Accumulation of Minor Alleles of Common SNPs in Schizophrenia

Pei He[1], Xiaoyun Lei[1], Dejian Yuan[1], Zuobin Zhu[2], and Shi Huang[1]*

[1]State Key Laboratory of Medical Genetics, School of life sciences, Central South University, 110 Xiangya Road, Changsha, Hunan, 410078, China

[2]Department of Genetics, Xuzhou Medical University, Xuzhou, Jiangsu 221004, China.

*Corresponding author at: State Key Laboratory of Medical Genetics, School of life sciences, Central South University, 110 Xiangya Road, Changsha, Hunan, 410078, China.

E-mail address: huangshi@sklmg.edu.cn (S. Huang).

Abstract

Schizophrenia is a common neuropsychiatric disorder with a lifetime risk of 1%. A number of large scale genome wide association studies have identified numerous individual risk single nucleotide polymorphisms (SNPs) whose precise roles in schizophrenia remain unknown. Accumulation of many of these risk alleles has been found to be a more important risk factor. Consistently, recent studies showed a role for enrichment of minor alleles (MAs) in complex diseases. Here we studied the role of MAs in general in schizophrenia using public datasets. Relative to matched controls, schizophrenia cases showed higher minor allele content (MAC), especially for the sporadic cases. By linkage analysis, we identified 82 419 SNPs that could be used to predict 2.2% schizophrenia cases with 100% certainty. Pathway enrichment analysis of these SNPs identified 17 pathways, 15 of which are known to be linked with Schizophrenia with the remaining 2 associated with other mental disorders. These results suggest a role for a collective effect of MAs in schizophrenia and provide a method to genetically screen for schizophrenia.

Key Words

Minor Alleles(MAs), SNPs, Schizophrenia, Accumulation, Collective, Risk prediction

Abbreviations:

MAs, minor alleles; MAC, minor allele content; MAF, minor allele frequency; AUC, under the curve; TPR ,True positive rate

1. Introduction

Schizophrenia is one of the most frequent neuropsychiatric disorders with a lifetime risk of 1% in the general population (McGrath et al., 2008; McGrath, 2007). This disease is often chronic and places a great burden on family and society. It is characterized by the occurrence of delusions, hallucinations, disorganized speech and behavior, impaired cognition, and mood symptoms (van Os and Kapur, 2009). Data from twin, family, and adoption studies showed strong evidence that schizophrenia is predominantly a genetic disorder with high heritability (Sullivan et al., 2003).

The precise mode of Schizophrenia inheritance is unclear and risk prediction using known genetic components is presently unrealistic. Based on investigating familial syndromes with schizophrenia-like phenotypes, two rare variants have been identified as associated with schizophrenia: the 22q11 deletion (Ivanov et al., 2003; Karayiorgou et al., 1995; Sporn A Fau - Addington et al., 2004) and a 1:11translocation (Blackwood et al., 2001; Hodgkinson et al., 2004). With the advent of copy number variants (CNVs) microarray technology, an increasing number of large rare deletions have been detected in schizophrenia patients (Levinson et al., 2011; Moreno-De-Luca et al., 2010; Walsh et al., 2008). However, the effect size associated with common CNVs is smaller than initially estimated (Wray and Visscher, 2010). In addition, many candidate genes for schizophrenia have been found by Genome-wide association studies (GWAS) (O'Donovan et al., 2008; Schizophrenia Psychiatric Genome-Wide Association Study, 2011). However, these SNPs are at frequencies of 20–80% in the general population and only account for a minimal increase in risk (Mulle, 2012; Tiwari et al., 2010). It is likely that schizophrenia may be related to accumulation of many risk alleles at thousands of loci (International Schizophrenia et al., 2009).

An allele can belong to either the major or the minor allele according to its frequency in the population and the minor allele (MA) has frequency (MAF) < 0.5. Most known risk alleles are MAs (Park et al., 2011). Our previous studies have shown that the collective effects of MAs may play a role in numerous traits and diseases (Yuan et al., 2014; Zhu et al., 2015a; Zhu et al., 2015b). Specifically, enrichment of genome wide common SNPs or MAs is associated with Parkinson's disease (Zhu et al., 2015b) and lower reproductive fitness in *C.elegans* and yeasts (Zhu et al., 2015a). We here studied the role of genome wide MAs as a collective whole in schizophrenia using previously published GWAS datasets.

2. Materials and Methods:

2.1 Subjects

Two GWAS datasets of Cases and controls (phs000021.v3.p2, phs000167.v1.p1 (International Schizophrenia et al., 2009; O'Donovan et al., 2008; Stefansson et al., 2009; Suarez et al., 2006) were downloaded from database of Genotypes and Phenotypes (dbGaP). All subjects we selected for analysis are European-American

ancestry population. The SNPs of all subjects were genotyped using AFFY_6.0 in genome-wide. Principal component analysis (PCA) using the GCTA tool was performed to analyze the genetic homogeneity of the subjects (Yang et al., 2011). Outliers were excluded through selection of the principal component values (Supplementary Table S1). Duplicated subjects were excluded, and the parents of cases were also excluded.

2.2 SNPs selection

All SNPs for analysis in this study are autosomal SNPs. In addition, we excluded SNPs showing departure from the Hardy-Weinberg equilibrium ($P < 0.01$), with missing data < 5%, and with MAF< 10^{-4}. After these filters, there were 512 673 SNPs remaining (Table 1).

