GWAS results for educational attainment aid in identifying genetic heterogeneity of schizophrenia

Authors: Vikas Bansal[1,2,3], Marina Mitjans[1,2,4], Casper A.P. Burik[5,6], Richard Karlsson Linnér[5,6], Aysu Okbay[5,6], Cornelius A. Rietveld[6,7], Martin Begemann[2,4,8], Stefan Bonn[3], Stephan Ripke[9,10,11], Ronald de Vlaming[5,6], Michel G. Nivard[12,13], Hannelore Ehrenreich[2,4,13], Philipp D. Koellinger[5,6,13]

Affiliations:

1 These co-authors contributed equally.

2 Clinical Neuroscience, Max Planck Institute of Experimental Medicine, Hermann-Rein-Straße 3, 37075, Göttingen, Germany

3 Research Group for Computational Systems Biology, German Center for Neurodegenerative Diseases (DZNE), Von-Siebold-Straße 3A, 37075 Göttingen, Germany

4 DFG Research Center for Nanoscale Microscopy and Molecular Physiology of the Brain (CNMPB), Humboldtallee 23, 30703, Göttingen, Germany

5 Department of Complex Trait Genetics, Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HV, Amsterdam, Netherlands

6 Erasmus University Rotterdam Institute for Behavior and Biology, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR, Rotterdam, Netherlands

7 Erasmus School of Economics, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR, Rotterdam, Netherlands

8 Department of Psychiatry & Psychotherapy, University of Göttingen, Von-Siebold-Straße 5, 37075, Göttingen, Germany

9 Analytic and Translational Genetics Unit, Massachusetts General Hospital, 02114 MA, Boston, USA

10 Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, 02142 MA, Cambridge, USA

11 Department of Psychiatry and Psychotherapy, Charité-Universitätsmedizin Berlin, Campus Mitte, Berlin, 10117, Germany

12 Department of Biological Psychology, Vrije Universiteit Amsterdam, van der Boechorststraat 1, 1081 BT, Amsterdam, Netherlands

13 These authors jointly directed this work.

Correspondence: Professor P.D. Koellinger, Complex Trait Genetics, Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HV, Amsterdam, Netherlands, E-mail: p.d.koellinger@vu.nl

ABSTRACT

Higher educational attainment (EA) is known to have a protective effect on the severity of schizophrenia (SZ). However, recent studies have found a small positive genetic correlation between EA and SZ. We investigated possible causes of this counterintuitive finding using genome-wide association (GWAS) results for EA and SZ (N = 443,581) and a replication cohort (1,169 controls and 1,067 cases) with high-quality SZ phenotypes. We found strong genetic dependence between EA and SZ that cannot be explained by chance, linkage disequilibrium, or assortative mating. Instead, several genes seem to have pleiotropic effects on EA and SZ, but without a clear pattern of sign concordance. Genetic heterogeneity in both phenotypes is the most likely explanation of this finding. This insight can be exploited by using a combination of EA and SZ GWAS results to improve the polygenic prediction of clinical symptoms and disease severity of SZ. In particular, although a polygenic score for SZ is robustly associated with case-control status, it does not predict any of the SZ symptoms or disease severity. In contrast, co-dependent polygenic scores that split the SZ score into two parts based on the sign concordance of SNPs for SZ and EA predict symptoms and disease severity in patients to some extent. Furthermore, using EA as a proxy-phenotype for SZ, we isolate *FOXO6* and *SLITRK1* as additional statistically plausible candidate genes for SZ.

INTRODUCTION

Schizophrenia (SZ) is the collective term used for a severe, highly heterogeneous and costly psychiatric disorder that is caused by environmental and genetic factors.[1–4] The latest genome-wide association study (GWAS) by the Psychiatric Genomics Consortium (PGC) identified 108 genomic loci that are associated with SZ.[5] These 108 loci jointly account for ≈3.4% of the variation on the liability scale to SZ,[5] while all single nucleotide polymorphisms (SNPs) that are currently measured by SNP arrays capture ≈64% (s.e. = 8%) in the variation in liability to the disease.[6] This implies that many genetic variants with small effect sizes contribute to the heritability of SZ, but most of them are unidentified as of yet. A polygenic score (PGS) based on all SNPs currently accounts for 4-15% of the variation on the liability scale to SZ.[5] Yet, this PGS does not predict any differences of symptoms or severity of the disease among SZ patients.[4] Partly, this could be due to the fact that the clinical disease classification of SZ spans across several different behavioral and cognitive traits that may not have identical genetic architectures. Therefore, identifying additional genetic variants and understanding through which pathways they are linked with the clinical diagnosis of SZ is an important step in understanding the etiologies of the 'schizophrenias'.[7] However, GWAS analyses of specific SZ symptoms would require very large sample sizes to be statistically well-powered, and the currently available datasets on deeply phenotyped SZ patients are not large enough yet for this purpose.

Here, we use an alternative approach to make progress with data that is readily available – by combining GWAS for SZ and educational attainment (EA). The GWAS sample sizes for EA are the largest for any cognition-related phenotype to date. Furthermore, previous studies suggest a complex relationship between EA and SZ[8,9] that may be used to gain additional insights into the genetic architecture of SZ and its symptoms. In particular, phenotypic data seem to suggest a *negative* correlation between EA and SZ.[10] For example, SZ patients with lower EA typically show an earlier age of disease onset, higher levels of psychotic symptomatology, and worsened global cognitive function.[10] In fact, EA has been suggested to be a measure of premorbid function and predictor of outcomes in SZ. Moreover, it has been forcefully argued that retarded intellectual development, global cognitive impairment during childhood, and bad school performance should be seen as core features of SZ that precede the

development of psychotic symptoms and differentiate SZ from bipolar disorder (BIP).[11–15] Furthermore, credible genetic links between SZ and impaired cognitive performance have been found.[16] In contrast to these findings, recent studies using large-scale GWAS results identified a small, but *positive genetic* correlation between EA and SZ ($\rho_{EA,SZ} = 0.08$)[8] and higher values of the PGS for SZ have been reported to be associated with creativity and greater educational attainment.[17] Other statistically well-powered studies found that high intelligence (IQ) has protective effects against SZ[18] and reported a negative genetic correlation between IQ and SZ ($\rho_{IQ,SZ} = -0.2$),[19] suggesting the possibility that genetic effects which contribute to EA *but not via IQ* are responsible for the observed positive genetic correlation between SZ and EA. Indeed, the latest GWAS on EA[8] already indicated that the genetic influence on higher schooling is not only mediated by IQ, but also by other factors such as behavioral inhibition and openness, which may be independently related to SZ and some of its symptoms.[20–22]

We propose that genetic heterogeneity in SZ is a possible reason for these partially contradictory results from previous studies. To explore this hypothesis and to discern it from alternative explanations, we performed a series of statistical genetic analyses using large-scale GWAS results for SZ and EA from non-overlapping samples. We find a non-monotonous genetic dependence between both traits that can be leveraged (i) to identify novel candidate loci for SZ and (ii) to improve the genetic prediction of SZ symptoms and disease severity. We annotate possible biological pathways, tissues, and cell types implied by genetic variants that are associated with both traits. Furthermore, we explore whether the genetic dependence between EA and SZ is due to pleiotropic effects of genes on both traits, linkage disequilibrium (LD) among SNPs, or assortative mating. In addition, we elucidate the relationship of EA and SZ with BIP and IQ. Finally, we test if our results are driven by individuals who were diagnosed with schizoaffective disorder (SD).

