Open

ORIGINAL ARTICLE

A genome-wide survey and functional brain imaging study identify *CTNNBL1* as a memory-related gene

A Papassotiropoulos[1,2,3], E Stefanova[4], C Vogler[1], L Gschwind[1], S Ackermann[1], K Spalek[5], B Rasch[1,5], A Heck[1], A Aerni[2,5], E Hanser[1,3], P Demougin[1,3], K-D Huynh[1,3], R Luechinger[6], M Klarhöfer[7], I Novakovic[8], V Kostic[4], P Boesiger[6], K Scheffler[9,10] and DJ-F de Quervain[2,5]

[1]Division of Molecular Neuroscience, Department of Psychology, University of Basel, Basel, Switzerland; [2]University of Basel, Psychiatric University Clinics, Basel, Switzerland; [3]Life Sciences Training Facility, Department Biozentrum, University of Basel, Basel, Switzerland; [4]Institute of Neurology, CCS, School of Medicine, University of Belgrade, Belgrade, Serbia; [5]Division of Cognitive Neuroscience, Department of Psychology, University of Basel, Basel, Switzerland; [6]Institute for Biomedical Engineering, University and ETH Zürich, Zürich, Switzerland; [7]Division of Radiological Physics, University Hospital Basel, University of Basel, Basel, Switzerland; [8]Institute of Biology, School of Medicine, University of Belgrade, Belgrade, Serbia; [9]MRC Department, MPI for Biological Cybernetics, Tübingen, Germany and [10]Department of Neuroimaging and MR-Physics, University of Tübingen, Tübingen, Germany

Unbiased genome-wide screens combined with imaging data on brain function may identify novel molecular pathways related to human cognition. Here we performed a dense genome-wide screen to identify episodic memory-related gene variants. A genomic locus encoding the brain-expressed beta-catenin-like protein 1 (*CTNNBL1*) was significantly ($P = 7 \times 10^{-8}$) associated with verbal memory performance in a cognitively healthy cohort from Switzerland ($n = 1073$) and was replicated in a second cohort from Serbia ($n = 524$; $P = 0.003$). Gene expression studies showed *CTNNBL1* genotype-dependent differences in beta-catenin-like protein 1 mRNA levels in the human cortex. Functional magnetic resonance imaging in 322 subjects detected *CTNNBL1* genotype-dependent differences in memory-related brain activations. Converging evidence from independent experiments and different methodological approaches suggests a role for *CTNNBL1* in human memory.
Molecular Psychiatry (2013) **18,** 255–263; doi:10.1038/mp.2011.148; published online 22 November 2011

Introduction

Searching for molecules related to human cognition is instrumental for understanding the biological mechanisms related to such complex traits as memory capacity and for identifying pathways possibly amenable to pharmacological interventions. Recent advances in the development of high-density genotyping platforms allow for high-resolution genome-wide association studies (GWAS) of genetically complex traits and have already led to a substantial increase in knowledge of the genetic underpinnings of physiological and pathological conditions of human cognition.[1] Such studies can be performed using both pooled and individual DNA samples.[2]

Pooled GWAS followed by individual genotyping of the most significant variants of the DNA pools are a cost-effective way to perform genome-wide surveys in large cohorts. Although pooled GWAS may be a worthwhile and fast approach as a preliminary screen, they lack the ability of retrospectively stratifying the genotyped pooled cohort by secondary phenotypic traits and control variables.[2]

Here we performed a GWAS in individual DNA samples in a homogenous cohort of Swiss healthy young adults ($n = 1198$) assessed for verbal episodic memory performance, as quantified by an unexpected delayed free-recall test of 30 previously learned words (see Materials and methods). Single-nucleotide polymorphisms (SNPs) surpassing genome-wide correction for multiple comparisons were analyzed in an independent sample of healthy young adults from Serbia ($n = 524$), who were also characterized for verbal episodic memory performance. To further validate the genetic findings of the behavioral studies we analyzed data from studies on gene expression in human post-mortem brain tissue and also used functional magnetic resonance imaging (fMRI), which can detect genotype-dependent differences in brain

activations during cognitive tasks.[3] We show that variants of *CTNNBL1* (encoding beta-catenin-like protein 1) are consistently associated with episodic memory performance, with mRNA expression levels of *CTNNBL1* in the human cortex, and with activation in episodic memory-related brain regions.

Materials and methods

Memory testing

Zurich sample (GWAS subsample 1): We recruited 709 healthy, young Swiss university students or age-matched employees/trainees (521 females, 188 males). Age was 21.9 ± 0.1 years (mean $\pm$ s.e.). After complete description of the study to the subjects, written informed consent was obtained. The ethics committee of the Canton of Zurich, Switzerland, approved the study protocol. Subjects viewed six series of five semantically unrelated nouns presented at a rate of one word per second with the instruction to learn the words for immediate free recall after each series. In addition, subjects underwent an unexpected delayed free-recall test of the learned words after 5 min (episodic memory). The number of correctly recalled words (hits) was the relevant output. Attention and concentration were assessed with the d2 cancellation test[4] and working memory with the digit span task (forward)[5] in a subset of the Zurich sample. In the d2 test, items consist of the letters *d* and *p* with one to four dashes, arranged either individually or in pairs above and below the letter. Subjects were instructed to scan, in a timed manner, across each letter line to identify and delete each *d* with two dashes. In the digit span test (forward), increasing numbers of orally presented digits at a rate of one per second had to be repeated in the same order. The test started with three digits. In every other trial the number of digits was increased by one. When errors in two consecutive trials were made, the test was ended and the number of correctly recalled trials was counted.

