



Multi-level genomic analyses suggest new genetic variants involved in human memory

Zijian Zhu¹ · Biqing Chen^{1,2} · Hongming Yan¹ · Wan Fang¹ · Qin Zhou³ · Shanbi Zhou⁴ · Han Lei³ · Ailong Huang³ · Tingmei Chen³ · Tianming Gao⁵ · Liang Chen⁶ · Jieyu Chen⁷ · Dongsheng Ni⁸ · Yuping Gu⁸ · Jianing Liu⁸ · Wenxia Zhang¹ · Yi Rao¹

Received: 20 June 2017 / Revised: 5 May 2018 / Accepted: 22 May 2018
© European Society of Human Genetics 2018

Abstract

Development of high-throughput genotyping platforms provides an opportunity to identify new genetic elements related to complex cognitive functions. Taking advantage of multi-level genomic analysis, here we studied the genetic basis of human short-term (STM, $n = 1623$) and long-term (LTM, $n = 1522$) memory functions. Heritability estimation based on single nucleotide polymorphism showed moderate (61%, standard error 35%) heritability of short-term memory but almost zero heritability of long-term memory. We further performed a two-step genome-wide association study, but failed to find any SNPs that could pass genome-wide significance and survive replication at the same time. However, suggestive significance for rs7011450 was found in the shared component of the two STM tasks. Further inspections on its nearby gene *zinc finger and at-hook domain containing* and SNPs around this gene showed suggestive association with STM. In LTM, a polymorphism within *branched chain amino acid transaminase 2* showed suggestive significance in the discovery cohort and has been replicated in another independent population of 1862. Furthermore, we performed a pathway analysis based on the current genomic data and found pathways including mTOR signaling and axon guidance significantly associated with STM capacity. These findings warrant further replication in other larger populations.

These authors contributed equally: Zijian Zhu, Biqing Chen.

These authors jointly supervised this work: Zijian Zhu, Biqing Chen.

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41431-018-0201-8>) contains supplementary material, which is available to authorized users.

✉ Zijian Zhu
zhuzijian0203@163.com

✉ Biqing Chen
bq_chen@pku.edu.cn

¹ PKU-IDG/McGovern Institute for Brain Research, Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Beijing Innovation Center for Genomics, Peking University, 100871 Beijing, China

² Central Laboratory, Affiliated Hospital of Nanjing University of Chinese Medicine, Jiangsu Province Hospital of Traditional Chinese Medicine, 210029 Nanjing, China

³ College of Laboratory Medicine, Chongqing Medical University, 400016 Chongqing, China

Introduction

Memory plays a pivotal role in human life. The memory system encodes, processes, and stores information from the outside world, which allows the information to serve for normal cognitive functions. Deficits in these processes could cause severe cognitive dysfunctions

⁴ University-Town Hospital of Chongqing Medical University, 401331 Chongqing, China

⁵ State Key Laboratory of Organ Failure Research, Guangdong Provincial Key Laboratory of Psychiatric Disorders, Collaborative Innovation Center for Brain Science, Department of Neurobiology, Southern Medical University, 510515 Guangzhou, China

⁶ Biostime institute of nutrition and care, 510623 Guangzhou, China

⁷ School of Traditional Chinese Medicine, Southern Medical University, 510515 Guangzhou, China

⁸ Division of Molecular Nephrology and Creative Training Center for Undergraduates, M.O.E. Key Laboratory of Medical Diagnostics, College of Laboratory Medicine, Chongqing Medical University, 400016 Chongqing, China

[1, 2]. Throughout the past decades, neuroscientists have identified brain regions and networks for memory functions, revealing the neural mechanisms of human memory at the brain level [3–5]. Electrophysiological and transgenic studies have further illustrated the molecular basis of memory functions [6, 7]. However, knowledge of the genetic basis of human memory functions is still limited.

There is plenty of evidence suggesting genetic basis of memory functions. First, substantial heritability of memory ability has been demonstrated in twins and family studies, which estimated the heritability of working memory or short-term memory (STM) to be 15%–72% [8–10] and long-term memory (LTM) to be moderately heritable (37%–55%) [11]. Using high-throughput single-nucleotide polymorphism (SNP) data, one recent study demonstrated SNP-based heritability for working memory being 31%–41% [12]. The non-zero heritability estimations implicate potential specific molecular loci that might contribute to memory functions. Recent advances in high-resolution genome-wide association studies (GWAS) of polygenic phenotypes provide an opportunity to identify genomic sequences that contribute to complicated cognitive processes, without a priori assumption. Papassotiropoulos et al. [13, 14] have first employed this technique in human memory functions. In their early study, they reported six genetic variations associated with human working memory performance and highlighted the role of *SCN1A* and *KIBRA* in memory performance [13, 14]. Later, they explored the genetic basis of human long-term episodic memory and suggested genetic associations of *CTNBL1* with LTM capacity [15]. Recently, their group has revealed the voltage-gated cation channel activity gene set, which is related to neuronal excitability, to be linked with working memory [3].

Employing a similar multi-level GWAS method, here we explored the genetic basis of both STM and LTM functions. We used two behavioral assays to test STM and one assay to test LTM in order to get a better understanding of memory encoding and storage in human brain. First, by estimating the proportion of phenotypic variance explained by all common SNPs across the genome, we obtained a rough estimation of heritability for STM and LTM. Then, we adopted a two-stage GWAS procedure to identify associated SNPs and genes. The first stage (discovery stage) intended to discover a small set of SNPs with suggestive significant ($p < 1 \times 10^{-4}$) association with memory capacity. At the second stage (replication stage), replication was performed in another two independent cohorts to confirm the association of these candidates found at the discovery stage.

Materials and methods

Participants

The GWAS discovery cohort consisted of ege students recruited from the Chongqing Medical University in China, with an average age of 18 ± 1 years (mean \pm standard deviation (S.D.)), 81% female, 94% Han People with the remainder from minor ethnic groups who did not significantly differ from Han People on the genetic structure (from population stratification analysis). The replication cohorts consisted of Chinese young adults recruited from the Southern Medical University at Guangzhou (cohort_GZ), from universities at Beijing (cohort_BJ) and from the Chongqing Medical University at Chongqing (cohort_CQ), China. In the current study, cohort_GZ and cohort_BJ were always combined together (cohort_GZ + BJ) for replication since there was no significant difference in terms of any demographic factor. Cohort_GZ + BJ aged 22 ± 3 years old, 46% female, 95% Han; cohort_CQ aged 18 ± 1 years old, 81% female, 94% Han. No significant difference in demographic factors between the discovery and replication cohorts was found. None of the subjects from the discovery or replication cohorts had reported neurological diseases. Informed consent was obtained from all participants. All procedures performed in the study were in accordance with the ethical standards of the institutional ethical committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Behavioral assays

Short-term memory

STM was estimated by two paradigms, testing digit span, and visuospatial memory capacity, respectively.

