SUPPLEMENTARY INFORMATION

Allelic differences between Europeans and Chinese for CREB1 SNPs and their implications in gene expression regulation, hippocampal structure and function and bipolar disorder susceptibility

Ming Li, Xiong-jian Luo, et al.

Correspondence to: Bing Su, Ph.D.; Tel.: +86-871-65120212; fax: +86-871-65193137;
Email: sub@mail.kiz.ac.cn

Members of the MooDS Bipolar Consortium

Jana Strohmaier[1], René Breuer[1], Sandra Meier[1], Thomas W. Mühleisen[2,3], Franziska A. Degenhardt[2,3], Per Hoffmann[2,3], Stefan Herms[2,3], Markus Schwarz[4], Helmut Vedder[4], Jutta Kammerer-Ciernioch[4], Andreas Reif[5], Johanna Sasse[6], Michael Bauer[6], Sandra Zwick[7], Martin Hautzinger[7], Adam Wright[8], Philip B. Mitchell[8], Janice M. Fullerton[9], Peter R. Schofield[9], Grant W. Montgomery[10], Nicholas G. Martin[10], Piotr M. Czerski[11], Joanna Hauser[11], Johannes Schumacher[3], Wolfgang Maier[12], Peter Propping[3]

[1]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, University of Heidelberg, D-68159 Mannheim, Germany

[2]Department of Genomics, Life & Brain Center, University of Bonn, D-53127 Bonn, Germany

[3]Institute of Human Genetics, University of Bonn, D-53127 Bonn, Germany

[4]Psychiatric Center Nordbaden, D-69168 Wiesloch, Germany

[5]Department of Psychiatry, University of Würzburg, D-97070 Würzburg, Germany

[6]Department of Psychiatry and Psychotherapy, University Hospital Carl Gustav Carus, D-01307 Dresden, Germany

[7]Department of Clinical and Developmental Psychology, Institute of Psychology, University of Tübingen, D-72074 Tübingen, Germany

[8]School of Psychiatry, University of New South Wales and Black Dog Institute, Sydney, New South Wales, Australia

[9]Prince of Wales Medical Institute, University of New South Wales, NSW 2031 Sydney, Australia

[10]Queensland Institute of Medical Research, Brisbane Qld 4006, Australia

[11]Department of Psychiatry, Poznan University of Medical Sciences, PL-60-572 Poznan, Poland

[12]Department of Psychiatry, University of Bonn, D-53127 Bonn, Germany

Members of the Swedish Bipolar Study Group

Lena Backlund[1], Louise Frisén[1], Catharina Lavebratt[2], Martin Schalling[2], Urban Ösby[1]

[1]Department of Clinical Neuroscience Neurogenetics Unit, Stockholm, Sweden

[2]Department of Molecular Medicine and Surgery, Karolinska Institute, Stockholm, Sweden

Replication Sample Information (see Table S2)

French sample

Patients with BD and controls were recruited as part of a large study on genetics of BD in France (Paris-Creteil, Bordeaux, Nancy) with a protocol approved by relevant IRBs and with written informed consent. Cases were of French descent for more than three generations and were all been assessed by a well-trained psychiatrist or psychologist with the DIGS[1] and the FIGS. Diagnoses were based on structured interviews supplemented by medical case notes, mood scales and a self rating questionnaire assessing dimensions. Genotyping of controls were provided by the Centre National de Génotypage (M Lathrop, Evry). Patients and controls were genotyped on the Illumina platform (HumanHap300, HumanHap550, HumanHap 610-quad).

Swedish sample

SBP Bipolar cases were recruited from St. Göran's Hospital in Stockholm, Sweden. All participants provided written informed consent to participate in a genetic study of BD, and the study was approved by the Regional Ethics Committee of Stockholm. Diagnoses were based on physician administered ADE[2] and MINI[3].

BD cases were identified from the Swedish Bipolar Quality Assurance Registry. Patient information within the registry includes disease sub-classification, psychosis, age at onset, number of manic and depressive episodes, number of hospitalizations and family history. Participants provided written informed consent to participate in a genetic study of psychiatric disease, and the study was approved by the Regional Ethics Committee of Stockholm.

Hospital Discharge Registry (HDR) Bipolar cases were identified from the Swedish Hospital Discharge Registry if they a) have at least two admissions with discharge diagnoses of BD and b) were born in Sweden or another Nordic country. The register contains a nearly complete record of all individuals hospitalized in Sweden since 1973. Diagnoses were established by an attending physician and were shown to have high sensitivity and specificity.[4] The study was approved by the Regional Ethics Committee of Stockholm. All participants provided written informed consent to participate in genetic studies of psychotic disorders and were interviewed by a research nurse about other medical conditions.

The SBP BD cases were recruited from the Stockholm County catchment area. All patients provided written informed consent to participate in a genetic study of BD, and the study was approved by the Regional Ethics Committee of Stockholm. Diagnoses were made according to the DSM-IV criteria.

Swedish control samples were obtained from the Swedish Hospital Discharge Registry on the condition they had never received discharge diagnoses of BD, schizophrenia and/or schizoaffective disorder.

German sample

Cases were ascertained from consecutive admissions to the inpatient units of the Department of Psychiatry and Psychotherapy at the University of Bonn and at the Central Institute for Mental Health in Mannheim, University of Heidelberg, as well as at other collaborating psychiatric university hospitals in Germany. DSM-IV lifetime diagnoses of BD were assigned using a consensus best-estimate procedure, based on all available information, including semi-structured interviews (AMDP), medical records, and family history. In addition, the OPCRIT system[5] was used for the detailed poly-diagnostic documentation of symptoms.

Controls were ascertained from the population-based Heinz Nixdorf Recall Study.[6] Study protocols were reviewed and approved in advance by Institutional Review Boards of the participating institutions. All subjects provided written informed consent and were genotyped using the Illumina platform.

Australian sample

Subjects were ascertained through two studies: 1) a BD pedigree sample (described in McAuley et al.[7]) and 2) a specialized Sydney Black Dog Institute BD clinic sample (described in Mitchell et al. 2009[8]). All subjects were interviewed by trained research staff using the DIGS or SCID, using best-estimate DSM-IV diagnoses derived from those instruments, medical records and FIGS. First, for the pedigree sample, only one BD subject per family was included in the case sample. Pedigrees were only included in the original genetic study if there was unilineal inheritance, and at least two BD subjects including at least one with bipolar I disorder. Subjects were ascertained through clinical presentations to the Mood Disorders Unit at the Prince of Wales Hospital in Sydney, direct referrals from Australian clinicians, and BD consumer

organizations. Second, for the clinic sample, subjects comprised consecutive subjects referred by psychiatrists or general practitioners for specialized clinical review. All patients provided written informed consent to participate in this study and the study was approved by the local ethics committee. Patients were included in the BOMA study and genotyped at the Life & Brain Centre in Bonn.

Australian controls were drawn from families participating in the Brisbane Longitudinal Twin Study, an unselected community sample recruited to take part in studies of melanoma risk factors, cognition, and other phenotypes. Subjects were not screened for any phenotype relevant to BD. The study was approved by the ethic committee and all proband gave written informed consent. All subjects were genotyped as a single project by deCODE and have been through an extensive QC process including exclusion for non-European ancestry. The sample is overwhelmingly of northern European origin, predominately from the British Isles.

Polish sample

All patients were recruited from consecutive hospital admissions and were directly interviewed with the Structured Clinical Interview for DSM-IV-TR-Axis I Disorders (Patient Edition). Information provided via medical records and interviews of family members was also used in a best estimate procedure of diagnosis on the basis of DSM-IV-TR criteria. The control samples were population-based, drawn from the same population as the patients. The ethnicity of the patients and control subjects was determined by genealogical investigation up to the grandparental generation.

