relative effects of polygenicity versus confounding on test statistic inflation that is more robust to sample size and heritability2. In the present study, we observe a mean chi-square of 2.18, and an LDSR intercept of 1.08 (SE 0.02), giving an attenuation ratio of 0.071 (SE 0.014). This is lower than the mean attenuation ratio observed across 23 ancestry restricted GWAS of the UKBiobank2 (attenuation ratio 0.078 (SE = 0.006) of anthropometric traits (e.g. height, weight, bone density), haematological measures (e.g. platelet count, white cell count), medical disorders (e.g. type 2 diabetes, respiratory disorders, allergies) and behavioural traits (e.g. neuroticism, education attainment, smoking). It is also less than the mean attenuation ratio obtained from linear mixed model analyses of the same traits in the full European sample of UK Biobank (0.082 (SE 0.005)), implying that confounders are well adjusted for in the present study and that polygenicity contributes to more than 90% of the genome-wide inflation in test statistics.

# **Sex based analyses**

## Autosomal

We performed separate GWAS of males and females and found the genetic correlation (EUR sample alone to avoid ancestry effects) was not significantly different from 1 (Rg M:F = 0.992 (SE 0.024)). Autosomal wide case-only male-vs-female association analysis identified neither genome wide significant findings nor genomic inflation (**Supplementary Table 23**, **Supplementary Figure 6**), and specific analysis of the GWS index SNPs identified no significant evidence of heterogeneity of effect size by sex (allowing for number of tests **Supplementary Table 4**). Together, the findings show that common variant genetic liability to schizophrenia is similar (and possibly identical) in males and females.

## Schizophrenia XWAS and Dosage Compensation

For SNP discovery using XWAS, the per-allele effect size of X-chromosome SNPs was estimated in males and females separately using logistic regression, and their summary statistics meta-analysed assuming either a full dosage compensation (FDC) or no dosage compensation (NDC) model4. All analyses were performed on a set of 230,230 common variants (MAF > 0.01), using EUR samples. We estimated the across-SNP average effective sample size of males and females for the X-chromosome as defined for autosomal analyses, and used an estimate of the effective number of markers, previously defined and estimated as Meff(X) = 1300 5. Subsequently, following previously described methods4,5, we estimated the dosage compensation ratio (DCR) and its SE from the sex-specific summary statistics. Using the same data and methods, we estimated the male-female genetic correlation on the X chromosome and derived its standard errors using a block jackknife method with 1000 blocks.

Heterogeneity in SNP effectsTo test the difference in the SNP effects between sexes we apply a heterogeneity test under assumption of full dosage compentation (DC). If $\hat{β}\_{m}$ and $\hat{β}\_{f}$ are the male and female per-allele effect estimates, and $\hat{SE}(\hat{β}\_{m})$ and $\hat{SE}(\hat{β}\_{f})$ are their corresponding standard errors, then the test statistic

$$T\_{d}= \frac{\left(\frac{1}{2}\hat{β}\_{m}-\hat{β}\_{f}\right)^{2}}{\frac{1}{4}\hat{SE}\left(\hat{β}\_{m}\right)^{2}+\hat{SE}\left(\hat{β}\_{f}\right)^{2}}$$

follows a χ2-distribution with one degree of freedom under the null hypothesis of no difference in estimates under full DC assumption.

Meta-analysisThe results from the sex-stratified analysis were meta-analysed using the inverse-variance weighted method to identify the top genome-wide significant loci (analyses performed in R). The choice of optimum meta-analysis of the sex-specific results depends on the genotype coding and assumptions of DC5. That is, under a no DC model, the joint estimates will be unbiased when using per-allele effect estimates in males, while under a full DC model, they are unbiased when the effect estimates in males are from an association analysis with diploid male genotype coding.

X-chromosome gene inactivation statusFor each top SNP we determined if it is physically located within a gene to infer the presumable gene and its inactivation status according to “Reported XCI status” from 6 **(Supplementary Table 29)**.

Results X chromosome analyses The estimated across-SNP average effective sample size was 58,591 for males and 42,235 for females **(Supplementary Table 29).** From the XWAS, in the male-female meta-analysis under full DC (FDC)we identified 34 SNPs with P< 5e-8, which reduce to 4 loci after clumping. Similarly, under no DC (NDC) we identified 37 SNPs with P< 5e-8 (the same 4 loci). The top 5 SNPs for these loci are presented in **Supplementary Table 29**, where 3 loci share the same top SNPs and for one locus the top-associated SNPs are different.

The estimate of the DCR for the entire X chromosome from the European samples was 2.12 (SE 0.68), consistent with the expectation of 2.0 under full dosage compensation, albeit with a large standard error. Consistent with the results from the autosomes (Rg = 0.992, SE = 0.024), the estimate of genetic correlation between males and females was not significantly different from 1 (Rg(X) =1.00, SE=0.09), and there was no indication of significant heterogeneity for any SNP (Mean Td= 1, Max Td= 17.1, in line with expectation under a χ2-distribution with one degree of freedom).

The top SNPs from clump1 and clump3 are physically located within genes, annotated to have “Variable escape” (*NLGN4X* and *IL1RAPL1*, respectively). SNPs in clump2 are near/in gene *CNKSR2*, also annotated with “Variable escape”. Clump 4 is near/in “Inactive” *PJA1*. In total 3 out of 4 signals are near/in “Variable escape” genes, which are preferentially expressed in the brain (GTEx portal). The mean effect size ratio for the three potential escape SNPs is 1.35 (Top SNPs from FDC meta-analysis) and 1.32 (Top SNPs from NDC meta-analysis). For the SNP near the inactive gene *PJA1*, the effect size ratio is 1.80.

