Trubetskoy, Vassily, Pardinas, Antonio , Qi, Ting, Panagiotaropoulou, Georgia, Awasthi, Swapnil, Bigdeli, Tim B., Bryois, Julien, Chen, Chia-Yen, Dennison, Charlotte A. , Hall, Lynsey, Lam, Max, Watanabe, Kyoto, Frei, Oleksandr, Ge, Tian, Harwood, Janet C., Koopmans, Frank, Magnusson, Sigurdur, Richards, Alexander L., Sidorenko, Julia, Wu, Yang, Zeng, Jian, Grove, Jakob, Kim, Minsoo, Li, Zhiqiang, Voloudakis, Georgios, Zhang, Wen, Adams, Mark, Agartz, Ingrid, Atkinson, Elizabeth G., Agerbo, Esben, Al Eissa, Mariam, Albus, Margot, Alexander, Madeline, Alizadeh, Behrooz Z., Alptekin, Köksal, Als, Thomas D., Amin, Farooq, Arolt, Volker, Arrojo, Manuel, Athanasiu, Lavinia, Azevedo, Maria Helena, Bacanu, Silviu A., Bass, Nicholas J., Begemann, Martin, Belliveau, Richard A., Bene, Judit, Benyamin, Beben, Bergen, Sarah E., Blasi, Giuseppe, Bobes, Julio, Bonassi, Stefano, Braun, Alice, Affonseca Bressan, Rodrigo, Bromet, Evelyn J., Bruggeman, Richard, Buckley, Peter F., Buckner, Randy L., Bybjerg-Grauholm, Jonas, Cahn, Wiepke, Cairns, Murray J., Calkins, Monica E., Carr, Vaughan J., Castle, David, Catts, Stanley V., Chambert, Kimberley D., Chan, Raymond C. K., Chaumette, Boris, Cheng, Wei, Cheung, Eric F. C., Chong, Siow Ann, Cohen, David, Consoli, Angele, Cordeiro, Quirino, Costas, Javier, Curtis, Charles, Davidson, Michael, Davis, Kenneth L., De Haan, Lieuwe, Degenhardt, Franziska, DeLisi, Lynn E., Demontis, Ditte, Dickerson, Faith, Dikeos, Dimitris, Dinan, Timothy, Djurovic, Srdjan, Duan, Jubao, Ducci, Giuseppe, Dudbridge, Frank, Eriksson, Johan G., Fananas, Lourdes, Faraone, Stephen V., Fiorentino, Alessia, Forstner, Andreas, Frank, Josef, Freimer, Nelson B., Fromer, Menachem, Frustaci, Alessandra, Gadelha, Ary, Genovese, Giulio, Gershon, Elliot S., Giannitelli, Marianna, Giegling, Ina, Giusti-Rodríguez, Paola, Godard, Stephanie, Goldstein, Jacqueline I., González Peñas, Javier, González-Pinto, Ana, Gopa, Srihari, Gratten, Jacob, Green, Michael F., Greenwood, Tiffany A., Guillin, Olivier, Gülöksüz, Sinan, Gur, Raquel E., Gur, Ruben C., Gutiérrez, Blanca, Hahn, Eric, Hakonarson, Hakon, Haroutunian, Vahram, Hartmann, Annette M., Harvey, Carol, Hayward, Caroline, Henskens, Frans A., Herms, Stefan, Hoffmann, Per, Howrigan, Daniel P., Ikeda, Masashi, Iyegbe, Conrad, Joa, Inge, Julià, Antonio, Kähler, Anna K., Kam-Thong, Tony, Kamatani, Yoichiro, Karachanak-Yankova, Sena, Kebir, Oussama, Keller, Matthew C., Kelly, Brian J., Khrunin, Andrey, Kim, Sung-Wan, Klovins, Janis, Kondratiev, Nikolay, Konte, Bettina, Kraft, Julia, Kubo, Michiaki, Kučinskas, Vaidutis, Ausrele Kučinskiene, Zita, Kusumawardhani, Agung, Kuzelov a-Ptackova, Hana, Land, Stefano, Lazzeroni, Laura C., Lee, Phil H., Legge, Sophie E., Lehrer, Douglas S., Lencer, Rebecca, Lerer, Bernard, Li, Miaoxin, Lieberman, Jeffrey, Light, Gregory A., Limborska, Svetlana, Liu, Chih-Min, Lönnqvist, Jouko, Loughland, Carmel M., Lubiski, Jan, Luykx, Jurjen J., Lynham, Amy , Macel Jr., Milan, Mackinnon, Andrew, Magnusson, Patrik K. E., Maher, Brion S., Maier, Wolfgang, Malaspina, Dolores, Mallet, Jacques, Marder, Stephen R., Marsal, Sara, Martin, Alicia R., Martorell, Lourdes, Mattheisen, Manuel, McCarley, Robert W., McDonald, Colm, McGrath, John J., Medeiros, Helena, Meier, Sandra, Melegh, Bela, Melle, Ingrid, Mesholam-Gately, Raquelle I., Metspalu, Andres, Michie, Patricia T., Milani, Lili, Milanova, Vihara, Mitjans, Marina, Molden, Espen, Molina, Esther, Molto, María Dolores, Mondelli, Valeria, Moreno, Carmen, Morley, Christopher P., Muntané, Gerard, Murphy, Kieran C., Myin-Germeys, Inez, Nenadić, Igor, Nestadt, Gerald, Nikitina-Zake, Liene, Noto, Cristiano, Neuchterlein, Keith H., O'Brien, N[illegible]uise, O'Neill, F. Anthony., Oh, Sang-Yun, Olincy, Ann, Ota, Vanessa Kiyomi, Pantelis, Christos, Papad[illegible] George N., Paredella, Mara, Paurio, Tiina, Pellegrino, Renata, Periyasamy, Sathish, Perkins, Di[illegible]a O., Pi[illegible]lmann, Bruno, Pietilainen, Olli, Pimm, Jonathan, Porteous, David, Powell, John, Quattrone, Diego, Q[illegible]ested, Digby, Radant, Allen D., Rampino, Antonio, Rapaport, Mark H., Rautanen, Anna,

Supplementary Information

Table of Contents

Consistency of Primary GWAS Results in an Independent Dataset **2**
Polygenicity versus other sources of Inflation **2**
Sex based analyses **3**
Autosomal 3
Schizophrenia XWAS and Dosage Compensation 3
Heterogeneity in SNP effects 3
Meta-analysis 4
X-chromosome gene inactivation status 4
Results X chromosome analyses 4
Conclusion 4
Outcome of 128 independent associations from PGC2 SCZ study (Nature, 2014) **4**
Heritability, SNP-based heritability, variance explained in out of sample prediction, and variance explained by genome-wide significant SNPs. **5**
Conditional Analysis **7**
FINEMAP **8**
Mutation intolerance metrics in fine-mapped genes 8
Summary Based Mendelian Randomisation **9**
Prioritising SMR Genes **10**
Combining eQTL and GWAS Fine-mapping 10
Combining SMR, FINEMAP and chromatin conformation analysis 10
Potential Heterogeneity due to Sample Ascertainment **11**
References ..... Error! Bookmark not defined.

SUPPLEMENTARY FIGURES **20**

Supplementary Figure 1 ..... **20**
Supplementary Figure 2 ..... **21**
Supplementary Figure 3 ..... **22**
Supplementary Figure 4 ..... **24**
Cohort: CLZ2A 24
Cohort: GAP 25
Cohort: XPFLA 26
Cohorts: XJRSA + XJR3A +XJR3B + XRJI6 27
Cohort: COGS 27
Supplementary Figure 5 ..... **29**
Supplementary Figure 6 ..... **30**
Supplementary Figure 7 ..... **31**
Supplementary Figure 8 ..... **32**
Supplementary Figure 9 ..... **33**
Supplementary Figure 10 ..... **34**
Supplementary Figure 11 ..... **36**
Supplementary Figure 12 ..... **38**

CASE-CONTROL SAMPLE DESCRIPTIONS **43**

References **63**

Supplementary Note

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 275.8 of the 308 SNPs (=89.5%) should have the same direction of effect if all the discovery associations are true positives. For details please refer to **Supplementary Table 23**.

Of the index SNPs that were genome wide significant in our primary GWAS, and available from deCODE (N=308), 87% had the same direction of effect in the deCODE cohort, a level greater than expected by chance (Binomial $P=6.8x10^{-43}$). This is also not significantly different (chi square 1df, P=0.40) 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, the phenomenon where discovery effect sizes are overestimated (and therefore so is power to replicate) in imperfectly powered studies. At a more relaxed significance range ($P=1x10^{-6}$ - $5x10^{-8}$) of 303 index SNPs available in the deCODE cohort, 226 showed congruent effects. Again, this is significantly different from the null (Binomial $P=1.89x10^{-18}$), 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 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 inflation[1], 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 confounding[2,3]. 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 heritability[2]. In the core PGC datasets (i.e. the 90 cohorts for which we can validly perform LDSR), 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 UKBiobank[2] (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 (R_g 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 24**, **Supplementary Figure 5**), and specific analysis of the Primary 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) model[4]. 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 $M_{eff(X)} = 1300$[5]. Subsequently, following previously described methods[4,5], 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 effects

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

$$T_d = \frac{\left(\frac{1}{2}\hat{\beta}_m - \hat{\beta}_f\right)^2}{\frac{1}{4}\widehat{SE}\left(\hat{\beta}_m\right)^2 + \widehat{SE}\left(\hat{\beta}_f\right)^2}$$

follows a χ^2-distribution with one degree of freedom under the null hypothesis of no difference in estimates under full DC assumption.

Meta-analysis

The 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 DC[5]. 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 status

For 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 24)**.

Results X chromosome analyses

The estimated across-SNP average effective sample size was 58,591 for males and 42,235 for females **(Supplementary Table 25).** 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 25**, 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 (R_g = 0.992, SE = 0.024), the estimate of genetic correlation between males and females was not significantly different from 1 ($R_{g(X)}$ =1.00, SE=0.09), and there was no indication of significant heterogeneity for any SNP (Mean T_d= 1, Max T_d= 17.1, in line with expectation under a χ^2-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.

Conclusion

We 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 26**. We found at least one genome-wide significant SNP in the current discovery analysis within 50kb of the index SNP defining 116 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 12 of the 128 GWS clumps PGC reported in its last primary study[7] (**Supplementary Figure 6)**. Five of these index SNPs) are subsumed into loci that are GWS in the present study after replication with summary statistics from deCODE Genetics (**Supplementary Table 26)**. Of the remaining 7 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[8], 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 (**Supplementary Table 26)**. 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 (**Online Methods** and **Supplementary Figure 6**).

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 studies[10] at 81% (95%CI 73-90%) or from Swedish national records[11] 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^2_{SNP}). 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^2_{SNP} has been estimated from SCZ GWAS data sets as they have become available. Even in the first analyses using PGC1 data[12] it was noted that the estimates were higher from individual cohorts than when cohorts were combined (International Schizophrenia Consortium (ISC): 0.27 (SE 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. In order to capture the correlation structure in the genome which induce correlations between SNP effect sizes these methods integrate GWAS summary statistics with a data file summarising LD. Since LD is population specific and depends on allele frequencies which differ between populations, SNP-based heritability methods can only be applied (currently) to GWAS summary statistics derived from a single ancestry. In this section we focus on the EUR ancestry summary statistics as this is the largest ancestry represented which therefore affords the most precise estimate. Here we The first such method was LDscore regression (LDSR)[1,13] . In the present study (PGG3), the LDSR estimates for EUR are h^2_{SNP} = 0.21 (SE 0.008) assuming lifetime risk of 1%, and h^2_{SNP} =0.19 (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 shown[13,14] to underestimate h^2_{SNP} . Despite biases in LDSR for h^2_{SNP} estimates, LDSR estimates of genetic correlation are robust, as are h^2_{SNP} enrichment analyses for genomic annotations e.g., exonic sequence[15] or cell-type-enriched expression[16]. More recent methods of estimating h^2_{SNP} from GWAS summary statistics do not provide downward biased estimates using simulated data[17,18], and provide higher estimates when applied to real data[19]. In simulated data sets it can be shown that h^2_{SNP} 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 where the GWAS is a meta-analysis of many cohorts of approximately similar ancestry (e.g. collectively European), the results seem relatively robust to the choice of LD reference. One method to estimate h^2_{SNP} from GWAS summary statistics is SBayesS[18]; when applied to PGC-SCZ data, using the GERA[20] LD reference, the estimates are PGC2[7] 0.21 (s.e. 0.006), and current PGC3 (EUR) 0.24 (0.007).

SBayesS also estimates other genetic architecture parameters: polygenicity (π) and selection (*S*) parameters, where π is the proportion of (HapMap3) SNPs estimated to be causal and *S* describes the effect size-MAF relationship. Using PGC3 (EUR) 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/disorders[18]. Relative to other traits, π of 4% is high, while S of -0.6 is similar to other traits.

While h^2_{SNP} measures the total variance associated with common SNPs, it does not reflect the variance attributable to the specific associations that we have been identified. While the h^2_{SNP} 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-V[21]). For PGC1+SWE[22], and PGC2[7], the number of LD independent (r^2<0.1) GWS associations are 24 and 108 respectively, while in the core PGC dataset, the number of conditionally independent GWS associations is 277. Applying a lifetime risk of 1% the variance obtained from summing the estimates of variance from the autosomal SNPs from each study are: 1.3%, 3.4%, and 6.8% representing an increase in PGC2 and PGC3 over PGC1 of 2.7 and 5.2-fold respectively. 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 endeavoured to mitigate this in PGC3 by using only SNPs that are independent in stepwise regression). Hence, we regard this current estimate of 6.8% 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 ways: 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 also known as C+T) which had its first application to SCZ data (ISC cohort[23]) and has been used in PGC1+SWE[21] and PGC2[7] 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 r^2 >0.1 with it), then the SNPs used to generate the PRS are selected on association p-value (P_T 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 this new GWAS was used to derive risk alleles for PRS analysis in the MGS cohort, and to calculate the variance explained in case-control status, which was reported as Nagelkerke's R^2. In the ISC study (2009), the maximum Nagelkerke's R^2 was 3.2%. In various waves of PGC data, this increased to 6.0% in 2013, 18.4% in 2014 and in the present study, 21.2% was the maximum variance explained (which reflected p-value thresholds in the GWAS; ISC:0.5, PGC1+SWE 0.10, PGC2: 0.05, PGC3 (ALL ancestries): 0.1). 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 R^2 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 R^2 can be difficult to interpret because for the same liability variance in a population sample, the Nagelkerke's R^2 increases as the proportion of cases included in the sample increases to 50% (then decreases again). Alternative evaluation statistics include 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, in 2014, we obtained in PGC2 a value of 0.72[7] and now in the present study (PGC3 (ALL ancestries)), we obtain an AUC of 0.74. To allow comparison with statistics such as h^2 and h^2_{SNP} 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 using PGC2 data was 8.4% and now using PGC3 (ALL ancestries) it is 9.9%. It is noteworthy that AUC 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 (which has been estimated after accounting for the other covariates) to AUC using normal distribution theory generates an AUC for MGS using the PGC2 GWAS of 0.71, and using PGC3 (ALL ancestries) of 0.74.

We also conducted leave-one-sample out PRS analyses, in which each of 98 cohorts from the core PGC dataset is removed from the full discovery meta-analysis (including the African American and Latino samples) in turn and PRS calculated in the left-out-cohort (**Online Methods**). Across all cohorts, the median p-value threshold that maximised the out-of-sample prediction in the left-out cohort was p_T=0.05. At this threshold, the median variance in liability explained is 7.3%, while genome-wide significant SNPs explain 2.4%. To allow comparison with earlier work, we also calculate using the leave-one out design, the median variance explained in the PGC2 cohorts. In our earlier study, when PGC2 is used as a discovery sample the median variance in liability explained was 7.0%[7]; this now increases to 8.5% in the PGC2 cohort when the present (PGC3 (ALL ancestries)) GWAS is used as the discovery sample. Note also that this is considerably higher than the median value of 7.3% across all PGC3 cohorts, likely as a result of the increased representation in PGC3 of samples of non EUR ancestry which have lower variance explained (**Extended Data Figure 1**) presumably due to the predominance of European ancestry cohorts in our discovery GWAS.

The relationship between liability to schizophrenia 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. Across all ancestries in the Primary GWAS, the top centile of PRS is associated with an OR for schizophrenia relative to the rest of the sample of 5.6 (CI 4.9-6.5) (**Supplementary Table 6).** This is larger than reported for a number of other common disorders[24], 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 correlates[25]. 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 is considerable using the Primary GWAS (OR across all ancestries, top v bottom centile = 39 (CI 29-53). Underscoring the need for better representation of populations of different ancestries in the GWAS, much lower values are obtained in the African American sample (OR=4.8) albeit with very wide CI (0.35-66). Full details of out-of-sample predictions are available (**Supplementary Table 6, Supplementary Table 27and Supplementary Table 5)** and summarised in **Extended Data Figure 2**).

Conditional Analysis

Using the same definitions as for the primary analysis, we defined 248 distinct associated genomic loci (5 on the X-chromosome) with in the core PGC dataset. 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 **Online Methods),** adding in the most significant SNP's dosages 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 248 regions, 225 regions had a single signal (5 on the X-chromosome), 20 showed two independent signals, 1 region 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 276 conditionally independent signals. The results are presented in **Supplementary Table 10b** and **Supplementary Figure 7** 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 three instances we found partial dependency across genomic regions, two in close proximity (chr12:122.8-123.1Mb and chr2:200.3-200.5Mb) and one spanning the centromere of chromosome 5 (chr5:46-50Mb). In **Supplementary Table** 28 we show all tested pairs of conditionally independent SNPs within a long-range conditional test (N= 271*270 = 73,170 autosomal pairs). In **Supplementary Figure 8** 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 7**).

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 proposed[26]. 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 v1.1[26] 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 size[27] (code available at: https://github.com/Pintaius/LDmergeFM). To avoid errors in the resulting LD structure caused by overlapping markers being improperly processed by any software we use, tri- and tetra-allelic SNPs were discarded in this stage. FINEMAP v1.4[28] was then applied to resolve each region, allowing for up to maximum of k=5 causal SNPs. For each region and most probable k model, 95% credible sets of causal SNPs were computed using the method implemented within the software, and posterior probabilities derived from the best-fitting causal model. Expected numbers of causal SNPs (K) for each region were derived from the FINEMAP output, specifically from the model-based probabilities of each evaluated number of causal (k) SNPs as follows:

$$K = \sum_{k=1}^{k_{max}} \left(k \times P(k)\right)$$

Where k = 1, … ,k_{max} ; k_{max} 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 patterns[26], differences in sample sizes for each SNP[30], low power to discern independent association signals[29] and associations driven by complex genetic variation (e.g. repetitive or structural variation)[30]. Thus, for downstream analysis, we focussed on 249 regions with low numbers of expected causal SNPs (K<3.5). Of these, 75% (N=186) had the same expected number of causal SNPs as there were conditional associations. For the other 25%, FINEMAP detected more complex association patterns with 1 (N=45), 2 (N=7) or 3+ (N=4) extra signals. At 5 regions, FINEMAP inferred 1 fewer causal variant than did the conditional analysis.

As a check of the assigned posterior probabilities for SNPs, we used a trans-ancestry method of fine-mapping[31]. As this method assumes a single causal variant for each locus instead of the multiple models that FINEMAP can accommodate, we compared its results with those from our primary analysis in K≤1.5 loci, obtaining highly consistent estimates of posterior probability (adjusted r^2=0.96; **Supplementary Figure 9**) and providing additional confidence in the mapping. Additionally, we also estimated each SNP causal posterior probability in an alternative manner, using the "posterior probability of inclusion" method[32], which is averaged across all FINEMAP causal models (**Supplementary Table 11d**). This alternate definition might be of use for further statistical modelling or variant prioritisation studies in schizophrenia, though we note it was highly correlated (r^2=0.97) with the metric we used to derive our main results.

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. To assess whether genes tagged by FINEMAP credible sets displayed greater signatures of mutation intolerance (here defined as 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 skewed[33], we used the "augmented" Mann-Whitney test proposed by Vermeulen and colleagues[34]. This test can be adjusted for covariates while retaining power and control of the type I 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) metric[33]) than the remaining protein-coding genes at the loci (**Supplementary Table 15a)**.

We also looked at the proportion of genes that meet a categorical definition of mutation intolerance in our FINEMAP set (LOEUF ≤ 0.35). Of genes within the genome-wide significant locus boundaries that were not tagged by FINEMAP, 15% were mutation intolerant as defined by this metric compared with 41% of the prioritised FINEMAP (OR. 3.03 [95%CI=1.64-5.60]; $P=3.93x10^{-4}$). The rate of mutation-intolerance in genes tagged by a FINEMAP credible SNP (k<3.5), but that did not meet the prioritisation criteria, was about half way between these two at 30% (OR=2.15 [95%CI=1.58-2.93]; $P=1.08x10^{-6}$), suggesting that as expected, this group remains enriched for schizophrenia susceptibility genes.

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 15b).**

Summary Based Mendelian Randomisation

In O'Brien et al[35] 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 set[36] 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 set[37] 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 16**. In fetal brain[35], the small sample size and therefore low power to discover eQTLs (N=120) meant that only 754 genes had significant ($P_{eQTL}<5.0x10^{-8}$) 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; $P_{SMR}<6.6x10^{-5}$), and none were rejected by the HEIDI test (**Supplementary Table 17a)**. The larger (N~1,500) PsychENCODE[36] 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 ($P_{SMR}<4.6\ x10^{-6}$) that were not rejected by the HEIDI test (**Supplementary Table 17b)**. Recognising that many eQTLs are shared across tissues[38], we next exploited the power of eQTL data from whole blood[37] (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 ($P_{SMR}<6.4\ x10^{-6}$) and were not rejected by HEIDI; 26 of these were significantly ($P_{SMR}<9.1\ x10^{-4}$) associated with the same direction of effect in the SMR analysis of brain (**Supplementary Table 17c).** 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 11).** In total, there were 116 unique genes identified through the SMR & HEIDI analyses above **(Supplementary Table 17d)**, 101 of which had eQTLs that mapped within the loci that were significant in the Extended GWAS.

To test if our results were robust to methodology, we applied FUSION[39] and EpiXcan[40] to the same brain eQTL datasets[35,36] we used for SMR and obtained excellent cross method consistency (**Online Methods** and **Supplementary Table 17e).** Of the 86 SMR associations from adult brain tissue (including those identified first in blood), 85 met the inclusion criteria for EpiXcan and/or FUSION (PsychENCODE and fetal tissue analyses) of which all 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 17e)**, 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 shown[41] 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 arrays[41], 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 loci[42]. Informed by previous findings demonstrating a positive correlation between Hi-C chromatin interacting SNPs and eQTLs for the same gene[43] we 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)[42]. Thus we sought to identify SMR genes that were supported by having high confidence Hi-C contacts from credible SNPs ($K \leq 3.5$) to the promoter of the same gene (**Supplementary Table 18**).

The Hi-C dataset we employed to annotate our credible SNPs[43] 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)[44] 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 $P<2.31\text{x}10^{-11}$ (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 interactions were identified using cortical functional data from the same developmental stage; detailed Methods available in Giusti-Rodríguez *et al*[43]. 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 18**).

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 SCID[45], SCAN[46], MINI[47], CASH[48], 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 $P_T = 0.05$. Given the considerable heterogeneity of the estimates as measured with the I^2 statistic[49] (all $I^2 > 75\%$), meta-analyses across the samples with the relevant defining characteristics were performed with the Der Simonian and Laird random effects model[50]. 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)
# 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 12**).

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* vol. 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. 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* vol. 2 413–433 (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. 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. 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).

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

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

17. 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).

18. 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.

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

20. 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).

21. 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).

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

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

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

25. 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).

26. 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).

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

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

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

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

31. 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.

32. Lam, M. *et al.* Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat. Genet.* **51**, 1670–1678 (2019).

33. 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.

34. 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).

35. 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).

