Check for updates

ARTICLE   OPEN

The impact of educational attainment, intelligence and intellectual disability on schizophrenia: a Swedish population-based register and genetic study

Jie Song [1], Shuyang Yao [1], Kaarina Kowalec [1,2], Yi Lu [1], Amir Sariaslan[3], Jin P. Szatkiewicz [4], Henrik Larsson[1,5], Paul Lichtenstein [1], Christina M. Hultman[1,6] and Patrick F. Sullivan [1,4 ✉]

Schizophrenia (SCZ) is highly heterogenous and no subtypes characterizing treatment response or longitudinal course well. Cognitive impairment is a core clinical feature of SCZ and a determinant of poorer outcome. Genetic overlap between SCZ and cognitive traits is complex, with limited studies of comprehensive epidemiological and genomic evidence. To examine the relation between SCZ and three cognitive traits, educational attainment (EDU), premorbid cognitive ability, and intellectual disability (ID), we used two Swedish samples: a national cohort (14,230 SCZ cases and 3,816,264 controls) and a subsample with comprehensive genetic data (4992 cases and 6009 controls). Population-based analyses confirmed worse cognition as a risk factor for SCZ, and the pedigree and SNP-based genetic correlations were comparable. In the genotyped cases, those with high EDU and premorbid cognitive ability tended to have higher polygenetic risk scores (PRS) of EDU and intelligence and fewer rare exonic variants. Finally, by applying an empirical clustering method, we dissected SCZ cases into four replicable subgroups characterized by EDU and ID. In particular, the subgroup with higher EDU in the national cohort had fewer adverse outcomes including long hospitalization and death. In the genotyped subsample, this subgroup had higher PRS of EDU and no excess of rare genetic burdens than controls. In conclusion, we found extensive evidence of a robust relation between cognitive traits and SCZ, underscoring the importance of cognition in dissecting the heterogeneity of SCZ.

INTRODUCTION

Schizophrenia (SCZ) is an often devastating psychiatric disorder associated with substantially elevated rates of impaired social functioning, morbidity, premature mortality, and personal and societal costs [1–4]. SCZ aggregates in families with a sibling recurrence risk ratio of 8.6 which is primarily due to shared genetic influences (twin/pedigree heritability 0.60–0.80) [5–8]. Approximately a third of the twin/pedigree heritability can be attributed to common single-nucleotide polymorphisms (SNP-heritability 0.24) [9, 10]. It is now firmly established that both common and rare genetic variation influence the risk of SCZ, as in most other complex diseases [11–15].

Impaired cognitive ability is an important clinical feature of SCZ [16] and a determinant of poorer outcome [17]. Lower premorbid cognitive ability is a risk factor for SCZ [18] and cognitive ability can decline after SCZ onset [19]. Intellectual disability (ID) is defined by marked impairment in cognitive ability and is an important comorbidity of SCZ [20].

The genetic relationship between cognitive ability and SCZ is complex. Common genetic variants contribute to both cognitive traits and rare severe neurodevelopmental disorders including ID.

Recent studies have found shared loci between intelligence and SCZ along with a negative SNP-based genetic correlation ($r_g = -0.21$) and Mendelian randomization analyses suggested bidirectional causal effects [21, 22]. Previous studies reported a positive genetic correlation between SCZ and EDU that was attributed to the genetic overlap between SCZ and bipolar disorder (BIP) [23, 24]. However, this correlation was zero by the most recent GWAS [9, 25].

Genetic overlap between SCZ and ID for rare genetic variants of strong effect also exists. Rare predicted loss-of-function (pLoF) exonic variants in *SETD1A* are associated with SCZ and developmental/cognitive delay [26]. Recent research also suggests the associations between pLoF variants (particularly in brain expressed genes) and SCZ [12] as well as educational attainment (EDU) [27]. The cumulative burden of rare pLoF variants is enriched in SCZ cases with comorbid ID, and can predict SCZ risk in individuals without ID [28]. Moreover, unaffected carriers of rare neuropsychiatric copy number variants (CNVs) had cognitive ability intermediate between controls and CNV carriers with SCZ [29].

Given that the extant data strongly hint at important interrelations between genetic risk for SCZ and cognitive traits,

[1]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [2]College of Pharmacy, University of Manitoba, Winnipeg, MB, Canada. [3]Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK. [4]Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, NC, USA. [5]School of Medical Sciences, Örebo University, Örebo, Sweden. [6]Icahn School of Medicine, Department of Psychiatry, Mt Sinai Hospital, New York, NY, USA. ✉email: patrick.sullivan@ki.se

we investigated whether cognitive ability and ID might usefully index the etiological or phenotypic heterogeneity of SCZ. We did this by studying two samples, one based on an entire country and second, using a large subset of that country with comprehensive genomic data. First, we evaluated the associations between SCZ, ID, and measures of cognition (EDU and premorbid cognitive ability) using Swedish national register data. Second, we estimated their heritabilities and genetic correlations via a Swedish national sibling cohort. Third, in the genotyped subsample, we assessed whether common and rare genetic variant burden measures (polygenetic risk scores or PRS, CNV burden, and rare exonic burden) usefully added to the results from the national sample [7, 12, 30–33]. Finally, we applied empirical clustering methods to cognitive-related factors to identify SCZ subgroups. To our knowledge, no prior report has considered the relation of cognitive ability, ID, and SCZ while incorporating multiple measures of common and rare genetic variation.

METHODS
Swedish National Sample
Statistics Sweden maintains national registries containing health service use and governmental data. Unique person numbers (assigned to all Swedish residents at birth or upon immigration [30]) allow linkage of individual data between registers. We were granted access to de-identified data after approval by an Ethical Committee at Karolinska Institutet. We established a national sample of SCZ cases defined as: (a) ≥2 inpatient hospitalizations or specialist outpatient visits with a diagnosis of SCZ or schizoaffective disorder from the National Patient Register; (b) born in Sweden from 1 January 1958 to 31 December 1993 (rationale is that there are incomplete data on older subjects and as we wanted subjects to have entered the core risk period for SCZ by the end of follow-up in 31 December 2013); and (c) excluded individuals with a plausible alternative primary diagnosis (Table S1). This definition of SCZ has been validated widely using clinical, epidemiological, and genetic analyses [6, 7, 11, 32]. We included demographic factors from linkage with other national registers (Supplementary Methods).

Genotyped subsample from the Swedish SCZ Study (S3)
The S3 genotyped subsample is a subset of the national sample. Full descriptions are in other papers [11, 12, 32–34]. Briefly, blood-derived DNA samples from SCZ cases and controls were collected from 2005–2013. Cases were defined as in the national sample. Controls selected at random from Swedish population registers and were never hospitalized for SCZ, schizoaffective disorder, or BIP and age ≥ 18 years. S3 was linked to Swedish registers, leaving 4992 cases and 6009 controls with validated status. Due to regulatory prohibitions, we could not remove S3 subjects from the de-identified national sample. All subjects provided informed consent and all procedures were approved by the relevant ethical committees.

