Miscellaneous

New data and an old puzzle: the negative association between schizophrenia and rheumatoid arthritis

S Hong Lee,[1,2] Enda M Byrne,[1] Christina M Hultman,[3] Anna Kähler,[3] Anna AE Vinkhuyzen,[1] Stephan Ripke,[4,5,6] Ole A Andreassen,[7,8] Thomas Frisell,[9] Alexander Gusev,[10,11,12] Xinli Hu,[13,14,15,16] Robert Karlsson,[3] Vasilis X Mantzioris,[1] John J McGrath,[1,17] Divya Mehta,[1] Eli A Stahl,[18] Qiongyi Zhao,[1] Kenneth S Kendler,[19,20,21] Patrick F Sullivan,[22] Alkes L Price,[10,11,12] Michael O'Donovan,[23,24] Yukinori Okada,[25,26,27,28] Bryan J Mowry,[1,29] Soumya Raychaudhuri[25,26,27,30,31] and Naomi R Wray[1]*; Schizophrenia Working Group of the Psychiatric Genomics Consortium and Rheumatoid Arthritis Consortium International

[1]The University of Queensland, Queensland Brain Institute, Brisbane, QLD, Australia, [2]School of Environmental and Rural Science, University of New England, Armidale, NSW, Australia, [3]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, [4]Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA, [5]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA, [6]Department of Psychiatry and Psychotherapy, Charité-Universitätsmedizin Berlin, Campus Mitte, Berlin, Germany, [7]NORMENT KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Oslo, Norway, [8]Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway, [9]Clinical Epidemiology Unit, Karolinska Institutet, Stockholm, Sweden, [10]Department of Epidemiology, [11]Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA, [12]Medical and Population Genetics Program, Broad Institute, Cambridge, MA, USA, [13]Division of Genetics, [14]Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA, [15]Medical and Population Genetics Group, Broad Institute, Cambridge, MA, USA, [16]Health Science and Technology MD Program, Harvard University and Massachusetts Institute of Technology, Boston, MA, USA, [17]Queensland Centre for Mental Health Research, Park Centre for Mental Health, Richlands, QLD, Australia, [18]Division of Psychiatric Genomics, Mt Sinai School of Medicine, New York, NY, USA, [19]Virginia Institute of Psychiatric and Behavioral Genetics, [20]Department of Human and Molecular Genetics, [21]Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA, [22]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, [23]Medical Research Council (MRC) Centre for Neuropsychiatric Genetics and Genomics, [24]Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff, UK, [25]Division of Rheumatology, Immunology, and Allergy, [26]Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA, [27]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA, [28]Department of Human Genetics and Disease Diversity, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan, [29]Queensland Centre for Mental

Health Research, Wacol, QLD, Australia, [30]Partners Center for Personalized Genetic Medicine, Boston, MA, USA and [31]Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK

*Corresponding author. The University of Queensland, Queensland Brain Institute, St Lucia, Brisbane, QLD 4072, Australia. E-mail: naomi.wray@uq.edu.au

Abstract

Background: A long-standing epidemiological puzzle is the reduced rate of rheumatoid arthritis (RA) in those with schizophrenia (SZ) and vice versa. Traditional epidemiological approaches to determine if this negative association is underpinned by genetic factors would test for reduced rates of one disorder in relatives of the other, but sufficiently powered data sets are difficult to achieve. The genomics era presents an alternative paradigm for investigating the genetic relationship between two uncommon disorders.

Methods: We use genome-wide common single nucleotide polymorphism (SNP) data from independently collected SZ and RA case-control cohorts to estimate the SNP correlation between the disorders. We test a genotype X environment (GxE) hypothesis for SZ with environment defined as winter- vs summer-born.

Results: We estimate a small but significant negative SNP-genetic correlation between SZ and RA (-0.046, s.e. 0.026, $P = 0.036$). The negative correlation was stronger for the SNP set attributed to coding or regulatory regions (-0.174, s.e. 0.071, $P = 0.0075$). Our analyses led us to hypothesize a gene-environment interaction for SZ in the form of immune challenge. We used month of birth as a proxy for environmental immune challenge and estimated the genetic correlation between winter-born and non-winter born SZ to be significantly less than 1 for coding/regulatory region SNPs (0.56, s.e. 0.14, $P = 0.00090$).

Conclusions: Our results are consistent with epidemiological observations of a negative relationship between SZ and RA reflecting, at least in part, genetic factors. Results of the month of birth analysis are consistent with pleiotropic effects of genetic variants dependent on environmental context.

Key words: Schizophrenia, rheumatoid arthritis, genetic relationship, pleiotropy

Key Messages

- The proportion of variance in liability attributable to common SNPs is 0.223 (s.e. 0.006) for schizophrenia and 0.194 (s.e. 0.007) for rheumatoid arthritis.
- The MHC region is proportionally more important for rheumatoid arthritis than for schizophrenia; after excluding this region, the proportion of variance attributable to common SNPs is reduced to 0.212 (s.e. 0.006) for schizophrenia and 0.137 (s.e. 0.007) for rheumatoid arthritis, respectively.
- The epidemiological observations of a negative relationship between SZ and RA reflects, at least in part, genetic factors.

Introduction

A long-standing epidemiological puzzle is the reduced prevalence of rheumatoid arthritis (RA) in those with schizophrenia (SZ) and vice versa. First identified nearly 80 years ago, it has been reported in nearly every observational study of the incidence of the two disorders. A meta-analysis[1] of nine studies found that risk of RA in SZ subjects was less than 29% of the risk in the general population. Under-reporting of somatic disease in those with severe psychiatric conditions may be a contributing factor,[2] but the prevalence of RA is not reduced in those with other psychiatric disorders.[1] Surprisingly, the reduced prevalences are observed despite the high prevalence of smoking in SZ [odds ratio (OR) = 5.3],[3] which is an established risk

factor for RA in general population samples (OR > 2).[4] Both disorders have similar lifetime risk (~1%), a waxing and waning pattern of symptoms and increased mortality. On the other hand, there are also dissimilarities including age at onset (16–30 years in SZ vs 25–55 years in RA)[5] and male:female ratio (nearly 3 females to 1 male for RA[5] and 1 female to 1.4 males for SZ[6]). Sex differences were not considered in early studies of the SZ-RA relationship, but recent population-based studies from Sweden[2] and Denmark[7] accounted for age and sex differences and still reported reduced risks of RA in SZ compared with those without SZ.

Both SZ and RA have a strong genetic component to their aetiology. For both, heritabilities estimated from national hospital records (SZ 64%[8] and RA 40%[9]) are lower than estimates from twin studies (SZ 81%[10] and RA 60%[11]). RA has two distinct subtypes classified on the presence (seropositive) or absence (seronegative) of antibodies to citrullinated protein antigen. Approximately two-thirds[12] of cases are seropositive, and of seronegative cases 4–11% have been estimated to reflect misdiagnosis (for example of ankylosing spondylitis) and 15–37% have been estimated to be undetected seropositive.[13] Using national data from Sweden, the heritability of seropositive RA was estimated as ~50% compared with ~20% for seronegative RA.[9]
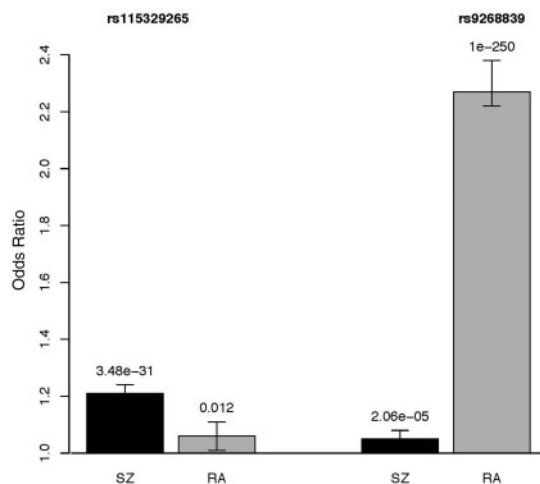
A number of hypotheses have been proposed to explain the SZ-RA protective relationship,[14] including abnormal tryptophan metabolism,[15] prostaglandin deficiency,[16–18] an imbalance in corticosteroids,[19] psychosocial factors[14] or consequence of medication.[1] Definitive evidence to support these hypotheses is lacking. There is evidence that factors influencing immune activation, including environmental insults such as infectious agents, are potential pathogenic mechanisms for both disorders.[5,20] For example, both RA and SZ have been linked, albeit with some controversy, to increased rates of infection by viruses such as Epstein-Barr virus and the parasite *Toxoplasma gondii* (for a review see[5]). SZ is considered to be a neurodevelopmental disorder, and immune activation in early life may be of particular importance, consistent with perinatal risk factors[21] such as infection and month of birth.[22] RA is an autoimmune disorder, and it has been suggested that there is an autoimmune component to SZ.[23,24] The autoimmune theory of SZ is supported by epidemiological evidence showing that whereas the relationship between SZ and RA is a negative one, the risk of many other autoimmune disorders is higher in SZ than in controls.[7,25,26] An analysis of Danish national records showed a dose-response relationship between risk of SZ and hospitalizations for infection and autoimmune disorders, where three or more infections and an autoimmune disease were associated with an

incidence rate ratio of 3.40 [95% confidence interval (CI) 2.91–3.94].[27] As for all autoimmune disorders,[28] the major histocompatibility complex (MHC) plays an important role in RA[29,30] but with different alleles being associated with seropositive cases compared with seronegative cases.[13] A role for the MHC in the aetiology of SZ has been proposed for decades,[31] but the empirical evidence has been less consistent than for RA. The first large genome-wide association studies (GWAS) for SZ identified the MHC locus as the most strongly associated locus.[32–34] Using the latest published GWAS results,[35,36] the MHC locus is the only locus that reaches genome-wide significance for both SZ and RA. The most associated single nucleotide polymorphism (SNP) for RA is associated with SZ and vice versa; contrary to expectation, given the negative SZ-RA association, the associated alleles are the same for the two disorders, albeit stronger for RA than SZ (Box 1). This unexpected result may reflect the well-recognised complexity of the MHC region.[37] The primary association for rheumatoid arthritis is within HLA-DRB1 in the class II MHC region, and has a large effect relative to the more modest and less clearly localized effect in schizophrenia. It may be the case that the role of HLA-mediated antigen recognition is simply different in the two diseases, playing a dominant role in rheumatoid arthritis and perhaps a more modest or absent role in schizophrenia. If the negative association between RA and SCZ reflects genetic factors, then it may be driven predominantly by non-HLA genetic factors that are related to immune activation rather than antigen recognition. Genome-wide significant variants explain 3.4% of the variance in liability to schizophrenia and 11.4% of the variance in liability of RA (Box 1). Genome-wide polygenic methods[38,39] have estimated that for SZ, ~23%[40] of the variance in liability is attributable to common SNPs (or SNP heritability), and 14%[41] to 18%[39] for RA excluding the contribution from the MHC region (~5%). These results imply that more associated loci will be identified for each disorder as sample size increases.

