

Cognition and Behavior

# Distinct Genetic Signatures of Cortical and Subcortical Regions Associated with Human Memory

Pin Kwang Tan,<sup>1</sup> Egor Ananyev,<sup>2</sup> and Po-Jang Hsieh<sup>3</sup><https://doi.org/10.1523/ENEURO.0283-19.2019>

<sup>1</sup>The N.1 Institute of Health, National University of Singapore, Singapore, Singapore 117456, <sup>2</sup>Department of Psychology, Nanyang Technological University, Singapore, Singapore 639798, and <sup>3</sup>Department of Psychology, National Taiwan University, Taipei City, Taiwan 10617

## Abstract

Despite the discovery of gene variants linked to memory performance, understanding the genetic basis of adult human memory remains a challenge. Here, we devised an unsupervised framework that relies on spatial correlations between human transcriptome data and functional neuroimaging maps to uncover the genetic signatures of memory in functionally-defined cortical and subcortical memory regions. Results were validated with animal literature and showed that our framework is highly effective in identifying memory-related processes and genes compared to a control cognitive function. Genes preferentially expressed in cortical memory regions are linked to memory-related processes such as immune and epigenetic regulation. Genes expressed in subcortical memory regions are associated with neurogenesis and glial cell differentiation. Genes expressed in both cortical and subcortical memory areas are involved in the regulation of transcription, synaptic plasticity, and glutamate receptor signaling. Furthermore, distinct memory-associated genes such as *PRKCD* and *CDK5* are linked to cortical and subcortical regions, respectively. Thus, cortical and subcortical memory regions exhibit distinct genetic signatures that potentially reflect functional differences in health and disease, and nominates gene candidates for future experimental investigations.

**Key words:** cognition; cortical; genetic; human; memory; neuroimaging

## Significance Statement

The anatomic and functional aspects of human memory are well characterized, but its biological mechanisms are poorly understood. Here, to uncover genetic signatures associated with human memory function, we analyzed spatial correlations between micro-scale gene expression and macro-scale neuroimaging maps to derive memory-related biological processes and genes in an unsupervised manner. We found the gene signatures of cortical and subcortical memory to be largely distinct and are associated with memory. We identified less characterized memory-associated genes as well. Furthermore, our framework demonstrated effectiveness and precision in identifying gene signatures related to memory versus another function as a control. Overall, our work provides a human-centric approach to understanding the genetics of cognition and identifies potential gene candidates for future experimental investigations.

## Introduction

Memory function is crucial for everyday life, ranging from mental arithmetic to long-term planning. Human

memory function is well characterized in terms of neural correlates associated with behavior and mental disorders. Insights from fMRI and lesion studies led to an under-

Received July 18, 2019; accepted November 19, 2019; First published December 9, 2019.

The authors declare no competing financial interests.

Author contributions: P.K.T., E.A., and P.-J.H. designed research; P.K.T. performed research; P.K.T., E.A., and P.-J.H. analyzed data; P.K.T., E.A., and P.-J.H. wrote the paper.

standing of cortical and subcortical memory regions as functionally distinct areas, subsumed under the broad umbrella of memory function (LaBar and Cabeza, 2006; Squire and Wixted, 2011). Yet, despite the fact that memory ability is highly heritable, with genetic risk factors for memory disorders, the genetic signature underlying human memory remains poorly understood (Papassotiropoulos and de Quervain, 2011; Kandel et al., 2014; Freudenberg-Hua et al., 2018). Our knowledge of human memory genes is largely based on interindividual variation in genomes [e.g., genome-wide association studies (GWAS)] and the short-term temporal dynamics of memory function (Berto et al., 2017). However, there is emerging interest in using the spatial dimension of gene expression to identify genetic profiles of functional networks, by integrating human brain transcriptomes and neuroimaging maps (Yarkoni et al., 2011; Hawrylycz et al., 2012; Ritchie et al., 2018). Such approaches based on spatial expression patterns may help answer a key question: Are there genes associated with general memory regions in the adult human brain? This may provide unprecedented insight into biological processes and genes associated with human memory, and propose potential candidates for further experimental investigation.

To identify such adult human genes associated with general memory, we rely on a spatial correlation method (Fox et al., 2014). The method identifies cognition-associated genes that have a high spatial correlation between its gene expression and a neuroimaging map that represents the relevance of each area for memory (Fig. 1A,B). This approach assumes that genes involved in memory should be highly expressed in the brain areas highly relevant for memory. For instance, this relationship was observed in the case of reward-associated gene *DRD2* in reward processing areas (Mengod et al., 1992; Pappata et al., 2002; Schott et al., 2008). For the genetic expression and functional maps, we used the Allen Human Brain Atlas (AHBA) transcriptome and the Neurosynth “memory” neuroimaging map.

We then identified memory-related genetic profiles in an unsupervised manner with gene set enrichment analysis (GSEA; Fig. 1C). Due to the correlational nature of the spatial correlation analysis, we drew on biological knowledge databases (i.e., enrichment analyses with the Gene Ontology library) to guide our identification of biological processes and genes associated with memory. To identify cortex-specific and subcortex-specific memory-associated genes, we compared the differences between their respective gene

profiles (Fig. 1D). As there are genes that are involved in general memory across cortical and subcortical regions (Gallo et al., 2018), we also characterized the overlap between cortical-subcortical genetic profiles (Fig. 1D). Furthermore, to identify candidate genes, we identified the top-10 genes most likely associated with memory with leading-edge analysis (LEA; Fig. 1E). We then validated our results by verifying that the genetic profiles corroborate with experimental literature. Finally, we assessed whether our approach was effective and precise (Fig. 1F,G).

Because of their common and critical involvement in general memory, we analyzed both cortical and subcortical areas involved in memory. Of note, previous studies on functional networks mainly focused on cortical or subcortical analyses due to disparate expression profiles. This may be because of marked differences in neuronal composition, number of layers, and connectivity (O’Leary and Koester, 1993; Yushkevich et al., 2009; Modha and Singh, 2010; Kim et al., 2015). At the genetic level, these differences are mirrored by distinct patterns of both gene expression intensity and variability (Hawrylycz et al., 2012; Richiardi et al., 2015; Fox et al., 2014). Thus, we analyzed their spatial correlations separately, as combining both in the spatial correlation analysis would capture the gross cortico-subcortical differences in expression intensity instead of meaningful inter-regional differences in genetic expression.

Despite a common involvement in general memory, we found largely distinct memory-related biological processes and genes across cortical and subcortical regions. Cortical processes included immune and epigenetic regulation; subcortical processes included neurogenesis and glial cell differentiation. Genes shared across cortical-subcortical regions were involved in the regulation of transcription, synaptic plasticity and glutamate receptor signaling. We show that our approach identified a greater number of memory genes in the memory analysis than expected by chance, and more memory genes than motor function genes. These results provide a better understanding of genetics associated with human memory, and nominate candidate genes for future experimental investigations.

## Materials and Methods

### AHBA transcriptome

The AHBA transcriptome was generated from the normalized mRNA microarray sampling of a combined 3702 sampling sites across six donor brains (Hawrylycz et al., 2012;  $N = 6$  left hemispheres,  $N = 2$  right hemispheres; Fig. 1A; see Extended Data Fig. 1-1 for an example visualization of a gene). The donors were three white males, two African-American males and one Hispanic female. Donor age ranged from 24 to 57, mean donor age was 42.5 years (SD = 11.2 years), Data from all six donors was horizontally concatenated into a .csv file, with one probe per row. For more details on the dataset and data collection procedures, see <http://help.brain-map.org/display/humanbrain/Documentation>.

This work was supported by the Duke-NUS Graduate Medical School and by the Yushan Young Scholar Program NTU-108V0202.

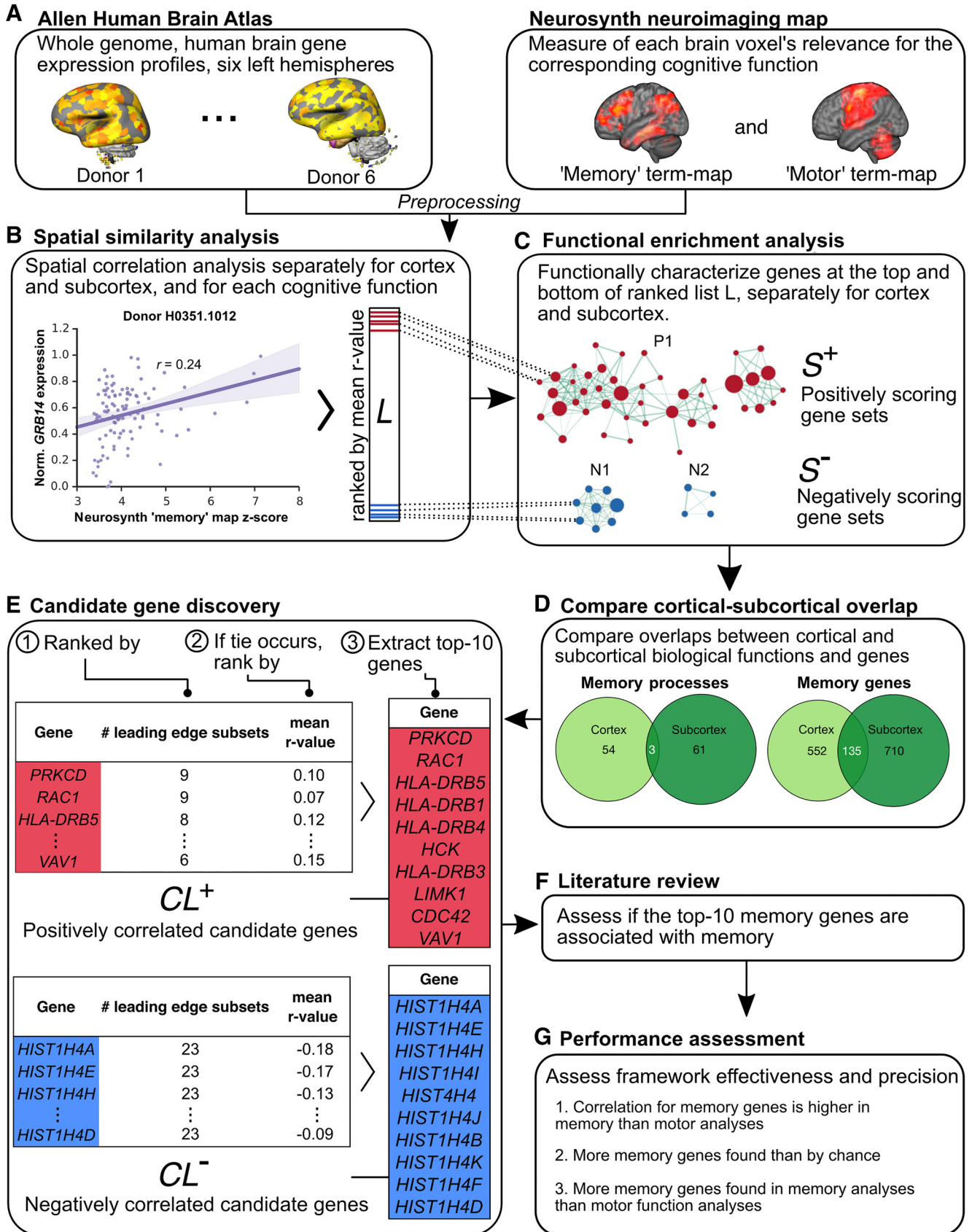
Acknowledgements: We thank Nilofar Husain and Wu Jinlu for their feedback on early versions of this manuscript. Data were provided by the Allen Institute for Brain Science and the Neurosynth repository.

Correspondence should be addressed to Pin Kwang Tan at [lsitpk@nus.edu.sg](mailto:lsitpk@nus.edu.sg).

<https://doi.org/10.1523/ENEURO.0283-19.2019>

Copyright © 2019 Tan et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.



