



Published in final edited form as:

*Schizophr Res.* 2014 April ; 154(0): 48–53. doi:10.1016/j.schres.2014.01.030.

## Genetic modifiers and subtypes in schizophrenia: Investigations of age at onset, severity, sex and family history

Sarah E. Bergen<sup>a,b,j,m,\*</sup>, Colm T. O'Dushlaine<sup>b</sup>, Phil H. Lee<sup>a,b,i</sup>, Ayman H. Fanous<sup>c,d,e,f,g</sup>, Douglas M. Ruderfer<sup>h,i</sup>, Stephan Ripke<sup>b,i</sup>, International Schizophrenia Consortium, Swedish Schizophrenia Consortium, Patrick F. Sullivan<sup>l</sup>, Jordan W. Smoller<sup>a,j</sup>, Shaun M. Purcell<sup>h,i</sup>, and Aiden Corvin<sup>k</sup>

<sup>a</sup>Psychiatric & Neurodevelopmental Genetics Unit, Center for Human Genetics Research, Massachusetts General Hospital, Boston, MA, USA <sup>b</sup>Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA <sup>c</sup>Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA <sup>d</sup>Washington VA Medical Center, Washington, DC, USA <sup>e</sup>Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA <sup>f</sup>Department of Psychiatry, Georgetown University School of Medicine, Washington, DC, USA <sup>g</sup>Department of Psychiatry, Keck School of Medicine of the University of Southern California, Los Angeles, CA, USA <sup>h</sup>Mount Sinai School of Medicine, New York, NY, USA <sup>i</sup>Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA <sup>j</sup>Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA <sup>k</sup>Neuropsychiatric Genetics Research Group, Department of Psychiatry and Institute of Molecular Medicine, Trinity College Dublin, Dublin 2, Ireland <sup>l</sup>Departments of Genetics, Psychiatry, and Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA <sup>m</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden

### Abstract

Schizophrenia is a genetically and clinically heterogeneous disorder. Genetic risk factors for the disorder may differ between the sexes or between multiply affected families compared to cases with no family history. Additionally, limited data support a genetic basis for variation in onset and severity, but specific loci have not been identified. We performed genome-wide association studies (GWAS) examining genetic influences on age at onset (AAO) and illness severity as well

© 2014 Elsevier B.V. All rights reserved.

\*Corresponding author at: Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Box 281, 171 77 Stockholm, Sweden. Tel.: +46 8 52486125; fax: +46 8 31 49 75. sbergen@gmail.com (S.E. Bergen).

### Contributors

SEB conducted most of the analyses and wrote the manuscript. CTO'D contributed to the genetic analyses. PHL ran the pathway analyses and wrote the methods for this. AHF provided information from the PGC AAO results. DMR imputed the genetic data. SR provided the PGC results for male-only analyses. PFS was instrumental in initial stages of compiling phenotypes across ISC sites. JWS offered analytic and manuscript advice. SMP advised on the genetic analysis methods. AC designed the study, harmonized the phenotypes across sites and contributed to the manuscript preparation. All authors have contributed to and approved the final manuscript.

### Conflict of interest statement

All authors have reported no financial interests or potential conflicts of interest.

as specific risk by sex or family history status using up to 2762 cases and 3187 controls from the International Schizophrenia Consortium (ISC).

Subjects with a family history of schizophrenia demonstrated a slightly lower average AAO that was not significant following multiple testing correction ( $p = .048$ ), but no differences in illness severity were observed by family history status ( $p = .51$ ). Consistent with prior reports, we observed earlier AAO ( $p = .005$ ) and a more severe course of illness for men ( $p = .002$ ). Family history positive analyses showed the greatest association with *KIF5C* ( $p = 1.96 \times 10^{-8}$ ), however, genetic risk burden overall does not differ by family history. Separate association analyses for males and females revealed no significant sex-specific associations. The top GWAS hit for AAO was near the olfactory receptor gene *OR2K2* ( $p = 1.52 \times 10^{-7}$ ). Analyses of illness severity (episodic vs. continuous) implicated variation in *ST18* ( $p = 8.24 \times 10^{-7}$ ). These results confirm recognized demographic relationships but do not support a simplified genetic architecture for schizophrenia subtypes based on these variables.

## Keywords

GWAS; Association; psychosis; Gender

---

## 1. Introduction

Schizophrenia is a complex genetic disorder for which heterogeneity at the genetic and phenotypic levels has led to inconsistent results across several domains of research. The growing number of subjects in genome-wide association studies (GWAS) enhances power not only for detection of susceptibility loci, but also for numerous other investigations such as modifier locus detection or identifying more homogeneous disease subgroups. With heritability estimates of 64–81% (Sullivan et al., 2003; Lichtenstein et al., 2009), the genetic contribution to schizophrenia is substantial. As with other common complex disorders, the identification of specific, replicated common risk loci has required large-scale collaborative efforts including many thousands of subjects (Schizophrenia Psychiatric GWAS Consortium, 2011; Bergen and Petryshen, 2012). Genetic variation almost certainly impacts variability at the phenotypic level, and genetic factors contributing to variation within a disease may overlap with risk loci (susceptibility-modifiers) or could be independent of disease risk (true modifiers) (Fanous and Kendler, 2005).

Genome-wide association studies have not yet specifically addressed the familial/sporadic distinction. Unlike some other complex diseases, familial transmission of highly penetrant genetic risk factors in families has not been identified for schizophrenia. Several reports of differential genetic associations by sex have been published for schizophrenia, but few of these findings have been replicated, and to our knowledge, sex-specific genetic associations also remain untested in a GWAS framework.