36. Gandal, M. J. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and

bipolar disorder. *Science (80-.).* **362**, (2018).

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

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

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

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

41. 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).

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

43. 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.

44. 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.

45. 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).

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

47. 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).

48. 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).

49. 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).

50. 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).

51. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*. (American Psychiatric Association, 1994).

52. World Health Organisation. The ICD-10 Classification of Mental and Behavioural Disorders Clinical descriptions and diagnostic guidelines. (World Health Organisation).

53. Nilsson, L.-Gör. *et al.* The betula prospective cohort study: Memory, health, and aging. *Aging, Neuropsychol. Cogn.* **4**, 1–32 (1997).

54. Athanasiu, L. *et al.* Gene variants associated with schizophrenia in a Norwegian genome-wide study are replicated in a large European cohort. *J. Psychiatr. Res.* **44**, 748–753 (2010).

55. Nesvåg, R. *et al.* The quality of severe mental disorder diagnoses in a national health registry as compared to research diagnoses based on structured interview. *BMC Psychiatry* **17**, 93 (2017).

56. First, M. B., Williams, J. B. ., Spitzer, R. L. & Gibbon, M. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Clinical Trials Version (SCID-CT)*. (Biometrics Research, New York State Psychiatric Institute, 2007).

57. Van Os, J. *et al.* Replicated evidence that endophenotypic expression of schizophrenia polygenic risk is greater in healthy siblings of patients compared to controls, suggesting gene-environment interaction. the EUGEI study. *Psychol. Med.* (2019) doi:10.1017/S003329171900196X.

58. Gayer-Anderson, C. *et al.* The EUropean Network of National Schizophrenia Networks Studying Gene–Environment Interactions (EU-GEI): Incidence and First-Episode Case–Control Programme. *Soc. Psychiatry Psychiatr. Epidemiol.* (2020) doi:10.1007/s00127-020-01831-x.

59. European Network of National Networks studying Gene-Environment Interactions in Schizophrenia (EU-GEI) *et al.* Identifying Gene-Environment Interactions in Schizophrenia: Contemporary Challenges for Integrated, Large-scale Investigations. *Schizophr. Bull.* **40**, 729–736 (2014).

60. Guloksuz, S. *et al.* Examining the independent and joint effects of molecular genetic liability and environmental exposures in schizophrenia: results from the EUGEI study. *World Psychiatry* **18**, 173–182 (2019).

61. Santoro, M. L. *et al.* Polygenic risk score analyses of symptoms and treatment response in an antipsychotic-naive first episode of psychosis cohort. *Transl. Psychiatry* **8**, 174 (2018).

62. Moretti, P. N. *et al.* Accessing Gene Expression in Treatment-Resistant Schizophrenia. *Mol. Neurobiol.* **55**, 7000–7008 (2018).

63. Swerdlow, N. R., Gur, R. E. & Braff, D. L. Consortium on the Genetics of Schizophrenia (COGS) assessment of endophenotypes for schizophrenia: An introduction to this Special Issue of schizophrenia research. *Schizophr. Res.* **163**, 9–16 (2015).

64. Nurnberger, J. I. *et al.* Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. *Arch. Gen. Psychiatry* **51**, 849–859 (1994).

65. Maxwell, M. Family Interview for Genetic Studies (FIGS): A Manual for FIGS. (NIMH: Clinical Neurogenetics Branch, 1992).

66. Glessner, J. T. *et al.* Strong synaptic transmission impact by copy number variations in schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 10584–9 (2010).

67. Jacquet, H. *et al.* Hyperprolinemia is a risk factor for schizoaffective disorder. *Mol. Psychiatry* **10**, 479–485 (2005).

68. Kay, S. R., Fiszbein, A. & Opler, L. A. The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. *Schizophr. Bull.* **13**, 261–276 (1987).

69. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (DSM-III-R)*. (American Psychiatric Association, 1987).

70. Muñoz-Negro, J. E. *et al.* A dimensional comparison between delusional disorder, schizophrenia and

schizoaffective disorder. *Schizophr. Res.* **169**, 248–254 (2015).

71. Cervilla, J. A. *et al.* Protocolo y metodología del estudio epidemiológico de la salud mental en Andalucía: PISMA-ep. *Rev. Psiquiatr. Salud Ment.* **9**, 185–194 (2016).

72. 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. *J. Clin. Psychiatry* **59 Suppl 20**, 22-33;quiz 34-57 (1998).

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

74. Richards, A. L. *et al.* The Relationship Between Polygenic Risk Scores and Cognition in Schizophrenia. *Schizophr. Bull.* (2019) doi:10.1093/schbul/sbz061.

75. Irish Schizophrenia Consortium & Wellcome Trust Case Control Consortium 2. Genome-Wide Association Study Implicates HLA-C*01:02 as a Risk Factor at the Major Histocompatibility Complex Locus in Schizophrenia. *Biol. Psychiatry* **72**, 620–628 (2012).

76. Di Forti, M. *et al.* Confirmation that the AKT1 (rs2494732) genotype influences the risk of psychosis in cannabis users. *Biol. Psychiatry* **72**, 811–6 (2012).

77. World Health Organisation. *Schedules for Clinical Assessment in Neuropsychiatry*. (World Health Organisation, 1992).

78. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR)*. (American Psychiatric Association., 2000).

79. First, M. B., Gibbon, M., Spitzer, R. L. & Williams, J. *Structured Clinical Interview for DSM-IV Axis I Disorders: Nonpatient Edition (SCID-I/NP)*. (New York State Psychiatric Institute, 1996).

80. Need, A. C. *et al.* A Genome-Wide Investigation of SNPs and CNVs in Schizophrenia. *PLoS Genet.* **5**, e1000373 (2009).

81. Blasi, G. *et al.* Association of *GSK-3β* Genetic Variation With GSK-3β Expression, Prefrontal Cortical Thickness, Prefrontal Physiology, and Schizophrenia. *Am. J. Psychiatry* **170**, 868–876 (2013).

82. Bertolino, A. *et al.* COMT Val158Met polymorphism predicts negative symptoms response to treatment with olanzapine in schizophrenia. *Schizophr. Res.* **95**, 253–255 (2007).

83. Rampino, A. *et al.* A Polygenic Risk Score of glutamatergic SNPs associated with schizophrenia predicts attentional behavior and related brain activity in healthy humans. *Eur. Neuropsychopharmacol.* **27**, 928–939 (2017).

84. Begemann, M. *et al.* Modification of Cognitive Performance in Schizophrenia by Complexin 2 Gene Polymorphisms. *Arch. Gen. Psychiatry* **67**, 879 (2010).

85. Metspalu, A., Köhler, F., Laschinski, G., Ganten, D. & Roots, I. The Estonian Genome Project in the context of European genome research. *Dtsch. Medizinische Wochenschrift* **129**, S25-8 (2004).

86. Shi, J. *et al.* Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460**, 753–757 (2009).

87. Sanders, A. R. *et al.* The Internet-Based MGS2 Control Sample: Self Report of Mental Illness. *Am. J. Psychiatry* **167**, 854–865 (2010).

88. Fedorenko, O. Y. *et al.* Opening up new horizons for psychiatric genetics in the Russian Federation: moving toward a national consortium. *Mol. Psychiatry* **24**, 1099–1111 (2019).

89. Stefansson, H. *et al.* Common variants conferring risk of schizophrenia. *Nature* **460**, 744–747 (2009).

90. Wing, J. K. *et al.* SCAN - schedules for the clinical assessment in neuropsychiatry. *Arch. Gen. Psychiatry* **47**, 589–593 (1990).

91. Kathiresan, S. *et al.* Genome-wide association of early-onset myocardial infarction with common single nucleotide polymorphisms, common copy number variants, and rare copy number variants. *Nat. Genet.* **41**, 334–341 (2009).

92. Guha, S. *et al.* Implication of a Rare Deletion at Distal 16p11.2 in Schizophrenia. *JAMA Psychiatry* **70**, 253 (2013).

93. Levinson, D. F. *et al.* Genome-Wide Association Study of Multiplex Schizophrenia Pedigrees. *Am. J. Psychiatry* **169**, 963–973 (2012).

94. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).

95. Lencz, T. *et al.* Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol. Psychiatry* **12**, 572–580 (2007).

96. Al Eissa, M. M. *et al.* Genetic association and functional characterization of MCPH1 gene variation in bipolar disorder and schizophrenia. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **180**, 258–265 (2019).

97. Spitzer, R. The schedule for affective disorders and schizophrenia, lifetime version. in (New York State Psychiatric Institute, 1977).

98. Spitzer, R. L., Endicott, J. & Robins, E. Research diagnostic criteria: rationale and reliability. *Arch. Gen. Psychiatry* **35**, 773–82 (1978).

99. McGuffin, P., Farmer, A. & Harvey, I. A polydiagnostic application of operational criteria in studies of psychotic illness. Development and reliability of the OPCRIT system. *Arch. Gen. Psychiatry* **48**, 764–70 (1991).

100. Morgan, V. A. *et al.* People living with psychotic illness in 2010: The second Australian national survey of psychosis. *Aust. N. Z. J. Psychiatry* **46**, 735–752 (2012).

101. Nimgaonkar, V. L. *et al.* Joint evaluation of serum C-Reactive Protein levels and polygenic risk scores as risk factors for schizophrenia. *Psychiatry Res.* **261**, 148–153 (2018).

102. O'Donovan, M. C. *et al.* Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat. Genet.* **40**, 1053–1055 (2008).

103. Jones, R. W. *et al.* The Atorvastatin/Donepezil in Alzheimer's Disease Study (LEADe): Design and baseline characteristics. *Alzheimer's Dement.* **4**, 145–153 (2008).

104. Golomb, B. A., Dimsdale, J. E., White, H. L., Ritchie, J. B. & Criqui, M. H. Reduction in Blood Pressure With Statins: Results From the UCSD Statin Study, a Randomized Trial</subtitle>. *Arch. Intern. Med.* **168**, 721 (2008).

105. Andreasen, N. Scale for the Assessment of Positive Symptoms (SAPS). in (University of Iowa, 1984).

106. Andreasen, N. Scale for the Assessment of Negative Symptoms (SANS). (University of Iowa, 1984).

107. Pato, M. T. *et al.* The genomic psychiatry cohort: Partners in discovery. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **162**, 306–312 (2013).

108. Semple, R. Diagnostic interview for Psychosis and Affective Disorders (DI-PAD). (2008).

109. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (DSM-V).* (American Psychiatric Association, 2013).

110. Krawczak, M. *et al.* PopGen: Population-Based Recruitment of Patients and Controls for the Analysis of Complex Genotype-Phenotype Relationships. *Public Health Genomics* **9**, 55–61 (2006).

111. Wichmann, H.-E., Gieger, C., Illig, T. & Group, null for the M. S. KORA-gen - Resource for Population Genetics, Controls and a Broad Spectrum of Disease Phenotypes. *Das Gesundheitswes.* **67**, 26–30 (2005).

112. Schmermund, A. *et al.* Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: Rationale and design of the Heinz Nixdorf RECALL Study. *Am. Heart J.* **144**, 212–218 (2002).

113. Lucae, S. *et al.* P2RX7, a gene coding for a purinergic ligand-gated ion channel, is associated with major depressive disorder. *Hum. Mol. Genet.* **15**, 2438–2445 (2006).

114. Corponi, F. *et al.* Genetic basis of psychopathological dimensions shared between schizophrenia and bipolar disorder. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **89**, 23–29 (2019).

115. Steinberg, S. *et al.* Common variant at 16p11.2 conferring risk of psychosis. *Mol. Psychiatry* **19**, 108–114 (2014).

116. Sullivan, P. F. *et al.* Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol. Psychiatry* **13**, 570–584 (2008).

117. Stroup, T. S. *et al.* The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Project: Schizophrenia Trial Design and Protocol Development. *Schizophr. Bull.* **29**, 15–31 (2003).

118. Lieberman, J. A. *et al.* Effectiveness of Antipsychotic Drugs in Patients with Chronic Schizophrenia. *N. Engl. J. Med.* **353**, 1209–1223 (2005).

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

120. Bergen, S. E. *et al.* Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. *Mol. Psychiatry* **17**, 880–886 (2012).

121. Korver, N., Quee, P. J., Boos, H. B. M., Simons, C. J. P. & de Haan, L. Genetic Risk and Outcome of Psychosis (GROUP), a multi site longitudinal cohort study focused on gene-environment interaction: objectives, sample characteristics, recruitment and assessment methods. *Int. J. Methods Psychiatr. Res.* **21**, 205–221 (2012).

122. Rees, E. *et al.* Analysis of copy number variations at 15 schizophrenia-associated loci. *Br. J. Psychiatry* **204**, 108–114 (2014).

123. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).

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

125. Carroll, L. S. *et al.* Mutation screening of the 3q29 microdeletion syndrome candidate genes DLG1 and PAK2 in schizophrenia. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **156**, 844–849 (2011).

126. 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).

127. Egan, M. *et al.* Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 6917–6922 (2001).

128. Li, Z. *et al.* Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nat. Genet.* **49**, 1576–1583 (2017).

129. Yu, H. *et al.* Common variants on 2p16.1, 6p22.1 and 10q24.32 are associated with schizophrenia in Han Chinese population. *Mol. Psychiatry* **22**, 954–960 (2017).

130. Shi, Y. *et al.* Common variants on 8p12 and 1q24.2 confer risk of schizophrenia. *Nat. Genet.* **43**, 1224–1227 (2011).

131. Yue, W.-H. *et al.* Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. *Nat. Genet.* **43**, 1228–31 (2011).

132. Bigdeli, T. B. *et al.* Contributions of common genetic variants to risk of schizophrenia among individuals of African and Latino ancestry. *Mol. Psychiatry* **25**, 2455–2467 (2020).

133. Castle, D. *et al.* The diagnostic interview for psychoses (DIP) : development, reliability and applications. *Psychol. Med.* **36**, 69–80 (2006).

134. GAIN Collaborative Research Group *et al.* New models of collaboration in genome-wide association studies: The Genetic Association Information Network. *Nature Genetics* vol. 39 1045–1051 (2007).

135. Calkins, M. E. *et al.* The consortium on the genetics of endophenotypes in schizophrenia: Model recruitment, assessment, and endophenotyping methods for a multisite collaboration. *Schizophr. Bull.* **33**, 33–48 (2007).

136. Aliyu, M. H. *et al.* Project among African-Americans to explore risks for schizophrenia (PAARTNERS): Recruitment and assessment methods. *Schizophr. Res.* **87**, 32–44 (2006).

Supplementary Figures

Supplementary Figure 1

For Supplementary Figure 1a see file "Supplementary Figure 1a - 282 autosomal Replication region plots - July2021.pdf", for 1b, see file "Supplementary Figure 1b - 5 x-chromosomal Replication region plots.pdf".

Supplementary Figure 1 legend:

Region plots: We show region plots (265 on the autosome in Figure (a) and 5 on ChrX in Figure (b)) for each of the associated loci with each page in the document containing a region plot for a single associated locus. The x-axis is chromosomal position (in kb) and the y-axis is the significance of association represented as –log10(p-value), a two-sided test from the fixed effects meta-analysis of variant effects, unadjusted for multiple comparisons. The green line shows the genome-wide significance level ($p=5\times10-8$). Dot size is proportional to LD between the plotted SNP and the index SNP defining the associated region.

Colour scheme: If only a single index SNP in the region, colour is based on degree of LD to the single index SNP as represented by R2. Legend for R2 is given in upper left corner. If a locus contains multiple independent index SNPs, each index SNP is denoted by a different colour. LD to each index SNP is denoted by the intensity of the same colour. Details of index SNPs are given in the upper right corner (labelled in blue): snp=SNP (or indel) name, p =association p-value, or = odds ratio for the minor allele, maf = minor allele frequency, info = imputation quality, direction denotes the consistency of the allelic association enrichment across all the studies N_one_direction –N_other_direction –missing (Note: for x-chromosomal regions we performed a meta analysis of meta-analyses, that's why this information adds up to only 5 in Fig 1b, please compare with Legend of Supplementary Figure 2). Black triangles denote p-values from the fixed effect meta-analysis additionally including replication data (also indicated by +rep) and are only shown for index SNPs. Genes (in green) in lower half of the plot: annotated based on UCSC (August 2017 freeze) with black vertical lines for exons. Arrowheads denote the direction of transcription. To minimize complexity, in some plots we exclude SNPs with fixed effect meta-analytic P-values above certain thresholds (indicated by filter: $p < x$). The blue line denotes regional recombination rates derived from HapMap. A) Autosomal B) X chromosome

Supplementary Figure 2

For Supplementary Figure 2a see file "Supplementary Figure 2a - 337 autosomal Replication forest plots.pdf", for 2b see file "Supplementary Figure 2b - 5 x-chromosomal Replication forest plots.pdf".

Supplementary Figure 2 Legend:

Forest plots: Listing all 329 index SNPs with genome wide significance after combination with replication summary statistics. 324 autosomal SNPs in Fig 2a, 5 SNPs on chromosome X in Fig 2b

First row: SNP identifier, Allele1/Allele2 and chromosome:position.

Second row gives the direction of the SNP in each study in the cohort collection meta-analysis. There are two cohort collections here “PGC_SCZ wave2 only” and “PGC_SCZ wave3 (excl. wave2)” described below. Here “het_P” and “het_I” stand for heterogeneity-significance and heterogeneity-strength between these cohort sets.

The left-hand side table lists details of distinct sub-meta analyses:

1. **ngt**: N of genotyped (not imputed) cohorts on this SNP.
2. **info**: Imputation quality (weighted over cohorts).
3. **p_value:** Significance.
4. **f_ca(n):** Frequency of A1 in cases (number of cases).
5. **f_co(n)**: Frequency of A1 in control (number of control).
6. **ln(OR)**: Natural log of odds ratio and **STDerr**; standard error.
7. **PGC_SCZ wave3 eur only:** Discovery, all European cohorts wave3 combined.
8. **PGC_SCZ wave3 asn only:** Discovery, all Asian cohort in wave3 combined. “(nan)” if SNP has MAF < 1% in the Asian meta-analysis and therefore does not contribute to the final meta analytic results.
9. **PGC_SCZ wave2 only:** Discovery, approximation of PGC SCZ wave2 including European and Asian) cohort collection.
10. **PGC_SCZ wave3 excl. wave2:** Wave3 — including European and Asian cohorts and excluding approximate PGC SCZ wave2 cohort collection.
11. **PGC_SCZ wave3 incl. wave2:** Final discovery meta-analysis.
12. **Decode − Replication:** Replication summary statistics in decode meta-analysis of two cohorts.
13. **PGC_SCZ wave3 incl. wave2 and replication:** Summary statistics from discovery and replication meta-analysis.

Plot on the right side: The x-axis shows natural log of odds ratio ln(OR). Each box and diamond along the y-axis show an effect size taken from the table on the left. The horizontal line passing through box and the outer edges of diamonds correspond to the 95% CI around the effect size. Size of box or diamond is proportional to the corresponding significance (more significant is larger) of each effect. Vertical red line shows no/zero effect (OR=1.0).

For SNPs on chromosome X the displayed sub-results are changed slightly:

1. **wave3 (incl. wave2) eur males:** Discovery, all European cohorts wave3 combined, male only.
2. **wave3 (incl. wave2) eur females:** Discovery, all European cohorts wave3 combined, female only.
3. **wave3 (incl. wave2) asn males:** Discovery, all Asian cohorts wave3 combined, male only.
4. **wave3 (incl. wave2) asn females:** Discovery, all Asian cohorts wave3 combined, female only.
5. **wave3 (incl. wave2) trios:** Discovery, wave3 combined, trios only.

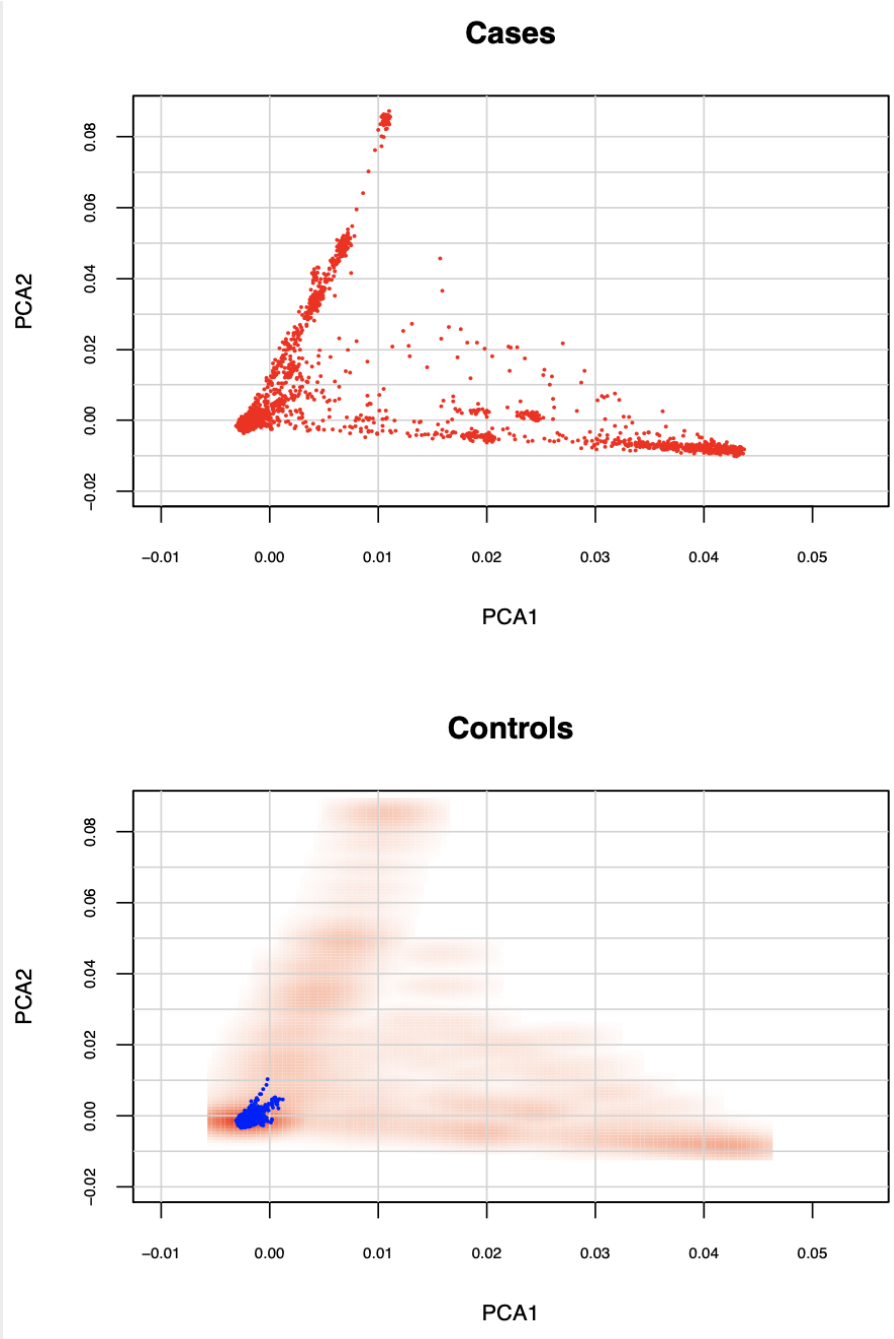
Supplementary Figure 3

Supplemetnary Figure 3 legend:

Associations between schizophrenia cell types from the mouse nervous system. The mean of strength of association evidence using two enrichment methods ($-\log_{10}P_{MAGMA}$, $-\log_{10}P_{LDSC}$) between schizophrenia and gene expression specificity derived from single cell RNA-seq[24] is shown for 265 mouse cell types tested. Details of all cell type definitions are given at URL: http://mousebrain.org/celltypes/. The bar colour indicates whether the cell type is significantly associated with both methods (i.e. MAGMA and LDSR), one method or none. The black vertical line represents the significance threshold corrected for the total number of tests. We also analysed previous waves of PGC schizophrenia GWAS[11,21] shown for comparison.