**Figure 1.** Overview of genetic signature discovery framework. **A**, The AHBA and Neurosynth neuroimaging maps, and their preprocessing and integration into a common neuroimaging template space. **B**, Calculation of spatial similarity between the maps

*continued*

separately for the cortical and subcortical regions, and for memory and motor functions, deriving a ranked gene list  $L$  per analysis (contains genes and mean  $r$  value). **C**, Functional characterization of each  $L$  with biologically meaningful gene sets with GSEA Pre-ranked analysis (dotted lines connecting  $L$  and gene sets represent the clustering of genes into enriched gene sets), yielding positively and negatively scoring gene sets  $S^+$  and  $S^-$ . **D**, Assessing differences and the overlap between cortical and subcortical memory genes. **E**, Identification of candidate genes associated with the cognitive function and brain region, operationalized as the subset of genes driving the enrichment score of the significantly enriched gene sets found using GSEA Pre-ranked analysis. This produced two candidate gene lists,  $CL^+$  and  $CL^-$ , containing highly positively and negatively correlated genes from  $S^+$  and  $S^-$ , respectively. **F**, Literature review of each  $CL$  quantifying the genes associated with the target or control cognitive function. **G**, Assessing framework validity and precision with each of eight  $CL$ s. See Extended Data Figure 1-1 for a visualization of *GRB14* gene expression in the AHBA, Extended Data Figure 1-2 for a visualization of the Neurosynth maps, and Extended Data Figure 1-3 for the cortical and subcortical regions used in the spatial correlation analysis.

### Neurosynth memory association map

Neurosynth ‘memory’ and ‘motor’ association maps [Montreal Neurologic Institute 152 (MNI152) space, thresholded FDR < 0.01] were used as neuroimaging data for the memory and motor functions (Fig. 1A; see Extended Data Fig. 1-2 for a visualization of the Neurosynth memory and motor maps). These cognitive functions were chosen as they were largely functionally and anatomically distinct and were constructed from a similar number of studies ( $N_{\text{memory}} = 2744$ ,  $N_{\text{motor}} = 2565$ ). Neurosynth quantifies the relevance of each voxel to the user-specified search terms (e.g., memory) based on a database of neuroimaging studies. In the example of the memory map, each voxel is assigned a  $z$  score that reflects the preferential association of that voxel with memory, instead of other functions. For instance, the large positive  $z$  score in the hippocampus means that studies whose abstracts include the word memory are more likely to report hippocampus engagement than studies that do not include the word memory (Yarkoni et al., 2011). Negative  $z$  scores indicate a higher correlation with other search terms unrelated to memory, and thus were excluded from our analyses. For broad cognitive function domains, single terms enable the generation of maps that approximate the target cognitive process reasonably well (Yarkoni et al., 2011). Therefore, we used memory and motor as our search terms to derive the memory and motor association maps. Note that in the example of memory, this approach resulted in inclusion of a broad range of subfunctions, such as working memory and long-term memory. This allowed for a broader and more inclusive definition of memory and motor function for the subsequent identification of their genetic signatures. In the generation of such maps, it is possible that the foci identified by the automatic coordinate extraction process of Neurosynth may be inaccurate due to different data formats of online neuroimaging journals and sites. However, when compared with the gold standard of manually curated activation foci in the Surface Management Systems Database (SumsDB), automatically extracted coordinates shows high sensitivity (84%) and specificity (97%; Yarkoni et al., 2011). As such, the memory and motor neuroimaging maps used are reasonable approximations of regions involved in general memory and motor function.

### Preprocessing of transcriptome

We followed preprocessing steps as outlined in Arnatkevičiūtė et al. (2019), including the brain atlas used

to delineate cortical and subcortical regions. Note that this atlas from Arnatkevičiūtė et al. (2019) differs from the modified Brodmann atlas used by AHBA, and the hippocampus was relabeled as a subcortical instead of a cortical region in line with human neuroimaging conventions (Hawrylycz et al., 2012; Ji et al., 2019; list of cortical and subcortical areas in Extended Data Fig. 1-3). In preprocessing, probes were first reannotated with the Reannotator package. We excluded the probes which had <50% samples exceeding the background expression level. For each gene, we selected the probe with the highest differential stability score, i.e., with the least spatial variability across donors. The AHBA data were normalized ( $z$  score) for each donor. Cortical and subcortical regions were normalized separately. This was done to account for individual and cortico-subcortical differences in gene expression (Hawrylycz et al., 2011). This returned a 15,625 gene-by-1285 brain sample matrix for the left cortex, and 15,625-by-497 matrix for the left subcortex, respectively. This gave an individual average of 214 left cortical (range: 175–259) and 83 left subcortical samples per donor (range: 59–115). In the subsequent step below, this was further restricted to the brain regions of interest in co-registration. With the usage of a different brain atlas from AHBA, it is possible that our re-annotation of regions as cortical or subcortical areas may be inaccurate and may affect cortical and subcortical analyses. However, as we retained the AHBA ontological labels and simply re-annotated the hippocampus as subcortex, this step is reasonable.

### Co-registration of AHBA and Neurosynth memory map

To allow a comparison of spatial similarity between neuroimaging and AHBA maps of differing resolutions, both maps were co-registered into a common 3D stereotaxic brain space (Fox et al., 2014). This was done by using the MNI coordinates provided by AHBA for representing the transcriptome sampling points in MNI152 template space. This was also the space used by the Neurosynth map. The Neurosynth map was used as a mask for the AHBA map, so that only the overlapping areas were included in the correlation analysis (Fox et al., 2014). Due to the limited availability of hemispheres sampled (six left and two right hemispheres), we used only the left hemispheres, separated into cortical and subcortical regions. In subsequent steps, the cortical and subcortical analyses were kept separate. Besides providing insight

into the separate cortical and subcortical genetic mechanisms, this also avoided confounds from their divergent transcriptional profiles (Richiardi et al., 2015). We then matched the smoothing of both maps by smoothing the AHBA with a 6mm radius sphere. At the end of this step, there remained on average 93 memory (range: 72–107) and 65 motor cortical (range: 55–76) data points, and on average 40 memory (range: 25–71) and 43 motor subcortical (range: 24–69) data points per individual. It is possible that the coregistration of AHBA and Neurosynth maps may be affected by errors introduced during the MRI to MNI coordinate transformation by AHBA. However, the Allen Institute transformed the MRI to MNI coordinates using standard methods for *in cranio* and *ex cranio* brains (four donor brains were imaged *ex cranio*), and we ensured reasonable coregistration by visually inspecting the resulting maps.

### Spatial correlation analysis of AHBA and Neurosynth data

To obtain spatial correlation values per gene, we relied on a tool that correlates the spatial AHBA and neuroimaging maps (Yarkoni et al., 2011; Fox et al., 2014). Each datapoint used in correlation is a point in space, with a normalized gene expression intensity value and a neuroimaging map  $z$  score. For a gene associated with memory, we would expect high spatial similarity between both AHBA and Neurosynth maps, i.e., a pattern of high gene expression within areas highly relevant for memory and vice versa. This would be reflected in a high mean correlation value for that gene. We applied the spatial analysis separately for cortical and subcortical regions (Fig. 1B). An approximate random effects analysis was used to account for individual gene expression variability and to counter the sparse cortical sampling in the AHBA maps. Donor regression slope and intercept were modeled individually. We subsequently obtain each gene's mean correlation value (averaged across the six donors), which was the statistic of interest. From this step, we obtained four lists  $L$  of 15,625 genes, for memory and motor function, and the respective cortical and subcortical regions.

### Identifying biological processes of cortical and subcortical memory

We used a gene set analysis tool (GSEA Pre-ranked, GenePattern module, version 6.0.5) to identify sets of genes associated with common biological functions (Fig. 1C). The four lists of genes  $L$  were ranked by mean correlation value (the ranking statistic used in this case) and passed to GSEA Pre-ranked. We analyzed each list  $L$  with GSEA Pre-ranked with the default parameters, including weighted scoring using the Gene Ontology Biological Process library (c5.bp.v6.0.symbols.gmt). GSEA Pre-ranked looks separately at the top and bottom of each list  $L$  for genes that overlap with each gene set in the database (Mootha et al., 2003; Subramanian et al., 2005). This overlap or gene set enrichment was assessed by weighted scoring based on mean correlation ( $r$  value). This returns a normalized enrichment score, a significance  $p$  value, and an FDR  $q$  value (across all gene sets tested) for each enriched gene set. From the top positively and

negatively correlated genes in each list  $L$ , we obtained separate sets  $S^+$  and  $S^-$  of positively and negatively enriched gene sets, respectively. For subsequent analyses, we only used all gene sets with FDR  $q < 0.05$ . Note that the motor cortical ( $-$ ) analysis (negatively correlated genes from the motor cortical analysis) did not have any gene sets surviving FDR  $< 0.05$ , and thus was not used in subsequent analyses, i.e., biological processes nor candidate genes as output. Thus, this effectively meant eight  $S^+$  and seven  $S^-$  usable sets for subsequent steps. For this analysis and subsequent steps, genes that are found in  $\geq 1$  significantly enriched gene set are termed memory genes, as opposed to the top-10 memory genes identified below.

### Visualization of significantly enriched gene sets

To identify the overall biological themes across gene sets, we grouped gene sets into networks by the genes that they share (Fig. 1C). For each pair of sets  $S^+$  and  $S^-$ , we input their gene sets into the Cytoscape network visualization software, and included the gene sets with FDR  $q < 0.05$ . We then used the Enrichment Map app to construct the gene set networks and annotated them with the Wordcloud extension for subsequent interpretation (Cline et al., 2007; Merico et al., 2010; Oesper et al., 2011). This was done using the default settings except for a custom FDR  $q$  value threshold of 0.05 (i.e., FDR  $< 0.05$ ). This step returned four annotated enrichment maps for the list  $L$  of each cognitive function and for each of cortical and subcortical areas.

### Functional annotation of overlapping genes

We used the ToppGene suite (with Gene Ontology Biological Process library) to functionally cluster memory genes (identified in the GSEA analysis) that are (1) cortex specific, (2) subcortex specific, (3) and shared between both (Fig. 1D; Chen et al., 2009). From the output, we thresholded biological process gene sets as those that satisfied FDR  $< 0.05$ . This returned three lists of gene sets, one for each type of gene above.

### Identifying candidate genes associated with cortical and subcortical memory

To identify the top-10 genes most likely to be relevant to the cognitive function, we identified genes frequently appearing across the gene sets with the LEA (Fig. 1E; Mootha et al., 2003; Subramanian et al., 2005). For the analysis of each cognitive function in cortical and subcortical regions, we input the respective gene sets with FDR  $q < 0.05$  (javaGSEA desktop application). LEA then identified the genes that appeared frequently across the leading-edge subset genes across gene sets in  $S^+$  or  $S^-$  (Subramanian et al., 2005; Fleming and Miller, 2016). We ranked genes by the number of leading-edge gene sets they enrich; in the case of a tie in the number of gene sets, we rank them by the mean spatial correlation value. The top-10 genes appearing most frequently in the positively and negatively enriched gene sets were designated as the candidate gene list  $CL$ . The outputs were seven candidate gene-cognition association lists  $CL$  of 10 genes each for all  $S^+$  and  $S^-$ .

## Literature review of genetic signatures

To quantify the number of candidate gene “hits” for the memory analysis, we conducted a literature review for each gene list *CL* and counted the number of gene-memory (i.e., true positives) or gene-motor function associations (i.e., false positives; Fig. 1F). This was done by reviewing experimental literature on Google Scholar, via a search query: [“gene name” AND (“memory” OR “amnesia” OR “Alzheimer’s” OR “dementia”)] and [“gene name” AND (“motor function” OR “motor coordination” OR “locomotor” OR “ataxia” OR “motor learning” OR “Parkinson’s” OR “Huntington’s”)], respectively. The same was repeated for the motor analysis for the respective true positives and false positives. The disorders were selected for keyword search because they prominently feature deficiencies in memory and motor functioning. Strong evidence included studies that employed *in vivo* gene manipulations, mutants and pharmacological interventions, while weak evidence included computational gene associations, *in vitro* studies, differential gene expression studies and human case studies. Literature evidence only counted as validation if it implicated the corresponding brain area, i.e., cortical or subcortical. As such, evidence of a given gene’s role solely in the non-analyzed brain region was not counted. For example, if a paper showed that the knock-out of gene A solely in the subcortex leads to memory deficits, it would not count as evidence for the analysis of cortical memory.

