Neurobiological substrates underlying the effect of genomic risk for depression on conversion of amnestic mild cognitive impairment

Author names and affiliations: Jiayuan Xu [1,8], Qiaojun Li [2,8], Wen Qin [1, 8], Mulin Jun Li [3], Chuanjun Zhuo [1, 4], Huaigui Liu [1], Feng Liu [1], Junping Wang [1], Gunter Schumann [5,6], Chunshui Yu [1, 7*], Alzheimer's Disease Neuroimaging Initiative.

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

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

[3] Collaborative Innovation Center of Tianjin for Medical Epigenetics, Tianjin Key Laboratory of Medical Epigenetics, Department of Pharmacology, Tianjin Medical University, Tianjin 300052, P.R. China.

[4] Department of Psychiatry Functional Neuroimaging Laboratory, Tianjin Mental Health Center, Tianjin Anding Hospital, Tianjin 300052, P.R. China

[5] Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE5 8AF, United Kingdom

[6] Medical Research Council Social, Genetic and Developmental Psychiatry Centre, London SE5 8AF, United Kingdom

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

[8] These authors contributed equally to this work.

***Correspondence to:** Chunshui Yu, Ph.D., Department of Radiology, Tianjin Medical University General Hospital, No. 154, Anshan Road, Heping District, Tianjin 300052, China, E-mail: chunshuiyu@tijmu.edu.cn. Phone: +86-22-60362026; Fax: +86-22-60362290

Content:

Supplementary Method

Discovery sample

This study included two discovery samples. The first discovery sample was used to calculate the PRS_{MDD}, including 9240 MDD patients and 9519 controls provided by the Psychiatric Genomics Consortium (PGC) (Sullivan, 2010). All patients met the criteria for a lifetime history of MDD based on the Diagnostic and Statistical Manual of Mental Disorders, 4th edition (DSM-IV). The genetic summary data from this sample were used to identify the MDD risk variants, reference allele, *P* values, and odd ratios (OR) (http://www.med.unc.edu/pgc/results-and-downloads) (Ripke *et al.*, 2013).

The second discovery sample was used for calculating PRS_{AD}. The International Genomics of Alzheimer's Project (IGAP) (Lambert *et al.*, 2013) is a large, two-stage study based on the GWAS of individuals with European ancestry. In this study, we only used the results of stage 1, which has been used the genotyped and imputed data of 7,055,881 SNPs to meta-analyze four previously published GWAS datasets consisting of 17,008 AD cases and 37,154 controls. The summary meta-GWAS statistics from the IGAP (stage 1) were used to identify the information of SNPs that are associated with the risk for AD (http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php).

Target sample

The target sample included 398 patients with aMCI provided by the first stage of Alzheimer's disease Neuroimaging Initiative (ADNI-1) (http://www.adni-info.org).

Diagnosis of aMCI was made according to the criteria by Petersen (Petersen *et al.*, 1999). After excluding 14 patients only with the baseline data and 9 patients who were diagnosed as normal during the follow-up evaluation, the remaindering 375 aMCI patients were included in the following analysis. According to the follow-up results in July 2015 released by ADNI, aMCI subjects were divided into the conversion (aMCI-C, N= 205) and stable (aMCI-S, N=170) groups. Here, the final diagnosis for patients with follow-up loss was based on the last clinical evaluation. The sample was used to test whether PRS could predict the conversion from aMCI to AD (up to 108 months follow up).

Quality control for individual level

For the 757 subjects from ADNI-1, the quality control was performed using the PLINK version 1.90 beta3 (http://www.cog-genomics.org/plink2/) (Purcell *et al.*, 2007; Chang *et al.*, 2015). We firstly removed 1 subject with a missing genotype rate of greater than 0.05. Then we removed 2 subjects with sex inconsistency based on the X chromosome information. After that, we identified individuals with possible relative relationships by using the estimate of pairwise identity-by-descent (IBD) to find pairs of individuals who had more similar genotypes than we would have expected by chance in a random sample and removed the one of each pair who had the greater missing genotype rate (3 subjects were excluded). The resulting SNP sets was then used to calculate multidimensional scaling (MDS) to assess the population stratification with HapMap phase 3 genetic data as the reference set, 54 participants were excluded from the sample as European population outliers. The first 4

components of MDS analysis were controlled in the subsequent analysis.

Quality control at SNP level

We applied SNP-level filtering to remove SNPs with missing call rates greater than 0.05, a minor allele frequency (MAF) less than 0.01, and significant deviation from Hardy-Weinberg Equilibrium (P < 5.00e-6). Strand ambiguous SNPs were also removed. After individual- and SNP-level quality control, we retained 697 individuals with European ancestry (419 males) with a genotyping rate of 99.66% for 521,695 SNPs.

Imputation

The MaCH (Li *et al.*, 2010) (http://www.sph.umich.edu/csg/abecasis/MACH) was used for haplotype phasing and the MiniMac (Howie *et al.*, 2012) (http://genome.sph.umich.edu/wiki/Minimac) was used for imputation with the 1000 Genomes Phase 1 version 3 CEU as the reference dataset (hg19). Finally, 7,747,882 autosomal SNPs with imputation quality score greater than 0.8 (R^2 > 0.8) were used for further analysis.

