

ORIGINAL ARTICLE

The SNP-set based association study identifies *ITGA1* as a susceptibility gene of attention-deficit/hyperactivity disorder in Han Chinese

L Liu^{1,2,8}, L Zhang^{3,4,5,8}, HM Li^{1,2}, ZR Wang⁶, XF Xie³, JP Mei³, JL Jin^{1,2}, J Shi⁶, L Sun^{1,2}, SC Li⁴, YL Tan⁶, L Yang^{1,2}, J Wang^{3,7}, HM Yang^{3,7}, QJ Qian^{1,2} and YF Wang^{1,2}

Genome-wide association studies, which detect the association between single-nucleotide polymorphisms (SNPs) and disease susceptibility, have been extensively applied to study attention-deficit/hyperactivity disorder (ADHD), but genome-wide significant associations have not been found yet. Genetic heterogeneity and insufficient genomic coverage may account for the missing heritability. We performed a two-stage association study for ADHD in the Han Chinese population. In the discovery stage, 1033 ADHD patients and 950 healthy controls were genotyped using both the Affymetrix Genome-Wide Human SNP Array 6.0 and the Illumina Infinium HumanExome BeadChip. The genotyped SNPs were combined to generate a powerful SNP set with better genomic coverage especially for the nonsynonymous variants. In addition to the association of single SNPs, we collected adjacent SNPs as SNP sets, which were determined by either genes or successive sliding windows, to evaluate their synergetic effect. The candidate susceptibility SNPs were further replicated in an independent cohort of 1441 ADHD patients and 1447 healthy controls. No genome-wide significant SNPs or gene-based SNP sets were found to be associated with ADHD. However, two continuous sliding windows located in *ITGA1* (P -value = $8.33E-7$ and P -value = $8.43E-7$) were genome-wide significant. The quantitative trait analyses also demonstrated their association with ADHD core symptoms and executive functions. The association was further validated by follow-up replications for four selected SNPs: rs1979398 (P -value = $2.64E-6$), rs16880453 (P -value = $3.58E-4$), rs1531545 (P -value = $7.62E-4$) and rs4074793 (P -value = $2.03E-4$). Our results suggest that genetic variants in *ITGA1* may be involved in the etiology of ADHD and the SNP-set based analysis is a promising strategy for the detection of underlying genetic risk factors.

Translational Psychiatry (2017) **7**, e1201; doi:10.1038/tp.2017.156; published online 15 August 2017

INTRODUCTION

Attention-deficit/hyperactivity disorder (ADHD) is one of the common neurodevelopmental disorders, which is characterized by inattention, hyperactivity and impulsivity.¹ The high heritability of ADHD (~0.76) indicates the significant influence of genetic effects in its pathogenesis.² In the past decades, numerous studies have attempted to explore and validate susceptibility genetic variants to explain the heritability of ADHD by either candidate gene association studies or genome-wide association studies (GWAS). Most of the candidate gene association studies focused on genes involved in the monoamine neurotransmitter systems; Nevertheless, they generally failed to yield consistent results.³ To avoid gene selection bias, GWAS were adopted as a hypothesis-free approach to identify single-nucleotide polymorphisms (SNPs) that associated with disease susceptibility.^{4,5}

Previous studies have shown that common variants contributed to ~40% of the heritability of ADHD.³ However, genome-wide significant associations have not been found in common variants,

even with a very large sample size or various types of microarrays.^{6–8} As an alternative approach, SNP-set based association analysis collects the adjacent SNPs and evaluates their joint association signal. This sophisticated method is complementary to single SNP association analysis in three aspects: (1) it can help explore the loci that failed to genotype the disease causal variant but contained multiple SNPs in linkage disequilibrium (LD) with it; (2) because the number of SNP sets should be much less than the number of SNPs, it can also alleviate the multiple testing burden; and (3) replicating the association signals in different cohorts can be possibly hindered by allelic heterogeneity, and SNP-set based association analysis alleviates the poor reproducibility of single SNP association.^{9–11} SNP-set based association analysis has demonstrated its effectiveness for ADHD.¹² On the basis of this approach, Mooney *et al.*¹³ detected a number of brain-relevant pathways and the pathways containing potassium channel genes that were suggestively associated with ADHD.

Traditional SNP microarrays, which were designed to optimize genome-wide coverage, have proved their advancement in GWAS;

¹Department of Child Psychiatry, Peking University Sixth Hospital/Institute of Mental Health, Beijing, China; ²National Clinical Research Center for Mental Disorders & Key Laboratory of Mental Health, Ministry of Health (Peking University), Beijing, China; ³BGI Genomics, BGI-Shenzhen, Shenzhen, China; ⁴Department of Computer Science, City University of Hong Kong, Hong Kong, China; ⁵Department of Computer Science, Stanford University, Stanford, CA, USA; ⁶Psychiatry Research Center, Beijing HuiLongGuan Hospital, Peking University, Beijing, China and ⁷James D. Watson Institute of Genome Sciences, Hangzhou, China. Correspondence: Professor QJ Qian or Professor YF Wang, Peking University Sixth Hospital/Institute of Mental Health, No. 51, Hua Yuan Bei Lu, Haidian District, Beijing 100191, China.

E-mail: qianqian@bjmu.edu.cn or wangyf@bjmu.edu.cn

⁸These authors contributed equally to this work.

