Comparative genetic architectures of schizophrenia in East Asian and European populations

Max Lam[1,2,3,4,5,6,72], Chia-Yen Chen[4,5,7,8,9,72], Zhiqiang Li[10,11], Alicia R. Martin[4,5,7], Julien Bryois[12], Xixian Ma[13], Helena Gaspar[14], Masashi Ikeda[15], Beben Benyamin[16,17,18], Brielin C. Brown[19,20], Ruize Liu[4,5], Wei Zhou[11,21], Lili Guan[22,23,24], Yoichiro Kamatani[25,26], Sung-Wan Kim[27], Michiaki Kubo[28], Agung A. A. A. Kusumawardhani[29], Chih-Min Liu[30,31], Hong Ma[22,23,24], Sathish Periyasamy[32,33], Atsushi Takahashi[26,34], Zhida Xu[35], Hao Yu[22,23,24], Feng Zhu[36,37,38], Schizophrenia Working Group of the Psychiatric Genomics Consortium[39], Indonesia Schizophrenia Consortium[39], Genetic REsearch on schizophreniA neTwork-China and the Netherlands (GREAT-CN)[39], Wei J. Chen[30,31,40], Stephen Faraone[41], Stephen J. Glatt[42], Lin He[11,43,44], Steven E. Hyman[5,45], Hai-Gwo Hwu[30,31,40], Steven A. McCarroll[5,46], Benjamin M. Neale[4,5,7], Pamela Sklar[47], Dieter B. Wildenauer[48], Xin Yu[22,23,24], Dai Zhang[22,23,24], Bryan J. Mowry[32,33], Jimmy Lee[49], Peter Holmans[50], Shuhua Xu[13,51,52,53], Patrick F. Sullivan[54], Stephan Ripke[4,5,55], Michael C. O'Donovan[50], Mark J. Daly[4,5,7,56], Shengying Qin[11,57,71], Pak Sham[58,59,71], Nakao Iwata[15,71], Kyung S. Hong[60,71], Sibylle G. Schwab[61,62,71], Weihua Yue[22,23,24,63,71*], Ming Tsuang[64,71*], Jianjun Liu[2,65,71*], Xiancang Ma[37,38,66,71*], René S. Kahn[67,68,69,71*], Yongyong Shi[1,10,11,70,71*] and Hailiang Huang[4,5,7,71*]

Schizophrenia is a debilitating psychiatric disorder with approximately 1% lifetime risk globally. Large-scale schizophrenia genetic studies have reported primarily on European ancestry samples, potentially missing important biological insights. Here, we report the largest study to date of East Asian participants (22,778 schizophrenia cases and 35,362 controls), identifying 21 genome-wide-significant associations in 19 genetic loci. Common genetic variants that confer risk for schizophrenia have highly similar effects between East Asian and European ancestries (genetic correlation = 0.98 ± 0.03), indicating that the genetic basis of schizophrenia and its biology are broadly shared across populations. A fixed-effect meta-analysis including individuals from East Asian and European ancestries identified 208 significant associations in 176 genetic loci (53 novel). Trans-ancestry fine-mapping reduced the sets of candidate causal variants in 44 loci. Polygenic risk scores had reduced performance when transferred across ancestries, highlighting the importance of including sufficient samples of major ancestral groups to ensure their generalizability across populations.

Schizophrenia is an often-disabling psychiatric disorder that occurs worldwide with a lifetime risk of about 1% (ref. [1]). It is well established that genetic factors contribute to the susceptibility of schizophrenia. Recently, 145 genetic loci were associated with schizophrenia in samples of primarily European ancestry[2,3] (EUR), but this still represents the tip of the iceberg with respect to common variant liability to the disorder: the highly polygenic nature of common variation underlying this disorder predicts that there are hundreds more loci to be discovered[4].

Most genetic studies of schizophrenia have been performed in EUR samples, with relatively few studies in other populations[5–8]. This is a substantial deficiency for multiple reasons, particularly as it greatly limits the discovery of biological clues about schizophrenia. For some causal variants, ancestry-related heterogeneity yields varying allele frequency and linkage disequilibrium (LD) patterns such that associations that can be detected in one population may not be readily detected in others. Examples include a nonsense variant in *TBC1D4*, which confers muscle insulin resistance and increases risk for type 2 diabetes, which is common in Greenland but rare or absent in other populations[9], several Asian-specific coding variants that influence blood lipids[10], a variant highly protective against alcoholism that is common in Asian populations but uncommon elsewhere[11], and two loci associated with major depression[12] that are more common in Chinese populations than in EUR samples[12,13] (rs12415800: 45 versus 2%; rs35936514: 28 versus 6%).

Even if alleles have similar frequencies across populations, the effects of alleles on risk might be specific to certain populations if there are prominent but local contributions of clinical heterogeneity or gene–environment or gene–gene interactions. In addition, there have been debates about differences in prevalence, symptomatology, etiology, outcome and course of illness across geographical regions[14–19]. Understanding the genetic architecture of schizophrenia across populations provides insights into whether any differences represent etiologic heterogeneity on the illness.

Finally, polygenic risk score (PRS) prediction is emerging as a useful tool for studying the effects of genetic liability, identifying

A full list of affiliations appears at the end of the paper.

more homogeneous phenotypes, and stratifying patients. However, previous studies have shown that prediction accuracy decays with increasing genetic divergence between the risk allele discovery and target datasets[20,21]. The risk predicted, measured as R^2 (coefficient of determination), was only 45% as accurate in East Asians (EAS) compared with in EUR individuals when computed from genome-wide association studies (GWASs) of Europeans[22]. These differences can be explained by ancestry-related differences in allele frequencies, LD and other factors[22]. Importantly, the applicability of training data from EUR studies to those of non-EUR ancestry has not been fully assessed, leaving uncertainty as to the biological relevance of discoveries made in EUR samples for non-Europeans[21].

Results

Schizophrenia genetic associations in East Asian (EAS) populations. This study combined multiple samples from individuals with schizophrenia across EAS to systematically examine the genetic architecture of schizophrenia in individuals of EAS ancestry. We compiled 22,778 schizophrenia cases and 35,362 controls from 20 sample collections from East Asia (Supplementary Table 1). Individual-level genotypes were available from 16 sample collections (Supplementary Table 1), on which we performed quality control, imputation and association tests (Methods and Supplementary Table 2). Two sample collections (TAI-1 and TAI-2) were trio-based, and pseudo-controls were used. Four sample collections made available summary statistics for 22,000–31,000 selected variants (Methods) that had been analyzed in published studies[7,8]. Compared with the latest study using only Chinese individuals[8], our study had about twice the sample size, and was much more diverse.

We used a two-stage study design (Supplementary Table 1a). Stage 1 included 13 sample collections for which we had individual genotype data (13,305 cases and 16,244 controls after quality control). Stage 2 incorporated the remaining seven sample collections: full genotype data from three sample collections that arrived after the stage 1 data freeze, and summary statistics (for selected variants) from four sample collections (Supplementary Table 1). Meta-analyses across stage 1 samples and across all EAS samples were conducted using a fixed-effect model with inverse-variance weighting. QQ plots (Extended Data Fig. 1) showed no inflation of test statistics (indicating that ancestry effects had been well controlled), with the genomic inflation factor (λ_{gc}) = 1.14, the genomic inflation factor scaled for an equivalent study of 1,000 cases and 1,000 controls ($\lambda_{1,000}$) = 1.01 and an LD score regression[23] (LDSC) intercept of 1.0145 ± 0.011 using stage 1 samples.

Combining stages 1 and 2, we found 21 genome-wide-significant associations at 19 loci (Table 1, Fig. 1, Supplementary Table 3 and Supplementary Datasets 1 and 2)—an additional 14 associations compared with the most recent schizophrenia genetic study of Chinese ancestry[8]. Most associations were characterized by marked differences in allele frequencies between the EAS and EUR samples: for 15 of 21 loci, the index variants had higher minor allele frequencies (MAFs) in EAS than EUR. The higher allele frequency potentially confers better power to detect associations in EAS. For example, we identified a locus (Supplementary Dataset 1) with the top association (rs374528934) having strong evidence in EAS ($P = 5 \times 10^{-11}$) but not in EUR using the stage 1 samples. rs374528934 has a MAF of 45% in EAS but only 0.7% in EUR. No other variant in this locus is significantly associated with schizophrenia in EUR. This locus contains *CACNA2D2* (encoding the calcium channel α2δ-2 subunit associated with childhood epilepsy[24,25], and to which the anticonvulsant medication gabapentin binds), suggesting a path for further therapeutic investigation[25]. This finding also adds new evidence to the calcium signaling pathway suggested to be implicated in psychiatric disorders[26,27].

Genetic effects are consistent across populations. For causal variants, heterogeneity of genetic effects across populations could arise from clinical heterogeneity, differences in pathophysiology, environmental differences that change the genetic effects (gene–environment interaction) or interaction with other genetic factors that may differ in frequency across populations (gene–gene interaction). Heterogeneity in estimating genetic effect sizes may also be a consequence of differential correlation across genetic markers in a region, when investigating variants that are tagging the causal variant but do not exert any influence on the trait in question. Such heterogeneity does not reflect biological differences, but is rather statistical in nature. While it is assumed that biological pathways underlying complex human disorders are generally consistent across populations, genetic heterogeneity has been observed in other genetically complex disorders[28]. The large EAS sample allowed us to systematically explore the heterogeneity of genetic effects influencing liability to schizophrenia across two major world populations.

Using LDSC[23] and common variants (MAF > 5%) outside of the major histocompatibility complex (MHC) region, we found that the single-nucleotide polymorphism (SNP) heritability of schizophrenia is very similar in EAS (0.23 ± 0.03) and EUR (0.24 ± 0.02) (Methods and Extended Data Fig. 2a). Using the same set of variants, we found that the genetic correlation (r_g) for schizophrenia between EAS and EUR was indistinguishable from 1 ($r_g = 0.98 \pm 0.03$) (using POPCORN[29]—a method designed for cross-ancestry comparisons). This finding indicates that the common variant genetic architecture of schizophrenia outside of the MHC region is highly consistent across EAS and EUR samples.

Genetic correlations between schizophrenia and 11 other psychiatric disorders and behavior traits also showed no significant differences when estimated within EUR and across EAS–EUR (Extended Data Fig. 2b). In agreement with recent reports[30–33], we observed significant positive genetic correlations for schizophrenia with bipolar disorder, major depressive disorder, anorexia nervosa, neuroticism, autism spectrum disorder and educational attainment. We observed significant negative correlations with general intelligence, fluid intelligence score, prospective memory and subjective well-being.

We used partitioned LDSC[23] to look for heritability enrichment in diverse functional genomic annotations defined and used in previous publications[34,35] (Methods and Extended Data Fig. 2c,d). Using EAS stage 1 samples, we observed significant enrichment (after Bonferroni correction) in regions conserved across 29 mammals (as described in Lindblad-Toh et al.[36]; 'Conserved LindbladToh'). No other annotations were significantly enriched, and there were no significant differences between EUR-only and EAS-only enrichments ($P = 0.16$, two-sided paired t-test).

We identified gene sets that are enriched for schizophrenia genetic associations using MAGMA[37] and gene set definitions from a recent schizophrenia exome sequencing study[38] (Methods). Despite large differences in sample size and genetic background, the gene sets implicated in EAS and EUR samples were highly consistent: we observed no significant differences between gene set ranks using the EAS samples and gene set ranks using the EUR samples ($P = 0.72$, two-sided Wilcoxon test). In addition, nine of the top ten gene sets identified using the EAS samples were also among the top ten gene sets identified using the EUR samples (Extended Data Fig. 3).

A study of EUR individuals suggested that common schizophrenia alleles are under strong background selection[3]. We performed two analyses and found that the natural selection signatures, including positive and background selections, are consistent in schizophrenia-associated loci across EAS and EUR populations. First, we compared the signatures in the top 100 associated loci in EAS with those in EUR. Among the selection signatures we calculated (Methods), none showed a significant difference across populations (Extended Data Fig. 4a; $P > 0.05$ for all panels, two-sided

Table 1 | Genome-wide-significant loci in the EAS populations

				Stage 1		Stage 2		Combined	
SNP	**Chromosome**	**BP**	**AL**	***P***	**OR**	***P***	**OR**	***P***	**OR**
rs4660761	1	44440146	A/G	3.6×10^{-6}	0.91	3.53×10^{-4}	0.92	5.08×10^{-9}	0.91
rs848293	2	58382490	A/G	3.7×10^{-10}	0.90	3.10×10^{-9}	0.87	9.87×10^{-18}	0.89
rs17592552	2	201176071	T/C	8.4×10^{-10}	0.86	2.68×10^{-5}	0.89	1.50×10^{-13}	0.88
rs2073499	3	50374293	A/G	1.1×10^{-9}	0.89	2.14×10^{-5}	0.91	1.33×10^{-13}	0.90
rs76442143	3	51043599	T/C	6.9×10^{-9}	1.14	1.03×10^{-2}	1.08	6.40×10^{-10}	1.12
rs10935182	3	136137422	A/G	1.3×10^{-6}	0.90	1.33×10^{-4}	0.90	7.08×10^{-10}	0.90
rs4856763	3	161831675	A/G	3.9×10^{-6}	0.92	8.54×10^{-6}	0.91	1.73×10^{-10}	0.92
rs13096176	3	180752138	T/C	3.1×10^{-7}	0.88	2.21×10^{-3}	0.90	3.35×10^{-9}	0.89
rs6832165	4	24270210	C/G	3.7×10^{-8}	1.12	3.70×10^{-1}	1.08	2.79×10^{-8}	1.12
rs13142920	4	176728614	A/C	9.5×10^{-5}	0.93	5.85×10^{-6}	0.89	4.85×10^{-9}	0.92
rs4479913	6	165075210	A/G	3.6×10^{-7}	1.13	9.98×10^{-5}	1.12	1.53×10^{-10}	1.12
rs320696	7	137047137	A/C	5.5×10^{-8}	0.90	1.07×10^{-2}	0.93	2.81×10^{-9}	0.91
rs11986274	8	38259481	T/C	5.1×10^{-4}	1.07	2.73×10^{-6}	1.11	1.44×10^{-8}	1.08
rs2612614	8	65310836	A/G	2.2×10^{-8}	1.14	4.51×10^{-2}	1.06	1.62×10^{-8}	1.11
rs4147157	10	104536360	A/G	6.6×10^{-10}	0.90	3.87×10^{-7}	0.89	1.32×10^{-15}	0.89
rs10861879	12	108609634	A/G	4.8×10^{-7}	1.09	5.00×10^{-3}	1.07	1.18×10^{-8}	1.08
rs1984658	12	123483426	A/G	5.1×10^{-11}	0.89	2.14×10^{-4}	0.92	8.62×10^{-14}	0.90
rs9567393	13	32763757	A/G	3.5×10^{-8}	1.11	4.37×10^{-3}	1.07	1.13×10^{-9}	1.09
rs9890128	17	1273646	T/C	3.5×10^{-8}	0.90	2.44×10^{-2}	0.91	2.61×10^{-9}	0.90
rs11665111	18	77622996	T/C	5.2×10^{-6}	1.08	6.89×10^{-4}	1.09	1.46×10^{-8}	1.09
rs55642704	18	77688124	T/C	1.1×10^{-6}	1.09	7.11×10^{-6}	1.10	3.76×10^{-11}	1.09

For EAS stage 1, $n=13{,}305$ cases and $n=16{,}244$ controls. For EAS stages 1 and 2, $n=22{,}778$ cases and $n=35{,}362$ controls. Fixed-effect inverse-variance meta-analysis was utilized to generate the *P* values. AL, reference and non-reference alleles; BP, genomic position in HG19; OR, odds ratio.

t-test). Next, we asked whether the population differentiation drives schizophrenia variants to have different effects in different populations. Using 295 autosomal variants that are genome-wide significant in EAS, EUR or combined EAS and EUR samples, we did not observe a correlation ($R^2=0.003$; Extended Data Fig. 4b) between the population differentiation (measured by fixation index (F_{ST})) and the heterogeneity of the effect size (measured by $-\log_{10}$[*P* value] from the heterogeneity test across EAS and EUR).