Given the substantial genetic contribution to both disorders, can the negative association between SZ and RA be attributed to genetic factors? This can be investigated from traditional epidemiological studies by measuring risk of RA in relatives of those with SZ compared with relatives of control subjects, and vice versa. As RA and SZ are relatively uncommon, very large cohorts of families with multiple family members measured for both disorders are needed, and this has not been achievable through the family study framework from which estimates of heritability are traditionally derived. National databases provide the only viable strategy to explore a genetic relationship through traditional epidemiological methods, but few countries have suitable national recording frameworks.

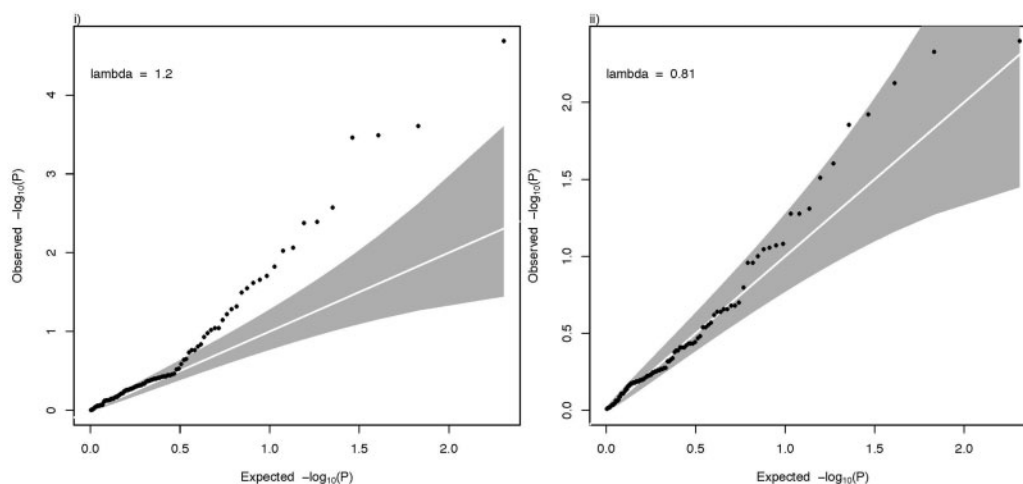**Box 1. Comparison of GWAS results of SZ[36] and RA.[35]**

a. MHC locus. Odds ratios for the two most highly associated SNPs for SZ, rs115329265 (aka rs1233578, hg19:chr 6: 28 712 247 bp), RAF = 0.85 and for RA, rs9268839 (aka rs116633882, hg19:chr 6: 32 428 772 bp) RAF = 0.45. Both are located in the MHC region and the LD r^2 of these SNPs is zero. We note that the RA allele tags the HLA DRB-1 allele, but that the SZ allele is not associated with any classical HLA allele, although it is reported[36] to be in LD r^2 = 0.32 with an eQTL SNP for HLA-A. The association *P*-values are listed above the error bars.
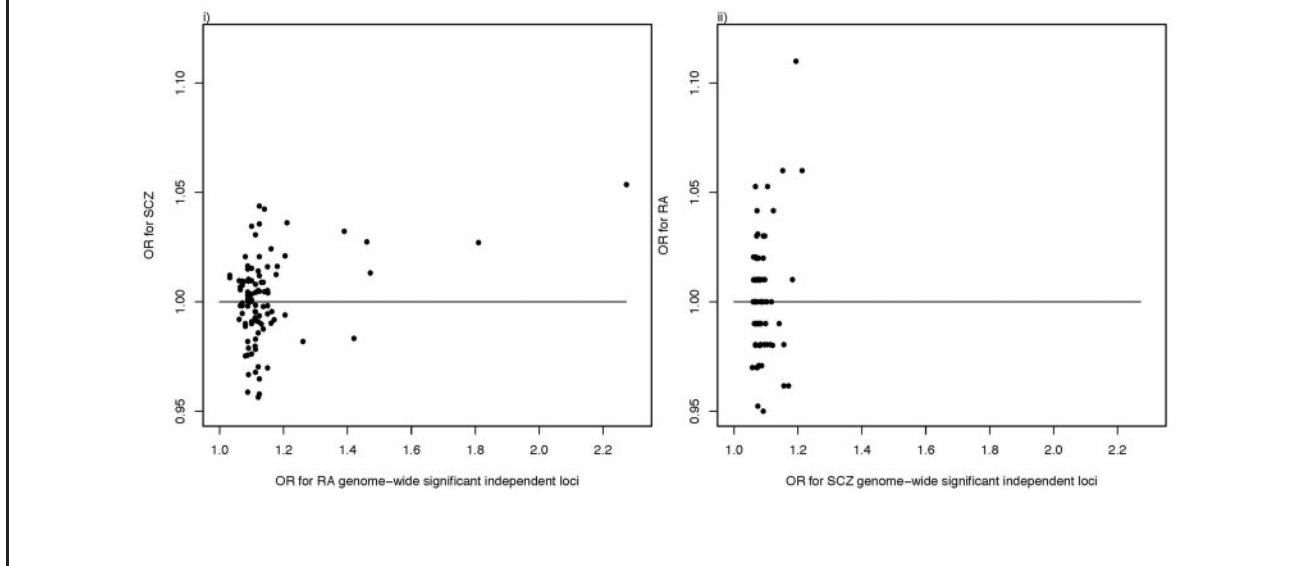

b. Variance explained by genome-wide significant (GWS) loci, reported as associated at $P < 5e\text{-}8$.

- For SZ 128, statistically independent GWS loci are reported.[36] Together these explain 3.4% of the variance in liability (calculated from reported RAF and OR using INDI-V[71] assuming lifetime risk of 1%); all SNPs associated with $P < 0.05$ explain 7% of variance in out of sample prediction.[36] Of these 128 SNPs, 102 could be matched to GWAS results for RA. The minimum *P*-value in RA of a SZ GWS locus was 0.004.
- For RA, 101 independent GWS loci are reported.[35] Together these explain 11.4% of the variance in liability (using INDI-V[71] assuming lifetime risk of 0.7%). All SNPs could be matched to GWAS results for SZ. The minimum *P*-value in SZ of an RA GWS locus was 2.06e-05 (see MHC locus above).

c. QQ plots: (i)- *P*-values from SZ GWAS for 101 RA GWS loci; (ii) *P*-values from RA GWAS for 102 SZ GWS loci.

d. OR plots: (i) RA GWS loci (correlation 0.35, $P = 4.0e\text{-}4$, becomes non-significant if SNPs with the two largest OR are excluded); (ii) SZ GWS loci (correlation 0.24, $P = 0.015$, which becomes non-significant if the MHC SNP is excluded).

A Danish national study[25] compared 7704 persons in Denmark diagnosed with SZ and their parents with a sample of matched subjects and their parents. Contrary to a hypothesis of a negative genetic association, this study found significantly increased rates of RA in parents of those with SZ compared with parents of control subjects. A Swedish national study reported that first-degree relatives of schizophrenia patients were not at reduced risk of RA, but the risk for seronegative RA was significantly decreased in children and siblings of SZ probands [hazard ratio (HR) = 0.13; 95% CI 0.02–0.95, and HR = 0.67; 95% CI 0.50–0.91, respectively].[2] These studies assumed that risk in first-degree relatives is only attributable to genetic factors, but sharing of environmental factors could also contribute.

The genomics era provides a new opportunity to investigate whether the SZ-RA relationship may be attributable to genetic factors, by determining whether common alleles conferring increased risk to SZ are protective against RA and vice versa. The hypothesis has been considered for candidate genes outside the MHC region, but no shared associations were found.[42,43] Comparison of the latest GWAS results for SZ[36] and RA[35] shows evidence for more association of the genome-wide significant (GWS) loci from each disease in the other disease than expected by chance (Box 1). Across GWS SNPs, there is a positive correlation between the OR of the two disorders (Box 1), although this relationship is dominated by the positive correlation between SNPs in the MHC region described above.