Cognition measures
EDU was derived from a national database coding the highest completed educational level [35]. We coded EDU according to the International Standard Classification of Education as in large Genome-wide association studies (GWAS) [25]. We standardized EDU with respect to birth year and sex into a Z-score. Premorbid cognitive ability, measured as premorbid intelligence quotient (IQ) scores, was obtained from the Conscription Register covering males aged 18–19 from 1967–2010. Individuals with a diagnosis of SCZ at this examination (144 in the national sample and 28 in S3) were excluded from the analyses for premorbid cognitive ability. IQ was Z-score standardized by birth year. ID was defined by medical records using the National Patient Register (Table S1).

Common and rare measures of genetic burden
Details of genome-wide SNP genotyping, PRS calculation, CNV assessment, and exome sequencing are in the Supplementary Methods. [11, 12, 32–34, 36] Training sets for PRS common variant burden were from the latest GWAS for SCZ, BIP, IQ, and EDU (after removing any Swedish samples) [10, 22, 25, 37]. Rare CNVs are defined by frequency < 0.01, size ≥ 100 kb, and spanning ≥ 15 probes. CNV burden, for duplications and deletions separately, was computed as CNV size (total KB affected by CNVs), total number of CNVs, and the number of

pathogenic CNVs (>50% overlap) associated with SCZ, autism, ID, or developmental delay [29, 38–40]. Rare exonic burden was the number of ultra-rare disruptive/damaging single nucleotide variations and indels not observed in Exome Aggregation Consortium study and in constrained genes (previously identified as 'missense-constrained' or 'loss-of-function intolerant') (Supplementary Methods) [12, 34]. In total, 4288 cases and 5305 controls had available data on all genetic profiles. Genetic burden measures were standardized to aid in interpretation.

Statistical analyses
We used the national sample to examine associations of SCZ with cognitive traits (i.e., EDU, premorbid cognitive ability and ID) via epidemiological and genetic epidemiological analyses. To assess the impact of the measures of cognitive ability on SCZ risk in the national sample, we fitted Cox regression models that accounted for time at risk. Subjects entered at 1-Jan-1973 and were followed to the date of emigration, death, or up to 31-Dec-2013. First, we examined associations between each cognitive measure and SCZ. Second, we examined the associations of SCZ with EDU and ID jointly. Finally, we examined the associations of SCZ with premorbid cognitive ability, EDU and ID jointly (males only). Relevant epidemiological covariates were adjusted in all models, including sex, birth year, parental EDU, parental age at birth, and whether the person was born in winter.

The national sample can be connected into pedigrees to enable population genetic epidemiological analyses through linkage of Multi-Generation Register. Using an extended twin-family design [41–44], we estimated pedigree heritability by fitting univariate quantitative genetic structural equation models (SEM) separately for SCZ, premorbid cognitive ability, EDU, and ID and decomposing phenotypic variance into additive genetic, shared environmental, and unique environmental components (Supplementary Methods). Sex and birth year were included as covariates to adjust for any group differences. We fitted bivariate quantitative genetic SEM to estimate the pedigree genetic correlations (r_g) for SCZ with cognitive traits.

We examined the effects of genetic variant burden measures (PRS with $P_T \leq 0.05$, CNV and rare exonic burden) on EDU and premorbid cognitive ability. PRS of BIP was tested as previous studies have suggested a positive genome-wide correlation between BIP and EDU [45] and a SCZ subtype resembling BIP and high IQ [23]. Separate models for each genetic burden were evaluated and interaction terms were added to examine whether the effects differed between SCZ cases and controls. Those showed significant associations were then included in a joint model. All statistical models were adjusted for ancestry principal components and genotyping waves.

Cluster analyses in SCZ cases
Regression methods may not detect the existence of natural groups of patients. Clinicians naturally seek categorical ways to understand patients, and empirical subtyping patients is of intense interest for "patient stratification" to optimize therapeutics. We applied unsupervised clustering to identify subgroups in the national sample. The input variables were cognition-related: EDU, parental EDU, ID, age at first SCZ diagnosis, and the number of BIP hospital contacts [23]. Except for ID, we regressed out birth year and sex for other input variables. As they were nominal and continuous variables, we used Gower's dissimilarity matrices as input [46]. We used the Uniform Manifold Approximation and Projection (UMAP) to project the embedding of the input matrix into a two-dimensional layout [47]. UMAP is a non-linear dimensional reduction algorithm that preserves data features in lower dimensions and is an effective feature extraction tool in various fields in life science, including population genetics and scRNA-seq [48–50]. We then applied the Density Based Spatial Clustering of Applications with Noise (DBSCAN) to identify clusters [51]. DBSCAN identifies clusters of arbitrary shape and handles outliers more effectively compared to other clustering methods. Clustering replication was evaluated within the national sample via a random 1:1 split into training and replication sets. We applied the same clustering procedures in both training data and replication data and evaluated cluster similarity. After confirming the similarity of the clustering results, we combined the training and replication sets, and fit Cox regression models to compare rates of adverse outcomes in the clusters. Treatment resistance (ever use of clozapine) was shown in proportions but not tested for rates since the data was only available from 2005. Tested adverse outcomes included suicidality (attempts and completed suicide, Table S1), first hospitalization > 200 days (i.e., the median length of hospitalization for those in top decile of hospitalization), and death. SCZ cases were followed from initial SCZ

Table 1. Demographic characteristics of the national sample and genotyped subsample.

Swedish national sample	Cases	Controls	Statistical comparison
Subjects	14,230	3,816,264	NA
Birth year, mean (SD)	1969 (8.6)	1975 (10.5)	$t_{3,830,492} = 72.3$, $P < 1 \times 10^{-300}$
Male sex, N (%)	8,762 (61.6%)	1,959,215 (51.3%)	$\chi_1^2 = 594.2$, $P = 3.12 \times 10^{-131}$
Premorbid cognitive ability (males), mean (SD)	−0.51 (1.07)	0.00 (1.00)	$t_{1,422,976} = 40.2$, $P < 1 \times 10^{-300}$
Educational attainment, mean (SD)	−0.59 (0.90)	0.00 (1.00)	$t_{3,706,109} = 69.5$, $P < 1 \times 10^{-300}$
Intellectual disability, N (%)	923 (6.5%)	23,692 (0.6%)	$\chi_1^2 = 7630.1$, $P < 1 \times 10^{-300}$
Genotyped subsample	Cases	Controls	Statistical comparison
Subjects	4,992	6,009	NA
Birth year, mean (SD)	1954 (11.8)	1952 (11.3)	$t_{10,999} = -9.7$, $P = 2.72 \times 10^{-22}$
Male sex, N (%)	3,021 (60.5%)	3,052 (50.8%)	$\chi_1^2 = 103.9$, $P = 2.11 \times 10^{-24}$
Premorbid cognitive ability (males), mean (SD)	−0.34 (0.97)	0.31 (0.91)	$t_{2,544} = 17.4$, $P = 2.60 \times 10^{-64}$
Educational attainment, mean (SD)	−0.34 (0.87)	0.28 (1.00)	$t_{10,770} = 33.7$, $P = 2.18 \times 10^{-237}$
Intellectual disability, N (%)	351 (7.0%)	5 (0.1%)	$\chi_1^2 = 418.2$, $P = 6.14 \times 10^{-93}$

Premorbid cognitive ability is Z-score standardized by birth year in each sample. Educational attainment is Z-score standardized by birth year and sex in each sample. All continuous variables are described by mean and standard deviation (SD). Categorical variables are described by sample size (N) and percentage (%). Statistical comparisons are t-test for continuous variables and chi-square test for categorical variables. All statistical comparisons exceed Bonferroni correction ($N = 10$, $P < 0.005$). NA: not applicable.

diagnosis to the date of emigration, death, or 31-Dec-2013. Relevant covariates were adjusted for each outcome (Table S5).

