Supplemental Methods

RNA sequencing

Further details have been previously described.[16,17] Total RNA was extracted from approximately 100 mg of tissue using the RNeasy kit (Qiagen) according to the manufacturer's protocol. The poly-A containing RNA molecules were purified from 1 µg DNAse treated total RNA and sequencing libraries were constructed using the Illumina TruSeq© RNA Sample Preparation v2 kit. Sequencing indices/barcodes were inserted into Illumina adapters allowing samples to be multiplexed across lanes in each flow cell. These products were then purified and enriched with PCR to create the final cDNA library for high throughput sequencing using an Illumina HiSeq 2000 with paired end 2x100bp reads. The Illumina Real Time Analysis (RTA) module performed image analysis, base calling, and the BCL Converter (CASAVA v1.8.2), generating FASTQ files containing the sequencing reads. These reads were aligned to the human genome (UCSC hg19 build) using the spliced-read mapper TopHat (v2.0.4) using the reference transcriptome to initially guide alignment, based on known transcripts of the previous Ensembl build GRCh37.67. Gene counts (e.g. FOLH1) were generated using the featureCounts tool based on the more recent Ensembl v75, which was the last stable release for the hg19 genome build, using single end read counting. We converted counts to RPKM values using the total number of aligned reads across the autosomal and sex chromosomes (dropping reads mapping to the mitochondria chromosome).

Selection of the FOLH1 missense mutation, rs202676

The known FOLH1 missense polymorphism, rs202676, was selected because of its putative functional variation and also due to its association with negative symptoms in patients with schizophrenia,[18] as well as its location within a sequence that is unique to FOLH1. FOLH1B is a duplication of the FOLH1 gene found

in humans, chimpanzees, and gorillas, but not other primates nor lower species like rats, making it difficult to study.[19] FOLH1B makes a protein called NAALAD2 or GCPIII, which shares 70% sequence homology to FOLH1.[20] The missense polymorphism, rs202676, was chosen because it is unique to FOLH1 alone and therefore the results are due to differences in FOLH1 (GCPII) alone.

Genotyping of the JHU Clinical Cohort

Genomic DNA was extracted using the QIAamp DNA Blood Mini Kit (Qiagen, Germantown, MD) from 200 µL of blood. Quality control of DNA concentrations was determined using the Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific, Waltham, MA). Typing of the FOLH1 SNP rs202676 involved PCR amplified from genomic DNA using the primers 5'-ACT CCT GCT CTA AAC CTC TGT AAT-3' and 5'-ATC TCG TTT ACA CCC ATT AGT TG -3', and presence of a product of the appropriate size was confirmed by visualization on agarose gel. 5 µL of the PCR reaction was cleaned using FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific, Waltham, MA) according to the manufacturer's instructions. Sanger sequencing was then performed on the cleaned product at the Genetic Resources Core Facility at Johns Hopkins School of Medicine, using the Applied Biosystems 3730xl DNA Analyzer. The clinical cohort was 33% Caucasian alone, 56% African American alone, and the remaining subjects were mixed races. The major allele in the African American population according to 1000Genomes is the G allele (0.6), and therefore it is not surprising that our clinical cohort has a higher frequency of the G allele (0.42). Hardy-Weinberg equilibrium for the clinical cohort is p=0.11.

Magnetic Resonance Spectroscopy Imaging

Magnetic resonance spectroscopy (MRS) data were collected on an 'Achieva' 7T scanner (Philips; Best, Netherlands) at the Kennedy Krieger Institute, using a 32-receive channel head coil (Nova Medical; Wilmington, DE). The STEAM pulse sequence was used to collect spectra in each subject from the bilateral