GMV calculation

All structural images were visually checked by two experimenters of radiology. In the 697 subjects with qualified genetic data, we removed 19 subjects because of poor image quality. Finally, a total of 322 aMCI patients were finally included in the voxel-based morphometry (VBM) analysis. The GMV maps were calculated using the VBM8 implemented in Statistical Parametric Mapping software package (SPM8, http://www.fil.ion.ucl.ac.uk/spm). In the segmentation of VBM8, an adaptive

Maximum A Posterior technique (Rajapakse *et al.*, 1997) and a Partial Volume Estimation (Tohka *et al.*, 2004) were used to estimate the fraction of each pure tissue type present in every voxel. After the structural images were segmented into gray matter (GM), white matter and cerebrospinal fluid, the individual's GM concentration map was normalized into the Dartel template in Montreal Neurological Institute (MNI) space (http://www.mni.mcgill.ca/). This template was derived from 550 healthy control subjects of the IXI-database (http://www.brain-development.org). In the modulated normalized process, we multiplied the individual's GM concentration map only by the non-linear determinants derived from the spatial normalization procedure. This step resulted in normalized GM density or relative GMV map for each subject. Here, the GMV of each voxel represents the fraction of GM present in each voxel, which preserves the local GM density while removing the confounding effect of variance in individual brain sizes. After that, we resliced the normalized GMV to a 1.5-mm cubic voxel. Finally, the GMV images were smoothed with a kernel of 8 × 8 × 8 mm^3 full width at half maximum. Then, the spatial pre-processing, normalized, modulated, and smoothed GMV maps were used for further analysis.

LD score regression and colocalization analysis

To validate the specificity of these index SNPs to MDD but not to AD, we performed LD score regression (Bulik-Sullivan *et al.*, 2015) and colocalization analyses (Giambartolomei *et al.*, 2014; Pickrell *et al.*, 2016). LD score regression was applied to quantify genetic correlation pattern (r_g) between GWAS summary statistics of the 1,806 MDD-specific index SNPs from PGC-MDD dataset and GWAS summary

statistics from IGAP-AD datasets. By estimating the Bayesian posterior probability, the colocalization analysis was used to test the hypothesis that the 1,806 MDD-specific index SNPs are associated with MDD, but not with AD. The Bayesian posterior probability > 0.90 was used as a cutoff threshold.

Mediation analysis and Mendelian Randomization (MR)

The SPSS macro (http://www.afhayes.com/spss-sas-and-mplus-macros-and-code.html) (Preacher and Hayes, 2008; Hayes, 2013) was used to perform the mediation analysis to test whether the GMV of each significant brain region mediates the association between the PRS_{sMDD} and the status of aMCI-C. The PRS_{sMDD} was defined as an independent variable, the GMV of each significant brain region as a mediator variable, and the aMCI group assignment (aMCI-S vs aMCI-C) as a binary dependent variable. The first step was to confirm that the independent variable (PRS_{sMDD}) was a predictor of the dependent variable (aMCI-S vs aMCI-C), which is known as the direct effect. The second step was to confirm that the independent variable (PRS_{sMDD}) was a predictor of the mediator (GMV). The third step was to confirm that the mediator (GMV) was a predictor of the dependent variable (aMCI-S vs aMCI-C), while controlling for the independent variable (PRS_{sMDD}). The indirect effect is the product of path coefficients of the last two steps. Then the bootstrapping method was used to assess the significance of the mediation effect. After 5000 bias-corrected bootstrapping, we could estimate the distribution of the indirect effect and calculate its 95% confidence intervals (CI). If zero does not fall between the resulting 95% confidence interval of the bootstrapping method, we could confirm the existence of a significant mediation effect ($P < 0.05$).

Mendelian randomization (MR) is a statistical technique that uses genetic variants associated with modifiable exposure as instrumental variables to infer the causal effect of the exposure on outcome under several assumptions, which can overcome confounders and reverse causality (Smith and Ebrahim, 2003; Davey Smith and Hemani, 2014). The pleiotropy may influence on the validity of causal estimates derived from MR methods. It has been suggested that the conventional MR analysis is valid when there is vertical pleiotropy or balanced horizontal pleiotropy. However, when there is unbalanced horizontal pleiotropy, the conventional MR analysis may generate false positive or false negative estimate. In this situation, MR-Egger method could correct for unbalanced horizontal pleiotropy and yield a valid estimate (see Figure 1 in (White *et al.*, 2016) for a recent pictorial description of vertical and balanced/unbalanced horizontal pleiotropy) (Bowden *et al.*, 2015).