Finally, we applied the same clustering procedures in S3 and evaluated the cluster similarity in this genetic subsample. We further examined whether common and rare genetic burdens differ across clusters and from controls (CNV duplications were not tested due to the null association with cognitive measures in previous analyses). Default parameters were used for UMAP algorithm except for specifying *n_neighbors* = 50 and a seed for random number generation (*random_state)* for reproducibility. Parameters for DBSCAN were set to *eps* = 1 and *MinPts* = 50.

Software and multiple testing corrections

All analyses were performed in R (v4.0.3) [52]. The quantitative genetic models for pedigree analyses were fitted with *OpenMx* (v2.18.1) [53]. Cluster analyses used R packages *cluster*, *umap*, and *fpc*. We performed multiple testing correction with Bonferroni method, which is a conservative correction and works in the worst-case scenario that all tests are independent. Correction on the total number of tests in the study would be overly rigorous and inappropriate. Therefore, we performed Bonferroni correction for groups of related statistical tests rather than the total number of tests performed across the study. This approach is appropriate here (and often found in the psychiatric genomics literature) given that these are distinctive sets of hypotheses and different from running the same analysis on data subsets (e.g., a GWAS for all subjects and then by sex). Here, sets of related statistical tests usually corresponded to the results in a table. The significance thresholds were Bonferroni-corrected to $P < 0.05$, and are given in table legends. Statistical tests were two-sided except for the comparison of SCZ-PRS, CNV, and rare exonic burden between SCZ cases in each cluster and controls, which were one-sided assuming higher rate in cases.

RESULTS

The Swedish national sample consisted of 14,230 SCZ cases and 3,816,264 controls. The lifetime prevalence of SCZ was 0.37% (95% CI 0.37–0.38%, similar to our 2006 report [6]). Table 1 shows demographic variables in population-level and in the S3 subsample with genetic data. Individuals in S3 were relatively old at recruitment. Both national and genetic sample had profound case-control differences in premorbid cognitive ability, EDU and ID.

Epidemiological analyses

In the national sample, we observed strong associations between cognition and risk of SCZ in separate and joint models (Table 2A).

Notably, lower premorbid cognitive ability, lower EDU, and the presence of ID were strongly associated with risk of SCZ.

Genetic epidemiological analyses

In total, 931,744 siblings were included and results were shown in Table 2B. The pedigree-heritability for SCZ was estimated as 0.70, and the estimates of pedigree-heritability for the cognition measures were 0.37 for EDU, 0.65 for premorbid cognitive ability, and 0.84 for ID. For pedigree-genetic correlations (r_g), SCZ had a negative pedigree-r_g with premorbid cognitive ability (−0.11; 95% CI: −0.15, −0.07), a positive pedigree-r_g with ID (0.50; 95%CI: 0.47, 0.52), and the pedigree-r_g with EDU was not significant (0.09; 95% CI: −0.04, 0.22). Intriguingly, these estimates of pedigree-r_g approximated those based on genotyping with SNP-r_g of SCZ with IQ (−0.21, assessed via neurocognitive tests), rare severe neurodevelopmental disorders (0.28, often comorbid including ID), and EDU (0.02) [22, 23, 54].

Genetic burden analyses

The association with SCZ was negative for PRS of IQ (OR = 0.88 (0.84–0.92), $P = 1.38 \times 10^{-8}$) but positive for PRS of EDU (OR = 1.08 (1.04, 1.13), $P = 1.12 \times 10^{-4}$). Results of associations between each burden measure and the two cognitive traits, EDU and premorbid cognitive ability, are shown in Table S2. The genetic burdens that showed significant associations were then included in the joint model, and the results are in Fig. 1 and Table S3. Since SCZ diagnosis modified the associations between EDU-PRS and EDU (Table S2), we examined the joint effect separately in cases and controls. PRS of EDU and IQ showed positive associations with cognitive traits in SCZ cases, among which the effect of EDU-PRS on EDU was lower than that in controls (0.13 vs. 0.19, P = 3.89 × 10^{-4} for test of SCZ diagnosis interaction). Rare exonic burden showed an inverse association with cognitive traits in SCZ cases (for EDU, −0.06, $P = 4.91 \times 10^{-6}$; for premorbid cognitive ability, −0.09, $P = 0.002$) but not in controls. SCZ-PRS had no associations with cognitive traits in SCZ cases.

Cluster analyses in SCZ cases

We conducted unsupervised cluster analyses on 13,647 cases with complete data available for the input clustering variables. In the training set ($N = 6823$), DBSCAN clustering identified four clusters after UMAP projection (Table 3A). Cluster 1 (56.6% cases) was

Table 2. Epidemiological and genetic epidemiological analyses in the national sample. A. Epidemiological analyses in the Swedish national sample. B. Heritability and genetic correlations.

A

Trait	Separate model		Joint model 1		Joint model 2 (males only)	
	HR [95% CI]	P-value	HR [95% CI]	P-value	HR [95% CI]	P-value
Premorbid cognitive ability	0.54 [0.52; 0.55]	$<1 \times 10^{-300}$	NA	NA	0.65 [0.63; 0.67]	5.87×10^{-169}
Educational attainment	0.43 [0.42; 0.44]	$<1 \times 10^{-300}$	0.47 [0.46; 0.48]	$<1 \times 10^{-300}$	0.65 [0.63; 0.67]	1.09×10^{-126}
Intellectual disability	13.81 [12.90; 14.79]	$<1 \times 10^{-300}$	7.54 [6.99; 8.14]	$<1 \times 10^{-300}$	12.56 [10.78; 14.65]	5.77×10^{-229}

B

Trait	Pedigree-heritability (95% CI)	SNP-heritability (95% CI)	Pedigree-r_g with SCZ (95% CI)	SNP-r_g with SCZ (95% CI)
SCZ	0.70 [0.63; 0.77]	0.24 [0.23; 0.25]	NA	NA
Premorbid cognitive ability	0.65 [0.62; 0.68]	0.19 [0.17; 0.21]	−0.11 [−0.15; −0.07]	−0.21 [−0.26; −0.16]
Educational attainment	0.37 [0.34; 0.39]	0.12 [0.12; 0.13]	0.09 [−0.04; 0.22]	0.02 [−0.01; 0.06]
Intellectual disability	0.84 [0.77; 0.91]	0.08 [0.04; 0.12]	0.50 [0.47; 0.52]	0.28 [0.15; 0.41]

