Title

A causal association of *ANKRD37* with human hippocampal volume

Author names and affiliations: Jiayuan Xu[1,#], Xianyou Xia[2,#], Qiaojun Li[3,#], Yan Dou[1,#], Xinjun Suo[1], Nana Liu[1], Yating Han[4], Xiaodi Sun[1], Yukun He[5], Wen Qin[1], Shijie Zhang[5], Tobias Banaschewski[6], Herta Flor[7,8], Antoine Grigis[9], Penny Gowland[10], Andreas Heinz[11], Rüdiger Brühl[12], Jean-Luc Martinot[13], Eric Artiges[14], Frauke Nees[6,7,15], Tomáš Paus[16], Luise Poustka[17], Sarah Hohmann[6], Henrik Walter[11], Pak Chung Sham[18], Gunter Schumann[19,20], Xudong Wu[2,21*], Mulin Jun Li[5,22*], Chunshui Yu[1,23,*], for the Alzheimer's Disease Neuroimaging Initiative[24] and IMAGEN Consortia[25].

[1] Department of Radiology and Tianjin Key Laboratory of Functional Imaging, Tianjin Medical University General Hospital, Tianjin 300052, P.R. China
[2] Department of Cell Biology, The Province and Ministry Co-sponsored Collaborative Innovation Center for Medical Epigenetics, Tianjin Medical University, Tianjin 300070, P.R. China
[3] College of Information Engineering, Tianjin University of Commerce, Tianjin 300052, P.R. China
[4] Department of Cell Biology, Tianjin Medical University, Tianjin 300070, P.R. China
[5] The Province and Ministry Co-sponsored Collaborative Innovation Center for Medical Epigenetics, Department of Pharmacology, Tianjin Key Laboratory of Medical Epigenetics, Tianjin Medical University, Tianjin 300070, P.R. China
[6] Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Square J5, 68159 Mannheim, Germany
[7] Department of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Square J5, Mannheim, Germany
[8] Department of Psychology, School of Social Sciences, University of Mannheim, 68131 Mannheim, Germany
[9] NeuroSpin, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France
[10] Sir Peter Mansfield Imaging Centre School of Physics and Astronomy, University of Nottingham, University Park, Nottingham, United Kingdom
[11] Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin,

Humboldt-Universität zu Berlin, and Berlin Institute of Health, Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Charitéplatz 1, Berlin, Germany
[12] Physikalisch-Technische Bundesanstalt (PTB), Abbestr. 2 - 12, Berlin, Germany
[13] Institut National de la Santé et de la Recherche Médicale, INSERM Unit 1000 "Neuroimaging & Psychiatry", University Paris Sud, University Paris Descartes - Sorbonne Paris Cité; and Maison de Solenn, Paris, France
[14] Institut National de la Santé et de la Recherche Médicale, INSERM Unit 1000 "Neuroimaging & Psychiatry", University Paris Sud, University Paris Descartes - Sorbonne Paris Cité; and Psychiatry Department 91G16, Orsay Hospital, France
[15] Institute of Medical Psychology and Medical Sociology, University Medical Center Schleswig-Holstein, Kiel University, Kiel, Germany
[16] Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital and Departments of Psychology and Psychiatry, University of Toronto, Toronto, Ontario, M6A 2E1, Canada
[17] Department of Child and Adolescent Psychiatry and Psychotherapy, University Medical Centre Göttingen, von-Siebold-Str. 5, 37075, Göttingen, Germany
[18] Centre for PanorOmic Sciences-Genomics and Bioinformatics Cores, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR 999077, P.R. China
[19] PONS Centre, Institute for Science and Technology of Brain-inspired Intelligence (ISTBI), Fudan University, Shanghai, P.R. China.
[20] PONS Centre, Charite Mental Health, Dept. of Psychiatry and Psychotherapy, CCM, Charite Universitaetsmedizin Berlin, Germany
[21] Department of Neurosurgery, Tianjin Medical University General Hospital, Tianjin 300052, P.R. China.
[22] Department of Bioinformatics, School of Basic Medical Sciences, Tianjin Medical University, Tianjin 300070, China
[23] CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai, 200031, P.R. China
[24] A complete list of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf
[25] A complete list authors and their affiliations of IMAGEN can be found at the end of the supplementary file.

[#] These authors contributed equally to this work

***Correspondence to:**
Chunshui Yu, M.D., Ph.D.
Department of Radiology, Tianjin Medical University General Hospital, No. 154, Anshan Road, Heping District, Tianjin 300052, China
E-mail: chunshuiyu@tmu.edu.cn
Phone (fax): 86-22-60362026

And
Mulin Jun Li, Ph.D.

The Province and Ministry Co-sponsored Collaborative Innovation Center for Medical Epigenetics, Department of Pharmacology, Tianjin Key Laboratory of Medical Epigenetics, Tianjin Medical University, Tianjin 300070, P.R. China
E-mail: mulin0424.li@gmail.com

And
Xudong Wu, Ph.D.
Department of Cell Biology, Tianjin Medical University and Department of Neurosurgery, Tianjin Medical University General Hospital, Tianjin 300070, P.R. China
E-mail: wuxudong@tmu.edu.cn

Abstract

Human hippocampal volume has been separately associated with single nucleotide polymorphisms (SNPs), DNA methylation and gene expression, but their causal relationships remain largely unknown. Here, we aimed at identifying the causal relationships of SNPs, DNA methylation, and gene expression that are associated with hippocampal volume by integrating cross-omics analyses with genome editing, overexpression and causality inference. Based on structural neuroimaging data and blood-derived genome, transcriptome and methylome data, we prioritized a possibly causal association across multiple molecular phenotypes: rs1053218 mutation leads to cg26741686 hypermethylation, thus leads to overactivation of the associated *ANKRD37* gene expression in blood, a gene involving hypoxia, which may result in the reduction of human hippocampal volume. The possibly causal relationships from rs1053218 to cg26741686 methylation to *ANKRD37* expression obtained from peripheral blood were replicated in human hippocampal tissue. To confirm causality, we performed CRISPR-based genome and epigenome-editing of rs1053218 homologous alleles and cg26741686 methylation in mouse neural stem cell differentiation models, and overexpressed *ANKRD37* in mouse hippocampus. These *in-vitro* and *in-vivo* experiments confirmed that rs1053218 mutation caused cg26741686 hypermethylation and *ANKRD37* overexpression, and cg26741686 hypermethylation favored *ANKRD37* overexpression, and *ANKRD37* overexpression reduced hippocampal volume. The pairwise relationships of rs1053218 with hippocampal volume, rs1053218 with cg26741686 methylation, cg26741686 methylation with *ANKRD37* expression, and *ANKRD37* expression with hippocampal volume could be replicated in an independent healthy young (n=443) dataset and observed in elderly people (n=194), and were more significant in patients with late-onset Alzheimer's disease (n=76). This study revealed

a novel causal molecular association mechanism of *ANKRD37* with human hippocampal volume, which may facilitate the design of prevention and treatment strategies for hippocampal impairment.

Introduction

The human hippocampus contributes to memory, navigation and cognition[1], and is vulnerable to stress[2]. Hippocampal impairment is commonly seen in various brain disorders, such as Alzheimer's disease (AD)[3], epilepsy[4], major depression[5], and schizophrenia[6]. As the most reliable neuroimaging measure, hippocampal volume has been extensively used to assess hippocampal impairment in brain disorders. For example, hippocampal atrophy becomes the most established neuroimaging feature of AD[7, 8] and the strongest predictor of progression to AD[9, 10]. Interindividual variability in hippocampal volume is determined by genetic variations, environmental exposures and their complex interactions. The human hippocampal volume has an estimated heritability of 0.62-0.74[11], and therefore a number of studies have investigated its genetic determinants.

Genetic determinants of human hippocampal volume include variation at genetic, epigenetic and gene expression levels. Genome-wide association studies (GWASs) have discovered a number of genetic loci related to human hippocampal volume[11-13]; however, the majority of the identified SNPs are located in non-coding genomic regions[14], indicative of an influence on transcriptional regulation rather than protein coding sequences. Since GWASs cannot accurately predict the genes mediating the effect of genetic variation on hippocampal volume, it is necessary to investigate the association between gene expression and hippocampal volume. Despite the lack of analyses throughout the entire transcriptome, several studies have reported specific associations between gene expression and hippocampal volume[15, 16]. In addition, epigenome-wide association studies (EWASs) have reported associations between blood DNA methylation and human hippocampal volume[17]. However, these studies were carried out separately to identify the association of genetic variation, gene

expression or DNA methylation with hippocampal volume, leaving the complex relationships of hippocampal volume with genetic, epigenetic and transcriptional variations unexplored. Moreover, these studies can only generate correlations rather than causal dependencies, which are more informative of efficacious therapies and diagnoses of brain disorders.

Non-coding functional genetic variants (S) may affect hippocampal volume (H) by regulating gene expression (E), which is defined as S→E→H, is the simplest and widely acknowledged causal model. Since DNA methylation (M) is under considerable genetic control[18, 19] and usually regulates gene expression by affecting chromatin state or the binding of transcription factors to DNA sequences[20], presumably, S→M→E→H could be a more detailed causal model where genetic variation initially regulates DNA methylation, and the latter affects hippocampal volume by modulating gene expression. Because gene expression and DNA methylation show both tissue-specific and shared patterns[21] as well as blood tissue is easier accessible than hippocampal tissue, identifying the S→M→E→H causal associations shared by blood and hippocampal tissues will create useful blood biomarkers for assessing hippocampal impairment in brain disorders.

In this study, we aimed at identifying S→M→E→H causal mechanisms shared by blood and hippocampal tissues. We firstly combined multi-omics analyses of GWAS[12] (S-H), genome-wide expression quantitative trait loci (*cis*-eQTL, S-E) and methylation quantitative trait loci (*cis*-mQTL, S-M) with causality approaches of summary data-based Mendelian Randomization (SMR)[22], Bayesian co-localization[23] and casual inference test (CIT)[24] to identify S→M→E→H causal mechanisms in human blood tissue. The identified possibly causal S→M→E associations were then replicated in

human hippocampal tissue, and the possibly causal S→M, S→E and M→E effects were confirmed by CRISPR-Cas9 genome and epigenome-editing techniques in the mouse neural stem cell differentiation models, and the possibly causal E→H effects were confirmed by gene overexpression in mouse hippocampus. We finally replicated the identified pairwise relationships of S-H, S-M, M-E, and E-H in an independent dataset of healthy young controls (HYC), and quantitatively compared these correlations in different populations, including HYC, healthy elderly controls (HEC) and patients with late-onset AD (LOAD). A schematic summary of the study design is demonstrated in Figure 1. Throughout the work, the symbol '-' demonstrated associations and '→' demonstrated possible causality.

Material and methods

Summary statistics of samples and datasets

In the discovery analysis, we used participants from ADNI dataset. In ADNI dataset, 808 subjects had qualified WGS data (10,142,241 SNPs), 744 subjects had qualified gene expression data (47,244 transcripts), and 649 subjects had qualified DNA methylation data (736,806 CpG sites). Structural neuroimaging data were acquired from 803 subjects at baseline, and Freesurfer version 5.1.0 was used to calculate bilateral hippocampal volumes and TIV. The mean value of the bilateral hippocampal volumes was defined as the hippocampal volume in this study. We used HYC participants from IMAGEN as replication sample. In IMAGEN dataset, 1,982 subjects had qualified genotyping data (506,932 SNPs), 570 subjects had qualified gene expression data (34,834 transcripts) and 1,290 subjects had qualified DNA methylation data (422,111 CpG sites). Structural neuroimaging data were acquired from 1,724 subjects at the age of 14, and the same method was used to calculate individual's hippocampal volume and TIV. The detailed preprocessing and quality control of genomics, transcriptome, methylation and neuroimaging data in ADNI and IMAGEN datasets were shown in Supplementary methods. Demographic information of the included participants for each statistical analysis is listed in Table 1.

Identifying causal S→M→E→H associations in blood tissue

Identifying S-E associations

In the blood sample, significant associations between SNPs and gene expression were identified by genome-wide *cis*-eQTL in 735 ADNI subjects with both blood-derived WGS and gene expression data using nominal pass function of QTLtools

software[25]. For each probe, Spearman's rank correlation was used to identify SNPs showing significant *cis*-correlations with this probe, controlling for the first five genetic principal components, 60 PEER factors of gene expression, blood cell-type composition, copy numbers of *APOE4*, age, gender, educational years and disease status as covariates in the rank correlation analysis (Supplementary methods and Supplementary Figure 1). The mapping window was defined as 1 Mb up and downstream of the transcription start site of the probe. The numbers of SNPs (n=10,142,241) and probes (n=47,244) were corrected by the Bonferroni method with a significant threshold of $P<0.05/10,142,241/47,244=1.04\times10^{-13}$.