As a further test, we examined whether the effect size estimates from EUR differed from those from EAS. We performed a heterogeneity test (Cochran's *Q*) for the most significant variants in the 108 published schizophrenia-associated loci[2]. Among them, seven variants showed significant heterogeneity after Bonferroni correction (Supplementary Table 4). Postulating that this might in part be driven by the inflation of EUR estimates as a result of the winner's curse, we applied a correction for the winner's curse[39], after which none of the variants showed evidence for significant heterogeneity, and the *P* values from the heterogeneity test followed a uniform distribution ($P=0.10$; two-tailed Kolmogorov–Smirnov test).

Lastly, we evaluated the heterogeneity of schizophrenia genetic effects within EAS samples. None of the EAS associations showed significant heterogeneity across EAS samples (Supplementary Table 3). Using their principal components, we further grouped the samples into Northeast Asian, Southeast Asian and Indonesian subpopulations (Methods). We then performed a heterogeneity test (Cochran's *Q*) and found no significant heterogeneity among the three subpopulations (Extended Data Fig. 5).

Schizophrenia genetic associations from the meta-analysis of EAS and EUR. As the genetic effects observed in EAS are largely consistent with those observed in EUR, we performed a meta-analysis including the EUR and EAS samples (stages 1 and 2), using a fixed-effect model with inverse-variance weighting[40]. The EUR + EAS samples in this analysis ($n=56{,}418$ cases and $n=78{,}818$ controls) included all samples of EUR ancestry ($n=33{,}640$ cases and $n=43{,}456$ controls) from a previous publication[2], with the exclusion of three samples of EAS ancestry and the deCODE samples ($n=1{,}513$ cases and $n=66{,}236$ controls), for which only summary statistics for selected variants were available. The three EAS samples (IMH-1, HNK-1 and JPN-1) excluded from the EUR samples were included in our EAS stage 1.

We identified 208 independent variants (both in EAS and EUR) associated with schizophrenia across 176 genetic loci (Fig. 2 and Supplementary Tables 5 and 6), among which 53 loci were novel (not reported in refs. [2,3,7,8]). Of the 108 schizophrenia-associated loci reported in the previous EUR study[2], 89 remained significant in this study (Supplementary Table 4). Using simulations with a correction for winner's curse[39], we found that this was consistent with an expected overestimation of the effect sizes due to the winner's curse in the previous study, rather than implying that the 19 loci no longer significant in this study were false positives (Supplementary Note). In addition, the deCODE samples ($n=1{,}513$ cases and $n=66{,}236$ controls) were not included in the present study, causing the power for loci that had low MAF in EAS to drop.

Population diversity improves fine-mapping. Causal variants in complex genetic disorders are defined as those that mechanistically contribute to the disorders, but this does not imply that the variant in isolation is likely to result in the disorder[41,42]. Due to LD, disease-associated loci from GWASs usually implicate genomic regions

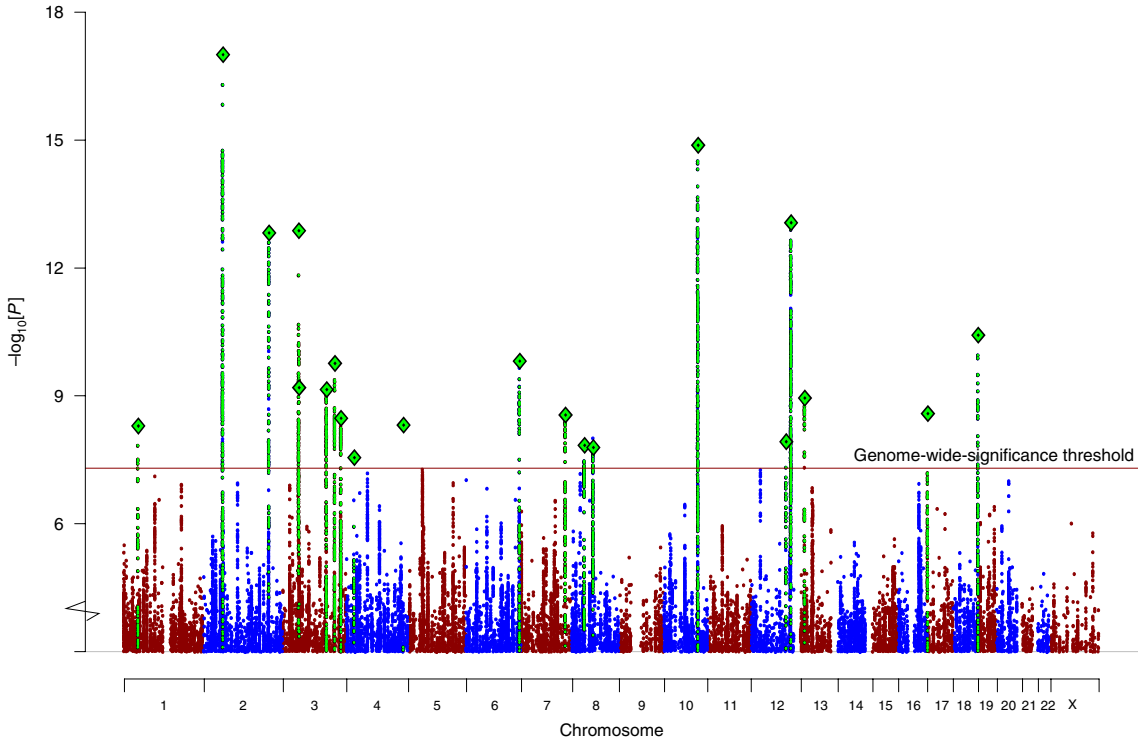


Fig. 1 | Genetic associations in EAS populations. Manhattan plot for schizophrenia genetic associations using EAS samples (stages 1 and 2; $n = 22{,}778$ cases; $n = 35{,}362$ controls). Red and blue dots refer to variants within odd and even chromosomes, respectively. Green dots refer to variants within genome-wide-significant loci. Diamonds represent index variants within genome-wide-significant loci. The genome-wide-significance threshold ($P < 5 \times 10^{-8}$) is represented as the horizontal red line.

containing many associated variants. A number of approaches allow for the associated variants to be refined to a smaller set of the most plausible (or credible) candidate causal variants[43–46]. Loci implicated in psychiatric disorders usually have small effect sizes, and as a result have generally poor performance using such approaches[2,3].

Diversity in genetic background across populations can be used to improve fine-mapping resolution[47]. Here, we demonstrate that resolution can be improved by exploiting differences in the patterns of LD between causal (directly associated) and non-causal (indirectly associated through LD) variants. Based on the premise that genetic effects are highly consistent across populations, the causal variants will have consistent effects across populations, whereas non-causal variants can have inconsistent effects due to population-specific LD patterns. We therefore expect causal variants to have greater statistical significance and less heterogeneity in the trans-ancestry meta-analysis compared with other alleles that are indirectly associated via LD (Extended Data Fig. 6). Using an algorithm based on this expectation (Methods), we fine-mapped 59 schizophrenia associations that reached genome-wide significance in the EUR and EAS stage 1 combined meta-analysis, had a MAF > 0.01 in both EAS and EUR, and for which we had >95% coverage of common variants (MAF > 1%) with imputation INFO > 0.6 (Supplementary Table 7). The MHC region was excluded from the fine-mapping analysis due to its long-range LD. EAS stage 2 samples were excluded because not all had full genome coverage, which confounds the fine-mapping outcome (Methods).

The results from this EAS and EUR trans-ancestry approach improved on those using only EUR, with 44 out of 59 loci mapped to a smaller number of candidate causal variants (Supplementary Table 7). For example, a locus on chromosome 1 (238.8–239.4 megabases (Mb)), which initially contained seven potentially causal variants based on a published fine-mapping method[43] and EUR samples only, was resolved to a single variant (rs11587347) with 97.6% probability (Fig. 3a). This variant showed strong association in both populations, while the other six variants were equally associated in EUR but not in EAS (Fig. 3b,c). Over all of the associations, the median size of the 95% credible set, defined as the minimum list of variants that were >95% likely to contain the causal variant, dropped from 49 to 30, and the number of associations mapped to ≤5 variants increased from two to seven (Fig. 3d). The number of associations mapped to a single variant with >50% probability increased from five to eight, and median size of the genomic regions the associations mapped decreased from 154 to 94 kilobases (kb).

Transferability of genetics across populations. For genome-wide-significant loci that individually explain >0.05% of the variance in schizophrenia liability in either ancestry, we compared the variance explained across EAS and EUR. Variance was approximated as $2f(1-f)\log[\mathrm{OR}]^2/(\pi^2/3)$ (ref. [48]) (Extended Data Fig. 7), where f represents the prevalence of the risk allele. Although these variants most often have comparable odds ratios across populations, their allele frequencies can differ. The variance explained when combining the effect size (odds ratio) and prevalence of the risk allele (f) can be regarded as an approximate measure of the importance of a causal variant in a population. In our analysis, most of the trans-ancestry differences in variance are explained by allele frequency differences. One of the implications of this observation, as suggested in recent studies[21,49,50], is that even if the risk alleles and effect sizes are primarily shared across populations, the disease predictive power of individual alleles, and of composite measures of those risk alleles such as PRS, may not be equivalent across populations.

Here, we evaluate this empirically. We assessed how much variation in schizophrenia risk can be explained in EAS using both EAS stage 1 and EUR training data. Using a standard clumping approach, we first computed PRS using a leave-one-out meta-analysis approach with EAS summary statistics (Methods), which explained ~3% of

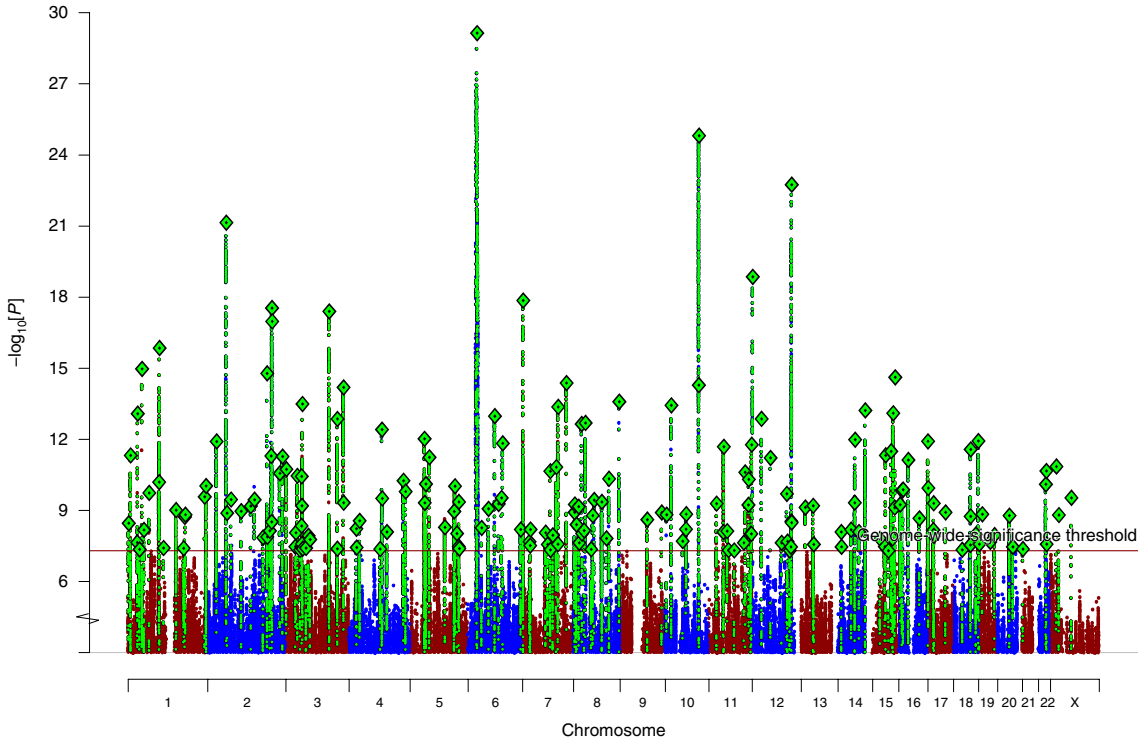


Fig. 2 | Schizophrenia associations in EUR and EAS samples. Manhattan plot for the schizophrenia genetic associations from the EAS (stages 1 and 2) and EUR meta-analysis ($n = 56{,}418$ cases; $n = 78{,}818$ controls). Red and blue dots refer to variantss within odd and even chromosomes, respectively. Green dots refer to variants within genome-wide-significant loci. Diamonds represent index variants within genome-wide-significant loci. Genome-wide-significance threshold ($P < 5 \times 10^{-8}$) is represented as the horizontal red line.

schizophrenia risk using genome-wide variants on the liability scale ($R^2 = 0.029$ at $P = 0.5$). In contrast, when EUR summary statistics were used to calculate PRS in the EAS samples, a maximum of only ~2% of schizophrenia risk was explained ($R^2 = 0.022$ at $P = 0.1$), despite a greater than threefold larger EUR effective sample size (Fig. 4 and Extended Data Fig. 8). The variance explained across various *P* value thresholds provides a proxy for the signal-to-noise ratio, which differs by training population—relative to the EUR training data, variants from the EAS training data with more permissive *P* values improve the EAS prediction accuracy. These results indicate that larger EAS studies will be needed to explain case/control variance similar to that currently explained in EUR individuals. Furthermore, although individual loci typically have the same direction and a similar magnitude across populations, aggregating variants that differentially tag causal loci across populations for genetic risk prediction results in considerable variability in prediction accuracy.

Discussion

To date, most large-scale psychiatric genetics studies have been based on samples of primarily EUR ancestry[6]. To increase global coverage, we compiled the largest non-EUR psychiatric genetics cohort to date and leveraged its size and diversity to provide new insights into the genetic architecture of schizophrenia. This study includes all available major genotyped schizophrenia samples of EAS ancestry, and presents analyses that have not previously been performed with sufficient power in psychiatric genetics. Although the first schizophrenia genetic associations from two much smaller studies of Chinese ancestry[51,52] were not genome-wide significant in the present EAS analysis, several loci from their subsequent better-powered studies[7,8] reached genome-wide significance. Consistent with a study using EUR samples[3], we note that this is consistent with the expected inflation of effect size from small studies, rather than suggesting that loci in previous studies are false positives.

When a single population is used to identify the disease-associated loci, the discovery is skewed towards disease-associated variants that have greater allele frequency in that population (Extended Data Fig. 9). When multiple populations are used, disease-associated variants are equally represented across the allele frequency spectrum in these populations (Extended Data Fig. 9). This shows that including global samples improves the power to find disease associations for which the power varies across populations. In this study, for example, more EUR than EAS samples would be required to detect around half of the new loci, as the MAF is higher in EAS than in EUR in these loci.

For traits such as body mass index and autoimmune diseases, we observed heterogeneity across populations in genetic effects[28,53], which may point to interactions between genetic associations and environmental factors and/or other genetic loci. In contrast, for schizophrenia, we did not find significant heterogeneity across EAS and EUR ancestries. Analyses of genetic heritability, genetic correlation, gene set enrichment and natural selection signatures converge on the conclusion that the schizophrenia biology is substantially shared across EAS and EUR ancestries (with MHC as a potential exception, as is discussed below). The remarkable genetic correlation ($r_g = 0.98$) shows that schizophrenia risk alleles operate consistently across different ethnic and cultural backgrounds—at least across EAS and EUR ancestries. Given that the main putative environmental risk factors (migration, urbanicity and substance misuse) differ across populations, this finding also suggests that any specific genetic liability to schizophrenia acting via these routes is minimal.

We note that a direct comparison of the effect sizes estimated in EAS with those estimated in EUR has reduced accuracy as we do not know the exact schizophrenia causal variants. This is further complicated by inflations in effect size estimates due to the winner's curse, which are of different magnitudes due to the sample size.

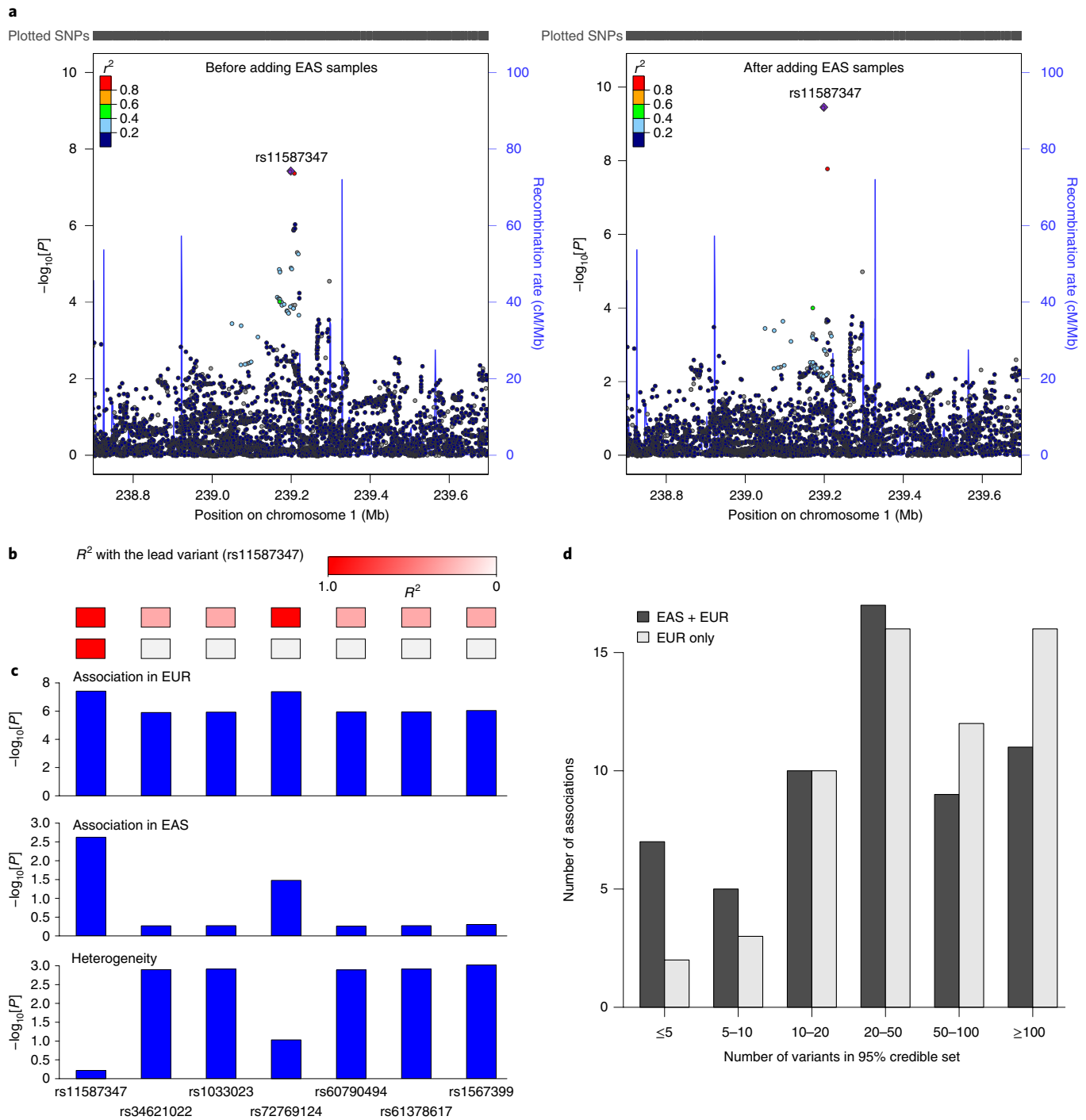


Fig. 3 | Trans-ethnic fine-mapping improves resolution. a, An association was mapped to a single variant (rs11587347) after adding EAS samples and using the trans-ancestry fine-mapping approach. Regional association plots were generated using http://locuszoom.org/ and LD from 1000 Genomes Project Phase 3 EUR subjects. Left, EUR-only samples. Right, Meta-analysis of EUR and EAS samples. **b**, LD with the lead variant (rs11587347). **c**, The lead variant (rs11587347) has strong association significance in both populations, and low heterogeneity across populations. In **a**–**c**, for EAS stage 1, $n = 13{,}305$ cases and $n = 16{,}244$ controls, and for EUR, $n = 33{,}640$ cases and $n = 43{,}456$ controls. **d**, Number of variants in the 95% credible set using the trans-ancestry (EAS + EUR) and published fine-mapping approaches (EUR only).

Increasing the sample size, especially in those of non-EUR ancestries, will reduce the bias and enable a better isolation of causal variants, leading to a more precise comparison of the genetic effect size across populations.

The MHC hosts the strongest schizophrenia association in EUR[54]. In this study, we did not find a significant schizophrenia association in MHC in EAS. An earlier EUR study[55] mapped the MHC associations to a set of variants (in LD) at both distal ends of the extended MHC (lead variant: rs13194504) and the complement component 4 (C4). None of these associations was significant in EAS in this study, which is consistent with previous studies of Chinese ancestry[7,8,51,52]. However, this does not necessarily

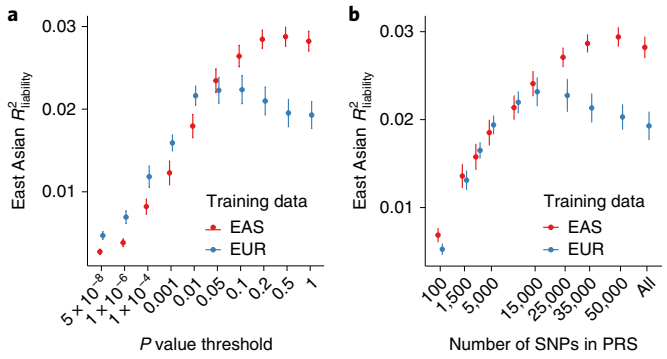


Fig. 4 | Genetic risk prediction accuracy in EAS from EAS or EUR training data. PRSs were computed with GWAS summary statistics from EAS and EUR populations as training sets. EAS risk alleles and weights were computed with a leave-one-out meta-analysis approach across the 13 stage 1 samples. Error bars indicate 95% confidence intervals. The LD panel for clumping is from the EUR and EAS 1000 Genomes Phase 3 samples. **a**, Case/control variance explained in EAS samples by variants from EAS and EUR training data with a *P* value more significant than the threshold. **b**, Case/control variance explained by the *n* most significant independent variants. In **a** and **b**, for EAS stage 1, $n=13{,}305$ cases and $n=16{,}244$ controls, and for EUR, $n=33{,}640$ cases and $n=43{,}456$ controls.

suggest population heterogeneity in their pathophysiological effect, as we attribute the disappearance of MHC signals partially to low frequencies. rs13194504 has a MAF < 1% in EAS, compared with 9% in EUR, and the C4-BS allele is extremely uncommon in samples from China and Korea[56,57]. Another reason may be the EUR-specific LD. In EUR, multiple protective alleles that contribute to the MHC associations are all on the same haplotype across about 6 Mb, due to an extremely long and EUR-specific haplotype that generates LD patterns at a 5-Mb scale. This may also be the reason that association signals span so many megabases of genome, and the aggregate association signal (at variants that are in partial LD to multiple signals) is stronger than the signals at the individual associations.

Two recent studies using much smaller samples with individuals of Chinese ancestry[7,8] reported variants in MHC significantly associated with schizophrenia (rs115070292 and rs111782145, respectively). The two studies did not replicate each other's findings, as the reported risk alleles were in very weak LD ($r^2=0.07$), nor did they replicate the EUR study[55], because the risk alleles were not in LD with the EUR MHC associations. rs115070292, from Yu et al.[7], is more frequent in EAS (12%) than in EUR (2%), with $P=10^{-9}$ using 4,384 cases and 5,770 controls of Chinese ancestry. This variant was not significantly associated in our study ($P=0.44$), even though some samples from the earlier study were included in the current study (BJM-1; 1,312 cases and 1,987 controls). The odds ratio estimated from these shared samples marginally differs from that estimated using all EAS samples ($P=0.018$), and this association showed marginally significant heterogeneity across all EAS samples ($P=0.039$). Similarly, we did not replicate the association at rs111782145 from Li et al.[8] ($P=0.47$) despite sample overlap (2,555 cases and 3,952 controls).

The lack of replication across all of these studies reflects the complexity of the MHC region and the limited power for the MHC signals in EAS. As shown in previous studies of complex disorders, it is still possible that when sample size increases for the EAS, genome-wide association within the MHC region could emerge. A study designed for the MHC region, such as in ref. [55], will be necessary to delineate the contribution of MHC to schizophrenia in EAS individuals.

Genetic associations usually implicate a large genomic region; thus, it can be challenging to map their molecular functions. We designed a novel algorithm to leverage the population diversity to fine-map schizophrenia associations to precise sets of variants. Using this algorithm, we reduced the number of candidate variants associated with schizophrenia and facilitated the functional interpretation of these associations. Our algorithm only maps the primary association signals in a locus because the power to fine-map signals beyond that, especially in the EAS samples, is still limited at the current sample size for schizophrenia. We also made an assumption that there is only one causal variant driving the primary association signal. In the scenario that there is a haplotypic effect driven by multiple variants in strong LD, our approach will split the posterior probability among these variants. We expect the causal variants to have non-trivial probability so that they will still be reported in the credible set for future studies. Imputation quality plays a key role in fine-mapping, as the power to map the causal variant decreases if it is poorly imputed. We restricted our study to genetic associations that have a MAF > 1% in both EAS and EUR populations to ensure the imputation quality. For these associations, we found no major change in the size of the credible sets when the EUR samples were imputed using the more powerful Haplotype Reference Consortium panel[58]. However, the Haplotype Reference Consortium reference panel, with its much larger sample size and better characterization of low frequency and rare variants, could improve fine-mapping resolution for variants with a MAF ≤ 1%[59].

Finally, this large-scale EAS sample allowed us to empirically evaluate the congruence of the genetic basis of schizophrenia between EAS and EUR. Despite a cross-population common variant genetic correlation being highly consistent, we found that polygenic risk models trained in one population have reduced performance in the other population due to different allele frequency distributions and LD structures. This highlights the importance of including all major ancestral groups in genomic studies, both as a strategy to improve the power to find disease associations, and to ensure that the findings have maximum relevance for all populations.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-019-0512-x.

Received: 9 January 2019; Accepted: 10 September 2019;
Published online: 18 November 2019

References

1. Stilo, S. A. & Murray, R. M. The epidemiology of schizophrenia: replacing dogma with knowledge. *Dialogues Clin. Neurosci.* **12**, 305–315 (2010).
2. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
3. Pardiñas, A. F. et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018).
4. Visscher, P. M. et al. 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
5. Li, Y. R. & Keating, B. J. Trans-ethnic genome-wide association studies: advantages and challenges of mapping in diverse populations. *Genome Med.* **6**, 91 (2014).
6. Popejoy, A. B. & Fullerton, S. M. Genomics is failing on diversity. *Nature* **538**, 161–164 (2016).
7. Yu, H. et al. Common variants on 2p16.1, 6p22.1 and 10q24.32 are associated with schizophrenia in Han Chinese population. *Mol. Psychiatry* **22**, 954–960 (2017).
8. Li, Z. et al. Genome-wide association analysis identifies 30 new susceptibility loci for schizophrenia. *Nat. Genet.* **49**, 1576–1583 (2017).

9. Moltke, I. et al. A common Greenlandic *TBC1D4* variant confers muscle insulin resistance and type 2 diabetes. *Nature* **512**, 190–193 (2014).
10. Tang, C. S. et al. Exome-wide association analysis reveals novel coding sequence variants associated with lipid traits in Chinese. *Nat. Commun.* **6**, 10206 (2015).
11. Clarke, T.-K. et al. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112 117). *Mol. Psychiatry* **22**, 1376–1384 (2017).
12. CONVERGE consortium. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* **523**, 588–591 (2015).
13. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
14. *Report of the International Pilot Study of Schizophrenia* (World Health Organization, 1973).
15. Carpenter, W. T. Jr, Strauss, J. S. & Bartko, J. J. Flexible system for the diagnosis of schizophrenia: report from the WHO International Pilot Study of Schizophrenia. *Science* **182**, 1275–1278 (1973).
16. Jablensky, A. et al. Schizophrenia: manifestations, incidence and course in different cultures. A World Health Organization ten-country study. *Psychol. Med. Monogr. Suppl.* **20**, 1–97 (1992).
17. McGrath, J. J. Variations in the incidence of schizophrenia: data versus dogma. *Schizophr. Bull.* **32**, 195–197 (2006).
18. Haro, J. M. et al. Cross-national clinical and functional remission rates: worldwide schizophrenia outpatient health outcomes (W-SOHO) study. *Br. J. Psychiatry* **199**, 194–201 (2011).
19. Gureje, O. & Cohen, A. Differential outcome of schizophrenia: where we are and where we would like to be. *Br. J. Psychiatry* **199**, 173–175 (2011).
20. Scutari, M., Mackay, I. & Balding, D. Using genetic distance to infer the accuracy of genomic prediction. *PLoS Genet.* **12**, e1006288 (2016).
21. Martin, A. R. et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am. J. Hum. Genet.* **100**, 635–649 (2017).
22. Martin, A. R. et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).
23. Bulik-Sullivan, B. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
24. Geisler, S., Schöpf, C. L. & Obermair, G. J. Emerging evidence for specific neuronal functions of auxiliary calcium channel $\alpha_2\delta$ subunits. *Gen. Physiol. Biophys.* **34**, 105–118 (2015).
25. Heyes, S. et al. Genetic disruption of voltage-gated calcium channels in psychiatric and neurological disorders. *Prog. Neurobiol.* **134**, 36–54 (2015).
26. Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381**, 1371–1379 (2013).
27. Gulsuner, S. et al. Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154**, 518–529 (2013).
28. Liu, J. Z. et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).
29. Brown, B. C., Asian Genetic Epidemiology Network Type 2 Diabetes Consortium, Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic genetic-correlation estimates from summary statistics. *Am. J. Hum. Genet.* **99**, 76–88 (2016).
30. Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
31. Davies, G. et al. Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N=112 151). *Mol. Psychiatry* **21**, 758–767 (2016).
32. Cross-Disorder Group of the Psychiatric Genomics Consortium et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
33. Anttila, V. et al. Analysis of shared heritability in common disorders of the brain. *Science* **360**, eaap8757 (2018).
34. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
35. Bryois, J. et al. Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. *Nat. Commun.* **9**, 3121 (2018).
36. Lindblad-Toh, K. et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).
37. De Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
38. Genovese, G. et al. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* **19**, 1433–1441 (2016).
39. Bowden, J. & Dudbridge, F. Unbiased estimation of odds ratios: combining genomewide association scans with replication studies. *Genet. Epidemiol.* **33**, 406–418 (2009).
40. Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. *Introduction to Meta-Analysis* (John Wiley & Sons, 2009).
41. Spain, S. L. & Barrett, J. C. Strategies for fine-mapping complex traits. *Hum. Mol. Genet.* **24**, R111–R119 (2015).
42. Schaid, D. J., Chen, W. & Larson, N. B. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* **19**, 491–504 (2018).
43. Huang, H. et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* **547**, 173–178 (2017).
44. Farh, K. K.-H. et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
45. Swaminathan, B. et al. Fine mapping and functional analysis of the multiple sclerosis risk gene *CD6*. *PLoS ONE* **8**, e62376 (2013).
46. Gaulton, K. J. et al. Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat. Genet.* **47**, 1415–1425 (2015).
47. Kichaev, G. & Pasaniuc, B. Leveraging functional-annotation data in trans-ethnic fine-mapping studies. *Am. J. Hum. Genet.* **97**, 260–271 (2015).
48. Pawitan, Y., Seng, K. C. & Magnusson, P. K. E. How many genetic variants remain to be discovered? *PLoS ONE* **4**, e7969 (2009).
49. Márquez-Luna, C., Loh, P.-R., South Asian Type 2 Diabetes (SAT2D) Consortium, SIGMA Type 2 Diabetes Consortium & Price, A. L. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet. Epidemiol.* **41**, 811–823 (2017).
50. Chen, C.-Y., Han, J., Hunter, D. J., Kraft, P. & Price, A. L. Explicit modeling of ancestry improves polygenic risk scores and BLUP prediction. *Genet. Epidemiol.* **39**, 427–438 (2015).
51. Yue, W.-H. et al. Genome-wide association study identifies a susceptibility locus for schizophrenia in Han Chinese at 11p11.2. *Nat. Genet.* **43**, 1228–1231 (2011).
52. Shi, Y. et al. Common variants on 8p12 and 1q24.2 confer risk of schizophrenia. *Nat. Genet.* **43**, 1224–1227 (2011).
53. Salinas, Y. D., Wang, L. & DeWan, A. T. Multiethnic genome-wide association study identifies ethnic-specific associations with body mass index in Hispanics and African Americans. *BMC Genet.* **17**, 78 (2016).
54. International Schizophrenia Consortium et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).
55. Sekar, A. et al. Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177–183 (2016).
56. Chen, J. Y. et al. Effects of complement C4 gene copy number variations, size dichotomy, and C4A deficiency on genetic risk and clinical presentation of systemic lupus erythematosus in East Asian populations. *Arthritis Rheumatol.* **68**, 1442–1453 (2016).
57. Hong, G. H. et al. Association of complement C4 and HLA-DR alleles with systemic lupus erythematosus in Koreans. *J. Rheumatol.* **21**, 442–447 (1994).
58. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
59. Mahajan, A. et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).

[1]Bio-X Institutes, Shanghai Jiao Tong University and Research Division, Institute of Mental Health Singapore, Singapore, Singapore. [2]Human Genetics, Genome Institute of Singapore, Singapore, Singapore. [3]Division of Psychiatry Research, the Zucker Hillside Hospital, Northwell Health, Glen Oaks, NY, USA. [4]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [5]Stanley Center for Psychiatric Research, the Broad Institute of MIT and Harvard, Cambridge, MA, USA. [6]Research Division, Institute of Mental Health Singapore, Singapore, Singapore. [7]Department of Medicine, Harvard Medical School, Boston, MA, USA. [8]Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [9]Biogen, Cambridge, MA, USA. [10]Biomedical Sciences Institute of Qingdao University, Qingdao Branch of Shanghai Jiao Tong University Bio-X Institutes and the Affiliated Hospital of Qingdao University, Qingdao, China. [11]Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education) and the Collaborative Innovation Center for Brain Science, Bio-X Institutes, Shanghai Jiao Tong University, Shanghai, China. [12]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [13]Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China. [14]Social Genetic and Developmental Psychiatry, King's College London, London, UK. [15]Department of Psychiatry, Fujita Health University School of Medicine, Toyoake, Japan. [16]Australian Centre for Precision Health, School of Health Sciences, University of South Australia Cancer Research Institute, Adelaide, South Australia, Australia. [17]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia. [18]South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia. [19]Data Science Institute, Columbia University, New York, NY, USA. [20]New York Genome Center, New York, NY, USA. [21]Shanghai Key Laboratory of Psychotic Disorders, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China. [22]Peking University Sixth Hospital and Institute of Mental Health, Beijing, China. [23]National Health Commission Key Laboratory of Mental Health (Peking University), Beijing, China. [24]National Clinical Research Center for Mental Disorders (Peking University Sixth Hospital), Beijing, China. [25]Laboratory of Complex Trait Genomics, Graduate School of Frontier Sciences, University of Tokyo, Tokyo, Japan. [26]Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [27]Department of Psychiatry, Chonnam National University Medical School, Gwangju, Republic of Korea. [28]RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. [29]Department of Psychiatry, Cipto Mangunkusumo General Hospital, Universitas Indonesia, Jakarta, Indonesia. [30]Department of Psychiatry, National Taiwan University Hospital, Taipei, Taiwan. [31]Department of Psychiatry, National Taiwan University College of Medicine, Taipei, Taiwan. [32]Queensland Brain Institute The University of Queensland, Brisbane, Queensland, Australia. [33]Queensland Center for Mental Health Research, The University of Queensland, Wacol, Queensland, Australia. [34]Department of Genomic Medicine, Research Institute, National Cerebral and Cardiovascular Center, Osaka, Japan. [35]Department of Psychiatry, University Medical Center Utrecht, Utrecht, the Netherlands. [36]Center for Translational Medicine, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China. [37]Department of Psychiatry, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China. [38]Center for Brain Science, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China. [39]A list of members and affiliations appears in the Supplementary Note. [40]Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan. [41]The State University of New York, Syracuse, NY, USA. [42]Psychiatric Genetic Epidemiology and Neurobiology Laboratory (PsychGENe lab), Department of Psychiatry and Behavioral Sciences, SUNY Upstate Medical University, Syracuse, NY, USA. [43]Shanghai Center for Women and Children's Health, Shanghai, China. [44]Baoan Maternal and Child Health Hospital, Jinan University, Shenzhen, China. [45]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA. [46]Department of Genetics, Harvard Medical School, Boston, MA, USA. [47]Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [48]University of Western Australia, Perth, Western Australia, Australia. [49]Institute of Mental Health, Singapore, Singapore. [50]MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, School of Medicine , Cardiff University, Cardiff, UK. [51]School of Life Science and Technology, ShanghaiTech University, Shanghai, China. [52]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China. [53]Collaborative Innovation Center of Genetics and Development, Shanghai, China. [54]Department of Genetics, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [55]Department of Psychiatry and Psychotherapy Charité – Universitätsmedizin, Berlin, Germany. [56]Institute for Molecular Medicine Finland (FIMM), Helsinki, Finland. [57]Collaborative Innovation Center, Jining Medical University, Jining, China. [58]State Key Laboratory of Brain and Cognitive Sciences, Centre for Genomic Sciences, The University of Hong Kong, Hong Kong, China. [59]Department of Psychiatry, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China. [60]Department of Psychiatry, Sungkyunkwan University School of Medicine, Samsung Medical Center, Seoul, Korea. [61]Molecular Horizons and School of Chemistry and Molecular Bioscience, University of Wollongong, Wollongong, New South Wales, Australia. [62]Illawarra Health and Medical Research Institute, Wollongong, New South Wales, Australia. [63]IDG/McGovern Institute for Brain Research at Peking University, Beijing, China. [64]Center for Behavioral Genomics, Department of Psychiatry, University of California, San Diego, San Diego, CA, USA. [65]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [66]Clinical Research Center for Mental Disease of Shaanxi Province, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China. [67]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [68]Behavioral Health System, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [69]Department of Psychiatry, Brain Center Rudolf Magnus, UMC Utrecht, Utrecht, the Netherlands. [70]Department of Psychiatry, First Teaching Hospital of Xinjiang Medical University, Ürümqi, China. [71]These authors jointly supervised this work: Shengying Qin, Pak Sham, Nakao Iwata, Kyung S. Hong, Sibylle G. Schwab, Weihua Yue, Ming Tsuang, Jianjun Liu, Xiancang Ma, René S. Kahn, Yongyong Shi, Hailiang Huang. [72]These authors contributed equally: Max Lam, Chia-Yen Chen. *e-mail: dryue@bjmu.edu.cn; mtsuang@ucsd.edu; liuj3@gis.astar.edu.sg; maxiancang@163.com; rene.kahn@mssm.edu; shiyongyong@gmail.com; hhuang@broadinstitute.org

Methods

Overview of samples. *EAS samples.* Full genome. Genome-wide genotype data were obtained from 16 samples from East Asia (Supplementary Table 1). Two of these samples (TAI-1 and TAI-2) had parent–offspring trios and were processed as case/pseudo-controls. The *Diagnostic and Statistical Manual of Mental Disorders* (fourth edition)[60] was used for diagnosing all schizophrenia cases in these samples, except for the trios (TAI-1 and TAI-2), for which the *Diagnostic Interview for Genetic Studies*[61] was used. All samples were processed according to the quality control procedures reported in ref. [2], with details reported in the following sections. After quality control, genotypes were phased and imputed against the 1000 Genomes Project Phase 3 reference panel[6]. Principal component analysis (PCA) was conducted across samples via imputed best-guess genotypes to identify and remove overlapping samples across datasets, cryptic related samples and population outliers. Eight principal components that were associated with case-control status were included in univariate logistic regression as covariates to control for the population stratification in each sample.

Selected variants. Summary statistics were obtained for a set of variants from four EAS samples (BJM-2, BJM-3, BJM-4 and BIX-5) that had been analyzed in published studies[7,8]. The summary statistics included odds ratios, standard errors, reference and tested alleles for variants that had $P < 10^{-5}$ in either stage 1 or the meta-analysis combining stage 1 and EUR samples. Between 22,156 and 31,626 variants were available after the exclusion of strand-ambiguous[62] variants (Supplementary Table 2).

EUR samples. Genotypes for EUR schizophrenia patients and controls were obtained from the Psychiatric Genomics Consortium, as reported in ref. [2]. All samples of EUR ancestry were included in this study except for the deCODE samples (1,513 cases and 66,236 controls). We also note that three sample collections of EAS ancestry reported in ref. [2] (IMH-1, HNK-1 and JPN-1) were not included in the EUR samples in our analysis, but were included in the EAS samples. The same procedures used in processing EAS samples were applied to the EUR samples.

EAS subpopulations. To investigate the heterogeneity of schizophrenia genetics effects within EAS, we grouped the samples based on their principal components. Other than Indonesians (UWA-1), who fall into their own subpopulation, samples were grouped into the Northeast Asian subpopulation if their average principal component 2 score was significantly greater than 0 (BIX-2, BJM-1, XJU-1, JPN-1 and KOR-1) and into the Southeast Asian subpopulation if their average principal component 2 score was significantly less than 0 (TAI-1, TAI-2, IMH-1, IMH-2, HNK-1 and BIX-3). The remaining samples (UMC-1, SIX-1, BIX-1 and BIX-4) were not included in the subpopulations. The heterogeneity test (Cochran's *Q*) across subpopulations (calculated pairwise and across all) was conducted using RICOPILI[63].

Quality control. Quality control procedures were carried out as part of the RICOPILI pipeline[63] with the following steps and parameters: (1) excluding variants with a call rate below 95%; (2) excluding subjects with a call rate below 98%; (3) excluding monomorphic variants; (4) excluding subjects with an inbred coefficient above 0.2 and below −0.2; (5) excluding subjects with a mismatch in their reported sex and chromosome X imputed sex; (6) excluding variants with missing rate differences >2% between cases and controls; (7) subsequent to step 6, excluding variants with a call rate <98%; and (8) excluding variants in violation of Hardy–Weinberg equilibrium ($P < 10^{-6}$ for controls or $P < 10^{-10}$ for cases). The numbers of variants or subjects removed in each step are reported in Supplementary Table 2.

Phasing and imputation. All datasets were phased using SHAPEIT[64] and IMPUTE2 (ref. [65]) using regular steps and parameters. Additional processing for trios (TAI-1 and TAI-2) was carried out such that case/pseudo-controls were identified and imputed. All samples were imputed to the 1000 Genomes Project Phase 3 reference panel[66] (2,504 subjects, including 504 EAS subjects). Imputation procedures resulted in dosage files and best-guess genotypes in PLINK[67] binary format. Dosage files were used for subsequent association analysis, while best-guess genotypes were used in the PCA and PRS analyses.

Sample overlaps, population outliers and population stratification. We used EIGENSTRAT[68] to calculate the principal components for all of the samples using the best-guess genotypes from imputation (Extended Data Fig. 10b). We computed the identity-by-descent matrix to identify intra- and inter- dataset sample overlaps. Samples with pi-hat > 0.2 were extracted, followed by a Fisher–Yates shuffle on all samples. The number of times each sample was related to another sample was tracked, and samples that were related to more than 25 samples were removed. When deciding which samples to retain, trios were preferred, followed by cases, and thereafter a random sample for each related pair was removed, resulting in the removal of 704 individuals.

To identify population outliers, *k*-means clustering was conducted using the first 20 principal components from PCA and covariates representing each of the 13 stage 1 samples. Guided by the results of *k*-means clustering and visual inspection of PCA plots, 46 individuals were identified as outliers and were excluded. Further population-level inspection was carried out by merging the 1000 Genomes Project Phase 1 reference samples with stage 1 samples and conducting PCA (Extended Data Fig. 10a). Using similar approaches to those reported above, no further samples were excluded as population outliers.

Eight principal components that were associated with case/control status with $P < 0.2$ were used as covariates for association analysis in each sample (principal components 1, 4, 5, 6, 8, 9, 15 and 19). QQ plots (Extended Data Fig. 1) showed that the population structure was well controlled.

Association analysis and meta-analysis. Association analysis was carried out for each sample using PLINK[67] and genotype dosage from imputation. Only variants with imputation INFO ≥ 0.6 and a MAF ≥ 1% were included in the analysis. We performed logistic regression with principal components identified in the previous subsection as covariates to control for population stratification within each study. Fixed-effect meta-analysis[69], weighted by inverse variance, was then used to combine association results across samples. A meta-analysis for EUR samples was conducted in the same manner. To find independent schizophrenia associations in both EUR and EAS populations (Supplementary Table 6), we performed LD clumping twice using the 1000 Genomes Project Phase 3 EUR and EAS reference panels, respectively (with default parameters in RICOPILI).

Chromosome X analysis. Chromosome X genotypes were processed separately from autosomal variants. Quality control was conducted separately for males and females, using similar quality control parameters as above. Cases and pseudo-controls were built out of the trios. Phasing and imputation were then performed on males and females separately for each sample, followed by logistic regression with the same principal components, and meta-analysis combining samples (same parameters as the autosomal analyses). The results were generated for EAS stage 1 samples and EUR and EAS combined (excluding BIX-1, BIX-2 and BIX-3). EAS stage 2, BIX-1, BIX-2 and BIX-3 samples did not have chromosome X data and were therefore not analyzed.

Genetic correlation and heritability. Schizophrenia heritabilities in the observed scale for samples of EUR and EAS ancestry were estimated from their summary statistics using LDSC[23]. We converted the heritabilities in the observed scale to a liability scale assuming a schizophrenia population prevalence of 1%. The LD scores were pre-computed from the 1000 Genomes Project Phase 3 reference panel in EUR and EAS, respectively (https://github.com/bulik/ldsc). Only autosomal variants with a MAF > 5% in their respective population were included in the analysis, and variants in the MHC region were not included due to the long-range LD.

We computed the genetic correlations between schizophrenia and other traits within EUR and across EUR and EAS. EUR and EAS (stage 1 only) summary statistics for autosomal variants from this study were used as schizophrenia genetic association inputs for their respective populations. The traits tested included schizophrenia[2], bipolar[70], major depression[71], anorexia nervosa[72], neuroticism and subjective well-being[73], autism spectrum disorder ('GWAS - 2015' release; available at http://www.med.unc.edu/pgc), attention deficit hyperactivity disorder ('European Ancestry GWAS'; available at http://www.med.unc.edu/pgc)[74], educational attainment[30], general intelligence[75], fluid intelligence score and prospective memory result (using individuals from UK Biobank; http://www.nealelab.is/uk-biobank). Only variants with a MAF > 5% were available and included. Variants in the MHC region were excluded from the analysis. Genetic correlations within EUR were computed using LDSC with LD scores pre-computed on the 1000 Genomes Project Phase 3 reference panel (503 EUR subjects). Genetic correlations across EUR and EAS were computed using POPCORN[29]. POPCORN uses a Bayesian approach that assumes that genotypes are drawn separately from each population and effect sizes follow the infinitesimal model. Genetic correlations in POPCORN were computed in the 'genetic effect' mode, which estimates the correlation based on the LD covariance scores and effect sizes from summary statistics.

Partitioned heritability. Partitioned LDSC[34] was conducted to look for heritability enrichment in diverse annotations using EAS (stage 1) and EUR autosomal variants (summary statistics), respectively. LD scores for each annotation were computed using a combination of PLINK[67] and LDSC[23], using the 1000 Genomes Project EAS and EUR subjects, respectively. We used baseline annotations[34] and additional annotations including chromatin accessibility in the brain dorsolateral prefrontal cortex, as determined via assay for transposase-accessible chromatin using sequencing peaks (Bryois et al.[35]; 'ATAC Bryois'), conserved regions (Lindblad-Toh et al.[36]; 'Conserved LindbladToh') located via the assay for transposase-accessible chromatin using sequencing peaks[35], and introgressed regions from Neanderthals[76] ('Neanderthal Vernot'). Variants can be included in multiple annotations. Multi-allelic variants were removed.

Gene set analysis. We performed gene- and gene set-based tests using MAGMA[37]. Genome-wide summary statistics for autosomal variants from EAS, EUR and EAS + EUR meta-analyses were used in this analysis. Variant-to-gene annotation

was performed using RefSeq NCBI37.3 with a window of 5 kb upstream and 1.5 kb downstream. LD was taken from the 1000 Genomes Project EAS, EUR and EUR + EAS panels, respectively. The gene-based *P* values were computed by *F*-test and multivariate linear modeling, and competitive tests were used for gene set analysis. A total of 70 gene sets were selected and tested in this study (Supplementary Table 8), including those from the Molecular Signatures Database[77], those related to psychiatric diseases[38,78,79] and those from gwaspipeline (https://github.com/freeseek/gwaspipeline/blob/master/makegenes.sh). Gene sets were ranked for EUR, EAS and EAS + EUR analyses, respectively. The top-ranking gene sets were compared across analyses to identify common schizophrenia pathways. Additionally, Wilcoxon signed-rank tests were conducted to compare the ranking of gene sets between the EUR and EAS datasets.

Natural selection analysis. We used the Han Chinese in Beijing (CHB) and Utah residents with ancestry from northern and western Europe (CEU) panels from the 1000 Genomes Project Phase 3 to investigate the natural selection signatures in schizophrenia-associated loci for the EAS and EUR populations, respectively. We used the following selection signatures, with their sensitivity to timeframes discussed in ref. [3].

Integrated haplotype score (iHS). The iHS captures the haplotype homozygosity at a given variant. We calculated iHS using the R rehh package[80]. Genetic distance between variants was determined using the HapMap phase II genetic map. Ancestral and derived alleles were obtained from the 1000 Genome project, which inferred the ancestral state using six primates on the Enredo–Pecan–Ortheus pipeline. Only biallelic variants that with a MAF ≥ 5% were included in the analysis.

Cross-population extended haplotype homozygosity (XPEHH). XPEHH[81] detects variants under selection in one population but not the other. We used CEU as the reference panel when calculating XPEHH for CHB, and vice versa.

F_{ST}. F_{ST} measures the population differentiation due to genetic structure. We estimated F_{ST} using the Weir and Cockerham approach[82], which is robust to sample-size effects.

Absolute derived allele frequency difference (|ΔDAF|). |ΔDAF| measures population differentiation between CHB and CEU populations.

Composite of multiple signals (CMS). CMS[83–86] combines iHS, XPEHH, F_{ST} and |ΔDAF|. As a result, CMS potentially has better power to detect the selection signature. For each variant, CMS $= \prod_{i=1}^{n} p_i$, in which p_i is the rank of the variant using method *i*, sorted by increasing *P* values, divided by the total number of variants.

B statistic. The B statistic measures the background selection. We calculated the B statistic as in ref. [85].

Trans-ethnicity fine-mapping. For a disease-associated genetic locus, fine-mapping defines a 'credible set' of variants that contains the causal variant with certain probability (for example, 99% or 95%). Bayesian fine-mapping approaches[2,43,87,88] have been widely used for studies of a single ancestry. Here, we extended a Bayesian fine-mapping approach[87] (defining credible sets; Methods) to studies of more than one ancestry. Intuitively, the extension was achieved through a prior calculated from the heterogeneity across ancestries, such that variants with different odds ratios across populations have a smaller prior probability to be the causal variant.

As in several previous studies[2,87], we restricted our fine-mapping analysis to the primary association signal in each locus. This was done by taking *P* variants that were in LD with the lead variant (the variant with the most significant *P* value), with $r^2 > 0.1$ in EUR or EAS. Assume that *D* represents the data including the genotype matrix *X* for the *P* variants and disease *Y* for *N* individuals, and β represents a collection of model parameters. We define the model, denoted by *A*, as the causal status for the *P* variants in locus $\mathbf{A} \equiv \{a_j\}$, in which a_j is the causal status for variant *j*. $a_j = 1$ if the variant *j* is causal, and $a_j = 0$ if it is not. For the primary association signal and under the presumption that the causal variant is the same across all ancestries, one and only one of the *P* variants is causal: $\sum_j a_j = 1$. For convenience, we define A_j as the model in which only variant *j* is causal, and A_0 as the model in which no variant is causal (the null model). The probability of model A_j (where variant *j* is the only causal variant in the locus), given the data (*D*), can be calculated using Bayes' rule:

$$\Pr(A_j|D) = \Pr(D|A_j)\frac{\Pr(A_j)}{\Pr(D)}$$

With the steepest descent approximation, the assumption of a flat prior on the model parameters (β), and the assumption of one causal variant per locus (equation (2) in ref. [87]), $\Pr(A_j|D)$ can be approximated as:

$$\Pr(A_j|D) \approx \Pr\left(D|A_j, \hat{\beta}_j\right) N^{-1/2} \frac{\Pr(A_j)}{\Pr(D)} \quad (1)$$

in which *N* is the sample size. We denote χ_j^2 as the χ^2 test statistic for variant *j*, which can be calculated from the *P* value from the meta-analysis combining the EAS and EUR samples. Using equation (3) in ref. [87], we have:

$$\Pr\left(D|A_j, \hat{\beta}_j\right) \approx \exp\left(\frac{\chi_j^2}{2}\right) \Pr\left(D|A_0, \hat{\beta}_0\right) \quad (2)$$

$\Pr(A_j)$ is the prior probability that variant *j* is causal. We have shown that schizophrenia causal variants have consistent genetic effects across populations. Therefore, we model the prior probability as a function of the heterogeneity measured in I^2:

$$\Pr(A_j) = 1 - I_j^2 \quad (3)$$

Using equations (2) and (3), $\Pr(A_j|D)$ in equation (1) can be calculated as:

$$\Pr(A_j|D) \approx \exp\left(\frac{\chi_j^2}{2}\right)\left(1 - I_j^2\right)\frac{N^{-1/2}}{\Pr(D)}\Pr\left(D|A_0, \hat{\beta}_0\right)$$

We only use stage 1 samples in fine-mapping, so the variants have the same sample size (assuming all variants have good imputation quality). Therefore, $N^{-1/2}$, $\Pr(D)$ and $\Pr\left(D|A_0, \hat{\beta}_0\right)$ can be regarded as constants:

$$\Pr(A_j|D) \propto \exp\left(\frac{\chi_j^2}{2}\right)\left(1 - I_j^2\right)$$

The normalized causal probability for variant *j* is then:

$$P(A_j) = \Pr(A_j|D) / \sum_k \Pr(A_k|D)$$

And the 95% credible set of variants is defined as the smallest set of variants, *S*, such that:

$$\sum_{A_j \in S} P(A_j) \geq 95\%$$

PRS analysis. We constructed PRSs using a pruning and thresholding approach in EAS individuals with training summary statistics from either EUR or EAS individuals. For EUR, we used summary statistics from all EUR individuals in this study, whereas for EAS, we used a leave-one-out meta-analysis approach across the 13 EAS stage 1 sample collections to build the PRS.

For the EUR training data, we extracted EUR individuals (FIN, GBR, CEU, IBS and TSI) from the 1000 Genomes Project[66] Phase 3 as an LD reference panel to greedily clump variants. For the EAS LD reference panel, we created two panels: (1) an analogous EAS panel (CDX, CHB, CHS, JPT and KHV) from the 1000 Genome Project[66] Phase 3 (Fig. 4 and Extended Data Fig. 8c,d); and (2) an LD panel from best-guess genotypes from each cohort in the study (Extended Data Fig. 8a,b,e,f). For both the EAS and EUR prediction sets, we filtered to variants with MAF > 1% in each respective population, and removed indels and strand-ambiguous variants. We subset each list of variants to those in the summary statistics with an imputation INFO > 0.9. We then selected approximately independent loci at varying *P* value thresholds or top-ranking *n* variants using an LD threshold of $r^2 \leq 0.1$ in a window of 500 kb in PLINK[67] with the --clump flag. We treated the MHC with additional caution to minimize overfitting by selecting only the most significant variant in this region. To profile variants, we multiplied the log[odds ratio] for selected variants by genotype dosages, and summed these values across the genome in PLINK[67] using the --score flag for each of the 13 EAS stage 1 samples. We assessed case/control variance explained by computing Nagelkerke's R^2 and a liability-scale pseudo-R^2, as in Lee et al.[89], by comparing a full model with the PRS and ten principal components with a model excluding the PRS. The results of PRS were presented in two ways. First, we selected SNPs based on GWAS *P* value thresholds (that is, 5×10^{-8}, 1×10^{-6}, 1×10^{-4}, 0.001, 0.01, 0.05, 0.1, 0.2, 0.5 and 1) and *P* value ranks. Second, top-ranked SNPs that existed between both EUR and EAS summary statistics were selected based on the SNP rank thresholds (that is, the top 100, 1,500, 5,000, 15,000, 25,000, 35,000 and 50,000, and all).

Ethics. The study protocols were approved by the institutional review board at each center involved with recruitment. Informed consent and permission to share the data were obtained from all subjects, in compliance with the guidelines specified by the recruiting center's institutional review board. Samples recruited in mainland China were processed and analyzed in a Chinese server to comply with the Interim Measures for the Administration of Human Genetic Resources (a regulation from the Ministry of Science and Technology of the People's Republic of China). We set up the computer codes on the Chinese server so that analyses performed on these samples were exactly the same as other samples. Summary statistics from these Chinese samples, with no individual-level data, were then shared and combined with the rest of the EAS samples.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Genome-wide summary statistics relating to the EAS samples, EUR samples ($n = 49$) and all samples combined (that is, EAS and EUR) can be downloaded from https://www.med.unc.edu/pgc/. Individual-level genotype data for EAS samples are available on request from the contact authors (Supplementary Note). Alternatively, requests can be made to the Psychiatric Genomics Consortium. In this case, access to individual-level genotypes from samples recruited outside of mainland China will go through the Psychiatric Genomics Consortium's fast-track approval system. Access to individual-level genotypes from samples recruited within mainland China has to be approved by the individual Chinese contact authors (Supplementary Note), and are subject to the policies and approvals from the Human Genetic Resource Administration, Ministry of Science and Technology of the People's Republic of China. Individual-level genotypes from samples recruited within mainland China are stored and kept in a server physically located in mainland China. Analyses were performed on these samples using the same computer codes as those used for other EAS and EUR samples, which are available in the 'Code availability' section.