Icelandic sample

The Icelandic sample consisted of 541 subjects with BD and 34,546 population controls. Patients and controls were Icelandic and were recruited throughout the country. Diagnoses were assigned according to RDC through the use of the SADS-L for 303 subjects. DSM-IV BD diagnoses were obtained through the use of the Composite International Diagnostic Interview (CIDI-Auto) for 82 subjects. In addition, there were 150 subjects with ICD-9 or ICD-10 BD diagnoses and nine subjects with DSM-III BD diagnoses.

The 34,546 controls were recruited as a part of various genetic programs at deCODE and were not screened for psychiatric disorders. Approval for the study was granted by the National Bioethics Committee of Iceland and the Icelandic Data Protection Authority and written informed consent was obtained for all participants.

British sample

The British sample consists of 1,636 cases of recurrent depression fulfilling DSM-IV or ICD-10 criteria of at least moderate severity, ascertained from three U.K. centers (London, Cardiff, Birmingham).[9] Probands from the Depression Network affected sibling pair linkage study[10,11] are cases of recurrent depression of at least moderate severity. The pharmacogenetic study, the Genome-Based Therapeutic Drugs for Depression (GENDEP) study, is comprised of subjects who have been investigated while in an episode of depression of at least moderate severity.[12,13]

All cases were interviewed with the Schedules for Clinical Assessment in Neuropsychiatry (SCAN),[14] focusing on their worst and second-worst episodes of depression in the studies and on their current episode in GENDEP, with study coordinators for all three studies trained by one investigator, giving a homogenous cohort of depression cases. Subjects were excluded if there was a history or family history of schizophrenia or bipolar disorder or if mood symptoms were related to alcohol or substance misuse.

The control subjects were 1,594 individuals contacted via the Medical Research Council general practice research framework and screened using a composite index of depressive and anxiety symptoms;[15] they were interviewed by telephone using the Past History Schedule.[16]

Munich-German sample

In total, 640 unipolar depressive inpatients were recruited for the Munich Antidepressant Response Signature (MARS) project at the Max Planck Institute of Psychiatry (MPIP) in Munich, Germany.[17,18] Briefly, patients were included in the study within 1-3 days of admission to the hospital and diagnosis was ascertained according to the Diagnostic and Statistical Manual of Mental Disorders (DSM) IV criteria. Patients fulfilling the criteria for at least a moderate depressive episode (HAM-D R 14 on the 21-item Hamilton Depression Rating Scale) were included in the analysis. All patients were of European descent. In total, 542 control subjects were matched to the patient sample for age, gender, and ethnicity from a randomly selected Munich-based community sample and underwent a strict screening procedure for the absence of psychiatric and severe somatic disease.[19] The overall inclusion rate of all contacted proband

was 50.3%. These subjects thus represent a group of individuals from the general population who have never been diagnosed as mentally ill. Age, sex, and ethnicity did not differ from the patient sample.

Functional Magnetic Resonance Imaging Analysis

Functional magnetic resonance images were obtained from healthy German participants (N=279) of European ancestry, as part of a tricentric study on the neurogenetic mechanisms of psychiatric disease.[20-22] The study was approved by the local ethics committee of the Universities of Bonn, Mannheim and Berlin. All subjects provided written informed consent to participate in the study.

Genotypes for rs2709370 and rs6785 were imputed based on the data from the 1000-Human-Genome reference samples[23] (2010 interim release in NCBI build 37 coordinates; the 1000-Human-Genome Project Consortium) and using IMPUTE2.[24] Post-imputation quality control was applied using an info score of 0.8 and a minor allele frequency of 0.01.

The allele frequencies for rs2709370 were in the Hardy-Weinberg equilibrium ($\chi^2 = 0.022$, $P = 0.88$). Genotype distributions did not differ between sites ($P = 0.77$). Sex, age, handedness and level of education, did not significantly differ between genotype groups (**Table S6**).

Blood oxygenation level-dependent (BOLD) functional magnetic resonance imaging (fMRI) was performed on a Siemens TRIO 3T scanner at each site (33 slices, slice thickness 2.4 mm + 0.6 mm gap, FOV 192 mm, TR 1.96 sec, TE 30 ms). Quality assurance (QA) measures were conducted on every measurement day at all sites according to a multicenter QA protocol revealing stable signals over time.[25] During scanning, participants completed three consecutive blocks of memory tasks (encoding, recall and recognition of face-profession pairs), as previously reported.[20-22] Image processing and statistical analyses were conducted using statistical parametric mapping methods as implemented in Statistical Parametric Mapping 8 (SPM8) and included realignment, slice timing, normalization to a standard EPI template and smoothing. The regression model at the first level consisted of two regressors (memory and control) convolved with the hemodynamic response function and six regressors describing residual motion. In the second level group, individual regionally specific effects of the contrast memory > contour were compared using a multiple regression model including the three allelic groups and age, sex and site as covariates. T-statistics for each voxel were threshold at $P < 0.05$ family wise error (FWE) corrected for multiple comparisons across the region of interest (hippocampus).

Cognitive Performance Analyses

Irish sample

This sample consisted of 88 healthy individuals recruited on the basis of responses to local media advertisements and were reported previously.[26,27] Participants were included only if they were aged 18 to 65 years and satisfied, based on clinical interview, the criteria of having no history of major mental health problems, intellectual disability, or acquired brain injury and also no self-reported history of substance misuse in the preceding 6 months. Participants were also excluded from the study if they reported having a first-degree relative with a history of psychosis. All the subject assessments were conducted in accordance with the relevant ethics committee approval from each participating site. All subjects were of Irish ancestry (*i.e.*, 4 grandparents born in Ireland), and all provided written informed consent.

Cognitive Assessment

General Cognitive Functioning (IQ). IQ was measured using selected subtests (vocabulary, similarities, block design, and matrix reasoning) from the Wechsler Adult Intelligence Scale, third edition, yielding full-scale, verbal, and performance IQs.

Episodic Memory recall. Verbal episodic memory and visual episodic memory were assessed using the logical memory and faces subtests respectively, from the Wechsler Memory Scale, third edition.

Working Memory. Verbal working memory and spatial working memory were assessed using the letter-number sequencing task from the Wechsler Memory Scale, third edition, and the spatial working memory task from the Cambridge Neuropsychological Test Automated Battery, Expedio Version.

Attention. Attention was assessed using the continuous performance test, identical pairs version.

Social cognition. Social cognition was measured using both the Eyes of the mind and the Hinting task to index theory of mind, while the Interpersonal and situational attribution task (IPSAQ) was used to index attributional style.

Statistical Analysis

To inform appropriate adjustments in the primary cognitive analyses, the associations of CREB1 with demographic variables were investigated using 1-way analysis of variance. There was no difference in age, years of education, or sex distribution between the genotype groups. Association of CREB1 with the phenotypes of IQ, episodic memory, working memory, attention and social cognition was conducted using a general factorial design in SPSS statistical software v14.0 (SPSS Inc, Chicago, Illinois). In a series of analyses of variance, scores for each neuropsychological subtest were entered as dependent variables, with age and sex included as covariates as appropriate.

