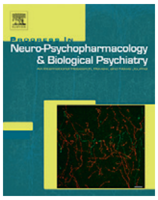
Progress in Neuro-Psychopharmacology & Biological Psychiatry 51 (2014) 140–145

Contents lists available at ScienceDirect

Progress in Neuro-Psychopharmacology & Biological Psychiatry

journal homepage: www.elsevier.com/locate/pnp



A hypothesis-driven pathway analysis reveals myelin-related pathways that contribute to the risk of schizophrenia and bipolar disorder

Hao Yu [a,b,1], Wenjian Bi [c,1], Chenxing Liu [d], Yanlong Zhao [c], Dai Zhang [a,b,d,e], Weihua Yue [a,b,*]

[a] *Institute of Mental Health, Peking University, 51 Hua Yuan Bei Road, Beijing 100191, China*
[b] *Key Laboratory of Mental Health, Ministry of Health, Institute of Mental Health, The Sixth Hospital, Peking University, China*
[c] *Academy of Mathematics and Systems Science, Chinese Academy Of Sciences, Beijing, China*
[d] *Peking-Tsinghua Center for Life Sciences, Beijing 100871, China*
[e] *PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing, China*

ARTICLE INFO

Article history:
Received 15 September 2013
Received in revised form 3 January 2014
Accepted 6 January 2014
Available online 19 January 2014

Keywords:
Bipolar disorder
Genome-wide association study
Myelin
Pathway-based analysis
Schizophrenia

ABSTRACT

Schizophrenia (SZ) and bipolar disorder (BD) are both severe neuropsychiatric disorders with a strong and potential overlapping genetic background. Multiple lines of evidence, including genetic studies, gene expression studies and neuroimaging studies, have suggested that both disorders are closely related to myelin and oligodendrocyte dysfunctions. In the current study, we hypothesized that the holistic effect of the myelin-related pathway contributes to the genetic susceptibility to both SZ and BD. We extracted pathway data from the canonical pathway database, Gene Ontology (GO), and selected a 'compiled' pathway based on previous literature. We then performed hypothesis-driven pathway analysis on GWAS data from the Psychiatric Genomics Consortium (PGC). As a result, we identified three myelin-related pathways with a joint effect significantly associated with both disorders: 'Myelin sheath' pathway ($P_{SZ} = 2.45E-7$, $P_{BD} = 1.22E-3$), 'Myelination' pathway ($P_{SZ} = 2.10E-4$, $P_{BD} = 2.53E-24$), and 'Compiled' pathway ($P_{SZ} = 4.57E-8$, $P_{BD} = 2.61E-9$). In comparing the SNPs and genes in these three pathways across the two diseases, we identified a substantial overlap in nominally associated SNPs and genes, which could be susceptibility SNPs and genes for both disorders. From these observations, we propose that myelin-related pathways may be involved in the etiologies of both SZ and BD.

1. Introduction

Schizophrenia (SZ) and bipolar disorder (BD) are common and complex psychiatric disorders affecting ~1.8% of the population worldwide (Sklar, 2002). Although the etiologies of the two disorders are unknown, family and twin studies have shown that they are both highly heritable neuropsychiatric disorders. Heritability for SZ and BD is 64% and 59%, respectively (Lichtenstein et al., 2009). Multiple lines of evidence have suggested that both disorders are closely related, including epidemiology studies (Lichtenstein et al., 2009), genetic studies (Craddock et al., 2006; Moskvina et al., 2009), and neuroimaging studies (Argyelan et al., 2013; Sussmann et al., 2009; Wang et al., 2013). However, the proof of the involvement of specific genes and variants remains elusive.

In the last dozen years, genetic studies have demonstrated the overlap between SZ and BD. In genetic linkage studies, two regions (18p11 and 22q11) of the human genome have been found to be implicated in both psychotic disorders (Berrettini, 2001). Recently, researchers have also used genome-wide association studies (GWAS) to successfully identify several susceptibility loci for the two diseases (Ripke et al., 2011; Shi et al., 2009; Yue et al., 2011), including some that are associated with both disorders (Purcell et al., 2009; Smoller et al., 2013; Steinberg et al., 2012). The traditional GWAS typically investigates the genetic effect of one single-nucleotide polymorphism (SNP) at a time and focuses on finding the strongest SNP that meets the genome-wide significance cutoff p-value of 5×10^{-8} for detecting significant markers. Due to the multifactorial-threshold model of genetic liability, it is unlikely that any single SNP or gene is sufficient to explain the etiology of these two complex diseases; the holistic effect of a gene set is

