# Genome-wide association study of psychosis proneness in the Finnish population

<table>
<thead>
<tr>
<th>Journal:</th>
<th><em>Schizophrenia Bulletin</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>SZBLTN-ART-16-0312.R2</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Regular Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>02-Jan-2017</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Ortega-Alonso, Alfredo; Institute for Molecular Medicine Finland; National Institute for Health and Welfare, Department of Health Ekelund, Jesper; National Institute for Health and Welfare, Department of Health; University of Helsinki, Department of Psychiatry; Vaasa Hospital District Sarin, Antti-Pekka; Institute for Molecular Medicine Finland; National Institute for Health and Welfare, Department of Health; University of Helsinki, Department of Public Health Miettunen, Jouko; University of Oulu, Center for Life Course Health Research; Medical Research Center Oulu, Oulu University Hospital and University of Oulu Veijola, Juha; Medical Research Center Oulu, Oulu University Hospital and University of Oulu; University of Oulu, Department of Psychiatry, Research Unit of Clinical Neuroscience; Oulu University Hospital, Department of Psychiatry Jarvelin, Marjo-Riita; University of Oulu, Center for Life Course Health Research; Imperial College London, Department of Epidemiology and Biostatistics, MRC–PHE Centre for Environment and Health, School of Public Health; University of Oulu, Biocenter Oulu; Oulu University Hospital, Unit of Primary Care Hennah, William; Institute for Molecular Medicine Finland; National Institute for Health and Welfare, Department of Health</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Genome-wide association study, Heritability, Finnish population, Psychoses proneness, Schizophrenia, Bipolar Disorder</td>
</tr>
</tbody>
</table>
Genome-wide association study of psychosis proneness in the Finnish population

Alfredo Ortega-Alonso, PhD 1,2; Jesper Ekelund, MD, PhD 2,3,4; Antti-Pekka Sarin, MSc 1,2,5; Jouko Miettunen, PhD 6,7; Juha Veijola, MD, PhD 7,8,9; Marjo-Riitta Jarvelin, MD, PhD 6,10,11,12; William Hennah, PhD 1,2

1 Institute for Molecular Medicine Finland-FIMM, Finland; 2 Department of Health, National Institute for Health and Welfare, Finland; 3 Department of Psychiatry, University of Helsinki, Finland; 4 Vaasa Hospital District, Finland; 5 Department of Public Health, University of Helsinki, Finland; 6 Center for Life Course Health Research, University of Oulu, Finland; 7 Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Finland; 8 Department of Psychiatry, Research Unit of Clinical Neuroscience, University of Oulu, Finland; 9 Department of Psychiatry, Oulu University Hospital, Finland; 10 Department of Epidemiology and Biostatistics, MRC–PHE Centre for Environment & Health, School of Public Health, Imperial College London, UK; 11 Biocenter Oulu, University of Oulu, Finland; 12 Unit of Primary Care, Oulu University Hospital, Finland

Corresponding Author: Alfredo Ortega-Alonso, Institute for Molecular Medicine Finland-FIMM, PO Box 20, FI-00014 University of Helsinki, Finland; Email: alfredo.ortega-alonso@helsinki.fi

Key words: Genome-wide association study; heritability; Finnish population; psychoses proneness; schizophrenia; bipolar disorder

Short Title: Genomic Approaches to Psychosis Proneness

Words in abstract: 246

Words in text: 4291

References: 71

Tables: 3

Figures: 1

Supplementary Text: 1

Supplementary Tables: 2

Supplementary Figures: 5
Abstract

The current study examined quantitative measures of psychosis proneness in a non-psychotic population, in order to elucidate their underlying genetic architecture and to observe if there is any commonality to that already detected in the studies of individuals with overt psychotic conditions, such as schizophrenia and bipolar disorder.

Heritability, univariate and multivariate genome-wide association tests, including a series of comprehensive gene-based association analyses, were developed in 4269 non-psychotic persons participating in the Northern Finland Birth Cohort 1966 study with information on the following psychometric measures: Hypomanic Personality, Perceptual Aberration, Physical and Social Anhedonia (a.k.a. Chapman’s Schizotypia scales), and Schizoidia scale. Genome-wide genetic data was available for ~9.84 million SNPs.

Heritability estimates ranged from 16% to 27%. Phenotypic, genetic and environmental correlations ranged from 0.04-0.43, 0.25-0.73, and 0.12-0.43, respectively. Univariate GWAS tests revealed an intronic SNP (rs12449097) at the TMC7 gene (16p12.3) that significantly associated ($p=3.485 \times 10^{-8}$) with the hypomanic scale. Bivariate GWAs tests including the hypomanic and physical anhedonia scales suggested a further borderline significant SNP (rs188320715; $p$-value=$5.261 \times 10^{-8}$, ~572kb downstream the ARID1B gene at 6q25.3).

Gene-based tests highlighted 20 additional genes of which 5 had previously been associated to schizophrenia and/or bipolar disorder: CSMD1, CCDC141, SLC1A2, CACNA1C and SNAP25. Altogether the findings explained from 3.7% to 14.1% of the corresponding trait heritability.

In conclusion, this study provides preliminary genomic evidence suggesting that qualitatively similar biological factors may underlie different psychosis proneness measures, some of which could further predispose to schizophrenia and bipolar disorder.
Introduction

Psychosis is a serious mental condition by which affected persons experience delusions, hallucinations and severe behavioral abnormalities usually leading to losing contact with reality. While acute forms of psychosis characterize disorders such as schizophrenia, schizoaffective disorder or psychotic bipolar disorder, subclinical types are also present in the general population in the form of odd behaviors, social withdrawal/anxiety, lack of feelings, perceptual abnormalities and magical ideation. These psychotic-like experiences, also termed “schizotypy”, are regarded as signs of an underlying predisposition (psychosis proneness) to undergo a clinically meaningful psychotic episode. Evidence from population data on this regard is steadily accumulating, and as a consequence there is a current consensus amongst clinicians and researchers on the idea that psychosis proneness is a continuum ranging from a normal dissociative to full-blown diagnosable primary psychotic disorders.

This continuum understanding of psychosis implies that individual differences exist in people’s vulnerability due to the combined effect of their personal genetic background and certain environmental stressors, and only those most susceptible would be pushed over a disease threshold. It is then possible that the genetic bases of these psychosis-like feelings and experiences (the lower and central portion of the continuum) are to a certain extent common to those for schizophrenia and other disorders with psychotic components (the upper end of the continuum, beyond the disease threshold). Consequently, unveiling the genes mediating psychosis proneness in non-psychotic individuals could help to understand part of the genetic architecture of disorders with psychotic components, and in future to accurately identify individuals at putative risk prior to any clinical manifestation.