Here we use linear mixed model methods, applied to genome-wide SNPs from case-control cohorts collected for GWAS, to explore the relationship between SZ and RA.[44]

By comparing additive genetic similarities between SZ and RA cases with their genetic similarities with controls, we quantify the relationship between the disorders by the SNP-genetic correlation.[44,45] Since this approach uses unrelated cases and controls to estimate the SNP correlation, estimates are less likely to be confounded with shared environmental factors that can bias estimates from family studies or population studies where individuals are measured for both phenotypes. We use this framework to explore the genetic relationship between SZ and RA.

Methods

Data

Three RA and two SZ GWAS data sets (see Supplementary Table 1, available as Supplementary data at *IJE* online) were made available to us. Briefly, the Stahl *et al.* ('Stahl') RA sample comprises 5441 seropositive cases and 22 532 controls of European ancestry from six independent case-control cohorts.[39,46] The Okada *et al.* ('Okada') RA sample has 3427 cases (1840 seropositive) and 6837 controls of European descent from five independent case-control cohorts,[35] including the Corrona RA cohort. The Epidemiological Investigation of Rheumatoid Arthritis (EIRA) sample comprises 770 seronegative RA cases (EIRA seropositive cases from this cohort were already included in the Stahl sample). The Psychiatric Genomics Consortium (PGC) for Schizophrenia Wave 1 sample comprises data from 17 GWAS cohorts[47] and a total of 9431 cases and 12 848 controls; and the Swedish (SWE) sample
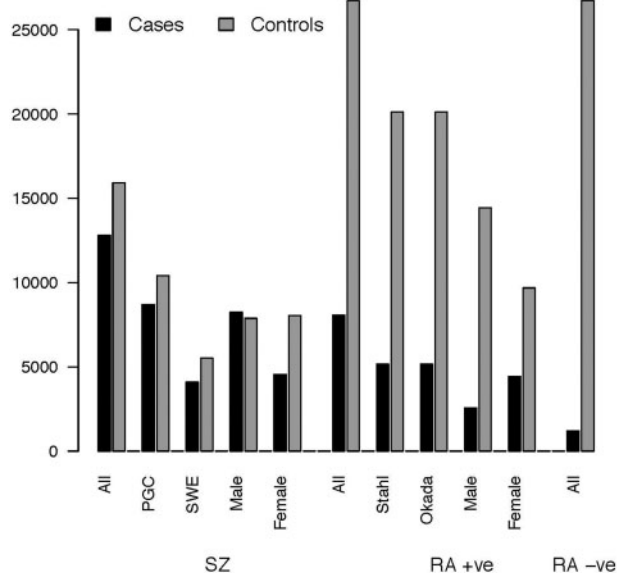
Figure 1. Sample sizes.

comprises 5193 case and 6391 controls[48] in four cohorts defined by genotyping platform which were independent of Swedish samples in PGC. All sample sizes are those after sample quality control (QC).

All data sets were processed through similar QC and imputation pipelines[47] using the CEU + TSI Hapmap Phase 3 data as the reference panel. We augmented these QC so that estimates of genetic variance would not be influenced by artefacts of genotyping.[38] SNPs with an imputation $r^2 > 0.6$ and an MAF > 0.01 in all cohorts were retained, resulting in 797 875 SNPs for analysis. Sex chromosome data were not available for all data sets and so were excluded. If any pair of individuals had an estimated similarity relationship coefficient[49] > |0.05|, one person was excluded at random so that all SZ cases, RA cases and controls were unrelated. The final analysis data set consisted of 8064 seropositive RA cases (including 1131 RA cases of unknown status of which at least two-thirds[12,13] are expected to be seropositive), 1197 seronegative RA cases and 26 737 controls plus and 12 793 SZ cases and 15 912 controls (Figure 1). Given accumulating evidence that seropositive and seronegative RA should be regarded as different clinical entities,[13] we did not combine seropositive and seronegative RA cases.

SNP heritabilityand SNP genetic correlation

The bivariate linear mixed model genomic relationship matrix (GRM) restricted maximum likelihood (GREML) approach[44] implemented in GCTA[50] was used to estimate SNP heritabilties, the SNP coheritability and the SNP correlation between the disorders. The standard error (s.e.) of each estimate was calculated by the delta method[51] which has been shown to agree well with s.e. expected from normal distribution theory.[52] We used the estimate and its s.e. to generate a Wald statistic to test hypotheses that SNP heritabilities were different from zero and that SNP correlations were less than zero; the directional hypothesis for the correlation is justified by the epidemiological data reported in the Methods section. The model of analysis estimates SNP heritabilities as the proportion of variance in case-control status attributable to genome-wide SNPs, but estimates of SNP heritabilites and coheritabilities are presented on the liability scale,[38,44] assuming population lifetime risk of 1% for SZ, 0.7% for seropositive RA and 0.3% for seronegative RA, so that they can be compared with estimates from epidemiological data. We note that when genetic relationships between individuals are small, the relationship between disease and liability scale is approximately linear and so the estimated genetic correlation is independent of scale.[44] A SNP correlation of zero is estimated if the genome-wide relationship between cases of one disorder is the same with the cases as with the controls of another disorder. A SNP correlation reflects the magnitude of the covariance term between the traits relative to the product of the standard deviations, and so can be high even when the covariance is low. A genome-wide SNP correlation could represent a uniform correlation across the genome or a weighted average of higher and lower correlations. Hence, we undertook genomic partitioning analyses which included multiple additive genetic random effects terms in the linear mixed model with multiple GRM constructed from non-overlapping SNP sets.[45,49] Cohort and the first 20 principal components were included as covariates in all analyses. Sex was included as a covariate for SZ and in some analyses for RA. SNP heritabilities are presented on the liability scale. Follow-up analyses were conducted by sex and considering functional annotation.

Sensitivity analysis

We explored the sensitivity of our results and sought to exclude the possibility that genetic outliers could explain our results. We tested this by restricting the coefficient of similarity between any pair of individuals to be < |0.025|. As one individual from a pair was excluded at random, we constructed 20 randomly drawn samples with restricted ancestry, and drew 20 random samples of the same size from the sample with coefficient of similarity < |0.05|.

Benchmarking with epidemiological observations

From epidemiological studies we can obtain estimates of the population risk of SZ and RA, K_{SZ} and $K_{RA,}$

respectively, and also for the probability of RA in those with SZ, $K_{RA|SZ}$. We assume that the phenotypic liabilities of SZ (l_{SZ}) and RA (l_{RA}) are distributed as bivariate normal with mean 0, standard deviation 1 and correlation R_{SZ-RA}:

$$V \begin{pmatrix} l_{SZ} \\ l_{RA} \end{pmatrix} = \begin{bmatrix} V(l_{SZ}) & Cov(l_{SZ}, l_{RA}) \\ Cov(l_{SZ}, l_{RA}) & V(l_{RA}) \end{bmatrix}$$

$$= \begin{bmatrix} 1 & R_{SZ-RA} \\ R_{SZ-RA} & 1 \end{bmatrix}$$

The variances and covariance of liabilities among those affected with SZ

$$V \begin{pmatrix} l_{SZ|D_{SZ}} \\ l_{RA|D_{SZ}} \end{pmatrix} = = \begin{bmatrix} 1-k_{SZ} & R_{SZ-RA}(1-k_{SZ}) \\ R_{SZ-RA}(1-k_{SZ}) & 1-R_{SZ-RA}^2 k_{SZ} \end{bmatrix}$$

where $k_{sz} = i_{sz}(i_{sz} - t_{sz})$ and reflects the proportional variance reduction as a consequence of ascertainment on SZ status,[50,51] with $mean(l_{SZ|D_{SZ}}) = i_{SZ} = \frac{d_{SZ}}{K_{SZ}}$, with d_{sz} the height of the normal curve at the threshold t_{sz} defined from $P(z > t_{SZ}) = K_{SZ}$. From $K_{RA|SZ}$ we define the normal distribution threshold[53,54] for RA in those with SZ as $t_{RA|SZ} = \frac{t_{RA} - i_{SZ}R_{SZ-RA}}{\sqrt{1 - k_{sz}R_{SZ-RA}^2}}$

Solving the quadratic for R_{SZ-RA} gives

$$R_{SZ-RA} = \frac{i_{SZ}t_{RA} - \sqrt{i_{SZ}^2 t_{RA}^2 - (t_{RA|SZ}^2 + i_{SZ}^2)(t_{RA}^2 - t_{RA|SZ}^2)}}{(t_{RA|SZ}^2 + i_{SZ}^2)}$$

(equation 1)

