Systematic Integration of Brain eQTL and GWAS Identifies *ZNF323* as a Novel Schizophrenia Risk Gene and Suggests Recent Positive Selection Based on Compensatory Advantage on Pulmonary Function

Xiong-Jian Luo[*,1,21], Manuel Mattheisen[2,3,4,21], Ming Li[5], Liang Huang[6], Marcella Rietschel[7], Anders D. Børglum[2,3,8], Thomas D. Als[2,3], Edwin J. van den Oord[9], Karolina A. Aberg[9], Ole Mors[3,10], Preben Bo Mortensen[3,11], Zhenwu Luo[12], Franziska Degenhardt[4], Sven Cichon[13,14], Thomas G. Schulze[15,16], Markus M. Nöthen[4], iPSYCH-GEMS SCZ working group[†], MooDS SCZ Consortium[‡], Bing Su[17], Zhongming Zhao[18], Lin Gan[19], and Yong-Gang Yao[1,20]

[1]Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences & Yunnan Province, Kunming Institute of Zoology, Kunming, Yunnan, China; [2]Department of Biomedicine and Centre for Integrative Sequencing (iSEQ), Aarhus University, 8000 Aarhus C, Denmark; [3]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus and Copenhagen, Denmark; [4]Department of Genomics, Life & Brain Center, and Institute of Human Genetics, University of Bonn, Bonn, Germany; [5]Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD; [6]First Affiliated Hospital of Gannan Medical University, Ganzhou, Jiangxi, China; [7]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty of Mannheim, University of Heidelberg, Mannheim, Germany; [8]Research Department, Psychiatric Hospital, Aarhus University Hospital, Aarhus, Denmark; [9]Center for Biomarker Research and Personalized Medicine, Virginia Commonwealth University; [10]Centre for Psychiatric Research, Aarhus University Hospital, Risskov, Denmark; [11]National Centre for Register-based Research, Aarhus University, Aarhus, Denmark; [12]Department of Microbiology and Immunology, Medical University of South Carolina, Charleston, SC; [13]Division of Medical Genetics, Department of Biomedicine, University Basel, Basel, Switzerland; [14]Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, Juelich, Germany; [15]Department of Psychiatry and Psychotherapy, University Medical Center Georg-August-Universität, 37075 Goettingen, Germany; [16]Institute of Psychiatric Phenomics and Genomics (IPPG), Ludwig-Maximilians-University Munich; [17]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China; [18]Departments of Biomedical Informatics and Psychiatry, Vanderbilt University School of Medicine, Nashville, TN 37232, USA; [19]Flaum Eye Institute and Department of Ophthalmology, University of Rochester, Rochester, NY 14642, USA; [20]CAS Center for Excellence in Brain Science, Chinese Academy of Sciences, Shanghai, 200031, China

[21]These authors contributed equally to this work.

[†]A full list of the members is provided in the supplementary data online.
[‡]A full list of the members is provided in the supplementary data online.

[*]To whom correspondence should be addressed; Key Laboratory of Animal Models and Human Disease Mechanisms, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China; tel: 86-18787077938, fax: 86-871-65180085, e-mail: luoxiongjian@mail.kiz.ac.cn

Genome-wide association studies have identified multiple risk variants and loci that show robust association with schizophrenia. Nevertheless, it remains unclear how these variants confer risk to schizophrenia. In addition, the driving force that maintains the schizophrenia risk variants in human gene pool is poorly understood. To investigate whether expression-associated genetic variants contribute to schizophrenia susceptibility, we systematically integrated brain expression quantitative trait loci and genome-wide association data of schizophrenia using *Sherlock*, a Bayesian statistical framework. Our analyses identified *ZNF323* as a schizophrenia risk gene ($P = 2.22 \times 10^{-6}$). Subsequent analyses confirmed the association of the *ZNF323* and its expression-associated single nucleotide polymorphism rs1150711 in independent samples (gene-expression: $P = 1.40 \times 10^{-6}$; single-marker meta-analysis in the combined discovery and replication sample comprising 44 123 individuals: $P = 6.85 \times 10^{-10}$). We found that the *ZNF323* was significantly downregulated in hippocampus and frontal cortex of schizophrenia patients ($P = .0038$ and $P = .0233$, respectively). Evidence for pleiotropic effects was detected (association of rs1150711 with lung function and gene expression of *ZNF323* in lung: $P = 6.62 \times 10^{-5}$ and $P = 9.00 \times 10^{-5}$, respectively) with the risk allele (T allele) for schizophrenia acting as protective allele for lung function. Subsequent population genetics analyses suggest that the risk allele (T) of rs1150711 might have undergone recent positive selection in human population. Our findings suggest that the *ZNF323* is a schizophrenia susceptibility gene whose expression may influence schizophrenia risk. Our study also illustrates a possible mechanism for maintaining schizophrenia risk variants in the human gene pool.

Introduction

Schizophrenia is a multifactorial and polygenic mental disorder with strong genetic heterogeneity and high heritability.[1,2] To date schizophrenia remains an evolutionary-genetic paradox as fecundity of schizophrenia patients (ie, strongly negative fitness) seemingly contradicts its stable prevalence of approximately 1% across all human cultures.[3–5] One potential explanation for this conundrum suggests recent positive selection based on compensatory effects on other human traits as (among others) the driving force in maintaining schizophrenia risk loci in the gene pool.[4–6] In line with this hypothesis, recent studies have identified pleiotropic effects of schizophrenia risk variants.[7,8]

Multiple promising risk variants and loci for schizophrenia have been identified by genome-wide association studies (GWAS) of schizophrenia.[1,9–11] Nevertheless, how these variants confer risk to schizophrenia remains elusive. Moreover, the identified susceptibility variants only explain a small fraction of the disease variance. A recent study has suggested that more than 8000 common loci are involved in the etiology of schizophrenia,[11] thus indicating that many common risk loci with potentially small effect sizes and/or with infrequent risk alleles are still to be detected.

Although more schizophrenia risk loci remain to be identified, accumulating evidence strongly suggests that changes in gene expression rather than alterations in protein function play a key role in the pathogenesis of schizophrenia.[12–14] Accordingly, recent studies have used integrative strategies to combine results from association studies and gene expression analyses.[15,16] The successful identification of schizophrenia susceptibility genes by these integrative methods supports the important roles of gene expression alteration in the pathogenesis of schizophrenia. However, currently there is no study that systematically integrated brain expression quantitative trait loci (eQTL) data and GWAS of schizophrenia to identify the potential schizophrenia risk genes.

In this study, we followed an integrated data analysis approach using *Sherlock*[17] and analyzed eQTL from normal human brain samples[18] together with association data from a large GWAS for schizophrenia.[11] To further characterize the potential role of our identified loci in schizophrenia etiology, we followed up and refined our results in independent gene expression and large-scale genetic datasets. Moreover, we subsequently tested for pleiotropic effects (ie, a simultaneous effect of our identified risk loci on other human traits) and tried to shed light on the potential presence of positive natural selection (as a consequence of compensatory effects on the traits implicated by the pleiotropic analyses).

Methods

Sherlock integration analysis

Considering that most of the disease-associated variants identified by recent GWAS are located in noncoding regions,[19] it is likely that these risk variants affect gene expression rather than protein function. To infer genes whose expression changes may contribute to the etiology of complex diseases, He et al[17] developed a statistical framework (named *Sherlock*) to identify the potential disease-associated genes through matching eQTL and single-nucleotide polymorphism (SNP) associations from GWAS. The underlying assumption is that the expression level of a specific gene(s) may influence the risk of a disease (eg, schizophrenia). Therefore, genetic variation (both in *cis*, ie, close to the gene, and in *trans*, ie, distant to the gene or on different chromosomes) that perturbs gene expression may affect the risk of this disease. For a given gene there may be several variants (expression-associated SNPs [eSNPs]) in the genome that act together to regulate the expression level of this gene. Genetic variation at these eSNPs would influence the expression level of this gene, which in turn might affect schizophrenia susceptibility. Hence, cumulative evidence from both eQTL and GWAS findings is indicative of association with schizophrenia. *Sherlock* first searches for all eSNPs of each gene using the whole genome eQTL data from the human frontal cortex.[18] For each eSNP, *Sherlock* will then evaluate its association with schizophrenia using genome-wide association (GWA) data of schizophrenia.[11] There are three different scenarios: (1) If the eSNP of a specific gene is also associated with schizophrenia in GWAS, a positive score would be given; (2) If the eSNP of this gene is not associated with schizophrenia, a negative score would be assigned; and (3) association only in GWAS (ie, non-eSNPs) does not alter the score. The total score of a gene increases along with the increase of the number of SNPs with combined evidence (SNPs that are associated with schizophrenia and expression). Finally, *Sherlock* identifies gene-disease associations through matching genetic signatures of gene expression traits to that of disease.[17] Statistical inference was performed using Bayes statistical framework[17] and Bayes factor (BF, the probability of the observed data under a specific model) of each SNP was calculated separately. For each gene, *Sherlock* computes individual logarithm of BF (LBF) for each SNP pair in the alignment, and the sum of these constitutes the final LBF score for the gene. The value of the LBF score of a gene reflects the strength of evidence (ie, a larger LBF represents higher probability that this gene is associated with the disease). *Sherlock* utilizes moderately associated *cis* and *trans* eSNPs (without introducing many false signals) instead of solely focusing on eSNPs that meet stringent cutoff criteria for both eQTL and GWAS results. *P* values of the BFs were calculated with simulation using Bayes/non-Bayes compromise.[20] False discovery rate (FDR) was calculated using Benjamini–Hochberg procedure[21] at *P*

threshold of 10^{-5}. More detailed information about the statistical model and the underlying algorithm is provided in the original paper by He et al.[17]

Brain Expression (eQTL) Data

Considering that schizophrenia is a mental disorder that mainly originates from abnormal brain function, it is reasonable to assume that the most appropriate tissue for an integrative identification of schizophrenia risk genes is the human brain. Therefore, we used brain expression data (human prefrontal cortex [PFC]) from a previous study on 193 neuropathologically normal human brain samples.[18] All individuals enrolled in this study were of self-reported European ancestry. Genotyping was performed on an Affymetrix Human Mapping 500K Array Set and expression profiles were identified using an Illumina HumanRefseq-8 Expression BeadChip. *Cis* eSNPs were defined as SNPs situated within the transcript of the respective gene or within 1 Mb of either the 5′ or the 3′ end of the transcript. More information about statistical analyses and procedures is provided in the original study.[18]

Schizophrenia GWAS Data

We used summary statistics from a large-scale schizophrenia GWAS.[11] In brief, the study represents a meta-analysis of a recent schizophrenia GWAS from the Psychiatric Genomics Consortium (PGC), hereafter referred to as "PGC1,"[9] and independent data from a Swedish schizophrenia GWAS ("SWE").[11] The study comprised a total of 13833 schizophrenia cases and 18310 controls ("SWE + PGC1")[11] and identified 22 risk loci for schizophrenia. Most of the subjects were of European ancestry. DNA was extracted from peripheral blood and samples were genotyped using Affymetrix (5.0 and 6.0) and Illumina OmniExpress chips according to the manufacturers' instructions. The association analysis was first performed in SWE samples (5001 cases and 6243 controls) with PLINK[22] using imputed SNP dosages and the principal components as covariates. Then the results from SWE samples were meta-analyzed with the results from PGC1 (8832 cases and 12067 controls) using an inverse-weighted fixed-effects model.[23] Genome-wide genetic association information (*P* values of a total of 9898078 SNPs) from SWE + PGC1 was used as input in this study. Detailed information about sample ascertainment, diagnosis, genotyping quality control, and statistical analyses can be found in the original study and PGC website (https://pgc.unc.edu).[11]

Replication and Refinement of Gene Expression Results

Four well-characterized expression databases were used to follow-up on the gene expression results that contributed to identification of *ZNF323* as our top schizophrenia risk gene. In addition to the replication, the follow-up comprised a deeper characterization of gene expression patterns for *ZNF323* in the brain (eg, healthy individuals

vs schizophrenia patients) and identification of additional eSNPs of *ZNF323*.

A brief description of the gene expression resources is provided below; more detailed information can be found in the original studies.[24-27] (1) BrainCloud.[24] BrainCloud contains genetic information and whole transcriptome expression data from postmortem dorsolateral PFC (DLPFC) of 269 normal human subjects (ie, without neuropathological or neuropsychiatric diagnosis). The data in BrainCloud is aimed at exploring temporal dynamics and genetic control of transcription in the DLPFC across the lifespan, ie, from fetal development through ageing. (2) The Genotype-Tissue Expression project (GTEx).[25] Compared with BrainCloud, which focuses on the DLPFC, the GTEx contains information at the level of both genetic variation and gene expression from a diverse set of human tissues including different brain regions. So far, 3797 tissues from 150 postmortem donors have been collected and subsequently analyzed using a RNA sequencing (RNA-seq)–based gene expression approach. (3) The Stanley Neuropathology Consortium Integrative Database (SNCID).[26] The SNCID contains gene expression information for different brain regions (including hippocampal tissue and PFC) of 61 individuals, including 20 bipolar disorder patients, 20 schizophrenia cases, and 21 healthy controls. (4) The Human Brain Transcriptome (HBT) database.[27] The HBT comprises the transcriptome of 16 regions: the cerebellar cortex, mediodorsal nucleus of the thalamus, striatum, amygdala, hippocampus, and 11 areas of the neocortex.