## Correlation difference in memory and motor analyses

If the method is valid, memory genes should have a higher average correlation value from the memory analysis compared with the motor analysis, and vice versa for motor genes and the motor analysis *r* value. For each gene, this was calculated by subtracting its motor function *r* value from the memory *r* value, with a positive difference counting toward the method’s effectiveness (Fig. 1G). Note that for the memory *r* values from the negative gene lists (e.g., memory cortical -), we multiply the *r* value difference with -1 to express this difference as a positive value, consistent with the positive memory gene lists. We then take the average of all genes for each set *S* that satisfy FDR  $q < 0.05$  (same threshold as enrichment map visualization) to obtain seven such values. As the number of genes per set *S* is different, we bootstrapped the number of correlation difference values used for calculating the average correlation difference value per set. This was done separately for the memory and motor analyses by repeatedly subsampling the correlation differences (10,000 iterations) to the minimum number of genes in memory ( $n = 231$ ) and motor sets ( $n = 146$ ), respectively. We visualized this as a boxplot for each of the seven sets, with the bootstrapped mean and 95th percentiles (whiskers) for memory and motor analyses. If the baseline does not fall within the 95th percentile distribution (i.e., whiskers do not overlap with the baseline of zero) the score is deemed significantly different from baseline ( $p < 0.05$ ).

## Assessing method effectiveness in identifying candidate genes

We quantified method effectiveness based on the prior literature review (Fig. 1G). To do so, we calculated the chance probability of obtaining *N* memory genes per gene list. This is done by selecting *N* memory genes (without replacement) from the pool of known memory or memory-related disorder genes ( $n = 644$ ) out of all 15,625 genes analyzed. For example, if 10 out of 10 genes in the gene list are memory genes, the chance probability of this occurring is  $1.32 \times 10^{-14}$ . The same was done accordingly for motor function and the motor function genes ( $n = 104$ ). These memory genes were compiled from three sources: (1) the literature review above; (2) the biological function gene sets “GO:0007611 Learning or memory,” from database AmiGO<sub>2</sub> (Carbon et al., 2009; version 2.4.26, release date 2016-08); and (3) van Cauwenberghe et al. (2016). The motor-related genes (motor or motor-related disorder) were obtained from (1) the literature review above, (2) the biological function gene sets “GO:0061743 motor learning” and “GO:0061744 motor behavior” from database AmiGO<sub>2</sub> (Carbon et al., 2009), and (3) Lin and Farrer (2014).

## Precision score for memory and motor analyses

We asked, of a given memory gene list with genes labeled as memory genes, how many of these are actually related to memory. We quantified this by calculating a precision score (Fig. 1G). We first determined the true positives (i.e., genes associated with memory from the literature review) and false positives (i.e., genes associated with motor function). The literature evidence was weighted such that for true positives, strong evidence and weak evidence (defined above) received a full point and half-point respectively. For each gene list, we then determined the method’s precision score by dividing “true positives” by the sum of true positives and false positives (Eqs. 1, 2). If the method is precise, for memory analyses, the memory precision scores should be above 0.5 and motor score below 0.5, and vice versa. We plotted the memory and motor precision scores for each gene list (ranging from 0 to 1), and the difference between these scores (ranging from -1 to 1). Ideally, the difference should be greater than zero. In the following equations,

$$\text{Memory precision score} = \frac{(0.5 \text{Memory}_w + \text{Memory}_s)}{((0.5 \text{Memory}_w + \text{Memory}_s) + (0.5 \text{Motor}_w + \text{Motor}_s))} \quad (1)$$

$$\text{Motor precision score} = \frac{(0.5 \text{Motor}_w + \text{Motor}_s)}{((0.5 \text{Memory}_w + \text{Memory}_s) + (0.5 \text{Motor}_w + \text{Motor}_s))} \quad (2)$$

$\text{Memory}_s$  = number of genes with strong evidence for its association with memory;  $\text{Memory}_w$  = number of genes with weak evidence for its association with memory;  $\text{Motor}_s$  = number of genes with strong evidence for its association with motor function; and  $\text{Motor}_w$  = number

of genes with weak evidence for its association with motor function.

### Data availability statement

All genetic and neuroimaging data used are available from the AHBA (<https://human.brain-map.org>) and Neurosynth (<https://www.neurosynth.org>). The scripts for pre-processing the transcriptome are available at <https://github.com/BMHLab/AHBAprocessing>. The correlation scripts and input data are available for non-commercial use in Extended Data 1 and at <https://github.com/PK-HQ/geneCognitionDiscovery>.

## Results

### AHBA and Neurosynth maps

For identifying whole-brain adult human memory genes, we first needed to conduct the spatial correlation analysis between 3D high-resolution neuroimaging and transcriptome maps of the adult human brain. As such, we used the high density, whole human brain AHBA transcriptome, and the Neurosynth memory association map of each voxel's association with memory in general as input datasets (Yarkoni et al., 2011).

The AHBA was derived from six donor brains, and contains whole genome human brain gene expression in the left cortical and subcortical regions ( $N = 6$ ; Fig. 1A; see example visualization in Extended Data Fig. 1-1; Hawrylycz et al., 2012). The Neurosynth memory association map is a meta-study map ( $N = 2744$ ) which represents each brain voxel's relevance for memory (as opposed to other cognitive functions), specified by positive  $z$  scores (Fig. 1A; see visualization of memory and motor function maps in Extended Data Fig. 1-2; Yarkoni et al., 2011). Note that the usage of memory here refers to memory in general, as the map was constructed from memory-related neuroimaging studies that employ multiple types of memory tasks (Yarkoni et al., 2011). We co-registered both maps into a common MNI152 space. The memory areas in the memory association map were used to define the usable AHBA samples for the subsequent spatial correlation analysis.

### Spatial similarity analysis

Using these datasets, we sought to isolate the genes with high spatial correlation values between their gene expression and memory term maps for subsequent analysis steps, as they are most likely related to memory (Fox et al., 2014). We conducted the spatial similarity analysis between the AHBA and Neurosynth association maps separately for cortical and subcortical regions due to their marked differences (see Introduction; the list of cortical and subcortical regions is available in Extended Data Fig. 1-3), and for memory and motor function (see an example of spatial correlation in Fig. 2). Each analysis yielded a list  $L$ , which contained the mean correlation values of 15,625 genes used for subsequent ranking (Fig. 1B).

We subsequently ranked each list  $L$ . A positive correlation indicates higher gene expression in areas relevant for memory, and a negative correlation implies lower expression in areas relevant for memory. The top-10 posi-

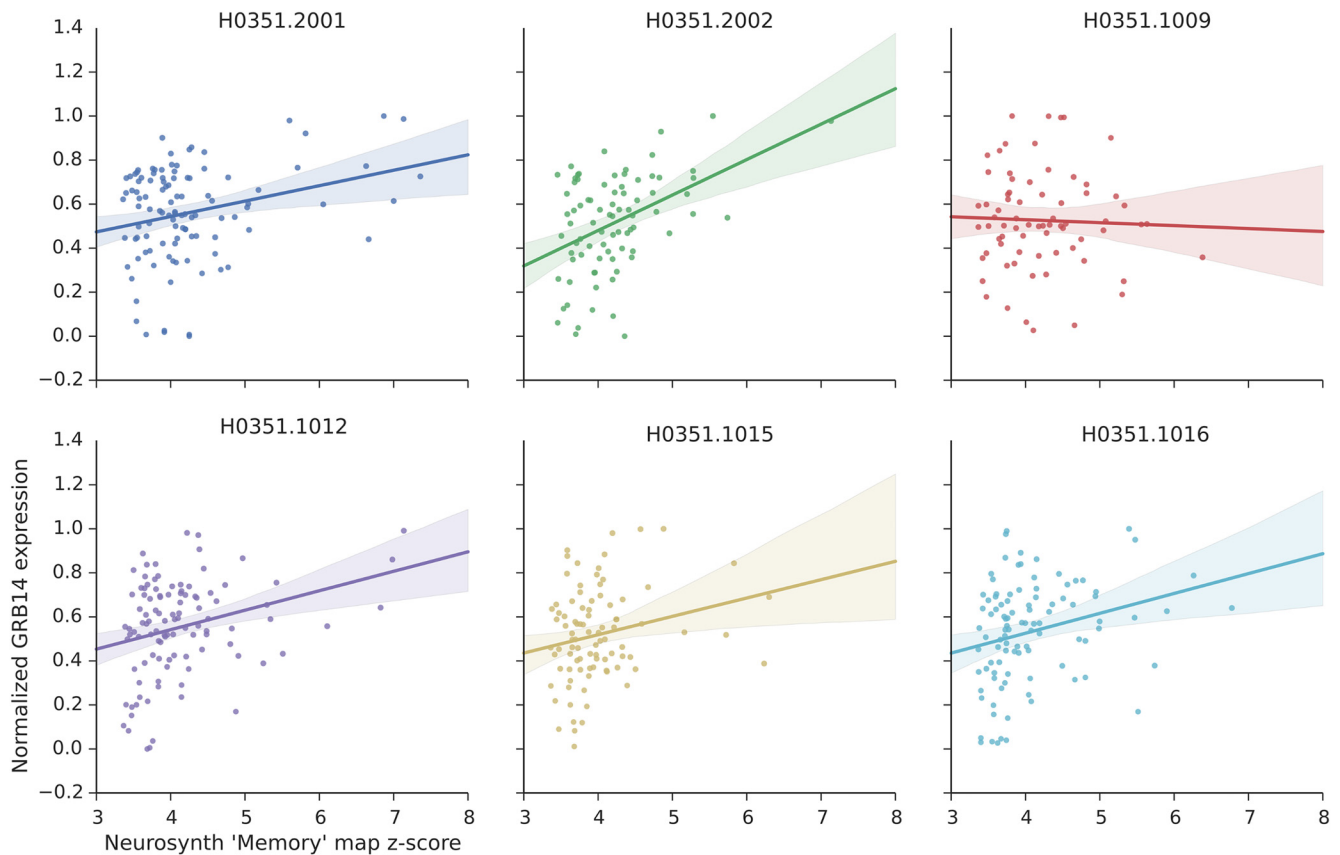
tively and negatively correlated genes for the memory cortical and subcortical analyses are shown in Table 1 (see the spatial correlation value of all genes in Extended Data Table 1-1). There were more negatively correlated genes than positively correlated genes for both cortical and subcortical analyses of memory (Extended Data Table 1-1). We found 8383 positively and 7243 negatively correlated genes for the cortical areas, and 7642 positively and 7984 negatively genes for the subcortical areas.

### Distinct gene expression profiles associated with cortical and subcortical memory

Following the spatial correlation analyses, we aimed to define the gene expression profiles related to cortical and subcortical memory in a comprehensive manner. To identify and characterize sets of genes that work toward a common biological function (i.e., gene sets), we analyzed each of the cortical and subcortical lists  $L$  with GSEA Pre-ranked (Fig. 1C). This yielded positively scoring and negatively scoring gene sets, derived from the positively and negatively correlated genes of  $L$ , respectively. These gene sets were then grouped into functionally related clusters, and automatically annotated with biological themes (Cline et al., 2007; Merico et al., 2010; Oesper et al., 2011).

Overall, the cortex and subcortex had distinct biological themes that were previously found associated with memory. For cortical memory, GSEA revealed 28 positive and 29 negative significantly enriched gene sets. Visualization of the enrichment network showed that these gene sets were grouped into five distinct clusters (Fig. 3; the complete GSEA results are in Extended Data Fig. 3-1), with gene sets within each cluster sharing enriched genes. These gene sets were found to be related to memory. The positive cluster P1 contained gene sets implicated in immune response and  $Fc\gamma$  receptor signaling (Fernandez-Vizarra et al., 2012; Marin and Kipnis, 2013). P2 was implicated in interferon gamma signaling (Litteljohn et al., 2014), P3 in transmembrane calcium ion transport and P4 in actin filament assembly (Krucker et al., 2000; Lamprecht, 2011). The negative cluster N2 contained gene sets involved in chromatin dynamics, epigenetic regulation, and immune cell differentiation (Kim and Kaang, 2017).