In addition to the behavioral tasks, all subjects were tested with a standardized mental health questionnaire for current and lifetime major psychiatric diseases, smoking and the use of any medication. Subjects also provided information on possible familiar aggregation of major psychiatric diseases, and female participants also gave information on the use of oral contraceptives and the actual menstrual cycle. None of these control variables affected the results presented herein. Blood was drawn by using 2×9 ml EDTA tubes (Sarstedt, Germany). Saliva was collected with the Oragene DNA sample collection kit (DNA Genotek, ON, Canada). DNA isolation was done according to standard protocols.

Basel sample (GWAS subsample 2): We recruited 489 healthy, young Swiss university students or age-matched employees/trainees (324 females, 165 males). Age was 22.5 ± 0.1 years. After complete description of the study to the subjects, written informed consent was obtained. The ethics committee

of the Canton of Basel, Switzerland, approved the study protocol. Subjects viewed six series of five semantically unrelated nouns presented at a rate of one word per second with the instruction to learn the words for immediate free recall after each series. The number of correctly recalled words (hits) after 5 min was the relevant output. All subjects were tested with a standardized mental health questionnaire for current and lifetime major psychiatric diseases, smoking and the use of any medication. In addition, subjects provided information on possible familiar aggregation of major psychiatric diseases, and female participants also gave information on the use of oral contraceptives and the actual menstrual cycle. In addition to the word-based episodic memory task, 872 subjects from GWAS samples 1 and 2 performed a picture-based episodic memory task (see this task's description in the fMRI experiment section). Saliva was collected with the Oragene DNA sample collection kit (DNA Genotek). DNA isolation was done according to standard protocols.

Serbian sample (replication sample): Study participants were 524 healthy young subjects (353 female and 171 male students of the School of Medicine of the University of Belgrade, age 20.3 ± 0.04 years) who agreed to participate in a study of the genetic basis of human cognition and gave informed consent. Free recall of a short story narrative (immediate and 30 min delayed recall of items from Story A of the Logical Memory test; Wechsler Memory Scale-revised)[5] was used to assess verbal memory performance. Working memory performance was assessed with the Digit Ordering Task,[6] verbal fluency with a category (fruits and vegetables) verbal fluency task[7] and visuospatial abilities with the Hooper Visual Organization Test.[8] Subjects also completed a questionnaire related to general health, living circumstances, parental education and subjective learning capacity.

GWAS statistics, population structure and imputation

GWAS and the replication study were run under the assumption of an additive model. False discovery rate (FDR) and correction for 582 514 comparisons were used to adjust for genome-wide multiple testing. The FDR and corrected significance levels were set to 5%. Golden Helix SNP and Variation Suite 7 (SVS7, version 7.3.1; Golden Helix, Bozeman, MT, USA) and SPSS 17 were used for statistical analyses and visualization of results.

Population stratification was assessed with EIGEN-STRAT[9] by analyzing all genome-wide, array-based autosomal SNPs passing quality control (QC) criteria. Principal component analysis was first applied to reduce genetic variation to a few dimensions. For principal component analysis, default parameters were used (that is, definition of ten principal components in five iterations, outlier criterion was 6 s.d.). We also applied the Genomic Control program that is implemented in the EIGENSTRAT package to compute the inflation factor λ[10] before and after removal of the individuals identified as outliers. In

the total GWAS sample, λ was low ($\lambda = 1.017$), while EIGENSTRAT identified 125 participants deviating from a large genetically homogenous population cluster. After outlier removal, λ indicated the absence of population stratification in both subsamples ($\lambda_{Zurich} = 1.001$, $\lambda_{Basel} = 1.000$; Supplementary Figure 1).

All SNPs in the *CTNNBL1* region reported in the latest release of the 1000 Genomes project (www.1000genomes.org) and in the CEU sample in HapMap Phase II were imputed and analyzed with IMPUTE,[11] which aligns SNPs between haplotype and genotype databases on the basis of base-pair position. The quality criterion for inclusion and subsequent analysis of an imputed SNP was set to proper_info $\geqslant 0.5$.[11]