2.3 Statistical analysis

The Hardy-Weinberg equilibrium, missing data, MAF and logistic regression analysis were performed using PLINK Tools (Purcell et al., 2007). MAC per subject means the ratio of the total number of MAs divided by the total number of SNPs scanned (non-informative NN SNPs were excluded). The script for MAC calculation was previously described (Zhu et al., 2015b). Risk coefficient of each SNP was calculated with logistic regression test (equal to coefficient logistic regression test). The weighted risk score of a MA was calculated as follows: for homozygous MA, the risk coefficient was 1 x the coefficient, for heterozygous MA, it was 0.5 x the coefficient, for homozygous major allele, the coefficient was 0. The total weighted risk score from all MAs in a subject was obtained by summing up the weighted risk coefficient of all MAs by the script as described previously (Zhu et al., 2015b).

2.4 Genetic risk models construction and evaluation

125 prediction models were obtained from different combinations of MAF and p-value (Zhu et al., 2015b). For external cross-validation, the phs000021.v3.p2 study was used as training dataset, and the phs000167.v1.p1 study as validation dataset. Receiver operating characteristic (ROC) curves were used to describe the ability to differentiate cases and controls. True positive rate (TPR) is the proportion of cases with weighted risk scores higher than all of the controls. Area under the curve (AUC) and the TPR were calculated for each model by prism5.

For the internal cross-validation, a 10 fold cross-validation was used to test the models with good performance in external cross-validation. The models with TPR > 2% and AUC > 0.58 were chosen for internal cross-validation. Subjects in phs000021.v3.p2 were divided into 10 sub-sets randomly. When a sub-set was used as the validation data, the other 9 sub-sets were used as the training data. The cross-validation process was repeated 10 times, and the mean AUC and TPR values were calculated from these 10 results.

2.5 SNPs annotation and functional enrichment analysis

ANNOVAR (http://annovar.openbioinformatics.org/) was used to annotate SNPs (Wang et al., 2010). For functional enrichment analysis, WebGestalt (http://bioinfo.vanderbilt.edu/webgestalt/) tools were used for gene ontology annotation and pathway analysis according to Kyoto Encyclopedia of Genes and Genes (KEGG) (http://www.genome.jp/kegg/) (Wang et al., 2013; Zhang et al., 2005).

3. Results

3.1 Collective effects of minor alleles in Schizophrenia

We made use of the published GWAS datasets (phs000021.v3.p2 and phs000167.v1.p1). We first cleaned these datasets by removing outliers in principle component analysis (PCA) plots (see Figure in Supplementary Table S1). The cleaned datasets contains 1 003 cases and 1 152 controls in phs000021.v3.p2 dataset, and 828 cases and 1 068 controls in phs000167.v1.p1 dataset (Table 1). MA status of each SNP was then obtained by using the control cohort using MAF < 0.5 as cutoff. MAC of each subject was calculated, and the mean MAC of cases and controls was compared. The results showed that the mean MAC of schizophrenia cases is significantly higher than that of controls in both the phs000021.v3.p2 data (mean MAC [mean ± stdev], cases vs controls is 0.2235 ± 0.0010 vs 0.2233 ± 0.0011, P = 2.71E-06, t test) and the phs000167.v1.p1 data (cases 0.2251± 0.0011 vs controls 0.2249 ± 0.0011, P = 9.21E-04, t test, Supplementary Table S2). MAC values of both cases and controls showed normal distribution but cases were shifted slightly to the right or higher MAC values (Figure 1A and B).

To study of the role of MAC in sporadic versus familial cases of schizophrenia, we further combined the subjects in the two datasets and calculated the MAC of each subject again. Then we compared the mean MAC of sporadic schizophrenia cases (n = 1217) to that of cases with family history of a psychotic illness (n = 493). We also compared these two groups of cases with the controls (n = 2220). The results showed that the mean MAC of sporadic schizophrenia cases (cases1) was slightly higher than cases with family history (cases2) (cases1 0.221505 ± 0.001039 vs cases2 0.221503 ± 0.001044 ,P = 0.49, one-way ANOVA). The MAC difference between sporadic cases and controls was significant (cases1 0.221505 ± 0.001039 vs controls 0.221412 ± 0.001054, P = 6.4E-03, one-way ANOVA), and more so than that between cases with family history and controls (cases2 0.221503 ± 0.001044 vs controls 0.221412 ± 0.001054, P = 0.04, one-way ANOVA, Supplementary Table S2). The results confirmed the expectation that collective effects of minor alleles should play more important role in sporadic Schizophrenia cases because familial cases may involve major effect mutations in a small number of genes.

We also calculated a risk coefficient score for each SNP by logistic regression analysis and obtained a weighted risk score based on the MA status and the risk coefficient score as previously described (Zhu et al., 2015b). The MAC of each individual was then converted into a weighted risk score by summing up the weighted

risk scores of each SNP. The mean weighted risk score of cases was found to be far greater than that of controls in both datasets (cases 290.43 ± 55.86 vs controls -253.59 ± 57.90 in phs000021.v3.p2, *P* = ~0; cases 294.37 ± 50.77 vs controls -188.19 ± 50.50 in phs000167.v1.p1, *P* = ~0, t-test, Supplementary Table S2). This was apparent on a distribution plot of the weighted risk score with clearly separated cases and controls (Figure 1C and D).

