Psychological Medicine, Page 1 of 9. © Cambridge University Press 2017
doi:10.1017/S0033291717002665

ORIGINAL ARTICLE

Genetic risk scores and family history as predictors of schizophrenia in Nordic registers

Y. Lu[1], J. G. Pouget[2], O. A. Andreassen[3], S. Djurovic[4,5], T. Esko[6], C. M. Hultman[1], A. Metspalu[6], L. Milani[6], T. Werge[7,8,9] and P. F. Sullivan[1,10]*

[1] *Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, SE-17177 Stockholm, Sweden*
[2] *Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Ontario, Canada*
[3] *NORMENT, KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo and Oslo University Hospital, 0424 Oslo, Norway*
[4] *Department of Medical Genetics, Oslo University Hospital, Oslo, Norway*
[5] *Department of Clinical Science, NORMENT, KG Jebsen Centre for Psychosis Research, University of Bergen, Bergen, Norway*
[6] *Estonian Genome Center, University of Tartu, Tartu, Estonia*
[7] *The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark*
[8] *Institute of Biological Psychiatry, MHC Sct. Hans, Mental Health Services Copenhagen, Denmark*
[9] *Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark*
[10] *Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, NC, 27599-7264, USA*

Background. Family history is a long-standing and readily obtainable risk factor for schizophrenia (SCZ). Low-cost genotyping technologies have enabled large genetic studies of SCZ, and the results suggest the utility of genetic risk scores (GRS, direct assessments of inherited common variant risk). Few studies have evaluated family history and GRS simultaneously to ask whether one can explain away the other.

Methods. We studied 5959 SCZ cases and 8717 controls from four Nordic countries. All subjects had family history data from national registers and genome-wide genotypes that were processed through the quality control procedures used by the Psychiatric Genomics Consortium. Using external training data, GRS were estimated for SCZ, bipolar disorder (BIP), major depression, autism, educational attainment, and body mass index. Multivariable modeling was used to estimate effect sizes.

Results. Using harmonized genomic and national register data from Denmark, Estonia, Norway, and Sweden, we confirmed that family history of SCZ and GRS for SCZ and BIP were risk factors for SCZ. In a joint model, the effects of GRS for SCZ and BIP were essentially unchanged, and the effect of family history was attenuated but remained significant. The predictive capacity of a model including GRS and family history neared the minimum for clinical utility.

Conclusions. Combining national register data with measured genetic risk factors represents an important investigative approach for psychotic disorders. Our findings suggest the potential clinical utility of combining GRS and family history for early prediction and diagnostic improvements.

Received 18 March 2017; Revised 14 August 2017; Accepted 15 August 2017

Key words: Family history, genetic risk score, genome-wide association, GWAS, schizophrenia.

Introduction

Schizophrenia (SCZ) is a debilitating psychiatric disorder that is associated with substantial health and economic burden for patients, families, and society (Chong *et al.* 2016). Its genetic component is substantial with estimated heritability of 77% from twin studies (Polderman *et al.* 2015) and shared environment accounts for a significant fraction of liability to SCZ (~10%) (Sullivan *et al.* 2003). These estimates are broadly consistent with national pedigree studies from Denmark and Sweden (Lichtenstein *et al.* 2009; Wray & Gottesman, 2012). These results underscore the importance of both genes and environment in the etiology of SCZ (Mortensen *et al.* 1999; van Os *et al.* 2005; March & Susser, 2006).

A family history of SCZ in first-degree relatives is a strong and robust risk factor for SCZ (Mortensen *et al.* 1999), with relative risk estimates ranging from 7.0 to 9.3 (Mortensen *et al.* 1999). However, given its low

* Address for correspondence: Dr P. F. Sullivan, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, SE-17177 Stockholm, Sweden and Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, NC, 27599-7264, USA.
(Email: patrick.sullivan@ki.se)

lifetime risk (<1%) (Lichtenstein *et al.* 2009), the absolute risks of a positive family history are low and the vast majority of SCZ cases (>95%) (Lichtenstein *et al.* 2006) do not have an immediate family history of SCZ (Yang *et al.* 2010).

Knowledge of the genetic basis of SCZ has advanced markedly in the past decade (Ripke *et al.* 2013). A substantial fraction of the heritability of SCZ is due to numerous common genetic variants each carrying small changes in risk (Lee *et al.* 2012*a*). One implication of these discoveries is the capacity to estimate individual genetic risk scores (GRS) (International Schizophrenia Consortium, 2009) which, for the first time, quantify genetic risks for an individual (McGrath *et al.* 2013). Case–control differences in SCZ GRS have been widely replicated (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Agerbo *et al.* 2015). GRS are a continuous measure of genetic liability and have been suggested to be a more powerful and more direct assessment of liability compared with a dichotomous family history measure (McGrath *et al.* 2013). Unlike family history, GRS does not suffer from information loss due to recall bias, reduced fertility, or incomplete knowledge of family history.

A major distinction between family history and GRS is the extent of information captured. Family history captures the overt results of shared genetics (common and rare genetic variants) and shared environmental risks among family members. GRS are comprised of common but probably non-causal genetic markers. Indeed, there may be low correlations between these two measures (Belsky *et al.* 2013*a*, *b*), which could be due to the small proportion of variation explained by current GRS for most traits. One study found that SCZ cases with a positive family history had higher GRS compared with family history negative cases (Bigdeli *et al.* 2016). However, there has been little research on the joint contribution of family history and GRS to liability for SCZ. One study found that the effect of family history was mediated through GRS (Agerbo *et al.* 2015).