ConclusionWe identified 4 genome-wide significant (GWS) loci in the European samples. The top-associated SNPs for 3 out of the 4 loci are located within or near genes that have been annotated to have variable escape from X-inactivation. These three genes (*NLGN4X*, *IL1RAPL1*, *CNKSR2*) that variably escape from X-inactivation are preferentially expressed in the brain (GTEx portal). For those 3 loci, the ratio of the effect size estimates in males and females (ratio = 1.35) is consistent with partial escape from X-inactivation. However, X-chromosome wide analysis was consistent with the absence of heterogeneity between the sexes and consistent with the effect of full dosage compensation creating more genetic variance in males than in females (dosage compensation ratio of 2.12, SE 0.68, when the expectation under FDC is 2.0).

# Outcome of 128 independent associations from PGC2 SCZ study (Nature, 2014)

Details of these results are provided in **Supplementary Table 24**. We found at least one genome-wide significant SNP in the current discovery analysis within 50kb of the index SNP defining 118 of the 128 genome-wide significant LD clumps we reported in 2014 7, of which 111 were more significant than the discovery index SNP in the previous study. We did not find a genome-wide significant index SNP for 10 of the 128 GWS clumps PGC reported in its last primary study7 (**Supplementary Figure 7)**. Two of these index SNPs (**Supplementary Figure 7**; chr1:243.5Mb and chr5:152.8Mb) were subsumed in our previous study into loci that remain GWS in the present study. Of the remaining 8 index SNPs, all but 1 (rs3768644; chr2:72.3Mb) retained SNPs associated at P< 5x10-6 within 50kb of the LD (R2 > 0.1) region of the index SNP suggesting they may still be true positives (**Supplementary Figure 7**). Loss of signal at a proportion of true associated loci is expected due to sampling variation8, and in the present study, additional random effects are expected due to re-matching controls to optimize case-control balance, the use of a different imputation reference panel, and a different balance of ancestries in the study. However, the previously reported index SNP, rs3768644, with no suggestively associated SNPs in the region in the present study is most likely a false positive, notwithstanding a GWS association within 1Mb (**Methods**, **Supplementary Figure 7**).

# **Heritability, SNP-based heritability, variance explained in out of sample prediction, and variance explained by genome-wide significant SNPs.**

Heritability of SCZ is defined as the proportion of variance in liability attributable to genetic factors and is estimated from the increased risk of SCZ in relatives of those with SCZ. Compared to many other psychiatric disorders there are good data to estimate heritability for SCZ. Any parameter estimated on the liability scale requires a scaling of the estimate which is based on measurements of case/control status, and the scaling requires an estimate of the lifetime risk of disease. The best estimates of lifetime risk of SCZ come in at less than 1% (0.7% in Saha *et al*9). The most commonly reported estimates of heritability of SCZ are from meta-analysis of relatively small studies10 at 81% (95%CI 73-90%) or from Swedish national records11 of 64% (95% CI 62-68%). Since, there are inherent assumptions in applying models to data, it is prudent to use approximate benchmark values, which we take to be lifetime risk of 1% and heritability ($h^{2}$) of 70%.

From GWAS data we can estimate the proportion of variance in liability associated with common SNPs, the so-called SNP-based heritability ($h\_{SNP}^{2}$). This is expected to be lower than $h^{2}$ because it only captures the variance associated with common SNPs measured; the correlation between the measured SNPs and causal variants (particularly those that are uncommon in the population) can be low. $h\_{SNP}^{2}$ has been estimated from SCZ GWAS data sets as they have become available. Even in the first analyses using PGC1 data12 it was noted that the estimates were higher from individual cohorts than when cohorts were combined ((International Schziophrenia Consortium (ISC): 0.27 (s.e. 0.02), Molecular Genetics of Schizophrenia (MGS): 0.31 (0.03), ISC+MGS: 0.25 (0.01); all other PGC1 cohorts combined; 0.27 (0.02), all PGC1 together : 0.23 (0.01)), implying real (e.g. population specific) or technical (e.g., genotyping array) contributions to these estimates.

Since 2014 it has become common to estimate SNP-based heritability from GWAS summary statistics using LDscore regression (LDSR)1,13, e.g. in LDHub14, because it is very quick to apply, even though the original authors did not recommend this use. In the present study (PGG3), the LDSR estimates for EUR are $h\_{SNP}^{2}$ = 0.207 (SE 0.008) assuming lifetime risk of 1%, and $h\_{SNP}^{2}=$0.191 (SE 0.007) for lifetime risk of 0.7% (demonstrating the small impact of the lifetime risk estimate in this range). However, LDSR has been shown13,15 to underestimate $h\_{SNP}^{2}$ . Despite biases in LDSR for $h\_{SNP}^{2}$ estimates, LDSR estimates of genetic correlation are robust, as are $h\_{SNP}^{2}$ enrichment analyses for genomic annotations e.g. exonic sequence16 or cell-type-enriched expression17. More recent methods of estimating $h\_{SNP}^{2}$ from GWAS summary statistics do not provide downward biased estimates using simulated data18,19, and provide higher estimates when applied to real data20. All methods applied to GWAS summary statistics require knowledge of the correlation structure between SNPs and so use LD reference samples to provide this information. In simulated data sets it can be shown that $h\_{SNP}^{2}$ estimates are less biased and more accurate (lower s.e.) when LD is calculated from the data that are used for calculating the GWAS summary statistics. In applications to real data, such as the PGC data in the present study, where the GWAS is a meta-analysis of many cohorts, the results seem relatively robust to the choice of LD reference. One method to estimate $h\_{SNP}^{2}$ from GWAS summary statistics is SBayesS19; when applied to PGC-SCZ data, using the GERA21 LD reference, the estimates are PGC2 7 0.21 (s.e. 0.006), PGC3 (EUR) 0.24 (0.007), PGC3 (All ancestries) 0.24 (0.007).