MATERIALS AND METHODS

All details are described in the **Supplementary Information**.

For our study, it is important to differentiate between genetic dependence and genetic correlation. In our context, genetic dependence means that the genetic variants associated with EA are more likely to be also associated with SZ than expected by chance. In contrast, genetic correlation is defined by the correlation of the (true) effect sizes of genetic variants on the two traits. Thus, genetic correlation implies a linear genetic relationship between two traits. Note that two traits can be genetically dependent even if they are not genetically correlated and *vice versa* (**Supplementary Section 1**). One possible cause of a non-linear genetic dependence is that at least one of the traits is genetically heterogeneous in the sense that it summarizes across endophenotypes (or symptoms) with non-identical genetic architectures.

GWAS

GWAS analyses on EA (n = 363,502)[8] and SZ (34,409 cases and 45,670 controls)[5] were carried out in non-overlapping samples of Europeans. The GRAS (Göttingen Research Association for Schizophrenia) replication sample[23] was not part of either GWAS (**Supplementary Section 2-3**).

Proxy-phenotype method (PPM)

We use the proxy-phenotype method (PPM)[24] to test if EA-associated loci are more likely to be associated with SZ than expected by chance. If the two traits are genetically dependent, PPM can increase the statistical power to identify loci associated with SZ because it reduces the multiple testing burden by using relevant information from the independent EA sample.[8,9,24] Our PPM analyses followed a pre-registered analysis plan (https://osf.io/dnhfk/). Analyses were carried out using 8,240,280 autosomal SNPs that passed quality controls in both GWAS and additional filters described in the **Supplementary Section 4**. We selected approximately independent lead-SNPs from the EA GWAS that passed the pre-defined significance threshold of $P_{EA} < 10^{-5}$ and looked up their SZ results. To test if EA-associated SNPs are more strongly associated with SZ than expected by chance, we conducted a Mann-Whitney test that compares the P_{SZ}-values of the EA-associated lead SNPs for SZ with the P_{SZ} -values of a set of randomly drawn, LD-independent SNPs with similar minor allele frequencies (**Supplementary Section 5-6**).

Distinguishing pleiotropy from LD

For each of the SNPs isolated by our PPM analysis, we looked at their neighboring SNPs within a +/- 500kb window and estimated their posterior probability of being causal for EA or SZ using PAINTOR.[25] We then selected two sets of SNPs, each of which contains the smallest number of SNPs which yields a cumulative posterior probability of 90% of containing the causal locus for EA and SZ, respectively. For each of these sets, we calculate the posterior probability that it contains the causal locus for the other trait. We classify the probability of a locus being pleiotropic as low (0-15%), medium (15-45%), or high (>45%) (**Supplementary Section 7**).

Biological annotations

To gain insights into possible biological pathways that are indicated by the PPM results, we applied DEPICT[8,26] using a false discovery rate (FDR) threshold of ≤ 0.05. To identify independent biological groupings, we used the affinity propagation method based on the Pearson distance matrix for clustering[27] (**Supplementary Section 8**).

LD-aware enrichment of PPM results across different traits

We developed an association enrichment test that corrects for the LD score of each SNP (**Supplementary Section 9**). We applied this test to the SNPs isolated by the PPM analyses. LD scores were obtained from the HapMap 3 European reference panel. We investigated SZ and 21 additional traits for which GWAS results were available in the public domain. Some of the traits were chosen because they are phenotypically related to SZ (e.g., BIP), while others were less obviously related to SZ (e.g., age at menarche) or served as negative controls (e.g., fasting insulin).

Replication of PPM results

Our replication sample, the GRAS data collection, is described in **Supplementary Section 10**. Following our pre-registered analysis plan (https://osf.io/dnhfk/), our replication uses a

PGS that is based on the 132 independent EA lead-SNPs that are also nominally associated with SZ ($P_{EA} < 10^{-5}$ and $P_{SZ} < 0.05$, **Supplementary Section 11**).

Polygenic prediction of schizophrenia symptoms in the GRAS sample

Our most direct test for genetic heterogeneity in SZ is based on PGS analyses that we carried out in our replication sample, the GRAS data collection, which contains exceptionally detailed measures of SZ symptoms.[4,7,23] We predicted years of education, age at prodrome, age at disease onset, premorbid IQ (approximated by a multiple-choice vocabulary test), global assessment of functioning (GAF), the clinical global impression of severity (CGI-S), as well as positive and negative symptoms (PANSS positive and negative, respectively) among SZ patients (*N* ranges from 903 to 1,039, see **Supplementary Sections 10 and 12**). If SZ would be a genetically homogenous trait, a PGS for SZ would be expected to predict symptoms and disease severity to some extent. If, however, SZ is genetically heterogeneous, there is potentially relevant information in the sign concordance of individual SNPs with EA traits that may improve the prediction of symptoms (see **Supplementary Section 1** for formal derivations). We use a simple way to do this here: First, we construct a PGS for SZ that contains one SNP per LD-block that is most strongly associated with SZ. Overall, this score (*SZ_all*) contains 349,357 approximately LD-independent SNPs. Next, we split *SZ_all* into two scores, based on sign-concordance of the SNPs with SZ and EA. More specifically, one score contains all estimated SZ effects of SNPs that have concordant signs for both traits (174,734 SNPs with ++ or -- on both traits, *Concordant*) while the other contains the estimated SZ effects of the remaining SNPs with discordant effects (174,623 SNPs with +- or -+, *Discordant*). We compare the predictive performance of models that include (i) only the *SZ_all* score, (ii) the *SZ_all* and the EA score (*EA_all*), and (iii) the *Concordant, Discordant*, and *EA_all* scores (**Supplementary Section 1.3.2**). As robustness a check, we repeated the analyses excluding patients that were diagnosed with schizoaffective disorder (*N* = 198). Furthermore, we also randomly split the *SZ_all* score 100 times into two equally large halves and tested if that improves prediction.

Distinguishing between schizophrenia and bipolar disorder

We further elucidate the genetic relationship between SZ and related traits (EA, BIP, childhood IQ, and neuroticism) by using Genome-Wide Inferred Statistics (GWIS)[28] to obtain GWAS regression coefficients and standard errors for SZ GWAS that are "purged" of their genetic correlation with BIP and *vice versa* (yielding "unique" $SZ_{(\text{min BIP})}$ and "unique" $BIP_{(\text{min SZ})}$ results, respectively). We computed genetic correlations of these GWIS results with EA, childhood IQ, and neuroticism using bivariate LD score regression[29] and compared the results to those obtained using ordinary SZ and BIP GWAS results (**Supplementary Section 13**).

