Trans-ancestry genome-wide association meta-analyses of hippocampal and subfield volumes

Chunshui Yu (chunshuiyu@tmu.edu.cn)  
Tianjin Medical University General Hospital  https://orcid.org/0000-0001-5648-5199

Nana Liu  
Tianjin Medical University General Hospital

Longjiang Zhang  
Jinling Hospital, Medical School of Nanjing University

Tian Tian  
Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

Jingliang Cheng  
The First Affiliated Hospital of Zhengzhou University

Bing Zhang  
Drum Tower Hospital, Medical School of Nanjing University

Shijun Qiu  
The First Affiliated Hospital of Guangzhou University of Traditional Chinese Medicine

Zuojun Geng  
The Second Hospital of Hebei Medical University

Guangbin Cui  
Tangdu Hospital, Air Force Medical University

Quan Zhang  
Characteristic Medical Center of Chinese People's Armed Police Force

Weihua Liao  
Xiangya Hospital

Yongqiang Yu  
The First Affiliated Hospital of Anhui Medical University

Hui Zhang  
The First Hospital of Shanxi Medical University

Bo Gao  
The Affiliated Hospital of Guizhou Medical University

Xiaojun Xu  
The Second Affiliated Hospital of Zhejiang University, School of Medicine

Tong Han  
Tianjin Huanhu Hospital

Zhenwei Yao
Junfang Xian  
Beijing Tongren Hospital, Capital Medical University

Jia-Hong Gao  
Peking University

Xiaochu Zhang  
University of Science & Technology of China

Mulin Jun Li  
Tianjin Medical University

Kai Xu  
The Affiliated Hospital of Xuzhou Medical University

Xi-Nian Zuo  
Beijing Normal University  https://orcid.org/0000-0001-9110-585X

Meiyun Wang  
Henan Provincial People's Hospital & Zhengzhou University People's Hospital

Zhaoxiang Ye  
Tianjin Medical University Cancer Institute and Hospital
Abstract

The hippocampus is critical for memory and cognition and neuropsychiatric disorders, and its subfields differ in cytoarchitecture, function and disease vulnerability. The genetic architectures of hippocampal and subfield volumes have been studied in individuals of European ancestry; however, the results cannot explain for inter-ancestry differences in their associations with cognitive abilities and brain disorders. Here, we conduct trans-ancestry genome-wide association meta-analyses in 65,821 individuals for hippocampal volume and 39,007 for subfield volumes, including 7,039 individuals of East Asian ancestry. We identify 539 locus-trait associations at $P < 5 \times 10^{-8}$ for 44 hippocampal volumetric traits, including 161 novel associations and 41 novel loci. Trans-ancestry analysis improves the precision of fine-mapping mainly caused by the inter-ancestry difference in linkage disequilibrium and can enhance the prediction performance of polygenic scores in under-represented population by including hundreds of individuals from this population. Genetic variants associated with hippocampal and subfield volumes are enriched for neurogenesis and Wnt signaling and also affect cognition, emotion and neuropsychiatric disorders, such as Alzheimer’s disease, schizophrenia, and depression. Our results highlight the value of trans-ancestry analysis in the investigation of genetic architectures of human traits.

Introduction

The hippocampus embedded in the medial temporal lobe of the human brain is critical for episodic memory, spatial navigation and other higher-level cognitive functions, and particularly sensitive to various stressful events. The structural and functional impairments of the hippocampus have been observed in numerous neuropsychiatric disorders with different pathogenic mechanisms, such as Alzheimer’s disease (AD), Parkinson’s disease, schizophrenia, major depressive disorder (MDD), anxiety, addiction, and posttraumatic stress disorder. The hippocampus is a heterogeneous structure consisting of subfields with distinct cytoarchitecture and function and vulnerability to neuropsychiatric disorders. For instance, the vulnerable subfield is CA1 in schizophrenia and dentate gyrus in MDD. The subiculum and CA1 are more prominently affected by AD than the dentate gyrus and CA3.

As a recognized neuroimaging biomarker for AD, hippocampal volume calculated from structural magnetic resonance imaging (MRI) is highly heritable with an estimated twin-based heritability of > 75%, which motivates the investigation of the genetic architecture of hippocampal volume. Prior genome-wide association studies (GWASs) have identified a number of genetic variants for hippocampal and subfield volumes in individuals of European ancestry (EUR), but non-EUR populations, such as East Asian ancestry (EAS), are severely under-represented. The results found in EUR cannot explain for the inter-ancestry difference in the association of hippocampal volume with memory. The variant-level differences in allele frequency, linkage disequilibrium (LD), and effect size across populations also prevent us from extrapolating genetic associations identified in EUR to non-EUR populations. For example, the genetic risk variants in CREB1 shared by bipolar disorder and MDD are associated with...
hippocampal volume in EUR, but their risk alleles are completely absent in EAS\textsuperscript{26}. Moreover, the genetic associations of hippocampal volume identified in EUR cannot be replicated in EAS\textsuperscript{27}. These results highlight the need of conducting GWASs of hippocampal volume in ancestrally diverse populations to uncover the ancestry-shared and ancestry-specific genetic associations of hippocampal volume and to ensure the resulting GWAS findings to be broadly applicable\textsuperscript{28}. Furthermore, trans-ancestry GWASs have found to benefit for discovering new genetic associations, improving fine-mapping of causal variants, estimating polygenic score (PGS), and addressing inter-ancestry health disparities\textsuperscript{29–33}. However, no trans-ancestry GWASs have been conducted for hippocampal and subfield volumes so far mainly due to the lack of neuroimaging genetics data of non-EUR populations.

The Chinese Imaging Genetics (CHIMGEN)\textsuperscript{34} project has collected genomic and neuroimaging data from 7,306 healthy Chinese Han participants, which provides a unique opportunity to investigate the genetic architectures of hippocampal and subfield volumes using trans-ancestry GWAS meta-analyses by incorporating genetic data from EUR. In this study, we aim to clarify the following questions: (a) whether and why trans-ancestry analysis can identify novel associations; (b) whether and why trans-ancestry analysis can improve the precision of fine-mapping; and (c) whether and how trans-ancestry analysis can promote the prediction ability of PGS in under-represented populations. We are also interested in functional and clinical associations of the identified genetic variants associated with hippocampal volumetric traits.

**Results**

**Trans-ancestry genetic associations of hippocampal and subfield volumes**

In this study, we aimed to identify genetic loci associated with hippocampal and subfield volumes, including 44 traits symmetrically distributed in the left and right hippocampus (Fig. 1a). To obtain GWAS summary statistics of these traits in EAS, we carried out GWASs for the 44 hippocampal volumetric traits at 6,830,146 single nucleotide polymorphisms (SNPs) in 7,039 EAS participants (Supplementary Table 1) from CHIMGEN. We identified 21 significant locus-trait associations at $P < 5 \times 10^{-8}$ (Extended Data Fig. 1), including 12 novel associations not reported in EUR-specific studies (Supplementary Table 2). The obtained GWAS summary statistics of hippocampal and subfield volumes in EAS were used for the following trans-ancestry meta-analyses.