For subcortical memory, GSEA revealed 50 positive and 14 negative significantly enriched gene sets. Visualization of the enrichment network showed that these gene sets were grouped into three distinct clusters (Fig. 4; the complete GSEA results are in Extended Data Fig. 4-1). Similarly, these gene sets were found to be related to memory. The positive cluster P1 is implicated in synaptic transmission and synaptic plasticity. It also included gene sets involved in endocytosis and exocytosis, neurotransmitter secretion, long-term potentiation (Stuchlik, 2014), glutamate receptor signaling, and neuron projection morphogenesis (Kasai et al., 2010). The negative cluster N1 is related to transcription and translation processes (Jarome and Helmstetter, 2014; Alberini and Kandel, 2015), and cluster N2 to glial cell and oligodendrocyte differentiation (Hertz and Chen, 2016; Pepper et al., 2018).



**Figure 2.** An example of spatial similarity analysis output. The expression levels of the top-correlated cortical gene, *GRB14*, is visualized as a function of the Neurosynth map’s voxel-wise relevance to memory function (z score). Normalized gene expression (y-axis) plotted against neuroimaging map z scores (x-axis). Each colored regression line represents the best-fit line for each of six donors (colors); the translucent band around each line represents the 95% confidence interval estimate.

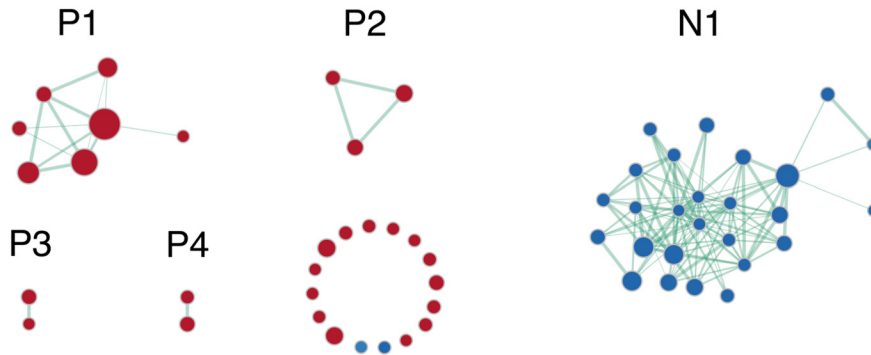
To identify differences and overlaps in the cortical and subcortical genetic profiles, we identified and characterized the different and shared (1) biological processes as shown in the enrichment maps, and (2) memory genes (i.e., all genes found in  $\geq 1$  enriched gene set; Fig. 1D). We found a low overlap of 2.5% of gene sets ( $N = 3$ ) and 9.6% of genes ( $N = 135$ ) between cortical and subcortical regions (Fig. 5; the complete list of distinct and overlapping genes is in Extended Data Fig. 5-1). The overlapping genes were involved in memory-related processes of protein transport, transcriptional regulation, synaptic plasticity and glutamate receptor signaling (Peng et al., 2011; Rosenberg et al., 2014; Alberini and Kandel, 2015; Table 2; full output of gene sets and genes from TopGene in Extended Data Table 2-1). These include genes involved in the Arp2/3 complex, GABA and AMPA ligand-gated ion channels which are critical for memory function (Gasbarri and Pompili, 2014; Basu et al., 2016; Takemoto et al., 2017; Extended Data Table 2-1). Cortex-specific genes were involved in memory-associated processes such as DNA repair, epigenetic regulation, immunity and  $IFN-\gamma$  signaling (Marin and Kipnis, 2013; Litteljohn et al., 2014; Kim and Kaang, 2017; Hou et al., 2018; Extended Data Table 2-1). Subcortex-specific genes are involved in neurogenesis, dendrite morphogenesis, glial cell differentiation and myelination (Hertz and Chen, 2016; Kao et al.,

**Table 1. Spatial correlation analysis for memory function**

| Correlation polarity | Cortical analysis |               | Subcortical analysis |               |
|----------------------|-------------------|---------------|----------------------|---------------|
|                      | Gene              | Mean <i>r</i> | Gene                 | Mean <i>r</i> |
| +                    | <i>GRB14</i>      | 0.24          | <i>NEUROD6</i>       | 0.66          |
|                      | <i>DYRK3</i>      | 0.21          | <i>NEUROD1</i>       | 0.66          |
|                      | <i>FILIP1</i>     | 0.21          | <i>NPTXR</i>         | 0.65          |
|                      | <i>SPHKAP</i>     | 0.21          | <i>PLEKHG5</i>       | 0.65          |
|                      | <i>TMTC1</i>      | 0.21          | <i>NNMT</i>          | 0.64          |
|                      | <i>TSPAN2</i>     | 0.21          | <i>LRRC2</i>         | 0.64          |
|                      | <i>S100A10</i>    | 0.21          | <i>C9orf16</i>       | 0.64          |
|                      | <i>HEYL</i>       | 0.20          | <i>MICAL2</i>        | 0.64          |
|                      | <i>FZD7</i>       | 0.20          | <i>SLC17A7</i>       | 0.64          |
|                      | <i>KCTD12</i>     | 0.20          | <i>DUSP4</i>         | 0.64          |
| -                    | <i>NRAP</i>       | -0.25         | <i>CRNDE</i>         | -0.64         |
|                      | <i>DLGAP1-AS4</i> | -0.22         | <i>FAM222A</i>       | -0.63         |
|                      | <i>CTNNAL1</i>    | -0.21         | <i>CRABP1</i>        | -0.62         |
|                      | <i>FGF18</i>      | -0.20         | <i>NTM-AS1</i>       | -0.62         |
|                      | <i>MIR124-2HG</i> | -0.19         | <i>SELENOP</i>       | -0.62         |
|                      | <i>HIST1H1D</i>   | -0.19         | <i>KIF19</i>         | -0.61         |
|                      | <i>TDRD1</i>      | -0.19         | <i>LOC100506725</i>  | -0.61         |
|                      | <i>SLC24A4</i>    | -0.18         | <i>CA14</i>          | -0.61         |
|                      | <i>CCDC144B</i>   | -0.18         | <i>ZFHX4</i>         | -0.61         |
|                      | <i>LINC00476</i>  | -0.18         | <i>LINC00844</i>     | -0.61         |

Top-10 positively and negatively correlated genes from the memory analysis, ranked by the mean correlation magnitude across six donor brains. The positively and negatively correlated genes are listed separately for cortical and subcortical areas. See Extended Data Table 1-1 for the complete list of genes and respective *r* values.





**Figure 3.** Enrichment map visualization for cortical memory. Clusters are labeled with P for positive, N for negative. Gene set clusters were found to be related to memory. Positive clusters were associated with immune signaling, calcium transport and actin filament assembly. The negative cluster contained gene sets involved in chromatin dynamics and epigenetic regulation. See Extended Data Figure 3-1 for the full output from GSEA Pre-ranked.

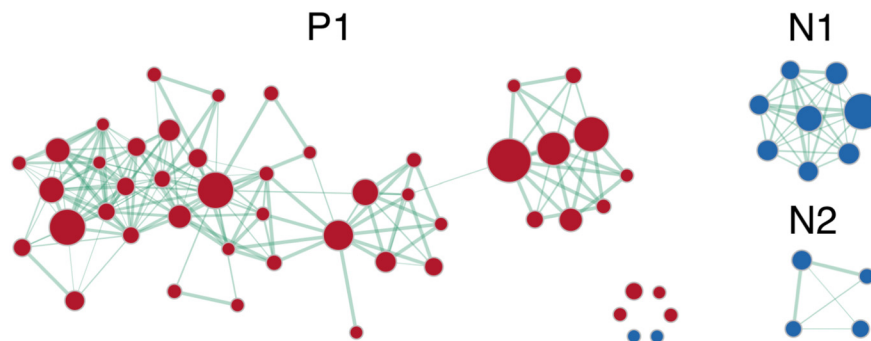
2018; Pepper et al., 2018; Extended Data Table 2-1). Note that the same gene set can appear both in the cortical-specific and subcortical-specific biological processes. For instance, the memory gene sets is enriched in both regions, but in each case, the gene set enrichment is driven by distinct genes (Extended Data Table 2-1). This is because different genes can be relevant for, and thus increase enrichment for the same biological process gene set.

**Core differentially expressed genes related to cortical and subcortical memory**

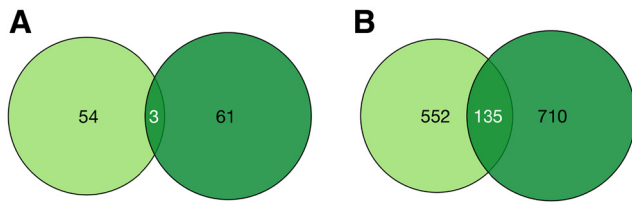
To identify the top-10 memory genes that are most likely linked to human memory function for future experimental investigation, we identified genes relevant for multiple gene sets obtained above with the LEA (Fig. 1E; Subramanian et al., 2005; Darby et al., 2016; Fleming and Miller, 2016). Previous work has shown that such genes that drive the enrichment of multiple gene sets are more likely related to the phenotype analyzed, i.e., memory function in this case (Subramanian et al., 2005; Darby et al., 2016; Fleming and Miller, 2016). The combination of GSEA and LEA were previously effective in identifying genetic signatures of cognitive functions (Thomassen et al., 2008; Erslund et al., 2012; Lee et al., 2013), including episodic and working memory (Heck et al., 2014;

Luksys et al., 2015). We applied LEA to the positively and negatively scoring gene sets above, followed by selecting the top-10 genes appearing most frequently across the leading-edge subsets of the gene sets. These genes were then validated with animal model literature, which were classified as strong or weak evidence supporting the link between the gene and memory function (Fig. 1F). Strong evidence was comprised of gene manipulation or drug treatment studies, e.g., gene knock-out leading to memory alteration. Weak evidence encompassed correlational or computational studies, such as gene upregulation that correlated with enhanced memory performance.

For cortical memory, nine out of 10 positively correlated genes were previously implicated in memory function (Table 3; full list of cortical memory genes and literature review Extended Data Table 3-1, complete LEA output in Extended Data Table 3-2). Genes *PRKCD* (Etcheberrigaray et al., 2004; Conboy et al., 2009), *RAC1* (Haditsch et al., 2009; Oh et al., 2010), *LIMK1* (Todorovski et al., 2015), and *CDC42* (Kim et al., 2014; Zhang et al., 2016) had strong associations with memory. For the corresponding negatively correlated candidate genes, all 10 genes had strong evidence supporting their role in memory. These were all genes encoding the histone H4 protein, which was linked to memory performance (Peleg et al., 2010). Deregulation of histone H4 acetylation in



**Figure 4.** Enrichment map visualization for subcortical memory. Clusters are labeled with P for positive, N for negative. Gene set clusters were found to be associated with memory. Positive clusters were associated with synaptic transmission, long-term plasticity, glutamate signaling, and neurite morphogenesis. Negative clusters included gene sets involved in transcription and translation, and glial cell differentiation. See Extended Data Figure 4-1 for the full output from GSEA Pre-ranked.



**Figure 5.** Overlap between cortical and subcortical memory gene sets and genes. **A**, Number of overlapping cortical and subcortical memory gene sets derived from GSEA. **B**, Number of overlapping cortical and subcortical memory genes derived from GSEA. Light green denotes cortical genes, dark green denotes subcortical genes. See Extended Data Figure 5-1 for the list of gene sets and genes that are shared or distinct across cortex and subcortex.

aged mice was linked to memory impairment, and reinstating this regulation improved their memory.