Table S1. Description of individual samples included in replication analysis

Sample	Ancestry	Cases	Case diagnosis	Controls	Genotyping Platform	λ	Ref.
France	French	451	BD1,BD2,BD-NOS	1,631	Illumina platform	1.03	28,29
Sweden	Swedish	836[a]	BD1,BD2,BD-NOS	2,093	Affymetrix 6.0	1.07	28
Germany	German	181	BD1,BD2,SAB,BD-NOS	527	Illumina platform	1.00	30,31
Australia	Australian	330	BD1,BD2,SAB,BD-NOS	1,811	Illumina platform	1.00	8,30,32,33
Poland	Polish	411	BD1,BD2	504	Illumina platform	/	30
Iceland	Icelandic	544[b]	BD1,BD2,BD-NOS	34,426	Affymetrix 6.0	1.11	34
U.K.	British	1,636	MDD	1,594	Illumina HumanHap610	1.02	35
Munich-Germany	German	640	MDD	542	Illumina HumanHap300	1.02	36
Total	/	5,029	/	43,128	/	/	/

[a] Diagnoses were made from hospital discharge records 226 of the samples, and thus DSM-IV subtypes are unavailable. Subtypes are given for 599 of the cases.
[b] DSM-IV diagnoses were not available for 150 subjects who had been collected under earlier diagnostic criteria and described above.
λ = genomic control lambda.
Abbreviations: BD1, bipolar disorder type 1; BD2, bipolar disorder type 2; BD-NOS, bipolar disorder not otherwise specified; SCZ, schizophrenia; MDD, major depressive disorder; SAB, schizoaffective disorder (bipolar type).

Table S2. Demographic details of samples in the Stanley Neuropathology Consortium Integrative Database

	Schizophrenia	Bipolar Disorder	Major Depression	Normal Controls
Age(y)	44.2 (25~62)	42.3 (25~61)	46.4 (30~65)	48.1 (29~68)
Sex	9M, 6F	9M, 6F	9M, 6F	9M, 6F
Race	13C, 2A	14C, 1AA	15C	14C, 1AA
PMI (hs)	33.7 (12~61)	32.5 (13~62)	27.5 (7~47)	23.7 (8~42)
pH	6.1 (5.8~6.6)	6.2 (5.8~6.5)	6.2 (5.6~6.5)	6.3 (5.8~6.6)
Side of Brain Frozen	6R, 9L	8R, 7L	6R, 9L	7R, 8L

Note:
M, male; F, female; C, Caucasian; AA, African American; R, right; L, left.

Table S3. Bioinformatics screen of the cis-eQTL SNPs in CREB1 using the Regulome DB tools

SNP	Position	Regulome DB score	Motifs		Protein binding		Chromatin structure		Histone modifications	
			Method	Motif	Method	Bound protein	Method	Chromatin structure	Method	Histone Mark
rs2709370	208382602	6	PWM	2 motifs	/	/	/	/	ChIP-seq	multiple
rs2709371	208384687	No data	/	/	/	/	/	/	/	/
rs2709373	208386024	6	PWM	2 motifs	/	/	FAIRE	1 chromatin structure	ChIP-seq	multiple
rs2709374	208386604	6	PWM	5 motifs	/	/	FAIRE	1 chromatin structure	ChIP-seq	multiple
rs2551660	208386799	No data	/	/	/	/	/	/	/	/
rs2709375	208388923	5	/	/	/	/	FAIRE DNase-seq	2 chromatin structures 2 chromatin structures	ChIP-seq	multiple
rs2551637	208392132	No data	/	/	/	/	/	/	/	/
rs2709377	208393907	4	/	/	ChIP-seq	4 bound proteins	FAIRE DNase-seq	4 chromatin structures multiple	ChIP-seq	multiple
rs2464976	**208394531**	**2b**	**PWM Footprinting**	**1 motif multiple**	**ChIP-seq**	**multiple**	**FAIRE DNase-seq**	**multiple multiple**	**ChIP-seq**	**multiple**
rs2551639	208396187	3a	PWM	1 motif	ChIP-seq	1 bound protein	FAIRE DNase-seq	8 chromatin structures 2 chromatin structures	ChIP-seq	multiple
rs889895	208398929	5	PWM Footprinting	2 motifs 3 motifs	/	/	DNase-seq	9 chromatin structures	ChIP-seq	multiple
rs2551641	208410267	5	/	/	ChIP-seq	1 bound protein	/	/	ChIP-seq	multiple
rs2709356	208412092	5	PWM	1 motif	/	/	FAIRE DNase-seq	2 chromatin structures 1 chromatin structure	ChIP-seq	multiple
rs2709357	208412532	6	PWM	multiple	/	/	/	/	ChIP-seq	multiple
rs2551642	208413497	6	PWM	1 motif	/	/	/	/	ChIP-seq	multiple
rs2709398	208417937	No data	/	/	/	/	/	/	/	/
rs2709399	208418101	No data	/	/	/	/	/	/	/	/
rs2709400	208418313	No data	/	/	/	/	/	/	/	/
rs2551917	208418332	No data	/	/	/	/	/	/	/	/
rs2709402	208420977	No data	/	/	/	/	/	/	/	/
rs2709403	208422263	6	PWM	2 motif	/	/	/	/	ChIP-seq	multiple
rs2709404	208422335	No data	/	/	/	/	/	/	/	/
rs2709378	208423429	No data	/	/	/	/	/	/	/	/
rs2709381	208429159	No data	/	/	/	/	/	/	/	/
rs2551919	208430383	No data	/	/	/	/	/	/	/	/
rs2551920	208431830	1f	/	/	/	/	DNase-seq	1 chromatin structure	ChIP-seq	multiple
rs2709385	208434645	No data	/	/	/	/	/	/	/	/
rs2551921	208435555	4	/	/	ChIP-seq	1 bound protein	FAIRE DNase-seq	1 chromatin structure 1 chromatin structure	ChIP-seq	multiple
rs2551643	208437433	6	PWM	1 motif	/	/	/	/	ChIP-seq	multiple
rs2551644	208440724	6	PWM	multiple	/	/	/	/	ChIP-seq	multiple
rs2551923	208441776	No data	/	/	/	/	/	/	/	/
rs2709387	208442095	6	/	/	/	/	/	/	ChIP-seq	multiple
rs2709388	208443011	6	PWM	5 motifs	/	/	FAIRE	1 chromatin structure	ChIP-seq	multiple
rs2551924	208443372	6	PWM	1 motif	/	/	/	/	ChIP-seq	multiple
rs2709389	208448995	5	PWM	3 motifs	/	/	DNase-seq	1 chromatin structure	ChIP-seq	multiple

rs34903544	208449146	No data	/	/	/	/	/	/	/	/
rs2551646	208451166	No data	/	/	/	/	/	/	/	/
rs2551647	208451238	No data	/	/	/	/	/	/	/	/
rs2709392	208454639	No data	/	/	/	/	/	/	/	/
rs2247053	208457063	No data	/	/	/	/	/	/	/	/
rs2952766	208457362	6	PWM	1 motif	/	/	/	/	ChIP-seq	multiple
rs2464978	208457918	No data	/	/	/	/	/	/	/	/
rs34775277	208458129	6	PWM	1 motif	/	/	/	/	ChIP-seq	multiple
rs2709362	208459822	No data	/	/	/	/	/	/	/	/
rs13029936	208465544	5	/	/	/	/	FAIRE DNase-seq	1 chromatin structure 1 chromatin structure	ChIP-seq	multiple
rs2551928	208465778	No data	/	/	/	/	/	/	/	/
rs1806584	208466270	6	PWM	4 motifs	/	/	/	/	ChIP-seq	multiple
rs1045780	208467150	No data	/	/	/	/	/	/	/	/
rs6785	208467997	6	/	/	/	/	/	/	ChIP-seq	multiple