Received 20 March 2017; revised 20 May 2017; accepted 7 June 2017

however, it is possible to lose signals due to their insufficient coverage of exonic variants. A recent study¹⁴ investigated the genomic coverage of 12 commonly used SNP microarrays and, surprisingly, observed that all of them exhibited poor genomic coverage of European and African genomes (< 50%). The Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix 6.0, Santa Clara, CA, USA) merely captures 31% of the genomic variants in Asian. The nonsynonymous SNPs, which alter translated protein products or structures, are more likely to cause diseases. The situation is more severe for these SNPs, and we found that most of them are weakly correlated with the common SNPs included in Affymetrix 6.0 (see Results and Supplementary Figure S1), which indicates that their association signals could be lost by previous studies that relied on traditional SNP microarray.

The high costs of exome sequencing limit its practical application on large sample size. The Illumina HumanExome BeadChip (Exome array) can be an alternative solution; it is designed to specifically capture the exonic variants that were previously identified by whole-exome and whole-genome sequencing projects. Exome arrays improve the exonic coverage for large cohort GWAS. Several studies have benefited from its ability to detect rare variants in exonic regions. Wessel *et al.*¹⁵ identified a nonsynonymous SNP rs10305492, that has low minor allele frequency in the *GLP1R* gene and is associated with low fasting glucose, type 2 diabetes risk and early insulin secretion. Zayats *et al.*⁸ genotyped 9356 individuals with European ancestry by Exome array and explored four loci harboring rare variants (*NTSDC1/COL10A1*, *SEC23IP*, *PSD* and *ZCCHC4*) that were genome-wide significant. The Exome array serves as a feasible solution for the genomic coverage bias of SNP microarrays by incorporating more independent exonic SNPs.

In this study, we collected 1033 ADHD patients and 950 healthy controls of Han Chinese ancestry and simultaneously genotyped them using both Affymetrix 6.0 and Exome array. In addition to SNP-based association analyses, the adjacent SNPs were also grouped into SNP sets in order to jointly capture their synergetic association signals. The SNP sets were defined both by genes and successive sliding windows to guarantee the inclusion of all coding and regulatory regions. We further replicated the candidate signals in an independent cohort of 1441 ADHD patients and 1447 healthy controls.

MATERIALS AND METHODS

Subjects

One thousand and thirty-three ADHD cases satisfying the DSM-IV criteria (870 males, 84.2%) and 950 normal controls (601 males, 63.3%) were genotyped by both the Affymetrix Genome-Wide Human SNP Array 6.0 (Affymetrix) and the Illumina Infinium HumanExome BeadChip (Illumina, San Diego, CA, USA). All samples were of Han Chinese ancestry. Most of these samples were included in our previous study.⁷ The cohort for replication consisted of 1441 ADHD patients (1168 males, 81.1%) and 1447 adult controls of Han Chinese ancestry (598 males, 41.3%). These patients were recruited from the child and adolescent psychiatric clinics of Peking University Sixth Hospital, and they were all between 6 and 16 years of age (average 10.9 ± 4.6 years old). The inclusion and exclusion criteria were the same as that in the discovery stage. The healthy controls had an average age of 40.7 ± 12.5 years and were recruited as volunteers from Beijing HuiLongGuan Hospital and neighboring community. The exclusion criteria were as follows: diagnosis or family history of ADHD, schizophrenia, affective disorders, pervasive development disorders and epilepsy; current or a history of substance abuse; and diagnosis of severe physical diseases. The detailed information for the discovery and replication stages is shown in Supplementary Table S1.

All subjects were treated in accordance with the Declaration of Helsinki. This study was approved by the Ethics Committee of Peking University Sixth Hospital. Written informed consent was obtained from all subjects or from the parents of the children.

SNP microarray genotyping

The collected genomic DNA samples from 1033 ADHD patients and 950 healthy controls were placed into 96-well plates, and the DNA was then genotyped on both the Affymetrix Genome-Wide Human SNP Array 6.0¹⁶ and the Illumina Infinium HumanExome-12v1 BeadChip.¹⁷ The SNP genotypes were called by BIRDSEEDv2 and GenomeStudio v2011.1. The SNPs were removed if they failed to satisfy standard quality controls (Supplementary Notes). The samples that passed the overall call rate of 98% and the identity by descent check were released. Common variants were the SNPs with a minor allele frequency (MAF) > 1% for both arrays and rare variants were those with $0.05\% < \text{MAF} < 1\%$ in the Exome array, which guaranteed the incorporation of rare deleterious nonsynonymous variants that were present in at least one sample. Strong concordance (> 95%) was calculated between the Affymetrix 6.0 and the Exome array by evaluating their overlapped SNPs (9127 SNPs).

Tagged SNPs in the Exome array

If a SNP in the Affymetrix 6.0 array (S_{Affy}) has sufficiently strong LD ($r^2 > 0.8$) with a SNP in the Exome array (S_{Exome}), the S_{Affy} was considered to be the tag SNP of S_{Exome} . For each S_{Exome} , we only selected tag SNPs from the top 10 nearest S_{Affy} . LD was calculated by PLINK,¹⁸ and the nonsynonymous SNPs were annotated by ANNOVAR¹⁹ on the HG19RefSeq.