We first performed trans-ancestry GWAS meta-analyses for the left and right hippocampal volumes in 65,821 participants (EAS: 7,039 CHIMGEN participants; EUR: $n = 58,782, 31,968$ from UKBB and 26,814 from ENIGMA), and identified 77 locus-trait associations at $P < 5 \times 10^{-8}$, including 46 novel associations (Fig. 1b) that have not been reported in previous EUR-specific GWASs. Then we conducted trans-ancestry GWAS meta-analyses for the 42 hippocampal subfield volumes in 39,007 participants (EAS: $n = 7,039$
from CHIMGEN; EUR: n = 31,968 from UKBB), and identified 462 locus-trait associations at $P < 5 \times 10^{-8}$, including 115 novel associations (Fig. 1b).

In the trans-ancestry GWAS meta-analyses for all 44 phenotypes of hippocampal and subfield volumes, we identified 539 locus-trait associations (Extended Data Fig. 2, Supplementary Table 3) at $P < 5 \times 10^{-8}$, including 161 novel associations (Fig. 1b) not reported in EUR-specific GWASs. The 539 locus-trait associations belonged to 118 loci, and 59 loci were correlated with more than one trait. The 161 novel associations belonged to 83 loci, including 41 novel loci that have not been reported in EUR-specific GWASs. For example, the locus with rs62430477 as the lead SNP was one of the novel loci identified by trans-ancestry meta-analyses for the right CA4 body volume (Fig. 1d). Since 80 of the 539 locus-trait associations showed nominal allelic effect heterogeneity (Cochran's $Q$-test: $P < 0.05$) between EAS and EUR (Supplementary Table 3), we also validated these findings using Han and Eskin's random effects model, which models the inter-study heterogeneity better. We found that 79 lead SNPs of the 80 associations reached genome-wide significance ($P < 5 \times 10^{-8}$) and the remaining one was also close to the significant level ($P = 5.40 \times 10^{-8}$) (Supplementary Table 3), indicating that the results of trans-ancestry meta-analyses are reliable.

It is an open question whether trans-ancestry meta-analysis itself can provide novel biological insight since the identified new associations by trans-ancestry meta-analyses may be attributed to the increase of the sample size. To answer the question, we compared the results derived from trans-ancestry meta-analysis and those from EUR-specific GWAS with comparable sample size. For hippocampal volumes, we compared trans-ancestry GWAS meta-analyses in 55,135 participants (7,039 from CHIMGEN, 21,282 from UKBB discovery and 26,814 from ENIGMA) and EUR-specific GWAS meta-analyses in 58,782 participants (31,968 from UKBB and 26,814 from ENIGMA). For hippocampal subfield volumes, we compared trans-ancestry GWAS meta-analyses in 28,321 participants (7,039 from CHIMGEN and 21,282 from UKBB discovery) and EUR-specific GWASs in 31,968 participants from UKBB. We found 60 locus-trait associations ($P < 5 \times 10^{-8}$) (Fig. 1c) that were discovered only through trans-ancestry meta-analyses (Extended Data Fig. 3 and Supplementary Table 4), indicating that ancestrally diverse populations with even smaller sample size than a single population can yield unique discoveries possibly due to the difference in genetic architectures between populations. As an example, the locus with rs2274692 as the lead SNP for the left CA1 body volume was identified in trans-ancestry meta-analysis rather than in EUR-specific GWAS with larger sample size (Fig. 1e).

To determine potential factors accounting for the improved power of trans-ancestry meta-analysis, we compared the LD, allele frequencies and effect sizes of the lead SNPs of the 60 locus-trait associations between EAS (CHIMGEN, n = 7,039) and EUR (UKBB, n = 31,968). Overall, the 60 associations had larger effects (Wilcoxon rank-sum test: $P = 5.42 \times 10^{-4}$ for positive effects and $P = 7.39 \times 10^{-8}$ for negative effects, Fig. 1f and g) on hippocampal and subfield volumes in EAS than in EUR. The 60 associations contained 50 SNPs, whose minor allele frequencies (MAFs) had no significant difference (Wilcoxon signed rank test, $P = 0.367$) between EAS and EUR (Fig. 1h). EAS-specific and EUR-specific LD values were
measured by $r^2$ using their respective 1000 Genomes Phase 3 reference datasets. For each lead SNP of a given locus, we separately calculated the $r^2$ of the SNP to all other SNPs in the locus in EAS and EUR, and compared the $r^2$ values greater than 0.01 in both ancestral populations. Overall, the $r^2$ of the 50 lead SNPs had no significant difference between EAS and EUR (Wilcoxon rank-sum test: $P = 0.23$, Extended Data Fig. 4), and the relationships between LD and genomic distance of EAS and EUR are shown in Fig. 1i. In this study, we found 21 associations whose lead SNPs had a similar effect allele frequency (EAF, difference within 5%) between EUR and EAS, 19/21 had larger effects on the hippocampal and subfield volumes in EAS than those in EUR (grey spheres in Fig. 1f). For instance, rs823385 had almost the same EAF values ($EAF_{EAS} = 0.581$, $EAF_{EUR} = 0.558$) in EAS and EUR, but its effect size on the right hippocampal volume was twice larger in EAS (beta = 0.0597) than that in EUR (beta = 0.0290). In another example, rs12495978 had similar EAF values ($EAF_{EAS} = 0.193$, $EAF_{EUR} = 0.151$) in the two populations, it was much more strongly associated with the left GC-ML-DG head volume (beta = 0.102) in EAS than that (beta = 0.0351) in EUR. Of the 39 associations whose lead SNPs had different EAFs between EAS and EUR, 35 associations had larger effect sizes on hippocampal and subfield volumes in EAS than those in EUR (Fig. 1f). For example, the effect allele of rs2235933 was more common in EAS than in EUR ($EAF_{EUR} = 0.271$, $EAF_{EAS} = 0.578$), and its effect on the right presubiculum body volume was larger in EAS (beta = -0.0647) than that in EUR (beta = -0.0312). Although the effect allele of rs1518085 was more common in EUR than in EAS ($EAF_{EUR} = 0.851$, $EAF_{EAS} = 0.417$), its effect on the left subiculum head volume in EAS (beta = -0.0866) was still larger than that in EUR (beta = -0.0401). Taken together, these findings indicate that the difference in effect size between ancestral populations, especially the relatively greater effect size in the under-represented population, is an important cause for novel discoveries in trans-ancestry meta-analysis.