For subcortical memory, all 10 positively correlated genes were previously implicated in memory function (Table 4; full list of subcortical memory genes and literature review in Extended Data Table 4-1; LEA results in Extended Data Table 4-2). Genes *CDK5*, *NLGN1*, *RAB3A*, *STX1A*, *SNCA*, *SYT1*, and *UNC13A* were strongly linked to memory (Fujiwara et al., 2006; Yang et al., 2007; Liu et al., 2009; Guan et al., 2011; Kokhan et al., 2012; Bie et al., 2014; Mishiba et al., 2014; Böhme et al., 2019). Seven out of 10 negatively correlated candidate genes had weak evidence implicating them in memory. These were genes encoding ribosomal subunits, which were differentially expressed in rodents that display better

**Table 2. Comparison of cortical and subcortical genes associated with memory**

|                                    | GO term                                     | <i>p</i>               | <i>q</i>               |
|------------------------------------|---|------------------------|------------------------|
| Common across cortex and subcortex | Protein targeting to ER                     | $8.2 \times 10^{-67}$  | $1.9 \times 10^{-63}$  |
|                                    | mRNA catabolic process                      | $3.5 \times 10^{-41}$  | $8.5 \times 10^{-39}$  |
|                                    | Regulation of synaptic plasticity           | $4.3 \times 10^{-9}$   | $1.1 \times 10^{-7}$   |
|                                    | Glutamate receptor signaling pathway        | $4.1 \times 10^{-9}$   | $1.1 \times 10^{-7}$   |
| Cortex only                        | DNA repair                                  | $6.2 \times 10^{-29}$  | $7.3 \times 10^{-27}$  |
|                                    | Regulation of gene expression, epigenetic   | $3.7 \times 10^{-27}$  | $3.8 \times 10^{-25}$  |
|                                    | Interferon-gamma-mediated signaling pathway | $3.4 \times 10^{-23}$  | $2.4 \times 10^{-21}$  |
|                                    | Regulation of neurogenesis                  | $8.9 \times 10^{-16}$  | $4.1 \times 10^{-12}$  |
| Subcortex only                     | Neurogenesis                                | $3.6 \times 10^{-117}$ | $5.6 \times 10^{-114}$ |
|                                    | Neuron projection morphogenesis             | $5.4 \times 10^{-96}$  | $2.4 \times 10^{-93}$  |
|                                    | Glial cell differentiation                  | $6.0 \times 10^{-42}$  | $7.5 \times 10^{-40}$  |
|                                    | Myelination                                 | $2.1 \times 10^{-30}$  | $1.6 \times 10^{-28}$  |

We identified biological processes linked to memory genes shared or distinct across cortex and subcortex. GO = gene ontology biological process library; *p* and *q* refer to *p* value and FDR *q* value, respectively. See Extended Data Table 2-1 for the complete list of enriched gene sets and genes from ToppGene.

**Table 3. Candidate gene lists from cortical analyses of memory**

| CL       | Gene            | # leading edge subsets | mean <i>r</i> | Associated cognitive function |                  |                  |                  |
|----------|-----------------|------------------------|---------------|-------------------------------|------------------|------------------|------------------|
|          |                 |                        |               | Mem <sub>s</sub>              | Mem <sub>w</sub> | Mot <sub>s</sub> | Mot <sub>w</sub> |
| Memory   | <i>PRKCD</i>    | 9                      | 0.10          | Y                             |                  |                  |                  |
| Cortical | <i>RAC1</i>     | 9                      | 0.07          | Y                             |                  |                  |                  |
| +        | <i>HLA-DRB5</i> | 8                      | 0.12          |                               | Y                |                  |                  |
|          | <i>HLA-DRB1</i> | 8                      | 0.09          |                               | Y                |                  |                  |
|          | <i>HLA-DRB4</i> | 8                      | 0.09          |                               | Y                |                  |                  |
|          | <i>HCK</i>      | 8                      | 0.09          |                               | Y                |                  |                  |
|          | <i>HLA-DRB3</i> | 8                      | 0.08          |                               |                  |                  |                  |
|          | <i>LIMK1</i>    | 7                      | 0.13          | Y                             |                  |                  | Y                |
|          | <i>CDC42</i>    | 7                      | 0.11          | Y                             |                  |                  |                  |
|          | <i>VAV1</i>     | 6                      | 0.15          |                               | Y                |                  |                  |
| Memory   | <i>HIST1H4A</i> | 23                     | -0.18         | Y                             |                  |                  | Y                |
| Cortical | <i>HIST1H4E</i> | 23                     | -0.17         | Y                             |                  |                  | Y                |
| -        | <i>HIST1H4H</i> | 23                     | -0.13         | Y                             |                  |                  | Y                |
|          | <i>HIST1H4I</i> | 23                     | -0.12         | Y                             |                  |                  | Y                |
|          | <i>HIST4H4</i>  | 23                     | -0.11         | Y                             |                  |                  | Y                |
|          | <i>HIST1H4J</i> | 23                     | -0.11         | Y                             |                  |                  | Y                |
|          | <i>HIST1H4B</i> | 23                     | -0.10         | Y                             |                  |                  | Y                |
|          | <i>HIST1H4K</i> | 23                     | -0.10         | Y                             |                  |                  | Y                |
|          | <i>HIST1H4F</i> | 23                     | -0.09         | Y                             |                  |                  | Y                |
|          | <i>HIST1H4D</i> | 23                     | -0.09         | Y                             |                  |                  | Y                |

Candidate gene lists for the memory analysis of cortical regions, from positively and negatively correlated gene lists. Genes are ranked by the number of leading-edge subsets they appear in, and subsequently by mean *r* value. CL: candidate gene list; # leading-edge subsets: number of leading-edge subsets that the gene was found in; Mem<sub>s</sub>: strong evidence for memory function; Mem<sub>w</sub>: weak evidence for memory function; Mot<sub>s</sub>: strong evidence for motor function; Mot<sub>w</sub>: weak evidence for motor function; +: positively correlated candidate gene list; -: negatively correlated candidate gene list. See Extended Data Table 3-1 for the literature review supporting the cortical gene-cognition associations and Extended Data Table 3-2 for the complete LEA output for cortical analyses.

**Table 4. Candidate gene lists from subcortical analyses of memory**

| CL                   | Gene          | # leading edge subsets | mean $r$ | Associated cognitive function |                  |                  |                  |
|----------------------|---------------|------------------------|----------|-------------------------------|------------------|------------------|------------------|
|                      |               |                        |          | Mem <sub>s</sub>              | Mem <sub>w</sub> | Mot <sub>s</sub> | Mot <sub>w</sub> |
| Memory subcortical + | <i>CDK5</i>   | 27                     | 0.26     | Y                             |                  |                  | Y                |
|                      | <i>NLGN1</i>  | 26                     | 0.51     | Y                             |                  |                  |                  |
|                      | <i>UNC13B</i> | 26                     | 0.38     |                               | Y                |                  |                  |
|                      | <i>RAB3A</i>  | 25                     | 0.40     | Y                             |                  | Y                |                  |
|                      | <i>STX1A</i>  | 24                     | 0.57     | Y                             |                  |                  |                  |
|                      | <i>SYT12</i>  | 23                     | 0.44     |                               | Y                |                  |                  |
|                      | <i>STX1B</i>  | 22                     | 0.45     |                               | Y                |                  |                  |
|                      | <i>SNCA</i>   | 21                     | 0.44     | Y                             |                  | Y                |                  |
|                      | <i>SYT1</i>   | 21                     | 0.39     | Y                             |                  |                  | Y                |
|                      | <i>UNC13A</i> | 20                     | 0.46     | Y                             |                  |                  | Y                |
| Memory subcortical - | <i>RPL34</i>  | 8                      | -0.54    |                               | Y                |                  |                  |
|                      | <i>RPS12</i>  | 8                      | -0.49    |                               |                  |                  |                  |
|                      | <i>RPS13</i>  | 8                      | -0.47    |                               | Y                |                  |                  |
|                      | <i>RPS15A</i> | 8                      | -0.44    |                               | Y                |                  |                  |
|                      | <i>RPS29</i>  | 8                      | -0.44    |                               |                  |                  |                  |
|                      | <i>RPL11</i>  | 8                      | -0.44    |                               | Y                |                  |                  |
|                      | <i>RPL37A</i> | 8                      | -0.44    |                               |                  |                  |                  |
|                      | <i>RPL10</i>  | 8                      | -0.44    |                               | Y                |                  |                  |
|                      | <i>RPS25</i>  | 8                      | -0.44    |                               | Y                |                  |                  |
|                      | <i>RPS27</i>  | 8                      | -0.43    |                               | Y                |                  |                  |

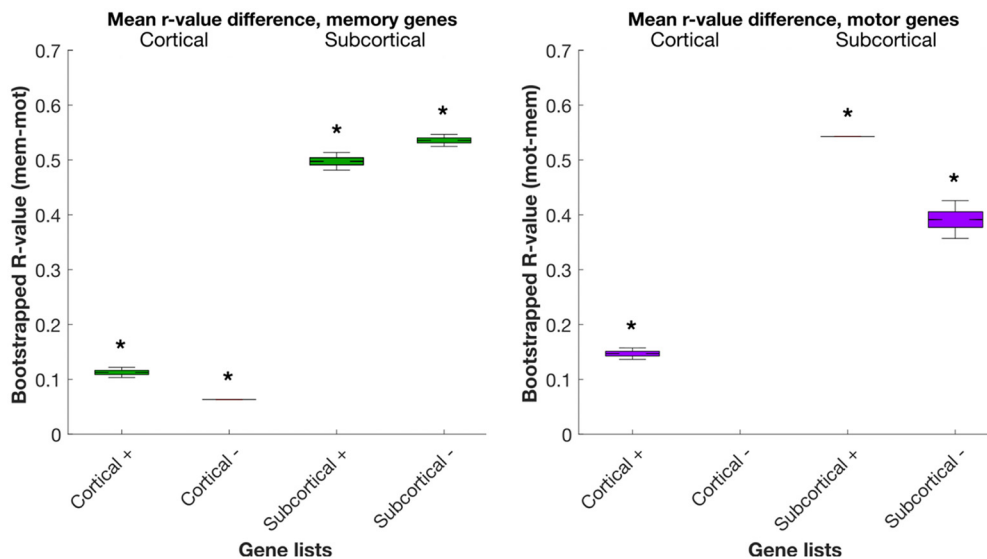
See Table 2 for notation, Extended Data Table 4-1 for the literature review supporting the subcortical gene-cognition associations, and Extended Data Table 4-2 for the complete LEA output for subcortical analyses.

memory performance (Wang et al., 2003; Kong et al., 2009; Winbush et al., 2012; Katz and Lamprecht, 2015; Oka et al., 2016; Zhang et al., 2018).

**Performance assessment of framework**

If our unsupervised approach is valid, for the memory analysis we expect that memory genes should have a

higher correlation value from the memory analysis compared with the motor analysis (i.e., sanity check; Figs. 1G, 6). Furthermore, in the top-10 memory genes, we expect that a greater number of memory genes in the memory analysis than expected by chance (i.e., statistical significance; Tables 5, 6), and that we find more memory genes than motor function genes (i.e., method precision; Fig. 7).



**Figure 6.** Bootstrapped correlation value differences for all cortical and subcortical candidate genes of memory and motor analysis. For a given memory gene, we calculated the difference between memory and motor analysis  $r$  values by subtracting motor  $r$  from memory  $r$ . If the memory  $r$  was negative, we took the negative of the difference (to get a positive value). Vice versa for the motor genes. For each cognitive function, we subsampled the number of genes used to the lowest number for calculating the bootstrapped mean difference (231 memory genes and 146 motor genes, respectively, 10,000 iterations). If the 95th percentile did not overlap with the baseline of zero, the bootstrapped difference is considered significant ( $p < 0.05$ ). Note that for the motor cortical analysis, no negatively correlated genes survived the threshold and thus no motor cortical (-) gene list is shown here. See Extended Data Figure 6-1 for the complete list of correlation value differences for genes used in the bootstrap analysis. \*denotes  $p < 0.05$ .