Digit-span STM test At the beginning, seven digits were presented on the screen sequentially, each for one second (s). Then, subjects were required to type out all the seven digits in the correct order. If subjects were correct, they would enter the next level, at which the number of digits was increased by one. If subjects failed, the task would repeat at the present difficulty level. The task ended when one failed three times at a given difficulty level, and then the digit span at that difficulty level was recorded as the subject's digit-span STM. The digits and their orders were generated randomly by a custom-made Matlab program.

Visuospatial STM test Visuospatial task employed a similar procedure. Twenty squares were presented on the screen. At first, six squares flashed sequentially, each for 1 s. Then, subjects were required to report the locations of the flashed

squares in the correct order by clicking the corresponding squares sequentially on the screen. Rules for increasing the difficulty level and ending of the experiment were the same as those for digit-span task (details in Supplementary Methods). The number of squares at the final difficulty level was recorded as the subject's visuospatial STM.

Long-term memory

LTM was tested by a delayed recognition task. Subjects first learnt 50 semantically unrelated Chinese words twice. During learning, the words were presented sequentially, each for 0.5 s. Immediately after learning, a distracting task with 20 arithmetic questions was given. Then, a recognition test for all learnt words was given with each word presented among seven distractors in each trial. Subject's LTM capacity was represented by the overall recognition accuracy.

Genotyping and quality control

Discovery cohort

DNA was extracted from peripheral blood of participants using the QuickGene Whole Blood Genome DNA Extract System (Kurabo Industries Ltd., Japan), and was genotyped for 894,517 SNPs using the HumanOmniZhongHua-8 Beadchip v1.1 (Illumina, Inc., San Diego, CA, USA) in the discovery cohort. Common quality control parameters were applied and retained 830,937 SNPs. The inclusion criteria for SNPs were as follows: call rate >0.95 , minor allele frequency (MAF) >0.01 and Hardy-Weinberg equilibrium test with $p > 1 \times 10^{-4}$. Individuals were excluded if their genotype call rate < 0.95 . Potential duplicates or close relatives were examined by identity-by-state (IBS) analysis, and none was excluded due to IBS distance <0.75 . Population stratification was examined with EIGENSTRAT [16]; and outliers were detected and excluded from subsequent analyses with the default mode. A total of 1623/1522 subjects with both phenotypic and genotypic data available for STM/LTM were included finally (Supplementary Table 1).

Replication cohorts

The replication sample was three independent cohorts. The two STM tasks were measured in cohort_GZ + BJ, which has 2790 individuals genotyped on the candidate SNPs via Sequenom iPLEX (Bio Miao Biological, Inc., Beijing, China). SNPs with call rate $<95\%$ (rs2469860, chr17:g.18880661T $>C$) and individuals with genotype rate $<75\%$ (47 subjects) were removed from subsequent analyses. LTM was measured in the cohort_CQ, which consisted of 1862 unrelated

individuals genotyped using the HumanOmniZhongHua-8 Beadchip v1.2. SNPs and individuals with call rate $<95\%$ were removed. No significant difference in allele frequencies was observed between the discovery and replication cohorts. Data have been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001002875; genotype data has been additionally deposited in the Genome Variation Map (GVM) in Big Data Center, Beijing Institute of Genomics, Chinese Academy of Science, under the accession number GVM000023.

Statistical genetic analyses

All traits were inverse normal transformed. Individuals with phenotypes outside four standard deviations of the population mean were removed from subsequent analyses (Supplementary Table 1).