Replication and Refinement of GWAS Results

In order to replicate the GWAS results that contributed to our integrated results from the discovery step and in order to follow-up on newly identified eSNPs in the eQTL replication step, we utilized four well-characterized GWAS datasets:

(1) Independent Samples From the German MooDS SCZ Working Group.[28]

This sample contained 1332 schizophrenia cases and 866 controls. It represents the nonoverlapping part (PGC, "SGENE-Bonn") of the MooDS SCZ Consortium sample that has been described in detail in study of Priebe et al.[28] Thus, there is no overlap between individuals from our German replication sample and subjects in the discovery step from the Ripke et al study (PGC1 + SWE).[11] The study was approved by the ethics committees of all study centers and written informed consent was obtained from all participants prior to inclusion. All patients were recruited from consecutive admissions to psychiatric inpatient units and of German descent according to self-reported ancestry. Venous blood samples were obtained from all participants and genotyping was performed using Illumina's BeadChips, including HumanHap550v3, Human610-Quadv1, and

Human660W-Quad. More detailed information on sample description, genotyping, quality control of controls utilized in this study can be found in the paper of Muglia et al.[29]

(2) Independent Family–Based Replication Sample From a Recent Study.[30] This sample consisted of 6298 individuals (including 3286 schizophrenia cases) from 1811 nuclear families. There were three independent family–based samples included in this family-based replication sample, the European family–based sample (2740 individuals, 794 families, and 1420 schizophrenia cases), the Asian family–based sample (2296 individuals, 579 families, and 1222 schizophrenia cases), and the African family–based sample (1262 individuals, 438 families, and 644 schizophrenia cases). Because three subsamples included in this family-based samples, genetic association analysis in nuclear families was performed within each ancestral group using UNPHASED program[31] to minimize the risk of population stratification effects. The results were then combined to obtain an overall *P* value. All of the individuals from this family-based sample do not overlap with the subjects from the discovery step of Ripke et al (PGC1 + SWE sample[11]). More detailed information about sample description (recruitment of schizophrenic patient and healthy controls, diagnosis), genotyping, quality control, population stratification analysis, and statistical analysis can be found in the original paper.[30]

(3) Two Independent Danish GWAS Datasets.[32] This sample composed of two datasets with a total of 1735 schizophrenia cases (876 and 859, respectively) and 1749 controls (871 and 878, respectively). There is no overlap between individuals from our Danish replication sample and subjects in the discovery step from the Ripke et al study ("PGC1 + SWE"[11]). A detailed description of the Danish study samples is provided elsewhere.[32] In brief, the samples used in this study here comprised the discovery step of a Danish schizophrenia GWAS[32] and a larger part of its initial replication samples that was in the meantime genotyped using Illumina's HumanCoreExome BeadChip. All involved samples were initially employed in two matched case-control designs (see above) and identified through the Danish Psychiatric Central Register.[33] The DNA was accessed through the Danish Newborn Screening Biobank[34] and was extracted from dried blood spots using Extract-N-Amp Blood PCR kit (Sigma Aldrich, Seelze, Germany) and subsequently whole genome-amplified in triplicates using the RepliG kit (Qiagen, Venlo, The Netherlands).[35] Subsequent genotyping for the first sample (discovery) was performed using Illumina's Human 610-quad beadchips (Illumina) and the second sample as described above. Diagnosis of all schizophrenia cases was based on ICD-10-DCR (The ICD-10 Classification of Mental and Behavioural Disorders Diagnostic Criteria for Research;

F20). Summary statistics of all four studies (comprising a total of 6353 schizophrenia patients) were meta-analyzed using PLINK.[22] There is no overlap between individuals from our replication samples and subjects in the PGC1 + SWE.

Pleiotropic Effects Analyses of the Identified Risk Loci

The National Human Genome Research Institute's (NHGRI) GWAS Catalog[36,37] comprises a curated list of results from all GWAS in humans to date (as of April 16, 2014, 1902 publications and 13 156 SNPs are listed) and therefore provides a valuable resource to study pleiotropic effects. We first explored whether *ZNF323* and its eSNPs were associated with other traits. We next investigated whether the observed associations of eSNPS of *ZNF323* with schizophrenia and lung function could have been the result of confounding. More detailed information about the pleiotropy analyses can be found in supplementary methods.

Bioinformatic Analysis and Functional Prediction of rs1150711

We examined the potential chromatin interactions between the genomic region containing rs1150711 and its neighboring regions using the Hi-C chromatin interactions (from the study of Dixon et al).[38] In addition, we performed bioinformatic analyses to predict the potential functional consequences of rs1150711 using Regulome DB (http://www.regulomedb.org).[39] Multiple types of data (eg, ChIP-seq, DNase-seq, and eQTLs) from the Encyclopedia of DNA Elements (ENCODE) project[40] and other sources were integrated into the Regulome DB to estimate the possible function of candidate SNPs.

Population Genetic Analyses

We utilized different approaches to investigate whether our top eSNPs and/or *ZNF323* underwent recent positive selection in human populations. (1) *Analysis of extended haplotype homozygosity (EHH)*: Haplotypes with unusually high EHH and high population frequency suggest recent positive selection that has driven rapid increase in frequency of new variants or haplotypes in the population.[41] We used Sweep,[41] the *rehh* package in R[42] and data from the HapMap project[43] and the 1000 Genomes project[44] to estimate the EHH of the region surrounding the eSNP of *ZNF323*. (2) *Calculation of integrated haplotype score (iHS)*: The iHS derived from EHH[45] represents an indicator of recent positive selection. Large negative or positive values indicate unusually long haplotypes (that carry the derived or the ancestral allele), whereas small values of iHS indicates no obvious recent positive selection. An |iHS| > 2.0 indicates a significant (corresponding to

$P < .05$) positive selection signal.[45] We used Haplotter[45] and data from the HapMap project[43] and the 1000 Genomes project[44] to calculate iHS. (3) *The Composite of Multiple Signals (CMS) test*[46]: The CMS integrates information from both EHH and iHS (together with Fst and two additional frequency-spectra–based tests) and is thought to have greater power compared with the two aforementioned tests. (4) *Level of allelic differentiation in different populations (Fst)*: Alleles, which experienced positive selection in a specific population or geographic region, have larger allelic differentiation (frequency differences) between populations compared with neutrally evolved alleles. We compared allele frequencies of our eSNPs in global populations by using genotype information from the Human Genome Diversity Project (HGDP) selection browser (http://hgdp.uchicago.edu/cgi-bin/gbrowse/HGDP/),[47] which contains the allele frequency data of 53 worldwide populations. To perform neutrality tests (including Tajima's D,[48] Fu and Li's D[49] tests, and Fay and Wu's H test[50]), we obtained sequencing data (including 99 Europeans, 103 Han Chinese, and 108 Africans) encompassing the entire *ZNF323* gene (chr6:28,292,515-28,304,152, hg19) from the 1000 Human Genomes Project.[44] Neutrality tests were carried out using DnaSP software (version 5).[51] Fst (a measure of population differentiation) of rs1150711 was calculated using the method described in study of Cheng et al.[52] For clarity, a workflow diagram illustrating these analyses is provided in supplementary figure S1.

Results

Sherlock Integrative Analyses (Discovery Step) Identified ZNF323 as a Novel Schizophrenia Risk Gene

Through systematic integration of brain eQTL[18] and SNP associations (a total of 9 898 078 SNPs) from a large GWAS of schizophrenia (PGC1 + SWE, including 13 833 schizophrenia cases and 18 310 controls)[11] using *Sherlock*[17] (see "Methods" section and supplementary figure S1), *ZNF323* (6p22.1) showed the most significant association with schizophrenia (LBF = 7.09, $P_{sher} = 2.2 \times 10^{-6}$, FDR = 3.69×10^{-3}) (table 1). One eSNP of *ZNF323* (rs2859365) showed significant association with *ZNF323* expression ($P_{eQTL} = 5.90 \times 10^{-8}$) and suggestive evidence for association with schizophrenia ($P_{GWAS} = 2.07 \times 10^{-7}$) (table 2). Homozygous carriers of the risk allele (A) of rs2859365 showed a reduced expression of *ZNF323*. Results for all genes with LBF > 5.0 and *P*-value < 10^{-4} are listed in table 1. Supporting SNPs for top predicted genes for schizophrenia are listed in supplementary table S1.

Table 1. Integrative Analysis (*Sherlock*) of Schizophrenia GWAS and Brain eQTL Reveals That *ZNF323* Had the Most Significant Association With Schizophrenia

Gene symbol	LBF[a]	*P*-value[b]	Supporting SNP[c] (*cis* or *trans*)	P_{eQTL}[d]	P_{GWAS}[e]	FDR[f]
ZNF323	7.09	2.22×10^{-6}	rs2859365 (*cis*)	5.90×10^{-8}	2.07×10^{-7}	3.69×10^{-3}
NAPSA	7.04	2.22×10^{-6}	rs10226475 (*trans*)	6.74×10^{-6}	2.02×10^{-10}	3.69×10^{-3}
LOC378135	6.76	2.22×10^{-6}	rs134169 (*trans*)	6.17×10^{-6}	2.50×10^{-5}	3.69×10^{-3}
EPB41L2	6.66	2.22×10^{-6}	rs7752195 (*trans*)	1.17×10^{-6}	1.82×10^{-5}	3.69×10^{-3}
GLT8D1	5.83	4.43×10^{-6}	rs2240920 (*cis*)	5.22×10^{-16}	7.30×10^{-6}	7.35×10^{-3}
RCN1	5.83	4.43×10^{-6}	rs3800917 (*trans*)	6.56×10^{-7}	6.27×10^{-10}	7.35×10^{-3}
ALMS1	5.56	1.11×10^{-5}	rs6753344 (*cis*)	1.37×10^{-9}	5.00×10^{-5}	1.84×10^{-2}
EIF4A2	5.39	1.11×10^{-5}	rs17002034 (*trans*)	4.00×10^{-7}	1.40×10^{-6}	1.84×10^{-2}
MPV17L2	5.30	1.11×10^{-5}	rs7752195 (*trans*)	1.54×10^{-6}	1.82×10^{-5}	1.84×10^{-2}
LSAMP	5.13	1.11×10^{-5}	rs4976976 (*trans*)	4.53×10^{-7}	6.23×10^{-6}	1.84×10^{-2}
CSNK2B	5.11	1.11×10^{-5}	rs707939 (*cis*)	5.35×10^{-6}	8.23×10^{-5}	1.84×10^{-2}
PLCD4	5.09	1.55×10^{-5}	rs2160567 (*trans*)	7.40×10^{-7}	8.29×10^{-6}	2.57×10^{-2}

[a]LBF (Log Bayes Factors for each gene) evaluates whether the combined evidence from GWAS and expression studies support a gene being associated with schizophrenia. For example, a LBF of 4.6 means that the alternative hypothesis (associated with Schizophrenia) is 100 times (*exp*[4.6] = 100) more likely than the null hypothesis (no association). High Bayes factors tend to correlate with low *P*-values. Several eSNPs may be associated with the expression of a gene. For each of these eSNP, if it is also associated with schizophrenia in GWAS, a positive score would be given; otherwise, a negative score. Association only in GWAS (ie, non-eSNPs) does not alter the score. The total score of a gene increases with the number of SNPs with combined evidence (SNPs that are associated with schizophrenia and expression). For each gene, *Sherlock* computes individual LBFs for each SNP pair in the alignment, and the sum of these constitutes the final LBF score for the gene.
[b]*P*-value from *Sherlock* integrative analysis.
[c]SNP with the highest LBF score.
[d]*P*-value for eQTL SNP from the gene expression study.[18]
[e]*P*-value for eQTL SNP from the GWAS of schizophrenia.[11]
[f]FDR was calculated at $P = 10^{-5}$ threshold. It should be noted that some genes with different LBF scores may have the same *P*-values. This is because that Sherlock computes *P*-values by finite permutations. If the number of permutation is not extremely high, it is possible that two different values of LBF correspond to the same *P*-values (their rankings are the same). Genes with LBF > 5.0 and *P*-value < 10^{-4} are listed.

Replication and Refinement of eQTL and GWAS Results for ZNF323

To replicate our results, we tested the association between rs2859365 and expression of *ZNF323* in an independent gene expression dataset, the BrainCloud.[24] The result confirmed the significant association of rs2859365 with *ZNF323* expression in the PFC ($P = 1.60 \times 10^{-7}$) (figure 1A). Again, homozygous risk allele carriers have a reduced *ZNF323* expression compared with the nonrisk allele individuals (figure 1A).