3.2 Evaluation of risk prediction models

In order to get an optimal MAs model or a subset of MAs for risk prediction, we divided the MAs into 5 groups according to MAF (<0.5, <0.4, <0.3, <0.2, and <0.1, Fig 2). We performed logistic regression analysis and obtained the p-values for each SNP. Based on these p-values, we divided each group into 25 subgroups and obtained a of 125 prediction models (Fig 2, Supplementary Table 3). We then performed external cross-validation and internal cross-validation analyses to test these models. In external cross-validation, we used phs000021.v3.p2 as the training dataset and phs000167.v1.p1 as the validation dataset. We then used the receiver operator characteristic (ROC) curve to examine the discriminatory capability or area under the curve (AUC) of each model in the testing dataset. We found 17 models with AUC > 0.58 and true positive rate (TPR) > 2%. Among these models, the best TPR is 2.78%, and the best AUC is 0.6 (Fig 2 and Supplementary Table S3).

A 10 fold internal cross-validation analysis with these 17 models was further performed using phs000021.v3.p2 dataset. Each model was analyzed 10 times, and the mean AUC and TPR were calculated. The best model had AUC 0.62 (95%CI, 0.5919-0.6367) and TPR 2.2% (95%CI, 0.8786%-3.4957%) in internal cross-validation analysis, and AUC 0.6 (95%CI, 0.5678-0.6374) and TPR 2.67% (95%CI, 1.672%-3.995%) in external cross-validation analysis. There were 82 419 SNPs in this model with MAF < 0.5, and each MA had a p-value < 0.16 (Figure 2 and Supplementary Table S3).

We next tested whether the set of 82 419 SNPs is relatively specific to schizophrenia. We compared these SNPs with the previously identified 37 564 SNPs specific for Parkinson's disease (Zhu et al., 2015b). Only 1 239 SNPs were found shared between these two sets, indicating that different diseases may be linked with different sets of SNPs.

3.3 SNPs annotation

We next examined the potential functions of the 82 419 SNPs by annotating them with the ANNOVAR software (Wang et al., 2010). There were 82 834 SNPs annotation results in total due to the fact that some SNPs may lie in between two genes and could hence generate two annotation results. We found 0.956% of SNPs in exonic regions (Table 2, Supplementary Table S4).

We mapped the 82 419 SNPs to gene loci using WebGestalt tools, and found 6 588

genes (Supplementary Table S4). These genes were characterized using Gene Ontology in WebGestalt according to biological process, molecular function, and cellular component. As shown in Table 3, most of these genes were related to cytoskeletal proteins, phospholipid, anion and actin biding, GTPase regulators, transmembrane receptor protein tyrosine kinases, transmembrane receptor protein kinases, small GTPase regulators, and mental ion transmembrane transporter activities, and nucleoside-triphosphatase regulators.

Pathway analysis was carried out on these 6 588 genes according to the Kyoto Encyclopedia of Genes and Genomes (KEGG) using WebGestalt tools. A total of 17 pathways were identified with $P < 0.05$ (after multiple test adjustment) (Table 4, and Supplementary Table S5). All of these signaling pathways have been shown to have a role in schizophrenia except the gastric acid secretion and bile secretion pathway that may play a role in autism, depression (Horvath et al., 1999; Padol et al., 2012), and Alzheimer's disease (Simpson et al., 1994; Winkler et al., 2015). The pathways linked with schizophrenia include focal adhesion(Fan et al., 2013), axon guidance (Chen et al., 2011), calcium signaling pathway (Berridge, 2013; Hertzberg et al., 2015; Lidow, 2003), ECM-receptor interaction (Lubbers et al., 2014), vascular smooth muscle contraction (Sakakibara et al., 2012), arrhythmogenic right ventricular cardiomyopathy (Kawasaki et al., 2015), regulation of actin cytoskeleton (Criscuolo and Balledux, 1996; Zhao et al., 2015), long-term potentiation (Frantseva et al., 2008; Hasan et al., 2011; Salavati et al., 2015), MAPK signaling pathway (Funk et al., 2012), ABC transporters (Akamine et al., 2016), neuroactive ligand-receptor interaction (Adkins et al., 2012), GnRH signaling pathway (Brambilla F Fau - Rovere et al., 1976), salivary secretion (Toone Bk Fau - Lader and Lader, 1979), cell adhesion molecules (CAMs) (Webster et al., 1999; Zhang et al., 2015) and dilated cardiomyopathy (Finsterer and Stollberger, 2016; Volkov Vs Fau - Volkov and Volkov, 2013).

4. Discussion

In this study, we showed enrichment of MAs in schizophrenia cases relative to matched controls. We also identified a set of 82 419 SNPs that can predict with certainty a fraction of schizophrenia cases. These results are consistent with previous work on other complex diseases and traits (Yuan et al., 2014; Zhu et al., 2015a; Zhu et al., 2015b).

There were reports of male bias in schizophrenia (Aleman et al., 2003). The ratio of males to females in cases of phs000021.v3.p2 data was 2.28 but was close to 1 in the other dataset. We however did not observe significant differences in MAC values between male and female cases in both datasets. Thus, MAC may play a similar role in both sexes.