To date, large meta-analyses of multinational case-control samples on psychotic disorders have successfully identified numerous common polymorphisms with moderately robust effects, mainly on individuals diagnosed with schizophrenia. However significant, many of these common polymorphisms showed very limited predictive power. This implies the need for complementary strategies capable of unveiling less common and population specific genetic polymorphisms of importance. Studying psychosis proneness in the general
population could make this feasible due to two main reasons. First, it allows incorporating standard psychiatric epidemiological data that provides information on carefully defined, quantifiable phenotypes and symptoms well below the disease threshold. Second, these data are normally gathered from well-characterized nation-wide registers and cohorts, which can take better advantage of population specific haplotype blocks and subtle linkage disequilibrium patterns to isolate less common, population specific variants with unanticipated roles that may well pass undetected in meta-analyses of multinational case-control samples.

Current understanding of the genetics underlying psychosis proneness in the general population derives principally from traditional quantitative genetic methods, applied to phenotypic data alone, and obtained from close relatives such as twins or parents-offspring. These have offered heritability estimates ranging from 15 to 65% depending upon symptom evaluated or population age. The most complete study thus far investigated phenotypic data on self-rated paranoia, hallucinations, cognitive disorganization, grandiosity, and anhedonia, as well as parent-rated negative symptoms, from 5059 adolescent twin pairs from England and Wales. Consistent with findings from smaller studies, their results showed genetic influences ranging between 15%-59%, and genetic correlations between 0.27-0.63. A relatively similar study in 3685 Australian young-adult twins focusing on Perceptual Aberration, Magical Ideation, Hypomania, Impulsivity, and Physical and Social Anhedonia, provided comparable genetic influences ranging from 24 to 50%.

Importantly, recent methodological and experimental studies have suggested that heritability estimates obtained through traditional quantitative genetic methods would represent an upper bound estimate of the overall genetic effects on the trait, and that applying current association methods to the available genomic data may well never reveal a significant part of those overall effects. This in turn imposes the need of finding distinctive, more conservative heritability estimates that can provide a sensible description of the extent to which psychoses proneness is affected by polymorphisms susceptible to be detected through the analyses of population genomic data. This type of lower level estimate can nowadays be obtained by applying modern statistical methods that rely on the joint analysis of phenotypic and genomic information from large samples of genetically unrelated individuals. At present, only the study
by Sieradzka at al.\textsuperscript{22} has considered this methodological approach, which resulted in heritability estimates of 20\% for anhedonia, 19\% for cognitive disorganization, 17\% for grandiosity and 14\% for paranoia, and an estimate of 0 for hallucinations.

Given the relevance of the genetic effects on psychosis proneness in the normal population, and the current availability of genome-wide association (GWAs) data in psychiatric research, looking for specific genes and markers affecting these traits is a fundamental and sensible action. However, to date only a few studies have provided suggestive evidence for genetic loci of potential interest, such as the TCF4, COMT and DISC1 genes\textsuperscript{23-26}. Therefore, we aim to elucidate the genetic architecture of psychosis proneness within a Finnish population-based cohort, maximizing the phenotypic and genomic information available in order to:

- compute an estimate of the heritability of psychoses proneness traits;
- disentangle whether shared genetic influences may account for the observed relation between different psychoses proneness scales; and
- examine genomic regions potentially harboring genes affecting one or more of the traits.
Methods

Participants

This study used data collected as part of the larger “Northern Finland Birth Cohort 1966” (NFBC66), a population based longitudinal cohort study comprising 12058 persons (~96.3% of all possible) with an expected date of birth in 1966 in the northern Finnish provinces of Oulu and Lapland. Data from the participants and their mothers were recorded during pregnancy and at birth with additional follow-up data being collected when they were 1, 14 and 31 years of age, by means of postal questionnaires. Supplementary data was obtained from hospital records, national registers and a physical examination at age 31 years. This investigation analyzed the data gathered at the 31-year follow-up. A detailed explanation of the study protocol for this follow-up may be found elsewhere. In brief, 8394 of the cohort members alive and living in Northern Finland or the Helsinki metropolitan area were initially contacted and invited to participate. A total of 6033 persons participated, of which 5122 provided answers to a series of standard psychometric instruments. Here we focus on the study of the summary scores from the following scales: Perceptual Aberration Scale (PAS), Hypomanic Personality Scale (HPS), Revised Social Anhedonia Scale (SAS), Revised Physical Anhedonia Scale (PHAS) (also known as Chapman’s Schizotypia Scales), and Schizoidia Scale (SCHS). PAS consists of 35 true/false items evaluating psychotic-like experiences including uncommon bodily discomforts, discontinuities and experiences (e.g., “My hearing is sometimes so sensitive that ordinary sounds become uncomfortable”). HPS includes 48 true/false items designed to identify hyperactive or exhibitionistic behaviors, feelings of euphoria, impulsivity, irritability, or mood swings (e.g. “I have often been so excited about an involving project that I didn't care about eating or sleeping”). PHAS comprises 61 true/false items assessing difficulties experiencing pleasure from physical stimuli that are usually pleasurable such as food, sex, visual or acoustic settings, etc. (e.g. “One food tastes as good as another to me”). SAS consists of 40 true/false items evaluating problems experiencing pleasure from non-physical stimuli and social interactions such as people's company, talking, etc. (e.g. “I prefer watching television to going out with other people.”).
Finally, SCHS includes 7 true/false items revealing key characteristics reckoned to be associated with schizotypal personality (e.g. “I am more sensitive than most other people”). These scales are well described and characterized in clinical and epidemiological psychiatric literature and their reliability and validity have been extensively examined previously, including in the current sample, and deemed acceptable (e.g. test-retest reliability values for the Chapman scales ranging between 0.75 to 0.85), which facilitates subsequent interpretation and understanding of the genetic results obtained. In addition a 12-item version of the Infrequency Scale was included to identify individuals offering random answers.

Exclusion procedures were applied to maximize the quality of the phenotypic data. First, when a participant left one or more items of a psychometric scale blank, the rest of the items for that scale were rejected, as full information was required to accurately build the corresponding score. Further, data from participants endorsing >2 items of the Infrequency Scale were disregarded as it was considered evidence of careless response. As we were interested in understanding the genetic predisposition to undergo psychosis-like experiences in the general population without overt psychotic disorders, individuals diagnosed between 1982 (when the cohort members turned 18 years old, the legal age of majority in Finland) to 1997 (when this last 31-year follow-up was completed) with schizophrenia or other psychotic disorders (e.g. schizophreniform disorder, schizoaffective disorder, delusional disorder, major depression with psychotic features, bipolar disorder with psychotic features, and psychotic disorders not otherwise specified) were excluded. This information was obtained from the Finnish Hospital Discharge Register. Finally, we disregarded participants who did not provide DNA or who showed any of the confounding characteristics for rejection of genomic information (see below). Supplementary Table 1 summarizes the numbers excluded at every step and the final sample sizes available for each trait.