dorsal anterior cingulate cortex (ACC, 20×30×20 mm^3), left centrum semiovale white matter (CSO, 20×40×15 mm^3), left dorsolateral prefrontal cortex (dlPFC, 20×25×20 mm^3), left orbitofrontal region (OFR, 20×20×20 mm^3), and bilateral thalamus (30×20×15 mm^3). Scan parameters were TR/TE/TM= 3s/14ms/33ms, 128 acquisitions with VAPOR water suppression,[34] 2 unsuppressed water acquisitions, scan time 6 min 30s per region. Prior to data acquisition, magnetic field homogeneity was optimized up to 2nd order using the FASTMAP routine,[35] and a localized radiofrequency power optimization routine was used.[36] In addition to the MRS measurements, we collected volumetric high-resolution T_1 and T_2-weighted MR images. After an initial survey acquisition (~30s) and transmit and receive B_1 maps ('reference scans', ~2 min), a sagittal 3D 0.8mm resolution MPRAGE scan was recorded for voxel placement (also reformatted into axial and coronal views). After MRS data collection, a high resolution MP2RAGE scan (0.6 mm isotropic) was recorded for segmentation purposes (partial volume corrections (i.e. % gray matter, white matter, and cerebrospinal fluid). High resolution multi-slice 2D T_2-weighted sequence were recorded.

All spectra were analyzed in the 'LCModel' software package (Version 6.3-0D)[37] using a basis set consisting of 19 different metabolites, including NAA, NAAG, glutamate (Glu), glutamine (Gln).[33] Spectra were fit between 0.2 and 4.0 ppm, using baseline correction and macromolecule peaks as described previously[33] NAAG levels were included in further statistical analyses only when their Cramer-Rao Lower Bounds (CRLB) were below 30%. CSF-corrected NAAG levels were expressed relative to the unsuppressed voxel water as 'institutional units' ('i.u.', approximately mM), i.e. for each metabolite: [M]corr = [M] * [100/(100 – %CSF)]. Using these acquisition and analysis methods, it has been previously shown that 7T MRS can reproducibly estimate brain NAAG concentrations with coefficients of variation within the range of 11-15%.[38]

Functional MRI replication dataset

A replication functional MRI dataset was collected at Johns Hopkins School of Medicine. All subjects provided written informed consent approved by the Johns Hopkins Institutional Review Board prior to study participation. Subjects were healthy, non-smoking, Caucasian, right-handed men 18 to 35 years old. Subjects were screened for current and past psychiatric illness using the Structural Clinical Interview for Diagnosis (SCID) of DSM-IV Disorders.[17] Subjects completed a history and physical examination, biochemical and hematological laboratory screens, MRI safety screen, and screened for drugs of abuse including alcohol. Subjects were nonsmoking, as confirmed with urine nicotine and cotinine screening. Subjects were excluded for past or current psychiatric or neurological illness, uncontrolled medical disorder, hypotension or uncontrolled hypertension, or head trauma with loss of consciousness or evidence of functional impairment after head trauma. Subjects did not take psychotropic medications within 3 months of the study or any enzyme-inducing or inhibiting agent within 1 month of the study. All subjects were right-handed as measured by the Edinburgh Inventory.[18] A blood sample was collected and processed using standardized methods to extract DNA. Genotyping was done on the Illumina Global Screening Array v1.0, and rs202676 was imputed using this data. Twenty-three subjects completed this fMRI study, which included 12 AA, 9 GA carriers, and 2 GG carriers of 202676.

Subjects were scanned on a Siemens 3T Trio MRI that was upgraded halfway through the study to the Siemens 3T Prisma MRI. All scanning parameters remained the same between the Trio and Prisma acquisitions, which were also consistent with the archival dataset presented in the main methods section. A structural MRI (T1-weighted 3D MPRAGE) was completed on each subject and reviewed by a neuroradiologist to ensure there were no structural abnormalities. Subjects completed the N-back working memory task, as described in the methods section in the main text. We acquired single-shot GRE echoplanar BOLD images using the Siemens 3D PACE MOCO (Prospective Acquisition CorrEction)

sequence.[22] Functional MR images were processed and analyzed using Statistical Parametric Mapping (SPM12), using the same process as the NIMH dataset. As was done in the analysis of the archival dataset in the main text, in the first-level analysis, linear contrasts were computed producing t-statistical parameter maps at each voxel (i.e. 2-back condition greater than 0-back condition). These statistical images were entered in a second-level regression comparing G carriers to AA carriers of rs202676. For this replication dataset, the statistical threshold of $p<0.01$ uncorrected with a voxel extent of $k=10$ was used because of the small sample size.

Supplemental Tables

Supplemental Table S1. FOLH1 rs202676 allele frequencies for the African American and Caucasian cohorts for the post-mortem brain samples.