Here, the conventional two-stage method of MR analysis was applied using PRS_{sMDD} as instrumental variable (G) to make causal inference between left hippocampal volume (X) and aMCI conversion (Y) (Burgess and Thompson, 2013; Burgess, 2014; Burgess and Thompson, 2015) (http://www.mendelianrandomization.com/index.php/software-code). A total of 1806 index SNPs specific to MDD (r^2 < 0.25 within 250 kb window) were included in calculation of the PRS_{sMDD} after excluding genetic variants common to PRS_{MDD} (P_T = 0.009) and PRS_{AD} (P_T = 0.352). In the first stage, the regression of X on G is fit, which creates a predicted value ($\hat{X}$) of X for each G using equation (1). In the second stage, a logistic regression is fit in which the binary outcome is Y (aMCI-S and

aMCI-C) and the independent variable is $\widehat{X}$ using equation (2).

$$X = \beta_{XG} G + e_X \quad (1)$$

$$\ln\big(P(Y = 1)\big) = \beta_1 \widehat{X} + e_Y \quad (2)$$

To investigate potential bias due to unbalanced horizontal pleiotropy of PRS_{sMDD} on the conversion of aMCI, MR Egger regression method was applied in sensitivity analysis (Bowden *et al.*, 2015) (http://www.mendelianrandomization.com/index.php/software-code). If the estimated intercept of this method differs from zero, this provides evidence that there is unbalanced horizontal pleiotropy. And the slope coefficient from MR Egger regression (β_{Egger}) estimated the causal effect even the presence of unbalanced horizontal pleiotropy (Burgess and Thompson, 2017).

Supplementary Results

PRS_{sMDD} and PRS_{tsAD} predict the conversion from aMCI to AD

After excluding the overlapping SNPs (n=4,711) between the PRS_{MDD} (P_T = 0.009) and PRS_{AD} (P_T = 0.009), the PRS_{sMDD} and PRS_{sAD} could significantly predict the aMCI-C (P = 0.002 and P = 1.05e-9). To balance the number of index SNPs used to construct the PRS, we also created the PRS_{AD} for the top 1,806 AD-specific index SNPs (PRS_{tsAD}). Although the PRS_{tsAD} could significantly predict the status of aMCI-C (P = 0.013), it could only explain 2.63% of variance for the aMCI-C (Table S5). In addition, under the same P_T (P_T = 0.009) as the PRS_{MDD}, the PRS_{AD} could also predict the aMCI-C (P = 0.046) and explained 1.80% of variance for the aMCI-C based on 2,554 index SNPs (Table S5). After excluding the overlapping SNPs (n=184) between the PRS_{MDD} and PRS_{AD} under the same P_T (P_T = 0.009), only the PRS_{sMDD} could significantly predict the aMCI-C (P = 7.51e-5), and explained 6.86% of variance for the aMCI-C (Table S5). These results indicate that the conversion from aMCI to AD is related to only a small number of MDD-specific genetic variants but to a large number of AD-specific genetic variants.

Ten-fold cross validation

For each PRS, we used ten-fold cross validation to test the unbiased prediction accuracy of the PRS on the conversion of aMCI. That is, we randomly divided the 322 aMCI patients into ten disjoint sets (eight sets: each included 32 patients; two set: each included 33 patients), and then used nine of the sets as training data (N =

289 or 290) and the remaining one set as test data (N = 33 or 32). We repeated this process ten times using different possible combinations of training and test sets. For each validation test, we calculated the accuracy of the test set (32 or 33 aMCI patients). For each PRS, the mean accuracy and coefficient of variance (CV) of validation tests are shown in Table S2.

Conversion rates of aMCI in the trisected PRS groups

When PRS_{sMDD} and PRS_{sAD} were trisected into the low, middle and high risk. There were significant differences in conversion rate among the 9 hierarchical PRS groups (P =3.23e-6). In the middle PRS_{sAD} group, aMCI patients with high PRS_{sMDD} showed significantly higher conversion rate than those with low PRS_{sMDD} (59.38% vs 31.58%, P = 0.022). In the high PRS_{sAD} group, the aMCI patients with high PRS_{sMDD} showed higher conversion rate than those with low PRS_{sMDD} (89.65% vs 72.97%, P = 0.047) (Table S6).

Enrichment analyses using genes fine-mapped based on physical position

We remapped the 8,762 SNPs calculated for PRS_{sMDD} into 1,608 genes only based on the physical position of each variant (within 5kb window). The specific enrichment results for the 1,608 genes are as the follows: in the annotation of gene ontology, 540/1,608 genes were enriched in the development process, 910/1,608 in the protein binding, and 783/1,608 in membrane part (Figure S2A and Table S7). Specifically, these genes mainly over-represented in biological processes of the anatomical structure morphogenesis (q_c = 5.27e-5, FDR-BH correction), development process (q_c = 1.83e-4, FDR-BH correction) and cellular developmental process (q_c = 7.48e-4,

FDR-BH correction) (Figure S2B), in the molecular function of the amyloid-beta binding (q_c = 9.40e-5, FDR-BH correction) (Figure S2C), and in the neuron part (q_c = 8.13e-11, FDR-BH correction) and neuron projection (q_c = 2.96e-5, FDR-BH correction) (Figure S2C). And in the KEGG pathway analysis, these genes were also significantly enriched in neuronal development-related axon guidance (q_c = 7.95e-3, FDR-BH correction) (Figure S2C and Table S9).