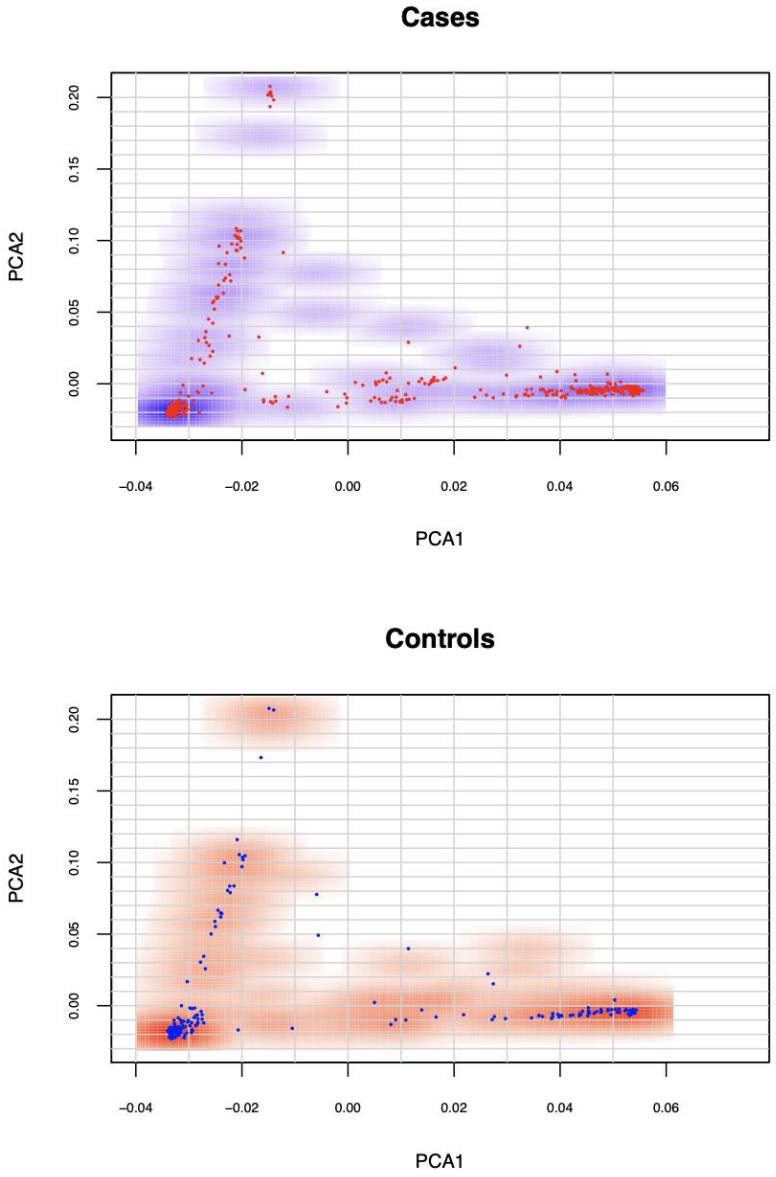
Supplementary Figure 4

Cohort: CLZ2A



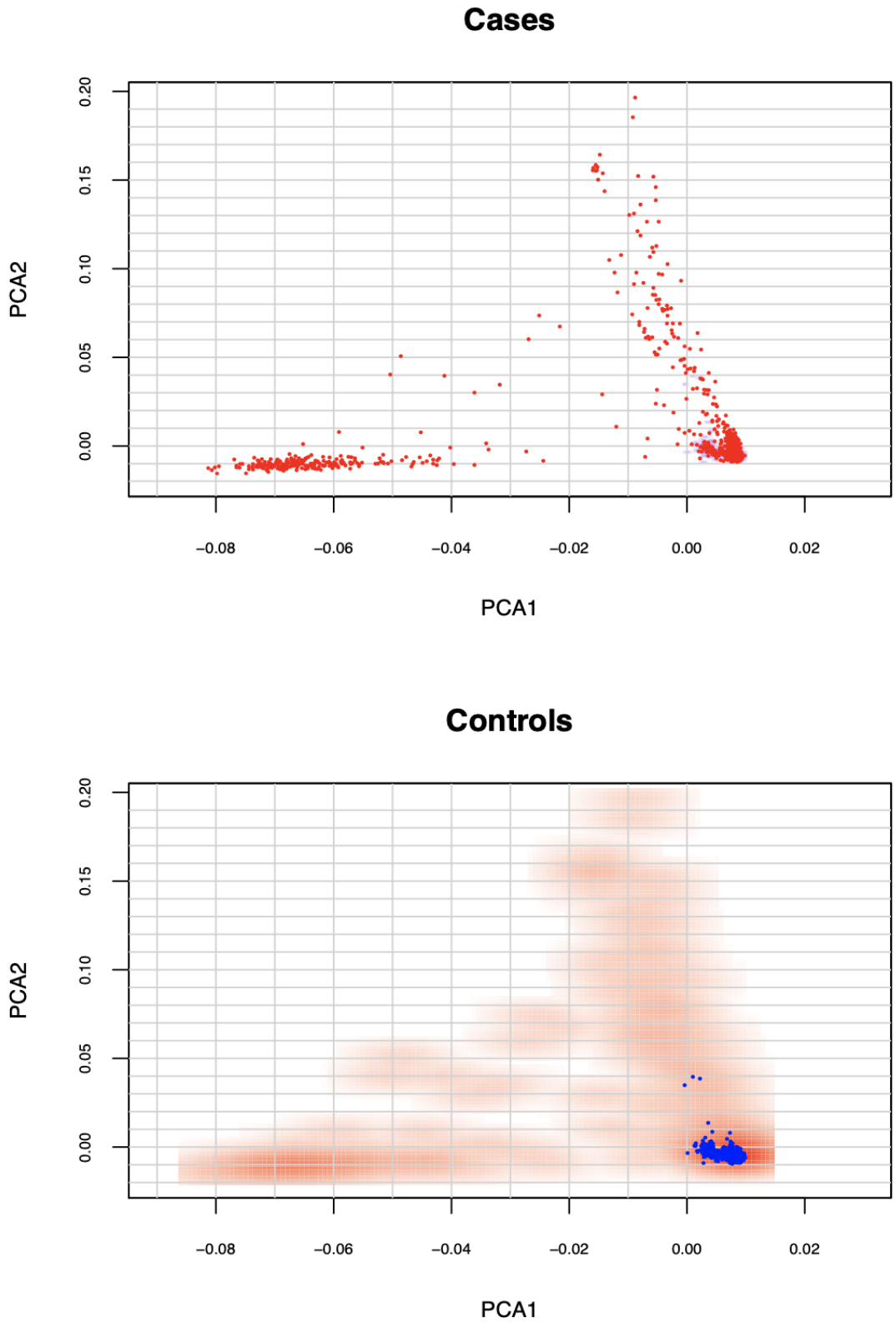
PCA analysis of CLZ2A case control sample: Cases included individuals of diverse ancestries, including admixed individuals, whereas the controls were essentially of European ancestry. Restricting to individuals with PCA1 < 0.0 and PCA2 < 0.01 - excluding 1553 cases and 60 controls in the process- led to a reduction in Lambda-QC from 45.45 to 1.4 before inclusion of PCA covariates when it further dropped to 1.136. The excluded samples are of multiple ancestries and do not form one or more cohorts of sufficient size relative to the total sample size to include in in a tightly controlled meta-analysis (https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0204056).

Cohort: GAP



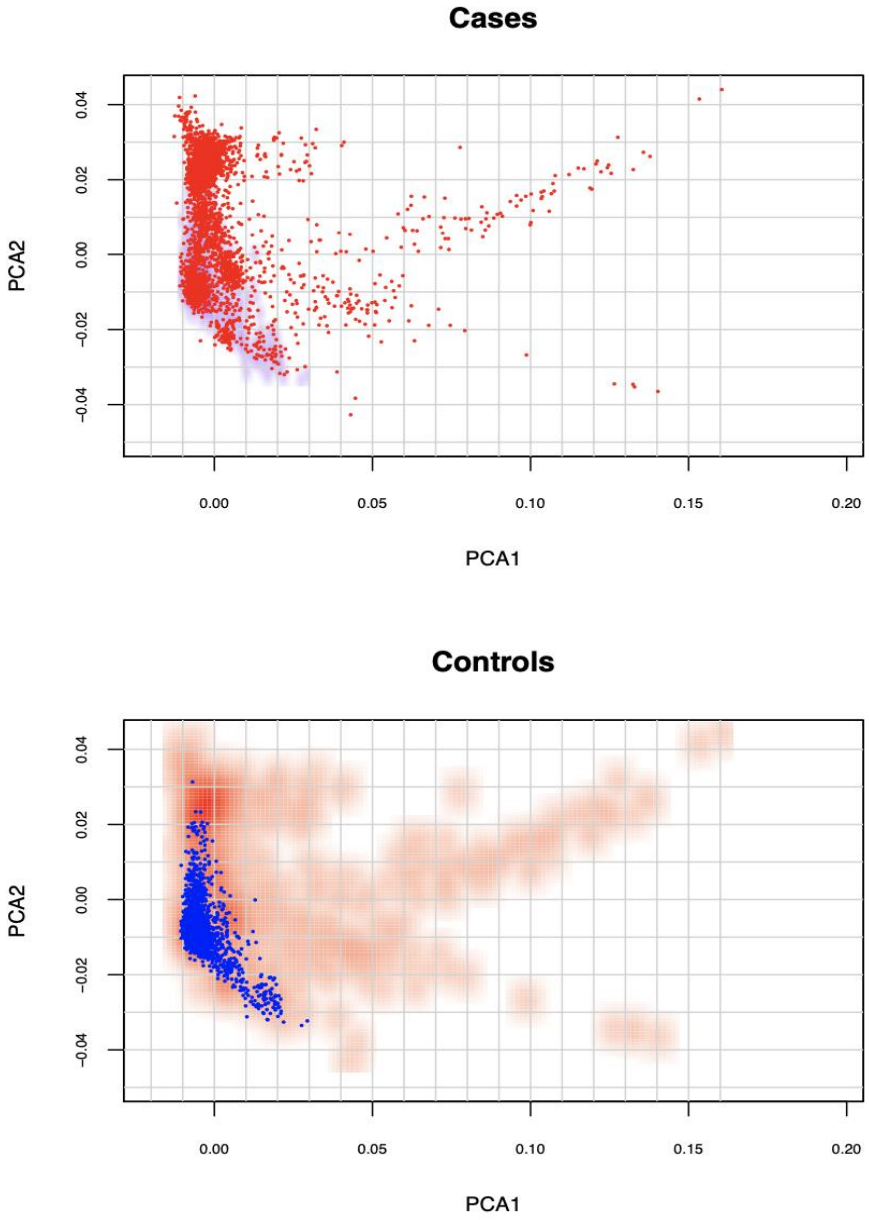
PCA analysis of GAP case control sample: Cases and controls both included individuals of diverse ancestry ancestries but the raw lambda-QC was high at 3.77. Restricting to individuals with PCA1 < -0.02 and PCA2 < 0.01 - excluding 300 cases and 102 controls - led to a reduction of lambda-QC to 1.22 before the inclusion of PCA covariates, which further reduced it to 1.046. The excluded samples are of multiple ancestries and do not form one or more cohorts of sufficient size relative to the total sample size to include in a tightly controlled meta-analysis ((https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0204056).

Cohort: XPFLA



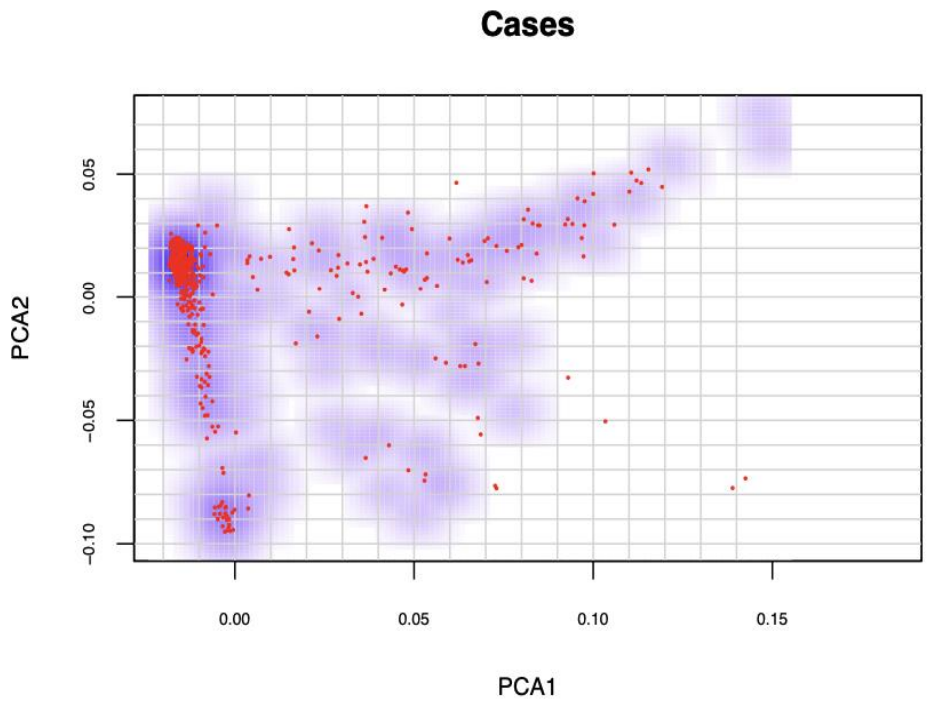
PCA analysis of XPFLA case control sample: The cases were of a much more diverse ancestry than the available controls genotyped on the same platform, which were essentially of European ancestry. Restricting to individuals with PCA1 > 0.00 and PCA2 < 0.01 - excluding 355 cases and 4 controls - led to a reduction in Lambda-QC from 13.3 down to 1.31 before inclusion of PCA covariates when it further reduced to 1.042.

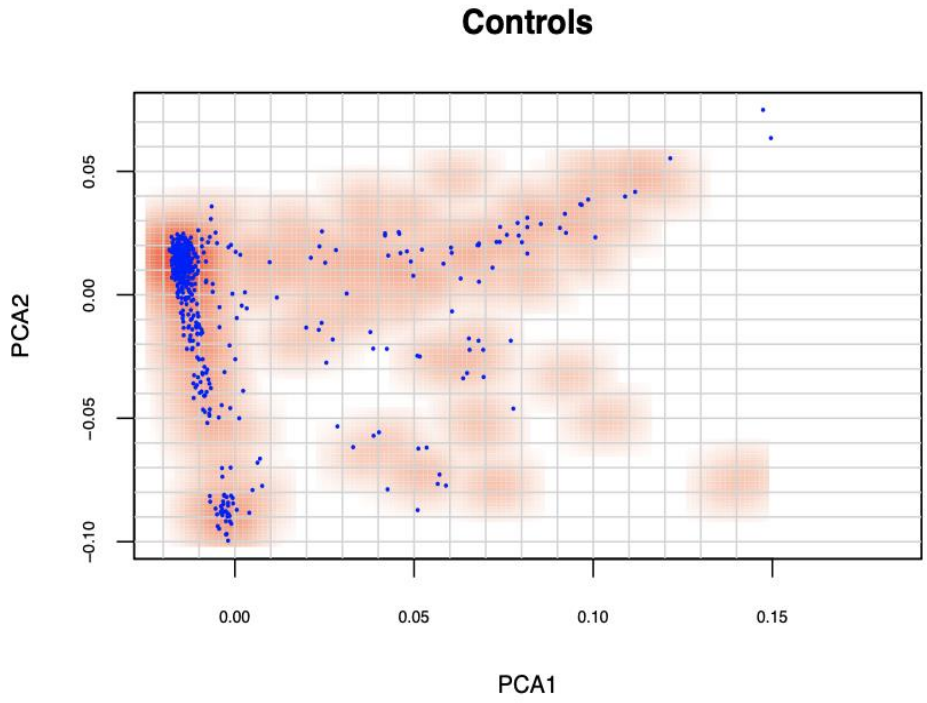
Cohorts: XJRSA + XJR3A +XJR3B + XRJI6



PCA analysis of J&J and Roche case control: We obtained 2744 cases from J&J and Roche (all self defined as "white") from 3 waves of case data which we matched to 10 merged control cohorts, resulting in lambda-QC of 4.67 for the primary association analysis (XJRSA, Supplementary table 22). Cases were more diverse than controls which were substantially of European ancestry. We performed an iterative process of creating new sub-cohorts of XJRSA by splitting samples from the cases and controls, remerging, and case-control matching, aiming to maximize case inclusion while minimizing lambda-QC. In doing so, we derived a smaller XJRSA cohort and 3 sub-cohorts; XJR3A +XJR3B + XRJI6. In total, 294 cases and 270 controls were excluded as PCA outliers and we retained a total of 2450 cases and 3407 controls. Across the 4 cohorts, lambda-QC ranged between 1.04 and 1.3, which was further improved by inclusion of PCA covariates (1.022-1.093) (**Supplementary Table 22**). The excluded samples are of multiple ancestries and do not form one or more cohorts of sufficient size relative to the total sample size to include in in a tightly controlled meta-analysis (https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0204056).

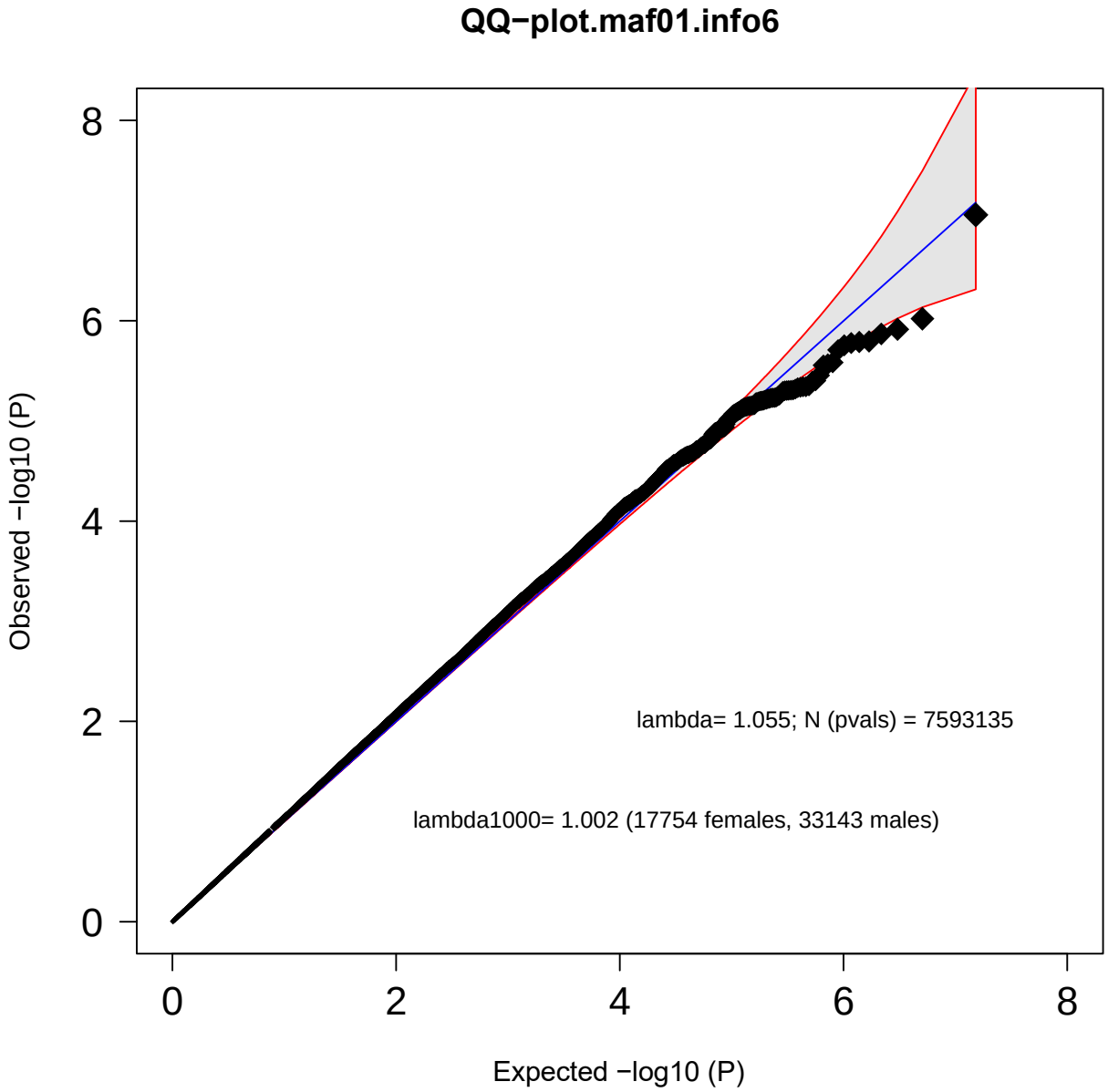
Cohort: COGS





PCA analysis of COGS case control sample: The COGS cases and controls are both highly stratified including individuals across many ancestries. Restricting to individuals with PCA1 < 0.01 - excluding 102 cases and 80 controls - led to a reduction in Lambda-QC from 1.04 down to 1.01 before inclusion of PCA covariates. The excluded samples are of multiple ancestries and do not form one or more cohorts of sufficient size relative to the total sample size to include in in a tightly controlled meta-analysis (https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0204056).

Supplementary Figure 5



Supplementary Figure 5 Legend:

A Quantile-Quantile plot from a GWAS with sex as an outcome. There is no sign of significant associations, nor inflation due to stratification from lambda or lambda1000. The x-axis shows the expected −log10(*P*) values for association under the null distribution given number of independent tests. The y-axis denotes the observed −log10(*P*). We truncate the Y-axis at −log10(*P*)=10. The shaded area surrounded by a red line indicates the 95% confidence interval under the null. Lambda is the observed median $\chi 2$ test statistic divided by the median expected $\chi 2$ test statistic under the null.

Supplementary Figure 6

For Supplementary Figure 6 see file "Supplementary Figure 6 - PGC wave 2 comparison region plots.pdf".

Supplementary Figure 6 legend:

12 region plots for regions that were previously reported as containing genome-wide significant associations in Ripke *et al*. 2014 wave2 meta-analysis, that do not contain genome-wide significant associations in the current wave3 meta-analysis. We show four plots for each region: using summary statistics from "wave2"-only cohort meta-analysis (left hand column), and summary statistics from the current meta-analysis, i.e. "wave3" (right hand column). Additionally, we show the region plus a 50kb flanking region (upper row) and a 2mb flanking region (bottom row).

The x-axis is chromosomal position (in kb) and the y-axis is the significance of association represented as –log10(P). The green line shows the genome-wide significance level (5×10^{-8}). Dot size is proportional to LD between the plotted SNP and the index SNP defining the associated region. A) Autosomal B) X chromosome.