Genotype	**AA cohort**	**CAUC cohort**	Total subjects
AA	63	120	183
AG	89	76	165
GG	46	12	58
Total subjects	198	208	406

*African American (AA), Caucasian (CAUC)

Supplemental Table S2. Demographics and FOLH1 rs202676 allele frequencies of the clinical subjects.

	AA	GA	GG	All
	n=45	n=51	n=26	n=122
Healthy, Psychosis	25, 20	29, 22	12, 14	65, 57
Age, mean (SD)	22.6 (3.05)	24.0 (4.85)	24.2 (4.27)	23.5 (4.20)
Female sex, No. (%)	15 (33%)	23 (45%)	14 (54%)	52 (43%)
IQ, mean (SD)	106.55 (12.05)	101.36 (9.40)	100.42 (8.98)	103.06 (10.71)

Supplemental Table S3. Demographics of the functional neuroimaging cohort

	AA	GA	GG	All
	n= 112	n= 62	n= 9	n= 183
Age, mean (SD)	27.79 (7.59)	27.90 (7.76)	25.11 (3.95)	27.70 (7.51)
Female sex, No. (%)	65 (58%)	31 (50%)	3 (33%)	99 (54%)
IQ, mean (SD)	110.13 (7.45)	110.56 (10.45)	107.67 (6.73)	110.16 (8.53)
Performance, mean (SD)	92.37 (5.96)	90.79 (5.85)	94.44 (5.80)	91.94 (5.95)

Supplemental Table S4. Linear regression of CSO NAAG levels with measures of cognition.

	IQ	Processing speed	Attention/working memory	Verbal memory	Visual Memory	Ideational fluency	Executive function	Composite score
Slope	7.35	9.87	4.86	6.95	13.65	1.34	6.79	5.25
R square	0.0345	0.0312	0.00882	0.0198	0.0716	0.000723	0.0212	0.0177
F	4.17	3.76	1.03	2.37	9.029	0.0839	2.53	2.08
P value	0.043	0.0548	0.312	0.127	0.0033	0.773	0.114	0.152

Supplemental Table S5. Linear regression of CSO NAA levels with measures of cognition.

	IQ	Processing speed	Attention/working memory	Verbal memory	Visual Memory	Ideational fluency	Executive function	Composite score
Slope	0.431	6.12	6.24	-4.02	3.60	2.91	1.25	3.62
R square	0.000257	0.0253	0.0324	0.0144	0.0108	0.00782	0.00155	0.0197
F	0.0303	3.06	3.92	1.72	1.29	0.923	0.183	2.33
P value	0.862	0.0828	0.0501	0.193	0.259	0.339	0.669	0.129

Supplemental Table S6. Linear regression of the ratio of NAAG/NAA in the CSO with measures of cognition.

	IQ	Processing speed	Attention/working memory	Verbal memory	Visual Memory	Ideational fluency	Executive function	Composite score
Slope	47.3	35.0	4.50	66.1	80.0	-7.92	41.0	18.3
R square	0.0258	0.007077	0.000137	0.0325	0.0446	0.000448	0.0140	0.00379
F	3.10	0.834	0.0158	3.93	5.46	0.0520	1.66	0.438
P value	0.0808	0.363	0.900	0.0499	0.0212	0.820	0.201	0.510

Supplemental Table S7. Linear regression of dlPFC NAAG levels with measures of cognition.