In the PPI network analysis, we mapped the PRS_{sMDD} fine-mapping 1,608 genes to the PPI network of the BIOGRID (315 unmatched genes were excluded). A tightly connected PPI sub-network was instituted by 928 genes from the remaining 1,293 genes. Using NTA, 928 seed genes and top ten neighboring genes were included in the construction of the final PPI network consisting of 938 genes. APP was also identified as the most functionally neighboring genes (Figure S2D). Table S10 showed that the 938 genes of the final PPI network were significantly enriched in various nervous system related development processes (Figure S2D), such as nervous system development (q_c = 9.91e-6, FDR-BH correction), neuron projection development (q_c = 3.78e-5, FDR-BH correction), neuron development (q_c = 1.01e-4, FDR-BH correction) and generation of neurons (q_c = 4.11e-4, FDR-BH correction). The PPI-based enrichment analysis further confirmed that the PRS_{sMDD} fine-mapping 1,608 genes also involved in the developmental process and amyloid beta binding.

We further explored in which the developmental periods these 1,608 genes were over-represented in the hippocampus. Under a pSI threshold of 0.05, 65 genes showed temporal-specific expression in the hippocampus in the middle-late fetal

developmental period (q_c = 0.002, FDR-BH correction) (Figure S2E). Under the most stringent threshold (pSI = 0.001), CDC20B also showed temporal-specific high expression in the hippocampus in the middle-late fetal stage, which is well consistent with the finding (CDC20B) from the 1,860 genes obtained based on expression patterns in the hippocampus (Figure S2E).

Supplementary Figures

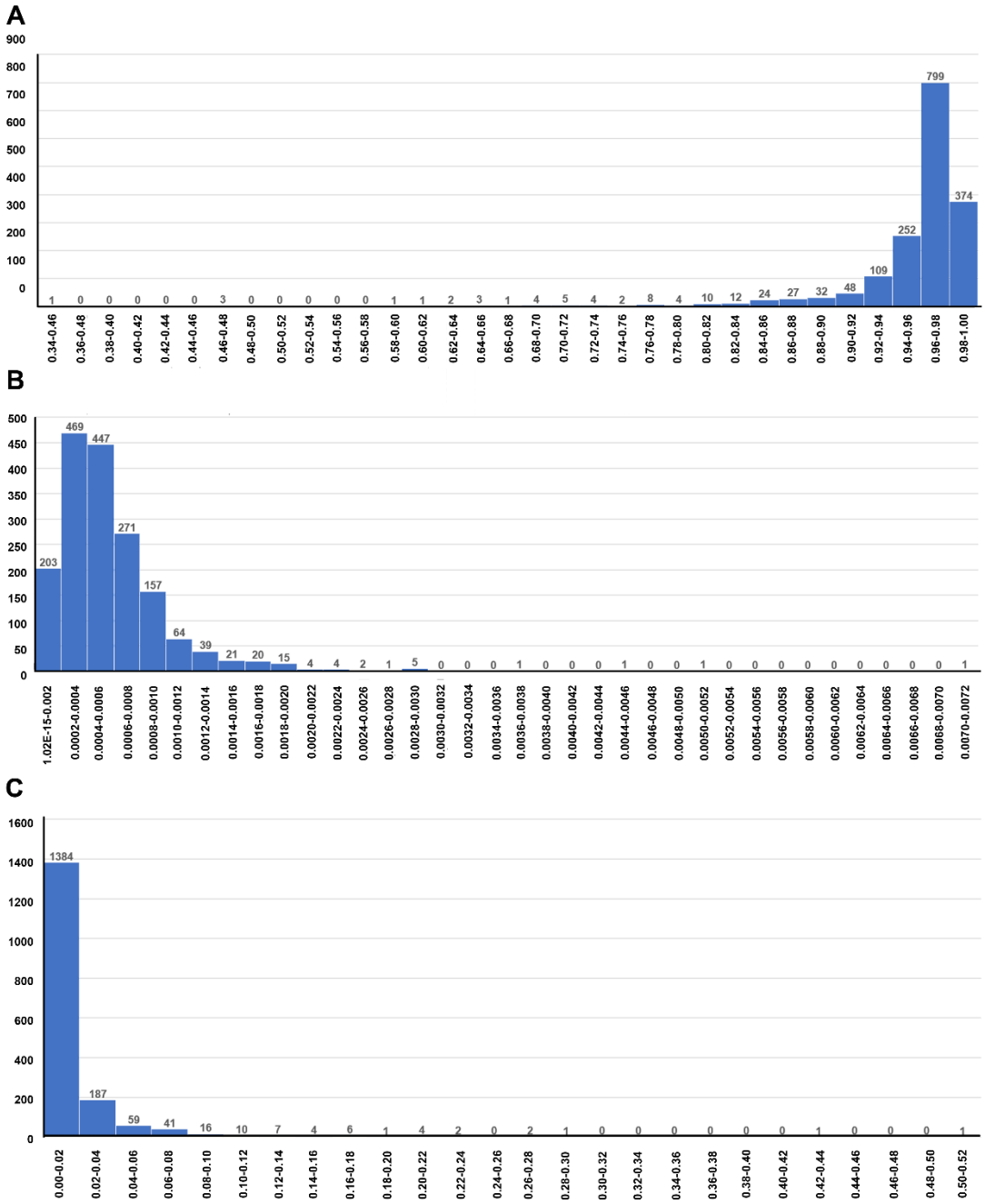

Figure S1. Histograms of Bayesian posterior probability for 1,806 index SNPs. Using PP = 0. 90 as a cutoff threshold, there are 1582/1,806 (87.5%) index SNPs are highly associated with MDD (A), but none SNPs are associated with AD (B), and none SNPs show colocalization between MDD and AD (C). The X-axis represents the intervals of Bayesian posterior probability and Y-axis denotes the number of index SNPs within each interval.