Note:
when the number of predicted transcription factor binding motifs (or protein binding sequences, chromatin structures, histone marks) is more than 10, we indicated it as “multiple” instead.
The methods for the bioinformatics screen on ENCODE data have been described previously, which can be found from Boyle et al study.[37]

Table S4. Association results of SNPs spanning CREB1 region with BD in the PGC BD GWAS samples and their allele frequencies in three major populations

SNP ID	Position	Location	Allele	*P*-values	OR	Allele Frequency		
						CEU	CHB	YRI
rs2709370	**208382602**	**5' near gene**	**C**	**1.80×10^{-3}**	**1.094**	**0.224**	**0**	**0.142**
rs2709371	**208384687**	**5' near gene**	**G**	**2.93×10^{-3}**	**1.091**	**0.212**	**0**	**0.148**
rs2709373	**208386024**	**5' near gene**	**C**	**2.15×10^{-3}**	**1.094**	**0.212**	**0**	**0.142**
rs2253206	208391978	5' near gene	A	0.704	1.010	0.494	0.402	0.324
rs2551639	**208396187**	**intron**	**G**	**9.15×10^{-4}**	**1.103**	**0.212**	**0**	**0.142**
rs889895	**208398929**	**intron**	**G**	**8.20×10^{-4}**	**1.104**	**0.212**	**0**	**0.142**
rs2551640	208407893	intron	A	0.529	0.985	0.376	0.402	0.33
rs2709356	**208412092**	**intron**	**T**	**9.35×10^{-4}**	**1.102**	**0.206**	**0**	**0.142**
rs2709399	**208418101**	**intron**	**T**	**9.84×10^{-4}**	**1.102**	**0.206**	**0**	**0.142**
rs2709402	**208420977**	**intron**	**A**	**9.98×10^{-4}**	**1.102**	**0.206**	**0**	**0.142**
rs10932201	208426257	intron	A	1.04×10^{-3}	0.925	0.463	0.604	0.127
rs11904814	208426798	intron	G	0.326	1.025	0.341	0.402	0.165
rs6740584	208429351	intron	C	9.07×10^{-4}	0.924	0.418	0.598	0.057
rs2551919	**208430383**	**intron**	**T**	**9.24×10^{-4}**	**1.103**	**0.206**	**0**	**0.148**
rs2551920	**208431830**	**intron**	**G**	**9.79×10^{-4}**	**1.102**	**0.206**	**0**	**0.142**
rs2551921	**208435555**	**intron**	**C**	**8.85×10^{-4}**	**1.103**	**0.206**	**0**	**0.142**
rs17811997	208440836	intron	C	3.09×10^{-3}	0.834	0.065	0	0
rs2551923	**208441776**	**intron**	**T**	**1.22×10^{-3}**	**1.101**	**0.206**	**0**	**0.148**
rs2709387	**208442095**	**intron**	**A**	**8.61×10^{-4}**	**1.103**	**0.206**	**0**	**0.205**
rs2254137	208444028	intron	A	0.448	0.982	0.629	0.598	0.392
rs2464978	**208457918**	**intron**	**A**	**1.00×10^{-3}**	**1.102**	**0.206**	**0**	**0.142**
rs2551928	**208465778**	**3'UTR**	**A**	**6.74×10^{-4}**	**1.106**	**0.212**	**0**	**0.244**
rs1806584	**208466270**	**3'UTR**	**G**	**5.54×10^{-4}**	**1.107**	**0.206**	**0**	**0.011**
rs1045780	**208467150**	**3'UTR**	**A**	**5.62×10^{-4}**	**1.107**	**0.206**	**0**	**0.142**
rs6785	**208467997**	**3'UTR**	**A**	**3.38×10^{-4}**	**1.111**	**0.206**	**0**	**0.142**

SNP position is defining according to the NCBI, GRCh37.p5 version.
The risk and cis-eQTL SNPs are marked in red color and bold.

Table S5. Association results of risk SNPs in CREB1 in each European replication sample

Study group (N cases/N controls)	rs2709370 [C]		rs6785 [A]	
	***P* value**	**OR (SE)**	***P* value**	**OR (SE)**
France (451/1,631)	0.4111	1.083 (0.0970)	0.5095	1.066 (0.0962)
Sweden (836/2,093)	0.5008	1.050 (0.0720)	0.5084	1.051 (0.0753)
Germany (181/527)	0.2414	1.180 (0.1532)	0.1445	1.235 (0.1521)
Australia (330/1,811)	*0.4666*	*0.927 (0.1058)*	*0.6623*	*0.956 (0.1084)*
Poland (411/504)	0.0536	1.208 (0.1077)	0.0891	1.186 (0.1081)
Iceland (544/34,426)	0.8618	1.012 (0.0735)	*0.8655*	*0.987 (0.0773)*
U.K. (1,636/1,594)	0.2897[1]	1.072 (0.0655)	0.3098[2]	1.069 (0.0658)
Munich-Germany (640/542)	0.0903	1.187 (0.1011)	0.1246	1.171 (0.1030)

Abbreviations:
OR, odds ratio; SE, standard error.
[1]rs2709373 (the 1000-Human-Genome, CEU r^2=0.93 with rs2709370)
[2]rs2551949 (the 1000-Human-Genome, CEU r^2=1.00 with rs6785)

Table S6. Association of rs2709370 with demographic characteristics in the German fMRI sample

	CC (N=14)	AC (N=99)	AA (N=166)	*P* value
Age (y)	33.5 (19~48)	34.0 (18~50)	34.8 (18~51)	0.77
Sex (M/F)	4M, 10F	46M, 53F	83M, 83F	0.29
Handedness (L/R)*	0/14	3/83	10/151	0.38
School education (y)	6.21 (4~7)	6.04 (3~7)	6.25 (3~7)	0.30

y=years, M=male, F=female, L=left, R=right,* 7 data points missing

Table S7. Analyses on cognitive performance by rs2709370 genotype in the Irish sample

Cognitive performance	Test or Subscale	N	Mean (SD)			$F_{Main\ effect}$	*P* value
			AA	AC	CC		
IQ	Abbreviated Full Scale IQ	87	124.4 (13.2)	125 (14.1)	126 (15.6)	0.027	0.974
	WTAR (Adult reading test)	87	110.2 (4.6)	110.5 (4.5)	109.5 (4.9)	0.112	0.894
	Verbal IQ	87	126.7 (14.6)	125.5 (14.9)	133.5 (16.2)	0.294	0.746
	Performance IQ	87	118.8 (15.6)	120.4 (17.5)	115 (8.5)	0.157	0.855
Working Memory	LN sequence	86	12.8 (2.8)	11.8 (3.5)	15 (2.8)	1.49	0.23
	CANTAB SWM (Between error)	87	0.19 (0.86)	0.3 (0.7)	0.8 (0.2)	0.678	0.511
Episodic Memory	Logical Memory Immediate	88	13.3 (2.7)	12.6 (2.7)	15 (1.4)	1.132	0.327
	Logical Memory Delayed	88	13.9 (2.6)	13.1 (3.47)	14.5 (.707)	0.81	0.448
	CANTAB PAL (Adjusted Standard score)	85	0.42 (0.8)	-0.03 (1.5)	0.13 (0.38)	1.627	0.203
	Faces 1	87	11.9 (2.8)	10.5 (2.6)	10 (0)	2.619	0.079
	Faces 2	87	11.5 (2.3)	10.6 (2.7)	9.5 (3.5)	1.614	0.205
Attentional Control	IDED (8 shapes adjusted)	84	9.1 (9.8)	5.1 (6.3)	7.5 (3.5)	1.85	0.164
	IDED (6 shapes adjusted))	84	0.87 (2.9)	0.62 (1.21)	1 (0)	0.091	0.913
Social Cognition	Hinting Task	82	16.8 (1.8)	16.7 (1.7)	17.5 (2.1)	0.171	0.84
	Reading the Eyes in the mind	82	26.6 (3.1)	25.2 (2.3)	26.5 (3.5)	2.3	0.1
	IPSAQ (externalizing bias)	82	0.26 (4.2)	1.2 (3.1)	1 (2.8)	0.52	0.59
	IPSAQ (personalizing bias)	82	0.59 (0.23)	0.59 (0.25)	0.65 (0.21)	0.056	0.95