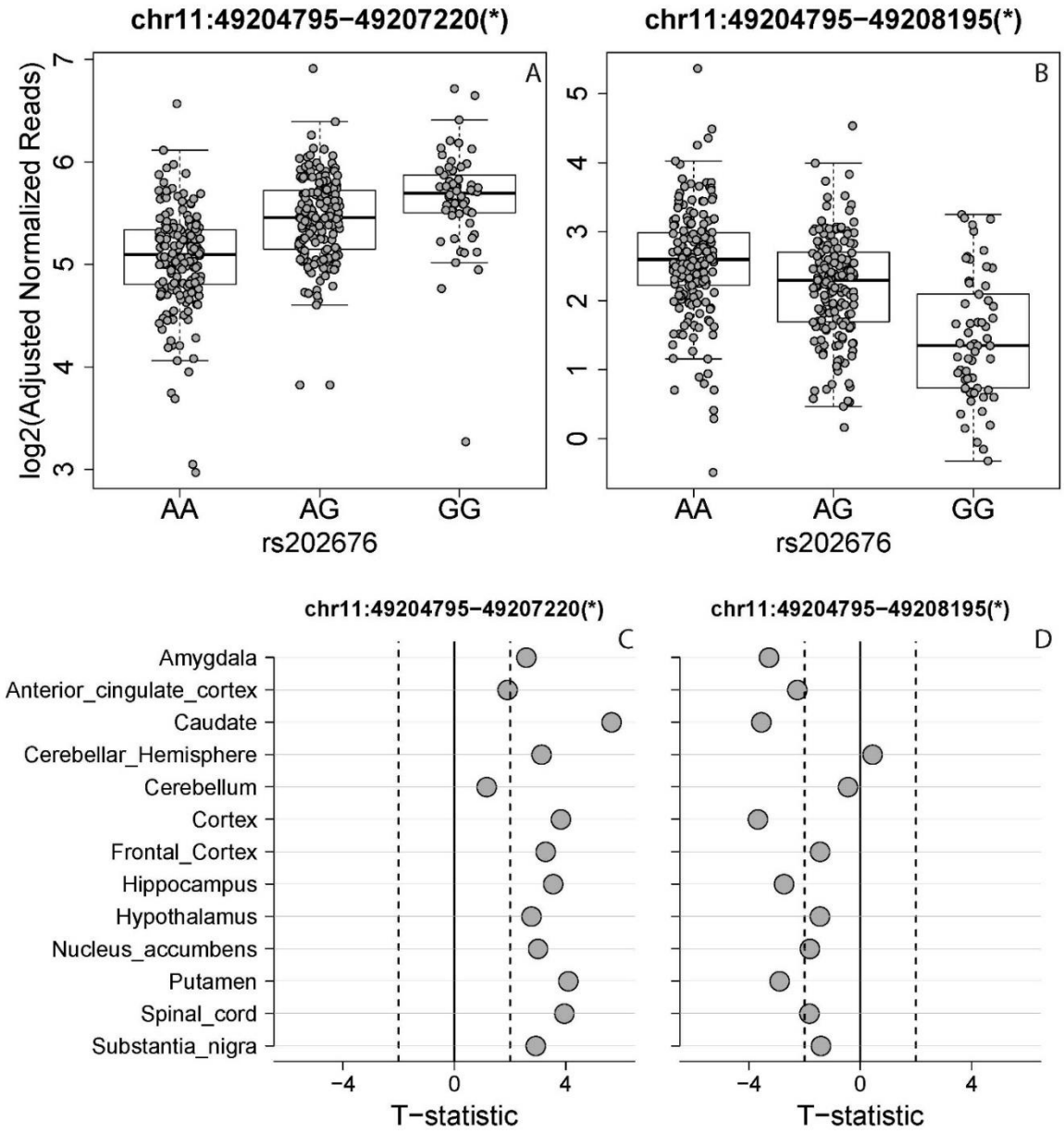
	IQ	Processing speed	Attention/working memory	Verbal memory	Visual Memory	Ideational fluency	Executive function	Composite score
Slope	-0.291	2.07	4.91	9.1	13.0	2.39	2.73	9.71
R square	5.4e-5	0.00148	0.00887	0.0339	0.0635	0.00231	0.00301	0.0585
F	0.00566	0.1416	0.931	3.69	7.11	0.241	0.317	6.40
P value	0.940	0.708	0.337	0.0576	0.0089	0.625	0.575	0.0129

Supplemental Table S8. Carriers of the G allele of FOLH1 rs202676 have higher cortical activity during working memory (2>0 back) compared to AA carriers in these regions in the JHU replication cohort.