Identifying S-M associations

In the blood sample, significant associations between SNPs and DNA methylation were identified by genome-wide *cis*-mQTL in 604 ADNI subjects with both WGS and CpG methylation data using nominal pass function of QTLtools software[25]. For each CpG site, Spearman's rank correlation was used to identify SNPs showing significant *cis*-correlations with this site, controlling for the first five genetic principal components, 60 PEER factors of DNA methylation, the first four methylation principal components, blood cell-type composition, copy numbers of *APOE4*, age, gender, educational years and disease status as covariates in the rank correlation analysis (Supplementary methods). The mapping window was defined as 1 Mb up and downstream of the CpG site. Multiple testing was corrected for the numbers of SNPs (n=10,142,241) and CpG sites (n=736,806) by the Bonferroni method ($P<0.05/10,142,241/736,806=6.69\times10^{-15}$).

Identifying possibly causal S→E→H associations

We used a summary data-based Mendelian Randomization (SMR) test[22] to identify S→E→H associations (http://cnsgenomics.com/software/smr/), where *y* was defined as phenotype (hippocampal volume), *x* as exposure (gene expression), *z* as instrumental variable (SNP), b_{zx} as the effect of *z* on *x*, b_{zy} as the effect of *z* on *y*, and $b_{xy}=b_{zy}/b_{zx}$ as the effect size of *x* on *y*. The b_{xy} was interpreted as the effect of *x* on *y* free of non-genetic confounders. The SNP effects on the hippocampal volume (b_{zy}) were estimated by GWAS summary statistics of the hippocampal volume from the ENIGMA consortium[12], and the SNP effects on gene expression (b_{zx}) were estimated by the summary data of our *cis*-eQTL analysis. By matching significant SNPs derived from the GWAS ($P<0.05$) and the *cis*-eQTL ($P<1.04\times10^{-13}$), we obtained 5,595 eSNPs corresponding to 7,062 independent eSNP-eProbe pairs. In the SMR analysis, for each candidate eSNP-eProbe pair, we tested for the association (b_{xy}) between the probe (*x*) and trait (*y*) at this eSNP, and the number of pairs were corrected by the Bonferroni method ($P_{SMR}<0.05/7{,}062=7.08\times10^{-6}$).

An observed association in an SMR test could be due to any one of the following three associations: causality (where the effect of a SNP on a trait is mediated by gene expression); pleiotropy (where a SNP shows direct effects on both a trait and gene expression); and linkage (where a SNP is in LD with two distinct causal variants, one impacting gene expression and the other impacting the trait) (Figure 1). Thus, a heterogeneity in dependent instruments (HEIDI) test was applied to multiple SNPs in a *cis*-eQTL region (± 1Mb from the center of the gene probe) to further exclude the linkage associations of less biological interest. Under the hypothesis of pleiotropy or causality, where gene expression and the hippocampal volume share the same causal variant, the b_{xy} values of any SNPs in LD with the causal variant are identical. Therefore,

testing against the null hypothesis that there is a single causal variant is equivalent to testing whether there is heterogeneity in the b_{xy} values estimated for the SNPs in the *cis*-eQTL region. For each probe that passed the significance threshold for the SMR test, we used the HEIDI method to test the heterogeneity in the b_{xy} values estimated for multiple SNPs in the *cis*-eQTL region. A $P_{HEIDI}>0.05$ for the HEIDI test was used since it is conservative for discovery by retaining fewer genes than correcting for multiple testing.

The effect size of GWAS summary statistics of hippocampal volume from the ENIGMA consortium[12] is large, however, a limitation of the above-mentioned analysis is that the GWAS summary data of the hippocampal volume and the *cis*-eQTL summary data came from different subjects, which may introduce bias to the SMR test. To exclude the false positive S→E→H associations identified by the above analysis, we re-performed the SMR and HEIDI tests based on the GWAS summary data of the hippocampal volume and the *cis*-eQTL summary data derived from 707 ADNI subjects with complete genomic, gene expression and hippocampal volume data. The SNP coding alleles were aligned between ENIGMA and ADNI datasets, and also set into the same build GRch37/hg19. The same statistical thresholds were applied ($P<0.05$ for the GWAS, $P<1.04\times10^{-13}$ for the *cis*-eQTL, $P_{SMR}<7.08\times10^{-6}$ for the SMR test, and $P_{HEIDI}>0.05$ for the HEIDI test). Only the S→E→H associations pairs surviving the statistical threshold and showing the consistent direction of effects in the subgroup of ADNI subjects were identified as replicated pairs and were included in the further analysis.

Identifying possibly causal S→M→H associations

We used the same approaches (SMR and HEIDI test) and GWAS data[12] to identify

causal S→M→H associations. By matching significant SNPs derived from the GWAS (P<0.05) and *cis*-mQTL ($P<6.69\times10^{-15}$), we obtained 65,417 meSNPs corresponding to 153,987 independent meSNP-CpG pairs. For each candidate meSNP-CpG pair, we tested the associations between CpG methylation and hippocampal volume by the SMR ($P<0.05/153{,}987=3.20\times10^{-7}$), and then the HEIDI test (P>0.05) was performed to filter out the linkage association.

Since the GWAS summary data of the hippocampal volume and the *cis*-mQTL summary data came from different subjects, we re-performed the SMR and HEIDI tests based on the GWAS summary data of the hippocampal volume and the *cis*-mQTL summary data derived from 585 ADNI subjects with complete genomic, CpG methylation and hippocampal volume data. The same statistical thresholds of P<0.05 for the GWAS, $P<6.69\times10^{-15}$ for the *cis*-mQTL, $P<3.20\times10^{-7}$ for the SMR test, and P>0.05 for the HEIDI test were applied here. Only the S→M→H associations pairs surviving the statistical threshold and showing the consistent direction of effects in the subgroup of ADNI subjects were identified as replicated pairs and were included in the further analysis.

Co-localizing S→M→H and S→E→H associations

Through the above-mentioned steps, we identified possibly causal S→M→H associations and S→E→H associations. To further identify the causal genetic variants (SNPs) that are associated with both gene expression and CpG methylation, we performed colocalization analysis using *coloc* R package with default prior parameters[23]. The colocalization analysis used a Bayesian framework to estimate the posterior probability that two GWAS traits (*cis*-eQTL and *cis*-mQTL) share a single casual variant (PP4) in the selected genome region. For each trait pair, SNPs within

250kb from the lead SNP in the *cis*-eQTL analysis and *cis*-mQTL analysis were included. The default prior parameters were 1.00×10^{-4} for *P1* (the probability of a SNP being associated with gene expression) and *P2* (the probability of a SNP being associated with CpG methylation), and 1.00×10^{-5} for *P12* (the probability of a SNP being associated with both gene expression and CpG methylation). With a Bayesian posterior probability ($P>0.80$), we can identify the possibly causal SNPs that are associated with both gene expression and CpG methylation.

Identifying possibly causal S→M→E→H associations

A casual inference test (CIT)[24] was applied to the co-localized S-M-E pairs in 590 ADNI subjects with complete genomic, transcriptomic and DNA methylation data. Here, we tested four possible relationship models: (1) S→M→E→H model: a SNP affects CpG methylation, then regulates gene expression, and finally impacts on hippocampal volume; (2) S→E→M→H model: a SNP affects gene expression, then regulates CpG methylation, and finally impacts on hippocampal volume; (3) Independent model (S→E→H or S→M→H): a SNP affects gene expression and then regulates hippocampal volume, or a SNP affects CpG methylation and then regulates hippocampal volume; (4) Unspecified model: the relationship among SNP, gene expression and CpG methylation cannot be specified (Figure 1).

The conditions for the establishment of a S→M→E→H model included: (1) a SNP is associated with gene expression; (2) the SNP is also associated with CpG methylation of this gene; (3) the SNP is associated with CpG methylation when adjusting for the expression of this gene; and (4) the SNP is independent of the gene expression after adjusting for the gene methylation (Figure 2). The conditions for the establishment of a S→E→M→H model were similar to the S→M→E→H model except for reversing the

role of gene expression and CpG methylation. If a SNP is only associated with gene expression or CpG methylation, the S→E→H or S→M→H was established. The *P* value was identified using the intersection-union test[24] because all the above four associations must be satisfied. Multiple comparisons were corrected for the number of tested associations using the Bonferroni at corrected $P<0.05$.

Replication of possibly causal S→M→E associations in hippocampus tissue

Due to the high tissue-specificity of gene expression and DNA methylation[21, 26], the identified associations of blood gene expression and CpG methylation with hippocampal volume may reflect either a shared feature of blood and hippocampal tissues or a paradoxical association with unknown biological relevance[21, 27]. If the former is correct, the gene expression and CpG methylation in blood tissue can be regarded as reliable biomarkers for hippocampal volume. Therefore, we first tested if the identified *cis*-mQTLs and *cis*-eQTLs in blood tissue can be observed in human hippocampus tissue. Therefore, we first tested if the identified *cis*-mQTLs and *cis*-eQTLs in blood tissue can be observed in human hippocampus tissue. The genome-wide *cis*-mQTL analysis was conducted in hippocampal biopsies from 110 European patients (58 males) with chronic pharmacoresistant temporal lobe epilepsy provided by Schulz's work[28]. After stringent quality control (Supplementary methods), 536,041 SNPs and 344,106 CpG sites were finally included in the *cis*-mQTL analysis, from which we detected 66,970 significant meSNP-CpG pairs at 14,118 CpG sites at FDR Pc<0.01. The genome-wide *cis*-eQTL analysis was performed in hippocampal tissues from 111 participants provided by GTEx v7 (https://gtexportal.org/home). After

stringent quality control (Supplementary methods), 10,526,813 SNPs and 23,737 genes were included in the *cis*-eQTLs analysis, from which we detected 221,877 significant eSNP-probe pairs of 3,262 genes at FDR $Pc<0.05$.

For the S-M and S-E associations replicated in human hippocampal tissue, we defined *y* as phenotype (gene expression), *x* as exposure (CpG methylation), and *z* as instrumental variable (SNP), and then we jointly performed a two-sample Mendelian randomization (2sMR) and a MR-Egger sensitivity analysis (http://www.mendelianrandomization.com/index.php/software-code)[29] to identify the causal S→M→E associations in hippocampal tissue.

Validation of S→M and M→E causal association in mouse neural stem cells

To confirm the causal effects of the rs1053218 on cg26741686 methylation and casual effects of cg26741686 methylation on *Ankrd37* expression, we performed genome editing of the rs1053218 in the mouse neural stem cell (NSC) line NE-4C. The detailed cell culture and neuronal differentiation, SNP editing, generation of Tetracycline-inducible DNA methylation system, DNA methylation detection and reverse transcription-quantitative polymerase chain reaction (RT-qPCR) analysis were shown in Supplementary methods.

Validation of E→H causal association in mouse hippocampus

To confirm the causal effect of *Ankrd37* overexpression on hippocampal volume, we designed an *in-vivo* experiment by evaluating hippocampal volume and related cognitive test after overexpressing *Ankrd37* in mice hippocampus (Figure 4a). The detailed *Ankrd37* overexpressed adeno-associated-virus packaging, animals and

grouping, *Ankrd37* overexpression in mouse hippocampus and mice cross-sectional and longitudinal neuroimaging acquisition and hippocampal volume calculation process were shown in Supplementary methods.

Replication in an independent dataset

To replicate the identified pairwise associations of *ANKRD37*, we performed the associations among the rs1053218 (S), cg26741686 methylation (M), *ANKRD37* expression (E), and hippocampal volume (H) in an independent 443 HYC. HYC subjects were from the IMAGEN cohort[30]. There are 443 subjects with quality-controlled genome, transcription and methylome data from the blood sample, and structural neuroimaging data. Spearman correlations were applied to test the pairwise relationships of S-H, S-M, M-E, and E-H in HYC from IMAGEN datasets.

These analyses included: S-H association (correlation between the numbers of risk allele of the candidate SNP and hippocampal volumes), controlling for the first five genetic principal components, copy numbers of *APOE4*, age, gender, educational years, TIV, and imaging centers; S-M association (correlation between the numbers of risk allele of the candidate SNP and the candidate DNA methylation levels), controlling for the first five genetic principal components, copy numbers of *APOE4*, 60 PEER factors of DNA methylation, the methylation principal components, blood cell-type composition, age, gender and educational years; M-E association (correlation between the candidate DNA methylation and gene expression levels), controlling for copy numbers of *APOE4*, 60 PEER factors of gene expression, 60 PEER factors of DNA

methylation, the methylation principal components, blood cell-type composition, age, gender and educational years; and E-H association (correlation between the candidate gene expression levels and the hippocampal volumes), controlling for copy numbers of *APOE4*, 60 PEER factors of gene expression, blood cell-type composition, age, gender, educational years, TIV and imaging centers.