We further tested whether rs2859365 is associated with *ZNF323* expression in brain regions other than the PFC in a third independent expression dataset (the GTEx).[25] We found that rs2859365 showed significant association with *ZNF323* expression in the hippocampus ($P = 5.0 \times 10^{-3}$) (figure 1B), the caudate basal ganglia ($P = 2.0 \times 10^{-2}$) (figure 1C), and

Table 2. Multiple eSNPs of *ZNF323* Are Significantly Associated With Schizophrenia in the PGC1 + SWE Sample (13 833 Cases and 18 310 Controls)

Gene symbol	SNP ID	Chromosome: Position	eQTL *P* value[a]	GWAS *P* value[b]
ZNF323	**rs1150711**	6:28208535	1.40×10^{-6}	2.76×10^{-9}
	rs2859365	6:28391465	1.60×10^{-7}	2.07×10^{-7}
	rs853684	6:28294550	9.52×10^{-7}	4.73×10^{-7}
	rs213240	6:28315875	2.89×10^{-6}	5.40×10^{-7}
	rs1997660	6:28269663	9.99×10^{-6}	6.23×10^{-8}
	rs853693	6:28282648	9.99×10^{-6}	6.04×10^{-8}
	rs12214383	6:28223731	1.48×10^{-5}	2.56×10^{-6}
	rs916403	6:28408258	4.63×10^{-5}	3.24×10^{-6}
	rs9380069	6:28203300	4.76×10^{-5}	2.31×10^{-7}
	rs9461446	6:28227633	5.77×10^{-6}	0.78
	rs9461448	6:28263721	3.87×10^{-5}	0.83

Boldface letters were used to highlight the genetic variants that we are interested in.
[a]eQTL *P* value shows the association significance between this SNP and *ZNF323* expression level in human brain (Brain eQTL data is from the study of Colantuoni et al[24]).
[b]GWAS *P* value represents the association significance between this SNP and schizophrenia (Based on association data from the PGC1 + SWE samples, included 13 833 schizophrenia cases and 18 310 controls[11]).
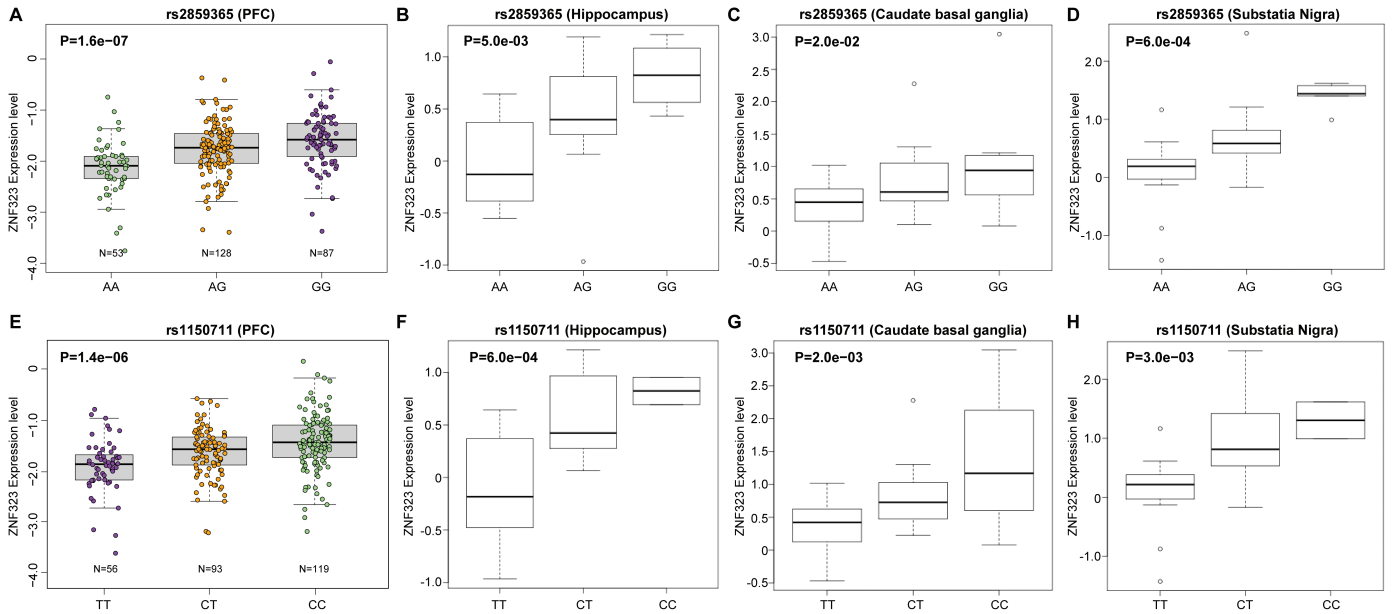
Fig. 1. rs2859365 and rs1150711 are significantly associated with *ZNF323* expression in human prefrontal cortex (PFC). (A) rs2859365 is significantly associated with the expression level of *ZNF323* in the PFC of normal human subjects ($n = 268$) across the lifespan ($P = 1.60 \times 10^{-7}$). The individuals with AA genotype have lower *ZNF323* expression level. (B–D) rs2859365 is also significantly associated with the *ZNF323* expression in the hippocampus (B), the caudate basal ganglia (C), and the substatia nigra (D). (E) rs1150711 is significantly associated with the expression level of *ZNF323* in the PFC of normal human subjects ($n = 268$) across the lifespan ($P = 1.40 \times 10^{-6}$). Individuals carrying the risk rs1150711 TT genotype (risk allele: T allele) have significant lower *ZNF323* expression level than CC genotype carriers. (F–H) rs1150711 is also significantly associated with the *ZNF323* expression in the hippocampus (F), the caudate basal ganglia (G), and the substatia nigra (H).

the substantia nigra ($P = 6.0 \times 10^{-4}$) (figure 1D). Consistent with the findings from the PFC, we found that homozygous risk allele carriers for rs2859365 did show a reduced expression of *ZNF323* in the hippocampus, caudate basal ganglia and the substantia nigra (figures 1B–D).

Similar to our validations on the eQTL results, we also attempted to replicate the GWAS signal (PGC1 + SWE study) for rs2859365 ($P = 2.07 \times 10^{-7}$) in independent schizophrenia datasets. Meta-analysis of six independent datasets comprising 6353 schizophrenia patients and 5618 controls yielded a *P*-value of .111 (overall combined *P*-value for PGC1 + SWE and the replication was 1.40×10^{-7}) with the same effect allele (A) and effect direction as in the PGC1 + SWE study (table 3).

Evidence for Additional eSNPs of ZNF323

Considering that the expression of *ZNF323* is significantly associated with schizophrenia, we hypothesized that other eSNPs of *ZNF323* might also be associated with schizophrenia. We therefore identified the other eSNPs of *ZNF323* by using BrainCloud.[24] Only eSNPs with a *P*-value less than 1.0×10^{-4} were considered. In total, we identified 11 SNPs that showed an association with *ZNF323* expression at the aforementioned *P*-value threshold (table 2). We further investigated the linkage disequilibrium (LD) pattern among the 11 eSNPs in Europeans based on genotype data from the 1000 Genomes project.[44] Our analysis revealed two LD blocks (supplementary figure S2) with the initial lead eSNP rs2859365 (that was identified through the integrated analysis) located in haplotype block 2. Among the eSNPs in haplotype block 1, rs1150711 showed association with *ZNF323* expression in the BrainCloud dataset ($P = 1.40 \times 10^{-6}$) (figure 1E) and reached genome-wide significance in the PGC1 + SWE GWAS data ($P = 2.76 \times 10^{-9}$) (table 2). Further investigation of rs1150711 in the GTEx expression dataset revealed that rs1150711 is significantly associated with *ZNF323* expression in the hippocampus ($P = 6.0 \times 10^{-4}$), the caudate basal ganglia ($P = 2.0 \times 10^{-3}$), and the substantia nigra ($P = 3.0 \times 10^{-3}$) (figures 1F–H). Consistent with results of rs2859365, homozygous risk allele (T) carriers have a reduced *ZNF323* expression compared with the nonrisk allele individuals (figures 1E–H). Additional support was gained from meta-analysis of our independent replication samples for rs1150711 ($P = 0.037$; overall combined *P*-value for PGC1 + SWE and the replication was 6.85×10^{-10}) (table 3). To explore if rs1150711 is a *cis* or *trans* eQTL of *ZNF323*, we examined the genomic location of rs1150711 and *ZNF323* and found that rs1150711 is located in ~84-kb downstream of *ZNF323*. As rs1150711 and *ZNF323* are located closely on same chromosome, according to the criteria described in previous papers,[17,53–55] rs1150711 is a *cis* eQTL of *ZNF323*.

SNP rs1150711 is located in the MHC region, considering the high LD among genes in this region, it is possible that the association between rs1150711 and schizophrenia is attributed to another causal variant that is highly linked with rs1150711. To test if there are SNPs that are

Table 3. Discovery and Replication Results of rs1150711 and rs2859365 in Schizophrenia Samples Comprising a Total of 44 123 Individuals

SNP ID	Sample (cases)	Allele[a]		OR[b]	*P* value[c]
		EA	OA		
rs2859365	European sample (1420)	A	G	1.030	0.659
	Asian sample (1222)	A	G	1.077	0.266
	African sample (644)	A	G	1.008	0.941
	German sample (1332)	A	G	1.155	0.025
	Danish sample 1 (876)	A	G	0.964	0.609
	Danish sample 2 (859)	A	G	1.027	0.702
	Combined replication (6 353)	*A*	*G*	*1.047*	*0.111*
	PGC1 + SWE (13 833)	A	G	1.095	2.07×10^{-7}
	Overall combined (20 186)	**A**	**G**	**1.082**	1.40×10^{-7}
rs1150711	European sample (1420)	T	C	1.109	0.140
	Asian sample (1222)	T	C	1.057	0.413
	African sample (644)	T	C	1.134	0.329
	German sample (1332)	T	C	1.209	0.007
	Danish sample 1 (876)	T	C	0.920	0.289
	Danish sample 2 (859)	T	C	1.009	0.908
	Combined replication (6353)	*T*	*C*	*1.068*	*0.037*
	PGC1 + SWE (13 833)	T	C	1.119	2.76×10^{-9}
	Overall combined (20 186)	**T**	**C**	*1.105*	6.85×10^{-10}

Boldface letters and italics were used to highlight the genetic variants that we are interested in.
[a]EA: effect allele, OA: other allele.
[b]OR (odds ratio) is based on effect allele.
[c]Two-tailed *P* values. Combined analyses are based on a fixed effects meta-analysis of individual study results.

in high linkage disequilibrium with rs1150711, we conducted LD analysis using genotype data from the 1000 Genomes project.[44] LD analysis revealed that 7 SNPs are highly linked ($r^2 > 0.80$) with rs1150711 (supplementary figure S3 and supplementary table S2). We investigated the association between these SNPs and schizophrenia and found that rs1150711 showed the most significant association with schizophrenia (supplementary figure S3 and supplementary table S2). Collectively, these results suggest that SNP rs1150711 may represent an independent schizophrenia-associated variant. However, given the degree of LD in this region (there are many SNPs showed moderate LD ($r^2 > 0.6$) with rs1150711), we could not completely rule out the potential effect of synthetic association.

Pleiotropic Analyses for rs1150711

Our above data showed a significant association between rs1150711 and *ZNF323* expression in human brain tissues. To further explore if rs1150711 is also associated with expression of other genes adjacent to *ZNF323*, we investigated the pleiotropic effects of rs1150711 using BrainCloud[24] and NCBI eQTL browser (which contains well-characterized brain eQTL dataset).[56] Both of these two datasets used brain tissues from the frontal cortex, a region associated with schizophrenia.[57,58] We found that rs1150711 showed the most significant association with *ZNF323* expression in BrainCloud database. SNP rs1150711 is absent in the NCBI eQTL browser, we therefore queried rs2859365. Again, we found the only eQTL gene for rs2859365 is *ZNF323* ($P = 2.93 \times 10^{-8}$). These results suggest that *ZNF323* may represent the most possible biological candidate gene behind rs1150711 (or rs2859365).

Bioinformatic Analysis of rs1150711 and Functional Prediction

To explore the potential mechanisms underlying the significant association between rs1150711 and *ZNF323* expression, we performed bioinformatic analyses. Recent studies showed that the chromatin interaction plays an important role in regulating gene expression.[38,59] We therefore extracted the chromatin interaction data generated by Dixon et al (http://yuelab.org/hi-c/).[38] We found there was chromatin interaction (topologically associated domains [TAD]) between the genomic region containing rs1150711 and its neighboring regions (supplementary figure S4). However, due to the fact that TAD spans about 700 kb and there are many genes in this region, more work is needed to elucidate if rs1150711 can directly regulate the expression of *ZNF323*. We also utilized the ENCODE data to predict the potential functional consequences of rs1150711 using Regulome DB (http://www.regulomedb.org).[39] We found that SNP rs1150711 have

RegulomeDB score of 3a (supplementary figure S5), indicating that at least three lines of evidence support this SNP is functional. First, ChIP-seq data revealed that rs1150711 is located in a transcription factor (TF) binding region. TFs MAFK and RFX3 can bind to the genomic sequence containing rs1150711. Second, rs1150711 is located in a well-defined motif region (Evi-1) and nucleotide substitution at rs1150711 may affect TFs binding. Third, DNase-seq data showed that rs1150711 is located in DNase peaks. These results implied that rs1150711 may have potential functional role in regulating *ZNF323* expression. However, more work is needed to test whether SNP rs1150711 is functional.

Significant Downregulation of ZNF323 in the Frontal Cortex and Hippocampus of Schizophrenia Patients

Sherlock integrative analysis revealed that the expression changes of *ZNF323* may contribute to schizophrenia risk. To further validate this, we tested whether *ZNF323* expression in the frontal cortex and/or hippocampus is dysregulated in schizophrenia patients using the SNCID data.[26] We found that *ZNF323* expression is significantly downregulated in schizophrenia patients compared with normal controls in both frontal cortex ($P = .023$) and hippocampus ($P = .00376$) (figure 2). Interestingly, we noticed that the T allele of rs1150711, which is associated with lower *ZNF323* expression in PFC and hippocampus, is significantly enriched in schizophrenia cases (table 3). We further examined if other neighboring genes of rs1150711 are differentially expressed in schizophrenia cases and controls using the data from the SNCID.[26] Genes located in a 100-kb window surrounding rs1150711 were examined. In addition to *ZNF323*, six genes (*ZNF192*, *ZNF193*, *ZNF187*, *ZNF427*, *NKAPL*, *PGBD1*) are located in the 100-kb window surrounding rs1150711. We found that these genes did not show differential expression in frontal cortex of schizophrenia cases and controls. Collectively, these findings provide further evidence that support the potential involvement of *ZNF323* expression in schizophrenia susceptibility.