Premorbid cognitive ability is Z-score standardized by birth year. Educational attainment (EDU) is Z-score standardized by birth year and sex. In Table 2A, Cox regression models are applied. Separate model tests for each cognitive trait are adjusted for sex (except for premorbid cognitive ability since it was assessed only in males), categorical birth year (1958–1962, 1963–1967, 1968–1974, 1975–1993), parental EDU (either mother's EDU or father's EDU if only one among them is available; if both mother's and father's EDU were available, take the mean), maternal age, paternal age and whether the person was born in winter (yes or no). Joint model 1 includes EDU, intellectual disability (ID) and other covariates listed as above. Joint model 2 includes premorbid cognitive ability, EDU, ID, and other covariates listed as above except for sex. All statistical comparisons exceed Bonferroni correction ($N = 8$, $P < 0.006$).
In Table 2B, for pedigree analyses, Wald confidence intervals (CI) are calculated by using the delta method. SNP-heritability and SNP-r_g are from the literature (SNP-r_g between SCZ and EDU was estimated using LDSC from the latest GWAS) [10, 22, 25, 54, 69]. Estimates and 95% CIs are shown. SNP-heritability and SNP-r_g in the second row refer to intelligence. SNP-heritability and SNP-r_g in the last row refer to severe neurodevelopmental disorders as a proxy for intellectual disability. NA: not applicable. Multiple testing correction is not applicable to this descriptive table.
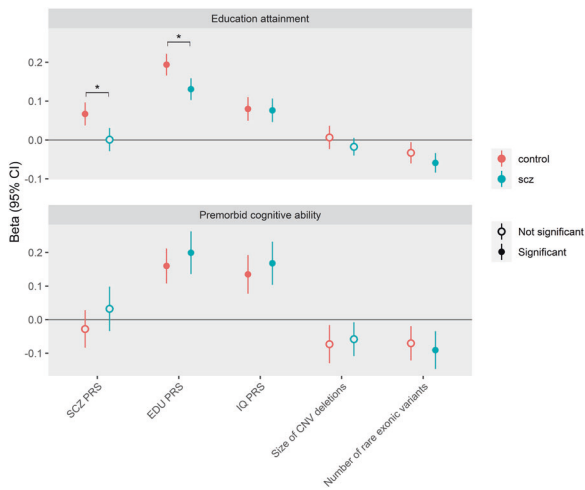

Fig. 1 Associations between genetic burden and cognitive measures in SCZ cases and controls. Genetic profiles include: (1) polygenetic risk scores (PRS) for schizophrenia (SCZ), intelligence quotient (IQ) and educational attainment (EDU); (2) size of copy number variants (CNV) deletions in KB; and (3) rare exonic burden, measured as number of disruptive and damaging ultra-rare variants in constrained genes. Burden measures were standardized. Cognitive measures include Educational attainment (EDU) and premorbid cognitive ability (measured by intelligence quotient (IQ) scores, IQ). EDU is Z-score standardized by birth year and sex. Premorbid cognitive ability is Z-score standardized by birth year. The analysis used linear regression models including all genetic burdens above and adjusted for the first 5 ancestry principle components and genotyping waves. Beta coefficient and 95% confidence intervals (CI) are reported. Estimates past significance threshold (corrected for 20 tests, $P < 0.0025$) are marked in solid circle. Asterisk indicates significant difference between SCZ cases and controls. The data for this figure are in Table S3.

characterized by moderate features compared to other groups. Cluster 2 (25.5% cases) was characterized by early age at first SCZ diagnosis, lower EDU, lower parental EDU, and fewer BIP contacts. Cluster 3 (11.7% cases) was characterized by later diagnosis, higher EDU/parental EDU, and more BIP contacts. Finally, Cluster 4 (6.2% cases) SCZ cases with ID, presenting lower parental EDU. In the replication set ($N = 6,824$), using same input variables and clustering algorithm, we also identified four clusters and the individual distributions and characteristics were similar to that of the training set (Table S4). Similar patterns for other characteristics were also observed for both sets (Tables 3A, S4). For example, Cluster 2 (low EDU) had the highest proportions of males and was more likely to have long hospitalizations. Cluster 3 (high EDU) had the fewest males, lowest mortality, and was less likely to have long hospitalizations (Tables S4, S5).

Applying the same clustering procedures to the genotyped subsample with complete data for the input clustering variables ($N = 3,674$), we found that the cluster distributions and characteristics were similar to the populational training set (Table 3B). Moreover, the clusters differed in EDU-PRS, IQ-PRS, size of CNV deletions, number of known pathogenic CNV deletions and rare exonic burden (Table 3B). Further, the cluster 3 with high EDU had higher EDU-PRS, no lower IQ-PRS and no excess burden of rare genetic variants when compared to controls (Fig. 2, Table S6).

DISCUSSION
We found evidence for a robust relation between SCZ and cognitive traits by combining comprehensive national registry and directly genomic assays. In populational analyses, we confirmed that indices of worse cognition were strong risk factors for SCZ, and the pedigree-r_g between SCZ and cognitive traits, including EDU, ID and premorbid cognitive ability, are comparable with the SNP-r_g from common genetic variants. In the genotyped sample, SCZ cases were likely to have higher EDU and premorbid cognitive

Table 3. SCZ case characteristics across cluster groups in: A. national training set, B. genotyped subsample.

A

Feature		Cluster 1 (medium EDU)	Cluster 2 (low EDU)	Cluster 3 (high EDU)	Cluster 4 (ID)	P
N (%)		3,865 (56.6%)	1,739 (25.5%)	798 (11.7%)	421 (6.2%)	–
Clustering input variables	Age at first SCZ diagnosis, mean (SD)	0.09 (0.96)	−0.23 (0.99)	0.35 (0.91)	−0.05 (1.10)	5.55×10^{-48} [a]
	ID, N (%)	0 (0%)	0 (0%)	0 (0%)	421 (100%)	–
	EDU, mean (SD)	0.12 (0.27)	−1.10 (0.27)	2.17 (0.19)	−0.70 (0.71)	$<1\times10^{-300}$ [a]
	Parental EDU, mean (SD)	0.03 (0.99)	−0.27 (0.90)	0.63 (1.06)	−0.43 (0.86)	3.08×10^{-117} [a]
	Number of BIP contacts, mean (SD)	0.01 (1.00)	−0.06 (0.76)	0.09 (1.28)	−0.03 (0.78)	0.002 [a]
Birth year, mean (SD)		1968 (8.61)	1969 (9.03)	1969 (7.61)	1969 (8.41)	0.001 [a]
Male sex, N (%)		2,444 (63.2%)	1,148 (66.0%)	408 (51.1%)	253 (60.1%)	7.04×10^{-12} [a]
Premorbid cognitive ability (males), mean (SD)		−0.43 (0.99)	−1.01 (0.91)	0.38 (0.90)	−1.75 (0.59)	6.38×10^{-130} [a]
Attempt/completed suicide, N (%)		497 (12.9%)	284 (16.3%)	66 (8.3%)	60 (14.3%)	4.40×10^{-7} [a]
Death, N (%)		322 (8.3%)	198 (11.4%)	28 (3.5%)	37 (8.8%)	1.53×10^{-9} [a]
Ever hospitalized for more than 200 days, N (%)		629 (16.3%)	487 (28.0%)	69 (8.6%)	76 (18.1%)	2.73×10^{-36} [a]
Ever use of clozapine, N (%)		879 (22.7%)	464 (26.7%)	141 (17.7%)	103 (24.5%)	2.61×10^{-7} [a]