SBayesS also estimates other genetic architecture parameters: polygenicity (𝜋) and selection (*S*) parameters, where 𝜋is the proportion of (HapMap3) SNPs estimated to be causal and *S* describes the effect size-MAF relationship. Using PGC3 the estimate of 𝜋 is 4% and an *S* value of -0.6. These values are best interpreted relative to the estimates for other traits/diseases/disorders19. Relative to other traits, 𝜋 of 4% is high, while S of -0.6 is similar to other traits.

While $h\_{SNP}^{2}$ measures the total variance associated with common SNPs, it does not reflect the variance attributable to the specific associations that have been identified. While the $h\_{SNP}^{2}$ is not expected to change with sample size (given the same SNP set), the number – and hence variance attributed to – specific identified associations is expected to increase with increasing sample size. One way to assess this is by considering the variance explained by genome-wide significant (GWS) SNPs. The variance explained by an individual associated locus can be approximated by using the effect size estimate and the allele frequency, and converting to the liability scale (for example, implemented in INDI-V22). For PGC1+SWE23, PGC2, PGC3 the number of LD independent (r2<0.1) GWS associations is 24, 108, 300 and the total variance from summing the estimates of variance from each SNP are: 1.3%, 3.4%, 7.2% representing an increase in PGC2 and PGC3 over PGC1+ of 2.7 and 5.6-fold. It is well-recognised that these estimates of variance explained will be biased upwards for two reasons. First, the effect sizes are estimated from within the data are biased upwards by winner’s curse. Second, although chosen to be approximately independent by local LD, they are unlikely to be independent if fitted jointly in a model, and this becomes more problematic as the number of GWS associations increase (although we have tried to avoid this by using only SNPs that are independent in stepwise regression). Hence, we regard this current estimate of 7.2% as inaccurate, inflated, but include it here to link to estimates reported in previous publications, and for transparency.

The most robust estimate of the variance explained by GWAS associated SNPs is to use out-of-sample prediction, estimating a polygenic risk score (PRS) as the weighted sum of risk alleles calculated in a (target) sample with known case/control status but independent of the GWAS (discovery) sample (**Methods**). There are now several methods available to calculate PRS and the methods differ in two key criteria: which SNPs (or other associated variants) to include in the score and what weights or effect sizes to allocate to them. Here, we use the basic p-value thresholding method (P+T) which had its first application to SCZ data (ISC cohort24) and has been used in PGC1+SWE and PGC2 publications. Briefly, the full SNP set (MAF > 0.10) is clumped (i.e., retains the most associated SNP in a region, but removes any SNP in LD r2 >0.1 with it), then the SNPs used to generate the PRS are selected on association p-value (PT threshold). In the ISC, PGC1+SWE and PGC2 publications the MGS cohort (2,687 cases and 2,656 controls) was removed from the discovery GWAS and the variance explained in case-control status. We then reported Nagelkerke’s R2. ISC: 3.2%, PGC1+SWE: 6%, PGC2: 18.4%, PGC3 (ALL): 22.1% as the maximum variance explained (which reflected p-value thresholds of ISC:0.5, PGC1+SWE 0.10, PGC2: 0.05, PGC3: 0.05). These values acknowledge that there is predictive association information in associations that do not pass the genome-wide significance threshold, and that PRS prediction is robust to the inclusion of some false positive associations. Reporting Nagelkerke’s R2 in this context can be informative because the comparison is made in a single cohort (MGS) to evaluate the effect of increasing power of the increasing sample size of discovery GWAS. However, when comparisons are made across different target cohorts the Nagelkerke’s R2 can be difficult to interpret because Nagelkerke’s R2 increases as the proportion of cases included in the sample increases to 50% (then decreases again). Alternative evaluation statistics are the AUC statistic which can be interpreted as the probability that a case ranks higher than a control when a randomly selected case and control are compared on their PRS, and hence is not dependent on proportion of cases in the sample. Again, benchmarking against the MGS, we obtain (PGC2: 0.72, PGC3: 0.75). To allow comparison with statistics such as $h^{2}$ and $h\_{SNP}^{2} $between different target samples, it is also useful to present result as variance explained by PRS on the liability scale, notwithstanding the need to assume a lifetime risk of SCZ to make the transformation (here we use 1%). On the liability scale the maximum variance explained for MGS is: PGC2: 8.4%, PGC3: 10.4%. It is noteworthy that AUC here is calculated without including ancestry principal components in the model, which are included in the models used to calculate all other statistics. However, the impact is likely small; converting the liability scale variance to AUC using normal distribution theory generates an AUC for MGS of PGC2: 0.71, PGC3:0.74.

