



The impact of clinical heterogeneity in schizophrenia on genomic analyses



Sherri G. Liang, Tiffany A. Greenwood *

Department of Psychiatry, University of California San Diego, La Jolla, CA, United States

ARTICLE INFO

Article history:

Received 19 August 2014

Received in revised form 12 November 2014

Accepted 17 November 2014

Available online 10 December 2014

Keywords:

Genome-wide association

Schizophrenia

Heterogeneity

Diagnostics

ABSTRACT

Though clinically useful, the diagnostic systems currently employed are not well equipped to capture the substantial clinical heterogeneity observed for most psychiatric disorders, as exemplified by the complex psychotic disorder(s) that Bleuler aptly labeled the “Group of Schizophrenias”. The clinical heterogeneity associated with schizophrenia has likely frustrated decades of attempts to illuminate the underlying genetic architecture, although recent genome-wide association studies have begun to provide valuable insight into the role of common genetic risk variants. Here we demonstrate the importance of using diagnostic information to identify a core form of the disorder and to eliminate potential comorbidities in genetic studies. We also demonstrate why applying a diagnostic screening procedure to the control dataset to remove individuals with potentially related disorders is critical. Additionally, subjects may participate in multiple studies at different institutions or may have genotype data released by more than one research group. It is thus good practice to verify that no identical subjects exist within or between samples prior to conducting any type of genetic analysis to avoid potential confounding of results. While the availability of genomic data for large collections of subjects has facilitated many investigations that would otherwise not have been possible, we clearly show why one must use caution when acquiring data from publicly available sources. Although the broad vs. narrow debate in terms of phenotype definition in genetic analyses will remain, it is likely that both approaches will yield different results and that both will have utility in resolving the genetic architecture of schizophrenia.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Schizophrenia (SZ) is a severe psychiatric disorder characterized by abnormalities in a patient's thoughts, perceptions, and behaviors, manifesting as hallucinations, delusions, and/or disorganized speech with significant social or occupational dysfunction (Andreasen, 1995). The substantial clinical heterogeneity associated with SZ, which Bleuler perhaps more appropriately labeled the “Group of Schizophrenias” (Bleuler, 1911), has likely combined with the inherent genetic heterogeneity to plague many attempts at identifying casual genetic variants (Karayiorgou and Gogos, 1997; Owen et al., 2007; Schork et al., 2007; Sanders et al., 2008). Although genome-wide association studies (GWAS) of increasingly large samples have finally begun to overcome this heterogeneity to provide valuable insight into the role of common genetic variants in SZ risk (O'Donovan et al., 2008; Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium, 2011; Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014; Shi et al., 2009), investigations of smaller samples

may suffer unnecessary power losses if the clinical heterogeneity is not appropriately accommodated.

Here we demonstrate the importance of using specific diagnostic criteria to identify the core features of psychiatric disorders in genetic studies. We further show why screening the control population for disorders genetically related to the disorder of interest may be critical to the success of the study. For this purpose, we use data from the subset of the Molecular Genetics of Schizophrenia (MGS) that was genotyped as part of the Genetic Association Information Network (GAIN). Finally, we emphasize that one must use caution when acquiring data from publicly available sources and always verify that no identical subjects exist within or between samples prior to conducting any type of genetic analysis to avoid the potential confounding of results.

2. Methods

All data for these analyses were obtained from the database of Genotypes and Phenotypes (dbGaP) (<http://www.ncbi.nlm.nih.gov/gap>). The Genome-Wide Association Study of Schizophrenia (SZ GAIN) was obtained via accession number phs000021.v3.p2 and contained 1343 cases and 1368 controls of European ancestry. Cases in this sample met criteria for SZ or schizoaffective disorder (SA) per the DSM-IV (American Psychiatric Association, 2000). Controls were screened

* Corresponding author at: Department of Psychiatry, University of California San Diego, 9500 Gilman Drive, MC 0689, La Jolla, CA 92093, United States. Tel.: +1 858 246 1897.

E-mail address: tgreenwood@ucsd.edu (T.A. Greenwood).

briefly and only excluded from the original sample if they endorsed a history of SZ, SA, or bipolar disorder (BP). The SZ GAIN sample was genotyped together with the Whole Genome Association Study of Bipolar Disorder (BP GAIN) sample with a common set of controls through collaboration with the Genetic Association Information Network (GAIN) for the Affymetrix Genomewide Human SNP Array 6.0 at the Broad Institute. An initial round of genotype processing and quality control was performed to remove subjects with genotype call rate <98.5% and SNPs with genotype call rates <95%, Hardy Weinberg Equilibrium p values < 10^{-6} in the controls, and minor allele frequencies <0.01 in the combined sample, resulting in the release of the filtered subjects above and 729,454 SNPs to dbGaP. Detailed information regarding the extensive quality control process and the clinical characteristics of this sample is provided elsewhere (Sanders et al., 2008; Shi et al., 2009), although we note that the SZ GAIN was ultimately published as a combined analysis with the complete MGS sample, rather than as an independent study (Shi et al., 2009).