Colour scheme: If only a single index SNP in the region, colour is based on degree of LD to the single index SNP as represented by R2. Legend for R2 is given in upper left corner. If a locus contains multiple independent index SNPs, each index SNP is denoted by a different colour. LD to each index SNP is denoted by the intensity of the same colour. Details of index SNPs are given in the upper right corner (labelled in blue): snp=SNP (or indel) name, p =association p-value, or = odds ratio for the minor allele, maf = minor allele frequency, info = imputation quality, direction denotes the consistency of the allelic association enrichment across all the studies N_one_direction –N_other_direction -missing. Black triangles denote p-values after inclusion of replication data (also indicated by +rep) and are only shown for index SNPs. Genes (in green) in lower half of the plot: annotated based on UCSC (August 2017 freeze) with black vertical lines for exons. Arrowheads denote the direction of transcription. To minimize complexity, in some plots we exclude SNPs with association P-values above certain thresholds (indicated by filter: p < x). The blue line denotes regional recombination rates derived from HapMap. A) Autosomal B) X chromosome

Supplementary Figure 7

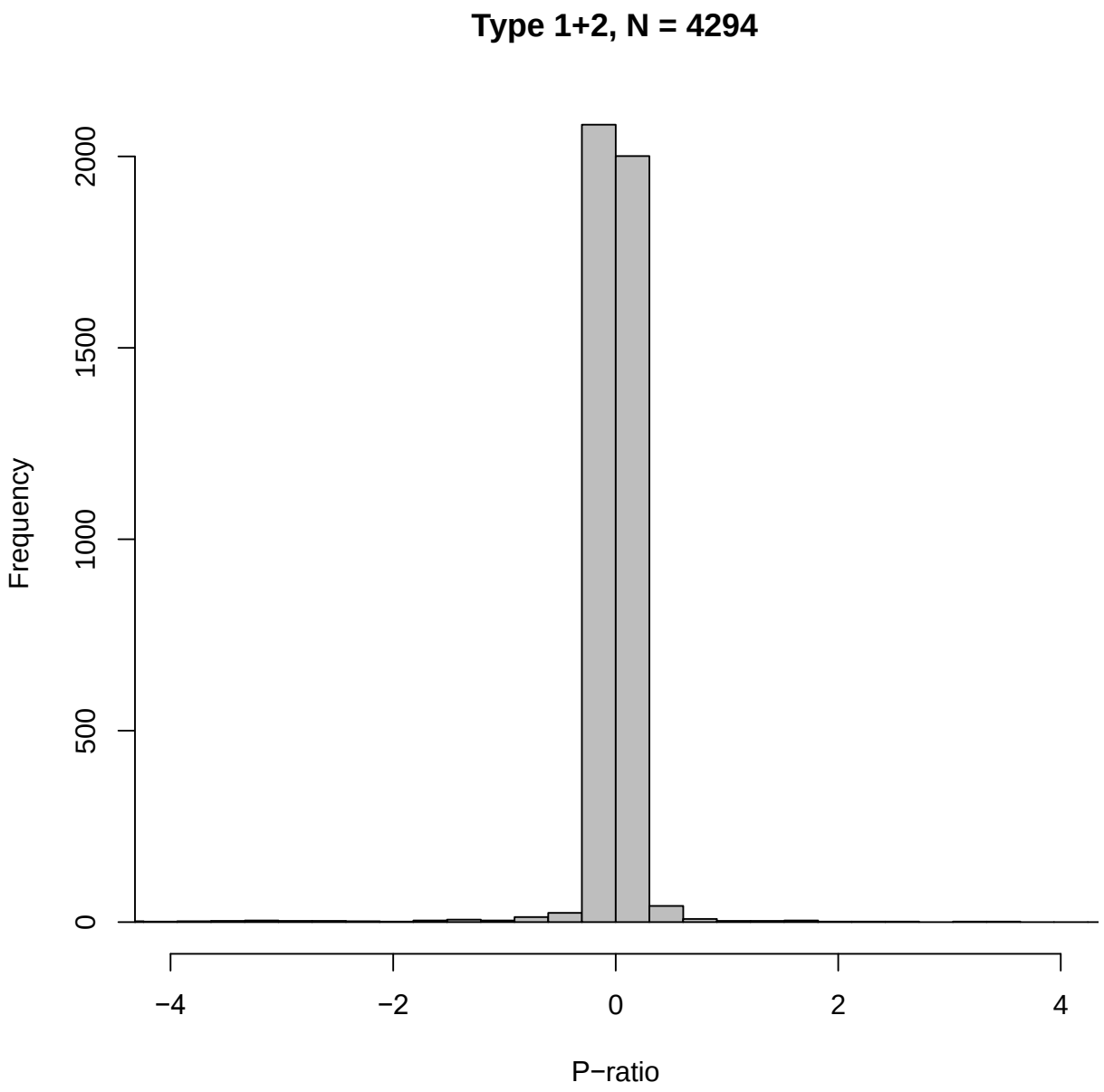
For Supplementary Figures 7 file "Supplementary Figure 7 - Conditional region plots.pdf".

Supplementary Figure 7 legend:

Region plots for conditional analyses performed on autosomal (8a) and chromosome X index (8b) variants. Region plots from the conditional analyses, ordered by significance of the index SNP in the main analysis. The sequence of plots for each locus starts with the unconditional results with the index SNP in the title "SCZ_uncond-rs58120505-…" , and proceeds by conditioning on that index SNP (e.g. "SCZ_cond-rs58120505-…"). As of note the unconditional results are slightly different to the main results, since we excluded the one cohort that only provided summary statistics. Conditioning is repeated as long there is an index SNP left with $P < 1.0*10^{-06}$ by adding the most significant SNP into the conditioning term.

For details about region plot: see Legend for Supplementary Figure 1.

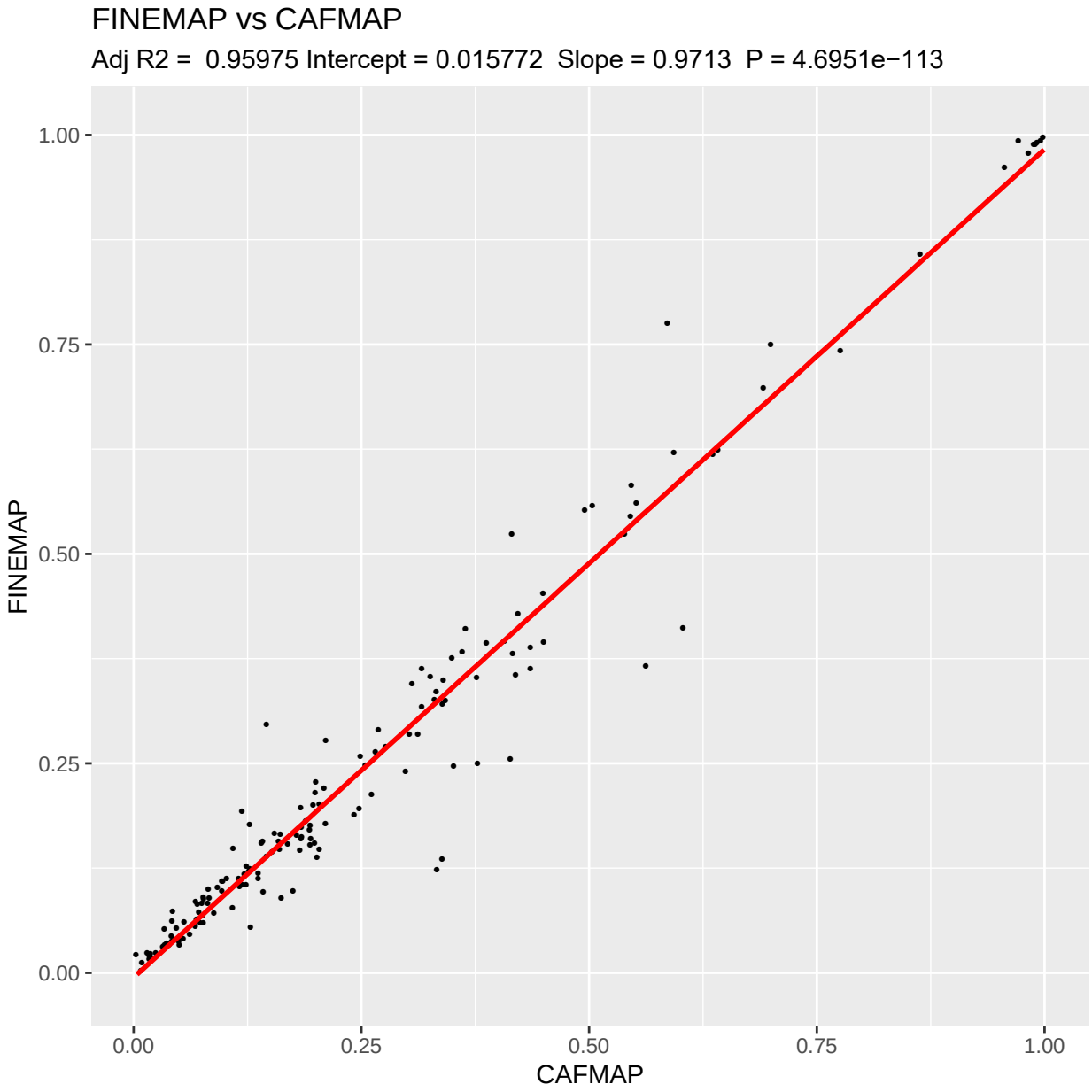
Supplementary Figure 8



Supplementary Figure 8 legend:

Here we present two histograms of all p-value changes (in orders of magnitude) separately for (i) 4,928 SNP pairs within chromosomes (Type 1+2) and (ii) 81,804 SNP pairs 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. Type=1:intra-region, 2:inter-region, within same chromosome, 3:across chromosomes, also compare with supplement table ST27. All SNP pairs independent (R^2<0.1) with imputation reference LD content. c$Pratio = orders of magnitude change before/after conditioning. Frequency = absolute number within histogram-bin.

Supplementary Figure 9



Supplementary Figure 9 legend:

Here we present two histograms of all p-value changes (in orders of magnitude) separately for (i) 4,928 SNP pairs within chromosomes (Type 1+2) and (ii) 81,804 SNP pairs 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. Type=1:intra-region, 2:inter-region, within same chromosome, 3:across chromosomes, also compare with supplement table ST27. All SNP pairs independent (R^2<0.1) with imputation reference LD content. c$Pratio = orders of magnitude change before/after conditioning. Frequency = absolute number within histogram-bin.

Supplementary Figure 10

For Supplementary Figure 10a see file "Supplementary Figure 10a - SMR plots.pdf".

Supplementary Figure 10a legend:

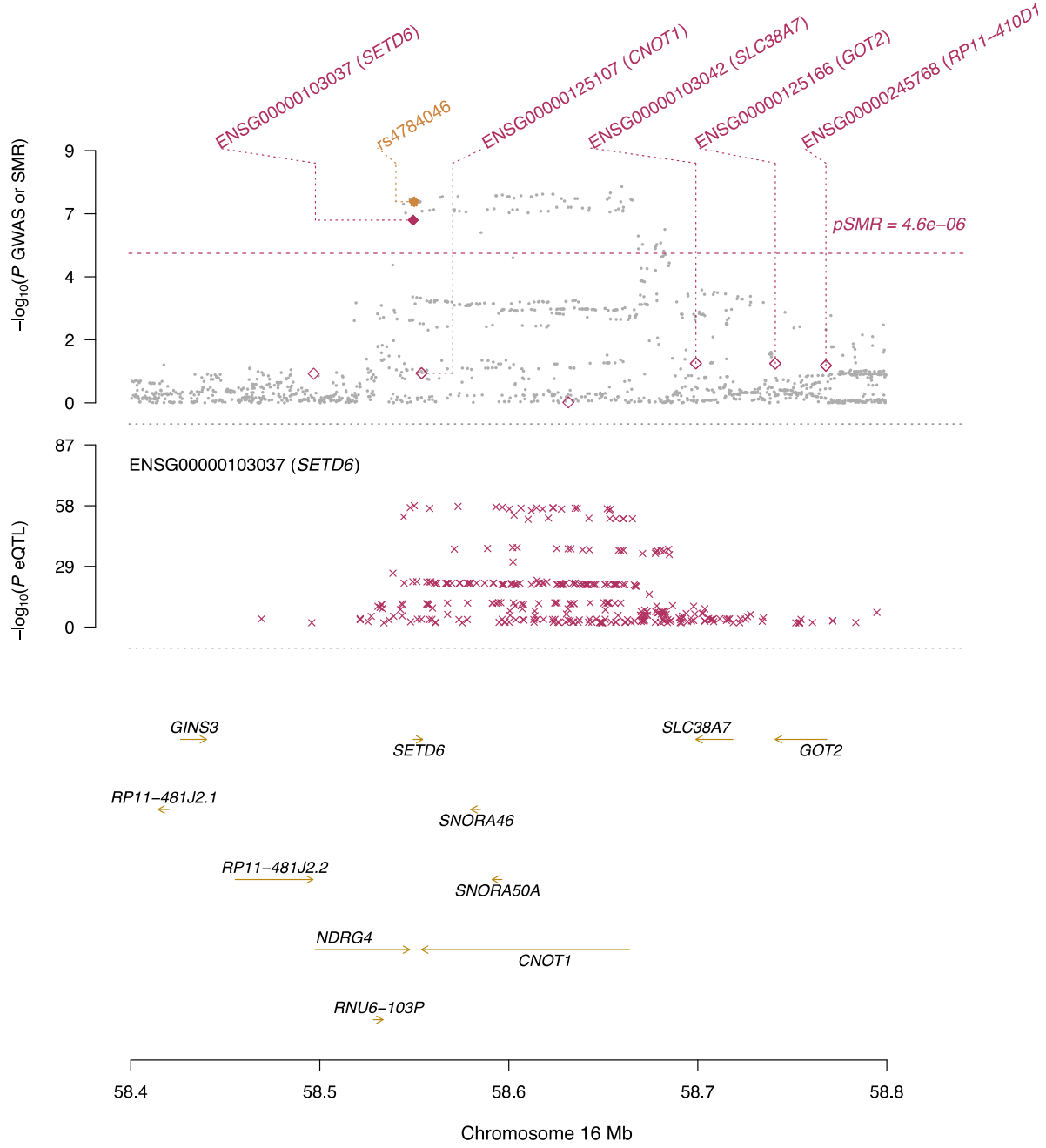
SMR+HEIDI plots.

All significant SMR results are provided in the supplementary data.

The first track shows -$\log_{10}$(P-value) of SNPs (grey dots) from the GWAS of schizophrenia (pre-extension data). Each red rhombus indicates the -$\log_{10}$(P-value) from the SMR tests for associations of gene expression with schizophrenia. A solid rhombus represents a probe not rejected by the HEIDI filtering. The yellow rhombus denotes the schizophrenia association of the SNP that is the top cis-eQTL. The second track shows -$\log_{10}$(P-value) of the SNP association with gene expression probes (ENSGXXX). The bottom track shows the genes underlying the genomic region.

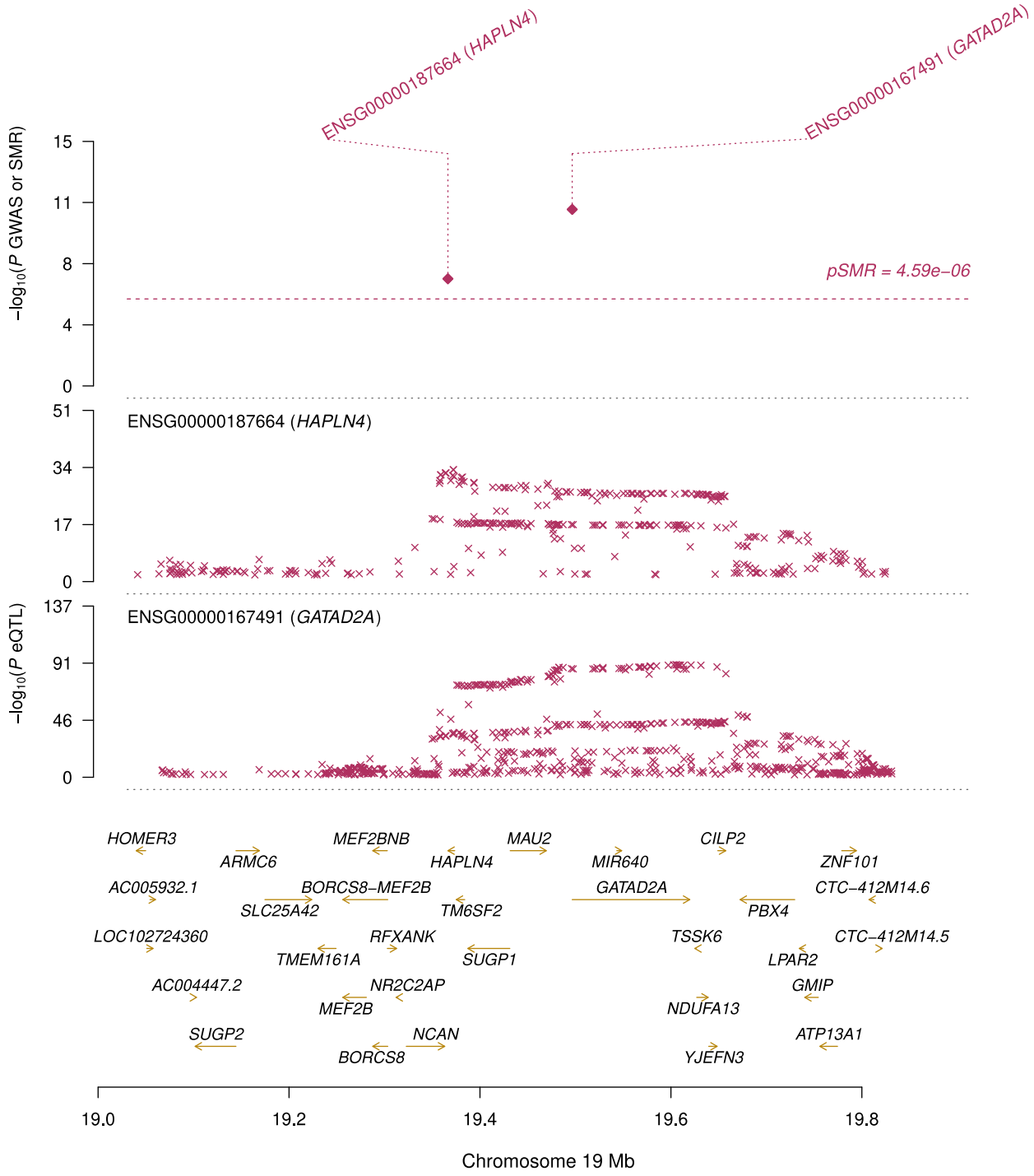
Supplementary Figure 10b Example SMR plot

SETD6: eQTL from PsychENCODE



SMR+HEIDI analysis combining data from GWAS and eQTL studies. All significant SMR results are provided in the supplementary data. In the above example, there is strong evidence that the schizophrenia association may be mediated through expression at a single gene *SETD6*.
The top track shows -$\log_{10}$(P-value) of SNPs (grey dots) from the GWAS of schizophrenia (pre-extension data). Each red rhombus indicates the -$\log_{10}$(P-value) from the SMR tests for associations of gene expression with schizophrenia. A solid rhombus represents a probe not rejected by the HEIDI filtering. The yellow rhombus denotes the schizophrenia association of the SNP that is the top cis-eQTL (rs4784046). The middle track shows -$\log_{10}$(P-value) of the SNP

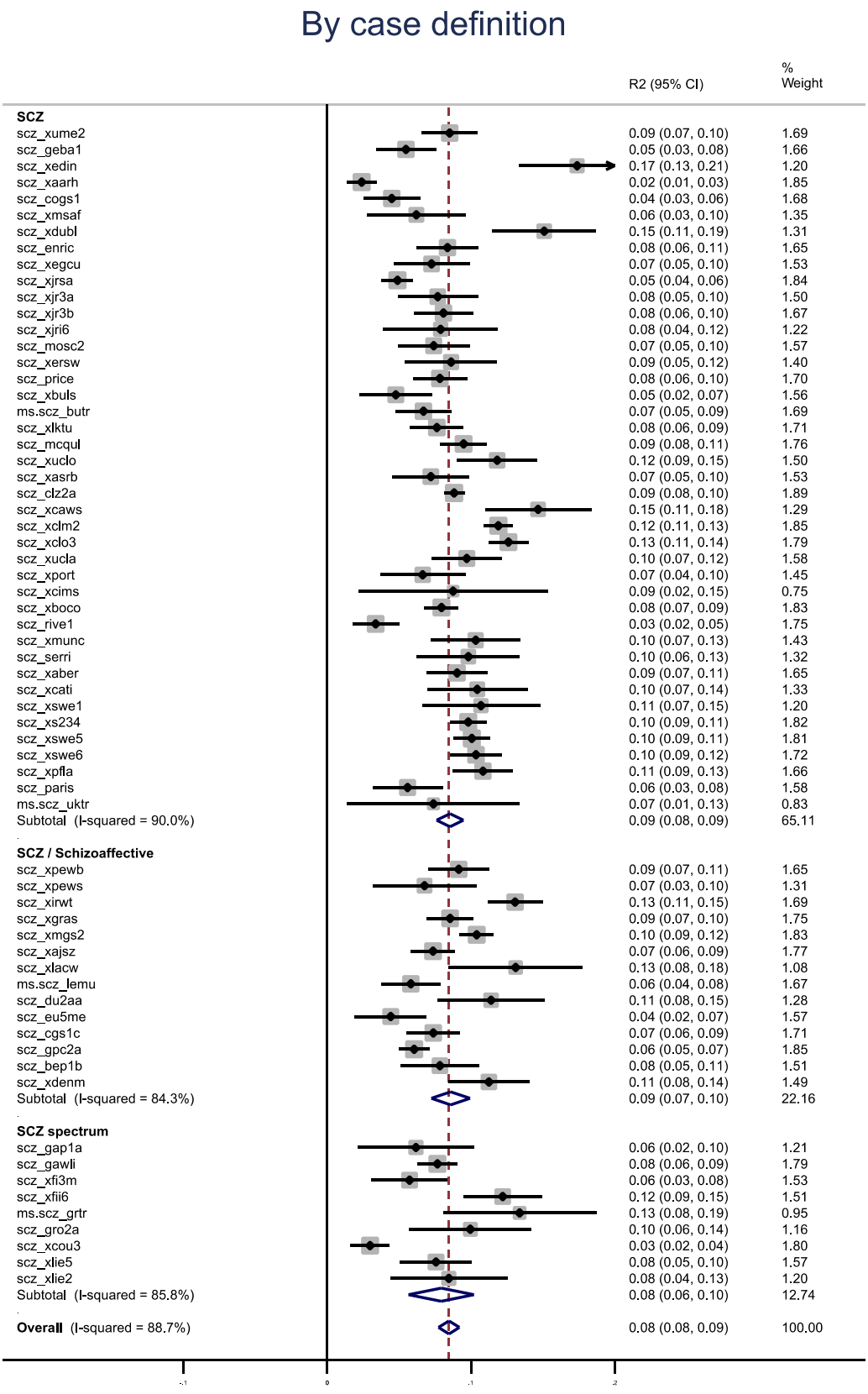
Supplementary Figure 11



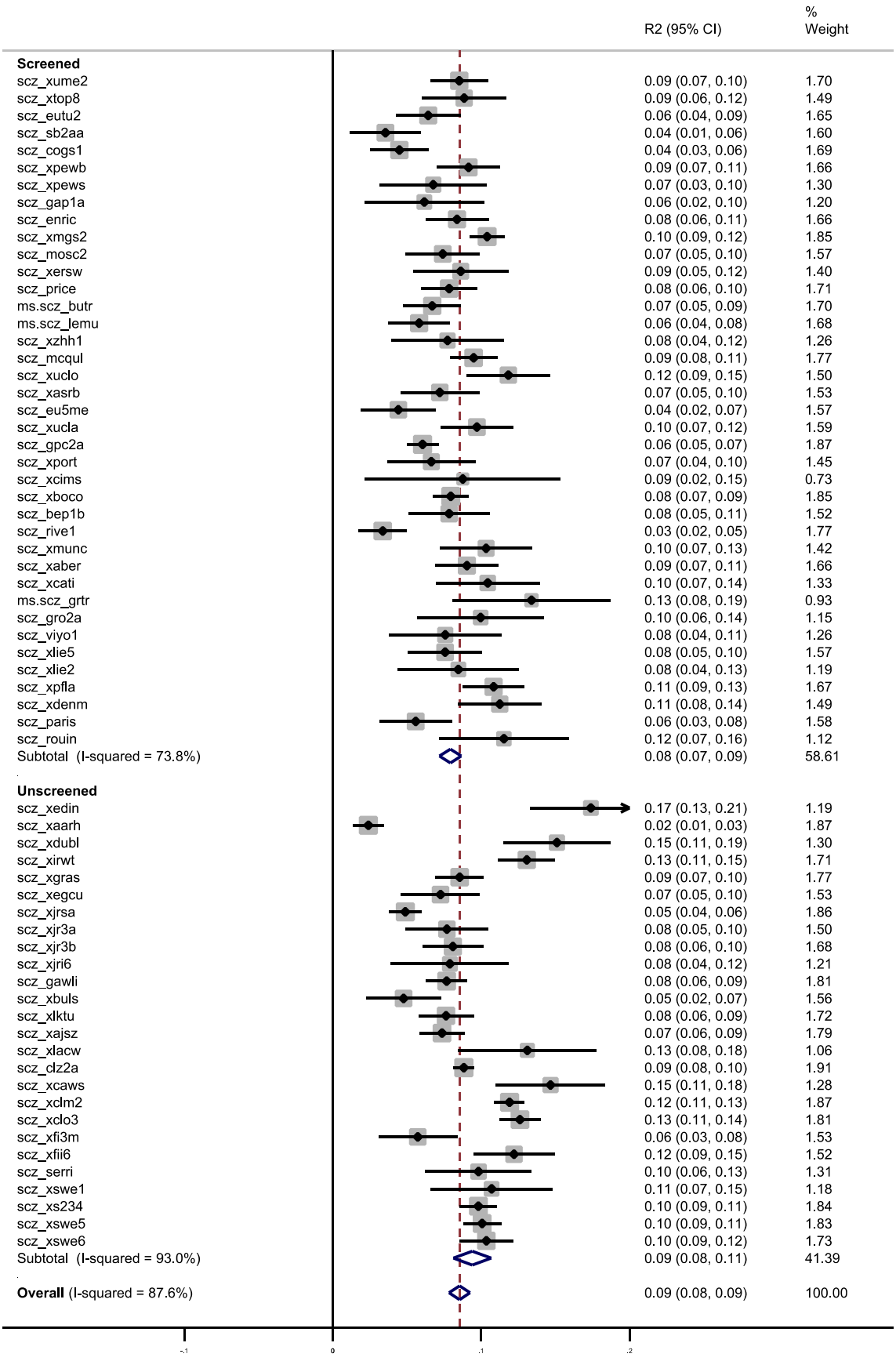
Supplementary Figure 11 Legend:

Exemplar plot of GCTA-COJO analysis for locus with SNP rs1858999, a locus with two genes showing significant SMR associations. There are three plots, with four tracks. Each plot is laid out with genomic location on the x-axis, and $-\log_{10}$(*P*-values) for a statistical test on the y-axis. The top track reports as grey dots for $-\log_{10}$(*P*-values) for association tests in the Primary SCZ meta-analysis, and red diamonds for SMR tests surviving the HEIDI filter. Each red diamond is labeled with the putative mediating gene Ensembl ID and gene name. Solid red diamonds report the SMR test $-\log_{10}$(*P*-value) that are above the SMR significance threshold, and unfilled red diamonds report SMR tests below this threshold, with the Bonferroni corrected significance threshold for SMR tests shown as a dotted red line. The second track shows the $-\log_{10}$(*P*-values) values of SNP eQTL associations with expression in ENSG00000167491 (*GATAD2A*), and the third track shows $-\log_{10}$(*P*-values) of SNP eQTL associations with expression in ENSG00000187664 (*HAPLN4*). The bottom track shows physical locations for genes in the region. The first plot shows unconditional associations for tests in each track. The second plot shows association tests, in all tracks, conditional on the top associated cis-eQTL SNP for *GATAD2A* (highlighted in yellow) using GCTA-COJO. Similarly, the third plot shows tests of association conditional on the top cis-eQTL SNP for *HAPLN4.*

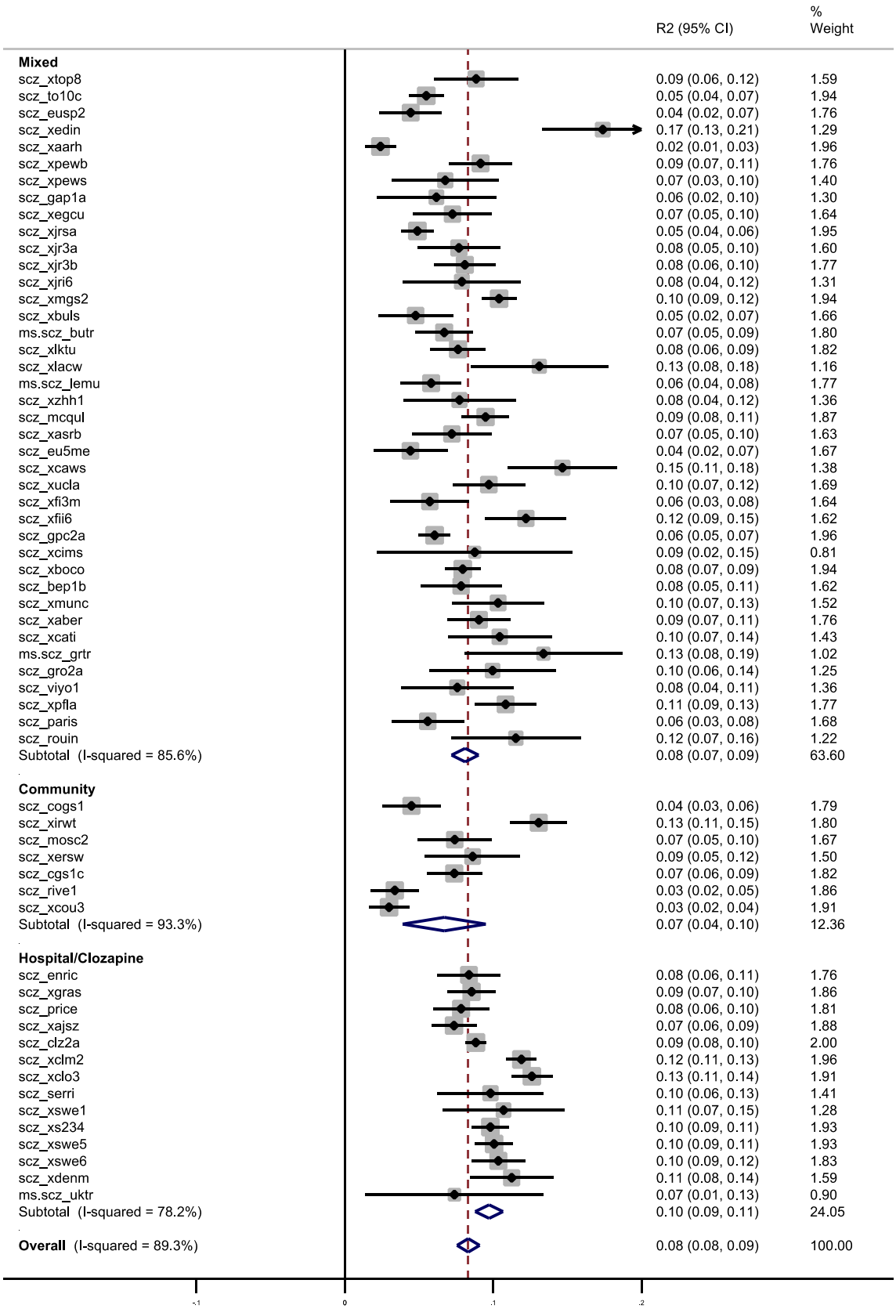
Supplementary Figure 12



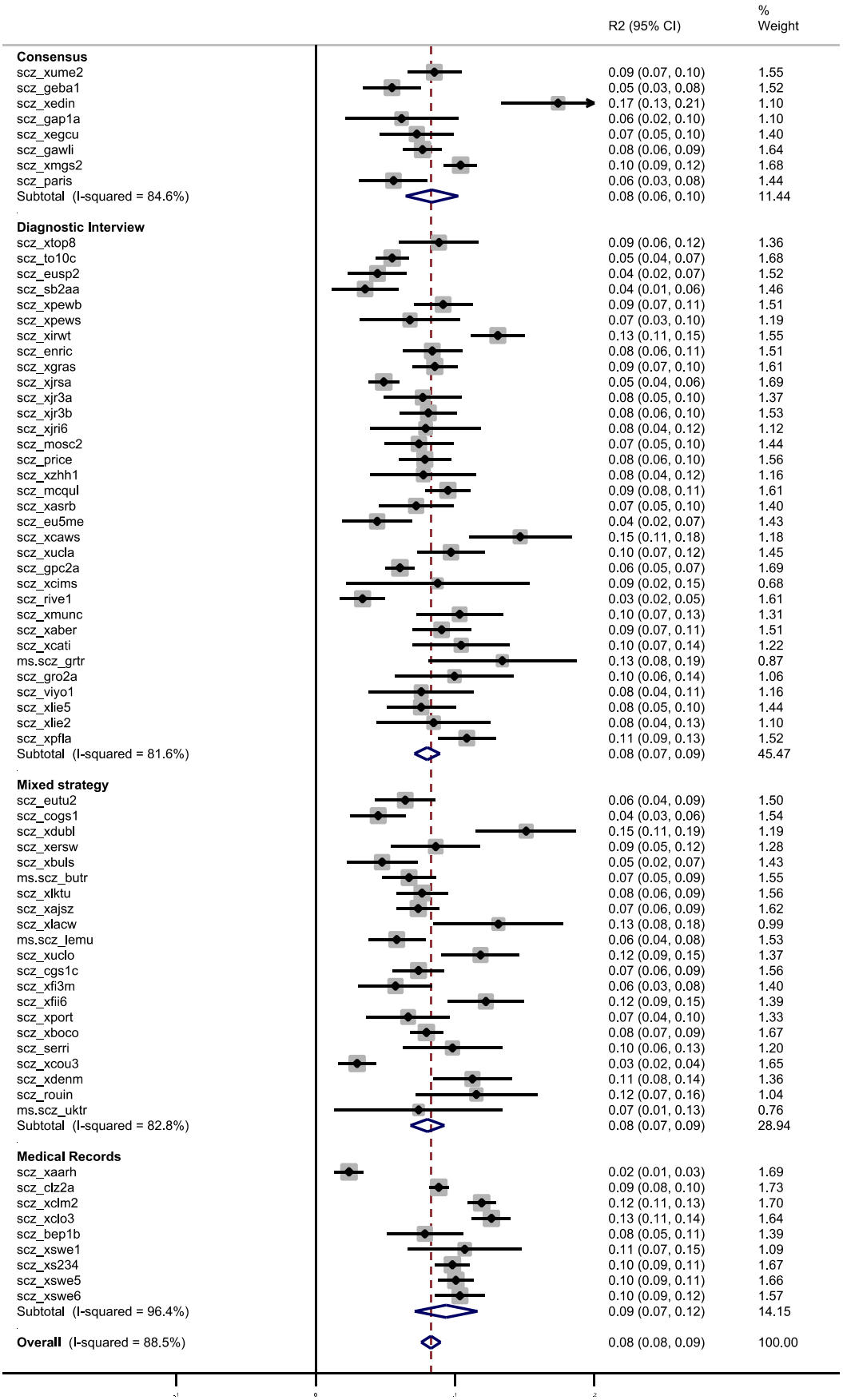
By control definition



By recruitment setting



By diagnostic strategy



Supplementary Figure 12 legend:

Forest plots of the variance explained (on the liability scale) of European ancestry samples divided according to postulated sources of heterogeneity.

Samples were sub-grouped according to the following criteria:

1) Case definition: schizophrenia (SCZ), schizophrenia or schizoaffective disorder (SCZ/Schizoaffective), schizophrenia spectrum disorder

2) Screened or unscreened controls for schizophrenia or other psychoses

3) Recruitment setting: Cases recruited from Mixed (ie Community or Hospital setting), Community only, or Hospital plus ascertainment for clozapine treatment.

4) Diagnostic strategy: Consensus diagnosis, research diagnostic interview, review of medical records, mixed strategy **(see Supplementary Note**)

For each sample, the mid-point of the box represents the point estimate of the variance explained (R2) and the horizontal line the 95% confidence intervals for that estimate. The shaded area of the box is proportional to the corresponding study weight, i.e. the cohort's contribution to the overall estimate of variance explained (R2) estimated with the DerSimonian & Laird random effects model. Heterogeneity is denoted by the I^2 statistic; measures above 75% indicating considerable heterogeneity among the samples within each subgroup. Unshaded diamonds represent the mean estimate for studies in the relevant section of the plot. The overall mean of the European sample is represented by the dashed line.

Details of all samples are given in the **Sample Supplementary note**.

Case-control sample descriptions

Adolfsson, R | Umeå, Sweden | scz_xume2_eur_sr-qc
Cases of European ancestry were ascertained from multiple different studies of schizophrenia (1992-2009). The diagnostic processes were similar between studies, and the final diagnosis is a best-estimate consensus lifetime diagnosis based on multiple sources of information such as clinical evaluation by research psychiatrists, different types of semi-structured interviews made by trained research nurses and research psychiatrists, medical records, course of the disease and data from multiple informants. Diagnosis was made in accordance with the Diagnostic and Statistical Manual of Mental Disorders-Version IV (DSM-IV)[51] or International Classification of Diseases, 10th Revision (ICD-10)[52] criteria. Controls were recruited from the Betula study, an ongoing longitudinal, prospective, population-based study from the same geographic area (North Sweden) that is studying aging, health, and cognition in adults[53]. All subjects (cases and controls) participated after giving written informed consent and the regional Ethical Review Board at the University of Umeå approved all original studies and participation in the PGC. GWAS genotyping was performed at Broad Institute.

Andreassen, O | Norway (TOP) | scz_xtop8_eur_sr-qc
Both the xtop8 and to10c samples include cases ascertained as part of the Thematically Organised Psychosis Research (TOP) study. In the TOP study, cases of European ancestry, born in Norway, were recruited from psychiatric hospitals in the Oslo region. Patients were diagnosed according to SCID and further ascertainment details have been reported[54]. Healthy control subjects were randomly selected from statistical records of persons from the same catchment area as the patient groups. All participants provided written informed consent and the human subjects protocol was approved by the Norwegian Scientific-Ethical Committee and the Norwegian Data Protection Agency.

Andreassen, O | Norway | scz_to10c_eur_sr-qc
The cases consisted of two groups; one group included participants from the TOP study. They were recruited from out-patient and in-patient psychiatric units at Oslo University Hospital and collaborating hospitals in Norway[55]. The criteria for inclusion were age between 18 and 65 years, IQ score above 70, meet the DSM-IV[51] criteria for schizophrenia, schizoaffective or schizophreniform disorder, and be willing and able to give informed consent. Diagnosis was established using the Structured Clinical Interview for DSM-IV-TR-axis I disorders[56]. The other group was included from the Therapeutic Drug Monitoring laboratory at Diakonhjemmet Hospital, Oslo. This laboratory is responsible for monitoring the majority of schizophrenia patients treated with clozapine in the region. We obtained DNA samples that were subsequently genotyped and used in the current study. The controls included blood donors and population controls from the Oslo University Hospital and healthy controls screened for psychiatric illness, all recruited from the same region. Data availability: genotype data are accessible via the THL Biobank. The access procedure is described at the following link: https://www.thl.fi/en/web/thl-biobank/for-researchers.

Arango, C | Spain | scz_eusp2_eur_sr-qc
This sample is part of the EUGEI and has been described elsewhere[57,58].

Arango, C | Spain | scz_celso_eur_sr-qc
Participants were recruited as part of the sample collection of CIBERSAM (Network Biomedical Research Centre in Mental Health), an institution based in Spain that recruits in-patients from psychiatric units at eight different hospitals in Spain. All patients recruited met the DSM-IV[51] diagnostic criteria for schizophrenia, schizoaffective or schizophreniform disorder. Healthy controls

were recruited after being screened for psychiatric illness. All participants provided written informed consent. Blood sample recruitment and preparation, participation in the PGC consortium and commitment to share the generated data in subsequent meta-analyses and further secondary proposals were approved by the different ethical committees at the hospitals involved in the recruitment.

Atbaşoğlu, EC; Saka, M | Turkey | scz_eutu2_eur_sr-qc
We recruited participants as part of the EUGEI (http://www.eu-gei.eu), a study based in the EU that recruits from schizophrenia patients. In this work package the patients came from the Netherlands, Turkey, Spain and Serbia. The sample was Caucasian ethnic origin. This sample has been described elsewhere[59,60]. Participants were interviewed by psychologists or trained research assistants. Inter-rater reliability was measured centrally by all interviewers rating videos and were found highly satisfactory. Participants were included if they met schizophrenia diagnosis and were excluded if they had an IQ <70 or had a CNS disease that would significantly affect their cognitive function. We recruited control participants from the population. Controls did not have psychosis, or a sibling with psychosis, but other psychiatric diagnosis were not excluded. All participants provided written informed consent. Participants were assessed for capacity to provide informed consent by their clinical team, and consent was also received from their caregiver or closest relative. The study had ethics approval granted by each centre's local ethics board, which permit inclusion of the data in meta-analyses. Blood was taken and samples were genotyped by Cardiff University. The study was funded by EU and Ankara University.

Baune, B | Germany | scz_geba1_eur_sr-qc
We recruited participants as part of the clinical research unit in schizophrenia at two study sites following the same study inclusion and exclusion criteria: The University of Münster and University of Lübeck, Germany, without any ancestry restriction in recruitment. Participants were interviewed by board accredited psychiatrists. Trained raters reviewed this interview, along with available clinical records, to determine a consensus lifetime DSM-IV[51] diagnosis. Participants were included if they met schizophrenia criteria and were excluded if a non-primary diagnosis of schizophrenia existed. All participants provided written informed consent prior to the conduct of any study procedures. Participants were assessed for capacity to provide informed consent by their clinical team. The study had ethics approval granted by the local ethics committee of the University of Münster and University of Lübeck.

Belangero, S | Brazil | scz_sb2aa_eur_sr-qc
We recruited participants as part of the "Programa de Esquizofrenia" (Schizophrenia Program) and "Primeiro Episódio Psicótico" (First Episode of Psychosis Program) both based at Universidade Federal de Sao Paulo (UNIFESP), Sao Paulo, Brazil, which recruits outpatients and inpatients, respectively. There was no ancestry restriction in recruitment. This sample has been described elsewhere[61,62]. Participants were interviewed using the Structured Clinical Interview of the DSM-IV (SCID-I)[56] by trained psychiatrists. Participants were included if they met criteria for schizophrenia, schizophreniform disorder, brief psychotic disorder or psychotic disorders not otherwise specified diagnoses and were excluded if they met criteria for bipolar disorder, substance-induced psychotic disorder or major depressive disorder diagnoses. We recruited control participants from volunteers, blood donors, UNIFESP employees or their relatives. They were interviewed with SCID-I and were excluded if they met current or previous psychiatric diagnoses or if they have a first-degree family history of psychotic disorders. The study had ethics approval granted by Research Ethics Committee of UNIFESP. Blood was collected from all participants.

Børglum, A | Denmark | scz_xaarh_eur_sr-qc

DNA samples for all subjects were collected from blood spots systematically collected by the Danish Newborn Screening Biobank), with case/control status established using the Danish Psychiatric Central Register. Cases were diagnosed clinically according to ICD-10 criteria. Controls were selected to match the cases by birth cohort. The Danish Data Protection Agency and the ethics committees in Denmark approved the human subjects protocol. Data availability: data access may be arranged through iPSYCH protocols: https://ipsych.dk/en/research/.

Braff, D | USA | scz_cogs1_eur_sr-qc
We recruited participants as part of the Consortium on the Genetics of Schizophrenia (COGS), a multi-site study based in the USA that recruited outpatients with schizophrenia and healthy control subjects[63]. Outpatient participants included patients with schizophrenia or schizoaffective disorder depressed type and were recruited from various sources including clinician referrals, outpatient facilities such as board and cares and clubhouses, local National Alliance for the Mentally Ill chapters, and advertising via the media. Diagnoses were established via the Diagnostic Interview for Genetics Studies (DIGS)[64], the Family Interview for Genetic Studies (FIGS)[65], and a Best-Estimate Final Diagnosis (BEFD) procedure based on Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)[51] criteria for schizophrenia. Other patients were interviewed using the Structural Clinical Interview for DSM-IV (SCID)[56] with additional items from the DIGS and FIGS by trained research assistants. The clinical team reviewed this interview in weekly clinical consensus meetings along with available clinical records to determine a consensus DSM-IV[51] diagnosis. Participants were included if they met diagnosis and were excluded if they had severe systemic illness that interferes with ability to be endophenotyped, significant head injury with loss of consciousness of 15 minutes or more, neurological disorder, positive illicit drug or alcohol screen, met the diagnosis of alcohol or substance abuse in the past month or alcohol or substance dependence in the last 6 months, or had an estimated premorbid IQ < 70 per Wide Range Achievement Test. We recruited healthy control participants from the community via advertising media. Healthy subjects were screened for psychiatric disorders and excluded for any history of psychosis in themselves or a family member. Additionally, healthy control subjects were excluded if they met any of the exclusion criteria listed above for the patients. The study had ethics approval granted by the local Institutional Review Board at each site.

Bramon | Seven countries (PEIC, WTCCC2) | scz_xpews_eur_sr-qc

Bramon | Spain (PEIC, WTCCC2) | scz_xpewb_eur_sr-qc

The Psychosis Endophenotypes International Consortium (PEIC) was part of WTCCC2. Samples were collected through seven centres in Europe and Australia (the Institute of Psychiatry, King's College London, London; GROUP (consisting of the University of Amsterdam, Amsterdam; the University of Groningen, Groningen; Maastricht University Medical Centre, Maastricht; and the University of Utrecht, Utrecht); the University of Western Australia, Perth; the Universidad de Cantabria, Santander; the University of Edinburgh, Edinburgh; Heidelberg University, Heidelberg and Ludwig-Maximilians-Universität München, Munich). To allow for a DSM-IV diagnosis to be ascertained or ruled out, all participants (including controls and unaffected family members) underwent a structured clinical interview with the Schedule for Affective Disorders and Schizophrenia (SADS), the Structured Clinical Interview for DSM Disorders (SCID), or the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). We included cases with schizophrenia and schizoaffective disorder. Participants in all groups were excluded if they had a history of neurological disease or head injury resulting in loss of consciousness. Data availability: Genotype data for both of these cohorts are accessible via the Wellcome Trust Case-Control Consortium (WTCCC2). https://www.wtccc.org.uk/ccc2/wtccc2_studies.html.

Buxbaum, J | New York, US & Israel | scz_xmsaf_eur_sr-qc
Samples contributed by Mount Sinai were derived from three cohorts. In all cohorts, ethical approval was obtained from all participating sites, and all subjects provided informed consent. Two of the cohorts were in a prior paper on copy number variation[66]. One of the cohorts was from the Mount Sinai brain bank, where DNA was extracted from post-mortem samples, and another comprised of patients ascertained in Israel. The third cohort included subjects more recently recruited through the Mount Sinai Conte Center.