We also conducted leave-one-sample out PRS analyses, in which each of 89 cohorts is removed from the meta-analysis in turn and PRS calculated in the left-out-cohort (**Methods**). Across all cohorts, the median p-value threshold that maximised the out-of-sample prediction in the left-out cohort was pT=0.05. At this threshold, the median variance in liability explained is 7.7%, while genome-wide significant SNPs explain 2.6%. We note that the median variance explained in the PGC2 cohorts from the PGC2 discovery sample is 7.0% 7, and the median variance explained in left-out PGC2 cohorts from the PGC3 discovery sample (that includes PGC2) is 8.6%. The lower median value across the full PGC3 cohorts results mostly from the inclusion of the ASN cohorts which have lower variance explained (**Extended Data Figure 5**), likely arising due to the predominance of European ancestry cohorts in our GWAS. The mean liability variance explained in the EUR ancestry cohorts (i.e. PGC3 excluding ASN cohorts) is 8.1% (pT=0.05).

The relationship between liability to schizophreia and PRS can also be expressed in terms of an odds ratio (OR), recognising that the magnitude of OR depends also on lifetime risk of the disorder. In the European cohorts, the top centile of PRS is associated with an OR for schizophrenia relative to the rest of the sample of 7.0 (CI 5.8-8.3) (**Supplementary Table 5).** This is larger than reported for a number of other common disorders25, but is insufficient for a diagnostic tool in general populations given the low lifetime risk for the disorder. However, PRS can be valuable in a research setting, e.g., identifying brain imaging risk correlates26. As the relevant research methods often depend on biomarkers derived from the application of expensive and time consuming technologies that are difficult to measure in the population as a whole, studies often depend on sampling from the extremes of liability in the population. In that context, we note that the OR for schizophrenia between the lowest and highest centiles across EUR and ASN datasets is considerable (OR top v bottom centile = 45 (95% CI=34-58). See also **Supplementary Table 5**.

Full details of all out-of-sample predictions are available (**Supplementary Table 25 and Supplementary Table 26)** and summarized for the full sample in **Extended Data Figure 5**).

# **Conditional Analysis**

We defined 270 distinct genomic regions (5 on the X-chromosome) with at least one genome-wide significant signal, after merging overlapping LD-clumps, defined using an LD-R2 > 0.1 in the imputation reference data. Since the LD information in the reference panel incompletely reflects the true genotypic correlations between SNPs in our study data, we performed a stepwise conditional approach. For that we repeated association testing and meta-analysis (see section “Association / Meta Analysis”) in the whole genomic region, adding the allele dosages of the most significant SNP (the index SNP) as a covariate. If the most-associated SNP in the resulting meta-analysis showed an association p-value of less than 1x10-6, we repeated the analysis, fitting the second best SNP as an additional covariate. We repeated this process until no SNP in the region achieved p<1x10-6. Out of the 270 regions, 247 regions had a single signal (5 on the X-chromosome), 18 regions showed two independent signals, 3 regions showed three independent signals and 2 regions showed four independent signals. No regions harboured more than 4 independent signals with the above definition, resulting in 300 conditionally independent signals. The results are presented in **Supplementary Table 10** and **Supplementary Figure 8a** shows region plots for the stepwise conditioning in the order of original significance.

We also searched for long range dependencies. Here we tested the independent autosomal signals for conditional independence, testing all pairs. We defined residual dependency as a loss of signal of more than two orders of magnitude when one region is conditioned against another. In two instances we found partial dependency across genomic regions, one on chr12:122.8-123.1Mb and the other spanning the centromere of chromosome 5 (chr5:46-50Mb). In **Supplementary Table 27** we show all tested pairs of conditionally independent SNPs within a long-range conditional test (N= 295\*294 = 86,730 autosomal pairs). In **Supplementary Figure 3** we present two histograms of all p-value changes separately for (i) within chromosomes (Type 1+2) and (ii) across chromosomes (Type 3), each with the full y-axis range and a zoomed version to demonstrate single events. With the cross-chromosomal distribution we demonstrate that a p-value change of more than two orders of magnitude is not observed, confirming the expectation of no cross-chromosome dependency.

On the X-chromosome, we did not observe any secondary signal, or long range dependency across regions (**Supplementary Figure 8b**).

# FINEMAP

While accurately fine-mapping results of a meta-analysis is possible using summary statistics, it has been shown that mismatches in LD patterns between the GWAS data and the LD reference panel can result in false positive causal SNPs being inferred, a situation which cannot be solved by shrinkage methods as previously proposed27. Given such mismatches are likely for a large multi-ancestry sample such as ours, we did not employ a reference panel for our analysis. We instead calculated LD matrices for each locus within each sample by applying LD-Store v1.127 to the same allelic dosage data employed for the GWAS. Individual matrices were then combined at each region by calculating a weighted average based on their effective sample size28. To avoid errors in the resulting LD structure caused by SNPs with the same position being improperly merged, tri- and tetra-allelic SNPs were discarded in this stage. FINEMAP v1.429 was then applied to resolve each individual region, allowing for up to maximum of k=5 causal SNPs. For each SNP, posterior probabilities were estimated taking into account all potential causal configurations. For each region and most probable k model, 95% credible sets of causal SNPs were computed using the method implemented within the software. Expected numbers of causal SNPs (K) per region were derived from the FINEMAP output, specifically from the model-based probabilities of each evaluated number of causal (k) SNPs as follows:

$$K=\sum\_{k=1}^{k\_{max}}\left(k×P\left(k\right)\right)$$

Where k = 1, … ,kmax ; kmax is the largest number of causal SNPs evaluated for the clump and P(k) is the probability that the clump contains exactly k causal SNPs.