B

Feature		Cluster 1 (medium EDU)	Cluster 2 (low EDU)	Cluster 3 (high EDU)	Cluster 4 (ID)	P
N (%)		2109 (57.4%)	1064 (29.0%)	266 (7.2%)	235 (6.4%)	–
Clustering input variables	Age at first SCZ diagnosis, mean (SD)	0.07 (0.91)	−0.21 (0.93)	0.23 (0.93)	−0.14 (0.91)	1.08×10^{-18}
	ID, N (%)	0 (0%)	0 (0%)	0 (0%)	235 (100%)	–
	EDU, mean (SD)	0.32 (0.31)	−0.98 (0.35)	2.29 (0.26)	−0.62 (0.73)	$<1\times10^{-300}$
	Parental EDU, mean (SD)	0.07 (1.01)	−0.21 (0.86)	0.66 (1.22)	−0.34 (0.84)	1.27×10^{-44}
	Number of BIP contacts, mean (SD)	0.00 (0.93)	−0.06 (0.93)	0.18 (1.53)	−0.002 (0.97)	0.007
Birth Year, mean (SD)		1958 (9.41)	1957 (10.45)	1957 (9.99)	1958 (11.17)	1.86×10^{-4}
Male, N (%)		1336 (63.3%)	701 (65.9%)	141 (53.0%)	139 (59.1%)	7.77×10^{-4}
N with available data on all genetic profiles (%)		1814 (57.3%)	913 (28.9%)	234 (7.4%)	203 (6.4%)	–
PRS, mean (SD)	SCZ	0.01 (1.00)	0.01 (1.01)	−0.10 (0.96)	−0.01 (1.02)	0.44
	EDU	0.06 (1.00)	−0.19 (0.96)	0.44 (0.92)	−0.19 (1.00)	2.25×10^{-20} [b]
	IQ	0.03 (0.99)	−0.09 (0.99)	0.34 (1.01)	−0.21 (1.00)	6.70×10^{-10} [b]
	BIP	0.03 (1.01)	−0.03 (0.96)	−0.07 (1.06)	−0.04 (0.97)	0.33
CNV deletions, mean (SD)	Size of CNVs	−0.01 (0.96)	−0.03 (0.87)	−0.08 (0.58)	0.31 (1.86)	7.81×10^{-5} [b]
	Number of CNVs	0.00 (1.01)	−0.03 (0.96)	−0.01 (1.04)	0.14 (1.06)	0.18
	Number of known pathogenic CNVs	−0.03 (0.85)	0.01 (1.01)	−0.11 (0.03)	0.33 (2.11)	9.13×10^{-6} [b]
Rare exonic burden, mean (SD)	Number of disruptive or damaging ultra-rare variants in constrained genes	−0.01 (1.00)	0.05 (1.02)	−0.23 (0.86)	0.18 (1.06)	8.88×10^{-5} [b]

Abbreviations: *SCZ*, schizophrenia; *BIP*, bipolar disorder; *EDU*, educational attainment; *ID*, intellectual disability; *IQ*, intelligence quotient; *PRS*, polygenic risk score; *CNV*, copy number variant.

Parental EDU is either from mother or from father if only one among them is available; if both mother's and father's EDU are available, take the mean. In Table 3A, age at first SCZ diagnosis, EDU, parental EDU, and number of BIP contacts are regressed on birth year and sex and then take the standardized residuals within the population case cohort. Premorbid cognitive ability is Z-score standardized by birth year in the whole population cohort. The hospitalization >200 days is the median length of hospitalization for those in the top decile of hospitalization. In Table 3B, except for ID, birth year and male sex, all other variables are the standardized residuals of regression models described as below: age at first SCZ diagnosis, EDU, parental EDU and number of BIP contacts are regressed on birth year and sex; CNV deletions were regressed on genotyping waves; Rare exonic burdens were regressed on the first 5 ancestry principle components (PC) and genotyping waves; PRS were regressed on PC1-PC20 estimated from whole exome sequencing and genotyping waves. Statistical comparisons are one-way ANOVA for continuous variables and chi-square test for categorical variables.

[a]Indicates results exceeding Bonferroni-corrected significance threshold ($N = 11$, $P < 0.004$) in Table 3A.

[b]Indicates results exceeding Bonferroni-corrected significance threshold ($N = 14$, $P < 0.003$) for genetic burden tests in Table 3B.
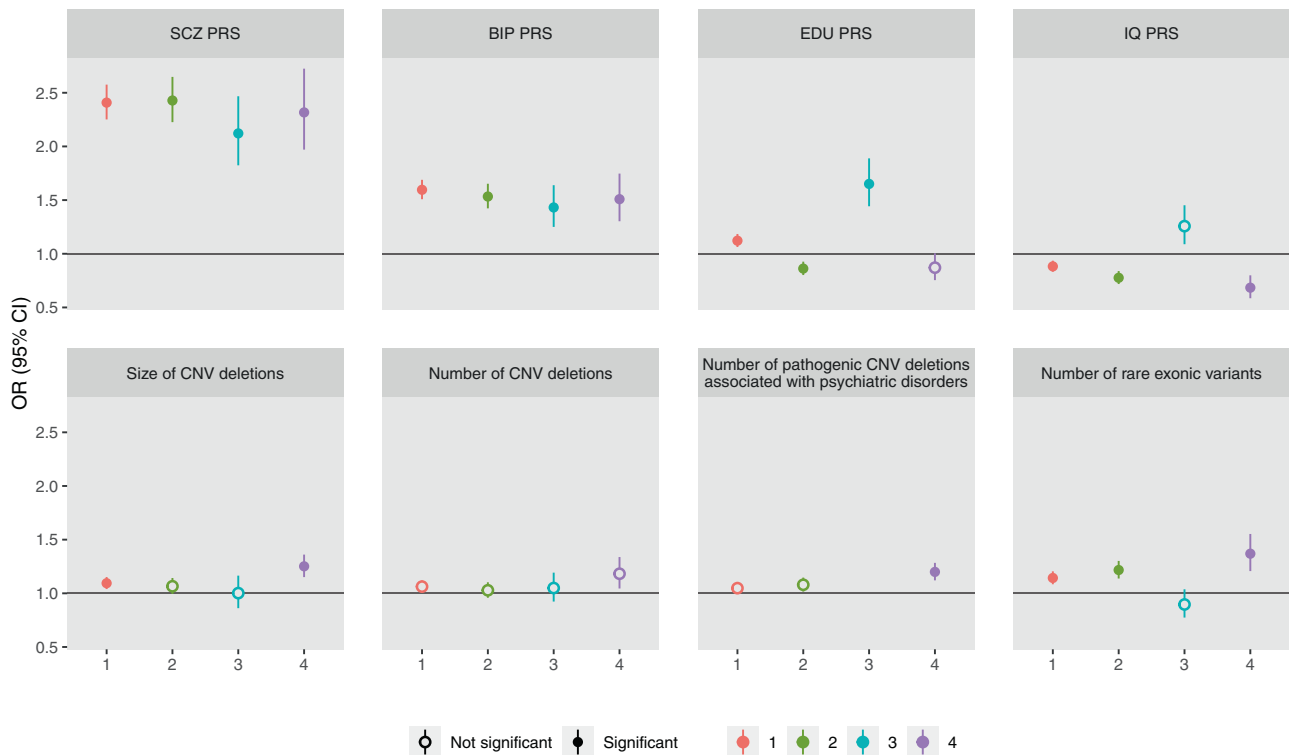
Fig. 2 Test of genetic burden between SCZ cluster groups and controls. Genetic profiles include: (1) polygenetic risk scores (PRS) for schizophrenia (SCZ), bipolar disorder (BIP), intelligence quotient (IQ) and Educational attainment (EDU); (2) copy number variants (CNV) deletions including size of CNVs in KB, count of CNVs, and count of CNVs in pathogenic regions associated with SCZ, autism, developmental delay and intellectual disability (defined as had > 50% overlap with the region (PLINK –cnv-region-overlap 0.5)); and (3) rare exonic burden, measured as number of disruptive and damaging ultra-rare variants in constrained genes. All genetic burden measures were standardized. All analysis used logistic regression model. For PRS, analyses were adjusted for the first 5 ancestry principle components (PC) and genotyping waves. For CNV, analyses were adjusted for genotyping waves. For rare exonic burden, the analysis was adjusted for PC1-PC20 estimated from whole exome sequencing data and genotyping waves. Odds ratios (OR) and 95% confidence intervals (CI) are reported. Estimates past significance threshold (corrected for 32 tests; $P < 0.0015$) are marked in solid circle. The data for this figure are in Table S6. The test for number of known pathogenic CNVs in Cluster 3 vs. controls is not applicable because no SCZ cases in Cluster 3 had known pathogenic CNVs (empty cell).