**Ancestry-shared and ancestry-specific genetic associations of hippocampal and subfield volumes**

Since no large-scale GWAS has been conducted for hippocampal and subfield volumes in non-EUR populations, the ancestry-shared and ancestry-specific genetic associations of hippocampal and subfield volumes between EUR and EAS remain unknown. To compare the SNP’s effect sizes on hippocampal and subfield volumes between EUR and EAS, we chose the SNPs with genome-wide significance ($P < 5 \times 10^{-8}$) in any of the four GWASs including trans-ancestry meta-analyses (n = 65,821), CHIMGEN (n = 7,039), UKBB (n = 31,968), or ENIGMA (n = 26,814) for hippocampal volume, and those in any of the three GWASs including trans-ancestry meta-analyses (n = 39,007), CHIMGEN (n = 7,039) or UKBB (n = 31,968) for hippocampal subfield volumes. After excluding duplicate associations and those with loci genotyped only in one population, we obtained 581 locus-trait associations (Supplementary Table 5) and compared the direction of effect of the lead SNPs between EAS (7,039 CHIMGEN participants) and EUR (31,968 UKBB participants). We observed 86.92% of these associations (505/581) with consistent direction of effect between EAS and EUR (Fig. 2a). In the tests for allelic effect heterogeneity, we found 462/581 (79.52%) ancestry-shared genetic associations with no evidence for allelic effect heterogeneity (Cochran's
Q-test: $P \geq 0.05$) between EAS and EUR (Fig. 2b). For instance, two loci with rs2363340 (chr10) and rs6496265 (chr15) as the lead SNPs showed significant effects on several phenotypes of hippocampal and subfield volumes in both EAS and EUR, such as the right hippocampal tail (Fig. 2c left) and left hippocampal body volumes (Fig. 2c right). We also found 119/581 (20.48%) potential ancestry-specific genetic associations with nominally significant (Cochran's Q-test: $P < 0.05$) in the allelic effect heterogeneity tests (Fig. 2b). Using a threshold of $P < 8.61 \times 10^{-5}$ to additionally correct for 581 associations, we found 23/581 (3.96%) reliable ancestry-specific genetic associations (Fig. 2b). The distribution of different types of genetic associations across the 44 traits of hippocampal and subfield volumes is shown in Fig. 2e. The left CA4 head volume had the largest number ($n = 3$) and ratio (23.08%) of reliable ancestry-specific genetic associations (Fig. 2e). For example, the locus with rs7315280 as the lead SNP, a previously identified locus for hippocampal volumetric traits in chromosome 12 in EUR-specific GWAS, showed strong allelic effect heterogeneity on the right hippocampal tail volume between EAS and EUR (Cochran's Q-test: $P = 2.14 \times 10^{-8}$). The locus affected the trait only in EUR (Fig. 2d left). In addition, there were loci that had a greater effect on hippocampal traits in EAS than in EUR, such as the locus with rs6463429 as the lead SNP, showed an ancestry-specific genetic association for left GC-ML-DG head volume (Cochran's Q-test: $P = 1.06 \times 10^{-6}$; Fig. 2d right), left CA4 head volume (Cochran's Q-test: $P = 1.48 \times 10^{-6}$) and left hippocampal head (Cochran's Q-test: $P = 4.35 \times 10^{-6}$). This was a novel locus identified in EAS for the three hippocampal traits, however, the locus had no impact for these traits in EUR.

Trans-ancestry analysis improves the resolution of fine-mapping

To test whether the resolution of fine-mapping based on trans-ancestry GWAS summary statistics is higher than that based on ancestry-specific GWAS summary statistics, we compared the fine-mapping results derived from trans-ancestry meta-analyses in 28,321 participants (7,039 from CHIMGEN and 21,282 from UKBB discovery) and those derived from EUR-specific GWASs in 21,282 individuals from UKBB discovery. After excluding 36 loci not present in CHIMGEN genetic data, the remaining 231 genome-wide significant locus-trait associations identified in UKBB discovery were used for comparing the fine-mapping performance based on GWAS summary statistics from trans-ancestry and single-ancestry analyses. For each locus, the range ($\pm$ 500kb) and SNPs were kept consistent between trans-ancestry and ancestry-specific datasets, and fine-mapping was performed using PAINTOR$^{37}$. The fine-mapping results are demonstrated in Fig. 3a and Supplementary Table 6. The fine-mapping resolution was improved in 121/231 (52.38%) locus-trait associations, and the number of the most precise fine-mapping (one variant in 95% credible set) was increased from 2 in single-ancestry analysis to 10 in trans-ancestry analysis. Since the fine-mapping resolution is increased with the sample size in GWASs$^{31}$, we also compared the fine-mapping results derived from trans-ancestry meta-analyses in 28,321 participants (7,039 from CHIMGEN and 21,282 from UKBB discovery) and those derived from single-ancestry GWASs in 31,968 EUR individuals from UKBB discovery ($n = 21,282$) and replication ($n = 10,686$). In the 121 locus-trait
associations with improved fine-mapping resolution in trans-ancestry analysis, the fine-mapping resolution of 82 loci (67.8%) was still higher in trans-ancestry analysis than in single-ancestry analysis with 12.9% larger sample size (Fig. 3a). Moreover, the number of the most precise fine-mapping was still 2 in single-ancestry analysis with a greater sample size, which was much smaller than 10 in trans-ancestry analysis. These findings indicate that the improved resolution of fine-mapping in trans-ancestry analysis cannot be simply explained by the increase in sample size. One example is the locus with rs1370938 as the lead SNP (Fig. 3b), the number of causal SNPs in 95% credible set shifted from three in EUR-specific analysis to one in trans-ancestry fine-mapping. The improved resolution of fine-mapping seems to be related to the difference in LD pattern of the locus between EAS and EUR (Fig. 3c), rs1370938 was highly linked to the other two SNPs included in 95% credible set in EUR, but it was not correlated with these two SNPs in EAS. In addition, the fine-mapping results from PAINTOR were further confirmed by approximate Bayesian approach. The sizes (the number of SNPs) of 95% credible sets were significantly correlated between the two fine-mapping methods (Pearson's r = 0.995), and the 95% credible sets with one causal SNP were identical (Supplementary Table 6).

To further explore which genetic features are associated with the improved fine-mapping in trans-ancestry analysis, we compared the inter-ancestry differences in LD, allele frequencies, and effect sizes of the peak SNPs of 82 locus-trait associations with improved fine-mapping resolution in trans-ancestry analysis relative to EUR-specific analyses. The 82 locus-trait associations contained 47 distinct lead SNPs, and their LD ($r^2$) values were calculated according to the above-mentioned approach. Overall, the $r^2$ values of 47 lead SNPs were much higher (Wilcoxon rank-sum test: $P = 3.78 \times 10^{-133}$) in EUR than in EAS (Fig. 3d). We further compared the $r^2$ grouped by distance (50kb as interval) and found consistently higher LD in EUR than in EAS within the range of 500kb (Fig. 3e). Despite the MAFs of these lead SNPs in EAS were less frequent (Wilcoxon signed rank test: $P = 2.51 \times 10^{-4}$) than that in EUR (Fig. 3f), their effect sizes did not show significant difference between trans-ancestry and EUR-specific analyses (Wilcoxon rank-sum test, $P = 0.88$) (Fig. 3g). These findings suggest that the LD difference between ancestries is a major contributor for the improved fine-mapping resolution in trans-ancestry analysis, and the contribution of the MAF difference needs to be further confirmed due to the moderate significance and large within-ancestry variance.

**Transferability for PGS of trans-ancestry meta-analysis**

So far, all large-scale GWASs for hippocampal and subfield volumes have been conducted in EUR individuals; however, the PGSs constructed only based on EUR individuals tend to underperform when predicting for individuals from non-EUR populations. In this study, we systematically assessed the applicability of PGSs constructed based on different combinations of individuals with distinct ancestries to predict traits of individuals from the under-represented population (EAS). In these analyses, CHIMGEN participants were divided into two groups according to the types of MRI scanners: 5,023 participants (CHIMGEN discovery) whose MRI data were acquired by the same type of scanner (GE MR750) and
parameters were regarded as the base dataset in the PGS analysis; and the remaining 2,016 participants whose MRI data were acquired by other nine types of MR scanners were regarded as the target dataset in this analysis (Extended Data Fig. 5). We compared the prediction performance of PGS constructed based on three types of GWAS summary statistics: (a) trans-ancestry meta-analysis in 26,305 participants (5,023 from CHIMGEN discovery and 21,282 from UKBB discovery); (b) EUR-specific GWAS in 31,968 participants from UKBB discovery (n = 21,282) and replication (n = 10,686); and (c) EAS-specific GWAS in 5,023 participants from CHIMGEN discovery. Because we were interested in the prediction performance of PGSs for traits of individuals from the under-represented population, we only included 2,016 EAS participants from CHIMGEN as the target dataset in all PGS analyses (Supplementary Table 7). We found that the prediction performance ($R^2_{\text{average}} = 0.011$) of PGSs derived from the trans-ancestry meta-analysis outperformed than those derived from EAS-specific GWAS ($R^2_{\text{average}} = 0.0056$; Wilcoxon rank-sum test: $P = 4.08 \times 10^{-11}$) and EUR-specific GWAS ($R^2_{\text{average}} = 0.0069$; Wilcoxon rank-sum test: $P = 6.09 \times 10^{-8}$).