Comparing pairwise associations in different populations

To further compare the pairwise associations of *ANKRD37* between HYC, HEC and LOAD, we performed the associations among the rs1053218 (S), cg26741686 methylation (M), *ANKRD37* expression (E), and hippocampal volume (H) in HEC (n=194) and LOAD (n=76) from ADNI datasets, respectively. And the pairwise associations of *ANKRD37* were compared between HYC, HEC and LOAD. All subjects had quality-controlled genome, transcription and methylome data from the blood sample, and structural neuroimaging data. Spearman correlations were applied to test the pairwise relationships of S-H, S-M, M-E, and E-H in HEC (n=194) and LOAD (n=76) from ADNI dataset, respectively. Then we compared the correlation coefficients among the HEC, HYC and LOAD groups using *cocor* 1.0-1 R package[31].

Results

Identifying possibly causal S→E→H associations in blood tissue

To find SNPs with *cis*-regulatory effects on gene expression (eSNPs), we performed genome-wide *cis*-eQTL by defining a window of 1 Mb implemented in QTLtools[25] in 735 ADNI subjects with both WGS (10,142,241 SNPs) and gene expression (47,244 transcripts) data. Using a Bonferroni corrected threshold *(Pc*<0.05 equals to uncorrected $P<0.05/10,142,241/47,244=1.04\times10^{-13}$), we detected 528,079 significant eSNP-probe pairs, including 261,062 unique eSNPs and 4,186 probes corresponding to 2,723 unique genes (eGenes) (Table 1 and Supplementary Figure 2a).

SMR was used to identify the possibly causal S→E→H associations based on the GWAS summary statistics of hippocampal volume from the ENIGMA consortium (n=30,717)[12] and the *cis*-eQTL summary statistics from this study. From a total of 7,062 candidate SNP-probe pairs, we identified 526 significant S→E→H associations ($P_{SMR}<0.05/7,062=7.08\times10^{-6}$) (Supplementary Figure 2b). An HEIDI test ($P_{HEIDI}>0.05$) was performed to exclude the linkage effect and finally resulted in 323 S→E→H associations corresponding to 229 eGenes (Supplementary Figure 2b and Supplementary Data 1). Among the 229 eGenes, 54 genes were replicated in S→E→H associations in 707 ADNI subjects with complete genomic, gene expression and hippocampal volume data. For example, the expression levels of probe 11763200_at tagging *FBXW8*, probe 11719186_a_at tagging *N4BP2L2*, probe 11721917_a_at tagging *ANKRD37* were possible causally associated with hippocampal volume (Supplementary Figure 2b).

Identifying possibly causal S→M→H associations in blood tissue

To identify SNPs with *cis*-regulatory effects on CpG methylation (meSNPs), we performed genome-wide *cis*-mQTL by defining a window of 1 Mb implemented in QTLtools[28] in 604 ADNI subjects with both WGS (10,142,241 SNPs) and DNA methylation (736,806 CpG sites) data. With a Bonferroni corrected threshold (Pc<0.05 equals to uncorrected P<0.05/10,142,241/736,806=6.69×10^{-15}), we detected 4,966,055 significant meSNP-CpG pairs, including 1,651,226 unique meSNPs and 114,625 CpG sites corresponding to 16,149 eGenes (Supplementary Figure 2c).

To identify possible causal S→M→H associations, SMR was performed by integrating the *cis*-mQTL summary statistics from this study and the GWAS summary statistics of hippocampal volume from the ENIGMA consortium[12]. From 153,987 candidate SNP-CpG pairs, we identified 6,853 S → M → H associations (P_{SMR}<0.05/153,987=3.20×10^{-7}). The HEIDI test (P_{HEIDI}>0.05) confirmed 5,330 S→M→H associations including 1,983 eGenes (Supplementary Figure 2d and Supplementary Data 2). In the 1,983 eGenes, 342 genes were replicated in S→M→H associations from independent 585 ADNI subjects with complete genomic, CpG methylation and hippocampal volume data. As an example, we found cg27632911 tagging *SPAG4* whose DNA methylation was possible causally associated with hippocampal volume (Table 1 and Supplementary Figure 2d).

Identifying candidate S-M-E-H associations in blood tissue

To construct candidate S-M-E-H associations, first, we should find eSNPs that affect hippocampal volume by regulating both gene expression and methylation. In the

identified eSNP→eProbe→H and meSNP→CpG→H associations, there are repeated SNPs or SNPs in high linkage disequilibrium (LD). Thus, we screened out 234 independent index eSNPs with significant S → E → H associations and 1,854 independent index meSNPs with significant S→M→H associations (LD clump r^2>0.8 in 250kb). The intersected SNPs (n=16) of these independent eSNPs and meSNPs were defined as SNPs affecting hippocampal volume via regulating both gene expression and methylation. Based on the associations of eSNP→eProbe→H and meSNP→CpG→H, these intersected SNPs corresponded to 30 eProbes and 44 CpGs, forming a total of 108 candidate SNP-CpG-eProbe pairs. The eSNP-eProbe pairs and meSNP-CpG pairs were input into a Bayesian co-localization test implemented in *coloc* R package[23] to estimate the probability for each intersected SNP that simultaneously affect eProbe and CpG. With a posterior probability (PP_{EM}) of > 0.80, we identified 25 co-localized SNP-CpG-eProbe pairs. The *cis*-eQTL and *cis*-mQTL effects and LD information of a representative co-localized SNP (rs1053218) are shown in Figure 2a-b.

Identifying possibly casual S→M→E→H associations in blood tissue

To establish the possibly causal S→M→E→H associations, a casual inference test (CIT)[24] was applied to the 25 co-localized SNP-CpG-eProbe pairs in 590 ADNI subjects with genome, transcriptome and methylome data. Here, we tested four possible relationship models (Figure 2c): (1) S→M→E→H model: a SNP first affects CpG methylation, then the CpG methylation regulates gene expression, and finally the gene expression impacts on hippocampal volume; (2) S→E→M→H model: a SNP first

affects gene expression, then the gene expression regulates CpG methylation, and finally the CpG methylation impacts on hippocampal volume; (3) Independent model (S→E→H or S→M→H): a SNP first affects gene expression and then the gene expression regulate hippocampal volume, or a SNP first affects CpG methylation and then the CpG methylation regulates hippocampal volume; (4) Unspecified model: the relationship among SNP, gene expression and CpG methylation cannot be specified.

At a Bonferroni corrected threshold of $Pc<0.05$ ($Pc<0.05$ equal to uncorrected $P<0.05/25=0.002$), 4 SNP-CpG-eProbe pairs were categorized into the S→M→E→H model, 3 pairs into the S→E→M→H model, 18 pairs into the S→E→H or S→M→H model and no pairs into the unspecified model (Figure 2c and Supplementary Data 3). For example, in S→M→E→H model, the SNP rs1053218 was found to be negatively correlated with hippocampal volume in ENIGMA data ($P=0.042$) and replicated in ADNI ($P=0.008$) and UK Biobank ($P=0.003$) data[32]; the SNP rs1053218 was positively correlated with 11721917_a_at probe of *ANKRD37* gene ($r=0.30$, variance explained (VE)=9.00%, $P<0.001$). The correlation was also replicated by using 1760650_a_at probe ($r=0.11$, VE=1.21%, $P=0.006$) and mean value of two probes ($r=0.16$, VE=2.56%, $P<0.001$) (Supplementary Figure 3); the SNP rs1053218 was positively with cg26741686 methylation of the same gene ($r=0.80$, VE=64.00%, $P<0.001$) (Figure 2c). Moreover, the rs1053218 was significantly positively correlated with the residual of cg26741686 methylation after adjusting for 11721917_a_at probe ($r=0.76$, VE=57.76%, $P<0.001$) (Figure 2c). However, there is no correlation between rs1053218 and residual of 11721917_a_at probe after adjusting for cg26741686 methylation ($r=0.08$, VE=0.64%, $P=0.07$) (Figure 2c). Finally, the cg26741686 methylation is positively correlated with the expression of 11721917_a_at probe

(r=0.27, VE=7.29%, P<0.001) (Figure 2c). These results indicate that this co-localized SNP-eProbe-CpG pair of *ANKRD37* gene satisfied an S→M→E→H model (Figure 2c). In this study, we identified four S→M→E→H possibly causal associations involved three independent genes: *ANKRD37 (*rs1053218 → cg26741686 → 11721917_a_at probe), *PCMT1 (*rs6244128 → cg22239180 → 11718778_s_at probe), and *SQRDL* (rs11633216→cg05747243→11718515_a_at probe and rs2733246→cg05747243→11718515_a_at probe).

Replication of possibly casual S→M→E associations in human hippocampal tissue

Based on the genome-wide *cis*-mQTLs data of human hippocampal tissue[28], we found 23 independent *cis*-mQTLs of the three identified genes in hippocampal tissue (4 *cis*-mQTLs at *ANKRD37*, 6 *cis*-mQTLs at *PCMT1* and 13 *cis*-mQTLs at *SQRDL*; Supplementary Data 4). In *ANKRD37*, rs10000869 showing strong LD (R^2=1.0) with rs1053218 was significantly associated with cg26741686 methylation (P=6.31×10^{-7}) in hippocampal tissue (Figure 2d). In *PCMT1,* rs7753812 having strong LD (R^2=0.96) with rs62441284 was associated with cg00933542 methylation (P=3.99×10^{-18}) and cg15181151 methylation (P=4.02×10^{-6}) in hippocampal tissue (Supplementary Figure 4a). In *SQRDL,* rs625466 demonstrating LD (R^2=0.90) with rs11633216 (P=1.94×10^{-6}) and rs187095 showing LD (R^2=0.97) with rs2733246 (P=1.21 × 10^{-21}) were correlated with cg16220294 methylation in hippocampal tissue (Supplementary Figure 4b). Based on the genome-wide *cis*-eQTLs data of human hippocampal tissue, only the association of rs1053218 with *ANKRD37* expression was replicated in human hippocampus tissue (P=6.20×10^{-16} and Figure 2e).

A two-sample Mendelian randomization (2sMR) and MR-Egger sensitivity analysis

were finally performed using rs1053218 as instrumental variable to make possibly causal inference between the cg26741686 methylation and *ANKRD37* expression in human hippocampal tissue. As a result, we replicated the possibly causal relationship of rs1053218 → cg26741686 methylation → *ANKRD37* expression in hippocampal tissue (P=1.54×10^{-12}).

Validation of S→M and M→E causal effects in mouse neural stem cells

Since the SMR and HEIDI tests are unable to distinguish between causality and pleiotropy, the identified S→M→E relationship may be just a reflection of pleiotropy rather than causality. To confirm the causal effects of the rs1053218 on cg26741686 methylation and causal effects of cg26741686 methylation on *ANKRD37* expression, we performed genome editing of the rs1053218 in the mouse neural stem cell (NSC) line NE-4C. The genome sequence around 20 kb of rs1053218 is generally conserved between mouse and human (Figure 3a), while the minor allele of rs1053218 is C in human but G in mouse. To mimic the rs1053218 genetic effect in humans, we used a CRISPR/Cas9 approach to generate TT genotype in the NE-4C cells (Figure 3a), followed by 10 days' exposure to retinoic acid for neuronal differentiation. The bisulphite sequencing analysis showed that the TT genotype resulted in cg26741686 hypermethylation both before (T=68.63, P=9.00×10^{-6}) (Figure 3b left) and after neuronal differentiation (T=5.16, P=0.04) (Figure 3c left). As shown by the RT-qPCR analysis with two pairs of specific primers (Primer 1 and 2), the TT genotype led to over-activation of *Ankrd37* gene expression specifically after neuronal differentiation (Primer 1: T=125.86, P=3.60×10^{-4}; Primer 2: T=3183.62, P=5.91×10^{-7}) (Figure 3b

and c right). These data indicate that the genetic effect mimicking rs1053218 (GG to TT) causally leads to cg26741686 hypermethylation in NSCs and subsequent hyperactivation of the associated gene *Ankrd37* during neurogenesis.