ability when they had higher EDU-PRS, higher IQ-PRS and less rare exonic burden. Finally, by applying an unsupervised clustering method, we found four clusters of SCZ cases characterized by EDU, age at first diagnosis, number of BIP contacts and ID in the national sample. The cases in the clusters with high EDU had less adverse outcomes including long hospitalization and death. When applying the same clustering analysis to the genetic subsample, the case cluster with high EDU presented higher PRS of EDU and no significant excess of rare genetic burdens than controls.

Multiple studies have found that lower premorbid cognitive ability is associated with multiple psychiatric disorders particularly SCZ [55, 56]. A Swedish national study found that decline in cognitive performance during the teenage years predicted psychosis in adulthood [57]. However, we cannot rule out reverse causation bias in the association between low EDU and SCZ (e.g., when onset of psychotic symptoms impaired school performance), and we note that >90% of cases achieved their highest level of education before first diagnosis of SCZ.

The pattern that the pedigree-r_g between SCZ and cognitive measures are comparable with previous reports and with the corresponding SNP-r_g presents a converging picture of the etiology for SCZ, cognitive traits and shared genetics between them [58]. We observed several interesting results. First, the SNP-heritability of severe neurodevelopmental disorders (including ID) is the smallest (0.08) [54] while its pedigree-heritability is the highest (0.84). This could be explained by current SNP genotyping arrays poorly capture rare variants with large effects which are an

important contributor to severe neurodevelopmental deficits like ID. Second, we detected a significant pedigree-r_g between SCZ and premorbid cognitive ability in males which is in line with previous reports and with the corresponding SNP-r_g. A Swedish twin-sibling study reported a negative genetic correlation between IQ and psychosis (−0.26), similar to the reported SNP-r_g (−0.21) [59]. The negative genetic correlation between premorbid cognitive ability and SCZ is also supported by our observation of positive genetic correlation between SCZ and ID. The modest negative genetic correlation (−0.11), along with a Swedish co-relative control analysis that found no attenuation in association between SCZ and intelligence in siblings, cousin pairs, and general population [18], suggests a role for non-shared environmental risk factors for lower IQ and SCZ. Third, the pedigree-r_g between SCZ and EDU was not significant, and was near zero SNP-r_g as estimated from the largest published GWAS [9, 25], despite a strong epidemiological association [24] and considerable overlap in causal variants [45].

For genetic burden analyses with cognitive traits in SCZ cases, PRS for both IQ and EDU showed positive relationship while SCZ-PRS showed no association. This is in line with a recent study that finds in SCZ cases, cognition is more strongly related with PRS that index cognitive traits in general than PRS for psychiatric disorders, suggesting the mechanisms of cognitive variation within SCZ is at least partly independent from that predisposes the illness [60]. However, unexpectedly, the EDU-PRS associated positively with SCZ risk, despite a small negative correlation between the two PRS

(−0.08 in cases and −0.03 in controls). Moreover, the SCZ-PRS was positively associated with EDU in controls. Such findings were in line with some of the previous studies [61], but they are contrast to a Danish study that reported higher SCZ-PRS associated with noncompletion of primary school in SCZ noncases [62]. Recent studies have shown evidence of shared genetic loci between SCZ and EDU, and their genetic dependence possibly related to SCZ subtypes [23, 63]. Taken together with our findings of the epidemiological and pedigree analyses, it is evident that the relationship between SCZ and EDU are complex.

Previous studies on identifying SCZ subgroups have employed clustering methods based on diverse measures of cognition [64–66], and have identified subtypes with different characteristics including real-world functioning, symptom severity, clinical pattern and neurocognitive features. Here we adopted cognitive-related variables on a population-level, and finally decided a four-cluster solution. This new clustering of SCZ cases identifies individuals at different level of cognition functioning, characterized with potential different mortality rate and hospitalization length, and were genetically validated by common and rare genetic burdens. The group with high EDU tended to have higher PRS of EDU and IQ (albeit non-significant) than controls, suggesting a subgroup differed from traditional SCZ. The variables we used (i.e., age at first SCZ diagnosis, EDU, parent EDU, number of BIP contacts and ID) were common features that were frequently recorded from registers and surveys, adding the generalizability to existing SCZ cohorts and patients in clinics.

In this paper, we combined multiple interlacing approaches (national-scale epidemiology and genetic epidemiology with multiple measures of common and rare genetic variation); this is a strength of our study and uncommonly presented in a single paper. We also had several limitations. First, although our findings of no association between CNVs, SCZ-PRS and cognitive traits in SCZ cases were in line with previous studies, it could also be due to lack of power. For rare CNVs existing in a small number of carriers and explaining only a small fraction of phenotype variance, the investigation of complex cognitive traits would require extremely large dataset to achieve sufficient power [67]. Second, the Swedish National Patient Register captures only a select minority of people who might have more severe ID and comorbidities of other diseases, which limits the generalizability of the findings [68]. Third, the associations between ID and SCZ could be overestimated, because patients diagnosed with one disease are likely to be in contact with physicians and are also more likely to receive other diagnoses. Fourth, the S3 genotyped subsample may have selection bias as it requires patients to survive and have the capacity to provide informed consent. Fifth, the clustering analysis was based on complicated analytical approaches with several proxy phenotypes, and risk of overfitting cannot be ruled out. Because the health care system in Sweden is tax-funded with universal access, the generalizability of these results to other places may be limited by differences in social welfare policies, resources and practices. Replication in independent samples are warranted in the future. Last but not least, the time-varying factors that affect EDU, such as socioeconomic status and cooccurrence of diseases, were not controlled and could influence assessment of EDU. Our work could be improved via integration of longitudinal measures of cognitive function in order to better understand the association between EDU and SCZ. As these are available only for younger Swedes, further exploration will await completion of our current expansion of genotyped cases to ~12 000. Moreover, EDU is perceived as a less precise indication of cognitive abilities, which also limits the generalizability. Future studies targeting broader measures of cognition might provide a way to dissect more features of this complex disorder.