Genomic data

The NFBC66 members participating at the 31-year follow-up were genotyped using the Illumina HumanHap CNV 370k array (Illumina Inc., San Diego, CA, USA). Quality control analyses were performed on raw genotype data: genotypes missing in >5% of the samples or samples missing >5% of genotypes, as well as samples showing signs of excessive genome-wide heterozygosity, gender discrepancies, or close consanguinity were excluded.
These cleaned genotypes formed the basis for imputing additional genomic variants, using the IMPUTE2 software\(^3\), considering the reference genomes from the 1000 Genomes Project (Phase I integrated variant set release from March 2012, Includes 93 Finnish individuals), and outlined as in the National Center for Biotechnology Information (NCBI) human genome assembly 37 (GRCh37). Quality control filters were subsequently applied to reduce the number of SNPs to a highly informative set of \(~9.8\) million SNPs. The exclusion criteria were: poorly imputed (an imputation informativeness INFO statistic from IMPUTE software <0.4), undefined genotypes (an allele posterior probability <0.95, genotypes set to “missing”), too rare (a minor allele frequency <0.001), poorly characterized (missing in >10% of the samples), or genetically unbalanced (Hardy-Weinberg Equilibrium test p-value \(<1 \times 10^{-6}\)).

**Statistical methods (extended in supplementary information)**

Descriptive statistics of the psychoses proneness traits were initially computed (Supplementary Table 2). Phenotypes were then corrected for sex and population substructure (10 principal components), and transformed to a Gaussian distribution by applying inverse normal transformation methods.

Estimates of the proportion of the phenotypic variances explained by the SNPs in autosomal chromosomes were obtained by genome-wide restricted maximum likelihood (GREML) models in the Genome-wide Complex Trait Analysis software (GCTA)\(^3\). The patterns of association between pairs of psychoses proneness traits were studied through examining their pairwise Pearson correlation coefficients. To understand the influences underlying such patterns observed at the phenotypic level, we further computed estimates of their genetic and environmental correlation through GCTA methods.

Phenotypes showing a significant heritability were examined in univariate association analyses using conventional linear regression modeling, while correlated phenotypes also showing significant heritability were entered into a series of bivariate association tests using canonical correlation analyses\(^4\). Both univariate and multivariate phenotype-SNP association tests were implemented through the corresponding extensions of the PLINK software\(^4, 4\).
Additional analysis of statistical power indicated that the sample size available here (N=4269, or the maximum number of individuals with phenotypic and genetic data) conferred 86.4% power to detect SNPs explaining ≥1% of the total variance in the phenotypes analyzed (Supplementary Figure 1), assuming a type-I error rate of $5 \times 10^{-8}$, complete LD between genotyped and causal markers, and a MAF ≥0.001 \cite{42, 43}.

Genomic regions harboring SNPs with a suggestive phenotype-SNP association p-value < $5 \times 10^{-6}$ were followed-up for gene-based association tests, using Versatile Gene-Based Test for Genome-wide Association methods within the VEGAS2 software \cite{44}. Because we are aware that this technique may have difficulties to pinpoint genes including one or very few significant SNPs (and depend somewhat on the LD structure of the region), we performed the analyses also taking into account only the top-10% most significant SNPs in each gene to partly correct for this limitation. Gene-based tests were consequently regarded in our study as an enhancement, rather than a replacement, to the single marker association tests \cite{45}. Genes offering p-values < 0.05 after Bonferroni correction taking into account the number of genes tested for each phenotype, were taken as significant.
Results

Genetic influences, as estimated from the ~9.6 million autosomal SNPs accounted for a significant part of the total variance for most of the phenotypes. The specific heritabilities (and SE) were as follows: $h^2_{HPS}=27.4\% (8.1)$; $h^2_{PAS}=16.6\% (6.9)$; $h^2_{PHAS}=26.6\% (7.9)$; $h^2_{SAS}=20.4\% (7.9)$; and $h^2_{SCHS}<0.1\% (6.9\%)$. Since the heritability for SCHS was extremely low, we disregarded the phenotype from further tests.

Phenotypic correlations between traits were low to moderate ($r_p=0.04-0.43$), while genetic ($r_g$) and environmental ($r_e$) correlations were moderate to high ($r_g=0.25-0.73$), and low to moderate ($r_e=0.12-0.43$), respectively (Table 1). The strongest of the pairwise phenotypic associations were $r_{p,PAS-HPS}=0.43$ and $r_{p,PHAS-SAS}=0.41$. However similar in magnitude, different patterns were detected in their underlying architecture, and genetic factors tend to account for a higher proportion of the association between PHAS-SAS than between PAS-HPS ($r_{g,PHAS-SAS}=0.73$ [SE $r_{g,PHAS-SAS}=0.20$] vs. $r_{g,PAS-HPS}=0.36$ [SE $r_{g,PAS-HPS}=0.22$]). Yet, given the wide standard error of these estimates this differential pattern would not be statistically significant.

The univariate GWAs analysis highlighted an intronic variant ($rs12449097$; p-value=$3.49 \times 10^{-8}$) at the TMC7 gene (16p12.3) associated to HPS (Table 2, Figure 1). Interestingly, this marker alone explained a meaningful 2.9% of the trait heritability. As this marker was imputed in our study sample, and was relatively uncommon, its imputation accuracy was verified by contrasting its minor allele frequency with respect to that observed in the Sequencing Initiative Suomi (SISu, www.sisuproject.fi) population reference exome set, which is based upon exome sequence data from 1918 Finnish individuals. Given that the marker was present in both populations at similar frequencies (MAF=0.003 here vs. MAF=0.004 in SISu), there are grounds to imply that $rs12449097$ was statistically imputed with reasonable precision in our study sample.

In total, 211 SNPs with suggestive p-value=$5 \times 10^{-6}$, in any of the four univariate traits tested, were mapped in or near to 108 different genes (HPS: 43 genes; PAS: 29 genes; PHAS: 23 genes; and SAS: 13 genes). The subsequent gene-based analyses pinpointed a series of loci harboring 6 significant genes for HPS, 4 genes for PAS, 1 gene for PHAS, and 3 genes for SAS (Table 3, and Supplementary Figures 2, 4 and 5). Altogether, these loci accounted for a meaningful proportion of the heritability of the corresponding psychosis proneness...
phenotypes: 13.9% of $h^2_{\text{HPS}}$, 10.4% of $h^2_{\text{PAS}}$, 3.7% of $h^2_{\text{PHAS}}$ and 7.1% of $h^2_{\text{SAS}}$. When taking into account the genetic variance explained by the SNPs pinpointed in marker-based GWAs tests, the genomic findings explained as much as 14.1% of $h^2_{\text{HPS}}$.

The bivariate GWAs analyses suggested a further SNP with a borderline genome-wide significant p-value ($rs188320715$; p-value=$5.261 \times 10^{-8}$) to HPS–PHAS. This SNP is located within an intergenic region at 6q25.3 with the nearest gene being $ARID1B$, ~572kb upstream (Table 2 and Supplementary Figure 2). A total of 263 SNPs showed a suggestive p-value<5 × 10^{-6} and they were mapped in or near to 125 different genes (HPS–PAS: 32 genes; HPS–PHAS: 57 genes; HPS–SAS: 40 genes; PAS–PHAS: 33 genes; PAS–SAS: 18 genes; PHAS–SAS: 38 genes). No signs of genomic inflation were detected (inflation factor $\lambda=1.01$-1.03, Supplementary Figure 3).