Validation of M→E causal effects in mouse neural stem cells

To confirm the causal effect of cg26741686 hypermethylation on *Ankrd37* hyperactivation, we developed an inducible system to establish locus-specific targeted DNA methylation in the NE-4C cells. Briefly, we generated a stable cell line expressing a Doxycycline (Dox)-inducible deactivated Cas9-DNA methyltransferase 3a (dCas9-Dnmt3a) in NE-4C cells. As detected by Western blot assay, the dCas9-Dnmt3a fusion protein expression was successfully induced by 24 hours' Dox treatment (2 μg/ml) (Figure 3d). First, we excluded the possibility that Dox treatment alone can affect the *Ankrd37* expression, as no statistically significant difference (P=0.33) in *Ankrd37* expression before and after Dox treatment was observed in NE-4C cells without sgRNA transductions (Figure 3d). Then we designed four single guide RNAs (sgRNAs) targeting the cg26741686 loci, cloned them into two lentiviral vectors (LV) by linear expression of two sgRNAs (LV1: sgRNA1+sgRNA3 and LV2: sgRNA2+sgRNA4) in each vector (Figure 3d) and transduced them into the inducible cells for further experiments. Through Bisulphite sequencing analysis, we confirmed the induced DNA hypermethylation by Dox treatment at the targeted locus before (LV1: T=420.46, $P=1.68\times10^{-9}$; LV2: T=227.82, $P=3.29\times10^{-8}$) (Figure 3e left) and after neuronal

differentiation (LV1: T=76.18, P=5.00×10^{-6}; LV2: T=38.59, P=1.00×10^{-4}) (Figure 3f left). The RT-qPCR analysis with Primer 1 demonstrated that *Ankrd37* expression was modestly but significantly activated before neuronal differentiation at the presence of Dox (LV1: T=39.08, P=3.34×10^{-3}; LV2: T=22.22, P=9.21×10^{-3}) (Figure 3e right). And the subsequent *Ankrd37* activation after neuronal differentiation was further significantly augmented in the hyper-methylated cells (LV1: T=2286.15, P=1.00×10^{-6}; LV2: T=263.42, P=8.40×10^{-5}) (Figure 3f right). The similar findings were also observed when measured by Primer 2 of *Ankrd37* before (LV1: T=1943.83, P=2.00×10^{-6}; LV2: T=142.41, P=2.82×10^{-4}) (Figure 3e right) and after neuronal differentiation (LV1: T=813.01, P=9.00×10^{-6}; LV2: T=922.42, P=7.00×10^{-6}) (Figure 3f right). Taken together, we have validated that the causal S→M and S→E association in NSCs and associated *Ankrd37* deregulation of M→E in the derived neurons.

Validation of E→H causal effects in mice hippocampus

To confirm the causal effect of *Ankrd37* overexpression on hippocampal volume, we designed an *in vivo* experiment by evaluating hippocampal volume and related cognitive test after overexpressing *Ankrd37* in mice hippocampus (Figure 4a and Supplementary Figure 5). Firstly, we constructed the *Ankrd37* overexpressed plasmid, which were packaged with AAV2/9, and then transfected it into the bilateral hippocampal CA1 regions of mice by stereotactic injection (Figure 4b). Two weeks later, the hippocampus was extracted for western blotting, and the results showed that *Ankrd37* expression in the AAV-*Ankrd37*-GFP injection group was significantly higher

than that in the sham group and the WT group, confirming that the model of *Ankrd37* overexpression in mouse hippocampus was constructed successfully (Figure 4c).

In vivo brain structural MRI was firstly performed to observe the changes in hippocampal volume caused by *Ankrd37* overexpression in hippocampus. The volumes of the entire hippocampus and its subregions (CA1, CA2, CA3 and DG) were calculated and showed significant intergroup differences (entire hippocampus: F=5.85, P=0.006, Figure 4d; CA1: F=5.49, P=0.008; CA2: F=6.75, P=0.003; CA3: F=7.02, P=0.002; and DG: F=5.79, P=0.006; Figure 4e) among the *Ankrd37* overexpressed group, the sham group and the WT group. Specifically, the *Ankrd37* overexpressed group exhibited consistent volumetric reduction than the other two groups, indicating that *Ankrd37* overexpression can lead to volumetric reduction in the hippocampus and its subregions.

To observe the longitudinal trajectories of hippocampal and subfield volumes, we performed five consecutive brain structural MRI scans (24-hours pre-injection, 7-, 14-, 21- and 28-days post-injection for the sham and *Ankrd37* mice and the same timeline except for no injection for the WT mice).

In the WT group (n=21), we did not find significant changes over time in any phenotypes (hippocampus, CA1, CA2, CA3 and DG volumes, all P>0.05; Supplementary Figure 6 and 7a). In the sham group (n=20), significant difference across time points was observed in the CA1 (F=4.008, P=0.005) and CA2 (F=3.308, P=0.015) volumes but not in other phenotypes (HP: F=2.429, P=0.054; CA3: F=2.444, P=0.053; and DG: F=1.487, P=0.214; Supplementary Figure 6 and Supplementary Figure 7b).

These findings indicate that CA1 injection reduces volumes of the CA1 and nearby CA2. In the *Ankrd37* group (n=20), all phenotypes showed significant volumetric reduction over time (all $P<0.05$; Supplementary Figure 6 and Supplementary Figure 7c), and the degrees of volumetric reduction in the post-injection time points of all phenotypes were greater than those in the sham group.

At pre-injection (MRI-1) and 7-days post-injection (MRI-2), we did not find any intergroup differences in all phenotypes (all P>0.05, Supplementary Table 1). At 14-days post-injection (MRI-3), there were significant differences in CA1 (F=4.200, P=0.020) and CA2 (F=3.148, P=0.050) volumes among the three groups and post hoc analyses showed that the *Ankrd37* group had significantly reduced CA1 (P=0.002) and CA2 (P=0.018) volumes than the WT group (Supplementary Figure 8a). At 21- and 28-days post-injection (MRI-4 and MRI-5), all phenotypes showed significant differences among the three groups (all $P<0.05$) (Supplementary Figure 8b and 8c). In the post hoc analyses, the *Ankrd37* group had significantly reduced volumes in all phenotypes than the WT and sham groups except for the CA1 volume differences (P=0.410 for MRI-4; P=0.217 for MRI-5) between the *Ankrd37* and sham groups (Supplementary Figure 8b and 8c). We also found that the sham group had significantly reduced CA1 volume at the two time points (P=0.029 for MRI-4; P=0.003 for MRI-5) than the WT group (Supplementary Figure 8b and 8c), indicating that injection has long-lasting impact on the CA1 volume.

The MWM test was performed to observe the changes in spatial learning and memory performance caused by *Ankrd37* overexpression in the hippocampus.

Swimming speeds of all mice before the MWM test were calculated and there were no significant differences (F=1.10, P=0.34) in swimming speeds among the three groups (Supplementary Figure 4). During the learning phase, the escape latency time of the *Ankrd37* overexpressed group were significantly longer (F=11.301, P<0.001) than those of the sham and WT groups, indicating that *Ankrd37* overexpression in hippocampus may cause reduced learning ability of mice (Figure 4f). During the probing phase, the mice in the *Ankrd37* overexpressed group swam aimlessly, whereas the mice in the sham and WT groups were more inclined to the platform quadrant (Figure 4g). Time spent in the target quadrant (F=3.59, P=0.031, Figure 4g) and the numbers of crossings over the platform (F=4.76, P=0.012, Figure 4g) were significantly decreased in the *Ankrd37* overexpressed group than those in the other two groups, indicating that *Ankrd37* overexpression in hippocampus may cause reduced memory ability of mice. We did not find significant differences in escape latency time (P=0.967), time spend in target quadrant (P=0.895) and numbers of crossings (P=0.906) between the sham group and the WT group, indicating that the obtaining results were not biased by the stereotactic injection procedure.

In summary, *Ankrd37* overexpression in hippocampus can reduce hippocampal volume and learning and memory performance in mice, thus validating the E→H causal association found by bioinformatics.

Replication of the pairwise associations of *ANKRD37* in IMAGEN

For the S-H association between rs1053218 and hippocampal volume, we found that the number of risk allele (T) of rs1053218 was negatively correlated with hippocampal volume (r=0.14, VE=1.96%, P=0.004). Specifically, the CT and TT genotypic groups showed significantly smaller hippocampal volume than the CC group (Figure 5a). For the S-M association between rs1053218 and cg26741686 methylation, we found that the number of risk allele (T) of rs1053218 was positively correlated with cg26741686 methylation in HYC (r=0.81, VE=65.61%, P<0.001). (Figure 5b). For the M-E association of cg26741686 methylation with *ANKRD37* expression, we found positively correlation between cg26741686 methylation and *ANKRD37* expression (r=0.14, VE=1.96%, P=0.003) (Figure 5c). For the E-H association of *ANKRD37* expression with hippocampal volume, we found significant negative correlation between *ANKRD37* expression and hippocampal volume (r=-0.14, VE=1.96%, P=0.004) (Figure 5d). Notably, *ANKRD37* expression was detected by different probes in IMAGEN (ILMN_1756417) and ADNI (11721917_a_at); however, these two probes are designed to demonstrate the same transcript NM_181726.2 of *ANKRD37* gene. These results suggest that the pairwise molecular associations of *ANKRD37* with human hippocampal volume is robust and could be extended in HYC.

Comparisons of the pairwise associations of *ANKRD37* in different populations

For the S-H association between rs1053218 and hippocampal volume, we found that the number of risk allele (T) of rs1053218 was negatively correlated with

hippocampal volume in HEC (r=0.15, VE=2.25%, P=0.04) and LOAD (r=0.27, VE=7.29%, P=0.02). Specifically, the CT and TT genotypic groups showed significantly smaller hippocampal volume than the CC group in HEC and LOAD (Figure 5a). For the S-M association between rs1053218 and cg26741686 methylation, we found that the number of risk allele (T) of rs1053218 was positively correlated with cg26741686 methylation in HEC (r=0.67, VE=44.89%, P<0.001) and LOAD (r=0.59, VE=34.81%, P<0.001) (Figure 5b). For the M-E association of cg26741686 methylation with *ANKRD37* expression, we found positively correlation between cg26741686 methylation and *ANKRD37* expression in HEC (r=0.31, VE=9.61%, P<0.001) and LOAD (r=0.54, VE=29.16%, P<0.001) (Figure 5c). More importantly, LOAD patients showed stronger associations than HEC (z=-2.06, P=0.04) and HYC (z=-3.68, P<0.001). HEC showed stronger associations than HYC (z=-2.10, P=0.04) (Figure 5c). For the E-H association of *ANKRD37* expression with hippocampal volume, we found significant negative correlation between *ANKRD37* expression and hippocampal volume in HEC (r=-0.29, VE=8.41%, P<0.001) and LOAD cases (r=-0.63, VE=39.69%, P<0.001) (Figure 5d). LOAD patients showed more significantly associations of *ANKRD37* expression with hippocampal volume than HEC (z=-3.18, P=0.002) and HYC (z=-4.74, P<0.001) (Figure 5d). There were no statistically significant differences of associations of *ANKRD37* expression with hippocampal volume between HYC and HEC. These results suggest that the pairwise molecular associations of *ANKRD37* with human hippocampal volume is highly reproducible and more significant in LOAD cases.

Discussion

To our knowledge, this is the first attempt to decipher the causal relationships among the genetic variation, DNA methylation, and gene expression associated with a neuroimaging phenotype of hippocampal volume by combining genome-wide association analyses of multi-omics data from both blood and hippocampal tissues with causal association tests. Based on hippocampal volume data from structural neuroimaging and genome, transcriptome and epigenome data from blood tissue, we identified a novel S→M→E→H possibly causal association mechanism where the rs1053218 genetic effect leads to cg26741686 hypermethylation, and hyperactivation of the associated *ANKRD37* gene expression, which may cause the reduction of hippocampal volume for yet unknown biological processes. The S→M→E causal association found in blood tissue was also observed in human hippocampal tissue, and the causal S→M, M→E and E→H effects were experimentally confirmed in *in vitro* mouse neuronal differentiation models and *in vivo* overexpressed mouse models. The S-H, S-M, M-E, and E-H associations were replicated in an independent dataset of healthy young people, and these correlations were much stronger in patients with LOAD. This study provides a plausible strategy to integrate the fragmental molecular associations of a given neuroimaging phenotype into a causal mechanism, which can remove false positive findings and provide more reliable molecular targets for the development of novel treatment and molecular markers for progression monitoring and prognosis prediction of hippocampal impairment in brain disorders.

The most important finding of this study is the association between *ANKRD37* and hippocampal volume and the causal relationships among the genetic effect and epigenome changes and the gene expression modulation of *ANKRD37*. The anchor

protein of *ANKRD37* contains an ankyrin repeat domain (ANKRD), a 33-amino acid motif mediating protein-protein interactions[33]. The ANKRD protein family includes many functionally diverse proteins such as enzymes, toxins, and transcription factors, and membrane receptors[33]. Although the function of *ANKRD37* is far from clear, much evidence indicates that *ANKRD37* is involved in the cell response to hypoxia[34, 35]. In the hypoxia-induced responses, hypoxia-inducible factor (HIF) family of transcription factor is commonly regulated. The HIF-1α subunit is degraded under normal oxygen conditions; however, in hypoxic conditions, it is stabilized, translocated to the nucleus and dimerizes with the constitutively expressed subunit HIF-1β to form HIF dimers that bind to and activate their target genes[36], such as *ANKRD37*.