Simulations of assortative mating

We conducted simulations to test if strong assortative mating on EA and SZ can induce a spurious genetic dependence between the two traits (**Supplementary Section 14**).

RESULTS

Proxy-phenotype analyses

Figure 1 presents an overview of the proxy-phenotype analyses. The first-stage GWAS on EA (**Supplementary Section 2**) identified 506 loci that passed our predefined threshold of $P_{EA} < 10^{-5}$ (https://osf.io/dnhfk/); 108 of them were genome-wide significant ($P_{EA} <$

5×10^{-8}, see **Supplementary Table 2**). Of the 506 EA lead-SNPs, 132 are associated with SZ at nominal significance ($P_{SZ} < 0.05$), and 21 of these survive Bonferroni correction ($P_{SZ} < \frac{0.05}{506} = 9.88 \times 10^{-5}$) (**Table 1**). LD score regression results suggest that the vast majority of the association signal in both the EA[8] and the SZ[5] GWAS are truly genetic signals, rather than spurious signals originating from uncontrolled population stratification. **Figure 2a** shows a Manhattan plot for the GWAS on EA highlighting SNPs that were also significantly associated with SZ (red crosses for $P_{SZ} < 0.05$, green crosses for $P_{SZ} = 9.88 \times 10^{-5}$).

A Q-Q plot of the 506 EA lead SNPs for SZ is shown in **Figure 2b**. Although the observed sign concordance of 52% is not significantly different from a random pattern ($P = 0.40$), we find 3.23 times more SNPs in this set of 506 SNPs that are nominally significant for SZ than expected given the distribution of the *P* values in the SZ GWAS results (raw enrichment $P = 6.87 \times 10^{-10}$, **Supplementary Section 6**). The observed enrichment of the 21 EA lead SNPs that pass Bonferroni correction for SZ ($P_{SZ} < \frac{0.05}{506} = 9.88 \times 10^{-5}$) is even more pronounced (27 times stronger, $P = 5.44 \times 10^{-14}$).

The effect sizes of these 21 SNPs on SZ are small, ranging from *Odds* = 1.02 (rs4500960) to *Odds* = 1.11 (rs4378243) after winner's curse correction (**Table 1**). However, Bayesian calculations with reasonable prior beliefs (e.g., 1% or 5%, **Supplementary Section 6**) suggest that most of these 21 SNPs are likely or virtually certain to be truly associated with SZ.

Prediction of future genome-wide significant loci for schizophrenia

Of the 21 variants we identified, 12 are in LD with loci previously reported by the PGC[5] and 2 are in the major histocompatibility complex (MHC) region on chromosome (chr) 6 and were therefore not separately reported in that study. Three of the variants we isolated (rs7610856, rs143283559, rs28360516) were independently found in a recent meta-analysis of the PGC results[5] with another large-scale sample.[30] We show in the **Supplementary Section 6** that using EA as a proxy-phenotype for SZ helped to predict the novel genome-wide significant findings reported in that study, which illustrates the power of the proxy-phenotype approach. Furthermore, two of the 21 variants (rs756912, rs7593947) are in LD with loci recently reported in a study that also compared GWAS findings from EA and SZ using smaller samples and a less conservative statistical approach.[31] The remaining 2 SNPs we identified (rs7336518 on chr13 and rs7522116 on chr1) add to the list of empirically plausible candidate loci for SZ.

Detection of shared causal loci

For eight of the 21 loci identified by the PPM analysis the credible set of causal loci for EA and SZ had a medium or high credibility to have direct causal effects on both EA and SZ (including rs7336518). Five of these loci have concordant effects on the two traits (i.e ++ or --) while three have a discordant effects (i.e +- or -+, **Supplementary Section 7**).

Biological annotations

Biological annotation of the 132 SNPs that are jointly associated with EA and SZ using DEPICT identified 111 significant reconstituted gene sets (**Supplementary Table 4.1**). Pruning these resulted in 19 representative gene sets including dendrites, axon guidance, transmission across chemical synapses, and abnormal cerebral cortex morphology (**Supplementary Table 4.2** and **Supplementary Figure 3.a**). All significantly enriched tissues are related to the nervous system and sense organs (**Supplementary Figure 3.b**).

Furthermore, "Neural Stem Cells" is the only significantly enriched cell-type (**Supplementary Table 4.3**). DEPICT prioritized genes that are known to be involved in neurogenesis and synapse formation (**Supplementary Table 4.4**). Some of the genes, including *SEMA6D* and *CSPG5*, have been suggested to play a potential role in SZ.[32,33] For the two novel candidate SNPs reported in this study (rs7522116 and rs7336518), DEPICT points to the *FOXO6* (Forkhead Box O6) and the *SLITRK1* (SLIT and NTRK Like Family Member 1) genes, respectively. *FOXO6* is predominantly expressed in the hippocampus and has been suggested to be involved in memory consolidation, emotion and synaptic function.[34,35] Similarly, *SLITRK1* is also highly expressed in the brain,[36] particularly localized to excitatory synapses and promoting their development,[37] and it has previously been suggested to be a candidate gene for neuropsychiatric disorders.[38]

LD-aware enrichment across different traits

Supplementary Figure 4 and **Supplementary Table 5.1** show the LD-aware enrichment of the SNPs that are jointly associated with EA and SZ across 22 traits. We find significant joint LD-aware enrichment for SZ, further confirming that the PPM results are not entirely driven by LD. We also find LD-aware enrichment for BIP, neuroticism, childhood IQ, inflammatory bowel disease, and age at menarche. However, we find no LD-aware enrichment for other brain-traits that are phenotypically related to SZ, such as depressive symptoms, subjective well-being, autism, and attention deficit hyperactivity disorder. We also do not find LD-aware enrichment for most traits that are less obviously related to the brain and our negative controls. Furthermore, one of the novel SNPs we isolated shows significant LD-aware enrichment both for SZ and for BIP (rs7522116). The results suggest that the loci identified by the PPM are not simply related to all (brain) traits. Instead, they show some degree of phenotype-specificity.

Replication in the GRAS sample

A PGS based on the 132 loci jointly associated with both EA and SZ (*SZ_132*) adds $\Delta R^2 = 7.54\% - 7.01\% = 0.53\%$ predictive accuracy for the SZ case-control status to a PGS (*SZ_all*) derived from the GWAS on SZ alone ($P = 1.7 \times 10^{-4}$, **Supplementary Table 7.2.a**, Model 3). The *SZ_132* score also significantly adds ($P = 3.4 \times 10^{-4}$) to the predictive accuracy of the SZ case-control status when all other scores we constructed are included as control variables (**Supplementary Table 7.2.a,** Model 9).