Models with large numbers of expected causal SNPs might reflect multiple causal associations, but they can also be artefacts of other features of the data including complex LD patterns27, differences in sample sizes for each SNP30, low power to discern independent association signals31 and associations driven by complex genetic variation (e.g. repetitive or structural variation)32.Thus, for downstream analysis, we focussed on 217 regions with low numbers of causal SNPs (K<3.5). Of these, 73% (N=158) had the same expected number of causal SNPs as there were conditional associations. For the other 27%, FINEMAP detected more complex association patterns with 1 (N=52) or 2 (N=4) extra signals; though simpler patterns with 1 less signal (N=3) were also inferred.

As a check of the assigned posterior probabilities for SNPs, we used a cross-ancestry method of fine-mapping as described33 and obtained highly consistent estimates of posterior probability (adjusted r2=0.83 for SNPs having the best probability in each locus) with those from our primary analysis (**Supplementary Figure 9**), providing additional confidence in the mapping.

# Mutation intolerance metrics in fine-mapped genes

We extracted all genes tagged by FINEMAP from the Ensembl VEP annotation, which expands the GENCODE boundaries by 5kb to account for upstream/downstream flanking regulatory regions. In order to assess whether genes tagged by FINEMAP credible sets displayed greater signatures of mutation intolerance (the upper boundary of the gnomAD O/E LoF statistic, “LOEUF”) than others, we used two approaches. For quantitative testing, since the LOEUF distribution is heavily skewed34, we used the “augmented” Mann-Whitney test proposed by Vermeulen and colleagues35. This test can be adjusted for covariates while retaining power and error rate. For binary testing, we used logistic regression and a definition of loss-of-function intolerant gene consistent with previous research (LOEUF≤0.35). A gene length covariate (in kilobases) was used in all analyses with either method. FINEMAP genes were more mutation intolerant (i.e. had a *lower* median loss-of-function (LOEUF) metric34 than the remaining protein coding genes at the loci (**Supplementary Table 14a)**. Specifically 33.1% of genes tagged by a FINEMAP credible SNP (k<3.5) were mutation intolerant compared with 15.8% of genes not tagged by FINEMAP within the genome-wide significant locus boundaries (OR for mutation intolerance = 2.22; p=7.9x10-8). In addition, 44.6% of the FINEMAP tagged genes we define as *prioritised* were mutation intolerant compared with 19.5% of the non-prioritised FINEMAP tagged genes (OR=2.65, p=5.3x10-4).

In order to ascertain the relationship between the FINEMAP posterior probability (defined for each SNP) and the LOEUF metric (defined for each gene), we calculated the cumulative posterior probability of all genes contained within the FINEMAP credible set (K<3.5), defined as the sum of the probabilities of every SNP within their boundaries. Due to the skewness of the LOEUF metric, we used a regression model with a gamma link function to estimate the association between LOEUF and cumulative posterior probability, using gene length and expected number of causal SNPs within each clump (K) as covariates. Within the set of FINEMAP genes, the proportion of the posterior probability captured by SNPs increased as a function of LOEUF **(Supplementary Table 14b)**.

# **Summary Based Mendelian Randomisation**

In O’Brien *et al* 36 the eQTL data set used gene expression levels measured by RNA-Seq in the human fetal brain (*n* = 120). It consists of ~5.8 million SNPs for 28,875 genes. The PsychENCODE adult brain data set37 was an eQTL data set with gene expression levels measured by RNA-Seq in tissue from predominantly the dorsolateral prefrontal cortex, a meta-analysis of 3 data sets. This data set consists of ~2.2 million SNPs for 24,560 genes. The eQTLGen data set38 was from a meta-analysis of *cis*-eQTL data sets, with gene expression levels measured by microarray in peripheral blood. It consists of ~10.8 million SNPs for 19,250 genes. *Cis*-eQTL effects were in standard deviation (SD) units of expression levels.

Full details of all SMR Results are given in **Supplementary Table 17**. In fetal brain36, the small sample size and therefore low power to discover eQTLs (N=120) meant that only 754 genes had significant (PeQTL<5.0x10-8 ) eQTLs for SMR testing **(Methods)** but nevertheless, fetal brain is of interest as gene expression at this time may be of high relevance to a disorder with a neurodevelopmental component. Using this eQTL dataset, we identified 21 genes with significant SMR associations (significance for 754 tests; PSMR<6.6x10-5), and none were rejected by the HEIDI test (**Supplementary Table 18a)**. The larger (N~1,500) PsychENCODE37 data set had 10,890 genes with significant eQTLs associated with gene expression in dorsolateral prefrontal cortex of the brain, excluding the extended MHC region. From these we identified 81 significant SMR associations (PSMR<4.6 x10-6) that were not rejected by the HEIDI test (**Supplementary Table 18b)**. Recognising that many eQTLs are shared across tissues39, we next exploited the power of eQTL data from whole blood38 (N ~32k). To retain relevance to brain, we used only eQTLs from the 7,803 genes that had eQTLs in both brain and blood. Fifty five of these genes had significant SMR associations (PSMR<6.4 x10-6) and were not rejected by HEIDI; 33 of these were significantly (PSMR<9.1 x10-4) associated with the same direction of effect in the SMR analysis of brain (**Supplementary Table 18c).** SMR locus plots are provided (**Supplementary Figure 10ab;** exemplar SMR plot for the *SETD6* locus, **Supplementary Figure 10b;** exemplar conditional analysis using GCTA-COJO, **Supplementary Figure 10c).** In total, there were 116 unique genes identified through the SMR & HEIDI analyses above **(Supplementary Table 18d)**.

