Title: Impact of a cis-associated gene expression SNP on chromosome 20q11.22 on bipolar disorder susceptibility, hippocampal structure and cognitive performance.
Authors: Li M, Luo XJ, Landén M, Bergen SE, Hultman CM, Li X, Zhang W, Yao YG, Zhang C, Liu J, Mattheisen M, Cichon S, Mühleisen TW, Degenhardt FA, Nöthen MM, Schulze TG, Grigoroiu-Serbanescu M, Li H, Fuller CK, Chen C, Dong Q, Chen C, Jamain S, Leboyer M, Bellivier F, Etain B, Kahn JP, Henry C, Preisig M, Kutalik Z, Castelao E, Wright A, Mitchell PB, Fullerton JM, Schofield PR, Montgomery GW, Medland SE, Gordon SD, Martin NG, MooDS Consortium, Swedish Bipolar Study Group, Rietschel M, Liu C, Kleinman JE, Hyde TM, Weinberger DR, Su B
Journal: The British journal of psychiatry : the journal of mental science
Year: 2016 Feb
Volume: 208
Issue: 2
Pages: 128-37
DOI: 10.1192/bjp.bp.114.156976

Impact of a *cis*-associated gene expression SNP in 20q11.22 on bipolar disorder susceptibility, hippocampal structure and cognitive performance

Ming Li[1,2,†] Ph.D.; **Xiong-jian Luo**[3,4,†] Ph.D.; **Mikael Landén**[5,6] M.D., Ph.D.; **Sarah E. Bergen**[5,7] Ph.D.; **Christina M. Hultman**[5] Ph.D.; **Xiao Li**[3,8] M.Sc.; **Wen Zhang**[3] Ph.D.; **Yong-Gang Yao**[3] Ph.D.; **Chen Zhang**[9] Ph.D.; **Jiewei Liu**[1,8] M.Sc.; **Manuel Mattheisen**[10] M.D.; **Sven Cichon**[11-13] Ph.D.; **Thomas W. Mühleisen**[12,13] Ph.D.; **Franziska A. Degenhardt**[12] M.D.; **Markus M. Nöthen**[12,14] M.D.; **Thomas G. Schulze**[15,16] M.D.; **Maria Grigoroiu-Serbanescu**[17] Ph.D.; **Hao Li**[18] Ph.D.; **Chris K. Fuller**[18] M.Sc.; **Chunhui Chen**[19,20] Ph.D.; **Qi Dong**[19,20] Ph.D.; **Chuansheng Chen**[21] Ph.D.; **Stéphane Jamain**[22-24] Ph.D.; **Marion Leboyer**[22-25] M.D., Ph.D.; **Frank Bellivier**[22,24,26,27] M.D., Ph.D.; **Bruno Etain**[22-25] M.D., Ph.D.; **Jean-Pierre Kahn**[24,28] M.D., Ph.D.; **Chantal Henry**[22-25] M.D., Ph.D.; **Martin Preisig**[29] M.D., M.P.H.; **Zoltán Kutalik**[30,31] Ph.D.; **Enrique Castelao**[29] M.Sc.; **Adam Wright**[32,33] Ph.D.; **Philip B. Mitchell**[32,33] M.D.; **Janice M. Fullerton**[34,35] Ph.D.; **Peter R. Schofield**[34,35] Ph.D.; **Grant W. Montgomery**[36] Ph.D.; **Sarah E. Medland**[36] Ph.D.; **Scott D. Gordon**[36] Ph.D.; **Nicholas G. Martin**[36] Ph.D.; MooDS Consortium[#], The Swedish Bipolar Study Group[#], **Marcella Rietschel**[15] M.D., Ph.D.; **Chunyu Liu**[37] Ph.D.; **Joel E. Kleinman**[2] M.D.; **Thomas M. Hyde**[2] M.D.; **Daniel R. Weinberger**[2] M.D.; **Bing Su**[1] Ph.D.;

[1]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China
[2]Lieber Institute for Brain Development, Johns Hopkins University, Baltimore, Maryland, USA
[3]Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences and Yunnan Province, Kunming Institute of Zoology, Kunming, Yunnan, China
[4]University of Rochester Flaum Eye Institute, University of Rochester, Rochester, New York, USA
[5]Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden
[6]Section of Psychiatry and Neurochemistry, Sahlgrenska Academy at Gothenburg University, Gothenburg, Sweden
[7]Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA
[8]Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, Yunnan, China
[9]Schizophrenia Program, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China
[10]Department of Biomedicine, Aarhus University, Aarhus C, Denmark
[11]Division of Medical Genetics, University of Basel, Basel, Switzerland
[12]Institute of Human Genetics and Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany
[13]Institute of Neuroscience and Medicine (INM-1), Structural and Functional

Organization of the Brain, Genomic Imaging, Research Centre Jülich, D-52425 Jülich, Germany
[14]German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany
[15]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim/University of Heidelberg, Mannheim, Germany
[16]Section on Psychiatric Genetics, Department of Psychiatry and Psychotherapy, University Medical Center, Georg-August-University, Göttingen, Germany
[17]Biometric Psychiatric Genetics Research Unit, Alexandru Obregia Clinical Psychiatric Hospital, Bucharest, Romania
[18]Department of Biochemistry and Biophysics, University of California at San Francisco, San Francisco, California, USA
[19]State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China
[20]Center for Collaboration and Innovation in Brain and Learning Sciences, Beijing Normal University, Beijing, China
[21]Department of Psychology and Social Behavior, University of California, Irvine, California, USA
[22]Inserm U 955, IMRB, Psychiatrie Génétique, Créteil, France
[23]Université Paris Est, Faculté de Médecine, Créteil, France
[24]Fondation Fondamental, Créteil, France
[25]AP-HP, Hôpital A. Chenevier - H. Mondor, Pôle de Psychiatrie, Créteil, France
[26]AP-HP, Groupe hospitalier Lariboisière - F. Widal, Pôle de Psychiatrie, Paris, France
[27]Université Paris Diderot, Paris, France
[28]Département de Psychiatrie et de Psychologie Clinique, CHU de Nancy, Hôpital Jeanne d'Arc, Toul, France
[29]Department of Psychiatry, Centre Hospitalier Universitaire Vaudois, Prilly, Switzerland
[30]Institute of Social and Preventive Medicine, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland
[31]Swiss Institute of Bioinformatics, Lausanne, Switzerland
[32]School of Psychiatry, University of New South Wales, Randwick, Australia
[33]Black Dog Institute, Prince of Wales Hospital, Randwick, Australia
[34]Neuroscience Research Australia, Randwick, Sydney, Australia
[35]School of Medical Sciences, Faculty of Medicine, University of New South Wales, Sydney, Australia
[36]Queensland Institute of Medical Research, Brisbane, Australia
[37]Department of Psychiatry, University of Illinois at Chicago, Chicago, Illinois, USA

[†]These authors contributed equally to this work.

[#]A list of the additional members is provided in the Appendix.

Correspondence

Ming Li or Bing Su, Kunming Institute of Zoology, Chinese Academy of Sciences, 32 East Jiao-chang Rd., Kunming 650223, Yunnan, People's Republic of China.

Email: limingkiz@gmail.com (M. Li), or sub@mail.kiz.ac.cn (B. Su)

Summary:

Bipolar disorder (BPD) is a highly heritable polygenic disorder. Recent enrichment analyses suggest that there may be true risk variants for BPD among the expression quantitative trait loci (eQTL) in the brain.

Aims: We sought to assess the impact of eQTL variants on BPD risk by combining data from both BPD genome-wide association study (GWAS) and brain eQTL.

Method: To detect single-nucleotide polymorphisms (SNPs) that influence expression levels of genes associated with BPD, we jointly analyzed data from a BPD GWAS (7,481 cases and 9,250 controls) and a genome-wide brain (cortical) eQTL (193 healthy controls) using a Bayesian statistical method, with independent follow-up replications. The identified risk SNP was then further tested for association with hippocampal volume (N=5,775) and cognitive performance (N=342) among healthy subjects.

Results: Integrative analysis revealed a significant association between a brain eQTL rs6088662 in 20q11.22 and BPD (Log Bayes Factor=5.48; BPD p-val=5.85×10^{-5}). Follow-up studies across multiple independent samples confirmed the association of the risk SNP (rs6088662) with gene expression and BPD susceptibility (p-val=3.54×10^{-8}). Further exploratory analysis revealed that rs6088662 is also associated with hippocampal volume and cognitive performance in healthy subjects.

Conclusions: Our findings suggest that 20q11.22 is likely a risk region for BPD, highlighting the informativeness of integrating functional annotation of genetic variants for gene expression in advancing our understanding of the biological basis underlying complex diseases such as BPD.

Declaration of interest: None

Bipolar disorder (BPD) is a severe, chronic psychiatric disorder with worldwide lifetime prevalence ranging from 0.5-1.5%.[1] BPD is characterized by a variety of profound mood symptoms including episodes of mania, hypomania and depression, and is often accompanied by psychotic features and cognitive deficits. To date, there has been a fair amount of data from family and twin studies to highlight a strong genetic predisposition for BPD.[1] That said, BPD is a highly polygenic disorder that can vary substantially from population to population. While linkage analyses and genetic association studies have yielded numerous candidate variants for BPD, only a few of these have been satisfactorily replicated across independent samples.[2,3]

With the advances in knowledge of human genetic variations—such as data generated by the HapMap and 1000-Human-Genome projects and several subsequent genome-wide association studies (GWASs) by a number of international collaborators—a wealth of novel susceptible variants for BPD have been reported, particularly SNPs in *CACNA1C*, *ANK3*, *ODZ4*, *NCAN* and *TRANK1*.[4-8] These GWAS-identified risk SNPs unfortunately only account for a small portion of the genetic risk for BPD, which suggests there should be additional loci contributing to the genetic susceptibility. Previous aggregated analyses indicated there may be valid risk loci underlying genetic markers passing only nominal significance in the GWASs,[9] a possibility confirmed by several later studies. For example, a number of schizophrenia (SCZ) and BPD susceptibility SNPs did not reach genome-wide significance in initial GWAS samples, but showed consistent replications in subsequent independent samples, thus implying that these loci might reflect weak but true risk signals.[10]

Genetic loci associated with clinical diagnosis also are expected to be related to so-called intermediate phenotypes implicated in the biology of genetic risk for BP

disorder. Previous studies have reported hippocampal dysfunction (e.g., memory impairment) in BPD patients and their unaffected relatives, implying that variation in hippocampal biology is an intermediate phenotype related to the genetic risk of BPD.[11] In addition, smaller hippocampal volume has been reported in BPD patients.[12,13] Meanwhile, functional neuroimaging studies have revealed that dysfunctions of hippocampus and its closely related regions underpin abnormal affective responses and dysfunctional emotion regulation in BPD.[14] Finally, postmortem studies further support the hypothesis that hippocampal abnormalities are relevant to the altered synaptic plasticity and diminished resilience in BPD.[15] Therefore, analysis of the BPD-associated SNPs on these hippocampus-related phenotypes may provide a plausible way to uncover their functions in neurodevelopment, and possibly, their involvements in disease susceptibility.

Recent successes in integrating disease GWAS and gene expression data for several other complex diseases have been promising,[16-18] and we wondered if such an approach may yield novel results for BPD. Predictably, several lines of evidence have suggested an enrichment of expression quantitative trait loci (eQTL) among BPD susceptibility SNPs in the brain,[19] further highlighting the importance of integrating functional annotation of genetic variants for gene expression to advance our understanding of the biological bases of BPD. In light of these findings, we integrated a BPD GWAS data from 16,731 individuals and a genome-wide eQTL data from 193 normal human brains, followed by a set of independent replications on both eQTL and disease associations.