Hypoxia can also facilitate the pathogenesis of LOAD through accelerating the accumulation of Aβ, increasing the hyperphosphorylation of tau, impairing the normal functions of blood-brain barrier, and promoting the degeneration of neurons[37, 38]. Despite of lacking direct evidence, these hypoxia-induced downstream events could occur in the hippocampus since it is particularly sensitive to hypoxia[39] and is severely impaired in LOAD[8]. The adverse effects of hypoxia on LOAD indicate a detrimental impact of the overexpression of *ANKRD37*, a molecule event of the hypoxic response[34], on the hippocampus. This hypothesis was supported by confirming a negative correlation between the *ANKRD37* expression and hippocampal volume, though the *ANKRD37* may be activated through different mechanisms: hypoxia in previous studies[34, 40] versus genetic variant and hyper-methylation in this study. The identified causal association chain from rs1053218 risk allele to cg26741686 hyper-methylation to *ANKRD37* hyperactivation and to reduced hippocampal volume, provides novel molecular targets for the development of the treatment and prevention strategies for the hippocampal impairment in LOAD, as well as novel molecular markers for monitoring

the progress and predicting the prognosis of the disorder.

In contrast to the association between the DNA methylation at CpG islands and gene inactivation[41], intergenic or intragenic DNA methylation has profound effects on gene expression through different mechanisms[42, 43]. In postnatal neural stem cells, Dnmt3a-dependent methylation at the non-proximal promoter regions facilitates expression of their target genes by functionally antagonizing Polycomb repression[44]. In this study, we found that the hyper-methylation of cg26741686 in NSCs resulted in the subsequent hyperactivation of *Ankrd37* gene expression. Notably, the cg26741686 locates in the gene body of *Ankrd37*. In this scenario, intragenic cg26741686 methylation may derestrict *Ankrd37* expression through preventing the occupancy of Polycomb group proteins. However, it requires further functional and mechanistic studies to understand how the multi-omics causal associations of *ANKRD37* affects the hippocampus volume and leads to the pathogenesis of LOAD.

Although the MR analysis demonstrated a causal effect of cg26741686 methylation on *ANKRD37* expression in human hippocampal tissue, we cannot confirm the across-subject correlation between cg26741686 methylation and *ANKRD37* expression in human hippocampal tissue because there were no available individual-level data of DNA methylation and gene expression in human hippocampal tissue from the same participants. In the independent replication of the pairwise associations of *ANKRD37*, we only replicated the associations in HYC. However, we cannot replicate these associations in HEC and LOAD because there are no available independent datasets of HEC and LOAD with blood-derived genetic, gene expression, DNA methylation and structural neuroimaging data. Future studies will replicate the pairwise associations of *ANKRD37* in the independent HEC and LOAD populations.

In conclusion, with a comprehensive strategy, we identified a novel causal

association mechanism of rs1053218 T allele → cg26741686 hypermethylation → overactivation of the *ANKRD37* gene expression → hippocampal volume reduction. This study not only provides a plausible approach to integrate the fragmental associations into a causal mechanism, but also novel targets for treatments and new biomarkers for prediction of LOAD.

Code availability

Custom code that supports the findings of this study is available from the corresponding author upon request.

References

1. Lisman J, Buzsáki G, Eichenbaum H, Nadel L, Ranganath C, Redish AD. Viewpoints: how the hippocampus contributes to memory, navigation and cognition. *Nat Neurosci* 2017; **20**: 1434-1447.

2. Kim EJ, Pellman B, Kim JJ. Stress effects on the hippocampus: a critical review. *Learn Mem* 2015; **22**: 411-416.

3. Moodley K, Chan D. The hippocampus in neurodegenerative disease. *Front Neurol Neurosci* 2014; **34**: 95-108.

4. Chatzikonstantinou A. Epilepsy and the hippocampus. *Front Neurol Neurosci* 2014; **34**: 121-142.

5. MacQueen G, Frodl T. The hippocampus in major depression: evidence for the convergence of the bench and bedside in psychiatric research? *Mol Psychiatry* 2011; **16**: 252-264.

6. Roeske MJ, Konradi C, Heckers S, Lewis AS. Hippocampal volume and hippocampal neuron density, number and size in schizophrenia: a systematic review and meta-analysis of postmortem studies. *Mol Psychiatry* 2021; **26**: 3524-3535.

7. Halliday G. Pathology and hippocampal atrophy in Alzheimer's disease. *Lancet Neurol* 2017; **16**: 862-864.

8. Bayram E, Caldwell JZ, Banks SJ. Current understanding of magnetic resonance imaging

biomarkers and memory in Alzheimer's disease. *Alzheimers Dement (N Y)* 2018; **4**: 395-413.

9. Henneman WJ, Sluimer JD, Barnes J, van der Flier WM, Sluimer IC, Fox NC et al. Hippocampal atrophy rates in Alzheimer disease: added value over whole brain volume measures. *Neurology* 2009; **72**: 999-1007.

10. Jack CR Jr, Petersen RC, Xu Y, O'Brien PC, Smith GE, Ivnik RJ et al. Rates of hippocampal atrophy correlate with change in clinical status in aging and AD. *Neurology* 2000; **55**: 484-490.

11. Stein JL, Medland SE, Vasquez AA, Hibar DP, Senstad RE, Winkler AM et al. Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet* 2012; **44**: 552-561.

12. Hibar DP, Stein JL, Renteria ME, Arias-Vasquez A, Desrivières S, Jahanshad N et al. Common genetic variants influence human subcortical brain structures. *Nature* 2015; **520**: 224-229.

13. Elliott LT, Sharp K, Alfaro-Almagro F, Shi S, Miller KL, Douaud G et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* 2018; **562**: 210-216.

14. Alexander RP, Fang G, Rozowsky J, Snyder M, Gerstein MB. Annotating non-coding regions of the genome. *Nat Rev Genet* 2010; **11**: 559-571.

15. Jasinska AJ, Zelaya I, Service SK, Peterson CB, Cantor RM, Choi OW et al. Genetic variation and gene expression across multiple tissues and developmental stages in a nonhuman primate. *Nat Genet* 2017; **49**: 1714-1721.

16. Kepa A, Martinez Medina L, Erk S, Srivastava DP, Fernandes A, Toro R et al. Associations of the intellectual disability gene MYT1L with helix–loop–helix gene expression, hippocampus volume and hippocampus activation during memory retrieval. *Neuropsychopharmacology* 2017; **42**: 2516-2526.

17. Jia T, Chu C, Liu Y, van Dongen J, Papastergios E, Armstrong NJ et al. Epigenome-wide meta-analysis of blood DNA methylation and its association with subcortical volumes: findings from the ENIGMA Epigenetics Working Group. *Mol Psychiatry* 2021; **26**: 3884-3895.

18. McRae AF, Powell JE, Henders AK, Bowdler L, Hemani G, Shah S et al. Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biol* 2014; **15**: 1-10.

19. Gutierrez-Arcelus M, Lappalainen T, Montgomery SB, Buil A, Ongen H, Yurovsky A et al. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife* 2013; **2**: e00523.

20. Wu H, Zhang Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* 2014; **156**, 45-68.

21. Blake LE, Roux J, Hernando-Herraez I, Banovich NE, Perez RG, Hsiao CJ et al. A comparison of gene expression and DNA methylation patterns across tissues and species. *Genome Res* 2020; **30**: 250-262.

22. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* 2016; **48**: 481-487.

23. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 2014; **10**: e1004383.

24. Millstein J, Zhang B, Zhu J, Schadt EE. Disentangling molecular relationships with a causal inference test. *BMC Genet* 2009; **10**: 1-15.

25. Delaneau O, Ongen H, Brown AA, Fort A, Panousis NI, Dermitzakis ET. A complete tool set for molecular QTL discovery and analysis. *Nat Commun* 2017; **8**: 15452.

26. Davies MN, Volta M, Pidsley R, Lunnon K, Dixit A, Lovestone S et al. Functional annotation of the human brain methylome identifies tissue-specific epigenetic variation across brain and blood. *Genome Biol* 2012; **13**: R43.

27. Montano C, Taub MA, Jaffe A, Briem E, Feinberg JI, Trygvadottir R et al. Association of DNA methylation differences with schizophrenia in an epigenome-wide association study. *JAMA Psychiatry* 2016; **73**: 506-514.

28. Schulz H, Ruppert AK, Herms S, Wolf C, Mirza-Schreiber N, Stegle O et al. Genome-wide mapping of genetic determinants influencing DNA methylation and gene expression in human hippocampus. *Nat Commun* 2017; **8**: 1511.

29. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM et al. Guidelines for performing Mendelian randomization investigations. *Wellcome Open Res* 2019; **4**: 186.

30. Schumann G, Loth E, Banaschewski T, Barbot A, Barker G, Büchel C et al. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Mol Psychiatry* 2010; **15**: 1128-1139.

31. Diedenhofen B, Musch J. cocor: a comprehensive solution for the statistical comparison of correlations. *PLoS One* 2015; **10**: e0121945.

32. Smith SM, Douaud G, Chen W, Hanayik T, Alfaro-Almagro F, Sharp K et al. An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nat Neurosci* 2021; **24**: 737-745.

33. Bork P. Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? *Proteins* 1993; **17**: 363-374.

34. Benita Y, Kikuchi H, Smith AD, Zhang MQ, Chung DC, Xavier RJ et al. An integrative genomics approach identifies Hypoxia Inducible Factor-1 (HIF-1)-target genes that form the core response to hypoxia. *Nucleic Acids Res* 2009; **37**: 4587-4602.

35. Galbraith MD, Allen MA, Bensard CL, Wang X, Schwinn MK, Qin B et al. HIF1A employs CDK8-mediator to stimulate RNAPII elongation in response to hypoxia. *Cell* 2013; **153**: 1327-1339.

36. Majmundar AJ, Wong WJ, Simon MC. Hypoxia-inducible factors and the response to hypoxic stress. *Mol Cell* 2013; **40**: 294-309.

37. Zhang X, Le W. Pathological role of hypoxia in Alzheimer's disease. *Exp Neurol* 2010; **223**: 299-303.

38. Sun X, He G, Qing H, Zhou W, Dobie F, Cai F et al. Hypoxia facilitates Alzheimer's disease pathogenesis by up-regulating BACE1 gene expression. *Proc Natl Acad Sci U S A* 2006; **103**, 18727-18732.

39. Lana D, Ugolini F, Giovannini MG. An Overview on the Differential Interplay Among Neurons–Astrocytes–Microglia in CA1 and CA3 Hippocampus in Hypoxia/Ischemia. *Front Cell Neurosci* 2020; **14**:585833.

40. Yasukochi Y, Shin S, Wakabayashi H, Maeda T. Transcriptomic changes in young Japanese males after exposure to acute hypobaric hypoxia. *Front Genet* 2020; **11**:559074.

41. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* 2012; **13**: 484-492.

42. Neri F, Rapelli S, Krepelova A, Incarnato D, Parlato C, Basile G et al. Intragenic DNA methylation prevents spurious transcription initiation. *Nature* 2017; **543**: 72-77.

43. Weinberg DN, Papillon-Cavanagh S, Chen H, Yue Y, Chen X, Rajagopalan KN et al. The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation

landscape. *Nature* 2019; **573**: 281-286.

44. Wu H, Coskun V, Tao J, Xie W, Ge W, Yoshikawa K et al. Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Nature* 2010; **329**: 444-448.

Acknowledgment

This work was partly supported by the National Key Research and Development Program of China (Grant No. 2018YFC1314301), National Natural Science Foundation of China (Grant No. 82001797), Tianjin Applied Basic Research Diversified Investment Foundation (Grant No. 21JCYBJC01360), Tianjin Health Technology Project (Grant No. TJWJ2021QN002), Science&Technology Development Fund of Tianjin Education Commission for Higher Education (2019KJ195), National Key Research and Development Program of China (Grant No. 2017YFA0604401 and 2017YFA0504102), National Natural Science Foundation of China (Grant No. 82030053, 32070675 and 81801687), Natural Science Foundation of Tianjin City (Grant No. 19JCJQJC63600 for M.J.L. and 18JCJQJC48200 for X.W.), Tianjin Key Medical Discipline (Specialty) Construction Project (Grant No. TJYXZDXK-001A). Further support was received by GS from the Horizon 2020 funded ERC Advanced Grant 'STRATIFY' (Brain network-based stratification of reinforcement-related disorders) (695313), the National Institute of Health (NIH) (R01DA049238, A decentralized macro and micro gene-by-environment interaction analysis of substance use behavior and its brain biomarkers), the Human Brain Project (SGA3; 945539), and the Chinese National High-end Foreign Expert Recruitment Plan. Healthy elderly controls and LOAD cases data collection and

sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Author information

Affiliations

Department of Radiology and Tianjin Key Laboratory of Functional Imaging, Tianjin Medical University General Hospital, Tianjin 300052, P.R. China

Jiayuan Xu, Yan Dou, Xinjun Suo, Nana Liu, Xiaodi Sun, Wen Qin & Chunshui Yu

CAS Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai, 200031, P.R. China

Chunshui Yu

Department of Cell Biology, The Province and Ministry Co-sponsored Collaborative Innovation Center for Medical Epigenetics, Tianjin Medical University, Tianjin 300070, P.R. China

Xianyou Xia & Xudong Wu

College of Information Engineering, Tianjin University of Commerce, Tianjin 300052, P.R. China

Qiaojun Li

Department of Cell Biology, Tianjin Medical University, Tianjin 300070, P.R. China

Yating Han

The Province and Ministry Co-sponsored Collaborative Innovation Center for Medical Epigenetics, Department of Pharmacology, Tianjin Key Laboratory of Medical Epigenetics, Tianjin Medical University, Tianjin 300070, P.R. China

Yukun He, Shijie Zhang & Mulin Jun Li

Department of Child and Adolescent Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University,

Square J5, 68159 Mannheim, Germany

Tobias Banaschewski, Frauke Nees & Sarah Hohmann

Department of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Square J5, Mannheim, Germany

Herta Flor

Department of Psychology, School of Social Sciences, University of Mannheim, 68131 Mannheim, Germany

Herta Flor

NeuroSpin, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

Antoine Grigis

Sir Peter Mansfield Imaging Centre School of Physics and Astronomy, University of Nottingham, University Park, Nottingham, United Kingdom

Penny Gowland

Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Department of Psychiatry and Psychotherapy, Campus Charité Mitte, Charitéplatz 1, Berlin, Germany

Andreas Heinz & Henrik Walter

Physikalisch-Technische Bundesanstalt (PTB), Abbestr. 2 - 12, Berlin, Germany

Rüdiger Brühl

Ecole Normale Supérieure Paris-Saclay, Université Paris-Saclay, CNRS; Centre

Borelli; INSERM U1299 "Trajectoires Développementales & Psychiatrie" Avenue des Sciences, 91190 Gif-sur-Yvette, France

Jean-Luc Martinot

Institut National de la Santé et de la Recherche Médicale, INSERM Unit 1000 “Neuroimaging & Psychiatry”, University Paris Sud, University Paris Descartes - Sorbonne Paris Cité; and Psychiatry Department 91G16, Orsay Hospital, France

Eric Artiges

Departments of Psychiatry and Neuroscience, Faculty of Medicine and Centre Hospitalier Universitaire Sainte-Justine, University of Montreal, Montreal, Quebec, Canada.

Tomáš Paus

Departments of Psychology and Psychiatry, University of Toronto, Toronto, Ontario, Canada.

Tomáš Paus

Department of Child and Adolescent Psychiatry and Psychotherapy, University Medical Centre Göttingen, von-Siebold-Str. 5, 37075, Göttingen, Germany

Luise Poustka

Centre for PanorOmic Sciences-Genomics and Bioinformatics Cores, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR 999077, P.R. China

Pak Chung Sham

Centre for Population Neuroscience and Stratified Medicine (PONS), Institute for Science and Technology of Brain-inspired Intelligence (ISTBI), Fudan University,

Shanghai, P.R. China.

Gunter Schumann

Centre for Population Neuroscience and Stratified Medicine (PONS), Charite Mental Health, Dept. of Psychiatry and Psychotherapy, CCM, Charite Universitätsmedizin Berlin, Germany

Gunter Schumann

Author contributions

J.X., M.J.L., X.W. and C.Y. designed the study. J.X., X.X, Y.D., M.J.L., X.W. and C.Y. wrote the manuscript. J.X., Q.L., X.X and Z.S. analysed the data. All authors critically reviewed the manuscript. X.S., N.L., Y.H., X.S., Y.H., W.Q. and S.Z. were the principal investigators. T.B., H.F., A.G., P.G., A.H., R.B., J.M., E.A., F.N., T.P., L.P., S.H., H.W., P.S. and G.S. acquired the data.

Conflict of interests

The authors declare no competing interests.

Figures legends

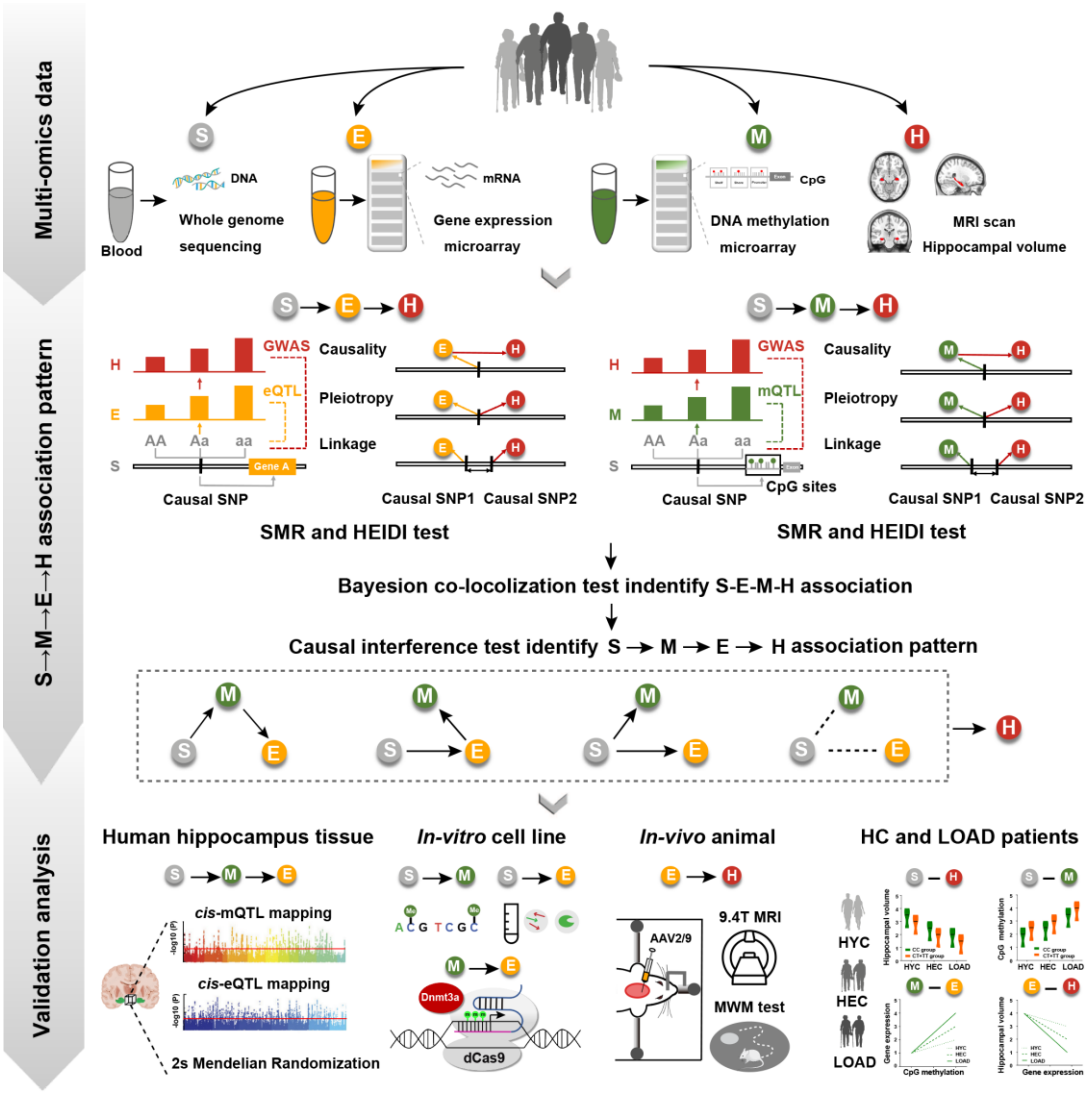

Figure 1. A schematic summary of the study design. Based on blood-derived genomic, transcriptomic and methylomic data and structural MRI data collected from ADNI dataset, we identify possibly causal S→M→E→H associations in blood tissue by jointly using multi-omics and causality analyses. We then replicate the identified possibly causal S→M→E association in human hippocampus tissue, and confirm the causal S→M, S→E and M→E effects in mouse neural stem cells by CRISPR-Cas9 genetic and epigenetic editing techniques, and the causal E→H effects by gene overexpression in mouse hippocampus. The pair-wise associations of different molecular traits are further confirmed in an independent dataset of HYC and compared between HYC, HEC and LOAD. 2s Mendelian Randomization, two-sample Mendelian

Randomization; eQTL, expression quantitative trait loci; E, gene expression; GWAS, genome-wide association analysis; H, hippocampal volume; HC, healthy controls; HEC, healthy elderly controls; HEIDI, heterogeneity in dependent instruments; HYC, healthy young controls; LOAD, Late onset Alzheimer's disease; M, DNA methylation; MWM, Morris water maze; S, single nucleotide polymorphisms; and SMR, summary data-based Mendelian Randomization.

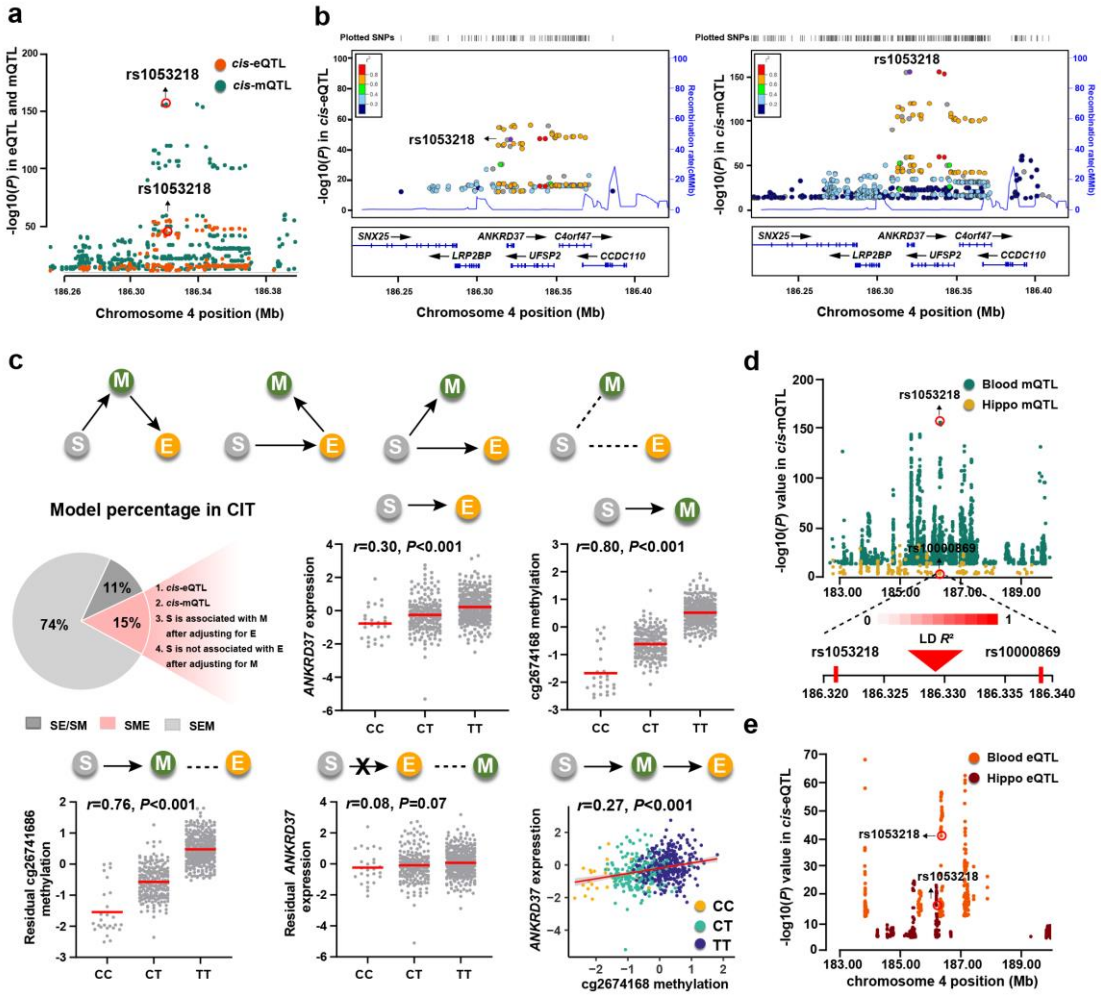

Figure 2. Identifying possibly casual S→M→E→H associations in blood and hippocampal tissues. a. The significance of a co-localizing SNP rs1053218 (red circle) in *cis*-eQTL (green) and *cis*-mQTL (orange). b. LD information of a co-localized SNP-CpG-eProbe pair tagging *ANKRD37*. The rs1053218 (purple dot) simultaneously affect the expression of 11721917_a_at (left) and the methylation of cg26741686 (right). c. First row. Four possible relationships (S→M→E, S→E→M, S→E or S→M, and unspecified) of the hippocampal volume (H)-related molecular phenotypes of SNP (S), gene expression (E) and CpG methylation (M); Second row. Left. Causal interference

test (CIT) reveals 4 S→M→E, 3 S→E→M, 18 S→E or S→M associations from the 25 co-localized SNP-CpG-eProbe pairs. The pink regions shows summary workflow for identifying S→M→E associations. An example of the identified S→M→E associations. The numbers of risk allele (T) of rs1053218 are positively correlated with *ANKRD37* expression (medium) and cg26741686 methylation (right). Third row. The rs1053218 is still positively correlated with cg26741686 methylation after adjusting *ANKRD37* expression (left), but not with *ANKRD37* expression after adjusting cg26741686 methylation (medium); and cg26741686 methylation is positively correlated with *ANKRD37* expression (right). The red line represents the mean. d. Replication of S→M associations in human hippocampal tissue. The rs1053218 is correlated with cg26741686 methylation in blood tissue (top, green) and the rs10000869 is correlated with cg26741686 methylation in hippocampal tissue (top, yellow), and rs1053218 and rs10000869 show strong LD (R^2=1.0). e. Replication of S→E associations in human hippocampal tissue. The rs1053218 is correlated with *ANKRD37* expression both in blood (orange) and hippocampal (red) tissues. *ANKRD37*, ankyrin repeat domain 37; *cis*-eQTL, *cis*-expression quantitative trait loci; *cis*-mQTL, *cis*-methylation quantitative trait loci; E, gene expression; Hippo, human hippocampal tissue; LD, linkage disequilibrium; M, CpG methylation; S, Single Nucleotide Polymorphisms; SEM, S→E→M association; SME, S→M→E association; SE/SM, S→E or S→M association.