In conclusion, we sought to comprehensively understand the relation of three cognitive traits and SCZ from both epidemiological and genetic perspectives. We confirmed a negative association between premorbid cognitive ability, ID and SCZ. The relationship between EDU and SCZ are complex and warrants further examination. The data-driven clustering results suggest that combined information from a few cognition-relevant variables might usefully index the heterogeneity of SCZ, which encourages the investigation of subtype-specific mechanisms and treatments in the future.

CODE AVAILABILITY
Custom written R scripts used for statistical analyses can be provided upon request.

REFERENCES
1. Chesney E, Goodwin GM, Fazel S. Risks of all-cause and suicide mortality in mental disorders: a meta-review. World Psychiatry. 2014;13:153–60.
2. Chong HY, Teoh SL, Wu DB, Kotirum S, Chiou CF, Chaiyakunapruk N. Global economic burden of schizophrenia: a systematic review. Neuropsychiatr Dis Treat. 2016;12:357–73.
3. World Health Organization. The Global Burden of Disease: 2004 Update. Geneva: WHO Press; 2008.
4. Knapp M, Mangalore R, Simon J. The global costs of schizophrenia. Schizophr Bull. 2004;30:279–93.
5. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch Gen Psychiatry. 2003;60:1187–92.
6. Lichtenstein P, Bjork C, Hultman CM, Scolnick EM, Sklar P, Sullivan PF. Recurrence risks for schizophrenia in a Swedish national cohort. Psychol Med. 2006;36:1417–26.
7. Lichtenstein P, Yip B, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. Lancet. 2009;373:234–9.
8. Wray NR. Gottesman, II. Using summary data from the Danish national registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder. Front Genet. 2012;3:118.
9. Pardinas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat Genet. 2018;50:381–9.
10. Ripke S, Walters JTR, O'Donovan MC, Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. medRxiv. 2020. https://doi.org/10.1101/2020.09.12.20192922.
11. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014;511:421–7.
12. Genovese G, Fromer M, Stahl EA, Ruderfer DM, Chambert K, Landen M, et al. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. Nat Neurosci. 2016;19:1433–41.
13. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB et al. Most genetic risk for autism resides with common variation. Nat Genet. 2014.
14. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, et al. The genetic architecture of type 2 diabetes. Nature. 2016;536:41–47.
15. Sullivan PF, Agrawal A, Bulik CM, Andreassen OA, Borglum AD, Breen G, et al. Psychiatric genomics: an update and an agenda. Am J Psychiatry. 2018;175:15–27.
16. Elvevag B, Goldberg TE. Cognitive impairment in schizophrenia is the core of the disorder. Crit Rev Neurobiol. 2000;14:1–21.
17. Lepage M, Bodnar M, Bowie CR. Neurocognition: clinical and functional outcomes in schizophrenia. Can J Psychiatry. 2014;59:5–12.
18. Kendler KS, Ohlsson H, Sundquist J, Sundquist KIQ. and schizophrenia in a Swedish national sample: their causal relationship and the interaction of IQ with genetic risk. Am J Psychiatry. 2015;172:259–65.
19. Khandaker GM, Barnett JH, White IR, Jones PB. A quantitative meta-analysis of population-based studies of premorbid intelligence and schizophrenia. Schizophr Res. 2011;132:220–7.
20. Morgan VA, Leonard H, Bourke J, Jablensky A. Intellectual disability co-occurring with schizophrenia and other psychiatric illness: population-based study. Br J Psychiatry. 2008;193:364–72.
21. Smeland OB, Bahrami S, Frei O, Shadrin A, O'Connell K, Savage J, et al. Genome-wide analysis reveals extensive genetic overlap between schizophrenia, bipolar disorder, and intelligence. Mol Psychiatry. 2020;25:844–53.
22. Savage JE, Jansen PR, Stringer S, Watanabe K, Bryois J, de Leeuw CA, et al. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. Nat Genet. 2018;50:912–9.
23. Bansal V, Mitjans M, Burik CAP, Linner RK, Okbay A, Rietveld CA, et al. Genome-wide association study results for educational attainment aid in identifying genetic heterogeneity of schizophrenia. Nat Commun. 2018;9:3078.
24. Adams CD. A multivariable Mendelian randomization to appraise the pleiotropy between intelligence, education, and bipolar disorder in relation to schizophrenia. Sci Rep. 2020;10:6018.

25. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet. 2018;50:1112–21.

26. Singh T, Kurki MI, Curtis D, Purcell SM, Crooks L, McRae J, et al. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. Nat Neurosci. 2016;19:571–7.

27. Ganna A, Genovese G, Howrigan DP, Byrnes A, Kurki MI, Zekavat SM, et al. Ultra-rare disruptive and damaging mutations influence educational attainment in the general population. Nat Neurosci. 2016;19:1563–5.

28. Singh T, Walters JTR, Johnstone M, Curtis D, Suvisaari J, Torniainen M, et al. The contribution of rare variants to risk of schizophrenia in individuals with and without intellectual disability. Nat Genet. 2017;49:1167–73.

29. 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. 2014;505:361–6.

30. Ludvigsson JF, Almqvist C, Bonamy AK, Ljung R, Michaelsson K, Neovius M, et al. Registers of the Swedish total population and their use in medical research. Eur J Epidemiol. 2016;31:125–36.

31. Pettersson E, Larsson H, Lichtenstein P. Common psychiatric disorders share the same genetic origin: a multivariate sibling study of the Swedish population. Mol Psychiatry. 2016;21:717–21.

32. Ripke S, O'Dushlaine C, Chambert K, Moran JL, Kähler A, Akterin S, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. Nat Genet. 2013;45:1150–9.

33. Szatkiewicz J, O'Dushlaine C, Chen G, Chambert K, Moran J, Neale B, et al. Copy number variation in schizophrenia in Sweden. Mol Psychiatry. 2014;19:762–73.

34. Exome Aggregation Consortium, Lek M, Karczewski K, Minikel E, Samocha K, Banks E, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536:285–91.

35. Ludvigsson JF, Svedberg P, Olen O, Bruze G, Neovius M. The longitudinal integrated database for health insurance and labour market studies (LISA) and its use in medical research. Eur J Epidemiol. 2019;34:423–37.