METHOD

Discovery brain eQTL and BPD GWAS datasets

The brain eQTL dataset used in this study was reported previously.[20] In brief, after excluding ethnic outliers and samples that were possibly related, a total of 193 independent healthy old (age>65) human cortex samples of European origin were included in the eQTL analysis. Detailed information about genotyping and expression profiling as well as statistical methods can be found in the **Supplemental Material** or the original publication.[20]

For the BPD GWAS data, the Psychiatric Genomics Consortium (PGC) BPD working group recently conducted a meta-analysis of large-scale genome-wide data on BPD among populations of European descent (PGC1 GWAS).[6] In this prior study, they opted to compare BPD patients that had experienced pathologically relevant episodes of elevated mood (mania or hypomania) and control subjects from the same geographic and ethnic populations. In sum, we utilized 2,117,872 SNPs across the genome from the GWAS samples (7,481 cases and 9,250 controls), and the association significance (*P*-value) for these SNPs were downloaded from the PGC1 data sharing website (https://pgc.unc.edu/Sharing.php#SharingOpp). Detailed descriptions of the samples, data quality, genotype imputation, genomic controls and statistical analyses can be found in the original GWAS report.[6]

Integrative analysis of eQTL and BPD GWAS data

We integrated the eQTL and BPD GWAS data using a Bayesian statistical framework. Statistical analyses for the eQTL and BPD GWAS was achieved by using the program *Sherlock* (http://sherlock.ucsf.edu/), which has been described elsewhere.[17] In brief, *Sherlock* is based on the rationale that a risk gene for the disease may have at least one eQTL, and these eQTL could alter gene expression, which in turn affects disease susceptibility. Given the probability that this is true, there should

be a significant overlap of the eQTL of a gene and the loci associated with the disorder, which would imply a likely functional role for the gene in that particular disease. At this juncture, *Sherlock* aligns the eQTL and BPD GWASs and only considers the shared SNPs in both datasets. *Sherlock*'s scoring rubric both increases the total gene score for overlapping SNPs and provides a penalty in the absence of an overlap, though associations found only in the BPD GWAS do not alter the score. *Sherlock* computes individual Log Bayes Factors (LBFs) for each SNP pair in the alignment, and the sum of these constitutes the final LBF score for each gene.

Brain eQTL data for replication analysis

Considering that bipolar disorder is a mental disorder that reasonably originates from abnormal brain functions, brain samples are presumably appropriate for replication test of the eQTL results. We first utilized a brain DLPFC (dorsolateral prefrontal cortex) sample (N=320) consisting of healthy controls in Caucasians and African Americans (named as “first replication sample”), in which the sample has been previously used to identify psychiatric risk transcripts.[21-23]

We also used other well-characterized brain expression databases for replication analysis of the eQTL associations. A brief description of the gene expression resources is provided below; more detailed information can be found in the original studies.[18,24-26] (1) BrainCloud: BrainCloud contains genetic information and whole transcriptome expression data from postmortem DLPFC of 261 normal human subjects in Caucasians and African Americans. The data in BrainCloud is aimed at exploring temporal dynamics and genetic control of transcription across lifespan.[24] Of note, there is partial overlap between BrainCloud data and our “first replication sample”. (2) Data from Webster et al: This report studied whole-genome

transcriptome and genome in a series of neuropathologically normal postmortem samples and a confirmed pathologic diagnosis of late-onset Alzheimer disease (LOAD; final N=188 controls, 176 cases), and suggested that studying the transcriptome as a quantitative endophenotype has greater power for discovering risk SNPs influencing expression than the use of discrete diagnostic categories such as presence or absence of disease.[25] It should be noted that the control sample in this study was the same as our discovery brain eQTL sample.[20] (3) SNPExpress: The authors analyzed genome-wide SNPs that were associated with gene expression in human primary cells at the exon level, using Affymetrix exon arrays, evaluating 93 autopsy-collected cortical brain tissue samples with no defined neuropsychiatric condition.[26] (4) Data from Zou et al: They measured expression levels of 24,526 transcripts in brain samples from the cerebellum and temporal cortex of autopsied subjects with Alzheimer's disease (AD, cerebellar n = 197, temporal cortex n = 202), and conducted a genome-wide expression association study (eGWAS) using 213,528 cis-SNPs within 6,100 kb of the tested transcripts.[18] Their results demonstrated the significant contributions of genetic factors to human brain gene expression, which are reliably detected across different brain regions, and also implicated that combined assessment of expression and disease GWAS may provide complementary information in discovery of human disease variants with functional implications.[18]

BPD samples for replication analysis

Replication analyses on BPD were conducted in two steps (replication-I and II), examining a total of 6,056 BPD cases and 46,614 controls from ten different geographic locations. Detailed information on each sample—including diagnostic assessment, genotyping method and quality control—are shown in the **Supplemental**

Data and **Table S1**.

Briefly, the BPD samples used in our replication included: (1) Germany II (181 cases and 527 controls);[5] (2) Germany III (490 cases and 880 controls);[5] (3) Australia (330 cases and 1,811 controls);[5] (4) France (451 cases and 1,631 controls);[2] (5) Sweden I (836 cases and 2,093 controls);[6] (6) Sweden II sample (1,415 cases and 1,271 controls);[6] and (7) Iceland (541 cases and 34,546 controls);[6] (8) Romania (244 cases and 174 controls),[5] and (9) China (350 cases and 888 controls).[27] For replication-II, we used a United Kingdom sample (1,218 cases and 2,913 controls).[28] The ten samples from replication-I and II showed no overlap with the PGC1 BPD samples.[6] Each of the original studies was conducted under appropriate ethical approvals, and written informed consents were obtained from all subjects.

Samples for analysis of hippocampal volume and cognitive performance

For analysis of hippocampal volume, we utilized the data from a recent GWAS conducted by the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) consortium.[29] The GWAS includes a total of 5,775 young healthy individuals (mean age: 34.8 years). Detailed information on the samples, imaging procedures, genotyping methods and statistical analysis can be found in the original GWAS report.[29]

For analysis of cognitive performance, we used a Chinese sample that included 342 healthy Chinese college students from Beijing Normal University who had self-reported no known history of any neurological or psychiatric disorders (197 females and 145 males, aged 18-23). Cognitive and behavioral measures included working memory, executive functions (as assessed with the Attention Network Test, the Wisconsin Card Sorting Task, and a reversal learning test), and motivation traits

etc., which were shown in **Table S2**. This experiment was approved by the Institutional Review Board of the State Key Laboratory of Cognitive Neuroscience and Learning at Beijing Normal University, China. Written informed consent was obtained from all participants following a full explanation of the study procedure.

Statistical analysis

For the replication analysis on BPD, genomic control was used to correct for relatedness and population stratification in each sample,[30] and association *P*-values and allele-specific odds ratios (ORs) for each individual sample were calculated by a logistic regression model with an additive effect using a lambda value (genomic control) as a covariate to adjust for potential population stratification. Meta-analyses were then conducted based on Z-scores by combining data from different samples in the R package (meta module) using the Mantel-Haenszel method under the fixed effects model. As described in previous GWAS meta-analysis,[6] *P*-values for replication samples are reported as one-tailed tests, while *P*-values for all combined samples are shown as two-tailed tests. We used a forest plot to graphically present the individual OR and their 95% confidence interval, i.e., each sample was represented by a square in the forest plot. For the analyses on cognitive performance, two-tailed t-tests were conducted with SPSS 16.0 (SPSS, Chicago, USA).

To explain the logic of the study design, a flow chart about the analytical methods and how variants were taken forward from one stage of analysis to the next was shown in **Figure 1**. All protocols and methods used in this study were approved by the institutional review board of Kunming Institute of Zoology, Chinese Academy of Sciences and adhere to all relevant national and international regulations.

RESULTS

Integrative analysis of eQTL and BPD GWAS data

The *Sherlock* identified a total 20,942 SNPs showing significant eQTL effects, and also having BPD data (e.g., p-value), and these SNPs were included for further analyses. Using a Bayesian statistical method to match the "signature" of genes from the brain eQTL with patterns of association in the BPD GWAS, we ranked the top candidate genes for BPD risk according to their LBF scores and *P*-values. Only genes with LBF scores higher than 5.00 were shown and included for further analyses.

The integrative analysis yielded four candidate risk genes (**Table S3**). The first gene was glycosyltransferase 8 domain containing 1 (*GLT8D1*, LBF=6.78), located on chromosome 3p21.1 that has been repeatedly reported for association with BPD.[31,32] Detailed analysis found that the significant association with this gene was mainly driven by a *cis*-associated SNP rs2251219. This SNP has already been reported in an earlier GWAS of BPD,[32] and was replicated in independent BPD samples (in which their samples overlapped with our replication samples).[33-35] The second top-ranked gene was chemokine (C-X-C motif) ligand 16 (*CXCL16*, LBF=6.16) on chromosome 17p13. To the best of our knowledge, this gene has never been reported in genetic association studies on BPD, and we observed two trans-associated SNPs showing moderate associations with BPD. The third top-ranked gene was *TRPC4AP* (LBF=5.57) on chromosome 20q11.22, with the significance mainly driven by a *cis*-associated SNP (rs6088662, p-val=5.85×10^{-5} with BPD). The last top-ranked gene was *TAF11* (TAF11 RNA polymerase II, TATA box binding protein-associated factor, 28kDa) on chromosome 6p21.31, with a trans-correlated SNP (rs4482754) showed significant association with BPD.

Replication of eQTL effects in diverse samples

Given the myriad confounders in single eQTL database, it is important and necessary to validate the eQTL associations in independent samples. The above four candidate genes and their *cis*- or *trans*- associated SNPs were followed-up in independent eQTL datasets.

For the *cis*-SNP rs2251219 and *GLT8D1*, we observed significant association in one replication sample of AD source (**Table S4**),[25] and a marginal significant association in the BrainCloud sample.[24] However, as demonstrated by a previous study,[32] the association of rs2251219 with *GLT8D1* expression in our discovery eQTL sample (Myers et al. study)[20] may be an artifact since the probes overlapped with other common SNPs, and it could not be replicated in the original cDNA samples of our discovery eQTL dataset by quantitative PCR using the probes not overlapping with known SNPs. In addition to *GLT8D1*, we also analyzed the expression of other nearby genes around rs2251219, however, no promising findings were observed (**Table S4**). For the significant *trans*- eQTL associations in our discovery sample, neither *CXCL16* nor *TAF11* could be validated in any of the replication samples (**Table S5**), implying they might be generated by chance.

For the *cis*-association between rs6088662 and *TRPC4AP* expression, in the discovery eQTL brain sample,[20] the risk allele [G] of rs6088662 showed significantly decreased gene expression ($p<1.0\times10^{-8}$, **Figure 2A**). This pattern was validated in one of the replication samples ($p<1.0\times10^{-8}$ in Webster et al. study),[25] but it should be noted that this replication data includes our discovery sample. We therefore re-analyzed the result using the non-overlapped AD patients, and it showed nominally significant association ($p=0.023$, **Figure 2B**). However, rs6088662 showed an opposite effect on

TRPC4AP expression in our "first replication sample" (risk allele [G] of rs6088662 showed increased gene expression), and in other replication samples, no significant association between rs6088662 and *TRPC4AP* was observed (**Table S6**).[18,26] These inconsistencies may not be surprising, given a prior report of low-to-moderate overlap between eQTL loci across eQTL studies (the percentage of overlapped eQTL is 0~35.4% between pairwise brain studies, shown in Table 4 of McKenzie et al. study).[36] In addition, with the use of several non-brain tissue eQTL databases,[37-39] we also observed significant and consistent associations between rs6088662 and *TRPC4AP* expression (the p-values range from 0.047 to 3.60×10^{-7}, **Figure S1-S3)**.

To further dissect if rs6088662 is also associated with the expression of other nearby genes, we screened 14 genes in the 20q11.22 region in both discovery and replication eQTL samples (**Table S6**). Intriguingly, we observed another gene *GGT7* showing significant association in the discovery sample ($p<1.0\times10^{-7}$, **Figure 3A**), and it remained significant in the "first replication sample" with the same direction of effect ($p<1.0\times10^{-8}$, **Figure 3B**). In other replication samples, the association is also significant (Webster et al.[25] and Zou et al.[18] studies, **Figure 3C** and **Table S6**) or marginal significant (BrainCloud[24]), except for Heinzen et al. study ($p=0.13$),[26] however, in the Heinzen et al. sample, rs6088662 still showed one of the strongest associations with *GGT7* among the genes in 20q11.22, and the SNP showed significant or marginal significant associations with the expression of several exons in *GGT7* (**Table S7**), which was not observed in most of other nearby genes.

For the other genes in 20q11.22, three of them (*ACSS2*, *MYH7B and EDEM2*) also showed associations in some of the eQTL samples, but the associations are not consistent, and these genes are unlikely the associated genes (**Table S6**). To summarize, from the eQTL analyses in both discovery and replication samples, we

demonstrated that rs6088662 is likely an authentic eQTL SNP, and we found two potential genes (*GGT7* and *TRPC4AP*) showing association with this risk SNP.

