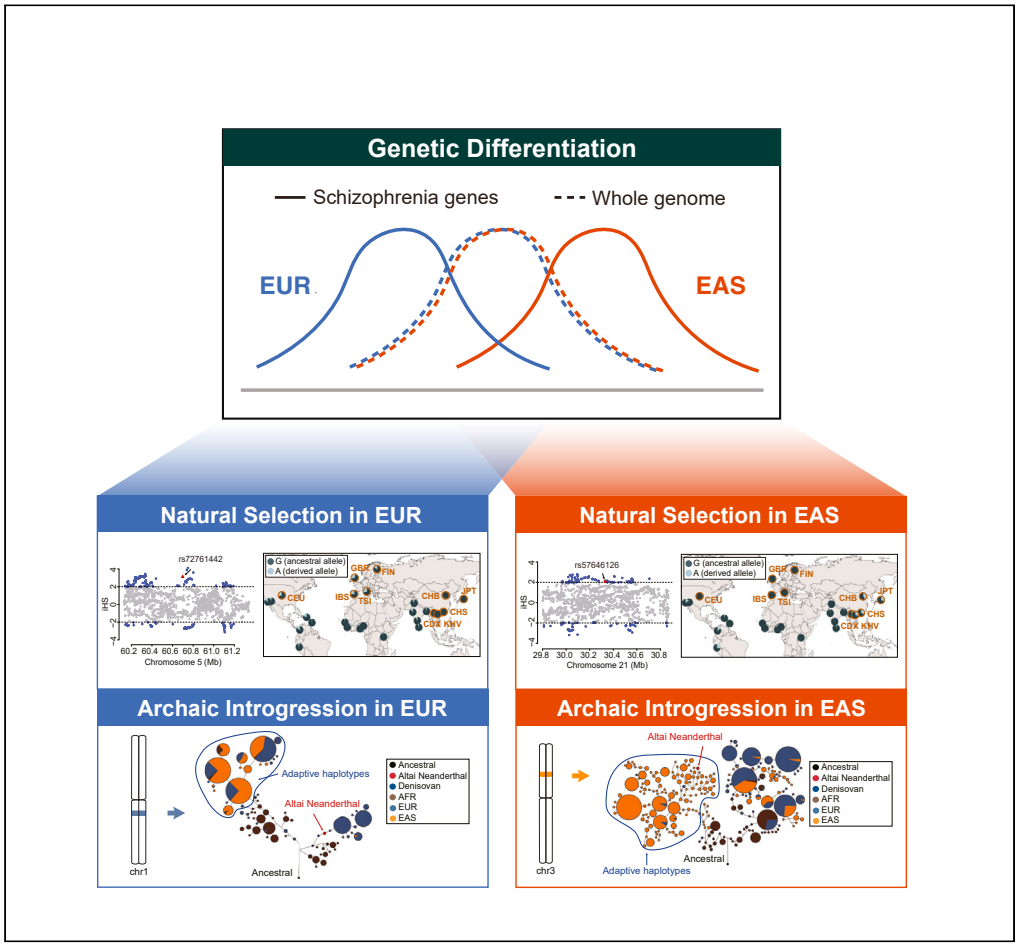
iScience

CellPress
OPEN ACCESS

Article

Asian-European differentiation of schizophrenia-associated genes driven by admixture and natural selection



Sihan Chen, Die Tang, Lian Deng, Shuhua Xu

denglian@fudan.edu.cn (L.D.)
xushua@fudan.edu.cn (S.X.)

Highlights

A higher west-east genetic differentiation in SCZ-genes than the genome average

Differentiation in SCZ-genes can be attributed to independent adaptative evolution

The SCZ-genes underlying stronger selection in East Asians than in Europeans

Opposite SCZ risks of Neanderthal-derived alleles between East Asians and Europeans

Chen et al., iScience 27, 109560
May 17, 2024
https://doi.org/10.1016/j.isci.2024.109560

iScience

CellPress
OPEN ACCESS

Article

Asian-European differentiation of schizophrenia-associated genes driven by admixture and natural selection

Sihan Chen,[1,3] Die Tang,[1,3] Lian Deng,[1,2,*] and Shuhua Xu[1,2,4,*]

SUMMARY

The European-centered genome-wide association studies of schizophrenia (SCZ) may not be well applied to non-European populations. We analyzed 1,592 reported SCZ-associated genes using the public genome data and found an overall higher Asian-European differentiation on the SCZ-associated variants than at the genome-wide level. Notable examples included 15 missense variants, a regulatory variant *SLC5A10*-rs1624825, and a damaging variant *TSPAN18*-rs1001292. Independent local adaptations in recent 25,000 years, after the Asian-European divergence, could have contributed to such genetic differentiation, as were identified at a missense mutation *LTN1*-rs57646126-A in Asians, and a non-risk allele *ZSWIM6*-rs72761442-G in Europeans. Altai-Neanderthal-derived alleles may have opposite effects on SCZ susceptibility between ancestries. Furthermore, adaptive introgression was detected on the non-risk haplotype at 1q21.2 in Europeans, while in Asians it was observed on the SCZ risk haplotype at 3p21.31 which is also potentially ultra-violet protective. This study emphasizes the importance of including more representative Asian samples in future SCZ studies.

INTRODUCTION

Schizophrenia (SCZ) is a severe chronic psychiatric disorder affecting 0.32% of the world's population and causes considerable individual and social burdens. The heritability of SCZ has been estimated to be as high as 80%.[1] The clinical and etiologic heterogeneity of SCZ across geographic regions have been widely recognized,[2–4] whereas the genetic architecture of this disease across populations is not currently fully understood despite great efforts focused on genome-wide association studies (GWASs) that have already substantially increased our knowledge over the past decades. According to the GWAS catalog (https://www.ebi.ac.uk/gwas/efotraits/MONDO_0005090) accessed in Feb 2023, 227 GWASs (including analyses of the "background traits data" and the "child trait data"), have been carried out for SCZ, 154 of which reported findings in populations of European ancestry. In the current collection of SCZ-associated loci from the GWAS catalog, 3,538 (88.7%) variants were reported in studies of Eurasian populations, including 1,622 (40.7%) variants mapped to 539 (33.9%) genes identified in European (EUR) populations but not in East Asian (EAS) populations, 141 (3.5%) variants mapped to 58 (3.6%) genes identified in EAS populations but not in any EUR populations, and 1,775 (44.5%) variants in 839 (52.7%) genes reported in multiple populations including EUR and EAS (Figure S1). The Psychiatric Genomics Consortium (PGC) conducted the largest GWAS on SCZ to date, incorporating nearly 70,000 cases and over 200,000 controls of EUR ancestry.[5–7] Although EAS populations have the greatest incidence and absolute increase of SCZ,[8] they were largely underrepresented in previous studies. Lam et al. reported the largest SCZ GWAS in individuals of EAS ancestries, and found several shared risk alleles between EUR and EAS.[9] However, the polygenic risk score could not be well-transferred across ancestries.[9] In particular, some of the robust candidate loci identified in EUR populations were not successfully replicated in EAS populations. One notable example is *C4A* localized to the major histocompatibility complex (MHC) class III region. The *C4A* variants have been repeatedly confirmed in EUR.[10] However, Yue et al. claimed different susceptibility variants in the MHC region in individuals of EUR and Chinese ancestry[11]; Lam et al. did not find a significant SCZ association in the MHC in the EAS populations.[9] These observations suggest that there are likely population-specific risk variants driven by different underlying genetic bases of SCZ across ancestries.

The evolutionary paradox of SCZ—it reduces reproduction but is highly prevalent in present-day human populations—brings about a long-standing discussion on the driving force of SCZ evolution in humans.[12–15] It has been hypothesized that SCZ represents, at least in part, a maladaptive by-product of adaptive changes during human evolution, as evidenced by the significant natural selection identified on genes potentially increasing the risk of SCZ.[16–19] Furthermore, natural selection signatures were consistently found in SCZ-associated

[1]State Key Laboratory of Genetic Engineering, Human Phenome Institute, Zhangjiang Fudan International Innovation Center, Center for Evolutionary Biology, School of Life Sciences, Department of Liver Surgery and Transplantation Liver Cancer Institute, Zhongshan Hospital, Fudan University, Shanghai 200032, China
[2]Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200438, China
[3]These authors contributed equally
[4]Lead contact
*Correspondence: denglian@fudan.edu.cn (L.D.), xushua@fudan.edu.cn (S.X.)
https://doi.org/10.1016/j.isci.2024.109560

loci across EAS and EUR populations,[9] whereas another study suggested that natural selection has driven genetic differentiation of SCZ-associated loci across global populations,[20] leading to the failed replication of the associations across ethnic populations. Notable examples of recent positive selection that occurred in specific populations include *SLC39A8* in EUR, and *TSPAN18* and *ST8SIA2* in the Asian populations,[21–23] supporting the hypothesis that local selective pressures acting on the pleiotropic genes may help to maintain SCZ risk alleles in the human gene pool. In addition, Neanderthal gene introgression, which contributed to around 2% of the modern human individual genome,[24] could have effects on the genetic architecture of SCZ. Current findings are largely based on the studies of EUR populations,[25,26] whereas the genetic footprints of archaic hominin introgression on the SCZ-associated loci in EAS have not been well characterized.

It has been well demonstrated and emphasized that a wide variety of population ancestries should be included when investigating the trait or disease mechanisms from the perspective of population genetics. In this work, we therefore attempted to provide a full characterization of the differentiated genetic predisposition of SCZ in EUR and EAS populations using the whole-genome sequencing data released by the 1000 Genomes Project Phase III and further investigated how the population admixture and local adaptations could have helped to shape the genetic heterogeneity of SCZ in modern humans. Our findings highlight the genetic differentiation of the SCZ-associated genes across ethnic populations and provide new insights into the driving forces of SCZ evolution in human populations.

RESULTS

Overview of the genetic differentiation of the SCZ-genes between EAS and EUR populations

We observed a clear separation between EAS and EUR populations on the 3,841 SCZ-associated single nucleotide polymorphisms (SCZ-SNPs) by principal component analysis (PCA) consistent with the whole-genome analysis (Figure S2). However, the measured genetic difference between EAS and EUR in this collection of SCZ-SNPs was larger compared to that at the genome-wide level (STAR methods). Compared with the strict control SNPs (sc-SNPs), the SCZ-SNPs showed larger nucleotide differences for pairwise haplotypes (1289.15 vs. 110.41 $\pm$ 5.98; $p < 2.2 \times 10^{-16}$, one-tailed Z test) and a higher F_{ST} weighted across variants on the population level (0.106 vs. 0.101 $\pm$ 0.007; $p = 1.85 \times 10^{-11}$, one-tailed Z test). Similar observations were obtained from the comparison between the SCZ-SNPs and the relaxed control SNPs (rc-SNPs) (Table S1; Figure S3). We conducted approximate pruning for the linkage disequilibrium (LD) among the 3,841 SCZ-SNPs, and then validated the aforementioned findings (Figure S4, STAR methods). Further validations were successfully conducted based on a set of confident SCZ-SNPs reported in two GWASs with large samples affected with schizophrenia (76,755 and 22,778 cases, respectively),[7,9] and 43 SCZ-SNPs identified in studies with comparable sample sizes for EAS (2,040, 2,413, and 4,384 cases, respectively)[27–29] and EUR populations (2,111, 2,413, and 4,528 cases, respectively)[30–32] (Table S2; Figure S5). As GWASs tend to identify common variants conferring disease risk in a population, we observed a biased minor allele frequency (MAF) spectrum at the SCZ-SNPs compared with the randomly selected SNPs representing the genome-wide MAF spectrum, and the EUR populations had fewer SCZ-SNPs at lower MAF than did the EAS populations (Figure S6A). However, the observed genetic differentiation at the SCZ-SNPs between EAS and EUR populations could not be fully attributed to the distinct MAF spectrum, as we have confirmed this result based on another set of random SNPs matching the SCZ-SNPs on the MAF spectrum ($p = 6.8 \times 10^{-157}$, two-sided Z test) (Figure S6B, STAR methods). This finding indicated that the population genetic background may affect the SCZ risk in EAS and EUR populations.

We next focused on the top 5% highly differentiated SNPs between EAS and EUR populations as indicated by the largest F_{ST} between EAS-EUR population pairs (max-$F_{ST(EAS\text{-}EUR)}$). We found that 1,052 of these SNPs were SCZ-associated (dSCZ-SNPs, abbreviated for differentiated SCZ-SNPs), making up 27.4% of the total SCZ-SNPs (Table S2). In total, 1,504 SCZ-genes encompassed 211,699 highly differentiated SNPs, and they were thus designated as dSCZ-genes (abbreviated for the differentiated SCZ-genes) (Figure 1A; Table S3; Figure S7). We further excluded the FIN and KHV populations that have been proven to be influenced by the population gene flow between Europe and Asia and found that 990 (94.1%) of the dSCZ-SNPs identified based on the comparison of five EUR and five EAS populations were confirmed (Table S4). Furthermore, functional annotations revealed that the dSCZ-genes were enriched in iron channel structure and function (e.g., regulation of postsynaptic cytosolic calcium ion concentration, odds ratio (OR) = 11.8, false discovery rate-corrected (FDR-corrected) $p = 0.0001$; neurotransmitter-gated ion channel clustering, OR = 7.1, FDR-corrected $p = 0.02$), synaptic component and function (e.g., retrograde *trans*-synaptic signaling, OR = 17.7, FDR-corrected $p = 0.004$; glutamatergic synapse, OR = 3.5, FDR-corrected $p = 4.53 \times 10^{-5}$), and hormone secretion (e.g., GnRH secretion, OR = 4.3, FDR-corrected $p = 0.0001$; insulin secretion, OR = 3.2, FDR-corrected p = 0.002) (Figure 1B; Table S5). Among the dSCZ-SNPs, 15 were missense variants (Table 1). The variant rs6083 showed larger differentiation between EAS and EUR populations ($F_{ST(EAS\text{-}EUR)}$ = 0.11–0.61, allele frequency difference (AFD) = 0.41) compared to the other missense variants, and it was reported to affect the cholesterol and triglyceride levels in the Han Chinese population.[33] In addition, we found two missense SNPs, rs3810449 and rs3810450, in perfect linkage to cf. schizophrenia risk,[7] and we found that they showed signals of positive selection (integrated haplotype score (iHS) = 2.25–3.72) in the EAS populations (Table 1). The top two dSCZ-SNPs showing the largest allele frequency difference between EAS and EUR populations were rs12602286 ($F_{ST(EAS\text{-}EUR)}$ = 0.63–0.80, AFD = 0.76) and rs959071 in *EPN2* ($F_{ST(EAS\text{-}EUR)}$ = 0.66–0.79, AFD = 0.76) (Figure 1A; Table S2). These two variants were in strong linkage disequilibrium in both EAS and EUR populations ($r^2 > 0.8$), and may affect the *EPN2* expression in the brain according to the Genotype-Tissue Expression (GTEx) data (Figures S8A and S8B). *EPN2* controls notch signaling activation via endocytosis,[34] and its high expression is fundamental in the signaling of the notch pathway,[35] which was reported to be associated with SCZ.[36] Other outstanding signals included rs732381 which overlapped a neuronal enhancer located within *CACNA1I* to cf. schizophrenia risk,[37] rs72639203 located in *SNORD3A* that encodes a long non-coding RNA implicated in neurogenesis,[38] and a *cis*-expression quantitative trait locus (eQTL) rs12911832 regulating *ADAM10* that was shown affected in patients with Alzheimer's disease.[39]

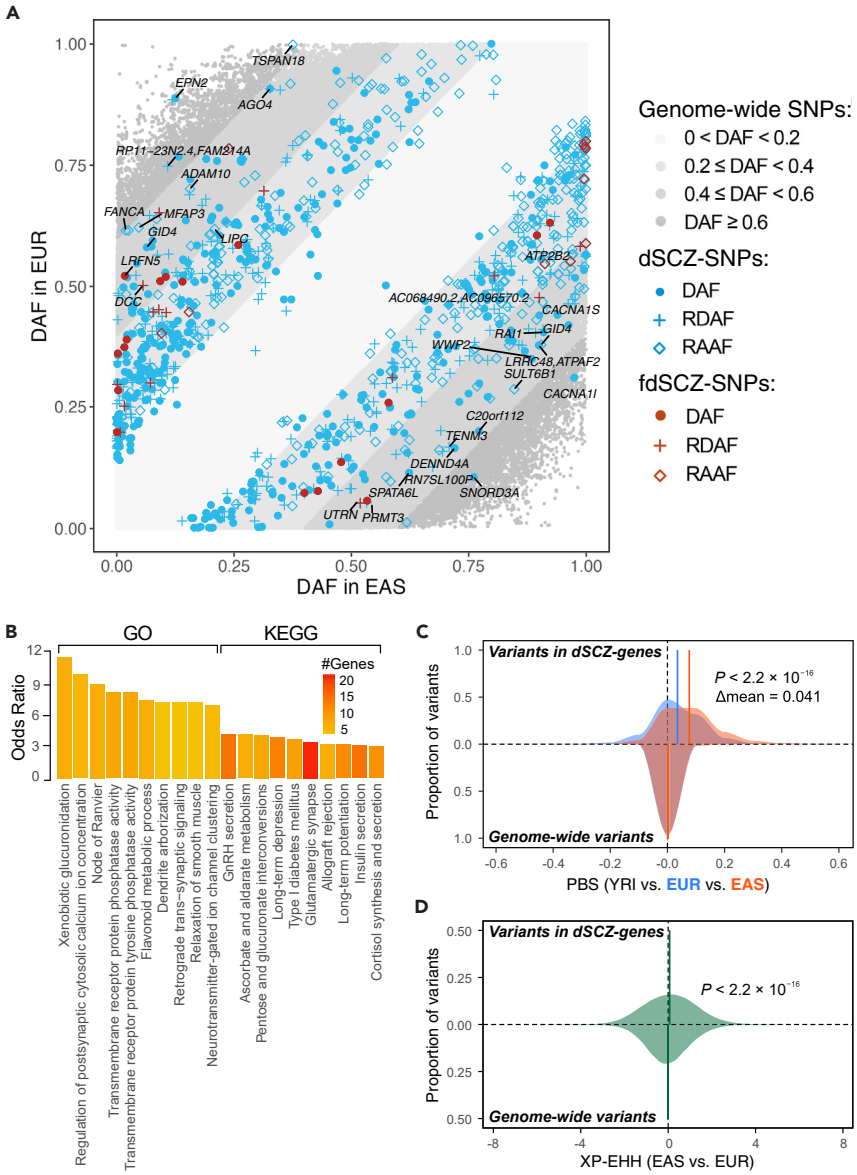


Figure 1. Overview of the genetic differentiation and differentiated selection of SCZ genes between EUR and EAS populations

(A) Allele frequency of the SCZ-SNPs in the EAS and EUR populations plotted on the background of genome-wide allele frequencies. DAF: derived allele; RDAF: risk derived allele; RAAF: risk ancestral allele.

(B) Functional enrichment of the dSCZ-genes conducted using clusterProfiler version 4.2.2.[81] The top 10 enriched categories of the gene ontology (GO) enrichment and the KEGG enrichment are shown in this plot, and a full list can be found in Table S5. Differentiated selection of the dSCZ-genes assessed between EUR and EAS populations based on (C) the population branch statistic (PBS), and (D) cross-population extended haplotype homozygosity (XP-EHH) calculated for EAS using EUR as the reference population. *P*-values were obtained by the one-tailed Students' t test.

Among the X-linked SCZ-SNPs, a missense variant *PJA1*-rs11539157 (c.1818G>T; p.E606D) showed the largest genetic differences between EAS and EUR populations (AFD = 0.24). The risk allele A[7] was more prevalent in EUR populations than in EAS populations (AF = 0.24 and 0 in EUR and EAS populations, respectively). *PJA1*, encoding ubiquitin ligase Ring-H2, is important for polyQ protein degradation and was abundantly expressed in the brain.[40] *PJA1* was reported to have protective potential against neurodegenerative diseases and variants are associated with numerous X-linked neurodevelopmental disorders.[41,42] We particularly checked for several SCZ-SNPs in chromosome X reported in the latest PGC GWAS research,[7] and found that all these variants showed substantial Asian-European differences (empirical *p* value <5% across the X chromosome) (Table S6).

Although allele-specific expression (ASE) was reported to have no substantial impact on the SCZ risk,[43] several examples have been identified; for example, rs4702-G downregulated *FURIN* in psychiatric disorders, and rs1625579-T linked to the miR-137 promoter and

Table 1. A list of 15 dSCZ-SNPs with missense variants

SNP	Gene	F_{ST}	AFD	Nucleotide and amino acids substitution	iHS	XP−EHH
rs6083[a,b]	*LIPC*	0.19/0.04/[0.11, 0.61]	0.41	c.644A>G: p.Asn215Ser	[−0.59, 0.76]/[−0.96, −0.23]	[−0.5, 0.55]
rs42945[a,b]	*NDRG4*	0.01/0.02/[0.3, 0.48]	0.41	c.51A>G: p.Gln18Arg	-/[0.79, 1.57]	[0.56, 0.87]
rs281377[a,b]	*FUT2*	0/0.03/[0.16, 0.36]	0.37	c.390C>T: p.Asn130Lys	[0.06, 0.63]/[−2.01, −1.15]	[0.41, 1.06]
rs698761[a,b]	*SLC3A1,PREPL*	0.06/0.05/[0.05, 0.37]	0.34	c.1854G>A: p.Met618Ile	[−1.14, 0.08]/[−0.17, 0.51]	[−0.47, 0.5]
rs2072736[a,b]	*ATXN1*	0/0.03/[0.18, 0.36]	0.33	c.474A>G: p.Ser149Gly	[−2.32, −1.49]/[−0.09, 0.07]	[−1.35, −0.81]
rs57646126[a,b]	*LTN1*	0.06/0/[0.23, 0.41]	0.31	c.2438C>T: p.Ala813Val	[2.08, 3.19]/-	[−1.53, −0.82]
rs1635[a,b]	*NKAPL*	0.04/0.02/[0.1, 0.34]	0.29	c.455C>A: p.Thr152Asn	[−1.48, 0.15]/[−2.86, −1.37]	[−2.82, −2.11]
rs3764002[a,b]	*WSCD2*	0.03/0.05/[0.03, 0.31]	0.29	c.797C>T: p.Thr266Ile	[0.65, 1.5]/[1.15, 1.52]	[1.7, 1.92]
rs11142[a,b]	*SORT1*	0.04/0.01/[0.11, 0.33]	0.24	c.594T>C: p.Phe198Leu	−/−	–
rs13306731[a,b]	*SOAT1*	0.02/0.01/[0.08, 0.26]	0.23	c.1577A>G: p.Gln526Arg	[0.5, 1.49]/[−0.26, 0.61]	[−0.95, 0.26]
rs950169[a,b]	*ADAMTSL3*	0.02/0/[0.11, 0.24]	0.22	c.4979C>T: p.Thr1660Ile	−/−	[−1.02, −0.13]
rs3810450[a,b]	*AC002116.7,THAP8*	0.02/0.01/[0.08, 0.24]	0.19	c.335A>G: p.Lys112Arg	[2.25, 3.72]/-	[−1.09, −0.7]
rs3810449[a,b]	*AC002116.7,THAP8*	0.02/0.01/[0.08, 0.24]	0.19	c.209G>A: p.Arg70His	[2.25, 3.72]/-	[−1.12, −0.7]
rs16897515[a,b]	*POM121L2*	0.08/0.1/[0, 0.23]	0.14	c.1738G>T: p.Gly580Cys	[−0.66, −0.66]/[−0.91, 0.07]	[0.57, 1.48]
rs1801133[a]	*MTHFR*	0.22/0.07/[0, 0.22]	0.07	c.788C>T: p.Ala263Val	[−1.03, 0.35]/[1.1, 2.27]	[0.77, 1.48]

AFD denotes the allele frequency difference between EUR and EAS calculated based on the combined dataset of populations from each continent. The max-$F_{ST(EAS)}$, max-$F_{ST(EUR)}$, and the range of $F_{ST(EAS\text{-}EUR)}$ are split by slashes (/) in the column named F_{ST}. The range of $iHS_{(EAS)}$ and that of $iHS_{(EUR)}$ are split by slashes (/) in the column named iHS. The range of $XP\text{-}EHH_{(EAS\text{-}CEU)}$ is shown in the column named XP-EHH. The superscribed letters a and b indicate that the cross-continental genetic differentiation at a variant can be validated using the high-coverage 1000 Genomes data and the HGDP data, respectively.

downregulated its expression.[44,45] In our data, rs4702 showed small genetic differentiation between EAS and EUR populations (max-$F_{ST(EAS\text{-}EUR)}$ = 0.026), whereas rs1625579 was identified as a dSCZ-SNP (max-$F_{ST(EAS\text{-}EUR)}$ = 0.161). Moreover, we identified three additional dSCZ-SNPs in the MHC region showing ASE, and the max-$F_{ST(EAS\text{-}EUR)}$ was estimated to be 0.121 at *BTN3A2*-rs12214031, 0.141 at *BTN3A2*-rs13218591, and 0.265 at *PBX2*-rs1004095 (Figure S9). We found rs12214031-T and the risk allele rs13218591-T[46] at *BTN3A2* were preferentially expressed in EAS and EUR populations (represented by Han Chinese and British, STAR methods). It was reported that the increased expression of *BTN3A2* might cf. risk for schizophrenia by altering excitatory synaptic function,[47] and we inferred that ASE could be one of the underlying mechanisms. However, the non-risk allele rs1004095-A[46] at *PBX2* had higher expression in the EUR populations than the risk allele. *PBX2* is a key transcription factor of multiple gene regulatory networks and regulates embryonic development, and the dysregulation of *PBX2* can lead to various diseases, such as cancer.[48] Therefore, the differential ASE may act as a predisposing factor for schizophrenia and show differences in EAS and EUR populations.

Screening for dSCZ-SNPs with small within-continent variations

When speculating the genetic differentiation within each continent, we found higher F_{ST} values of the SCZ-SNPs across EAS populations as compared to the EUR populations (average F_{ST} = 0.008 vs. 0.006, p = 0.026, one-tailed Wilcoxon rank-sum test; weighted F_{ST} = 0.009 vs. 0.006) (Figure S10). To determine possible contributors to the genetic heterogeneity of SCZ between EAS and EUR populations, we controlled for the within-continental variations by excluding SNPs showing large population differentiation in either continent (STAR methods). Only 43 of the 1,052 dSCZ-SNPs showed small within-continental variations (designated as filtered dSCZ-SNPs; fdSCZ-SNPs), in which 26 are with known risk alleles (Table S2), but 38.9% of the dSCZ-genes (n = 586) contained 7,392 SNPs showing large between-continent differentiation but small within-continent differentiation, and they were thus designated as filtered dSCZ-genes (fdSCZ-genes) (Figure 1A). Following Yang et al.,[49] we performed a two-sided Fisher's exact test for genome-wide variations and found that all the fdSCZ-SNPs could tell the between-continent differences in allele frequency ($p < 0.05$ after adjusting for the false discovery rate). The enriched functional categories of the 586 fdSCZ-genes were similar to those of the 1,504 dSCZ-genes (Figure S11; Table S7).

In the protein-coding genes, the top fdSCZ-SNP was rs1950829, located in *LRFN5* (max-$F_{ST(EAS)} = 1.63 \times 10^{-4}$, max-$F_{ST(EUR)} = 5.66 \times 10^{-4}$, $F_{ST(EAS\text{-}EUR)}$ = 0.422–0.528, AFD = 0.502), which is a brain-specific gene needed for synaptic development and plasticity[50] (Table 1). Among the other fdSCZ-genes, *SLC5A10* was the most outstanding, attributing to rs1624825 ($F_{ST(EAS\text{-}EUR)}$ = 0.7–0.82; AFD = 0.79) (Table 2). *SLC5A10* encodes a member of the solute carrier superfamily, and it plays a vital role in neuropsychiatric mechanisms by affecting transport expression, function, and regulation in the neurotransmitter systems.[51] Importantly, rs1624825 located at the 3′-UTR region is a microRNA target site, and mounting evidence has implicated the importance of microRNAs in the pathology of schizophrenia.[52] In addition, we found the likely damaging variant rs1001292 ($F_{ST(EAS\text{-}EUR)}$ = 0.49–0.58; AFD = 0.6; Combined Annotation Dependent Depletion (CADD) = 17.6; Genomic Evolutionary Rate Profiling (GERP) = 3.1) in *TSPAN18* of the candidates variants. *TSPAN18* was previously shown to be the top dSCZ-gene

Table 2. Top 10 protein-coding dSCZ genes with small within-continental variation

Gene	Top SNP	AFD	PBS_{EAS}	PBS_{EUR}	iHS	XP-EHH	Gene function
SLC5A10[a]	rs1624825	0.79	0.07	0.61	–/[1.41, 1.79]	[−1.45, −0.33]	It encodes the solute carrier (SLC) superfamily which plays a vital role in neurodegenerative disorders.[51]
IQGAP2[a]	rs10037956	0.68	0.53	−0.08	[1.24, 1.62]/ [−0.34, 0.03]	[1.68, 1.68]	It acts as a key regulator of dendritic spine number, and plays an important role during neuronal signal transmission and neuron connectivity.[83]
SLC8A1	rs7590169	0.67	0.08	0.35	[−0.65, 0.82]/ [0.17, 0.64]	[−0.54, −1.14]	It encodes the solute carrier (SLC) superfamily which plays a vital role in neurodegenerative disorders.[51]
DACH1[a]	rs1981738	0.66	0.09	0.31	–/–	–/–	It encodes a transcription factor acting as a neurogenic cell-fate determining factor.[84]
UTRN	rs4305737	0.64	0	0.38	[−0.51, 0.45]/ [1.93, 2.7]	[−1.82, −1.94]	It involves in neuronal cytoskeleton organization and intracellular transport.[85]
SMYD3	rs985919	0.62	−0.01	0.39	[−0.67, −0.27]/ [−2.92, −2.58]	[−1.86, −1.47]	The gene is a methyltransferase responsible for methylation of H3K4 which has been investigated in postmortem brain tissues.[86]
ATP2B2[a]	rs2111751	0.61	0.49	−0.09	-/[−1.71, −1.19]	[0.26, 0.74]	It encodes the plasma membrane calcium-transporting ATPase 2 which plays an essential role in intracellular calcium homeostasis and is reported to confer schizophrenia.[61]
CNTNAP5[a]	rs34728680	0.61	0.54	−0.14	-/[−0.69, −0.5]	[1.12, 1.89]	It encodes proteins functioning in the nervous system as cell adhesion molecules and receptors, and is reported to confer susceptibility to autism spectrum disorder.[87]
FOXP1	rs1053797	0.59	0.07	0.31	-/[−2.03, −0.68]	[−1.09, 0.18]	It shows association signals in a cross-disorder meta-analysis of ASD and genome-wide association studies in SCZ.[46]
PRICKLE2[a]	rs61868826	0.56	0.25	0.11	[−0.31, 0.68]/ [1.22, 2.06]	[−2.47, −2.12]	The gene is involved in human neuronal development and that pathogenic variants in the gene cause neurodevelopmental delay, behavioral difficulties and epilepsy in humans.[88]

The range of $iHS_{(EAS)}$ and that of $iHS_{(EUR)}$ are split by slashes (/) in the column named iHS. The range of $XP\text{-}EHH_{(EAS\text{-}CEU)}$ is shown in the column named XP-EHH. The superscribed letters a indicate that a dSCZ-gene with small within-continental variation can be validated using the high-coverage 1000 Genomes data.

and the target of selection in the EAS populations, but only intronic variants with possibly benign consequences were identified to be dSCZ-SNPs in this gene.[21] Based on the GTEx data, we found these two variants, rs1624825 and rs1001292, showed correlations with the expression of *SLC5A10* and *TSPAN18*, respectively, in the brain cortex (Figures S8C and S8D). Despite that all of the fdSCZ-SNPs were predicted to be non-coding variants, we found one missense variant rs76082815 (max-$F_{ST(EAS)}$ = 0.001, max-$F_{ST(EUR)}$ = 0.006, min-$F_{ST(EAS\text{-}EUR)}$ = 0.052) by examining the top 5% of SNPs with larger min-$F_{ST(EAS\text{-}EUR)}$ values (>0.048) than the other genome-wide SNPs. The missense Pro-Leu variant rs76082815 (c.2393C>T; P798L) located in *LMO7* at 13q22.2 was identified to be associated with the intermediate phenotypes for psychotic disorders.[53] These findings again highlight the essential role of genetic ancestry in susceptibility to schizophrenia, and demonstrate the potential of genetic ancestry prediction based on the schizophrenia genetic architecture.

Stronger selection signals on the SCZ genes in EAS than in EUR populations

Previous analyses showed that gene loci associated with SCZ are significantly more prevalent in genomic regions that are likely to have undergone recent positive selection in humans.[18] Consistently, we observed a significant excess of SCZ-SNPs in the genome-wide candidate adaptive regions indicated by iHS over those associated with several other traits, e.g., Alzheimer's disease ($p = 2.67 \times 10^{-5}$, the Fisher's exact test), bipolar disorder ($p = 5.39 \times 10^{-6}$), Body Mass Index ($p = 2.53 \times 10^{-6}$), and cardiovascular disease ($p = 2.50 \times 10^{-4}$) (Table S8, STAR methods). To investigate the magnitude of allele frequency changes of the variants in the dSCZ-genes in the EAS and EUR populations

Table 3. The SCZ-SNPs show outstanding signals of selection measured by iHS in EAS and EUR, respectively

		EAS					EUR				
SNP	Gene	CHB	CHS	CDX	JPT	KHV	CEU	GBR	TSI	IBS	FIN
rs142972412	-	−6.26*	−6.27*	**−8.71***	−5.46*	−4.72*	−8.01*	−7.95*	−7.00*	−7.66*	**−9.52***
rs9274657[a]	-	−1.75	−1.04	2.80*	**−5.11***	−1.28	−3.72*	−3.48*	−0.90	0.93	−3.49*
rs191239160	*HLA-DRB6*	−0.98	−1.75	**−5.05***	−3.99*	−1.62	−3.67*	−2.25*	−0.78	−1.07	−3.40*
rs9274623[a,b]	*HLA-DQB1*	3.77*	**4.78***	2.37*	–	3.94*	0.78	3.71*	3.30*	2.94*	1.42
rs9274390	*HLA-DQB1*	1.14	**4.27***	0.35	2.60*	2.79*	**−5.84***	−1.02	−1.38	0.36	−1.17
rs9269271	-	−1.02	−1.50	**−4.21***	−0.96	0.01	−1.90	−1.05	−0.26	−0.32	−2.29*
rs77296290[a]	*HLA-DQB1*	0.43	3.16*	1.40	**4.13***	0.30	0.67	3.61*	2.04*	0.04	−1.84
rs140849564	*HLA-DRB5*	−2.42*	−3.48*	**−4.11***	−2.27*	−0.29	−1.47	0.16	1.04	0.56	**−6.81***
rs9270074	*HLA-DRB1*	0.17	−0.97	**−4.07***	−1.30	0.32	−1.94	−0.12	−0.11	0.30	−2.49*
rs184538485	*HLA-DRB5*	−0.38	−2.28*	**−3.95***	−2.67*	−0.46	−4.39*	−5.20*	−0.76	−1.52	**−8.54***
rs9274299	*HLA-DQB1*	−1.32	–	–	–	–	**−4.50***	0.90	−1.02	−1.12	–
rs4075330[a,b]	*ALG12*	−2.17*	−2.54*	−1.65	−2.88*	−3.17*	**−4.46***	−3.90*	−3.69*	−4.27*	−3.35*
rs5769765	*ZBED4*	−2.07*	−2.25*	−2.02*	−2.98*	−3.23*	−3.10*	−3.40*	−3.50*	**−4.11***	−3.31*
rs2596500	-	–	–	–	–	–	**3.97***	3.86*	3.48*	3.32*	2.77*
rs2760981	-	−3.10*	−2.66*	−3.82*	−2.14*	−1.14	−2.49*	−2.55*	−2.17*	−1.41	**−3.94***
rs7719676	*ZSWIM6*	–	–	–	–	–	**3.89***	3.33*	3.21*	3.54*	3.29*

The top ten SNPs with the highest maximum iHS value in the EAS or EUR populations are shown in this table, and a full list can be found in Table S9. The most significant iHS values among the EAS or EUR populations are bolded, and those that reach the significance level at $p < 0.05$ are indicated with asterisks. The superscribed letters a and b indicate that natural selection at an SCZ-SNP can be validated using the high-coverage 1000 Genomes data and the HGDP data, respectively.

relative to their divergence from the African population, we calculated the population branch statistic[54] (PBS) for each continental group using the African YRI population as an outgroup. Interestingly, the EAS populations showed a significantly greater excess of PBS compared with the EUR populations ($mean_{EAS}-mean_{EUR} = 0.041$; $p < 2.2 \times 10^{-16}$, one-tailed Student's t test) when compared with the genome-wide level ($mean_{EAS}-mean_{EUR} = 0.002$; $p < 2.2 \times 10^{-16}$, one-tailed Student's t test), indicating that these loci may be under selective pressure in the EAS populations (Figure 1C). Moreover, this finding was further confirmed by the haplotype-based estimation of recent selective sweeps. Although the 3,841 SCZ-SNPs collectively showed a similar extent of recent positive selection in the EAS and EUR populations (Figure S12), the 211,699 highly differentiated variants in the dSCZ-genes showed higher cross-population extended haplotype homozygosity (XP-EHH) values for the EAS populations when using CEU as a reference population (Figure 1D). This result was supported by the analyses on a subset of the dSCZ-SNPs, which were reported in the latest GWASs of SCZ with a large sample size[7,9] (Figure S13). We did not find any preference for natural selection acting on functional variants or non-functional variants associated with SCZ (Figure S14).

In general, selective sweeps identified in the SCZ genes were quite different across populations in either continent, and most of the adaptations likely occurred in a single or a few populations (Table 3; Figures S15A and S15B; Table S9). For example, rs11038172 at *TSPAN18* was reported to have experienced recent positive selection in the combined EAS population of the 1000 Genomes dataset.[21] However, we tested for each EAS population separately and found no sign of adaptation at this variant in CHS (Figure S15C; Table S9). We also found that *CACNA1I*-rs8139773 showed a signature of selection in three of the EAS populations (iHS = −2.04 to −2.01 for CHB, CHS, and JPT; max-$F_{ST(EAS-EUR)} = 0.167$). *CACNA1I* is involved in the calcium channel functions and might contribute to the risk of schizophrenia in the Chinese population.[55] *CADM2* was detected with signals of positive selection in the southern Chinese populations (CHS and CDX) (Figure S15D). This gene encodes a mediator of synaptic signaling enriched in the brain. It was reported to be a potent regulator of systemic energy homeostasis,[56] and showed possible influence on vitamin D concentrations in animal studies.[56] We did not find signals of positive selection on the haplotypes at *EPN2*, the top gene showing the largest differentiation between EAS and EUR populations. However, this gene showed an excess of rare variants over common variants in EUR populations (Tajima's D = −2.27 and $p = 5.42 \times 10^{-3}$), which also indicated footprints of directional selection (Figure S16). All of these results suggested that local adaptation might play a role in shaping the difference in the genetic basis of SCZ between EAS and EUR populations.

Validation of the dSCZ-SNPs using the high-coverage genomes from the 1000 Genomes Project and the Human Genome Diversity Project

We further validated the dSCZ-SNPs primarily identified in the 1000 Genomes Project data sequenced with a mean depth of 7.6× by using the updated genome sequences with a mean depth of 30×. Most of the SCZ-SNPs showed similar allele frequencies between the two versions of

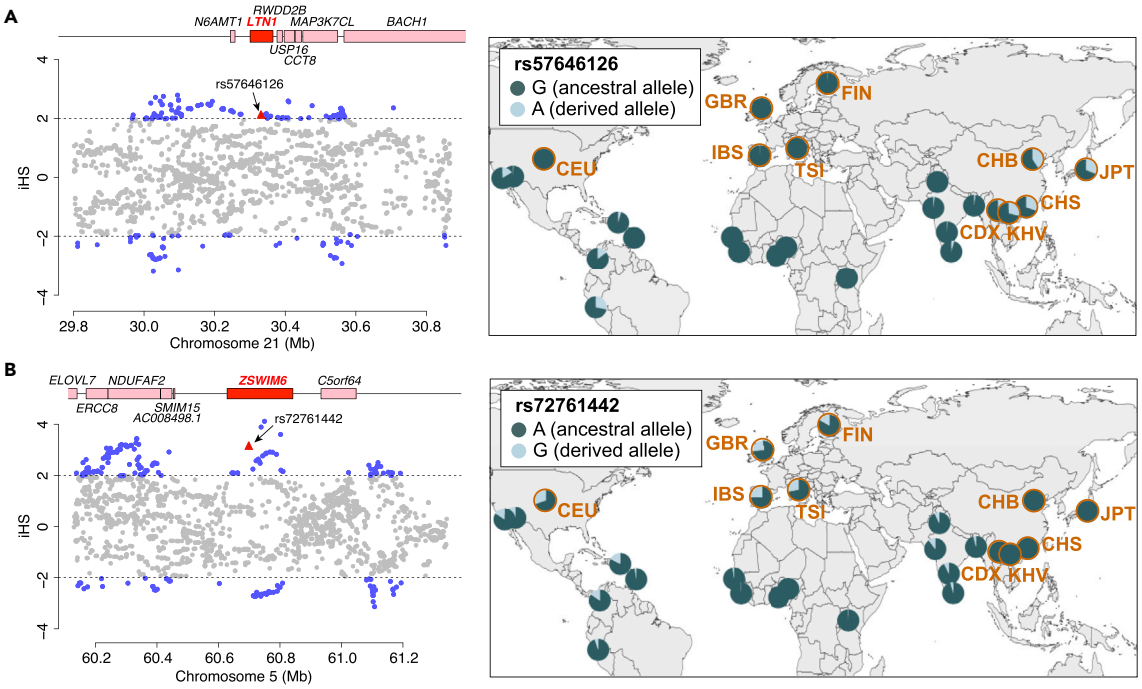


Figure 2. Signatures of recent selection at *LTN1* and *ZSWIM6*

(A) Signals of recent selective sweeps in the EAS popultions and allele frequency distribution at *LTN1*-rs57646126.. (B) Signals of recent selective sweeps in the EUR populations and allele frequency distribution at *ZSWIM6*-rs72761442. In each plot, the left panel shows the estimates of iHS at each SNP (blue dots: SNPs with |iHS| > 2; red triangles: the key adaptive SNP associated with SCZ), and the right panel shows the allele frequency distribution of the key adaptive SNPs obtained via PGG.SNV (https://pog.fudan.edu.cn/pggsnv/).

the 1000 Genomes data (Figure S1). We found 990 in the 1,052 dSCZ-SNPs could be supported by the high-coverage genome data, including the 15 missense dSCZ-SNPs and the top dSCZ-SNPs (rs12602286 and rs959071 in *EPN2*) (Table S2). We also referred to the population data sequenced to 30× on average released by the Human Genome Diversity Project (HGDP) as they represent greater EAS and EUR population diversity than the 1000 Genomes Project, and confirmed 931 dSCZ-SNPs including 14 missense variants. More importantly, two additional missense dSCZ-SNPs, *SECISBP2L*-rs11854184 and *DMKN*-rs12460932, were identified in both validation datasets. SECISBP2L is specifically expressed in differentiating oligodendrocytes and the autonomous regulation of oligodendrocytes differentiation is mediated by the SECISBP2L-DIO2-T3 pathway during myelin development.[57] Using the high-coverage 1000 Genomes data, we found two eQTLs in *VPS45* to be dSCZ-SNPs ($F_{ST(EAS-EUR)}$ = 0.16–0.22, AFD = 0.20 for rs12138231; $F_{ST(EAS-EUR)}$ = 0.13–0.20, AFD = 0.17 for rs2318763). *VPS45* encodes proteins that have roles in early to late endosome maturation[58] and in recycling back to the plasma membrane.[59] Endosomal trafficking is critical for processes thought to be disrupted in SCZ as alterations in endocytic trafficking can have dramatic effects on postsynaptic function and plasticity.[60] Regarding the 43 fdSCZ-SNPs, 19 were validated in the high-coverage 1000 Genomes data (Table S2), among which *ATP2B2*-rs9879311 showed the largest differentiation between EAS and EUR ($F_{ST(EAS-EUR)}$ = 0.36–0.42). *ATP2B2* encodes the plasma membrane calcium-transporting ATPase 2, which plays an essential role in intracellular calcium homeostasis and is reported to cf. schizophrenia.[61] In particular, the aforementioned missense variant *SECISBP2L*-rs11854184 and regulatory variant *VPS45*-rs12138231 showed as fdSCZ-SNPs in this data. Moreover, we confirmed the top signal of between-continent differentiation at *EPN2* in the HGDP data. The HGDP EUR population showed signals of positive selection (represented by rs959071 with iHS = 2.27 and $F_{ST(EAS-EUR)}$ = 0.664), consistent with our primary findings (Tajima's D = −2.27 and p = 5.42 × 10^{-3} in EUR).

Differential natural selection at *LTN1* and *ZSWIM6*

Some dSCZ-genes showed consistent selective sweeps in all of the populations from one continent but were neutral in the other continental populations. One notable example is *LTN1*, which plays a role in ribosome quality control and was reported to have a significantly higher expression in the peripheral blood in SCZ patients than in the healthy controls.[62] We observed an excess of haplotype extension at a 61.7 kb region around *LTN1* in the EAS populations (Figures 2A and S17A). The frequency of the adaptive haplotype reached >0.23 in the EAS populations but was very rare in the EUR populations. We found that the adaptive haplotype carried a missense SCZ-SNP rs57646126 (c.2438C>T; A813V) reported to be associated with SCZ in EUR populations with large population differences between EAS and EUR populations ($F_{ST(EAS-EUR)}$ = 0.225–0.410). The derived allele frequency (DAF) at rs57646126 was higher in CHB (0.408) compared with other EAS

Table 4. Differentiated Altai Neanderthal introgression detected at the SCZ-related loci

Region (Mb)	Length (Kb)	Genes	SCZ-SNP	AFD	max-$F_{ST(EAS)}$	max-$F_{ST(EUR)}$	max-$F_{ST(EAS-EUR)}$	Potential effects on SCZ	Modern human detected with introgression	TMRCA (KYA, mean ± sd)
Chr1:150.00–150.15	149	*VPS45, PLEKHO1*	rs140505938	0.168	9.38×10^{-5}	6.63×10^{-3}	0.198	Non-risk	EUR	51.56 ± 5.09
Chr1:150.27–150.51	244.2	*MRPS21, RPRD2, PRPF3*	rs11587682	0.137	9.38×10^{-5}	0.018	0.181	Unknown	EUR	68.23 ± 5.56
Chr3:50.19–50.42	230.2	*SEMA3F, RASSF1, GNAT1, GNAI2, SEMA3B, HYAL3, HYAL2, ZMYND10, CACNA2D2, IFRD2, NPRL2, SLC38A3, TUDC2, NAA80, NAT6, HYAL1, TMEM115, LSMEM2 CYB561D2*	rs9882618	0.607	0.068	0.043	0.692	Risk	EAS	54.67 ± 8.32
Chr10:65.01–65.26	249.1	*JMJD1C*	rs7901794	0.205	0.112	0.015	0.320	Risk	EAS	68.95 ± 1.90
Chr2:68.36–68.50	138	*WDR92, PPP3R1*	rs12052801	0.026	0.030	0.070	0.00	Non-risk	EAS, EUR	35.27 ± 6.21

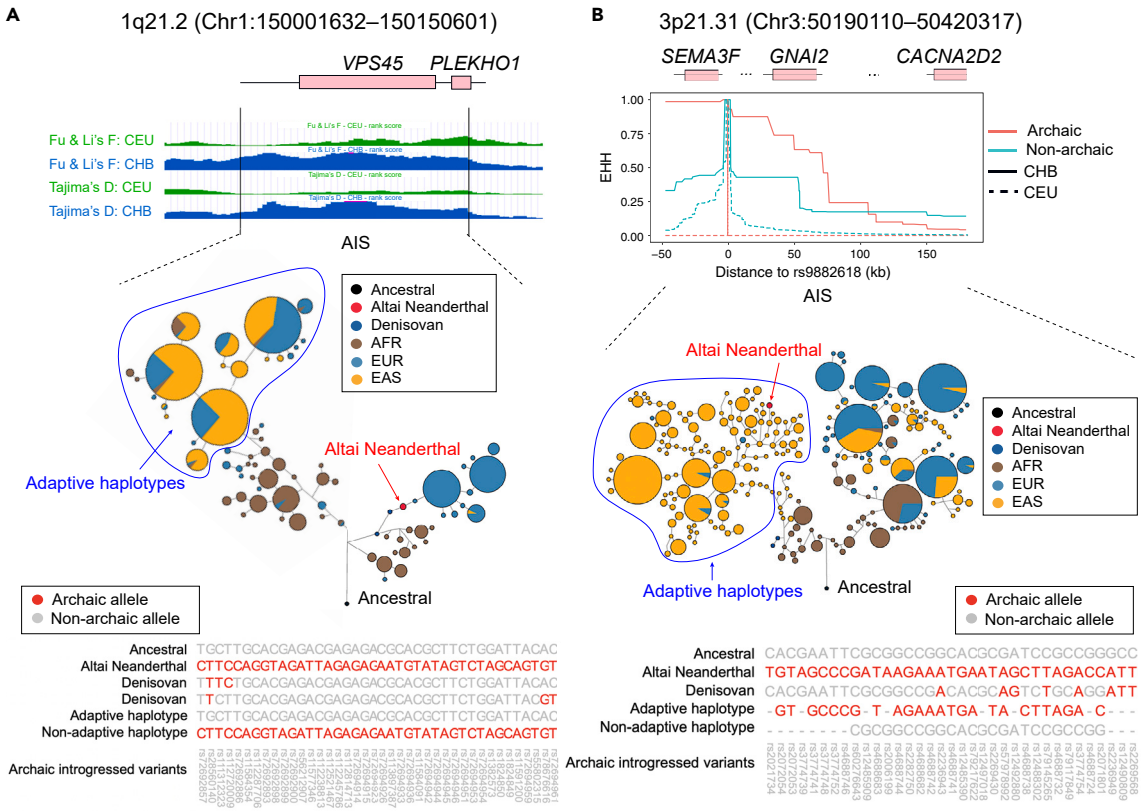
populations (0.237–0.308). We then estimated the iHS statistic to detect the presence of a signal of recent selection in EAS populations. The SCZ-SNP rs57646126 showed consistent signals of positive selection in all five EAS populations (iHS = 2.07–3.19), indicating that the derived A allele had an unusually long extended haplotype as compared to the ancestral allele G. And the onset of selection at this region was inferred to be 23,759 (95% confidence interval (CI): 21,872–25,812) years ago, as estimated from the missense SCZ-SNP.

Furthermore, we found a EUR-specific adaptive haplotype spanning 213.9 kb at *ZSWIM6* (Figures 2B and S17B). Knockout mouse studies have shown that *ZSWIM6* was expressed in the key brain regions involved in the pathogenesis of schizophrenia,[63] and it might regulate the morphology of striatal neurons and motor function.[64] The adaptive haplotype was carried by 43.9% of the EUR individuals but was very rare in EAS individuals. We found a non-risk allele rs72761442-G in the adaptive haplotype. The association between rs72761442 and SCZ was reported in populations with European ancestry.[46] It showed a large genetic differentiation difference between EAS and EUR ($F_{ST(EAS-EUR)}$ = 0.158–0.306). *ZSWIM6* also showed consistent signals of positive selection in all five EUR populations, and the non-risk allele rs72761442-G could be one of the targets (iHS = 2.78–3.22). Strikingly, this candidate adaptive allele was confirmed in all the EUR populations in the high-coverage 1000 Genomes data and the combined EUR population in the HGDP data (Figure S18) The onset of selection of this variant was estimated to be 14,152 years ago (95% CI: 12,316–15,955), after the split of EUR and EAS populations.

Adaptive introgression increases the SCZ risk in EAS

According to the Altai Neanderthal introgressed segments (spanning a total of 439.5 Mb in EUR, and 451.3 Mb in EAS) and alleles (n = 519,546 in EUR; n = 592,014 in EAS) detected using ArchaicSeeker 2.0,[65] we found 179 (6.9%) and 205 (7.9%) LD-pruned SCZ-SNPs affected by the introgressed segments in the EUR and EAS populations, respectively, in which only 12 (6.7%) and 7 (3.4%) were identified to be introgressed loci (STAR methods) (Table S10). We quantified the relationship between SCZ and Altai Neanderthal introgression using stratified-LD score regression but observed no enriched heritability for SCZ in either population (enrichment = 0.77, p = 0.414) or EUR (enrichment = 0.78, p = 0.229) groups. To assess the direction of effects on SCZ across the Altai Neanderthal-derived variants that have survived in the modern human genome, we applied the signed LD profile regression. We predicted the protective effects of the Altai Neanderthal introgressed variants on SCZ in the EUR populations (functional correlation = −0.0037, $p = 1.1 \times 10^{-5}$) as was reported in a previous study,[26] whereas the effect in the EAS populations tended to be in the opposite direction but it did not reach the significance level (functional correlation = 0.0007, p = 0.64).

We intersected the SCZ-SNPs and their linked loci ($r^2 > 0.8$) with the introgressed variants and found some examples of the inconsistent effect of Altai Neanderthal introgression on SCZ in different continental populations (Table 4; Figure S19). For example, we identified a region spanning around 150 kb at 1q21.2 (Chr1:150001632-150150601) with a large genetic differentiation between EUR and EAS populations represented by the SCZ-SNP rs140505938 (max-$F_{ST(EAS)} = 9.38 \times 10^{-5}$, max-$F_{ST(EUR)} = 6.63 \times 10^{-3}$, $F_{ST(EAS-EUR)}$ = 0.121–0.198; AFD = 0.168). The European version of haplotypes in this region was likely inherited from the Altai Neanderthal introgression, and the archaic allele was in strong linkage disequilibrium (r^2 = 0.8–1) with the rs140505938 allele that predicted to reduce the risk of SCZ[6] (Figure 3A). However, the Asian version



Ancestral TGCTTGCACGAGACGAGAGACGCACGCTTCTGGATTACAC
Altai Neanderthal CTTCCAGGTAGATTAGAGAGAATGTATAGTCTAGCAGTGT
Denisovan TTTCTGCACGAGACGAGAGACGCACGCTTCTGGATTACAC
Denisovan TTCTTGCACGAGACGAGAGACGCACGCTTCTGGATTACGT
Adaptive haplotype TGCTTGCACGAGACGAGAGACGCACGCTTCTGGATTACAC
Non-adaptive haplotype CTTCCAGGTAGATTAGAGAGAATGTATAGTCTAGCAGTGT
Archaic introgressed variants

Ancestral CACGAATTCGCGGCCGGCACGCGATCCGCCGGGCC
Altai Neanderthal TGTAGCCCGATAAGAAATGAATAGCTTAGACCATT
Denisovan CACGAATTCGCGGCCGACACGCAGTCTGCAGGATT
Adaptive haplotype -GT-GCCCG-T-AGAAATGA-TA-CTTAGA-C---
Non-adaptive haplotype ---------CGCGGCCGGCACGCGATCCGCCGG---
Archaic introgressed variants

Figure 3. Examples of differentiated archaic introgression at the dSCZ-genes

The top to the bottom panels show signals of natural selection, haplotype networks, and nucleotide sequences at two Altai Neanderthal introgressed regions: (A) 1q21.2 (Chr1:150001632–150150601) and (B) 3p21.31 (Chr3:50190110–50420317). Statistic tests for neutrality in these regions were provided by the 1000 Genomes Selection Browser 1.0.[82] Haplotype networks were constructed based on the archaic introgressed variants and the linked variants ($r^2 > 0.8$). AIS: archaic introgressed segments. Nucleotide polymorphisms are indicated by the dashes ("–").

of haplotypes in this region carrying the non-archaic alleles showed strong signals of selective sweep indicated by Fu and Li's F* (−7.44 to −5.45, p = 2.51 × 10^{-4}–4.17 × 10^{-3}) and Tajima's D (−2.73 to −2.27, p = 7.59 × 10^{-5}–8.91 × 10^{-3}) (Figure 3A). *VPS45* and *PLEKHO1* were affected by the Altai Neanderthal introgression in this region. *PLEKHO1* is a previously identified ubiquitination-related molecule that specifically targets the linker region between the WW domains of Smurf1 to promote the ubiquitination of Smad1/5,[66] which disrupts osteoblastic bone formation.[67] It has been suggested that increased PLEKHO1 suppressed Smad-dependent BMP signaling to inhibit bone formation during aging. In addition, we observed Altai Neanderthal-like adaptive haplotypes at 3p21.31 (Chr3:50190110–50420317) specifically in the EAS populations (Figure 3B). The Altai Neanderthal-like haplotypes also carried an SCZ risk allele rs9882618-T.[68] We found that the risk allele frequency at rs9882618 reached 0.486 in the EAS populations but remained very rare (<0.045) in the EUR populations. Among the five EAS populations, the frequency of rs9882618-T was lower in JPT (0.486) than in the other four populations (0.602–0.677). Local adaptation at 3p21.31 was reported in the EAS populations, possibly driven by the latitude-dependent ultraviolet-B irradiation.[69] Interestingly, we found the SCZ risk allele was in strong linkage with the reported adaptive alleles (r^2 = 0.98) and showed signals of positive selection in the HGDP EAS population (iHS = −2.48) (Figure S20). Consistent with a previous study,[69] the extended haplotype homozygosity (EHH) analysis showed that the haplotype carrying the risk allele had decayed much more slowly than the haplotype carrying the non-risk allele (Figure 3B). We inferred the onset of positive selection at this region to be around 6,000 years in EAS populations, according to the allele frequency trajectories in the ancient genomes (Figure S21). We also detected Altai Neanderthal-like haplotypes at another region of 1q21.2 (Chr1:150265619–150509820) in the EUR populations carrying the SCZ-SNP rs11587682[70] (Figure S22). The archaic haplotypes in this region showed EHH, indicating possible positive selection. The positive selection of the archaic haplotypes may contribute to the genetic differentiation between EAS and EUR populations at this locus (max-$F_{ST(EAS)}$ = 9.38 × 10^{-5}, max-$F_{ST(EUR)}$ = 0.018, $F_{ST(EAS-EUR)}$ = 0.089–0.182; AFD = 0.137). Among the genes encompassed by this introgression segment, *RPRD2* encodes a constitutively expressed protein that potently restricts HIV replication during reverse transcription[71]; abnormal *PRPF3* expression is potentially associated with carcinogenesis, such as hepatocellular carcinoma.[72] Recapitulating the aforementioned findings, we propose that the evolution of SCZ in humans could have been paralleled with positive selections acting on some other human traits, such as several protective traits in response to latitude-dependent ultraviolet-B irradiation, immunological disease, and cancer.

DISCUSSION

Our analyses indicated considerable heterogeneity of SCZ in human populations based on the reported genotype-phenotype associations. We observed a substantial excess of cross-continental population differentiation in the SCZ-SNPs compared to the genome-wide level. Furthermore, we revealed 1,052 SCZ-SNPs and 1,504 SCZ genes that were highly differentiated between EAS and EUR populations, and most of them were validated with the high-coverage 1000 Genomes data and the HGDP data. Notable examples include the top-ranked gene *EPN2* under directional selection in the EUR populations, missense variants in *LTN1* and *THAP8* showing consistent selection signals in EAS populations, and several ASEs in *BTN3A2 and PBX2*. After adjusting for the within-continental variation, we further screened 43 SCZ-SNPs and 7,349 other SNPs in the SCZ genes with the potential of indicating population genomic ancestry and distinguishing the EAS and EUR populations. These loci might have experienced parallel or convergent evolution among populations from either continent, and could be considered common signatures for disease risk prediction or therapeutic targets on the continent level. In particular, some of the variants that had no reported association with SCZ were highlighted in our results, such as the regulatory variant *SLC5A10*-rs1624825 showing the most outstanding west-east divergence, and the damaging variant *TSPAN18*-rs1001292 that provides novel insights into a previously identified gene with evolutionary significance in EAS populations.[21] In addition to the predicted functional importance of the highly differentiated variants in these reported SCZ-genes, these two signals also showed potential regulatory effects according to the gene expression in brain tissues in the GTEx database. ASE was identified at several variants but was not observed at the top signals. Considering that the ASE analysis was conducted based on the gene expression data in whole blood, we would expect more findings when nervous tissues are analyzed in the future.

We further explored the evolutionary driving forces of the observed genetic differentiation. We inferred that the differentiated natural selection on the SCZ-SNPs between EAS and EUR populations, especially the selective preference acting on the SCZ risk allele in the EAS populations, could be driven by the trade-off in the evolution of other human traits or diseases. Moreover, we hypothesized that ultra-violet radiation or its related phenotypes influenced SCZ evolution, as evidenced by two adaptive loci in EAS. One is the Altai Neanderthal introgressed region at 3p21.31. The adaptive haplotype may increase the SCZ risk but also affects the cellular response to the ultraviolet-B irradiation.[69] The other is *CADM2*, at which we observed latitude-related local adaptations. This gene plays a role in regulating the vitamin D concentration, which is closely related to ultraviolet-B irradiation.[56] Ultraviolet-B irradiation was demonstrated to be a confounding variable of human disease severity and seasonal variance, including SCZ, by causing damage to DNA in an embryo.[73] Other research suggested that vitamin D status is highly correlated with the manifestation of clinical symptoms and treatment response in patients with psychiatric disorders,[74,75] possibly via its action on the regulation of inflammatory and immunological processes.[76] We should also note that SCZ has correlations with various traits, especially some non-cognitive traits, such as body mass index and obesity,[77] and we cannot rule out their potential interactions with SCZ in this study.

The extreme divergence of the SCZ-SNPs between EUR and EAS populations suggests that differential population histories may serve as one of the primary reasons for failed replications of the reported GWAS-significant SNPs for SCZ among ethnic populations, and this reasoning can be applied to multiple psychiatric disorders. Some of the dSCZ-genes identified have pleiotropic effects on SCZ and other psychiatric disorders, including *TRANK1*, which is related to bipolar disorder (represented by a *cis*-eQTL of *TRANK1*, rs3732386, $F_{ST(EAS\text{-}EUR)} = 0.25$), and *FURIN*, which is related to autism spectrum disorder (represented by rs8032315 associated with autism spectrum disorder, $F_{ST(EAS\text{-}EUR)} = 0.20$).

Limitations of the study

Several factors may have limited our understanding of the differentiated genetic architecture of SCZ across population ancestries. The collected SCZ-SNPs were mainly reported in the EUR-based GWASs, and the transferability of these associations across ancestries remains unclear. These priori candidates are enriched with common variants, which could explain at most 24% of the genetic liability to SCZ in EUR populations.[19] However, the rare variants were not fully investigated in this study. Furthermore, it is challenging to link the genetic variations to the phenotypic diversity without any functional validation. We explicitly demonstrated evidence of positive selection and archaic hominin gene flow that occurred in the SCZ-SNPs, whereas the causal variants and their phenotypic contributions remain ambiguous, and the real scenario would be much more complicated than the cases we show here. Particularly for the complex genomic regions, such as the human MHC region and those located with structural variations, advanced sequencing technologies, and functional experiments may be essential for further explorations of their contributions to SCZ. Although efforts are made in this study to uncover the SCZ genetic diversity in multiple geographical populations from Eurasia, we acknowledge that the coverage of population genetic diversity in the 1000 Genomes dataset is still insufficient.[78] In particular, population stratification in East Asia is very complex.[79,80] There is thus a need for additional efforts incorporating diverse ethnic groups and fine maps of high-quality sequence data to determine the genetic mechanisms of SCZ in humans to further facilitate precision and personalized medicine.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.109560.

ACKNOWLEDGMENTS

We thank Dr. Kai Yuan for his help in detecting the archaic introgression in modern human genomes. This study is supported by the National Key Research and Development Program of China (no. 2023YFC2605400), the National Natural Science Foundation of China (NSFC) grants (32030020, 32288101, and 32270665), the Shanghai Science and Technology Commission Program (23JS1410100), the UK Royal Society-Newton Advanced Fellowship (NAF\R1\191094). L.D. also gratefully acknowledges the support of the SANOFI Scholarship Program. The computational work in this study was supported by the CFFF Computing Platform and the Human Phenome Data Center of Fudan University. The funders had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

S.X. conceived and designed the study. S.X. and L.D. supervised the project. D.T. and S.C. collected the data and analyzed the genetic differentiation, natural selection, and archaic introgression. S.C. analyzed the ancient human genomes. L.D., D.T., and S.C. drafted the manuscript. S.X. and L.D. revised the manuscript. All authors discussed the results and implications and commented on the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 22, 2023
Revised: December 29, 2023
Accepted: March 22, 2024
Published: March 26, 2024

REFERENCES

1. Marshall, C.R., Howrigan, D.P., Merico, D., Thiruvahindrapuram, B., Wu, W., Greer, D.S., Antaki, D., Shetty, A., Holmans, P.A., Pinto, D., et al. (2017). Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nat. Genet. *49*, 27–35.
2. Tsuang, M.T. (1975). Heterogeneity of schizophrenia. Biol. Psychiatry *10*, 465–474.
3. Garver, D.L. (1997). The etiologic heterogeneity of schizophrenia. Harv. Rev. Psychiatry *4*, 317–327.
4. Kremen, W.S., Seidman, L.J., Faraone, S.V., Toomey, R., and Tsuang, M.T. (2004). Heterogeneity of schizophrenia: a study of individual neuropsychological profiles. Schizophr. Res. *71*, 307–321.
5. Schizophrenia Psychiatric Genome-Wide Association Study GWAS Consortium (2011). Genome-wide association study identifies five new schizophrenia loci. Nat. Genet. *43*, 969–976.
6. Ripke, S., Neale, B.M., Corvin, A., Walters, J.T.R., Farh, K.-H., Holmans, P.A., Lee, P., Bulik-Sullivan, B., Collier, D.A., Huang, H., et al. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421–427.
7. Trubetskoy, V., Pardiñas, A.F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T.B., Bryois, J., Chen, C.Y., Dennison, C.A., Hall, L.S., et al. (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature *604*, 502–508.
8. Charlson, F.J., Ferrari, A.J., Santomauro, D.F., Diminic, S., Stockings, E., Scott, J.G., McGrath, J.J., and Whiteford, H.A. (2018). Global epidemiology and burden of schizophrenia: findings from the global burden of disease study 2016. Schizophr. Bull. *44*, 1195–1203.
9. Lam, M., Chen, C.Y., Li, Z., Martin, A.R., Bryois, J., Ma, X., Gaspar, H., Ikeda, M., Benyamin, B., Brown, B.C., et al. (2019). Comparative genetic architectures of schizophrenia in East Asian and European populations. Nat. Genet. *51*, 1670–1678.
10. Sekar, A., Bialas, A.R., de Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N., Tooley, K., Presumey, J., Baum, M., Van Doren, V., et al. (2016). Schizophrenia risk from complex variation of complement component 4. Nature *530*, 177–183.
11. Yue, W.-H., Wang, H.-F., Sun, L.-D., Tang, F.-L., Liu, Z.-H., Zhang, H.-X., Li, W.-Q., Zhang, Y.-L., Zhang, Y., Ma, C.-C., et al. (2011). Genome-wide association study identifies a susceptibility locus for

schizophrenia in Han Chinese at 11p11.2. Nat. Genet. *43*, 1228–1231.
12. Crow, T.J. (2000). Schizophrenia as the price that homo sapiens pays for language: a resolution of the central paradox in the origin of the species. Brain Res. Brain Res. Rev. *31*, 118–129.
13. Burns, J.K. (2006). Psychosis: A costly by-product of social brain evolution in Homo sapiens. Prog. Neuro-Psychopharmacol. Biol. Psychiatry *30*, 797–814.
14. Uher, R. (2009). The role of genetic variation in the causation of mental illness: an evolution-informed framework. Mol. Psychiatry *14*, 1072–1082.
15. Pearlson, G.D., and Folley, B.S. (2008). Schizophrenia, psychiatric genetics, and Darwinian psychiatry: an evolutionary framework. Schizophr. Bull. *34*, 722–733.
16. Crespi, B., Summers, K., and Dorus, S. (2007). Adaptive evolution of genes underlying schizophrenia. Proc. Biol. Sci. *274*, 2801–2810.
17. Xu, K., Schadt, E.E., Pollard, K.S., Roussos, P., and Dudley, J.T. (2015). Genomic and network patterns of schizophrenia genetic variation in human evolutionary accelerated regions. Mol. Biol. Evol. *32*, 1148–1160.
18. Srinivasan, S., Bettella, F., Mattingsdal, M., Wang, Y., Witoelar, A., Schork, A.J., Thompson, W.K., Zuber, V.; Schizophrenia Working Group of the Psychiatric Genomics Consortium The International Headache Genetics Consortium, and Winsvold, B.S., et al. (2016). Genetic markers of human evolution are enriched in schizophrenia. Biol. Psychiatry *80*, 284–292.
19. Pardiñas, A.F., Holmans, P., Pocklington, A.J., Escott-Price, V., Ripke, S., Carrera, N., Legge, S.E., Bishop, S., Cameron, D., Hamshere, M.L., et al. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat. Genet. *50*, 381–389.
20. Guo, J., Wu, Y., Zhu, Z., Zheng, Z., Trzaskowski, M., Zeng, J., Robinson, M.R., Visscher, P.M., and Yang, J. (2018). Global genetic differentiation of complex traits shaped by natural selection in humans. Nat. Commun. *9*, 1865.
21. Liu, J., Li, M., and Su, B. (2016). GWAS-identified schizophrenia risk SNPs at TSPAN18 are highly diverged between Europeans and East Asians. Am. J. Med. Genet. B Neuropsychiatr. Genet. *171*, 1032–1040.
22. Li, M., Wu, D.D., Yao, Y.G., Huo, Y.X., Liu, J.W., Su, B., Chasman, D.I., Chu, A.Y., Huang, T., Qi, L., et al. (2016). Recent positive selection drives the expansion of a schizophrenia risk nonsynonymous variant at SLC39A8 in Europeans. Schizophr. Bull. *42*, 178–190.
23. Fujito, N.T., Satta, Y., Hane, M., Matsui, A., Yashima, K., Kitajima, K., Sato, C., Takahata, N., and Hayakawa, T. (2018). Positive selection on schizophrenia-associated ST8SIA2 gene in post-glacial Asia. PLoS One *13*, e0200278.
24. Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P.H., de Filippo, C., et al. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. Nature *505*, 43–49.
25. McCoy, R.C., Wakefield, J., and Akey, J.M. (2017). Impacts of Neanderthal-introgressed sequences on the landscape of human gene expression. Cell *168*, 916–927.e12.
26. McArthur, E., Rinker, D.C., and Capra, J.A. (2021). Quantifying the contribution of Neanderthal introgression to the heritability of complex traits. Nat. Commun. *12*, 4481.
27. Yu, H., Yan, H., Wang, L., Li, J., Tan, L., Deng, W., Chen, Q., Yang, G., Zhang, F., Lu, T., et al. (2018). Five novel loci associated with antipsychotic treatment response in patients with schizophrenia: a genome-wide association study. Lancet Psychiatr. *5*, 327–338.
28. Lu, Z., Zhang, Y., Yan, H., Su, Y., Guo, L., Liao, Y., Lu, T., Yu, H., Wang, L., Li, J., et al. (2022). ATAD3B and SKIL polymorphisms associated with antipsychotic-induced QTc interval change in patients with schizophrenia: a genome-wide association study. Transl. Psychiatry *12*, 56.
29. Yu, H., Yan, H., Li, J., Li, Z., Zhang, X., Ma, Y., Mei, L., Liu, C., Cai, L., Wang, Q., et al. (2017). Common variants on 2p16.1, 6p22.1 and 10q24.32 are associated with schizophrenia in Han Chinese population. Mol. Psychiatry *22*, 954–960.
30. Bergen, S.E., O'Dushlaine, C.T., Ripke, S., Lee, P.H., Ruderfer, D.M., Akterin, S., Moran, J.L., Chambert, K.D., Handsaker, R.E., Backlund, L., et al. (2012). Genome-wide association study in a Swedish population yields support for greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. Mol. Psychiatry *17*, 880–886.
31. Stefansson, H., Ophoff, R.A., Steinberg, S., Andreassen, O.A., Cichon, S., Rujescu, D., Werge, T., Pietiläinen, O.P.H., Mors, O., Mortensen, P.B., et al. (2009). Common variants conferring risk of schizophrenia. Nature *460*, 744–747.
32. Ruderfer, D.M., Fanous, A.H., Ripke, S., McQuillin, A., Amdur, R.L., Schizophrenia Working Group of the Psychiatric Genomics Consortium; Bipolar Disorder Working Group of the Psychiatric Genomics Consortium; Cross-Disorder Working Group of the Psychiatric Genomics Consortium, Gejman, P.V., O'Donovan, M.C., et al. (2014). Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. Mol. Psychiatry *19*, 1017–1024.
33. Long, T., Lu, S., Li, H., Lin, R., Qin, Y., Li, L., Chen, L., Zhang, L., Lv, Y., Liang, D., et al. (2018). Association of APOB and LIPC polymorphisms with type 2 diabetes in Chinese Han population. Gene *672*, 150–155.
34. Sen, A., Madhivanan, K., Mukherjee, D., and Aguilar, R.C. (2012). The epsin protein family: coordinators of endocytosis and signaling. Biomol. Concepts *3*, 117–126.
35. Chen, H., Ko, G., Zatti, A., Di Giacomo, G., Liu, L., Raiteri, E., Perucco, E., Collesi, C., Min, W., Zeiss, C., et al. (2009). Embryonic arrest at midgestation and disruption of Notch signaling produced by the absence of both epsin 1 and epsin 2 in mice. Proc. Natl. Acad. Sci. USA *106*, 13838–13843.
36. Wei, J., and Hemmings, G.P. (2000). The NOTCH4 locus is associated with susceptibility to schizophrenia. Nat. Genet. *25*, 376–377.
37. Bryois, J., Calini, D., Macnair, W., Foo, L., Urich, E., Ortmann, W., Iglesias, V.A., Selvaraj, S., Nutma, E., Marzin, M., et al. (2022). Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. Nat. Neurosci. *25*, 1104–1112.
38. Qureshi, I.A., and Mehler, M.F. (2012). Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. Nat. Rev. Neurosci. *13*, 528–541.
39. Wang, S., Chen, H., Kong, W., Wu, X., Qian, Y., and Wei, K. (2023). A modified FGL sparse canonical correlation analysis for the identification of Alzheimer's disease biomarkers. Electronic Research Archive *31*, 882–903.
40. Yu, P., Chen, Y., Tagle, D.A., and Cai, T. (2002). PJA1, encoding a RING-H2 finger ubiquitin ligase, is a novel human X chromosome gene abundantly expressed in brain. Genomics *79*, 869–874.
41. Watabe, K., Kato, Y., Sakuma, M., Murata, M., Niida-Kawaguchi, M., Takemura, T., Hanagata, N., Tada, M., Kakita, A., and Shibata, N. (2020). Praja1 RING-finger E3 ubiquitin ligase suppresses neuronal cytoplasmic TDP-43 aggregate formation. Neuropathology *40*, 570–586.
42. Suzuki, T., Suzuki, T., Raveau, M., Miyake, N., Sudo, G., Tsurusaki, Y., Watanabe, T., Sugaya, Y., Tatsukawa, T., Mazaki, E., et al. (2020). A recurrent PJA1 variant in trigonocephaly and neurodevelopmental disorders. Ann. Clin. Transl. Neurol. *7*, 1117–1131.
43. Gulyás-Kovács, A., Keydar, I., Xia, E., Fromer, M., Hoffman, G., Ruderfer, D., Sachidanandam, R., and Chess, A. (2018). Unperturbed expression bias of imprinted genes in schizophrenia. Nat. Commun. *9*, 2914.
44. Hou, Y., Liang, W., Zhang, J., Li, Q., Ou, H., Wang, Z., Li, S., Huang, X., and Zhao, C. (2018). Schizophrenia-associated rs4702 G allele-specific downregulation of FURIN expression by miR-338-3p reduces BDNF production. Schizophr. Res. *199*, 176–180.
45. Warburton, A., Breen, G., Bubb, V.J., and Quinn, J.P. (2016). A GWAS SNP for schizophrenia is linked to the internal MIR137 promoter and supports differential allele-specific expression. Schizophr. Bull. *42*, 1003–1008.
46. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium (2017). Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. Mol. Autism. *8*, 21.
47. Wu, Y., Bi, R., Zeng, C., Ma, C., Sun, C., Li, J., Xiao, X., Li, M., Zhang, D.-F., Zheng, P., et al. (2019). Identification of the primate-specific gene BTN3A2 as an additional schizophrenia risk gene in the MHC loci. EBioMedicine *44*, 530–541.
48. Liu, Y., Ao, X., Zhou, X., Du, C., and Kuang, S. (2022). The regulation of PBXs and their emerging role in cancer. J. Cell Mol. Med. *26*, 1363–1379.
49. Yang, H.C., Chen, C.W., Lin, Y.T., and Chu, S.K. (2021). Genetic ancestry plays a central role in population pharmacogenomics. Commun. Biol. *4*, 171.
50. Lybaek, H., Robson, M., de Leeuw, N., Hehir-Kwa, J.Y., Jeffries, A., Haukanes, B.I., Berland, S., de Bruijn, D., Mundlos, S., Spielmann, M., and Houge, G. (2022). LRFN5 locus structure is associated with autism and influenced by the sex of the individual and locus conversions. Autism Res. *15*, 421–433.

51. Ayka, A., and Şehirli, A.Ö. (2020). The role of the SLC transporters protein in the neurodegenerative disorders. Clin. Psychopharmacol. Neurosci. *18*, 174–187.
52. Caputo, V., Ciolfi, A., Macri, S., and Pizzuti, A. (2015). The emerging role of MicroRNA in schizophrenia. CNS Neurol. Disord. Drug Targets *14*, 208–221.
53. Lencer, R., Mills, L.J., Alliey-Rodriguez, N., Shafee, R., Lee, A.M., Reilly, J.L., Sprenger, A., McDowell, J.E., McCarroll, S.A., Keshavan, M.S., et al. (2017). Genome-wide association studies of smooth pursuit and antisaccade eye movements in psychotic disorders: findings from the B-SNIP study. Transl. Psychiatry *7*, e1249.
54. Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z.X.P., Pool, J.E., Xu, X., Jiang, H., Vinckenbosch, N., Korneliussen, T.S., et al. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. Science *329*, 75–78.
55. Xu, W., Liu, Y., Chen, J., Guo, Q., Liu, K., Wen, Z., Zhou, Z., Song, Z., Zhou, J., He, L., et al. (2018). Genetic risk between the CACNA1I gene and schizophrenia in Chinese Uygur population. Hereditas *155*, 5.
56. Yan, X., Wang, Z., Schmidt, V., Gauert, A., Willnow, T.E., Heinig, M., and Poy, M.N. (2018). Cadm2 regulates body weight and energy homeostasis in mice. Mol. Metab. *8*, 180–188.
57. Dai, Z.M., Guo, W., Yu, D., Zhu, X.J., Sun, S., Huang, H., Jiang, M., Xie, B., Zhang, Z., and Qiu, M. (2022). SECISBP2L-Mediated Selenoprotein Synthesis Is Essential for Autonomous Regulation of Oligodendrocyte Differentiation. J. Neurosci. *42*, 5860–5869.
58. Frey, L., Ziętara, N., Łyszkiewicz, M., Marquardt, B., Mizoguchi, Y., Linder, M.I., Liu, Y., Giesert, F., Wurst, W., Dahlhoff, M., et al. (2021). Mammalian VPS45 orchestrates trafficking through the endosomal system. Blood *137*, 1932–1944.
59. Rahajeng, J., Caplan, S., and Naslavsky, N. (2010). Common and distinct roles for the binding partners Rabenosyn-5 and Vps45 in the regulation of endocytic trafficking in mammalian cells. Exp. Cell Res. *316*, 859–874.
60. Plooster, M., Brennwald, P., and Gupton, S.L. (2022). Endosomal trafficking in schizophrenia. Curr. Opin. Neurobiol. *74*, 102539.
61. Tempel, B.L., and Shilling, D.J. (2007). The plasma membrane calcium ATPase and disease. Subcell. Biochem. *45*, 365–383.
62. Farhang, S., Sabaie, H., Gharesouran, J., Asadi, M.R., Arsang-Jang, S., Ghafouri-Fard, S., Taheri, M., and Rezazadeh, M. (2022). Expression analysis of ermin and listerin E3 ubiquitin protein ligase 1 genes in the periphery of patients with schizophrenia. J. Mol. Neurosci. *72*, 246–254.
63. Chang, C.C., Kuo, H.Y., Chen, S.Y., Lin, W.T., Lu, K.M., Saito, T., and Liu, F.C. (2021). Developmental characterization of schizophrenia-associated gene ZSWIM6 in mouse forebrain. Front. Neuroanat. *15*, 669631.
64. Tischfield, D.J., Saraswat, D.K., Furash, A., Fowler, S.C., Fuccillo, M.V., and Anderson, S.A. (2017). Loss of the neurodevelopmental gene ZSWIM6 alters striatal morphology and motor regulation. Neurobiol. Dis. *103*, 174–183.
65. Yuan, K., Ni, X., Liu, C., Pan, Y., Deng, L., Zhang, R., Gao, Y., Ge, X., Liu, J., Ma, X., et al. (2021). Refining models of archaic admixture in Eurasia with ArchaicSeeker 2.0. Nat. Commun. *12*, 6232.
66. Lu, K., Yin, X., Weng, T., Xi, S., Li, L., Xing, G., Cheng, X., Yang, X., Zhang, L., and He, F. (2008). Targeting WW domains linker of HECT-type ubiquitin ligase Smurf1 for activation by CKIP-1. Nat. Cell Biol. *10*, 994–1002.
67. Liu, J., Liang, C., Guo, B., Wu, X., Li, D., Zhang, Z., Zheng, K., Dang, L., He, X., Lu, C., et al. (2017). Increased PLEKHO1 within osteoblasts suppresses Smad-dependent BMP signaling to inhibit bone formation during aging. Aging Cell *16*, 360–376.
68. Li, Z., Chen, J., Yu, H., He, L., Xu, Y., Zhang, D., Yi, Q., Li, C., Li, X., Shen, J., et al. (2017). Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. Nat. Genet. *49*, 1576–1583.
69. Ding, Q., Hu, Y., Xu, S., Wang, J., and Jin, L. (2014). Neanderthal introgression at chromosome 3p21.31 was under positive natural selection in East Asians. Mol. Biol. Evol. *31*, 683–695.
70. Cross-Disorder Group of the Psychiatric Genomics Consortium (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet *381*, 1371–1379.
71. Gibbons, J.M., Marno, K.M., Pike, R., Lee, W.Y.J., Jones, C.E., Ogunkolade, B.W., Pardieu, C., Bryan, A., Fu, R.M., Warnes, G., et al. (2020). HIV-1 accessory protein Vpr interacts with REAF/RPRD2 to mitigate its antiviral activity. J. Virol. *94*, e01591-19.
72. Liu, Y., Yang, Y., Luo, Y., Wang, J., Lu, X., Yang, Z., and Yang, J. (2020). Prognostic potential of PRPF3 in hepatocellular carcinoma. Aging (Albany NY) *12*, 912–930.
73. Davis, G.E., Jr., Davis, M.J., and Lowell, W.E. (2022). The effect of ultraviolet radiation on the incidence and severity of major mental illness using birth month, birth year, and sunspot data. Heliyon *8*, e09197.
74. Humble, M.B., Gustafsson, S., and Bejerot, S. (2010). Low serum levels of 25-hydroxyvitamin D (25-OHD) among psychiatric out-patients in Sweden: relations with season, age, ethnic origin and psychiatric diagnosis. J. Steroid Biochem. Mol. Biol. *121*, 467–470.
75. Valipour, G., Saneei, P., and Esmaillzadeh, A. (2014). Serum vitamin D levels in relation to schizophrenia: a systematic review and meta-analysis of observational studies. J. Clin. Endocrinol. Metab. *99*, 3863–3872.
76. Chiang, M., Natarajan, R., and Fan, X. (2016). Vitamin D in schizophrenia: a clinical review. Evid. Based. Ment. Health *19*, 6–9.
77. Subramaniam, M., Lam, M., Guo, M.E., He, V.Y.F., Lee, J., Verma, S., and Chong, S.A. (2014). Body mass index, obesity, and psychopathology in patients with schizophrenia. J. Clin. Psychopharmacol. *34*, 40–46.
78. Lu, D., and Xu, S. (2013). Principal component analysis reveals the 1000 Genomes Project does not sufficiently cover the human genetic diversity in Asia. Front. Genet. *4*, 127.
79. Gao, Y., Zhang, C., Yuan, L., Ling, Y., Wang, X., Liu, C., Pan, Y., Zhang, X., Ma, X., Wang, Y., et al. (2020). PGG.Han: the Han Chinese genome database and analysis platform. Nucleic Acids Res. *48*, D971–D976.
80. Xu, S., Yin, X., Li, S., Jin, W., Lou, H., Yang, L., Gong, X., Wang, H., Shen, Y., Pan, X., et al. (2009). Genomic dissection of population substructure of Han Chinese and its implication in association studies. Am. J. Hum. Genet. *85*, 762–774.
81. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS *16*, 284–287.
82. Pybus, M., Dall'Olio, G.M., Luisi, P., Uzkudun, M., Carreño-Torres, A., Pavlidis, P., Laayouni, H., Bertranpetit, J., and Engelken, J. (2014). 1000 Genomes Selection Browser 1.0: a genome browser dedicated to signatures of natural selection in modern humans. Nucleic Acids Res. *42*, D903–D909.
83. Hedman, A.C., Smith, J.M., and Sacks, D.B. (2015). The biology of IQGAP proteins: beyond the cytoskeleton. EMBO Rep. *16*, 427–446.
84. Honsa, P., Pivonkova, H., and Anderova, M. (2013). Focal cerebral ischemia induces the neurogenic potential of mouse Dach1-expressing cells in the dorsal part of the lateral ventricles. Neuroscience *240*, 39–53.
85. Costas, J., Suárez-Rama, J.J., Carrera, N., Paz, E., Páramo, M., Agra, S., Brenlla, J., Ramos-Ríos, R., and Arrojo, M. (2013). Role of DISC1 interacting proteins in schizophrenia risk from genome-wide analysis of missense SNPs. Ann. Hum. Genet. *77*, 504–512.
86. Huang, H.S., Matevossian, A., Whittle, C., Kim, S.Y., Schumacher, A., Baker, S.P., and Akbarian, S. (2007). Prefrontal dysfunction in schizophrenia involves mixed-lineage leukemia 1-regulated histone methylation at GABAergic gene promoters. J. Neurosci. *27*, 11254–11262.
87. Pagnamenta, A.T., Bacchelli, E., de Jonge, M.V., Mirza, G., Scerri, T.S., Minopoli, F., Chiocchetti, A., Ludwig, K.U., Hoffmann, P., Paracchini, S., et al. (2010). Characterization of a family with rare deletions in CNTNAP5 and DOCK4 suggests novel risk loci for autism and dyslexia. Biol. Psychiatry *68*, 320–328.
88. Bayat, A., Iqbal, S., Borredy, K., Amiel, J., Zweier, C., Barcia, G., Kraus, C., Weyhreter, H., Bassuk, A.G., Chopra, M., et al. (2021). PRICKLE2 revisited-further evidence implicating PRICKLE2 in neurodevelopmental disorders. Eur. J. Hum. Genet. *29*, 1235–1244.
89. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74.
90. Byrska-Bishop, M., Evani, U.S., Zhao, X., Basile, A.O., Abel, H.J., Regier, A.A., Corvelo, A., Clarke, W.E., Musunuri, R., Nagulapalli, K., et al. (2022). High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios. Cell *185*, 3426–3440.e19.
91. Bergström, A., McCarthy, S.A., Hui, R., Almarri, M.A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S., Hallast, P., Kamm, J., et al. (2020). Insights into human genetic variation and population history from 929 diverse genomes. Science *367*, eaay5012.
92. Reich, D., Green, R.E., Kircher, M., Krause, J., Patterson, N., Durand, E.Y., Viola, B., Briggs, A.W., Stenzel, U., Johnson, P.L.F., et al. (2010). Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature *468*, 1053–1060.

93. Mallick, S., Micco, A., Mah, M., Ringbauer, H., Lazaridis, I., Olalde, I., Patterson, N., and Reich, D. (2024). The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human genomes. Sci. Data *11*, 182.
94. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al. (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585.
95. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.
96. Szpiech, Z.A., and Hernandez, R.D. (2014). selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. Mol. Biol. Evol. *31*, 2824–2827.
97. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience *4*, 7.
98. Smith, J., Coop, G., Stephens, M., and Novembre, J. (2018). Estimating time to the common ancestor for a beneficial allele. Mol. Biol. Evol. *35*, 1003–1017.
99. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. *47*, 1228–1235.
100. Reshef, Y.A., Finucane, H.K., Kelley, D.R., Gusev, A., Kotliar, D., Ulirsch, J.C., Hormozdiari, F., Nasser, J., O'Connor, L., van de Geijn, B., et al. (2018). Detecting genome-wide directional effects of transcription factor binding on polygenic disease risk. Nat. Genet. *50*, 1483–1493.
101. Weir, B.S., and Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. Evolution *38*, 1358–1370.
102. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl variant effect predictor. Genome Biol. *17*, 122.
103. Castel, S.E., Mohammadi, P., Chung, W.K., Shen, Y., and Lappalainen, T. (2016). Rare variant phasing and haplotypic expression from RNA sequencing with phASER. Nat. Commun. *7*, 12817.
104. Ning, Z., Tan, X., Yuan, Y., Huang, K., Pan, Y., Tian, L., Lu, Y., Wang, X., Qi, R., Lu, D., et al. (2023). Expression profiles of east-west highly differentiated genes in Uyghur genomes. Natl. Sci. Rev. *10*, nwad077.
105. Voight, B.F., Kudaravalli, S., Wen, X., and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. PLoS Biol. *4*, e72.
106. Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. Nature *449*, 913–918.
107. Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics *123*, 585–595.
108. Fu, Y.X., and Li, W.H. (1993). Statistical tests of neutrality of mutations. Genetics *133*, 693–709.
109. Schaschl, H., Göllner, T., and Morris, D.L. (2022). Positive selection acts on regulatory genetic variants in populations of European ancestry that affect ALDH2 gene expression. Sci. Rep. *12*, 4563.
110. Bandelt, H.J., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. *16*, 37–48.
111. Polzin, T., and Daneshmand, S.V. (2003). On Steiner trees and minimum spanning trees in hypergraphs. Oper. Res. Lett. *31*, 12–20.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
NHGRI-EBI GWAS Catalog (As of 2023-02-08)	NHGRI-EBI	https://www.ebi.ac.uk/gwas/efotraits/MONDO_0005090
1000 Genomes Project Phase III (low-coverage)	EMBL-EBI[89]	https://ftp.ncbi.nlm.nih.gov/1000genomes/ftp/release/20130502/
1000 Genomes Project Phase III (high-coverage)	EMBL-EBI[90]	http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/working/20190425_NYGC_GATK/
Human Genome Diversity Project	Bergström et al.[91]	ftp://ngs.sanger.ac.uk/production/hgdp/hgdp_wgs.20190516/
Altai Neanderthal genome	Prüfer et al.[24]	https://www.ebi.ac.uk/ena/browser/view/ERP002097
Denisovan genome	Reich et al.[92]	https://www.ebi.ac.uk/ena/browser/view/ERP000318
Ancient DNA data (As of 2022-11-21)	Mallick et al.[93]	https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data
Genotype-tissue expression data	The GTEx Consortium[94]	https://gtexprotal.org
1000 Genomes Selection Browser 1.0	Pybus et al.[82]	https://hsb.upf.edu
Software and algorithms		
VCFtools	Danecek et al.[95]	https://vcftools.sourceforge.net/
ClusterProfiler v4.2.2	Yu et al.[81]	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
Python 3.6.10	Python Software Foundation	https://python.org/
Selscan v1.2.0	Szpiech et al.[96]	https://github.com/lasugden/selscan
PLINK v1.9	Chang et al., 2015[97]	https://www.cog-genomics.org/plink/1.9/
Startmrca	Smith et al.[98]	https://github.com/jhavsmith/startmrca
ArchaicSeeker 2.0	Yuan et al.[65]	https://github.com/Shuhua-Group/ArchaicSeeker2.0
LDSC (LD Score) v1.0.1	Finucane et al.[99]	http://www.github.com/bulik/ldsc
SLDP (Signed LD profile) regression	Reshef et al.[100]	http://www.github.com/yakirr/sldp
NETWORK v10	Fluxus Technology Ltd	https://www.fluxus-engineering.com/

RESOURCE AVAILABILITY

Lead contact

Further information about this manuscript and requests for resources will be fulfilled by the lead contact, Shuhua Xu (xushua@fudan.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data: All data analyzed in this study are publicly available with accession numbers provided in the key resources table.
- Code: This study does not report any original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

This study did not use any experimental model or subject.