Genotype x environment analysis

Our analyses led us to a postulate a hypothesis of genotype x environment interaction for SZ. Specifically we hypothesized that the SNP correlation would be less than 1 for coding and regulatory SNPs in a bivariate analysis in which the two traits are winter-born and non-winter born SZ cases and controls. We undertook a bivariate GREML analysis to test this hypothesis. Month of birth was only available for the SWE SZ sample, which comprised winter-born cases (born January to April[22], $n = 1511$) and winter-born controls ($n = 2036$), as well as non-winter cases ($n = 2962$) vs non-winter controls ($n = 3772$); 199 individuals did not have month of birth recorded. To visualize the interaction, we identified 47318 SNPs associated with schizophrenia[36] at $P < 0.05$ (the threshold that maximized out of sample prediction across multiple cohorts;[36] the SNPs were quasi-independent with minor allele frequency > 0.05, pairwise linkage disequilibrium $r^2 < 0.25$ in a 250-kb window). We identified the risk alleles of the SNPs that defined the odds ratio to be greater than 1. We undertook association analysis
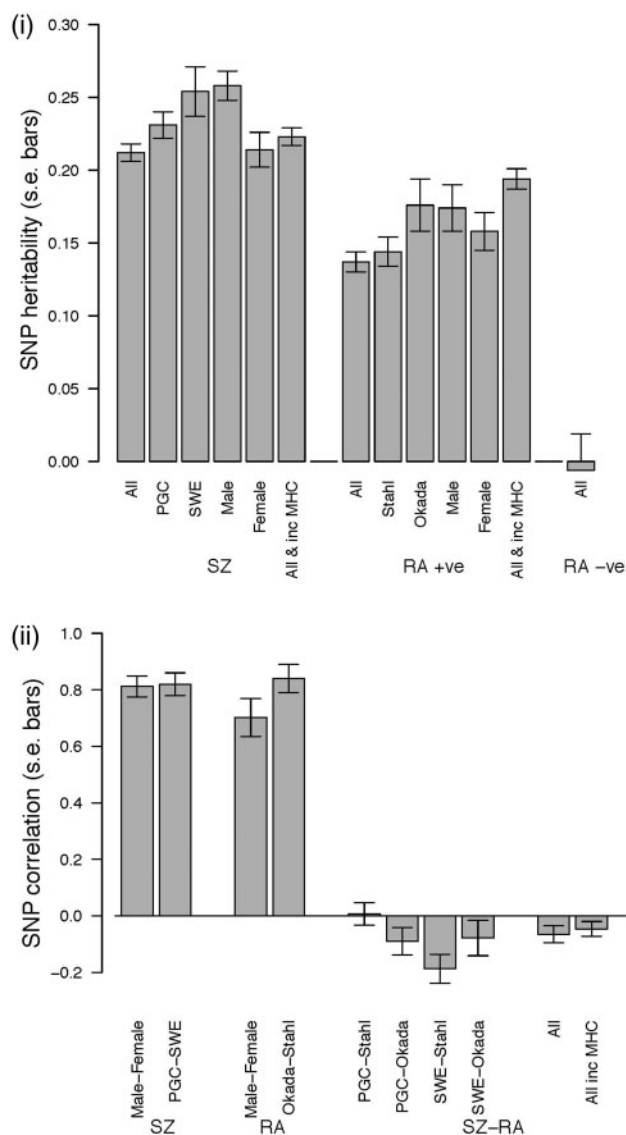
(logistic regression with 20 principal components, cohort and sex as covariates) in the Swedish sample that had season of birth recorded. We estimated the OR of the risk alleles and compared mean OR for SNPs annotated as C&R (coding/regulatory, the genomic region showing strong negative SNP correlation between SZ and RA, 2820 SNPs, 6%) and not coding/regulatory, testing the hypothesis H$_0$: Mean OR for winter-born sample = Mean OR for other sample.

Results

SNP heritability and SNP genetic correlation

The estimated SNP heritability for RA seronegative cases was indistinguishable from zero (−0.006, s.e. 0.025, $P = 0.98$). Despite the smaller sample size for seronegative cases, the s.e. shows that it was powered to detect SNP heritability > 5%. Given there was no evidence of contribution to risk of common variants for this sample, detection of a genetic relationship between these cases and SZ was not possible. Hence, all reported analyses are for seropositive RA cases only. Given the major contribution of the MHC region in RA (which may violate underlying assumptions of GREML[55,56]), we undertook analyses using as the phenotype residuals after adjusting for the 550 SNPs (the number after pruning for SNPs with linkage disequilibrium $r^2 > 0.99$) located within the MHC region (29–34Mb in chromosome 6) and for the other covariates. For SZ, the estimated SNP heritabilities were 0.223 (s.e. 0.006) including the MHC region and 0.212 (s.e. 0.006) after correcting for the MHC region. For RA, the estimated SNP heritabilities were 0.194 (s.e. 0.007) including the MHC region and 0.137 (s.e. 0.007), after correcting for the MHC region. The estimated SNP genetic correlations were −0.046 (s.e. 0.026) and −0.065 (s.e. 0.030) for including the MHC and after correcting for the MHC, respectively, which were significantly less than zero (*P*-values = 0.036 and 0.015, respectively) (Figure 2; Supplementary Table 2, available as Supplementary data at *IJE* online). We confirmed that the method to estimate *P*-values was robust, by checking the *P*-value from a likelihood ratio test comparing models with and without genome-wide SNP effects.

To aid interpretation and comparisons, we also present the SNP genetic covariance or coheritability, the latter represents the relationship between the disorders on the same scale as the heritability (Supplementary Table 2). Subset analyses (PGC-SZ/Stahl-RA, PGC-SZ/Okada-RA, SWE-SZ/Stahl-RA and SWE-SZ/Okada-RA) showed negative genetic correlations estimated for all combinations except PGC-SZ/Stahl-RA (Figure 2, Supplementary Table 2). We explored the sensitivity of our results and sought to exclude the possibility that genetic outliers could explain our results

Figure 2. Genome-wide estimates: (i) SNP heritability; (ii) SNP correlation (using RA + ve only). Estimates are from analyses using as the phenotype residuals after adjusting SNPs in the MHC region unless otherwise stated.

(see Methods). We found that the SNP correlation between SZ and RA was significantly ($P = 0.0049$) more negative in 20 samples drawn from our data when ancestry was more restricted (similarity relatedness coefficient $< |0.025|$) $[-0.054$, standard deviation over replicates (s.d.) 0.002 vs -0.047 s.d. 0.001] (Supplementary Figure 1). This sensitivity analysis implies that the negative correlation is not driven by ancestry artefacts and provides confidence that SNP correlation between SZ and RA is negative.
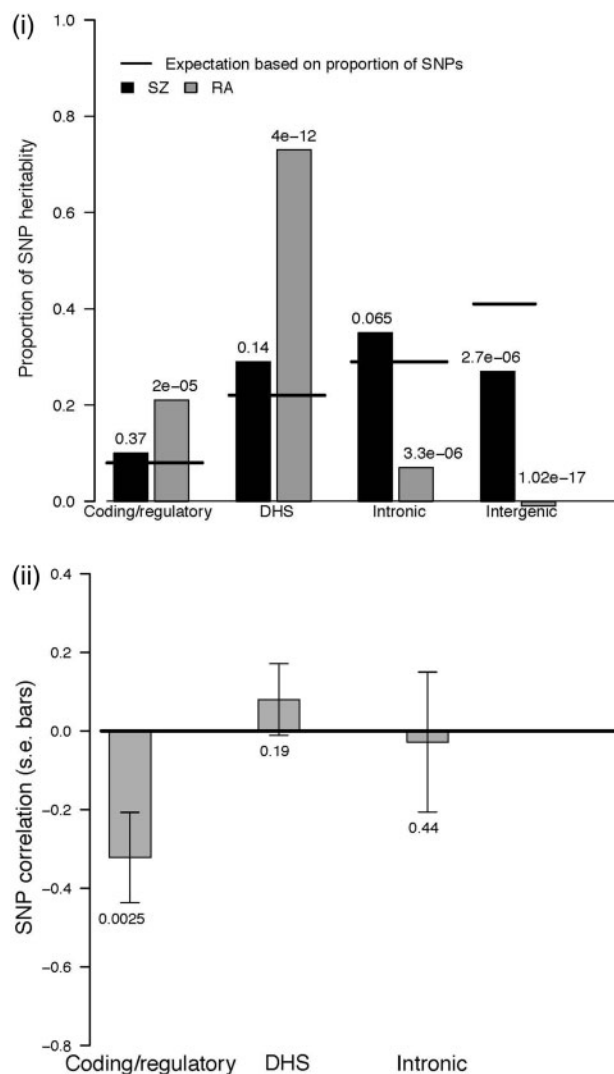
Sex analyses

Given that risk of RA is higher in females and risk of SZ is higher in males, we undertook SZ/RA analyses stratified by sex (i.e. four-trait multivariate GREML, in which the four traits were SZ-male, SZ-female, RA-male, RA-female each matched by their sex-specific control set) to determine if SNP correlations (based on autosomal SNPs) were sex dependent. Sex information was missing for $\sim11\%$ of the RA sample who were excluded in the analysis. Based on reported male:female population ratios,[5,6] we assumed the male and female baseline risks were 0.42% and 0.98% for seropositive RA, and 1.15% and 0.85% for SZ respectively. SNP heritabilities were significantly greater when estimated from males compared with females for SZ (male 0.258, s.e. 0.010, female 0.214, s.e. 0.012, $P = 0.0053$) but not for RA (male 0.174, s.e. 0.016, female 0.158, s.e. 0.013, $P = 0.43$) (Figure 2, Supplementary Table 2); these estimates must be interpreted with caution, recognizing that they are dependent on the lifetime risk of disease chosen for each sex. SNP correlations between sexes were high but were significantly different from 1 for both RA ($P = 6.1$e-06) and SZ ($P = 2.4$e-07, Supplementary Table 2). All SNP correlation point estimates are negative between male/female SZ/RA analyses.