Rs6088662 is associated with bipolar disorder across cohorts

Given the replication of significant associations between rs6088662 and *TRPC4AP* expression, we opted to pursue further analysis of this SNP on BPD risk. In the stage I replication analysis, rs6088662 was analyzed in nine independent case-control samples. Although the association between rs6088662 and BPD did not achieve even nominal significance (p=0.05) in any single cohort, it does show a trend of association in Germany II and Sweden II samples (p=0.08 and p=0.07, respectively). In a Chinese sample, there is no difference in allele frequencies of this SNP between Han Chinese and Europeans (0.165 versus 0.171 for G allele), and the effect size (OR) in the Chinese sample was even higher than in our discovery sample (1.17 versus 1.12), the non-significant result being likely due to the limited sample size. When all the replication-I samples were combined, the association p-value reached nominal significance level ($p=4.95\times10^{-2}$), with the OR being 1.06 (95% CI=0.99-1.13), consistent with the discovery PGC1 GWAS. There is no significant heterogeneity among the replication-I samples (p=0.77), and detailed results for each individual sample were shown in **Table 1**. The forest plot of the meta-analysis on all replication-I samples is presented in **Figure 4**.

Notably, a previous study[28] has reported a significant association of a proxy SNP of rs6088662 (rs13041792, $r^2=1.00$ with rs6088662 in Europeans) with BPD in an independent UK sample (1,218 cases and 2,913 controls), which is in concordant with our results and was also included in our analysis, denoted as “replication-II” sample. Meta-analysis by combining PGC1 GWAS, replication-I and replication-II samples

yielded a genome-wide significant association of rs6088662 with BPD ($p=3.54\times10^{-8}$, OR=1.12, 95% CI=1.07-1.16, Table 1). We used the fixed effect model for meta-analysis because there was no significant heterogeneity among the samples ($p>0.05$).

Considering the genetic overlap between BPD and other psychiatric disorders [1], we also tested the association of rs6088662 with two other mental disorders, SCZ and major depressive disorder (MDD). It showed nominally significant association with SCZ in the latest PGC2 GWAS (p=0.0037, OR=1.04, 95% CI=1.00-1.08, N=35,476/46,839),[40] however, it did not show any significant associations with MDD when using data from the PGC1 MDD GWAS plus Colaus study samples (10,541/11,208) (**Table S8**),[41,42] implying that rs6088662 is likely a psychosis risk SNP rather than a risk SNP for a broader spectrum of mood disorders.

A proxy search for SNPs in high LD with rs6088662 was performed on the SNAP website (http://www.broadinstitute.org/mpg/snap/ldsearch.php) with the European panel from the 1000 Genomes Project (pilot 1) dataset. This identified 43 SNPs in high LD ($r^2>0.8$) with rs6088662, all of which are located within *MYH7B* and *TRPC4AP* regions (**Figure 5**). Among these, there are 1 non-synonymous SNP, 3 synonymous SNPs, 1 SNP in the 5' untranslated region (UTR), and 1 SNP located in the non-coding RNA (ncRNA) region (**Table S9**). However, to identify causal variants for BPD, further studies are needed.

Rs6088662 is associated with hippocampal volume and cognitive performance

To move beyond statistical association with clinical diagnosis and to obtain convergent evidence for association between rs6088662 and BPD-related biology, we also performed a series of convergent experiments testing risk-associated SNPs on

several intermediate biological phenotypes. The hippocampus is a subcortical brain region frequently reported to show dysfunction among BPD patients.[18,23] We therefore hypothesized that if the identified risk-associated SNP (e.g., rs6088662) affects the anatomy or function of this brain region, then related cognitive deficits, regardless of illness status should be associated with it. In an exploratory manner, we tested the effects of rs6088662 on the biological phenotypes related to the hippocampus (hippocampal volume and cognitive performance) in healthy subjects.

In the ENIGMA sample, rs6088662 was significantly associated with hippocampal volume across multiple cohorts (p=0.00063, β=27.29 mm^3, **Table S10**), supporting the prior speculation that the BPD-associated SNPs will likely affect hippocampal structure, but detailed analysis found that the risk allele [G] led to larger volume. As a *post hoc* exploratory test, we then investigated the potential impacts of rs6088662 on cognitive performance, and found that rs6088662 showed nominally significant association with executive functions (the alert attention task) (p=0.0094, **Table S11**) and language abilities (visual-auditory) (p=0.012, **Table S11**). Again, the risk allele [G] indicated a better cognitive performance.

Analysis on BPD-related phenotypes further confirmed the role of the risk SNPs in BPD susceptibility and implied it may be functional in the brain. However, as the association results on these intermediate phenotypes (especially for cognitive performance) may not survive multiple correction, further validation in larger samples are needed. In addition, the discrepancy of allelic directionality between clinical diagnosis and intermediate phenotypes suggests that the molecular mechanism at work may be more complicated than what we had initially expected when undertaking this study.

DISCUSSION

Findings relating to 20q11.22 region

In this study, with an integrative analysis on both expression and BPD data, we identified a potential risk region 20q11.22 for BPD, although it is still unclear which SNPs are actually responsible. This genomic region contains an extensive area of high LD spanning ~276 kb, including at least 5 protein coding genes (**Figure 5**). Of the 43 common SNPs in high LD (r^2>0.8) with rs6088662, there is one non-synonymous SNP, three synonymous SNPs, one SNP in the 5' UTR area of genes, and one SNP located in the ncRNA region, all of which are potentially functional but as of yet unknown roles (**Table S9**).

We found a nominally significant association of BPD-risk SNPs with hippocampal volume and cognitive performance, which is consistent with the prevalent perspective that many BPD-related genes also affect brain structures and cognitive functions. Rather perplexingly though, the risk allele of rs6088662 actually seemed to be associated with larger hippocampal volume and better cognition, running entirely opposite to the conventional view that risk alleles generally lead to smaller hippocampal volume and worse cognition. One potential speculative explanation is that the risk genes (*GGT7* and/or *TRPC4AP*) may play diverse roles in neural development, and the SNP has pleiotropic effects--some detrimental and some beneficial. Another possible explanation is that gene-behavior association differs by diagnosis status, as previous studies also reported other similar situations: for example, the psychosis risk allele of rs1344706 in *ZNF804A* is associated with better cognitive performance in SCZ patients seen in two independent samples.[43,44] Likewise, another psychosis risk SNP (rs1006737) in *CACNA1C* was shown to be associated with larger gray matter volume for those with the risk allele.[45,46]

Additional evidence of GGT7 and TRPC4AP in bipolar disorder

TRPC4AP is known to be a substrate-specific adapter of a DCX (DDB1-CUL4-X-box) E3 ubiquitin-protein ligase complex required for cell cycle control, and *GGT7* is a member of a gene family that encodes enzymes involved in metabolism of glutathione and in the transpeptidation of amino acids, however, their roles in susceptibility to BPD are still unclear. Here we studied the spatial expression profiling of *GGT7* and *TRPC4AP* in multiple human tissues to see if they are enriched in brain tissues, as BPD is a mental disorder that mainly originates from abnormal brain function, and if these genes are preferentially expressed in brain, which would make more sense of taking them as potential risk genes for BPD. We used the expression data from Genotype-Tissue Expression project (GTEx),[47] in which 3,797 tissues from 150 post-mortem donors have been collected and subsequently analyzed using a RNA sequencing (RNA-seq) based gene expression approach. Notably, we found *GGT7* is abundantly expressed in human brain tissues, such as cerebellum (**Figure S4-A**), while the expression level of *GGT7* is generally low in non-neural tissues. However, the expression of *TRPC4AP* in brains is relatively lower than other tissues (**Figure S4-B**), but this gene has been previously reported in association with Alzheimer's disease (AD),[48,49] a neurological disorder showing a high comorbidity with affective disorders (such as BPD and MDD) in geriatric populations.[50]

Implications

Alongside our specific findings for genetic susceptibility to BPD, our results highlight several advantages of convergent analysis using BPD and eQTL GWAS datasets (**Figure 1**) over conventional analytical strategies aimed at uncovering

susceptibility genes. First, analyses such as ours may identify genes that may be missed by traditional univariate analytical approaches, because these genes tend to be authentic risk genes but with small effects. Second, the identification of eQTL effects of the risk SNPs could provide insights for future focused studies, since conventional analyses often observed a large LD region containing numerous genes showing association with the illness, but actually determining which one is the susceptibility gene is difficult at best. Third, significant association between eQTL and illness has been consistently replicated across independent datasets, providing convergent validity for findings and suggesting potentially higher reproducibility for this kind of system-level analysis. Given these advantages, it is likely that further studies using similar methods will strengthen the case for such studies in trying to uncover genetic risk factors for psychiatric diseases.

Study limitations

While this study offers some interesting observations, it should be noted that the present evidence is limited, and we are cautious in interpreting these results. (1) In the integrative analysis on BPD and eQTL GWAS data, we arbitrarily selected genes that were scored higher than 5.0 (LBF scores). As such, it is possible that some genes that may contribute to BPD risk but did not meet our selection criteria could have been missed. (2) Similarly, while we used GWAS data in our analysis, the SNP coverage is still relatively low and other true risk SNPs may have been missed. Due to the dearth of functional data, it is difficult to identify the causative variant(s). (3) Likewise, we cannot exclude the possibility that the positive association signal was actually caused by the hitch-hiking effect of rare missense mutations, copy number variations or variants in a distant region. Further focused studies may provide a more complete

survey. (4) The SNPs in the discovery eQTL sample were not imputed, thus reducing the overlap between eQTL and GWAS data sets and the power of our method, although we believe the obtained results are valuable. (5) The gene expression coverage in the discovery eQTL dataset is relative low, and we cannot exclude the possibility of other missing risk genes during the integrative analyses, although we conducted a comprehensive replication and fine-mapping analyses to localize the actual risk genes. Further studies using a high-coverage array or RNA-sequencing are warranted. (6) It also should be acknowledged that the eQTL databases that we used are highly variable, in terms of expression platforms and tissue quality, age, and diagnoses. It is highly likely that biological factors mediating eQTL associations, such as epigenetic regulation, transcription factor binding, and microRNA dynamics will vary across age and diagnosis. (7) We also would note that our results reached genome wide significance in the final meta-analysis of our ten new samples added to the public BPD dataset. Our understanding of the association of rs6088662 with BPD and with gene expression and hippocampal biology might have started first with the combined GWAS result, but this was not our strategy.

Conclusions

In conclusion, our data from large-scale samples support that SNPs in a region on chromosome 20q11.22 are significantly associated with BPD. We observed associations with *GGT7* and *TRPC4AP* mRNA expression, hippocampal volume and cognitive performance. Although the actual risk gene(s) for BPD in this genomic region are yet to be determined, future studies may give a more compelling picture on the association between these potential risk factors and genetic susceptibility to BPD.

Supplemental Data

Supplementary material cited in this article is available online.

Funding

This work was supported by grants from the National 973 project of China (2011CBA00401), the National Natural Science Foundation of China (U1202225, 31130051, 31071101 and 31221003), the German Federal Ministry of Education and Research (BMBF), the National Genome Research Network (NGFN), the Integrated Genome Research Network (IG) MooDS (grant 01GS08144 to SC and MMR, grant 01GS08147 to MR and TGS), 111 Project (B07008) of the Ministry of Education of China, the Strategic Priority Research Program (B) of the Chinese Academy of Sciences (XDB02020000), the National Authority for Scientific Research, Bucharest, Romania (UEFISCDI - PN-II-89/2012) and Personal Genetics SRL, Bucharest, Romania.

Acknowledgements

We would like to acknowledge the efforts of the Bipolar Disorder Working Group of Psychiatric GWAS Consortium in their contribution to this study. We are deeply grateful for Stacy Steinberg, Hreinn Stefansson, Kari Stefansson, Thorgeir Thorgeirsson (deCODE genetics, Reykjavik, Iceland) and Engilbert Sigurdsson (Landspitali University Hospital, Reykjavík, Iceland) for their results in Iceland samples. We are also grateful to Andrew Willden (Kunming Institute of Zoology) for language editing of the manuscript.

References

1 Craddock N, Jones I. Genetics of bipolar disorder. *J Med Genet*. 1999; **36**: 585-94.

2 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**: 748-55.

3 Li M, Luo XJ, Rietschel M, Lewis CM, Mattheisen M, Muller-Myhsok B, et al. Allelic differences between Europeans and Chinese for CREB1 SNPs and their implications in gene expression regulation, hippocampal structure and function, and bipolar disorder susceptibility. *Mol Psychiatry*. 2014; **19**: 452-61.

4 Chen DT, Jiang X, Akula N, Shugart YY, Wendland JR, Steele CJ, et al. Genome-wide association study meta-analysis of European and Asian-ancestry samples identifies three novel loci associated with bipolar disorder. *Mol Psychiatry*. 2011; **18**: 195-205.

