Schizophrenia Research xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Schizophrenia Research

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

Regional enrichment analyses on genetic profiles for schizophrenia and bipolar disorder

Jingyu Liu [a,b,*], Jiayu Chen [a], Nora I. Perrone-Bizzozero [c], Jessica A. Turner [a,d], Vince D. Calhoun [a,b,c]

[a] The Mind Research Network, Albuquerque, NM, USA
[b] Dept. of Electrical Engineering, University of New Mexico, Albuquerque, NM, USA
[c] Dept. of Neurosciences, University of New Mexico School of Medicine, Albuquerque, NM, USA
[d] Psychology Department and Neuroscience Institute, Georgia State University, Atlanta, GA, USA

ARTICLE INFO

Article history:
Received 3 October 2016
Received in revised form 14 April 2017
Accepted 16 April 2017
Available online xxxx

Keywords:
Schizophrenia
Bipolar disorder
Genetic association
Regional enrichment
European ancestry
African ancestry

ABSTRACT

Both schizophrenia (SZ) and bipolar disorder (BD) are highly heritable psychiatric disorders. The significant genomic risk loci are of great importance but with no guarantee of known functional impact and they cannot totally explain the genetic inheritance. In this study we present regional enrichment analyses across the genome, aiming to strike a balance between individual risk loci and integrated regional effects. Chromosomes were partitioned into 2 million base-pair regions (indicated by an underscore sign in the cytogenetic bands) on which enrichment tests are performed. In the discovery phase, we leverage the Psychiatric Genomics Consortium SZ and BD initial association test results for European Ancestry (EA) population and dbGAP SNP data for African Ancestry (AA) population. 78 and 48 regions show significantly enriched associations with SZ and BD respectively in the EA population, and nine are in common including MHC, 3p21.1, 7p22.3_2, 2q32.3_2, 8q24.3_4, and 19q13.33_1. The most unique SZ associated region is 1p21.3_3, while the most unique BD associated region is 6q25.2_1. For the AA population fewer regions are discovered with only 10% overlapping with that of EA population. A replication test using Wellcome Trust Case Control Consortium data for EA population verified 9% of the SZ enriched regions and 40% of the BD enriched regions. In summary, we showed that regional enrichment analyses produce reliable genetic association profiles using about one tenth of samples compared to single base-pair genome wide association approach. The identified association regions will be useful for further genetic functional studies.

1. Introduction

Schizophrenia (SZ) and bipolar disorder (BD) have long been debated for their uniqueness given shared clinical symptoms and neurological aberrance, common treatment, and genetic heritability (Ayalew et al., 2012; Potash and Bienvenu, 2009). This viewpoint has been strengthened by recent genetic discoveries from genome-wide association studies (GWAS) (Ripke et al., 2011; Ruderfer et al., 2014; Sklar et al., 2011). For example, polygenetic risk scores generated using SNPs associated with SZ can explain about 15% of the variance for BD disorder (Lee et al., 2013). Combined analyses on five psychiatric disorders, including SZ and BD, have identified SNPs in association with either SZ or BD, as well as both disorders (Smoller et al., 2013). Notably, the genome-wide significant top risk SNPs, though of great interest, do not guarantee direct functional impact or causality (Franke et al., 2016). And the polygenetic nature of these two disorders (Purcell et al., 2009; Stefansson et al., 2009) cannot exclude the importance of SNPs with nominal risks yet possibly with functional impact.

In parallel, enrichment test of relevance to a disease on preselected genes, pathways or networks is designed to strike a balance between significant individual mutations and interactive nature of biological function (Holden et al., 2008; Liu et al., 2010). Gene ontology categories, pathways defined by KEGG, protein-protein-interaction networks have all been tested for both BD and SZ (Holmans et al., 2009; Jia et al., 2010; Pedroso et al., 2012), and brain signaling pathways and cellular basic functions have all shown enrichment (Jia et al., 2010; Pedroso et al., 2012; Torkamani et al., 2008). Yet, the Psychiatric Genomics Consortium (PGC) SZ working group found that many pathways previously reported to be associated with SZ were not significantly enriched for SZ risk in the large patients vs. control data (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). It likely signals the limitation of such enrichment tests that rely heavily on knowledge base of gene annotation, pathway function, and protein interaction, all of which are not yet complete and undergo rapid changes (Khatri et al., 2012). Recently, enrichment analyses have extended

* Corresponding author at: The Mind Research Network, Albuquerque, NM, USA.
E-mail address: jliu@mrn.org (J. Liu).

http://dx.doi.org/10.1016/j.schres.2017.04.033
0920-9964/

to sets of loci derived from genomic or methylomic studies, many of which are not in coding gene regions but with promising regulation effects, and are tested for enrichment of diseases' association (GTEx-Consortium, 2015; Jaffe et al., 2016; Nicolae et al., 2010).

To our knowledge no enrichment test has been done on natural consecutive chromosome regions of the genome. The reason for regional enrichment tests is twofold. First, genetic loci within a close distance are potentially functionally related, not limited to within genes or linkage disequilibrium (LD) blocks, as documented most of regulatory *cis*-effects (Maston et al., 2006; Siepel and Arbiza, 2014). Second, increasing evidence supports the functional impact of SNPs outside protein coding regions (Maurano et al., 2012; Roussos et al., 2014). In fact SNPs associated with diseases are enriched within non-coding functional elements (ENCODE, 2012). Thus, we hypothesize that SNPs within a close distance, not limited to coding regions, might contribute to a disease in an integrative manner. Our research interest here is not to pinpoint the top risk carrying individual loci. Instead we aim to identify genetic regions with disease association, from where further functional studies such as association with gene expression or brain structural and functional variation can be launched.

Most genetic association studies for both disorders are based on European Ancestry (EA) population, less on African Ancestry (AA) population. Here we leveraged public genetic databases and PGC initial test results, and conducted regional enrichment test for SZ and BD in both EA and AA populations separately, followed by a verification using the Wellcome Trust Case Control Consortium (WTCCC) dataset. Then, we compared regional association profiles between SZ and BD, and between EA and AA populations.

2. Methods

2.1. Samples and data

We studied four cohorts in the discovery phase including EA SZ cohort, EA BD cohort, AA SZ cohort, and AA BD cohort, and two cohorts in the replication phase including a replication EA SZ cohort and a replication EA BD cohort. For the EA population we started our analyses with the genomic individual SNP association results reported by PGC1 SZ and BD working groups derived from large EA samples (Ripke et al., 2011; Sklar et al., 2011). For the AA population, SNP data from three projects in the dbGaP database were used: Genome-Wide Association Study of Schizophrenia, Molecular Genetics of Schizophrenia–nonGAIN, and Whole Genome Association Study of Bipolar Disorder. For replication WTCCC data were used. Sample and SNP information for each cohort are listed in Table 1 and the preprocessing of SNP data is explained in Supplementary text. Note that part of WTCCC data have been used in the PGC analyses. PGC SZ and BD analyses used 2333 and 3502 samples from WTCCC project, respectively. We used 4618 WTCCC samples for SZ association and 3301 WTCCC samples for BP association. Thus, given PGC analyses leveraged much large samples, our replication is more of a subset verification.

2.2. Analysis approaches

Enrichment tests on a continuous segment of chromosome were implemented. To partition the whole genome into chromosomal regions, we first leveraged the natural cytogenetic band structure. According to genome build hg19 reference (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/cytoBand.txt.gz), there are 862 cytogenetic bands across the whole genome and 811 in autosomes with average size 3.5 million (M) base-pairs (bps). Considering the size of 108 SZ risk regions reported by PGC (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014) ranging from 331 bps to 857 kbps, we tested 1 Mbp and 2 Mbp windows as a chromosome region to cover a possible continuous risk region within a cytoband. The results from 1 Mbp- and 2 Mbp-windows are very similar and results from 2 Mbp windows are reported here (see Supplementary text for results of 1 Mbp windows). Specifically, for any cytoband larger than 3 Mbps within chromosomes 1–22, we applied a moving window of 2 Mbps with 1 Mbps overlapping. And any cytoband of less than 3 Mbps was treated as one window. Windows with >5 SNPs available in our data, leading to averaged 1938 regions, were tested for enrichment. See Table 1 for region information of each cohort.

Enrichment test requires: 1) univariate association tests to get individual SNPs' p values, and 2) statistical tests on enrichment of a preselected set of SNPs. For univariate association tests, we applied standard logistic regression on disease status to each SNP with covariates for the top 3 population structure factors in MATLAB, i.e. logit (disease status) $= b_0 + b_1$ SNP $+ b_{2-5}$ population structures, similar model as used in PGC1 SZ and BD studies (Ripke et al., 2011; Sklar et al., 2011). No regression for age, sex or sub-study was included. We implemented univariate tests on dbGaP AA data and WTCCC replication data, and used PGC initial univariate test results. While many enrichment statistical tests have been proposed, we compared three algorithms: a versatile gene-based association enrichment test (VEGAS) (Liu et al., 2010), a gene score from gene-set enrichment analyses (Holden et al., 2008), and a Fisher's exact test. After comparison we chose to report VEGAS results as it properly controls for gene (region) size and LD structure based on nonparametric simulation (Mirina et al., 2012). Two levels of significance, $p < 0.01$ uncorrected and false discovery rate (FDR) corrected $p < 0.05$, were applied to capture the significant disease associated chromosome regions as well as the promising highly suggestive ones. Here we used FDR correction instead of Bonferroni correction due to relatedness of nearby windows.

In replication significance $p < 0.05$ was applied to verify regions derived from the discovery EA population. In addition, we also compared our results with the 108 SZ-associated regions reported in PGC2 SZ project that leveraged data from 36,989 SZ cases and 113,075 controls (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014).

3. Results

Regional genetic association profiles for SZ and BD in EA and AA populations are plotted in Figs. 1-4. As showed in Fig. 1, enrichment tests for

Table 1
Samples and SNPs of each cohort investigated.

	Discovery data: European ancestry	Discovery data: African ancestry
SZ	1,252,901 SNPs from 12,462 HCs & 9394 SZ patients; 1929 chromosomal regions with an average of 927 SNPs per region	809,498 SNPs from 957 HCs &1256 SZ patients (Affymetrix 6.0); 1932 chromosomal regions with an average of 602 SNPs per region
BD	2,427,220 SNPs from 9250 HCs & 7481 BD patients; 1928 chromosomal regions with an average of 1806 SNPs per region	809,498 SNPs from 957 HCs & 140 BD patients (Affymetrix 6.0); 1932 chromosomal regions with an average of 602 SNPs per region
Replication WTCCC data: European Ancestry		
SZ	623,669 SNPs from 2491 HCs & 2127 SZ patients (Affymetrix 6.0); 1941 chromosomal regions with an average of 435 SNPs per region	
BD	348,374 SNPs from 1456 HCs & 1845 BD patients (Affymetrix 500 K); 1930 chromosomal regions with an average of 256 SNPs per region	

Please cite this article as: Liu, J., et al., Regional enrichment analyses on genetic profiles for schizophrenia and bipolar disorder, Schizophr. Res. (2017), http://dx.doi.org/10.1016/j.schres.2017.04.033