The subsequent gene-based analyses showed that genes at 1q32.1, 8q24.13 and 9p21.3 primarily associated to HPS also had residual effects on PAS, PHAS and SAS. Similarly, the gene $CACNA1C$ (12p13.33) primarily associated to HPS also had residual effects on PHAS, $MRC2$ (17q23.2) associated with PAS also had effects on HPS, PHAS and SAS; while $CCR4$ (3p22.3) associated to SAS also had effects on PHAS. These analyses revealed additional new loci harboring significant genes that passed undetected to all previous tests: $SNAP25$ (20p12.2) and $ADAM18$ (8p11.22) associated to both HPS and SAS; $CERKL$ (2q31.3) to both PAS and SAS; and $SLC1A2$ (11p13) to both PHAS and SAS (Table 3, and Supplementary Figures 2, 4 and 5).
Discussion

The initial part of this study provides heritability estimates for traits that relate to psychosis proneness, based on the analysis of ~9.6 million autosomal SNPs available from a large sample of unrelated individuals without overt psychotic disorders. It was earlier hypothesized that estimates obtained through this method would be lower than those detected in phenotypic analyses of twin samples. Our results provide support to this hypothesis as the estimates here are either similar or lower than those from previous twin studies with matching psychometric instruments (current vs. earlier): HPS: 27.4% vs. 28%; PAS: 16.6% vs. 33–49%; PHAS: 26.6% vs. 36–57%; and SAS: 20.4% vs. 45–67%. Accordingly, they constitute an upper estimate of genetic influences possible to identify by current genome wide association methods.

Other results in this study provide further relevant conclusions. First, by employing multiple continuous measures of psychosis proneness and utilizing statistical methods that allowed scrutinizing genomic regions of interest, we were able to identify a series of genetic loci using a much smaller sample size than has been required to pinpoint most loci in earlier psychiatric case-control GWAs settings. These loci support some earlier findings from case based GWAs studies, but also provide novel insights on the molecular pathways potentially contributing to the development of clinical psychosis, such as schizophrenia or bipolar disorder, as well as other behavioral problems usually associated with psychoses proneness (e.g. substance abuse, post-traumatic stress disorder, or some personality disorders). For example, the locus at 1q32.1 showed multiple genes with significant p-values in gene-based association tests involving HPS. A detailed examination of the LD structure at this locus in relation to the leading SNP suggested that KISS1, a gene highly expressed in different brain regions is likely to be the candidate gene for the locus and explains a meaningful 3.5% of the trait heritability (2.3% accounted for by the leading SNP in the locus alone). The neuropeptide coded by this gene (kisspeptin) modulates neuronal calcium homeostasis in hypothalamic neurons involved in the production of gonadotropin-release hormone and initiates the hypothalamic–pituitary–gonadal axis. Importantly, studies have recognized that gonadotropin alterations may prompt mood and behavioral changes. Hypothalamic-
pituitary dysfunctions have been earlier reported in relation to non-clinical psychotic feelings and experiences\textsuperscript{55}, and are also characteristic of schizophrenia patients\textsuperscript{56-58}. This finding thus provides preliminary molecular understanding on observations detected in earlier clinical studies among both psychiatrically healthy and ill populations.

Second, it is noticeable that our analyses also detected a series of genes previously linked to some extent either to schizophrenia alone or to both schizophrenia and bipolar disorder (\textit{CSMD1}, \textit{CCDC141}, \textit{SLC1A2}, \textit{CACNA1C} and \textit{SNAP25})\textsuperscript{7, 8, 59-61}. These genes accounted here for 2.8\% of the heritability of HPS, 3.8\% of PHAS and 1.8 \% of SAS. Particularly noteworthy is the effect on PHAS of \textit{CCDC141} alone (3.8\% of the heritability), a gene that interacts directly with the schizophrenia candidate gene \textit{DISC1}\textsuperscript{61}, and had been identified previously in this cohort in a GWAs conditioned on \textit{DISC1} genotypes\textsuperscript{24}. Altogether, these findings have important implications as they provide suggestive evidence of a genomic link between psychosis proneness in healthy adults and schizophrenia / bipolar disorder.

Previous positive evidence on this genomic link is limited. The study by Sieradzka et al. investigated a large community sample of 2152 16-year-olds, utilizing the Specific Psychotic Experiences Questionnaire in relation to polygenic risk scores for schizophrenia and bipolar disorder, as well as 28 individual SNPs previously associated with schizophrenia, and reported two SNPs at the \textit{TCF4} gene significantly associated with the paranoia subscale\textsuperscript{25}. Smaller-sized candidate gene studies have also reported positive findings for some markers at the \textit{COMT} gene associated with psychosis proneness scores in healthy samples of European and Chinese origin\textsuperscript{26, 62, 63}.

Our results here provide preliminary support to the idea that a biological susceptibility to disorders such as schizophrenia or bipolar disorder may still be partly found even in individuals without a psychosis diagnosis. Still, observational studies have shown that very few of the persons who score high on psychosis proneness scales go on to develop psychotic disorders\textsuperscript{64}, with subclinical negative psychotic symptoms (e.g. physical and social anhedonia) carrying stronger predictive power than others\textsuperscript{65, 66}. Altogether this would imply that only those individuals most susceptible, who were further affected by key genetic and / or environmental influences (e.g. an increased burden of large and rare chromosomal
abnormalities \(^67\), would finally develop those diseases. Further evidence on this respect is nevertheless necessary.

A series of methodological considerations need to be acknowledged here. This study ensues from earlier recommendations on psychiatric genetics encouraging new investigations to explore approaches in which traits, rather than being classified as dichotomized entities, are deconstructed into a spectrum of lower-order, highly characterized set of traits with clinical relevance \(^68\). This would initiate a dynamic process, in which refined phenotypes would allow pinpointing of new genetic signals and in turn generating further phenotype refinements leading to even lower-order traits \(^69\). Considering the findings in this study, in which the identified loci explained up to 14.1% of the estimated trait heritability, we believe that our methodological approach was indeed an advantage. The results reported here from an array of inter-related and carefully characterized traits represent only the first step in the genetic dissection of lower-order, high-resolution phenotypes regarding psychosis proneness, and new studies considering further phenotypic refinements are warranted to discover additional genomic regions of importance.