5 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**: 372-81.

6 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.

7 Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM, Jones L, et al. Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. *Nat Genet*. 2008; **40**: 1056-8.

8 Muhleisen TW, Leber M, Schulze TG, Strohmaier J, Degenhardt F, Treutlein J, et al. Genome-wide association study reveals two new risk loci for bipolar disorder. *Nat Commun*. 2014; **5**: 3339.

9 Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009; **460**: 748-52.

10 Steinberg S, de Jong S, Andreassen OA, Werge T, Borglum AD, Mors O, et al. Common variants at VRK2 and TCF4 conferring risk of schizophrenia. *Hum Mol Genet*. 2011; **20**: 4076-81.

11 Quraishi S, Walshe M, McDonald C, Schulze K, Kravariti E, Bramon E, et al. Memory functioning in familial bipolar I disorder patients and their relatives. *Bipolar Disord*. 2009; **11**: 209-14.

12 Rimol LM, Hartberg CB, Nesvag R, Fennema-Notestine C, Hagler DJ, Jr., Pung CJ, et al. Cortical thickness and subcortical volumes in schizophrenia and bipolar disorder. *Biol Psychiatry*. 2010; **68**: 41-50.

13 Haukvik UKW, L. T.; Mørch-Johnsen, L.; Jørgensen, K. N.; Lange, E. H.; Dale, A. M.; Melle, I.; Andreassen, O. A.; Agartz, I. In vivo hippocampal subfield volumes in schizophrenia and bipolar disorder. *Biol Psychiatry*. 2014; **DOI: 10.1016/j.biopsych.2014.06.020**.

14 Phillips ML, Ladouceur CD, Drevets WC. A neural model of voluntary and automatic emotion regulation: implications for understanding the pathophysiology and neurodevelopment of bipolar disorder. *Mol Psychiatry*. 2008; **13**: 829, 33-57.

15 Frey BN, Andreazza AC, Nery FG, Martins MR, Quevedo J, Soares JC, et al. The role of hippocampus in the pathophysiology of bipolar disorder. *Behav Pharmacol*. 2007; **18**: 419-30.

16 Conde L, Bracci PM, Richardson R, Montgomery SB, Skibola CF. Integrating GWAS and expression data for functional characterization of disease-associated SNPs: an application to follicular

lymphoma. *Am J Hum Genet*. 2013; **92**: 126-30.
17 He X, Fuller CK, Song Y, Meng Q, Zhang B, Yang X, et al. Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS. *Am J Hum Genet*. 2013; **92**: 667-80.
18 Zou F, Chai HS, Younkin CS, Allen M, Crook J, Pankratz VS, et al. Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genet*. 2012; **8**: e1002707.
19 Gamazon ER, Badner JA, Cheng L, Zhang C, Zhang D, Cox NJ, et al. Enrichment of cis-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. *Mol Psychiatry*. 2012; **18**: 340-6.
20 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**: 1494-9.
21 Morita Y, Callicott JH, Testa LR, Mighdoll MI, Dickinson D, Chen Q, et al. Characteristics of the cation cotransporter NKCC1 in human brain: alternate transcripts, expression in development, and potential relationships to brain function and schizophrenia. *J Neurosci*. 2014; **34**: 4929-40.
22 Nakata K, Lipska BK, Hyde TM, Ye T, Newburn EN, Morita Y, et al. DISC1 splice variants are upregulated in schizophrenia and associated with risk polymorphisms. *Proc Natl Acad Sci U S A*. 2009; **106**: 15873-8.
23 Tao R, Li C, Newburn EN, Ye T, Lipska BK, Herman MM, et al. Transcript-specific associations of SLC12A5 (KCC2) in human prefrontal cortex with development, schizophrenia, and affective disorders. *J Neurosci*. 2012; **32**: 5216-22.
24 Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R, Leek JT, et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature*. 2011; **478**: 519-23.
25 Webster JA, Gibbs JR, Clarke J, Ray M, Zhang W, Holmans P, et al. Genetic control of human brain transcript expression in Alzheimer disease. *Am J Hum Genet*. 2009; **84**: 445-58.
26 Heinzen EL, Ge D, Cronin KD, Maia JM, Shianna KV, Gabriel WN, et al. Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS Biol*. 2008; **6**: e1.
27 Zhang X, Zhang C, Wu Z, Wang Z, Peng D, Chen J, et al. Association of genetic variation in CACNA1C with bipolar disorder in Han Chinese. *J Affect Disord*. 2013; **150**: 261-5.
28 Green EK, Hamshere M, Forty L, Gordon-Smith K, Fraser C, Russell E, et al. Replication of bipolar disorder susceptibility alleles and identification of two novel genome-wide significant associations in a new bipolar disorder case-control sample. *Mol Psychiatry*. 2013; **18**: 1302-7.
29 Stein JL, Medland SE, Vasquez AA, Hibar DP, Senstad RE, Winkler AM, et al. Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet*. 2012; **44**: 552-61.
30 Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999; **55**: 997-1004.
31 Scott LJ, Muglia P, Kong XQ, Guan W, Flickinger M, Upmanyu R, et al. Genome-wide association and meta-analysis of bipolar disorder in individuals of European ancestry. *Proc Natl Acad Sci U S A*. 2009; **106**: 7501-6.
32 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**: 128-31.
33 Breen G, Lewis CM, Vassos E, Pergadia ML, Blackwood DH, Boomsma DI, et al. Replication of association of 3p21.1 with susceptibility to bipolar disorder but not major depression. *Nat Genet*. 2011; **43**: 3-5; author reply

34 Kondo K, Ikeda M, Kajio Y, Saito T, Iwayama Y, Aleksic B, et al. Genetic variants on 3q21 and in the Sp8 transcription factor gene (SP8) as susceptibility loci for psychotic disorders: a genetic association study. *PLoS One*. 2013; **8**: e70964.
35 Vassos E, Steinberg S, Cichon S, Breen G, Sigurdsson E, Andreassen OA, et al. Replication study and meta-analysis in European samples supports association of the 3p21.1 locus with bipolar disorder. *Biol Psychiatry*. 2012; **72**: 645-50.
36 McKenzie M, Henders AK, Caracella A, Wray NR, Powell JE. Overlap of expression quantitative trait loci (eQTL) in human brain and blood. *BMC Med Genomics*. 2014; **7**: 31.
37 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**: 1246-50.
38 Nica AC, Parts L, Glass D, Nisbet J, Barrett A, Sekowska M, et al. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet*. 2011; **7**: e1002003.
39 Stranger BE, Montgomery SB, Dimas AS, Parts L, Stegle O, Ingle CE, et al. Patterns of cis regulatory variation in diverse human populations. *PLoS Genet*. 2012; **8**: e1002639.
40 Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014; **511**: 421-7.
41 Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, Breen G, et al. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry*. 2012; **18**: 497-511.
42 Preisig M, Waeber G, Vollenweider P, Bovet P, Rothen S, Vandeleur C, et al. The PsyCoLaus study: methodology and characteristics of the sample of a population-based survey on psychiatric disorders and their association with genetic and cardiovascular risk factors. *BMC Psychiatry*. 2009; **9**: 9.
43 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.
44 Chen M, Xu Z, Zhai J, Bao X, Zhang Q, Gu H, et al. Evidence of IQ-modulated association between ZNF804A gene polymorphism and cognitive function in schizophrenia patients. *Neuropsychopharmacology*. 2012; **37**: 1572-8.
45 Wang F, McIntosh AM, He Y, Gelernter J, Blumberg HP. The association of genetic variation in CACNA1C with structure and function of a frontotemporal system. *Bipolar Disord*. 2011; **13**: 696-700.
46 Perrier E, Pompei F, Ruberto G, Vassos E, Collier D, Frangou S. Initial evidence for the role of CACNA1C on subcortical brain morphology in patients with bipolar disorder. *Eur Psychiatry*. 2011; **26**: 135-7.
47 Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013; **45**: 580-5.
48 Poduslo SE, Huang R, Huang J. The frequency of the TRPC4AP haplotype in Alzheimer's patients. *Neurosci Lett*. 2009; **450**: 344-6.
49 Poduslo SE, Huang R, Huang J, Smith S. Genome screen of late-onset Alzheimer's extended pedigrees identifies TRPC4AP by haplotype analysis. *Am J Med Genet B Neuropsychiatr Genet*. 2009; **150B**: 50-5.
50 Teipel SJ, Walter M, Likitjaroen Y, Schonknecht P, Gruber O. Diffusion tensor imaging in Alzheimer's disease and affective disorders. *Eur Arch Psychiatry Clin Neurosci*. 2014.

Table 1. Summary of logistic regression results for rs6088662 across cohorts.

Sample	Ethnicity	Cases	Controls	Effect Allele	Additive *P*-value	Odds ratio	95% CI	Data source
Discovery								
PGC1	Europeans	7,481	9,250	G	**5.85×10^{-5}**	1.12	1.06-1.19	[6]
Replication-I								
Germany II	German	181	527	G	0.08	1.23	0.91-1.65	this study
Germany III	German	490	880	G	0.16	1.09	0.90-1.33	this study
Australia	Australian	330	1,811	G	0.29	1.07	0.85-1.33	this study
France	French	451	1,631	G	0.42	1.02	0.85-1.22	this study
Sweden I	Swedish	836	2,093	G	0.37	1.02	0.89-1.18	this study
Sweden II	Swedish	1,415	1,271	G	0.07	1.12	0.97-1.29	this study
Iceland	Icelandic	541	34,426	G	0.19	0.93	0.79-1.10	this study
Romania	Romanian	244	174	G	0.42	1.04	0.74-1.46	this study
China	Han Chinese	350	888	G	0.11	1.17	0.91-1.50	this study
All replication-I samples		4,838	43,701	G	**4.95×10^{-2}**	1.06	0.99-1.13	
Replication-II								
UK	British	1,218	2,913	G	**1.06×10^{-6}**	1.34	1.19-1.51	[28]
Discovery + Replications		13,537	55,864	G	**3.54×10^{-8}**	1.12	1.07-1.16	

Test of heterogeneity
All replication-I cohorts: p=0.77, I^2=0%; meta-analysis was conducted under fixed effect model
Discovery + Replication samples: p=0.07, I^2=41.8%; meta-analysis was conducted under fixed effect model

Abbreviations:
CI, confidence interval.

Note:
P-values are two-sided for the discovery cohort and combined analysis; one-sided *P*-values are listed for the replication-I samples

FIGURE LEGENDS

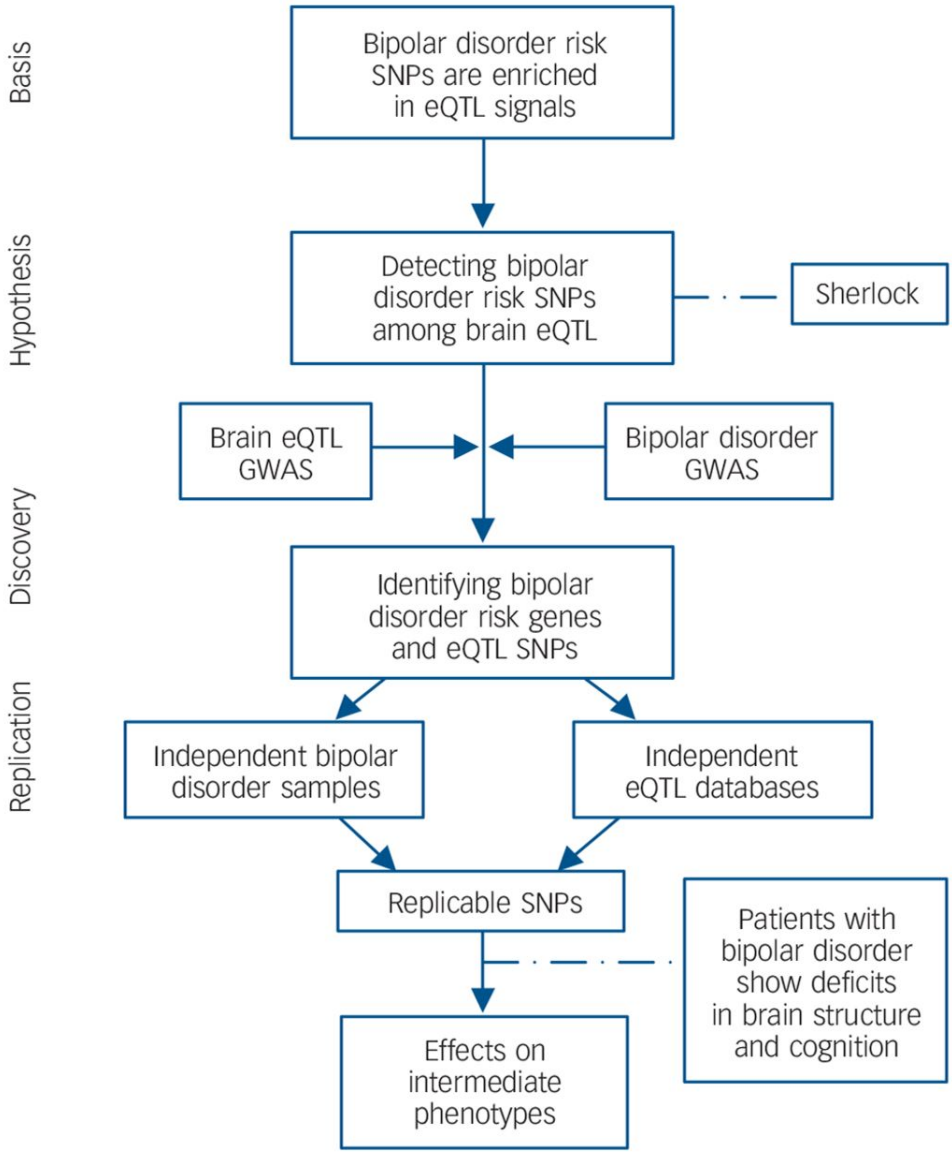


Figure 1. Flow chart of the present study.