Association analysis

The association of autosomal SNPs was evaluated by logistic regression, including sex and the first two eigenvectors as confounders. For the association of SNPs in the X chromosome, males and females were calculated separately and then combined using Fisher's method. Because males have one copy of the X chromosome, all SNPs in the X chromosome of males were considered to be homozygous.²⁰ The *P*-values from the discovery and replication stages were merged by fixed effect meta-analysis in METAL.²¹ Conditional logistic regression was performed to capture the independent signals from the same loci.

For the SNP-set based association analysis, the SNP sets were defined by genes or by a 10 kb sliding window (5 kb overlap). The genes were defined by extending ± 50 bp to the boundaries extracted from the HG19RefSeq. We chose the SKAT_Common_Rare²² function in the SKAT package (SNP-set Kernel Association Test) to jointly consider both common and rare variants in autosomal loci. SKAT_Common_Rare separately calculated the contribution of common and rare variants and weighted their aggregated *P*-values. For non-pseudoautosomal regions of the X chromosome, the SNP-set associations were identified by applying a modified version of VEGAS²³ with the truncated tail strength method²³ in XWAS.²⁰

Quantitative trait analysis

ADHD symptoms were evaluated using the ADHD RS-IV scale, which had been translated into Chinese and showed good reliability and validity.²⁴ Executive functions (EFs) have been suggested to be potential endophenotypes and may promote the functional exploration of susceptibility genetic variants for ADHD.²⁵ We used the Behavior Rating Inventory of Executive Function (BRIEF) to evaluate the ecological EF of children with ADHD.²⁶ For the analyses of dimensional symptoms, gender and age were adjusted. For the analysis of EF, gender, age, IQ and ADHD subtypes were adjusted to control the influence of potential confounding factors. The SKAT_Common_Rare function was also utilized to evaluate the candidate SNP-set quantitative trait association. To correct for the multiplicity of testing, Bonferroni corrections were performed. Considering the evaluation of the two loci on 15 variables, the adjusted significance threshold was set at a *P*-value < 0.0017.

Replication

We designed a custom Illumina GoldenGate Genotyping assay to replicate the selected SNPs on an independent cohort of 1441 ADHD patients and 1447 healthy controls of Han Chinese ancestry. We carefully designed 97 SNPs for the custom array (Supplementary Table S2), including the SNPs in *ITGA1* ($n = 4$), other signals with marginal association with ADHD from the Affymetrix 6.0 or the Exome array ($n = 73$), and ancestral informative markers ($n = 20$). Raw data were analyzed by Illumina GenomeStudio V2011.1.

Table 1. Significant ADHD associated windows in *ITGA1* from SNP-set (sliding windows) based association analysis

Locus	Gene	Chr	Number of SNPs	Start	End	P-value	q-value
L1	<i>ITGA1</i>	5	9	52191000	52201000	8.33E-7	0.03
L2	<i>ITGA1</i>	5	13	52186000	52196000	8.43E-7	0.03

Abbreviation: ADHD, attention-deficit/hyperactivity disorder; SNP, single-nucleotide polymorphism. Note: L1 and L2 are two adjacent sliding windows in *ITGA1*, whose P-values are calculated by SKAT and the multiple testing is corrected by false discovery rate with q-values.

RESULTS

Increased genomic coverage by combining SNPs from the Affymetrix 6.0 and the Exome array

Increased genomic coverage was expected by merging the SNPs from the Affymetrix 6.0 and the Exome array together. The SNPs loci in the Affymetrix 6.0 (S_{Affy}) were carefully designed to cover the human genome based on the SNPs found in HapMap;²⁷ however, they poorly covered the exonic SNPs (S_{Exome}) included on the Exome array, especially nonsynonymous SNPs (Supplementary Figure S1). More than half of the S_{Exome} were independent with S_{Affy} even when the tag SNPs were defined by marginal LD ($r^2 > 0.1$; Supplementary Figure S1a). The tag S_{Affy} can be found for only 27.49% of S_{Exome} ($r^2 > 0.8$), which indicated the insufficient coverage of the Affymetrix 6.0 in exonic regions. The coverage problem was more severe for nonsynonymous SNPs (Supplementary Figure S1b), and more than 90% of the nonsynonymous S_{Exome} were independent.

ITGA1 as a novel susceptibility gene of ADHD

At the discovery stage, we genotyped 1033 patients and 950 controls of Han Chinese ancestry using both the Affymetrix 6.0 and the Exome arrays. After genotype and sample quality control (see Materials and Methods), we combined the genotyped SNPs in these two microarrays and performed genome-wide association analysis on 7 17 417 SNPs from 1983 individuals.

The principal component analysis (Supplementary Figure S2) showed a good geographical match between the cases and controls. The inflation factor (Supplementary Figure S3) shown in the Q-Q plot indicated an insignificant deviation from null distribution. Unfortunately, no SNPs were genome-wide significant ($P\text{-value} < 5E-8$) in either the Affymetrix 6.0 or the Exome array (Supplementary Figure S4). To investigate the loci with multiple marginal signals, the adjacent SNPs were grouped as SNP sets (defined by sliding windows or genes), and their synergetic associations were evaluated. The SNP set phenotypic association was evaluated by SKAT²² and adjusted by the Benjamini and Hochberg false discovery rate.²⁸ Rather than defining the SNP set by gene, the two successive sliding windows from *ITGA1* were genome-wide significant (Supplementary Figure S5), designated as L1 (chr5: 52 191 000-52 201 000; $P\text{-value} = 8.33E-7$, $q\text{-value} = 0.03$) and L2 (chr5: 52 186 000-52 196 000; $P\text{-value} = 8.43E-7$, $q\text{-value} = 0.03$) (Table 1).