In this study, we aimed to investigate the interplay of the conventional measure of family history and GRS. The Nordic countries have multiple advantages for population-based research. These countries have universal health care and a long tradition of collecting high-quality and comprehensive health care data (i.e. diagnostic data from hospital admissions and specialist outpatient treatment contacts). This makes it possible to recruit representative samples, unbiased by sociodemographic factors or insurance plans. These factors are important for generating large samples and facilitating psychiatric genetics research. The current study builds on these advantages by combining large samples across the Nordic countries. We expand upon prior research by combining data across four Nordic countries, incorporating systematically captured family history, and the use of GRS for multiple traits. We sought to determine if the effects of family history could be reduced or eradicated by adding GRS based on direct measures obtained from the genomes of cases and controls. In other words, can a direct measure of genetic liability like GRS replace cruder family history measures used in epidemiological and clinical studies?

Methods

Samples

We evaluated data from Denmark, Estonia, Norway, and Sweden (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), all of which have universal public health care systems. Case/control status and family history of SCZ and other psychiatric disorders were extracted from national health registers that systematically capture International Classification of Diseases (ICD) diagnoses from inpatient admissions and outpatient specialist treatment contacts. As in previous studies (Schizophrenia Psychiatric Genome-Wide Association Study Consortium, 2011; Ripke *et al.* 2013; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), the case definition included individuals with SCZ or schizoaffective disorder. For family history, the first-degree relatives of cases and controls were identified via linkage with multi-generational registers.

Danish cases were ascertained through psychiatric departments in the Copenhagen area and were of Danish parentage. Controls of Danish parentage were collected from the Danish Blood Donor Corps in the same catchment area. All subjects gave written informed consent and the Danish Data Protection Agency and the ethics committees of Denmark approved the human subjects protocol. The *Estonian* cohort comes from the population-based Estonian Genome Project (Metspalu *et al.* 2004). The project was conducted according to the Estonian Gene Research Act and all participants provided informed consent. Participation was voluntary and no gender balance was controlled for. In total, 52 000 individuals aged 18 years or older participated. General practitioners and physicians in hospitals recruited participants but SCZ was diagnosed by psychiatrists according to ICD-10 criteria. Controls were drawn from a larger pool of genotyped biobank samples by matching on sex, age, and genetic ancestry. All controls were population-based and have not been sampled for any specific disease. *Norwegian* cases of European ancestry, born in Norway, were recruited from psychiatric hospitals in the Oslo region and diagnosed using the SCID (Athanasiu *et al.* 2010). Healthy controls were

randomly selected from statistical records of persons from the same catchment area as cases. Controls were required to have no family history of any psychiatric disorder. All participants provided written informed consent and the human subjects protocol was approved by the Norwegian Scientific-Ethical Committee and the Norwegian Data Protection Agency. Cases and controls from the *Swedish* Schizophrenia Study (Ripke *et al.* 2013) were identified via the Swedish Hospital Discharge Register which captures all public and private inpatient hospitalizations. Case inclusion criteria included $\geqslant 2$ hospitalizations with a discharge diagnosis of SCZ or schizoaffective disorder and age $\geqslant 18$ years. Case exclusion criteria included hospital register diagnosis of any medical or psychiatric disorder mitigating a confident diagnosis of SCZ as determined by expert review. The validity of this case definition of SCZ was strongly supported by clinical, epidemiological, genetic epidemiological and genetic evidence (Ripke *et al.* 2013). Controls were selected at random from Swedish population registers. Control inclusion criteria included never being hospitalized for SCZ or bipolar disorder (BIP) and age $\geqslant 18$ years. All procedures were approved by ethical committees and all subjects provided written informed consent (or legal guardian consent and subject assent).

GRS were estimated for all target Nordic subjects using best practices (online Supplementary Fig. S1) (International Schizophrenia Consortium, 2009; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). We derived GRS for SCZ, BIP, major depressive disorder (MDD), autism spectrum disorder (ASD), body mass index (BMI), and educational attainment (EDU). These traits all have significant genetic correlations with SCZ (Cross-Disorder Group of the Psychiatric Genomics Consortium, 2013; Okbay *et al.* 2016; Zheng *et al.* 2016). We obtained genome-wide association study (GWAS) results for SCZ, BIP, ASD, and MDD from the Psychiatric Genomics Consortium (PGC) (URLs), and reran meta-analysis of case–control data from European samples but excluding Nordic sites. The results of meta-analyses from these non-Nordic samples were considered as discovery sets to select SNPs and to weight the risk alleles during the calculation of GRS. We found limited evidence for sample overlap with our target sets (online Supplementary Table S1). Genome-wide association (GWA) summary statistics for BMI were obtained from GIANT (Locke *et al.* 2015). GWA summary statistics for the total number of years of education (a proxy for general intelligence) from SSGAC were used to derive GRS (after excluding Swedish studies) (Okbay *et al.* 2016). Meta-analyses were performed using METAL (Willer *et al.* 2010). The genotype data in the Nordic target sets were processed using the PGC Ricopilli pipeline for quality control (QC) and imputation. Cryptic relatedness was checked in the Ricopilli pipeline, where one in each pair of related individuals ($\hat{\pi} > 0.2$) was excluded. Variants were excluded for the following reasons: not biallelic, lacking The Short Genetic Variations database (dbSNP) name, strand ambiguous, allele frequency <0.05 or >0.95, or poor imputation quality (R^2 from MACH (Scott *et al.* 2007) or INFO from IMPUTE (Howie *et al.* 2009) < 0.8). To select a relatively independent set of SNPs for the calculation of GRS, we ran linkage disequilibrium (LD clumping) ($r^2 < 0.1$ in 1Mb window) on the overlapping SNPs using 1000 Genomes Project European samples as LD reference. Online Supplementary Table S2 has details on the number of SNPs used for calculation of GRS.

