Schizophrenia Research xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Schizophrenia Research

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

Two-stage replication of previous genome-wide association studies of *AS3MT-CNNM2-NT5C2* gene cluster region in a large schizophrenia case–control sample from Han Chinese population

Fanglin Guan [a,b,d,*,1], Tianxiao Zhang [c,**,1], Lu Li [a,b], Dongke Fu [b,d], Huali Lin [e], Gang Chen [b], Teng Chen [a,b,d]

[a] Department of Forensic Psychiatry, School of Medicine & Forensics, Xi'an Jiaotong University, Xi'an, China
[b] Key Laboratory of National Ministry of Health for Forensic Sciences, School of Medicine & Forensics, Xi'an Jiaotong University, Xi'an, China
[c] Department of Psychiatry, School of Medicine, Washington University, Saint Louis, MO, USA
[d] Key Laboratory of Environment and Genes Related to Diseases, Ministry of Education, Xi'an, China
[e] Xi'an Mental Health Center, Xi'an, China

ARTICLE INFO

Article history:
Received 19 December 2015
Received in revised form 29 June 2016
Accepted 4 July 2016
Available online xxxx

Keywords:
Schizophrenia
Common variants
AS3MT
CNNM2
NT5C2

ABSTRACT

Schizophrenia is a devastating psychiatric condition with high heritability. Replicating the specific genetic variants that increase susceptibility to schizophrenia in different populations is critical to better understand schizophrenia. *CNNM2* and *NT5C2* are genes recently identified as susceptibility genes for schizophrenia in Europeans, but the exact mechanism by which these genes confer risk for schizophrenia remains unknown. In this study, we examined the potential for genetic susceptibility to schizophrenia of a three-gene cluster region, *AS3MT-CNNM2-NT5C2*. We implemented a two-stage strategy to conduct association analyses of the targeted regions with schizophrenia. A total of 8218 individuals were recruited, and 45 pre-selected single nucleotide polymorphisms (SNPs) were genotyped. Both single-marker and haplotype-based analyses were conducted in addition to imputation analysis to increase the coverage of our genetic markers. Two SNPs, rs11191419 (OR = 1.24, $P = 7.28 \times 10^{-5}$) and rs11191514 (OR = 1.24, $P = 0.0003$), with significant independent effects were identified. These results were supported by the data from both the discovery and validation stages. Further haplotype and imputation analyses also validated these results, and bioinformatics analyses indicated that *CALHM1*, which is located approximately 630 kb away from *CNNM2*, might be a susceptible gene for schizophrenia. Our results provide further support that *AS3MT*, *CNNM2* and *CALHM1* are involved with the etiology and pathogenesis of schizophrenia, suggesting these genes are potential targets of interest for the improvement of disease management and the development of novel pharmacological strategies.

1. Introduction

Schizophrenia (SCZ) is a devastating mental disorder often characterized by holding false beliefs about reality, confusion, auditory hallucinations, reduced social engagement, reduced emotional expression, and a lack of motivation (van Os and Kapur, 2009). A systematic review indicates that the lifetime prevalence and incidence of SCZ are 0.30–0.66% and 10.2–22.0 per 100,000 person-years, respectively (McGrath et al., 2008). SCZ is a debilitating illness associated with significant costs, which place a great burden on the families of the patients and our entire society (Wu et al., 2005). It is a complex disorder, and genetics have played an important role in the onset of SCZ. A previous meta-analysis has indicated that its heritability is 81% (95% confidence interval, 73%–90%), making SCZ one of the most heritable psychiatric disorders (Sullivan et al., 2003). In the past decade, several susceptible genes, including DRD3 (Baritaki et al., 2004), DRD2 (Kukreti et al., 2006), DTNBP1 (Funke et al., 2004, Vilella et al., 2008), SYN2 (Lee et al., 2005), and MTHFR (Lewis et al., 2005), have been identified as conferring risk for the onset of SCZ. A recently published large scale GWAS study recently reported 108 significant conservatively defined loci associated with SCZ (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). Nevertheless, most of the common susceptible variants only moderately effects on the onset of SCZ, and only a small portion of the heritability can be explained by these variants. Many more variants with small effects remain undetected.