Due to the genomic link that our data suggested between psychosis proneness and schizophrenia / bipolar disorder, we believe that evaluating psychosis proneness even in non-psychotic individuals is an advisable strategy to understand part of the genetic basis of severe mental disorders with psychotic components. This conclusion however is in contrast to that of Zammit et al. \(^70\) from an earlier GWAs study evaluating psychosis proneness in 3483 non-psychotic adolescents. That previous study did not pinpoint any markers associated with psychotic-like experiences, and variants known to be associated with schizophrenia were also far from significant. The authors concluded that psychosis proneness in population-based samples may not share a comparable genetic architecture to schizophrenia, and thus utilizing a broader more common phenotype of psychotic experiences may not be an efficient approach to increase understanding of the genetic etiology of schizophrenia. This discrepancy with our conclusion is likely to be accounted for by two key methodological differences between the studies. Firstly our investigation evaluated a significantly larger pool of genetic markers (~9.8 million here vs. ~2.5 million there). This in turn may explain the greater capability of our investigation to detect important loci, as variants that are in high LD
with causal ones would be better captured. Secondly, and most importantly, we investigated here psychosis proneness utilizing summary (quantitative) scales built upon an ample number of answers provided to an array of questionnaires dissecting psychosis proneness into specific psychotic features. However, Zammit et al. carried out a pre hoc categorization on participants as “cases and controls” depending on whether they displayed any type of psychotic-like experience during adolescence. Given that this categorization resulted in 424 “cases” (or 12.2% of the participants) and 3057 “controls”, their statistical power to detect relevant loci was considerably lower compared to ours. To reach a reasonable statistical power similar to the one in our study (e.g. ≥80%) with their strategy of categorizing individuals as “cases and controls”, at least 7307 cases and as many as 52682 controls would be required to detect variants significant genome-wide ($P < 5 \times 10^{-8}$) with a moderate effect (relative risk ≥ 1.5), not rare in the population (risk allele frequency ≥ 0.01), assuming a 12.2% prevalence, and 7.21 cases-to-controls ratio (as provided by Zammit et al.). This illustrates further the high implications of utilizing quantitative endophenotypes in psychiatric genetic studies of persons not reaching a positive clinical diagnosis of schizophrenia or bipolar disorder.

Finally, it should be acknowledged that our study investigated a sample of adult individuals, while the two already mentioned genetic studies on the topic to date (by Sieradzka et al., and Zammit et al.) investigated adolescents. Despite psychosis proneness features being more common among adolescents, it has been suggested that when observed at that period it may constitute a developmental phenomenon, carrying a somewhat different connotation and weight than when observed in adulthood \(^7\). It is very likely however, that most genetic influences involved in adolescence psychotic-like experiences may still be present during adulthood, although a considerable proportion of them may well be age-specific, as genetic studies on many other complex diseases seem to indicate. Further longitudinal studies to clearly determine transitory and persistent genetic and environmental causes are thus warranted. In addition, despite that our descriptive statistics showed significant sex differences within the psychoses proneness scores, our genomic tests were limited to sex-corrected data, as sex-stratified analyses to detect potential sex-heterogeneity of the genetic effects would have resulted in a substantial reduction of the statistical power of detection.
In summary, this study highlights the value of using population-based data without overt psychiatric disorders for genetic association analyses concerning predisposition to psychoses and their related psychiatric conditions. Our results confirm that the genetic predisposition to psychosis proneness is moderate for most traits evaluated. We provide further evidence for overlap in genetic risk between psychiatric endophenotypes related to both schizophrenia and bipolar disorder. We also found association of several of these traits with 5 loci previously reported in connection with schizophrenia and/or bipolar disorder, and provide initial evidence for 15 additional loci that had not previously been connected to any psychiatric trait.
Acknowledgements

Dr. Nelson Freimer, Professor of Psychiatry at University of California, Los Angeles (USA), is gratefully acknowledged for his contribution to this study. The Academy of Finland supported this work (grant numbers 265097, and 259589). NFBC1966 received financial support also from the Academy of Finland (project grants [104781, 120315, 129269, 268336, 1114194], Center of Excellence in Complex Disease Genetics and SALVE); University Hospital Oulu, Biocenter, University of Oulu, Finland [75617]; the European Commission (EURO-BLCS, Framework 5 award [QLG1-CT-2000-01643]); National Heart, Lung and Blood Institute (NHLBI) grant [5R01HL087679-02] through the STAMPEED program [1RL1MH083268-01]; National Institute of Health/ National Institute of Mental Health [5R01MH63706:02]; ENGAGE project and grant agreement [HEALTH-F4-2007-201413]; and the Medical Research Council, UK [G0500539, G0600705, PrevMetSyn/SALVE]; EU Framework Programme 7 small-scale focused research collaborative project [EurHEALTHAgeing 277849]. The DNA extractions, sample quality controls, biobank up-keeping and aliquotting was performed in the National Public Health Institute, Biocemedic Helsinki, Finland and supported financially by the Academy of Finland and Biocentrum Helsinki.
References


42. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* Jan 2003;19(1):149-150.


46. Institute for Molecular Medicine Finland (FIMM); University of Helsinki (Finland). Sequencing Initiative Suomi project (SISu). Available at: [http://sisuproject.fi](http://sisuproject.fi).


Table 1: Phenotypic, genetic and environmental correlations between psychoses proneness traits

<table>
<thead>
<tr>
<th>Trait 1</th>
<th>Trait 2</th>
<th>N (pairwise)</th>
<th>r_p (SE)</th>
<th>P-value</th>
<th>r_g (SE)</th>
<th>P-value</th>
<th>r_e (SE)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPS</td>
<td>PAS</td>
<td>3824</td>
<td>0.432 (0.015)</td>
<td>&lt;1.00E-16</td>
<td>0.358 (0.217)</td>
<td>0.084</td>
<td>0.429 (0.056)</td>
<td></td>
</tr>
<tr>
<td>HPS</td>
<td>PHAS</td>
<td>3787</td>
<td>-0.204 (0.016)</td>
<td>&lt;1.00E-16</td>
<td>-0.435 (0.197)</td>
<td>0.023</td>
<td>-0.119 (0.076)</td>
<td></td>
</tr>
<tr>
<td>HPS</td>
<td>SAS</td>
<td>3837</td>
<td>-0.041 (0.016)</td>
<td>1.11E-02</td>
<td>-0.297 (0.262)</td>
<td>0.122</td>
<td>-0.140 (0.074)</td>
<td></td>
</tr>
<tr>
<td>PAS</td>
<td>PHAS</td>
<td>3812</td>
<td>0.040 (0.016)</td>
<td>1.14E-02</td>
<td>0.256 (0.261)</td>
<td>0.148</td>
<td>-0.122 (0.069)</td>
<td></td>
</tr>
<tr>
<td>PAS</td>
<td>SAS</td>
<td>3868</td>
<td>0.229 (0.016)</td>
<td>&lt;1.00E-16</td>
<td>0.337 (0.263)</td>
<td>0.121</td>
<td>0.205 (0.062)</td>
<td></td>
</tr>
<tr>
<td>PHAS</td>
<td>SAS</td>
<td>3820</td>
<td>0.410 (0.015)</td>
<td>&lt;1.00E-16</td>
<td>0.730 (0.203)</td>
<td>0.005</td>
<td>0.352 (0.062)</td>
<td></td>
</tr>
</tbody>
</table>

Note: r_p, r_g and r_e refers to estimates of phenotypic, genetic and environmental correlations (pair-wise); SE refers to standard errors of estimates; P-value obtained from tests on correlation estimates equal to 0 as null hypothesis.