Table S8. Analyses on cognitive performance by rs6785 genotype in the Irish sample

Cognitive function	Test or Subscale	N	Mean (SD)			$F_{Main\ effect}$	*P* value
			AA	AG	GG		
IQ	Abbreviated Full Scale IQ	87	126 (5.6)	125 (14)	124 (13.2)	0.048	0.95
	WTAR (Adult reading test)	87	109.5 (4.9)	110.2 (4.6)	110.2 (4.6)	0.029	0.97
	Verbal IQ	87	133.5 (16.2)	124 (15.1)	127.1 (14.4)	0.62	0.54
	Performance IQ	87	115 (8.5)	122 (17.2)	118.2 (15.6)	0.57	0.56
Working	LN sequence	86	15 (2.8)	119 (3.6)	12.7 (2.8)	1.29	0.27
Memory	CANTAB SWM (Between error)	87	0.81 (0.2)	0.37 (0.64)	0.17 (0.87)	1.03	0.36
Episodic Memory	Logical Memory Immediate	88	15 (1.4)	12.5 (2.6)	13.4 (2.7)	1.48	0.232
	Logical Memory Delayed	88	14.5 (0.71)	12.9 (3.5)	14 (2.6)	1.3	0.27
	CANTAB PAL (Adjusted Standard score)	85	0.13 (0.38)	-0.02 (1.6)	0.39 (0.8)	1.2	0.29
	Faces 1	87	10 (0)	10.7 (2.6)	11.4 (2.5)	1.1	0.35
	Faces 2	87	9.5 (3.5)	10.7 (2.6)	11.4 (2.5)	1.1	0.35
Attentional	IDED (8 shapes adjusted)	84	7.5 (3.5)	5.3 (6.7)	8.7 (9.6)	1.2	0.304
Control	IDED (6 shapes adjusted))	84	1 (0)	0.67 (1.3)	0.84 (2.8)	0.049	0.952
Social Cognition	Hinting Task	82	17.5 (2.1)	16.7 (1.7)	16.8 (1.7)	0.19	0.83
	Reading the Eyes in the mind	82	26.5 (3.5)	25.2 (2.3)	26.5 (3.1)	1.5	0.22
	IPSAQ (externalizing bias)	82	1 (2.8)	1.3 (3.1)	0.23 (4.1)	0.76	0.46
	IPSAQ (personalizing bias)	82	0.65 (0.21)	0.56 (0.24)	0.6 (0.23)	0.36	0.69

Table S9. Results of SNP association analyses with CREB1 expression in brain samples of the Stanley Neuropathology Consortium Integrative Database

	Frontal Cortex	Cerebellum	Thalamus	Hippocampus
rs2709370	3.63×10^{-8}	4.22×10^{-6}	/	2.22×10^{-9}
rs2709373	1.23×10^{-7}	/	2.34×10^{-6}	1.62×10^{-9}
rs6785	7.45×10^{-7}	/	2.34×10^{-6}	5.17×10^{-10}
rs2551949	7.45×10^{-7}	/	2.34×10^{-6}	5.17×10^{-10}

Figure S1. Linkage disequilibrium (LD) structure of the risk SNPs spanning CREB1 in Europeans, and the LD value of tested SNPs were calculated using the r^2 algorithm.

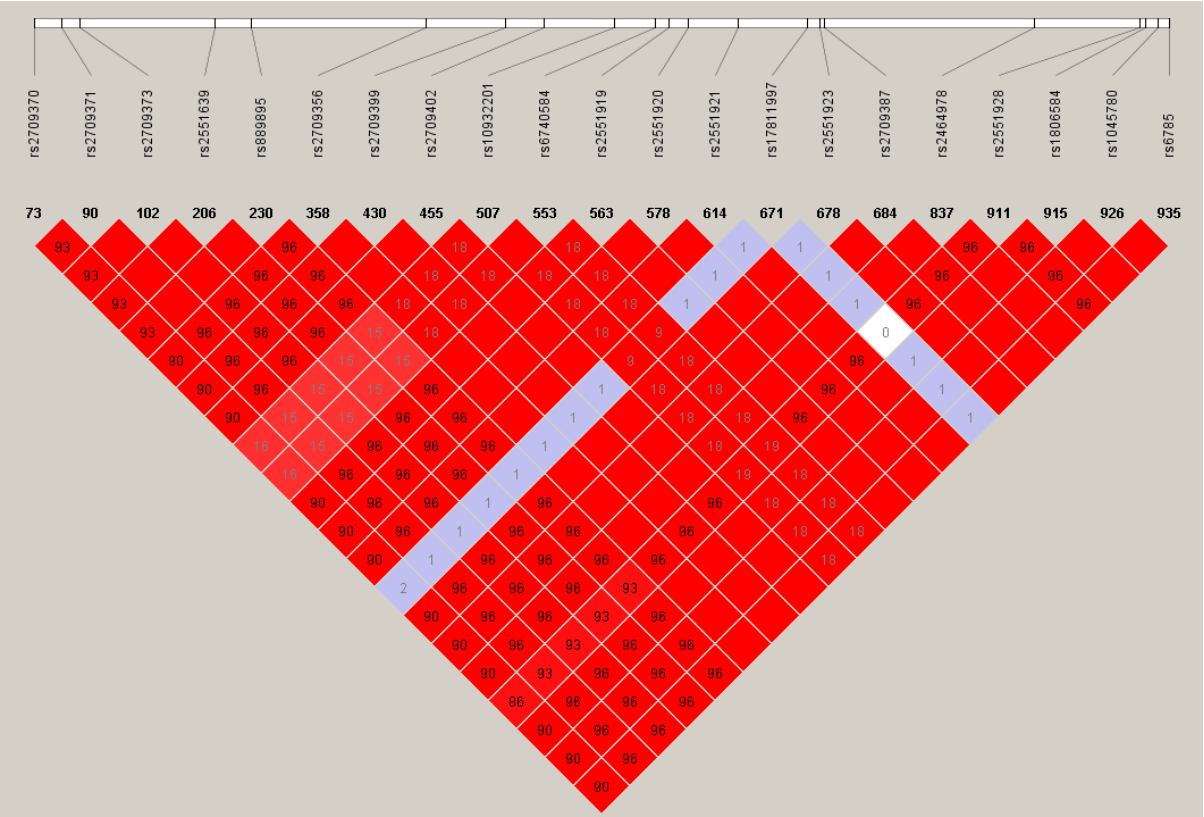
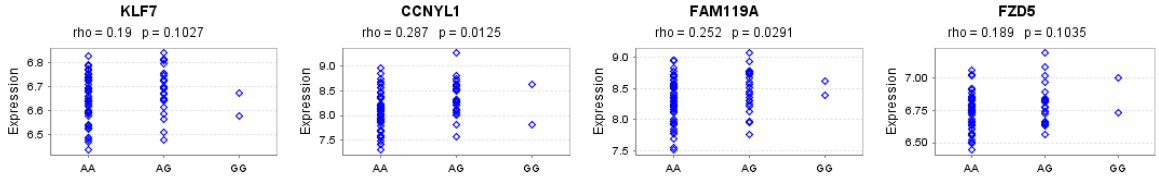