Pleiotropic Analyses for ZNF323

We explored whether *ZNF323* is also associated with other human traits. A query of the NHGRI GWAS catalog (accession date April 01, 2014) for *ZNF323* (or its alias names) revealed that the only entry listed was a GWAS for lung function.[60] The trait(s) analyzed in this GWAS were the ratio of "forced expiratory volume in 1 second (FEV_1)" to "forced vital capacity (FVC)" as well as FEV_1. The strongest association for a SNP in *ZNF323* in the aforementioned GWAS was rs6903823 (FEV_1/FVC: $P = .001$ for joint meta-analysis of all stages; FEV_1: $P = 2.18 \times 10^{-10}$) (supplementary table S3). Analyses of genetic correlation between rs6903823 and rs2859365/
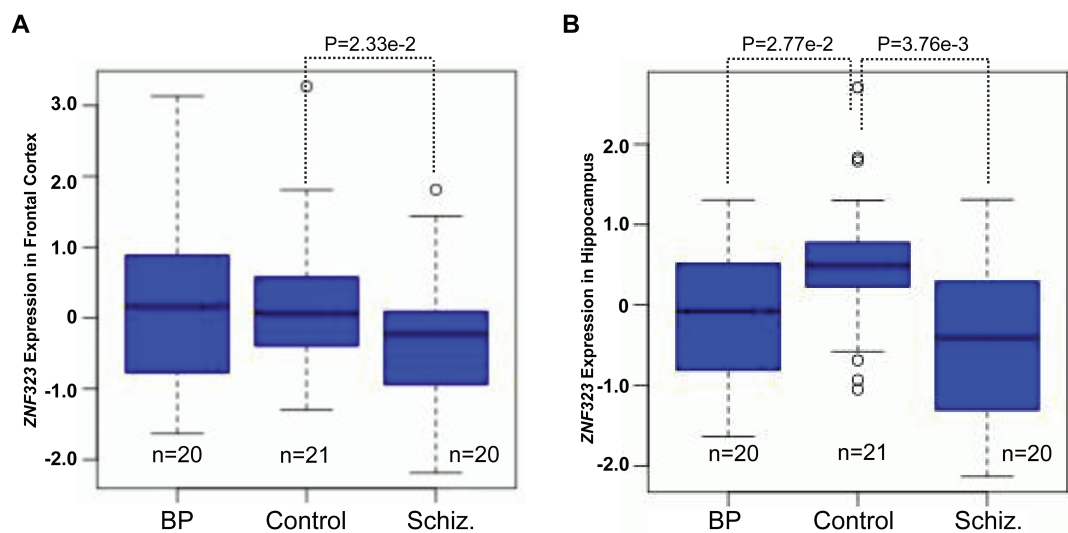
Fig. 2. *ZNF323* is significantly downregulated in the frontal cortex and hippocampus of schizophrenia patients. (A) Compared with healthy controls, *ZNF323* expression was significantly downregulated in frontal cortex of schizophrenia patients. (B) Expression of *ZNF323* was significantly decreased in hippocampus of schizophrenia cases. BP, patients with bipolar disorder, Schiz., patients with schizophrenia.

rs1150711 based on 1000 Genomes data (phase 1)[44] in CEU-revealed moderate LD (r^2: 0.480 and 0.448, respectively; D': 0.871 and 1.000, respectively). We observed evidence for association of rs1150711 in the aforementioned lung function GWAS (FEV_1: $P = 6.62 \times 10^{-5}$) (supplementary table S3). The association for rs6903823 in the PGC1+SWE schizophrenia GWAS was significant ($P = 2.29 \times 10^{-10}$). It is of note that in both SNPs (rs1150711 and rs6903823) the risk allele for schizophrenia (T and A, respectively) shows a protective effect in the lung function GWAS[60] (supplementary table S3).

Encouraged by this finding and the observation that *ZNF323* is expressed in the lung, we tested whether our eSNPs for *ZNF323* in the brain also showed association with *ZNF323* expression in the lung. Based on 119 individuals included in the GTEx dataset,[25] we found significant association for both rs2859365 and rs1150711 with *ZNF323* expression in the lung ($P = .008$ and $P = 9 \times 10^{-5}$) (supplementary figure S6). Furthermore, we observed that for both markers the carriers of the schizophrenia risk alleles (rs2859365:A; rs1150711:T), ie, carriers of the allele that confers protective effects to lung function, did show a reduced expression of *ZNF323* (in brain and lung) (supplementary figure S6).

Evidence for simultaneous association of genetic variation with different traits (pleiotropy) is sometimes based on confounding factors (hidden factors associated with the genetic variation and at the same time with the seemingly pleiotropic traits). Utilizing results from the TAG GWAS on smoking behavior (see "Methods" section of this manuscript for the analysis rationale), we found no evidence for association of our top eSNPs with smoking behavior as a potential confounding factor (for rs2859365 and rs1150711, all corrected P-value > .05; supplementary table S4).

Population Genetics Analyses for rs1150711

Based on our findings for schizophrenia and lung function, we investigated the evolutionary pattern of rs1150711 and *ZNF323* in human populations using population genetic analyses. We first detailed the distribution of rs1150711 allele frequency in worldwide populations using data from the HGDP[47] and the 1000 Genomes project.[44] We found that the frequency of the T allele of rs1150711 (derived and risk allele) is highly variable in world-wide populations ranging from lower frequencies in most African populations to higher frequencies in other populations (figure 3). At the extremes we found that rs1150711 is fixed (either for the ancestral or the derived allele) in some populations (figure 3). Fst analysis also indicates a high degree of population differentiation in African, European, and Asian populations for rs1150711 (Fst = 0.44).

The dramatic differences in allelic frequency strongly suggested that rs1150711 might have experienced recent positive selection in human populations. To test this, we calculated the iHS[45] of rs1150711 in world populations. The iHS analysis (figure 4A and supplementary table S5) revealed that the derived allele of rs1150711 (T) has experienced strong recent positive selection in Europeans (CEU; iHS = −2.034, $P < .05$) and Asians (iHS = −2.206, $P < .05$), while we did not observe evidence for significant positive selection in Africans (iHS = -0.617, $P > .05$). This finding was supported by the EHH[41] analyses across all three studied populations, the haplotypes carrying the T allele (derived allele, is also the risk allele) of rs1150711 decay slower for the derived alleles than for the ancestral alleles (figures 4B, 4E, and 4H).

We further plotted haplotype bifurcation diagrams[41] for both alleles at rs1150711 (figures 4C, 4D, 4F, 4G, 4
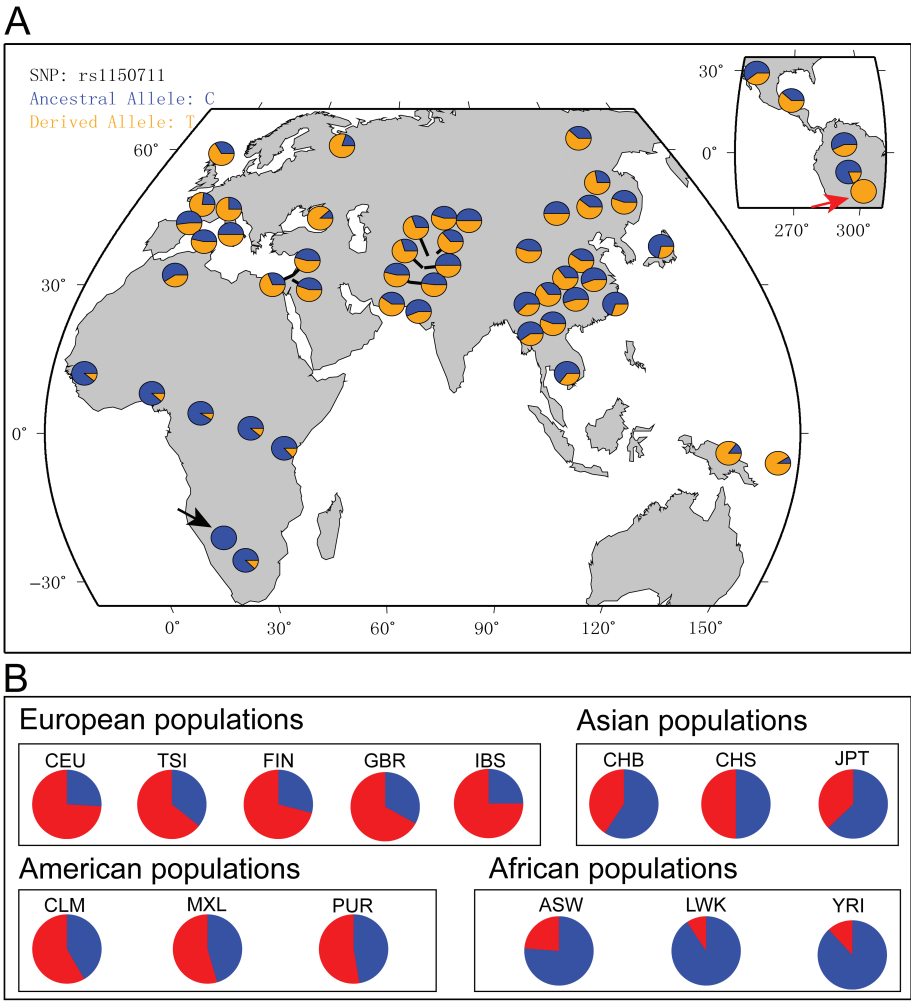
Fig. 3. Global distribution of the rs1150711 risk allele (T allele) in world populations. (A) Global distribution of the risk allele (T allele) of rs1150711 in world populations. The frequency of the risk allele is highly variable in world populations. For example, it is absent in San population from Africa (black arrow), nevertheless, the risk allele (T allele) is completely fixed (frequency = 1.0) in Surui population from Brazil (red arrow). (B) Allele frequency distribution of the risk allele (T allele) of rs1150711 in four continental populations from the 1000 genomes projects. The frequency of the risk allele of rs1150711 is low in African populations. However, the risk allele is prevalent in other populations, including European, Asian, and American populations.

I, and 4J). Consistent with our EHH and iHS results, we found that the bifurcation of haplotypes containing the derived allele (T) was slower in contrast to haplotypes with the ancestral allele (C). Haplotypes carrying the derived allele (T) of rs1150711 have unusual high EHH and high population frequency, indicating that recent positive selection operated on the derived allele and as a result led to a faster increase in its frequency in the human population when compared with a model under neutral evolution. Further analyses using haplotter[45] confirmed our findings (supplementary figure S7). Collectively, these results indicate that the derived allele of rs1150711 (T) has experienced recent positive selection in human populations.

We also explored the recent positive selection on rs2859365. Consistent with the results of rs1150711, iHS analysis revealed that the risk allele of rs2859365

(A allele) might have experienced significant recent positive selection in Europeans (iHS = −2.023, $P <$ .05). Nevertheless, we did not observe significant recent positive selection in Chinese and African populations (iHS = −0.547 and −1.051, respectively, P > .05) (supplementary table S6 and supplementary figure S8). The results of EHH showed similar patterns (supplementary table S6 and supplementary figure S8). Tajima's D, Fu and Li's D and F tests, and Fay and Wu's H test indicated that the evolution of *ZNF323* did not deviate from neutrality significantly (supplementary table S7), which further suggesting that the positive selection on rs1150711 and *ZNF323* is a relatively recent event.

It is conceivable that the positive selection of the derived allele of rs1150711 (T) in the human populations was driven by other SNPs in LD with rs1150711 ("genetic