METHOD DETAILS

Genome data collection of present-day and ancient samples

We downloaded the genome data of 1,007 unrelated individuals from five EAS populations and five EUR populations released by the 1000 Genomes Project Phase III (https://ftp.ncbi.nlm.nih.gov/1000genomes/ftp/release/20130502/).[89] The EAS populations included CDX (Chinese Dai in Xishuangbanna, China; n = 93), CHB (Han Chinese in Beijing, China; n = 103), CHS (Southern Han Chinese, China; n = 105), JPT (Japanese in Tokyo, Japan; n = 104), and KHV (Kinh in Ho Chi Minh City, Vietnam; n = 99); the EUR populations included CEU (Utah residents with northern and western European ancestry; n = 99), GBR (British in England and Scotland; n = 91), FIN (Finnish in Finland; n = 99), TSI (Toscani in Italy; n = 107), and IBS (Iberian population in Spain; n = 107). The African YRI (Yoruba in Ibadan, Nigeria; n = 108) population was used as an outgroup. We filtered the multiallelic sites and used 77,818,183 biallelic sites in subsequent analyses.

We also obtained two high-coverage genome datasets for validation analyses, including the updated 1000 Genomes data with a mean depth of 30× (http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/data_collections/1000G_2504_high_coverage/working/20190425_NYGC_GATK/),[90] and the HGDP data (ftp://ngs.sanger.ac.uk/production/hgdp/hgdp_wgs.20190516/).[91] The HGDP data analyzed included 241 EAS individuals from China, Japan, Cambodia, and Siberia, and 161 EUR individuals from France, Russia, Italy, and Orkney islands. Totally 91,796,551 and 63,540,915 biallelic variants in these two datasets, respectively, were used in subsequent analyses.

To investigate the allele dynamics in human history, we obtained the high-coverage genomes of an Altai Neanderthal[24] and a Denisovan,[92] and 8,872 worldwide Neolithic-to-Paleolithic ancient human genomes (Allen Ancient DNA Resource, https://reich.hms.harvard.edu/allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data,[93] last accessed November 21, 2022, version 54.1).[93] Samples from each continent were grouped according to the date to ensure a sufficient sample size (>40 for each group) for allele frequency estimation. The ancestral alleles were determined according to the GRCh37 e71 ancestral sequence. We used the reference alleles to represent the ancestral alleles for single nucleotide polymorphisms (SNPs) without determining ancestral or derived allele states.

A compiled list of the SCZ-associated genes

We obtained 3,987 SCZ-associated SNPs (SCZ-SNPs) on the autosomes under the trait term "schizophrenia" from the GWAS catalog (https://www.ebi.ac.uk/gwas/efotraits/MONDO_0005090), including the "background traits data" and the "child trait data". We found that 3,841 of the 3,987 SNPs were genotyped in the 1000 Genomes dataset used in the primary analyses.[7,9] According to the Ensembl 96 genome annotation (GRCh37), the SCZ-SNPs are located in 1,592 genes, which were thus considered SCZ-associated genes (SCZ-genes). A full description of the SCZ genes and SNPs can be found in Figure S1. In the high-coverage 1000 Genomes data and the HGDP data, 3,634 SCZ-SNPs located in 1,278 genes and 3,770 SCZ-SNPs located in 1,235 genes were detected, respectively. In this study, we analyzed all of the SNPs located in the SCZ-genes (collectively known as SCZ-related SNPs), and focused on 2,401 SNPs with determined phenotypic effects (risk or non-risk for SCZ) in previous association studies (Table S11). In addition, we extracted a subset of SCZ-SNPs, including 442 SCZ-SNPs reported in two latest GWASs of SCZ conducted on large samples,[7,9] to validate the genetic differentiation of SCZ between population ancestries.

We also generated sets of random SNPs (with a comparable number of random SNPs to that of the SCZ-SNPs in each set) using two strategies (100 SNP sets by each strategy) to evaluate the genetic diversity of the SCZ-genes in a genome-wide context. For the first "relaxed" strategy, we randomly selected SNPs from the whole genome, and these sets of SNPs were designated as relaxed control SNPs (rc-SNPs). For the second "strict" strategy, we first selected random genes with matching sizes and biological types of the SCZ-genes, and then conducted variant sampling from these strict control genes (sc-genes) according to the SCZ-SNP density distribution in the SCZ-genes. These SNPs were thus designated as strict control SNPs (sc-SNPs). The two sampling strategies are illustrated in Figure S1. In addition, we randomly sampled a set of SNPs that matched the SCZ-SNPs on the minor allele frequency (MAF) spectrum in EUR and EAS populations, with a frequency bin size of 0.05.

Principal component analysis (PCA) and estimation of the genetic differentiation

We carried out approximate pruning for possible linkage disequilibrium (LD) of the SCZ-related SNPs by selecting an SNP in each of the 200-kb non-overlapping windows, and PCA was performed on the remaining 1,899 SNPs using PLINK version 1.9[97] with all of the options set to the default. As a measure of genetic differentiation on the individual level, we calculated the nucleotide difference for pairwise individuals between or within continents; on the population level, we measured the overall genetic divergence for pairwise populations by calculating the SNP-based F_{ST}[101] for pairwise populations and the combined continental groups using VCFtools.[95] All of these analyses were carried out independently for the target and control sets of SNPs.

Identification and annotation of the highly differentiated SCZ-related SNPs between EAS and EUR populations

For each SNP in the autosomes, we calculated the F_{ST}[101] for all of the Eurasian population pairs, and determined a candidate SNP if the maximum F_{ST} between EAS and EUR populations (max-$F_{ST(EAS-EUR)}$) at this locus reached the top 5% significance level (empirical $p < 5\%$) of the autosomal SNPs. Based on these candidate SNPs, we further screened for the SNPs that were genetically similar within each continent by calculating the largest F_{ST} value within each continent (denoted as max-$F_{ST(EAS)}$ and max-$F_{ST(EUR)}$ for EAS and EUR, respectively) at each locus. We selected candidate SNPs with lower max-$F_{ST(EAS)}$ and max-$F_{ST(EUR)}$ compared to the genome-wide average in EAS and EUR populations, respectively. In addition, SNPs with a lower max-$F_{ST(EAS-EUR)}$ (lower than the genome-wide average max-$F_{ST(EAS-EUR)}$) were considered to be less differentiated between EAS and EUR populations. Functional consequences and pathogenicity of the candidate SNPs were further

assessed using Variant Effect Predictor (VEP).[102] Combined Annotation Dependent Depletion (CADD) could score the deleteriousness of SNPs and insertion or deletion variants. Genomic Evolutionary Rate Profiling (GERP) identifies constrained elements in multiple alignments by quantifying substitution deficits. Gene functional enrichment was carried out using ClusterProfiler version 4.2.2.[81]

We independently analyzed the X chromosome. We calculated the allele frequency of the X-linked SNPs using PLINK version 1.9[97] given the gender information, and determined the candidate highly differentiated X-linked SNPs between EAS and EUR populations using the top 5% cutoff.

Gene expression analysis

We explored possible regulatory effects of the highly differentiated SCZ-related SNPs using the gene expression data in brain tissues released by the Genotype-Tissue Expression (GTEx, https://gtexprotal.org) Project. In addition, we analyzed the allele-specific expression (ASE) at the SCZ-genes. The SNP-based ASE data for the EAS and EUR populations, represented by 39 Han Chinese samples and 94 GBR samples, respectively, were generated using phASER v3.22.0[103] in a previous study.[104] One ASE locus was defined as a false discovery rate (FDR) < 5% in the binomial test and allelic imbalance $\geq$ 0.2. Homozygous sites and heterozygous sites with <10 reads were excluded. The allelic imbalance was calculated as follows:

$$\left| \frac{ref}{ref+alt} - 0.5 \right|.$$

where *ref* and *alt* are the numbers of mappable reads of the reference and alternative alleles, respectively.

Testing for natural selection

To detect recent or ongoing positive selection, we calculated the integrated haplotype score (iHS)[105] for the polymorphic sites in each of the EAS and EUR populations, and estimated the cross-population extended haplotype homozygosity (XP-EHH)[106] for EUR-EAS population pairs. Both analyses were performed using Selscan version 1.2.0,[96] and the outputs were normalized across 100 frequency bins. Finally, we determined the candidate adaptive SNPs using a criterion of |iHS| > 2 or |XP-EHH| > 2. To determine the candidate genome regions of a selective sweep, we calculated the proportions of candidate adaptive SNPs for a sliding window (50 kb in length, stepping 25 kb, $\geq$20 SNPs) across the genome. Candidate regions of positive selection with empirical $p < 0.005$ were selected from each group.

To examine whether the SCZ-SNPs are more prevalent in candidate regions of positive selection than those associated with other traits, we downloaded SNPs associated with Alzheimer's disease, bipolar disorder, body mass index, and cardiovascular disease from the GWAS Catalog, and tested for the enrichment of trait-associated SNPs in the candidate regions of positive selection using one-sided Fisher's exact test. We also tested for the enrichment of candidate regions of positive selection at the functional or non-functional sites. The functional variants included missense variants and conserved variants with CADD >15 or GERP >2 annotated using VEP, and eQTLs reported by the GTEx Project. The p-values were also obtained by the one-sided Fisher's exact test.

We also calculated estimators of nucleotide diversity ($\theta\pi$), number of segregating sites (θ_K), and Tajima's D[107] to test for neutrality with a sliding window of 50 kb along the chromosome, stepping by 10 kb. The empirical p-value was estimated by ranking the statistical values of each gene among all of the regions across the genome. For some local genomic regions of interest, we referred to the 1000 Genomes Selection Browser[82] (https://hsb.upf.edu) for signatures of natural selection by comprehensive statistics, including Tajima's D[107] and Fu and Li's F*.[108]

Estimating the selection time

We used startmrca[98] to estimate the onset of selection of a beneficial allele by investigating its 1-Mb surrounding region. A mutation rate of 1.6×10^{-8} per site per generation was used in accordance with a previous study.[109] We ran 15,000 iterations and retained the last 6,000 iterations as the final results.

Quantifying the effects of Altai Neanderthal introgression on the SCZ-genes

The introgressed segments and alleles from Altai Neanderthal in the EAS and EUR genomes inferred using ArchaicSeeker 2.0 were reported by Yuan et al.[65] To quantify the proportion of SCZ-associated loci potentially influenced by the Altai Neanderthal introgression, we obtained 2,579 LD-pruned blocks with a between-SNP-distance <50 kb for any adjacent SCZ-SNPs in one block. If a block contained an SCZ-SNP with likely archaic origin, it was determined to be influenced by the archaic introgression. To test whether the introgressed variants were enriched for the heritability of SCZ, we applied the Stratified-LD Score Regression v1.0.1[99] to the summary statistics of the association test in the EAS and EUR populations, respectively[9] (downloaded from https://www.med.unc.edu/pgc/), following the procedures and parameters used by McArthur et al.[26] We then applied signed LD profile regression[100] to quantify the directional effect of archaic introgression on SCZ. The signed annotations were generated by calculating the maximum LD (measured by r^2) between the Altai Neanderthal introgressed variants and others in the 1000 Genomes Phase III dataset.

Constructing the haplotype networks

We constructed and visualized the haplotype networks using NETWORK v10 (https://www.fluxus-engineering.com/). For the archaic introgressed regions, the haplotype networks were constructed based on the archaic introgressed variants and the linked variants ($r^2 > 0.8$). We first calculated the median-joining network,[110] and then used the maximum-parsimony algorithm[111] to identify and remove the unnecessary median vectors and links.

Estimating the time to the most recent common ancestor (TMRCA)

We estimated the TMRCA of the archaic-like haplotypes in the introgressed regions of populations. We conducted 100 replicated runs, each conducted on 80% of the sequences, to obtain the confidence interval (CI) of the estimates. For a given segment from *a* to *b*, we first calculated the average pairwise nucleotide differences of sequences ($\overline{\pi}$) in the region as follows:

$$\overline{\pi} = \frac{2 \times \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \pi_{ij}}{(n+1) \times n}$$

Where n is the number of sequences analyzed in this region, and π_{ij} is the nucleotide difference between the i^{th} sequence and j^{th} sequence. We then estimated the local mutation rate (μ_{ab}) in this region as:

$$\mu_{ab} = \frac{d_{Hum-AncHumChimp}}{l_{ab} \times T_{Hum-HumChimp}}$$

where the l_{ab} is the segment length, $d_{Hum-AncHumChimp}$ is the nucleotide difference between the human reference genome and that of the most recent common ancestor of humans and chimpanzees in this region, and $T_{Hum-HumChimp}$ is the divergence time between humans and chimpanzees, which is assumed to be 13 million years in this study. The TMRCA can be estimated as:

$$TMRCA = \frac{\overline{\pi}}{2 \times \mu_{ab} \times l_{ab}}$$