Recent studies have shown that a much larger than expected portion of the human genome may be functional (Fung et al., 2014; Hu et al., 2013; Sauna and Kimchi-Sarfaty, 2011). Genetic diversities are at saturation levels as indicated by the

observation that higher fractions of fast evolving SNPs are shared between different human groups (Yuan et al., 2017). This raises the question of what selection forces are keeping genetic diversity levels from increasing with time. By linking the total amount of SNPs or MAs in an individual to complex diseases and traits, it is clear that complex diseases could serve as a negative selection mechanism to prevent abnormal increase in SNP numbers in an individual. It is intuitively obvious that the overall property of the genome as a whole should be linked with the wellbeing of an organism. Our results here on schizophrenia further confirmed the hypothesis we put forward before that a highly complex and ordered system such as the human brain must have an optimum limit on the level of randomness or entropy in its building parts or DNAs (Zhu et al., 2015b).

It has been difficult to use any genetic markers or combinations of them to predict risk of schizophrenia. We here identified a set of 82 419 SNPs that could predict 2.2% cases with 100% specificity. Although this is still a low percentage, it may still prove valuable for prenatal diagnosis of schizophrenia. The set of 82 419 SNPs specific for schizophrenia was highly linked with pathways known to be involved in the disease, thereby validating our method of looking for disease specific set of SNPs. This set is much larger than any known from previous studies (International Schizophrenia et al., 2009). This large collection of risk alleles is not unexpected if most genome sequences are functional as explained by the maximum genetic diversity (MGD) theory (Huang, 2008; Huang, 2009; Huang, 2016), which inspired this work in the first place. Future studies using larger sample sizes may help identify a more specific set of risk SNPs that could predict higher fraction of cases.

Contributors

Shi Huang and Zuobin Zhu designed the study and wrote the protocol. Pei He, Xiaoyun Lei and Dejian Yuan managed the literature searches and analyses. Pei He undertook the statistical analysis and wrote the first draft of the manuscript. All authors contributed to and have approved the final manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

Role of the funding source

This work was supported by the National Natural Science Foundation of China grant 81171880 and the National Basic Research Program of China grant 2011CB51001 (S.H.).

Acknowledgments

We thank NINDS dbGaP GWAS Data Repository, the datasets (phs000021.v3.p2, phs000167.v1.p1) necessary for our analysis were obtained from it; we also thank all the Contributing Investigator of raw data.

References

Adkins, D.E., Khachane, A.N., McClay, J.L., Aberg, K., Bukszar, J., Sullivan, P.F., et al., 2012. SNP-based analysis of neuroactive ligand-receptor interaction pathways implicates PGE2 as a novel mediator of antipsychotic treatment response: data from the CATIE study. Schizophr Res 135(1-3), 200-201.

Akamine, Y., Sugawara-Kikuchi, Y., Uno, T., Shimizu, T., Miura, M., 2016. ANNALS EXPRESS: Quantification of the steady-state plasma concentrations of clozapine and N-desmethylclozapine in Japanese patients with schizophrenia using a novel HPLC method and the effects of CYPs and ABC transporters polymorphisms. Ann Clin Biochem(1758-1001 (Electronic)).

Aleman, A., Kahn, R.S., Selten, J.P., 2003. Sex differences in the risk of schizophrenia: evidence from meta-analysis. Arch Gen Psychiatry 60(6), 565-571.

Berridge, M.J., 2013. Dysregulation of neural calcium signaling in Alzheimer disease, bipolar disorder and schizophrenia. Prion 7(1), 2-13.

Blackwood, D.H., Fordyce A Fau - Walker, M.T., Walker Mt Fau - St Clair, D.M., St Clair Dm Fau - Porteous, D.J., Porteous Dj Fau - Muir, W.J., Muir, W.J., 2001. Schizophrenia and affective disorders--cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. Am, J. Hum Genet 69(2), 428-433.

Brambilla F Fau - Rovere, C., Rovere C Fau - Guastalla, A., Guastalla A Fau - Guerrini, A., Guerrini A Fau - Riggi, F., Riggi, F., 1976. Gonadotropin response to synthetic gonadotropin hormone-releasing hormone (GnRH) in chronic schizophrenia. Acta Psychiatr Scand. 54(2), 131-145.

Chen, S.Y., Huang, P.H., Cheng, H.J., 2011. Disrupted-in-Schizophrenia 1-mediated axon guidance involves TRIO-RAC-PAK small GTPase pathway signaling. Proc Natl Acad Sci U S A 108(14), 5861-5866.

Criscuolo, G.R., Balledux, J.P., 1996. Clinical neurosciences in the decade of the brain: hypotheses in neuro-oncology. VEG/PF acts upon the actin cytoskeleton and is inhibited by dexamethasone: relevance to tumor angiogenesis and vasogenic edema. Yale J Biol Med. 69(4), 337-355.

Fan, Y., Abrahamsen, G., Mills, R., Calderon, C.C., Tee, J.Y., Leyton, L., et al., 2013. Focal adhesion dynamics are altered in schizophrenia. Biol Psychiatry 74(6), 418-426.

Finsterer, J., Stollberger, C., 2016. Noncompaction and Dilated Cardiomyopathy in a Patient with Schizophrenia. Case Rep Cardiol 2016(2090-6404 (Print)), 7384264.

Frantseva, M.V., Fitzgerald, P.B., Chen, R., Moller, B., Daigle, M., Daskalakis, Z.J., 2008. Evidence for impaired long-term potentiation in schizophrenia and its relationship to motor skill learning. Cereb Cortex 18(5), 990-996.

Fung, K.L., Pan, J., Ohnuma, S., Lund, P.E., Pixley, J.N., Kimchi-Sarfaty, C., et al., 2014. MDR1 synonymous polymorphisms alter transporter specificity and protein stability in a stable epithelial monolayer. Cancer Res 74(2), 598-608.