**Table 5. Probability of observing the number of memory or motor genes by chance**

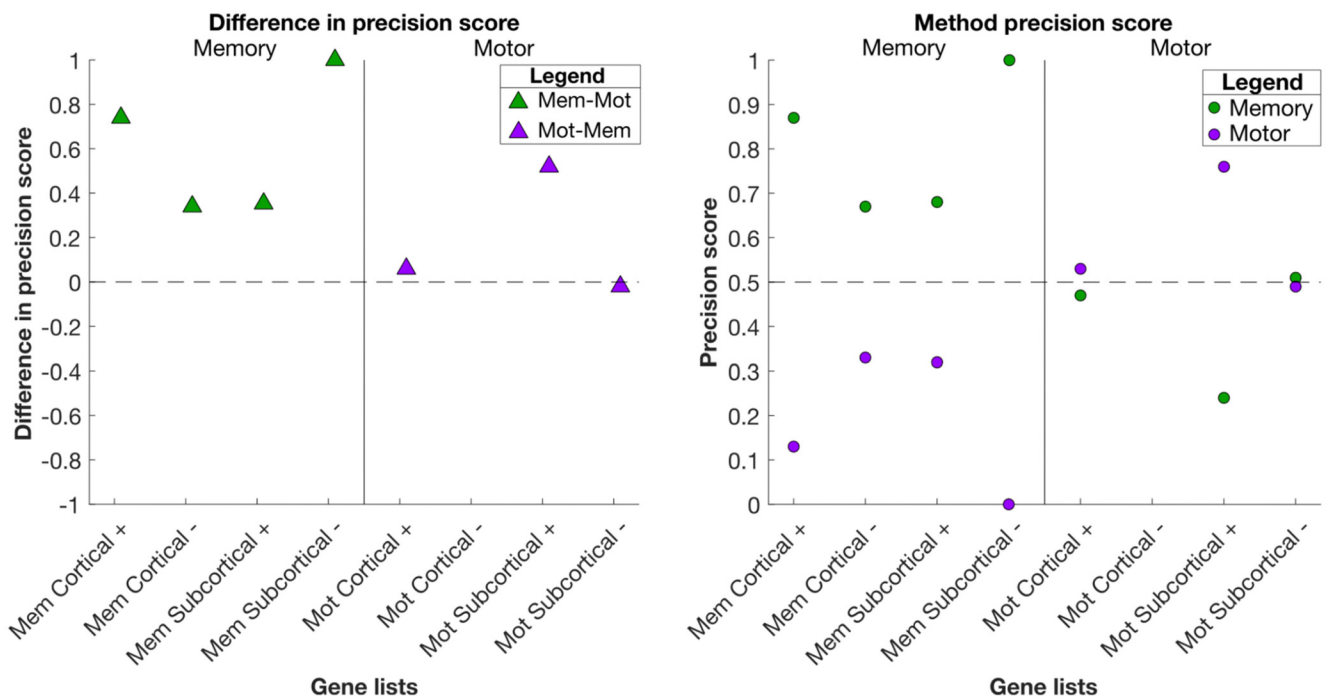
| Cognitive function | Brain region | Gene list | # cognitive function genes (of 10) | <i>p</i>              |
|--------------------|--------------|-----------|------------------------------------|-----------------------|
| Memory             | Cortical     | +         | 9                                  | $3.3 \times 10^{-13}$ |
|                    |              | -         | 10                                 | $1.3 \times 10^{-14}$ |
|                    | Subcortical  | +         | 10                                 | $1.3 \times 10^{-14}$ |
|                    |              | -         | 7                                  | $2.0 \times 10^{-10}$ |
| Motor              | Cortical     | +         | 6                                  | $7.5 \times 10^{-14}$ |
|                    |              | -         | n/a                                | n/a                   |
|                    | Subcortical  | +         | 8                                  | $2.9 \times 10^{-18}$ |
|                    |              | -         | 10                                 | $1.1 \times 10^{-22}$ |

We calculated the chance probability of obtaining *N* memory genes per gene list (without replacement), by using the proportion of known memory genes out of the 15,625 genes analyzed. Vice versa for motor genes. Note that for the motor cortical (-) analysis, no genes survived the threshold, and thus, no gene list is shown here. See Extended Data Table 5-1 for the known memory and motor function genes and derived effectiveness scores across all gene lists.

**Table 6. Statistical table**

| Results section            | Data structure   | Type of test   | Power            |
|----------------------------|--|--|------------------|
| <a href="#">Figs. 3, 4</a> | Gene set enrichment analysis                                       | Kolmogorov–Smirnov test  | FDR < 0.05       |
| <a href="#">Figs. 5, 6</a> | Distribution of bootstrapped correlation <i>r</i> value difference | Overlap of 95th percentiles with baseline of zero                                  | <i>p</i> < 0.05  |
| <a href="#">Table 4</a>    | Distribution of known memory and motor genes out of 15,625 genes   | Probability of obtaining <i>n</i> memory/motor genes out of 10 without replacement | <i>p</i> < 0.001 |

Statistical tests for the gene set enrichment analysis, bootstrapped correlation *r* value differences between the memory and motor analyses and the probability of obtaining *n* memory/motor genes by chance from a known pool of memory and motor genes.



**Figure 7.** Precision scores for top-10 cortical and subcortical candidate genes of memory and motor analysis. For a given memory gene list, we calculated the memory and motor precision scores with Equations 1, 2 and their difference. Ideally, memory gene lists should obtain a memory score above 0.5, and a motor score below 0.5, and vice versa for the motor genes. Note that for the motor cortical analysis, no negatively correlated genes survived the threshold and thus no motor cortical (-) gene list is shown. See Extended Data Figure 7-1 for the candidate genes of each analysis and the derived method precision score for each gene list.

Using the candidate gene correlation values, we show that the memory genes displayed a significant positive difference between memory analysis *r* value and motor analysis *r* value, as the 95th percentile (whiskers) did not overlap with zero (Fig. 6; all gene correlation values used

in the bootstrap analysis in Extended Data Fig. 6-1). As such, our approach performs as expected.

We found that the method was highly effective. For all memory cortical and subcortical gene lists, the probability of obtaining the number of memory genes observed was

significantly above chance (Table 5; full list of memory-related and motor function-related genes that constitute the chance probability in Extended Data Table 5-1). Likewise, for all motor cortical and subcortical gene lists, the probability of deriving the number motor genes observed was highly significant as well.

Using the putative gene functions inferred from the literature review, we also found that the method had high precision, as the difference in top-10 candidate gene list precision scores are non-negative [except for motor subcortex (-), Fig. 7; calculation of precision values in Extended Data Fig. 7-1]. These results suggest that the method is valid and specific in identifying genes associated with memory and motor function.

## Discussion

Taken together, our results show that cortical and subcortical regions involved in human memory possess distinct genetic signatures. These genetic signatures are in agreement with prior research in animal models of memory, and were dissociable from the control of motor function. Thus, we show that the strong similarities between the spatial patterns of human brain transcriptome and the functional neuroimaging map of memory can be exploited to highlight candidate biological processes and genes associated with human memory for future experimental investigations. This may contribute to our knowledge of the functional differences of cortical and subcortical regions in healthy human memory function and memory disorders.

Presently, human memory evidence is generally derived from popular non-invasive methods such as GWAS (Wellcome Trust Case Control Consortium, 2007), which identifies links between gene variants and cognition (Heck et al., 2014). However, GWAS ignores the spatially distributed gene expression in the brain by solely analyzing gene variants in relation to brain or behavioral measures (Hawrylycz et al., 2012; Mahfouz et al., 2017). Our approach relies on spatial pattern of gene expression and identifies genetic profiles related to human memory. Crucially, our unsupervised approach is versatile as it can reveal unprecedented insights into any human cognitive function of interest, e.g., decision making. This insight may be especially useful in the case of functions that are clinically relevant but with a genetic basis that is less understood, e.g., attention (ADHD) and language (dyslexia).

To identify general human memory genes that function across the brain, we compared the differences and the overlap between cortical and subcortical memory genes (Fig. 1). Particularly, this overlap comparison is supported by the existence of genes underlying memory function as a whole, as in the case of neuronally-expressed immediate-early genes (IEGs) involved in memory function (Gallo et al., 2018). IEGs are a broad class of genes that are expressed in a rapid, transient manner in response to a plethora of cellular stimuli. Of the neuron-specific IEGs, *c-fos*, *Egr1*, and *arc* are broadly associated with various facets of memory across both cortical and subcortical areas. For example, the blockade of hippocampal *c-Fos* negatively impacted spatial long-term memory (Kemp et al., 2013), and its blockade in either the hippocampus or retro-

splenial cortex induced deficits in the consolidation of fear memory (Katche et al., 2010; Katche and Medina, 2017). Such genes are relevant for different subtypes of memory across both cortical and subcortical areas, which we term whole-brain general memory genes.

If there are such general memory genes whose function in memory spans the whole brain, both cortical and subcortical analyses should show overlapping genes. We found that the cortical and subcortical areas possess largely distinct genetic profiles, as identified by gene-functional spatial correlation (Fig. 5). There was no overlap in the top-10 cortical and subcortical memory genes, with some overlap for memory genes (9.6% out of 1397 genes) and biological process gene sets (2.5% out of 118 gene sets).

At the biological process level, we found differences in cortical and subcortical memory. In the cortex, the identified gene sets included epigenetic regulation and immune signaling. The latter received recent interest as a central factor in the onset and progression of dementia (Litteljohn et al., 2014; Kim and Kaang, 2017; Hammond et al., 2019). In the subcortex, the identified genes are involved in neurogenesis and glial cell differentiation. Furthermore, we identified gene sets with a less understood link to memory as well. For instance, astrocytes and oligodendrocytes were recently discovered to be involved in linking glial-mediated potassium homeostasis and myelination to memory deficits (Hertz and Chen, 2016; Pepper et al., 2018). It is still unclear how myelinating oligodendrocytes may enable plasticity in memory (Pepper et al., 2018). Our work suggests that glial cell differentiation may play a complementary role in memory function, and should be further investigated for a comprehensive understanding of cellular contributions to memory. Overall, this may suggest inherent differences in the biological processes supporting cortical and subcortical memory regions. Future work may look into the interplay of these processes and clarify their differential contributions toward cortical and subcortical memory function.

At the gene level, enriched genes for cortical and subcortical memory were similarly distinct. Of the enriched genes that are associated with the biological processes above (in sets  $S^+$  and  $S^-$ ), a small proportion of genes (9.6%, or 135 genes) were shared between cortical and subcortical regions (Fig. 5). These genes are related to the Arp2/3 complex, GABA and AMPA ligand-gated ion channels, and srp-dependent protein localization to the membrane. The Arp2/3 complex is necessary for the maturation of dendritic spines, hippocampal and extra-hippocampal AMPA receptors are involved in excitatory ion channels in memory, and GABA receptor subunits are part of inhibitory ion channels in memory function (Collinson et al., 2002; Freudenberg et al., 2016; Spence et al., 2016). As such, this recapitulates known literature, and hints at basic requirements for general memory function. Overall, this may suggest differences due to gross cortico-subcortical differences in transcriptome profiles and function in healthy memory function and disease (Huber et al., 1986; Salmon and Filoteo, 2007). Future work may look into how the convergence and divergence between cortical and sub-

cortical genetic profiles and how those enable cortico-subcortical-specific functions in memory.

Additionally, our approach also identified memory-associated genes with poorly understood relations to memory. For example, the *MIS18BP1* gene was identified in the subcortical memory genes (Extended Data Table 2-1). This gene is required for recruitment of centromere proteins to centromeres and allow normal chromosome segregation during mitosis (Moree et al., 2011). It is unclear whether such cell division genes play a role in memory across subcortical areas. However, the gene has been linked to hippocampal neurogenesis, which is critical for hippocampal function in memory (Shin et al., 2015; Gonçalves et al., 2016). Such lesser known genes constitute a crucial contribution of our framework, as their immediate link to memory yet to be established, and should be examined in future research.