Since we analyzed only a subset of the originally genotyped samples (i.e., the SZ cases and controls), we reapplied similar quality control thresholds, resulting in 724,067 SNPs for analysis. We note that initial quality control thresholds were applied to a subset that excluded 335 control subjects with recurrent major depression (MDD-R) as a primary hypothesis, consistent with the BP GAIN (Smith et al., 2009), and these subjects were later added for the evaluation of clinical heterogeneity in controls. Identity by descent (IBD) was calculated for all pairs of individuals to identify potential cryptic relatedness, and the genetic

homogeneity of the sample was assured by multidimensional scaling (MDS). No related individuals or population outliers were detected. Association analyses were conducted using logistic regression, and the minimal genomic inflation identified was accommodated through the use of a genomic control correction of the p values. All analyses were performed in PLINK v.1.07 or v.109 (Purcell et al., 2007).

3. Results

3.1. Primary analysis of the SZ GAIN sample

The analysis of the complete SZ GAIN sample of 1343 cases and 1368 controls is provided for comparison in Fig. 1A. The best result was found for rs11789399 on chromosome 9q33.1 with a p value of 3.0×10^{-6} . Another SNP, rs11789407, located just 260 bp from rs11789399 on 9q33.1 and in nearly complete linkage disequilibrium with it produced a p value of 5.3×10^{-6} . The best independent findings were for rs12745968 on chromosome 1p22 ($p = 3.2 \times 10^{-6}$) and for rs271876 on chromosome 6q21 ($p = 4.5 \times 10^{-6}$).

3.2. Evaluation of the effect of clinical heterogeneity in the controls

Although epidemiological and family studies indicate a genetic overlap between MDD-R and SZ (Maier et al., 1993; Somnath et al., 2002), the SZ GAIN sample includes a substantial number ($N = 335$) of control subjects with a history of MDD-R (Sanders et al., 2008; Shi et al., 2009).

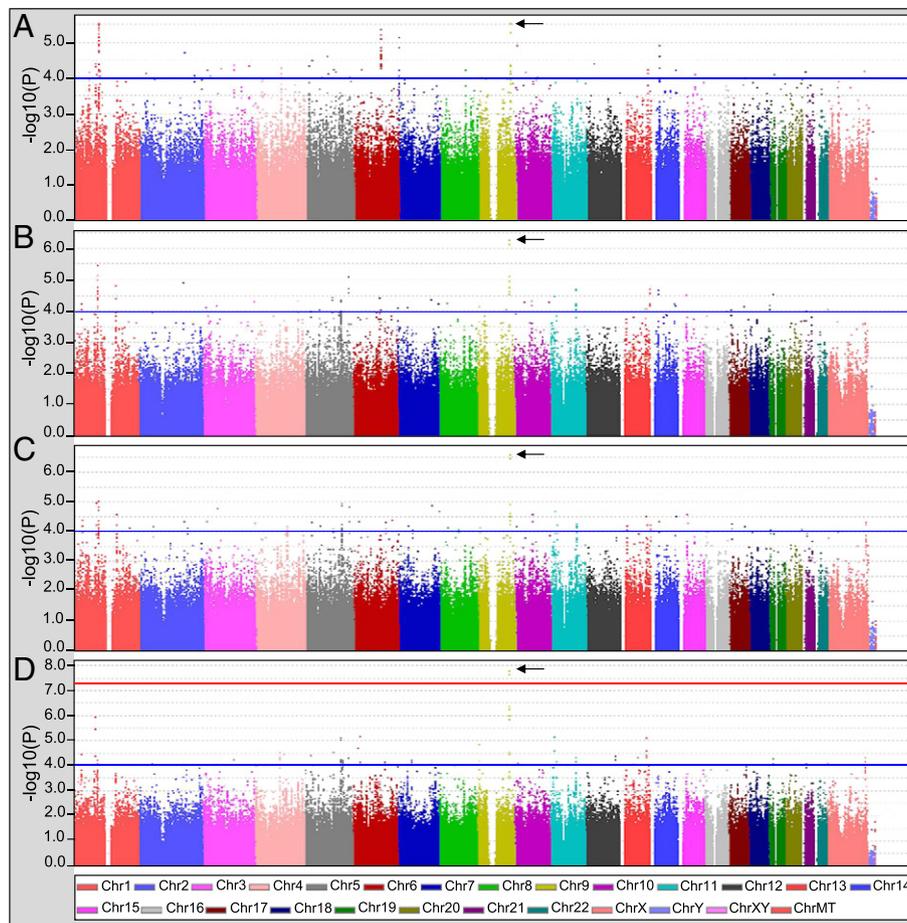


Fig. 1. Genome-wide association results for the SZ GAIN sample demonstrating the effects of clinical heterogeneity in both cases and controls. (A) All cases ($N = 1343$) vs. all controls ($N = 1368$) following genotyping quality control procedures only. (B) All cases ($N = 1343$) vs. controls screened for depression ($N = 1033$). (C) Cases screened for SA ($N = 1191$) vs. controls screened for depression ($N = 1033$). (D) Cases screened for SA and secondary mood diagnoses (see Table 1, $N = 814$) vs. controls screened for depression ($N = 1033$). In each panel, a blue line indicates $p < 10^{-4}$, a red line indicates genome-wide significance at $p = 5 \times 10^{-8}$, and an arrow indicates the peak SNP, rs11789399.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات