Mol Neurobiol
DOI 10.1007/s12035-016-0041-x

Identification of a Bipolar Disorder Vulnerable Gene *CHDH* at 3p21.1

Hong Chang[1] · Lingyi Li[1] · Tao Peng[1] · Maria Grigoroiu-Serbanescu[2] · Sarah E. Bergen[3,4] · Mikael Landén[3,5] · Christina M. Hultman[3] · Andreas J. Forstner[6,7] · Jana Strohmaier[8] · Julian Hecker[6,9] · Thomas G. Schulze[10] · Bertram Müller-Myhsok[11,12,13] · Andreas Reif[14] · Philip B. Mitchell[15,16] · Nicholas G. Martin[17] · Sven Cichon[6,7,18,19] · Markus M. Nöthen[6,7] · Stéphane Jamain[20,21,22] · Marion Leboyer[20,21,22,23] · Frank Bellivier[20,22,23,24,25] · Bruno Etain[20,21,22,23] · Jean-Pierre Kahn[22,26] · Chantal Henry[20,21,22,23] · Marcella Rietschel[8] · The Swedish Bipolar Study Group · MooDS Consortium · Xiao Xiao[1] · Ming Li[1]

Received: 8 April 2016 / Accepted: 5 August 2016

Abstract Genome-wide analysis (GWA) is an effective strategy to discover extreme effects surpassing genome-wide significant levels in studying complex disorders; however, when sample size is limited, the true effects may fail to achieve genome-wide significance. In such case, there may be authentic results among the pools of nominal candidates, and an alternative approach is to consider nominal candidates but are replicable across different samples. Here, we found that mRNA expression of the choline dehydrogenase gene (*CHDH*) was uniformly upregulated in the brains of bipolar disorder (BPD) patients compared with healthy controls across different studies. Follow-up genetic analyses of

Hong Chang, Lingyi Li and Tao Peng contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s12035-016-0041-x) contains supplementary material, which is available to authorized users.

✉ Xiao Xiao
xiaoxiao.whu05@gmail.com

✉ Ming Li
limingkiz@mail.kiz.ac.cn

[1] Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences and Yunnan Province, Kunming Institute of Zoology, Kunming, Yunnan, People's Republic of China

[2] Biometric Psychiatric Genetics Research Unit, Alexandru Obregia Clinical Psychiatric Hospital, Bucharest, Romania

[3] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

[4] Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA

[5] Section of Psychiatry and Neurochemistry, Sahlgrenska Academy at Gothenburg University, Gothenburg, Sweden

[6] Institute of Human Genetics, University of Bonn, Bonn, Germany

[7] Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany

[8] Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Heidelberg, Germany

[9] Institute of Genomic Mathematics, University of Bonn, Bonn, Germany

[10] Institute of Psychiatric Phenomics and Genomics, Ludwig-Maximilians-University Munich, Munich, Germany

[11] Max Planck Institute of Psychiatry, Munich, Germany

[12] Munich Cluster for Systems Neurology (SyNergy), Munich, Germany

[13] Institute of Translational Medicine, University of Liverpool, Liverpool, UK

[14] Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, University Hospital Frankfurt, Frankfurt, Germany

[15] School of Psychiatry, University of New South Wales, Sydney, Australia

[16] Black Dog Institute, Sydney, Australia

Published online: 25 August 2016

CHDH variants in multiple independent clinical datasets (including 11,564 cases and 17,686 controls) identified a risk SNP rs9836592 showing consistent associations with BPD ($P_{meta} = 5.72 \times 10^{-4}$), and the risk allele indicated an increased *CHDH* expression in multiple neuronal tissues (lowest $P = 6.70 \times 10^{-16}$). These converging results may identify a nominal but true BPD susceptibility gene *CHDH*. Further exploratory analysis revealed suggestive associations of rs9836592 with childhood intelligence ($P = 0.044$) and educational attainment ($P = 0.0039$), a "proxy phenotype" of general cognitive abilities. Intriguingly, the *CHDH* gene is located at chromosome 3p21.1, a risk region implicated in previous BPD genome-wide association studies (GWAS), but *CHDH* is lying outside of the core GWAS linkage disequilibrium (LD) region, and our studied SNP rs9836592 is ~1.2 Mb 3′ downstream of the previous GWAS loci (e.g., rs2251219) with no LD between them; thus, the association observed here is unlikely a reflection of previous GWAS signals. In summary, our results imply that *CHDH* may play a previously unknown role in the etiology of BPD and also highlight the informative value of integrating gene expression and genetic code in advancing our understanding of its biological basis.

Keywords Bipolar disorder · Gene expression · *CHDH* · Genetic evidence · Expression quantitative trait loci (eQTL) · Cognitive ability

Introduction

Bipolar disorder (BPD) has a lifetime prevalence of 0.5–1.5 % in the general population [1], and a series of family, twin, and adoption studies have established a degree of heritability averaging about 70 % [2]. Previous studies identified several potential candidates as BPD risk factors, such as genetic variants in *SLC6A4*, *DRD4*, *DAOA*, *TPH2*, *BDNF*, and *GSK3* [3], while recent genome-wide association studies (GWAS) and other large-scale meta-analyses further implicated several novel susceptibility genomic loci, including *CACNA1C*, *ANK3*, *ODZ4*, *TRANK1*, and *NCAN*, etc. [4–12]. Despite progress in elucidating specific genetic risk factors, both the etiology and pathogenesis of BPD still remain largely unknown.

Alongside genetic risk architectures, genes that are differentially expressed between BPD patients and healthy controls may also play key roles in the etiology of BPD. Recent advances in gene microarray and RNA sequencing (RNA-seq) techniques have allowed researchers to begin investigating the hypothesis that gene expression alterations may be involved in the pathogenesis of BPD [13–19], and several genome-wide significantly differentially expressed genes (DEGs) between BPD patients and normal controls have been reported. However, only a few overlapping DEGs were identified across different studies, and many of those identified DEGs were rarely implicated in association studies, while their genetic contributions to BPD susceptibility remain unclear, at best [13].

An integrative analysis combining both genetic and gene expression data may help better understand the risk structures of BPD [20–22]. In this study, we performed a series of analyses involving both genetic and gene expression approaches to identify BPD susceptibility genes. First, we attempted to discover potential vulnerability genes that are consistently differentially expressed in the brains of BPD patients across expression studies, and only genes differentially expressed across all included studies were considered viable candidates for further analysis, wherein we then conducted association analyses of the single nucleotide polymorphisms (SNPs) in the DEGs to identify their potential genetic contributions to BPD susceptibility. We further assessed the genotypic effects of the risk SNPs on gene expression to uncover the underlying genetic mechanisms that may explain the aberrant gene expression in BPD patients. Finally, genetic loci associated with clinical diagnosis are also expected to be related to the so-called intermediate phenotypes implicated in the biology of genetic risk for BPD [23]. Previous studies have reported deficits in cognitive abilities in patients with BPD and their unaffected relatives [24], implying that variation in cognition is an intermediate phenotype related to the genetic risk of BPD. We therefore also tested the effects of the BPD risk SNPs on cognitive abilities.

Materials and Methods

To explain the logic of the study design, a flowchart summarizing the analytical methods and showing how genes/variants were taken forward from one stage of analysis to the next is shown in Fig. 1. All protocols and methods used in this study were approved by the institutional review board of the

[17] QIMR Berghofer Medical Research Institute, Brisbane, Australia

[18] Division of Medical Genetics, University Hospital Basel and Department of Biomedicine, University of Basel, Basel, Switzerland

[19] Institute of Neuroscience and Medicine (INM-1), Structural and Functional Organization of the Brain, Genomic Imaging, Research Centre Jülich, 52425 Jülich, Germany

[20] Inserm U 955, IMRB, Psychiatrie Translationnelle, Créteil, France

[21] Faculté de Médecine, Université Paris Est, Créteil, France

[22] Fondation Fondamental, Créteil, France

[23] AP-HP, Hôpitaux Universitaires Henri Mondor, DHU Pepsy, Pôle de Psychiatrie, Créteil, France

[24] AP-HP, Groupe Hospitalier Lariboisière - F. Widal, Pôle de Psychiatrie, Paris, France

[25] Université Paris Diderot, Paris, France

[26] Département de Psychiatrie et de Psychologie Clinique, CHU de Nancy, Hôpital Jeanne d'Arc, Toul, France

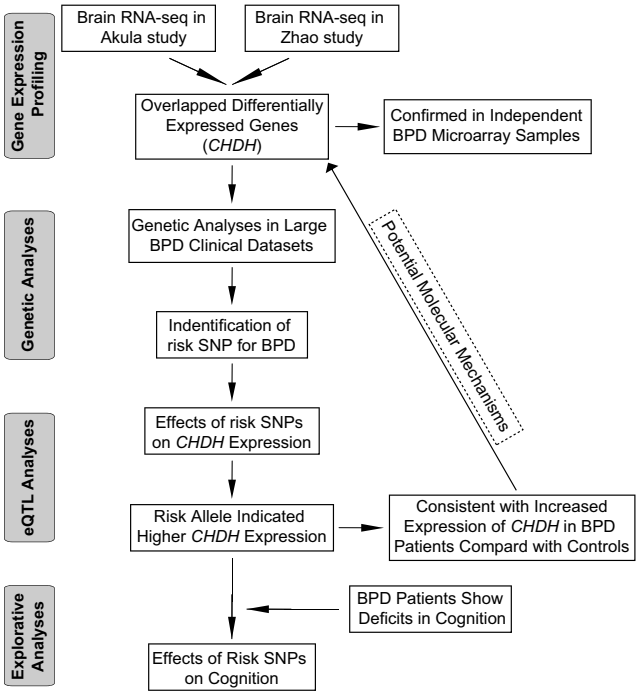

Fig. 1 Flowchart of the present study. Based on the analytic results of differentially expressed genes between BPD patients and healthy controls, we systematically studied the associations of the risk genes with BPD in genetic case–control samples. The risk variants were then tested for their effects on risk gene expression and explored for their associations with cognitive abilities

Kunming Institute of Zoology, Chinese Academy of Sciences, and adhere to all relevant national and international regulations.

Expression Data Used in This Study

For the gene expression comparisons, we used two-step analyses. The discovery analysis includes two RNA-seq studies performed by Akula et al. [13] and Zhao et al. [19], respectively. In the Akula et al.'s [13] study, RNA-seq analysis was performed in the dorsolateral prefrontal cortex (DLPFC; Brodmann area 46) of postmortem brain tissues among 11 BPD cases and 8 psychiatrically healthy controls from the Stanley Medical Research Institute (www.stanleyresearch.org). In their study, a total of 25,017 genes were analyzed, with 1225 genes having a nominal P value <0.05 and 298 genes showing P value <0.01. In the Zhao et al.'s [19] study, RNA-seq analysis was conducted in postmortem brain samples from 26 BPD patients and 26 healthy controls that explored the anterior cingulated cortex (Brodmann area 24)—a brain region known to be involved in learning and executive functions and psychiatric disorders. In their study, a total of 15,294 genes were analyzed, identifying 789 genes nominally differentially expressed between cases and controls ($P < 0.05$) and 416 genes showing P value <0.01.

The genes showing P value <0.01 in both discovery samples were subject to replications in another two microarray expression studies. The first replication microarray analyses were also performed by Akula et al. [13], but the DLPFC samples (22 BPD patients and 26 controls) used for microarray analyses were completely independent from those of the RNA-seq analyses. The second replication dataset was from "Metamoodics," an online database (http://psychiatry.som.jhmi.edu/metamoodics/) that consists of a systematic meta-analysis of ten genome-wide microarray gene expression studies on human brains including 57 BPD patients and 60 healthy controls [17]. Of note is that there are partial overlaps in sampling between "Metamoodics" and the other samples used in this study.

In sum, a total of four gene expression datasets were used in this study. All of the included subjects were of European origin. Detailed information including sample descriptions, data quality control, and statistical analysis can be found in the original studies [13, 17, 19].

Genetic Association Samples

The Psychiatric Genomics Consortium (PGC) BPD group recently conducted a meta-analysis of large-scale genome-wide data on BPD in populations of European ancestry [8], wherein they compared BPD patients who had experienced pathologically relevant episodes of elevated mood (mania or hypomania) with control subjects from the same geographic and ethnic populations. In this study, we used summary statistics from these primary GWAS samples that consisted of 7481 cases and 9250 controls. Detailed descriptions of the samples, data quality, genomic controls, and statistical analyses can be found in the original GWAS [8].

Here, replication analyses were performed in seven independent BPD samples that included 4083 patients and 8436 controls, and no overlap was found with the discovery samples. Detailed information on individual samples—including diagnostic assessment, genotyping, and quality control—are shown in the Supplemental Data and Table S1. Most of these replication samples were previously reported in earlier large-scale collaborative studies, where they were found to be effective in detecting genetic risk variants for BPD [5, 7, 12, 25]. Each of the original sample subjects was recruited under relevant ethical and legal guidelines for their respective areas, and all provided written informed consents prior to their inclusion in the earlier studies. In brief, the origin and sizes of the replication BPD samples are as follows: (1) Romania (380 cases and 223 controls) [5]; (2) Sweden I (836 cases and 2093 controls) [25]; (3) Sweden II (1415 cases and 1271 controls) [12]; (4) France (451 cases and 1631 controls) [25]; (5) Germany II (181 cases and 527 controls) [7]; (6) Germany III (490 cases and 880 controls) [7]; and (7) Australia (330 cases and 1811 controls) [7].

Healthy Subjects for Expression Quantitative Trait Loci Analysis

To identify the impact of risk SNPs on mRNA expression, we utilized two well-characterized gene expression databases. A brief description of the gene expression resources is provided below; more detailed information can be found in the original studies. (1) GTEx (Genotype-Tissue Expression project) [26]: GTEx contains both genetic variation and RNA-seq gene expression data from a diverse set of human tissues. (2) BrainCloud [27]: BrainCloud contains genetic information and whole transcriptome microarray expression data from postmortem DLPFC of 261 normal human subjects (i.e., without neuropsychiatric diagnosis). Raw genotype data were obtained. Expression data and demographic information such as RNA integrity number (RIN), race, sex, and age were also obtained. Prenatal subjects were removed from the analysis since *CHDH* mRNA is significantly differentially expressed between fetal and postnatal subjects. Statistical analysis was conducted using linear regression, with RIN, sex, race, and age as covariates. The GTEx and BrainCloud data are primarily used to explore the regulation of gene expression in the human tissues and are valuable resources in exploring functional follow-ups of disease-associated variants.

Cognitive Analysis

We used educational attainment as a "proxy phenotype" for cognitive function. Although it is not a direct cognitive measure, educational attainment is correlated with cognitive ability ($r \sim 0.5$) and some personality traits related to persistence and self-discipline [28]. Educational attainment is strongly associated with social outcomes, and there is a well-documented health-education gradient. Estimates suggest that around 40 % of the variance in educational attainment is explained by genetic factors. We used two well-characterized measurements of educational attainment, a binary variable for college completion ("College") and a quantitative variable defined as an individual's years of schooling ("EduYears"). College may be more comparable across countries, whereas EduYears contains more information about individual differences within countries. Briefly, educational attainment was measured at an age at which participants were very likely to have completed their education (more than 95 % of the sample were at least 30 years old). On average, participants have 13.3 years of schooling, and 23.1 % have a college degree. Recently, a GWAS on these "educational attainment" phenotypes has been performed on 101,069 European individuals [28], and we obtained the statistical results of BPD risk SNPs with educational attainment from their study. Detailed information on the samples, genotyping methods, and statistical analyses can be found in the original report [28].

In addition to the "educational attainment," we also used the phenotype of "childhood intelligence," which is measured by psychometric cognitive tests (intelligence quotient (IQ)-type tests). Childhood intelligence is a strong predictor of many important life outcomes, such as educational attainment [29] and is also associated with various psychiatric disorders, including schizophrenia, BPD, and major depression [30, 31]. Results from twin, family, and adoption studies of childhood intelligence are consistent with general intelligence being highly heritable and genetically stable throughout the life course [32]. We utilized a recent GWAS of childhood intelligence including 12,441 children of European ancestry [33]. In brief, the age of the children ranged between 6 and 18 years, and the best available measure of general cognitive ability (*g*) or IQ derived from diverse tests that assess both verbal and non-verbal ability was used. Detailed information on the cohorts, intelligence measurements, genotyping methods, and statistical analyses can be found in the original GWAS [33].

SNP Selection and Statistical Analysis

For initial screening in the discovery sample, a total of 52 SNPs from the *CHDH* gene were selected. The *CHDH* gene and its surrounding genomic regions (~15 kb upstream and downstream, respectively) were screened (spanning a total of ~63.6 kb). The previous GWAS (i.e., our discovery sample) has analyzed a total of 52 SNPs in this region [8], and we have chosen all of them without any selection bias in this study. The linkage disequilibrium (LD) structure of these 52 SNPs in Europeans was constructed with the R package "snp.plotter" [34], and the LD relationship between paired SNPs was determined using r^2 confidence interval algorithm and was shown by gradient color. The genomic structures of the *CHDH* gene, the locations of the tested SNPs, and their LD patterns in European populations are shown in Fig. 2; SNP information is shown in Table S2. For replication analyses, significant SNPs among the PGC discovery samples were analyzed in the Romania case–control sample and other replication samples.

Both Illumina (San Diego, CA, USA) and Affymetrix platforms were used for SNP genotyping in a bulk of the replication samples (details shown in Supplemental Data). Control subjects were tested for deviation from Hardy–Weinberg equilibrium (HWE), and none of the tested SNPs were found to deviate from HWE in any sample. Association *P* values and allele-specific odds ratios (ORs) for each individual sample were calculated using a logistic regression model with an additive effect. Meta-analyses were then conducted based on *Z* scores combining data from all samples within the R package (*Meta module*) using the inverse variance weighted method under the fixed effects model, which subsequently yielded the combined *P* values and ORs. Before pooling, Cochran's (*Q*) χ^2 test of heterogeneity was performed to ensure that each

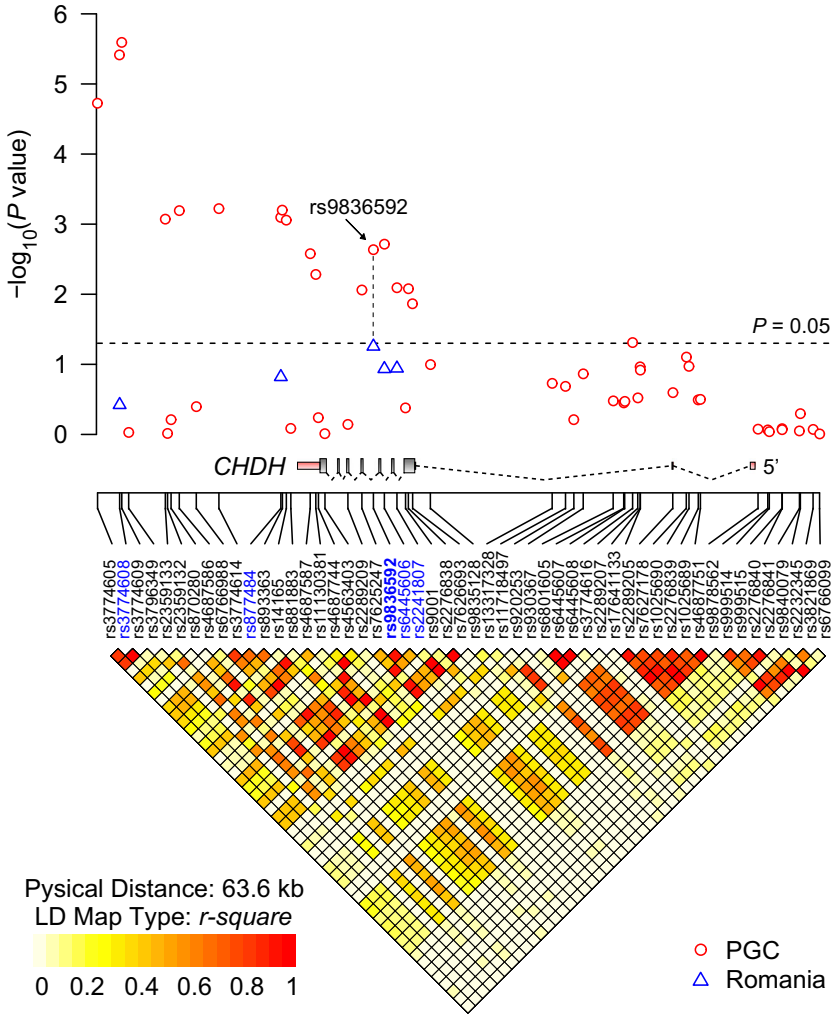

Fig. 2 Genetic association of *CHDH* with risks of BPD. A physical map of the region is given and depicts known genes within the region. *Bottom*, the linkage disequilibrium structure of the tested markers for 490 unrelated healthy control subjects of European descent depicted as r^2

sample population was suitable for meta-analysis. As described in a previous GWAS meta-analysis [8], *P* values for replication samples are reported as one-tailed tests and *P* values for all combined samples are shown as two-tailed tests. We used a forest plot to graphically present the individual ORs and their 95 % confidence intervals, wherein each sample was represented by a square in the forest plot.

Results

Identification of Nominally Significant DEGs Across Different Studies

In recent years, advances in sequencing platforms have allowed analyzing genome-wide gene expression profiling between BPD patients and normal controls, and different sets of genome-wide DEGs have been reported in each individual studies, but most of the DEGs were not overlapped across samples. Typically, people applied stringent multiple corrections in genome-wide analyses, which allowed them to identify extremely significant effects, but at the cost of causing false negative results (especially in small samples); that is to say, there may still be true effects among nominally significant associations. Before we are able to increase the sample size to have a sufficient statistical power of discovering all true associations in the genome-wide significant level, an alternative approach is to consider the DEGs that show consistently nominal significance across different studies. This hypothesis has been validated in the genetic analyses of schizophrenia, a psychiatric disorder that shared phenotypic and genetic risk with BPD—in the latest GWAS of schizophrenia with a giant sample size [35], many genes surpassed genome-wide significance in this largest GWAS, but only showed nominal associations in previous small samples, such as *DRD2*, *SRR*, and *GRM3* [36–38].

We believe this hypothesis is also suitable for gene expression comparison analyses between BPD patients and healthy controls. To identify the DEGs that show nominal significance but consistently replicated across samples, we used several well-characterized gene expression studies. In the discovery phase analysis, Akula et al. [13] performed genome-wide

RNA-seq analysis in the DLPFC tissues from 11 BPD cases and 8 healthy individuals; a total of 25,017 genes were analyzed, and 298 DEGs showed a *P* value of lower than 0.01. In parallel, Zhao et al. [19] conducted RNA-seq analysis in the anterior cingulated cortex from 26 BPD patients and 26 normal controls; a total of 15,294 genes were analyzed, and 416 DEGs revealed a *P* value lower than 0.01. There were eight consistently dysregulated genes exhibiting the same direction of disease-associated regulation across the two studies (Table S3). Among these eight DEGs, five genes (*ALDH4A1*, *PBXIP1*, *GALM*, *CHDH*, and *TP53BP2*) showed uniformly elevated expressions among individuals with BPD as compared to healthy controls, while three genes (*VIP*, *HIVEP2*, and *FAM49A*) were consistently downregulated in BPD patients compared with normal subjects (Table S3).

We tend to replicate these eight DEGs in two additional genome-wide microarray gene expression samples. Among these eight DEGs, seven (only excluded *GALM*) showed nominal significance in at least one of the replication samples (Table S3), suggesting the discovery analysis is reliable although the effect size is relatively small. Of note is that *CHDH*, which encodes the choline dehydrogenase, is the only gene showing consistently nominal significance in all of the four expression datasets (including discovery and replication samples; Table S3). Although the *P* values for *CHDH* did not survive genome-wide correction in either study, which is likely caused by the limited statistical power of individual small sample size, the gene is uniformly upregulated in the studies of Akula et al. [13] ($P = 8.11 \times 10^{-3}$) and Zhao et al. [19] ($P = 2.51 \times 10^{-3}$), as well as two replication samples (*P* values of 4.97×10^{-3} and 3.61×10^{-2}, respectively), and seems a true susceptibility gene with a relatively low to moderate effect. To test whether *CHDH* is a real BPD-related gene, we performed next-step analyses.

A SNP rs9836592 in CHDH Is Consistently Associated with BPD in Independent Samples

To date, several GWAS have been conducted in various BPD samples, identifying a few genome-wide significant genes, such as *ANK3*, *CACNA1C*, *ODZ4*, *NCAN*, *TRANK1*, and the like [4–12]. However, even these risk genes can only explain a small portion of the genetic liability to BPD, and the missing heritability is still unclear. GWAS is indeed an effective strategy to discover risk genes as it could scan the genome with hundreds of thousands of genetic variations and employ the necessary rigid statistical correction as no prior probability of any variant being positive. This strategy has the appeal of a level of statistical significance being clear and incontrovertible and has been demonstrated successful during the applications of several complex disorders when the sample size is quite large, such as schizophrenia [35], type 2 diabetes [39], and blood pressure [40]. However, when the sample size is not large enough, the stringent corrections in GWAS might preclude the authentic risk genes that only reached nominal statistical associations, but obviously showed sufficient biological interest in the illness. This notion is also supported by recent aggregated analyses [41] which indicated that there may be satisfactory replications in independent samples among those markers passing nominal significance in the initial GWAS of psychiatric disorders, such as *CMYA5* [42], *FGFR2* [43], *CAMKK2* [44], and *CREB1* [25].

Based on this rationale, with the use of a published BPD GWAS sample (7481 cases and 9250 controls) [8] conducted by the PGC, we analyzed 52 SNPs in *CHDH* spanning ~63.6 kb for associations with BPD and identified 18 SNPs in moderate to high LD showing nominal associations with BPD ($P < 0.05$; Fig. 2 and Table S2). We then tested five SNPs that reflected majority of the discovery association signals in our first replication Romania sample (380 cases and 223 controls). However, the most significant SNP in this region in the PGC discovery sample, rs3774609 ($P = 2.56 \times 10^{-6}$, OR = 0.888), showed no sign of association with BPD in the Romania sample ($P = 0.318$, OR = 1.038), and the direction of the allelic effect was opposite between the discovery and Romania samples. The non-significant replication of rs3774609 in our Romania sample was confirmed in other independent European samples (including 4496 cases and 42,422 controls) reported by PGC GWAS (the replication result of rs3774609 in their samples was shown in their Table 3; $P = 0.107$, OR = 0.970) [8]. We therefore disregarded rs3774609 from further analyses.

Alternatively, we found another SNP, rs9836592, which is in low LD with rs3774609 ($r^2 = 0.078$ in Europeans) showed nominal significance in the PGC discovery sample ($P = 0.00232$, OR = 1.080) and was also marginally associated with BPD among our Romania sample with the same direction of the allelic effect ($P = 0.0556$, OR = 1.217). The other three SNPs in *CHDH* were not associated with BPD in the Romania sample (rs877484: $P = 0.156$, OR = 0.886; rs6445606: $P = 0.116$, OR = 0.860; rs2241807: $P = 0.114$, OR = 0.869) and were excluded from further analyses. Further genetic analyses of rs9836592 in additional replication samples showed that this SNP again manifested marginal associations with BPD in the Germany II ($P = 0.0836$, OR = 1.209, including 181 cases and 527 controls) and Germany III samples ($P = 0.0756$, OR = 1.128, a total of 490 cases and 880 controls). The marginal associations in these small individual replication samples were likely caused by the limited statistical power as the effect sizes (i.e., OR) in these replication samples were even larger than those in the discovery sample. When we combined all the replication samples together through meta-analysis (including a total of 4083 cases and 8436 controls), the SNP (rs9836592) showed nominal significant association with BPD ($P_{meta} = 0.0441$, OR = 1.055), which is consistent with the results in the discovery sample. There is no

heterogeneity among these replication samples ($P = 0.404$, $I^2 = 2.9$ %). To increase the statistical power, we combined the discovery and replication samples (including 11,564 cases and 17,686 controls), and meta-analysis revealed a stronger association ($P_{meta} = 5.72 \times 10^{-4}$, OR = 1.070). Forest plot for the meta-analysis of rs9836592 on all BPD samples is presented in Fig. 3; the results for each sample are shown in Table S4.

We noticed that the association *P* values between rs9836592 and BPD did not achieve the conventional genome-wide level of statistical significance ($P = 5.0 \times 10^{-8}$), and given the herein observed OR (1.070), the results would not become genome-wide significant until the sample size increases to 79,179 cases and 79,179 controls (have a power of >80 %). However, as mentioned in the previous aggregated analyses [41], BPD is a polygenic disorder involving hundreds of thousands of genes in it, with each gene showing a minor effect, and there might be true findings among those markers only passing nominal significance in the initial GWAS, but later were confirmed in independent samples. Of note is that we did not set up any *in prior* hypothesis of choosing or dropping any SNPs within *CHDH*, but selected all the analyzed SNPs ($n = 52$) from previous GWAS [8] in this genomic region without any bias. Although the result did not achieve genome-wide significance, it does survive multiple correction according to the number of tested SNPs ($n = 52$) in this study (corrected $P = 0.0297$).

Notably, *CHDH* is located in chromosome 3p21.1, a genomic region that has been implicated in the genetic risk of BPD among previous GWAS [4, 9] and individual replication studies [45–47]. However, the SNP (rs9836592) identified here is located about 1.2 Mb away from the previous GWAS loci (e.g., rs2251219) and is not in the core GWAS LD region (Fig. S1); thus, our results are unlikely a reflection of previous GWAS signals.

The Risk SNP rs9836592 Is Associated with CHDH Expression

The observed changes of *CHDH* expression in patients and the genetic associations of rs9836592 with BPD among multiple samples strengthened the hypothesis of the status of *CHDH* as a potential susceptibility gene for BPD. However, these findings do not identify whether the altered *CHDH* expression was related to any specific genetic risk. Considering the known enrichment of expression quantitative trait loci (eQTL) among BPD risk loci in human brains [48], it is thus possible that alterations of *CHDH* expression in BPD patients are (at least partially) influenced by the genetic changes of rs9836592. To explore this possibility, we examined the RNA-seq gene expression database GTEx and found that, among European samples, rs9836592 is significantly (or marginally) associated with *CHDH* expression in several brain tissues, such as the cerebellum ($P = 5.70 \times 10^{-5}$; Fig. 4) and cerebellar hemisphere ($P = 0.0033$). In the tibial nerve tissue, rs9836592 showed a stronger association with *CHDH* expression ($P = 6.70 \times 10^{-16}$; Fig. 4), and even among the non-neural tissues, this SNP was still associated with *CHDH* expression ($P < 0.05$; Fig. 4). Across these tested GTEx samples, subjects carrying the BPD risk allele (T) of rs9836592 manifested higher *CHDH* expression compared to those with the protective alleles (C; as shown in Fig. 4).

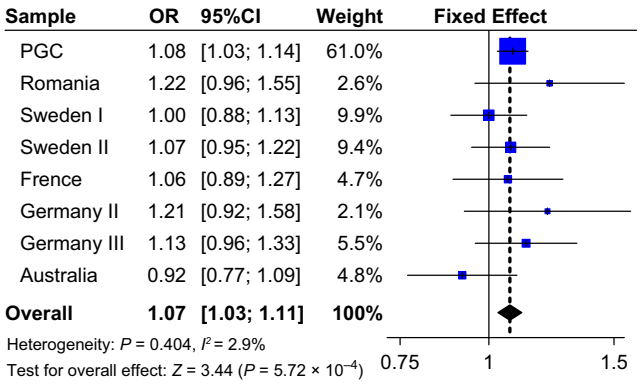

Fig. 3 Forest plot of odds ratios with 95 % confidence interval for BPD samples included in a meta-analysis of rs9836592

It should be noted that as the GTEx datasets include few subjects with brain data (as shown in Fig. 4), we then tested whether rs9836592 was associated with *CHDH* expression in the DLPFC tissues using BrainCloud [27], which contains more brain-derived samples. We again found that risk (T) allele carriers of rs9836592 showed increased *CHDH* expression compared with the protective allele individuals in this sample ($P = 0.040$; Fig. S2). We further analyzed whether rs9836592 was associated with the expression of other genes near to *CHDH*. In the BrainCloud samples, *CHDH* was the only gene showing eQTL association with the risk SNP; the other genes were not significant (all $P > 0.5$; Fig. S2). In the multiple GTEx samples, although rs9836592 showed associations with the expression of some nearby genes in several tissues, they are less significant and less consistent than those with *CHDH* (Table S5), implying that *CHDH* is the major gene in relation to the risk SNP in this region. Collectively, these results implied that the observed upregulation of *CHDH* expression among individuals diagnosed with BPD may be explained by some previously uncharacterized genetic mechanisms underlying risk variants (e.g., rs9836592), and these converging data uncovered a potential pathogenic mechanism for BPD with both gene expression and genetic evidence.