Campion, D; Laurent-Levinson, C | France | scz_rouin_eur_sr-qc
The cohort included two independent French schizophrenia cohorts. Firstly, case participants were unrelated Caucasian in- or out-patients who were recruited for the Etude de l'hyperprolémie chez les patients avec schizophrénie (Study of hyperprolinemia in schizophrenic patients), in the Normandy region of France. This sample has been described previously[67]. Cases were interviewed using the Positive and Negative Schizophrenia Scale (PANSS)[68] and appropriate sections of the Schedule for Affective Disorders and Schizophrenia[3], by licensed psychiatrists. Final DSM-III-R[69] diagnoses were assigned by a group of trained psychiatrists based on the interview and available clinical records. Individuals with alcohol abuse were excluded.
We recruited control participants mainly from staff members. Prospective control subjects with a history of psychiatric or metabolic disorder or with first-degree relatives with history of psychiatric disorder were not included. All controls were drug-free, except oral contraceptives in women. All partici-pants provided written informed consent. Participants were assessed for capacity to provide informed consent by their clinical team. The study had ethics approval granted by local IRB ethic committee of Rouen (Comité Consultatif pour la Protection des Personnes se prêtant à la Recherche Biomédicale (CCPPRB) which permit inclusion of the data in meta-analyses. Blood samples were taken for DNA extraction.
A subset of the Rouen-Pitié cohort consisted of 24 cases recruited separately in a study of childhood-onset schizophrenia in the Paris region, the "Etude de protéines candidates par techniques d'immuno-affinité et de spectrométrie de masse lors d'un premier épisode psychotique" (Study of candidate proteins by immunoaffinity and mass spectrometry techniques during a first psychotic episode). Cases were recruited from in-patient and outpatient settings. There was no ancestry restriction, but most were Caucasian; ancestry outliers were excluded during quality control procedures. Cases with age at onset between 7-17 years were interviewed using the French version of the DIGS 2.0[64] by licensed psychiatrists. Final DSM-IV[51] diagnoses were assigned by a group of trained psychiatrists based on the interview and available clinical records. Participants were included in the genotyping cohort if the final DSM-IV[51] diagnosis was schizophrenia or schizoaffective disorder depressed type, and the patient and family agreed to provide a blood specimen for genetic studies. Exclusion criteria included psychotic disorders judged to have been caused by somatic pathologies, drug treatments, or the abuse of alcohol or drugs, moderate or severe intellectual disability (IQ<50), diagnosis of bipolar disorder, or patients under guardianship or curatorship. The study was approved by the relevant ethics committee (CPP IdF6). The approval permits inclusion of the data in meta-analyses.

Cervilla, M | Granada | scz_sanch_eur_sr-qc
We recruited patients as part of the GENIMS and PISMA studies. GENIMS was a cross-sectional clinical study in which participating patients were consecutive attendees to psychiatric outpatient clinics[70]. All were in a stable stage of their disorder and on antipsychotic medication. Patients were all diagnosed by fully trained psychiatrists using the Structured Clinical Interview for DSM-IV Axis I disorder (SCID-I)[56]. Trained raters reviewed these interviews along with available clinical records to determine a consensus lifetime DSM-IV[51] diagnosis of schizophrenia. Additional assessments included sociodemographic and clinical variables such as sex, age, educational level, employment,

marital status, and years after onset. To estimate each participant's premorbid intelligence quotient (IQ), a Spanish version of the Barona index was used. The Spanish version of the PANSS[68] was used to measure psychopathology, since PANSS is the standard scale valid and reliable for this purpose. Global functioning was assessed using the GAF. Inclusion criteria were as follows: 1) meet DSM-IV[4] diagnostic criteria for SZ; 2) be older than 18 years; and 3) agree to participate. Exclusion criteria were as follows: 1) mental retardation and 2) any type of dementia. We recruited control participants from the PISMA study, which has been reported elsewhere[71]. This was a cross-sectional study targeting a large representative stratified sample of community-dwelling Andalusian adults between 18 and 75 years of age. All provinces in the Andalusian community were included. Participants were administered the MINI[72] by trained psychologists, which generated both DSM-IV[51] and ICD-10[52] diagnoses. A saliva sample was obtained from each participant.

Corvin, A | Ireland | scz_xdubl_eur_sr-qc
The case sample was collected primarily in the Dublin area and the ascertainment procedure has been previously described[73]. The controls were recruited, from the same region through the Irish Blood Transfusion Services. All participants gave written, informed consent and the collections were approved through the Federated Dublin Hospitals and Irish Blood Transfusion Services Research Ethics Committees, respectively. DNA samples were genotyped at the Broad Institute.

Corvin, A; Morris, D | Ireland | scz_du2aa_eur_sr-qc
This sample has been described in a previous publication[74].

Corvin, A; Riley, B | Ireland (WTCCC2) | scz_xirwt_eur_sr-qc
The case sample was recruited from the Republic of Ireland and Northern Ireland. All cases had four Irish grandparents and ascertainment details have been reported elsewhere[75]. Ethics approval was obtained from all participating hospitals and centres. Controls were blood donors from the Irish Blood Transfusion Service, whose Ethics Committee approved the human subjects protocol. All participants gave written informed consent. Samples were genotyped at Affymetrix (Santa Clara, California, US) laboratory as part of the WTCCC2 genotyping pipeline.

Di Forti, M | London, UK | scz_gap1a_eur_sr-qc
Participants were part of the Genetic and Psychotic Disorders Study case-control project that approached all patients aged 18 to 65 years who presented with their first episode of psychosis to the Lambeth, Southwark, and Croydon adult inpatient units of the South London and Maudsley National Health Service Foundation Trust between December 2005 and October 2010[76]. Patients who met ICD-10 criteria for a diagnosis of nonorganic psychosis (F20 – F29 and F30–F33)[52], validated by administering the Schedules for Clinical Assessment in Neuropsychiatry (SCAN)[77], were invited to participate in the study. DNA was obtained from all participants that completed the SCAN. Seventy-five percent of DNA samples used originated from blood and 25% from cheek swabs.

Domenici, E | Italy | scz_enric_eur_sr-qc
The sample included schizophrenia patients and healthy volunteers of Caucasian ancestry for at least two generations, from the Lombardy and Apulia regions of Italy. All subjects provided written informed consent before entering the study according to the institutional guidelines of local Ethical Committees. Patients admitted to the Unit of Psychiatry of Brescia IRCCS Fatebenefratelli and to the Unit of Psychiatry of Bari "Ospedale Policlinico Consorziale" were enrolled if they had a DSM-IV-TR[78] diagnosis of schizophrenia. Diagnosis was made using the Structured Clinical Interview for the DSM-IV Axis I disorders[56], which was administered by expert psychiatrists. Control subjects were randomly recruited from different sources (hospital visitors, cultural and elderly associations, trade unions) and using different strategies (word of mouth, newspaper advertising,

internet and social networks). Subjects enrolled at Brescia were screened for DSM-IV Axis I disorders[51] by expert psychologists using the Mini-International Neuropsychiatric Interview[72]. Only healthy volunteers without a history of drug or alcohol abuse or dependence and without a personal or first-degree family history of psychiatric disorders were enrolled in the study. Subjects who obtained a score lower than 27/30 in the Mini Mental State Examination were excluded as well. Control subjects enrolled at Bari were screened using the Non-Patient Structured Clinical Interview for DSM-IV[79] to ensure they were unaffected by any psychiatric condition. In addition, subjects were screened using the Family Interview for Genetic Studies[65] to ensure absence of psychotic disorders in their first-degree relatives. Both patients and control subjects were excluded if they had: a significant history of drug or alcohol abuse; active drug abuse in the previous year; or experienced a head trauma with a loss of consciousness. Part of the cohort was described in previous publications[80–83].

Ehrenreich, H | Germany (GRAS) | scz_xgras_eur_sr-qc
The Gottingen Research Association for Schizophrenia (GRAS) collection included cases recruited across 23 German hospitals. Controls were unscreened blood donors recruited at the Georg-August-University according to national blood donation guidelines[84]. Cases completed a structured clinical interview and were diagnosed with DSM-IV schizophrenia or schizoaffective disorder. The study was approved by the Georg-August-University ethics committee and local internal review boards of the participating centres. All participants gave written informed consent.

Esko, T | Estonia (EGCUT) | scz_xegcu_eur_sr-qc
The Estonian cohort comes from the population-based biobank of the Estonian Genome Project of University of Tartu (EGCUT)[85]. The project was conducted according to the Estonian Research Act and all participants provided informed consent (www.biobank.ee). In total, 52,000 individuals aged 18 years or older participated in this cohort (33% men, 67% women). The population distributions of the cohort reflect those of the Estonian population (83% Estonians, 14% Russians and 3% other). General practitioners (GP) and physicians in the hospitals randomly recruited the participants. A Computer-Assisted Personal interview was conducted over 1-2 ours at doctors' offices. Data on demographics, genealogy, educational and occupational history, lifestyle and anthropometric and physiological data were assessed. Schizophrenia was diagnosed prior to the recruitment by a psychiatrist according to ICD-10[52] criteria and identified from the Estonian Biobank phenotype database. Controls were drawn from a larger pool of genotyped biobank samples by matching on gender, age and genetic ancestry. All the controls were population-based and have not been sampled for any specific disease.

Esko, T; Li, Q; Malhotra D | &J and Roche cases, EGCUT controls | scz_xjr3a_eur_sr-qc
Esko, T; Li, Q; Malhotra D | J&J and Roche cases, EGCUT controls | scz_xjr3b_eur_sr-qc
Esko, T; Li, Q; Malhotra D | J&J and Roche cases, EGCUT controls | scz_xjri6_eur_sr-qc
Esko, T; Li, Q; Malhotra D | J&J and Roche cases, EGCUT controls | scz_xjrsa_eur_sr-qc
Cases were collected by Johnson and Johnson (J&J) and Roche as part of clinical collaborations with hospitals and outpatient centres. Cases were diagnosed according to DSMIV criteria, with medical record review by a trained psychiatrist. There were reliability trials across centres for the J&J studies. The J& J cases were mostly collected in Eastern Europe, with most coming from Estonian and Russia (>100); intermediate numbers from Austria, the Czech Republic, Latvia, Lithuania, and Spain (50-100); and smaller collections from Bulgaria, Hungary, and Poland (<50). The Roche cases were assessed with a structured psychiatric assessment by trained interviewers. Most of the Eastern European controls were from the Estonian Biobank project (EGCUT)[85] and were ancestrally matched with cases from the J&J sample. Data availability: genotype data are accessible through the Estonian Biobank. The procedure for accessing them is described in detail here: https://genomics.ut.ee/en/access-biobank.

Gareeva, A; Khusnutdinova, E | Ufa, Russia | scz_price_eur_sr-qc
We recruited participants as part of "GWAS analysis in Russian SZ sample from the Volga Ural region", a study based in the Volga-Ural region of Russia that recruits in patients from Republican Mental Hospital of Public Health Ministry of Bashkiria (RMH PHMB), Ufa, Russia. The sample is comprised of individuals of Caucasian ancestry aged 15-74 years old. All patients met the ICD10[52] criteria for paranoid schizophrenia (ICD-10 code F20.xx: F20.006, F20.004, F20.016, F20.014). For each patient, through an individual survey and analysis of the medical history, a specially designed questionnaire was filled out, in which passport data of patients, complaints, clinical and anamnestic data, and all ongoing general clinical and special research methods were entered. All participants were interviewed by trained psychiatrists and trained research assistants and the diagnosis was made by psychiatrists of RMH PHMB. The control group consisted of mentally healthy unrelated volunteers, matched for age, gender, and ethnicity. All controls had no family history of neuropsychiatric disorders. For all individuals, key phenotypic information has been collected including information about sex, age, ethnicity, age at onset and family history of psychiatric disorders. The study has been approved by the local bioethical committee of the Institute of Biochemistry and Genetics of Ufa Federal Research Center of the Russian Academy of Sciences (IBG UFRC RAS). Peripheral blood was taken from all participants of the study. DNA was extracted from peripheral blood by the phenol and chloroform method.

Gawlik, M | Germany | scz_gawli_eur_sr-qc
We recruited participants as part of "Molekulargenetische Untersuchungen bei phasischen und schizophrenen Psychosen" (Genetics of Psychoses), a study based in Germany that recruits from the Department of Psychiatry, Psychosomatics and Psychotherapy, University Hospital Wuerzburg. Participants were interviewed by trained psychiatrists. Trained raters, Prof Gerald Stöber, Dr Micha Gawlik, and Prof Bruno Pfuhlmann reviewed participants along with available clinical records, to determine a consensus lifetime DSM-IV[51] diagnosis. Participants were included if they met diagnoses as stated and were excluded with any other psychiatric or neurological illness. We recruited control participants from healthy blood donors. All participants or legal representatives provided written informed consent. Participants were assessed for capacity to provide informed consent by their clinical team. The study had ethics approval granted by local ethics board. Blood samples were taken for DNA.

Gejman, P | US, Australia (MGS) | scz_xmgs2_eur_sr-qc
European ancestry case samples were collected by the Molecular Genetics of Schizophrenia (MGS) collaboration across multiple sites in the USA and Australia as described in detail elsewhere[86]. Cases gave written informed consent, and IRBs at each collecting site approved the human subjects protocol. A survey company (Knowledge Networks, under MGS guidance) collected the European ancestry control sample and ascertainment is described in detail elsewhere[87]. DNA samples were genotyped at the Broad Institute.

Golimbet, V | Moscow | scz_mosc2_eur_sr-qc
Information on this sample has been published elsewhere[88]. Briefly, the sample comprises cases with schizophrenia and age matched controls. Phenotypic information including age, gender, ethnicity, age at onset, PANSS[68], Hamilton depression scale scores, personality traits scores, and cognitive tests results were collected.

Jönsson, E | Sweden (Hubin) | scz_xersw_eur_sr-qc
Cases were recruited from north-western Stockholm County and ascertainment has been described previously[89]. Cases gave informed consent and the human subjects protocol was approved by the ethical committees of the Karolinska Hospital and the Stockholm Regional Ethical Committee. Controls were recruited either among subjects previously participating in biological research at the

Karolinska Institute or drawn from a representative register of the population of Stockholm County. All participants provided informed consent.

Kirov, G | Bulgaria | scz_xbuls_eur_sr-qc
All cases were recruited from Bulgaria and had a history of hospitalization for treatment of schizophrenia. Controls were recruited from the two largest cities in Bulgaria as previously described[73]. All participants gave written informed consent and the study was approved by local ethics committees at the participating centres.

Kirov, G; Owen, M | Bulgaria | ms.scz_xuktr_eur_sr-qc
All cases and parents were recruited from UK and had a history of hospitalization for treatment of schizophrenia. Diagnosis was confirmed following a SCAN[90] interview and review of case notes followed by consensus diagnosis according to DSM-IV[51] criteria. The samples were genotyped at the Broad Institute. All participants gave written informed consent and the study was approved by local ethics committees at the participating centres. The samples were genotyped at the Broad Institute.

Kirov, G; Owen M | Bulgaria | ms.scz_xbutr_eur_sr-qc
Families from Bulgaria were recruited if a proband had schizophrenia or schizoaffective disorder, both parents were available, and all members of the trio agreed to participate in the study. Recruitment took place between 1999 and 2004 in several psychiatric hospitals in Bulgaria. Ethical Committee approval was obtained from each of these hospitals. All probands and all parents received an Information Sheet and signed Informed Consent Forms. All participants had attended mainstream schools, which at the time in Bulgaria, excluded people with mental retardation. Probands were either in- or out-patients at the time of the study but each had a history of hospitalization. A team of psychiatrists was trained in using the rating scales and methods of the study. We used the SCAN instrument to perform an interview for psychotic and mood symptoms. This instrument has been translated into Bulgarian and validated by one of its authors (A. Jablensky). Consensus diagnoses were made according to DSM-IV[51] criteria on the basis of an interview and inspection of hospital notes by two clinicians. If consensus was not attained, the patient was re-interviewed by a research interview trained clinician and was excluded if consensus could still not be reached. In addition, approximately 23% of the sample was selected at random and re-interviewed by a research interview trained clinician. Hospital notes were also collected for affected relatives in order to confirm diagnoses.

Kennedy JL, Collier DA | Canada, US(Lilly), US (MIGen)| scz_xlktu_eur_sr-qc
Toronto cases were recruited by referral and advertisement. Diagnoses were made according to DSM-III or DSM-IV criteria following interview and medical record review. US cases were recruited from schizophrenia clinical trials in a range of settings as part of a trial with Eli Lilly. Diagnoses were made according to DSM-III or DSM-IV criteria following interview by psychiatrist and medical record review. No controls were sampled as part of the study, and ancestrally-matched controls were chosen from the Myocardial Infarction Genetics Consortium (MIGen, dbGaP ID phs000294.v1.p1) that was genotyped with the same SNP array[91].

Krebs, M | France | scz_paris_eur_sr-qc
We recruited participants as part of the PSYDEV collection ("Etude familiale et génétique des aspects développementaux des maladies psychiatriques"), a study based in Paris (France) that recruits both inpatients and outpatients as well as controls. The samples included in the analysis were limited to Caucasian individuals. Participants were interviewed by board accredited psychiatrists. Trained raters reviewed this interview, along with available clinical records, to determine a consensus lifetime DSM-IV[51] diagnosis. Participants were included if they met criteria

for chronic schizophrenia. Control participants were volunteers recruited from advertisement in paramedical schools. They were screened for absence of any psychiatric illnesses by trained psychiatrists and psychologists. The screening was declarative by face-to-face medical interview. Participants were assessed for capacity to provide informed consent by the clinical team. All participants provided written informed consent. Oral consent was asked for individuals under the age of 18 or under guardianship but the written consent was obtained from their legal representatives. The study had ethics approval granted by CPP Ile de France IV which permit inclusion of the data in meta-analyses. Either blood or saliva have been taken.

Lencz, T; Darvasi A | Israel | scz_xajsz_eur_sr-qc
Cases and controls were sampled from an Ashkenazi Jewish repository (Hebrew University Genetic Resource, http://hugr.huji.ac.il). Patients were recruited from hospitalized inpatients at 7 medical centres in Israel and were diagnosed with DSM-IV schizophrenia or schizoaffective disorder. Controls were sampled through the Israeli Blood Bank and did not report any chronic disease or regularly prescribed medication at the time of assessment. Full ascertainment details have previously been reported[92]. Local ethics committees and the National Genetic Committee of the Israeli Ministry of Health approved the studies and all participants gave informed, written consent.

Levinson, D | 22885689 | Six countries, WTCCC controls | scz_xlacw_eur_sr-qc
Cases collected as part of a larger pedigree-based study[93] were partitioned into two subsamples. Cases with two genotyped parents were analysed as trios (see PI Levinson, ms.scz_xlemu_eur_sr-qc below). Unrelated cases who could not be used as part of a trio were included as a separate case-control analysis, using independent controls, matched by ancestry and genotyping array, from the Wellcome Trust Case Control Consortium[94]. Cases were identified from different clinical settings (e.g. inpatients, outpatients and community facilities) in six countries (Australia, France, Germany, Ireland, UK, and the US). Diagnoses were established using semi-structured interviews, psychiatric records and informant reports. Case subjects were diagnosed with schizophrenia or schizoaffective disorder according to DSM-III-R criteria. All protocols were approved by loci IRBs, and all cases provided written informed consent.

Levinson, D | Six countries | ms.scz_xlemu_eur_sr-qc
Schizophrenia cases were included from the family sample of European-ancestry pedigrees described by Levinson et al.[93] Participants and their families In this trio study, probands were ascertained and recruited from different clinical settings (e.g. inpatients, outpatients and community facilities) in six countries (Australia, France, Germany, Ireland, UK, and the US)[93]. (Unrelated individuals were includes as part of a case-control design, see Levinson, D, scz_xlacw_eur_sr-qc above.) Diagnoses were established using semi-structured interviews, psychiatric records and informant reports. Case probands were diagnosed with schizophrenia or schizoaffective disorder according to DSM-III-R criteria[69]. The trio-based analysis included families where there was at least one affected proband and two available parents. Each affected sibling in such families was included, with the parents, as an independent trio. All protocols were approved by loci IRBs, and all cases provided written informed consent.

Malhotra, A | New York, US | scz_xzhh1_eur_sr-qc
The case and control subjects were recruited in the New York metropolitan area and ascertainment methods have been described previously[95]. All participants gave written, informed consent and the IRB of the North Shore-Long Island Jewish Health System approved the human subjects protocols. DNA was genotyped at Zucker Hillside.

McIntosh, A | Edinburgh, UK | scz_xedin_eur_sr-qc

Cases and controls were recruited from the southeast of Scotland, and ascertainment has been previously described as part of the International Schizophrenia Consortium studies[73]. All participating subjects gave written, informed consent and the human subjects protocol was approved by the Scotland A Research Ethics Committee. DNA samples were genotyped at the Broad Institute. Ex PI D Blackwood

McQuillin, A | United Kingdom| scz_mcqul_eur_sr-qc
We recruited participants as part of the DNA Polymorphisms in Mental Illness (DPIM) study based in the United Kingdom that recruits both inpatient and outpatient research participants. Recruitment of participants was restricted to people of UK and Irish ancestry. Participants were required to have three grand-parents of UK or Irish ancestry with the fourth grandparent of western European origin. This sample has been described elsewhere[96]. Participants with an ICD10[52] diagnosis of schizophrenia were interviewed by trained research assistants using the Schedule for Affective Disorders and Schizophrenia-Lifetime Version (SADS-L)[97] to confirm the diagnosis according to Research Diagnostic Criteria (RDC)[98]. Research interviews were supplemented with information from clinical records. Case participants were also rated with the 90-item Operational Criteria Checklist (OPCRIT)[99]. Recruitment of control participants was also from recruitment centres across the UK. Control subjects were interviewed by trained research assistants with the SADS-L[97] to exclude a personal history of mental illness. They were also excluded if they reported a family history of mental illness or reported personal substance misuse including harmful alcohol consumption. All participants provided written informed consent. The control sample was supplemented with DNA from 480 ECACC Human Random Control (HRC) samples obtained from Public Health England. The donors for this sample were all healthy blood donors. The study had ethics approval granted by the UK National Health Service Metropolitan Multi-centre Research Ethics Committee. Genomic DNA was obtained from blood or saliva samples.

McQuillin, A | London, UK | scz_xuclo_eur_sr-qc
All cases and controls were collected by University College London and had both parents from England, Scotland or Wales. All participants gave written informed consent and the U.K. National Health Service multicentre and local research ethics committee approved the human subjects protocol. Further details on ascertainment are available elsewhere[73]. The samples were genotyped at the Broad Institute. Data availability: genotype data for this study are available as a part of the Schizophrenia 20.21 resource at the NIMH Repository & Genomics Resource: https://www.nimhgenetics.org/download-tool/SZ.

Menezes, P; Belangero, S | Brazil | scz_braz2_eur_sr-qc

This sample is part of the EUGEI and has been described elsewhere[58].

Mowry, B; Morgan, V | Australia | scz_xasrb_eur_sr-qc
These subjects were part of the Australian Schizophrenia Research Bank. The case sample was recruited in five Australian States (New South Wales, Queensland, Western Australia, South Australia and Victoria) through hospital inpatient units, community mental health services, outpatient clinics and non-government mental illness support organizations as part of the Survey of High Impact Psychosis (SHIP) study, as described elsewhere[100] as well as, in the initial stages, through a large-scale, national, multi-media advertising campaign. Briefly, the SHIP study used a two-phase design to recruit a representative sample of 1825 participants aged 18-64 years who were in contact with mental health services across Australia. Participants were screened for psychosis, and a random subset of those who screened positively for psychosis were invited to interview. Data on symptoms, substance use, functioning, service utilization, medication use, education, employment, housing, and physical health were collected, and a diagnosis was made using ICD-10 criteria[52]. This sample is comprised of 509 cases from larger metropolitan centres of

Brisbane, Newcastle, Sydney, Melbourne, Perth and Adelaide. Cases gave written informed consent, and the human subjects protocol was initially approved by the Hunter New England Area Health Research Committee and subsequently approved by relevant Institutional Ethics Committees in Brisbane, Sydney, Melbourne, Perth and Adelaide. Healthy controls were recruited through multi-media advertisements, and other sources. Controls were from the metropolitan centres of Brisbane, Newcastle, Sydney, Melbourne, and Perth. Controls gave written informed consent, and the human subjects protocol was approved by the Hunter New England Area Health Research Committee and Institutional Ethics Committees in Brisbane, Sydney, Melbourne and Perth. The samples were genotyped in two stages at the Hunter Medical Research Institute, University of Newcastle, Newcastle, Australia.