To test if our results were robust to methodology, we applied FUSION40 and EpiXcan41 to the same brain eQTL datasets36,37 we used for SMR and obtained excellent cross method consistency (**Methods and Supplementary Table 18e).** Of the 101 SMR associations from adult brain tissue (including those identified first in blood), 100 met the inclusion criteria for EpiXcan and/or FUSION (PsychENCODE and fetal tissue analyses) of which 99 were genome-wide significantly associated (or for those detected first in blood, significant corrected for multiple testing), all with the same direction of effect. Of 21 SMR significant SMR genes in fetal brain, 19 of 20 that were also testable in fetal brain with FUSION were genome-wide significantly associated.

Moreover, we found strong consistency of effects across development; 17 of the SMR genes from fetal brain were testable in adult brain by FUSION or EpiXcan of which 16 were significant (corrected for 34 tests) in both methods. All significant associations were in the same direction using adult or fetal brain eQTLs **(Supplementary Table 18e)**, with the notable exception of *ABCB9*, where our analyses predict schizophrenia is associated with higher expression in adult brain, but lower expression in fetal brain.

# **Prioritising SMR Genes**

## Combining eQTL and GWAS Fine-mapping

It has been shown42 that most causal variants are physically close to, and in strong LD with, the top associated GWAS signals, and that causal and the top associated variants typically have small minor allele frequency (MAF) differences. It has also been shown that the mapping precision of GWAS increases with increased association test-statistic. If a gene is associated with schizophrenia through the same causal variant, then the mapping precision of the top eQTL for the gene site is generally expected to be higher than that of the GWAS top SNP because the test-statistic of the former is often larger than the latter. For each schizophrenia-associated gene identified from the SMR and HEIDI analysis above, we used the top *cis*-eQTL as the focal variant and selected sequence variants from the 1000 Genomes Project (European sample) as a credible set of causal variants using criteria which have been reported to capture around 75% of the causal variants underlying index associations detected using GWAS arrays42, physical distance < 100Kb, MAF difference < 0.05, and LD > 0.8. We then used the FINEMAP data to sum the posterior probabilities of all these SNPs to estimate the posterior probability that these candidate causal eSNPs includes a variant that is also causal for disorder.

## Combining SMR, FINEMAP and chromatin conformation analysis

Previous studies have demonstrated the value of credible SNP annotation using chromatin interactome data to prioritise candidate genes implicated by GWAS loci43. Informed by previous findings demonstrating a positive correlation between Hi-C chromatin interacting SNPs and eQTLs for the same gene43we chose to use Hi-C to add weight to the confidence of the SMR results and thus prioritise SMR genes on this basis. Indeed previous reports suggest that such SNPs (that are in Hi-C contacts and eQTLs) show stronger associations than QTLs in promoters or exons (as indicated by the QTL *P* value)43. Thus we sought to identify SMR genes that were supported by having high confidence Hi-C contacts from credible SNPs (K ≤ 3.5) to the promoter of the same gene (**Supplementary Table 19**).

The Hi-C dataset we employed to annotate our credible SNPs44 has shown highly consistent results against other brain Hi-C datasets but includes data of a greater depth. This dataset and chromatin interactome map was derived using “easy Hi-C” (eHi-C)45 of post-mortem human brain (*N*=3 adult temporal cortex and *N*=3 fetal cortex) to generate 1.32 billion high-confidence regulatory chromatin interactions defined as those interactions with *P*<2.31x10-11 (Bonferroni correction of 0.001 for 43,222,677 possible interactions) that intersected open chromatin, active histone marks, or brain-expressed transcription start sites consistent with enhancer-promoter (E-P) or promoter-promoter (P-P) interactions. E-P and P-P interacations were identified using cortical functional data from the same developmental stage; detailed Methods available in Giusti-Rodríguez *et al*44. We identified 10kb resolution E-P and P-P contacts which ranged between 20kb and 2Mb apart.

We prioritised all genes identified both by Hi-C and SMR regardless of developmental stage, given the small sample size and relative low power of the fetal brain expression dataset. We would stress that we consider this methodologically independent evidence from two functional annotation sources to be sufficient for prioritisation purposes of SMR-identified genes, but that this approach should not be used to relegate or rule out other SMR genes, since there are technical differences in the genomic distance ranges covered by Hi-C and eQTL methods and hence not all SMR-identified SNP-genes pairs are covered by the Hi-C data. Finally for the Hi-C SNP and SMR gene interactions we also calculated the cumulative posterior probability, and the proportion of total posterior probability in the clump, of the credible SNPs in the Hi-C anchor (10kb resolution) which interacts with the promoter of the SMR gene for both adult and fetal derived data (**Supplementary Table 19**).