Functional annotation analyses

Previous studies[40,41,57] have demonstrated that contributions to SNP heritabilities are not distributed equally over the genome. We therefore set out to test if the SNP-correlation between SZ and RA was dependent on SNP annotation. We undertook genomic partitioning analyses in which multiple additive genetic random effects terms were considered in the linear mixed model,[45,49] with multiple GRM each constructed from SNPs grouped by a functional annotation. Following Gusev *et al.*,[57] SNPs were classified as being in coding/regulatory (in exons, 3' UTR, 5'UTR, 1-kb region up- and downstream of transcription start and end site and noncoding RNA), DNase I hypersensitivity sites (DHS) and intronic or intergenic regions. SNPs with multiple annotations were allocated with hierarchical preference of coding/regulatory, over DHS and over intronic. For SZ, all annotations had estimates of SNP heritability that were significantly greater than zero (Figure 3; Supplementary Table 3, available as Supplementary data at *IJE* online), although the proportion of total variance allocated to intergenic SNPs was significantly less than expected given the proportion of all SNPs annotated to that group (Figure 3; Supplementary Table 3). For RA, the coding/regulatory and DHS annotations had SNP heritability estimates that were both significantly different from zero and were higher than expected based on the proportion of SNPs in those functional partitions (Figure 3; Supplementary Table 3). As a consequence in RA, and in contrast to SZ, neither intronic nor intergenic variants

Figure 3. Genomic partitioning analyses; (i) Percentage of SNP heritability attributed to each functional annotation class for SZ and RA, compared with the percentage of SNPs attributed to each class. Adjacent to the bars are the *P*-values for H0: percentage of variation attributed to annotation class = percentage of SNPs attributed to the annotation class; (ii) SNP correlation between SZ and RA based on annotation. Estimates are from analyses using as the phenotype residuals after adjusting SNPs in the MHC region. When the MHC is included, the coding/regulatory correlation is −0.174, s.e. 0.071, *P* = 0.0075.

made a significant contribution to SNP heritability, a finding that is underscored by the fact that the proportion of all SNP heritability contributed by these types of variant is significantly less than expected based upon the proportion of SNPs of these classes. Given that in RA, SNP heritability is essentially restricted to SNPs in coding/regulatory and DHS regions, these are the only classes of SNP for which meaningful estimates of SNP correlation between SZ and RA can be made. We note that currently available DHS annotation may be biased towards cell types of relevance to RA compared with SZ. We estimated a significant and stronger genetic correlation for coding/regulatory region

regions (−0.322, s.e. 0.115, *P* = 0.003) (Figure 3; Supplementary Table 3) than for the whole genome.

Immune related pathway analyses

Since RA is an autoimmune disease and since the epidemiological negative relationship between SZ and RA has contributed to the autoimmune hypothesis of SZ, we set out to test if the SZ-RA SNP correlation is more negative in SNPs in immune-related pathways. To avoid multiple testing, we selected a single immune gene set based on previous work using the Stahl GWAS results of Hu *et al.*[58] Using gene expression from 223 murine immune cell types, they reported over-representation of RA-associated SNPs in genes expressed specifically in CD4+ effector memory T cells, with strongest over-representation in genes expressed in the subcutaneous lymph node subset named T.4Mem44h62l.LN.[58] We selected the top 4000 genes expressed in the T.4Mem44h62l.LN cells to be in the 'T4Mem' set. The arbitrary threshold of 4000 genes generated an SNP set of about 10% of the total SNPs analysed. We tested if the T4Mem genes make an enriched contribution to SNP heritabilities and the SNP correlation, partitioning the coding/regulatory, DHS and intronic partitions into T4Mem and non-T4Mem classes (Supplementary Table 4, available as Supplementary data at *IJE* online). As expected, there was a significant enrichment for variance attributable to the T4Mem class for RA (28% of the SNP heritability compared with only 11.5% of SNPs, 2.5 fold enrichment, *P* = 2.0e-11). Interestingly, there was also significant enrichment of variance attributable to T4Mem regulatory SNPs in SZ (16% of SNP heritability, *P* = 2.4e-04). We did not find evidence that SNPs in this group of genes were more negatively correlated than those in the rest of the genome, although the patterns of correlations are difficult to interpret, and the size of the standard errors means that the sample is underpowered (Supplementary Table 4).

Benchmarking with epidemiological observations

We can benchmark our estimated SNP correlation between RA and SZ of −0.046 relative to expectation from epidemiological data. Using the meta-analysis result that the risk of RA in those with SZ is 29% of the risk in the general population,[1] we estimate that this implies a phenotypic correlation between the disorders of −0.15 (equation 1; $K_{SZ} = 0.01$ and $K_{RA} = 0.01$, respectively, $K_{RA|SZ} = 0.29*K_{RA}$). More recent epidemiological studies[2,7] imply a substantially smaller negative phenotypic correlation (∼−0.05). Phenotypic correlation is considered a reasonable benchmark for genetic correlation,[59] and

therefore our small negative estimate of the genetic correlation is consistent with the epidemiological data. However, the genetic relationship between SZ and RA is complex. As an autoimmune disease, the MHC region contains risk factors for RA which alone explain ∼5% of the variance in liability to RA, whereas the most significant individual SNP association for SZ, also in the MHC region, explains only ∼0.1% of variance in liability (ref[60] and consistent with Table 1 with and without MHC region). These MHC risk alleles are positively correlated in SZ and RA, although clearly the effect sizes are very different. In analyses in which we first removed the contribution to variance of the MHC region, the magnitude of negative correlation between SZ and RA increased (−0.065, s.e. 0.015, $P = 0.015$), and increased further still when considering only SNPs in coding and regulatory regions of the genome (−0.322, s.e. 0.115, $P = 0.003$), indicating that regions of the genome other than the MHC region contribute to the epidemiological observations.

Genotype x environment analysis

Although RA is well recognized as an immune disorder, the contribution of immune activation to SZ is open to debate (see Introduction). To add further insight to the complex relationship between SZ and RA, we postulated that if there is any interplay in risk for SZ between environmental risk factors associated with immune challenge and genes that are relevant to the immune response, it is likely to occur at the coding and regulatory regions. This is based on the rationale that these are the sets of variants that both capture the SNP heritability of RA (i.e. are likely to be most enriched for genes with influences on immune activation) and are negatively correlated with SZ (i.e. are both immune activation and SZ relevant). If our hypothesis is correct, we predict that the apparent effect sizes in schizophrenia at these loci will be greater in those exposed to a relevant immune challenge than in those who are not. Consequently, we predict that the negative correlation between SZ and RA will be larger at these loci in cases exposed to an immune challenge. At present, our ability to test this G x E hypothesis is limited by availability of samples that are both genetically informative and recorded for environmental risks. We therefore sought a proxy for immune challenge in SZ.
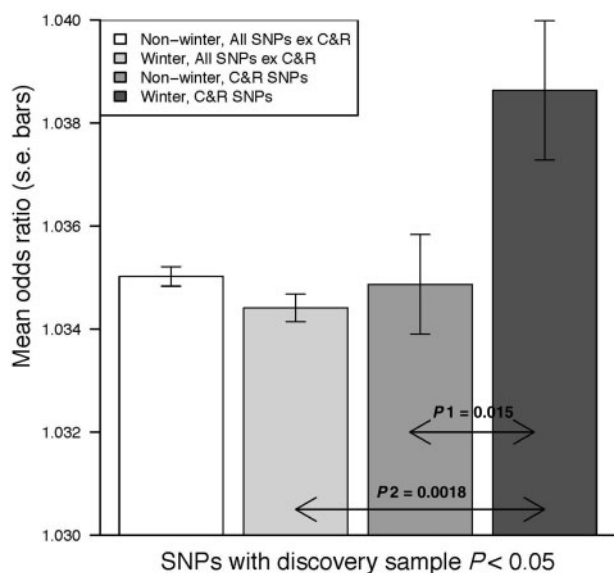
A robust epidemiological finding in SZ research is that people with the disorder are more likely to be born in winter or spring than summer or autumn[61] (odds ratio 1.07, 95% CI 1.05–1.08, estimated from a meta-analysis of 27 studies[62]). Winter/spring birth is also associated with recognized immune-mediated disorders including RA.[63] Candidate exposures underlying this finding include

seasonally varying factors such as prenatal vitamin D, or maternal/fetal exposure to infections. Both factors can impact on the immune system.[64–66] In the absence then of direct measures of immune activation, we used season of birth as our proxy measure; in doing so, we are aware this proxy measure is only likely to be weakly correlated with exposure and hence its use will adversely affect power.

Month of birth was only available for the SWE SZ sample. We selected quasi-independent SNPs associated at $P < 0.05$ and minor allele frequency > 0.05 (47 318 SNPs) from the largest published schizophrenia meta-analysis.[36] We next divided them into coding/regulatory (2820 SNPs, 6%) and non coding/regulatory sets. We undertook a bivariate GREML analysis of the SWE datasets in which the two traits were winter-born (January to April) vs non-winter born SZ cases and controls; this division was justified by studies using Swedish data.[22] The correlation between winter/non-winter born was significant for coding/regulatory SNPs (0.56, s.e. 0.14, $P = 0.0009$) but not so for other SNPs (0.95, s.e. 0.05, $P = 0.15$) (Supplementary Table 5, available as Supplementary data at *IJE* online). To visualize this interaction, and to demonstrate that effects sizes of the coding/regulatory SNP set are increased in the winter-born cohort, we present the mean association OR for the different season by annotation classes estimated in the SWE SZ sample (Figure 4). We confirmed that results were robust to the *P*-value threshold used for selection of associated SNPs (Supplementary Table 6, available as Supplementary data at *IJE* online) and we checked the seasonal trend by using sliding window definitions of 4-month season definitions (Supplementary Table 7, available as Supplementary data at *IJE* online). The results are consistent with our G x E hypothesis of immune-related disruption proxied by winter birth, in concert with risk variants in coding/regulatory regions increasing risk of schizophrenia.

Strengths and limitations

The strength of our methodological approach to explore the genetic relationship between SZ and RA is that it is based on genome-wide genotype data and uses independently collected data for the two diseases studied. While under review, two studies have investigated the relationship between SZ and RA based partially[67] or fully[68] on GWAS summary statistics; one reported a non-significant small positive relationship[67] and the other a non-significant small negative relationship.[68] Use of genotype data is computationally more demanding but is considered more definitive than methods based on summary statistics,[57] and such an approach is needed when the relationship between the disorders is benchmarked by a weak phenotypic correlation of −0.05 to −0.15. The results here indicate that there is a subset of

Figure 4. Relationship between season of birth and association odds ratios of SNPs associated with SZ at $P < 0.05$ identified in the GWAS of the Psychiatric Genomics Consortium.