Code availability

Computer code relating to this study includes: RICOPILI (quality control, PCA, pre-phasing, imputation, association test and meta-analysis; https://github.com/Nealelab/ricopili/wiki); The following code is embedded within RICOPILI (EIGENSTRAT; https://github.com/DReichLab/EIG/tree/master/EIGENSTRAT; SHAPEIT, https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html; EAGLE, https://github.com/poruloh/Eagle; IMPUTE, https://mathgen.stats.ox.ac.uk/impute/impute_v2.html; Minimac, https://genome.sph.umich.edu/wiki/Minimac); POPCORN (trans-ancestry genetic correlation; https://github.com/brielin/Popcorn); LDSC (heritability, partitioned heritability and within-ancestry genetic correlation; https://github.com/bulik/ldsc); MAGMA (pathway analysis; https://ctg.cncr.nl/software/magma); fine-mapping (fine-mapping and PAINTOR; https://github.com/hailianghuang/FM-summary and https://github.com/gkichaev/PAINTOR_V3.0, respectively); rehh (selection; https://cran.r-project.org/web/packages/rehh/index.html); B score (background selection; http://www.phrap.org/othersoftware.html); and PRS analyses (https://github.com/armartin/pgc_scz_asia).

References

60. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* 4th edn (American Psychiatric Association, 2000).
61. Nurnberger, J. I. Jr. et al. Diagnostic interview for genetic studies: rationale, unique features, and training. NIMH Genetics Initiative. *Arch. Gen. Psychiatry* **51**, 849–859 (1994).
62. Sand, P. G. A lesson not learned: allele misassignment. *Behav. Brain Funct.* **3**, 65 (2007).
63. Lam, M. et al. RICOPILI: Rapid Imputation for COnsortias PIpeLIne. *Bioinformatics* https://doi.org/10.1093/bioinformatics/btz633 (2019).
64. O'Connell, J. et al. A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet.* **10**, e1004234 (2014).
65. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
66. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
67. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
68. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
69. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
70. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat. Genet.* **43**, 977–983 (2011).
71. Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium et al. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* **18**, 497–511 (2013).
72. Boraska, V. et al. A genome-wide association study of anorexia nervosa. *Mol. Psychiatry* **19**, 1085–1094 (2014).
73. Okbay, A. et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624–633 (2016).
74. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
75. Sniekers, S. et al. Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat. Genet.* **49**, 1107–1112 (2017).
76. Juric, I., Aeschbacher, S. & Coop, G. The strength of selection against Neanderthal introgression. *PLoS Genet.* **12**, e1006340 (2016).
77. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
78. Turner, T. N. et al. denovo-db: a compendium of human de novo variants. *Nucleic Acids Res.* **45**, D804–D811 (2017).
79. Pirooznia, M. et al. High-throughput sequencing of the synaptome in major depressive disorder. *Mol. Psychiatry* **21**, 650–655 (2016).
80. Gautier, M. & Vitalis, R. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* **28**, 1176–1177 (2012).
81. Sabeti, P. C. et al. Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
82. Weir, B. S. & Cockerham, C. C. Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
83. Grossman, S. R. et al. A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* **327**, 883–886 (2010).
84. Grossman, S. R. et al. Identifying recent adaptations in large-scale genomic data. *Cell* **152**, 703–713 (2013).
85. Ma, Y. et al. Properties of different selection signature statistics and a new strategy for combining them. *Heredity* **115**, 426–436 (2015).
86. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet.* **5**, e1000471 (2009).
87. Gormley, P. et al. Meta-analysis of 375,000 individuals identifies 38 susceptibility loci for migraine. *Nat. Genet.* **48**, 856–866 (2016).
88. Wellcome Trust Case Control Consortium et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44**, 1294–1301 (2012).
89. Lee, S. H., Goddard, M. E., Wray, N. R. & Visscher, P. M. A better coefficient of determination for genetic profile analysis. *Genet. Epidemiol.* **36**, 214–224 (2012).

Acknowledgements

We thank K. Kendler, J. McGrath, J. Walters, D. Levinson and M. Owen for helpful discussions. We thank S. Awathi, V. Trubetskoy and G. Panagiotaropoulou for support with the RICOPILI analysis pipeline. We thank SURFsara and Digital China Health for computing infrastructure for this study. M.L. acknowledges the National Medical Research Council Research Training Fellowship award (grant number: MH095:003/008-1014). A.R.M. acknowledges funding from K99MH117229. B.B. acknowledges funding support from the National Health and Medical Research Council (NHMRC; funding numbers 1084417 and 1079583). L.G. acknowledges the National Key Research and Development Program of the Ministry of Science and Technology of China (2016YFC1306802). M.I., Y.K. and N.I. acknowledge the Strategic Research Program for Brain Sciences (grant number JP19dm0107097) and Advanced Genome Research and Bioinformatics Study toFacilitate Medical Innovation (GRIFIN) of Platform Program for Promotion of Genome Medicine (P3GM) (grant numbers JP19km0405201 and JP19km0405208) from the Japan Agency for Medical Research and Development. Y.K., M.K. and A.T. acknowledge the BioBank Japan Project from the Ministry of Education, Culture, Sports, Science and Technology of Japan. S.-W.K. acknowledges a grant of the Korean Mental Health Technology Research and Development Project (HM15C1140). W.J.C. acknowledges the Ministry of Education, Taiwan ('Aim for the Top University Project' to National Taiwan University; 2011–2017), Ministry of Science and Technology, Taiwan (MOST 103-2325-B-002-025), National Health Research Institutes, Taiwan (NHRI-EX104-10432PI), NIH/NHGRI grant U54HG003067, NIMH grant R01 MH085521 and NIMH grant R01 MH085560. S.J.G. acknowledges R01 MH08552. B.J.M. acknowledges Australian NHMRC grant 496698. H.-G.H. acknowledges the Ministry of Education, Taiwan ('Aim for the Top University Project' to National Taiwan University; 2011–2017), MOST, Taiwan (MOST 103-2325-B-002-025), NIH/NHGRI grant U54HG003067, NIMH grants R01 MH085521 and R01 MH085560, and NHRI, Taiwan (NHRI-EX104-10432PI). P. Sklar acknowledges U01MH109536. B.J.M. acknowledges Australian NHMRC grant 496698. J. Lee acknowledges the National Medical Research Council Translational and Clinical Research Flagship Programme (grant number: NMRC/TCR/003/2008) and the National Medical Research Council under the Centre Grant Programme (grant number: NMRC/CG/004/2013). P.H. acknowledges funding support from the Medical Research Council (MR/L010305/1). S.X. acknowledges funding support from the Strategic Priority Research Program (XDB13040100) and Key Research Program of Frontier Sciences (QYZDJ-SSW-SYS009) of the Chinese Academy of Sciences, as well as the National Natural Science Foundation of China (NSFC; grants 91731303, 31525014 and 31771388), UK Royal Society–Newton Advanced Fellowship (NAF/R1/191094), Program of Shanghai Academic Research Leader (16XD1404700), National Key Research and Development Program (2016YFC0906403) and Shanghai Municipal Science and Technology Major Project (2017SHZDZX01). P.F.S. acknowledges PGC funding from U01 MH109528 and U01 MH1095320. M.J.D. acknowledges NIH/NIMH 5U01MH109539-02. M.C.O.'D. acknowledges funding from MRC (G0800509) and NIMH (reference: 1U01MH109514-01). S.Q. acknowledges the National Key Research and Development Program of China (2016YFC0905000 and 2016YFC0905002) and the Shanghai Key Laboratory of Psychotic Disorders (13dz2260500). K.S.H. acknowledges a grant from the National

Research Foundation of Korea (2015R1A2A2A01002699), funded by the Ministry of Science, ICT and Future Planning. D.B.W. and S.G.S. acknowledge funding support from the NHMRC (grant 513861). W.Y. acknowledges the National Key Research and Development Program of China (2016YFC1307000), NSFC (81571313) and the Peking University Clinical Scientist Program, supported by 'the Fundamental Research Funds for the Central Universities' (BMU2019LCKXJ012). M.T. acknowledges R01 MH085560 (Expanding Rapid Ascertainment Networks of Schizophrenia Families in Taiwan). J. Liu acknowledges funding support from the Agency for Science, Technology and Research, Singapore. Xiancang Ma acknowledges (and was principal investigator on) the National NSFC Surface project (81471374). Y.S. acknowledges the National Key Research and Development Program of China (2016YFC0903402), NSFC (31325014, 81130022 and 81421061) and 973 Program (2015CB559100). H.H. acknowledges support from NIH K01DK114379, NIH R21AI139012, the Zhengxu and Ying He Foundation and the Stanley Center for Psychiatric Research.

Author contributions

M.L. and H.H. performed the genotype quality control and PCA. M.L. and C.-Y.C. performed the association analysis. M.L. and B.C.B. performed the heritability and genetic correlation. Xixian Ma, C.-Y.C. and S.X. investigated the effects of natural selection. J.B. investigated the effects of partitioned heritability. Z.L., M.L. and H.G. performed the gene set analysis. A.R.M., C.-Y.C. and R.L. calculated the PRSs. H.H. performed the fine-mapping. P.H. performed the replication analysis. Data acquisition, generation, quality control and analysis were performed by: M.L., J. Lee and J. Liu (IMH-1 and IMH-2); P. Sham (HNK-1); A.T., Y.K., M.K., M.I. and N.I. (JPN-1); Z.L., L.H. and Y.S. (BIX-1, BIX-3 and BIX-5); W.Z., L.H., S.Q., F.Z. and Xiancang Ma (XJU-1 and BIX-4); L.G., H.M., Z.X., P. Sklar, X.Y., R.S.K. and the Genetic REsearch on schizophreniA neTwork-China and Netherlands (UMC-1 and SIX-1); B.B., A.K., D.B.W., S.G.S. and the Indonesia Schizophrenia Consortium (UWA-1); H.Y., D.Z. and W.Y. (BJM-1, BJM-2, BJM-3 and BJM-4); C.-M.L., W.J.C., S.F., S.J.G., H.-G.H., S.A.M., B.M.N. and M.T. (TAI-1 and TAI-2); S.-W.K. and K.S.H. (KOR-1); and the Schizophrenia Working Group of the Psychiatric Genomics Consortium (EUR samples). M.L., C.-Y.C., S.G.S., M.C.O'D., M.J.D. and H.H. drafted the primary manuscript, with major contributions from A.R.M., S.J.G., B.B., S.P., B.J.M., K.S.H., M.T., J. Lee, W.Y., H.-G.H., J.B., S.R. and P.F.S. H.H., Y.S., R.S.K., X.C.M., J. Liu., M.T., W.Y., S.G.S., K.S.H., N.I., P. Sham, S.Q., B.M.N., M.J.D., M.C.O'D., S.R., P. Sklar, L.H. and S.E.H. conceived, designed, supervised and implemented the project. All authors reviewed and approved the final draft.

Competing interests

B.M.N. is a member of the Deep Genomics Scientific Advisory Board. He also serves as a consultant for the Camp4 Therapeutics Corporation, Takeda Pharmaceutical and Biogen. M.J.D. is a founder of Maze Therapeutics. The remaining authors declare no competing interests.

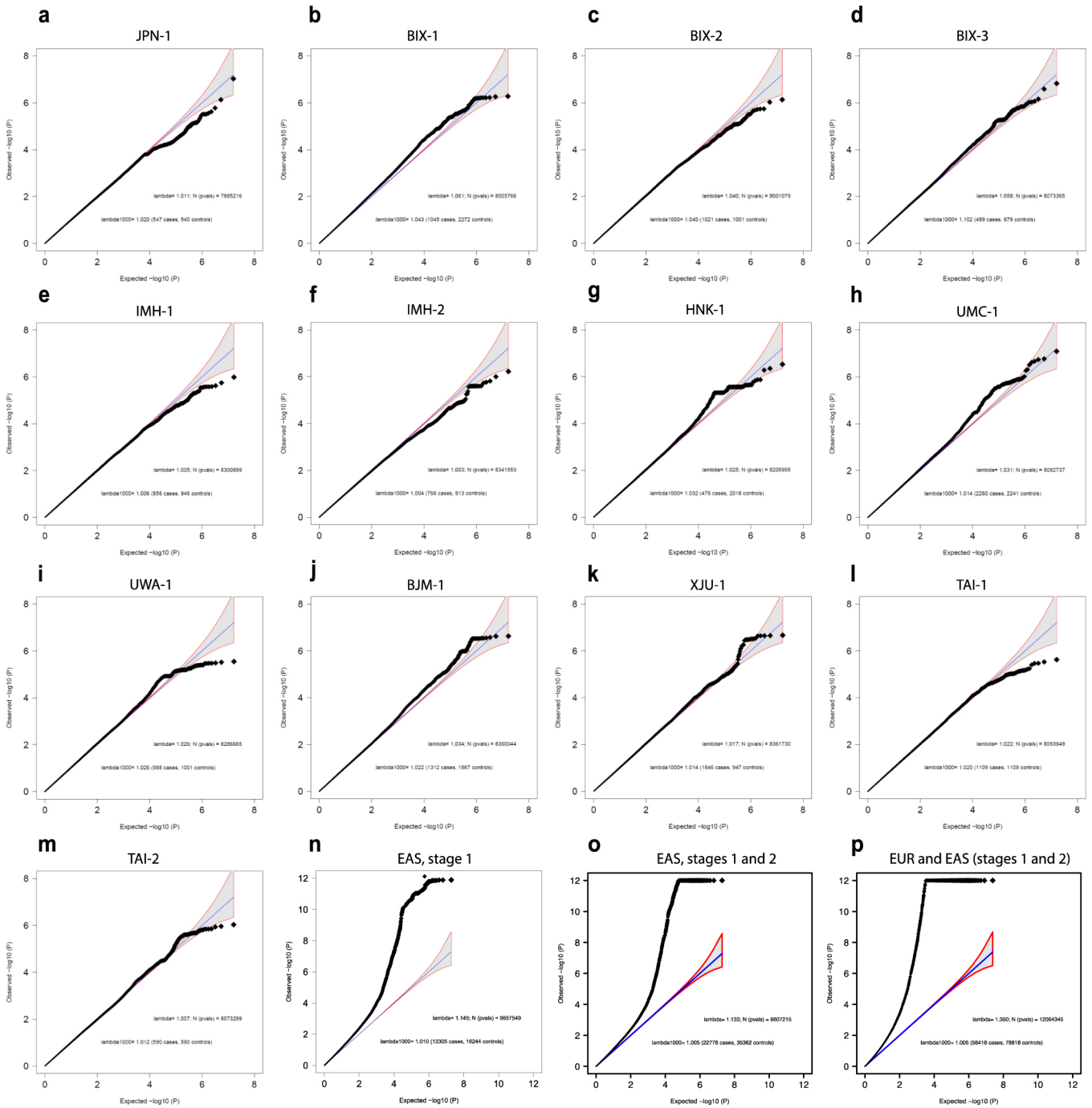
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-019-0512-x.