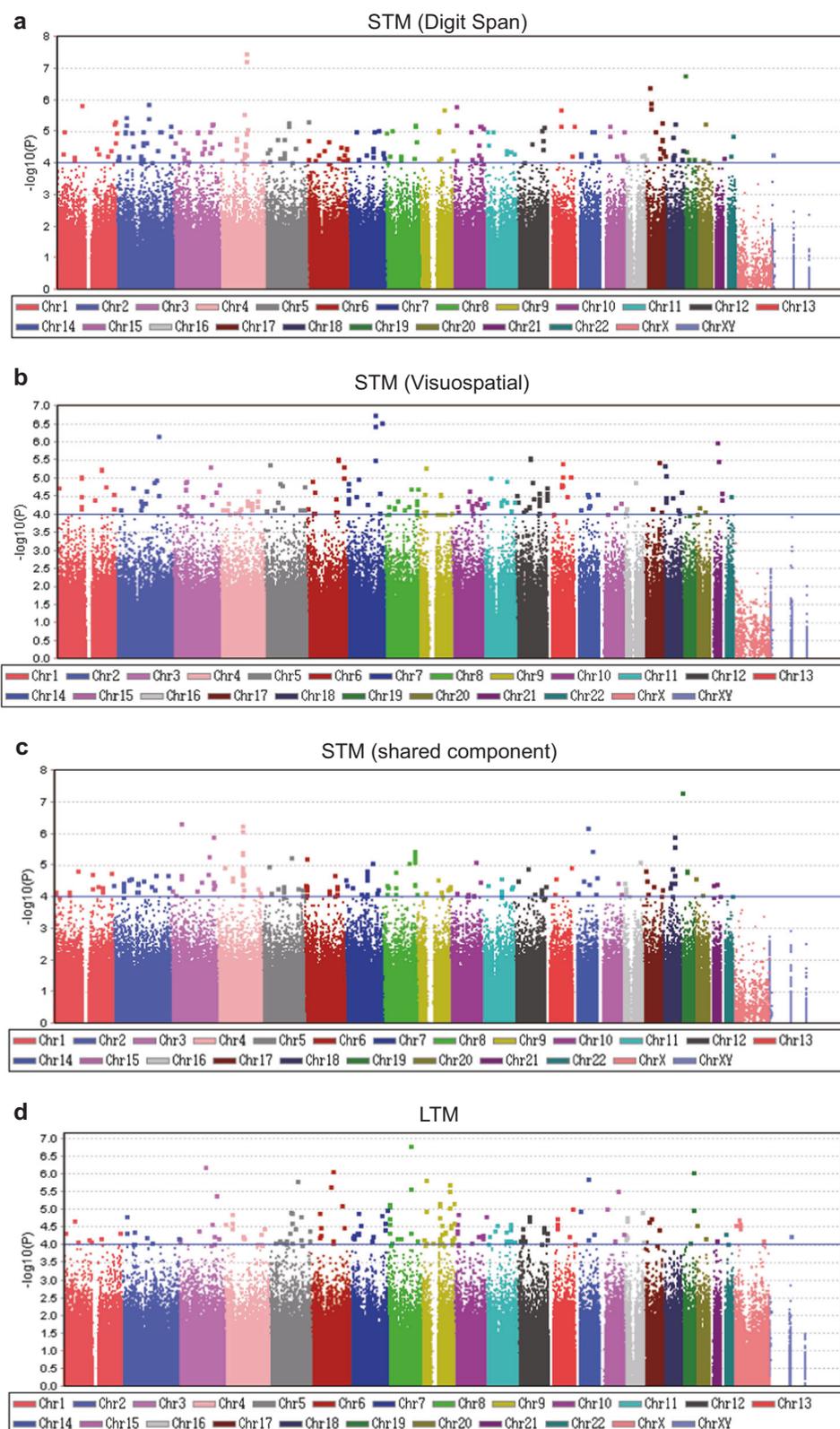
Heritability estimation

SNP-based heritability estimation was conducted by GCTA version 1.24 [17]. The genetic relationship matrix (GRM) was estimated using all autosomal markers (MAF >0.01). One individual in each pair with genetic relatedness >0.025 was excluded [18]. 1604, 1569, and 3326 individuals were retained for digit-span STM, visuospatial STM, and LTM, respectively. Then, the proportion of phenotypic variance explained by all common SNPs was estimated by restricted maximum likelihood (REML) algorithm, with the first 20 eigenvectors from GCTA-PCA included as covariates. The GRM-REML method was adopted to quantify the SNP-based heritability, assuming an additive model. Statistical power was calculated via the online GCTA Power Calculator [19].

GWAS

GWAS was performed in the discovery cohort using PLINK [20]. Quantitative traits were assumed and GWAS was run under a full linear model assumption, testing both additive genetic effect and dominance component (dominance deviation from additivity). Effects of demographic factors, i.e., gender, age, and ethnicity, were tested using analysis of variance (ANOVA). Finally, gender and age were included as covariates for all STM-related phenotypes. The full set of p -values that emerged from association analysis was loaded and visualized in Haploview v4.2 [21] to generate Manhattan plots. Basic statistical analyses, the genomic inflation factor λ , quantile-quantile (Q-Q) plots, and the replication sample size were generated by R v3.2.1. The post-hoc power was calculated by Quanto v1.2 (<http://hydra.usc.edu/gxe>). We additionally did GWAS

Fig. 1 Manhattan plots of memory-related phenotypes. **a** Manhattan plot for STM measured in digit span task; **b** Manhattan plot for STM measured in visuospatial memory task; **c** Manhattan plot for the shared component of STM measured in digit span task and in visuospatial memory task; **d** Manhattan plot for LTM measurements. Results plotted are based on association tests considering both additive and dominant models; the one with smaller p -value is plotted. Chromosomes are shown in different colors for clarity. The blue line indicates suggestive significance level ($p = 1 \times 10^{-4}$). Plots were generated by Haploview



on the imputed genotypes for the STM-shared component (the first principal component from principle component analysis of the two STM tests) in the discovery cohort and

LTM in the combined cohort. Imputation and subsequent association tests were performed using SHAPEIT [22, 23], IMPUTE v2.3.1 [24], and SNPTEST v2.5 [25], based on

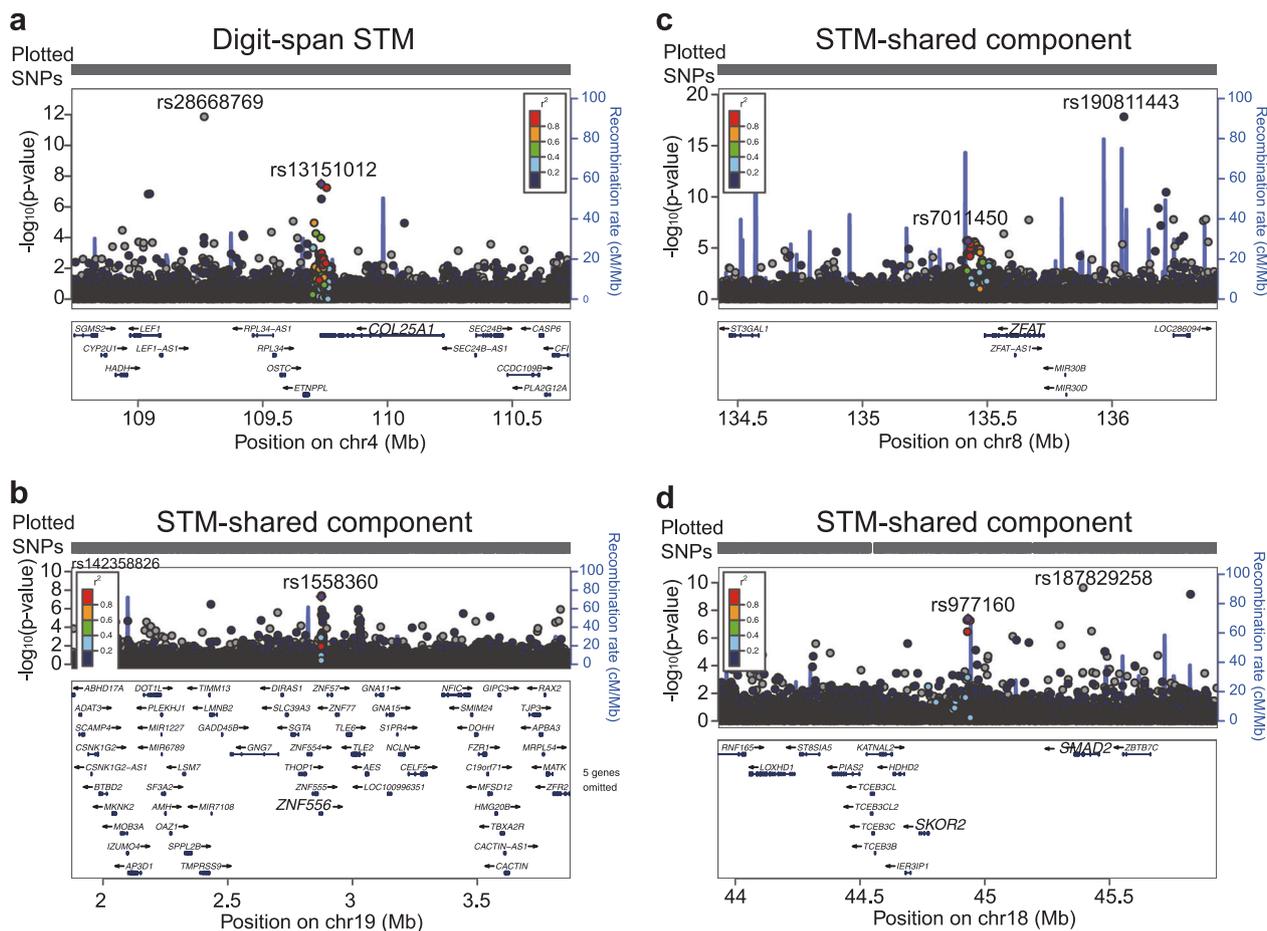


Fig. 2 Regional association plots of top associated loci for short-term memory. **a** rs13151012 (the gene *COL25A1*) associated with digit-span STM; **b** rs1558360 (the gene *ZNF556*), **c** rs7011450 (the gene *ZFAT*), and **d** rs977160 and rs5824676 (the gene *SKOR2*) associated with STM-shared component. Imputed genotypes in the discovery cohort were used in association tests for regional plots. Genome Build