$P1$ H_0: OR Winter C&R SNPs = OR for Non-Winter C&R SNPs. H_A: OR Winter C&R SNPs > OR for Non-Winter C&R SNPs. $P2$ H_0: OR Winter C&R SNPs = OR for Winter All SNPs except C&R. H_A: OR Winter C&R SNPs > OR for Winter All SNPs except C&R. C&R = Coding and Regulatory (2820 SNPs). All SNPs except C&R = All SNPs except Coding and Regulatory (44498 SNPs). Winter-born cases ($N = 1511$) and controls ($N = 2036$) were born January to April; $N = 2962$ for non-winter cases and $N = 3772$ for non-winter controls.

genetic variants that are risk alleles for SZ and protective for RA, and vice versa. Furthermore, as sample sizes increase, more fine-scaled partitioning of variance by annotation and genomic region will become possible and specific variants involved will be identified. Our analyses indicate that these pleiotropic loci are located in the coding/regulatory regions of the genome, hinting at directly functional variants. Pleiotropy could help explain why common risk variants are maintained in the population, since selection operates on the net effect of fitness.

As limitations, we first note that cases and controls for each disorder were not screened for the other disorder. However, since both disorders are uncommon, the impact is expected to be small. Second, we were unable to investigate the relationship between SZ and seronegative RA because we found no evidence of SNP heritability for this case group. An analysis of seronegative RA is of particular interest since a recent epidemiological study found that decreased risk of RA in relatives of those with SZ was only apparent for seronegative RA.[2] The stringent QC needed for case-control GREML analyses may remove true signals[38] and this may play a role in the seronegative samples available here. Nonetheless, a lower SNP heritability for seronegative RA vs seropositive RA is consistent with recent

estimates of heritability based on family data,[9] and also consistent with smaller OR for seronegative genetic associations than for seropositive associations.[13,69] A SNP correlation of 0.98, s.e. 0.165, was estimated in a Han Chinese sample;[69] we note that the proportion of seronegative cases in this study was high (519/952 = 55%) and any misclassification could serve to inflate the correlation.[70] Third, the non-availability of sex chromosome data meant that we could only explore sex differences based on autosomal SNPs. Fourth, we did not have the power to break down the signal attributed to coding/regulatory regions into more finely defined functional categories. Last, given the limitations on the data available to test the G x E hypothesis, alternative explanations for our results cannot be excluded.

Discussion

In summary, we have applied a mixed linear model method to estimate the genetic correlation between RA and SZ. Epidemiological evidence has demonstrated decreased prevalence rates of RA in SZ cases, consistent with a phenotypic correlation of liabilities of up to -0.15. We show that there is a small but significant negative correlation across the genome and the signal is stronger for SNPs annotated as coding and regulatory. Given that RA is an immune-related disorder and that a role for immune activation has long been hypothesized for SZ, a negative genetic correlation could imply that variants in immune response pathways have different roles in different tissues and/or in response to different challenges. The immune activation hypothesis of SZ is partly founded on an increased risk for SZ associated with month of birth. Our hypothesis that increased effect sizes for SZ-associated SNPs in the coding/regulatory SNP set for a winter-born case-control set was supported by our analyses, although other explanations for these results may be possible. Most importantly, if the complexity of SZ is to be unraveled, then data sets that are informative for both genetic and environmental risk factors are essential. Since SZ is an adult-onset disorder, and yet perinatal and childhood experience, especially infections, are known environmental risk factors, then prospective gathering of data in nationally accessible repositories is needed.

Acknowledgements

Author list continued:

Schizophrenia Working Group of the Psychiatric Genomics Consortium authors: William Byerley,[1,2] Wiepke Cahn,[3] Rita M Cantor,[4] Sven Cichon,[5,6,7] Paul Cormican,[8] David Curtis,[9] Srdjan Djurovic,[10] Valentina Escott-Price,[11,12] Pablo V Gejman,[13] Lyudmila Georgieva,[11,14] Ina Giegling,[15] Thomas F Hansen,[16,17] Andrés Ingason,[16,17] Yunjung Kim,[18] Bettina Konte,[15] Phil H Lee,[19] Andrew McIntosh,[20,21] Andrew McQuillin,[22] Derek W Morris,[23] Markus M Nöthen,[5,24] Colm O'Dushlaine,[25] Ann Olincy,[26] Line Olsen,[16,17] Carlos N Pato,[27] Michele T Pato,[27] Benjamin S Pickard,[20,21] Danielle Posthuma,[28] Henrik B Rasmussen,[16,17] Marcella Rietschel,[29] Dan Rujescu,[15] Thomas G Schulze,[30,31] Jeremy M Silverman,[32] Srinivasa Thirumalai[33] and Thomas Werge.[16,17,34]

Affiliations: [1]Department of Psychiatry, University of California, San Francisco, CA, USA, [2]NCIRE (Northern California Institute of Q Research and Education), San Francisco, CA, USA [3]Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center, Utrecht, The Netherlands, [4]David Geffen School of Medicine, University of California, Los Angeles, CA, USA, [5]Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany, [6]Division of Medical Genetics, Department of Biomedicine, University of Basel, Basel, Switzerland, [7]Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, Juelich, Germany, [8]Department of Psychiatry, Trinity College Dublin, Dublin, Ireland, [9]UCL Genetics Institute, University College London, London, UK, [10]NORMENT, KG Jebsen Centre for Psychosis Research, Department of Clinical Science, University of Bergen, Bergen and Department of Medical Genetics, Oslo University Hospital, Oslo, Norway, [11]Medical Research Council (MRC) Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Cardiff, UK, [12]Biostatistics and Bioinformatics Unit, Cardiff University, Cardiff, UK, [13]Department of Psychiatry and Behavioral Sciences, North Shore University Health System and University of Chicago, Evanston, IL, USA, [14]Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff, UK, [15]Department of Psychiatry, University of Halle, Halle, Germany, [16]Institute of Biological Psychiatry, Copenhagen University Hospital, Roskilde, Denmark.,[17]Lundbeck Initiative for Integrative Psychiatric Research, iPSYCH, Roskilde, Denmark, [18]Department of Genetics, University of North Carolina at Chapel Hill, NC, USA, [19]Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA, USA, [20]Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK, [21]Molecular Medicine Centre, University of Edinburgh, Edinburgh, UK, [22]Mental Health Sciences Unit, University College London, London, UK, [23]Cognitive Genetics and Therapy Group, Discipline of Biochemistry and School of Psychology, National University of Ireland, Galway, Ireland, [24]Institute of Human Genetics, University of Bonn, Bonn, Germany, [25]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA, [26]Department of Psychiatry, University of Colorado Denver, Aurora, CO, USA, [27]Department of Psychiatry, Zilkha Neurogenetic Institute, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA, [28]Department of Complex Trait Genetics, VU University, Amsterdam, The Netherlands, [29]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany, [30]Department of Psychiatry & Behavioral Sciences, Johns Hopkins University, Baltimore, MD, USA, [31]Department of Psychiatry & Psychotherapy, University of Göttingen, Göttingen, Germany, [32]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA, [33]Oxford Health NHS Foundation Trust, Marlborough House Secure Unit, Milton Keynes, UK and [34]Faculty of Health and Medical Science, University of Copenhagen, Copenhagen, Denmark.

Schizophrenia Working Group of the Psychiatric Genomics Consortium collaborators: Ingrid Agartz,[1,2] Farooq Amin,[3] Maria H Azevedo,[4] Nicholas Bass,[5] Donald W Black,[6] Douglas H R Blackwood,[7] Richard Bruggeman,[8] Nancy G Buccola,[9] Khalid Choudhury,[5] C Robert Cloninger,[10] Aiden Corvin,[11] Nicholas Craddock,[12,13] Mark J Daly,[14,15] Susmita Datta,[16] Gary J Donohoe,[17] Jubao Duan,[18] Frank Dudbridge,[19] Ayman Fanous,[20,21] Robert Freedman,[22] Nelson B Freimer,[23] Marion Friedl,[24] Michael Gill,[11] Hugh Gurling,[5] Lieuwe De Haan,[25] Marian L Hamshere,[12,26] Annette M Hartmann,[24] Peter A Holmans,[12,26] René S Kahn,[27] Matthew C Keller,[28] Elaine Kenny,[11] George K Kirov,[12,13] Lydia Krabbendam,[29] Robert Krasucki,[5] Jacob Lawrence,[5] Todd Lencz,[30,31,32] Douglas F Levinson,[33] Jeffrey A Lieberman,[34] Dan-Yu Lin,[35] Don H Linszen,[36] Patrik KE Magnusson,[37] Wolfgang Maier,[38] Anil K Malhotra,[30,31,32] Manuel Mattheisen,[39,40,41,42] Morten Mattingsdal,[1,43] Steven A McCarroll,[15] Helena Medeiros,[44] Ingrid Melle,[1,45] Vihra Milanova,[46] Inez Myin-Germeys,[29] Benjamin M Neale,[14,15] Roel A Ophoff,[23,47,48] Michael J Owen,[12,13] Jonathan Pimm,[5] Shaun M Purcell,[14,15,49] Vinay Puri,[5] Digby J Quested,[50] Lizzy Rossin,[15] Douglas Ruderfer,[49] Alan R Sanders,[18] Jianxin Shi,[51] Pamela Sklar,[49] David St. Clair,[52] T Scott Stroup,[53] Jim Van Os,[29] Peter M Visscher,[54,55] Durk Wiersma[8] and Stanley Zammit.[12,13]