Array-based SNP genotyping, SNP filtering
Samples were processed as described in the Genome-Wide Human SNP Nsp/Sty 6.0. User Guide (Affymetrix, Santa Clara, CA, USA). Briefly, genomic DNA concentration was determined by using a Nano-Drop ND-1000 and adjusted to $50 \, ng \, \mu l^{-1}$ in water. A quantity of $250 \, ng$ of DNA was digested in parallel with 10 units of *Sty*I and *Nsp*I restriction enzymes (New England Biolabs, Beverly, MA, USA) for 2 h at 37 °C. Enzyme-specific adaptor oligonucleotides were then ligated onto the digested ends with T4 DNA ligase for 3 h at 16 °C. After adjustment to $100 \, \mu l$ with water, $10 \, \mu l$ of the diluted ligation reactions were subjected to PCR. Three PCR reactions of $100 \, \mu l$ were performed for *Sty*-digested products and four PCR reactions for Nsp. PCR was performed with Titanium Taq DNA Polymerase (Clontech, Mountain View, CA, USA) in the presence of $4.5 \, \mu M$ PCR primer 002 (Affymetrix), $350 \, \mu M$ each dNTP (Clontech), $1 \, M$ G-C Melt (Clontech), and $1 \times$ Titanium Taq PCR buffer (Clontech). Cycling parameters were as follows: initial denaturation at 94 °C for 3 min, amplification at 94 °C for 30 s, 60 °C for 45 s and extension at 68 °C for 15 s repeated a total of 30 times, final extension at 68 °C for 7 min. Reactions were then verified to migrate at an average size between 200 and 1100 bp using 2% Tris-borate-EDTA (TBE) gel electrophoresis. PCR products were combined and purified with the Filter Bottom Plate (Seahorse Bioscience, North Billerica, MA, USA) using Agencourt Magnetic Beads (Beckman Coulter, Fullerton, CA, USA). Purified PCR products were quantified on a Zenith 200rt microplate reader (Anthos-Labtec, Cambridge, UK). A quantity of $4–5 \, \mu g \, \mu l^{-1}$ were obtained on average for each sample. From this stage on, the SNP Nsp/Sty 5.0/6.0 Assay Kit (Affymetrix) was used. Around $250 \, \mu g$ of purified PCR products were fragmented using 0.5 units of DNAse I at 37 °C for 35 min. Fragmentation of the products to an average size less than 180 bp was verified using 4% TBE gel electrophoresis. Following fragmentation, the DNA was end labeled with 105 units of terminal deoxynucleotidyl transferase at 37 °C for 4 h. The labeled DNA was then hybridized onto Genome-Wide Human SNP 6.0 Array at 50 °C for 18 h at 60 r.p.m. The hybridized array was washed, stained and scanned

according to the manufacturer's (Affymetrix) instructions using Affymetrix GeneChip Command Console (version 3.0.1.1214). Generation of SNP calls and Array QC were performed using the command line programs of the Affymetrix Power Tools package (version: apt-1.12.0). According to the manufacturer's recommendation, contrast QC was chosen as QC metric, using the default value of greater or equal than 0.4. Mean call rate for all samples averaged 98.5%. All samples passing QC criteria were subsequently genotyped using the Birdseed (v2) algorithm. A total of 909 600 SNPs were genotyped and imported into Golden Helix SNP and Variation Suite 7 (version 7.3.0). For association testing, 19 162 markers with fewer than two alleles, 38 218 markers with call rate < 0.95, 220 877 markers with minor allele frequency < 0.05 and 113 918 markers with Hardy–Weinberg equilibrium $P < 0.05$ were skipped leaving a total of 582 514 markers to be analyzed.

Singleplex SNP genotyping
Singleplex genotyping for rs16986890 was done with Pyrosequencing on a PyroMarkID System (QIAGEN, Hilden, Germany). Forward primer: 5′-TCCTGGATCT TGGCCTTTACAT-3′; reverse primer (5′-biotinylated): 5′-GTGCAACCAACATATTTGCATTT-3′; sequencing primer: 5′-TCTATTGTATCTGAGGGTTC-3′. The chip-based SNP_A-1909171 (rs16986890) was also genotyped by Pyrosequencing in the Zurich sample. The level of convergence between array- and pyro-sequencing-based genotype calls was 100%.

Cortical gene expression
Data are based on the survey of genetic human cortical gene expression published by Myers *et al.*[12] Gene expression studies of 178 samples from the cerebral cortex of neurpathologically normal brains were done with the Illunima (San Diego, CA, USA) HumanRefseq-8 Expression BeadChip. For genome-wide genotyping, the Affymetrix GeneChip Human Mapping 500 K Array Set was used. The complete data files were downloaded from http://labs.med.-miami.edu/myers/. *CTNNBL1* transcript probe was GI_29570786 (NM_030877.3), and expression levels of GI_29570786 were used as dependent variable.

fMRI experiment
Participants. A total of 322 healthy young subjects (mean age 22.5 ± 0.2 years, range 18–34 years; 193 females; 290 *AA* genotype carriers of SNP rs16986890; 29 *AG* genotype carriers of SNP rs16986890, 3 *GG* genotype carriers of SNP rs16986890) were included in the study. All subjects were right-handed, free of any lifetime neurological or psychiatric illness, and did not take any medication at the time of the experiment. The experiments were approved by the ethics committee of the University of Basel. Written informed consent was obtained from all subjects prior to participation.

Picture task. Stimuli consisted of 72 pictures that were selected from the International Affective

258

Picture System (IAPS)[13] as well as from in-house standardized picture sets that allowed us to equate the pictures for visual complexity and content (for example, human presence). On the basis of normative valence scores (from 1 to 9), pictures were assigned to emotionally negative (2.3 ± 0.6), emotionally neutral (5.0 ± 0.3) and emotionally positive (7.6 ± 0.4) conditions, resulting in 24 pictures for each emotional valence. Four additional pictures showing neutral objects were used to control for primacy and recency effects in memory. Two of these pictures were presented in the beginning and two at the end of the picture task. They were not included in the analysis. In addition, 24 scrambled pictures were used. The background of the scrambled pictures contained the color information of all pictures used in the experiment (except primacy and recency pictures), overlaid with a crystal and distortion filter (Adobe Photoshop CS3, Adobe Systems Inc., San Jose, CA, USA). In the foreground, a mostly transparent geometrical object (rectangle or ellipse of different sizes and orientations) was shown. Pictures were presented in the scanner using MR-compatible LCD goggles (Visuastim XGA, Resonance Technology, Los Angeles, CA, USA). Eye correction was used when necessary.