Figure S2. Association of risk SNPs with adjacent gene mRNA expression in Europeans in Dimas *et al.* study.[38]

rs2709373



rs2551949

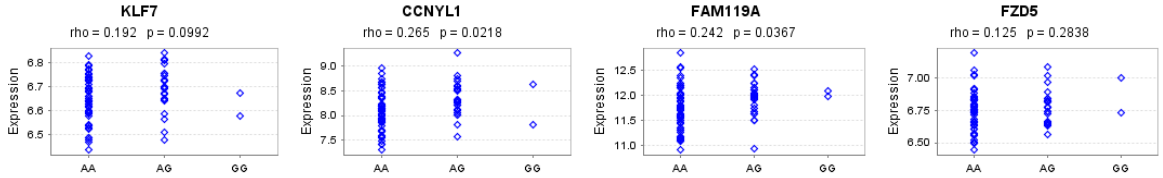


Figure S3. Prediction of binding affinity with microRNA using the mirSNP database.[39]

rs2551928

Gene	miRNA	Snp	mirSVR	CEU	Effect	Allele	Score	Energy	Conservation	Start	End	binding
CREB1	hsa-miR-618	rs2551928		0.199	create	A						
						G	142.00	-13.15	0.322	2	31	Query: 3' ugAGUCUUCCUG-UU-C------AUCUCAAa 5' Ref: 5' atTAAGTAGGGCTAATGTATCTTAGAGTTa 3'

rs1806584

Gene	miRNA	Snp	mirSVR	CEU	Effect	Allele	Score	Energy	Conservation	Start	End	binding
CREB1	hsa-miR-4325	rs1806584	-0.019	0.183	break	A	148.00	-11.75	0.000	10	27	Query: 3' agUGACUCUGUUCACGUu 5' Ref: 5' aaAATAACTCAAGTGCAt 3'
						G						
CREB1	hsa-miR-491-3p	rs1806584	-0.018	0.183	break	A	148.00	-16.92	0.007	1	30	Query: 3' cauCUUCCC--UUA--GA-----ACGUAUUc 5' Ref: 5' attGAAGGGAAAATAACTCAAGTGCATAAt 3'
						G						
CREB1	hsa-miR-521	rs1806584		0.183	create	A						
						G	146.00	-13.77	0.001	7	28	Query: 3' ugugagauUUCCCUUCACGCAa 5' Ref: 5' gggaaaatAACTCAAGTGCGTa 3'
CREB1	hsa-miR-656	rs1806584	-0.010	0.183	break	A	149.00	-5.64	0.010	13	33	Query: 3' ucuccaacUGACAUAUUAUAa 5' Ref: 5' ataactcaAGTGCATAATATt 3'
						G						

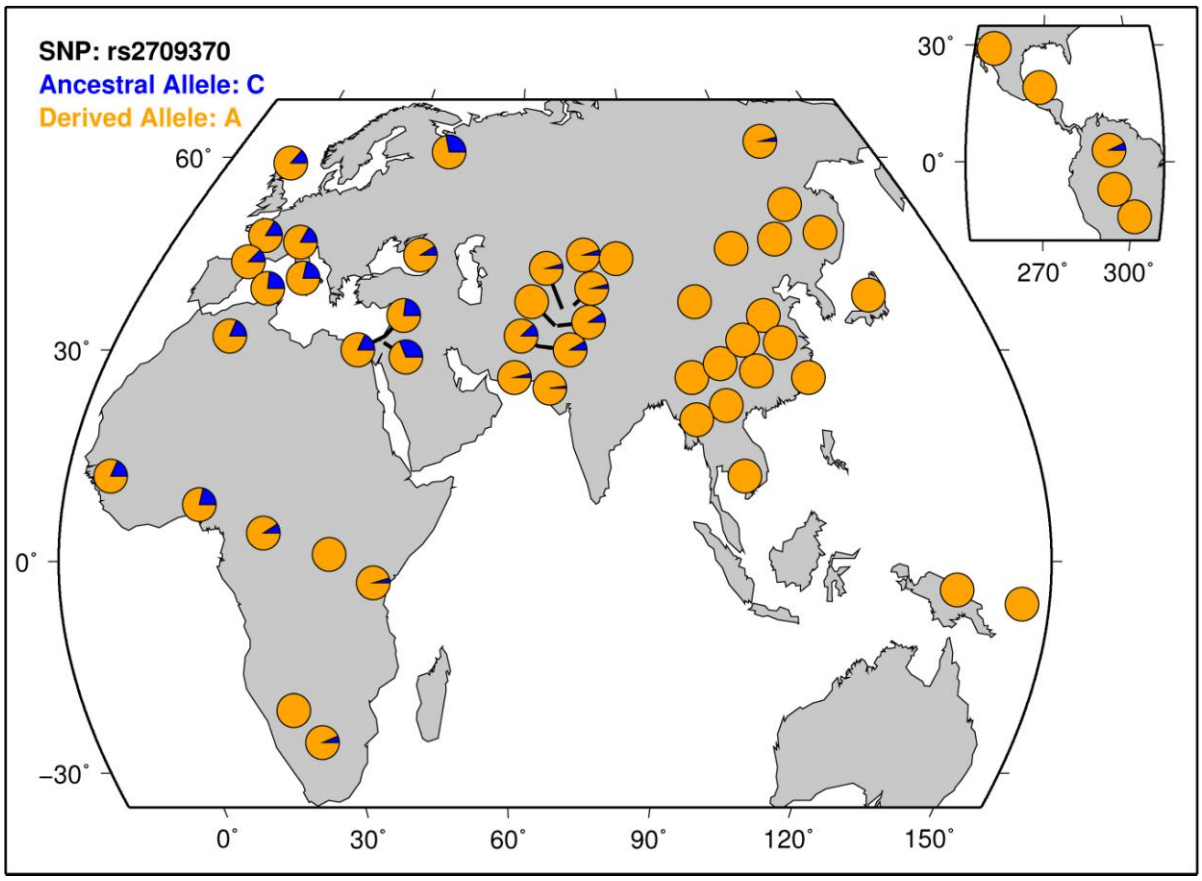
rs1045780

Gene	miRNA	Snp	mirSVR	CEU	Effect	Allele	Score	Energy	Conservation	Start	End	binding
CREB1	hsa-miR-4445-5p	rs1045780		0.195	break	A	161.00	-13.25	0.199	9	29	Query: 3' acgUGCCGUUUUCUUUGUUAGa 5' Ref: 5' taaAAGG-AAAATGAACAATCt 3'
						G						
CREB1	hsa-miR-548u	rs1045780		0.195	create	A						
						G	167.00	-16.25	0.026	8	32	Query: 3' gcgUUUUCAUUAAC--GUCAGAAAc 5' Ref: 5' gtaAAAGGAAAATGAACAGTCTTTt 3'