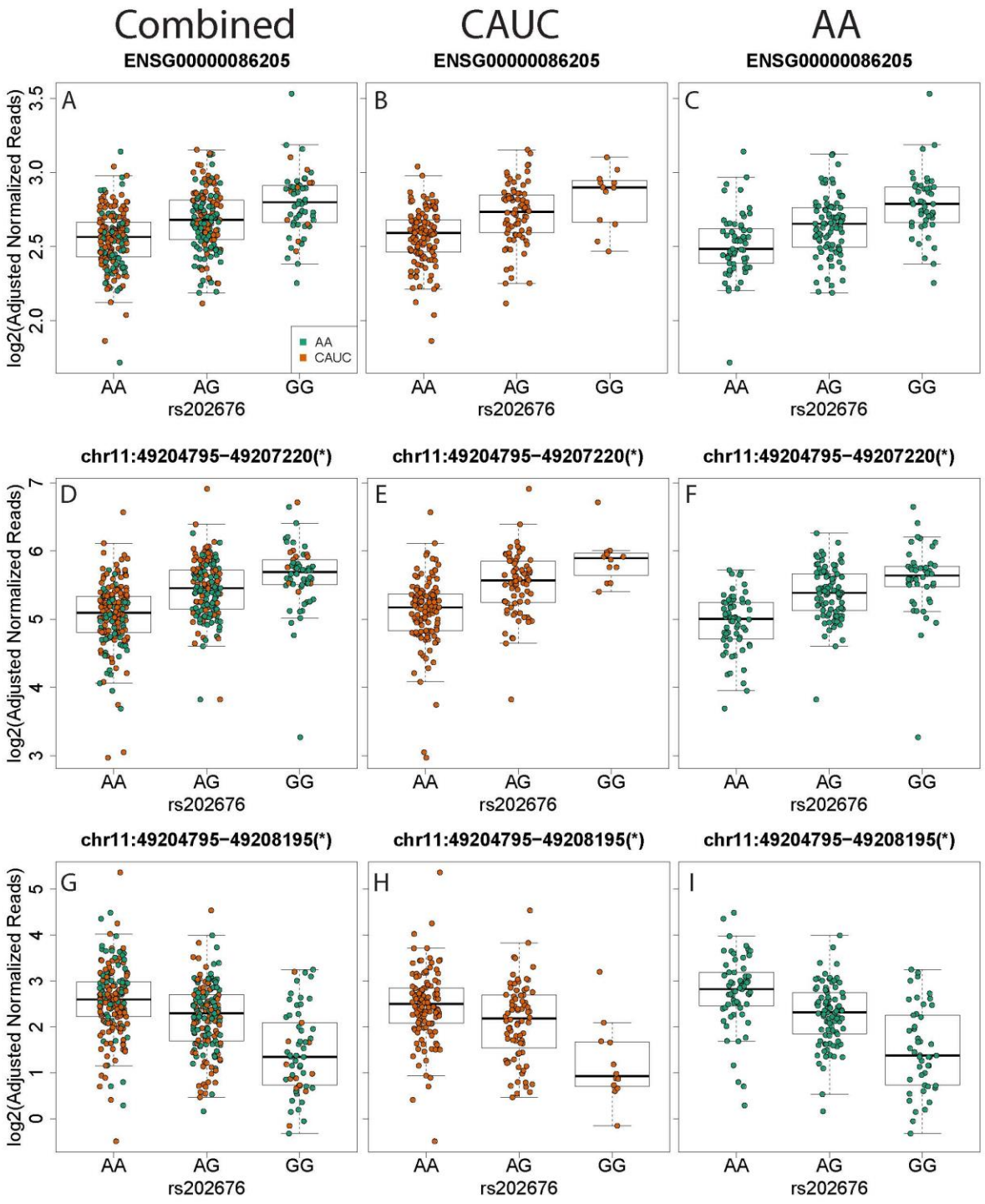
Region	Coordinates	Voxels	T (peak)	P (peak)
Inferior parietal cortex	-51, -64, 41	29	3.69	0.001
Superior temporal gyrus	-60, -10, -4	18	3.62	0.001
Cuneus	-9, -73, 23	19	3.44	0.001
Inferior parietal cortex	57, -58, 32	27	3.44	0.001
Middle temporal gyrus	-63, -28, -7	13	3.32	0.002
Superior frontal gyrus	24, -4, 71	22	3.22	0.002
Anterior cingulate gyrus	3, 26, -10	18	2.99	0.004
Posterior cingulate gyrus	6, -37, 35	10	2.94	0.004
Precuneus	6, -67, 26	10	2.87	0.005