Encoding phase. The pictures were presented for 2.5 s in a quasi-randomized order so that at maximum four pictures of the same category occurred consecutively. A fixation-cross appeared on the screen for 500 ms before each picture presentation. Trials were separated by a variable intertrial period of 9–12 s (jitter) that was equally distributed for each stimulus category. During the intertrial period, participants subjectively rated the picture showing scenes according to valence (negative, neutral, positive) and arousal (large, medium, small) on a three-point scale (Self-Assessment Manikin) by pressing a button with a finger of their dominant hand. For scrambled pictures, participants rated form (vertical, symmetric or horizontal) and size (large, medium, small) of the geometrical object in the foreground. The encoding phase of the picture had a total duration of 22 min. Participants were not told that they had to remember the pictures for later recall. Participants were instructed and trained on the picture task before being positioned in the scanner. Training consisted of presentation and rating of five pictures including scenes and scrambled pictures, which were not used during scanning. *Free recall*: 10 min after picture presentation, memory performance was tested using a free-recall task, which required participants to write down a short description of the previously seen pictures. Remembered primacy and recency pictures as well as training pictures were excluded from the analysis. No time limit was set for this task. Two trained investigators independently rated the descriptions for recall success (inter-rater reliability >99%).

Procedure. After receiving general information about the study and giving their informed consent,

participants were instructed and then trained on the picture task they later performed in the scanner. After training, they were positioned in the scanner. The participants received earplugs and headphones to reduce scanner noise. Their head was fixated in the coil using small cushions, and they were told not to move their heads. Functional MR images were acquired during the performance of the picture task in two separate sessions (total scanning time $\sim$30 min). After finishing the tasks, participants left the scanner and were taken to a separate room for free recall of the pictures. Finally, participants filled out questionnaires, gave saliva for genotype analysis and were debriefed. The total length of the experimental procedure was $\sim$3 h. Participants received 25 Swiss Francs (CHF) per hour for participation.

fMRI data acquisition and processing. Measurements were performed on a Siemens (Munich, Germany) Magnetom Verio 3T whole-body MR unit equipped with a twelve-channel head coil. Functional time series were acquired with a single-shot echo-planar sequence using parallel imaging (GRAPPA). We used the following acquisition parameters: TE (echo time) = 35 ms, FOV (field of view) = 22 cm, acquisition matrix = 80×80, interpolated to 128×128, voxel size: $2.75 \times 2.75 \times 4$ mm^3, GRAPPA acceleration factor $R = 2.0$. Using a midsaggital scout image, 32 contiguous axial slices were placed along the anterior–posterior commissure plane covering the entire brain with a TR = 3000 ms ($\alpha = 82°$). The first two acquisitions were discarded due to T1 saturation effects.

Preprocessing and data analysis was performed using SPM5 (Statistical Parametric Mapping, Wellcome Department of Cognitive Neurology, London, UK; http://www.fil.ion.ucl.ac.uk/spm/) implemented in Matlab 2008a (The Mathworks, Natick, MA, USA). Volumes were slice-time corrected to the first slice, realigned to the first acquired volume, normalized into standard stereotactic space (template provided by the Montreal Neurological Institute), and smoothed using a 8-mm full-width-at-half-maximum Gaussian kernel. A 128 s cut-off high-pass filter was added to the confound partition of the design matrix to account for low-frequency drifts, and a correction for intrinsic autocorrelations was included in the analysis. For each subject, evoked hemodynamic responses to event types were modeled with a delta (stick) function corresponding to presentation of each stimulus category (negative, positive, neutral and scrambled pictures, respectively) convolved with a canonical hemodynamic response function within the context of a general linear model. The pictures accounting for possible primacy and recency effects as well as button presses during valence and arousal ratings were modeled separately. In addition, six movement parameters from spatial realigning were included as regressors of no interest. The contrast between brain activity during encoding of pictures subsequently remembered versus pictures subsequently forgotten was calculated individually

using a fixed-effects model (first-level analysis). The resulting contrast parameters were then used for genotype-dependent analyses in a random-effects model (second-level analysis). Specifically, we used a regression model to analyze gene–dose-dependent differences in brain activity (with the number of *G*-alleles as covariate) and two-sample *t*-tests to analyze genotype-dependent differences under the dominant model (*GG* vs [*AG* and *AA*]).

For genotype-independent analyses, we used a threshold of $P < 0.05$, Family-Wise Error (FWE)-corrected for multiple comparisons in the whole brain. For genotype-dependent analyses, we used an uncorrected threshold of $P < 0.001$ in a minimum number of 10 adjacent voxels ($k = 10$) for an exploratory analysis of the whole brain.

Results

Genome-wide survey and replication study
All samples ($n = 1198$) were processed on the Affymetrix Genome-Wide Human SNP Array 6.0 in one centralized microarray facility. After standard QC and correction for minor allele frequency and deviation from Hardy–Weinberg equilibrium (see Materials and methods), a total of 582 514 out of the 909 600 array SNPs were used for association analyses under an additive model. Analysis of genetic background in the GWAS sample showed that the level of population stratification was low ($\lambda = 1.017$). Principal component analysis using all available autosomal array markers passing QC identified 125 participants deviating from one large population cluster (see Materials and methods, Supplementary Figure 1). Omission of these subjects from the analysis led to an ideal population structure ($\lambda = 1.000$). Therefore, genome-wide association analyses were done with the remaining 1073 participants. *CTNNBL1* marker SNP rs16986890 reached the highest level of statistical significance for association with verbal episodic memory and survived both FDR correction and correction for 582 514 multiple comparisons ($P_{\text{nominal}} = 7.7 \times 10^{-8}$, $P_{\text{FDR}} = 0.002$, $P_{\text{corrected}} = 0.046$; Table 1; Figure 1; Supplementary Tables 1 and 4).