Prediction of schizophrenia measures in the GRAS sample of patients

Phenotypic correlations in the GRAS sample show that higher education is associated with later age at prodrome, later onset of disease, and less severe disease symptoms among SZ patients (**Supplementary Section 12**, **Supplementary Table 8.1 and Supplementary Figure 5**). The *EA_all* PGS is associated with years of education ($P = 1.0 \times 10^{-6}$) and premorbid IQ ($P = 2.7 \times 10^{-4}$) among SZ patients (**Supplementary Section 12** and **Table 2**). Consistent with earlier results,[4] we find that none of the SZ measures can be predicted by the PGS for SZ (*SZ_all*, **Table 2**). Yet, splitting the PGS for SZ based on the sign-concordance of SNPs with EA (*Concordant* and *Discordant*) increases predictive accuracy significantly for severity of disease (GAF ($p_F = 0.023$)) and symptoms (PANSS negative ($p_F = 0.007$)) (**Table 2**). This increase in predictive accuracy is evidence for genetic heterogeneity in SZ (**Supplementary Section 1**). Specifically, our results indicate that those with a high genetic propensity for EA have better assessments of global functioning (GAF) and less severe negative symptoms (PANSS negative). However, if the high genetic predisposition for EA is primarily due to loci that also increase the risk for SZ (i.e. high values on the *Concordant* score), this protective effect is attenuated. The highest gain in

predictive accuracy for SZ symptoms from splitting the *SZ_all* by sign concordance with EA is currently observed for PANSS negative ($\Delta R^2 = 0.63\%$). We repeated these analyses excluding patients who were diagnosed with schizoaffective disorder (SD) and found similar results, implying that our findings are not only due to the presence of patients with SD (**Supplementary Section 12**, **Supplementary Table 8.4.a**). Randomly splitting the *SZ_all* score does not yield any gains in predictive accuracy (**Supplementary Section 12 and Supplementary Table 8.5)**.

Controlling for the genetic overlap between schizophrenia and bipolar disorder

The genetic correlations between $SZ_{(\text{min BIP})}$ with EA and IQ are negative and significant ($r_g = -0.16$, $P = 3.88\times10^{-04}$ and $r_g = -0.31$, $P = 6.00\times10^{-03}$ respectively), while the genetic correlation of $BIP_{(\text{min SZ})}$ with EA and IQ remain positive ($r_g \approx 0.3$) (**Figure 3, Supplementary Table 9.2**).

Simulations of assortative mating

Our simulation results suggest it is unlikely that assortative mating is a major cause for the genetic dependence we observe between EA and SZ (**Supplementary Figure 7**).

DISCUSSION

We demonstrate strong genetic dependence between EA and SZ, but without a clear pattern of sign concordance. Our results cannot be explained by chance, linkage disequilibrium, or assortative mating. Using EA as a proxy-phenotype, we isolated 21 genetic loci for SZ and two novel candidate genes, *FOXO6* and *SLITRK1*. Eight of the 21 loci seem to have true pleiotropic effects on EA and SZ. Furthermore, we showed that EA GWAS results help to predict future GWAS findings for SZ, despite the low genetic correlation between both traits. Moreover, splitting the PGS for SZ into two scores based on the sign concordance of SNPs with EA enables the prediction of disease symptoms and severity from genetic data for the first time to some extent. This result is not driven by patients with SD and it cannot be repeated by randomly splitting the SZ score. Finally, we showed that the small positive genetic correlation of EA and SZ seems to be driven by loci that also increase the risk for BIP, whereas the genetic correlation between EA and "unique" $SZ_{(\text{min BIP})}$ is actually negative. This is consistent with Kraeplin's original description of dementia praecox and with more recent accounts which suggest that the main difference between SZ and BIP is that SZ is a neurodevelopmental disorder, whereas BIP is not.[11–15]

Our results are most consistent with the idea that EA[8] and SZ are both genetically heterogeneous traits. Specifically, our results raise the possibility that the clinical diagnosis of SZ aggregates over at least two subtypes with non-identical symptoms and genetic architectures: One part resembles BIP and high IQ (possibly associated with *Concordant* SNPs and personality traits such as openness or behavioral inhibition), while the other part is a cognitive disorder that is independent of BIP (possibly influenced by *Discordant* SNPs). Furthermore, our results point to possible side-effects of pharmacogenetic interventions that may aim to target pleiotropic genes.

AUTHOR CONTRIBUTIONS

P.D.K. designed and oversaw the study and conducted proxy-phenotype analyses. V.B. and M.M. carried out analyses in the GRAS sample. R.V., C.A.P.B., M.N, and P.D.K. developed statistical methods. V.B. conducted bioinformatics and computed the LD-aware enrichment tests. C.A.P.B. and R.V. conducted simulation analyses. M.N. computed GWIS results, genetic correlations, and carried out pleiotropy analyses. R.K.L. assisted with biological annotation and visualization of results. P.D.K., V.B., M.M., and H.E. made especially major contributions to writing and editing. All authors contributed to and critically reviewed the manuscript.

ACKNOWLEDGMENTS

This research was carried out under the auspices of the Social Science Genetic Association Consortium (SSGAC), including use of the UK Biobank Resource. We thank all research consortia that provide access to GWAS summary statistics in the public domain. Specifically, we acknowledge data access from the Psychiatric Genomics Consortium (PGC), the Genetic Investigation of ANthropometric Traits Consortium (GIANT), the International Inflammatory Bowel Disease Genetics Consortium (IIBDGC), the International Genomics of Alzheimer's Project (IGAP), the CARDIoGRAMplusC4D Consortium, the Reproductive Genetics Consortium (ReproGen), the Tobacco and Genetics Consortium (TAG), the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC), the ENIGMA Consortium, and the Childhood Intelligence Consortium (CHIC). We would like to thank the customers and employees of 23andMe for making this work possible as well as Joyce J. Tung, Nick. A. Furlotte, and David. A Hinds from the 23andMe research team. This study was supported by funding from an ERC Consolidator Grant (647648 EdGe, Philipp D Koellinger), the Max Planck Society, the Max Planck Förderstiftung, the DFG (CNMPB), EXTRABRAIN EU-FP7, the Niedersachsen-Research Network on Neuroinfectiology (N-RENNT), and EU-AIMS. Michel G Nivard was supported by Royal Netherlands Academy of Science Professor Award to Dorret I Boomsma (PAH/6635). Additional acknowledgements are provided in the Supplementary Information.

CONFLICT OF INTEREST

The authors declare no conflict of interests.

REFERENCES

1 Knapp M, Mangalore R, Simon J. The global costs of schizophrenia. *Schizophr Bull* 2004; **30**: 279–293.

2 Sullivan PF, Kendler KS, Neale MC, KS K, SB T, DB P *et al.* Schizophrenia as a complex trait. *Arch Gen Psychiatry* 2003; **60**: 1187.