Affiliations: [1]KG Jebsen Centre for Psychosis Research, Institute of Clinical Medicine, University of Oslo, Oslo, Norway, [2]Department

of Research, Diakonhjemmet Hospital, Oslo, Norway, [3]Department of Psychiatry and Behavioral Sciences, Atlanta Veterans Affairs Medical Center, Emory University, Atlanta, GA, USA, [4]Faculty of Medicine, University of Coimbra, Coimbra, Portugal, [5]Mental Health Sciences Unit, University College London, London, UK, [6]Department of Psychiatry, University of Iowa, Iowa City, IA, USA, [7]Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK. [8]Department of Psychiatry, University Medical Center Groningen, Groningen, The Netherlands, [9]School of Nursing, Louisiana State University Health Sciences Center, New Orleans, LA, USA, [10]Department of Psychiatry, Washington University School of Medicine, St Louis, MO, USA, [11]Department of Psychiatry, Trinity College Dublin, Dublin, Ireland, [12]Medical Research Council (MRC) Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Cardiff, UK, [13]Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff, UK, [14]Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA, [15]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA, [16]Genetics Institute, University College London, London, UK, [17]Cognitive Genetics and Therapy Group, Discipline of Biochemistry and School of Psychology, National University of Ireland Galway, Ireland, [18]Department of Psychiatry and Behavioral Sciences, North Shore University Health System and University of Chicago, Evanston, IL, USA, [19]Department of Non-Communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, UK, [20]Department of Psychiatry, Georgetown University School of Medicine, Washington, DC, USA, [21]Virginia Institute of Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA, [22]Department of Psychiatry, University of Colorado Denver, Aurora, CO, USA, [23]Center for Neurobehavioral Genetics, University of California, Los Angeles, CA, USA, [24]Department of Psychiatry, University of Halle, Halle, Germany, [25]Department of Psychiatry, Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands, [26]Biostatistics and Bioinformatics Unit, Cardiff University, Cardiff, UK, [27]Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center, Utrecht, The Netherlands, [28]Department of Psychology, University of Colorado, Boulder, CO, USA, [29]Department of Psychiatry and Neuropsychology, Maastricht University Medical Centre, South Limburg Mental Health Research and Teaching Network, Maastricht, The Netherlands, [30]Department of Psychiatry, Long Island Jewish Health System, Glen Oaks, New York, NY, USA, [31]Center for Psychiatric Neuroscience, Feinstein Institute of Medical Research, Manhasset, New York, NY, USA, [32]Department of Psychiatry and Behavioral Science, Albert Einstein College of Medicine of Yeshiva University, Bronx, NY, USA, [33]Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA, [34]New York State Psychiatric Institute, Columbia University, New York, NY, USA, [35]Department of Biostatistics, University of North Carolina at Chapel Hill, NC, USA, [36]Department of Psychiatry, Academic Medical Centre University of Amsterdam, Amsterdam, The Netherlands, [37]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, [38]Department of Psychiatry, [39]Institute of Human Genetics, University of Bonn, Bonn, Germany, [40]Lundbeck Initiative for Integrative Psychiatric Research, iPSYCH, Roskilde, Denmark, [41]Department of Biomedicine, Aarhus University, Aarhus, Denmark, [42]Department of Genomic Mathematics, University of Bonn, Bonn, Germany, [43]Sørlandet Hospital, Kristiansand, Norway, [44]Department of Psychiatry, Zilkha Neurogenetic Institute, University of Southern California, Los Angeles, CA, USA, [45]Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway, [46]Department of Psychiatry, Alexander University Hospital, Sofia, Bulgaria, [47]Department of Human Genetics, University of California, Los Angeles, CA, USA, [48]Department of Psychiatry, University Medical Center Utrecht, Utrecht, The Netherlands, [49]Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA, [50]Academic Department of Psychiatry, University of Oxford, Oxford, UK, [51]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA, [52]Institute of Medical Sciences, University of Aberdeen, Aberdeen, UK, [53]Department of Psychiatry, Columbia University, New York, NY, USA, [54]Queensland Brain Institute, [55]Diamantina Institute, University of Queensland, Brisbane, QLD, Australia.

Rheumatoid Arthritis Consortium International authors: S Louis Bridges Jr,[1] Hyon K Choi,[2,3,4] Marieke JH Coenen,[5] Niek de Vries,[6] Philippe Dieudé,[7,8] Jeffrey D Greenberg,[9] Tom WJ Huizinga,[10] Leonid Padyukov,[11] Katherine A Siminovitch,[12,13,14] Paul P Tak[15,16,17] and Jane Worthington.[18,19]

Affiliations: [1]Division of Clinical Immunology and Rheumatology, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA, [2]Channing Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA, [3]Section of Rheumatology, [4]Clinical Epidemiology Research and Training Unit, Boston University School of Medicine, Boston, MA, USA, [5]Department of Human Genetics, Radboud University Medical Centre, Nijmegen, The Netherlands, [6]Amsterdam Rheumatology and Immunology Center, Amsterdam Medical Center, Amsterdam, The Netherlands, [7]Service de Rhumatologie Hôpital Bichat Claude Bernard, Assistance Publique des Hôpitaux de Paris, Paris, France, [8]Rheumatology Department, Université Paris Diderot, Paris, France, [9]New York University Hospital for Joint Diseases, New York, NY, USA, [10]Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands, [11]Rheumatology Unit, Karolinska Institutet, Stockholm, Sweden, [12]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada, [13]Toronto General Research Institute, Toronto, ON, Canada, [14]Department of Medicine, University of Toronto, Toronto, ON, Canada, [15]Academic Medical Centre, University of Amsterdam, Amsterdam, The Netherlands, [16]School of Clinical Medicine, University of Cambridge, Cambridge, UK, [17]Immunoinflammation Therapy Unit, GlaxoSmithKline, Stevenage, UK, [18]Arthritis Research UK Epidemiology Unit, Centre for Musculoskeletal Research, University of Manchester, Manchester, UK and [19]Musculoskeletal Biomedical Research Unit, Manchester Academic Health Sciences Centre, UK.

Rheumatoid Arthritis Consortium International collaborators: Philip L De Jager,[1,2,3] Joshua C Denny,[4,5] Peter K Gregersen,[6] Lars Klareskog,[7] Xavier Mariette,[8] Robert M Plenge,[1,2,9] Mart van Laar[10] and Piet van Riel.[11]

Affiliations: [1]Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA, [2]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA,

[3]Program in Translational Neuropsychiatric Genomics, Department of Neurology, Brigham and Women's Hospital, Boston, MA, USA, [4]Department of Biomedical Informatics, [5]Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA, [6]Feinstein Institute for Medical Research, North Shore–Long Island Jewish Health System, Manhasset, New York, NY, USA, [7]Rheumatology Unit, Karolinska Institutet, Stockholm, Sweden, [8]Institut National de la Santé et de la Recherche Médicale, Rhumatologie, Hôpitaux Universitaires Paris-Sud, Paris, France, [9]Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA, [10]Department of Rheumatology and Clinical Immunology, University Twente & Medisch Spectrum Twente, Enschede, The Netherlands and [11]Scientific Institute for Quality of Healthcare, Radboud University Medical Centre, Nijmegen, The Netherlands.

References

1. Oken RJ, Schulzer M. At issue: schizophrenia and rheumatoid arthritis: the negative association revisited. *Schizophr Bull* 1999;**25**:625–38.

2. Sellgren C, Frisell T, Lichtenstein P, Landen M, Askling J. The association between schizophrenia and rheumatoid arthritis: a nationwide population-based Swedish study on intraindividual and familial risks. *Schizophr Bull* 2014;**40**:1552–59.

3. de Leon J, Diaz FJ. A meta-analysis of worldwide studies demonstrates an association between schizophrenia and tobacco smoking behaviors. *Schizophr Res* 2005;**76**:135–57.

4. Bang SY, Lee KH, Cho SK, Lee HS, Lee KW, Bae SC. Smoking increases rheumatoid arthritis susceptibility in individuals carrying the HLA-DRB1 shared epitope, regardless of rheumatoid factor or anti-cyclic citrullinated peptide antibody status. *Arthritis Rheum* 2010;**62**:369–77.

5. Torrey EF, Yolken RH. The schizophrenia-rheumatoid arthritis connection: infectious, immune, or both? *Brain Behav Immun* 2001;**15**:401–10.

6. Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. *PLoS Med* 2005;**2**:e141.

7. Benros ME, Pedersen MG, Rasmussen H, Eaton WW, Nordentoft M, Mortensen PB. A nationwide study on the risk of autoimmune diseases in individuals with a personal or a family history of schizophrenia and related psychosis. *Am J Psychiatry* 2014;**171**:218–26.

8. Lichtenstein P, Yip BH, Bjork C et al. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet* 2009;**373**:234–39.

9. Frisell T, Holmqvist M, Kallberg H, Klareskog L, Alfredsson L, Askling J. Familial risks and heritability of rheumatoid arthritis: role of rheumatoid factor/anti-citrullinated protein antibody status, number and type of affected relatives, sex, and age. *Arthritis Rheum* 2013;**65**:2773–82.

10. 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.

11. MacGregor AJ, Snieder H, Rigby AS et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum* 2000;**43**:30–37.

12. de Vries RR, van der Woude D, Houwing JJ, Toes RE. Genetics of ACPA-positive rheumatoid arthritis: the beginning of the end? *Ann Rheum Dis* 2011;**70**(Suppl 1):i51–54.