The trans-ancestry meta-analysis could improve the predictive performance in 42/44 traits relative to EAS-specific GWAS and in 40/44 traits than EUR-specific GWAS (Fig. 4a and b). The improved prediction performance of PGSs derived from trans-ancestry meta-analysis cannot simply attribute to the sample size because the sample size of the EUR-specific GWAS (n = 31,968) was 21.5% greater than that of the trans-ancestry meta-analysis (n = 26,305).

Consistent with previous studies, we found that PGSs derived from the same ancestral population (EAS, 5,023 from CHIMGEN discovery) could achieve similar prediction performance (Wilcoxon rank-sum test of $R^2$, $P = 0.46$) relative to that from different ancestral population with much larger sample size (EUR, 21,282 from UKBB discovery) (Fig. 4b). We also found that PGS derived from EUR (n = 31,968) performed even better (Wilcoxon rank-sum test of $R^2$, $P = 0.012$) than that derived from the same under-represented population (EAS, n = 5,023) (Fig. 4b), indicating that PGS derived from large-sample EUR population can be used for predicting traits of individuals from the under-represented population when there was no GWAS summary statistics of the population with enough large sample size.

We further assessed the prediction performance of the PGSs derived from the trans-ancestry meta-analyses that included different numbers of EAS participants in the base datasets, which were extracted according to random sampling within the 5,023 participants from CHIMGEN discovery, and found that the inclusion of limited sample size of EAS participants can improve the prediction performance of PGS (Fig. 4c). When we constructed PGS based on trans-ancestry GWAS results of 500 EAS participants from CHIMGEN discovery and 21,282 EUR participants from UKBB discovery, its prediction performance was improved compared with that of PGS constructed using UKBB discovery participants (Wilcoxon rank-sum test of $R^2$, $P = 0.02$) and was comparable with the prediction performance of PGS constructed using all UKBB participants (n = 31,968) (Wilcoxon rank-sum test of $R^2$, $P = 0.42$), which was better than that constructed using UKBB discovery due to the increased sample size (Wilcoxon rank-sum test of $R^2$, $P = 0.0031$). Compared with PGSs constructed using all UKBB participants (n = 31,968), the prediction performance was improved by constructing PGS using 3,000 EAS from CHIMGEN discovery and 21,282...
EUR participants from UKBB discovery (Wilcoxon rank-sum test of $R^2$, $P=3.55 \times 10^{-4}$) (Fig. 4d). Taken together, these results suggest that the construction of PGSs based on trans-ancestry GWAS results could improve the prediction performance for traits of individuals from the under-represented population by adding a moderate sample size of individuals from the population.

**Functional annotation of genetic loci associated with hippocampal and subfield volumes**

To understand functional consequences of genetic loci associated with hippocampal and subfield volumes, we performed functional annotation of significant SNPs ($P<5 \times 10^{-8}$) derived from trans-ancestry analysis using an integrative web-based platform (FUMA)\textsuperscript{42}. Gene functional consequences of these SNPs were investigated using ANNOVAR\textsuperscript{43} and combined annotation-dependent depletion (CADD) scores\textsuperscript{44}, and effects of SNPs on gene expression were explored by expression quantitative trait loci (eQTL) mapping of hippocampal tissue from GTEx\textsuperscript{45}. Of the 4,844 SNPs, 4,833 were annotated by ANNOVAR, and they were mainly located in intronic (56.76%) and intergenic (27.21%) areas (Fig. 5a). We also identified 20 SNPs of non-synonymous mutations in exons (Supplementary Table 8) including four novel mutations in genes ($NR1H3$, rs2279238; $MADD$, rs1051006 and rs2290148; $APOE$, rs429358). $NR1H3$ encodes liver X receptor alpha and is linked to neurogenesis in the hippocampus\textsuperscript{46} and AD\textsuperscript{47}. The $MADD$ expression reduces in the hippocampus with AD pathology, which contributes to neuronal cell death\textsuperscript{48}. The rs429358 (CADD score = 12.64) is a missense mutation in $APOE$, which has been associated with hippocampal volume\textsuperscript{49} and AD\textsuperscript{50}. The rs4076700 (a missense mutation in $FBXW8$) had the highest CADD score (26.8), and $FBXW8$ knockdown can lead to abnormal morphology of Golgi apparatus and reduced dendrite length in hippocampal neurons\textsuperscript{51}.

In the hippocampal tissue, 835 distinct SNPs were eQTLs of 21 protein-coding genes (1,412 SNP-gene pairs) at a 5% false discovery rate (FDR) threshold (Fig. 5b and Supplementary Table 9). The rs11158037 located in a novel locus for the left CA1 body volume ($P=1.40 \times 10^{-8}$, beta = 0.043) was an eQTL of $ATG14$ ($P=1.11 \times 10^{-7}$, normalized effect size = 0.29), an autophagy-specific regulator\textsuperscript{52} associated with neural differentiation, axon growth, and neuronal signaling\textsuperscript{53}. The rs2311528 located in another novel locus for the right presubiculum body volume ($P=3.83 \times 10^{-8}$, beta = 0.043) was correlated with the expression of $CROCC$ ($P=4.07 \times 10^{-7}$, normalized effect size = 0.29), an encoding gene of the major structural component of the ciliary rootlet and being linked to brain development\textsuperscript{54}. Ancestry-shared significant SNPs in chr10:126427037–126562724 for hippocampal and subfield volumes were eQTLs of $LHPP$ ($P=1.21 \times 10^{-6}$, normalized effect size = 0.19) and $METTL10$ ($P=2.41 \times 10^{-11}$, normalized effect size = -0.42). The former is a risk gene of MDD\textsuperscript{55} and the latter is related to the process of translation\textsuperscript{56}.

All 118 loci (4,844 SNPs) were mapped to 166 protein-coding genes based on location, among which the 41 novel loci were aligned to 49 protein-coding genes (Fig. 5b). We performed the statistical over-
representation analysis with WebGestalt online tool to identify enriched gene ontology (GO) terms of biological process for these genes. The 166 genes were enriched for 71 GO biological process terms ($q_c < 0.05$, Benjamini-Hochberg FDR corrected) (Fig. 5c and Supplementary Table 10), mainly for the neuron-related biological processes, such as neurogenesis ($P = 1.13 \times 10^{-4}$), neuron development ($P = 3.20 \times 10^{-5}$) and differentiation ($P = 3.59 \times 10^{-5}$), axonogenesis ($P = 4.39 \times 10^{-5}$), and axon extension ($P = 7.50 \times 10^{-5}$) and guidance ($P = 1.42 \times 10^{-4}$). Of these over-represented genes, nine (RAP1A, PTCH1, STRN, SPTBN1, RIPOR2, VEGFA, SFRP1, PARD3, and APOE) were mapped by the novel loci. We also found significant enrichment for Wnt signaling (e.g. negative regulation of the Wnt signaling, $P = 1.37 \times 10^{-6}$), and the enriched genes included those (STRN, IGFBP2, SFRP1, WWOX, and APOE) mapped by the novel loci.