In addition to the analyses for dichotomous phenotype, we also evaluated the association between these two windows and quantitative phenotypes, including ADHD core symptoms and executive functions that have been suggested to be potential endophenotypes. Our results showed that both L1 and L2 were significantly associated with ADHD core symptoms (evaluated by ADHD RS-IV: inattentive, hyperactive/impulsive and total symptom) and ecological executive functions (evaluated by the BRIEF scale: inhibit, shift, emotional control, initiate, working memory, plan, organization and monitor; and the consequent Behavioral Regulation Index, Metacognition Index and total score), after adjusting for the potential confounders (Table 2). These data also

Table 2. The quantitative trait association of *ITGA1* with ADHD core symptoms and ecological executive functions

Phenotype	N	L1	L2
		P-value	P-value
<i>Core symptoms (ADHD RS-IV)</i>			
Inattentive	986	6.75 E-6	7.37 E-6
Hyperactive/impulsive	991	1.60E-5	1.34E-5
Total	959	7.57E-6	8.37E-6
<i>Executive function (BRIEF)</i>			
Inhibit	676	6.22 E-4	2.32 E-4
Shift	660	5.48 E-5	2.11 E-5
Emotional control	663	9.03 E-4	3.47 E-4
Initiate	665	1.44 E-4	6.06 E-5
Working memory	664	1.38 E-4	8.91E-5
Plan	658	7.42 E-4	4.17 E-4
Organization	663	6.48 E-4	3.55 E-4
Monitor	670	2.65 E-4	1.47 E-4
Behavioral Regulation Index	628	2.40 E-4	7.88 E-5
Metacognition Index	606	2.01 E-4	1.79 E-4
Total	569	3.07 E-4	1.92 E-4

Abbreviations: ADHD, attention-deficit/hyperactivity disorder; ADHD RS-IV, ADHD rating scale-IV; BRIEF, Behavior Rating Inventory of Executive Function. Note: N is the number of samples involved in the analysis. The threshold of genome-wide significance is $P\text{-value} = 1.7E-3$.

indicated that *ITGA1* might participate in the pathogenesis of ADHD.

To validate the association of *ITGA1*, we carefully selected four SNPs to replicate the initial findings (Figure 1): rs1979398 ($P\text{-value} = 1.92E-6$) and rs16880453 ($P\text{-value} = 1.81E-5$) were the two most significant SNPs in *ITGA1*, rs1531545 ($P\text{-value} = 5.64E-4$) was genotyped by the Exome array and rs4074793 ($P\text{-value} = 1.12E-4$) showed marginal LD with the other three SNPs (Supplementary Figure S6). These four SNPs were replicated on an independent cohort of 1441 patients and 1447 controls of Han Chinese ancestry using the Illumina GoldenGate Assay. The rs1979398 and rs16880453 were genotyped by Affymetrix 6.0 and exhibited moderate LD ($r^2 = 0.63$, Supplementary Figure S6), and a consistent trend of association between the discovery and replication stages (combined $P\text{-value} = 2.64E-6$ and $3.58E-4$, respectively; Table 3). The SNP rs1531545 was selected from the Exome array and showed a consistent trend of association (combined $P\text{-value} = 7.62E-4$; Table 3). All the four SNPs were with the same risk alleles in the two stages. To investigate whether independent signals exist in *ITGA1*, we examined rs4074793, which had marginal LD with rs1979398 ($r^2 = 0.05$), rs16880453 ($r^2 = 0.06$) and rs1531545 ($r^2 = 0.05$). The results suggested that rs4074793 was also associated with ADHD (combined $P\text{-value} = 2.03E-4$; Table 3) as an independent contributor by controlling the most significant SNP rs1979398 ($P\text{-value} = 4.93E-3$; Table 4).