The effect size (ES) of SNP rs16986890 was small and thereby within the expected range of small ESs observed in genetic studies of polygenic behavioral traits (Table 1). This SNP was additionally genotyped on a different, singleplex platform to prevent false interpretations due to possible array-related genotyping errors (see Materials and methods). The level of convergence between array- and singleplex-based genotype calls was 100%. Because the GWAS population consisted of two subsamples, each recruited independently in a different Swiss city (Zurich, $n = 625$; Basel, $n = 448$), we performed a *post-hoc* analysis to rule out possible recruitment location-related bias. The additive effect on episodic memory was observed in both subsamples (Table 1) and showed no significant SNP $\times$ sample interaction ($P = 0.9$).

The significance of association rapidly decreased with increasing distance from the *CTNNBL1* locus and did not extend to nearby loci (Figure 1; Supplementary Table 1). To better characterize the profile of the association signal we used data from the 1000 Genomes project (www.1000genomes.org) and reinvestigated this region using imputation, which allowed for analysis of virtually all common SNPs in this region and offered a sevenfold increase in marker density over the 6.0 array SNPs. Imputation analysis confirmed the initially observed pattern of association with highly significant intragenic SNPs and SNPs adjacent to *CTNNBL1* and rapid decrease in significance with increasing distance from the *CTNNBL1* locus (Supplementary Figure 2).

To study possible pleiotropic effects of rs16986890 on cognition, we additionally analyzed this SNP's effects on immediate recall, working memory and attention (see Materials and methods; Supplementary Table 2). No significant associations with the outcome of these cognitive tasks were observed. These findings suggest that the *CTNNBL1* genotype–dependent differences in episodic memory processes were not biased by genotype effects on attention and short-term memory.

Quantile–quantile plotting and analysis further supported the robustness of the initial finding in

Table 1 Genotype-dependent performance in word-based episodic memory tasks[a] in the GWAS sample (Switzerland, $n = 1073$) and the replication sample (Serbia, $n = 524$)

Genotype (rs16986890)	GWAS total sample (Switzerland, n = 1073)		GWAS subsample 1 (Zurich, n = 625)		GWAS subsample 2 (Basel, n = 448)		Replication sample (Serbia, n = 524)	
	n	Mean ± s.e.m.	n	Mean ± s.e.m.	n	Mean ± s.e.m.	n	Mean ± s.e.m.
AA	961	−0.05 ± 0.03	564	−0.04 ± 0.04	397	−0.07 ± 0.04	434	−0.14 ± 0.06
AG	106	0.41 ± 0.11	59	0.39 ± 0.14	47	0.43 ± 0.17	88	0.37 ± 0.16
GG	6	1.06 ± 0.37	2	0.75 ± 0.00	4	1.21 ± 0.57	2	0.84 ± 0.70
	$P = 0.00000007$		$P = 0.0005$		$P = 0.00003$		$P = 0.003$	
	ES: $d = 0.33$; $r^2 = 0.027$		ES: $d = 0.28$; $r^2 = 0.019$		ES: $d = 0.39$; $r^2 = 0.038$		ES: $d = 0.26$; $r^2 = 0.017$	

Abbreviations: d, Cohen's d; ES, effect size; GWAS, genome-wide association study.
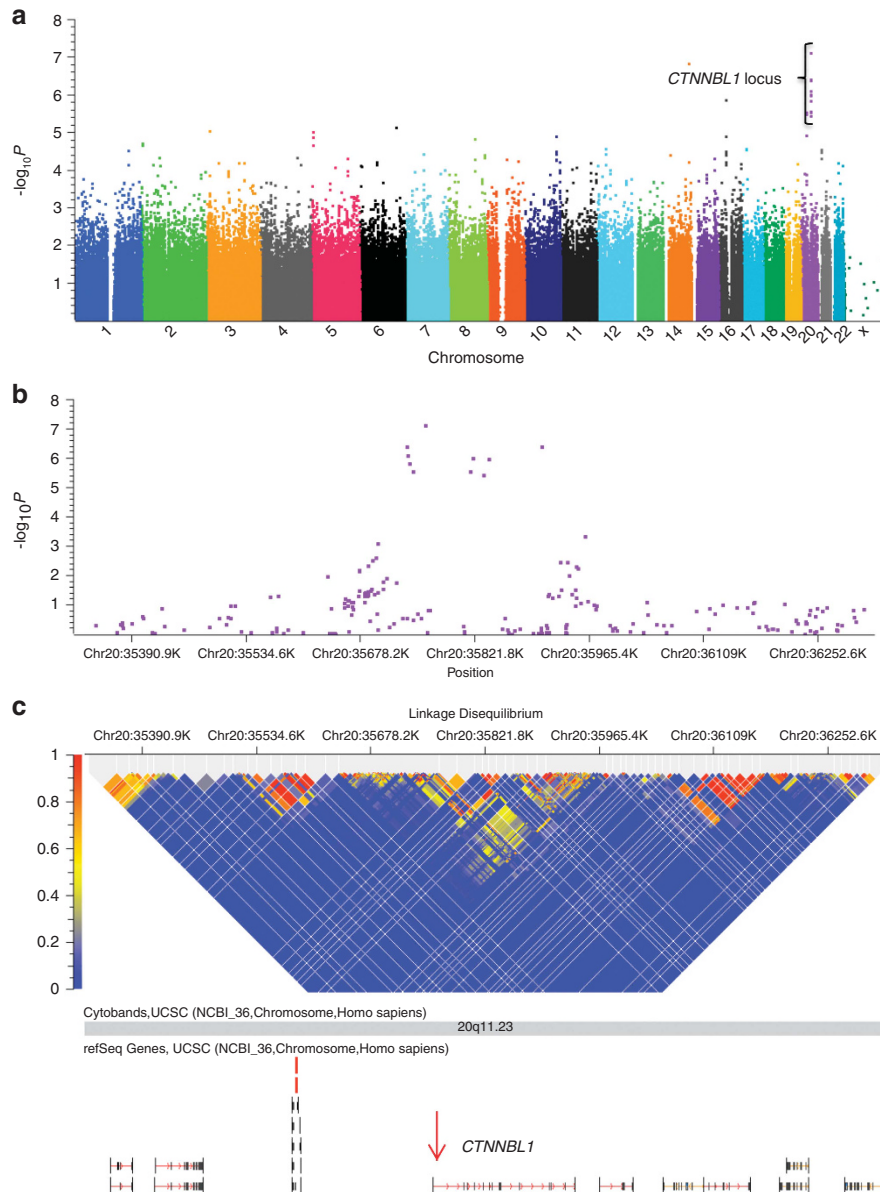[a]NOTE: Values are *z*-transformed to allow direct comparison between samples.