The Risk SNP rs9836592 Is Associated with Cognitive Abilities

To move beyond statistical association with clinical diagnosis and to obtain convergent evidence for an association between rs9836592 and BPD-related biology, we also performed a series of convergent analyses testing the risk SNPs on several biological phenotypes. Considering that cognitive functions

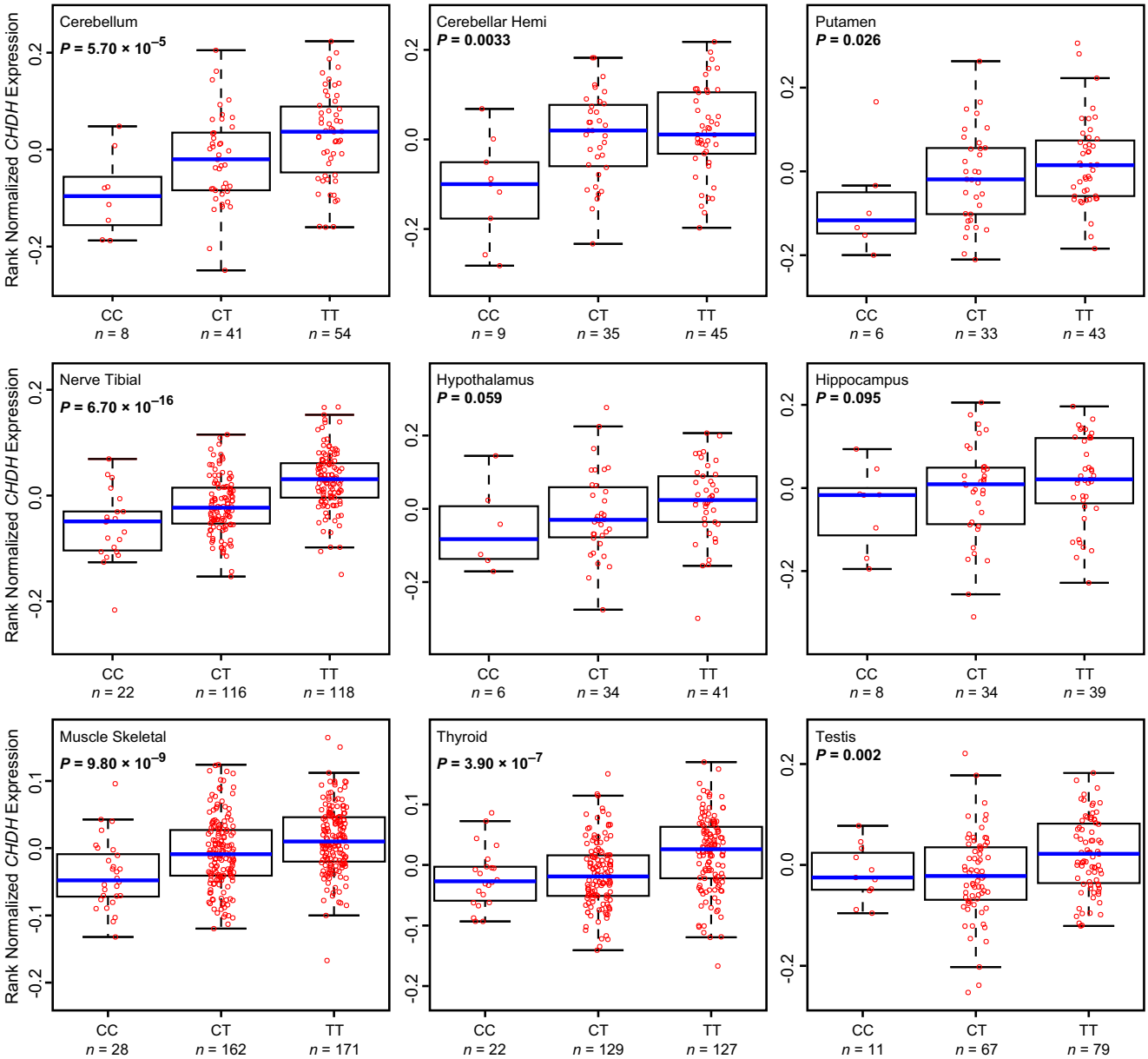

Fig. 4 rs9836592 is significantly associated with *CHDH* mRNA expression in diverse tissues compiled in the GTEx dataset

were frequently impaired in patients with BPD [24], and accumulating data indicated that many genetic loci associated with BPD were also related to cognitive functions in humans, we therefore hypothesized whether the identified risk-associated SNP (i.e., rs9836592) also affects cognitive abilities.

We firstly tested the effects of rs9836592 on educational attainment, a "proxy phenotype" of general cognitive abilities ($r \sim 0.5$) [28]. The harmonized measurements of educational attainment were coded by study-specific measures using the International Standard Classification of Education (1997) scale [49] and included a binary variable for college completion ("College") and a quantitative variable defined as an individual's years of schooling ("EduYears"). The sample comprised 95,427 individuals for "College" and 101,069 for "EduYears." In this explorative analysis, rs9836592 is significantly associated with "College" (OR = 1.025 for T allele, $P = 0.013$) and "EduYears" (SE = 0.012 for T allele, $P = 3.87 \times 10^{-3}$), further confirming the hypothesis that BPD risk SNPs are also expected to influence cognitive abilities. However, the risk allele predicts better cognitive function, implicating a more complex neurocognitive mechanism than expected.

In an independent sample, we also studied the impact of rs9836592 on childhood intelligence in 12,441 children of European ancestry. Intelligence is heritable and gives humans the cognitive abilities to learn, form concepts, understand and reason, etc. Childhood intelligence is a significant predictor of

cognitive change in later life [50] and is also associated with psychiatric disorders [30, 31]. Within this analysis, rs9836592 is nominally associated with childhood intelligence ($P = 0.044$), and the risk allele carriers showed worse performance.

Analysis of cognitive-related phenotypes further confirmed the role of the risk SNPs in BPD susceptibility and implied that it may be functional in the brain. However, as the association results on these cognitive phenotypes may not survive multiple correction, further validation in larger samples is needed.

Discussion

In this study, we showed several convergent lines of evidence supporting the role of *CHDH* as a BPD susceptibility gene. Across the various gene expression studies we conducted, *CHDH* was consistently upregulated in the brain tissues of BPD patients compared with healthy subjects. Likewise, the genetic risk variant rs9836592 in the *CHDH* gene was significantly associated with BPD across large-scale sample sets, and this allele was found to be strongly associated with an elevated expression of *CHDH* in the human brain. Together, these results strongly suggest that rs9836592 may be linked with a regulatory element that ultimately contributes to the dysfunction of *CHDH* observed among BPD patients. To date, *CHDH* is known to encode the choline dehydrogenase that localizes to the mitochondrion, and RNA-seq analysis in the GTEx database showed that it is abundantly expressed in brain tissues (Fig. S3). Further temporal expression analysis showed that the expression level of *CHDH* is relatively low at the early developmental stages (fetal age); as development progresses, the expression of *CHDH* gradually increases in human brain (Figs. S4 and S5). A previous $Chdh^{-/-}$ mouse study reported abnormal mitochondrial morphology in sperm, suggesting the function of this gene in sperm motility and fertility [51]. Unfortunately, the precise function of *CHDH* in the brain remains unclear at best, but our present findings provide a strong case for further follow-up study to explore its function and more fully elucidate its potential roles in BPD.