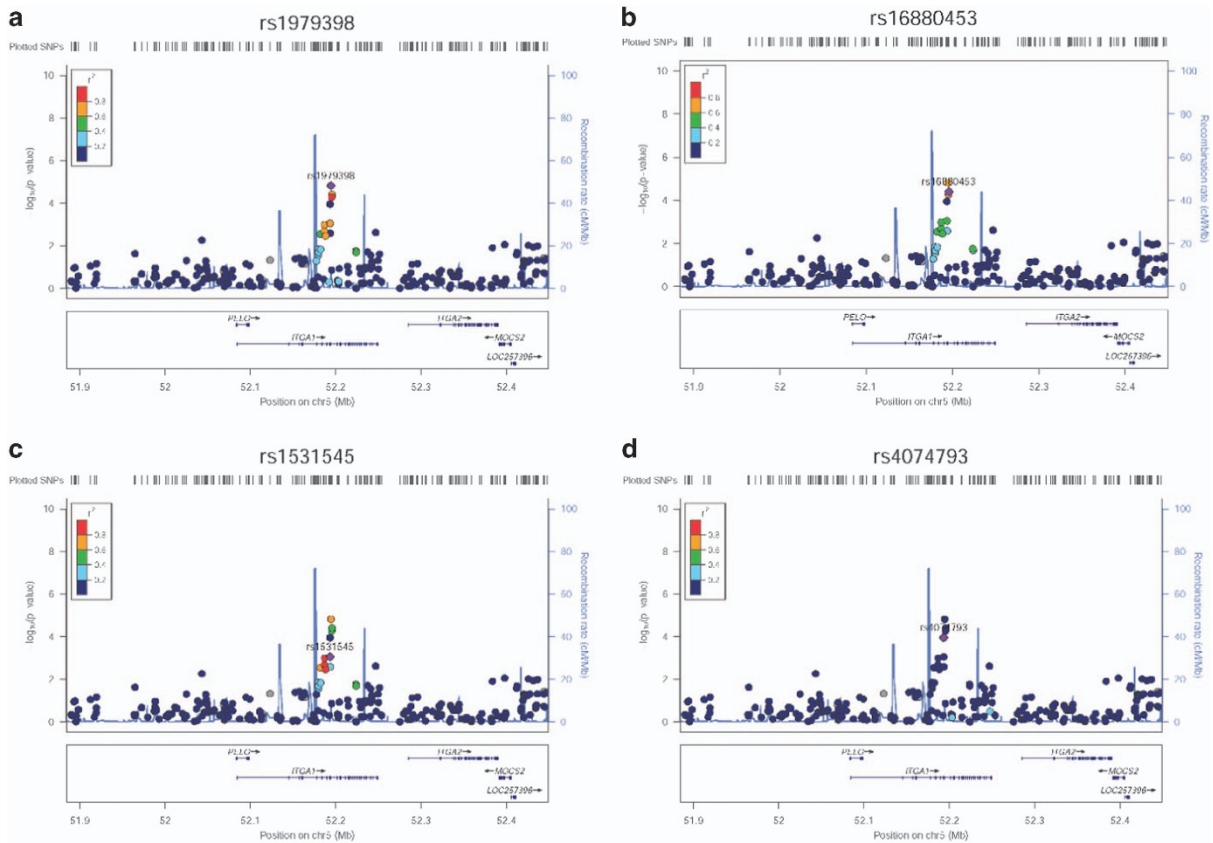


Figure 1. Fine mapping of ± 200 kb regions of rs1979398 (a), rs16880453 (b), rs4074793 (c) and rs1531545 (d) in *ITGA1*. The Locus Zoom plots showing the association significance ($-\log_{10}(P\text{-value})$) and local linkage disequilibrium in r^2 (color-coded).

Table 3. The association results of the four SNPs in *ITGA1*

SNP	Chr	Position	Minor allele	Gene	Discovery stage		Replication stage		Combined	
					P-value	β	P-value	β	P-value	β
rs1979398	5	52194327	A	<i>ITGA1</i>	1.92E-6	-0.319	0.037	-0.121	2.64E-6	-0.207
rs16880453	5	52195507	G	<i>ITGA1</i>	1.81E-5	-0.290	0.305	-0.060	3.58E-4	-0.157
rs1531545	5	52193287	C	<i>ITGA1</i>	5.64E-4	-0.234	0.142	-0.086	7.62E-4	-0.149
rs4074793	5	52193125	G	<i>ITGA1</i>	1.12E-4	0.499	0.131	0.177	2.03E-4	0.323

Abbreviation: SNP, single-nucleotide polymorphism. Note: The *P*-values of discovery and replication stages were calculated by logistic regression, and combined by METAL. Both of discovery and replication stages have the consistent trend and risk allele for associations of the four SNPs. β is calculated for the minor allele. If the number is positive, the minor allele is a risk allele, otherwise it is a protective allele.

Table 4. The replication result of rs16880453, rs1531545 and rs4074793 conditioning on rs1979398

SNP	Chr	Position	Minor allele	Gene	Discovery stage		Replication stage		Combined	
					P-value	β	P-value	β	P-value	β
rs16880453	5	52195507	G	<i>ITGA1</i>	0.377	-0.098	0.185	0.144	0.692	0.029
rs1531545	5	52193287	C	<i>ITGA1</i>	0.705	0.040	0.569	0.056	0.5	0.049
rs4074793	5	52193125	G	<i>ITGA1</i>	4.14E-3	0.380	0.235	0.144	4.93E-3	0.251

Abbreviation: SNP, single-nucleotide polymorphism. Note: The *P*-values of discovery and replication stages were calculated by logistic regression, and combined by METAL. The SNP rs4074793 is an independent contributor with other three SNPs. β is calculated for the minor allele. If the number is positive, the minor allele is a risk allele, otherwise it is a protective allele.

In addition to these four SNPs, we also replicated 73 other SNPs (Supplementary Table S2) that were marginally associated with ADHD in the discovery stage. Unfortunately, these SNPs were not genome-wide significant even with a larger sample size.

DISCUSSION

ADHD is a neurodevelopmental disorder that is mainly influenced by genetic factors and exhibits high heritability. In the recent decades, there has been much effort in exploring and defining the susceptibility genetic variants for ADHD. Many studies have attempted to identify the susceptibility genetic variants from large amounts of samples; unfortunately, genetic risk variants have not yet been observed to be significantly associated with ADHD. ADHD is a complex polygenic disorder, and numerous genetic variants with minor effects may contribute to its etiology. However, these minor effects are hard to explore by examining single SNP association in a limited sample size. Therefore, a combination of numerous adjacent SNPs may promote the investigation.