In addition to the wide range of symptoms displayed by people with schizophrenia, several other features of this illness are highly variable. Schizophrenia may be diagnosed at any time from childhood through old age, and the severity of illness can range from mild with few psychotic episodes to severe and unremitting. Although similar numbers of males and

females are diagnosed with schizophrenia, males often have an earlier age at onset (AAO), increased illness severity, and less favorable prognoses (Lewine, 1980; Angermeyer and Kuhn, 1988; Hafner et al., 1993; Gorwood et al., 1995; Esterberg et al., 2010). Pubertal timing is reported to influence AAO in women suggesting sex-specific biological mechanisms influencing disease onset but not risk (Cohen et al., 1999). Heritability of AAO has been estimated at 33% indicating moderate genetic basis (Hare et al., 2010). Increased genetic risk loading has been related to earlier AAO for systemic lupus erythematosus (SLE), cancer, and Alzheimer's disease and may also be true for schizophrenia (Beyer et al., 2005; Chen et al., 2009; Webb et al., 2011). Some individual genes have been associated with measures of severity in schizophrenia such as *COMT* (Wright et al., 2012), *RGS2* and *RGS5* (Campbell et al., 2008), but the genetic relationship to course of illness has not yet been assessed on a genomic scale.

In this study, we explore the relationships between AAO, severity, sex, and family history, and investigate modifying genetic influences on age at onset (AAO) and severity as well as whether different genetic risk factors exist between males and females or by family history status.

## 2. Methods

### 2.1. Subjects

All participating adults were drawn from the International Schizophrenia Consortium (ISC); data collection, genotyping, and quality control steps have been previously described (International Schizophrenia Consortium, 2008; 2009). Analyses were restricted to subjects of European ancestry from the six sites with available phenotypic information (Aberdeen, Cardiff, Dublin, Edinburgh, Portugal, and London) resulting in up to 2762 cases and 3187 controls remaining for analyses. All subjects gave written informed consent.

### 2.2. Phenotypes

AAO was defined as the age at diagnosis for schizophrenia. Cases were diagnosed according to DSM-IV (American Psychiatric Association, 2000) or ICD-10 (Janca et al., 1993) criteria as described in prior publications (International Schizophrenia Consortium, 2008; 2009). AAO values were log transformed by site to approximate a normal distribution. Subjects with any first degree relatives diagnosed with a psychotic disorder were deemed family history positive (FH+). Course of illness/severity was coded 1 (remitting) to 5 (severe, chronic) according to the Operational Criteria Checklist (OPCRIT) (McGuffin et al., 1991; Williams et al., 1996). Subjects from Edinburgh were removed from the severity analyses due to lack of variance for this measure. Genetic sex defined the male–female distinction. The number of subjects with available information for each analysis across all sites is shown in Table 1.

### 2.3. Phenotypic analysis methods

Two-tailed t-tests, conducted in R (R Development Core Team, 2012), were used to investigate 1) AAO differences by FH, 2) course by sex, 3) AAO by sex, and 4) AAO and

illness severity. Conservative adjustment presuming independence of the four tests conducted yielded a Bonferroni-corrected p-value threshold of .0125.

## 2.4. Genetic analyses

Most subjects were genotyped using Affymetrix 5.0 and 6.0 arrays, however, control subjects from London were genotyped using the Affymetrix 500k array. Single nucleotide polymorphism (SNP) calls were made using Birdsuite software (Korn et al., 2008). Quality control steps were fully reported previously (International Schizophrenia Consortium, 2009), but briefly: SNPs were excluded for low call rates or significantly different call rates between cases and controls, mapping to multiple locations, lack of variation (monomorphic), departure from Hardy–Weinberg equilibrium, or minor allele frequency  $<.01$ . Subjects were removed for low genotyping rates, having a second degree or closer relative in the dataset, population outlier status, or suspected sample contamination. Beagle software (Browning and Browning, 2007) was used to impute genotypes from HapMap2 (International HapMap Consortium, 2007) reference files containing ~2.4 million markers. Autosomal markers were imputed, and markers with high confidence scores (Info  $>.8$ ) were retained for analyses.

**2.4.1. GWAS analyses**—All analyses were conducted in PLINK (Purcell et al., 2007) using linear regression (AAO) or logistic regression (sex, FH, course) and four multidimensional scaling covariates to control for genetic population variation. Sex was included as a covariate when not specifically tested. Most analyses were conducted by site and meta-analyzed. For severity, however, some sites contained few subjects with a mild, episodic course, therefore, subjects across all sites were analyzed together. Additionally, severity was dichotomized (mild 1–3 or severe 4–5) since this binary measure is less sensitive to potential rating differences across sites. After a Bonferroni correction to adjust the genome-wide association significance threshold ( $p < 5 \times 10^{-8}$ ) for the six analyses performed, the significance threshold is  $p < 8.33 \times 10^{-9}$ .

We sought replication for our top AAO results in the Psychiatric Genomics Consortium schizophrenia data; removing the overlapping subjects yielded results from 5079 additional cases. Proxy SNPs with  $r^2 > .8$  were determined for markers not directly represented in both samples using the SNAP tool (Johnson et al., 2008). Family history and illness severity were not available measures for replication.

In addition to the separate GWAS analyses for AAO and sex, the genetic basis for the observed differences in AAO by sex was examined using the  $G \times E$  test in PLINK which necessitated using the directly genotyped markers and no covariates.

**2.4.2. Pathway analyses**—To investigate whether nominally significant association signals cluster in genes involved in common biological pathways, top results from each association analysis were tested against Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and synaptic functional gene-groups using INRICH (Lee et al., 2012). The KEGG database is a collection of molecular interaction and reaction network maps, manually defined and drawn by field experts. The synaptic gene database is comprised of 18 groups of genes involved in synaptic functioning defined based on data mining,

experimentation, and experts' curation (Lips et al., 2012). Details of the pathway analysis methods are located in the supplementary material, Appendix A.