rs6785

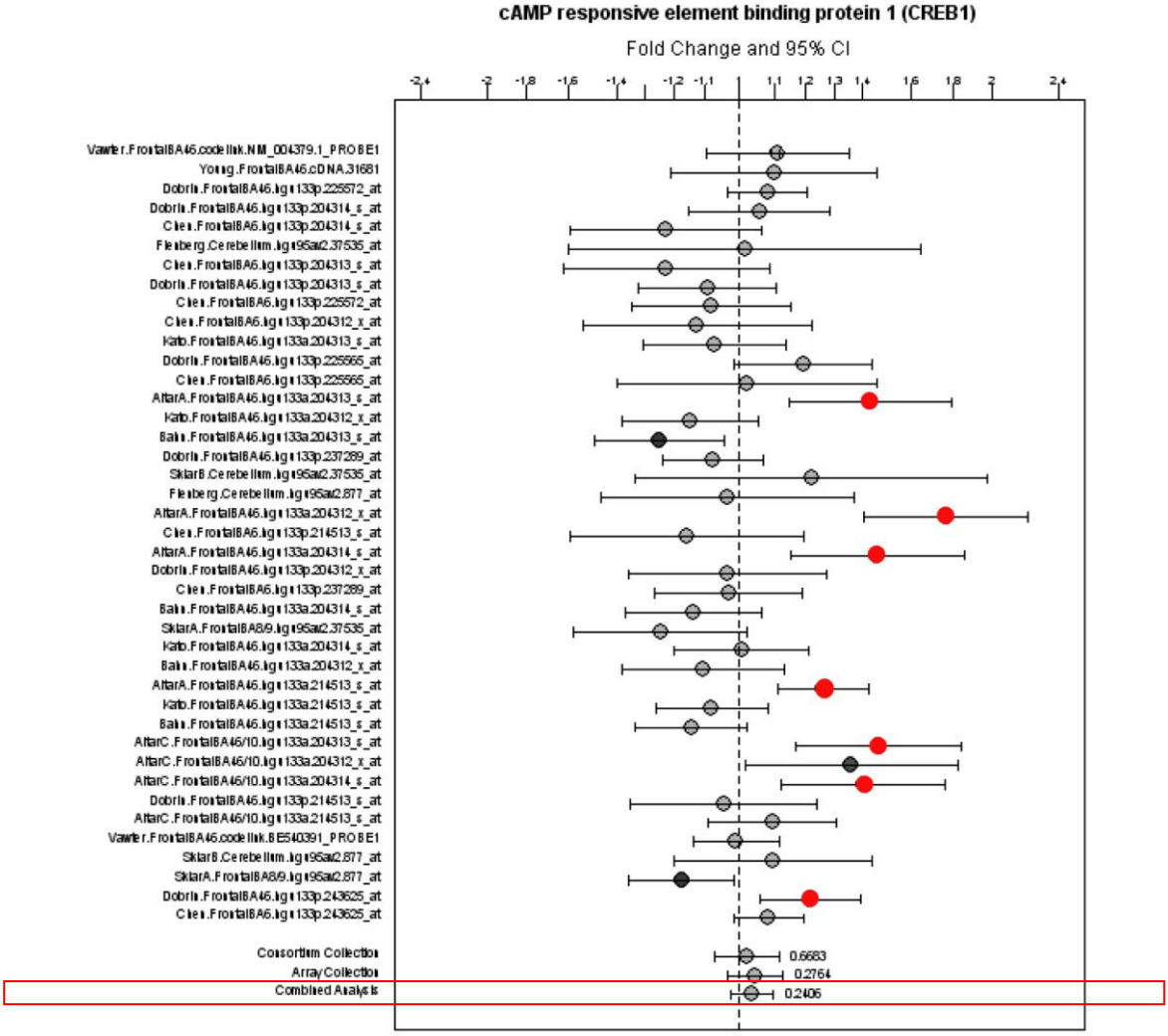
Gene	miRNA	Snp	mirSVR	CEU	Effect	Allele	Score	Energy	Conservation	Start	End	binding
CREB1	hsa-miR-26a-1-3p	rs6785	-0.024	0.186	break	A	145.00	-13.36	1.000	8	28	Query: 3' gcacguuCAUUGGUUCUUAUCc 5' Ref: 5' ctgctttGTGTTC-AGAATAGt 3'
						G						
CREB1	hsa-miR-26a-2-3p	rs6785	-0.024	0.186	break	A	149.00	-10.34	1.000	8	28	Query: 3' cuuuguuCAUUAGUUCUUAUCc 5' Ref: 5' ctgctttGTGTTC-AGAATAGt 3'
						G						
CREB1	hsa-miR-330-3p	rs6785	-0.061	0.186	decrease	A	150.00	-14.35	1.000	22	42	Query: 3' agagaCGUCCGGCACACGAAACg 5' Ref: 5' gaataGTA-GCAGT-TGCTTTGt 3'
						G	140.00	-8.58	1.000	1	16	Query: 3' agagacguccggcacACGAAACg 5' Ref: 5' -------ataatagcTGCTTTGt 3'
CREB1	hsa-miR-330-3p	rs6785	-0.061	0.186	enhance	A	140.00	-8.58	1.000	1	16	Query: 3' agagacguccggcacACGAAACg 5' Ref: 5' -------ataatagcTGCTTTGt 3'
						G	150.00	-14.52	1.000	22	42	Query: 3' agagaCGUCCGGCACACGAAACg 5' Ref: 5' gaatgGTA-GCAGT-TGCTTTGt 3'
CREB1	hsa-miR-524-5p	rs6785	-0.054	0.186	decrease	A	153.00	-13.92	1.000	20	44	Query: 3' cuCUU-UCA-CG-AAGGGAAACAUc 5' Ref: 5' caGAATAGTAGCAGTTGCTTTGTAt 3'
						G	149.00	-13.47	1.000	20	44	Query: 3' cuCUU-UCA-CG-AAGGGAAACAUc 5' Ref: 5' caGAATGGTAGCAGTTGCTTTGTAt 3'

Figure S4. Global distribution of a risk SNP rs2709370 allele in major world populations



*The ancestral allele C is the risk allele

Figure S5. mRNA expression analyses of CREB1 in the brain of bipolar disorder patients and controls in an online database (https://www.stanleygenomics.org**)**



Spots indicate the mean fold change of CREB1 mRNA expression in bipolar disorder patients as compared with normal controls in individual studies and the combined analysis (labeled by the red box). Among the 41 individual studies, 19 studies reported up-regulation of CREB1 in bipolar disorder patients as compared with normal controls, with 7 of them reaching statistical significance ($P < 0.01$ [red spot]), and 21 studies found down-regulations of CREB1 in bipolar disorder patients compared with normal controls, but none of them reached statistical significance ($P < 0.01$). The combined analysis indicated a slight increase in CREB1 expression in BD patients than controls, though not significantly (combined $P = 0.2406$).