We calculated GRS in the target sets as the sum of the SNP dosages weighted by the effect from the discovery set (logistic regression β for the case/control psychiatric disorders or multiple regression β for BMI and EDU) across all SNPs under a specified p value threshold. All β values were recoded to be positive. GRS were calculated for eight p value cutoffs (P_T): $\leqslant 5 \times 10^{-8}$, $\leqslant 1 \times 10^{-5}$, $\leqslant 1 \times 10^{-3}$, $\leqslant 0.01$, $\leqslant 0.05$, $\leqslant 0.1$, $\leqslant 0.5$, and $\leqslant 1$. We used PLINK (v1.9) for the calculation of GRS (Chang *et al.* 2015).

Statistical methods

We used logistic regression to test the association of family history and different GRS on SCZ case/control status. Family history was dichotomized as negative (i.e. no first-degree relative with SCZ or schizoaffective disorder) or positive (i.e. at least one first-degree relative with the condition).

The associations between the GRS of four psychiatric disorders with SCZ were first tested within each Nordic dataset. We included 20 ancestry principal components (PCs). The variance in SCZ [Nagelkerke's (Nagelkerke, 1991) pseudo-R^2] explained by GRS was calculated as the difference in R^2 from the full model including the GRS and the baseline model with PCs only (online Supplementary Fig. S2). Assuming a life-time risk of 1%, we reported variance explained on the liability scale adjusted for case–control ascertainment (Lee *et al.* 2012*b*; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). GRS were standardized within each target set prior to the main analyses to account for variation in numbers of SNPs used for GRS calculation across studies. We used the GRS from the SNP set that was selected from each of the discovery sets at $P_T \leqslant 0.05$, given that variance in SCZ explained by most of GRS plateaued at this cut-off (online Supplementary Fig. S2). We included the GRS of all four psychiatric disorders due to the known moderate to strong genetic overlap among these disorders (Cross-Disorder Group of the Psychiatric Genomics

Consortium, 2013) (online Supplementary Table S3). We included GRS for BMI and EDU given their genetic correlations with SCZ (Zheng *et al.* 2016). GRS were included as continuous variables. To aid the comparison of ORs from GRS and from family history, we also divided the GRS of SCZ into three subgroups: low genetic risk (GRS in the first quintile), intermediate genetic risk (second to fourth quintile), and high genetic risk (fifth quintile).

The main analyses included: (a) fitting separate logistic regression models for family history (adjusting for study: SCZ diagnosis ~ family history + study indicator) and for the GRS of four psychiatric disorders and of BMI and EDU (adjusting for study and PCs: SCZ diagnosis ~ GRS_SCZ + GRS_BIP + GRS_MDD + GRS_ASD + GRS_BMI + GRS_EDU + study indicator + PCs); and (b) joint estimation of family history and all GRS (adjusting for the studies and PCs: SCZ diagnosis ~ family history + GRS_SCZ + GRS_BIP + GRS_MDD + GRS_ASD + GRS_BMI + GRS_EDU + study indicator + PCs). We also tested the significance of the interaction term of family history and the GRS of SCZ in addition to the main effects. All analyses were performed using *R* (v3.3.2).

Results

This study included 5959 SCZ cases and 8717 controls from four Nordic countries. Table 1 shows sample characteristics for individual studies. The prevalence of a positive family history was ~1% in controls (except for Norwegian controls ascertained for the absence of family history) and 10–14% in cases (Table 1). SCZ cases had more first-degree relatives with BIP or other mental disorders than controls (Table 1). Consistent with epidemiological estimates (Lichtenstein *et al.* 2006), a positive family history was associated with an OR of 9.70 (95% CI 7.99–11.79, Table 2), which explained 3.5% of variance on the liability scale of SCZ calculated from a logistic regression model and adjusted for case–control ascertainment.

The mean values of GRS for all four psychiatric disorders were higher among cases than controls within each study site (Fig. 1, online Supplementary Table S4). Case–control differences were most pronounced for GRS for SCZ and BIP. Across all four study sites, the variance on the liability scale explained by single GRS was 5.3, 1.8, 0.2, and 0.02% for SCZ, BIP, MDD, and ASD. When GRS for all four psychiatric disorders were included in the model, the total variance explained was 6.0%, which was less than the sum of variances explained by individual GRS due to significant correlations among the GRS (online Supplementary Table S5). An increase in SCZ GRS of one standard deviation was associated with 83% greater risk for SCZ (OR 1.83, 95% CI 1.76–1.91), and the corresponding figures were 26% for BIP GRS, 3% for ASD GRS, and 3% for MDD GRS (Table 2). GRS for BMI and EDU were associated with SCZ at borderline significance (OR 0.97 for GRS of BMI, and 1.04 for GRS of EDU).

Cases with a positive family history had a significantly higher mean value of the GRS of SCZ when compared with those with a negative family history (mean = 0.47 as compared with 0.32; $p = 7 \times 10^{-5}$; online Supplementary Fig. S3). In the absence of a family history, the risk to SCZ was increased by 2.48-fold (95% CI 2.21–2.77) and by 5.49-fold (95% CI 4.79–6.30) for individuals with intermediate and high genetic risk when compared with the low GRS group.

Joint estimation of the risks of family history and all GRS yielded a lower OR associated with family history (OR 8.37, 95% CI 6.85–10.23; Table 2) as compared with the results from the separate models. The effects from all the GRS were essentially unchanged. Collinearity was not severe owing to a low correlation between GRS and family history (Spearman correlation of 0.11 among all samples). Family history and multiple GRS, together, explained 8.9% of variance on the liability scale. The interaction between family history and the GRS was not significant (p = 0.11) when tested together with the main effects. Thus, in the joint model, the effect of family history remained significant despite inclusion of multiple GRS that assessed common variant liability directly.

Discussion

We leveraged the unique resources of genomic and medical register data from four Nordic countries, and provided evidence that genetic risk profile and family history are strong predictors of SCZ, even when considered jointly.