is *hg19/1000 Genomes Nov 2014 ASN*. The gray dots represent SNPs that are not in linkage disequilibrium with the SNP hit. Figures are plotted using the web-based LocusZoom program. The mean LD in terms of R^2 between top SNPs deriving from GWAS of directly typed data and those from GWAS of imputed data from the same genomic regions (all with p -value $< 1 \times 10^{-4}$) was ~ 0.8

1000 Genomes, assuming an additive model. Only SNPs with $INFO > 0.6$ and $MAF > 0.1$ were considered. Regional associations were plotted on the imputed data using the web-based LocusZoom [26] (details in Supplementary Methods).

Candidate selection

We selected the top 10 or 20 most significantly associated SNPs in the discovery cohort as candidates, i.e., 10 for each STM, 20 for LTM because the top 10 SNPs showed no significant association. SNPs with a more significant candidate nearby (within 30 kilo-base (kb)) were not considered for further replication. SNPs with $MAF < 0.1$ were excluded since our sample was small so that there were not enough sample in each genotype group for SNPs with $MAF < 0.1$. Replication tests assumed a general linear

regression, taking the minor allele as the effect allele. The combined p -values for meta-analysis were calculated by Stouffer’s Z -score method. Successful replication was defined as false discovery rate corrected p -value (q) < 0.05 in the replication cohort and in the discovery and replication combined sample. Genome reference in this paper is human genome assembly *hg19*.

Gene-based and pathway analyses

All SNPs with their association results from the previous single-marker based GWAS were used in the gene-based and pathway-based analyses. Gene-based analysis was performed by VEGAS [27]. SNP to gene mapping was based on *hg18*. Gene boundaries were set at 50 kb. Linkage disequilibrium (LD) patterns were estimated based on HapMap2 CHB + JPT. Pathway analysis

Table 1 Associations of SNPs selected for replication in different cohorts

SNP	Position	Gene	Alle	Freq	β_{discover}	β_{rep}	P_{discover}	P_{rep}	P_{meta}
<i>Digit-span STM</i>									
rs13151012	4:109733172	<i>COL25A1</i>	C/T	0.44, 0.45	-0.14 ± 0.02	0.004 ± 0.03	3.29E-08	0.87	8.00E-03
rs1558360	19:2873323	<i>ZNF556</i>	T/C	0.14, 0.14	-0.24 ± 0.05	-0.008 ± 0.04	1.61E-07	0.83	4.70E-03
rs16842477	2:133608499	<i>NCKAP5</i>	C/T	0.11, 0.11	-0.23 ± 0.05	0.02 ± 0.04	1.29E-06	0.65	3.94E-02
rs3806237	1:100715782	<i>DBT</i>	C/T	0.11, 0.11	0.26 ± 0.05	0.04 ± 0.04	1.39E-06	0.36	1.30E-03
rs11257948	10:12747672	<i>CAMK1D</i>	A/G	0.35, 0.35	-0.13 ± 0.03	-0.0001 ± 0.03	1.50E-06	1	1.51E-02
rs2472716	17:18879761	<i>FAM83G/SLC5A10</i>	A/G	0.24, 0.24	-0.16 ± 0.03	-0.06 ± 0.03	1.77E-06	0.04	2.93E-05
<i>Visuospatial STM</i>									
rs199750188	7:115850196	<i>TES</i>	T/C	0.29, 1.00	-0.15 ± 0.03	0.0001 ± 0.02	1.74E-07	0.99	8.50E-03
rs2969363	2:177483123	<i>MTX2</i>	A/G	0.22, 0.18	-0.18 ± 0.04	0.02 ± 0.03	6.35E-07	0.45	6.29E-02
rs2830960	21:28839569	– ^a	T/G	0.14, 0.14	-0.21 ± 0.04	-0.04 ± 0.04	1.01E-06	0.26	6.00E-04
rs2204679	7:115943514	<i>TES</i>	A/G	0.19, 0.21	-0.17 ± 0.04	0.04 ± 0.03	2.99E-06	0.17	2.42E-01
<i>The shared component of Digit-span and Visuospatial STM</i>									
rs1558360	19:2873323	<i>ZNF556</i>	T/C	0.14, 0.14	-0.25 ± 0.05	0.02 ± 0.04	4.97E-08	0.64	1.88E-02
rs3886330	4:105772094	<i>TET2</i>	T/C	0.49, 0.47	-0.12 ± 0.02	-0.03 ± 0.03	5.29E-07	0.27	5.00E-04
rs9637619	4:105845011	<i>TET2</i>	A/G	0.47, 0.47	-0.12 ± 0.02	-0.007 ± 0.03	7.80E-07	0.8	6.70E-03
rs4456603	18:48810587	<i>MEX3C</i>	A/G	0.17, 0.16	0.18 ± 0.04	0.01 ± 0.04	2.40E-06	0.73	7.60E-03
rs7011450	8:135419683	<i>ZFAT</i>	G/A	0.38, 0.39	0.12 ± 0.03	-0.02 ± 0.03	3.49E-06	0.47	8.73E-02
rs7140168	14:81801538	<i>STON2</i>	C/T	0.46, 0.46	-0.11 ± 0.02	0.02 ± 0.03	3.50E-06	0.56	6.58E-02

The genomic reference sequence used here is *hg19*. Position, genomic position in the form of chromosome: basepair. Alle, listed are in the form of variant allele / reference allele, which is the same in the discovery and replication cohorts. Freq, effect allele frequency in the discovery (former) and replication (latter) cohorts. β , effect size at each stage, in the form of mean \pm standard error

rep replication, *meta* meta-analysis

^aThere is no gene within 500 kb nearby the SNP. The sample size for discovery cohort was 1621/1623 for digit-span STM /visuospatial STM, while for replication cohort, the number was 2788/2790 for digit-span STM /visuospatial STM. The cohort GZ + BJ was used for replication

adopted MAGENTA [28], taking Gene Ontology, KEGG, Reactome, BioCarta, PANTHER, and INGENUITY databases (3216 in total) (details in Supplementary Methods).

accuracy for LTM was 56% (S.D. = 13%) with a range of 11% ~ 98%. The estimated SNP-based heritability for LTM was zero (s.e. = 17.5%).