# **Mendelian randomisation analyses of causal effects across schizophrenia and complex traits and diseases**

Mendelian randomisation allows potential causal relationships between traits to be tested using summary level genomic data, and in particular, well-powered but non-significant results can usefully reject hypotheses with confidence, although results consistent with causal relationships require evaluation using alternative methods to ensure they are not the result of bias and violation of the assumptions inherent in the method.

Mendelian randomisation (MR) provides estimates of causal effects from one trait/disease (exposure) to another (outcome) by leveraging genetic instrumental variables (IV)46,47. In 2-sample MR, the causal effect estimate is obtained through estimating the change of the IV’s effect on the outcome in proportion to the IV’s effect on the exposure. Two-sample MR leverages summary statistics from independent GWAS of the exposure and the outcome, which are often publicly available. Two-sample MR relies on 3 assumptions: 1) the genetic IV is associated with the exposure; 2) the genetic IV is independent of all confounders of the exposure-outcome relationship; and 3) the genetic IV is independent of the outcome conditional on the exposure and all confounders of the exposure-outcome association (i.e. exclusion restriction criterion or horizontal pleiotropy). Violation these assumptions will resulted in biased causal effect estimates; however, existing MR methods can control for or mitigate biases from violation of assumption48–51.

We identified 58 GWAS appropriate for bidirectional 2-sample Mendelian Randomization (MR)48–52 with schizophrenia both as the exposure and outcome (**Methods** and **Supplementary Table 6**).

# **Potential Heterogeneity due to Sample Ascertainment**

To explore the observed differences between cohorts in the predictive ability of PRS, we performed a series of meta-analyses within subgroups. Based on the sample descriptions, cohorts were organised into subgroups according to 4 criteria: 1) Case definition: Cases defined as (i) schizophrenia, (ii) schizophrenia or schizoaffective disorder, (iii) schizophrenia spectrum disorder (including non-affective psychosis). 2) Controls definition: Whether controls were (i) screened or (ii) unscreened for schizophrenia or other psychoses. 3) Recruitment setting: Whether sample collection was performed in (i) hospital/inpatient setting, including treatment-resistant cases treated with the antipsychotic clozapine, (ii) community or outpatient clinics, and (iii) mixed including inpatient and outpatient recruitment. 4) Diagnostic strategy: Whether final diagnosis of cases was ascertained through (i) consensus between psychiatrists according to DSM or ICD criteria, (ii) diagnostic interviews (including SCID53, SCAN54, MINI55, CASH56, structured psychiatric assessment), (iii) review of medical records or hospital registers, and (iv) a mixed strategy using a combination of the previous methods. Cohorts with missing or inconclusive information were excluded from the relevant subgroup analysis.

Analyses were restricted to cohorts of European ancestry, given the impact of ancestry on PRS. For each sample, we calculated the variance explained on the liability scale using PT = 0.05. Given the considerable heterogeneity of the estimates as measured with the I2 statistic57 (all I2 > 75%), meta-analyses across the samples with the relevant defining characteristics were performed with the Der Simonian and Laird random effects model58. We compared the pooled variance explained and its confidence intervals for each subgroup with the pooled estimate of all the remaining cohorts.

Algorithm in R:

SE <- sqrt(SEg^2+SEo^2)

z <- (Mg-Mo)/SE

P.z <- pnorm(z, lower.tail = TRUE)

# Mg, SEg the mean and standard error of the pooled estimate for the subgroup

# Mo, SEo the mean and standard error of the pooled estimate for all the other cohorts, excluding the tested subgroup

PRS had increased predictive ability in samples which by ascertainment are likely to be enriched for the most severe cases, i.e. hospitalized patients including those treated with clozapine (liability scale variance explained 0.10, 95% CI 0.09-0.11; p = 0.003 for the difference from remaining cohorts). Pooled effects of subgroups clustered by case definition, screening of controls, or diagnostic strategy did not differ from the overall mean (**Supplementary Figure 4**).

References

1. Bulik-Sullivan, B. *et al.* LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

2. Loh, P. R., Kichaev, G., Gazal, S., Schoech, A. P. & Price, A. L. Mixed-model association for biobank-scale datasets. *Nature Genetics* **50**, 906–908 (2018).

3. Yengo, L., Yang, J. & Visscher, P. M. Expectation of the intercept from bivariate LD score regression in the presence of population stratification. *bioRxiv* 310565 (2018). doi:10.1101/310565

4. Sidorenko, J. *et al.* The effect of X-linked dosage compensation on complex trait variation. *Nat. Commun.* **10**, 3009 (2019).

5. Lee, J. J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet.* **50**, 1112–1121 (2018).

6. Tukiainen, T. *et al.* Landscape of X chromosome inactivation across human tissues. *Nature* **550**, 244–248 (2017).

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

8. Hamshere, M. L. *et al.* Genome-wide significant associations in schizophrenia to ITIH3/4, CACNA1C and SDCCAG8, and extensive replication of associations reported by the Schizophrenia PGC. *Mol. Psychiatry* **18**, 708–712 (2013).

9. Saha, S., Chant, D., Welham, J. & McGrath, J. A systematic review of the prevalence of schizophrenia. *PLoS Medicine* **2**, 0413–0433 (2005).

10. Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a Complex Trait: Evidence from a Meta-analysis of Twin Studies. *Arch. Gen. Psychiatry* **60**, 1187–1192 (2003).