Figure 1 Genome-wide association results. (**a**) Genome-wide significances (Y axis, $-\log_{10}P$) are shown in chromosomal order for individually genotyped SNPs that were tested for association with word-based episodic memory performance in 1073 healthy Swiss young adults. (**b**) Zoom-in of the association results (Y axis, $-\log_{10}P$) to a region of chromosome 20q11.23 harboring the *CTNNBL1* locus. Chromosomal positions were retrieved from the March 2006 UCSC genome browser assembly. (**c**) LD structure (r^2-values) of the chromosomal region shown in panel **b** as calculated in the entire sample of 1073 healthy Swiss young adults. Known transcripts in this region are visualized in the lower part. The red arrow indicates the position of rs16986890, which is located in the first intron of *CTNNBL1*.

the GWAS sample. Over 99.999% of the *P*-values distributed in accordance with random expectation, suggesting that the GWAS results are most likely to be attributable to true genetic variation rather than to potential bias, such as genotyping errors or population stratification (Supplementary Figure 3). The significant association between *CTNNBL1* SNP rs16986890 and episodic memory performance was replicated in a sample of 524 healthy young adults from Belgrade, Serbia (*P* = 0.003; additive model and direction of effect as in the GWAS sample; Table 1). Also in this sample, rs16986890 was not associated

with performance in tasks quantifying working memory performance, verbal fluency and visuospatial abilities (all *P*-values > 0.3; see Materials and methods; Supplementary Table 3). In addition to SNP rs16986890, the nine most significant SNPs of the Swiss sample, which were not in tight linkage disequilibrium (LD) (that is, $r^2 < 0.8$) with rs16986890, were tested for replication in the Serbian sample (Supplementary Table 5). Four out of ten tested SNPs (including rs16986890) were nominally significantly associated with episodic memory also in the Serbian sample ($P_{uncorrected} < 0.05$; Supplementary Table 5).

Genetic association with cortical gene expression

SNP rs16986890 is located in the 5′ end of *CTNNBL1* exon 1 and is tightly linked to 5′ untranslated region variants (Figure 1, Supplementary Figure 1). To study the possible relation of *CTNNBL1* polymorphisms with mRNA expression levels, we examined the association between *CTNNBL1* SNPs and cortical expression of the *CTNNBL1* transcript GI_29570786 (NM_030877.3) in the brains of 178 non-demented deceased subjects (see Materials and methods). A subset of *CTNNBL1* SNPs reported in our genetic study of human memory performance is also represented on the lower marker density 500 k SNP Array set, which was used in the study of cortical gene expression. SNP rs16986890 was not analyzed due to insufficient minor allele frequency (<0.05) in this sample. We therefore analyzed SNP rs7363432, which was the closest, tightly linked SNP (expectation-maximization (EM) $r^2 = 0.89$) with minor allele frequency $\geqslant 0.05$ (154 *AA* carriers, 23 *AG* carriers, 1 *GG* carrier; Hardy–Weinberg equilibrium, $P = 0.88$). Comparison between genotype groups revealed significantly elevated levels of *CTNNBL1* mRNA in *AG* compared with *AA* carriers (312.7 ± 11.8 vs 289.2 ± 4.0, respectively; $T = 2.1$, d.f. $= 175$, $P = 0.037$; due to low number of observations ($n = 1$) *GG* genotypes were omitted from statistical analysis). Thus, in the post-mortem gene expression sample, RNA expression levels of *CTNNBL1* were increased in minor allele carriers of rs7363432. Interestingly, SNP rs7363432 was among the most significant SNPs in the Swiss GWAS sample (episodic memory, $n = 1073$, $P = 0.0000015$; Supplementary Table 1), with minor allele carriers showing better memory performance.