**2.4.3. Polygenic profile scoring**—Polygenic scoring was used to evaluate whether FH+ subjects have increased genetic risk burden compared to FH- subjects and whether subjects with earlier AAO have increased genetic risk loading compared to later onset subjects. Scoring was conducted for each group separately using a Swedish schizophrenia dataset (2111 cases, 2535 controls) as the discovery sample (Bergen et al., 2012; International Schizophrenia Consortium, 2009). To implement this method, SNPs are weighted on the basis of the strength of their association (e.g. log OR) in a discovery sample, and these weights are applied to the SNPs in a target sample and aggregated to yield polygenic profiles scores for each individual. Quantitative scores were calculated for each subject based on the  $p_T$  (p-value threshold), the proportion of SNPs with p-values  $< p_T$  in the discovery sample. For each SNP set defined by  $p_T$ , we calculated the proportion of variance explained ( $R^2$ ) by subtracting the Nagelkerke's  $R^2$  attributable to ancestry covariates alone from the  $R^2$  for polygenic scores plus covariates (International Schizophrenia Consortium, 2009). Sample size is one determinant of the proportion of variance explained in polygenic scoring. Therefore, a random subset of the FH-sample was selected to be the same size as the FH+ sample. We also tested for possible differences in polygenic risk for early and late AAO by dividing the ISC (target set) cases into similarly-sized early (AAO  $\leq 21$ , N = 1100) and late (AAO  $> 21$ , N = 1283) groups. T-tests were then performed for the resulting scores between subgroups across each p-value threshold examined (.01, .05, .1, .2, .3, .4, and .5).

### 3. Results

#### 3.1. Phenotypic analysis results

We observed later average AAO in women (24.7 yrs) compared with men (23.7 yrs;  $t = -2.80$ ,  $p = .005$ ) and a more severe course of illness for men ( $t = 3.15$ ,  $p = .002$ ). Regardless of sex, a lower AAO was significantly related to a more severe disease course ( $t = -3.90$ ,  $p = 9.93 \times 10^{-5}$ ). Family history positive subjects demonstrated a lower average AAO ( $t = 1.98$ ,  $p = .048$ ), but this was not significant following multiple testing correction. There were no significant differences in illness severity between family history groups ( $t = -.66$ ,  $p = .51$ ).

#### 3.2. Genetic results

**3.2.1. GWAS results**—Top results for all GWAS analyses are shown in Table 2. The strongest association signal for AAO analyses (rs11999864,  $p = 1.52 \times 10^{-7}$ ) spanned two genes: an olfactory receptor, *OR2K2*, and a scaffolding protein, *KIAA0368*. We also investigated the top loci for our AAO analyses in the Psychiatric Genomics Consortium results. Restricting analyses to the samples not overlapping with this study and markers with  $r^2 > .8$  yielded no additional support for the loci identified by this study.

Analyses of disease severity highlighted variation near the *ST18* gene as showing the greatest difference between subjects with a relatively mild episodic course compared to subjects with severe, chronic forms of illness (rs12548481,  $p = 8.24 \times 10^{-7}$ ). This gene

encodes a siRNA which is most highly expressed in brain and regulates apoptotic and inflammatory pathways (Yang et al., 2008).

In the sex-specific GWAS, the top result for the female-only analyses (rs17115481,  $p = 1.90 \times 10^{-6}$ ) also extended across two genes, *FAM114A2*, a brain expressed gene, and *MFAP3*, a microfibrillar associated protein. The male-specific analyses yielded the strongest signal in *JMJD2C* (rs7037878,  $p = 8.07 \times 10^{-7}$ ), a histone demethylase which regulates androgen receptor function (Wissmann et al., 2007).

The AAO by sex test implicated variation in the *GATA5* gene (rs6121620,  $6.54 \times 10^{-7}$ ). The strength of the association signal for this SNP is fairly similar in males and females, but it is in different directions (males: beta = .052,  $p = 1.33 \times 10^{-5}$ ; females: beta = -.073,  $p = 4.76 \times 10^{-4}$ ). This gene has known functions in cardiac and smooth muscle, but its role in the brain is not yet known.

In the FH+ analyses, a strong, but not significant, signal (rs10929935,  $p = 1.96 \times 10^{-8}$ ) spanned two genes: *KIF5C*, a microtubule-associated protein involved in organelle transport, and *LYPD6B* which is brain expressed but otherwise uncharacterized. Family history negative results revealed the strongest association signal in *PTBP2* (rs11165690,  $p = 9.96 \times 10^{-8}$ ), a regulator of exon splicing and neuronal precursor cell differentiation.

**3.2.2. Pathway analysis results**—Two pathways from the synaptic database and none from KEGG remained significant following correction for all pathways tested within each set of association results (Supplementary Tables 1A–D). The “neurotransmitter metabolism” set was significantly enriched in the FH+ analyses ( $p = .045$ ), and “intracellular signal transduction” was significant in the AAO results ( $p = .023$ ). For the FH–, severity, male-only, and female-only results, no pathways were significantly enriched.

**3.2.3. Polygenic profile scoring results**—FH+ and FH– groups showed no significant differences at any p-value threshold tested. Similarly, subjects with AAO 21 did not show increased polygenic risk burden (Supplementary Table 2).

## 4. Discussion

We tested genome-wide genotype data for association with AAO and severity of schizophrenia and whether differences in genetic risk factors were evident between FH+ and FH– groups or between males and females. Aside from one study of AAO (Wang et al., 2011), these measures have been untested in a GWAS framework. Although results did not surpass genome-wide thresholds of significance, several strong association signals were observed.