Based on the hypothesis that BPD risk variants are enriched among eQTL, we systematically integrated BPD GWAS and genome-wide brain eQTL data by using *Sherlock*. The top genes identified by *Sherlock* were then replicated in independent BPD samples and eQTL datasets. Finally, the successfully replicated SNP (rs6088662) was further tested for the association with BPD biological phenotypes including hippocampal volume and cognitive performance.

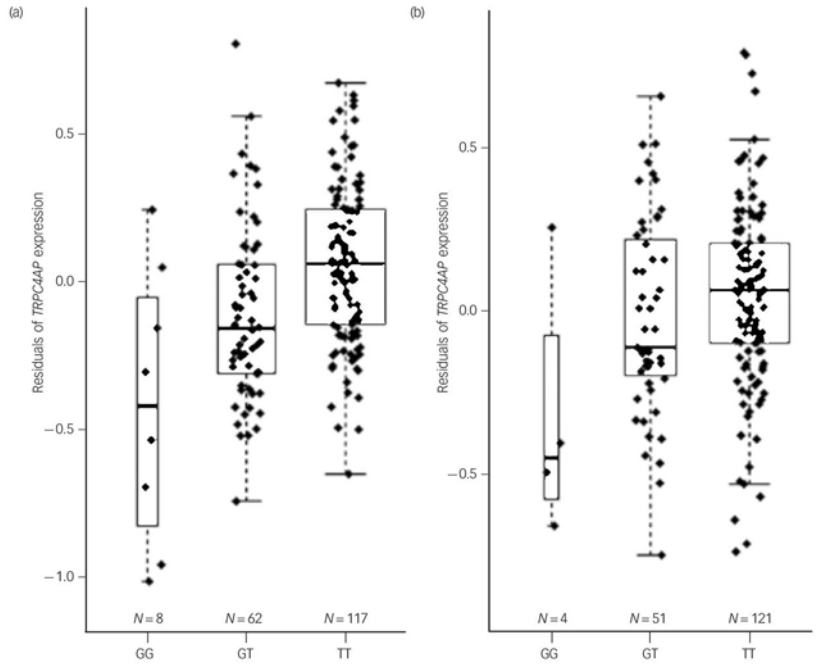


Figure 2. Rs6088662 is significantly associated with *TRPC4AP* mRNA expression.

(A) Results in 193 neuropathologically normal human brain (cortical) samples of European subjects. (B) Results in 176 Alzheimer's disease human brain (cortical) samples of European subjects.

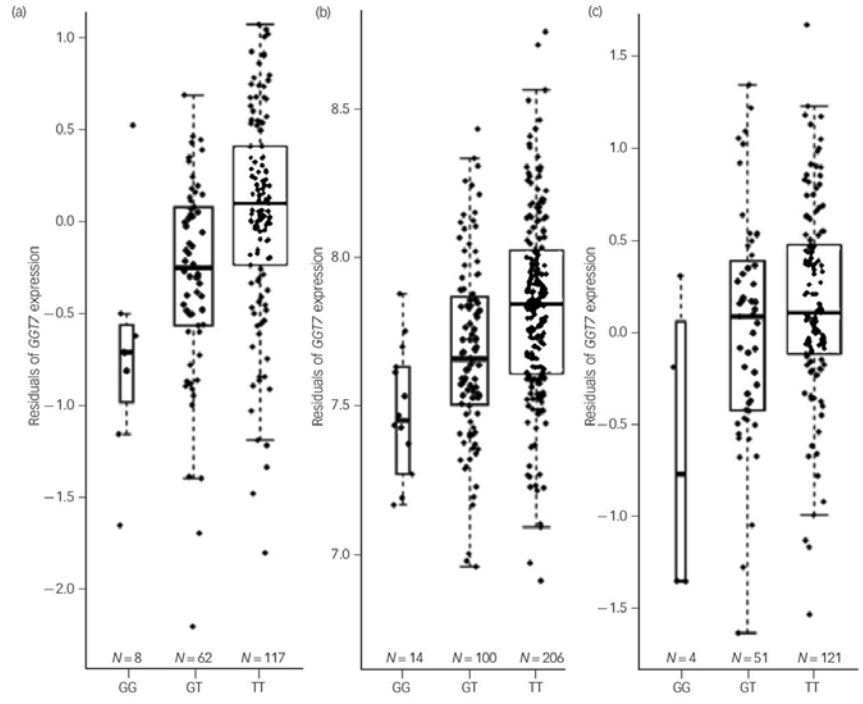


Figure 3. Rs6088662 is significantly associated with *GGT7* mRNA expression.

(A) Results in 193 neuropathologically normal human brain (cortical) samples of

European subjects. (B) Results in 320 healthy human brain DLPFC samples of Caucasian and African American individuals. (C) Results in 176 Alzheimer's disease human brain (cortical) samples of European subjects.

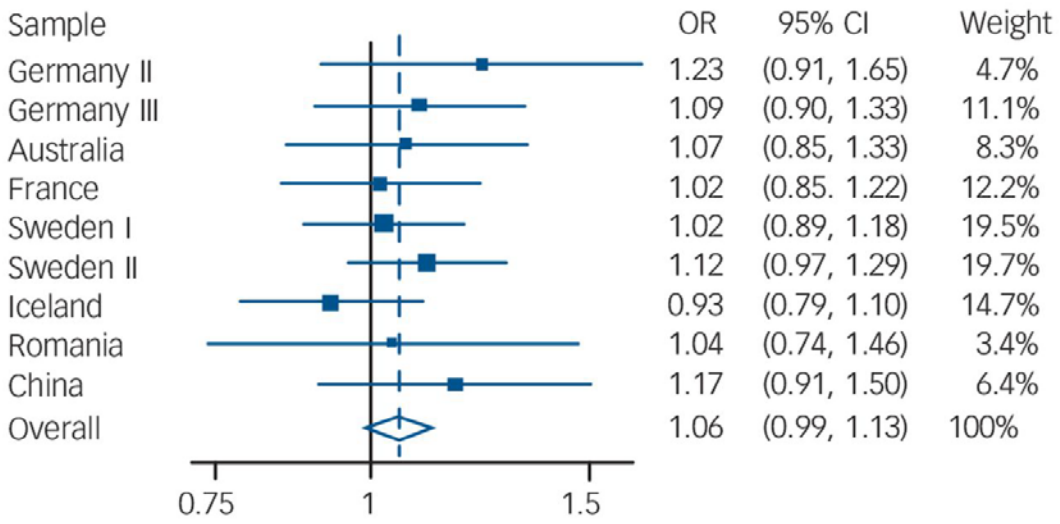


Figure 4. Forest plot of odds ratios with 95% confidence interval for total replication-I bipolar disorder samples included in meta-analysis of rs6088662. The G allele of rs6088662 is overrepresented in BPD cases in all of the tested cohorts (except for the Iceland sample).

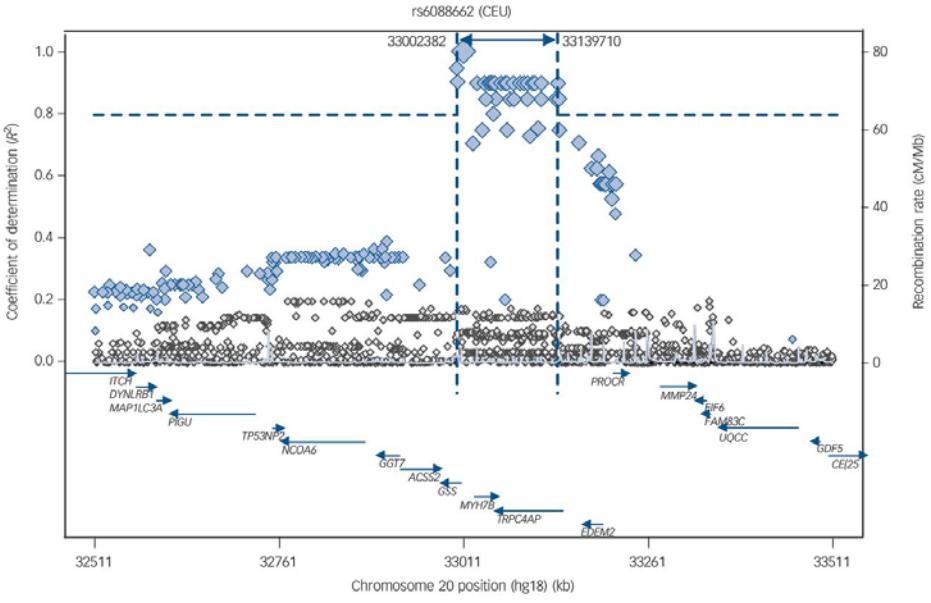


Figure 5. Plot of chromosome region showing a genomic area of high linkage disequilibrium with rs6088662 in European populations.

Online supplement to Ming Li, Xiong-jian Luo, *et al*. Impact of a *cis*-associated gene expression SNP on chromosome 20q11.22 on bipolar disorder susceptibility, hippocampal structure and cognitive performance. Br J Psychiatry doi: 10.1192/bjp.bp.114.156976

Additional members from the Bipolar Disorder Group of the MooDS Consortium

Jana Strohmaier,[1] René Breuer,[1] Sandra Meier,[1] Per Hoffmann,[2-5] Stefan Herms,[2-4] Markus Schwarz,[6] Helmut Vedder,[6] Jutta Kammerer-Ciernioch,[6] Andreas Reif,[7] Johanna Sasse,[8] Michael Bauer,[8] Sandra Zwick,[8] Martin Hautzinger,[9] Piotr M. Czerski,[10] Joanna Hauser,[10] Johannes Schumacher,[3] Wolfgang Maier,[11] Peter Propping[3]

[1]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, University of Heidelberg, Mannheim, Germany
[2]Division of Medical Genetics, University of Basel, Basel, Switzerland
[3]Institute of Human Genetics, University of Bonn, Bonn, Germany
[4]Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany
[5]Institute of Neuroscience and Medicine (INM-1), Structural and Functional Organization of the Brain, Genomic Imaging, Research Centre Jülich, Jülich, Germany
[6]Psychiatric Center Nordbaden, Wiesloch, Germany
[7]Department of Psychiatry, University of Würzburg, Würzburg, Germany
[8]Department of Psychiatry and Psychotherapy, University Hospital Carl Gustav Carus, Dresden, Germany
[9]Department of Clinical and Developmental Psychology, Institute of Psychology, University of Tübingen, Tübingen, Germany
[10]Department of Psychiatry, Poznan University of Medical Sciences, Poznan, Poland
[11]Department of Psychiatry, University of Bonn, 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

Description about discovery eQTL database

The brain eQTL dataset used in this study was reported previously.[1] In brief, human cortex samples were collected from the National Institute of Aging Alzheimer Centers and the Miami Brain Bank and the original subjects met several criteria: (a) self-defined as being ethnically of European descent; (b) had no clinical history of stroke, cerebrovascular disease, Lewy bodies or co-morbidity with neurological disease; (c) were assessed by board certified neurologists who made a determination on their condition; and (d) had an age at death greater than 65 years. After excluding ethnic outliers and samples that were possibly related, a total of 193 independent subjects' samples remained for subsequent analysis.