---

Abbreviations: PIK3C2G, phosphatidylinositol-4-phosphate 3-kinase, catalytic subunit type 2 gamma; PIK3R1, phosphoinositide-3-kinase, regulatory subunit 1 (alpha); INPP1, inositol polyphosphate-1-phosphatase; INPP4B, inositol polyphosphate-4-phosphatase, type II, 105kDa; INPP5D, inositol polyphosphate-5-phosphatase, 145kDa; PLCG1, phospholipase C, gamma 1; PLCB1, phospholipase C, beta 1 (phosphoinositide-specific); KGG, Knowledge-based mining system for Genome-wide Genetic studies; PIK4A, phosphatidylinositol 4-kinase, catalytic, alpha; PIK3C2B, phosphatidylinositol-4-phosphate 3-kinase, catalytic subunit type 2 beta; PIK3C3, phosphatidylinositol 3-kinase, catalytic subunit type 3; PIP4K2B, phosphatidylinositol-5-phosphate 4-kinase, type II, beta; PLCE1, phospholipase C, epsilon 1; PLP1, proteolipid protein 1; MAG, myelin-associated glycoprotein; ERBB3, v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 3; ERBB4, v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 4; MAL, mal, T-cell differentiation protein; TF, transferrin; GSN, gelsolin; CNP, 2′,3′-cyclic nucleotide 3′ phosphodiesterase; QKI, QKI, KH domain containing, RNA binding; BDNF, brain-derived neurotrophic factor; OMG, oligodendrocyte myelin glycoprotein; FGFR1, fibroblast growth factor 1 (acidic); ARHGEF10, Rho guanine nucleotide exchange factor (GEF) 10; PSAP, prosaposin; MBP, myelin basic protein; PMP22, peripheral myelin protein 22; UGT8, UDP glycosyltransferase 8; NRG1, neuregulin 1; RXRG, retinoid X receptor, gamma; NGF, nerve growth factor (beta polypeptide).

[*] Corresponding author at: Institute of Mental Health, Peking University, 51 Hua Yuan Bei Road, Beijing 100191, China. Tel.: +86 10 8280 5307; fax: +86 10 6201 7114.
E-mail address: dryue@bjmu.edu.cn (W. Yue).

[1] Yu H and Bi WJ contributed to this work equally.

0278-5846/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.pnpbp.2014.01.006

expected to have an effect larger than the individual effects of each gene. Therefore, pathway-based analytical (PBA) methods have been developed to complement the GWAS data in identifying candidate genes and providing functional links to bridge the knowledge gap between the genetic variants and the phenotypes (Sun, 2012; Wang et al., 2010). Combining prior biological knowledge with the results from GWAS, PBA methods can assess whether a group of genes or pathways with related functions are jointly associated with a trait of interest and generate specific hypotheses for follow-up experimental studies. By using PBA approaches, researchers have demonstrated the biological connection between ion channels and SZ (Askland et al., 2012) and found that the cell adhesion molecule pathway contributes to SZ and BD susceptibility (O'Dushlaine et al., 2011). Although PBA methods have been widely used for exploratory analyses of GWAS datasets, they have less commonly been employed for hypothesis-driven analyses, which have several advantages (Askland et al., 2012). First, when complex diseases are studied, it is important to manually compile candidate pathways using expert knowledge, as public databases may not have well-annotated pathways of biological function, especially for neuropsychiatric diseases (Wang et al., 2010). Hypothesis testing could test and verify the association between compiled pathways and diseases. Second, with a priori hypothesis, interpretation of the results relies on scientific deduction rather than speculation. Finally, hypothesis-driven analysis frees the researcher from the burden of extensive hypothesis testing correction.

Over recent years, the importance of myelin pathology in SZ and BD has become widely recognized. Myelin provides the basis for rapid impulse conduction in the central nervous system. It acts as electrical insulation for the ensheathed axon, which both helps to preserve the amplitude and increase the conduction velocity of the propagating axon potential. Given its role as the primary infrastructure for long-distance communication in the brain, it is perhaps not surprising that damage to the myelin dysfunction has been implicated in both SZ and BD (Whitford et al., 2012). Temporal onset of psychotic disorders in late adolescence or early adulthood coincides with the concluding myelination of the prefrontal cortex (Lebel et al., 2008; Tamnes et al., 2010). Additionally, multiple lines of evidence have suggested that myelin and oligodendrocyte dysfunction underlie the neurobiology of both SZ and BD. Imaging studies have found that white matter abnormalities in both disorders are associated with abnormal myelination (Du et al., 2013; Lagopoulos et al., 2013). Gene expression studies have demonstrated that several myelin-related genes are differentially expressed in the brain tissue of patients compared to that of unaffected controls (Barley et al., 2009; Tkachev et al., 2003) and genetic studies have identified several myelin-related genes that are associated with both diseases (Ayalew et al., 2012; Jitoku et al., 2011; Jungerius et al., 2008). However, most of the genetic studies have focused on the association of a single myelin-related gene, so the combined genetic effects of genes in myelin-related pathways on SZ and BD epidemiology have gone largely unexamined. To fill in this gap in the previous work, the current study performed pathway analysis.