In this study, we genotyped the ADHD patients and controls of Han Chinese ancestry on both the Affymetrix 6.0 and the Exome array to explore the susceptibility SNPs. In addition to single SNP association analyses, SNP sets were defined by genes and successive sliding windows to integrate the association across multiple adjacent SNPs. Although the gene-based SNP-set analysis failed to detect any associations, two adjacent sliding windows located in *ITGA1* were found to be significantly associated with ADHD. They were replicated by choosing four SNPs (rs1979398, rs16880453, rs1531545 and rs4074793) in an independent replication cohort. Haplotype-based association analysis, which is performed by defining haplotypes according to LDs and the physical locations of SNPs, is another efficient approach to boost the association power of SNPs in the same LD block. However, it does not consider independent signals, such as rs4074793 here, which was identified in the study.

The association of these four SNPs with ADHD or other psychiatric disorders has not been previously reported. Only rs4074793 has been previously identified to be associated with liver enzyme concentration.²⁹ However, existing evidence could support the involvement of *ITGA1* in the genetic etiology of ADHD. One previous study has explored the association of genes at chromosome 5p13-q11 with ADHD, including *ITGA1*. Although significant associations were not found with dichotomous phenotype, analyses on dimensional symptoms yielded a marginal association of rs10513003 with both inattentive and hyperactive/impulsive symptoms.³⁰ Unfortunately, this SNP was not genotyped in the Affymetrix 6.0 or the Exome Array. This SNP is not located in the locus identified in our study and is ~30 kb away. The SNP rs10513003 showed marginal LD with the four SNPs identified in the present study ($r^2 = 0.009\text{--}0.213$) based on CHB in HapMap, indicating that the signal found in our study is independent from the one found by Laurin *et al*. As indicated in our previous report, the interval-based enrichment analysis tool (INRICH)³¹ showed that the neuron projection morphogenesis pathway (GO:0048812; including *ITGA1* and *GJA1* from our data) was suggestively associated with ADHD.⁷ In addition, *ITGA1* and its paralogs (*ITGA11* and *ITGAE*) were the top candidates of a previous ADHD GWAS.^{6,32} We have examined the SNP associations of *ITGA1* from the GWAS performed by Psychiatric Genetics Consortium (PGC).⁶ This study was also unable to find any genome-wide significant association in *ITGA1* (Supplementary Figure S7). In the PGC study, the most significant SNPs in the loci that were identified in our study were rs7735139 ($P\text{-value} = 0.059$, located in *L1*) and rs1110350 ($P\text{-value} = 0.045$, located in *L2*; not genotyped in the Affymetrix 6.0 or the Exome Array, but was in strong LD with rs7735139 with $r^2 = 0.961$ and $r^2 = 1$ based on the HapMap CEU and CHB, respectively). The analysis of the SNP rs7735139 in our

GWAS data showed a $P\text{-value}$ of 0.655 and different associated allele from PGC (allele A vs allele G, Supplementary Table S3). All the four candidate SNPs (Table 3) were insignificant in PGC study (Supplementary Table S3). And the associated orientation was contradictory for rs1179398 and rs4074793 between the two studies, whereas the other two SNPs were the same. In addition, the most significant locus found in PGC was far away (~137 kb from *L2*) and should be the independent signal with ours ($r^2 < 0.5$ in HapMap CHB, Supplementary Figure S7). These implied the genetic heterogeneity of ADHD in different populations.

Integrins are ubiquitously expressed adhesion molecules and may mediate cell–cell interactions. Aberrant cell adhesion has been shown and proven to participate in the etiology of psychiatric disorders, including ADHD. One important cell adhesion-related candidate gene for ADHD was *CDH13*.³³ *CDH13* was significantly associated with ADHD in the study by Lasky-Su *et al*,³⁴ whereas it only showed a trend toward association in a later research study.⁵ *ITGA1* encodes the α subunits of integrin. The α and β subunits of integrin form a heterodimer, which forms a cell-surface receptor for collagen and laminin and play an important role in the development and maturation of nervous system by mediating neural cell migration, histogenesis, neurite out-growth³⁵ and synaptic plasticity.³⁶ In addition, Murase and Hayashi³⁶ examined the distribution of neurons expressing integrin $\alpha 1$ in adult mouse. They found that most of the INTa1-positive neurons also expressed tyrosine hydroxylase (TH), which was the classification marker for catecholaminergic neurons. Dysfunction of the catecholaminergic system has been suggested to be involved in the pathogenesis of ADHD. In summary, the abovementioned evidence suggests that, integrins may participate in the etiology of ADHD by affecting cell adhesion, migration of neurons and/or nerve fiber outgrowth or by influencing the functions of the catecholaminergic system. However, the explicit function of these two loci of *ITGA1* that were found in the present study is still unknown. We analyzed the expression quantitative trait loci (eQTL) of these four reported SNPs based on the data generated from the UK Brain Expression Cohort (UKBEC, <http://www.braineac.org/>) and found that these markers may affect the expression of *ITGA1* in the hippocampus (rs4074793, $P\text{-value} = 0.0016$), thalamus (rs4074793, $P\text{-value} = 0.0048$), substantia nigra (rs1531545, $P\text{-value} = 0.0071$) and occipital cortex (rs16880453, $P\text{-value} = 0.0071$; Supplementary Figure S8). Aberrant gene expression may change the structure of brain regions, but additional neurobiological experiments and genetic imaging studies are required to confirm this hypothesis.