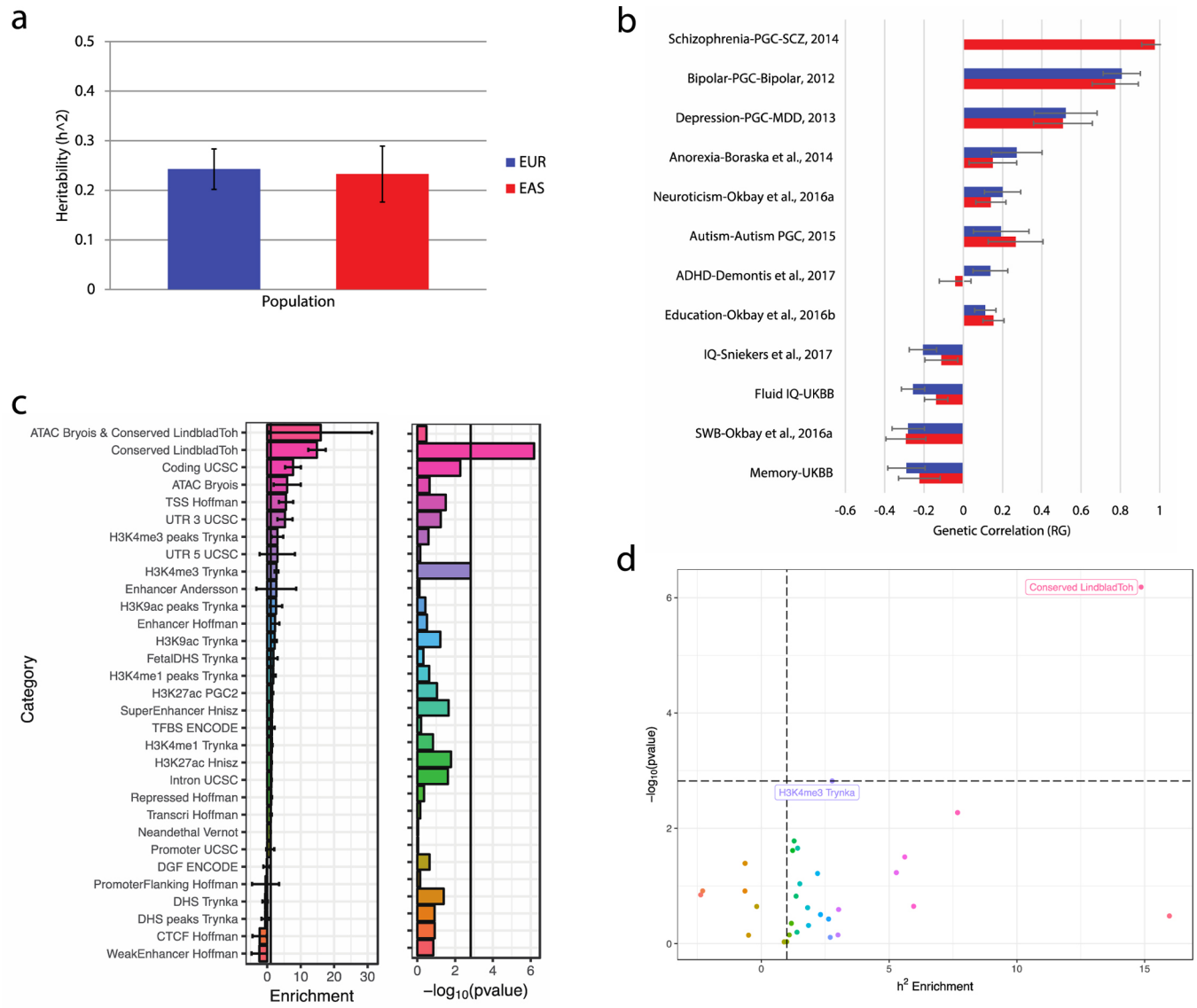
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0512-x.

Correspondence and requests for materials should be addressed to W.Y., M.T., J.L., X.M., R.S.K., Y.S. or H.H.

Reprints and permissions information is available at www.nature.com/reprints.

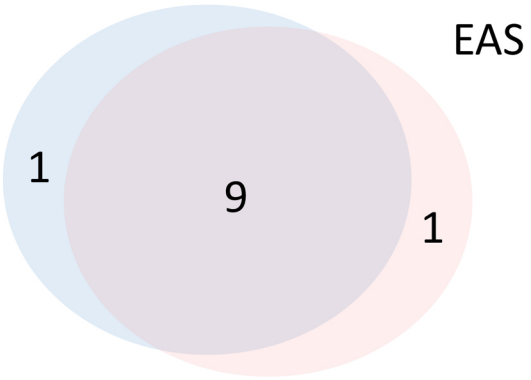


Extended Data Fig. 1 | Quantile-quantile (QQ) plots. a-p, QQ plots for two-tailed logistic regression in each EAS stage 1 sample (**a-m**) and fixed effect inverse variance meta-analyses including all EAS stage 1 samples (**n**), stages 1 and 2 samples (**o**), and all EUR and EAS (stages 1 and 2) samples (**p**). Blue line indicates the expected null distribution, and the shaded area indicates the 95% confidence interval of the null distribution. Legend: "lambda" = genomic inflation factor; "lambda1000" = genomic inflation factor for an equivalent study of 1,000 cases and 1,000 controls; "N(pvals)" = number of variants used in the plot. Autosomal variants that have minor allele frequency ≥ 1% and INFO ≥ 0.6 from imputation were included. Observed *P*-values were capped at 10^{-12} for visualization purposes.



Extended Data Fig. 2 | Heritability and genetic correlation. a, Heritability (h^2) for the EAS stage 1 ($n = 13,305$ cases; 16,244 controls) and EUR samples ($n = 33,640$ cases; 43,456 controls). Sample description applies to **b-d**. Error bars indicate the 95% confidence interval. **b**, Genetic correlation between schizophrenia and other traits within EUR (blue) and across EAS and EUR (red). Error bars indicate the 95% confidence interval. **c**, Enrichment and its corresponding significance for heritability partitioned based on various annotations. Error bars indicate the 95% confidence interval. **d**, Scatterplot showing the enrichment versus the significance for heritability partitioned based on various annotations. More details are available in Methods.

a

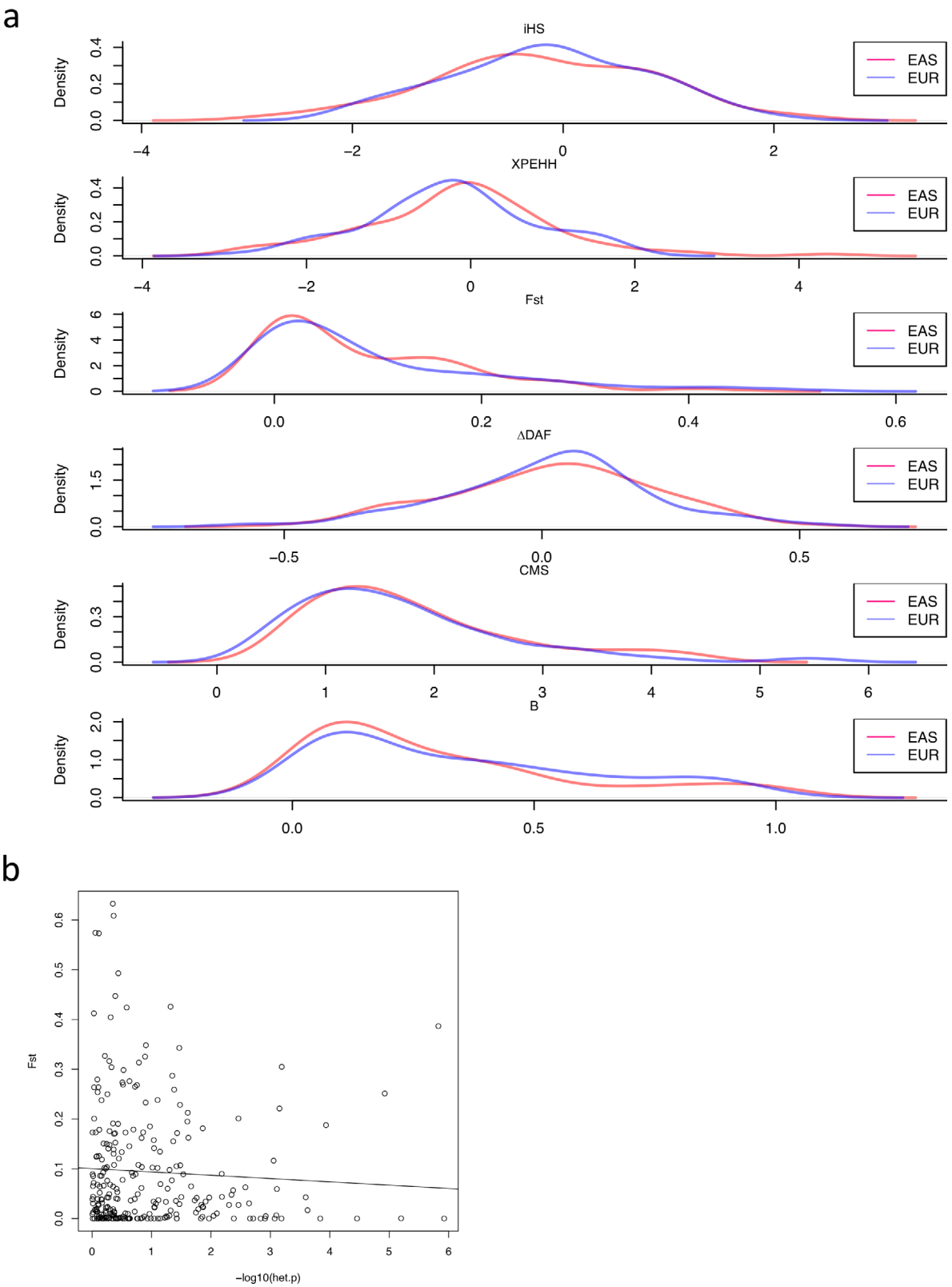


b

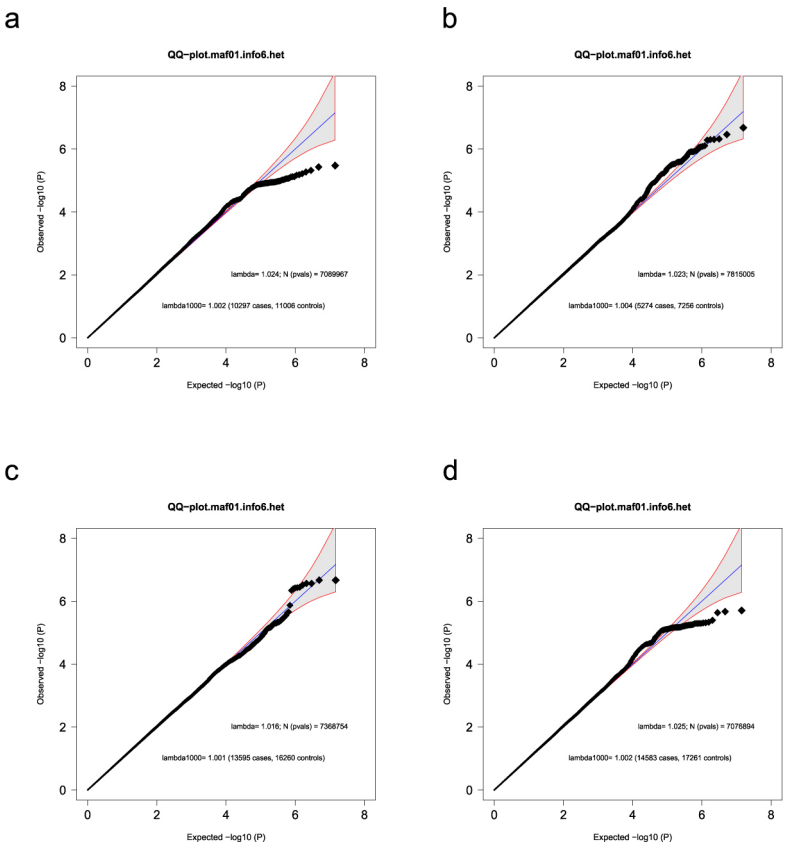
Top 10 EUR and EAS Pathways

			EAS+EUR	EUR	EAS
EAS ∩ EUR	9	PGC_SCZ_P10-4	101.39	112.54	6.05
		RBFOX1_RBFOX3	19.13	14.48	4.87
		POTENTIALLY_SYNAPTIC_ALL	17.20	11.90	4.44
		PLI09	14.60	11.75	3.96
		RBFOX2	14.09	12.45	3.37
		CHD8_HNSC	12.02	11.06	3.83
		FMRP	13.58	10.45	2.52
		CELF4	10.58	7.13	2.87
		CHD8_HNSC+HUMAN_BRAIN	7.44	6.83	2.32
EUR	1	CONSTRAINED	6.88	7.68	1.01
EAS	1	MIR-137	3.47	2.62	2.31