1303
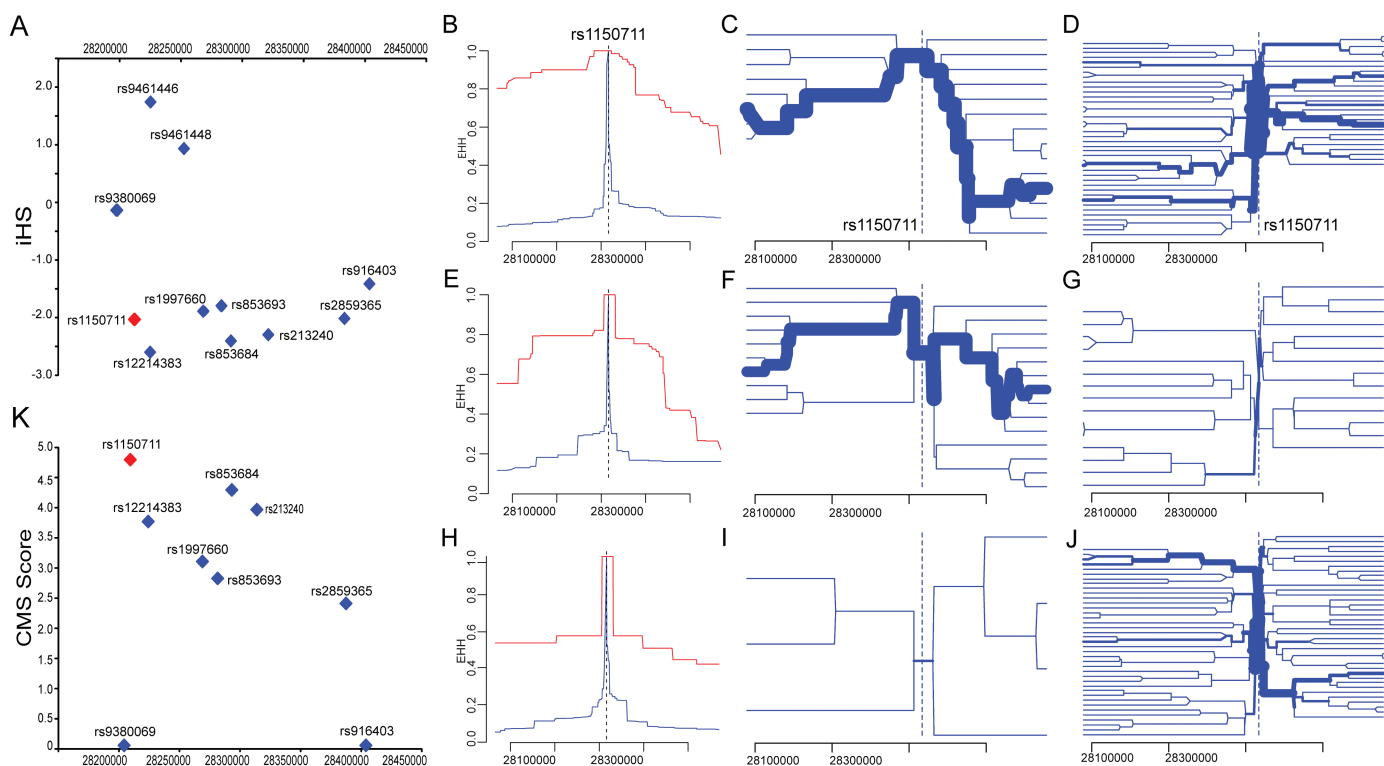
Fig. 4. Recent positive selection of rs1150711 risk allele in human populations. (A) iHS value of the 11 single nucleotide polymorphisms (SNPs) that showed significant association with schizophrenia and *ZNF323* expression in Europeans. (B, C, D) extended haplotype homozygosity (EHH) and haplotype bifurcation plots of rs1150711 in Han Chinese. (E, F, G) EHH and haplotype bifurcation plots of rs1150711 in Europeans. (H, I, J) EHH and haplotype bifurcation plots of rs1150711 in Africans. (B, E, H) Decay of EHH for the risk allele (T allele) of rs1150711 in three representative world populations (B, CHB, Han Chinese; E, CEU, Europeans; H, YRI, Africans) over physical distance. In each population, the decay of haplotype homozygosity for the risk allele (red) occurs much more slowly than for the protective allele (blue), indicating strong positive selection acting on haplotypes containing this risk allele. (C, D, F, G, I, J) Haplotype bifurcation graphs show that the bifurcation of haplotypes is also slower for the risk allele (T allele) of rs1150711 (C, F, I) compared with the protective allele (C allele) (D, G, J). In all of studied populations (CHB, CEU, and YRI), bifurcation of haplotypes for the risk allele (T allele) of rs1150711 is significantly lower than for the protective allele (C allele). The haplotype thicknesses represent the haplotype frequency. Haplotypes with unusually high EHH and a high population frequency (thicknesses) suggest the presence of recent positive selection that drives rapid increase in frequency of new variants or haplotypes in the population. Usual long and thick haplotypes are observed for the risk allele of rs1150711, but not for the protective allele, indicating recent positive selection of the risk allele. (K) Composite of multiple signals (CMS) score of the 11 SNPs that showed significant association with schizophrenia and *ZNF323* expression in Europeans. rs1150711 (marked by red) has the highest CMS score, suggesting it is likely represents the variant that under positive selection. The CMS score of rs9461446 and rs9461448 was not available.

hitchhiking"). To test this hypothesis we utilized the CMS method[46] and found that rs1150711 has the highest score among the 11 brain eSNPs for *ZNF323* that were identified in the BrainCloud data (figure 4K), suggesting that rs1150711 is likely the site that drives the positive selection.

As our focus is *ZNF323* in this study, therefore, we did not follow other top predicted genes. The reasons that we focused on *ZNF323* are as follows: (1) Sherlock integration analysis revealed that *ZNF323* showed the most significant association with schizophrenia (with the lowest LBF). (2) We noticed that among the top four predicted genes, only the supporting SNP (rs2859365) of *ZNF323* is located near the *ZNF323* gene (*cis* eQTL). Generally speaking, *cis* eQTL is more likely to be functional. Nevertheless, other genes are also deserved to be investigated in the future.

Discussion

Though genetic studies have revealed multiple promising risk variants that show robust association with schizophrenia, how these risk variants contribute to schizophrenia susceptibility remains largely unknown. Given that most of the identified risk variants are located in noncoding regions, it is likely that these risk variants alter the expression of schizophrenia-associated genes rather than protein function. Our study systematically integrated brain eQTL data[18] and genetic association findings from a large GWAS of schizophrenia (PGC1 + SWE, with a total of 32 143 subjects).[11] Our findings revealed that *ZNF323* is significantly associated with schizophrenia. We validated the association between eSNPs of *ZNF323* (rs2859365 and rs1150711) at both the gene expression level and at the level of the association with schizophrenia. It should

1304

be noted that our replication results for association of rs2859365 and rs1150711 with schizophrenia are in line with the findings of a recently published study ("PGC2 SCZ") that reported schizophrenia-association of 108 genetic loci.[61] Although our own results unfortunately could only provide moderate further evidence for association, "PGC2 SCZ" put the evidence for association of rs2859365 and rs1150711 beyond reasonable doubt (1.68×10^{-11} and 2.62×10^{-13}). In fact, the *P* values of rs2859365 and rs1150711 were steadily diminished with the increase of sample size (supplementary table S8), strongly suggesting that rs2859365 and rs1150711 are authentic risk variants for schizophrenia. Please note that all but one ("Danish2") of the studies used for our analyses were included in the bigger "PGC2 SCZ" study. We also performed *Sherlock* integrative analysis using the PGC2 SCZ data set and found that *ZNF323* still ranked in the top 5, which further suggests that *ZNF323* may represent an authentic candidate gene.

It should be noted that the *P* value of rs2859365 did not reach a significant level and *P* value of rs1150711 showed marginally significant association ($P = .037$) with schizophrenia in the combined replication samples. Following two reasons may lead to these results: First, compared with other replication samples, we noticed that the *Danish sample 1* has a different risk allele, which increased the *P* value when all of the replication samples were combined. In fact, when the *Danish sample 1* was excluded, rs2859365 and rs1150711 showed significant associations with schizophrenia in the replication samples. Second, considering that the sample size of the combined replication samples is still relatively small ($N = 6353$ cases) and the association trend (the affect allele) in the combined replication samples is same as the discovery sample, we may obtain a more significant *P* value if the sample size is increased in the replication stage.

ZNF323 (also known as *ZSCAN31*) encodes SCAN/(Cys)2(His)2 zinc-finger TF that plays an active role in human embryonic development.[62] Expression analysis revealed that *ZNF323* is extensively expressed in human tissues, including lung, liver, kidney, brain, heart, and pancreas,[62] suggesting that is may be involved in the development of multiple embryonic organs (including brain). Recent studies reveal the important role of zinc-finger TFs in schizophrenia. For example, *ZNF804A*, the first identified gene that showed genome-wide significant association with schizophrenia,[63] also encodes a zinc finger protein. In addition to schizophrenia, significant association between *ZNF804A* and bipolar disorder also has been reported.[64] These lines of evidence suggest that zinc-finger TF may play a pivotal role in psychiatric disorders.

ZNF323 is located on the MHC region. This region on chromosome 6 is known to harbor a large number of disease associated variants for traits such as mental disorders (eg, schizophrenia),[10] traits related to immune processes (eg, rheumatoid arthritis),[65] and many others.[66,67]

Due to the extended LD in this region, pinpointing the genes of interest remained a challenge in most of the studies. Our integrated analysis design potentially could help to identify more disease relevant genes in the MHC region. Although eQTL studies in general suffer from similar limitations that GWA studies do (large amount of correlated measurements), integrating several layers of gene expression with genetic data and other data might help to pinpoint disease associated genes. In our study only gene expression of *ZNF323* was affected by all the 11 eSNPs in the brain (table 2). Moreover, we found evidence for a significantly reduced *ZNF323* expression in brains of schizophrenia patients (in brain regions relevant to schizophrenia). More importantly, we noted that lower *ZNF323* expression-associated alleles at rs2859365 and rs1150711 were enriched in schizophrenia cases (OR = 1.082 and 1.105, respectively). In addition to *ZNF323*, several other genes were also be identified by *Sherlock* (table 1). We noticed that *EIF4A2* was reported to be downregulated in brains (anterior temporal lobe) of schizophrenia patients.[68] In addition, genetic variants in *LSAMP* showed significant association with schizophrenia[69] and several other mental disorders, including major depressive disorder and panic disorder,[70] and male suicide.[71] These lines of evidence strongly suggest the validity of *Sherlock* in identifying disease susceptibility genes.

Not only did we identify a new potential risk gene for schizophrenia but also we could find evidence for recent positive selection for rs1150711 and thus potentially provide a possible explanation to the evolutionary enigma of schizophrenia.[4] Although in general this is of no surprise for genetic variation located in the MHC region,[11] we were also able to link the evidence for recent positive selection with a reasonable candidate mechanism that could have provided compensatory advantage. Gene expression of *ZNF323* in brain and lung is reduced as a consequence of the presence of the derived allele of rs1150711, which at the same time confers risk to schizophrenia and provides a protective effect on lung function. Considering that the protective allele on lung function (rs1150711-T) has experienced positive selection in Europeans, it is possible that lung function (measured as FEV1 and FCV) may show difference among different populations. Interestingly, one study showed that compared with European children, the lung function (FEV$_1$ and FVC) of African and Asian children were significantly lower.[72] These results suggest that lung function might have experienced positive selection in Europeans. It is of note that the protective effect on lung function is probably only one of multiple mechanisms that could have helped the risk allele of rs1150711 to escape extinction from the human gene pool (as a consequence of its involvement in schizophrenia susceptibility). Further work is warranted to investigate the mechanism underlying the expansion of the risk allele of rs1150711.

Supplementary Material

Supplementary material is available at http://schizophreniabulletin.oxfordjournals.org.

Funding

Acknowledgments

References

1. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460:748–752.

2. Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry*. 2003;60:1187–1192.

3. Power RA, Kyaga S, Uher R, et al. Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. *JAMA Psychiatry*. 2013;70:22–30.

4. Crespi B, Summers K, Dorus S. Adaptive evolution of genes underlying schizophrenia. *Proc Biol Sci*. 2007;274:2801–2810.

5. Pearlson GD, Folley BS. Schizophrenia, psychiatric genetics, and Darwinian psychiatry: an evolutionary framework. *Schizophr Bull*. 2008;34:722–733.

6. Huxley J, Mayr E, Osmond H, Hoffer A. Schizophrenia as a genetic morphism. *Nature*. 1964;204:220–221.

7. Andreassen OA, Harbo HF, Wang Y, et al. Genetic pleiotropy between multiple sclerosis and schizophrenia but not bipolar disorder: differential involvement of immune-related gene loci. *Mol Psychiatry.* 2015; 20:207–214.

8. Andreassen OA, Djurovic S, Thompson WK, et al. Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. *Am J Hum Genet*. 2013;92:197–209.

9. Ripke S, Sanders AR, Kendler KS, et al. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet*. 2011;43:969–976.

10. Stefansson H, Ophoff RA, Steinberg S, et al. Common variants conferring risk of schizophrenia. *Nature*. 2009;460:744–747.

11. Ripke S, O'Dushlaine C, Chambert K, et al. Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nat Genet*. 2013;45:1150–1159.

12. Bray NJ. Gene expression in the etiology of schizophrenia. *Schizophr Bull*. 2008;34:412–418.

13. Harrison PJ, Weinberger DR. Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Mol Psychiatry*. 2005;10:40–68; image 45.

14. Bacanu SA, Chen J, Sun J, et al. Functional SNPs are enriched for schizophrenia association signals. *Mol Psychiatry*. 2014;19:276–277.

15. Ayalew M, Le-Niculescu H, Levey DF, et al. Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction. *Mol Psychiatry*. 2012;17:887–905.

16. Schadt EE, Lamb J, Yang X, et al. An integrative genomics approach to infer causal associations between gene expression and disease. *Nat Genet*. 2005;37:710–717.

17. He X, Fuller CK, Song Y, et al. Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS. *Am J Hum Genet*. 2013;92:667–680.

18. Myers AJ, Gibbs JR, Webster JA, et al. A survey of genetic human cortical gene expression. *Nat Genet*. 2007;39:1494–1499.

19. Hindorff LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*. 2009;106:9362–9367.

20. Servin B, Stephens M. Imputation-based analysis of association studies: candidate regions and quantitative traits. *PLoS Genet*. 2007;3:e114.

21. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*. 1995;57:289–300.

22. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–575.

23. de Bakker PI, Ferreira MA, Jia X, et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet*. 2008;17:R122–R128.

24. Colantuoni C, Lipska BK, Ye T, et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature*. 2011;478:519–523.

25. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45:580–585.

26. Kim S, Webster MJ. The stanley neuropathology consortium integrative database: a novel, web-based tool for exploring neuropathological markers in psychiatric disorders and the biological processes associated with abnormalities of those markers. *Neuropsychopharmacology*. 2009;35:473–482.

27. Kang HJ, Kawasawa YI, Cheng F, et al. Spatio-temporal transcriptome of the human brain. *Nature*. 2011;478:483–489.

28. Priebe L, Degenhardt F, Strohmaier J, et al. Copy number variants in German patients with schizophrenia. *PLoS One*. 2013;8:e64035.

29. Muglia P, Tozzi F, Galwey NW, et al. Genome-wide association study of recurrent major depressive disorder in two European case-control cohorts. *Mol Psychiatry*. 2010;15:589–601.