We first replicated previous findings that GRS of SCZ and related disorders were robustly associated with SCZ, together accounting for 6% of variation on the liability scale after adjustment of case–control ascertainment (Lee *et al.* 2012*b*). The variance explained was smaller than reported before (Schizophrenia Working Group of the Psychiatric Genomics, 2014) because, by necessity, we had to exclude all Nordic samples from the discovery samples. This highlighted that the accuracy of GRS will improve as sample sizes amassed for discovery continue to increase. We used Nordic registers to accurately identify family history, and found that a positive family history (defined as $\geqslant$1 first-degree relatives with SCZ or schizoaffective disorder) was associated with a nearly 10-fold risk to SCZ, and subjects with a positive family history carried higher GRS burden compared with those without family history. The risk estimate was highly consistent with the estimates of relative risks in the literature (Lichtenstein *et al.* 2006). However, a comprehensive

Table 1. *Sample characteristics*

	Denmark		Estonia		Norway		Sweden	
	Case	Control	Case	Control	Case	Control	Case	Control
N	463	956	233	1153	339	401	4924	6207
Age, mean	50.7	49.9	46.2	46.4	32.7	34.8	53.9	56.3
% Male	58.3%	61.8%	26.2%	26.8%	55.8%	50.1%	59.7%	51.2%
Family history: number of first-degree relative with SCZ or schizoaffective disorder (%)								
0	363 (78.4)	875 (91.5)	N.A.	N.A.	278 (82.0)	401 (100)	4001 (81.3)	6015 (96.9)
1	33 (7.1)	11 (1.2)	N.A.	N.A.	40 (11.8)		581 (11.8)	106 (1.7)
>1	13 (2.8)	1 (0.1)	N.A.	N.A.	3 (0.9)		94 (1.9)	5 (0.1)
Family history: number of first-degree relative with bipolar disorder (%)								
0	389 (84.0)	880 (92.1)	N.A.	N.A.	300 (88.5)	401 (100)	4102 (83.3)	5829 (93.9)
1	20 (4.3)	7 (0.7)	N.A.	N.A.	19 (5.6)		490 (10.0)	272 (4.4)
>1	0 (0)	0 (0)	N.A.	N.A.	1 (0.3)		84 (1.7)	25 (0.4)
Family history: number of first-degree relative with other mental disorders (%)								
0	234 (50.5)	626 (65.5)	N.A.	N.A.	157 (46.3)	401 (100)	2474 (50.2)	4130 (66.5)
1	132 (28.5)	222 (23.2)	N.A.	N.A.	111 (32.7)		1204 (24.5)	1279 (20.6)
>1	43 (9.3)	39 (4.1)	N.A.	N.A.	52 (15.3)		998 (20.3)	717 (11.6)

The percentages in family history measures do not add up to 1 due to missing data; no sufficient information to determine the number of first-degree relatives with psychiatric disorders in the Estonian cohort; controls in the Norwegian cohort were required to have no family history of any psychiatric disorder.

Table 2. *Separate and joint estimations of family history and GRS*

	Separate estimation		Joint estimation	
	OR (95% CI)	*p* value	OR (95% CI)	*p* value
Family history	9.70 (7.99–11.79)	3×10^{-116}	8.37 (6.85–10.23)	5×10^{-96}
GRS of four psychiatric disorders $P_T \leqslant 0.05$				
SCZ	1.83 (1.76–1.91)	5×10^{-191}	1.82 (1.74–1.90)	2×10^{-159}
BIP	1.26 (1.22–1.31)	5×10^{-34}	1.25 (1.20–1.30)	1×10^{-26}
MDD	1.03 (0.99–1.07)	0.11	1.02 (0.98–1.06)	0.26
ASD	1.03 (1.00–1.07)	0.06	1.04 (1.00–1.08)	0.04
GRS of BMI and EDU, $P_T \leqslant 0.05$				
BMI	0.97 (0.93–1.00)	0.08	0.97 (0.93–1.01)	0.17
EDU	1.04 (1.00–1.08)	0.03	1.04 (1.00–1.08)	0.08

Results based on the dataset excluding Norwegian study were provided in online Supplementary Table S6.

measure of family history explained around half of the variance explained by directly measured genetic risk. In accordance with simulations in a previous report (Do *et al.* 2012), GRS outperforms traditional family history, and will likely improve further with larger studies now in analysis (Sullivan *et al.* 2017).

The effect of family history was attenuated but remained significant when jointly analyzed with GRS measures for SCZ, BIP, MDD, ASD, EDU, and BMI. This indicates that family history contains information not captured by these multiple GRS. Intuitively, the reduction in the effect of family history was due to shared genetic information captured by both measures. We expect that the evolving GRS with better accuracy will continue to capture a higher proportion of SCZ risk associated with family history, although it will be inherently capped at a level determined by the SNP heritability, i.e. the phenotypic variance attributable to common genetic variants that are tagged on the genotyping arrays. The independent effects captured by family history could be due to common or rare genetic variation not well captured by the SNP arrays used in these GWAS as well as shared environmental risk factors. On the other hand, few individuals would have a positive family history, which may limit its utility as a predictive measure.

