

Research Article

# Integrative genomics analysis identifies five promising genes implicated in insomnia risk based on multiple omics datasets

Haozhen Sun<sup>1,\*</sup>, Jianhua Zhang<sup>2,\*</sup>,  Yunlong Ma<sup>3,4</sup> and Jingjing Liu<sup>5</sup>

<sup>1</sup>Department of Clinical Pharmacy, the First Affiliated Hospital, Zhejiang University School of Medicine, 79 Qingchun Road, Hangzhou 310003, P. R. China; <sup>2</sup>Department of psychological consulting, the logistics Service Center of Municipal Building, Hangzhou 310019, P. R. China; <sup>3</sup>Institute of Biomedical Big Data, Wenzhou Medical University, Wenzhou 325027, Zhejiang, China; <sup>4</sup>School of Biomedical Engineering, School of Ophthalmology and Optometry and Eye Hospital, Wenzhou Medical University, Wenzhou 325027, Zhejiang, China; <sup>5</sup>Department of Nursing, the First Affiliated Hospital, Zhejiang University School of Medicine, 79 Qingchun Road, Hangzhou 310003, P. R. China

**Correspondence:** Jingjing Liu (1202042@zju.edu.cn) or Yunlong Ma (glb-biotech@zju.edu.cn)



In recent decades, many genome-wide association studies on insomnia have reported numerous genes harboring multiple risk variants. Nevertheless, the molecular functions of these risk variants conveying risk to insomnia are still ill-studied. In the present study, we integrated GWAS summary statistics ( $N=386,533$ ) with two independent brain expression quantitative trait loci (eQTL) datasets ( $N=329$ ) to determine whether expression-associated SNPs convey risk to insomnia. Furthermore, we applied numerous bioinformatics analyses to highlight promising genes associated with insomnia risk. By using *Sherlock* integrative analysis, we detected 449 significant insomnia-associated genes in the discovery stage. These identified genes were significantly overrepresented in six biological pathways including Huntington's disease ( $P=5.58 \times 10^{-5}$ ), Alzheimer's disease ( $P=5.58 \times 10^{-5}$ ), Parkinson's disease ( $P=6.34 \times 10^{-5}$ ), spliceosome ( $P=1.17 \times 10^{-4}$ ), oxidative phosphorylation ( $P=1.09 \times 10^{-4}$ ), and wnt signaling pathways ( $P=2.07 \times 10^{-4}$ ). Further, five of these identified genes were replicated in an independent brain eQTL dataset. Through a PPI network analysis, we found that there existed highly functional interactions among these five identified genes. Three genes of *LDHA* ( $P=0.044$ ), *DALRD3* ( $P=5.0 \times 10^{-5}$ ), and *HEBP2* ( $P=0.032$ ) showed significantly lower expression level in brain tissues of insomniac patients than that in controls. In addition, the expression levels of these five genes showed prominently dynamic changes across different time points between behavioral states of sleep and sleep deprivation in mice brain cortex. Together, the evidence of the present study strongly suggested that these five identified genes may represent candidate genes and contributed risk to the etiology of insomnia.

## Introduction

Insomnia is characterized by persistent dissatisfaction with sleep, which may play central roles in the etiology of physical and mental health [1–3], including suicide [2], depression [4], and post-traumatic stress disorder [5]. In the general population, it is estimated that the prevalence of insomnia is approximately 10–20% [6–8]. Previous twin studies have documented that insomnia and sleep characteristics are highly influenced by genetic factors [9–11]. The heritability rates were estimated to be 59% in females, and 38% in males [12]. Thus, growing studies show considerable interest in identifying the genetic basis of insomnia.

With the advance of technique, genome-wide association study (GWAS) is widely applied and considered as an effective method that could simultaneously examine the genetic association signals from millions of SNPs with complex traits of interest. In recent years, numerous genetic variants based on multiple

\*These authors contributed equally to this work.

Received: 05 April 2020  
Revised: 15 August 2020  
Accepted: 21 August 2020

Accepted Manuscript online:  
24 August 2020  
Version of Record published:  
02 September 2020

GWAS [13–17], have been identified to be associated with insomnia complaints and insomnia symptoms. There were more than 200 genomic loci from these reported GWAS including a wealth of information that are capable of revealing new risk genes and biological pathways implicated in insomnia risk [13–15]. Nevertheless, these biological effects of significant genetic variants from GWAS studies on insomnia risk remain unclear. Meanwhile, in light of the stringent genome-wide threshold of statistical significance employed, there were a great number of genetic loci with small effect size difficult to be identified in a single GWAS study. Therefore, more relevant investigations on uncovering the biological mechanism of genetic variants with small-to-modest effect size may contribute to understanding the missing heritability of insomnia.