To move beyond statistical association with clinical diagnosis and to obtain convergent evidence for association between *CHDH* and BPD-related biology, we have performed a series of convergent analyses testing risk-associated SNPs on several related biological phenotypes, e.g., cognitive abilities. Although we believe our deductive, hypothesis-driven strategy minimizes serendipity, it does involve a number of tests. Thus, it is necessary to consider the potential for spurious association because of multiple testing. The SNP which showed significant associations in a meta-analysis of diverse independent clinical datasets was tested for associations with molecular regulation of *CHDH* mRNA and to biological phenotypes related to risk of BPD. A consistent pattern of gene expression diagnostic comparison, allelic association, involving clinical diagnosis and cognition, was found in independent samples and in the expression of the *CHDH* gene in brain tissues from healthy controls. The likelihood that by chance the same risk-associated alleles would predict variation in each of these independent phenotypes across these diverse samples and always in the direction of abnormality associated with illness is remote.

Although this study presents several interesting results, the current evidence is limited, and we are cautious in extrapolating our results further. First, the observed differences of *CHDH* expression between BPD patients and controls did not survive multiple corrections in any single study. Though showing consistently nominal significance, the effect size of *CHDH* should be deemed as low to moderate (fold changes in BPD patients versus controls range from 1.08 to 1.30). Second, the association between *CHDH* SNPs and BPD in genetic case–control samples is far to reach genome-wide statistical significance, and the effect size of rs9836592 in BPD genetic risk is relatively small (OR = 1.070). Power analysis showed that the present sample size (11,564 cases and 17,686 controls) is underpowered to observe genome-wide significant association assuming such effect size (power = 0.002), and a power estimation suggests that analyses including at least 79,179 cases and 79,179 controls may have a power of higher than 0.8 to observe genome-wide significance. Thus, further genetic analyses in much larger samples are required. Finally, although we have identified the effects of rs9836592 on *CHDH* mRNA expression, we cannot exclude the possibility of other genetic loci having an effect on the changes of *CHDH* expression in BPD patients, and we were unable to estimate in this study the effects of non-genetic factors, such as durations of illness or drug medications, on *CHDH* expression.

In summary, our study presents a series of convergent lines of evidence that support *CHDH* as a candidate BPD vulnerable gene while also providing potentially significant insights into the pathogenesis of BPD and outlining a novel model of identifying further risk genes for BPD.

Acknowledgments We would like to acknowledge the efforts of the Psychiatric Genomics Consortium Bipolar Disorder Working Group for their contributions to this study. We are grateful to Andrew Willden (Kunming Institute of Zoology) for language editing of the manuscript. This work was supported by CAS Pioneer Hundred Talents Program (to M.L.). This work was also supported by the German Federal Ministry of Education and Research (BMBF) through the Integrated Genome Research Network (IG) MooDS (Systematic Investigation of the Molecular Causes of Major Mood Disorders and Schizophrenia; grant 01GS08144 to SC and MMN and grant 01GS08147 to MR), under the auspices of the National Genome Research Network plus (NGFNplus), and through the Integrated Network IntegraMent (Integrated Understanding of Causes and Mechanisms in Mental Disorders), under the auspices of the e:Med Programme (grant 01ZX1314A to SC and MMN and grant 01ZX1314G to MR). MMN is a member of the DFG-funded Excellence-Cluster ImmunoSensation. The Romanian sample

recruitment and genotyping was funded by UEFISCDI, Bucharest, Romania, grant no. 89/2012 to M.G.S., and by the German Federal Ministry of Education and Research (BMBF), MooDS Project, grant no. 01GS08144 to S.C. and M.M.N. Funding for the Swedish collection was provided by the Stanley Center for Psychiatric Research, Broad Institute, from a grant from Stanley Medical Research Institute. We also wish to thank the BBMRI.se and KI Biobank at Karolinska Institutet for the professional biobank service.