Results

SNP-based heritability estimation

Participants' memory performance varied dramatically among individuals (see Supplementary Results). The mean digit span was 10.6 (S.D. = 1.6) digit numbers, while the mean visuospatial span was 7.2 (S.D. = 1.1) Corsi blocks. The two STM measures were significantly correlated with each other, both phenotypically (Pearson correlation coefficient $r = 0.27$, $p < 0.001$) and genetically (genetic correlation $r = 0.48$ (s.e. = 0.28)). Results showed that 60.6% (standard error (s.e.) = 34.7%, $p = 0.04$) of the variance in digit-span STM could be explained by common SNPs; for visuospatial STM, the result was 24.4% (s.e. = 25.8%, $p = 0.17$); for the STM-shared component, it was 32.0% (s.e. = 25.8%, $p = 0.11$). These non-significant results might be due to the relatively small sample size. The mean recall

GWAS and gene-based analyses for STM

The genomic inflation factor λ was 1.00 in all association tests, indicating that the significance reported herein was not affected by population stratification (Supplementary Figure 1). Association results across the whole genome are shown in 'Manhattan' plots (Fig. 1). For digit-span STM, one SNP (rs13151012, chr4:g.109733172T > C), located in an intron of *collagen type XXV alpha 1 chain (COL25A1)* and predicted to be in open chromatin region (Supplementary Table 2, by 3DSNP: <http://cbportal.org/3dsnp>), reached $p = 3.3 \times 10^{-8}$ (Fig. 2a), but it was not replicated ($p_{\text{replication}} > 0.05$, Table 1). Besides this SNP, we selected another nine most significant SNPs as candidates. Two SNPs were close to each other so one was discarded from subsequent analyses. We genotyped nine SNPs for replication, but three of them failed quality control at the replication stage, thus finally six SNPs have replication results. rs2472716 (chr17:g.18879761G > A), in an intron of *family with*

Table 2 Top SNPs associated with LTM task

SNP	Position	Gene	Alle	Freq	$\beta_{\text{discovery}}$	$\beta_{\text{replication}}$	$P_{\text{discovery}}$	P_{rep}	P_{meta}
rs4805097	19:35302088	<i>ZNF599</i>	C/T	0.265	0.16 ± 0.03	0.03 ± 0.03	4.30E-07	0.271	6.91E-06
rs11087617	20:4063234	<i>SMOX</i>	C/T	0.246	-0.16 ± 0.03	-0.01 ± 0.03	5.10E-07	0.865	4.30E-04
rs17204340	9:113080316	<i>TXNDC8</i>	T/G	0.426	0.13 ± 0.03	-0.02 ± 0.02	1.56E-06	0.509	8.11E-03
rs959692	8:117899211	<i>RAD21</i>	G/A	0.455	-0.12 ± 0.03	-0.02 ± 0.02	3.83E-06	0.305	9.11E-05
rs72634650	10:53080211	<i>PRKG1</i>	A/G	0.374	0.12 ± 0.03	0 ± 0.03	4.80E-06	0.867	3.06E-03
rs4823400	22:45253926	<i>ARHGAP8</i>	A/G	0.432	0.12 ± 0.03	-0.01 ± 0.02	8.36E-06	0.833	6.50E-04
rs33375	5:171066113	<i>SMIM23</i>	T/C	0.295	-0.14 ± 0.03	-0.01 ± 0.03	9.60E-06	0.661	1.20E-03
rs9369426	6:43811268	<i>VEGFA</i>	T/C	0.437	-0.12 ± 0.03	-0.01 ± 0.02	1.02E-05	0.774	2.21E-03
rs837642	19:49307999	<i>BCAT2</i>	G/A	0.413	0.12 ± 0.03	0.08 ± 0.02	1.32E-05	0.001	3.68E-07
rs9369228	6:40698365	<i>LRFN2</i>	T/C	0.173	-0.18 ± 0.04	0 ± 0.04	1.38E-05	0.953	1.60E-03
rs12655793	5:18764710	- ^a	A/G	0.24	-0.15 ± 0.03	0.04 ± 0.03	1.59E-05	0.224	1.94E-02
rs10908431	1:154716883	<i>KCNN3</i>	T/C	0.417	-0.11 ± 0.03	-0.01 ± 0.02	1.76E-05	0.642	5.25E-04
rs10820865	9:94279402	<i>NFIL3</i>	C/T	0.394	0.12 ± 0.03	-0.03 ± 0.02	1.80E-05	0.148	5.54E-02

The genomic reference sequence used here is *hg19*. Position, genomic position in the form of chromosome: basepair. Alle, listed are in the form of variant allele / reference allele, which is the same in the discovery and replication cohorts. Freq, effect allele frequency in the discovery and replication cohorts. $\beta_{\text{discovery}}/\beta_{\text{replication}}$, effect size at the discovery / replication stage, in the form of mean ± standard error

rep replication; *meta* meta-analysis

^aThere is no gene within 500 kb nearby the SNP. The sample size for discovery cohort was 1522, while for replication cohort, the number was 1862. The cohort CQ was used for replication

sequence similarity 83 member G (*FAM83G*) and solute carrier family 5 member 10 (*SLC5A10*) and predicted to be in enhancers (Supplementary Table 2), reached uncorrected significance in replication ($p_{\text{discovery}} = 1.8 \times 10^{-6}$, $p_{\text{replication}} = 0.04$, $p_{\text{joint}} = 2.9 \times 10^{-5}$, Table 1). Gene-based analysis gave consistent results with gene-level association $p = 1.5 \times 10^{-5} / 1.3 \times 10^{-4}$ for *FAM83G / SLC5A10* in the combined sample.

For visuospatial STM, none of the tested markers passed the acceptable genome-wide significance in the discovery cohort. In case of insufficient sensitivity, the 10 most significant SNPs were picked as candidates; according to the candidate selection rules, seven from the 10 SNPs were genotyped for replication. Three SNPs failed replication quality control. None of the remaining four tested SNPs was significant in the replication tests (Table 1).