Previous GWAS have been unable to detect common variants that contribute to ADHD, and although supplementary approaches, such as gene-based and pathway-based analyses, have been utilized, validated biomarkers have not yet been explored. One possible reason is insufficient or biased genomic coverage, especially for nonsynonymous variants that are underestimated by the SNP microarrays that were previously used. Although signals in the intergenic regions are difficult to interpret, more attention should be directed toward those regulatory regions. Our study integrated the Exome array with Affymetrix 6.0 to achieve more power for exonic regions and adopted SNP-set analysis to include both genic and intergenic SNPs into the analysis. In addition to the two described SNP sets (*L1* and *L2*), there is another window ($P\text{-value} = 1.07E - 6$, $q\text{-value} = 0.035$) in chromosome 18 that was genome-wide significant (Supplementary Figure S5a). This window is located at the intergenic region between *RP11-638L3.1* and *TMX3*. However, the replications of the most significant SNP, rs17232800, showed a contradictory trend (Supplementary Table S2).

In addition to common SNPs, rare variants, such as low frequency single-nucleotide variants, small insertions/deletions (indels) may also be involved in the etiology of ADHD. Whole-exome or whole-genome sequencing is an attractive solution in

which all nucleotides can be sequenced without any variant selection bias. Their contribution to the advancement of understanding neurodevelopmental disorders has been shown by several family-based studies.^{37–39} However, the cost for sequencing a large number of samples is still unaffordable. We adopted an alternative approach and genotyped ~2000 samples by the Exome array, which was designed based on the findings from previous large sample exome sequencing. However, we found that most of the variants in the Exome array (77.11%) were appeared infrequently (MAF < 0.05%) in the Han Chinese population. The phenomenon was consistent with a recent publication regarding the Korean population.⁴⁰ The Exome array may be more appropriate for the European population than the Asian population. Nevertheless, it can be used to validate the association signals from GWAS microarrays (see Results).

Some limitations must be noted here. First, the sample size for the present study was still small and insufficient for the identification of single SNPs with a marginal effect size. The genetic power calculated by CaTS⁴¹ was only 44%, assuming a disease prevalence of 0.05, minor allele frequency of 0.3, significance level of 5E–8 and genotypic relative risk of 1.5. Second, the genotypes are permanent and unrelated to human age. However, the adult controls used in the replication study may exhibit some recall bias for their psychiatric status. Age-matched controls would be better, but the collection of children as controls is very difficult. Finally, although the results of the replication study provided further evidence of the association between *ITGA1* and ADHD, we should note that the significance was only marginal and could not survive stringent Bonferroni corrections. Further replication in different populations is thus required.

In conclusion, the results of our study using SNP-set analyses of both dichotomous and quantitative phenotypes and subsequent SNP based replications suggested the involvement of *ITGA1* in the genetic etiology of ADHD. In the future, additional work on gene function is required to illustrate how gene functions change due to the genetic variants and to determine the biological alteration caused by the associated genetic variants. In addition, our results support the effectiveness of SNP-set analyses for the study of complex polygenic disease, such as ADHD.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

This study was funded by the National Basic Research Program of China (2014CB846104), the National Key Technology Research and Development Program of the Ministry of Science and Technology of China (2015BAI13B01), the National Natural Science Foundation of China (81641163; 81571340 and 81301171) and Beijing Natural Science Foundation (7172245). We thank all participants in our present study.

REFERENCES

- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5 Task Force*, 5th edn, vol. xlv. American Psychiatric Association: Washington, DC, 2013, 947 p.
- Faraone SV, Mick E. Molecular genetics of attention deficit hyperactivity disorder. *Psychiatr Clin North Am* 2010; **33**: 159–180.
- Faraone SV, Asherson P, Banaschewski T, Biederman J, Buitelaar JK, Ramos-Quiroga JA *et al*. Attention-deficit/hyperactivity disorder. *Nat Rev Dis Primers* 2015; **1**: 15020.
- Hunter DJ, Altshuler D, Rader DJ. From Darwin's finches to canaries in the coal mine—mining the genome for new biology. *N Engl J Med* 2008; **358**: 2760–2763.
- Licinio J. Messages from hypothesis-driven genotyping: the case of schizoaffective disorder, bipolar type. *Mol Psychiatry* 2010; **15**: 113–114.
- Neale BM, Medland SE, Ripke S, Asherson P, Franke B, Lesch KP *et al*. Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 2010; **49**: 884–897.