**Phenome-wide association studies of genetic loci associated with hippocampal and subfield volumes**

We queried the 4,844 genome-wide significant ($P < 5 \times 10^{-8}$) SNPs associated with hippocampal and subfield volumes in GWAS Catalog by FUMA. We only focused on brain-related traits and neuropsychiatric disorders. We found that 66 SNPs from 19 genomic regions associated with hippocampal and subfield volumes also affected the focused phenotypes at $P < 5 \times 10^{-8}$, including neuropsychiatric disorders (AD and schizophrenia), cognitive-related traits (educational attainment, math ability, cognitive performance, verbal declarative memory, intelligence, and reaction time), mental health (neuroticism, worry, depressive symptoms, life satisfaction, and general risk tolerance), and smoking/drinking (smoking and alcohol consumption) (Fig. 5d and Supplementary Table 11). For example, 13 SNPs in 2q24.2 (SLC4A10 and DPP4) associated with hippocampal volume and 30 subfield volumes were also correlated with cognitive-related traits and smoking/drinking traits. SLC4A10 encodes a Na$^+$-coupled HCO$_3^-$ exchanger, which is abundant in CA3 pyramidal cells and regulates neuronal excitability and synaptic short-term plasticity. DPP4 is highly expressed in the hippocampus and is correlated with cognitive impairment. Another example, two SNPs located in 10q26.13 associated with hippocampal volume and 22 subfield volumes that regulate METTL10 expression in the hippocampus were also correlated with cognitive-related traits. The rs73405293 (chr12) marked the most significant locus for hippocampal volume was also associated with educational attainment. This SNP is located in the intron of TESC encoding tescalcin with a neuroprotective effect. In the novel loci identified by trans-ancestry analysis, three SNPs (PTCH1, a gene correlated with neurogenesis in the hippocampus) associated with the left CA1 body volume were also correlated with cognitive-related traits. Notably, we found that 5 SNPs located in 17q21.31 associated with four hippocampal subfield volumes but not with the whole hippocampal volume were also correlated with cognitive-related traits and mental health, indicating that the subfield-specific genetic architecture may contribute to the inter-subfield differences of the hippocampus in cytoarchitecture and function and vulnerability to brain disorders.
Conclusions

In this study, we conducted the first trans-ancestry GWAS meta-analyses for 44 traits of hippocampal and subfield volumes in up to 65,821 individuals. The results improve our understanding of the genetic architecture of the whole and regional hippocampal volumes by providing 161 novel associations and 41 novel loci. Furthermore, we proved that trans-ancestry analysis with even smaller samples relative to ancestry-specific analysis could still yield new genetic discoveries, which mainly benefited from the relatively greater effect size in under-represented population rather than the inter-ancestry difference in allele frequency or LD. Taking together, these results highlight the value of a more global representation of populations in genetic studies. Despite most of the identified genetic associations with hippocampal volumetric traits were shared by ancestries, we also found ancestry-specific genetic associations, which may account for the inter-ancestry discrepancy in hippocampus-related findings.

In addition to confirming the previously identified advantages of trans-ancestry analysis in fine-mapping and genetic risk prediction\textsuperscript{29,30}, we provided additional insight into these points. In terms of fine-mapping, we found that the LD difference between ancestries contributed more to the improved resolution of fine-mapping by trans-ancestry analysis than the inter-ancestry difference in allele frequency and effect size. In genetic risk prediction, we proved that the prediction performance could be significantly improved by including only hundreds of individuals from the under-represented population in trans-ancestry analysis. This finding is exciting in the field of genetic risk prediction because small sample-size genetic data of the under-represented population are already available for a large number of human traits and disorders.

Beyond the discovery of novel genetic associations for hippocampal and subfield volumes, we provided new biological insight into the relation of hippocampus with neuropsychiatric disorders. On the one hand, we provided the first evidence for the effects of several AD-related genes (\textit{NR1H3}, \textit{MADD} and \textit{APOE}) on the hippocampus based on the discovered associations of non-synonymous mutations in exons of these genes with hippocampal and subfield volumes. On the other hand, we identified two inter-related biological convergence points (neurogenesis and Wnt signaling) between hippocampus and neuropsychiatric disorders since they were main enrichment terms for genetic loci associated with hippocampal and subfield volumes and critical mechanisms in neuropsychiatric disorders including AD, MDD, schizophrenia, and autism spectrum disorder\textsuperscript{65}. The shared biological substrates support hippocampal and subfield volumes to be used as biomarkers for these disorders, and novel agents targeting the aberrant Wnt signaling in the hippocampus may benefit for patients with neuropsychiatric disorders with similar abnormality in the signal pathway.

Declarations

Funding

This work was supported by the National Key Research and Development Program of China (2018YFC1314300) to C.Y., the National Natural Science Foundation of China to CY (82030053,
The authors report no biomedical financial interests or potential conflicts of interest. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests

The authors have declared that no competing interests exist.

Acknowledgements

We are grateful to the UKBB and ENIGMA for providing GWAS summary statistics of hippocampal and subfield volumes.

Author contributions

Chunshui Yu and Nana Liu designed the study and wrote the manuscript. All authors critically reviewed the manuscript. Nana Liu, Longjiang Zhang, Tian Tian, Jingliang Cheng, Bing Zhang, Shijun Qiu, Mulin Jun Li, Kai Xu, Xi-Nian Zuo, Meiyun Wang, Zhaoxiang Ye, Wen Qin, Feng Liu, Meng Liang, Qiang Xu, Jilian Fu, Jiayuan Xu and Chunshui Yu were the principal investigators. Longjiang Zhang, Tian Tian, Jingliang Cheng, Bing Zhang, Shijun Qiu, Zuojun Geng, Guangbin Cui, Quan Zhang, Weihua Liao, Yongqiang Yu, Hui Zhang, Bo Gao, Xiaojun Xu, Tong Han, Zhenwei Yao, Wenzhen Zhu, Peng Zhang, Wei Li, Dapeng Shi, Caihong Wang, Su Lui, Zhihan Yan, Feng Chen, Jiance Li, Jing Zhang, Dawei Wang, Wen Shen, Yanwei Miao, Junfang Xian, Jia-Hong Gao, Xiaochu Zhang, Kai Xu, Xi-Nian Zuo, Meiyun Wang, Zhaoxiang Ye and Chunshui Yu acquired the data. Chunshui Yu, Zhaoxiang Ye and Meiyun Wang supervised this work. Nana Liu analyzed the data.

Additional information

Supplementary Information is available for this paper.

Correspondence and requests for materials should be addressed to Chunshui Yu, Zhaoxiang Ye and Meiyun Wang.

Data availability

The raw brain imaging and genotype data are available from the authors upon reasonable request and with permissions of the CHIMGEN consortia. The GWAS summary statistics used in this work from two publicly available datasets: the UKBB study (https://open.win.ox.ac.uk/ukbiobank/big40/) and the ENIGMA study (http://enigma.ini.usc.edu/). All trans-ancestry meta-analysis GWAS summary statistics of 44 hippocampal and subfield volumes can be found at http://www.mulinlab.org/Hipp_img_TA_meta/.

Code availability
We made use of publicly available software and tools. All codes used to generate results that are reported in this paper are available upon request.