Table 2: Relevant loci detected in univariate and bivariate marker-based association analyses

<table>
<thead>
<tr>
<th>Trait</th>
<th>SNP</th>
<th>Chr</th>
<th>Position</th>
<th>Gene</th>
<th>Allele</th>
<th>INFO</th>
<th>MAF</th>
<th>Missing</th>
<th>HWE</th>
<th>Test Statistic</th>
<th>P-value</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPS</td>
<td>rs12449097</td>
<td>16</td>
<td>19005186</td>
<td>TMC7</td>
<td>A/T</td>
<td>0.606</td>
<td>0.003</td>
<td>0.045</td>
<td>1</td>
<td>-5.527</td>
<td>3.485E-08</td>
<td>-1.13</td>
</tr>
<tr>
<td>PAS-PHAS</td>
<td>rs1803320715</td>
<td>6</td>
<td>156526681</td>
<td>ARID1B</td>
<td>C/T</td>
<td>0.600</td>
<td>0.005</td>
<td>0.065</td>
<td>0.117</td>
<td>16.830</td>
<td>5.261E-08</td>
<td>-0.616 / 0.609</td>
</tr>
</tbody>
</table>

Note: Chr refers to chromosome; Position refers to marker position (NCBI human genome assembly 37, GRCh37); Gene refers to suggested candidate gene in the SNP region; INFO refers to imputation informativeness INFO statistic from IMPUTE2 software; MAF refers to minor allele frequency; missing refers to percentage of individuals with missing genotype; HWE refers to p-values from Hardy-Weinberg Equilibrium tests; Test statistic obtained the corresponding univariate (t-statistic) or bivariate (F-statistic) tests; P-value from the corresponding univariate or bivariate test; Beta from independent univariate association tests.
Table 3: Results from gene-based association analyses, including the most likely candidate genes for each region

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Location</th>
<th>Gene</th>
<th>Top10%</th>
<th>Top1%</th>
<th>Top0.1%</th>
<th>Top0.01%</th>
<th>Top0.001%</th>
<th>Top0.0001%</th>
<th>Top0.00001%</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPS</td>
<td>20415468-20416569</td>
<td>SAS</td>
<td>0.559%</td>
<td>0.314%</td>
<td>0.036%</td>
<td>0.928%</td>
<td>0.642%</td>
<td>0.604%</td>
<td>0.756%</td>
</tr>
<tr>
<td>HPS-PAS</td>
<td>20415468-20416569</td>
<td>SAS</td>
<td>0.559%</td>
<td>0.314%</td>
<td>0.036%</td>
<td>0.928%</td>
<td>0.642%</td>
<td>0.604%</td>
<td>0.756%</td>
</tr>
<tr>
<td>HPS-PHAS</td>
<td>20415468-20416569</td>
<td>SAS</td>
<td>0.559%</td>
<td>0.314%</td>
<td>0.036%</td>
<td>0.928%</td>
<td>0.642%</td>
<td>0.604%</td>
<td>0.756%</td>
</tr>
<tr>
<td>HPS-SAS</td>
<td>20415468-20416569</td>
<td>SAS</td>
<td>0.559%</td>
<td>0.314%</td>
<td>0.036%</td>
<td>0.928%</td>
<td>0.642%</td>
<td>0.604%</td>
<td>0.756%</td>
</tr>
</tbody>
</table>

Note: Start-End BP refers to base-pair position for the gene start and end (NCBI human genome assembly 37, GRCh37); N SNP refers to the number of SNPs mapped to the gene region and used in gene-based association tests; Sim GENE indicates the number of simulations reached during the Vegas test for all markers mapped within the gene; Sim Top10% indicates the number of simulations reached during the Vegas test for the top 10% most relevant markers; VEGAS GENE indicates the VEGAS statistic for the test all SNPs mapped within the gene; VEGAS Top10% indicates the VEGAS statistic for the test the top 10% most relevant SNPs within the gene;
$P_{value_{GENE}}$ refers to the P-value obtained in VEGAS tests all SNPs mapped within the gene; $P_{value_{Top10\%}}$ refers to the P-value obtained in VEGAS tests including the top 10% most relevant SNPs within the gene (significant P-values after Bonferroni correction are in **bold**); Position BP refers to the chromosomal base-pair position of the most relevant marker within the gene tested; $P_{value_{SNP}}$ refers to the P-value obtained in the corresponding univariate or bivariate SNP-based test; $h^2_{GENE}$ refers to the heritability (genetic variance) explained by the gene region evaluated in VEGAS tests, using GCTA methods; $h^2_{SNP}$ refers to the heritability (genetic variance) explained by the best SNP within the gene evaluated in VEGAS tests.
Figure 1: Regional Manhattan plot (16p12.3) based on p-values from an univariate model including HPS. The candidate gene in the locus is marked with a star (*). The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAs findings (with p-value <5e-8) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov
**Supplementary Table 1: Evaluation of data participants before and after exclusions**

<table>
<thead>
<tr>
<th>Scale</th>
<th>Raw data</th>
<th>Step 1</th>
<th>Step 2</th>
<th>Step 3</th>
<th>Step 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Questions</td>
<td>N&lt;sub&gt;0&lt;/sub&gt;</td>
<td>Missing All / Part</td>
<td>N&lt;sub&gt;1&lt;/sub&gt;</td>
<td>Infreq. Scale &gt; 2</td>
</tr>
<tr>
<td>HPS</td>
<td>4522</td>
<td>5122</td>
<td>19 / 422</td>
<td>4073</td>
<td>95</td>
</tr>
<tr>
<td>PAS</td>
<td>2228</td>
<td>5122</td>
<td>20 / 2599</td>
<td>4083</td>
<td>99</td>
</tr>
<tr>
<td>PHAS</td>
<td>2212</td>
<td>5122</td>
<td>18 / 395</td>
<td>4709</td>
<td>91</td>
</tr>
<tr>
<td>SAS</td>
<td>2226</td>
<td>5122</td>
<td>20 / 335</td>
<td>4747</td>
<td>96</td>
</tr>
<tr>
<td>SCHS</td>
<td>7</td>
<td>5122</td>
<td>28 / 143</td>
<td>4951</td>
<td>103</td>
</tr>
</tbody>
</table>

HPS: Hypomanic Personality Scale; PAS: Perceptual Aberration Scale; PHAS: Revised Physical Anhedonia Scale; SAS: Revised Social Anhedonia Scale; SCHS: Schizoidia Scale.