Our analyses of gene expression and neuroimaging maps are not without limitations. These include the limited sample size, the validity of a text-mining-like approach with GSEA and Gene Ontology library, and the spatial resolution of the AHBA. First, the limited donor sample size and reduced genome coverage after preprocessing may contribute to a reduced power, but not statistical precision, of our approach. Although future increase in sample size may identify more genes using this method, we found the current results to be robust as our results are significantly better than chance (i.e., statistical significance). Furthermore, the identified genes were specific to memory, as demonstrated by the precision of our framework. Second, GSEA utilizes the Gene Ontology library to identify enriched gene sets, and associates these enriched genes with the library's ontological terms, e.g., synaptic plasticity. We concede that the Gene Ontology library is continually being extended with manual curation efforts, and thus is vulnerable to being outpaced by the deluge of recent experimental findings (Baumgartner et al., 2007; Dutkowski et al., 2013; Gaudet and Dessimoz, 2017). As such, it is possible that the database is incomplete and does not reflect all biological functions associated with each gene. This may lead to false negatives, where we miss genes that should be considered enriched. Nevertheless, our approach demonstrates high effectiveness (as seen in the top-10 memory and motor function genes) and the results are in concordance with known experimental literature independent of ontology libraries. Additionally, unsupervised methods of identifying candidate genes always require manual curation and selection of these genes for further investigation. Third, this approach is also limited by the spatial resolution of the human brain transcriptome. Despite being the most appropriate human transcriptional atlas with its whole-genome and high resolution whole-brain coverage, the AHBA map still has a lower resolution compared to functional imaging maps, especially in the cortex (Hawrylycz et al., 2011). As such, we expect the precision and statistical power of our approach to grow as the spatial resolution and sample size of the AHBA database increases. Furthermore, as the translation of gene mRNA into a functional product is subject to regulation, donor

brain proteomes may be complementary in identifying genes linked to memory (Lubec et al., 2003; Park et al., 2006; Sjöstedt et al., 2015).

## Conclusion

Here, using the Allen Institute brain transcriptional atlas and Neurosynth neuroimaging maps, we demonstrate that cortical and subcortical memory regions have distinct genetic signatures. These genetic signatures provide novel biological processes and molecular targets for understanding of human memory function. Crucially, we hope that our unsupervised and spatially guided approach may help guide researchers toward productive gene and biological process candidates for understanding how complex cognitive functions such as memory may be enabled by the molecular components of the brain.

## References

- Alberini CM, Kandel ER (2015) The regulation of transcription in memory consolidation. *Cold Spring Harb Perspect Biol* 7:a021741.
- Amatkevičiūtė A, Fulcher BD, Fornito A (2019) A practical guide to linking brain-wide gene expression and neuroimaging data. *Neuroimage* 189:353–367.
- Basu S, Kustanovich I, Lamprecht R (2016) Arp2/3 and VASP are essential for fear memory formation in lateral amygdala. *eNeuro* 3:ENEURO.0302-16.2016.
- Baumgartner WA, Cohen KB, Fox LM, Acquaaah-Mensah G, Hunter L (2007) Manual curation is not sufficient for annotation of genomic databases. *Bioinformatics* 23:i41–i48.
- Berto S, Wang G-Z, Germi J, Lega BC, Konopka G (2017) Human genomic signatures of brain oscillations during memory encoding. *Cereb Cortex* 89:1–16.
- Bie B, Wu J, Yang H, Xu JJ, Brown DL, Naguib M (2014) Epigenetic suppression of neuroligin 1 underlies amyloid-induced memory deficiency. *Nat Neurosci* 17:223–231.
- Böhme MA, McCarthy AW, Grasskamp AT, Beuschel CB, Goel P, Jusyte M, Laber D, Huang S, Rey U, Petzoldt AG, Lehmann M, Göttfert F, Haghghi P, Hell SW, Oswald D, Dickman D, Sigrist SJ, Walter AM (2019) Rapid active zone remodeling consolidates pre-synaptic potentiation. *Nat Commun* 10:1085.
- Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S; AmiGO Hub; Web Presence Working Group (2009) AmiGO: online access to ontology and annotation data. *Bioinformatics* 25:288–289.
- Chen J, Bardes EE, Aronow BJ, Jegga AG (2009) ToppGene suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res* 37:W305–W311.
- Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, et al. (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* 2:2366–2382.
- Collinson N, Kuenzi FM, Jarolimek W, Maubach KA, Cothliff R, Sur C, Smith A, Otu FM, Howell O, Atack JR, McKernan RM, Seabrook GR, Dawson GR, Whiting PJ, Rosahl TW (2002) Enhanced learning and memory and altered GABAergic synaptic transmission in mice lacking the alpha 5 subunit of the GABAA receptor. *J Neurosci* 22:5572–5580.
- Conboy L, Foley AG, O'Boyle NM, Lawlor M, Gallagher HC, Murphy KJ, Regan CM (2009) Curcumin-induced degradation of PKC delta is associated with enhanced dentate NCAM PSA expression and spatial learning in adult and aged Wistar rats. *Biochem Pharmacol* 77:1254–1265.

- Darby MM, Yolken RH, Sabuncian S (2016) Consistently altered expression of gene sets in postmortem brains of individuals with major psychiatric disorders. *Transl Psychiatry* 6:e890.
- Dutkowski J, Kramer M, Surma MA, Balakrishnan R, Cherry JM, Krogan NJ, Ideker T (2013) A gene ontology inferred from molecular networks. *Nat Biotechnol* 31:38–45.
- Ersland KM, Christoforou A, Stansberg C, Espeseth T, Mattheisen M, Mattingsdal M, Hardarson GA, Hansen T, Fernandes CP, Giddaluru S, Breuer R, Strohmaier J, Djurovic S, Nöthen MM, Rietschel M, Lundervold AJ, Werge T, Cichon S, Andreassen OA, Reinvang I, et al. (2012) Gene-based analysis of regionally enriched cortical genes in GWAS data sets of cognitive traits and psychiatric disorders. *PLoS One* 7:e31687.
- Etcheberrigaray R, Tan M, Dewachter I, Kuiperi C, Van der Auwera I, Wera S, Qiao L, Bank B, Nelson TJ, Kozikowski AP, Van Leuven F, Alkon DL (2004) Therapeutic effects of PKC activators in Alzheimer's disease transgenic mice. *Proc Natl Acad Sci USA* 101:11141–11146.
- Fernandez-Vizarrá P, Lopez-Franco O, Mallavia B, Higuera-Matas A, Lopez-Parra V, Ortiz-Muñoz G, Ambrosio E, Egidio J, Almeida OFX, Gomez-Guerrero C (2012) Immunoglobulin G Fc receptor deficiency prevents Alzheimer-like pathology and cognitive impairment in mice. *Brain* 135:2826–2837.
- Fleming DS, Miller LC (2016) Leading edge analysis of transcriptomic changes during pseudorabies virus infection. *Genom Data* 10:104–106.
- Fox AS, Chang LJ, Gorgolewski KJ, Yarkoni T (2014) Bridging psychology and genetics using large-scale spatial analysis of neuroimaging and neurogenetic data. *bioRxiv*. Available at <https://www.biorxiv.org/content/10.1101/012310v1>.
- Freudenberg-Hua Y, Li W, Davies P (2018) The role of genetics in advancing precision medicine for Alzheimer's disease—A narrative review. *Front Med (Lausanne)* 5:108.
- Freudenberg F, Resnik E, Kolléker A, Celikel T, Sprengel R, Seeburg PH (2016) Hippocampal GluA1 expression in *Gria1*<sup>−/−</sup> mice only partially restores spatial memory performance deficits. *Neurobiol Learn Mem* 135:83–90.
- Fujiwara T, Mishima T, Kofuji T, Chiba T, Tanaka K, Yamamoto A, Akagawa K (2006) Analysis of knock-out mice to determine the role of HPC-1/syntaxin 1A in expressing synaptic plasticity. *J Neurosci* 26:5767–5776.
- Gallo FT, Katche C, Morici JF, Medina JH, Weisstaub NV (2018) Immediate early genes, memory and psychiatric disorders: focus on c-Fos, *Egr1* and *Arc*. *Front Behav Neurosci* 12:79.
- Gasbarri A, Pompili A (2014) The role of GABA in memory processes. In: Identification of neural markers accompanying memory, pp 47–62. Waltham, MA: Elsevier.
- Gaudet P, Dessimoz C (2017) Gene ontology: pitfalls, biases, and remedies, pp 189–205. New York: Humana Press.
- Gonçalves JT, Schafer ST, Gage FH (2016) Adult neurogenesis in the hippocampus: from stem cells to behavior. *Cell* 167:897–914.
- Guan JS, Su SC, Gao J, Joseph N, Xie Z, Zhou Y, Durak O, Zhang L, Zhu JJ, Clauser KR, Carr SA, Tsai LH (2011) *Cdk5* is required for memory function and hippocampal plasticity via the camp signaling pathway. *PLoS One* 6:e25735.
- Haditsch U, Leone DP, Farinelli M, Chrostek-Grashoff A, Brakebusch C, Mansuy IM, McConnell SK, Palmer TD (2009) A central role for the small GTPase *Rac1* in hippocampal plasticity and spatial learning and memory. *Mol Cell Neurosci* 41:409–419.
- Hammond TR, Marsh SE, Stevens B (2019) Immune signaling in neurodegeneration. *Immunity* 50:955–974.
- Hawrylycz M, Ng L, Page D, Morris J, Lau C, Faber S, Faber V, Sunkin S, Menon V, Lein E, Jones A (2011) Multi-scale correlation structure of gene expression in the brain. *Neural Netw* 24:933–942.
- Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, van de Lagemaat LN, Smith KA, Ebbert A, Riley ZL, Abajian C, Beckmann CF, Bernard A, Bertagnolli D, Boe AF, Cartagena PM, Chakravarty MM, Chapin M, Chong J, Dalley RA, et al. (2012) An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489:391–399.
- Heck A, Fastenrath M, Ackermann S, Auschra B, Bickel H, Coynel D, Gschwind L, Jessen F, Kaduszkiewicz H, Maier W, Milnik A, Pentzek M, Riedel-Heller SG, Ripke S, Spalek K, Sullivan P, Vogler C, Wagner M, Weyerer S, Wolfsgruber S, et al. (2014) Converging genetic and functional brain imaging evidence links neuronal excitability to working memory, psychiatric disease, and brain activity. *Neuron* 81:1203–1213.
- Hertz L, Chen Y (2016) Editorial: all 3 types of glial cells are important for memory formation. *Front Integr Neurosci* 10:31.
- Hou Y, Lautrup S, Cordonnier S, Wang Y, Croteau DL, Zavala E, Zhang Y, Moritoh K, O'Connell JF, Baptiste BA, Stevnsner TV, Mattson MP, Bohr VA (2018) NAD<sup>+</sup> supplementation normalizes key Alzheimer's features and DNA damage responses in a new AD mouse model with introduced DNA repair deficiency. *Proc Natl Acad Sci USA* 115:E1876–E1885.
- Huber SJ, Shuttlesworth EC, Paulson GW, Bellchambers MJ, Clapp LE (1986) Cortical vs subcortical dementia. *Arch Neurol* 43:392.
- Jarome TJ, Helmstetter FJ (2014) Protein degradation and protein synthesis in long-term memory formation. *Front Mol Neurosci* 7:61.
- Ji JL, Spronk M, Kulkarni K, Repovš G, Anticevic A, Cole MW (2019) Mapping the human brain's cortical-subcortical functional network organization. *Neuroimage* 185:35–57.
- Kandel ER, Dudai Y, Mayford MR (2014) The molecular and systems biology of memory. *Cell* 157:163–186.
- Kao YC, Wang IF, Tsai KJ (2018) miRNA-34c overexpression causes dendritic loss and memory decline. *Int J Mol Sci* 19.
- Kasai H, Fukuda M, Watanabe S, Hayashi-Takagi A, Noguchi J (2010) Structural dynamics of dendritic spines in memory and cognition. *Trends Neurosci* 33:121–129.
- Katche C, Bekinschtein P, Slipczuk L, Goldin A, Izquierdo IA, Cammarota M, Medina JH (2010) Delayed wave of c-Fos expression in the dorsal hippocampus involved specifically in persistence of long-term memory storage. *Proc Natl Acad Sci USA* 107:349–354.
- Katche C, Medina JH (2017) Requirement of an early activation of BDNF/c-Fos cascade in the retrosplenial cortex for the persistence of a long-lasting aversive memory. *Cereb Cortex* 27:1060–1067.
- Katz IK, Lamprecht R (2015) Fear conditioning leads to alteration in specific genes expression in cortical and thalamic neurons that project to the lateral amygdala. *J Neurochem* 132:313–326.
- Kemp A, Tischmeyer W, Manahan-Vaughan D (2013) Learning-facilitated long-term depression requires activation of the immediate early gene, c-fos, and is transcription dependent. *Behav Brain Res* 254:83–91.
- Kim S, Kaang BK (2017) Epigenetic regulation and chromatin remodeling in learning and memory. *Exp Mol Med* 49:e281.
- Kim IH, Wang H, Soderling SH, Yasuda R (2014) Loss of *Cdc42* leads to defects in synaptic plasticity and remote memory recall. *Elife* 3:e02839.
- Kim EJ, Juavinett AL, Kyubwa EM, Jacobs MW, Callaway EM (2015) Three types of cortical layer 5 neurons that differ in brain-wide connectivity and function. *Neuron* 88:1253–1267.
- Kokhan VS, Afanasyeva MA, Van'kin GI (2012)  $\alpha$ -Synuclein knockout mice have cognitive impairments. *Behav Brain Res* 231:226–230.
- Kong W, Mou X, Liu Q, Chen Z, Vanderburg CR, Rogers JT, Huang X (2009) Independent component analysis of Alzheimer's DNA microarray gene expression data. *Mol Neurodegener* 4:5.
- Krucker T, Siggins GR, Halpain S (2000) Dynamic actin filaments are required for stable long-term potentiation (LTP) in area CA1 of the hippocampus. *Proc Natl Acad Sci USA* 97:6856–6861.
- LaBar KS, Cabeza R (2006) Cognitive neuroscience of emotional memory. *Nat Rev Neurosci* 7:54–64.
- Lamprecht R (2011) The roles of the actin cytoskeleton in fear memory formation. *Front Behav Neurosci* 5:39.
- Lee MY, Yu JH, Kim JY, Seo JH, Park ES, Kim CH, Kim H, Cho SR (2013) Alteration of synaptic activity-regulating genes underlying functional improvement by long-term exposure to an enriched environment in the adult brain. *Neurorehabil Neural Repair* 27:561–574.