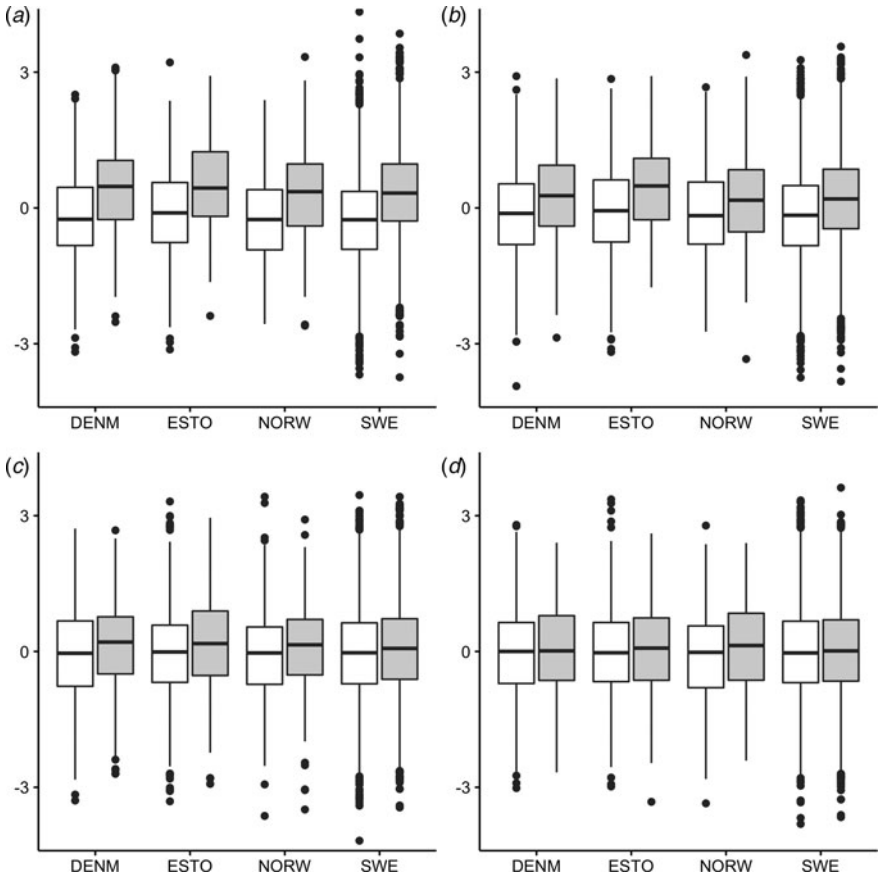

Fig. 1. Boxplots for GRS mean differences between cases/controls ($P_T \leqslant 0.05$) (*a*) GRS of schizophrenia; (*b*) GRS of bipolar disorder; (*c*) GRS of major depressive disorder; (*d*) GRS of autism spectrum disorder in the four Nordic studies ('DENM' Denmark, 'ESTO' Estonia, 'NORW' Norway, and 'SWE' Sweden). Controls in white and cases in grey.

We note that, in the absence of a family history, there is a risk gradient associated with different profiles of GRS. The combination of GRS and comprehensive family history explained 9% of variance in SCZ liability – better discrimination accuracy can be achieved by jointly considering GRS and family history. This has proven useful in other diseases. For prostate cancer, integrating family history, prostate cancer GRS, and protein biomarkers led to a clinically important prediction algorithm that is now being implemented in routine clinical care (Gronberg *et al.* 2015). In our data, joint modeling both measures yielded an area under the receiver operating characteristic (AUC) curve of 0.73. This level of discriminative accuracy is close to the AUC threshold of 0.75, which is regarded as the minimum for clinical utility (Janssens *et al.* 2007; Wray *et al.* 2010). Future studies should evaluate the impact of additional epidemiological and genetic risk factors (including the burden of rare copy number variants and exon variation). Together with other risk factors, a reasonable predictive utility may be attainable.

In conclusion, by integrating register data and carefully curated common-variant SNP data from multiple Nordic countries, we determined that 'family history' contains information that is not captured by GRS for multiple psychiatric disorders and EDU. By combining family history with these multiple GRS, we obtained levels of discriminative accuracy that – with replication in other studies and in samples with a lower prior probability of SCZ than in these case–control cohorts – could prove to be clinically useful.

Supplementary material

The supplementary material for this article can be found at https://doi.org/10.1017/S0033291717002665.

Acknowledgements

This project was supported by 'Tryggve' project (https://wiki.neic.no/wiki/Tryggve) designed to establish a Nordic platform for collaboration on sensitive data. Support from the NeIC (Nordic e-Infrastructure Collaboration, https://neic.nordforsk.org) Tryggve project and NBIS (National Bioinformatics Infrastructure Sweden) for the use of the Mosler system for sensitive

data is gratefully acknowledged. PFS gratefully acknowledges support from the Swedish Research Council (Vetenskapsrådet, award D0886501). Creation of the Sweden schizophrenia study data was supported by NIMH R01 MH077139 and the Stanley Center of the Broad Institute. The PGC is supported by NIMH U01 MH109528. The Norwegian sample was included with support from Research Council of Norway (RCN 223273, 213837) and KG Jebsen Stiftelsen.

Declaration of Interest

PFS reports the following potentially competing financial interests: Lundbeck (advisory committee), Pfizer (Scientific Advisory Board member), and Roche (grant recipient, speaker reimbursement). PFS is a scientific advisor for Pfizer, Inc. OAA received speaker honoraria from Lundbeck.

URLs

MaTCH, http://match.ctglab.nl. Ricopili, https://github.com/Nealelab/ricopili. Psychiatric Genomics Consortium, https://pgc.unc.edu

References

Agerbo E, Sullivan PF, Vilhjalmsson BJ, Pedersen CB, Mors O, Borglum AD, Hougaard DM, Hollegaard MV, Meier S, Mattheisen M, Ripke S, Wray NR, Mortensen PB (2015). Polygenic risk score, parental socioeconomic status, family history of psychiatric disorders, and the risk for schizophrenia: a Danish population-based study and meta-analysis. *JAMA Psychiatry* **72**, 635–641.

Athanasiu L, Mattingsdal M, Kahler AK, Brown A, Gustafsson O, Agartz I, Giegling I, Muglia P, Cichon S, Rietschel M, Pietilainen OPH, Peltonen L, Bramon E, Collier D, Clair DS, Sigurdsson E, Petursson H, Rujescu D, Melle I, Steen VM, Djurovic S, Andreassen OA (2010). Gene variants associated with schizophrenia in a Norwegian genome-wide study are replicated in a large European cohort. *Journal of Psychiatric Research* **44**, 748–753.