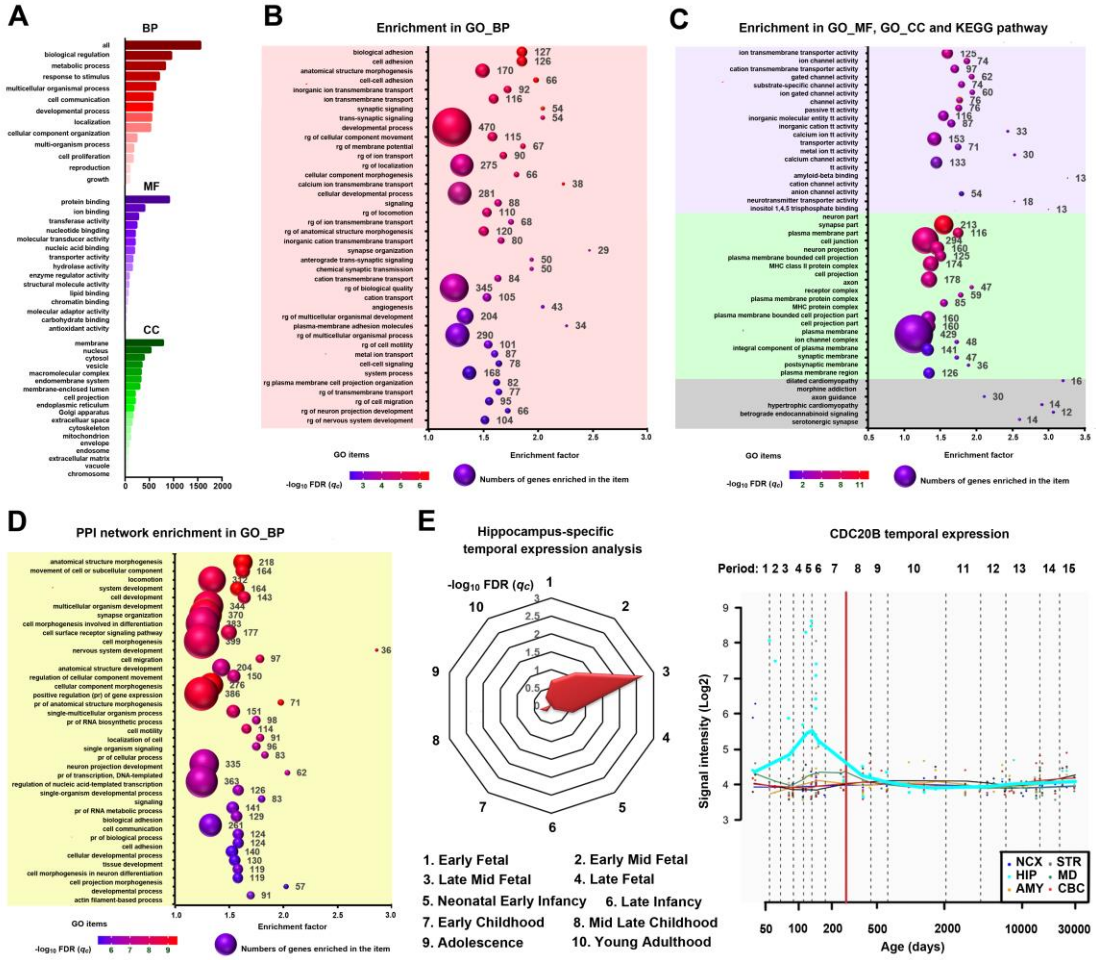

Figure S2. Gene enrichment of the PRS_{sMDD} fine-mapping 1,608 genes. (*A*) Enrichment of the PRS_{sMDD} genes in GO items. The X-axis shows the numbers of genes enriched in each item (Y-axis). The red, purple, green bars denote the biological process, molecular function and cellular component, respectively. (*B*) Top 40 significant enriched GO BP items of the PRS_{sMDD} genes. The X-axis shows enrichment factor of each GO item (Y-axis). The size of balls shows the numbers of genes enriched in each item (numbers of genes are labeled beside the balls). The color of balls demonstrates the significance of the enrichment analysis. (*C*) Top 20 significant enriched GO MF and CC items, and all significant KEGG pathway items of the PRS_{sMDD} genes. The purple, green and grey ground colors show the MF, CC and KEGG pathway, respectively. (*D*) Top 40 significant enriched GO BP items of PPI network. (*E*) Left, PRS_{sMDD} genes were highly expressed in the middle-late fetal developmental period in the hippocampus; right, CDC20B was highly expressed in the early and late mid-fetal development periods in the hippocampus. Period 1-15

have been described in Table S11. AMY, amygdala; BP, biological process; CBC, cerebellar cortex; CC, cellular component; CDC20B, cell division cycle 20B; cp, cell projection; GO, gene ontology; HIP, hippocampus; MF, molecular function; MD, mediodorsal nucleus of the thalamus; NCX, neocortex; PPI, protein-protein interaction; pr, positive regulation; rg, regulation; STR, striatum; tt, transmembrane transporter.