The locus 10q24.32-q24.33 was first associated with SCZ in a genome-wide study in 2011 (The Schizophrenia Psychiatric Genome-

* Correspondence to: F. Guan, Key Laboratory of National Ministry of Health for Forensic Sciences, School of Medicine & Forensics, Xi'an Jiaotong University, 76 Yanta West Road, Xi'an 710061, China.
** Correspondence to: T. Zhang, Department of Psychiatry, School of Medicine, Washington University, Saint Louis, MO 63110, USA.
E-mail addresses: fanglingguan@163.com (F. Guan), t.zhang@wustl.edu (T. Zhang).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.schres.2016.07.004
0920-9964/

Wide Association Study (GWAS) Consortium, 2011). Multiple subsequent studies (Aberg et al., 2013, Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014, Guan et al., 2012) identified several susceptible genetic polymorphisms located near the three-gene cluster of *AS3MT-CNNM2-NT5C2*. Although these three genes are clustered within a genomic region, they are unrelated functionally. The gene *AS3MT* encodes a protein that catalyzes the transfer of a methyl group from S-adenosyl-L-methionine (AdoMet) to a trivalent arsenic, and it has been reported to be related to arsenic metabolism (Lin et al., 2002). *CNNM2* encodes the member of a protein family that contains a cyclin box motif. The encoded protein may play an important role in magnesium homeostasis. Disruption of this gene has been reported to be associated with renal hypomagnesaemia (Stuiver et al., 2011). The proteins encoded by *NT5C2* are involved in cellular purine metabolism, and mutations in this gene have been reported to be linked to childhood acute lymphoblastic leukemia (Meyer et al., 2013). Previous studies have identified several significant SNPs associated with SCZ within this three-gene cluster, including rs3740390 (*AS3MT*) (Aberg et al., 2013), rs11191499 (*CNNM2*) (Aberg et al., 2013, Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), rs11191514 (*CNNM2*) (Aberg et al., 2013, Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), rs7914558 (Aberg et al., 2013, Ripke et al., 2013) and rs17094683 (*NT5C2*) (Aberg et al., 2013). However, the functional differences in these three genes and the complexity of linkage disequilibrium between the associated genetic markers in different populations within this genomic region indicate that the effects of these previously identified SNPs might not be independent and that most of these SNPs, if not all, could be surrogates of some variants with true functional effects related to SCZ. In addition, most of these susceptible variants are located within the non-coding regions, and how these variants confer risk for SCZ remains elusive. More research is definitely needed, especially for this specific genomic region, to clarify these unsolved problems.

In our study, we aimed to thoroughly evaluate the genomic region of this three-gene cluster (*AS3MT-CNNM2-NT5C2*) in individuals with SCZ from the Han Chinese population. Although previous studies that focused on this region were based on Han Chinese samples (Guan et al., 2012), the marker density was too limited to capture the sufficient information of the underlying SNPs that may be associated with increased susceptibility to SCZ. Guan's previous study insufficiently covered this 360 kb locus because only 6 SNPs across the entire region were examined. In this study, we analyzed 45 pre-selected SNPs across this three-gene locus from 8218 individuals (2538 cases and 5680 healthy controls) with Han Chinese ancestry. To maximize the study efficiency, we used a two-stage strategy. Furthermore, imputation was utilized to increase the coverage of these genetic markers.

2. Methods

2.1. Study subjects

We implemented a two-stage experimental strategy comprising the following stages: 1) a discovery stage in which we genotyped a relatively larger set of markers in a smaller subject subset of 1117 patients with SCZ (536 males and 581 females) and 1815 healthy controls (873 males and 942 females) and 2) a validation stage in which we only genotyped a smaller set of markers in the same linkage disequilibrium (LD) block with certain markers that passed the screening *P* value threshold (0.05) in a relatively large subject subset of 1421 patients with SCZ (759 males and 662 females) and 3865 healthy controls (1863 males and 2002 females). Characteristic information about the study subjects is shown in Table 1. All patients were recruited from the inpatient and outpatient clinical services of the psychiatric unit at the Xi'an Mental Health Center and were diagnosed by at least two experienced psychiatrists based on the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) criteria for SCZ. All unrelated healthy controls were selected from a combination of local volunteers; those with a personal family history of mental illness in the previous three generations and with current or past evidence of psychoses were excluded from the present study. All subjects enrolled from the city of Xi'an in the Shaanxi Province were of Han descent, and we excluded anyone who was not born locally or whose immediate family members from the previous three generations were not born locally. This research was performed in accordance with the ethical guidelines of the Declaration of Helsinki (version 2002) and was approved by the Medical Ethics Committee of Xi'an Jiaotong University. All participants completed written informed consent forms.

Table 1
The characteristic information for study subjects.

	Case	Control	Total
Stage (%)			
Discovery	1117 (38.1)	1815 (61.9)	2932
Validation	1421 (26.9)	3865 (73.1)	5286
Gender (%)			
Male	1295 (31.1)	2875 (68.9)	4170
Female	1243 (30.7)	2805 (69.3)	4048
Age (SD)	35.8 (8.18)	36.4 (8.44)	36.19 (8.36)

SD, standard deviation.

2.2. SNP genotyping

2.2.1. Power calculation and SNP selection

Statistical power was evaluated under an additive genetic model, using a genetic power calculator (Purcell et al., 2003) with disease prevalence at 1% and statistical significance set at $P < 0.05$. We conducted a comprehensive power analysis, and the results are shown in Table 2. As shown in the table, based on our sample size level of discovery stage, 78% statistical power for detecting a relative risk of 1.2 could be achieved for SNPs with minor allele frequency (MAF) between 0.15 and 0.3, including 19 out of those 45 selected SNPs (42%). In addition, we could also achieve 75% statistical power for detecting a relative risk of 1.3 or higher when the risk allele frequency was between 0.15 and 0.4, which contained 34 out of the 45 selected SNPs (74%).

We implemented a two-stage study design. In the first stage, we selected SNPs for genotyping from a 350 kb genomic region (Chr10: 104,600,000–104,950,000) that covered the *AS3MT-CNNM2-NT5C2* genomic region in a discovery sample composed of 2932 individuals (Fig. 1). Values of MAF ≥ 0.01 with pair-wise tagging and $r^2 \geq 0.9$ were used as the cutoff criteria during tag SNP selection, resulting in 40 tag SNPs, and another 5 SNPs reported to be significant in previous studies (rs11191419 and rs55833108) (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014); rs4409766, rs12221064 and rs11191499 (Aberg et al., 2013) were also included. A total of 45 SNPs were selected for genotyping in the first stage (Supplemental Table S1). Single-marker-based association analyses were performed for each SNP in the discovery stage, and SNPs with nominal significance ($P \geq 0.05$) were genotyped in another validation sample comprising 5286 individuals.

2.2.2. Genotyping and quality control

Genomic DNA of all participants was extracted from peripheral leukocytes according to the standard protocol of the DNA Isolation Kit for Mammalian Blood (Tiangen Biotech Co., Ltd, Beijing, China). The

Table 2
Full results of comprehensive power analysis.

Relative risk/MAF	0.05	0.1	0.15	0.2	0.3	0.4
1.1	0.0723	0.1356	0.2699	0.3608	0.3109	0.1805
1.2	0.1443	0.4124	0.7755	0.9013	0.8341	0.5375
1.3	0.2724	0.7497	0.9804	0.9979	0.9902	0.8484

MAF, minor allele frequency.

Please cite this article as: Guan, F., et al., Two-stage replication of previous genome-wide association studies of *AS3MT-CNNM2-NT5C2* gene cluster region in a large schizophrenia case–control sample..., Schizophr. Res. (2016), http://dx.doi.org/10.1016/j.schres.2016.07.004