3 Polderman TJC, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet* 2015; **47**: 702–709.

4 Stepniak B, Papiol S, Hammer C, Ramin A, Everts S, Hennig L *et al.* Accumulated environmental risk determining age at schizophrenia onset: a deep phenotyping-based study. *The Lancet Psychiatry* 2014; **1**: 444–453.

5 Ripke S, Neale BM, Corvin A, Walters JTR, Farh K-H, Holmans PA *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**: 421–427.

6 Bhatia G, Gusev A, Loh P, Vilhjálmsson BJ, Ripke S, PGC *et al.* Haplotypes of common SNPs can explain missing heritability of complex diseases. 2015http://dx.doi.org/10.1101/022418.

7 Ehrenreich, H; Mitjans, M; Van der Auwera, S; Centeno, TP; Begemann, M; Grabe, HJ; Bonn, S; Nave K-A. OTTO: a new strategy to extract mental disease-relevant combinations of GWAS hits from individuals. *Mol Psychiatry* 2017. doi:10.1038/mp.2016.208.

8 Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* 2016; **533**: 539–542.

9 Okbay A, Baselmans BML, Neve J-E De, Turley P, Nivard MG, Fontana MA *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet* doi:10.1038/ng.3552.

10 Swanson CL, Gur RC, Bilker W, Petty RG, Gur RE. Premorbid educational attainment in schizophrenia: association with symptoms, functioning, and neurobehavioral measures. *Biol Psychiatry* 1998; **44**: 739–747.

11 Kahn RS, Keefe RSE, JD H, B E, GM K, H D *et al.* Schizophrenia is a cognitive illness. *JAMA Psychiatry* 2013; **70**: 1107.

12 Kraepelin E. *Psychiatrie: Ein Lehrbuch für Studierende und Ärzte*. 4th ed. Verlag von Johann Ambrosius Barth: Leipzig, Germany, 1893.

13 Trotta A, Murray R, MacCabe J. Do premorbid and post-onset cognitive functioning differ between schizophrenia and bipolar disorder? A systematic review and meta-analysis. *Psychol Med* 2015; **45**: 381–394.

14 Murray RM, Sham P, Van Os J, Zanelli J, Cannon M, McDonald C. A developmental model for similarities and dissimilarities between schizophrenia and bipolar disorder. *Schizophr Res* 2004; **71**: 405–416.

15 Murray RM, O'Callaghan E, Castle DJ, Lewis SW. A Neurodevelopmental Approach to the Classification of Schizophrenia. *Schizophr Bull* 1992; **18**: 319–332.

16 Stefansson H, Meyer-Lindenberg A, Steinberg S, Magnusdottir B, Morgen K,

Arnarsdottir S *et al.* CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature* 2013; **505**: 361–366.

17 Power RA, Steinberg S, Bjornsdottir G, Rietveld CA, Abdellaoui A, Nivard MM *et al.* Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nat Neurosci* 2015; **18**: 953–955.

18 Kendler KS, Ohlsson H, Sundquist J, Sundquist K. IQ and Schizophrenia in a Swedish National Sample: Their Causal Relationship and the Interaction of IQ With Genetic Risk. *Am J Psychiatry* 2015; **172**: 259–265.

19 Sniekers S, Stringer S, Watanabe K, Jansen PR, Coleman JRI, Krapohl E *et al.* Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat Genet* 2017. doi:doi:10.1038/ng.3869.

20 Camisa KM, Bockbrader MA, Lysaker P, Rae LL, Brenner CA, O'Donnell BF. Personality traits in schizophrenia and related personality disorders. *Psychiatry Res* 2005; **133**: 23–33.

21 Beauchamp M-C, Lecomte T, Lecomte C, Leclerc C, Corbière M. Do people with a first episode of psychosis differ in personality profiles? 2006 doi:10.1016/j.schres.2006.03.026.

22 Bagby RM, Young LT, Schuller DR, Bindseil KD, Cooke RG, Dickens SE *et al.* Bipolar disorder, unipolar depression and the Five-Factor Model of personality. *J Affect Disord* 1996; **41**: 25–32.

23 Ribbe K, Friedrichs H, Begemann M, Grube S, Papiol S, Kästner A *et al.* The cross-sectional GRAS sample: A comprehensive phenotypical data collection of schizophrenic patients. *BMC Psychiatry* 2010; **10**: 91.

24 Rietveld CA, Esko TT, Davies G, Pers TH, Turley PA, Benyamin B *et al.* Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc Natl Acad Sci U S A* 2014; **111**: 13790–13794.

25 Kichaev G, Roytman M, Johnson R, Eskin E, Lindström S, Kraft P *et al.* Improved methods for multi-trait fine mapping of pleiotropic risk loci. *Bioinformatics* 2017; **33**: 248–255.

26 Pers TH, Karjalainen JM, Chan Y, Westra H-JH-J, Wood AR, Yang J *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* 2015; **6**: 5890.

27 Frey BJ, Dueck D. Clustering by passing messages between data points. *Science (80-)* 2007; **315**.

28 Nieuwboer HA, Pool R, Dolan CV, Boomsma DI, Nivard MG. GWIS: Genome-wide inferred statistics for functions of multiple phenotypes. *Am J Hum Genet* 2016; **99**: 917–927.

29 Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Consortium R *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* 2015; **47**: 1236–1241.

30 Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and maintained by background selection. 2016http://dx.doi.org/10.1101/068593.

31 Le Hellard S, Wang Y, Witoelar A, Zuber V, Bettella F, Hugdahl K *et al.*

Identification of gene loci that overlap between schizophrenia and educational attainment. *Schizophr Bull* 2016. doi:10.1093/schbul/sbw085.

32 So H-C, Fong PY, Chen RYL, Hui TCK, Ng MYM, Cherny SS *et al.* Identification of neuroglycan C and interacting partners as potential susceptibility genes for schizophrenia in a Southern Chinese population. *Am J Med Genet Part B Neuropsychiatr Genet* 2010; **153B**: 103–113.

33 Arion D, Horváth S, Lewis DA, Mirnics K. Infragranular gene expression disturbances in the prefrontal cortex in schizophrenia: signature of altered neural development? *Neurobiol Dis* 2010; **37**: 738–46.

34 Salih DAM, Rashid AJ, Colas D, de la Torre-Ubieta L, Zhu RP, Morgan AA *et al.* FoxO6 regulates memory consolidation and synaptic function. *Genes Dev* 2012; **26**: 2780–2801.

35 Maiese K. FoxO Proteins in the Nervous System. *Anal Cell Pathol* 2015; **2015**: 1–15.

36 Aruga J, Yokota N, Mikoshiba K. Human SLITRK family genes: genomic organization and expression profiling in normal brain and brain tumor tissue. *Gene* 2003; **315**: 87–94.

37 Beaubien F, Raja R, Kennedy TE, Fournier AE, Cloutier J-F, Ichtchenko K *et al.* Slitrk1 is localized to excitatory synapses and promotes their development. *Sci Rep* 2016; **6**: 27343.