Supplementary Tables

Table S1. Quality control of aMCI patients

Steps	Exclusion No.	Reasons for exclusion	Remainder No.
0	0	The total number of aMCI patients	398
1	14	Only with the baseline data	384
2	9	With a normal diagnosis at the follow-up	375
3	9	Without genotyping data	366
4	1	Sex inconsistence	365
5	30	European population outliers	335
6	13	Poor image quality	322

Table S2. The predictive accuracies of PRS measures on aMCI conversion in ten-fold cross validation

Fold	PRS_{AD} (N = 322)		PRS_{sAD} (N = 322)		PRS_{MDD} (N = 322)		PRS_{sMDD} (N = 322)		$PRS_{sMDD+AD}$ (N = 322)	
No.	Training*	Test	Training*	Test	Training*	Test	Training*	Test	Training*	Test
1	71.25%	66.26%	68.91%	67.24%	65.69%	64.38%	61.38%	60.75%	71.72%	70.63%
2	71.72%	67.10%	68.07%	65.63%	63.45%	60.42%	62.76%	56.88%	78.13%	70.24%
3	70.02%	70.15%	72.45%	63.13%	65.34%	64.34%	61.72%	60.62%	74.24%	70.13%
4	68.17%	68.56%	72.79%	66.67%	65.17%	62.21%	60.55%	59.23%	74.17%	68.97%
5	71.88%	70.36%	72.79%	68.97%	65.34%	60.34%	61.03%	62.88%	69.97%	68.63%
6	70.13%	66.34%	72.38%	68.97%	65.34%	59.75%	62.76%	62.52%	69.97%	69.34%
7	69.31%	65.62%	68.41%	65.63%	65.34%	64.34%	62.41%	59.13%	79.31%	70.63%
8	70.69%	63.56%	68.14%	65.63%	62.41%	60.13%	62.41%	60.13%	74.69%	69.34%
9	70.21%	67.65%	68.14%	68.28%	62.07%	59.98%	62.76%	60.13%	72.28%	69.70%
10	70.24%	66.89%	68.36%	68.70%	64.59%	60.58%	62.63%	60.61%	74.24%	72.63%
M	70.36%	67.25%	70.04%	66.89%	64.47%	61.65%	62.04%	60.29%	73.87%	70.02%
CV	0.02	0.02	0.03	0.03	0.02	0.03	0.01	0.03	0.04	0.02

* The 322 aMCI patients are divided into ten sets, 8 sets include 32 patients for each and 2 sets include 33 patients for each. Thus, the training data include 289 or 290 aMCI patients and the corresponding test data include 33 or 32 patients.

AD, Alzheimer's disease; aMCI, amnestic mild cognitive impairment; CV, coefficient of variance; MDD, major depressive disorder; PRS, polygenic risk scores; PRS_{AD}, PRS for AD-related genetic variants; PRS_{sAD}, PRS for AD-specific genetic variants; PRS_{MDD}, PRS for MDD-related genetic variants; PRS_{sMDD}, PRS for MDD-specific genetic variants; $PRS_{sMDD+AD}$, PRS for MDD-specific and AD-related genetic variants.

Table S3. Top ten known AD locus from AD GWAS meta-analysis [a]

Rank	Gene	SNP	Ethnicity	OR (95% CI)	*P* value	Sample size
1	APOE	APOE_ e2/3/4	All	3.685 (3.30-4.12)	<1.00E-50	4,167
2	BIN1	rs744373	All	1.166 (1.13-1.20)	1.59E-26	49,650
3	CLU	rs11136000	Caucasian	0.879 (0.86-0.90)	3.37E-23	72,432
4	ABCA7	rs3764650	All	1.229 (1.18-1.28)	8.17E-22	60,569
5	CR1	rs3818361	Caucasian	1.174 (1.14-1.21)	4.72E-21	47,052
6	PICALM	rs3851179	Caucasian	0.879 (0.86-0.92)	2.85E-20	44,358
7	MS4A6A	rs610932	All	0.904 (0.88-0.93)	1.81E-11	63,026
8	CD33	rs3865444	All	0.893 (0.86-0.93)	2.04E-10	37,767
9	MS4A4E	rs670139	All	1.079 (1.05-1.11)	9.51E-10	64,577
10	CD2AP	rs9349407	All	1.117 (1.08-1.16)	2.75E-09	35,840

AD, Alzheimer's disease; CI, confidence interval; SNP, single-nucleotide polymorphisms; OR, odd ratios; GWAS, genome-wide association analysis.

[a] These data are from http://www.alzgene.org/TopResults.asp.

Table S4. The predictive effects of PRS on the conversion of aMCI (N=322) after removing SNPs located in the top ten AD loci

	P_T	SNPs	iSNPs	*P*	R^2	Sp	Se	Ac	AUC
PRS_{AD}	0.352	2,622,829	49,829	7.93e-10	18.44%	71.84%	87.17%	80.75%	0.72
PRS_{MDD}	0.009	13,462	2,559	7.49e-5	6.86%	68.89%	83.43%	77.33%	0.65
PRS_{sAD}	NA	2,618,124	49,504	1.05e-9	18.09%	70.24%	85.63%	80.12%	0.70
PRS_{sMDD}	NA	8,756	1,806	1.74e-3	4.19%	65.11%	81.74%	76.01%	0.62
$PRS_{sMDD+AD}$	NA	2,631,585	50,523	2.26e-10	18.71%	70.37%	88.22%	80.86%	0.75

Ac, accuracy; AD, Alzheimer's disease; aMCI, amnestic mild cognitive impairment; AUC, area under curve of receiver operating characteristic curve; iSNPs, numbers of index single-nucleotide polymorphisms that constitute PRS; MDD, major depressive disorder; NA, not applicable; PRS, polygenic risk scores; PRS_{AD}, PRS for AD-related genetic variants; PRS_{MDD}, PRS for MDD-related genetic variants; PRS_{sMDD}, PRS for MDD-specific genetic variants; $PRS_{sMDD+AD}$, PRS for MDD-specific and AD-related genetic variants; P_T, *P* values threshold of genome-wide association studies*;* R^2, Nagelkerke's pseudo R^2 of logistic regression; Se, sensitivity; SNPs, numbers of single-nucleotide polymorphisms that constitute PRS; Sp, specificity.

Table S5. The predictive effects of PRS measures on the conversion of aMCI

	P_T	No. of SNPs	No. of iSNPs	*P*	R^2
PRS_{MDD}	0.009	13,472	2,559	***7.49e-5***	***6.86%***
PRS_{AD}	0.009	100,088	2,554	***4.62e-2***	***1.80%***
PRS_{tsAD}	NA	NA	1,806	***1.30e-2***	***2.63%***
PRS_{sMDD}	NA	13,286	2,535	***7.51e-5***	***6.86%***
PRS_{sAD}	NA	99,904	3,254	1.05e-1	1.10%

aMCI, amnestic mild cognitive impairment; AD, Alzheimer's disease; iSNPs, index SNPs; MDD, major depressive disorder; NA, not applicable; PRS, polygenic risk scores; PRS_{AD}, PRS for AD-related genetic variants; PRS_{MDD}, PRS for MDD-related genetic variants; $PRS_{sAD,}$ PRS for AD-specific genetic variants; $PRS_{sMDD,}$ PRS for MDD-specific genetic variants. $PRS_{tsAD,}$ PRS for the top 1,806 AD-specific genetic variants; P_T, *P* values threshold of genome-wide association studies; R^2, Nagelkerke's pseudo R^2 of logistic regression; and SNP, single-nucleotide polymorphism.

Table S6. Conversion rates of aMCI in the trisected PRS groups

PRS groups	aMCI-S (N)	aMCI-C (N)	Conversion rate
Low PRS_{sAD} and low PRS_{sMDD}	22	11	33.33%
Low PRS_{sAD} and middle PRS_{sMDD}	19	14	42.42%
Low PRS_{sAD} and high PRS_{sMDD}	19	21	52.50%
Middle PRS_{sAD} and low PRS_{sMDD}	26	12	31.58%
Middle PRS_{sAD} and middle PRS_{sMDD}	16	21	56.76%
Middle PRS_{sAD} and high PRS_{sMDD}	13	19	59.38%
High PRS_{sAD} and low PRS_{sMDD}	10	27	72.97%
High PRS_{sAD} and middle PRS_{sMDD}	7	36	83.72%
High PRS_{sAD} and high PRS_{sMDD}	3	26	89.65%
All	135	187	58.07%

AD, Alzheimer's disease; aMCI, amnestic mild cognitive impairment; aMCI-C, conversion from aMCI to AD; aMCI-S, aMCI patients with a stable diagnosis; PRS, polygenic risk scores; $PRS_{sAD,}$ PRS for AD-specific genetic variants; and $PRS_{sMDD,}$ PRS for MDD-specific genetic variants.