30. Aberg KA, Liu Y, Bukszár J, et al. A comprehensive family-based replication study of schizophrenia genes. *JAMA Psychiatry*. 2013;70:573–581.

31. Dudbridge F. Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Hum Hered*. 2008;66:87–98.

32. Børglum AD, Demontis D, Grove J, et al. Genome-wide study of association and interaction with maternal cytomegalovirus infection suggests new schizophrenia loci. *Mol Psychiatry*. 2014;19:325–333.

33. Mors O, Perto GP, Mortensen PB. The Danish psychiatric central research Register. *Scand J Public Health*. 2011;39:54–57.

34. Nørgaard-Pedersen B, Hougaard DM. Storage policies and use of the Danish Newborn Screening Biobank. *J Inherit Metab Dis*. 2007;30:530–536.

35. Hollegaard MV, Grauholm J, Børglum A, et al. Genome-wide scans using archived neonatal dried blood spot samples. *BMC Genomics*. 2009;10:297.

36. Hindorff L, MacArthur J, Morales J, et al. A Catalog of Published Genome-Wide Association Studies. Available at: www.genome.gov/gwastudies.

37. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 2014;42:D1001–D1006.

38. Dixon JR, Selvaraj S, Yue F, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012;485:376–380.

39. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res*. 2012;22:1790–1797.

40. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489:57–74.

41. Sabeti PC, Reich DE, Higgins JM, et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature*. 2002;419:832–837.

42. Gautier M, Vitalis R. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics*. 2012;28:1176–1177.

43. International HapMap Consortium. The International HapMap Project. *Nature*. 2003;426:789–796.

44. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467:1061–1073.

45. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol*. 2006;4:e72.

46. Grossman SR, Shlyakhter I, Karlsson EK, et al. A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science*. 2010;327:883–886.

47. Pickrell JK, Coop G, Novembre J, et al. Signals of recent positive selection in a worldwide sample of human populations. *Genome Res*. 2009;19:826–837.

48. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989;123:585–595.

49. Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics*. 1993;133:693–709.

50. Fay JC, Wu CI. Hitchhiking under positive Darwinian selection. *Genetics*. 2000;155:1405–1413.

51. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009;25:1451–1452.

52. Cheng F, Chen W, Richards E, Deng L, Zeng C. SNP@ Evolution: a hierarchical database of positive selection on the human genome. *BMC Evol Biol*. 2009;9:221.

53. Bryois J, Buil A, Evans DM, et al. Cis and trans effects of human genomic variants on gene expression. *PLoS Genet*. 2014;10:e1004461.

54. Wittkopp PJ. Genomic sources of regulatory variation in cis and in trans. *Cell Mol Life Sci*. 2005;62:1779–1783.

55. Li H, Deng H. Systems genetics, bioinformatics and eQTL mapping. *Genetica*. 2010;138:915–924.

56. Gibbs JR, van der Brug MP, Hernandez DG, et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet*. 2010;6:e1000952.

57. Perlstein WM, Carter CS, Noll DC, Cohen JD. Relation of prefrontal cortex dysfunction to working memory and symptoms in schizophrenia. *Am J Psychiatry*. 2001;158:1105–1113.

58. Knable MB, Weinberger DR. Dopamine, the prefrontal cortex and schizophrenia. *J Psychopharmacol*. 1997;11:123–131.

59. Pope BD, Ryba T, Dileep V, et al. Topologically associating domains are stable units of replication-timing regulation. *Nature*. 2014;515:402–405.

60. Soler Artigas M, Loth DW, Wain LV, et al. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat Genet*. 2011;43:1082–1090.

61. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511:421–427.

62. Pi H, Li Y, Zhu C, et al. A novel human SCAN/(Cys)2(His)2 zinc-finger transcription factor ZNF323 in early human embryonic development. *Biochem Biophys Res Commun*. 2002;296:206–213.

63. O'Donovan MC, Craddock N, Norton N, et al. Identification of loci associated with schizophrenia by genome-wide association and follow-up. *Nat Genet*. 2008;40:1053–1055.

64. Riley B, Thiselton D, Maher BS, et al. Replication of association between schizophrenia and ZNF804A in the Irish Case-Control Study of Schizophrenia sample. *Mol Psychiatry*. 2010;15:29–37.

65. Raychaudhuri S, Sandor C, Stahl EA, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet*. 2012;44:291–296.

66. Su Z, Gay LJ, Strange A, et al. Common variants at the MHC locus and at chromosome 16q24.1 predispose to Barrett's esophagus. *Nat Genet*. 2012;44:1131–1136.

67. Yu XQ, Li M, Zhang H, et al. A genome-wide association study in Han Chinese identifies multiple susceptibility loci for IgA nephropathy. *Nat Genet*. 2012;44:178–182.

68. Martins-de-Souza D, Gattaz WF, Schmitt A, et al. Alterations in oligodendrocyte proteins, calcium homeostasis and new potential markers in schizophrenia anterior temporal lobe are revealed by shotgun proteome analysis. *J Neural Transm*. 2009;116:275–289.

69. Koido K, Janno S, Traks T, et al. Associations between polymorphisms of LSAMP gene and schizophrenia. *Psychiatry Res*. 2014;215:797–798.

70. Koido K, Traks T, Balõtšev R, et al. Associations between LSAMP gene polymorphisms and major depressive disorder and panic disorder. *Transl Psychiatry*. 2012;2:e152.

71. Must A, Tasa G, Lang A, et al. Association of limbic system-associated membrane protein (LSAMP) to male completed suicide. *BMC Med Genet*. 2008;9:34.

72. Johnston ID, Bland JM, Anderson HR. Ethnic variation in respiratory morbidity and lung function in childhood. *Thorax*. 1987;42:542–548.

SUPPLEMENTARY INFORMATION

Systematic integration of brain eQTL and GWAS identifies *ZNF323* as a novel schizophrenia risk gene and suggests recent positive selection based on compensatory advantage on pulmonary function

Xiong-jian Luo, Manuel Mattheisen et al.

Correspondence to Xiong-jian Luo, Ph.D., Key Laboratory of Animal Models and Human Disease Mechanisms, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650223, China; Tel: +86-871-65180085, Fax: +86-871-65180085, E-mail:

luoxiongjian@mail.kiz.ac.cn.

1. **BYLINE AUTHORS**

a) **MooDS SCZ Consortium**

Jana Strohmaier[1], Franziska A. Degenhardt[2], Sandra Meier[1], Josef Frank[1], Stephanie H. Witt[1], Andreas J. Forstner[2], Buket F. Basmanav[2], Andrea Hofman[2], Peter Propping[2], Manuel Mattheisen[2,3,4], Christoph Lange[5], Markus Leber[6], Thomas G. Schulze[7], C. Christoph Schultz[8], Ralf G.M. Schlösser[8], Igor Nenadic[8], Heinrich Sauer[8], Rainald Mössner[9], Wolfgang Maier[9], Ina Giegling[10], Dan Rujescu[10], Bertram Müller-Myhsok[11], Sven Cichon[2,12,13], Markus M. Nöthen[2], Marcella Rietschel[1]

[1]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, University of Mannheim, Mannheim, Germany

[2]Department of Genomics, Life & Brain Center, and Institute of Human Genetics, University of Bonn, Bonn, Germany

[3]Department of Biomedicine and Centre for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark

[4]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus and Copenhagen, Denmark

[5]Department of Genomic Mathematics, University of Bonn, Bonn, Germany

[6]Institute for Medical Biometry, Informatics, and Epidemiology, University of Bonn, Bonn, Germany

[7]Department of Psychiatry and Psychotherapy, University Medical Center Georg-August-Universität, Goettingen, Germany

[8]Department of Psychiatry and Psychotherapy, Friedrich-Schiller-University Jena, Jena, Germany

[9]Department of Psychiatry, University of Bonn, Bonn, Germany

[10]Department of Psychiatry, University of Halle-Wittenberg, Halle/Saale, Germany

[11]Department of Statistical Genetics, Max-Planck-Institute of Psychiatry, Munich, Germany

[12]Division of Medical Genetics, Department of Biomedicine, University Basel, Basel, Switzerland

[13]Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, Juelich, Germany

b) iPSYCH-GEMS SCZ working group

Mors O[a,b], Mortensen PB[b,c], Hougaard DM[d], Pedersen CB[b,c], Demontis D[b,e], Grove J[b,e,f], Mattheisen M[b,e,g], Børglum AD[a,b,e]

[a]Centre for Psychiatric Research, Aarhus University Hospital, Risskov, Denmark

[b]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus and Copenhagen, Denmark

[c]National Centre for Register-based Research, Aarhus University, Aarhus, Denmark

[d]Section of Neonatal Screening and Hormones, Statens Serum Institute, Copenhagen, Denmark

[e]Department of Biomedicine and Centre for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark

[f]Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark

[g]Department of Genomics, Life & Brain Center, and Institute of Human Genetics, University of Bonn, Bonn, Germany

2. SUPPLEMENTARY TABLES

Supplementary Table S1. Supporting SNPs for Some of the Top Predicted Genes of Schizophrenia

Gene	SNP	SNP Location	Proximity	eQTL p Value	GWAS p Value	LBF
ZNF323	rs2859365	chr6:28499444	*cis*	5.90e-08	2.07e-07	7.33
	rs4498913	chr10:97266482	*trans*	3.95e-06	1.31e-02	0.55
NAPSA	rs10226475	chr7:2192688	*trans*	6.74e-06	2.02e-10	4.65
	rs6822958	chr4:133767613	*trans*	3.14e-06	6.92e-05	2.98
LOC378135	rs134169	chr22:26942399	*trans*	6.17e-06	2.50e-05	4.41
	rs2888019	chr7:82341345	*trans*	1.05e-07	8.31e-04	2.69
EPB41L2	rs7752195	chr6:25527073	*trans*	1.17e-06	1.82e-05	5.24
	rs17111984	chr14:94951179	*trans*	2.43e-06	3.06e-03	1.75

A SNP is called in cis of a gene if it is located within 1 Mb of the transcription start site of this gene. Each SNP in the table actually represents a block of adjacent eSNPs (chosen to be the one with the highest LBF). The number in parentheses in the Gene column shows the total number of putative eSNP blocks (defined as $p < 10^{-5}$ in the eQTL data) of that gene. The LBF column shows the LBF of individual SNP. Only SNPs with LBF greater than 0.5 are shown.

Supplementary Table S2. Association between SNPs that are highly linked with rs1150711 and schizophrenia

SNP	Distance from rs1150711	R^2	GWAS P value
rs1150711	0	1.00	2.76×10^{-9}
rs1997660	61128	0.879	6.23×10^{-8}
rs9468333	62663	0.879	6.77×10^{-8}
rs853693	74113	0.879	6.05×10^{-8}
rs853690	76947	0.879	5.63×10^{-8}
rs707907	82705	0.879	6.51×10^{-8}
rs853683	86504	0.879	5.94×10^{-8}
rs4713158	69236	0.840	6.38×10^{-8}

Supplementary Table S3. The schizophrenia-associated risk alleles at rs1150711 and rs6903823 have protective effect on pulmonary function

SNP ID	Alleles	Trait	Beta-coefficient	P-value[a]	P-value[b]	Risk allele (PF)[c]	Risk allele (SZ)[d]
rs1150711	C/T	FEV_1	0.031	6.62×10^{-5}	2.82×10^{-10}	C	T
rs6903823	A/G	FEV_1	0.037	2.10×10^{-10}	2.29×10^{-10}	G	A

[a]P-value indicates the association significance between the SNP and pulmonary function. [b]P-value indicates the association significance between the SNP and schizophrenia. [c]Risk allele for pulmonary function (PF). [d]Risk allele for schizophrenia (SZ).

Supplementary Table S4. Association significance between eSNPs of *ZNF323* and smoking

SNP ID	Chr:Position	Cigarettes smoked per day[a]	Ever smoked[b]	Former smoker[c]	Age-onset[d]
rs853684	6:28294550	0.671	0.552	0.0551	0.259
rs2859365	6:28391465	0.889	0.470	0.165	0.0717
rs1150711	6:28208535	0.561	0.331	0.0547	0.165
rs213240	6:28315875	0.864	0.490	0.0554	0.205
rs1997660	6:28269663	0.469	0.250	0.00852	0.144
rs853693	6:28282648	0.412	0.292	0.0108	0.193
rs12214383	6:28223731	0.445	0.676	0.0702	0.367
rs916403	6:28408258	0.667	0.825	0.267	0.170
rs9380069	6:28203300	0.794	0.0466	0.00798	0.136
rs9461446	6:28227633	0.248	0.0791	0.0293	0.921
rs9461448	6:28263721	0.334	0.00388	0.145	0.957

[a]P-value indicates the association significance between the SNP and smoking quantity (number of cigarettes smoked per day. [b]P-value indicates the association significance between the SNP and smoking initiation (ever versus never been a regular smoker. [c]P-value indicates the association significance between the SNP and smoking cessation (former versus current smokers). [d]P-value indicates the association significance between the SNP and age of smoking initiation.

Supplementary Table S5. Recent positive selection of rs1150711 and *ZNF*323 in human populations

Gene or SNP	iHS[a]			*P* value[b]			Fst
Population	Europeans	Asians	Africans	Europeans	Asians	Africans	
rs1150711	-2.034*	-2.206*	-0.617	N/A	N/A	N/A	0.440
*ZNF*323	N/A	N/A	N/A	0.148	**0.031**	0.171	N/A

[a]iHS, integrated haplotype score, is an indicator for recent positive selection. Large negative iHS value indicates unusually long haplotypes carrying the derived allele, while large positive value indicates long haplotypes carrying the ancestral allele. [b]Gene-based *P* value for recent positive selection (iHS test). *$P < 0.05$. N/A, not applicable.