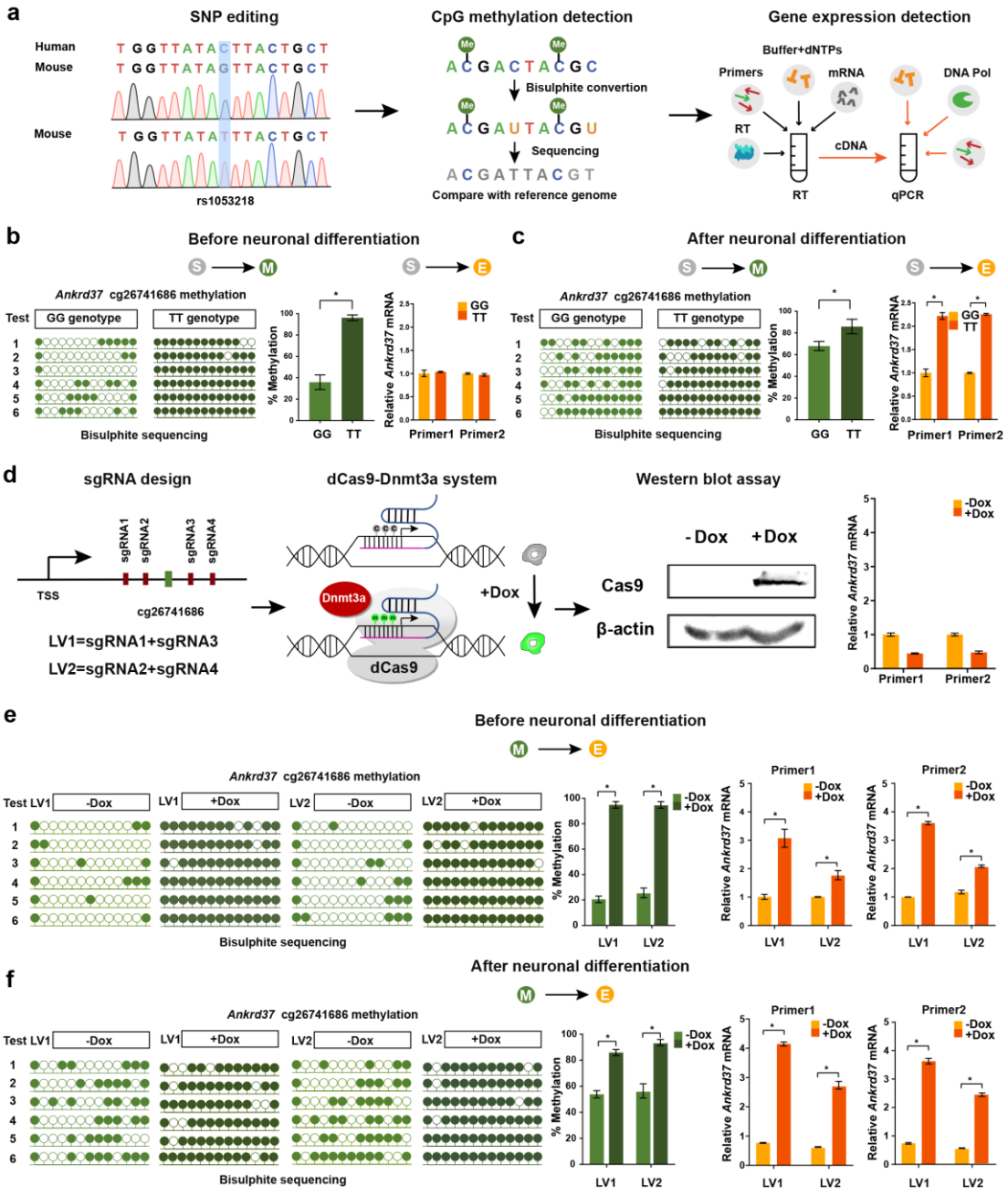

Figure 3. Validation of S→M, S→E and M→E causal effects in mouse neural stem cell (NSC). a. Left: The genome sequences around 20 kb of rs1053218 are conserved between mouse and human, while the minor allele of rs1053218 is C in human but G in mouse. Sequencing peaks confirms the rs1053218 TT genotype in NE-4C cells; Medium: CpG methylation is detected by bisulphite sequencing; Right: Gene expression is detected by RT-qPCR. b. Before neuronal differentiation of NSCs, bisulphite sequencing shows that rs1053218 TT genotype leads to significant cg26741686 hypermethylation and RT-qPCR demonstrates that this variant alone is not sufficient for *Ankrd37* activation. c. After neuronal differentiation, rs1053218 TT

genotype leads to both cg26741686 hypermethylation and *Ankrd37* hyperactivation. d. Left: sgRNAs (red) are located around the targeted cg26741686 site (green). The sgRNA1 and sgRNA3 were cloned into the first lentiviral vector (LV1) and sgRNA2 and sgRNA4 into the second (LV2); Medium: Schematic representation of deactivated Cas9 (dCas9) fused with Dnmt3a for de novo methylation of targeted cg26741686 site; Right: Western blot assay confirms the inducible expression of dCas9-Dnmt3a fusion protein. In the NE-4C cells without sgRNA transductions, RT-qPCR analysis shows that Dox treatment does not significantly affect the *Ankrd37* expression levels. e. Before neuronal differentiation of NSCs, Dox treatment results in significant cg26741686 hypermethylation and modest *Ankrd37* hyperactivation. f. After neuronal differentiation, Dox treatment leads to significant cg26741686 hypermethylation and *Ankrd37* hyperactivation. cDNA, complementary deoxyribonucleic acid; dCas9-Dnmt3a, deactivated Cas9-DNA methyltransferase 3a; dNTPs, deoxyribonucleoside triphosphate; DNA pol, DNA polymerase; Dox, doxycycline; E, gene expression; M, CpG methylation; mRNA, messenger ribonucleic acid; qPCR, quantitative polymerase chain reaction; RT, reverse transcription; S, Single Nucleotide Polymorphisms; sgRNA, single guide ribonucleic acid.

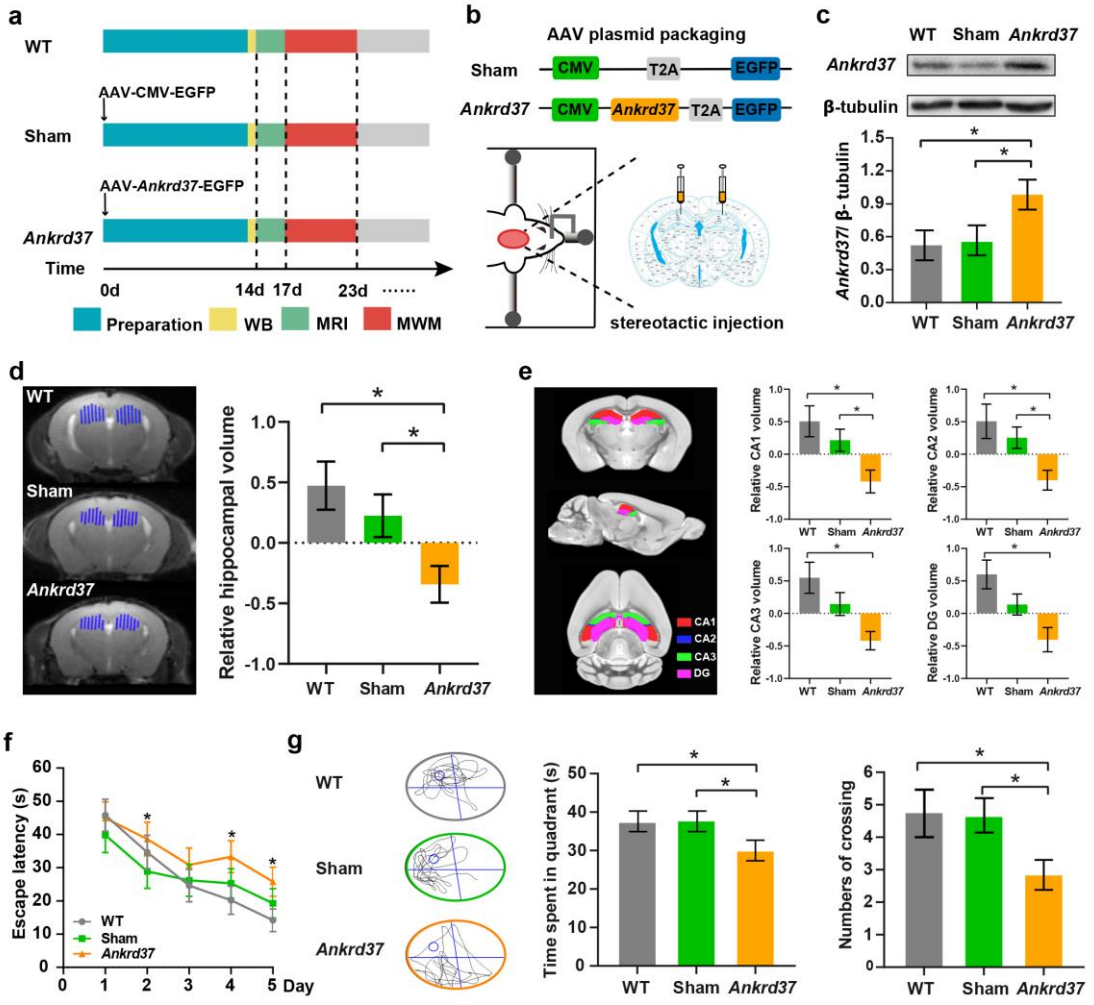

Figure 4. Validation of E→H causal association in mice hippocampus. a. Schematic design for animal experiments. b. Schematic of AAV plasmid construction and stereotactic injection site in mice. c. Top. Western blot assay confirms the overexpression of *ANKRD37* in the hippocampus. Bottom. *Ankrd37* expression in the AAV-*Ankrd37*-EGFP injection group is significantly higher than those in the sham and WT groups, confirming the successful construction of *Ankrd37* overexpression in mouse hippocampus. d. Left. Representative mice bilateral hippocampal segmentation derived from T2-weighted MR images among the three groups. Right. The relative hippocampal volume is significantly reduced in the *Ankrd37* overexpressed group than those in the sham and WT groups. e. Left. Mice bilateral hippocampal subregions including CA1, CA2, CA3 and DG. Right. The bilateral relative CA1, CA2, CA3 and DG volumes are significantly decreased in *Ankrd37* overexpressed group than those in the other two groups. f. During the learning phase of the MWM test, escape latency

time in the *Ankrd37* overexpressed group is longer than those in the sham and WT groups. g. Left. During the probing phase, representative swimming paths of mice among the three groups. The mice in the *Ankrd37* overexpressed group swim aimlessly, whereas the mice in the sham and WT groups are more inclined to the platform quadrant. Medium and right. Time spent in the target quadrant (medium) and the numbers of crossings (right) over the platform in the *Ankrd37* overexpressed group are decreased than those in the other two groups. CA, Cornu Ammonis; DG, dentate gyrus; EGFP, Enhanced Green Fluorescent Protein; WT, wild type; MWM, Morris Water Maze; WB, western blot; MRI, Magnetic resonance imaging; EF1a, EF1 alpha promoter.