Funk, A.J., McCullumsmith, R.E., Haroutunian, V., Meador-Woodruff, J.H., 2012. Abnormal activity of the MAPK- and cAMP-associated signaling pathways in frontal cortical areas in

postmortem brain in schizophrenia. Neuropsychopharmacology 37(4), 896-905.
Hasan, A., Nitsche, M.A., Rein, B., Schneider-Axmann, T., Guse, B., Gruber, O., et al., 2011. Dysfunctional long-term potentiation-like plasticity in schizophrenia revealed by transcranial direct current stimulation. Behav Brain Res 224(1), 15-22.
Hertzberg, L., Katsel, P., Roussos, P., Haroutunian, V., Domany, E., 2015. Integration of gene expression and GWAS results supports involvement of calcium signaling in Schizophrenia. Schizophr Res 164(1-3), 92-99.
Hodgkinson, C.A., Goldman, D., Jaeger, J., Persaud, S., Kane, J.M., Lipsky, R.H., et al., 2004. Disrupted in schizophrenia 1 (DISC1): association with schizophrenia, schizoaffective disorder, and bipolar disorder. Am J Hum Genet 75(5), 862-872.
Horvath, K., Papadimitriou, J.C., Rabsztyn, A., Drachenberg, C., Tildon, J.T., 1999. Gastrointestinal abnormalities in children with autistic disorder. J Pediatr 135(5), 559-563.
Hu, T., Long, M., Yuan, D., Zhu, Z., Huang, Y., Huang, S., 2013. The genetic equidistance result: misreading by the molecular clock and neutral theory and reinterpretation nearly half of a century later. Sci China Life Sci 56(3), 254-261.
Huang, S., 2008. Histone methylation and the initiation of cancer, in: T. Tollefsbol (Ed.),, Cancer Epigenetics.
Huang, S., 2009. Inverse relationship between genetic diversity and epigenetic complexity. Nature Precedings.
Huang, S., 2016. New thoughts on an old riddle: What determines genetic diversity within and between species? Genomics 108(1), 3-10.
International Schizophrenia, C., Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., et al., 2009. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460(7256), 748-752.
Ivanov, D., Kirov, G., Norton, N., Williams, H.J., Williams, N.M., Nikolov, I., et al., 2003. Chromosome 22q11 deletions, velo-cardio-facial syndrome and early-onset psychosis. Molecular genetic study. Br J Psychiatry 183, 409-413.
Karayiorgou, M., Morris, M.A., Morrow, B., Shprintzen, R.J., Goldberg, R., Borrow, J., et al., 1995. Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11. Proc Natl Acad Sci U S A 92(17), 7612-7616.
Kawasaki, K., Miyaji, K., Kodera, S., Suzuki, Y., Kanda, J., Ikeda, M., 2015. Arrhythmogenic right ventricular cardiomyopathy in a patient with schizophrenia. Clin Case Rep 3(5), 308-314.
Levinson, D.F., Duan, J., Oh, S., Wang, K., Sanders, A.R., Shi, J., et al., 2011. Copy number variants in schizophrenia: confirmation of five previous findings and new evidence for 3q29 microdeletions and VIPR2 duplications. Am J Psychiatry 168(3), 302-316.
Lidow, M.S., 2003. Calcium signaling dysfunction in schizophrenia: a unifying approach. Brain Res Brain Res Rev 43(1), 70-84.
Lubbers, B.R., Smit, A.B., Spijker, S., van den Oever, M.C., 2014. Neural ECM in addiction, schizophrenia, and mood disorder. Prog Brain Res 214(1875-7855 (Electronic)), 263-284.
McGrath, J., Saha, S., Chant, D., Welham, J., 2008. Schizophrenia: a concise overview of incidence, prevalence, and mortality. Epidemiol Rev 30(0193-936X (Print)), 67-76.
McGrath, J.J., 2007. The surprisingly rich contours of schizophrenia epidemiology. Arch Gen Psychiatry 64(1), 14-16.
Moreno-De-Luca, D., Consortium, S., Mulle, J.G., Simons Simplex Collection Genetics, C.,

Kaminsky, E.B., Sanders, S.J., et al., 2010. Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. Am J Hum Genet 87(5), 618-630.

Mulle, J.G., 2012. Schizophrenia genetics: progress, at last. Curr Opin Genet Dev 22(3), 238-244.

O'Donovan, M.C., Craddock, N., Norton, N., Williams, H., Peirce, T., Moskvina, V., et al., 2008. Identification of loci associated with schizophrenia by genome-wide association and follow-up. Nat Genet 40(9), 1053-1055.

Padol, I.T., Wang, C., Hunt, R.H., 2012. Altered physiology of acid secretion in depression-prone Flinders rats results in exacerbated NSAID and stress-induced gastric damage. Neurogastroenterol Motil 24(2), 154-163, e189.

Park, J.H., Gail, M.H., Weinberg, C.R., Carroll, R.J., Chung, C.C., Wang, Z., et al., 2011. Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. Proc Natl Acad Sci U S A 108(44), 18026-18031.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., et al., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81(3), 559-575.

Sakakibara, E., Nishida, T., Sugishita, K., Jinde, S., Inoue, Y., Kasai, K., 2012. Acute psychosis during the postictal period in a patient with idiopathic generalized epilepsy: postictal psychosis or aggravation of schizophrenia? A case report and review of the literature. Epilepsy Behav 24(3), 373-376.