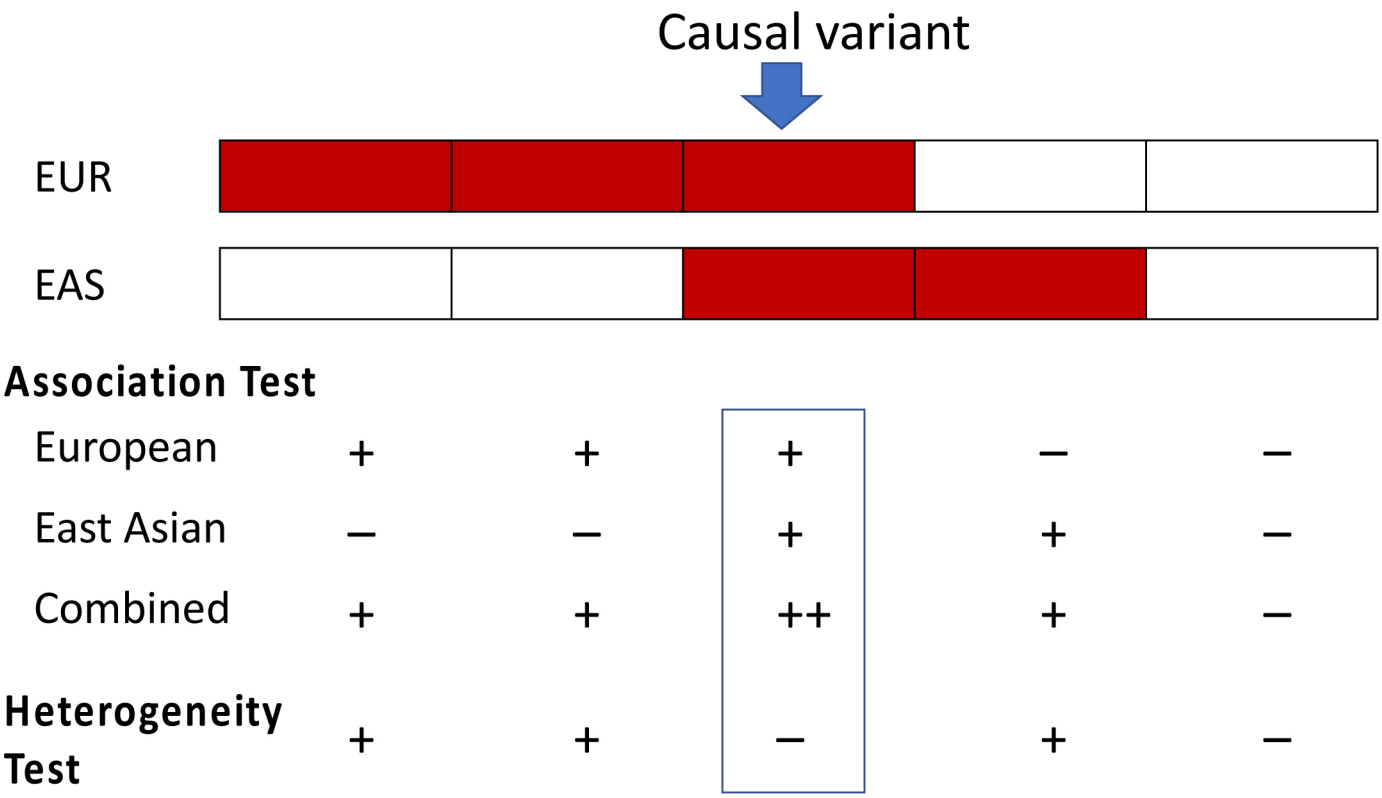
Extended Data Fig. 3 | Gene-sets implicated by schizophrenia genetic associations. a, Overlap of implicated gene-sets across EAS stage 1 ($n = 13,305$ cases; 16,244 controls) and EUR samples ($n = 33,640$ cases; 43,456 controls). **b**, List of the top 10 gene-sets implicated in the EAS and EUR samples and their MAGMA Gene-Set Analysis *P*-values in -$\log_{10}$ scale. Descriptions of the gene-sets are available in Supplementary Table 8.



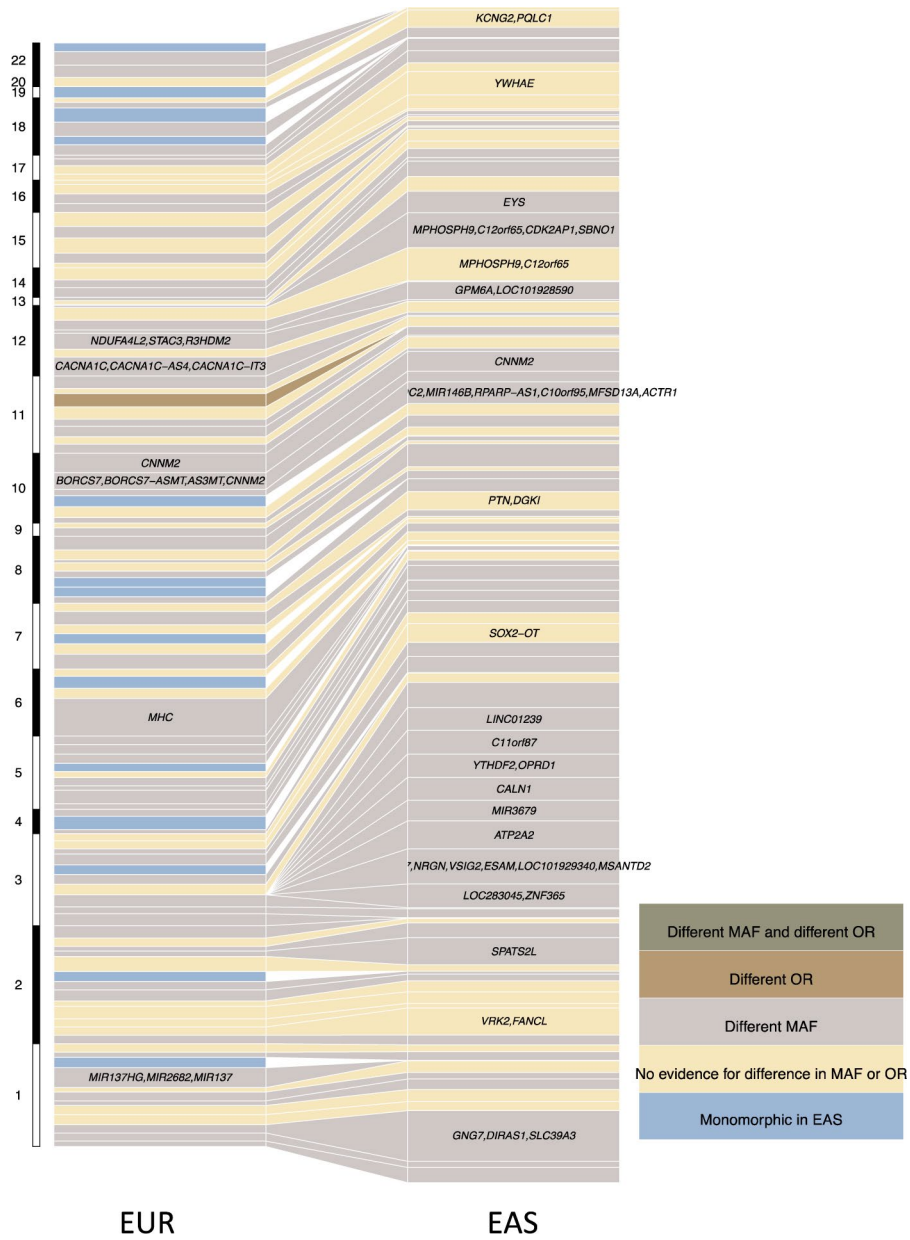
Extended Data Fig. 4 | Natural selection signals in EAS and EUR. a, Distributions of natural selection signals in the top 100 schizophrenia associations in EAS (red) and EUR (blue). **b**, Scatterplot of *Fst* versus the heterogeneity of effect size for schizophrenia associations. More details are available in Methods.



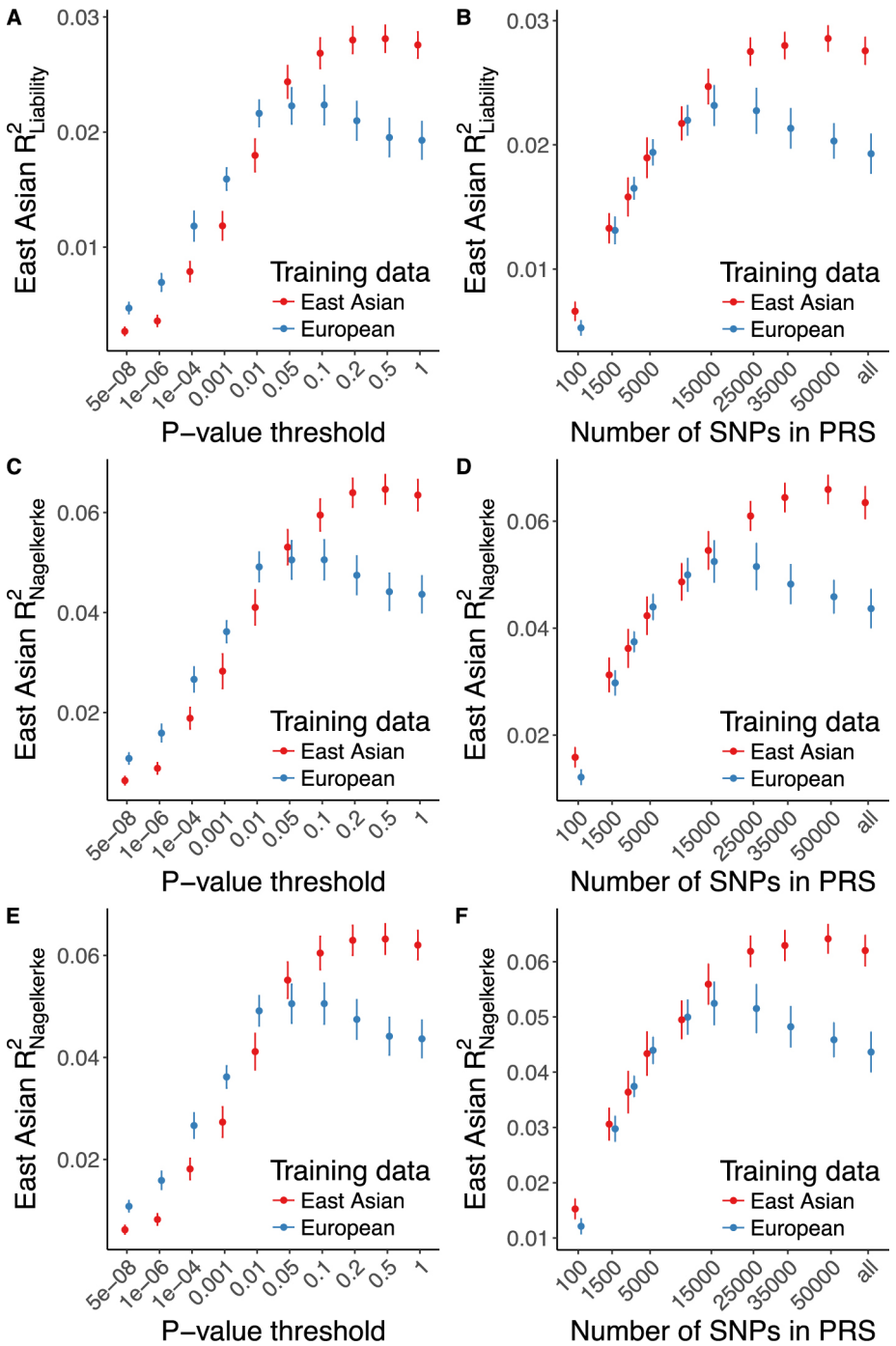
Extended Data Fig. 5 | Quantile-quantile (QQ) plots for heterogeneity within EAS. a, Heterogeneity QQ-plot across Northeast Asian and Indonesian samples. **b**, Heterogeneity across Southeast Asian and Indonesian samples. **c**, Heterogeneity QQ-plot across Northeast Asian and Southeast Asian samples. **d**, Heterogeneity QQ-plots across all three subpopulations. Cochran Q-test used to compute heterogeneity effects (**a-d**).



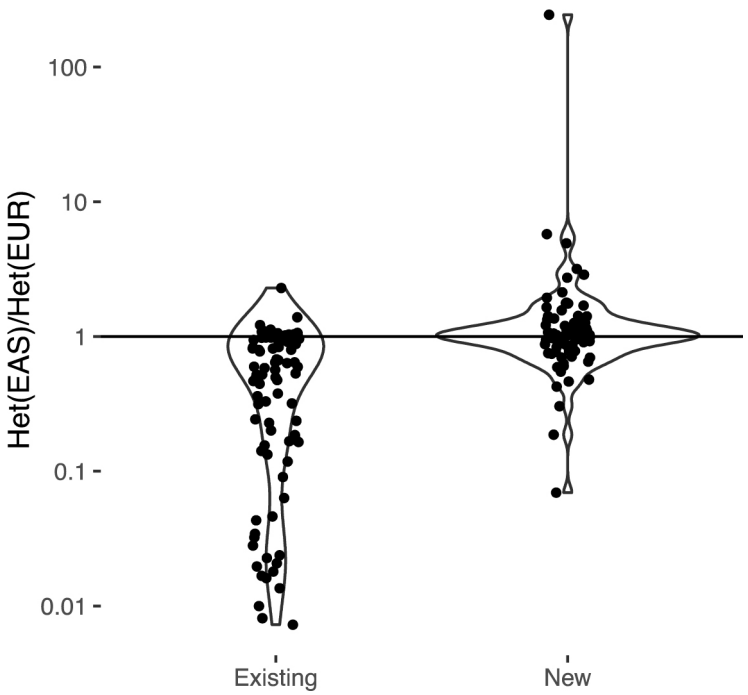
Extended Data Fig. 6 | Trans-ethnicity fine-mapping. Illustration of the fine-mapping method.



Extended Data Fig. 7 | Variance explained for schizophrenia associations across EUR and EAS samples. Genome-wide significant associations that have variance explained greater than 0.05% in either EAS or EUR samples were plotted. One locus can host multiple independent associations. Different MAF is defined as $Fst > 0.01$, and different OR is defined as heterogeneity test P-value < 0.05 after Bonferroni correction. Nearest genes to the associations were used as labels for associations when the text space is available, with the exception that the MHC locus was labeled as "MHC".



Extended Data Fig. 8 | Genetic risk prediction accuracy in EAS from EAS or EUR training data. As in Fig. 4, PRS shows case/control variance explained with EUR and EAS samples using a leave-one-out meta-analysis approach for the EAS samples. Error bars indicate the 95% confidence intervals. **a,b**, Liability-scale variance explained when LD panel for clumping is from EUR 1000 Genomes Phase 3 samples and best-guess genotypes are from each EAS cohort. **c,d**, Nagelkerke's R^2 for EAS prediction accuracy when LD panel for clumping is from EUR and EAS 1000 Genomes Phase 3 samples. **e,f**, Nagelkerke's R^2 for EAS prediction accuracy when LD panel for clumping is from EUR 1000 Genomes Phase 3 samples and best-guess genotypes are from each EAS cohort. **a-f**, EAS stage 1 (n = 13,305 cases; 16,244 controls) and EUR samples (n = 33,640 cases; 43,456 controls).



Extended Data Fig. 9 | Ratio of the heterozygote rate in EAS to that in EUR for existing and new loci. Het(EAS) and Het(EUR), calculated as $2f(1-f)$, are the heterozygote rates for a variant in EAS and EUR respectively, in which f is the variant allele frequency in EAS or EUR. Power to identify genetic associations increases with the expected non-centrality parameter for the association, which is proportional to the heterozygote rate. Therefore, we use the ratio of the heterozygote rate in EAS to that in EUR as a measure of the relative power to identify genetic association of the same effect size in the two populations. A ratio greater than 1 means that EAS samples have more power to identify the association and vice versa. Existing loci are those that are genome-wide significant in the previous study of European ancestry[2], and new loci are those that are genome-wide significant just in this study combining EAS and EUR samples. Sample sizes utilized were EAS stage 1 (n = 13,305 cases; 16,244 controls) and EUR samples (n = 33,640 cases; 43,456 controls).