Genetic variability within CTNNBL1 is linked to human brain function

The above-described behavioral genetics experiments used episodic memory tests, which consisted of delayed free recall of verbal information. In 872 Swiss participants (a subsample of the above-described GWAS sample), an additional picture-based episodic memory task was performed. In this sample, *CTNNBL1* SNP rs16986890 was also associated with performance in this picture-based memory task, albeit with lower ES compared with the word-based task. Specifically, we observed a significant additive effect of rs16986890 with a positive correlation between the number of minor alleles and performance during free recall of pictures ($P = 0.027$; ES: $d = 0.15$, $r^2 = 0.006$). In 322 subjects, this picture-based episodic memory task was performed during fMRI. This allowed us to investigate the impact of *CTNNBL1* genetic variability on activity of memory-related brain regions. These regions were identified by comparing activity during encoding of subsequently remembered versus subsequently forgotten pictures (that is, subsequent memory analysis). The genotype-independent subsequent memory analysis revealed several memory-related brain regions (Supplementary Figure 4), including prefrontal and parietal regions as well as the hippo-
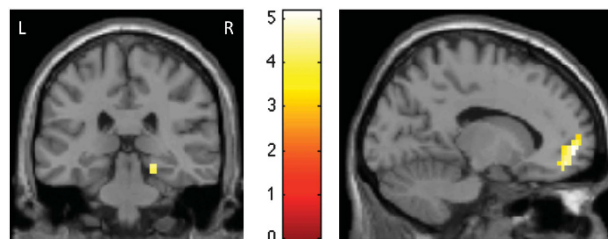

Figure 2 *CTNNBL1* rs16986890 genotype-depended differences in memory-related brain activity (subsequent memory analysis, see main text) in 322 individuals. Displayed are gene dose-dependent (with increasing number of *G*-alleles) increases in activity in the parahippocampal gyrus (left figure, coronal section, peak activation at [19 −30 −12]; $P_{\text{whole-brain FDR−corrected}} = 0.04$, $P_{\text{nominal}} < 0.001$) and in the medial frontal gyrus/anterior cingulate (right figure, sagittal section, peak activation at [−14 55 0]; $P_{\text{whole-brain FDR−corrected}} = 0.01$, $P_{\text{nominal}} < 0.001$). Activations were overlaid on sections of a T1-weighted magnetic resonance image of SPM5, displayed at an uncorrected threshold of $P = 0.001$ and using color-coded *t*-values.

campus and parahippocampal gyrus—brain regions known to have an important role in episodic memory.[14,15]

In the fMRI sample, there were 290 *AA* carriers, 29 *AG* carriers and 3 *GG* carriers of *CTNNBL1* SNP rs16986890. As expected due to the lower number of subjects in the fMRI sample ($n = 322$) than in the behavioral genetics sample ($n = 872$), we did not observe significant ($P > 0.05$) genotype-dependent differences on the behavioral level. Genotype-dependent, subsequent memory analyses of the fMRI data revealed significant, whole-brain FDR-corrected, gene dose-dependent (with increasing number of *G*-alleles) increases in activity in the parahippocampal gyrus (peak activation at [19 −30 −12]; $P_{\text{whole-brain FDR−corrected}} = 0.04$, $P_{\text{nominal}} < 0.001$) and in the medial frontal gyrus/anterior cingulate (peak activations at [−14 55 0]; [−30 47 −12]; [22 44 −4]; $P_{\text{whole-brain FDR−corrected}} \leqslant 0.03$, $P_{\text{nominal}} < 0.001$) (Figure 2). There were no other significantly increased brain activations in this analysis and there were no supra-threshold activations when calculating the negative correlation (with decreasing number of *G*-alleles). Reanalysis under the dominant model (*GG* vs [*AG* and *AA*]) revealed similar findings.

Discussion

Here we present evidence suggesting that the gene encoding beta-catenin-like protein 1 (*CTNNBL1*) has a significant role in human episodic memory. Specifically, marker SNPs in the *CTNNBL1* locus were associated with episodic memory performance after correction for multiple comparisons in a homogenous and carefully phenotyped sample of Swiss healthy young adults, and were replicated in an independent sample of healthy young adults recruited in Belgrade, Serbia. Furthermore, gene expression analysis in human post-mortem brain tissue showed that a

262

CTNNBL1 marker associated with better episodic memory was also associated with higher mRNA levels of *CTNNBL1* in the human cortex. In addition, fMRI experiments during episodic memory tasks revealed *CTNNBL1* genotype-dependent differences in brain activations in regions typically involved in episodic memory processes. The significant marker SNPs are located either in intronic or in upstream, putatively regulatory regions of *CTNNBL1*, which also harbor the hypothetical *LOC100287792*.

Recent studies reporting associations of candidate genes with specific memory-related phenotypes observed medium to large ESs (for example, SNPs in *BDNF*,[16] *GRM3*,[17] *ADRA2B*[18]). The GWAS sample of 1073 participants, which was used for unbiased genome-wide screening in this study, was sufficiently powered also for the detection of small ES. Assuming a genetic marker with rather small ES ($0.01 \leqslant r^2 \leqslant 0.05$),[19] the power of our GWAS sample to detect associations at a genome-wide-corrected α-error probability ($P_{FDR} \leqslant 0.05$) was $\sim 90\%$. The power of the replication sample ($n = 524$) to detect associations with a similarly small ES at an α-error probability $\leqslant 0.05$ was $> 95\%$. The actual ESs (r^2) in this study varied between 0.017 and 0.038 (Table 1), hence they were within the expected range of small ESs observed in GWAS dealing with polygenic cognitive traits.