The strongest association signal for the AAO genetic analyses (rs11999864,  $p = 8.85 \times 10^{-7}$ ) did not show association in the main case–control analyses ( $p = .96$ ) indicating potential modifier properties at this locus rather than susceptibility-modifier influences. Attempts to replicate the top loci for our AAO analyses in the Psychiatric Genomics Consortium yielded no additional support. The prior GWAS of AAO in a European–American sample reported association signals at 8q24.22, 4q25 in *COL25A1*, and 4p16.1 near *RAF1P1* (Wang et al.,

2011). Results for two of the reported SNPs in these regions were available in the present study, but failed to support association at these loci (rs7819815 at 8q24.22,  $p = .103$ ; rs17407555 at 4p16.1,  $p = .616$ ).

Early AAO could be due to modifying genetic or environmental influences or increased number or penetrance of genetic risk factors. A test of polygenic risk burden between early and late onset groups, divided at an AAO of 21, revealed no significant differences, suggesting that increased genetic risk due to common variants is not a major contributor to variation in onset timing. It is possible, however, that defining early and late onset differently could alter the results. Differential onset timing by sex may be mediated by *GATA5*, but the mechanism by which this might occur is not immediately clear pending elucidation of the role of this gene in the brain.

Next, we conducted association analyses of disease severity. Here, the marker showing the greatest association was, rs12548481 ( $p = 8.24 \times 10^{-7}$ ), but this marker showed no association in the main case-control results ( $p = .86$ ). As in the AAO results, this points toward possible modifying properties without affecting risk *per se*. The regulatory functions of the nearest gene, *ST18*, on inflammation and apoptotic pathways may hold clues to the mechanism by which illness severity is genetically mediated.

Although men and women exhibit differences in the presentation of schizophrenia, the similar rates of disease between the sexes do not suggest the presence of sex-specific risk factors. However, some distinct genetic risk loci are possible, and the top hit in the male-only association analyses, *JMJD2C* ( $p = 8.07 \times 10^{-7}$ ), might be considered a plausible candidate considering its prior association to autism spectrum disorders which are predominantly diagnosed in males and share some genetic and phenotypic features with schizophrenia (Kantolarvi et al., 2010). In Psychiatric Genomics Consortium schizophrenia analyses, including the ISC samples, this marker showed no evidence of association in male-only GWAS results ( $p = .314$ ; Stephan Ripke, personal communication).

We also considered family history as a potential stratifying variable. Under some disease models, FH+ subjects would be expected to have increased genetic risk burden compared to subjects without a family history of disease, however, we did not detect any differences in polygenic risk scores between these groups. The familial-sporadic distinction is fraught with problems for highly polygenic disorders such as schizophrenia (Kendler, 1987). Most “sporadic” cases actually have inherited genetic risk factors circulating in their families (Yang et al., 2010). Nevertheless, family history status may offer some utility. Having a close relative with schizophrenia is currently the best predictor of this diagnosis. Furthermore, some genetic risk factors may be related to family history status. For example, single nucleotide *de novo* mutations appear enriched in sporadic cases of schizophrenia (Xu et al., 2011), and the CNVs previously associated with this illness are far more likely to arise *de novo* (Rees et al., 2011).

Uncovering the mechanisms by which genetic changes exert their behavioral effects is vital to understanding schizophrenia etiology and pathology. To that end, the pathway analyses revealing “intracellular signal transduction” as modulating AAO offer some insight.

Interestingly, one of the top genes driving the significance of this pathway, *RYR2*, encodes a calcium channel and was also a top hit in the prior GWAS of AAO for the female-only and AAO  $\times$  gender analyses (Wang et al., 2011). The “neurotransmitter metabolism” pathway, significant for the FH+ analyses, was driven by signals in several neurotransmitter systems.

We confirmed previously reported relationships for four widely-studied clinical variables of schizophrenia: AAO, severity, sex and family history. Earlier AAO and more severe disease course were observed in male subjects, and for all subjects, earlier AAO was associated with increased severity. A link between early-onset illness and increased severity has been observed for several psychiatric conditions such as obsessive compulsive disorder, (Taylor, 2011) bipolar disorder (Leverich et al., 2007), alcoholism (Hingson et al., 2006), and schizophrenia (Rabinowitz et al., 2006). A slightly lower AAO for FH+ subjects was not significant following correction for multiple tests, and no difference in disease severity was detected between FH+ and FH- groups. In a meta-analysis of 12 studies, earlier AAO in FH + subjects was observed (Esterberg et al., 2010), and several explanations for this relationship are possible. For example, an increase in genetic risk in familial cases could precipitate earlier onset. Alternatively, the emotional distress of having an affected family member could lead to an earlier psychotic break in susceptible individuals, or familial awareness of this disease could facilitate recognition of prodromal symptoms and expedite diagnosis.

In conclusion, we replicated prior observations between the variables under investigation, but no significant results emerged from the GWAS of AAO, severity or disease associations by FH or sex. For AAO and severity, greater power from larger sample sizes may yield significant results since these are almost certainly genetically mediated to some extent. Since illness severity may increase over time, duration of illness should be taken into account when possible. In a highly polygenic disorder such as schizophrenia, the familial/sporadic dichotomy has little meaning since most “sporadic” cases will have genetic risk factors circulating in their families. Sex-specific genetic risk factors for schizophrenia are theoretically possible, however, the lack of significant results by sex in this study supports the idea that no strong sex-specific genetic risk factors exist for schizophrenia. Therefore, a simplified genetic architecture for schizophrenia predicated on subgrouping by FH or sex is not supported.