- Lin MK, Farrer MJ (2014) Genetics and genomics of Parkinson's disease. *Genome Med* 6:48.
- Litteljohn D, Nelson E, Hayley S (2014) IFN- $\gamma$  differentially modulates memory-related processes under basal and chronic stressor conditions. *Front Cell Neurosci* 8:391.
- Liu YF, Chen H, Wu CL, Kuo YM, Yu L, Huang AM, Wu FS, Chuang JI, Jen CJ (2009) Differential effects of treadmill running and wheel running on spatial or aversive learning and memory: roles of amygdalar brain-derived neurotrophic factor and synaptotagmin I. *J Physiol* 587:3221–3231.
- Lubec G, Krappfenbauer K, Fountoulakis M (2003) Proteomics in brain res: potentials and limitations. *Prog Neurobiol* 69:193–211.
- Luksys G, Fastenrath M, Coyne D, Freytag V, Gschwind L, Heck A, Jessen F, Maier W, Milnik A, Riedel-Heller SG, Scherer M, Spalek K, Vogler C, Wagner M, Wolfsgruber S, Papassotiropoulos A, de Quervain DF (2015) Computational dissection of human episodic memory reveals mental process-specific genetic profiles. *Proc Natl Acad Sci USA* 112:E4939–E4948.
- Mahfouz A, Huisman SMH, Lelieveldt BPF, Reinders MJT (2017) Brain transcriptome atlases: a computational perspective. *Brain Struct Funct* 222:1557–1580.
- Marin I, Kipnis J (2013) Learning and memory . . . and the immune system. *Learn Mem* 20:601–606.
- Mengod G, Villaró MT, Landwehrmeyer GB, Martínez-Mir MI, Niznik HB, Sunahara RK, Seeman P, O'Dowd BF, Probst A, Palacios JM (1992) Visualization of dopamine D1, D2 and D3 receptor mRNAs in human and rat brain. *Neurochem Int* 20:33–43.
- Merico D, Isserlin R, Stueker O, Emili A, Bader GD (2010) Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* 5:e13984.
- Mishiba T, Tanaka M, Mita N, He X, Sasamoto K, Itoharu S, Ohshima T (2014) Cdk5/p35 functions as a crucial regulator of spatial learning and memory. *Mol Brain* 7:82.
- Modha DS, Singh R (2010) Network architecture of the long-distance pathways in the macaque brain. *Proc Natl Acad Sci USA* 107:13485–13490.
- Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstråle M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, et al. (2003) PGC-1 $\alpha$ -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* 34:267–273.
- Moree B, Meyer CB, Fuller CJ, Straight AF (2011) CENP-C recruits M18BP1 to centromeres to promote CENP-A chromatin assembly. *J Cell Biol* 194:855–871.
- O'Leary DD, Koester SE (1993) Development of projection neuron types, axon pathways, and patterned connections of the mammalian cortex. *Neuron* 10:991–1006.
- Oesper L, Merico D, Isserlin R, Bader GD (2011) WordCloud: a Cytoscape plugin to create a visual semantic summary of networks. *Source Code Biol Med* 6:7.
- Oh D, Han S, Seo J, Lee JR, Choi J, Groffen J, Kim K, Cho YS, Choi HS, Shin H, Woo J, Won H, Park SK, Kim SY, Jo J, Whitcomb DJ, Cho K, Kim H, Bae YC, Heisterkamp N, et al. (2010) Regulation of synaptic Rac1 activity, long-term potentiation maintenance, and learning and memory by BCR and ABR Rac GTPase-activating proteins. *J Neurosci* 30:14134–14144.
- Oka S, Leon J, Sakumi K, Ide T, Kang D, LaFerla FM, Nakabeppu Y (2016) Human mitochondrial transcriptional factor A breaks the mitochondria-mediated vicious cycle in Alzheimer's disease. *Sci Rep* 6:37889.
- Papassotiropoulos A, de Quervain DJF (2011) Genetics of human episodic memory: dealing with complexity. *Trends Cogn Sci* 15:381–387.
- Pappata S, Dehaene S, Poline JB, Gregoire MC, Jobert A, Delforge J, Frouin V, Bottlaender M, Dolle F, Di Giambardino L, Syrota A (2002) In vivo detection of striatal dopamine release during reward: a PET study with [ $^{11}$ C]raclopride and a single dynamic scan approach. *Neuroimage* 16:1015–1027.
- Park YM, Kim JY, Kwon K-H, Lee SK, Kim YH, Kim S-Y, Park GW, Lee JH, Lee B, Yoo JS (2006) Profiling human brain proteome by multi-dimensional separations coupled with MS. *Proteomics* 6:4978–4986.
- Peleg S, Sananbenesi F, Zovoillis A, Burkhardt S, Bahari-Javan S, Agis-Balboa RC, Cota P, Wittnam JL, Gogol-Doering A, Opitz L, Salinas-Riester G, Dettenhofer M, Kang H, Farinelli L, Chen W, Fischer A (2010) Altered histone acetylation is associated with age-dependent memory impairment in mice. *Science* 328:753–756.
- Peng S, Zhang Y, Zhang J, Wang H, Ren B (2011) Glutamate receptors and signal transduction in learning and memory. *Mol Biol Rep* 38:453–460.
- Pepper RE, Pitman KA, Cullen CL, Young KM (2018) How do cells of the oligodendrocyte lineage affect neuronal circuits to influence motor function, memory and mood? *Front Cell Neurosci* 12:399.
- Richiardi J, Altmann A, Milazzo AC, Chang C, Chakravarty MM, Banaschewski T, Barker GJ, Bokde AL, Bromberg U, Büchel C, Conrod P, Fauth-Bühler M, Flor H, Frouin V, Gallinat J, Garavan H, Gowland P, Heinz A, Lemaître H, Mann KF, et al. (2015) Brain networks. Correlated gene expression supports synchronous activity in brain networks. *Science* 348:1241–1244.
- Ritchie J, Pantazatos SP, French L (2018) Transcriptomic characterization of MRI contrast with focus on the T1-w/T2-w ratio in the cerebral cortex. *Neuroimage* 174:504–517.
- Rosenberg T, Gal-Ben-Ari S, Dieterich DC, Kreutz MR, Ziv NE, Gundelfinger ED, Rosenblum K (2014) The roles of protein expression in synaptic plasticity and memory consolidation. *Front Mol Neurosci* 7:86.
- Salmon D, Filoteo J (2007) Neuropsychology of cortical versus subcortical dementia syndromes. *Semin Neurol* 27:7–21.
- Schott BH, Minuzzi L, Krebs RM, Elmenhorst D, Lang M, Winz OH, Seidenbecher CI, Coenen HH, Heinze H-J, Zilles K, Düssel E, Bauer A (2008) Mesolimbic functional magnetic resonance imaging activations during reward anticipation correlate with reward-related ventral striatal dopamine release. *J Neurosci* 28:14311–14319.
- Shin J, Berg DA, Zhu Y, Shin JY, Song J, Bonaguidi MA, Enikolopov G, Nauen DW, Christian KM, Ming G, Song H (2015) Single-cell RNA-Seq with waterfall reveals molecular cascades underlying adult neurogenesis. *Cell Stem Cell* 17:360–372.
- Sjöstedt E, Fagerberg L, Hallström BM, Häggmark A, Mitsios N, Nilsson P, Pontén F, Hökfelt T, Uhlén M, Mulder J (2015) Defining the human brain proteome using transcriptomics and antibody-based profiling with a focus on the cerebral cortex. *PLoS One* 10:e0130028.
- Spence EF, Kanak DJ, Carlson BR, Soderling SH (2016) The Arp2/3 complex is essential for distinct stages of spine synapse maturation, including synapse unsilencing. *J Neurosci* 36:9696–9709.
- Squire LR, Wixted JT (2011) The cognitive neuroscience of human memory since H.M. *Annu Rev Neurosci* 34:259–288.
- Stuchlik A (2014) Dynamic learning and memory, synaptic plasticity and neurogenesis: an update. *Front Behav Neurosci* 8:106.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 102:15545–15550.
- Takemoto K, Iwanari H, Tada H, Suyama K, Sano A, Nagai T, Hamakubo T, Takahashi T (2017) Optical inactivation of synaptic AMPA receptors erases fear memory. *Nat Biotechnol* 35:38–47.
- Thomassen M, Tan Q, Kruse TA (2008) Gene expression meta-analysis identifies metastatic pathways and transcription factors in breast cancer. *BMC Cancer* 8:394.
- Todorovski Z, Asrar S, Liu J, Saw NMN, Joshi K, Cortez MA, Snead OC, Xie W, Jia Z (2015) LIMK1 regulates long-term memory and synaptic plasticity via the transcriptional factor CREB. *Mol Cell Biol* 35:1316–1328.
- Van Cauwenbergh C, Van Broeckhoven C, Sleegers K (2016) The genetic landscape of Alzheimer disease: clinical implications and perspectives. *Genet Med* 18:421–430.



- Wang G, Zhang Y, Chen B, Cheng J (2003) Preliminary studies on Alzheimer's disease using cDNA microarrays. *Mech Ageing Dev* 124:115–124.
- Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–678.
- Winbush A, Reed D, Chang PL, Nuzhdin SV, Lyons LC, Arbeitman MN (2012) Identification of gene expression changes associated with long-term memory of courtship rejection in *Drosophila* males. *G3 (Bethesda)* 2:1437–1445.
- Yang S, Farias M, Kapfhamer D, Tobias J, Grant G, Abel T, Bućan M (2007) Biochemical, molecular and behavioral phenotypes of Rab3A mutations in the mouse. *Genes Brain Behav* 6:77–96.
- Yarkoni T, Poldrack RA, Nichols TE, Van Essen DC, Wager TD (2011) Large-scale automated synthesis of human functional neuroimaging data. *Nat Methods* 8:665–670.
- Yushkevich PA, Avants BB, Pluta J, Das S, Minkoff D, Mechanic-Hamilton D, Glynn S, Pickup S, Liu W, Gee JC, Grossman M, Detre JA (2009) A high-resolution computational atlas of the human hippocampus from postmortem magnetic resonance imaging at 9.4 T. *Neuroimage* 44:385–398.
- Zhang Q, Ma C, Gearing M, Wang PG, Chin L-S, Li L (2018) Integrated proteomics and network analysis identifies protein hubs and network alterations in Alzheimer's disease. *Acta Neuropathol Commun* 6:19.
- Zhang X, Li Q, Wang L, Liu ZJ, Zhong Y (2016) Cdc42-dependent forgetting regulates repetition effect in prolonging memory retention. *Cell Rep* 16:817–825.