38 Proenca CC, Gao KP, Shmelkov S V, Rafii S, Lee FS, Aruga J *et al.* Slitrks as emerging candidate genes involved in neuropsychiatric disorders. *Trends Neurosci* 2011; **34**: 143–53.

Table 1: SNPs significantly associated with schizophrenia after Bonferroni correction.

	SNP-ID	*EA* (beta)	*signs concordant?*	*SZ-R^2* (adj)	*SZ-Odds* (adj)	*EAF*	Power ($\alpha = 0.05/506$)	Posterior probability of true association with SZ, Prior belief (π) 0.1%	1.0%	5.0%	10.0%
1	rs79210963	-0.016	yes	0.021%	0.931	0.89	22.9%	75.0%	96.8%	99.3%	99.7%
2	rs7610856	0.013	no	0.022%	0.955	0.41	22.8%	74.9%	96.8%	99.3%	99.7%
3	rs10896636	0.012	no	0.020%	0.956	0.67	17.8%	68.7%	95.6%	99.1%	99.5%
4	rs756912	-0.015	yes	0.022%	0.956	0.51	22.7%	74.8%	96.7%	99.3%	99.7%
5	rs6449503	0.018	no	0.020%	0.961	0.51	12.9%	60.0%	93.7%	98.7%	99.3%
6	**rs7336518**	-0.016	yes	**0.014%**	**0.964**	**0.13**	**1.5%**	**13.4%**	**60.6%**	**88.5%**	**93.9%**
7	rs143283559	0.014	no	0.017%	0.965	0.72	4.6%	32.8%	83.0%	96.1%	98.0%
8	rs11210935	0.015	no	0.014%	0.973	0.77	1.2%	10.9%	55.1%	86.0%	92.5%
9	rs77000541	-0.014	yes	0.018%	0.974	0.33	1.6%	14.1%	62.2%	89.2%	94.3%
10	rs2819344	0.014	no	0.017%	0.983	0.62	0.3%	3.0%	23.3%	60.4%	75.3%
11	rs4500960	-0.013	no	0.017%	1.017	0.47	0.3%	3.0%	23.3%	60.4%	75.3%
12	rs28360516	-0.012	no	0.013%	1.027	0.70	1.4%	12.6%	59.0%	87.8%	93.5%
13	**rs7522116**	0.011	yes	**0.015%**	**1.029**	**0.56**	**3.0%**	**23.8%**	**75.8%**	**94.0%**	**96.9%**
14	rs7593947	0.014	yes	0.018%	1.040	0.51	12.5%	59.1%	93.5%	98.6%	99.3%
15	rs11694989	0.011	yes	0.021%	1.044	0.43	17.9%	68.8%	95.7%	99.1%	99.5%
16	rs320700	0.013	yes	0.024%	1.054	0.65	36.4%	85.3%	98.3%	99.7%	99.8%
17	rs3957165	0.015	yes	0.020%	1.056	0.83	14.7%	63.6%	94.6%	98.9%	99.4%
18	rs10791106	0.011	yes	0.026%	1.056	0.54	46.9%	89.9%	98.9%	99.8%	99.9%
19	rs2992632	0.016	yes	0.025%	1.060	0.74	36.8%	85.5%	98.3%	99.7%	99.8%
20	rs10773002	0.022	yes	0.043%	1.087	0.28	91.0%	99.0%	99.9%	100.0%	100.0%
21	rs4378243	0.019	yes	0.044%	1.112	0.85	91.5%	99.1%	99.9%	100.0%	100.0%

Notes: The SNPs in the table are order by their *Odds* ratio on schizophrenia. Effect sizes for schizophrenia (in R^2 and *Odds*) are downward adjusted for the winner's curse. R^2 was approximated from the winner's curse adjusted *Odds* ratios, using the formulas described in **Supplementary Section 6.2**. The winner's curse adjustment took into account that only SNPs with $P = 0.05/506$ were selected. Power calculations assumed that the available GWAS sample size for schizophrenia for each SNP consisted of 34,409 cases and 45,670 controls. EAF is the effect allele frequency in the schizophrenia GWAS data. SNPs highlighted in bold are associations for schizophrenia that have not been emphasized in the previous literature. EA (beta) is the beta value from educational attainment GWAS. SNPs with concordant effects on both schizophrenia and educational attainment are marked as "yes" in sign concordance column.

Table 2: Polygenic risk prediction of schizophrenia outcomes in the GRAS sample of schizophrenia patients.

		Years of education[1]	Age at prodrome	Age at disease onset	Premorbid IQ[1]	GAF[2]	CGI-S[2]	PANSS positive[2]	PANSS negative[2]
Baseline Model									
SZ_all	*standardized beta*	0.001	-0.041	-0.056	-0.063	-0.024	0.041	0.033	0.043
	P value	0.976	0.297	0.129	0.090	0.510	0.249	0.364	0.253
EA_all	*standardized beta*	0.182**	0.005	-0.002	0.149**	0.068*	-0.057	0.001	-0.051
	P value	$4.4x10^{-09}$	0.884	0.961	$7.2x10^{-06}$	0.029	0.065	0.981	0.107
	Adj. R^2	0.0612	0.0023	0.0047	0.0417	0.0655	0.0816	0.0711	0.0243
	Δ Adj. R^2#	0.0312	-0.0010	-0.0009	0.0209	0.0035	0.0023	-0.0010	0.0015
Split Model									
Concordant	*standardized beta*	-0.013	-0.019	-0.031	-0.043	-0.096*	0.050	0.079	0.125**
	P value	0.751	0.665	0.456	0.326	0.022	0.232	0.059	0.0036
Discordant	*standardized beta*	0.014	-0.030	-0.035	-0.034	0.066	<0.001	-0.039	-0.072
	P value	0.730	0.515	0.409	0.437	0.112	0.996	0.351	0.090
EA_all	*standardized beta*	0.191**	0.002	-0.002	0.153**	0.122**	-0.074	-0.039	-0.118**
	P value	$1.0x10^{-06}$	0.965	0.953	$2.7x10^{-04}$	0.002	0.058	0.319	0.003
	Adj. R^2	0.0604	0.0012	0.0037	0.0406	0.0694	0.0811	0.0728	0.0306
	Δ Adj. R^2#	0.0304	-0.0021	-0.0019	0.0198	0.0074	0.0018	0.0007	0.0078
	n	1,039	915	1,043	903	1,010	1,014	1,009	1,002
ΔR^2 (Baseline Model – Split Model)		-0.0008	-0.0011	-0.0010	-0.0011	0.0039	-0.0005	0.0017	0.0063
P value from F-test°		0.698	0.907	0.968	0.891	0.023*	0.479	0.098	0.007**

Notes: Linear regression using the first 10 genetic principal components as control variables. [1]: Age of onset was included as covariate. [2]: Medication was included as covariate. # Change in *Adj. R^2* of the models compared to a model that only contains the *SZ_all* score and the control variables. ° *P value* from *F-test* refers to improvement in split model compared to baseline model. *denotes significance at $P < 0.05$. **denotes significance at $P < 0.01$.

FIGURE LEGENDS

Figure 1: Workflow of the proxy-phenotype analyses.

Notes: Educational attainment (EA) and schizophrenia (SZ) GWAS results are based on the analyses reported in ref. [5,8]. All cohorts that were part of the SZ GWAS were excluded from the meta-analysis on EA. The GRAS data collection was not included in either the SZ or the EA meta-analysis. Proxy-phenotype analyses were conducted using 8,240,280 autosomal SNPs that passed quality control. Genetic outliers of non-European decent (N = 13 cases) were excluded from the analysis in the GRAS data collection.

Figure 2: Results of the proxy-phenotype analyses.

Notes: Panel a: **Manhattan plot for educational attainment (EA) associations (n = 363,502).** The x axis is the chromosomal position, and the y axis is the significance on the $-\log_{10}$ scale (2-sided). The black dashed line shows the suggestive significance level of 10^{-5} that we specified in our preregistered analysis plan. Red and green crosses identify EA-associated lead-SNPs that are also associated with SZ at nominal or Bonferroni-adjusted significance levels, respectively.
Panel b: **Q–Q plot of the 506 EA-associated SNPs for schizophrenia (SZ) (n = 34,409 cases and n = 45,670 controls).** SNPs with concordant effects on both phenotypes are pink, and SNPs with discordant effects are blue. SNPs outside the grey area (21 SNPs) pass the Bonferroni-corrected significance threshold that corrects for the total number of SNPs we tested ($P < 0.05/506 = 9.88\times10^{-5}$) and are labelled with their rs numbers. Observed and expected P values are on the $-\log_{10}$ scale. For the sign concordance test: P = 0.40, 2-sided.

Figure 3: Genetic correlations of GWAS and GWIS results that are central to the relationship between SZ and EA.

Notes: The heatmap displays the genetic correlations across 7 sets of GWAS or GWIS summary statistics. Genetic correlations were estimated with LD score regression.[29] The color scale represents the genetic correlations ranging from –1 (red) to 1 (blue). Asterisk denotes that the genetic correlation is significant at P value < 0.01.