Considering the possible existence of confounding factors that might affected estimations of the pure memory capacity in the two STM tests (see Supplementary Results), another GWAS was done with the STM-shared component as a phenotype. Results showed rs1558360 (chr19: g.2873323C > T), in an intron of *zinc finger protein 556* (*ZNF556*) and predicted to be in enhancers (Supplementary Table 2), reached $p = 5.0 \times 10^{-8}$ in the discovery cohort (Fig. 2b). However, its association was not replicated ($p_{\text{replication}} = 0.64$, Table 1). Among the 10 most significant SNPs, three failed candidate selection rules and were removed from further tests; seven SNPs were genotyped in the replication cohort; one SNP failed quality control at the replication stage; finally, six SNPs were tested for

replication. None was replicated with STM-shared component (Table 1). However, we found rs7011450 (chr8: g.135419683 A > G) reaching uncorrected significance in association with the visuospatial STM in the replication cohort (rs7011450 associations in visuospatial STM: $p_{\text{discovery}} = 8.4 \times 10^{-5}$, $p_{\text{replication}} = 0.01$, $p_{\text{joint}} = 2.5 \times 10^{-5}$, Supplementary Table 3), and its nearby gene is *zinc finger and at-hook domain containing* (*ZFAT*). Further inspection showed that five SNPs within 70 kb around *ZFAT* reached suggestive significance ($p < 1 \times 10^{-4}$) in association with both visuospatial STM and STM-shared component at the discovery stage and were significantly ($p < 0.05$) associated with visuospatial STM at the replication stage (Supplementary Table 3). After imputation, more variants around this locus were found to be associated with STM-shared component (Supplementary Table 4) and one imputed variant in an intron of *ZFAT*, rs182335917 (chr8: g.135662957G > A), reached genome-wide significance in the discovery cohort ($p = 1.8 \times 10^{-8}$) (Fig. 2c). Gene-based analysis showed that *ZFAT* was significantly associated with both STM-shared component and visuospatial STM, with $p = 0.0084 / 0.018$, respectively.

We further performed imputation and subsequent association tests to improve the coverage of genomic data (Supplementary Table 4). Two variants displayed genome-wide significant associations with $\text{INFO} > 0.8$ and $0.05 < \text{MAF} < 0.1$ in the discovery cohort, namely, rs977160 (chr18: g.44929658T > C, major/minor allele = T/C, $\beta = 0.37 \pm 0.07$, $p = 4.0 \times 10^{-8}$) and rs5824676 (chr18: g.44931008:44931009insGGG, major/minor allele = A/AGGG, $\beta = 0.37 \pm 0.07$,

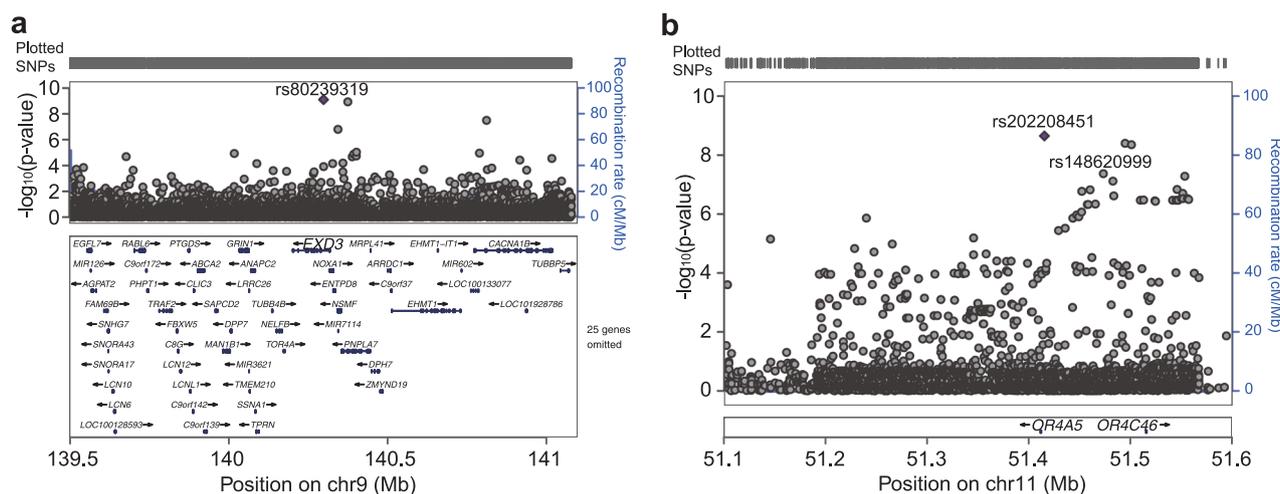


Fig. 3 Regional association plots for genome-wide significant loci in association with long-term memory. **a** rs80239319 and **b** rs148620999. Imputed genotypes in the discovery and replication combined cohorts were used. Genome Build is *hg19/1000 Genomes*

Nov 2014 ASN. The gray dots represent SNPs that are not in linkage disequilibrium with the SNP hit. Figures are plotted using the web-based LocusZoom program

Table 3 Top pathways associated with memory performance

Gene set	Phenotype	OBS/EXP	<i>p</i> -Value
glioma	DG	9/3	2.64E-02
mTOR signaling pathway	DG	6/3	3.83E-02
axon guidance	DG	11/6	3.64E-02
regulation of autophagy	CB	6/1	2.74E-02
mRNA end processing and stability	CB	2/1	2.20E-02
Ephrin receptor signaling	DG-CB shared	3/2	4.94E-02
Olfaction	LTM	8/6	7.80E-03

OBS/EXP observed number of genes versus expected number of genes, *DG* digit-span STM, *CB* visuospatial STM based on Corsi Block, *DG-CB shared* the shared component between DG and CB, *LTM* long term memory

$p = 4.2 \times 10^{-8}$); they are located ~155 kb upstream from *SKI family transcriptional corepressor 2 (SKOR2)* (Fig. 2d) and predicted to be in enhancers (Supplementary Table 2). No SNP that passed imputation quality control reached genome-wide significance, thus no replication was conducted.

GWAS and gene-based analyses for LTM

None of the tested SNPs reached genome-wide significance for LTM. We first took the top 10 most significant SNPs as candidates and genotyped those that fit our candidate selection rules, but none of them remained significant after multiple testing corrections. Thus, the second 10 most significant SNPs were picked and only those that fit our selection rules described in the Methods were genotyped. Finally, 13 SNPs (Table 2) were indeed genotyped for replication. rs837642 (chr19:g.49307999G>A), in an intron of *branched chain amino acid transaminase 2 (BCAT2)*, was successfully replicated with $q < 0.05$ ($p_{\text{replication}} = 0.001$, Table 2). Further gene-based analysis showed that

BCAT2 was significantly associated with LTM ($p = 1.2 \times 10^{-4}$). Additionally, two imputed SNPs with $\text{INFO} > 0.8$ and $\text{MAF} > 0.001$ reached genome wide significance in the combined sample (Supplementary Table S4) but they were not accepted for further analyses since they did not match the selection criteria. They were rs80239319 (chr9:g.140298162G>A), in an intron of *exonuclease 3'-5' domain containing 3*, predicted to be in enhancers and alter transcription factor binding motifs, $p = 7.9 \times 10^{-10}$ (Fig. 3a); and rs148620999 (chr11:g.51473048delC), $p = 4.3 \times 10^{-8}$ (Fig. 3b).