Nimgaonkar, V | USA | scz_viyo1_eur_sr-qc
The inclusion criteria for cases were either a diagnosis of schizophrenia or schizoaffective disorder or, at Baltimore, schizophreniform disorder, according to DSM-IV criteria[51]. The cases in Baltimore were recruited from inpatient and day hospital programs of Sheppard Pratt Hospital and from affiliated psychiatric rehabilitation programs. The cases in Pittsburgh were recruited from Western Psychiatric Institute and Clinic, Pittsburgh and additional psychiatric treatment facilities in a 500-mile radius of Pittsburgh[101]. Patients referred by their therapists were evaluated using structured diagnostic instruments (Structured Clinical Interview for DSM-IV Axis 1 Disorders[56] for Baltimore patients, Diagnostic Interview for Genetic Studies[64] for Pittsburgh patients). This information was synthesized with available information from the patients' medical records (both sites) and additional details from informants, usually relatives (Pittsburgh). The clinical information was synthesized and presented to board certified psychiatrists / psychologists, followed by consensus diagnosis based on DSM-IV[51] criteria. The control group was recruited from posted announcements at local health care facilities and universities in the same geographic area and settings where the schizophrenia participants were recruited. At Baltimore, the primary inclusion criterion for the control individuals was the absence of a current or past psychiatric disorder. The diagnosis of each control participant was made by a board-certified psychiatrist based on the Structured Clinical Interview for Diagnosis for Axis I disorders[56]. At Pittsburgh, the control individuals were assessed in the same manner as the cases and were screened for absence of psychosis. Participants in both groups fulfilled following additional criteria: age 20-60 years (Baltimore); proficient in English; absence of any history of intravenous substance abuse; absence of mental retardation; absence of self-reported HIV infection; absence of serious medical disorder that would confound a diagnosis of schizophrenia; absence of a primary diagnosis of alcohol or substance use disorder. All participants provided written informed consent and the study was approved by the Institutional Review Boards of Sheppard Pratt, the University of Pittsburgh School of Medicine, and the Johns Hopkins School of Medicine following established guidelines.

O'Donovan, M; Owen, M | Cardiff, UK | scz_xcaws_eur_sr-qc
The case sample included European ancestry schizophrenia cases recruited in the British Isles and described previously[102]. All cases gave written informed consent to. The study was approved by the Multicentre Research Ethics Committee in Wales and Local Research Ethics Committees from all participating sites. The control sample used the Wellcome Trust CaseControl Consortium (WTCCC) sample described elsewhere[94], but included similar numbers of individuals from the 1958 British Birth Cohort and a panel of consenting blood donors (UK Blood Service). Samples were genotyped at Affymetrix service lab (San Francisco, USA).

Ophoff, R | Netherlands | scz_xucla_eur_sr-qc
The case sample consisted of inpatients and outpatients recruited through psychiatric hospitals and institutions throughout the Netherlands. Cases with DSM-IV schizophrenia were included in the analysis. Further details on ascertainment are provided elsewhere[89]. Controls came from the University Medical Centre Utrecht and were volunteers with no psychiatric history. Ethical

approval was provided by local ethics committees and all participants gave written informed consent.

Paciga, S | Pfizer | Multiple countries | scz_xpfla_eur_sr-qc
Pfizer contributed anonymized individual genotypes for cases from seven multi-centre randomized, double-blind efficacy and safety clinical trials (A1281063, A1281134, A1281148, A245-102, NRA7500001, NRA7500002, NRA7500003, and NRA7500004) as well as a set of purchased samples (NRA9000099). Also included in this cohort were a sample non-Irish non-trio probands from the Six Countries cohort, who were well-matched with the Pfizer cohort for genotypic ancestry and genotyping platform. Trial samples were collected for antipsychotic medications across outpatient and inpatient treatment settings. All participating cases had a diagnosis of schizophrenia and were assessed using a structural clinical interview by trained interviewers, with systematic procedures to quality-control diagnostic accuracy and reliability trials across participating sites in the United States and internationally. Purchased blood samples were obtained from PrecisionMed International by Pharmacia and Upjohn Corporation, and were collected from diagnosed subjects with schizophrenia and schizoaffective disorder. All studies were reviewed by both central and local institutional review boards, depending on the study site, before recruitment of subjects started. Protocol amendments were approved while the study was in progress and before the data were unblinded. The studies were conducted in conformity with the U.S. Food and Drug Administration Code of Federal Regulations (21CFR, Part 50) and the Declaration of Helsinki and its amendments and were consistent with Good Clinical Practice and the applicable regulatory requirements. Participants provided written informed consent before enrolment. An optional blood sample was collected from clinical trial subjects for pharmacogenetic analysis to investigate potential associations between genetic variant drug response and general characteristics of schizophrenia and related disorders. Sample collection was not required for participation in the original clinical trials. The controls (A9011027) were recruited in a multi-site, cross-sectional, non-treatment prospective trial to collect data, including DNA, from cognitive normal and free of psychiatric diseases elderly subjects in the US. Subjects were specifically recruited to match the gender, age, and ethnicity information from the LEADe[103] and UCSD MCI[104] studies. The study described here is within the scope of patient consent.

Palotie, A | Finland | scz_xfi3m_eur_sr-qc
Palotie, A | Finland | scz_xfii6_eur_sr-qc
We recruited participants as part of the THL Psychiatric Family Collections, a Finnish nationwide collection of families with schizophrenia spectrum disorders. The control participants are from several different population-based cohorts: the Finnish Health 2000 survey, the National FINRISK Study, Northern Finland Birth Cohort (NFBC) and Helsinki Birth Cohort Study (HBCS). For the analysis the controls were selected based on PCA matching. The Finnish Schizophrenia Family Study sample was identified from nationwide Finnish health care registers. Affected subjects were ascertained using three nationwide registers: the Finnish Hospital Discharge Register, the Finnish Pension Register, and the Finnish Medication Reimbursement Register. The family members of affected individuals were subsequently identified from the Finnish Population Information System, and health care register information was obtained for them as well. For the diagnostic assessment, medical records were collected from all in- and outpatient mental health treatment contacts. The probands were contacted through their treating psychiatrist, and the rest of the sample was contacted with the proband's permission. The lifetime diagnoses were evaluated according to DSM-IV[51] criteria. Each case was evaluated independently by two psychiatrists, who were blind to register diagnoses and family structure, and in case of disagreement, a third psychiatrist evaluated the lifetime diagnosis and a consensus was made. One of the psychiatrists completed the Operational Criteria Checklist (OPCRIT)[99] based on lifetime review of symptoms. In addition, one third of the sample participated in a clinical assessment, consisting of the Structured Clinical

Interview for DSM-IV (SCIDI and SCID-II)[56], symptom assessment using the Scale for the Assessment of Positive Symptoms (SAPS)[105] and the Scale for the Assessment of Negative Symptoms (SANS)[106] and neuropsychological testing. These assessments were done by psychiatrists, trained psychologists or psychiatric nurses. Blood samples were collected from all participants. The study was approved by the Ministry of Social Affairs and Health of Finland, the Ethics Committee of the National Public Health Institute of Finland (since 1 January 2009 National Institute for Health and Welfare), and the Ethics Committee of the Hospital District of Helsinki and Uusimaa.

Pato, C | Portugal | scz_xport_eur_sr-qc
Cases and controls lived in Portugal, the Azorean and Madeiran islands, or were the direct (first or second-generation) Portuguese immigrant population in the US, as previously described[73]. Controls were not biologically related to cases. All participants gave written informed consent and the IRB of SUNY Upstate Medical University approved the protocol. The samples were genotyped at the Broad Institute.

Pato, C | Multiple sites | scz_gpc2a_eur_sr-qc
We recruited participants as part of the Genomic Psychiatry Cohort (GPC), a study based in the Los Angeles that recruited controls and cases living and being treated in local communities and healthcare delivery systems. This sample has been described elsewhere[107]. Case participants were interviewed using the Diagnostic Interview for Psychosis and Affective Disorders (DI-PAD)[108], a semi-structured clinical interview administered by mental health professionals. Inclusion criteria for cases included meeting lifetime diagnostic criteria for schizophrenia or schizoaffective disorder (any subtype) in accordance with the OPCRIT[99] algorithms for DSM-IV and/or ICD-10 criteria, and/or DSM-5[51,52,109]. Individuals reporting no lifetime symptoms indicative of psychosis or mania and who had no first-degree relatives with these symptoms were included as control participants. Exclusion criteria included any premorbid organic mental disorders and premorbid history of significant drug or alcohol dependence by DSM-IV/5[51,109] that confounds the diagnosis of schizophrenia. DNA was extracted from whole blood. All participants gave written informed consent and the IRB of the participating institutions approved the protocol.

Petryshen, T | Boston, US (CIDAR) | scz_xcims_eur_sr-qc
Cases were recruited from inpatient and outpatient settings in the Boston area by clinician referral, through review of medical records, or through advertisements in local media. Cases were diagnosed with DSM-IV schizophrenia through a structured clinical interview (SCID) by trained interviewers with review of medical records and a best estimate diagnostic procedure including reliability trials across interviewers. A psychiatrist or a PhD-level mental health professional made the final diagnostic determination. Controls were ascertained through local advertisements from the same geographical area. Ethical approval was provided by local ethics committees and all participants gave written informed consent.

Rietschel, M; Rujescu, D; Nöthen, M | Bonn/Mannheim, Germany | scz_xboco_eur_sr-qc
These German samples were collected by separate groups within the MooDS Consortium in Mannheim, Bonn, Munich and Jena. For the PGC analyses, the samples were combined by chip and ancestry. In Bonn/Mannheim, cases were ascertained as previously described[89]. Controls were drawn from three population-based epidemiological studies (PopGen)[110], the Cooperative Health Research in the Region of Augsburg (KORA) study[111], and the Heinz Nixdorf Recall (HNR) study[112]. All participants gave written informed consent and the local ethics committees approved the human subjects protocols. Additional controls were randomly selected from a Munich-based community sample and screened for the presence of anxiety and affective disorders using the

Composite International Diagnostic Screener[113]. Only individuals negative for the above mentioned disorders were included in the sample.

Ripke, S | Berlin | scz_bep1b
The Berlin Psychosis Study (BePS) is a case-control sample initiated in greater Berlin aiming to facilitate the discovery of novel genetic variants associated with schizophrenia. The current sample consists of individuals of Caucasian ancestry. Participants were primarily recruited from inpatient and outpatient facilities of eight collaborating hospitals and secondarily from the general population. Lifetime diagnosis of schizophrenia or schizoaffective disorder according to ICD-10[52] was ensured by referral from the attending clinicians and/or access to medical records. In addition, we recruited control subjects into the study via local advertisement and participant databases. Control subjects were excluded if they had ever been diagnosed with schizophrenia, schizoaffective disorder or bipolar disorder. Ethical approval for the study was granted by the local ethics committee at Charité Universitätsmedizin, Berlin. Saliva samples were collected from each participant using OraGene-510 DNA-Self-Collection Kits (DNA Genotek, Ottawa, Ontario, Canada).

Rujescu, D | Munich, Germany | scz_xmunc_eur_sr-qc
For the Munich sample, cases were ascertained from the Munich area of Germany, as described previously[89]. The controls were unrelated volunteers randomly selected from the general population of Munich. All were screened to exclude a history of psychosis/central neurological disease either personally or in a first-degree relative. All participants gave written informed consent and the local ethics committees approved the human subjects protocols.

Serretti, A | Italy | scz_serri_eur_sr-qc
Unrelated subjects with chronic psychotic disorders admitted to participating psychiatric units with a DSM-IV-TR[78] diagnosis of schizophrenia were recruited. All patients were of Italian origin. Selected participants were inpatients receiving care from the national mental health service at the Psychiatric Unit of the San Filippo Neri Hospital (Rome, Italy) and at the nursing home (RSA) San Raffaele “Villa dei Fiori" (Rome, Italy) between 2011 and 2012. A subsample was previously described elsewhere[114]. All patients were under antipsychotic treatment. Subjects were included if they had capacity to give informed consent, sufficient Italian language skills to complete the study measures, were not intellectually disabled, and were not considered a risk to the safety of research staff. Healthy controls were recruited and included in the context of a medical screening, no formal psychiatric interview was administered but the absence of major and invalidating psychiatric disorder was recorded. The study was approved by the San Raffaele Pisana and by ASL RME Ethics Committees, and all participants provided written informed consent.

St Clair, D | Aberdeen, UK | scz_xaber_eur_sr-qc
Ascertainment and inclusion/exclusion criteria for cases and controls have been previously described[73]. All participating subjects were born in the UK (95% Scotland) and gave written informed consent. Both local and multiregional academic ethical committee approved the human subjects protocol. The samples were genotyped at the Broad Institute.

Stefánsson, K | Iceland (SGENE+, deCODE) | NA
Stefánsson, K | Non-Icelandic (SGENE+, deCODE) | NA
This sample had two components. The first consisted of cases and controls recruited in Iceland. Diagnoses were assigned either 1) according to Research Diagnostic Criteria (RDC)[98] using the Schedule for Affective Disorders and Schizophrenia Lifetime Version (SADS-L)[97] as described previously[89] or 2) using ICD-10[52] criteria for schizophrenia (F20) or schizoaffective disorder (F25).

Controls were recruited as part of various genetic programs at deCODE and were not screened for psychiatric disorders. The non-Icelandic sample included cases and controls from Hungary, Italy, Georgia, Macedonia, Russia and Serbia. Recruitment and diagnosis for the individuals from Italy, Georgia, Macedonia, Russia and Serbia has been described[89,115]. Hungarian inpatients and outpatients with a DSM-IV[51] diagnosis of schizophrenia were recruited from the Department of Psychiatry and Psychotherapy, Semmelweis University and the Psychiatry Unit of Szent János Hospital, both in Budapest. Criteria for exclusion were severely disorganized behaviour that prevented patient cooperation and severe comorbidity, such as neurological disorders, head trauma, mental retardation, or substance-abuse. The DSM-IV diagnosis for schizophrenia excluding schizoaffective disorder was validated using the MINI Neuropsychiatric Interview[72]. Healthy controls were recruited from the employees and students of Semmelweis University and outpatients of the Department of Internal Medicine after screening for psychiatric disorders. All studies were approved by local ethics committees, and all participants provided written, informed consent. Genotyping was carried out at deCODE Genetics.

Sullivan, PF | US (CATIE) | scz_xcati_eur_sr-qc
Cases were collected as part of the Clinical Antipsychotics Trials of Intervention Effectiveness (CATIE) project and ascertainment was previously described[116–118]. Participants were recruited from multiple sites in the USA with informed written consent and approval from the IRBs at each CATIE site and the University of North Carolina (Chapel Hill). The control subjects were collected by MGS (described above) and gave online informed consent and were fully anonymized. There was no overlap with controls included in the MGS collaboration sample. Data availability: genotype data for this study are available as a part of the Schizophrenia 20.21 resource, study 17, at the NIMH Repository & Genomics Resource: https://www.nimhgenetics.org/download-tool/SZ.

Sullivan, PF; Sklar P; Hultman C | Sweden | scz_xswe1_eur_sr-qc
Sullivan, PF; Sklar P; Hultman C | Sweden | scz_xs234_eur_sr-qc
Sullivan, PF; Sklar P; Hultman C | Sweden | scz_xswe5_eur_sr-qc
Sullivan, PF; Sklar P; Hultman C | Sweden | scz_xswe6_eur_sr-qc
Samples from the Swedish Schizophrenia Study were collected in a multi-year project and genotypes in six batches (sw1-6). Reference [119] is the main report for this study but, in order to further progress in the field, sw1-2 were included in reference [73] and sw1-4 in reference[120]. All procedures were approved by ethical committees at the Karolinska Institutet and the University of North Carolina, and all subjects provided written informed consent (or legal guardian consent and subject assent). All samples were genotyped at the Broad Institute. Cases with schizophrenia were identified via the Swedish Hospital Discharge Register which captures all public and private inpatient hospitalizations. The register is complete from 1987 and is augmented by psychiatric data from 1973-1986. The register contains International Classification of Disease discharge diagnoses made by attending physicians for each hospitalization. Case inclusion criteria included ≥2 hospitalizations with a discharge diagnosis of schizophrenia, both parents born in Scandinavia and age ≥18 years. Case exclusion criteria included hospital register diagnosis of any medical or psychiatric disorder mitigating a confident diagnosis of schizophrenia as determined by expert review. The validity of this case definition of schizophrenia was strongly supported by clinical, epidemiological, genetic epidemiological and genetic evidence (see the Supplementary Note in reference [119]). Controls were selected at random from Swedish population registers, with the goal of obtaining an appropriate control group and avoiding ‘super-normal’ controls. Control inclusion criteria included never being hospitalized for schizophrenia or bipolar disorder (given evidence of genetic overlap with schizophrenia), both parents born in Scandinavia and age of ≥18 years. Data availability: genotype data for these studies are available as a part of the Schizophrenia 20.21 resource at the NIMH Repository & Genomics Resource: https://www.nimhgenetics.org/download-tool/SZ.

Van Os, J | Netherlands and Belgium | ms.scz_grtr1_eur_sr-qc
Van Os, J | Netherlands and Belgium | scz_gro2a_eur_sr-qc
We recruited participants as part of the GROUP STUDY, a study based in The Netherlands and (Dutch speaking part of) Belgium. Patients were identified through clinicians working in regional psychosis departments or academic centres, whose caseload was screened for inclusion criteria. Subsequently, a group of patients presenting consecutively at these services either as outpatients or inpatients were recruited for the study. This sample has been described in the GROUP method paper[121]. Participants were interviewed by trained research assistants, psychologists, psychiatrists, nurses and PhD students. Psychiatric diagnosis was established according to criteria of DSM-IV[51]. Inclusion criteria for patients were the following: (1) age range of 16 to 50 years, (2) a diagnosis of non-affective psychotic disorder according to DSM-IV[51] criteria, (3) good command of the Dutch language, and (4) able and willing to give written informed consent. Inclusion criteria for healthy controls were the following: (1) age range of 16 and 50 years, (2) no lifetime psychotic disorder, (3) no first degree family member with a lifetime psychotic disorder (4) good command of the Dutch language, and (5) able and willing to give written informed consent. The healthy control participants underwent similar structured and unstructured interviewing to the cases, to screen for both psychiatric and non-psychiatric illness.

Van Os, J; O'Donovan, M | EUGEI | scz_eu5me_eur_sa-qc
This sample is part of the EUGEI and has been described previously[57,58].

Walters, J | UK | scz_cgs1c_eur_sr-qc
These samples derived from the CardiffCOGS2 study and recruitment was via secondary care, mainly outpatient, NHS mental health services in Wales and England. These patients were not exclusively taking clozapine at the time of their recruitment. All cases underwent a SCAN interview[77] and case note review followed by consensus research diagnostic procedures and were included if they had a DSM-IV[51] schizophrenia or schizoaffective disorder-depressive type diagnosis, as previously reported[122,123]. The CardiffCOGS samples were recruited and genotyped in two waves: CardiffCOGS1 (see immediately below), included in a previous GWAS[124], and CardiffCOGS2, this current sample, both following the same procedures and protocol.

Walters, J | Cardiff, UK | scz_xcou3_eur_sr-qc
These samples constituted the CardiffCOGS1 study. Cases were recruited from community mental health teams in Wales and England on the basis of a clinical diagnosis of schizophrenia or schizoaffective disorder (depressed sub-type) as described previously[125]. Diagnosis was confirmed following a SCAN[90] interview and review of case notes followed by consensus diagnosis according to DSM-IV[51] criteria. The samples were genotyped at the Broad Institute. The UK Multicentre Research Ethics Committee (MREC) approved the study and all participants provided valid informed consent.

Walters, J; O'Donovan, M; Owen, M | CLOZUK | scz_clz2a_eur_sr-qc
We collected blood samples from those with treatment-resistant schizophrenia (TRS) in the UK through the mandatory clozapine blood-monitoring system for those taking clozapine, an antipsychotic licensed for TRS. Following national research ethics approval and in line with UK Human Tissue Act regulations we worked in partnership with the commercial companies that manufacture and monitor clozapine in the UK. We ascertained anonymous aliquots of the blood samples collected as part of the regular blood monitoring that takes place whilst taking clozapine due to a rare haematological adverse effect, agranulocytosis. The sample was assembled in

collaboration with Leyden Delta (Nijmegen, Netherlands), a major company involved in the supply Treatment Access System (ZTAS), provided whole-blood samples and anonymised phenotypic information. Both Clozaril® and Zaponex® are bioequivalent brands of clozapine licensed in the UK. The sample has been described previously[123].

Walters, J; O'Donovan, M; Owen M | UK (CLOZUK) | scz_xclm2_eur_sr-qc
Walters, J; O'Donovan, M; Owen M | UK (CLOZUK) | scz_xclo3_eur_sr-qc
CLOZUK cases were taking the antipsychotic clozapine and had received a clinical diagnosis of treatment-resistant schizophrenia. Patients taking clozapine provide blood samples to allow detection of adverse drug-effects. Through collaboration with Novartis (the manufacturer of a proprietary form of clozapine, Clozaril), we acquired blood from people with treatment-resistant schizophrenia according to the clozapine registration forms completed by treating psychiatrists as previously reported[126]. The samples were genotyped at the Broad Institute. The UK Multicentre Research Ethics Committee (MREC) approved the study. The controls were drawn from the WTCCC2 control samples (~3,000 from the 1958 British Birth Cohort and ~3,000 samples from the UK Blood Service Control Group). An additional 900 controls, held by Cardiff University, were recruited from the UK National Blood Transfusion Service. They were not specifically screened for psychiatric illness. All control samples were from participants who provided informed consent.

Weinberger, D | NIMH CBDB | scz_xlie2_eur_sr-qc
Weinberger, D | NIMH CBDB | scz_xlie5_eur_sr-qc
Subjects were recruited from the Clinical Brain Disorders Branch of the NIMH 'Sibling Study' as previously described[127]. In brief, cases and controls gave informed consent and only participants of European ancestry were included in the current analysis. Cases completed a structured clinical interview and were diagnosed with schizophrenia-spectrum disorders. Samples were genotyped at the NIMH. Data availability: genotype data for this study are available as a part of the Schizophrenia 20.21 resource at the NIMH Repository & Genomics Resource: https://www.nimhgenetics.org/download-tool/SZ.

Werge, T | Denmark | scz_xdenm_eur_sr-qc
Cases were ascertained through psychiatric departments and twin pair studies and were of Danish parentage for at least the prior three generations. The controls were collected at the University of Aarhus, and included 500 medical students, all of Danish parentage for at least three generations. All subjects gave written informed consent and the Danish Data Protection Agency and the ethics committees of Denmark approved the human subjects protocol.

Asian samples
All fourteen Asian samples (bix1, bix2, bix3, cno1[bjm1], hku1, imh1, imh2, jpn1, kor1, tai1, tai2, umc1, uwa1, xju1) in this study have been described in previous publications[31,128–131].

African-American and Latino samples
Cases and controls of African-American and Latino ancestry were ascertained from the Genomic Psychiatry Cohort (GPC)[132]. GPC participants were drawn from cases and controls living and being treated in local communities and healthcare delivery systems. All participants enrolled as probable cases were interviewed using the Diagnostic Interview for Psychosis and Affective Disorders (DI-PAD), a semi-structured clinical interview administered by mental health professionals. The DI-PAD was developed specifically for the GPC study using the same principles as were applied in the development of the Diagnostic Interview for Psychosis – Diagnostic Module (DIP-DM)[133], and incorporates questions developed for the Diagnostic Interview for Genetic Studies (DIGS)[64]. Inclusion criteria for cases include meeting lifetime diagnostic criteria for schizophrenia or schizoaffective disorder (any subtype) in accordance with the OPCRIT algorithms for DSM-IV[78]

and/or ICD-10[52] criteria, and/or DSM-5[109]. Individuals reporting no lifetime symptoms indicative of psychosis or mania and who have no first-degree relatives with these symptoms are included as control participants. Exclusion criteria included any premorbid organic mental disorders (i.e., epilepsy, CNS infection, significant head trauma, mental retardation), and premorbid history of significant drug or alcohol dependence by DSM IV/5 that confounds the diagnosis of schizophrenia. All participants gave written informed consent and the IRB of the participating institutions approved the protocol.
Blood-derived DNA for additional cases and controls meeting the same inclusion and exclusion criteria were obtained from the NIMH Repository, including the Molecular Genetics of Schizophrenia (MGS)[86,87,134], COGS[135], and PAARTNERS[136] studies.