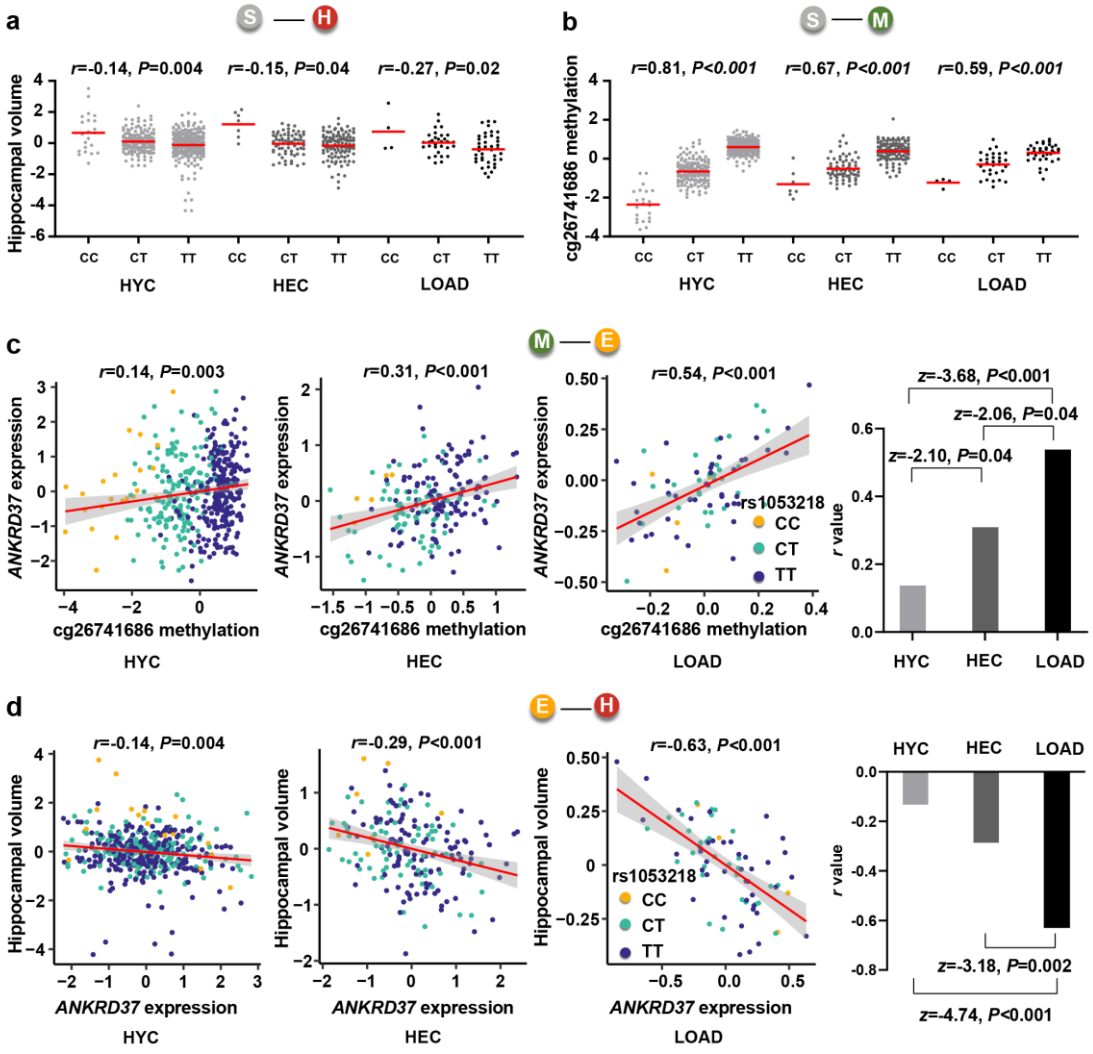

Figure 5. Pairwise replication of the S→M→E→H associations of *ANKRD37* in HYC and pairwise comparisons in different populations. a. S-H: The numbers of risk allele (T) of rs1053218 is negatively correlated with hippocampal volume in HYC, HEC and LOAD. b. S-M: The numbers of risk allele (T) of rs1053218 is positively

correlated with cg26741686 methylation in HYC, HEC and LOAD. c. M-E: The cg26741686 methylation is positively correlated with *ANKRD37* expression in HYC, HEC and LOAD. More importantly, LOAD patients show stronger associations than HEC and HYC. d. E-H: The *ANKRD37* expression is negatively correlated with hippocampal volume in HYC, HEC and LOAD. LOAD patients show more significant associations than HYC and HEC. *ANKRD37*, ankyrin repeat domain 37; E, gene expression; HEC, healthy elderly controls; H, hippocampal volume; HYC, healthy young controls; LOAD, late-onset Alzheimer's disease; M, CpG methylation; S, Single Nucleotide Polymorphisms.

Supplementary Figure 1. Identification of the optimal number of PEER factors for hidden covariate correction during *cis*-eQTL and *cis*-mQTL analyses. The optimization was performed in *cis*-eQTL and *cis*-mQTL analyses, with increments of 5 PEER factors as confounding covariates. The number of PEER factors was chosen to maximize the numbers of eGenes in *cis*-eQTL and *cis*-mQTL results. We chose 60 PEER factors because numbers of eGenes plateaued at 60 in *cis*-eQTL and *cis*-mQTL analysis, respectively.

Supplementary Figure 2. Possibly causal S→E→H and S→M→H associations in blood tissue. a. Manhattan plots of *cis*-eQTLs (Bonferroni Pc<0.05). b. Manhattan plots show 7,062 SNP-eProbe pairs (left) included in SMR analysis, 526 significant S →E→H associations (medium) found in SMR analysis (*Pc*<0.05), and 323 significant associations (right) passing HEIDI test (*P*>0.05). The top seven associations found in SMR analysis are marked out and only three (black) pass HEIDI test. c. Manhattan plot displays candidate *cis*-mQTLs (Bonferroni *Pc*<0.05) and the red line indicates the significant threshold (Bonferroni Pc<0.05); d. Manhattan plots show 153,987 SNP-CpG site pairs (left) included in SMR analysis, 6,853 significant S→M→H associations (medium) found in SMR analysis (*Pc*<0.05), and 5,330 significant S → E → H associations (right) passing HEIDI test (*P*>0.05). The top seven associations found in SMR analysis are marked out and only one (black) passes HEIDI test. *cis*-eQTL, *cis*-expression quantitative trait loci; *cis*-mQTL, *cis*-methylation quantitative trait loci;

HEIDI, heterogeneity in dependent instruments; LD, linkage disequilibrium; SMR, summary data-based Mendelian Randomization.

Supplementary Figure 3. Genetic associations with blood *ANKRD37* expression detected by different probes. a. Kernel density plot shows the distributions of *ANKRD37* expression detected by 11721917_a_at probe (yellow), 11760650_a_at probe (orange) and mean value of the two probes (red). b. The numbers of risk allele (T) of rs1053218 are positively correlated with *ANKRD37* expression detected by 11721917_a_at probe (left), 11760650_a_at probe (medium) and mean value of the two probes (right).

Supplementary Figure 4. Linkage disequilibrium (LD) block between meSNPs of *PCMT1* (A) and *SQRDL* (B) from blood and hippocampal tissues. The SNPs in red represent the meSNPs of *cis*-mQTL from blood sample. The SNPs in green represent the high LD meSNPs of *cis*-mQTL from hippocampal tissue.

Supplementary Figure 5. Mean swimming speed before MWM test among three groups. We found there were no significant differences of mean swimming speed between WT, sham and *Ankrd37* overexpressed groups (F=1.10, P=0.34). WT, wild type; MWM, Morris water maze.

Supplementary Figure 6. Longitudinal trajectories of hippocampal and subfield volumes before and after injection. In the WT group (grey line), all phenotypes (HP, CA1, CA2, CA3 and DG) show negligible changes over time. In the sham group (green line), all phenotypes show a trend toward volumetric reduction over time and the most drastic change is observed in CA1 volume. In the *Ankrd37* group (orange line), all phenotypes show significant volumetric reduction over time. Before injection, all phenotypes show negligible differences among the three groups. After injection, the differences among the three groups become larger and larger and the most significant differences are observed in the latter two time points (MRI-4 and MRI-5). CA, Cornu

Ammonis; DG, dentate gyrus; HP, total hippocampus; WT, wild type; MRI, magnetic resonance imaging. MRI-1 to MRI-5 demonstrate the time point of 24 hours pre-injection, 7-, 14-, 21- and 28- days post-injection for mice.

Supplementary Figure 7. Comparisons of phenotype differences across time points in each experimental group. a. In the WT group, all phenotypes show non-significant differences among the five time points. b. In the sham group, CA1 and CA2 volumes show significant differences among the five time points and only CA1 volume shows progressive volumetric reduction over time. c. In the *Ankrd37* group, all phenotypes show significant differences among the five time points and demonstrate progressive volumetric reduction over time. CA, Cornu Ammonis; DG, dentate gyrus; HP, total hippocampus; MRI, magnetic resonance imaging; WT, wild type; MRI-1 to MRI-5 demonstrate the time point of 24 hours pre-injection, 7-, 14-, 21- and 28- days post injection for mice.

Supplementary Figure 8. Comparisons of hippocampal phenotype differences across groups at the latter three time points. a. At 14-days post-injection (MRI-3), there are significant differences in CA1 (F=4.200, P=0.020) and CA2 (F=3.148, P=0.050) volumes among the three groups and post hoc analyses showed that the *Ankrd37* group has significantly reduced CA1 (P=0.002) and CA2 (P=0.018) volumes than the WT group. b and c. At 21-days post-injection (MRI-4; b) and 28-days post-injection (MRI-5; c), all phenotypes show significant differences among the three groups (all P<0.05). In the post hoc analyses, the *Ankrd37* group shows significantly reduced volumes in all phenotypes than the WT and sham groups except for the CA1 volume differences (P=0.410 for MRI-4; P=0.217 for MRI-5) between the *Ankrd37* and sham groups. The sham group also shows significantly reduced CA1 volume at the two time points (P=0.029 for MRI-4; P=0.003 for MRI-5) than the WT group. CA, Cornu Ammonis; DG, dentate gyrus; HP, total hippocampal volume; MRI, magnetic resonance imaging; WT, wild type; MRI-3, MRI-4 and MRI-5 demonstrate the time point of 14-, 21- and 28- days post-injection for mice.

1

Table

Table 1. Demographics of participants used in specific statistical analysis.

Analysis (data sources)	Required data	Sample size (n)	Age (years)	Gender (Male/Female)	SNP	eProbe/CpG	eGene	Pair
***cis*-eQTL (ADNI)**	S, E	735	73.13 (7.06)	402/333	261,062	4,186 eProbes	2,723	528,079
***cis*-mQTL (ADNI)**	S, M	604	73.29 (7.04)	336/268	1,651,226	114,625 CpGs	16,149	4,966,055
SMR and HEIDI [a]								
S→E→H (ENIGMA)	S, E, H	735	73.13 (7.06)	402/333	260	323 eProbes	229	323
S→E→H (ADNI)	S, E, H	707	73.19 (7.07)	395/312	274	315 eProbes	249	315
S→M→H (ENIGMA)	S, M, H	604	73.29 (7.04)	336/268	2,218	5,330 CpGs	1,983	5,330
S→M→H (ADNI)	S, M, H	585	73.35 (7.05)	331/254	3,223	12,831 CpGs	3,448	12,831
Bayesian coloc test [b]	S, E, M	590	73.71 (6.99)	329/261	16	12 eProbes and 17 CpGs	12	25
CIT (ADNI) [c]	S, E, M	590	73.71 (6.99)	329/261	4	3 eProbes and 3 CpGs	3	4
Spearman correlation								
HYC (IMAGEN)	S, E, M, H	443	14.45 (0.42)	236/207	-	-	-	-
HEC (ADNI)	S, E, M, H	194	74.46 (5.75)	101/93	-	-	-	-
LOAD (ADNI)	S, E, M, H	76	74.07 (7.65)	46/30	-	-	-	-

ADNI, Alzheimer's Disease Neuroimaging Initiative; Bayesian coloc test, Bayesian co-localization test; cis-eQTL, cis-expression quantitative trait loci; cis-mQTL, cis-methylation quantitative trait loci; ENIGMA, Enhancing NeuroImaging Genetics through Meta-Analysis; E, gene expression; H, hippocampal volume; HEC, healthy elderly controls; HEIDI, heterogeneity in dependent instruments; HYC, healthy young controls; LOAD, Late onset Alzheimer's disease; M, CpG methylation; S, Single Nucleotide Polymorphisms; SMR, summary data-based Mendelian Randomization.

[a] *SMR and HEIDI are performed twice: the first is based on GWAS summary data of hippocampal volume from ENIGMA[12] and cis-eQTL and cis-mQTL results from ADNI (the first two rows) and the second is based on GWAS summary data of hippocampal volume, cis-eQTL and cis-mQTL results from ADNI;*

[b] *Only 25 pairs with $PP_{EM}>0.8$ in the Bayesian co-localization test are shown in the table;*

[c] *Only 4 pairs in the $S \rightarrow M \rightarrow E \rightarrow H$ association model of CIT test are shown in the table;*