The Full List of CHIMGEN Consortium Authors

Department of Radiology and Tianjin Key Laboratory of Functional Imaging, Tianjin Medical University General Hospital

Chunshui Yu; Quan Zhang; Wen Qin; Feng Liu; Junping Wang; Qiang Xu; Jiayuan Xu; Xue Zhang; Xinjun Suo; Jilian Fu; Congcong Yuan; Yuan Ji; Hui Xue; Tianying Gao; Junpeng Liu; Yanjun Li; Xi Guo; Lixue Xu; Jiajia Zhu; Huaimgui Liu; Fangshi Zhao; Jie Sun; Yongjie Xu; Huanhuan Cai; Jie Tang; Yaodan Zhang; Yongqin Xiong; Xianting Sun; Nannan Pan; Xue Zhang (Junior); Jiayang Yang; Nana Liu; Ya Wen; Dan Zhu; Bingjie Wu; Wenshuang Zhu; Qingqing Diao; Yujuan Cao; Bingbing Yang; Lining Guo; Yingying Xie; Jiahui Lin; Zhimin Li; Yan Zhang; Kaizhong Xue; Zirui Wang; Junlin Shen

School of Medical Imaging, Tianjin Medical University

Meng Liang; Xuejun Zhang; Hao Ding; Qian Su

Department of Bioinformatics, The Province and Ministry Co-sponsored Collaborative Innovation Center for Medical Epigenetics, School of Basic Medical Sciences, Tianjin Medical University

Mulin Jun Li; Shijie Zhang

Department of Radiology, Tianjin Medical University Cancer Institute and Hospital

Zhaoxiang Ye; Peng Zhang; Wei Li

Department of Radiology, Henan Provincial People's Hospital & Zhengzhou University People's Hospital

Meiyun Wang; Dapeng Shi; Lun Ma; Yan Bai; Min Guan; Wei Wei

Department of Magnetic Resonance Imaging, The First Affiliated Hospital of Zhengzhou University

Jingliang Cheng; Caihong Wang; Peifang Miao; Fuhong Duan; Yafei Guo; Weijian Wang

Department of Medical Imaging, Jinling Hospital, Medical School of Nanjing University

Longjiang Zhang; Lijuan Zheng; Lilin; Yunfei Wang; Han Zhang; Xinyuan Zhang

Department of Radiology, Drum Tower Hospital, Medical School of Nanjing University

Bing Zhang; Zhao Qing; Sichu Wu; Junxia Wang; Yi Sun; Yang He

Institute of Psychology, Chinese Academy of Sciences
Zhenwei Yao; Yue Wu

Department of Radiology, the Center for Medical Imaging, West China Hospital of Sichuan University

Su Lui

References


11. Kühn, S. *et al.* Plasticity of hippocampal subfield volume cornu ammonis 2+3 over the course of withdrawal in patients with alcohol dependence. *JAMA Psychiatry* **71**, 806-811,


**method**

**Study populations**

**CHIMGEN**

The Chinese Imaging Genetics (CHIMGEN) study (http://chimgen.tmu.edu.cn/) has collected genomic, neuroimaging, environmental, and behavioral data from 7,306 healthy Chinese Han participants aged 18-30 years. Participants were recruited from 32 centers and neuroimaging data were acquired by 30 MR scanners. (Supplementary Table 12). The inclusion and exclusion criteria are presented in Supplementary Table 13, and the recruitment processes, standard operating procedures, and data quality control can be found in the website (http://chimgen.tmu.edu.cn/). Informed consent was obtained from all participants and the study was approved by local ethical committees of all research institutions.

**Genetic data processing**

In the 7,306 CHIMGEN participants, 7,195 were genotyped by Illumina ASA-750K (Asian Screening Array), which was specially designed for EAS individuals. The genotyped data were aligned to the human reference genome (GRCh37/hg19). The quality control of the CHIMGEN genetic data was performed by PLINK (http://zzz.bwh.harvard.edu/plink/). In the variant-level quality control, we excluded duplicated variants and variants with call rate < 95%, minor allele frequency (MAF) < 0.001, and Hardy-Weinberg equilibrium (HWE) \( P < 1 \times 10^{-6} \). In the sample-level quality control, we excluded participants with sex mismatch between self-reported and genotyping data, close relations estimated by identity by descent (IBD > 0.1875), excess heterozygosity (> mean ± 5SD), missing genotypes > 3%, and outliers relative to EAS identified by principal components analysis (PCA). The numbers of variants and participants removed in each step are reported in Extended Data Fig. 6. After the variant- and sample-level quality control, a total of 7,163 individuals and 549,309 variants were included in the further imputation. The obtained SNPs were pre-phased by SHAPEIT2 and imputed by IMPUTE2 with the merged reference panel from 1000 Genomes Project (1KGP) and SG10K project. Based on the criteria of MAF ≥ 0.01 and IMPUTE2 info quality score ≥ 0.9, 6,830,146 imputed SNPs were finally included in GWAS.

**MRI data processing**

Structural MRI data were acquired by 10 types of 3.0-Tesla MRI scanners, and scanning parameters are presented in Supplementary Table 14. The image quality of each scanner and consistency across
scanners were assessed\textsuperscript{34}, and quality control was performed to exclude participants with unqualified imaging data, details please see the website (http://chimgen.tmu.edu.cn/en/). Of the 7,163 participants with qualified genetic data, 23 participants without structural MRI data and 79 participants with unqualified structural MRI data were excluded. Hippocampal and subfield volumes were estimated by a subfield segmentation algorithm provided by FreeSurfer v7.0\textsuperscript{70} (https://surfer.nmr.mgh.harvard.edu/). This algorithm can automatically segment out hippocampal subfields based on a probabilistic atlas, which was built with ultra-high-resolution \textit{ex vivo} MRI data\textsuperscript{14}. According to the atlas, the left or right hippocampus can be segmented into 19 subfields, including the parasubiculum, hippocampus-amygdala-transition-area (HATA), fimbria, hippocampal tail, hippocampal fissure, and heads and bodies of the presubiculum, subiculum, CA1, CA3 (including the CA2 and CA3 subfields), CA4, granule cell and molecular layer of the dentate gyrus (GC-ML-DG), and molecular layer of hippocampus proper (molecular layer HP). We also included the volumes of the entire hippocampus, hippocampal head, and hippocampal body as three additional phenotypes, and thus we finally generated 44 traits for hippocampal and subfield volumes. After excluding 22 participants with segmentation failure, we included 7,039 participants with qualified data of the hippocampal and subfield volumes obtained from the 30 MR scanners (Supplementary Fig. 5). The ComBat method was used to harmonize hippocampal and subfield volumes data across scanners, which can remove between-scanner variation and preserve biological variability\textsuperscript{71}. Then the rank-based inverse Gaussian transformation was applied to enhance Gaussianity and to reduce the influence of outliers.

UKBB

The UK Biobank (UKBB) cohort is a prospective study of 500,000 individuals aged between 40-69 years old at recruitment\textsuperscript{72}. UKBB has collected a wide variety of phenotypes and biological samples, a portion of the participants have performed brain MRI examinations\textsuperscript{73}. With UKBB data, two GWASs have included hippocampal volume or hippocampal subfield volumes as the studied traits, and provided GWAS summary statistics\textsuperscript{21,23}. One study included 8,411 EUR participants as the discovery dataset and provided GWAS summary statistics for hippocampal volume\textsuperscript{21}. Another study included 31,968 EUR participants (21,282 for discovery) and provided GWAS summary statistics for both hippocampal volume and hippocampal subfield volumes\textsuperscript{23} (https://open.win.ox.ac.uk/ukbiobank/big40/). In the latter study, the hippocampus and hippocampal subfields were also segmented by FreeSurfer v7.0, and thus GWAS summary statistics of this study\textsuperscript{23} were used for the trans-ancestry meta-analysis.