The goal of this study was to test whether myelin-related pathways were critical to both SZ and BD. In the present study, we first hypothesized that the holistic effect of myelin-related genes contributes to the genetic susceptibility to both SZ and BD. Then, to verify our hypothesis, we selected three pathways with various myelin-related functions using the current literature and bioinformatics database searching. Finally, we applied the hybrid gene set-based test (HYST), a PBA method, to analyze SZ and BD GWAS datasets respectively.

2. Methods

2.1. GWAS data

The Psychiatric Genomics Consortium (PGC) (pgc.unc.edu/index.php), which was established in order to conduct meta-analyses of genome-wide genetic data for psychiatric disease, has identified many robust and replicable associations. In addition, PGC has been committed to full and open sharing of data and results since its inception. The current study used PGC GWAS datasets for SZ and BD. The SZ meta-analysis combined GWAS data from 17 separate studies (with a total of 9394 cases and 12,462 controls). The analysis provided data on 1,252,901 autosomal SNPs, with λ_{1000} equals 1.02, a metric that standardized the degree of inflation by sample size. The samples included in the analysis were all from unrelated European populations (Ripke et al., 2011). The BD GWAS data included 7481 cases and 9250 controls. This study analyzed imputed SNP dosages from 2,415,422 autosomal SNPs, with a genomic control inflation factor of 1.148. The samples included in the analysis were also all from unrelated European populations (Sklar et al., 2011). Complete details of two GWAS datasets were provided elsewhere.

2.2. Selection of myelin-related pathway

We extracted pathway data from the canonical pathway database, Gene Ontology (GO) (www.geneontology.org), and selected gene product associations to ‘Myelin sheath’ (GO: 0043209), ‘Myelination’ (GO: 0042552), and their subsets. In addition, based on a search of the current literature, we also selected a set of 138 myelin-related genes compiled by Jungerius et al. (2008). These genes have various myelin-related functions, including structural, compositional, developmental, and maintenance functions. For convenience, we refer to this pathway as the ‘compiled’ pathway.

2.3. Hybrid gene set-based test (HYST)

In the current study, we used a PBA approach proposed by Li et al. (2012), in which the HYST is conducted with the Java software that they developed, called ‘KGG’. KGG offers a couple key advantages to our study. First, HYST uses all SNPs in a pathway and combines the SNP-related biological knowledge with the p-values of single marker analysis to produce optimal weights, which could maximize the potential power of pathway-based analysis while controlling false positive discoveries. Second, HYST is versatile and able to quickly produce valid pathway-based p-values without time-consuming permutation tests.

The HYST consisted of three main steps. First, the SNPs were mapped to their respective genes based on their coordinates from the University of California Santa Cruz (UCSC) database. When mapping, we also included the regions 5 kb upstream and 5 kb downstream of each gene to account for variants in potential gene control regions. Second, for each gene, a gene-level statistics were computed using the improved Simes test formula (Li et al., 2011). This step constructed weights for the gene-based tests according to a priori knowledge of a particular set of SNPs found to be associated with each disease, using the categorical information of available GWAS hits in the Human Genome Epidemiology (HuGE) Navigator literature database. Third, the pathway-level statistics were calculated by combining all gene-based p-values for association.

2.4. Gene expression profiling analysis

To investigate whether the genes involved in the three pathways are expressed in the brain, we assessed the genes with the online Gene Enrichment Profiler (http://xavierlab2.mgh.harvard.edu/EnrichmentProfiler/) (Benita et al., 2010). This profiler computed the expression and enrichment of any set of query genes on the basis of a reference set obtained from 126 normal tissues and cell types. In this dataset, the gene expression intensities have been converted to enrichment scores, which reflect the enrichment of a gene based on its expression in all tissues. For each tissue, the enrichment scores of the genes in the three pathways were compared to the