With the exception of one GWAS of AAO in schizophrenia, this is the first reported study of genetic associations for these characteristics in schizophrenia on this scale, and additional work remains to be done. For example, further explorations of whether the proportion of variance explained by common SNPs depends on FH, age of onset or gender would be useful. Environmental influences modifying the manifestation or risk for schizophrenia also need to be incorporated into investigations of heterogeneity. Furthermore, distinct modifying influences – genetic and environmental – may exist in different populations, and exploring these questions in other samples would be a worthwhile endeavor. Interventional efforts delaying onset or reducing the severity of schizophrenia could profoundly impact the lives and families of individuals with this disorder. Since the vast heterogeneity observed in schizophrenia is at least partially genetically mediated, a better understanding of the specific



genetic factors impacting this variability is a vital step in developing new prevention and treatment strategies.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

### Role of funding source

This work was supported by a National Institute of Mental Health grant, R01 MH079799, to JWS, a Science Foundation Ireland grant to AC, and a National Alliance for Research on Schizophrenia and Depression (NARSAD) Young Investigator grant to SEB.

## References

- American Psychiatric Association. Diagnostic Criteria From DSM-IV-TR. American Psychiatric Association; Washington, D.C: 2000.
- Angermeyer MC, Kuhn L. Gender differences in age at onset of schizophrenia. An overview. *Eur Arch Psychiatry Neurol Sci.* 1988; 237(6):351–364. [PubMed: 3053193]
- Bergen SE, Petryshen TL. Genome-wide association studies of schizophrenia: does bigger lead to better results? *Curr Opin Psychiatry.* 2012; 25(2):76–82. [PubMed: 22277805]
- Bergen SE, O’Dushlaine CT, Ripke S, Lee PH, Ruderfer DM, Akterin S, Moran JL, Chambert KD, Handsaker RE, Backlund L, Osby U, McCarroll S, Landen M, Scolnick EM, Magnusson PK, Lichtenstein P, Hultman CM, Purcell SM, Sklar P, Sullivan PF. Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. *Mol Psychiatry.* 2012; 17:880–886. [PubMed: 22688191]
- Beyer K, Lao JI, Latorre P, Ariza A. Age at onset: an essential variable for the definition of genetic risk factors for sporadic Alzheimer’s disease. *Ann N Y Acad Sci.* 2005; 1057:260–278. [PubMed: 16399900]
- Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet.* 2007; 81(5):1084–1097. [PubMed: 17924348]
- Campbell DB, Lange LA, Skelly T, Lieberman J, Levitt P, Sullivan PF. Association of RGS2 and RGS5 variants with schizophrenia symptom severity. *Schizophr Res.* 2008; 101(1–3):67–75. [PubMed: 18262772]
- Chen L, Hsu L, Malone K. A frailty-model-based approach to estimating the age-dependent penetrance function of candidate genes using population-based case–control study designs: an application to data on the BRCA1 gene. *Biometrics.* 2009; 65(4):1105–1114. [PubMed: 19210733]
- Cohen RZ, Seeman MV, Gotowiec A, Kopala L. Earlier puberty as a predictor of later onset of schizophrenia in women. *Am J Psychiatry.* 1999; 156(7):1059–1064. [PubMed: 10401452]
- Purcell SM, Wray NR, Stone JL, Visscher PM, O’Donovan MC, Sullivan PF, Sklar P. International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature.* 2009; 460(7256):748–752. [PubMed: 19571811]
- Esterberg ML, Trotman HD, Holtzman C, Compton MT, Walker EF. The impact of a family history of psychosis on age-at-onset and positive and negative symptoms of schizophrenia: a meta-analysis. *Schizophr Res.* 2010; 120(1–3):121–130. [PubMed: 20303240]
- Fanous AH, Kendler KS. Genetic heterogeneity, modifier genes, and quantitative phenotypes in psychiatric illness: searching for a framework. *Mol Psychiatry.* 2005; 10(1):6–13. [PubMed: 15618952]
- Gorwood P, Leboyer M, Jay M, Payan C, Feingold J. Gender and age at onset in schizophrenia: impact of family history. *Am J Psychiatry.* 1995; 152(2):208–212. [PubMed: 7840353]

- Hafner H, Maurer K, Loffler W, Riecher-Rossler A. The influence of age and sex on the onset and early course of schizophrenia. *Br J Psychiatry*. 1993; 162(1):80–86. [PubMed: 8425144]
- Hare E, Glahn DC, Dassori A, Raventos H, Nicolini H, Ontiveros A, Medina R, Mendoza R, Jerez A, Munoz R, Almasy L, Escamilla MA. Heritability of age of onset of psychosis in schizophrenia. *Am J Med Genet B Neuropsychiatr Genet*. 2010; 153B(1):298–302. [PubMed: 19350535]
- Hingson RW, Heeren T, Winter MR. Age at drinking onset and alcohol dependence: age at onset, duration, and severity. *Arch Pediatr Adolesc Med*. 2006; 160(7):739–746. [PubMed: 16818840]
- Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, Leal SM, Pasternak S, Wheeler DA, Willis TD, Yu F, Yang H, Zeng C, Gao Y, Hu H, Hu W, Li C, Lin W, Liu S, Pan H, Tang X, Wang J, Wang W, Yu J, Zhang B, Zhang Q, Zhao H, Zhao H, Zhou J, Gabriel SB, Barry R, Blumenstiel B, Camargo A, Defelice M, Faggart M, Goyette M, Gupta S, Moore J, Nguyen H, Onofrio RC, Parkin M, Roy J, Stahl E, Winchester E, Ziaugra L, Altshuler D, Shen Y, Yao Z, Huang W, Chu X, He Y, Jin L, Liu Y, Shen Y, Sun W, Wang H, Wang Y, Wang Y, Xiong X, Xu L, Wayne MM, Tsui SK, Xue H, Wong JT, Galver LM, Fan JB, Gunderson K, Murray SS, Oliphant AR, Chee MS, Montpetit A, Chagnon F, Ferretti V, Leboeuf M, Olivier JF, Phillips MS, Roumy S, Sallee C, Verner A, Hudson TJ, Kwok PY, Cai D, Koboldt DC, Miller RD, Pawlikowska L, Taillon-Miller P, Xiao M, Tsui LC, Mak W, Song YQ, Tam PK, Nakamura Y, Kawaguchi T, Kitamoto T, Morizono T, Nagashima A, Ohnishi Y, Sekine A, Tanaka T, Tsunoda T, Deloukas P, Bird CP, Delgado M, Dermitzakis ET, Gwilliam R, Hunt S, Morrison J, Powell D, Stranger BE, Whittaker P, Bentley DR, Daly MJ, de Bakker PI, Barrett J, Chretien YR, Maller J, McCarroll S, Patterson N, Pe'er I, Price A, Purcell S, Richter DJ, Sabeti P, Saxena R, Schaffner SF, Sham PC, Varilly P, Altshuler D, Stein LD, Krishnan L, Smith AV, Tello-Ruiz MK, Thorisson GA, Chakravarti A, Chen PE, Cutler DJ, Kashuk CS, Lin S, Abecasis GR, Guan W, Li Y, Munro HM, Qin ZS, Thomas DJ, McVean G, Auton A, Bottolo L, Cardin N, Eyheramendy S, Freeman C, Marchini J, Myers S, Spencer C, Stephens M, Donnelly P, Cardon LR, Clarke G, Evans DM, Morris AP, Weir BS, Tsunoda T, Mullikin JC, Sherry ST, Feolo M, Skol A, Zhang H, Zeng C, Zhao H, Matsuda I, Fukushima Y, Macer DR, Suda E, Rotimi CN, Adebamowo CA, Ajayi I, Aniagwu T, Marshall PA, Nkwodimmah C, Royal CD, Leppert MF, Dixon M, Peiffer A, Qiu R, Kent A, Kato K, Niikawa N, Adewole IF, Knoppers BM, Foster MW, Clayton EW, Watkin J, Gibbs RA, Belmont JW, Muzny D, Nazareth L, Sodergren E, Weinstock GM, Wheeler DA, Yakub I, Gabriel SB, Onofrio RC, Richter DJ, Ziaugra L, Birren BW, Daly MJ, Altshuler D, Wilson RK, Fulton LL, Rogers J, Burton J, Carter NP, Clee CM, Griffiths M, Jones MC, McLay K, Plumb RW, Ross MT, Sims SK, Willey DL, Chen Z, Han H, Kang L, Godbout M, Wallenburg JC, L'Archeveque P, Bellemare G, Saeki K, Wang H, An D, Fu H, Li Q, Wang Z, Wang R, Holden AL, Brooks LD, McEwen JE, Guyer MS, Wang VO, Peterson JL, Shi M, Spiegel J, Sung LM, Zacharia LF, Collins FS, Kennedy K, Jamieson R, Stewart J. International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature*. 2007; 449(7164):851–861. [PubMed: 17943122]
- International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*. 2008; 455(7210):237–241. [PubMed: 18668038]
- Janca A, Ustun TB, Early TS, Sartorius N. The ICD-10 symptom checklist: a companion to the ICD-10 classification of mental and behavioural disorders. *Soc Psychiatry Psychiatr Epidemiol*. 1993; 28(5):239–242. [PubMed: 8284737]
- Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*. 2008; 24(24):2938–2939. [PubMed: 18974171]
- Kantojarvi K, Onkamo P, Vanhala R, Alen R, Hedman M, Sajantila A, Nieminen-von Wendt T, Jarvela I. Analysis of 9p24 and 11p12–13 regions in autism spectrum disorders: rs1340513 in the JMJD2C gene is associated with ASDs in Finnish sample. *Psychiatr Genet*. 2010; 20(3):102–108. [PubMed: 20410850]
- Kendler KS. Sporadic vs familial classification given etiologic heterogeneity: I. Sensitivity, specificity, and positive and negative predictive value. *Genet Epidemiol*. 1987; 4(5):313–330. [PubMed: 3692132]
- Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, Hubbell E, Veitch J, Collins PJ, Darvishi K, Lee C, Nizzari MM, Gabriel SB, Purcell S, Daly MJ, Altshuler D. Integrated

- genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet.* 2008; 40(10):1253–1260. [PubMed: 18776909]
- Lee PH, O’Dushlaine C, Thomas B, Purcell SM. INRICH: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics.* 2012; 28(13):1797–1799. [PubMed: 22513993]
- Leverich GS, Post RM, Keck PE Jr, Altshuler LL, Frye MA, Kupka RW, Nolen WA, Suppes T, McElroy SL, Grunze H, Denicoff K, Moravec MK, Luckenbaugh D. The poor prognosis of childhood-onset bipolar disorder. *J Pediatr.* 2007; 150(5):485–490. [PubMed: 17452221]
- Lewine RR. Sex differences in age of symptom onset and first hospitalization in schizophrenia. *Am J Orthopsychiatry.* 1980; 50(2):316–322. [PubMed: 7361879]
- Lichtenstein P, Yip BH, Bjork C, Pawitan Y, Cannon TD, Sullivan PF, Hultman CM. Common genetic determinants of schizophrenia and bipolar disorder in Swedish families: a population-based study. *Lancet.* 2009; 373(9659):234–239. [PubMed: 19150704]
- Lips ES, Cornelisse LN, Toonen RF, Min JL, Hultman CM, Holmans PA, O’Donovan MC, Purcell SM, Smit AB, Verhage M, Sullivan PF, Visscher PM, Posthuma D. the International Schizophrenia C. Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. *Mol Psychiatry.* 2012; 17(10):996–1006. [PubMed: 21931320]
- McGuffin P, Farmer A, Harvey I. A polydiagnostic application of operational criteria in studies of psychotic illness. Development and reliability of the OPCRIT system. *Arch Gen Psychiatry.* 1991; 48(8):764–770. [PubMed: 1883262]
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; 81(3):559–575. [PubMed: 17701901]
- Rabinowitz J, Levine SZ, Hafner H. A population based elaboration of the role of age of onset on the course of schizophrenia. *Schizophr Res.* 2006; 88(1–3):96–101. [PubMed: 16962742]
- Rees E, Moskvina V, Owen MJ, O’Donovan MC, Kirov G. De novo rates and selection of schizophrenia-associated copy number variants. *Biol Psychiatry.* 2011; 70(12):1109–1114. [PubMed: 21855053]
- R Development Core Team. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2012.
- Schizophrenia Psychiatric Genome-Wide Association Study Consortium. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet.* 2011; 43(10):969–976. [PubMed: 21926974]
- Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry.* 2003; 60(12):1187–1192. [PubMed: 14662550]
- Taylor S. Early versus late onset obsessive–compulsive disorder: evidence for distinct subtypes. *Clin Psychol Rev.* 2011; 31(7):1083–1100. [PubMed: 21820387]
- Wang KS, Liu X, Zhang Q, Aragam N, Pan Y. Genome-wide association analysis of age at onset in schizophrenia in a European–American sample. *Am J Med Genet B Neuropsychiatr Genet.* 2011; 156B(6):671–680. [PubMed: 21688384]
- Webb R, Kelly JA, Somers EC, Hughes T, Kaufman KM, Sanchez E, Nath SK, Bruner G, Alarcon-Riquelme ME, Gilkeson GS, Kamen DL, Richardson BC, Harley JB, Sawalha AH. Early disease onset is predicted by a higher genetic risk for lupus and is associated with a more severe phenotype in lupus patients. *Ann Rheum Dis.* 2011; 70(1):151–156. [PubMed: 20881011]
- Williams J, Farmer AE, Ackenheil M, Kaufmann CA, McGuffin P. A multicentre inter-rater reliability study using the OPCRIT computerized diagnostic system. *Psychol Med.* 1996; 26(4):775–783. [PubMed: 8817712]
- Wissmann M, Yin N, Muller JM, Greschik H, Fodor BD, Jenuwein T, Vogler C, Schneider R, Gunther T, Buettner R, Metzger E, Schule R. Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. *Nat Cell Biol.* 2007; 9(3):347–353. [PubMed: 17277772]
- Wright GE, Niehaus DJ, van der Merwe L, Koen L, Korkie LJ, Kinnear CJ, Drogemoller BI, Warnich L. Association of MB-COMT polymorphisms with schizophrenia-susceptibility and symptom severity in an African cohort. *Prog Neuropsychopharmacol Biol Psychiatry.* 2012; 39(1):163–169. [PubMed: 22705295]

- Xu B, Roos JL, Dexheimer P, Boone B, Plummer B, Levy S, Gogos JA, Karayiorgou M. Exome sequencing supports a de novo mutational paradigm for schizophrenia. *Nat Genet.* 2011; 43(9): 864–868. [PubMed: 21822266]
- Yang J, Siqueira MF, Behl Y, Alikhani M, Graves DT. The transcription factor ST18 regulates proapoptotic and proinflammatory gene expression in fibroblasts. *FASEB J.* 2008; 22(11):3956–3967. [PubMed: 18676404]
- Yang J, Visscher PM, Wray NR. Sporadic cases are the norm for complex disease. *Eur J Hum Genet.* 2010; 18(9):1039–1043. [PubMed: 19826454]

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.schres.2014.01.030>.

Table 1

Sample characteristics by phenotype and site.

Phenotype	Collection site							Total
	Aberdeen	Cardiff	Dublin	Edinburgh	Portugal	London		
Total cases <sup>a</sup>	693	528	178	358	257	523	2537	
Age at onset	662	519	174	302	250	480	2387	
Mean (sd)	24.2 (8.2)	24.3 (7.4)	24.5 (8.1)	24.0 (8.1)	23.0 (6.8)	23.6 (7.6)	24	
Range	8 to 62	13 to 52	14 to 64	10 to 55	13 to 53	8 to 60	8 to 64	
Severity, total	668	457	178	358	167	462	2290	
Mild/chronic	25/643	222/235	118/60	4/354 <sup>b</sup>	11/156	120/342	496/1436	
Sex, total	1422	1139	1141	656	563	1028	5949	
Males/females	985/437	562/577	449/692	416/240	294/269	585/443	3291/2658	
Cases/controls	720/702	528/611	275/866	369/287	347/216	523/505	2762/3187	
Family history	612	449	177	235	257	458	2188	
% FH+	0.17	0.18	0.22	0.65 <sup>c</sup>	0.19	0.22	0.24	

<sup>a</sup>Cases with phenotypic information beyond genetic sex.<sup>b</sup>Removed from analysis due to lack of variance.<sup>c</sup>Preferentially ascertained on the basis of positive family history.

**Table 2**

Regions of association for all GWAS analyses with  $p < 1 \times 10^{-5}$ . Allele = tested allele for each SNP, alphabetically ordered in the imputed data; Beta values are reported for the AAO results and ORs (odds ratios) for all other results; N = number of SNPs in associated region; KB = number of kilobases encompassed by the associated region.

SNP	Allele	Beta/OR	N	P	Chr	Position	KB	Genes
<i>Age at onset (AAO)</i>								
rs11999864	A	-0.126	22	$1.52 \times 10^{-7}$	9	112,956,770-113,109,094	152	OR2K2
rs8010374	C	-0.119	17	$1.89 \times 10^{-6}$	14	34,040,125-34,146,736	107	SNX6, EAPP, C14orf147
rs375621	C	-0.296	8	$3.28 \times 10^{-6}$	21	45,335,839-45,352,080	16	ADARB1
rs2310739	A	0.098	25	$3.76 \times 10^{-6}$	5	105,825,590-105,887,483	62	no genes
rs11847024	A	0.083	78	$4.68 \times 10^{-6}$	14	95,096,203-95,176,040	80	TCL6, TCLIB, GLRX5
rs218106	C	0.076	35	$5.81 \times 10^{-6}$	7	31,437,434-31,484,547	47	CCDC129
rs9551900	A	-0.077	10	$7.43 \times 10^{-6}$	13	29,843,921-29,869,421	26	No genes
rs6449781	C	0.073	72	$7.44 \times 10^{-6}$	5	64,529,388-64,797,072	268	ADAMTS6
rs2493275	A	0.190	1	$7.52 \times 10^{-6}$	1	3,297,894-3,297,894	0	PRDM16
<i>Severity</i>								
rs12548481	A	2.282	43	$8.24 \times 10^{-7}$	8	53,094,129-53,172,107	78	ST18
rs12526194	C	2.309	72	$4.55 \times 10^{-6}$	6	53,689,464-53,939,420	250	LRRCL1
rs11635207	A	2.660	13	$7.38 \times 10^{-6}$	15	64,258,249-64,658,268	400	ZWILCH, TIPIN, SNAPC5, RPL4, MEGF11, MAP2K1, LCTL, DIS3L
rs1820124	A	0.438	19	$9.12 \times 10^{-6}$	5	56,766,448-56,874,876	108	ACTBL2
<i>Males</i>								
rs7037878	C	1.702	36	$8.07 \times 10^{-7}$	9	7,125,540-7,182,313	57	JMJD2C
rs10929935	C	0.586	18	$3.47 \times 10^{-6}$	2	149,558,655-149,622,402	64	LOC130576, KIF5C
rs17106652	C	3.084	19	$6.82 \times 10^{-6}$	11	106,637,571-106,845,912	208	CWF19L2, ALKBH8
rs7440402	C	0.577	24	$6.92 \times 10^{-6}$	4	24,417,416-24,497,630	80	SOD3, CCDC149
rs11124307	C	1.638	5	$7.08 \times 10^{-6}$	2	33,189,351-33,199,797	10	LTBP1
rs6536692	A	1.641	43	$7.93 \times 10^{-6}$	4	164,018,560-164,049,966	31	No genes
rs6573239	A	0.410	36	$8.24 \times 10^{-6}$	14	58,285,994-58,631,695	346	No genes
rs4711338	C	1.608	45	$8.27 \times 10^{-6}$	6	33,719,329-33,826,396	107	MLN, LEMD2, ITPK3, IHPK3, C6orf125
<i>Females</i>								

SNP	Allele	Beta/OR	N	P	Chr	Position	KB	Genes
rs17115481	A	1.944	142	$1.90 \times 10^{-6}$	5	153,170,191–153,516,339	346	MFAP3, GRIAL1, GALNT10, FAM114A2
rs10430786	C	0.556	64	$2.24 \times 10^{-6}$	10	16,034,396–16,104,140	70	No genes
rs2038521	C	0.523	60	$4.38 \times 10^{-6}$	9	17,620,646–17,725,707	105	SH3GL2
rs802457	A	0.488	47	$8.20 \times 10^{-6}$	7	86,040,601–86,293,472	253	GRM3
rs2895202	C	0.554	73	$8.70 \times 10^{-6}$	7	1,878,583–2,239,462	361	SNX8, NUDT1, MAD1L1, FTS2
<i>Family history negative (FH-)</i>								
rs11165690	C	1.875	38	$9.96 \times 10^{-8}$	1	96,712,707–97,219,771	507	PTBP2
rs747709	A	1.588	14	$2.07 \times 10^{-6}$	2	43,179,074–43,212,565	33	No genes
rs7517755	C	0.470	1	$4.85 \times 10^{-6}$	1	151,063,073–151,063,073	0	LCE6A, LCE1D, LCE1F, LCE1E, LCE1B, LCE1C, LCE1A
rs2039521	A	0.475	15	$5.19 \times 10^{-6}$	6	8,005,533–8,175,770	170	TXNDC5, MUTED, EEF1E1
rs11959425	A	0.594	12	$5.39 \times 10^{-6}$	5	109,564,482–109,609,016	45	No genes
<i>Family history positive (FH+)</i>								
rs10929935	C	0.400	16	$1.96 \times 10^{-8}$	2	149,558,655–149,614,988	56	LOC130576, KIF5C
rs10094997	C	2.703	11	$1.25 \times 10^{-7}$	8	114,864,190–114,922,054	58	No genes
rs884888	C	1.953	6	$2.86 \times 10^{-6}$	9	100,342,898–100,369,715	27	GABBR2
rs4515592	C	0.166	24	$4.24 \times 10^{-6}$	8	65,237,815–65,379,408	142	No genes
rs7314262	A	2.184	23	$6.21 \times 10^{-6}$	12	55,608,759–55,674,465	66	ZBTB39, TAC3, SDR-O, RDH16, GPR182
rs12253911	C	3.015	11	$7.28 \times 10^{-6}$	10	55,209,960–55,260,244	50	PCDH15
rs10930016	C	2.168	2	$7.79 \times 10^{-6}$	2	149,620,796–149,622,402	2	LOC130576
rs748262	C	0.256	15	$8.37 \times 10^{-6}$	19	16,331,677–16,537,327	206	SLC35E1, MED26, EPS15L1, CHERP, CALR3, C19orf44
rs7692151	C	4.416	3	$9.62 \times 10^{-6}$	4	6,530,962–6,549,557	19	PPP2R2C