Pathway analysis

In addition, pathway analysis by MAGENTA found three pathways (glioma, mTOR signaling pathway, axon guidance) associated with digit-span STM with nominal significance, two pathways (regulation of autophagy, mRNA end-processing and stability) for visuospatial STM, one pathway (ephrin receptor signaling) for the

STM-shared component, and one pathway (olfaction) for LTM (Table 3).

Discussion

This study systematically investigated the genetic basis of both STM and LTM capacity in a large Chinese population. Common SNP-based heritability estimation suggested moderate heritability (61%, s.e. = 35%) for the digit-span STM, consistent with previous classic twin heritability studies [8–10]. Common SNPs failed to reveal significant non-zero heritability for visuospatial STM and LTM. However, the low heritability could be caused by the small sample size or due to the polygenic component of the memory score using these particular tests in this sample is small. It could also be due to that low frequency SNPs accounting for major effects in our LTM measure. Further two-step GWAS of altogether ~4500 individuals suggested *ZFAT* to be associated with STM performance, as rs7011450 near and five other SNPs in this gene were significantly associated. GWAS of ~3380 individuals suggested *BCAT2* to be related with LTM.

Three *ZFAT* nearby SNPs were predicted to be in enhancer state (Supplementary Table 2). *ZFAT*, suggesting to be related with STM, encodes a nuclear zinc-finger protein that binds DNA and functions as a transcriptional regulator involved in apoptosis and cell survival [29]. *ZFAT* can recognize histone H3 acetylation, which is involved in inflammation-mediated epigenetic modulation of memory [30]; another target gene of *ZFAT* is *bromodomain and PHD finger containing 1* [30], which is important for brain development [31] and is associated with schizophrenia [32]. Previous research has found that deficits in *ZFAT* were associated with autoimmune thyroid disease [33] and a *ZFAT* variant was associated with multiple sclerosis that sometimes involves memory deficits [34]. So far, its relationship with memory has not been well studied. Given that GWAS found significant contribution of SNPs within *ZFAT* to STM and gene-based association test on this gene gave positive results, it suggests a role of *ZFAT* in STM capacity. Further studies are needed to explore the detailed mechanism.

The SNP rs837642 is an expression quantitative trait locus for *BCAT2* (from Lieber Brain Institute RNAseq project) and *ribosomal protein S11* (from GTEx Portal); it was predicted to be in DNaseI-hypersensitive site and enhancers, and interacted with 25 genes via three-dimensional chromatin loops (Supplementary Table 2). Gene-based analysis revealed *BCAT2* to be significantly related to LTM performance. *BCAT2* is involved in leucine-related pathways and plays a role in hormone regulation and glutamate metabolism in brain [35], but its

relationship with memory has not been established before. The current study suggests several new gene targets for future research to understand the molecular basis of human memory.

GWAS on imputed data discovered variants near *SKOR2* to be related with STM. *SKOR2* is specifically expressed in neuronal tissues [36] and has been regarded as a biomarker for Purkinje cells [37]; a recent GWAS study has found correlations of this gene with cognitive performance [38]. Future replications on other independent cohorts are needed to valid this discovery. Furthermore, pathway analysis revealed several nervous system related pathways significantly associated with memory performance. Among them, glioma, mTOR signaling pathway, axon guidance, and Ephrin receptor signaling have been related with memory functions in previous literature [39–42]; it is of note that mTOR signaling in hippocampus is necessary for memory formation [40], while other associated pathways are somewhat linked with mTOR signaling pathway, which further supports the involvement of these pathways in memory functions.

Despite the fact that the *post hoc* power reached 80% for SNPs with MAF > 0.08 and effect size > 0.18, we did not find any genome-wide significant SNPs for visuospatial STM and LTM (Supplementary Discussion). Due to small effect size of the SNPs and no replication data available for imputation analyses, the current findings are only suggestive and warrant future replication. For STM, given that common SNP heritability is moderate, each SNP might only account for a very small percentage of total variance. STM capacity might be influenced by a large group of SNPs but each with a very small effect size. Therefore, a much larger sample is needed to identify the missing heritability. For LTM, its nearly zero heritability requests further replication studies in larger samples and suggests future studies focusing on testing rare variants contributions. Nevertheless, newly emerged technologies combined with larger samples and meta-analyses are needed to improve our understanding of the genetic architecture of complex traits.

We also tested associations of previously discovered genes and SNPs in the current study (Supplementary Table 5). Among these genes that have been associated with human memory performance, consistent results in the current study were found by gene-based analyses for *ODZ2* [13], *SCN1A*, *P2RY6*, *TFF2*, *TTC21B*, *TBC1D8* [14], *BINI* [43], *APBA1*, *CADM2*, *EXOC4* [44], *RASGRF2*, *PLCG2*, *LMO1*, and *PRKG1* [45]. For the previously discovered loci on human memory, some were replicated in this study, namely, rs2278729 ($p = 0.019$, STM) [14], rs401758 ($p = 0.013$, STM) [45], rs2469383 ($p = 0.038$ / 0.024, STM / LTM) [45], and rs7898516 ($p = 0.0083$, LTM) [45]. In accordance to Heck et al. [3], we found a nominally significant ($p = 0.028$) association of the voltage-gated

calcium channel activity pathway with STM-shared component.