Salavati, B., Rajji, T.K., Price, R., Sun, Y., Graff-Guerrero, A., Daskalakis, Z.J., 2015. Imaging-based neurochemistry in schizophrenia: a systematic review and implications for dysfunctional long-term potentiation. Schizophr Bull 41(1), 44-56.

Sauna, Z.E., Kimchi-Sarfaty, C., 2011. Understanding the contribution of synonymous mutations to human disease. Nat Rev Genet 12(10), 683-691.

Schizophrenia Psychiatric Genome-Wide Association Study, C., 2011. Genome-wide association study identifies five new schizophrenia loci. Nat Genet 43(10), 969-976.

Simpson, I.A., Chundu, K.R., Davies-Hill, T., Honer, W.G., Davies, P., 1994. Decreased concentrations of GLUT1 and GLUT3 glucose transporters in the brains of patients with Alzheimer's disease. Ann Neurol 35(5), 546-551.

Sporn A Fau - Addington, A., Addington A Fau - Reiss, A.L., Reiss Al Fau - Dean, M., Dean M Fau - Gogtay, N., Gogtay N Fau - Potocnik, U., Potocnik U Fau - Greenstein, D., et al., 2004. 22q11 deletion syndrome in childhood onset schizophrenia: an update. Mol Psychiatry. 9(3), 225-226.

Stefansson, H., Ophoff, R.A., Steinberg, S., Andreassen, O.A., Cichon, S., Rujescu, D., et al., 2009. Common variants conferring risk of schizophrenia. Nature 460(7256), 744-747.

Suarez, B.K., Duan J Fau - Sanders, A.R., Sanders Ar Fau - Hinrichs, A.L., Hinrichs Al Fau - Jin, C.H., Jin Ch Fau - Hou, C., Hou C Fau - Buccola, N.G., et al., 2006. Genomewide linkage scan of 409 European-ancestry and African American families with schizophrenia: suggestive evidence of linkage at 8p23.3-p21.2 and 11p13.1-q14.1 in the combined sample. Am J Hum Genet. 78(2), 315-333.

Sullivan, P.F., Kendler Ks Fau - Neale, M.C., Neale, M.C., 2003. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Arch Gen Psychiatry. 60(12), 1187-1192.

Tiwari, A., Zai, C., Müller, D., Kennedy, J., 2010. Genetics in schizophrenia: where are we and

what next? Dialogues Clin Neurosci 12(3), 289-303.

Toone Bk Fau - Lader, M.H., Lader, M.H., 1979. Salivary secretion in the affective disorders and schizophrenia. Acta Psychiatr Scand. 59(5), 529-535.

van Os, J., Kapur, S., 2009. Schizophrenia. lancet 374, 635-645.

Volkov Vs Fau - Volkov, V.P., Volkov, V.P., 2013. Dilated cardiomyopathy in patients with schizophrenia. Ter, Arkh 85(10), 43-46.

Walsh, T., McClellan, J.M., McCarthy, S.E., Addington, A.M., Pierce, S.B., Cooper, G.M., et al., 2008. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. Science 320(5875), 539-543.

Wang, J., Duncan, D., Shi, Z., Zhang, B., 2013. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. Nucleic Acids Res 41(Web Server issue), W77-83.

Wang, K., Li, M., Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38(16), e164.

Webster, M.J., Vawter, M.P., Freed, W.J., 1999. Immunohistochemical localization of the cell adhesion molecules Thy-1 and L1 in the human prefrontal cortex patients with schizophrenia, bipolar disorder, and depression. Mol Psychiatry 4(1), 46-52.

Winkler, E.A., Nishida, Y., Sagare, A.P., Rege, S.V., Bell, R.D., Perlmutter, D., et al., 2015. GLUT1 reductions exacerbate Alzheimer's disease vasculo-neuronal dysfunction and degeneration. Nat Neurosci 18(4), 521-530.

Wray, N.R., Visscher, P.M., 2010. Narrowing the boundaries of the genetic architecture of schizophrenia. Schizophr Bull 36(1), 14-23.

Yang, J., Lee, S.H., Goddard, M.E., Visscher, P.M., 2011. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88(1), 76-82.

Yuan, D., Lei, X., Gui, Y., Zhu, Z., Wang, D., Yu, J., et al., 2017. Modern human origins: multiregional evolution of autosomes and East Asia origin of Y and mtDNA. bioRxiv.

Yuan, D., Zhu, Z., Tan, X., Liang, J., Zeng, C., Zhang, J., et al., 2014. Scoring the collective effects of SNPs: association of minor alleles with complex traits in model organisms. Sci China Life Sci 57(9), 876-888.

Zhang, B., Kirov S Fau - Snoddy, J., Snoddy, J., 2005. WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids, Res 1(33), 741-748.

Zhang, Z., Yu, H., Jiang, S., Liao, J., Lu, T., Wang, L., et al., 2015. Evidence for Association of Cell Adhesion Molecules Pathway and NLGN1 Polymorphisms with Schizophrenia in Chinese Han Population. PLoS One 10(12), e0144719.

Zhao, Z., Xu, J., Chen, J., Kim, S., Reimers, M., Bacanu, S.A., et al., 2015. Transcriptome sequencing and genome-wide association analyses reveal lysosomal function and actin cytoskeleton remodeling in schizophrenia and bipolar disorder. Mol Psychiatry 20(5), 563-572.