11. Lichtenstein, P. *et al.* Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* **373**, 234–239 (2009).

12. The Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nat. Genet.* **43**, 969–978 (2011).

13. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).

14. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).

15. Evans, L. M. *et al.* Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nat. Genet.* **50**, 737–745 (2018).

16. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

17. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).

18. Ge, T., Chen, C. Y., Ni, Y., Feng, Y. C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, (2019).

19. Zeng, J. *et al.* Bayesian analysis of GWAS summary data reveals differential signatures of natural selection across human complex traits and functional genomic categories. *bioRxiv* 752527 (2019). doi:10.1101/752527

20. Revez, J. A. *et al.* Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin D concentration. *Nat. Commun.* **11**, 1–12 (2020).

21. Banda, Y. *et al.* Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort. *Genetics* **200**, 1285–1295 (2015).

22. Witte, J. S., Visscher, P. M. & Wray, N. R. The contribution of genetic variants to disease depends on the ruler. *Nat. Rev. Genet.* **15**, 765–776 (2014).

23. Ripke, S. *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat. Genet.* **45**, 1150–1159 (2013).

24. International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).

25. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature Genetics* **50**, 1219–1224 (2018).

26. Lancaster, T. M. *et al.* Structural and Functional Neuroimaging of Polygenic Risk for Schizophrenia: A Recall-by-Genotype-Based Approach. *Schizophr. Bull.* **45**, 405–414 (2019).

27. Benner, C. *et al.* Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-wide Association Studies. *Am. J. Hum. Genet.* **101**, 539–551 (2017).

28. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

29. Benner, C. *et al.* FINEMAP: Efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32**, 1493–1501 (2016).

30. Dadaev, T. *et al.* Fine-mapping of prostate cancer susceptibility loci in a large meta-analysis identifies candidate causal variants. *Nat. Commun.* **9**, 1–19 (2018).

31. Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nature Reviews Genetics* **19**, 491–504 (2018).

32. Chiang, C. *et al.* The impact of structural variation on human gene expression. *Nat. Genet.* **49**, 692–699 (2017).

33. Lam, M. *et al.* Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat. Genet.* (2019). doi:10.1038/s41588-019-0512-x

34. Karczewski, K. J. *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* 531210 (2019). doi:10.1101/531210

35. Vermeulen, K., Thas, O. & Vansteelandt, S. Increasing the power of the Mann-Whitney test in randomized experiments through flexible covariate adjustment. *Stat. Med.* **34**, 1012–1030 (2015).

36. O’Brien, H. E. *et al.* Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. *Genome Biol.* **19**, 194 (2018).

37. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science (80-. ).* **362**, (2018).

38. Võsa, U. *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv* 447367 (2018). doi:10.1101/447367

39. Qi, T. *et al.* Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun.* **9**, (2018).

40. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).

41. Zhang, W. *et al.* Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits. *Nat. Commun.* **10**, (2019).

42. Wu, Y., Zheng, Z., Visscher, P. M. & Yang, J. Quantifying the mapping precision of genome-wide association studies using whole-genome sequencing data. *Genome Biol.* **18**, (2017).

43. Wang, D. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science (80-. ).* **362**, eaat8464 (2018).

44. Giusti-Rodríguez, P. *et al.* Schizophrenia and a high-resolution map of the three-dimensional chromatin interactome of adult and fetal cortex. *bioRxiv* 406330 (2018). doi:10.1101/406330

45. Lu, L., Liu, X., Peng, J., Li, Y. & Jin, F. Easy Hi-C: A simple efficient protocol for 3D genome mapping in small cell populations. *bioRxiv* 245688 (2018). doi:10.1101/245688

46. Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Smith, G. D. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27**, 1133–1163 (2008).

47. Pingault, J.-B. *et al.* Using genetic data to strengthen causal inference in observational research. *Nat. Rev. Genet.* **19**, 566–580 (2018).

48. Bowden, J., Smith, G. D. & Burgess, S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).

49. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).

50. Bowden, J. *et al.* Improving the accuracy of two-sample summary-data Mendelian randomization: moving beyond the NOME assumption. *Int. J. Epidemiol.* **48**, 728–742 (2019).

51. Verbanck, M., Chen, C. Y., Neale, B. & Do, R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat. Genet.* **50**, 693–698 (2018).

52. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7**, (2018).

53. First, M. B., Spitzer, R. L., Gibbon, M. & Williams, J. B. W. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition. (SCID-I/P)*. (Biometrics Research, New York State Psychiatric Institute, 2002).

54. World Health Organization. Division of Mental Health. *Schedules for clinical assessment in neuropsychiatry (Version 2.1). 2.1. edn*. (WHO Assessment, Classification and Epidemiology, 1999).

55. Sheehan, D. V. *et al.* The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10 - PubMed. *J. Clin. Psychiatry* **59**, 22–33 (1998).

56. Andreasen, N. C., Flaum, M. & Arndt, S. The Comprehensive Assessment of Symptoms and History (CASH): An Instrument for Assessing Diagnosis and Psychopathology. *Arch. Gen. Psychiatry* **49**, 615–623 (1992).

57. Higgins, J. P. T., Thompson, S. G., Deeks, J. J. & Altman, D. G. Measuring inconsistency in meta-analyses. *British Medical Journal* **327**, 557–560 (2003).

58. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control. Clin. Trials* **7**, 177–188 (1986).