Belsky DW, Moffitt TE, Caspi A (2013*a*). Genetics in population health science: strategies and opportunities. *American Journal of Public Health* **103** (Suppl. 1), S73–S83.

Belsky DW, Sears MR, Hancox RJ, Harrington H, Houts R, Moffitt TE, Sugden K, Williams B, Poulton R, Caspi A (2013*b*). Polygenic risk and the development and course of asthma: an analysis of data from a four-decade longitudinal study. *Lancet Respiratory Medicine* **1**, 453–461.

Bigdeli TB, Ripke S, Bacanu SA, Lee SH, Wray NR, Gejman PV, Rietschel M, Cichon S, St Clair D, Corvin A, Kirov G, Mcquillin A, Gurling H, Rujescu D, Andreassen OA, Werge T, Blackwood DH, Pato CN, Pato MT, Malhotra AK, O'donovan MC, Kendler KS, Fanous AH, Schizophrenia Working Group of the Psychiatric Genomics C (2016). Genome-wide association study reveals greater polygenic loading for schizophrenia in cases with a family history of illness. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics* **171B**, 276–289.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7.

Chong HY, Teoh SL, Wu DB, Kotirum S, Chiou CF, Chaiyakunapruk N (2016). Global economic burden of schizophrenia: a systematic review. *Neuropsychiatric Disease and Treatment* **12**, 357–373.

Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics* **45**, 984–994.

Do CB, Hinds DA, Francke U, Eriksson N (2012). Comparison of family history and SNPs for predicting risk of complex disease. *PLoS Genetics* **8**, e1002973.

Gronberg H, Adolfsson J, Aly M, Nordstrom T, Wiklund P, Brandberg Y, Thompson J, Wiklund F, Lindberg J, Clements M, Egevad L, Eklund M (2015). Prostate cancer screening in men aged 50–69 years (STHLM3): a prospective population-based diagnostic study. *The Lancet Oncology* **16**, 1667–1676.

Howie BN, Donnelly P, Marchini J (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* **5**, e1000529.

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

Janssens AC, Moonesinghe R, Yang Q, Steyerberg EW, Van Duijn CM, Khoury MJ (2007). The impact of genotype frequencies on the clinical validity of genomic profiling for predicting common chronic diseases. *Genetics in Medicine* **9**, 528–535.

Lee SH, Decandia TR, Ripke S, Yang J, Schizophrenia Psychiatric Genome-Wide Association Study C, International Schizophrenia C, Molecular Genetics of Schizophrenia C, Sullivan PF, Goddard ME, Keller MC, Visscher PM, Wray NR (2012*a*). Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nature Genetics* **44**, 247–250.

Lee SH, Goddard ME, Wray NR, Visscher PM (2012*b*). A better coefficient of determination for genetic profile analysis. *Genetic Epidemiology* **36**, 214–224.

Lichtenstein P, Bjork C, Hultman CM, Scolnick EM, Sklar P, Sullivan PF (2006). Recurrence risks for schizophrenia in a Swedish national cohort. *Psychological Medicine* **36**, 1417–1426.

Lichtenstein P, Yip B, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, Hultman CM (2009). Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* **373**, 234–239.

Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, Powell C, Vedantam S, Buchkovich ML, Yang J, Croteau-Chonka DC, Esko T, Fall T, Ferreira T, Gustafsson S, Kutalik Z, Luan J, Magi R, Randall JC, Winkler TW, Wood AR, Workalemahu T, Faul JD, Smith JA, Hua Zhao J, Zhao W, Chen J, Fehrmann R, Hedman AK, Karjalainen J, Schmidt EM, Absher D, Amin N, Anderson D, Beekman M, Bolton JL, Bragg-Gresham JL,