Genotyping of the 193 cortex samples was conducted using Affymetrix GeneChip Human Mapping 500K Array Set, and mRNA expression measurements were performed using Illumina HumanRefseq-8 Expression BeadChip using standard manufacturer's protocols. The PLINK program was used to carry out as a one-degree-of-freedom allelic test of association, and the associations results were further separated into *cis* and *trans* significantly associated SNP-transcript pair sets. *Cis* SNPs were defined as SNPs within either 1 Mb of the 5' or 3' end of the transcript and within the transcript. Sherlock considers both *cis* and *trans* eQTL SNPs. Detailed information about genotyping and expression profiling as well as statistical methods can be found in the original publication.[1]

Description about non-brain tissue replication eQTL databases

The non-brain tissue eQTL databases were retrieved through Genevar,[2] which have ever been reported by Nica *et al*,[3] Dimas *et al*,[4] and Stranger *et al*[5] In brief, Nica *et al* explored in depth the roles of genetic variation on gene expression in three human tissues: lymphoblastoid cell lines (LCL), skin, and adipose, and the samples (156 LCL, 160 skin, 166 adipose) derived simultaneously from a subset of healthy female twins of the MuTHER resource;[3] Dimas *et al* conducted the genome-wide expression analysis in three types of cells (fibroblast, LCL and T-cell) from 75 Geneva GenCord Caucasian individuals;[4] Stranger *et al* analyzed genome-wide gene expression in LCL from 8 global populations of the HapMap3 project and correlated gene expression levels with HapMap3 SNPs located in cis to the genes. We used the data from the Caucasian samples (N=109) reported by Stranger *et al*[5] In these three datasets, all the statistical analysis between SNPs and gene expression were conducted using Spearman's correlation.

Replication-I sample information (see Table DS1)

Germany II and III sample

Cases for Germany II and III samples were again 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 bipolar disorder were assigned using a consensus best-estimate procedure, based on all available information, including structured interviews (SCID-I, SADS-L; Germany III) or semi-structured interviews (AMDP; Germany II), medical records, and the family history method. In addition, the OPCRIT system was used for the detailed polydiagnostic documentation of symptoms.[6]

Controls for Germany II were ascertained from the population-based Heinz Nixdorf Recall Study.[7] Study protocols were reviewed and approved in advance by Institutional Review Boards of the participating institutions. All subjects provided written informed consent. This includes a clause that all data may be shared with collaborating partners such as the PGC. However, consents do not include permission for depositing of de-identified individual GWAS genotype and phenotype data into the NIMH genetics initiative repository, although these data may be used in specific collaborations for studies of neuropsychiatric disorders. All subjects were genotyped using the Illumina platform.

The controls for Germany III were recruited at the Max Planck Institute of Psychiatry in Munich, Germany, and were selected randomly from a Munich-based community sample. They were collected in the course of genetic studies of major depression, and were therefore screened for the presence of anxiety and affective disorders using the Composite International Diagnostic Screener (WHO-CIDI). Only individuals negative for the above-named disorders were included in the sample. All included controls were Caucasian, 93.04% were of German origin. These subjects thus represent a group of healthy individuals with regard to depression and anxiety. The study was approved by the ethics committee of the Ludwig Maximilians University in Munich, Germany, and written informed consent was obtained from all subjects.

Australia sample

Subjects were ascertained through two studies: (a) a bipolar disorder pedigree sample (described in McAuley *et al*)[8] and (b) a specialized Sydney Black Dog Institute bipolar disorder clinic sample (described in Mitchell *et al* 2009).[9] 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 bipolar disorder 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 bipolar disorder 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 bipolar disorder 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.

Australia 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 bipolar disorder. 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.

France sample

Patients with bipolar disorder and controls were recruited as part of a large study on genetics of bipolar disorder 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[10] 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).

Sweden I 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 bipolar disorder, and the study was approved by the Regional Ethics Committee of Stockholm. Diagnoses were based on physician administered ADE[11] and MINI.[12]

Bipolar disorder 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 bipolar disorder 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.[13] 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 bipolar disorder cases were recruited from the Stockholm County catchment area. All patients provided written informed consent to participate in a genetic study of bipolar disorder, and the study was approved by the Regional Ethics Committee of Stockholm. Diagnoses were made according to the DSM-IV criteria.

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

Sweden II sample

This sample consisted of 1415 patients with bipolar disorder (62.5% female, age ± s.d. = 53 ± 14, bipolar disorder type I =578, bipolar disorder type II= 517, NOS=281, SAB = 39, unknown subtype = 4), and 1271healthy controls (50.3% female, age ± s.d.= 59 ± 11 years). All subjects were unrelated to each other and ethnically Swedish. Patients with bipolar disorder were collected from the Swedish National Quality Assurance Registry for bipolar disorder (BipoläR), to which all patients with a DSM-IV diagnosis of bipolar I, II, NOS, or schizoaffective disorder are considered for registration at the participating clinics.[14] There were no other inclusion or exclusion criteria. Diagnoses were made by the treating physician with longitudinal access to all available clinical information. Controls were also identified from national population registers, and had never received a discharge diagnosis of SCZ or bipolar disorder. Controls were contacted directly in a similar procedure as the cases, gave written informed consent, were interviewed about other medical conditions and visited their family doctor or local hospital laboratory for blood donation. Patients and controls were genotyped on the Illumina Omni Express array, and the genomic inflation factor (lambda) is 1.03.

Iceland sample

The Iceland sample consisted of 541 subjects with bipolar disorder 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 bipolar disorder diagnoses and nine subjects with DSM-III bipolar disorder 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.

Romania 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 Version (SCID-I, 1994) and the Diagnostic Interview for Genetic Studies (DIGS) version 3.0 (1999). Information provided by 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 sample was population-based, drawn from the same population as the patients, and was screened for major psychiatric disorders. The

ethnicity of the patients and control subjects was determined by genealogical investigation to the grandparental generation. Only the patient sample was previously reported in other collaborative studies.[15-17] The 174 controls were genotyped on Illumina OMNI-Express chips in Bonn, and the patients were also genotyped on Illumina chips (partly on Quad Omni-1).

China sample

The patients who met DSM-IV criteria for bipolar disorder type 1or type 2 were recruited from the Division of Mood Disorders at Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine between November 2006 and October 2010. Each patient was independently interviewed and diagnosed by a consensus of at least two experienced psychiatrists. Diagnoses were further confirmed with an Extensive Clinical Interview and a Structured Clinical Interview for DSM-IV Axis/Disorders, Patient Version (SCID-P) given by a research psychiatrist. Subjects with comorbid diagnosis of other psychiatric disorders or chronic physical illness were excluded in this study to mitigate the potential for compounding factors during our analysis. The Extensive Clinical Interview contains items to assess demographics, mental status, and ages at onset for the bipolar disorder patients. To avoid the biases due to the low reliability of retrospective evaluation of prodromal symptoms, we defined age at onset as the first reliably diagnosed hypo/manic or depressive episode according to DSM-IV criteria.

Control subjects were enrolled from hospital staff and students of the School of Medicine in Shanghai that were interviewed by a specialized psychiatrist with SCID-P. Subjects with any psychiatric disorder and chronic physical disease were excluded from our analysis. All subjects were of Han Chinese origin and provided written informed consent before any study-related procedures were performed. This sample has been reported in a previous study.[18]

Replication-II sample information (see Table DS1)

UK sample

The cases consisted of 1218 individuals of which 29% were male. The mean age of recruitment was 46 (s.d.=12) years, with a mean age at first impairment because of bipolar disorder of 22 (s.d.=9) years. A lifetime diagnosis was made according to Research Diagnostic Criteria and the 1218 individuals were categorized as follows: bipolar I disorder/mania: 63% cases, bipolar II disorder/hypomania: 29% cases, schizoaffective disorder, bipolar type: 8% cases. Of those individuals for whom we were able to make a definite rating, 65% of the cases had a lifetime experience of psychotic symptoms (defined as a score over 9 on the Bipolar Affective Disorder Dimension Scale (BADDS)) and 25% had a lifetime experience of predominantly mood-incongruent psychotic symptoms (defined as a score over 29 on the BADDS mood incongruence scale). There were 2913 controls in the independent sample, of which 47% were male. This sample has been reported in a previous study.[19]

Analysis of hippocampal volume and cognitive performance

To analyse hippocampal volume, we used the data from a recent large-scale GWAS conducted by the Enhancing Neuro Imaging Genetics through Meta-Analysis (ENIGMA) consortium.[20] The GWAS comprised 17 samples of European ancestry of which genome-wide SNP data and hippocampal volume data were collected, including a total of 5,775 young healthy individuals (mean age: 34.8 years). Evidence for potential association was assessed using the allelic dosage of the SNP and covariates controlling for population stratification (four MDS components), intracranial volume, age, age^2, sex and the interactions between age and sex, and age^2 and sex. Detailed information on the samples, imaging procedures, genotyping methods and statistical analysis can be found in the original GWAS report.[20]

For cognitive analysis, we utilized a Chinese sample that included 342 healthy Chinese college students from Beijing Normal University who had self-reported no known history of any neurological or psychiatric disorders (197 females and 145 males, aged 18-23). Cognitive and behavioral measures included working memory, executive functions (as assessed with the Attention Network Test, the Wisconsin Card Sorting Task, and a reversal learning test), and motivation traits etc. Detailed cognitive functions examined in this study are listed in Table DS2. This cognitive sample was previously used in several studies and shown to be effective in detecting authentic risk effects.[21-23] Genotyping was performed by Affymetrix 6.0 array using standard protocols. Since homozygotes for the rs6088662 minor allele (GG) are rare in this sample, we combined GG and GA genotypes as a single group denoted 'G carrier', and statistical analysis using two-tailed t-test was done with SPSS 16.0 (SPSS, Chicago, USA). This particular experiment was approved by the Institutional Review Board of the State Key Laboratory of Cognitive Neuroscience and Learning at Beijing Normal University, China. Written consent form was obtained from all participants following a full explanation of the study procedure.

Table DS1 Description of individual samples included in this study

Sample	Cases	Case diagnosis	Diagnosis	Interview	Controls	Genotyping	λ	Ref.
Discovery								
PGC1	7,481	BPD1,BPD2,SAB,BPD-NOS	DSMIIR,DSM-IV,RDC	multiple	9,250	multiple	1.15	24
Replication-I								
Germany II	181	BPD1,BPD2	DSM-IV	AMDP	527	Illumina	1.05	17,25
Germany III	490	BPD1,BPD2,SAB,BPD-NOS	DSM-IV	SCID-I,SADS-L	880	Illumina	1.00	17,25
Australia	330	BPD1,BPD2,SAB,BPD-NOS	DSM-IV	SCID,DIGS	1,811	Illumina	1.00	9,26
France	451	BPD1,BPD2,BPD-NOS	DSM-IV	DIGS	1,631	Illumina	1.03	27,28
Sweden I	836	BPD1,BPD2,BPD-NOS	DSM-IV	ADE,MINI	2,093	Affymetrix 6.0	1.07	27
Sweden II	1,415	BPD1,BPD2,SAB,BPD-NOS	DSM-IV	/	1,271	Affymetrix 6.0	1.03	14
Iceland	541	BPD1,BPD2,BPD-NOS	DSM-IV,ICD-10,	CID-I,SADS-L	34,426	Affymetrix 6.0	1.11	17
Romania	244	BPD1	DSM-IV	SCID-I-P/DIGS	174	Illumina	/	15-17
China	350	BPD1,BPD2	DSM-IV	SCID-P	888	SNaPShot	/	18
Total	**4,838**				**43,701**			
Replication-II								
UK	1,218	BPD1,BPD2,SAB	RDC	/	2,913	ImmunoChip	1.02	19
Grand Total	**13,537**				**55,864**			

BPD1, bipolar disorder type 1; BPD2, bipolar disorder type 2; BPD-NOS, bipolar disorder not otherwise specified; SCZ, schizophrenia; SAB, schizoaffective disorder (bipolar type); λ = genomic control lambda.

We primarily used the Illumina (San Diego, CA, USA), Affymetrix and SNaPShot platforms to genotype rs6088662. For the genotyping in UK and Romania samples we used proxy SNP rs13041792 in UK (r^2=1.00 with rs6088662 in Europeans using data from 1000-Human-Genome) and rs6088667 in Romania samples (r^2=0.90 with rs6088662) instead, as rs6088662 is not covered.