ENIGMA

The enhancing neuroimaging genetics through meta-analysis (ENIGMA) consortia (http://enigma.ini.usc.edu/) focuses on neuroimaging genetics by pooling worldwide data\textsuperscript{74}. The consortia have conducted a collaborative large-scale GWAS of hippocampal volume to identify genetic associations\textsuperscript{20}. The GWAS summary statistics of the mean hippocampus volume was derived from a meta-analysis of 26,814 EUR participants collected from the ENIGMA consortium and the cohort
for heart and aging research in genomic epidemiology (CHARGE). The hippocampal volume was segmented using FMRIB Software Library (FSL)\textsuperscript{75} and FreeSurfer, details about genotyping, quality control, and genome-wide meta-analysis can be found in the original paper\textsuperscript{20}.

**GWAS of hippocampal and subfield volumes in CHIMGEN**

We used BGENIE v1.3\textsuperscript{21,76} (https://jmarchini.org/bgenie/) to perform GWASs with an additive model, which tested linear associations between qualified genetic variants and hippocampal volumetric traits (n = 44) in the 7,039 CHIMGEN participants, while controlling for the effects of sex, age, age\textsuperscript{2}, age × sex, total intracranial volume, and top 3 genetic principal components from PCA.

**Trans-ancestry meta-analysis of GWAS summary statistics**

Trans-ancestry meta-analysis was performed based on GWAS summary statistics of hippocampal and subfield volumes in EUR participants from UKBB and ENIGMA and in EAS participants from CHIMGEN. All the base positions of these datasets were based on GRCh37/hg19 human reference genome. For GWAS summary data from UKBB, we removed genetic variants with MAF < 0.01. Indel variants were removed because GWAS summary statistics of ENIGMA did not have specific allele information. To handle tri-allelic variants, we matched each variant across the three datasets based on the chromosome, base position (hg19), and alleles. We applied the METASOFT v.2\textsuperscript{35} (http://genetics.cs.ucla.edu/meta/) to perform the meta-analysis. METASOFT can perform meta-analysis based on either the fixed-effects model or the random-effects model. The former estimates the inverse variance weighted effect size, and the latter detects genetic associations under heterogeneity by Han and Eskin's Random Effects model. Because most of the downstream analyses in this study needed beta value and standard error that can only be obtained from the fixed-effects model, we reported the results derived from the fixed-effects model as the main results and validated these results by the Han and Eskin's Random Effects model. Cochran's Q test was used to assess the degree of heterogeneity between EAS and EUR GWAS by estimating the effect size for each genetic variant. Cochran's Q is based on the test with a null hypothesis that homogeneity existed between studies and an alternative hypothesis that heterogeneity existed between studies. The genome-wide significance threshold was set at $P < 5 \times 10^{-8}$. In a given genomic region, a lead SNP was defined as the SNP with the lowest p-value, and its locus was defined as the ±500 kb region centered at the lead SNP. A locus was considered novel if the lead SNP was ±500 kb away from any known loci for the hippocampal volumetric traits of interest.

**Comparing linkage disequilibrium between ancestries**

LD was measured using $r^2$ between SNPs by PLINK with 1000 Genomes Phase 3 reference dataset. For each locus, we calculated the $r^2$ between the lead SNP and all remaining SNPs in corresponding locus in EAS and EUR. The most commonly used threshold of $r^2$ is 0.1 in estimating LD and LD decay\textsuperscript{77}. We compared the $r^2$ between SNPs that both greater than 0.01 in the two populations.
Statistical fine-mapping

Statistical fine-mapping was carried out using the probabilistic annotation integrator (PAINTOR)\(^{37}\) (https://github.com/gkichaev/PAINTOR_V3.0/wiki). For each included SNP, PAINTOR makes a causal inference by generating posterior probability based on the Bayes theorem with the effect of associations and the LD pattern. For each locus, we set a window of 1 Mb centered at the lead SNP and one causal variant was assumed. For GWAS summary data from a single ancestral population (EUR or EAS), the LD value of each SNP was estimated based on 1000 genomes phase 3 reference data of EUR or EAS. For GWAS summary data from the trans-ancestry analysis, we first estimated the LD values with the reference data for each ancestry, and then calculated a weighted LD value for each SNP by including the sample size of each population in trans-ancestry meta-analysis as respective weight\(^{33}\). The 95% credible set of a locus was calculated based on the posterior probabilities of SNPs within the locus. Specifically, the posterior probabilities of the SNPs within one locus were ordered from the largest to the smallest, and then we accumulated the posterior probabilities from the largest one until the joint probability \(\geq 95\%\), the list of SNPs formed the 95% credible set. To validate the fine-mapping results derived from PAINTOR, we also calculated approximate Bayes factors (aBF)\(^{38}\) for each SNP within the locus using its beta and SE values, and then divided aBF of the SNP by the accumulated aBF for all SNPs within the locus to calculate the posterior probability of the SNP. The 95% credible set for each locus was determined by the joint probability \(\geq 95\%\) based on the ordering of the posterior probabilities of SNPs within the locus.

Polygenic score (PGS)

The PGS was calculated using PRSice v.1.25 software\(^{78}\) (https://www.prsice.info/). PRSice removed the ambiguous variants and matched SNPs between base and target datasets, applied LD-pruning on the target dataset (clump-kb 250kb, clump-r\(^2\) 0.1). PRSice performed optimization of the p-value threshold for the most accurate prediction, as measured by r-squared (r\(^2\)), the explained variance from a multivariate regression model. We used the r\(^2\) to estimate the predictive accuracy. The effects of sex, age, age\(^2\), age \(\times\) sex, total intracranial volume, and top 3 genetic principal components from PCA were controlled during the analysis.

Annotation of the identified SNPs

FUMA\(^{42}\) (https://fuma.ctglab.nl/) was used to perform functional annotation of SNPs identified in the trans-ancestry meta-analysis. FUMA is an online platform for annotating and prioritizing genetic variants by integrating comprehensive annotation data sources. We included all SNPs with genome-wide significance \((P < 5 \times 10^{-8})\) in functional annotation and did not use the LD information. The SNPs were mapped to the databases of functional annotations, including ANNOVAR categories\(^{43}\), combined annotation-dependent depletion (CADD) scores\(^{44}\) and eQTL dataset\(^{45}\) of hippocampal tissue provided by GTEx. ANNOVAR annotates the SNPs based on the SNP’s genic position, such as exon, intron, UTR3, and intergenic region. CADD scores prioritize deleterious and pathogenic variants. A SNP is considered potentially pathogenic when the CADD score was above 12.37\(^{79}\). The eQTL dataset describes the
relationships between SNPs and gene expression in hippocampal tissue, an FDR of 0.05 was used to define the genome-wide significant eQTL associations. The SNPs were also mapped to GWAS catalog\textsuperscript{58} (https://www.ebi.ac.uk/gwas/) of brain-related phenotypes and neuropsychiatric diseases by FUMA.