Buyske S, Demirkan A, Deng G, Ehret GB, Feenstra B, Feitosa MF, Fischer K, Goel A, Gong J, Jackson AU, Kanoni S, Kleber ME, Kristiansson K, Lim U, Lotay V, Mangino M, Mateo Leach I, Medina-Gomez C, Medland SE, Nalls MA, Palmer CD, Pasko D, Pechlivanis S, Peters MJ, Prokopenko I, Shungin D, Stancakova A, Strawbridge RJ, Ju Sung Y, Tanaka T, Teumer A, Trompet S, Van Der Laan SW, Van Setten J, Van Vliet-Ostaptchouk JV, Wang Z, Yengo L, Zhang W, Isaacs A, Albrecht E, Arnlov J, Arscott GM, Attwood AP, Bandinelli S, Barrett A, Bas IN, Bellis C, Bennett AJ, Berne C, Blagieva R, Bluher M, Bohringer S, Bonnycastle LL, Bottcher Y, Boyd HA, Bruinenberg M, Caspersen IH, Ida Chen YD, Clarke R, Daw EW, De Craen AJ, Delgado G, Dimitriou M, Doney ASF, Eklund N, Estrada K, Eury E, Folkersen L, Fraser RM, Garcia ME, Geller F, Giedraitis V, Gigante B, Go AS, Golay A, Goodall AH, Gordon SD, Gorski M, Grabe HJ, Grallert H, Grammer TB, Grassler J, Gronberg H, Groves CJ, Gusto G, Haessler J, Hall P, Haller T, Hallmans G, Hartman CA, Hassinen M, Hayward C, Heard-Costa NL, Helmer Q, Hengstenberg C, Holmen O, Hottenga JJ, James AL, Jeff JM, Johansson A, Jolley J, Juliusdottir T, Kinnunen L, Koenig W, Koskenvuo M, Kratzer W, Laitinen J, Lamina C, Leander K, Lee NR, Lichtner P, Lind L, Lindstrom J, Lo KS, Lobbens S, Lorbeer R, Lu Y, Mach F, Magnusson PKE, Mahajan A, McArdle WL, McLachlan S, Menni C, Merger S, Mihailov E, Milani L, Moayyeri A, Monda KL, Morken MA, Mulas A, Muller G, Muller-Nurasyid M, Musk AW, Nagaraja R, Nothen MM, Nolte IM, Pilz S, Rayner NW, Renstrom F, Rettig R, Ried JS, Ripke S, Robertson NR, Rose LM, Sanna S, Scharnagl H, Scholtens S, Schumacher FR, Scott WR, Seufferlein T, Shi J, Smith AV, Smolonska J, Stanton AV, Steinthorsdottir V, Stirrups K, Stringham HM, Sundstrom J, Swertz MA, Swift AJ, Syvanen AC, Tan ST, Tayo BO, Thorand B, Thorleifsson G, Tyrer JP, Uh HW, Vandenput L, Verhulst FC, Vermeulen SH, Verweij N, Vonk JM, Waite LL, Warren HR, Waterworth D, Weedon MN, Wilkens LR, Willenborg C, Wilsgaard T, Wojczynski MK, Wong A, Wright AF, Zhang Q, Brennan EP, Choi M, Dastani Z, Drong AW, Eriksson P, Franco-Cereceda A, Gadin JR, Gharavi AG, Goddard ME, Handsaker RE, Huang J, Karpe F, Kathiresan S, Keildson S, Kiryluk K, Kubo M, Lee JY, Liang L, Lifton RP, Ma B, McCarroll SA, McKnight AJ, Min JL, Moffatt MF, Montgomery GW, Murabito JM, Nicholson G, Nyholt DR, Okada Y, Perry JRB, Dorajoo R, Reinmaa E, Salem RM, Sandholm N, Scott RA, Stolk L, Takahashi A, Tanaka T, van'tHooft FM, Vinkhuyzen AAE, Westra HJ, Zheng W, Zondervan KT, Heath AC, Arveiler D, Bakker SJL, Beilby J, Bergman RN, Blangero J, Bovet P, Campbell H, Caulfield MJ, Cesana G, Chakravarti A, Chasman DI, Chines PS, Collins FS, Crawford DC, Cupples LA, Cusi D, Danesh J, deFaire U, denRuijter HM, Dominiczak AF, Erbel R, Erdmann J, Eriksson JG, Farrall M, Felix SB, Ferrannini E, Ferrieres J, Ford I, Forouhi NG, Forrester T, Franco OH, Gansevoort RT, Gejman PV, Gieger C, Gottesman O, Gudnason V, Gyllensten U, Hall AS, Harris TB, Hattersley AT, Hicks AA, Hindorff LA, Hingorani AD, Hofman A, Homuth G, Hovingh GK, Humphries SE, Hunt SC, Hypponen E, Illig T, Jacobs KB, Jarvelin MR, Jockel KH, Johansen B, Jousilahti P, Jukema JW, Jula AM, Kaprio J, Kastelein JJP, Keinanen-Kiukaanniemi SM, Kiemeney LA, Knekt P, Kooner JS, Kooperberg C, Kovacs P, Kraja AT, Kumari M, Kuusisto J, Lakka TA, Langenberg C, Marchand LL, Lehtimaki T, Lyssenko V, Mannisto S, Marette A, Matise TC, McKenzie CA, McKnight B, Moll FL, Morris AD, Morris AP, Murray JC, Nelis M, Ohlsson C, Oldehinkel AJ, Ong KK, Madden PAF, Pasterkamp G, Peden JF, Peters A, Postma DS, Pramstaller PP, Price JF, Qi L, Raitakari OT, Rankinen T, Rao DC, Rice TK, Ridker PM, Rioux JD, Ritchie MD, Rudan I, Salomaa V, Samani NJ, Saramies J, Sarzynski MA, Schunkert H, Schwarz PEH, Sever P, Shuldiner AR, Sinisalo J, Stolk RP, Strauch K, Tonjes A, Tregouet DA, Tremblay A, Tremoli E, Virtamo J, Vohl MC, Volker U, Waeber G, Willemsen G, Witteman JC, Zillikens MC, Adair LS, Amouyel P, Asselbergs FW, Assimes TL, Bochud M, Boehm BO, Boerwinkle E, Bornstein SR, Bottinger EP, Bouchard C, Cauchi S, Chambers JC, Chanock SJ, Cooper RS, deBakker PIW, Dedoussis G, Ferrucci L, Franks PW, Froguel P, Groop LC, Haiman CA, Hamsten A, Hui J, Hunter DJ, Hveem K, Kaplan RC, Kivimaki M, Kuh D, Laakso M, Liu Y, Martin NG, Marz W, Melbye M, Metspalu A, Moebus S, Munroe PB, Njolstad I, Oostra BA, Palmer CNA, Pedersen NL, Perola M, Perusse L, Peters U, Power C, Quertermous T, Rauramaa R, Rivadeneira F, Saaristo TE, Saleheen D, Sattar N, Schadt EE, Schlessinger D, Slagboom PE, Snieder H, Spector TD, Thorsteinsdottir U, Stumvoll M, Tuomilehto J, Uitterlinden AG, Uusitupa M, vanderHarst P, Walker M, Wallaschofski H, Wareham NJ, Watkins H, Weir DR, Wichmann HE, Wilson JF, Zanen P, Borecki IB, Deloukas P, Fox CS, Heid IM, O'Connell JR, Strachan DP, Stefansson K, vanDuijn CM, Abecasis GR, Franke L, Frayling TM, McCarthy MI, Visscher PM, Scherag A, Willer CJ, Boehnke M, Mohlke KL, Lindgren CM, Beckmann JS, Barroso I, North KE, Ingelsson E, Hirschhorn JN, Loos RJF, Speliotes EK (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206.