Zhu, Z., Man, X., Xia, M., Huang, Y., Yuan, D., Huang, S., 2015a. Collective effects of SNPs on transgenerational inheritance in Caenorhabditis elegans and budding yeast. Genomics 106(1), 23-29.

Zhu, Z., Yuan, D., Luo, D., Lu, X., Huang, S., 2015b. Enrichment of Minor Alleles of Common SNPs and Improved Risk Prediction for Parkinson's Disease. PLoS One 10(7), e0133421.

Table Legends:

Table 1.Description of datasets used in this study.

	phs000021.v3.p2		phs000167.v1.p1	
	cases	controls	cases	controls
Number of Subjects	1003	1152	828	1068
Number of SNPs	512673	512673	512673	512673
Sex (male:female)	697:306	526:626	534:534	567:261

Table 2. Statistics of SNPs distribution.

Locus	**Number of SNPs**	**Percentage%**
Exonic	792	0.956
Intergenic	44440	53.649
Intronic	30742	37.113
UTR3	821	0.991
UTR5	80	0.097
Upstream	357	0.431
Downstream	536	0.647
upstream,downstream	8	0.010
Splicing	4	0.005
ncRNA-intronic	4735	5.716
ncRNA-exonic	319	0.385

Table 3. Categories of top-10 significantly enriched genes.

Terms	Count	P
Biological Process		
signaling	1859	2.56E-14
single organism signaling	1859	2.56E-14
cell communication	1904	2.56E-14
nervous system development	747	1.49E-12
signal transduction	1647	3.63E-12
cell adhesion	442	7.23E-12
biological adhesion	442	8.20E-12
single-organism process	2873	8.20E-12
cell projection organization	447	1.54E-10
single-multicellular organism process	2149	1.75E-10
Molecular Function		
cytoskeletal protein binding	322	3.07E-09
phospholipid binding	245	6.20E-07
anion biding	1010	3.11E-06
actin binding	179	1.20E-05
GTPase regulator activity	215	1.20E-05
transmembrane receptor protein tyrosine kinase activity	48	1.37E-05
nucleoside-triphosphatase regulator	219	1.50E-05
transmembrane receptor protein kinase activity	60	2.03E-05
small GTPase regulator activity	150	5.10E-05
mental ion transmenbrane transporter activity	190	5.10E-05
Cellular Component		
cell projection	580	4.27E-13
membrane	2824	9.06E-13
neuron projection	329	7.13E-12
cell periphery	1628	8.53E-12
plasma membrane part	813	9.65E-12
synapse	260	2.08E-11
plasma membrane	1588	7.56E-11
cell junction	371	7.80E-11
cell projection part	318	5.03E-10
intrinsic to plasma membrane	539	3.84E-09

Count: the number of genes in pathway;

P: p value adjusted by the multiple test adjustment.

Table 4. Significantly enriched KEGG pathways from WebGestalt Toolkit.

KEGG pathway	O	E	R	P
Focal adhesion	105	76.3	1.38	0.0027
Axon guidance	70	47.84	1.46	0.0027
Calcium signaling pathway	86	62.27	1.38	0.0049
ECM-receptor interaction	47	31.76	1.48	0.0112
Vascular smooth muscle contraction	58	40.42	1.44	0.0112
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	43	28.46	1.51	0.0112
Gastric acid secretion	39	25.98	1.5	0.0173
Regulation of actin cytoskeleton	99	76.71	1.29	0.0173
Long-term potentiation	38	25.16	1.51	0.0173
MAPK signaling pathway	118	94.03	1.25	0.0178
ABC transporters	27	16.91	1.60	0.0264
Neuroactive ligand-receptor interaction	114	91.97	1.24	0.0316
GnRH signaling pathway	48	34.64	1.39	0.0395
Salivary secretion	39	27.22	1.43	0.0398
Cell adhesion molecules (CAMs)	61	46.19	1.32	0.0461
Dilated cardiomyopathy	45	32.58	1.38	0.0474
Bile secretion	38	26.81	1.42	0.0485

O: the number of genes in our gene set and also in the category; E: the expected gene number in the category; R: ratio of enrichment; P: p value adjusted by the multiple test adjustment.