REFERENCES

1. Nurnberger JI, Jr., Blehar MC, Kaufmann CA, York-Cooler C, Simpson SG, Harkavy-Friedman J *et al.* Diagnostic interview for genetic studies. Rationale, unique features, and training. NIMH Genetics Initiative. *Arch Gen Psychiatry* 1994; **51:** 849-859; discussion 863-864.
2. Spitzer RL, Williams JB, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description. *Arch Gen Psychiatry* 1992; **49:** 624-629.
3. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E *et al.* The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998; **59 Suppl 20:** 22-33;quiz 34-57.
4. Sellgren C, Landen M, Lichtenstein P, Hultman CM, Langstrom N. Validity of bipolar disorder hospital discharge diagnoses: file review and multiple register linkage in Sweden. *Acta Psychiatr Scand* 2011; **124:** 447-453.
5. McGuffin P, Farmer A, Harvey I. A polydiagnostic application of operational criteria in studies of psychotic illness. Development and reliability of the OPCRIT system. *Arch Gen Psychiatry* 1991; **48:** 764-770.
6. Schmermund A, Mohlenkamp S, Stang A, Gronemeyer D, Seibel R, Hirche H *et al.* Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study. Risk Factors, Evaluation of Coronary Calcium and Lifestyle. *Am Heart J* 2002; **144:** 212-218.
7. McAuley EZ, Fullerton JM, Blair IP, Donald JA, Mitchell PB, Schofield PR. Association between the serotonin 2A receptor gene and bipolar affective disorder in an Australian cohort. *Psychiatr Genet* 2009; **19:** 244-252.
8. Mitchell PB, Johnston AK, Corry J, Ball JR, Malhi GS. Characteristics of bipolar disorder in an Australian specialist outpatient clinic: comparison across large datasets. *Aust N Z J Psychiatry* 2009; **43:** 109-117.
9. Cohen-Woods S, Gaysina D, Craddock N, Farmer A, Gray J, Gunasinghe C *et al.* Depression Case Control (DeCC) Study fails to support involvement of the muscarinic acetylcholine receptor M2 (CHRM2) gene in recurrent major depressive disorder. *Hum Mol Genet* 2009; **18:** 1504-1549.
10. Farmer A, Breen G, Brewster S, Craddock N, Gill M, Korszun A *et al.* The Depression Network (DeNT) Study: methodology and sociodemographic characteristics of the first 470 affected sibling pairs from a large multi-site linkage genetic study. *BMC Psychiatry* 2004; **4:** 42.
11. McGuffin P, Knight J, Breen G, Brewster S, Boyd PR, Craddock N *et al.* Whole genome linkage scan of recurrent depressive disorder from the depression network study. *Hum Mol Genet* 2005; **14:** 3337-3345.
12. Uher R, Maier W, Hauser J, Marusic A, Schmael C, Mors O *et al.* Differential efficacy of escitalopram and nortriptyline on dimensional measures of depression. *Br J Psychiatry* 2009; **194:** 252-259.
13. Uher R, Huezo-Diaz P, Perroud N, Smith R, Rietschel M, Mors O *et al.* Genetic predictors of response to antidepressants in the GENDEP project. *Pharmacogenomics J* 2009; **9:** 225-233.
14. Wing JK, Babor T, Brugha T, Burke J, Cooper JE, Giel R *et al.* SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Arch Gen Psychiatry* 1990; **47:** 589-593.
15. Sham PC, Sterne A, Purcell S, Cherny S, Webster M, Rijsdijk F *et al.* GENESiS: creating a composite index of the vulnerability to anxiety and depression in a community-based sample of siblings. *Twin Res* 2000; **3:** 316-322.
16. McGuffin P, Katz R, Aldrich J. Past and present state examination: the assessment of 'lifetime ever' psychopathology. *Psychol Med* 1986; **16:** 461-465.
17. Hennings JM, Owashi T, Binder EB, Horstmann S, Menke A, Kloiber S *et al.* Clinical characteristics and treatment outcome in a representative sample of depressed inpatients - findings from the Munich Antidepressant Response Signature (MARS) project. *J Psychiatr Res* 2009; **43:** 215-229.
18. Ising M, Lucae S, Binder EB, Bettecken T, Uhr M, Ripke S *et al.* A genomewide association study points to multiple loci that predict antidepressant drug treatment outcome in depression. *Arch Gen Psychiatry* 2009; **66:** 966-975.
19. Heck A, Lieb R, Ellgas A, Pfister H, Lucae S, Erhardt A *et al.* Polymorphisms in the angiotensin-converting enzyme gene region predict coping styles in healthy adults and depressed patients. *Am J Med Genet B Neuropsychiatr Genet* 2009; **150B:** 104-114.
20. Walter H, Schnell K, Erk S, Arnold C, Kirsch P, Esslinger C *et al.* Effects of a genome-wide supported psychosis risk variant on neural activation during a theory-of-mind task. *Mol Psychiatry* 2010; **16:** 462-470.

21. Esslinger C, Walter H, Kirsch P, Erk S, Schnell K, Arnold C *et al.* Neural mechanisms of a genome-wide supported psychosis variant. *Science* 2009; **324:** 605.
22. Erk S, Meyer-Lindenberg A, Schnell K, Opitz von Boberfeld C, Esslinger C, Kirsch P *et al.* Brain function in carriers of a genome-wide supported bipolar disorder variant. *Arch Gen Psychiatry* 2010; **67:** 803-811.
23. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012; **491:** 56-65.
24. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; **5:** e1000529.
25. Friedman L, Glover GH. Report on a multicenter fMRI quality assurance protocol. *J Magn Reson Imaging* 2006; **23:** 827-839.
26. Walters JT, Corvin A, Owen MJ, Williams H, Dragovic M, Quinn EM *et al.* Psychosis susceptibility gene ZNF804A and cognitive performance in schizophrenia. *Arch Gen Psychiatry* 2010; **67:** 692-700.
27. Donohoe G, Walters J, Morris DW, Quinn EM, Judge R, Norton N *et al.* Influence of NOS1 on verbal intelligence and working memory in both patients with schizophrenia and healthy control subjects. *Arch Gen Psychiatry* 2009; **66:** 1045-1054.
28. Sklar P, Ripke S, Scott LJ, Andreassen OA, Cichon S, Craddock N *et al.* Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet* 2011; **43:** 977-983.
29. Etain B, Dumaine A, Mathieu F, Chevalier F, Henry C, Kahn JP *et al.* A SNAP25 promoter variant is associated with early-onset bipolar disorder and a high expression level in brain. *Mol Psychiatry* 2009; **15:** 748-755.
30. Cichon S, Muhleisen TW, Degenhardt FA, Mattheisen M, Miro X, Strohmaier J *et al.* Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. *Am J Hum Genet* 2011; **88:** 372-381.
31. Muglia P, Tozzi F, Galwey NW, Francks C, Upmanyu R, Kong XQ *et al.* Genome-wide association study of recurrent major depressive disorder in two European case-control cohorts. *Mol Psychiatry* 2008; **15:** 589-601.
32. Medland SE, Nyholt DR, Painter JN, McEvoy BP, McRae AF, Zhu G *et al.* Common variants in the trichohyalin gene are associated with straight hair in Europeans. *Am J Hum Genet* 2009; **85:** 750-755.
33. McAuley EZ, Blair IP, Liu Z, Fullerton JM, Scimone A, Van Herten M *et al.* A genome screen of 35 bipolar affective disorder pedigrees provides significant evidence for a susceptibility locus on chromosome 15q25-26. *Mol Psychiatry* 2009; **14:** 492-500.
34. McMahon FJ, Akula N, Schulze TG, Muglia P, Tozzi F, Detera-Wadleigh SD *et al.* Meta-analysis of genome-wide association data identifies a risk locus for major mood disorders on 3p21.1. *Nat Genet* 2010; **42:** 128-131.
35. Lewis CM, Ng MY, Butler AW, Cohen-Woods S, Uher R, Pirlo K *et al.* Genome-wide association study of major recurrent depression in the U.K. population. *Am J Psychiatry* 2010; **167:** 949-957.
36. Erhardt A, Czibere L, Roeske D, Lucae S, Unschuld PG, Ripke S *et al.* TMEM132D, a new candidate for anxiety phenotypes: evidence from human and mouse studies. *Mol Psychiatry* 2010; **16:** 647-663.
37. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012; **22:** 1790-1797.
38. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, Attar-Cohen H *et al.* Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* 2009; **325:** 1246-1250.
39. Liu C, Zhang F, Li T, Lu M, Wang L, Yue W *et al.* MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs and eQTLs. *BMC Genomics* 2012; **13:** 661.