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Craddock N, Jones I (1999) Genetics of bipolar disorder. J Med Genet 36:585–594
2. Smoller JW, Finn CT (2003) Family, twin, and adoption studies of bipolar disorder. Am J Med Genet C: Semin Med Genet 123C:48–58
3. Serretti A, Mandelli L (2008) The genetics of bipolar disorder: genome 'hot regions', genes, new potential candidates and future directions. Mol Psychiatry 13:742–771
4. Chen DT, Jiang X, Akula N, Shugart YY, Wendland JR et al (2013) Genome-wide association study meta-analysis of European and Asian-ancestry samples identifies three novel loci associated with bipolar disorder. Mol Psychiatry 18:195–205
5. Cichon S, Muhleisen TW, Degenhardt FA, Mattheisen M, Miro X et al (2011) Genome-wide association study identifies genetic variation in neurocan as a susceptibility factor for bipolar disorder. Am J Hum Genet 88:372–381
6. Ferreira MA, O'Donovan MC, Meng YA, Jones IR, Ruderfer DM et al (2008) Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder. Nat Genet 40:1056–1058
7. Muhleisen TW, Leber M, Schulze TG, Strohmaier J, Degenhardt F et al (2014) Genome-wide association study reveals two new risk loci for bipolar disorder. Nat Commun 5:3339
8. Psychiatric Genomics Consortium Bipolar Disorder Working Group (2011) Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat Genet 43:977–983
9. McMahon FJ, Akula N, Schulze TG, Muglia P, Tozzi F et al (2010) Meta-analysis of genome-wide association data identifies a risk locus for major mood disorders on 3p21.1. Nat Genet 42:128–131
10. Baum AE, Akula N, Cabanero M, Cardona I, Corona W et al (2008) A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. Mol Psychiatry 13:197–207
11. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447:661–678
12. Li M, Luo XJ, Landen M, Bergen SE, Hultman CM et al (2015) Impact of a *cis*-associated gene expression SNP on chromosome 20q11.22 on bipolar disorder susceptibility, hippocampal structure and cognitive performance. Br J Psychiatry 208:128–137. doi:10.1192/bjp.bp.114.156976
13. Akula N, Barb J, Jiang X, Wendland JR, Choi KH et al (2014) RNA-sequencing of the brain transcriptome implicates dysregulation of neuroplasticity, circadian rhythms and GTPase binding in bipolar disorder. Mol Psychiatry 19:1179–1185
14. Choi KH, Higgs BW, Wendland JR, Song J, McMahon FJ et al (2011) Gene expression and genetic variation data implicate PCLO in bipolar disorder. Biol Psychiatry 69:353–359
15. Elashoff M, Higgs BW, Yolken RH, Knable MB, Weis S et al (2007) Meta-analysis of 12 genomic studies in bipolar disorder. J Mol Neurosci 31:221–243
16. Matigian N, Windus L, Smith H, Filippich C, Pantelis C et al (2007) Expression profiling in monozygotic twins discordant for bipolar disorder reveals dysregulation of the WNT signalling pathway. Mol Psychiatry 12:815–825
17. Seifuddin F, Pirooznia M, Judy JT, Goes FS, Potash JB et al (2013) Systematic review of genome-wide gene expression studies of bipolar disorder. BMC Psychiatry 13:213
18. Shao L, Vawter MP (2008) Shared gene expression alterations in schizophrenia and bipolar disorder. Biol Psychiatry 64:89–97
19. Zhao Z, Xu J, Chen J, Kim S, Reimers M et al (2015) Transcriptome sequencing and genome-wide association analyses reveal lysosomal function and actin cytoskeleton remodeling in schizophrenia and bipolar disorder. Mol Psychiatry 20:563–572
20. Niculescu AB (2013) Convergent functional genomics of psychiatric disorders. Am J Med Genet B Neuropsychiatr Genet 162B:587–594
21. Le-Niculescu H, Kurian SM, Yehyawi N, Dike C, Patel SD et al (2009) Identifying blood biomarkers for mood disorders using convergent functional genomics. Mol Psychiatry 14:156–174
22. Le-Niculescu H, McFarland MJ, Mamidipalli S, Ogden CA, Kuczenski R et al (2007) Convergent functional genomics of bipolar disorder: from animal model pharmacogenomics to human genetics and biomarkers. Neurosci Biobehav Rev 31:897–903
23. Meyer-Lindenberg A, Weinberger DR (2006) Intermediate phenotypes and genetic mechanisms of psychiatric disorders. Nat Rev Neurosci 7:818–827
24. Robinson LJ, Ferrier IN (2006) Evolution of cognitive impairment in bipolar disorder: a systematic review of cross-sectional evidence. Bipolar Disord 8:103–116
25. Li M, Luo XJ, Rietschel M, Lewis CM, Mattheisen M et al (2014) 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 19:452–461
26. GTEx Consortium (2013) The Genotype-Tissue Expression (GTEx) project. Nat Genet 45:580–585
27. Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R et al (2011) Temporal dynamics and genetic control of transcription in the human prefrontal cortex. Nature 478:519–523
28. Rietveld CA, Medland SE, Derringer J, Yang J, Esko T et al (2013) GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science 340:1467–1471
29. Deary IJ (2012) Intelligence. Annu Rev Psychol 63:453–482
30. Koenen KC, Moffitt TE, Roberts AL, Martin LT, Kubzansky L et al (2009) Childhood IQ and adult mental disorders: a test of the cognitive reserve hypothesis. Am J Psychiatry 166:50–57
31. Batty GD, Mortensen EL, Osler M (2005) Childhood IQ in relation to later psychiatric disorder: evidence from a Danish birth cohort study. Br J Psychiatry 187:180–181
32. Deary IJ, Johnson W, Houlihan LM (2009) Genetic foundations of human intelligence. Hum Genet 126:215–232
33. Benyamin B, Pourcain B, Davis OS, Davies G, Hansell NK et al (2014) Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. Mol Psychiatry 19:253–258
34. Luna A, Nicodemus KK (2007) snp.plotter: an R-based SNP/haplotype association and linkage disequilibrium plotting package. Bioinformatics 23:774–776
35. Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014) Biological insights from 108 schizophrenia-associated genetic loci. Nature 511:421–427

36. Glatt SJ, Faraone SV, Tsuang MT (2003) Meta-analysis identifies an association between the dopamine D2 receptor gene and schizophrenia. Mol Psychiatry 8:911–915
37. Labrie V, Fukumura R, Rastogi A, Fick LJ, Wang W et al (2009) Serine racemase is associated with schizophrenia susceptibility in humans and in a mouse model. Hum Mol Genet 18:3227–3243
38. Egan MF, Straub RE, Goldberg TE, Yakub I, Callicott JH et al (2004) Variation in GRM3 affects cognition, prefrontal glutamate, and risk for schizophrenia. Proc Natl Acad Sci U S A 101:12604–12609
39. Voight BF, Scott LJ, Steinthorsdottir V, Morris AP, Dina C et al (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 42:579–589
40. International Consortium for Blood Pressure Genome-Wide Association Studies, Ehret GB, Munroe PB, Rice KM, Bochud M et al (2011) Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature 478:103–109
41. International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM et al (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460:748–752
42. Chen X, Lee G, Maher BS, Fanous AH, Chen J et al (2011) GWA study data mining and independent replication identify cardiomyopathy-associated 5 (CMYA5) as a risk gene for schizophrenia. Mol Psychiatry 16:1117–1129
43. O'Donovan MC, Norton N, Williams H, Peirce T, Moskvina V et al (2009) Analysis of 10 independent samples provides evidence for association between schizophrenia and a SNP flanking fibroblast growth factor receptor 2. Mol Psychiatry 14:30–36
44. Luo XJ, Li M, Huang L, Steinberg S, Mattheisen M et al (2014) Convergent lines of evidence support CAMKK2 as a schizophrenia susceptibility gene. Mol Psychiatry 19:774–783
45. Breen G, Lewis CM, Vassos E, Pergadia ML, Blackwood DH et al (2011) Replication of association of 3p21.1 with susceptibility to bipolar disorder but not major depression. Nat Genet 43:3–5, author reply 5
46. Vassos E, Steinberg S, Cichon S, Breen G, Sigurdsson E et al (2012) Replication study and meta-analysis in European samples supports association of the 3p21.1 locus with bipolar disorder. Biol Psychiatry 72:645–650
47. Kondo K, Ikeda M, Kajio Y, Saito T, Iwayama Y et al (2013) Genetic variants on 3q21 and in the Sp8 transcription factor gene (SP8) as susceptibility loci for psychotic disorders: a genetic association study. PLoS One 8:e70964
48. Gamazon ER, Badner JA, Cheng L, Zhang C, Zhang D et al (2013) Enrichment of *cis*-regulatory gene expression SNPs and methylation quantitative trait loci among bipolar disorder susceptibility variants. Mol Psychiatry 18:340–346
49. UNESCO (1997) International Standard Classification of Education—ISCED 1997. November 1997, UNESCO, Paris
50. Bourne VJ, Fox HC, Deary IJ, Whalley LJ (2007) Does childhood intelligence predict variation in cognitive change in later life? Personal Individ Differ 42:1551–1559
51. Johnson AR, Craciunescu CN, Guo Z, Teng YW, Thresher RJ et al (2010) Deletion of murine choline dehydrogenase results in diminished sperm motility. FASEB J 24:2752–2761