March D, Susser E (2006). Invited commentary: taking the search for causes of schizophrenia to a different level. *American Journal of Epidemiology* **163**, 979–981.

Mcgrath JJ, Mortensen PB, Visscher PM, Wray NR (2013). Where GWAS and epidemiology meet: opportunities for the simultaneous study of genetic and environmental risk factors in schizophrenia. *Schizophrenia Bulletin* **39**, 955–959.

Metspalu A, Kohler F, Laschinski G, Ganten D, Roots I (2004). The Estonian genome project in the context of European genome research. *Deutsche medizinische Wochenschrift* **129** (Suppl. 1), S25–S28.

Mortensen PB, Pedersen CB, Westergaard T, Wohlfahrt J, Ewald H, Mors O, Andersen PK, Melbye M (1999). Effects of family history and place and season of birth on the risk of schizophrenia. *New England Journal of Medicine* **340**, 603–608.

Nagelkerke NJD (1991). A note on a general definition of the coefficient of determination. *Biometrika* **78**, 691–692.

Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, Turley P, Chen GB, Emilsson V, Meddens SF, Oskarsson S, Pickrell JK, Thom K, Timshel P, De Vlaming R, Abdellaoui A, Ahluwalia TS, Bacelis J, Baumbach C, Bjornsdottir G, Brandsma JH, Pina Concas M, Derringer J, Furlotte NA, Galesloot TE, Girotto G, Gupta R, Hall LM, Harris SE, Hofer E, Horikoshi M, Huffman JE, Kaasik K, Kalafati IP, Karlsson R, Kong A, Lahti J, Van Der Lee SJ, Deleeuw C, Lind PA, Lindgren KO, Liu T, Mangino M, Marten J, Mihailov E, Miller MB, Van Der Most PJ, Oldmeadow C, Payton A, Pervjakova N, Peyrot WJ, Qian Y, Raitakari O, Rueedi R, Salvi E, Schmidt B, Schraut KE, Shi J, Smith AV, Poot RA, St Pourcain B, Teumer A, Thorleifsson G, Verweij N, Vuckovic D, Wellmann J, Westra HJ, Yang J, Zhao W, Zhu Z, Alizadeh BZ, Amin N, Bakshi A, Baumeister SE, Biino G, Bonnelykke K, Boyle PA, Campbell H, Cappuccio FP, Davies G, De Neve JE, Deloukas P, Demuth I, Ding J, Eibich P, Eisele L, Eklund N, Evans DM, Faul JD, Feitosa MF, Forstner AJ, Gandin I, Gunnarsson B, Halldorsson BV, Harris TB, Heath AC, Hocking LJ, Holliday EG, Homuth G, Horan MA *et al*. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542.

Polderman TJ, Benyamin B, De Leeuw CA, Sullivan PF, Van Bochoven A, Visscher PM, Posthuma D (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics* **47**, 702–709.

Ripke S, O'dushlaine C, Chambert K, Moran JL, Kähler A, Akterin S, Bergen S, Collins AL, Crowley J, Fromer M, Kim Y, Lee SH, Magnusson PK, Sanchez N, Stahl EA, Williams S, Wray N, Xia K, Bettella F, Børglum AD, Bulik-Sullivan BK, Cormican P, Craddock N, De Leeuw C, Durmishi N, Gill M, Golimbet V, Hamshere ML, Holmans P, Hougaard DM, Kendler KS, Lin K, Morris DW, Mors O, Mortensen PB, Neale B, O'neill FA, Owen MJ, Pejovic Milovancevic M, Posthuma D, Powell J, Richards AL, Riley BP, Ruderfer D, Rujescu D, Sigurdsson E, Silagadze T, Smit AB, Stefansson H, Steinberg S, Suvisaari J, Tosato S, Walters JT, Verhage M, Multicenter Genetic Studies of Schizophrenia Consortium, Psychosis Endophenotypes Consortium, Wellcome Trust Case-Control Consortium2, Bramon E, Corvin AP, O'donovan MC, Stefansson K, Scolnick E, Purcell S, Mccarroll S, Sklar P, Hultman C, Sullivan PF (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genetics* **45**, 1150–1159.

Schizophrenia Psychiatric Genome-Wide Association Study Consortium (2011). Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics* **43**, 969–976.

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

Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M (2007). A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345.

Sullivan PF, Agrawal A, Bulik C, Andreassen OA, Borglum A, Breen G, Cichon S, Edenberg H, Faraone SV, Gelernter J, Mathews CA, Nievergelt CM, Smoller J, O'donovan M (2017). Psychiatric genomics: an update and an Agenda. *bioRxiv*.

Sullivan PF, Kendler KS, Neale MC (2003). Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Archives of General Psychiatry* **60**, 1187–1192.

Van Os J, Krabbendam L, Myin-Germeys I, Delespaul P (2005). The schizophrenia envirome. *Current Opinion in Psychiatry* **18**, 141–145.

Willer CJ, Li Y, Abecasis GR (2010). METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191.

Wray NR, Gottesman II (2012). Using summary data from the Danish national registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder. *Frontiers in Genetics* **3**, 118.

Wray NR, Yang J, Goddard ME, Visscher PM (2010). The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genetics* **6**, e1000864.

Yang J, Visscher PM, Wray NR (2010). Sporadic cases are the norm for complex disease. *European Journal of Human Genetics* **18**, 1039–1043.

Zheng J, Erzurumluoglu AM, Elsworth BL, Kemp JP, Howe L, Haycock PC, Hemani G, Tansey K, Laurin C, Early G, Lifecourse Epidemiology Eczema C, Pourcain BS, Warrington NM, Finucane HK, Price AL, Bulik-Sullivan BK, Anttila V, Paternoster L, Gaunt TR, Evans DM, Neale BM (2016). LD hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279.