CTNNBL1, which was identified and cloned recently, spans over 178 kb on human chromosome 20q11.23 (UCSC Genome Browser; GRCh36/hg18 assembly) and contains 16 exons. The encoded protein (beta-catenin-like protein 1) is a widely expressed nuclear protein and is highly conserved in mammals (96% amino acid identity with the mouse ortholog). Its biological function is not well understood, however, it seems to be a pleiotropic protein involved in such important processes as apoptosis and antibody diversification, and is part of the spliceosomal complex.[20] Moreover, beta-catenin-like protein 1 shares significant structural similarity with the armadillo repeat family of proteins, including *Drosophila melanogaster* armadillo protein and its mammalian ortholog beta-catenin. Beta-catenin is a pleitropic adherens junction protein, which regulates numerous cellular processes and also functions as a component of the Wnt-signaling pathway. Beta-catenin regulates numerous neuronal functions; for example, it is a mediator of dendritic development through enhancing dendritic arborization.[21] In rodents, beta-catenin is required for memory consolidation.[22] Interestingly, the results reported herein show that *CTNNBL1* genotypes associated with better episodic memory performance were also associated with increased brain activation during encoding of information and with increased mRNA levels of *CTNNBL1* in the human brain. It is tempting to speculate that *CTNNBL1* acts on human memory via beta-catenin-related mechanisms. However, it is hitherto completely unclear, whether beta-catenin-like protein 1 shares any similar functions with beta-catenin, especially because both proteins show only structural, rather than sequence similarity.

Taken together, the present high-density GWAS followed by a replication study and supported by gene expression data and fMRI experiments suggests a role for the beta-catenin-like protein 1 in human episodic memory. These results demonstrate the usefulness of unbiased genome-wide approaches and their potential to identify important and novel molecular pathways related to human cognition.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

References

1 Green AE, Munafo MR, Deyoung CG, Fossella JA, Fan J, Gray JR. Using genetic data in cognitive neuroscience: from growing pains to genuine insights. *Nat Rev Neurosci* 2008; **9**: 710–720.

2 Pawlowski TL, Huentelman MJ. Identification of a common variant affecting human episodic memory performance using a pooled genome-wide association approach: a case study of disease gene identification. *Methods Mol Biol* 2011; **700**: 261–269.

3 Rasch B, Papassotiropoulos A, de Quervain DF. Imaging genetics of cognitive functions: focus on episodic memory. *Neuroimage* 2010; **53**: 870–877.

4 Brickenkamp R, Zillmer E. *The d2 Test of Attention*. Hogrefe & Huber: Seattle, Washington, 1998.

5 Wechsler D. *Wechsler Memory Scale-Revised Manual*. Psychological Corp: New York, NY, 1987.

6 Cooper JA, Sagar HJ, Jordan N, Harvey NS, Sullivan EV. Cognitive impairment in early, untreated Parkinson's disease and its relationship to motor disability. *Brain* 1991; **114**(Pt 5): 2095–2122.

7 Benton AL. Differential behavioural effects in frontal lobe disease. *Neuropsychologia* 1968; **6**: 53–60.

8 Lezak MD. *Neuropsychological Assessment*. 2 edn. Oxford University Press: New York, Oxford 1983.

9 Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 2006; **38**: 904–909.

10 Devlin B, Roeder K. Genomic control for association studies. *Biometrics* 1999; **55**: 997–1004.

11 Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007; **39**: 906–913.

12 Myers AJ, Gibbs JR, Webster JA, Rohrer K, Zhao A, Marlowe L *et al.* A survey of genetic human cortical gene expression. *Nat Genet* 2007; **39**: 1494–1499.

13 Lang PJ, Bradley MM, Cuthbert BN. *International Affective Pictures System (IAPS): Affective Ratings of Pictures and Instruction Manual*. University of Florida: Gainesville, FL, 2008.

14 Cabeza R, Nyberg L. Imaging cognition II: an empirical review of 275 PET and fMRI studies. *J Cogn Neurosci* 2000; **12**: 1–47.

15 Squire LR, Alvarez P. Retrograde amnesia and memory consolidation: a neurobiological perspective. *Curr Opin Neurobiol* 1995; **5**: 169–177.

16 Egan MF, Kojima M, Callicott JH, Goldberg TE, Kolachana BS, Bertolino A *et al.* The BDNF val66met polymorphism affects activity-dependent secretion of BDNF and human memory and hippocampal function. *Cell* 2003; **112**: 257–269.

17 Egan MF, Straub RE, Goldberg TE, Yakub I, Callicott JH, Hariri AR *et al.* Variation in GRM3 affects cognition, prefrontal glutamate, and risk for schizophrenia. *Proc Natl Acad Sci USA* 2004; **101**: 12604–12609.

18 de Quervain DJ, Kolassa IT, Ertl V, Onyut PL, Neuner F, Elbert T *et al.* A deletion variant of the alpha2b-adrenoceptor is related to emotional memory in Europeans and Africans. *Nat Neurosci* 2007; **10**: 1137–1139.

19 Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2 edn. Lawrence Erlbaum Associates: Mahwah, NJ, 1988.

20 Makarova OV, Makarov EM, Urlaub H, Will CL, Gentzel M, Wilm M *et al.* A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *EMBO J* 2004; **23**: 2381–2391.

21 Yu X, Malenka RC. Beta-catenin is critical for dendritic morphogenesis. *Nat Neurosci* 2003; **6**: 1169–1177.

22 Maguschak KA, Ressler KJ. Beta-catenin is required for memory consolidation. *Nat Neurosci* 2008; **11**: 1319–1326.

Supplementary Information accompanies the paper on the Molecular Psychiatry website (http://www.nature.com/mp)