Figure 1

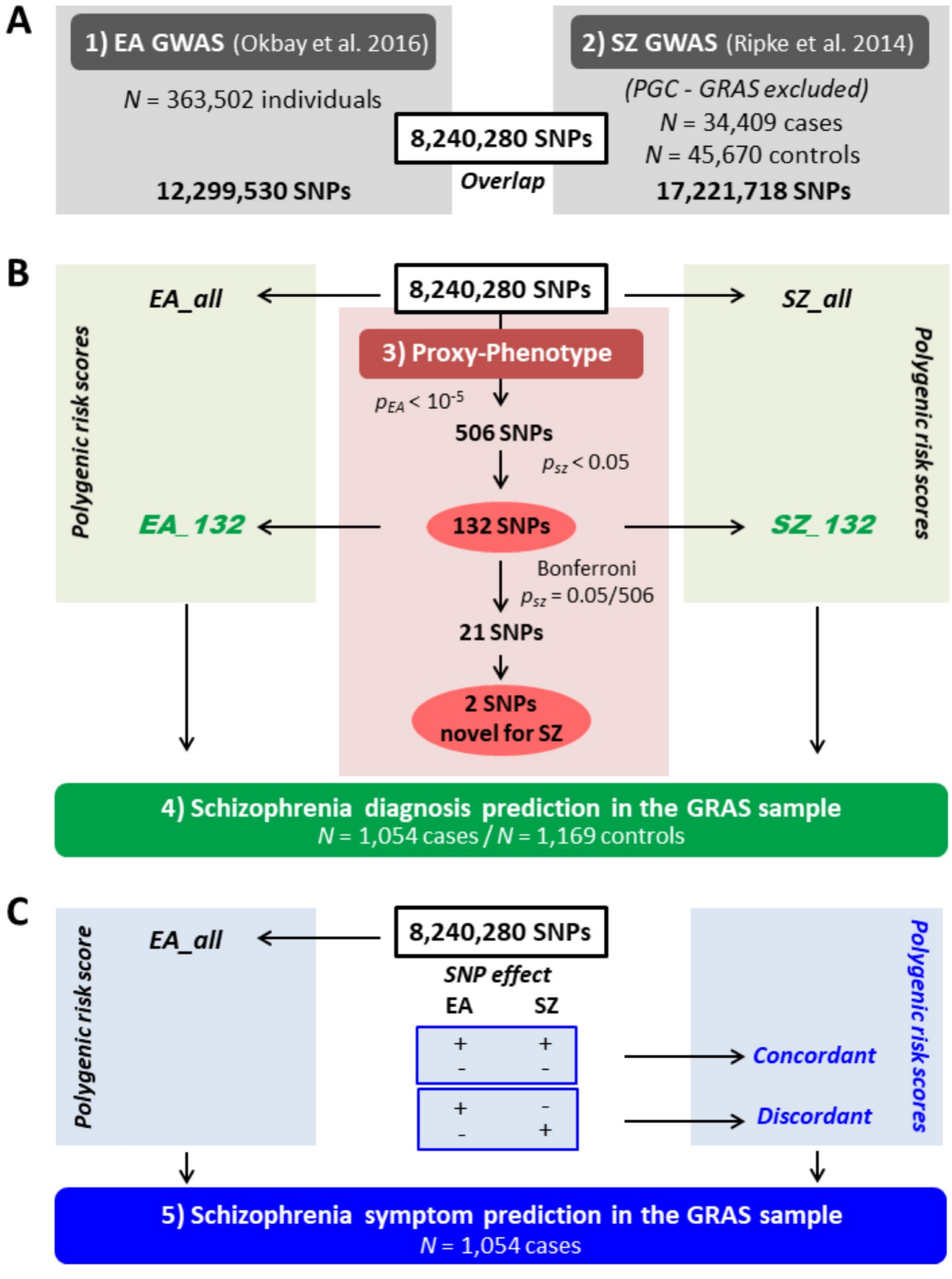

Figure 2a

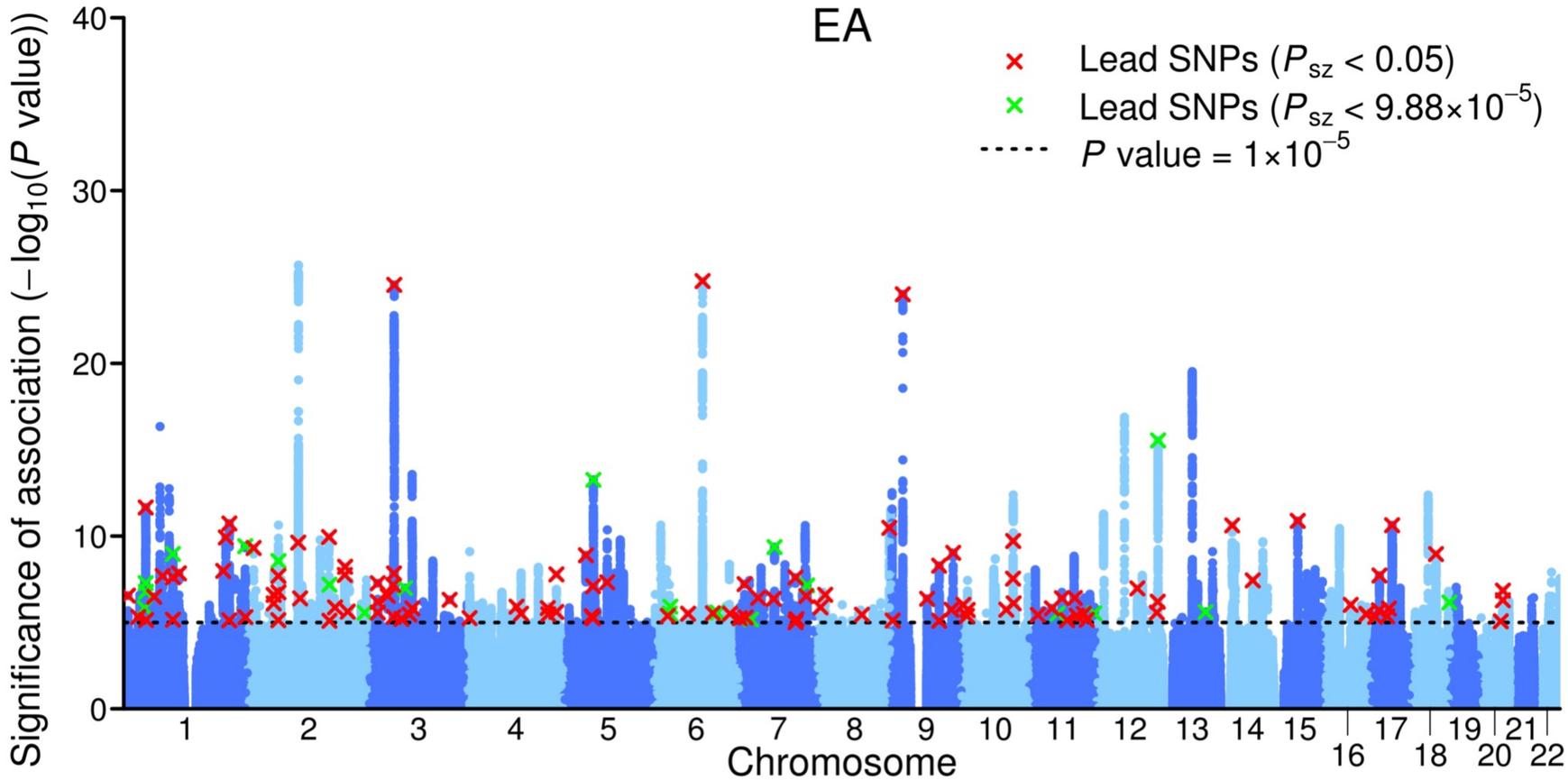

Figure 2b

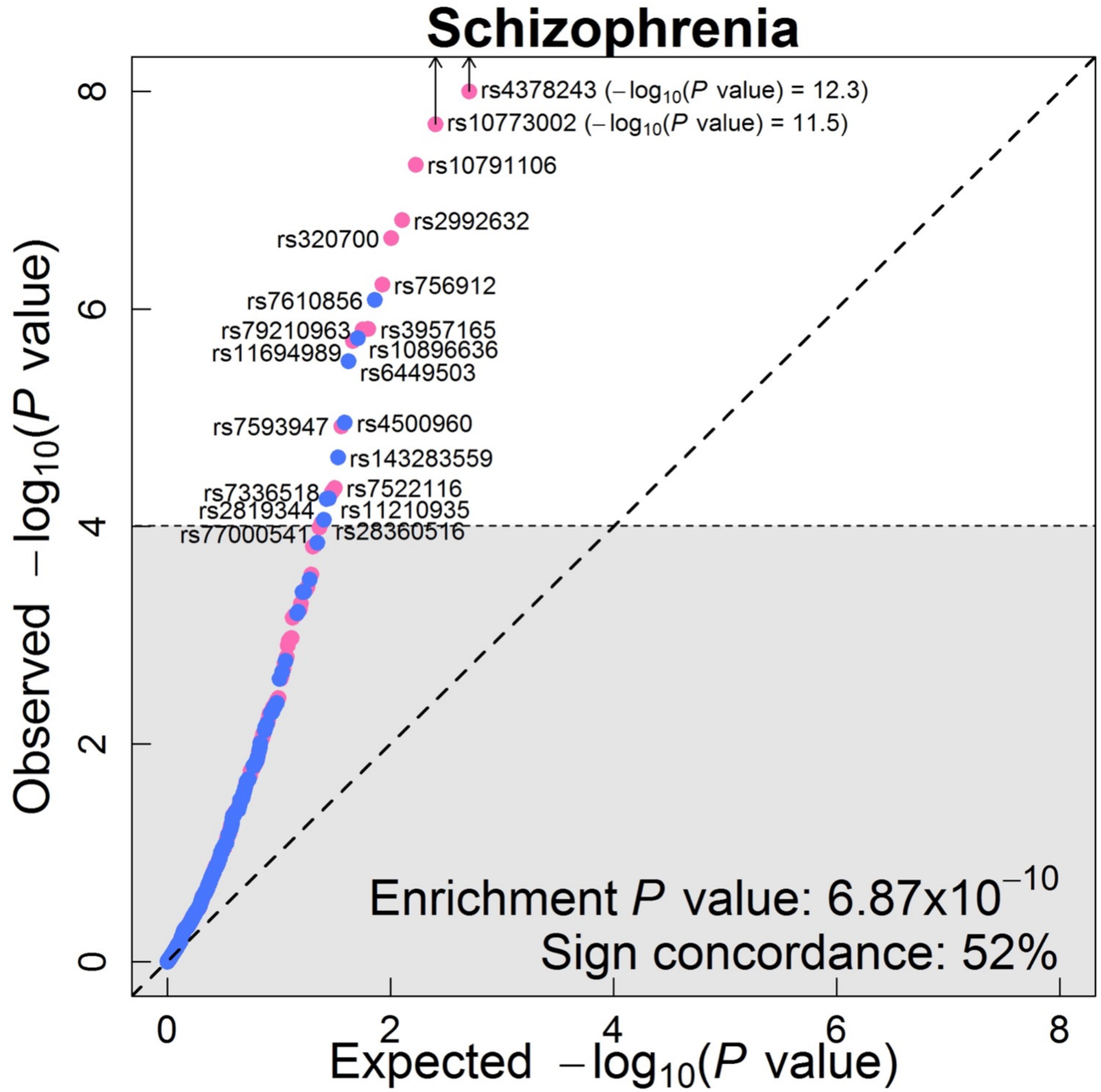

Figure 3