Table S7. Numbers of PRS_{sMDD} fine-mapping 1,860 genes and PRS_{sMDD} fine-mapping 1,608 genes observed in the GO category

Items	Description	PRS_{sMDD} fine-mapping 1,860 genes [a]	PRS_{sMDD} fine-mapping 1,608 genes [b]
GO: biological process	biological regulation	901	954
	metabolic process	777	825
	response to stimulus	661	700
	multicellular organismal process	590	627
	cell communication	542	571
	developmental process	505	540
	localization	523	558
	cellular component organization	480	522
	multi-organism process	148	158
	cell proliferation	160	168
	reproduction	83	92
	growth	80	85
	unclassified	561	237
GO: molecular function	protein binding	855	910
	ion binding	381	400
	transferase activity	192	204
	nucleotide binding	189	202
	molecular transducer activity	140	143
	nucleic acid binding	257	273
	transporter activity	144	157
	hydrolase activity	181	189
	enzyme regulator activity	77	79
	structural molecule activity	61	71
	lipid binding	58	58
	chromatin binding	38	41
	molecular adaptor activity	23	24
	carbohydrate binding	19	17
	antioxidant activity	10	11
	unclassified	560	238
GO: cellular component	membrane	738	783
	nucleus	491	525
	cytosol	266	295
	vesicle	302	327
	macromolecular complex	362	388
	endomembrane system	320	341

membrane-enclosed lumen	300	319
cell projection	194	200
endoplasmic reticulum	146	155
Golgi apparatus	127	134
extracelluar space	97	104
cytoskeleton	154	172
mitochondrion	100	103
envelope	69	74
endosome	58	62
extracellular matrix	56	56
vacuole	55	60
chromosome	45	51
unclassified	528	201

GO, gene ontology; PRS_{sMDD}, PRS for MDD-specific genetic variants.

[a] PRS_{sMDD} genetic variants were fine-mapped into 1,860 genes based on the hippocampal-specific regulatory probability between eQTLs and epigenomic features (within a 5kb window)

[b] PRS_{sMDD} genetic variants were fine-mapped into 1,608 genes based on physical position of each variant (within a 5kb window)

Table S8. Gene enrichment analysis of PRS_{sMDD} fine-mapping 1,860 genes

(See accompanying Excel file)

Table S9. Gene enrichment analysis of PRS_{sMDD} fine-mapping 1,608 genes

(See accompanying Excel file)

Table S10. Gene enrichment analysis of the PPI network from PRS_{sMDD} fine-mapping 1,860 genes and PRS_{sMDD} fine-mapping 1,608 genes

(See accompanying Excel file)

Table S11. Period 1-15 in the temporal expression analysis

Period	Description	Age
1	Embryonic	4PCW-8PCW
2	Early fetal	8PCM-10PCW
3	Early fetal	10PCM-13PCW
4	Early mid-fetal	13PCW-16PCW
5	Early mid-fetal	16PCW-19PCW
6	Late mid-fetal	19PCW-24PCW
7	Late fetal	24PCW-38PCW
8	Neonatal and early infancy	0M-6M
9	Late infancy	6M-12M
10	Early childhood	1Y-6Y
11	Middle and late childhood	6Y-12Y
12	Adolescence	12Y-20Y
13	Young adulthood	20Y-40Y
14	Middle adulthood	40Y-60Y
15	Late adulthood	60Y-

M, postnatal months; PCW, post-conceptional weeks; Y, postnatal years.

1

References

Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol 2015; 44(2): 512-25.

Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet 2015; 47(3): 291-5.

Burgess S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. Int J Epidemiol 2014; 43(3): 922-9.

Burgess S, Thompson SG. Use of allele scores as instrumental variables for Mendelian randomization. Int J Epidemiol 2013; 42(4): 1134-44.

Burgess S, Thompson SG. Mendelian Randomization: Methods for Using Genetic Variants in Causal Estimation. Routledge Chapman & Hall 2015.

Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. Eur J Epidemiol 2017; 32(5): 377-89.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 2015; 4: 7.

Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet 2014; 23(R1): R89-98.

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(5): e1004383.

Hayes AF. Introduction to mediation, moderation, and conditional process analysis: A regression-based approach: Guilford Press; 2013.

Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 2012; 44(8): 955-9.

Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 2013; 45(12): 1452-8.

Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 2010; 34(8): 816-34.

Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E. Mild cognitive impairment: clinical characterization and outcome. Arch Neurol 1999; 56(3):

303-8.

Pickrell JK, Berisa T, Liu JZ. Detection and interpretation of shared genetic influences on 42 human traits. Nat Genet 2016; 48(7): 709-17.

Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Methods 2008; 40(3): 879-91.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81(3): 559-75.

Rajapakse JC, Giedd JN, Rapoport JL. Statistical approach to segmentation of single-channel cerebral MR images. IEEE Trans Med Imaging 1997; 16(2): 176-86.

Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, Breen G, *et al.* A mega-analysis of genome-wide association studies for major depressive disorder. Mol Psychiatry 2013; 18(4): 497-511.

Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 2003; 32(1): 1-22.

Sullivan PF. The psychiatric GWAS consortium: big science comes to psychiatry. Neuron 2010; 68(2): 182-6.

Tohka J, Zijdenbos A, Evans A. Fast and robust parameter estimation for statistical partial volume models in brain MRI. Neuroimage 2004; 23(1): 84-97.

White J, Swerdlow DI, Preiss D, Fairhurst-Hunter Z, Keating BJ, Asselbergs FW, *et al.* Association of Lipid Fractions With Risks for Coronary Artery Disease and Diabetes. JAMA Cardiol 2016; 1(6): 692-9.