Table DS2 Cognitive performance assessment in Chinese sample

Domain	Task	Brief description	Index
Memory	Wechsler Memory Scale -3rd Edition (WMS-III)	Two subscales: Picture recall (Subjects were showed pictures of 20 simple objects for 30 seconds and then asked to recall as many as possible) and picture recognition (Subjects were showed pictures of 8 simple objects for 30 seconds and then asked to pick them out from 28 pictures).	Number of items correctly recalled or recognized
	Working memory	In the 2-back working memory task, subjects judged whether the current item was the same (or related) to the one presented two trials earlier. Three sessions involved morphological, phonological, and semantic judgment.	Overall accuracy
Executive function	Attention network test	Subjects saw several small arrows on the computer screen and had to judge the direction of the arrow in the middle (left or right). The 6 peripheral arrows can either in the same or inverse direction to the middle one. There were also cues to alert subjects or point to the position where arrows will be presented	Alert, orientation, conflict
	Wisconsin card sort task	Subjects had to select one from four cards that fits a rule. Rules included color, form, and amount of items on the cards, and rules changed during the experiment	Preserved error (Nelson)
Personality	Temperament and Character Inventory-Re	7 aspects of personality: Novelty Seeking, Harm Avoidance, Reward Dependence, Persistence, Self-Directedness, Cooperativeness, Self-Transcendence	7 subscales scores
Language abilities	Visual-auditory learning, from Woodcock Reading Mastery test Revised, Forms G.	This task consists of several sessions. In each session, subjects were asked to learn a few symbol-word pairs. Afterwards, they were asked to read out some sentences written in symbols using corresponding words they just learned.	Number of correct responses

Table DS3 Results of integrative analysis using brain eQTL and bipolar disorder GWAS data

Gene	Gene LBF	Gene p-val	SNP	Location	Proximity	eQTL p-val	BPD p-val	SNP LBF
GLT8D1	6.78	2.22e-06						
			rs2251219	3:52559827	cis	2.84e-17	5.45e-07	6.95
			rs17073273	6:144330243	trans	8.55e-06	0.73	-0.093
			rs2070968	10:73251566	trans	6.20e-06	0.67	-0.075
CXCL16	6.16	2.22e-06						
			rs12634640	3:187552259	trans	6.92e-07	1.63e-03	2.38
			rs810517	10:80612626	trans	1.14e-06	2.12e-04	3.78
TRPC4AP	5.57	8.89e-06						
			rs9883745	3:133715013	trans	2.86e-06	0.46	-0.13
			rs10501340	11:55439371	trans	6.73e-06	2.75e-02	0.28
			rs11049310	12:28100068	trans	8.31e-06	0.77	-0.056
			rs6088662	20:33011294	cis	5.44e-09	5.85e-05	5.48
TAF11	5.52	1.11e-05						
			rs4482754	4:87230328	trans	7.57e-06	3.62e-06	4.40
			rs7263316	20:19632036	trans	5.94e-07	2.37e-02	1.12

Table DS4 Association of rs2251219 with gene expression in 3p21.1 region

Author (Ref.)	**Myers *et al*[1]**	**Unpublished data**	**Colantuoni *et al*[29]**	**Webster *et al*[30]**		**Zou *et al*[31]**		**Heinzen *et al*[32]**
Region	Brain	DLPFC	DLPFC	Brain		Cerebellar	Temporal cortex	Frontal cortex
Dx (Number)	Control (N=193)	Control (N=320)	Control (N=261)	Control + AD (N=369)	AD (N=176)	AD (N=197)	AD (N=202)	Control (N=93)
STAB1	n.s.	/	n.s.	n.s.	n.s.	n.s.	n.s.	*0.095*
NT5DC2	/	/	n.s.	/	/	n.s.	n.s.	n.s.
PBRM1	/	n.s.	*0.085*	/	/	n.s.	n.s.	n.s.
GNL3	/	/	*0.017*	/	/	n.s.	n.s.	/
GLT8D1	**$<1.0\times10^{-16}$**	n.s.	*0.048*	**$<1.0\times10^{-30}$**	**$<1.0\times10^{-18}$**	n.s.	n.s.	n.s.
SPCS1	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
NEK4	/	*0.051*	n.s.	/	/	n.s.	n.s.	n.s.
ITIH1	/	/	n.s.	/	/	n.s.	n.s.	*0.055*
ITIH3	/	/	n.s.	/	/	n.s.	n.s.	n.s.
ITIH4	**$<1.0\times10^{-3}$**	/	**$<1.0\times10^{-2}$**	**$<1.0\times10^{-4}$**	**$<1.0\times10^{-2}$**	n.s.	n.s.	n.s.
TMEM110	/	*0.083*	/	/	/	n.s.	n.s.	n.s.

N.A., not available; Dx, diagnosis; AD, Alzheimer's disease; n.s., not significant; DLPFC, Dorsolateral prefrontal cortex

Table DS5 Replication of *trans* eQTL association in different samples

Author (Ref.)		**Myers *et al*[1]**	**Unpublished data**	**Colantuoni *et al*[29]**	**Webster *et al*[30]**		**Zou *et al*[31]**		**Heinzen *et al*[32]**
Region		Brain	DLPFC	DLPFC	Brain		Cerebellar	Temporal cortex	Frontal cortex
Dx (Number)		Control (N=193)	Control (N=320)	Control (N=261)	Control + AD (N=369)	AD (N=176)	AD (N=197)	AD (N=202)	Control (N=93)
CXCL16	rs12634640	$<1.0\times10^{-6}$	n.s.	n.s.	$<1.0\times10^{-2}$	n.s.	n.s.	n.s.	n.s.
	rs810517	$<1.0\times10^{-5}$	n.s.	n.s.	0.023	n.s.	n.s.	n.s.	n.s.
TAF11	rs4482754	$<1.0\times10^{-5}$	/	n.s.	$<1.0\times10^{-2}$	n.s.	n.s.	n.s.	n.s.

Dx, diagnosis; AD, Alzheimer's disease; n.s., not significant; DLPFC, Dorsolateral prefrontal cortex

Table DS6 Association of rs6088662 with gene expression in 20q11.22 region

Author (Ref.)	**Myers *et al*[1]**	**Unpublished data**	**Colantuoni *et al*[29]**	**Webster *et al*[30]**		**Zou *et al*[31]**		**Heinzen *et al*[32]**
Region	Brain	DLPFC	DLPFC	Brain		Cerebellar	Temporal cortex	Frontal cortex
Dx (Number)	Control (N=193)	Control (N=320)	Control (N=261)	Control + AD (N=369)	AD (N=176)	AD (N=197)	AD (N=202)	Control (N=93)
AHCY	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
ITCH	n.s.	*0.067*	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
DYNLRB1	/	n.s.	n.s.	/	/	n.s.	n.s.	n.s.
PIGU	/	n.s.	n.s.	/	/	n.s.	n.s.	n.s.
NCOA6		n.s.	*0.036*	n.s.	n.s.	n.s.	n.s.	n.s.
GGT7	**$<1.0\times10^{-7}$**	**$<1.0\times10^{-8}$**	*0.054*	**$<1.0\times10^{-7}$**	**$<1.0\times10^{-2}$**	n.s.	**$<1.0\times10^{-2}$**	0.13
ACSS2	/	n.s.	*0.098*	/	/	**$<1.0\times10^{-2}$**	**$<1.0\times10^{-5}$**	n.s.
GSS	n.s.	/	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
MYH7B	n.s.	n.s.	**$<1.0\times10^{-3}$**	n.s.	n.s.	**$<1.0\times10^{-15}$**	n.s.	n.s.
TRPC4AP	**$<1.0\times10^{-8}$**	<0.005	/	**$<1.0\times10^{-8}$**	*0.023*	n.s.	n.s.	n.s.
EDEM2	*0.040*	*0.067*	**$<1.0\times10^{-2}$**	**$<1.0\times10^{-2}$**	*0.010*	n.s.	n.s.	n.s.
PROCR	/	/	n.s.	/	/	n.s.	n.s.	n.s.
MMP24	/	*0.098*	n.s.	/		n.s.	n.s.	0.11
UQCC1	/	n.s.	n.s.	/	/	n.s.	n.s.	n.s.

N.A., not available; Dx, diagnosis; AD, Alzheimer's disease; n.s., not significant; DLPFC, Dorsolateral prefrontal cortex

Table DS7 Association of rs6088662 with GGT7 exon expression in Heinzen *et al*[32]

Probe_Type	Gene_Symbol	Transcript	Transcript_Probe_ID	Exon_Probe_ID	Start	End	P-val
transcript	*GGT7*	NM_178026	3903598	-	32884010	32924318	0.128
exon	*GGT7*	NM_178026	3903598	3903603	32896517	32896683	0.2386
exon	*GGT7*	NM_178026	3903598	3903604	32896819	32896919	0.4164
exon	*GGT7*	NM_178026	3903598	3903606	32901439	32901524	**0.02122**
exon	*GGT7*	NM_178026	3903598	3903610	32902247	32902276	0.2765
exon	*GGT7*	NM_178026	3903598	3903611	32902349	32902375	0.2827
exon	*GGT7*	NM_178026	3903598	3903613	32902716	32902761	0.2265
exon	*GGT7*	NM_178026	3903598	3903614	32902789	32902817	0.2122
exon	*GGT7*	NM_178026	3903598	3903616	32903626	32903723	*0.07463*
exon	*GGT7*	NM_178026	3903598	3903618	32903901	32903942	*0.08682*
exon	*GGT7*	NM_178026	3903598	3903619	32903951	32903990	0.6926
exon	*GGT7*	NM_178026	3903598	3903620	32906001	32906080	0.2708
exon	*GGT7*	NM_178026	3903598	3903621	32906278	32906381	0.5625
exon	*GGT7*	NM_178026	3903598	3903623	32908286	32908357	0.6493
exon	*GGT7*	NM_178026	3903598	3903624	32910932	32911062	0.8358
exon	*GGT7*	NM_178026	3903598	3903626	32911433	32911460	*0.05013*
exon	*GGT7*	NM_178026	3903598	3903629	32911733	32911766	0.1856
exon	*GGT7*	NM_178026	3903598	3903633	32912864	32912890	0.1768
exon	*GGT7*	NM_178026	3903598	3903634	32912953	32913023	0.5577
exon	*GGT7*	NM_178026	3903598	3903636	32914307	32914408	0.2789
exon	*GGT7*	NM_178026	3903598	3903638	32914781	32914926	0.7815
exon	*GGT7*	NM_178026	3903598	3903643	32924120	32924243	0.938
exon	*GGT7*	NM_178026	3903598	3903644	32924280	32924317	0.2398

Table DS8 Association of rs6088662 with schizophrenia and major depression

Disorder	Sample	Cases	Controls	Allele	P-value	Odds ratio	95% CI
Schizophrenia	PGC2[33]	35,476	46,839	G	0.0037	1.04	1.00-1.08
Depression	PGC1[34]	9,240	9,519	G	0.27	1.03	0.98-1.08
	PsyCoLaus study[35]	1,301	1,689	G	0.90	0.99	0.88-1.12

Table DS9 SNPs in the LD area with potentially functional role on genes

SNP	Position	Distance (bp)	R^2	MAF	Function	Gene Name
rs3746444	33041912	30618	0.848	0.20	ncRNA	MIR499A
rs7268266	33045550	34256	0.898	0.20	cds-synon	MYH7B
rs3746436	33049854	38560	0.898	0.20	cds-synon	MYH7B
rs3746435	33050859	39565	0.898	0.20	missense	MYH7B
rs36003887	33052768	41474	0.898	0.20	cds-synon	MYH7B
rs8501	33054245	42951	0.898	0.20	3’ UTR	TRPC4AP

Table DS10 Association of rs6088662 with hippocampal volume in Europeans[20]

SNP	Position	Allele	Frequency	β (mm^3)	SE (mm^3)	P-value
rs6088662	20:33547633	G	0.1937	27.29	7.99	0.00063

SE, standard error; β represents the difference in hippocampal volumes per copy increase of effect allele.

The association analysis in 5,775 healthy European subjects was corrected for intracranial volume, sex, age, age^2, sex × age, sex × age^2 and four MDS components.