N<sub>0</sub>: Initial number of individuals that consented to participate and provided any answer to the overall questionnaire

N<sub>1</sub>: Number of individuals after excluding those with incomplete data for a given scale

N<sub>2</sub>: Number of individuals after removing unreliable data (individuals with >2 answers endorsed in a 12-items Infrequency scale)

N<sub>3</sub>: Number of individuals after excluding persons with a medical diagnosis of schizophrenia, schizophrenia spectrum disorders, psychosis and psychoses spectrum disorders

N<sub>4</sub>: Number of individuals after excluding persons with missing genetic information

**Supplementary Table 2: Description of the sample values for the different psychological scales evaluated**

<table>
<thead>
<tr>
<th>Scale</th>
<th>N All / Women</th>
<th>Mean Score (95% CI)</th>
<th>% Deviant Cases *</th>
<th>Men All / Women</th>
<th>Mean (95% CI)</th>
<th>Women All / Women</th>
<th>Mean (95% CI)</th>
<th>Sex differences **</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPS</td>
<td>3967/2188</td>
<td>11.211 (10.994 - 11.428)</td>
<td>2.285/24 /0</td>
<td>10.491 (10.169 - 10.813)</td>
<td>11.791 (11.505 - 12.088)</td>
<td>0.201</td>
<td>2.26E-10</td>
<td>0.010</td>
</tr>
<tr>
<td>PAS</td>
<td>4057/2228</td>
<td>2.263 (2.162 - 2.36)</td>
<td>4.634/1081</td>
<td>1.96 (1.325 - 2.095)</td>
<td>2.513 (2.377 - 2.648)</td>
<td>0.496</td>
<td>4.78E-12</td>
<td>0.012</td>
</tr>
<tr>
<td>PHAS</td>
<td>3988/2212</td>
<td>14.975 (14.761 - 15.19)</td>
<td>4.413/0.827</td>
<td>17.809 (17.476 - 18.142)</td>
<td>12.7 (12.459 - 12.942)</td>
<td>-0.735</td>
<td>6.90E-127</td>
<td>0.134</td>
</tr>
<tr>
<td>SCHS</td>
<td>4191/2330</td>
<td>2.549 (2.506 - 2.591)</td>
<td>2.768/1.885</td>
<td>2.259 (2.196 - 2.322)</td>
<td>2.781 (2.725 - 2.836)</td>
<td>0.362</td>
<td>2.61E-34</td>
<td>0.035</td>
</tr>
</tbody>
</table>

HPS: Hypomanic Personality Scale; PAS: Perceptual Aberration Scale; PHAS: Revised Physical Anhedonia Scale; SAS: Revised Social Anhedonia Scale; SCHS: Schizoidia Scale.

* % Deviant cases * refers to percentage of cases with values > 2 standard deviations from the mean score within the population

** Sex-differences evaluated for each scale through linear regression modeling of normalized and transformed values.
**Supplementary Methods**

**Extended description of statistical methods**

Descriptive statistics of the psychoses proneness traits were initially computed (Supplementary Table 2). Phenotypes were corrected for sex and population substructure, by obtaining standardized residuals from linear regression models entering as covariates individuals’ gender and the first 10 eigenvectors of the genetic relatedness matrix (obtained from a subset of ~248K genotyped autosomal markers with linkage disequilibrium $r^2$<0.2 and MAF>0.05). Since the probability distribution functions of these regression residuals were still highly skewed, and thus could lead to spurious results in later analyses, they were further transformed to a normal (Gaussian) distribution by applying inverse normal transformation methods.

Estimates of the proportion of the phenotypic variances explained by the SNPs in autosomal chromosomes were obtained by genome-wide restricted maximum likelihood (GREML) models in the Genome-wide Complex Trait Analysis software (GCTA) \(^1\). GCTA first allowed computing participants’ genetic relatedness matrix (GRM) utilizing the genomic information available from autosomal chromosomes, through which we excluded close relatives (e.g. a genomic relatedness score >0.025, that is 2.5% similarity or approximately a familial connection of third or fourth cousins) \(^2\). Subsequently, this GRM was introduced as a random effects term within a model that decomposed the phenotypic variance into genetic and environmental influences using residual maximum likelihood methods. A narrow estimate of the heritability was finally calculated as the ratio of the additive genetic effects to the total phenotypic variance. An estimate of environmental influences was also obtained through an analogous algorithm. Standard errors of estimates were computed from the estimated errors of the variances using a first-order Taylor expansion.

The patterns of association between pairs of psychoses proneness traits were studied through examining their pairwise Pearson correlation coefficients. To understand the influences underlying such patterns observed at the phenotypic level, we further computed estimates of their genetic and environmental correlation through GCTA methods. The phenotypic covariance between two traits was decomposed as a function of their estimated genetic and environmental covariances through a model that incorporated the sample’s GRM
as a random effects term. These genetic and environmental covariances were later used to obtain the specific genetic correlation as the ratio of the genetic covariance to the square root of the addition of the specific genetic variances estimated for each trait. Environmental correlations were also obtained alike. As in heritability analyses, standard errors of estimates were computed using a first-order Taylor expansion.

Phenotypes showing a significant heritability were examined in univariate association analyses using conventional linear regression modeling, while correlated phenotypes also showing significant heritability were entered into a series of bivariate association tests using canonical correlation analyses (CCA). These CCA, applied to genomic data, allows testing for linear combinations of phenotypes that correlate maximally with linear combinations of SNPs. In the present study, we investigated pairs of psychoses proneness phenotypes that correlated maximally with a single genetic marker. Both univariate and multivariate phenotype-SNP association tests were implemented through the corresponding extensions of the PLINK software. Additional analysis of statistical power indicated that the sample size available here (N=4269, or the maximum number of individuals with phenotypic and genetic data) conferred 86.4% power to detect SNPs explaining ≥1% of the total variance in the phenotypes analyzed (Supplementary Figure 1), assuming a type-I error rate of $5 \times 10^{-8}$, complete LD between genotyped and causal markers, and a MAF ≥0.001.

Genomic regions harboring SNPs with a suggestive phenotype-SNP association p-value < $5 \times 10^{-6}$ were followed-up for gene-based association tests, using Versatile Gene-Based Test for Genome-wide Association methods within the VEGAS2 software. If one or more of those suggestive SNPs were within ±50kb of a gene’s 5’ and 3’ UTRs, all regional p-values obtained from the corresponding univariate or bivariate association scans were entered in VEGAS2. This technique first maps the p-values provided to each gene and then converts them to upper-tail chi-squared statistics with one degree of freedom to sum them and obtain a gene-based statistic. Since these p-values are partly dependent on the LD structure within the region, VEGAS2 implements Monte-Carlo simulations of the LD structure based on pre-compiled 1000 Genomes data (specifying the Finnish subsample as reference) in a stepwise algorithm starting with 1000 simulations and reaching a specified maximum of 10 million simulations.
We are aware that this technique presents some limitations. For instance, besides disregarding intergenic SNPs by definition, this procedure may have difficulties to pinpoint genes including one or very few significant SNPs, as their effect would be diluted by the inclusion of a greater number of non-significant markers. This number of non-significant SNPs may partly depend on the underlying structure and LD patterns within the gene region. To partly correct this limitation, we performed the analyses also taking into account the top-10% most significant SNPs in each gene. Gene-based tests were consequently regarded in our study as an enhancement, rather than a replacement, to the single marker association tests. Genes offering p-values < 0.05 after Bonferroni correction taking into account the number of genes tested for each phenotype, were taken as significant.