Supplemental Figures

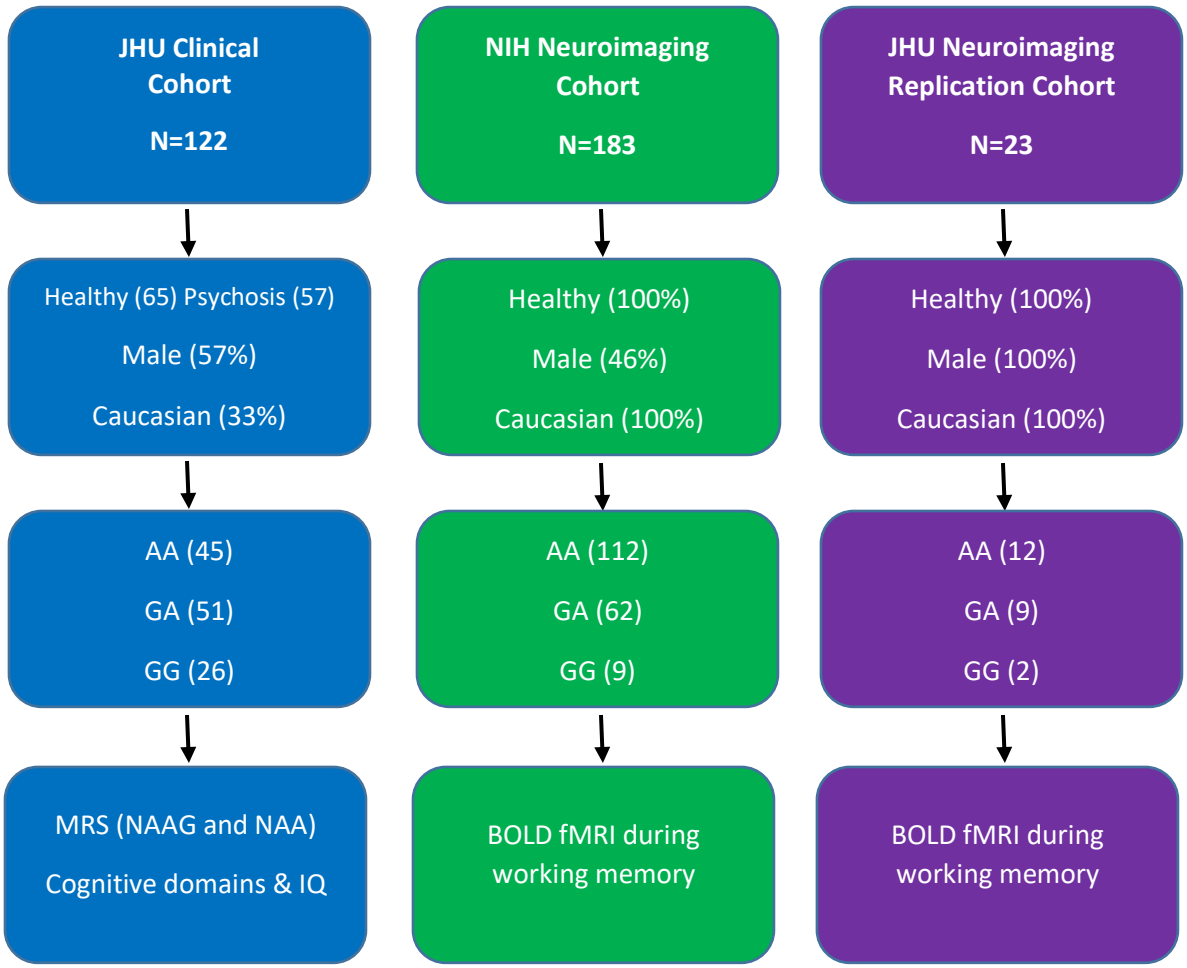
Supplemental Figure S1. A feature-level analysis focusing on the association between FOLH1rs202676 genotype and the the exon-exon splice junction (chr11:49204795-49207220) corresponding to the major protein-coding isoform (ENST00000256999.6, which encodes the 750 amino acid GCPII protein, figure S1A), and a splice junction in a transcript encoding a shorter protein (ENST00000525826, figure S1B), which is replicated in many of the brain regions in GTEx (figures S1C, S1D).



Supplemental Figure S2. FOLH1 mRNA expression by rs202676 genotype, separated by racial cohorts (gene level: panels A-C; junction level: panels D-I).



Supplemental Figure S3. Flowchart of study cohort composition and measurements.



Supplemental Figure S4. Carriers of the G allele of FOLH1 rs202676 have higher cortical activity during working memory (2>0 back) compared to AA carriers in the JHU replication cohort (thresholded at p<0.05 uncorrected for visualization purposes).