Figure Legends

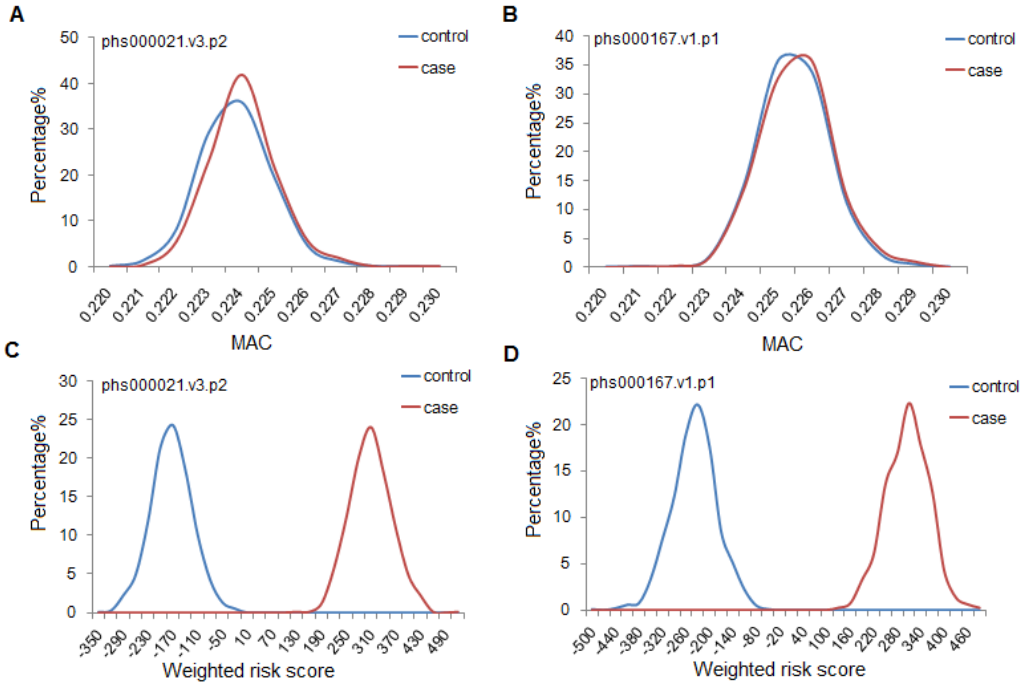

Figure 1. Minor allele distribution in cases and controls. Distribution of MAC (A, B) and weighted risk score (C, D) of case and control subjects in two different datasets.

MAC：Minor allele content of SNPs with MAF < 0.5.

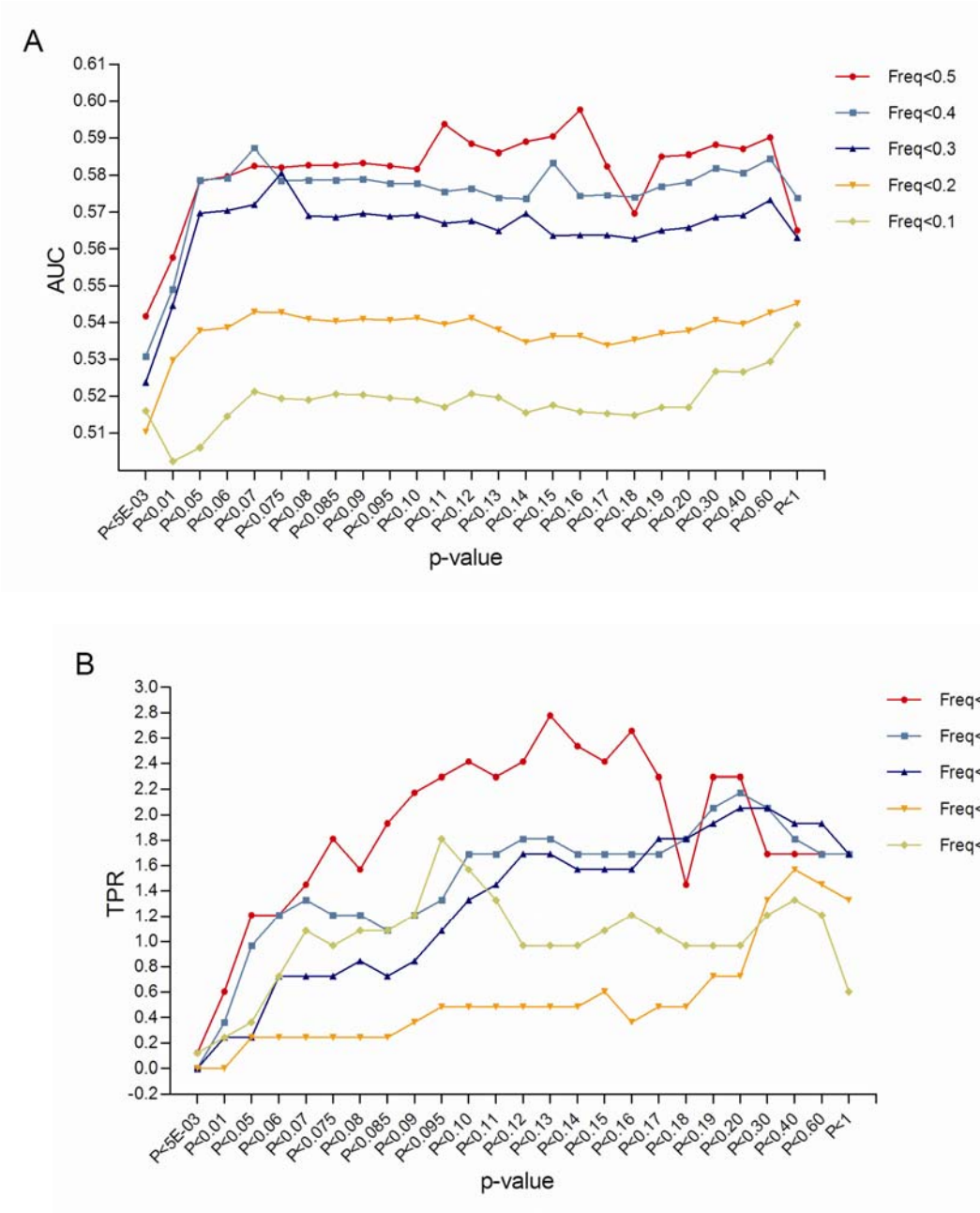

Figure 2. Discriminatory ability of different prediction models. SNPs were divided into 5 groups based on MAF, each group was further divided into 25 subgroups based on p-values from the logistic regression test and 125 prediction models were obtained. AUC (A) and TPR (B) were calculated using a training dataset and a validation dataset to evaluate the discriminatory ability.