36. Purcell SM, Moran JL, Fromer M, Ruderfer D, Solovieff N, Roussos P, et al. A polygenic burden of rare disruptive mutations in schizophrenia. Nature. 2014;506:185–90.

37. Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z, et al. Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat Genet. 2021;53:817–29.

38. Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nat Genet. 2017;49:27–35.

39. Bragin E, Chatzimichali EA, Wright CF, Hurles ME, Firth HV, Bevan AP, et al. DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. Nucleic Acids Res. 2014;42:D993–D1000.

40. Coe BP, Witherspoon K, Rosenfeld JA, van Bon BW, Vulto-van Silfhout AT, Bosco P, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat Genet. 2014;46:1063–71.

41. Sandin S, Lichtenstein P, Kuja-Halkola R, Larsson H, Hultman CM, Reichenberg A. The familial risk of autism. JAMA. 2014;311:1770–7.

42. Viktorin A, Meltzer-Brody S, Kuja-Halkola R, Sullivan PF, Landen M, Lichtenstein P, et al. Heritability of Perinatal Depression and Genetic Overlap With Nonperinatal Depression. Am J Psychiatry. 2016;173:158–65.

43. Yao S, Larsson H, Norring C, Birgegard A, Lichtenstein P, D'Onofrio BM, et al. Genetic and environmental contributions to diagnostic fluctuation in anorexia nervosa and bulimia nervosa. Psychol Med. 2021;51:62–9.

44. Skoglund C, Tiger A, Ruck C, Petrovic P, Asherson P, Hellner C, et al. Familial risk and heritability of diagnosed borderline personality disorder: a register study of the Swedish population. Mol Psychiatry. 2021;26:999–1008.

45. Frei O, Holland D, Smeland OB, Shadrin AA, Fan CC, Maeland S, et al. Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. Nat Commun. 2019;10:2417.

46. Gower JC. A general coefficient of similarity and some of its properties. Biometrics. 1971;27:857–74.

47. McInnes L, Healy J, Melville J, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. ArXiv e-prints 1802.03426v3 2020. https://doi.org/10.48550/arXiv.1802.03426.

48. Dorrity MW, Saunders LM, Queitsch C, Fields S, Trapnell C. Dimensionality reduction by UMAP to visualize physical and genetic interactions. Nat Commun. 2020;11:1537.

49. Wang J, Zhang S, Qiao H, Wang J. UMAP-DBP: an improved DNA-binding proteins prediction method based on uniform manifold approximation and projection. Protein J. 2021;40:562–75.

50. Diaz-Papkovich A, Anderson-Trocme L, Gravel S. A review of UMAP in population genetics. J Hum Genet. 2021;66:85–91.

51. Martin E, Kriegel HP, Sander J, Xu X, A density-based algorithm for discovering clusters in large spatial databases with noise. AAAI Press 1996; 226–31.

52. Team RC. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria URL https://wwwR-projectorg/ 2015.

53. Neale MC, Hunter MD, Pritikin JN, Zahery M, Brick TR, Kirkpatrick RM, et al. OpenMx 2.0: extended structural equation and statistical modeling. Psychometrika. 2016;81:535–49.

54. Niemi MEK, Martin HC, Rice DL, Gallone G, Gordon S, Kelemen M, et al. Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. Nature. 2018;562:268–71.

55. Addington J, Barbato M. The role of cognitive functioning in the outcome of those at clinical high risk for developing psychosis. Epidemiol Psychiatr Sci. 2012;21:335–42.

56. David AS, Zammit S, Lewis G, Dalman C, Allebeck P. Impairments in cognition across the spectrum of psychiatric disorders: evidence from a Swedish conscript cohort. Schizophr Bull. 2008;34:1035–41.

57. MacCabe JH, Wicks S, Lofving S, David AS, Berndtsson A, Gustafsson JE, et al. Decline in cognitive performance between ages 13 and 18 years and the risk for psychosis in adulthood: a Swedish longitudinal cohort study in males. JAMA Psychiatry. 2013;70:261–70.

58. Polderman TJ, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, et al. Fifty years of twin studies: a meta-analysis of the heritability of human traits. Nat Genet. 2015;47:702–9.

59. Fowler T, Zammit S, Owen MJ, Rasmussen F. A population-based study of shared genetic variation between premorbid IQ and psychosis among male twin pairs and sibling pairs from Sweden. Arch Gen Psychiatry. 2012;69:460–6.

60. Richards AL, Pardinas AF, Frizzati A, Tansey KE, Lynham AJ, Holmans P, et al. The relationship between polygenic risk scores and cognition in schizophrenia. Schizophr Bull. 2020;46:336–44.

61. Mallet J, Le Strat Y, Dubertret C, Gorwood P, Polygenic risk scores shed light on the relationship between schizophrenia and cognitive functioning: review and meta-analysis. J Clin Med. 2020;9:341.

62. Sorensen HJ, Debost JC, Agerbo E, Benros ME, McGrath JJ, Mortensen PB, et al. Polygenic risk scores, school achievement, and risk for schizophrenia: a Danish population-based study. Biol Psychiatry. 2018;84:684–91.

63. 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. 2017;43:654–64.

64. Rocca P, Galderisi S, Rossi A, Bertolino A, Rucci P, Gibertoni D, et al. Social cognition in people with schizophrenia: a cluster-analytic approach. Psychol Med. 2016;46:2717–29.

65. Martinuzzi E, Barbosa S, Daoudlarian D, Bel Haj Ali W, Gilet C, Fillatre L, et al. Stratification and prediction of remission in first-episode psychosis patients: the OPTiMiSE cohort study. Transl Psychiatry. 2019;9:20.

66. Chand GB, Dwyer DB, Erus G, Sotiras A, Varol E, Srinivasan D, et al. Two distinct neuroanatomical subtypes of schizophrenia revealed using machine learning. Brain. 2020;143:1027–38.

67. Mannik K, Magi R, Mace A, Cole B, Guyatt AL, Shihab HA, et al. Copy number variations and cognitive phenotypes in unselected populations. JAMA. 2015;313:2044–54.

68. Hirvikoski T, Boman M, Tideman M, Lichtenstein P, Butwicka A. Association of intellectual disability with all-cause and cause-specific mortality in Sweden. JAMA Netw Open. 2021;4:e2113014.

69. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47:291–5.

AUTHOR CONTRIBUTIONS
PFS conceived the study idea, designed and supervised the implementation, interpreted the results and revised the manuscript. JS conducted the analyses, interpreted the results, drafted and revised the manuscript. SY conducted the pedigree analyses and contributed to the application of clustering methods. KK and

YL contributed to the interpretation of the results and revised the manuscript. AS provided scripts for epidemiological analyses. JPS contributed to the CNV analyses. HL, PL and CMH discussed and commented on the manuscript.

COMPETING INTERESTS
PFS reports potentially competing financial interests: Lundbeck (grant recipient), Neumora Therapeutics (advisory committee, shareholder). HL has served as a speaker for Evolan Pharma and Shire and has received research grants from Shire; all outside the submitted work. The other authors report no financial COI.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41380-022-01500-2.

Correspondence and requests for materials should be addressed to Patrick F. Sullivan.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.