- Yang L, Neale BM, Liu L, Lee SH, Wray NR, Ji N *et al*. Polygenic transmission and complex neuro developmental network for attention deficit hyperactivity disorder: genome-wide association study of both common and rare variants. *Am J Med Genet B Neuropsychiatr Genet* 2013; **162B**: 419–430.
- Zayats T, Jacobsen KK, Kleppe R, Jacob CP, Kittel-Schneider S, Ribases M *et al*. Exome chip analyses in adult attention deficit hyperactivity disorder. *Transl Psychiatry* 2016; **6**: e923.
- Wu MC, Kraft P, Epstein MP, Taylor DM, Chanock SJ, Hunter DJ *et al*. Powerful SNP-set analysis for case-control genome-wide association studies. *Am J Hum Genet* 2010; **86**: 929–942.
- Lu ZH, Zhu H, Knickmeyer RC, Sullivan PF, Williams SN, Zou F *et al*. Multiple SNP set analysis for genome-wide association studies through Bayesian latent variable selection. *Genet Epidemiol* 2015; **39**: 664–677.
- Cai M, Dai H, Qiu Y, Zhao Y, Zhang R, Chu M *et al*. SNP set association analysis for genome-wide association studies. *PLoS ONE* 2013; **8**: e62495.
- Aebi M, van Donkelaar MM, Poelmans G, Buitelaar JK, Sonuga-Barke EJ, Stringaris A *et al*. Gene-set and multivariate genome-wide association analysis of oppositional defiant behavior subtypes in attention-deficit/hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* 2016; **171**: 573–588.
- Mooney MA, McWeeney SK, Faraone SV, Hinney A, Hebebrand J, IMAGE2 Consortium *et al*. Pathway analysis in attention deficit hyperactivity disorder: an ensemble approach. *Am J Med Genet B Neuropsychiatr Genet* 2016; **171**: 815–826.
- Ha NT, Freytag S, Bickeboeller H. Coverage and efficiency in current SNP chips. *Eur J Hum Genet* 2014; **22**: 1124–1130.
- Wessel J, Chu AY, Willems SM, Wang S, Yaghootkar H, Brody JA *et al*. Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. *Nat Commun* 2015; **6**: 5897.
- Edenberg HJ, Liu Y. Laboratory methods for high-throughput genotyping. *Cold Spring Harb Protoc* 2009; **2009**: pdb top62.
- Grove ML, Yu B, Cochran BJ, Haritunians T, Bis JC, Taylor KD *et al*. Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PLoS ONE* 2013; **8**: e68095.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; **38**: e164.
- Gao F, Chang D, Biddanda A, Ma L, Guo Y, Zhou Z *et al*. XWAS: a software toolset for genetic data analysis and association studies of the X chromosome. *J Hered* 2015; **106**: 666–671.
- Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics* 2010; **26**: 2190–2191.
- Ionita-Laza I, Lee S, Makarov V, Buxbaum JD, Lin X. Sequence kernel association tests for the combined effect of rare and common variants. *Am J Hum Genet* 2013; **92**: 841–853.
- Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM *et al*. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet* 2010; **87**: 139–145.
- Su LY, Geng YG, Wang H, Du YS, Sun L, Zhao FT, Fan F. [Norm of ADHD diagnostic scale-parent version in Chinese urban children]. *Zhong Guo Shi Yong Er Ke Za Zhi* 2006; **21**: 833–836.
- Doyle AE, Faraone SV, Seidman LJ, Willcutt EG, Nigg JT, Waldman ID *et al*. Are endophenotypes based on measures of executive functions useful for molecular genetic studies of ADHD? *J Child Psychol Psychiatry* 2005; **46**: 774–803.
- Qian Y, Wang YF. [Reliability and validity of behavior rating scale of executive function parent form for school age children in China]. *Beijing Da Xue Xue Bao Yi Xue Ban* 2007; **39**: 277–283.
- International HapMap C. The International HapMap Project. *Nature* 2003; **426**: 789–796.
- Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 2001; **125**: 279–284.
- Chambers JC, Zhang W, Sehmi J, Li X, Wass MN, Van der Harst P *et al*. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nat Genet* 2011; **43**: 1131–1138.
- Laurin N, Lee J, Ickowicz A, Pathare T, Malone M, Tannock R *et al*. Association study for genes at chromosome 5p13-q11 in attention deficit hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* 2008; **147B**: 600–605.
- Lee PH, O'Dushlaine C, Thomas B, Purcell SM. INRICH: interval-based enrichment analysis for genome-wide association studies. *Bioinformatics* 2012; **28**: 1797–1799.
- Lesch KP, Timmesfeld N, Renner TJ, Halperin R, Roser C, Nguyen TT *et al*. Molecular genetics of adult ADHD: converging evidence from genome-wide association and extended pedigree linkage studies. *J Neural Transm* 2008; **115**: 1573–1585.
- Rivero O, Sich S, Popp S, Schmitt A, Franke B, Lesch KP. Impact of the ADHD-susceptibility gene CDH13 on development and function of brain networks. *Eur Neuropsychopharmacol* 2013; **23**: 492–507.

- 34 Lasky-Su J, Neale BM, Franke B, Anney RJ, Zhou K, Maller JB *et al*. Genome-wide association scan of quantitative traits for attention deficit hyperactivity disorder identifies novel associations and confirms candidate gene associations. *Am J Med Genet B Neuropsychiatr Genet* 2008; **147**: 1345–1354.
- 35 Bamdad M, Volle D, Dastugue B, Meinie A. Alpha1beta1-integrin is an essential signal for neurite outgrowth induced by thrombospondin type 1 repeats of SCO-spondin. *Cell Tissue Res* 2004; **315**: 15–25.
- 36 Murase S, Hayashi Y. Integrin alpha1 localization in murine central and peripheral nervous system. *J Comp Neurol* 1998; **395**: 161–176.
- 37 Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ *et al*. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012; **485**: 237–241.
- 38 De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE *et al*. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014; **515**: 209–215.
- 39 Iossifov I, O’Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D *et al*. The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 2014; **515**: 216–221.
- 40 Park TJ, Heo L, Moon S, Kim YJ, Oh JH, Han S *et al*. Practical calling approach for exome array-based genome-wide association studies in Korean population. *Int J Genom* 2015; **2015**: 421715.
- 41 Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 2006; **38**: 209–213.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

© The Author(s) 2017

Supplementary Information accompanies the paper on the Translational Psychiatry website (<http://www.nature.com/tp>)