Supplementary Table S6. Recent positive selection on rs2859365

Gene or SNP	iHS[a]			Fst
Population	Europeans	Asians	Africans	
rs2859365	-2.023 (P<0.05)	-0.547 (P>0.05)	-1.051 (P>0.05)	0.222

[a]iHS, integrated haplotype score, is an indicator for recent positive selection. Large negative iHS value indicates unusually long haplotypes carrying the derived allele, while large positive value indicates long haplotypes carrying the ancestral allele.

Supplementary Table S7. Neutrality tests for the *ZNF323* gene in human populations.

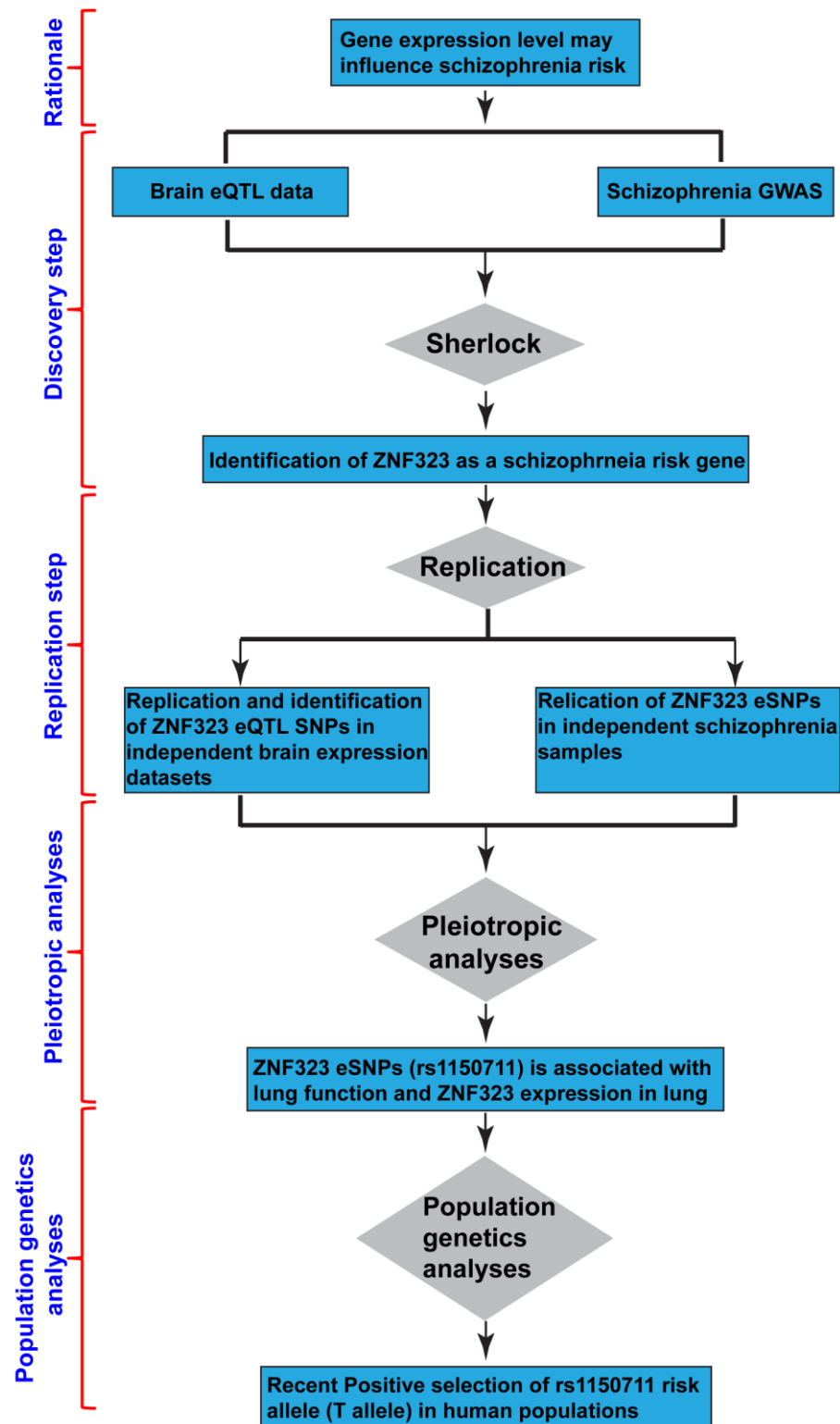| *ZNF323* (11,637bp)	Nucleotide diversity	Tajima's D	Fu and Li's D	Fu and Li's F	Fu and Li's D*	Fu and Li's F*	Fay and Wu's H
CHB (206)[#]	1.34×10^{-5}	-0.258	1.66*	1.74	-2.14	-1.47	-2.64
CEU (198)	1.05×10^{-5}	-0.878	1.02	0.576	-2.13	-1.83	-6.97
YRI (216)	1.25×10^{-5}	-0.436	N/A	N/A	-0.792	-0.738	N/A

*$P < 0.05$. [#]The sample size refers to the number of chromosomes used in the analysis.

Supplementary Table S8. The *P* values of rs2859365 and rs1150711 are steadily diminished with the increase of schizophrenia sample size

P value	Sample (Cases/Controls)[a]			
	PGC1 (9394/12462)	PGC1+SWE (13833/18310)	PGC1+SWE+follow-up (20186/23928)	PGC2 (35476/46839)
$P_{rs2859365}$	2.45×10^{-6}	2.07×10^{-7}	1.40×10^{-7}	1.68×10^{-11}
$P_{rs1150711}$	3.84×10^{-7}	2.76×10^{-9}	6.85×10^{-10}	2.62×10^{-13}

[a]Data were from the Schizophrenia Working Group of the Psychiatric Genomics Consortium (PGC).

3. SUPPLEMENTARY Figures

Rationale

Gene expression level may influence schizophrenia risk

Discovery step

Brain eQTL data

Schizophrenia GWAS

Sherlock

Identification of ZNF323 as a schizophrneia risk gene

Replication step

Replication

Replication and identification of ZNF323 eQTL SNPs in independent brain expression datasets

Relication of ZNF323 eSNPs in independent schizophrenia samples

Pleiotropic analyses

Pleiotropic analyses

ZNF323 eSNPs (rs1150711) is associated with lung function and ZNF323 expression in lung

Population genetics analyses

Population genetics analyses

Recent Positive selection of rs1150711 risk allele (T allele) in human populations

Supplementary Figure S1. Flowchart of the analyses. Based on the rationale that gene expression level may influence disease susceptibility, we systematically integrated brain eQTL data and SNP findings from GWAS using *Sherlock*. We replicated the results in independent brain expression and schizophrenia datasets. Finally, pleiotropic and population genetics were performed.
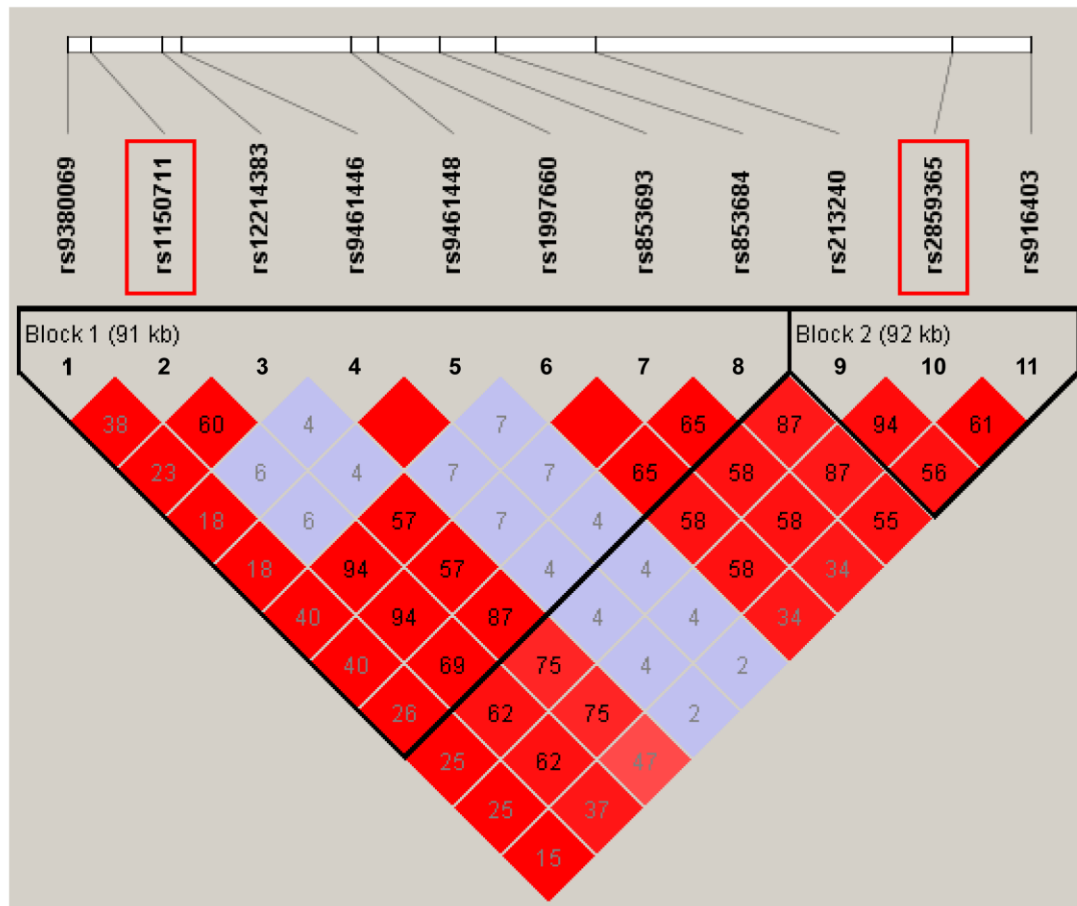
Supplementary Figure S2. Linkage disequilibrium (LD) pattern of the 9 SNPs that showed significant association with *ZNF323* expression. Haplotype blocks were defined by four gamete rule implemented in Haploview program (1). Values in the square represent LD values (r^2). Rs1150711 and rs2859365 are marked by red box.

Supplementary Figure S3. Linkage disequilibrium (LD) between rs1150711 and its surrounding SNPs. Genotype data of Europeans (CEU) from the 1000 genomes project was used to construct the LD map. Only 7 SNPs showed high LD with rs1150711 (r^2>0.80).

Supplementary Figure S4. Chromatin interaction (Topologically associated domains, TAD) between the genomic region containing rs1150711 and its neighboring regions.

Two-dimensional heat map shows the chromatin interaction frequency (The heat map is viewed in a manner similar to a linkage disequilibrium plot, where the interaction between two loci is viewed as the point where the diagonals originating from each locus intersect on the heat map). Below the heat map is data from the UCSC genome browser containing information on domain, the directionality index, and ChIP-Seq data. Black arrow shows the genomic position of rs1150711 and red arrow points to *ZNF323*. Black box showed the chromatin interaction between the genomic region containing rs1150711 and *ZNF323*.

Protein Binding					Filter:
Method	Location	Bound Protein	? Cell Type	Additional Info	Reference
ChIP-seq	chr6:28208313..28208593	MAFF	K562		ENCODE
ChIP-seq	chr6:28208335..28208591	MAFK	HepG2		ENCODE
ChIP-seq	chr6:28208335..28208605	MAFK	IMR90		ENCODE
ChIP-seq	chr6:28208342..28208622	MAFK	K562		ENCODE
ChIP-seq	chr6:28208033..28209031	RFX3	K562		20378718
ChIP-seq	chr6:28208314..28208590	MAFF	HepG2		ENCODE
ChIP-seq	chr6:28208339..28208579	MAFK	HepG2		ENCODE

Motifs					Filter:
Method	Location	Motif	? Cell Type	PWM	Reference
PWM	chr6:28208528..28208539	Evi-1			16381825
PWM	chr6:28208528..28208539	Evi-1			16381825
PWM	chr6:28208528..28208539	Evi-1			16381825