13. Han B, Diogo D, Eyre S et al. Fine mapping seronegative and seropositive rheumatoid arthritis to shared and distinct HLA alleles by adjusting for the effects of heterogeneity. *Am J Hum Genet* 2014;**94**:522–32.

14. Eaton WW, Hayward C, Ram R. Schizophrenia and rheumatoid arthritis: a review. *Schizophr Res* 1992;**6**:181–92.

15. Taylor WM. Schizophrenia, rheumatoid arthritis and trytophan metabolism. *J Clin Psychiatry* 1978;**39**:499–503.

16. Horrobin DF. Schizophrenia as a prostaglandin deficiency disease. *Lancet* 1977;**1**:936–37.

17. Horrobin DF, Ally AI, Karmali RA, Karmazyn M, Manku MS, Morgan RO. Prostaglandins and schizophrenia: further discussion of the evidence. *Psychol Med* 1978;**8**:43–48.

18. Horrobin DF. Prostaglandins and schizophrenia. *Lancet* 1979;**1**:1031–32.

19. Trevathan RD, Tatum JC. Rarity of concurrence of psychosis and rheumatoid arthritis in individual patients; report of a case. *J Nerv Ment Dis* 1954;**120**:83–84.

20. Blomstrom A, Karlsson H, Svensson A et al. Hospital admission with infection during childhood and risk for psychotic illness - a population-based cohort study. *Schizophr Bull* 2014;**40**:1518–25.

21. Brown AS. Epidemiologic studies of exposure to prenatal infection and risk of schizophrenia and autism. *Dev Neurobiol* 2012;**72**:1272–76.

22. Hultman CM, Sparen P, Takei N, Murray RM, Cnattingius S. Prenatal and perinatal risk factors for schizophrenia, affective psychosis, and reactive psychosis of early onset: case-control study. *BMJ* 1999;**318**:421–26.

23. Strous RD, Shoenfeld Y. Schizophrenia, autoimmunity and immune system dysregulation: a comprehensive model updated and revisited. *J Autoimmun* 2006;**27**:71–80.

24. Jones AL, Mowry BJ, Pender MP, Greer JM. Immune dysregulation and self-reactivity in schizophrenia: do some cases of schizophrenia have an autoimmune basis? *Immunol Cell Biol* 2005;**83**:9–17.

25. Eaton WW, Byrne M, Ewald H et al. Association of schizophrenia and autoimmune diseases: linkage of Danish national registers. *Am J Psychiatry* 2006;**163**:521–28.

26. Chen SJ, Chao YL, Chen CY et al. Prevalence of autoimmune diseases in in-patients with schizophrenia: nationwide population-based study. *Br J Psychiatry* 2012;**200**:374–80.

27. Benros ME, Nielsen PR, Nordentoft M, Eaton WW, Dalton SO, Mortensen PB. Autoimmune diseases and severe infections as risk factors for schizophrenia: a 30-year population-based register study. *Am J Psychiatry* 2011;**168**:1303–10.

28. Fernando MM, Stevens CR, Walsh EC et al. Defining the role of the MHC in autoimmunity: a review and pooled analysis. *PLoS Genet* 2008;**4**:e1000024.

29. Stastny P. Association of the B-cell alloantigen DRw4 with rheumatoid arthritis. *N Engl J Med* 1978;**298**:869–71.

30. Newton JL, Harney SM, Wordsworth BP, Brown MA. A review of the MHC genetics of rheumatoid arthritis. *Genes Immun* 2004;**5**:151–57.

31. Wright P, Nimgaonkar VL, Donaldson PT, Murray RM. Schizophrenia and HLA: a review. *Schizophr Res* 2001;**47**:1–12.

32. Purcell SM, Wray NR, Stone JL *et al*. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009;**460**:748–52.

33. Stefansson H, Ophoff RA, Steinberg S *et al*. Common variants conferring risk of schizophrenia. *Nature* 2009;**460**:744–47.

34. Shi J, Levinson DF, Duan J *et al*. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* 2009;**460**:753–57.

35. Okada Y, Wu D, Trynka G *et al*. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 2014;**506**:376–81.

36. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Ripke S, Neale BM *et al*. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014;**511**:421–27.

37. de Bakker PI, Raychaudhuri S. Interrogating the major histocompatibility complex with high-throughput genomics. *Hum Mol Genet* 2012;**21**:R29–36.

38. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* 2011;**88**:294–305.

39. Stahl EA, Wegmann D, Trynka G *et al*. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat Genet* 2012;**44**:483–89.

40. Lee SH, DeCandia TR, Ripke S *et al*. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet* 2012;**44**:247–50.

41. Gusev A, Bhatia G, Zaitlen N *et al*. Quantifying missing heritability at known GWAS loci. *PLoS Genet* 2013;**9**:e1003993.

42. de la Fontaine L, Schwarz MJ, Riedel M *et al*. Investigating disease susceptibility and the negative correlation of schizophrenia and rheumatoid arthritis focusing on MIF and CD14 gene polymorphisms. *Psychiatr Res* 2006;**144**:39–47.

43. Watanabe Y, Nunokawa A, Kaneko N *et al*. Two-stage case-control association study of polymorphisms in rheumatoid arthritis susceptibility genes with schizophrenia. *J Hum Genet* 2009;**54**:62–65.

44. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* 2012;**28**:2540–42.

45. Lee SH, Ripke S, Neale BM *et al*. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet* 2013;**45**:984–94.

46. Stahl EA, Raychaudhuri S, Remmers EF *et al*. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet* 2010;**42**:508–14.

47. Ripke S, Sanders AR, Kendler KS *et al*. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* 2011;**43**:969–76.

48. Ripke S, O'Dushlaine C, Chambert K *et al*. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet* 2013;**45**:1150–59.

49. Yang J, Manolio TA, Pasquale LR *et al*. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet* 2011;**43**:519–25.

50. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011;**88**:76–82.

51. Lynch M, Walsh B. *Genetics and Analysis of Quantitative Traits*. Sunerland, MA: Sinauer Associates, 1998.

52. Visscher PM, Hemani G, Vinkhuyzen AAE *et al*. Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet* 2014;**10**:e1004269.

53. Falconer D. The inheritance of liability to certain diseases, estimates from the incidence among relatives. *Ann Hum Genet* 1965;**29**:51–76.

54. Reich T, Morris CA, James JW. Use of multiple thresholds in determining mode of transmission of semi-continuous traits. *Ann Hum Genet* 1972;**36**:163–84.

55. Speed D, Hemani G, Johnson MR, Balding DJ. Improved heritability estimation from genome-wide SNPs. *Am J Hum Genet* 2012;**91**:1011–21.

56. Lee SH, Yang J, Chen GB *et al*. Estimation of SNP heritability from dense genotype data. *Am J Hum Genet* 2013;**93**:1151–55.

57. Gusev A, Lee SH, Trynka G *et al*. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am J Hum Genet* 2014;**95**:535–52.

58. Hu X, Kim H, Stahl E, Plenge R, Daly M, Raychaudhuri S. Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. *Am J Hum Genet* 2011;**89**:496–506.

59. Cheverud JM. A comparison of genetic and phenotypic correlations. *Evolution* 1988;**42**:958–68.

60. Visscher PM, Goddard ME, Derks EM, Wray NR. Evidence-based psychiatric genetics, AKA the false dichotomy between common and rare variant hypotheses. *Mol Psychiatr* 2012;**17**:474–85.

61. Torrey EF, Miller J, Rawlings R, Yolken RH. Seasonality of births in schizophrenia and bipolar disorder: a review of the literature. *Schizophr Res* 1997;**28**:1–38.

62. Davies G, Welham J, Chant D, Torrey EF, McGrath J. A systematic review and meta-analysis of Northern Hemisphere season of birth studies in schizophrenia. *Schizophr Bull* 2003;**29**:587–93.

63. Disanto G, Chaplin G, Morahan JM *et al*. Month of birth, vitamin D and risk of immune-mediated disease: a case control study. *BMC Med* 2012;**10**:69.

64. Chun RF, Liu PT, Modlin RL, Adams JS, Hewison M. Impact of vitamin D on immune function: lessons learned from genome-wide analysis. *Front Physiol* 2014;**5**:151.

65. Yang CY, Leung PS, Adamopoulos IE, Gershwin ME. The implication of vitamin D and autoimmunity: a comprehensive review. *Clin Rev Allergy Immunol* 2013;**45**:217–26.

66. Mora JR, Iwata M, von Andrian UH. Vitamin effects on the immune system: vitamins A and D take centre stage. *Nat Rev Immunol* 2008;**8**:685–98.

67. Euesden J, Breen G, Farmer A, McGuffin P, Lewis CM. The relationship between schizophrenia and rheumatoid arthritis revisited: genetic and epidemiological analyses. *Am J Med Genet B Neuropsychiatr Genet* 2015;**168**:8188.

68. Bulik-Sullivan B, Finucane H, Antilla V *et al*. An atlas of genetic correlations across human diseases and traits. *BioRXiv*. http://dx.doi.org/10.1101/014498.

69. Jiang L, Yin J, Ye L *et al*. Novel risk loci for rheumatoid arthritis in Han Chinese and congruence with risk variants in Europeans. *Arthritis Rheum* 2014;**66**:1121–32.

70. Wray NR, Lee SH, Kendler KS. Impact of diagnostic misclassification on estimation of genetic correlations using genome-wide genotypes. *Eur J Hum Genet* 2012;**20**:668–74.

71. Witte JS, Visscher PM, Wray NR. The contribution of genetic variants to disease depends on the ruler. *Nat Rev Genet* 2014;**15**:765–76.