The identified genetic variants by the trans-ancestry meta-analysis were mapped to genes based on physical distance; each SNP was mapped to the nearest gene in the human reference assembly (GRCh37/hg19). The generated genes were investigated by performing a statistical over representation analysis using WebGestalt\textsuperscript{57} online tool (http://www.webgestalt.org/) based on GO biological processes. We used the BH-FDR correction for multiple testing ($q_c < 0.05$).

**References of Method**


74 Thompson, P. M. \textit{et al.} ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. \textit{Transl Psychiatry} \textbf{10}, 100, doi:10.1038/s41398-020-0705-1 (2020).


**Figures**
Figure 1

**Trans-ancestry meta-analyses of hippocampal and subfield volumes.**

- **a,** The spatial distribution of the 44 hippocampal phenotypes. **b,** The ideogram shows genomic locations (lead SNPs) of the 161 novel associations with hippocampal and subfield volumes identified by trans-ancestry meta-analyses with all available samples. **c,** The ideogram shows genomic locations (lead SNPs) of the 60 novel associations identified by trans-ancestry meta-analyses with comparable sample size to EUR-specific GWASs. **d,** The regional plots of a novel locus for the right CA4 body volume only identified by trans-ancestry meta-analysis with all available samples. **e,** The regional plots of a locus for the left CA1 body volume only identified by trans-ancestry meta-analysis with comparable sample size to EUR-specific GWAS. **f,** Effect
sizes and allele frequencies of the 60 novel associations in EAS and EUR as well as representative examples. The distribution indicates that the inter-ancestry difference in effect size rather than allele frequency is the main cause for new discoveries in trans-ancestry meta-analyses. The x-axis shows the effect sizes of the 60 associations in EUR (UKBB, n = 31,968), and the y-axis shows those in EAS (CHIMGEN, n = 7,039). \( g \), In the 60 novel associations, their lead SNPs show larger effects (Wilcoxon rank-sum test: \( P = 5.42 \times 10^{-4} \) for positive effects in the left column and \( P = 7.39 \times 10^{-8} \) for negative effects in the right column) on hippocampal and subfield volumes in EAS (CHIMGEN, n = 7,039) than in EUR (UKBB, n = 31,968). \( h \), The plot shows no difference in the MAFs of the 50 SNPs from the 60 novel associations between EAS and EUR (Wilcoxon signed rank test: \( P = 0.367 \)). \( i \), The Gaussian fitting curves show the relationship between average LD (50kb as interval, y-axis) and genomic distance (x-axis) in EAS and EUR. CA, cornu ammonis; EAS, East Asian ancestry; EUR, European ancestry; GC-ML-DG, the granule cell and molecular layer of the dentate gyrus; GWAS, genome-wide association study; HATA, hippocampus-amygdala-transition-area; HP, hippocampus proper; LD, linkage disequilibrium; MAFs, minor allele frequencies; SNP, single nucleotide polymorphism.
Figure 2

Ancestry-shared and ancestry-specific genetic associations of hippocampal and subfield volumes. **a.** The ideogram shows genomic locations of the 581 locus-trait associations for hippocampal and subfield volumes. The color of the dots demonstrates the directional consistency of effect between EAS and EUR. **b.** The distribution of ancestry-shared and ancestry-specific genetic associations of hippocampal and subfield volumes across the genome. **c.** The regional plots of two loci with ancestry-shared genetic associations between EUR and EAS. **d.** The regional plots of two loci with ancestry-specific genetic associations between EUR and EAS. **e.** The distribution of ancestry-shared and ancestry-specific genetic associations of hippocampal and subfield volumes across the 44 hippocampal volumetric traits. The compound bar chart shows the counts of ancestry-shared associations, and potential and reliable
ancestry-specific genetic associations of hippocampal and subfield volumes. EAS, East Asian ancestry; EUR, European ancestry; GWAS, genome-wide association study.

Figure 3

**Trans-ancestry fine-mapping.** **a,** Comparison of the 95% credible sets between trans-ancestry and EUR-specific fine-mapping (left: 231 loci and 21,282 EUR subjects; right 121 loci and 31,968 EUR subjects). **b** and **c,** Trans-ancestry fine-mapping (**b**) and LD pattern (**c**) of a locus associated with the right CA4 head volume. The number of SNPs in 95% credible set reduces from three in EUR-specific analysis to one in trans-ancestry fine-mapping (**b**), possibly due to the much lower LD of rs1370938 in EAS than in EUR (**c**). **d,** The bar plot shows much lower LD of the 47 SNPs in EAS than in EUR (Wilcoxon rank-sum test: \( P = 3.78 \times 10^{-133} \)). **e,** The Gaussian fitting curves show the relationship between average LD (50kb as interval, y-axis) and genomic distance (x-axis) in EAS and EUR. **f,** The plot shows lower MAFs of the 47 SNPs in EAS than in EUR (Wilcoxon signed rank test: \( P = 2.51 \times 10^{-4} \)). **g,** The plot shows no difference (Wilcoxon rank-sum test, \( P = 0.88 \)) in effect sizes of the 82 associations between trans-ancestry meta-analysis (\( n = 28,321 \)) and EUR GWASs (\( n = 31,968 \)). EAS, East Asian ancestry; EUR, European ancestry; GWAS, genome-wide association study; LD, linkage disequilibrium; MAF, minor allele frequencies; PIP, posterior inclusion probability; SNP, single nucleotide polymorphism.
Prediction performance of polygenic scores (PGSs) constructed by different schemes for hippocampal volumetric traits. **a,** Prediction performance of the four constructed PGSs for each trait (hippocampal volume or subfield volume). The y-axis shows the variance ($R^2$) of each trait explained by each PGS. The PGSs are constructed based on GWAS summary statistics derived from four different combinations of participants: EAS-specific GWAS in CHMGEN discovery ($n = 5,023$), EUR-specific GWASs in UKBB discovery ($n = 21,282$) and total ($n = 31,968$), and trans-ancestry GWAS meta-analysis in CHMGEN+UKBB discovery ($n = 26,305$). **b,** Comparison of overall prediction performance of the four PGSs on hippocampal volumetric traits using Wilcoxon rank-sum test. **c,** Prediction performance of the seven constructed PGSs on each trait. The PGSs are constructed based on GWAS summary statistics derived from seven different combinations of participants: EUR-specific GWASs in UKBB discovery ($n = 21,282$).
and total (n = 31,968), and trans-ancestry GWAS meta-analyses in UKBB discovery (n = 21,282) and five different numbers of CHMGEN discovery (n = 500, 1000, 2000, 3000, and 4000). d, Comparison of overall prediction performance of the seven PGSs on hippocampal volumetric traits using Wilcoxon rank-sum test.

Figure 5

Annotation of genetic loci associated with hippocampal and subfield volumes. a, Categorizing the 4,833 SNPs associated with hippocampal and subfield volumes based on genomic location and functional consequence. b, The yellow circle shows protein-coding genes mapped by the identified SNPs based on...
location, and the gene names in red represent those mapped by the novel loci; the gray circle shows the mapped genes with eQTLs in hippocampal tissue. **c**, The bubble plot shows the enriched gene ontology terms of biological process for protein-coding genes. The x-axis shows the level of significance of each term on the y-axis. The bubble size reflects the number of genes enriched in a given term. The bubble color demonstrates the significance of each term from the statistical over-representation analysis. **d**, Ideogram of the significant associations between the 66 SNPs correlated with hippocampal and subfield volumes and brain-related traits and neuropsychiatric diseases.

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- [ExtendedDataFigures.docx](#)
- [SupplementaryTables.xlsx](#)