Supplementary Figure S5. Bioinformatic analyses of rs1150711 using Regulome DB.
Transcription factors (MAFF and RFX3) can bind to the genomic sequence containing rs1150711. In addition, rs1150711 is located in a well-defined motif (Evi-1).
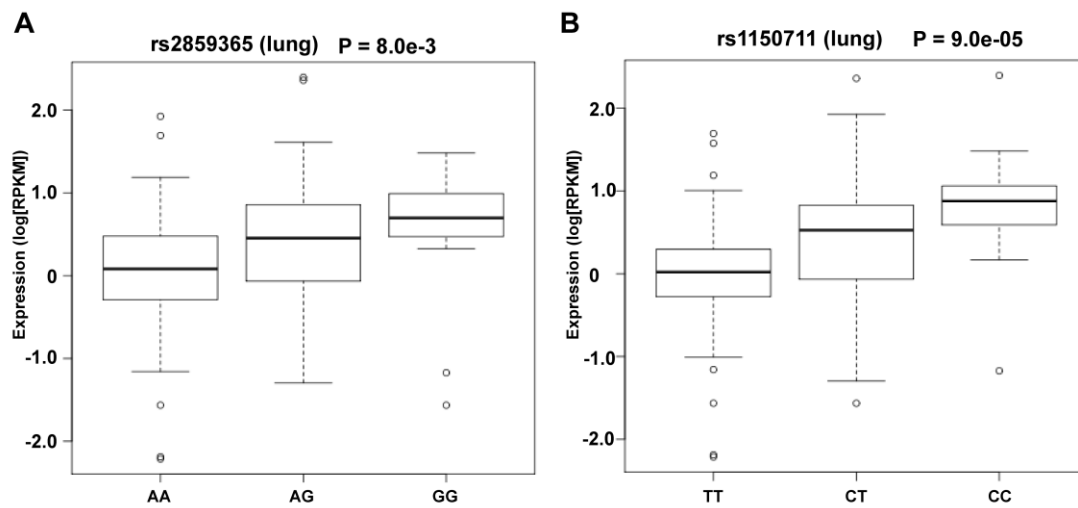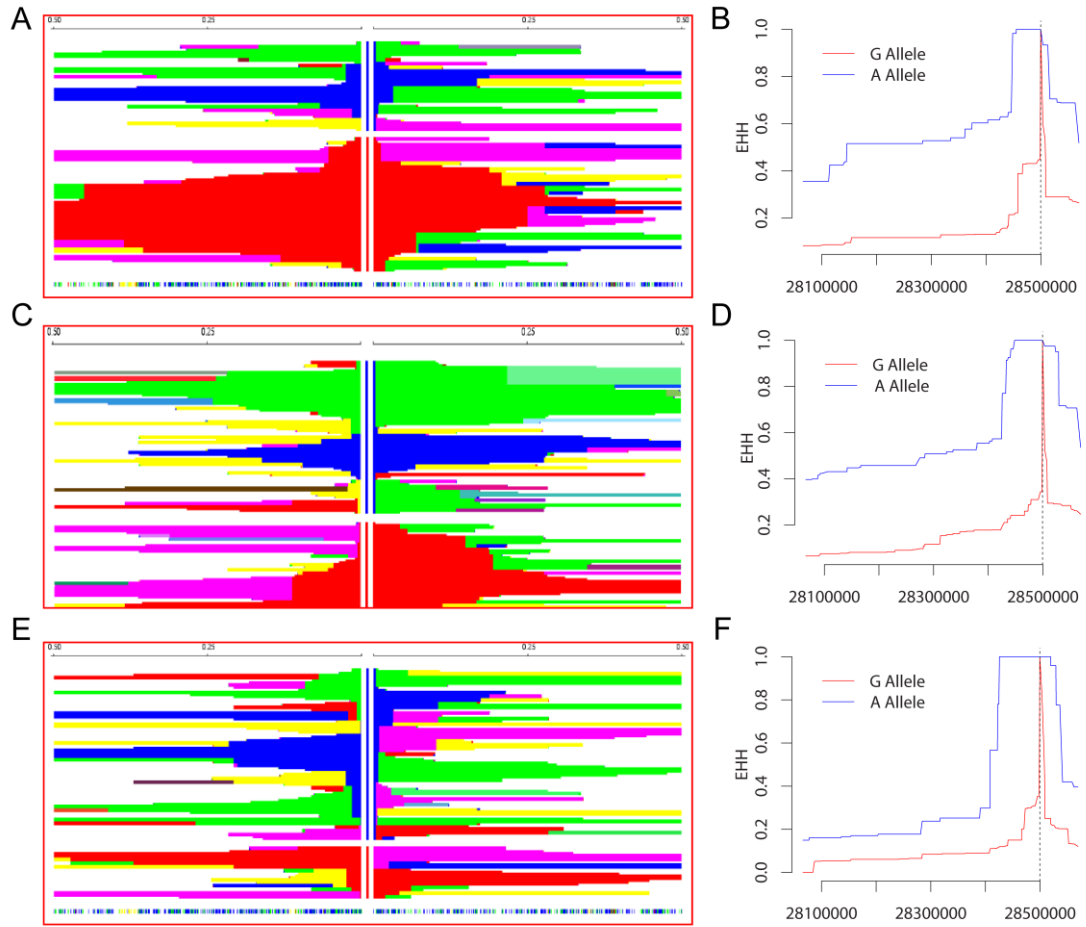
Supplementary Figure S6. Rs2859365 and rs1150711 are significantly associated with *ZNF323* **expression in lung.**
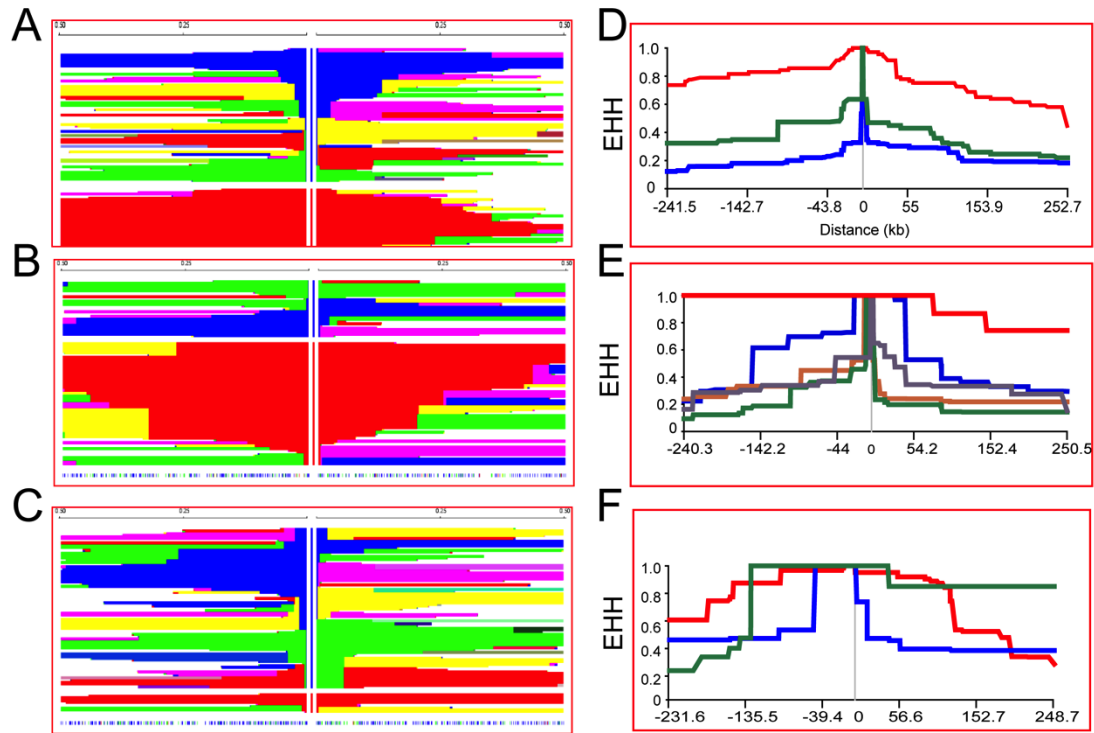
Supplementary Figure S7. Recent positive selection of rs1150711 risk allele (T allele) in Europeans and Chinese. **(A-C)** Decay of haplotypes carrying the rs1150711 risk allele (T allele) and protective allele (C allele) in Chinese (CHB), Europeans (CEU) and Africans (YRI). Horizontal lines are haplotypes, rs1150711 is located in the center and SNP positions are marked below the haplotype plot. These plots are divided into two parts: the upper portion shows haplotypes with the non-risk C allele at rs1150711 (blue), and the lower portion shows haplotypes with the risk T allele at rs1150711 (red). For a given SNP, adjacent haplotypes with the same color carry identical genotypes everywhere between that SNP and the central (selected) site. The left- and right-hand sides are plotted separately. Haplotypes are no longer plotted beyond the points at which they become unique. In CHB and CEU, the haplotype homozygosity surrounding the risk allele (T allele) (red) is longer than the C allele, indicating positive selection driven the rapidly increasing of the frequency of haplotypes carrying the rs1150711 risk allele. **(D-F)** Decay of extended haplotype homozygosity for the rs1150711 risk allele in CHB, CEU and YRI. In each population, the decay of haplotype homozygosity for the protective C allele (blue) occurs much more quickly than for the risk T allele (red), suggesting strong positive selection acting on haplotypes carrying the risk allele.

Supplementary Figure S8. Recent positive selection of rs2859365 risk allele (T allele) in Europeans. (A, B) Decay of haplotypes carrying the rs2859365 risk allele (A allele) and protective allele (G allele) in Europeans (CEU). Horizontal lines are haplotypes, rs2859365 is located in the center and SNP positions are marked below the haplotype plot. These plots are divided into two parts: the upper portion shows haplotypes with the non-risk G allele at rs2859365 (blue), and the lower portion shows haplotypes with the risk G allele at rs2859365 (red). For a given SNP, adjacent haplotypes with the same color carry identical genotypes everywhere between that SNP and the central (selected) site. The left- and right-hand sides are plotted separately. Haplotypes are no longer plotted beyond the points at which they become unique. **(C, D)** Decay of haplotypes carrying the rs2859365 risk allele (A allele) and protective allele (G allele) in Han Chinese (CHB). **(E, F)** Decay of haplotypes carrying the rs2859365 risk allele (A allele) and protective allele (G allele) in Africans (YRI).

4. <u>SUPPLEMENTARY MATERIALS AND METHODS</u>

Pleiotropic effects analyses of the identified risk loci

To explore the pleiotropic effects of ZNF323 and its eSNPs, we used the data from the National Human Genome Research Institute's (NHGRI) GWAS Catalog. In the first step, we searched for association of other human traits with *ZNF323*, the top schizophrenia risk gene identified in our study. In the second step, we investigated whether brain eSNPs for *ZNF323* were also associated with other human traits. For this step we utilized the NHGRI GWAS Catalog entries and the data provided by the group that originally reported the association for this trait (2). In the third step, we investigated whether the brain eSNPs for *ZNF323* that showed association with schizophrenia and lung function (the only other trait found in the NHGRI GWAS Catalog to be associated with our eSNPs) also acted as eSNPs for *ZNF323* in lung tissue. Analyses in this step utilized data from the GTEx (3). Finally, we investigated whether the observed associations of our eSNPS with schizophrenia and lung function could have been the result of confounding. Reasonable candidates for such confounding are smoking behavior traits. First, smoking is significantly more prevalent among schizophrenia patients compared with controls. Second, lung function studies were typically carried out in patients with chronic lung diseases, such as obstructive pulmonary disease (COPD), for which smoking behavior acts as strong environmental risk factor. We therefore utilized a recent GWAS on smoking behavior (4) conducted by the Tabacco and Genetics Consortium (TAG), which studied 74,053 individuals and for which results are publicly available at the PGC's website (https://pgc.unc.edu), to test whether the eSNPs for *ZNF323* are also associated with smoking behavior.

5. **REFERENCES**

1. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21:263-265.

2. Soler Artigas M, Loth DW, Wain LV, Gharib SA, Obeidat M, Tang W, Zhai G, Zhao JH, Smith AV, Huffman JE, Albrecht E, Jackson CM, Evans DM, Cadby G, Fornage M, Manichaikul A, Lopez LM, Johnson T, Aldrich MC, Aspelund T, Barroso I, Campbell H, Cassano PA, Couper DJ, Eiriksdottir G, Franceschini N, Garcia M, Gieger C, Gislason GK, Grkovic I, Hammond CJ, Hancock DB, Harris TB, Ramasamy A, Heckbert SR, Heliovaara M, Homuth G, Hysi PG, James AL, Jankovic S, Joubert BR, Karrasch S, Klopp N, Koch B, Kritchevsky SB, Launer LJ, Liu Y, Loehr LR, Lohman K, Loos RJ, Lumley T, Al Balushi KA, Ang WQ, Barr RG, Beilby J, Blakey JD, Boban M, Boraska V, Brisman J, Britton JR, Brusselle GG, Cooper C, Curjuric I, Dahgam S, Deary IJ, Ebrahim S, Eijgelsheim M, Francks C, Gaysina D, Granell R, Gu X, Hankinson JL, Hardy R, Harris SE, Henderson J, Henry A, Hingorani AD, Hofman A, Holt PG, Hui J, Hunter ML, Imboden M, Jameson KA, Kerr SM, Kolcic I, Kronenberg F, Liu JZ, Marchini J, McKeever T, Morris AD, Olin AC, Porteous DJ, Postma DS, Rich SS, Ring SM, Rivadeneira F, Rochat T, Sayer AA, Sayers I, Sly PD, Smith GD, Sood A, Starr JM, Uitterlinden AG, Vonk JM, Wannamethee SG, Whincup PH, Wijmenga C, Williams OD, Wong A, Mangino M, Marciante KD, McArdle WL, Meibohm B, Morrison AC, North KE, Omenaas E, Palmer LJ, Pietilainen KH, Pin I, Pola Sbreve Ek O, Pouta A, Psaty BM, Hartikainen AL, Rantanen T, Ripatti S, Rotter JI, Rudan I, Rudnicka AR, Schulz H, Shin SY, Spector TD, Surakka I, Vitart V, Volzke H, Wareham NJ, Warrington NM, Wichmann HE, Wild SH, Wilk JB, Wjst M, Wright AF, Zgaga L, Zemunik T, Pennell CE, Nyberg F, Kuh D, Holloway JW, Boezen HM, Lawlor DA, Morris RW, Probst-Hensch N, Kaprio J, Wilson JF, Hayward C, Kahonen M, Heinrich J, Musk AW, Jarvis DL, Glaser S, Jarvelin MR, Ch Stricker BH, Elliott P, O'Connor GT, Strachan DP, London SJ, Hall IP, Gudnason V, Tobin MD. Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. Nat Genet. 2011;43:1082-1090.

3. Consortium GT. The Genotype-Tissue Expression (GTEx) project. Nature genetics. 2013;45:580-585.

4. Tobacco, Genetics C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nature genetics. 2010;42:441-447.