Table DS11 Association analysis between rs6088662 and cognitive performance in the Chinese sample

Cognitive function	Test or subscale	Mean (s.d.)		t	P-value
		G carrier	TT		
Executive function	Attention alert	0.013 (0.027)	0.0047 (0.025)	2.612	0.0094
Language abilities	Visual-auditory	124.81 (7.76)	121.73 (10.56)	2.539	0.012

Before performing two-tailed t-test, F-test was conducted to compare the variances between two genotype groups.

F-test in the analysis of visual-auditory was significant ($p<0.005$), i.e., assuming the two groups do not have equal standard deviations, thus we used unpaired t-test with Welch's correction.

F-test in the analysis of attention alert was not significant ($p>0.3$), i.e., assuming both groups have the same standard deviation, we used unpaired t-test with no correction.

Fig. DS1 Association of rs6088662 with *TRPC4AP* mRNA expression in Europeans in Dimas *et al* study (*n*=75).[4]

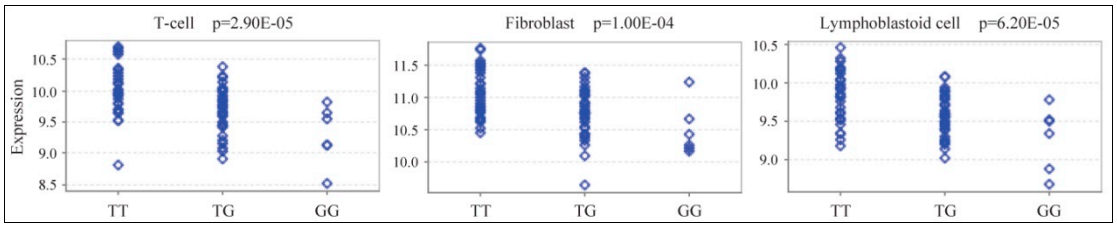


Fig. DS2 Association of rs6088662 with *TRPC4AP* mRNA expression in Europeans in Nica *et al* study (*n*=160).[3]

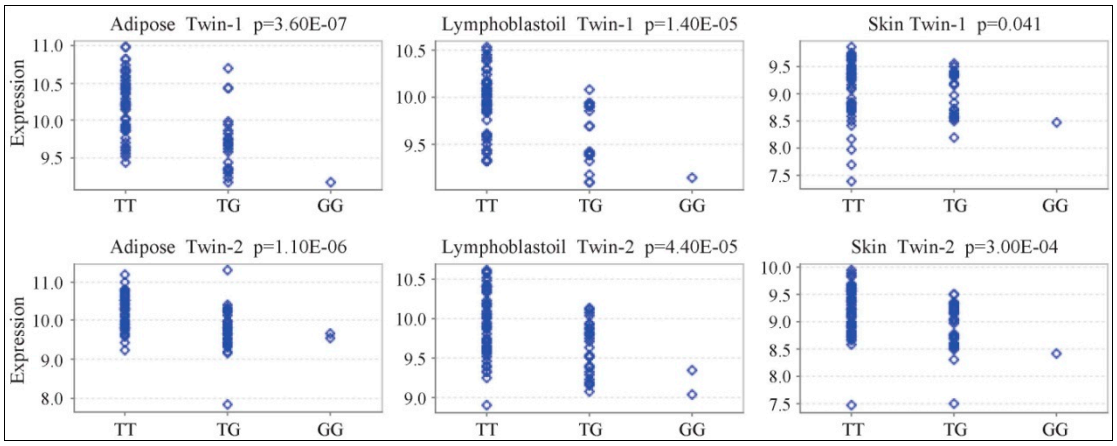


Fig. DS3 Association of rs6088662 with *TRPC4AP* mRNA expression in Europeans in Stranger *et al* study (*n*=109).[5]

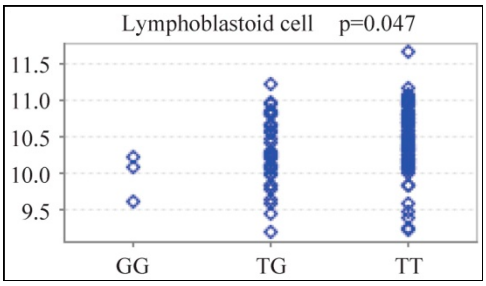
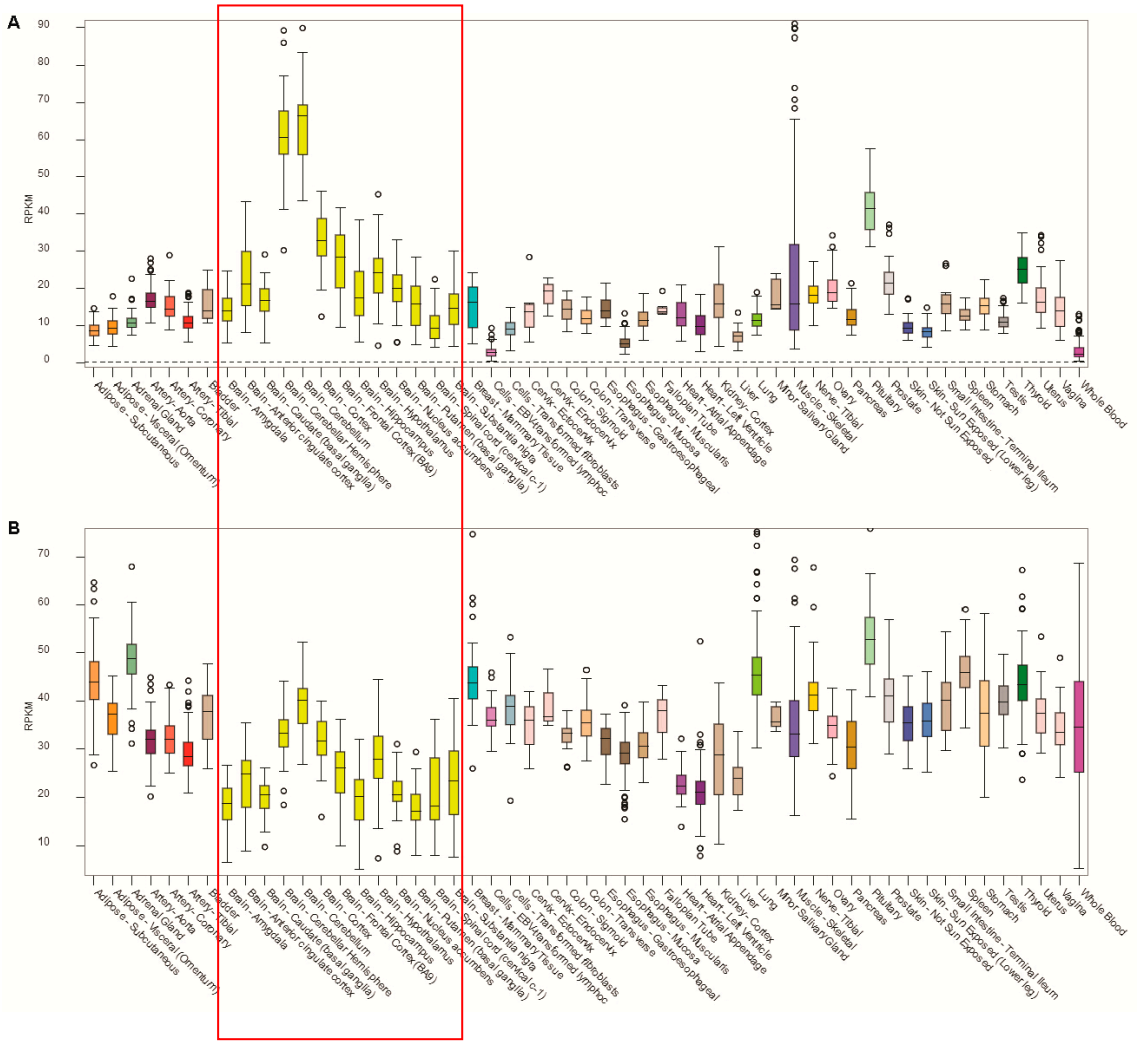


Fig. DS4 Spatial expression profiling of *GGT7* (A) and *TRPC4AP* (B) in human tissues. The results in brain tissues were marked in red rectangle.



References

1 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**: 1494-9.

2 Yang TP, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, Stranger BE, et al. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics*. 2010; **26**: 2474-6.

3 Nica AC, Parts L, Glass D, Nisbet J, Barrett A, Sekowska M, et al. The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet*. 2011; **7**: e1002003.

4 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**: 1246-50.

5 Stranger BE, Montgomery SB, Dimas AS, Parts L, Stegle O, Ingle CE, et al. Patterns of cis regulatory variation in diverse human populations. *PLoS Genet*. 2012; **8**: e1002639.

6 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**: 764-70.

7 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**: 212-8.

8 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**: 244-52.

9 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**: 109-17.

10 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**: 849-59; discussion 63-4.

11 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**: 624-9.

12 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 4-57.

13 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**: 447-53.

14 Sellgren C, Landén M, Lichtenstein P, Hultman CM, Långström N. Validity of bipolar disorder hospital discharge diagnoses: file review and multiple register linkage in Sweden. *Acta Psychiatr Scand*. 2011; **124**: 447-53.

15 Vassos E, Steinberg S, Cichon S, Breen G, Sigurdsson E, Andreassen OA, et al. Replication study

and meta-analysis in European samples supports association of the 3p21.1 locus with bipolar disorder. *Biol Psychiatry*. 2012; **72**: 645-50.

16 Hammer C, Cichon S, Muhleisen TW, Haenisch B, Degenhardt F, Mattheisen M, et al. Replication of functional serotonin receptor type 3A and B variants in bipolar affective disorder: a European multicenter study. *Transl Psychiatry*. 2012; **2**: e103.

17 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**: 372-81.

18 Zhang X, Zhang C, Wu Z, Wang Z, Peng D, Chen J, et al. Association of genetic variation in CACNA1C with bipolar disorder in Han Chinese. *J Affect Disord*. 2013; **150**: 261-5.

19 Green EK, Hamshere M, Forty L, Gordon-Smith K, Fraser C, Russell E, et al. Replication of bipolar disorder susceptibility alleles and identification of two novel genome-wide significant associations in a new bipolar disorder case-control sample. *Mol Psychiatry*. 2013; **18**: 1302-7.

20 Stein JL, Medland SE, Vasquez AA, Hibar DP, Senstad RE, Winkler AM, et al. Identification of common variants associated with human hippocampal and intracranial volumes. *Nat Genet*. 2012; **44**: 552-61.

21 Chen M, Xu Z, Zhai J, Bao X, Zhang Q, Gu H, et al. Evidence of IQ-modulated association between ZNF804A gene polymorphism and cognitive function in schizophrenia patients. *Neuropsychopharmacology*. 2012; **37**: 1572-8.

22 Zhang Q, Shen Q, Xu Z, Chen M, Cheng L, Zhai J, et al. The effects of CACNA1C gene polymorphism on spatial working memory in both healthy controls and patients with schizophrenia or bipolar disorder. *Neuropsychopharmacology*. 2011; **37**: 677-84.

23 Zhu X, Gu H, Liu Z, Xu Z, Chen X, Sun X, et al. Associations between TCF4 gene polymorphism and cognitive functions in schizophrenia patients and healthy controls. *Neuropsychopharmacology*. 2012; **38**: 683-9.

24 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.

25 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.

26 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.

27 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**: 977-83.

28 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**: 748-55.

29 Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R, Leek JT, et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature*. 2011; **478**: 519-23.

30 Webster JA, Gibbs JR, Clarke J, Ray M, Zhang W, Holmans P, et al. Genetic control of human brain transcript expression in Alzheimer disease. *Am J Hum Genet*. 2009; **84**: 445-58.

31 Zou F, Chai HS, Younkin CS, Allen M, Crook J, Pankratz VS, et al. Brain expression genome-wide association study (eGWAS) identifies human disease-associated variants. *PLoS Genet*. 2012; **8**: e1002707.

32 Heinzen EL, Ge D, Cronin KD, Maia JM, Shianna KV, Gabriel WN, et al. Tissue-specific genetic control of splicing: implications for the study of complex traits. *PLoS Biol*. 2008; **6**: e1.

33 Schizophrenia Working Group of the Psychiatric Genomics C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014; **511**: 421-7.

34 Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, Breen G, et al. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry*. 2012; **18**: 497-511.

35 Preisig M, Waeber G, Vollenweider P, Bovet P, Rothen S, Vandeleur C, et al. The PsyCoLaus study: methodology and characteristics of the sample of a population-based survey on psychiatric disorders and their association with genetic and cardiovascular risk factors. *BMC Psychiatry*. 2009; **9**: 9.