Control cohorts descriptions

Wellcome Trust Case-Control Consortium 2
Wellcome Trust Case-Control Consortium unscreened controls from the UK Blood Bank and 1958 Birth Cohort (NCDS).

Generation Scotland
Samples from individuals recruited by the Generation Scotland Scottish Family Health Study. While in the original design there was no selection on the basis of medical status or history, these controls have been screened for psychiatric disorders using SCID criteria.

dbGAP We obtained control genotypes from dbGAP, accession numbers: phs000021.v3.p2, phs000294.v1.p1

Sample Description Table. Information on the genotyping chip, number of cases and controls with breakdown by sex, and number of genotyped SNPs is provided for each sample.

Dataset	Ancestry	Platform	Cases	Controls	N all	Female	Male	No sex	N SNPs
scz_xume2_eur_sr-qc	EUR	OMEX	595	1638	2233	1152	1079	2	625174
scz_xtop8_eur_sr-qc	EUR	A6.0	377	403	780	364	416	0	694404
scz_to10c_eur_sr-qc	EUR	OMEX	970	5040	6010	2946	3064	0	605462
scz_eusp2_eur_sr-qc	EUR	COEX	338	490	828	291	537	0	360801
scz_celso_eur_sr-qc	EUR	PSYC	2030	1517	3547	1377	2170	0	296859
scz_eutu2_eur_sr-qc	EUR	COEX	393	690	1083	513	570	0	292450
scz_geba1_eur_sr-qc	EUR	PSYC	397	703	1100	578	522	0	225788
scz_sb2aa_eur_sr-qc	EUR	OMEX	250	237	487	179	308	0	479655
scz_xedin_eur_sr-qc	EUR	A6.0	368	284	652	239	413	0	686101
scz_xaarh_eur_sr-qc	EUR	I650	883	873	1756	799	957	0	570800
scz_cogs1_eur_sr-qc	EUR	PSYC	428	476	904	362	542	0	395587
scz_xpewb_eur_sr-qc	EUR	I1M	641	1892	2533	1121	1412	0	730963
scz_xpews_eur_sr-qc	EUR	I1M	150	236	386	160	226	0	758373
scz_xmsaf_eur_sr-qc	EUR	A6.0	327	139	466	180	286	0	772796
scz_rouin_eur_sr-qc	EUR	PSYC	204	185	389	169	220	0	346567
scz_xdubl_eur_sr-qc	EUR	A6.0	272	860	1132	686	446	0	684202
scz_du2aa_eur_sr-qc	EUR	COEX	345	245	590	233	357	0	311797
scz_xirwt_eur_sr-qc	EUR	A6.0	1309	1022	2331	893	1438	0	691537
scz_gap1a_eur_sr-qc	EUR	COEX	152	164	334	133	201	0	318675
scz_enric_eur_sr-qc	EUR	PSYC	700	574	1274	539	735	0	379548
scz_xgras_eur_sr-qc	EUR	AXI	1086	1232	2318	829	1489	0	593545
scz_xegcu_eur_sr-qc	EUR	omni	239	1177	1416	1037	379	0	610089

scz_xjrsa_eur_sr-qc	EUR	I1M	1181	2313	3494	1728	1766	0	813668
scz_xjr3a_eur_sr-qc	EUR	I317	362	325	687	332	355	0	277187
scz_xjr3b_eur_sr-qc	EUR	I317	647	639	1286	682	604	0	282011
scz_xjri6_eur_sr-qc	EUR	I610	260	130	390	202	188	0	479354
scz_price_eur_sr-qc	EUR	PSYC	841	727	1568	681	887	0	296858
scz_gawli_eur_sr-qc	EUR	PSYC	1255	1555	2810	1235	1575	0	287202
scz_xmgs2_eur_sr-qc	EUR	A6.0	2681	2653	5334	2200	3134	0	653445
scz_mosc2_eur_sr-qc	EUR	I650	410	433	843	401	442	0	543041
scz_xuclo_eur_sr-qc	EUR	A6.0	521	494	1015	434	581	0	281420
kor1	ASN	Korean BioBank chip (affy)	688	492	1180	573	607	0	638808
jpn1	ASN	A5.0	547	540	1087	544	543	0	355120
scz_xersw_eur_sr-qc	EUR	omni	322	332	654	250	404	0	728357
umc1	ASN	PSYC	2328	2380	4708	2176	2532	0	372575
scz_xbuls_eur_sr-qc	EUR	A6.0	195	608	803	422	381	0	697382
ms.scz_xuktr_eur_sr-qc	EUR	omni	70	140	210	85	125	0	605797
ms.scz_xbutr_eur_sr-qc	EUR	A6.0	741	1156	1947	970	977	0	657466
scz_xlktu_eur_sr-qc	EUR	A6.0	322	332	654	250	404	0	728357
scz_paris_eur_sr-qc	EUR	PSYC	316	390	706	314	392	0	378285
scz_xajsz_eur_sr-qc	EUR	omni	896	1595	2491	744	1747	0	761416
scz_xlacw_eur_sr-qc	EUR	I550	157	466	623	51	572	0	512787
ms.scz_xlemu_eur_sr-qc	EUR	I650	585	0	1548	687	861	0	569126
imh1	ASN	I1M	898	996	1894	1197	648	49	973108
imh2	ASN	I1M	821	956	1777	937	835	5	861016
xju1	ASN	Zhonghua-8	1902	1009	2911	1224	1732	0	850021
scz_xzhh1_eur_sr-qc	EUR	A500	191	190	381	161	220	0	264937
scz_mcqul_eur_sr-qc	EUR	PSYC	1351	1310	2661	1241	1420	0	282320
scz_braz2_eur_sr-qc	EUR	COEX	110	334	444	199	245	0	354156
scz_xasrb_eur_sr-qc	EUR	I650	509	310	819	327	492	0	561022
tai1	ASN	PSYC	1123	2243	3366	1807	1559	0	332892
tai2	ASN	PSYC	593	1190	1783	593	1190	0	327109
scz_viyo1_eur_sr-qc	EUR	PSYC	356	131	487	201	286	0	282551
scz_xcaws_eur_sr-qc	EUR	A500	424	306	730	300	430	0	365639
scz_xclm2_eur_sr-qc	EUR	I1M	3466	4297	7763	3078	4685	0	429056
scz_xclo3_eur_sr-qc	EUR	omni	2150	2083	4233	1638	2595	0	659685
scz_xucla_eur_sr-qc	EUR	I550	705	637	1342	499	843	0	520895
scz_xfi3m_eur_sr-qc	EUR	I317	186	930	1116	542	574	0	287517
scz_xfii6_eur_sr-qc	EUR	I550	361	1082	1443	775	668	0	511059
scz_xport_eur_sr-qc	EUR	A6.0	346	216	562	269	293	0	344466
scz_gpc2a_eur_sr-qc	EUR	PSYC	1957	2062	4019	1722	2297	0	294169
scz_xcims_eur_sr-qc	EUR	ill	71	69	140	34	106	0	670122
scz_xboco_eur_sr-qc	EUR	I550	1847	2170	4017	2050	1967	0	434891
scz_bep1b	EUR	GSA	294	573	867	510	357	0	525759
scz_sanch_eur_sr-qc	EUR	PSYC	335	1209	1544	693	851	0	296387
scz_xmunc_eur_sr-qc	EUR	I317	437	351	788	340	448	0	305068
uwa1	ASN	PSYC	996	1047	2043	1277	764	2	351136
scz_serri_eur_sr-qc	EUR	PSYC	217	238	455	213	242	0	345297
hku1	ASN	I610	476	2018	2494	992	1501	1	457012
bix1	ASN	A6.0	1047	2301	3348	1453	1895	0	660360
bix2	ASN	A6.0	1021	1001	2022	919	1113	0	672911
bix3	ASN	A6.0	492	679	1171	656	515	0	629614
scz_xaber_eur_sr-qc	EUR	A6.0	720	699	1419	435	984	0	373076
scz_xcati_eur_sr-qc	EUR	A500	409	392	801	187	614	0	419565
scz_xswe1_eur_sr-qc	EUR	A5.0	221	214	435	208	227	0	387864
scz_xs234_eur_sr-qc	EUR	A6.0	2077	2341	4418	1990	2428	0	746747
scz_xswe5_eur_sr-qc	EUR	omni	1801	2617	4418	1975	2443	0	642179

scz_xswe6_eur_sr-qc	EUR	omni	1094	1219	2313	1062	1251	0	631239
ms.scz_grtr1_eur_sr-qc	EUR	COEX	145	290	435	171	264	0	308001
scz_gro2a_eur_sr-qc	EUR	COEX	329	277	606	220	386	0	318633
scz_eu5me_eur_sa-qc	EUR	COEX	615	182	797	388	408	1	259434
scz_cgs1c_eur_sr-qc	EUR	OMEX	524	3115	3639	1766	1873	0	539647
scz_xcou3_eur_sr-qc	EUR	omni	540	693	1233	557	676	0	700613
scz_clz2a_eur_sr-qc	EUR	OMEX	5370	6940	12310	5527	6779	4	556861
scz_xlie5_eur_sr-qc	EUR	I550	509	389	898	335	563	0	495531
scz_xlie2_eur_sr-qc	EUR	O25	137	269	406	184	222	0	1503067
scz_xpfla_eur_sr-qc	EUR	I550	681	1174	1855	856	999	0	471418
scz_xdenm_eur_sr-qc	EUR	I650	492	458	950	396	554	0	555486
cno1[bjm1]	ASN	Zhonghua-8	1332	2036	3368	1582	1786	0	869413

References

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*. (American Psychiatric Association, 1994).

2. World Health Organisation. *The ICD-10 Classification of Mental and Behavioural Disorders Clinical descriptions and diagnostic guidelines*. (World Health Organisation).

3. Nilsson, L.-Gör. *et al.* The betula prospective cohort study: Memory, health, and aging. *Aging, Neuropsychol. Cogn.* **4**, 1–32 (1997).

4. Athanasiu, L. *et al.* Gene variants associated with schizophrenia in a Norwegian genome-wide study are replicated in a large European cohort. *J. Psychiatr. Res.* **44**, 748–753 (2010).

5. Nesvåg, R. *et al.* The quality of severe mental disorder diagnoses in a national health registry as compared to research diagnoses based on structured interview. *BMC Psychiatry* **17**, 93 (2017).

6. First, M. B., Williams, J. B. ., Spitzer, R. L. & Gibbon, M. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Clinical Trials Version (SCID-CT)*. (Biometrics Research, New York State Psychiatric Institute, 2007).

7. Van Os, J. *et al.* Replicated evidence that endophenotypic expression of schizophrenia polygenic risk is greater in healthy siblings of patients compared to controls, suggesting gene-environment interaction. the EUGEI study. *Psychol. Med.* (2019). doi:10.1017/S003329171900196X

8. Gayer-Anderson, C. *et al.* The EUropean Network of National Schizophrenia Networks Studying Gene–Environment Interactions (EU-GEI): Incidence and First-Episode Case–Control Programme. *Soc. Psychiatry Psychiatr. Epidemiol.* (2020). doi:10.1007/s00127-020-01831-x

9. European Network of National Networks studying Gene-Environment Interactions in Schizophrenia (EU-GEI) *et al.* Identifying Gene-Environment Interactions in Schizophrenia: Contemporary Challenges for Integrated, Large-scale Investigations. *Schizophr. Bull.* **40**, 729–736 (2014).

10. Guloksuz, S. *et al.* Examining the independent and joint effects of molecular genetic liability and environmental exposures in schizophrenia: results from the EUGEI study. *World Psychiatry* **18**, 173–182 (2019).

11. Santoro, M. L. *et al.* Polygenic risk score analyses of symptoms and treatment response in an antipsychotic-naive first episode of psychosis cohort. *Transl. Psychiatry* **8**, 174 (2018).

12. Moretti, P. N. *et al.* Accessing Gene Expression in Treatment-Resistant Schizophrenia. *Mol. Neurobiol.* **55**, 7000–7008 (2018).

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

14. Swerdlow, N. R., Gur, R. E. & Braff, D. L. Consortium on the Genetics of Schizophrenia (COGS) assessment of endophenotypes for schizophrenia: An introduction to this Special Issue of schizophrenia research. *Schizophr. Res.* **163**, 9–16 (2015).

15. Nurnberger, J. I. *et al.* Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. *Arch. Gen. Psychiatry* **51**, 849–859 (1994).

16. Maxwell, M. *Family Interview for Genetic Studies (FIGS): A Manual for FIGS.* (NIMH: Clinical Neurogenetics Branch, 1992).

17. Glessner, J. T. *et al.* Strong synaptic transmission impact by copy number variations in schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 10584–9 (2010).

18. Jacquet, H. *et al.* Hyperprolinemia is a risk factor for schizoaffective disorder. *Mol. Psychiatry* **10**, 479–485 (2005).

19. Kay, S. R., Fiszbein, A. & Opler, L. A. The Positive and Negative Syndrome Scale (PANSS) for Schizophrenia. *Schizophr. Bull.* **13**, 261–276 (1987).

20. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (DSM-III-R)*. (American Psychiatric Association, 1987).

21. Muñoz-Negro, J. E. *et al.* A dimensional comparison between delusional disorder, schizophrenia and schizoaffective disorder. *Schizophr. Res.* **169**, 248–254 (2015).

22. Cervilla, J. A. *et al.* Protocolo y metodología del estudio epidemiológico de la salud mental en Andalucía: PISMA-ep. *Rev. Psiquiatr. Salud Ment.* **9**, 185–194 (2016).

23. 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. *J. Clin. Psychiatry* **59 Suppl 20**, 22-33;quiz 34-57 (1998).

24. Richards, A. L. *et al.* The Relationship Between Polygenic Risk Scores and Cognition in Schizophrenia. *Schizophr. Bull.* (2019). doi:10.1093/schbul/sbz061

25. Irish Schizophrenia Consortium & Wellcome Trust Case Control Consortium 2. Genome-Wide Association Study Implicates HLA-C*01:02 as a Risk Factor at the Major Histocompatibility Complex Locus in Schizophrenia. *Biol. Psychiatry* **72**, 620–628 (2012).

26. Di Forti, M. *et al.* Confirmation that the AKT1 (rs2494732) genotype influences the risk of psychosis in cannabis users. *Biol. Psychiatry* **72**, 811–6 (2012).

27. World Health Organisation. *Schedules for Clinical Assessment in Neuropsychiatry*. (World Health Organisation, 1992).

28. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR)*. (American Psychiatric Association., 2000).

29. First, M. B., Gibbon, M., Spitzer, R. L. & Williams, J. *Structured Clinical Interview for DSM-IV Axis I Disorders: Nonpatient Edition (SCID-I/NP)*. (New York State Psychiatric Institute, 1996).

30. Need, A. C. *et al.* A Genome-Wide Investigation of SNPs and CNVs in Schizophrenia. *PLoS Genet.* **5**, e1000373 (2009).

31. Blasi, G. *et al.* Association of *GSK-3β* Genetic Variation With GSK-3β Expression, Prefrontal Cortical Thickness, Prefrontal Physiology, and Schizophrenia. *Am. J. Psychiatry* **170**, 868–876 (2013).

32. Bertolino, A. *et al.* COMT Val158Met polymorphism predicts negative symptoms response to treatment with olanzapine in schizophrenia. *Schizophr. Res.* **95**, 253–255 (2007).

33. Rampino, A. *et al.* A Polygenic Risk Score of glutamatergic SNPs associated with schizophrenia predicts attentional behavior and related brain activity in healthy humans. *Eur. Neuropsychopharmacol.* **27**, 928–939 (2017).

34. Begemann, M. *et al.* Modification of Cognitive Performance in Schizophrenia by Complexin 2 Gene Polymorphisms. *Arch. Gen. Psychiatry* **67**, 879 (2010).

35. Metspalu, A., Köhler, F., Laschinski, G., Ganten, D. & Roots, I. The Estonian Genome Project in the context of European genome research. *Dtsch. Medizinische Wochenschrift* **129**, S25-8 (2004).

36. Shi, J. *et al.* Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* **460**, 753–757 (2009).

37. Sanders, A. R. *et al.* The Internet-Based MGS2 Control Sample: Self Report of Mental Illness. *Am. J. Psychiatry* **167**, 854–865 (2010).

38. Fedorenko, O. Y. *et al.* Opening up new horizons for psychiatric genetics in the Russian Federation: moving toward a national consortium. *Mol. Psychiatry* **24**, 1099–1111 (2019).

39. Stefansson, H. *et al.* Common variants conferring risk of schizophrenia. *Nature* **460**, 744–747 (2009).

40. Wing, J. K. *et al.* SCAN - schedules for the clinical assessment in neuropsychiatry. *Arch. Gen. Psychiatry* **47**, 589–593 (1990).

41. Kathiresan, S. *et al.* Genome-wide association of early-onset myocardial infarction with common single nucleotide polymorphisms, common copy number variants, and rare copy number variants. *Nat. Genet.* **41**, 334–341 (2009).

42. Guha, S. *et al.* Implication of a Rare Deletion at Distal 16p11.2 in Schizophrenia. *JAMA Psychiatry* **70**, 253 (2013).

43. Levinson, D. F. *et al.* Genome-Wide Association Study of Multiplex Schizophrenia Pedigrees. *Am. J. Psychiatry* **169**, 963–973 (2012).

44. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).

45. Lencz, T. *et al.* Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Mol. Psychiatry* **12**, 572–580 (2007).

46. Al Eissa, M. M. *et al.* Genetic association and functional characterization of MCPH1 gene variation in bipolar disorder and schizophrenia. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **180**, 258–265 (2019).

47. Spitzer, R. The schedule for affective disorders and schizophrenia, lifetime version. in (New York State Psychiatric Institute, 1977).

48. Spitzer, R. L., Endicott, J. & Robins, E. Research diagnostic criteria: rationale and reliability. *Arch. Gen. Psychiatry* **35**, 773–82 (1978).

49. McGuffin, P., Farmer, A. & Harvey, I. A polydiagnostic application of operational criteria in studies of psychotic illness. Development and reliability of the OPCRIT system. *Arch. Gen. Psychiatry* **48**, 764–70 (1991).

50. Morgan, V. A. *et al.* People living with psychotic illness in 2010: The second Australian national survey of psychosis. *Aust. N. Z. J. Psychiatry* **46**, 735–752 (2012).

51. Nimgaonkar, V. L. *et al.* Joint evaluation of serum C-Reactive Protein levels and polygenic risk scores as risk factors for schizophrenia. *Psychiatry Res.* **261**, 148–153 (2018).

52. O'Donovan, M. C. *et al.* Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat. Genet.* **40**, 1053–1055 (2008).

53. Jones, R. W. *et al.* The Atorvastatin/Donepezil in Alzheimer's Disease Study (LEADe): Design and baseline characteristics. *Alzheimer's Dement.* **4**, 145–153 (2008).

54. Golomb, B. A., Dimsdale, J. E., White, H. L., Ritchie, J. B. & Criqui, M. H. Reduction in Blood Pressure With Statins: Results From the UCSD Statin Study, a Randomized Trial</subtitle>. *Arch. Intern. Med.* **168**, 721 (2008).

55. Andreasen, N. Scale for the Assessment of Positive Symptoms (SAPS). in (University of Iowa, 1984).

56. Andreasen, N. *Scale for the Assessment of Negative Symptoms (SANS)*. (University of Iowa, 1984).

57. Pato, M. T. *et al.* The genomic psychiatry cohort: Partners in discovery. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **162**, 306–312 (2013).

58. Semple, R. Diagnostic interview for Psychosis and Affective Disorders (DI-PAD). (2008).

59. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (DSM-V)*. (American Psychiatric Association, 2013).

60. Krawczak, M. *et al.* PopGen: Population-Based Recruitment of Patients and Controls for the Analysis of Complex Genotype-Phenotype Relationships. *Public Health Genomics* **9**, 55–61 (2006).

61. Wichmann, H.-E., Gieger, C., Illig, T. & Group, null for the M. S. KORA-gen - Resource for Population Genetics, Controls and a Broad Spectrum of Disease Phenotypes. *Das Gesundheitswes.* **67**, 26–30 (2005).

62. Schmermund, A. *et al.* Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: Rationale and design of the Heinz Nixdorf RECALL Study. *Am. Heart J.* **144**, 212–218 (2002).

63. Lucae, S. *et al.* P2RX7, a gene coding for a purinergic ligand-gated ion channel, is associated with major depressive disorder. *Hum. Mol. Genet.* **15**, 2438–2445 (2006).

64. Corponi, F. *et al.* Genetic basis of psychopathological dimensions shared between schizophrenia and bipolar disorder. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **89**, 23–29 (2019).

65. Steinberg, S. *et al.* Common variant at 16p11.2 conferring risk of psychosis. *Mol. Psychiatry* **19**, 108–114 (2014).

66. Sullivan, P. F. *et al.* Genomewide association for schizophrenia in the CATIE study: results of stage 1. *Mol. Psychiatry* **13**, 570–584 (2008).

67. Stroup, T. S. *et al.* The National Institute of Mental Health Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Project: Schizophrenia Trial Design and Protocol Development. *Schizophr. Bull.* **29**, 15–31 (2003).

68. Lieberman, J. A. *et al.* Effectiveness of Antipsychotic Drugs in Patients with Chronic Schizophrenia. *N. Engl. J. Med.* **353**, 1209–1223 (2005).

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

70. Bergen, S. E. *et al.* Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. *Mol. Psychiatry* **17**, 880–886 (2012).

71. Korver, N., Quee, P. J., Boos, H. B. M., Simons, C. J. P. & de Haan, L. Genetic Risk and Outcome of Psychosis (GROUP), a multi site longitudinal cohort study focused on gene-environment interaction: objectives, sample characteristics, recruitment and assessment methods. *Int. J. Methods Psychiatr. Res.* **21**, 205–221 (2012).

72. Rees, E. *et al.* Analysis of copy number variations at 15 schizophrenia-associated loci. *Br. J. Psychiatry* **204**, 108–114 (2014).

73. Pardiñas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).

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

75. Carroll, L. S. *et al.* Mutation screening of the 3q29 microdeletion syndrome candidate genes DLG1 and PAK2 in schizophrenia. *Am. J. Med. Genet. Part B Neuropsychiatr. Genet.* **156**, 844–849 (2011).

76. 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).

77. Egan, M. *et al.* Effect of COMT Val108/158 Met genotype on frontal lobe function and risk for schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 6917–6922 (2001).

78. 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

79. Li, Z. *et al.* Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nat. Genet.* **49**, 1576–1583 (2017).

80. Yu, H. *et al.* Common variants on 2p16.1, 6p22.1 and 10q24.32 are associated with schizophrenia in Han Chinese population. *Mol. Psychiatry* **22**, 954–960 (2017).

81. Shi, Y. *et al.* Common variants on 8p12 and 1q24.2 confer risk of schizophrenia. *Nat. Genet.* **43**, 1224–1227 (2011).

82. Yue, W.-H. *et al.* Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. *Nat. Genet.* **43**, 1228–31 (2011).

83. Bigdeli, T. B. *et al.* Contributions of common genetic variants to risk of schizophrenia among individuals of African and Latino ancestry. *Mol. Psychiatry* **25**, 2455–2467 (2020).

84. Castle, D. *et al.* The diagnostic interview for psychoses (DIP) : development, reliability and applications. *Psychol. Med.* **36**, 69–80 (2006).

85. GAIN Collaborative Research Group *et al.* New models of collaboration in genome-wide association studies: The Genetic Association Information Network. *Nature Genetics* **39**, 1045–1051 (2007).

86. Calkins, M. E. *et al.* The consortium on the genetics of endophenotypes in schizophrenia: Model recruitment, assessment, and endophenotyping methods for a multisite collaboration. *Schizophr. Bull.* **33**, 33–48 (2007).

87. Aliyu, M. H. *et al.* Project among African-Americans to explore risks for schizophrenia (PAARTNERS): Recruitment and assessment methods. *Schizophr. Res.* **87**, 32–44 (2006).