To conclude, classical GWAS helps understand the genetic basis of complicate cognitive functions in humans. Using this technique, this paper revealed the heritability of short-term and long-term memories, and identified genes related to different types of human memory. However, more studies and larger samples are always needed to obtain more stable and reliable results.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Project 31421003); the Beijing Innovation Center for Genomics at Peking University; the Applied Development Program from the Science and Technology Committee of Chongqing (cstc2014yykfb10003 and cstc2015shms-ztx10006); and the Program of Mass Creativities Workshops from the Science and Technology Committee of Chongqing. We are grateful to Zhangyan Guan and Huizhen Yang for help with DNA preparation. Zijian Zhu thanks the Chinese Scholarship Council (CSC, No. 201709920075) and the German Academic Research Foundation (DAAD, No. 91658524) for financial support.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Skodzik T, Holling H, Pedersen A. Long-term memory performance in adult ADHD: a meta-analysis. *J Atten Disord*. 2017;21:267–83.
- Wang Y, Zhang YB, Liu LL, et al. A meta-analysis of working memory impairments in autism spectrum disorders. *Neuropsychol Rev*. 2017;27:46–61.
- Heck A, Fastenrath M, Ackermann S, et al. Converging genetic and functional brain imaging evidence links neuronal excitability to working memory, psychiatric disease, and brain activity. *Neuron*. 2014;81:1203–13.
- Bettencourt K, Xu Y. Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nat Neurosci*. 2016;19:150–7.
- Muller NG, Knight RT. The functional neuroanatomy of working memory: contributions of human brain lesion studies. *Neuroscience*. 2006;139:51–58.
- Tang Y, Shimizu E, Dube GR, et al. Genetic enhancement of learning and memory in mice. *Nature*. 1999;401:63–69.
- Voss JL, Paller KA. An electrophysiological signature of unconscious recognition memory. *Nat Neurosci*. 2009;12:349–55.
- McClearn GE, Johansson B, Berg S, et al. Substantial genetic influence on cognitive abilities in twins 80 or more years old. *Science*. 1997;276:1560–3.
- Goldberg HX, Lemos GS, Fananas SL. A systematic review of the complex organization of human cognitive domains and their heritability. *Psicothema*. 2014;26:1–9.
- Jensen AR, Marisi DQ. Note on the heritability of memory span. *Behav Genet*. 1979;9:379–87.
- Volk HE, McDermott KB, Roediger HL III, Todd RD. Genetic influences on free and cued recall in long-term memory tasks. *Twin Res Hum Genet*. 2006;9:623–31.
- Vogler C, Gschwind L, Coynel D, et al. Substantial SNP-based heritability estimates for working memory performance. *Transl Psychiatry*. 2014;4:e438.
- Papassotiropoulos A, Stephan DA, Huentelman MJ, et al. Common *KIBRA* alleles are associated with human memory performance. *Science*. 2006;314:475–8.
- Papassotiropoulos A, Henke K, Stefanova E, et al. A genome-wide survey of human short-term memory. *Mol Psychiatry*. 2011;16:184–92.
- Papassotiropoulos A, Stefanova E, Vogler C, et al. A genome-wide survey and functional brain imaging study identify *CTNBL1* as a memory-related gene. *Mol Psychiatry*. 2013;18:255–63.
- Price AL, Patterson N, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38:904–9.
- Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88:76–82.
- Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*. 2010;42:565–U131.
- Visscher PM, Hemani G, Vinkhuyzen AA, et al. Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. *PLoS Genet*. 2014;10:e1004269.
- Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–75.
- Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21:263–5.
- Delaneau O, Marchini J, Zagury J. A linear complexity phasing method for thousands of genomes. *Nat Methods*. 2012;9:179–81.
- 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;44:955–9.
- Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet*. 2009;5:e1000529.
- Marchini J, Howie B, Myers B, Mcvean G, Donnelly P. A new multipoint method for genome-wide association studies via imputation of genotypes. *Nat Genet*. 2007;39:906–13.
- Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26:2336–7.
- Liu JZ, McRae AF, Nyholt DR, et al. A versatile gene-based test for genome-wide association studies. *Am J Hum Genet*. 2010;87:139–45.
- Segrè AV, DIAGRAM Consortium, MAGIC investigators, et al. Common inherited variation in mitochondrial genes is not enriched for associations with Type 2 diabetes or related glycemic traits. *PLoS Genet*. 2010;6:e1001058.
- Fujimoto T, Doi K, Koyanagi M, et al. ZFAT is an antiapoptotic molecule and critical for cell survival in MOLT-4 cells. *FEBS Lett*. 2009;583:568–72.
- Ishikura S, Tsunoda T, Nakabayashi K, et al. Molecular mechanisms of transcriptional regulation by the nuclear zinc-finger protein Zfat in T cells. *Biochim Biophys Acta*. 2016;1859:1398–410.
- You L, Zou J, Zhao H, et al. Deficiency of the chromatin regulator BRPF1 causes abnormal brain development. *J Biol Chem*. 2015;290:7114–29.
- Kushima I, Aleksic B, Ikeda M, et al. Association study of bromodomain-containing 1 gene with schizophrenia in Japanese population. *Am J Med Genet B Neuropsychiatr Genet*. 2010;153B:786–91.

33. Shirasawa S, Harada H, Furugaki K, et al. SNPs in the promoter of a B cell-specific antisense transcript, SAS-ZFAT, determine susceptibility to autoimmune thyroid disease. *Hum Mol Genet.* 2004;13:2221–31.
34. Bourguiba-Hachemi S, Ashkanani TK, Kadhem FJ, Almawi WY, Alroughani R, Fathallah MD. ZFAT gene variant association with multiple sclerosis in the Arabian Gulf population: A genetic basis for gender-associated susceptibility. *Mol Med Rep.* 2016;14:3543–50.
35. Hull J, Hindy ME, Kehoe PG, Chalmers K, Love S, Conway ME. Distribution of the branched chain aminotransferase proteins in the human brain and their role in glutamate regulation. *J Neurochem.* 2012;123:997–1009.
36. Arndt S, Poser I, Schubert T, Moser M, Bosserhoff AK. Cloning and functional characterization of a new Ski homolog, Fussel-18, specifically expressed in neuronal tissues. *Lab Invest.* 2005;85:1330–41.
37. Minaki Y, Nakatani T, Mizuhara E, Inoue T, Ono Y. Identification of a novel transcriptional corepressor, Corl2, as a cerebellar Purkinje cell-selective marker. *Gene Expr Patterns.* 2008;8:418–23.
38. Rietveld CA, Esko T, Davies G, et al. Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc Natl Acad Sci USA.* 2014;111:13790–4.
39. Mu YG, Huang LJ, Li SY, et al. Working memory and the identification of facial expression in patients with left frontal glioma. *Neuro Oncol.* 2012;14:81–89.
40. Bekinschtein P, Katze C, Slipczuk LN, et al. mTOR signaling in the hippocampus is necessary for memory formation. *Neurobiol Learn Mem.* 2007;87:303–7.
41. Nakahara S, Miyake S, Tajinda K, Ito H. Mossy fiber mis-pathfinding and semaphorin reduction in the hippocampus of α -CaMKII hKO mice. *Neurosci Lett.* 2015;598:47–51.
42. Dines M, Grinberg S, Vassiliev M, Ram A, Tamir T, Lamprecht R. The roles of Eph receptors in contextual fear conditioning memory formation. *Neurobiol Learn Mem.* 2015;124:62–70.
43. Zhang X, Yu JT, Li J, et al. Bridging Integrator 1 (BIN1) genotype effects on working memory, hippocampal volume, and functional connectivity in young healthy individuals. *Neuropsychopharmacology.* 2015;40:1794–803.
44. Davies G, Marioni RE, Liewald DC, et al. Genome-wide association study of cognitive functions and educational attainment in UK Biobank (N = 112 151). *Mol Psychiatry.* 2016;21:758–67.
45. Xiang B, Wu JY, Ma XH, et al. Genome-wide association study with memory measures as a quantitative trait locus for schizophrenia. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi.* 2012;29:255–9.