References


Supplementary Figure 1: Statistical power of detection in the current study for markers explaining between 0.1 and 3% of the trait heritability. The current heritability confers 86.41% power to detect SNPs explaining 1% of the trait heritability.
Supplementary Figure 2: Genome-wide Manhattan plots for univariate and bivariate models computed. Significant SNPs in marker-based tests are highlighted with a red square. The leading SNPs in loci offering significant p-values in gene-based analyses are highlighted with a green diamond.

Supplementary Figure 2.1: Genome-wide Manhattan plot for HPS

Supplementary Figure 2.2: Genome-wide Manhattan plot for HPS-PAS
Supplementary Figure 2.3: Genome-wide Manhattan plot for HPS-PHAS

Supplementary Figure 2.4: Genome-wide Manhattan plot for HPS-SAS
Supplementary Figure 2.5: Genome-wide Manhattan plot for PAS

Supplementary Figure 2.6: Genome-wide Manhattan plot for PAS-PHAS
Supplementary Figure 2.7: Genome-wide Manhattan plot for PAS-SAS

Supplementary Figure 2.8: Genome-wide Manhattan plot for PHAS
Supplementary Figure 2.9: Genome-wide Manhattan plot for PHAS-SAS

Supplementary Figure 2.10: Genome-wide Manhattan plot for SAS
Supplementary Figure 3: Regional Manhattan plot (6q25.3) based on p-values from a bivariate model for PAS-PHAS. The candidate gene in the locus is marked with a star (*). The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAs findings (with p-value <5e-8) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov
Supplementary Figure 4: Quintile-Quintile (QQ) plots for univariate and bivariate models computed.

Lambda refers to the genomic inflation factor (λ) or the ratio of the observed median p-value to its expected value.
Supplementary Figure 5: Regional Manhattan plots of loci of interest

Supplementary Figure 5.1: Regional Manhattan plot (1q32.1) based on p-values from an univariate model for HPS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according to their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37: GRCh37. Information on previous GWAs findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.2: Regional Manhattan plot (8q24.13) based on p-values from an univariate model for HPS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according to their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37: GRCh37. Information on previous GWAs findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.3: Regional Manhattan plot (9p21.3) based on p-values from an univariate model for HPS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according to their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37: GRCh37. Information on previous GWAs findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.4: Regional Manhattan plot (12p13.33) based on p-values from an univariate model for HPS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according to their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37: GRCh37. Information on previous GWAs findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov
Supplementary Figure 5.5: Regional Manhattan plot (13q32.1) based on p-values from a univariate model for HPS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.6: Regional Manhattan plot (1q32.1) based on p-values from a bivariate model for HPS-PAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.7: Regional Manhattan plot (8q24.13) based on p-values from a bivariate model for HPS-PAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.8: Regional Manhattan plot (9p21.3) based on p-values from a bivariate model for HPS-PAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

http://www.schizophreniabulletin.oupjournals.org
Supplementary Figure 5.9: Regional Manhattan plot (13q12.2) based on p-values from a bivariate model for HPS-PAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs colored according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.10: Regional Manhattan plot (17q23.2) based on p-values from a bivariate model for HPS-PAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs colored according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.11: Regional Manhattan plot (1q32.1) based on p-values from a bivariate model for HPS-PHAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs colored according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.12: Regional Manhattan plot (2q31.2) based on p-values from a bivariate model for HPS-PHAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs colored according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov
Supplementary Figure 5.13: Regional Manhattan plot (6q22.32) based on p-values from a bivariate model for HPS-PHAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.14: Regional Manhattan plot (8q24.13) based on p-values from a bivariate model for HPS-PHAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.15: Regional Manhattan plot (9p21.3) based on p-values from a bivariate model for HPS-PHAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.16: Regional Manhattan plot (12p13.33) based on p-values from a bivariate model for HPS-PHAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov
Supplementary Figure 5.17: Regional Manhattan plot (1q32.1) based on p-values from a bivariate model for HPS-PAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37. GRCh37. Information on previous GWAs findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.18: Regional Manhattan plot (3p22.3) based on p-values from a bivariate model for HPS-SAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37. GRCh37. Information on previous GWAs findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.19: Regional Manhattan plot (8q24.13) based on p-values from a bivariate model for HPS-SAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37. GRCh37. Information on previous GWAs findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.20: Regional Manhattan plot (9p21.3) based on p-values from a bivariate model for HPS-SAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37. GRCh37. Information on previous GWAs findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov
Supplementary Figure 5.21: Regional Manhattan plot (20p12.2) based on p-values from a bivariate model for HPS-SAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.22: Regional Manhattan plot (1p36.12) based on p-values from a bivariate model for PAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.23: Regional Manhattan plot (12q24.33) based on p-values from a univariate model for PAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.24: Regional Manhattan plot (16p13.11) based on p-values from a univariate model for PAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov
Supplementary Figure 5.25: Regional Manhattan plot (17q23.2) based on p-values from a univariate model for PAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.26: Regional Manhattan plot (8p11.22) based on p-values from a bivariate model for PAS-PHAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.27: Regional Manhattan plot (17q23.2) based on p-values from a bivariate model for PAS-PHAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.28: Regional Manhattan plot (2q31.3) based on p-values from a bivariate model for PAS-SAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

http://www.schizophreniabulletin.oupjournals.org
Supplementary Figure 5.29: Regional Manhattan plot (3p22.3) based on p-values from a bivariate model for PAS-SAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 27, GRCh37. Information on previous GWAs findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.30: Regional Manhattan plot (17q23.2) based on p-values from a bivariate model for PAS-SAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 27, GRCh37. Information on previous GWAs findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.31: Regional Manhattan plot (2q31.2) based on p-values from a bivariate model for PHAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 27, GRCh37. Information on previous GWAs findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov

Supplementary Figure 5.32: Regional Manhattan plot (3p22.3) based on p-values from a bivariate model for PHAS-SAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic position as in the National Center for Biotechnology Information (NCBI) human genome assembly 27, GRCh37. Information on previous GWAs findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov
Supplementary Figure 5.33: Regional Manhattan plot (11p13) based on p-values from a bivariate model for PHAS-SAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic positions as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov.

Supplementary Figure 5.34: Regional Manhattan plot (3p22.3) based on p-values from a bivariate model for SAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic positions as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov.

Supplementary Figure 5.35: Regional Manhattan plot (8p23.2) based on p-values from an univariate model for SAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic positions as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov.

Supplementary Figure 5.36: Regional Manhattan plot (12q14.3) based on p-values from an univariate model for SAS. Genes with significant p-values in gene-based tests are in orange font while the candidate gene is in red font. The leading SNP within the locus is highlighted in green; other SNPs coloured according their linkage disequilibrium (LD) with it. Genomic positions as in the National Center for Biotechnology Information (NCBI) human genome assembly 37, GRCh37. Information on previous GWAS findings (with p-value <5e-6) retrieved from the National Human Genome Research Institute (NHGRI): http://www.genome.gov.

http://www.schizophreniabulletin.oupjournals.org