Although there are many insomnia-related genetic loci to be detected, accumulating studies with strong evidence have reported that abnormal expression in risk genes plays an important role in the pathogenesis of complex diseases [13,18–21], including insomnia. In addition, multiple studies [22–25] have recently applied systematically integrative methods to combine expression quantitative trait loci (eQTL) data and GWAS summary statistics for identifying the underlying regulatory effect of the reported risk SNPs from previous GWAS. A recent study reported by He and his co-workers [22] demonstrated a Bayesian statistical inference method called *Sherlock* to systematically uncover the cis- and trans-regulatory effects of susceptibility genes on complicate disorders based on GWAS summary statistics data and eQTL data. Based on this effective and powerful approach, a growing number of studies on different phenotypes, such as major depressive disorders [26,27], gout disease [28], and schizophrenia [29], have identified numerous new susceptibility genes, which cannot be detected in GWAS alone.

To the best of our knowledge, there is no systematically integrative study integrating the large-scale GWAS and eQTL data to reveal the insomnia-associated risk genes. Thus, the primary aim of the current investigation is to determine whether expression-associated SNPs could confer risk to insomnia and detect insomnia-associated risk genes by using the *Sherlock* approach based on both eQTL and GWAS data. Furthermore, we adopted quantities of bioinformatics tools based on multiple independent omics data to validate our findings.

## Methods

### Insomnia GWAS summary data

We applied a large-scale insomnia-related GWAS summary dataset from the UK Biobank database [13] for identifying susceptibility SNPs and genes. There is a total of 386,533 individuals with 109,402 cases and 277,131 controls of European descent included in the current investigation. Since the number of 40,000 samples allows for more than 90% high-power detection of the small effect sizes of genetic variants [30], we infer the present study has sufficient power for identifying risk genetic variants. In this chosen population, the insomnia prevalence was 28.3%. All included subjects signed informed consent. Both phenotypic information and DNA samples were collected from participants. The phenotype of insomnia was recorded according to the following question: “Do you have trouble falling asleep at night or do you wake up in the middle of the night?” The answers of the question were provided to subjects for choosing: “A: Never/rarely, B: Sometimes, C: Usually, D: Prefer not to answer. Genotype data were based on the combined UK10K and 1000 Genome Projects reference panel (hg19) for imputation. In total, there were a number of 10,862,567 genetic variants and related *P* values employed as input in current *Sherlock* Bayesian inference analysis.

### GWAS data based on Null phenotype

In order to avoid the influence of random events, we employed published GWAS data ( $N=3960$ ) [31] to construct a fake insomnia-based GWAS data as a negative control. With regard to the constructed GWAS dataset, we randomly assigned the disease status (namely, insomnia or control) into each individual of 3960 samples by using the method of `RANDBETWEEN` (“insomnia”, “control”) in the Microsoft Excel. We called the randomly assigned insomnia as Null trait. The statistical analysis utilized the widely used tool, namely PLINK v1.07 [32], based on the logistic regression model. In light of the assumption that there were no true genetic effects of the GWAS on Null phenotype, the relatively small sample size is not an issue.

### Brain eQTL data

Considering that insomnia is the second most prevalent mental disorder that may be susceptible to the aberrant function of the brain, it is plausible to assume that brain tissue is the most suitable sample for integrative analysis of identifying insomnia-associated risk genes. Thus, we employed human brain prefrontal cortex tissue-based expression data from a previously reported study by Myers and co-workers [33] to perform an eQTL analysis. There were 193 neuropathologically normal brain samples without clinical history of psychiatric phenotypes or other neurologic traits included in subsequent analysis. All enrolled participants were European origins based on self-report.

The expression profiles were obtained with the adoption of an Illumina HumanRefseq-8 Expression BeadChip, and the genome-wide genotype data were genotyped with the use of an Affymetrix Human Mapping 500K Array Set.

## The Bayesian statistical method of Sherlock analysis

Since the vast majority of disease-associated genetic variants identified by GWAS are mapped in non-coding genomic regions [34], it is reasonable to infer that these identified risk variants in non-coding regions are more likely to affect the expression level of a specific gene rather than the molecular function of its protein. Thus, we employed a *Sherlock* Bayesian-based inference analysis to integrate GWAS summary data on insomnia from Jansen and co-workers [13]. The procedures of *Sherlock* inference approach are described as following steps: first, the *Sherlock* tool will identify all SNPs that show significant association with gene expression (called as eSNP) from chosen eQTL data based on brain cortical samples reported by Myers and co-workers [33]. After the definition of eSNP, *Sherlock* will examine the association between eSNPs and insomnia with the use of GWAS summary data from Jansen et al. [13]. A positive score would be given to an eSNP if this eSNP is also significantly associated with insomnia based on GWAS data. A negative score would be recorded if this eSNP shows a non-significant association with insomnia. There would be no score to be recorded if the SNP only shows significant association with insomnia but no prominent signal for alterations in gene expression. The total score of a specific gene is based on the score of each eSNP by combining evidence from GWAS and eQTL data. The logarithm of the Bayes factor (LBF) for a specific gene is a crucial indicator to determine whether the gene contributes to insomnia risk. The LBF is computed with the use of *Sherlock* Bayesian-based analysis by integrating the evidence from GWAS summary statistics and eQTL. The larger value of LBF represents the higher probability that the gene can convey risk to insomnia. Considering that existing traditional analyses often ignore SNPs with moderate effect size, *Sherlock* Bayesian-based analysis is an effective approach to systematically integrate SNP with moderate-to-strong effect size from GWAS and eQTL data. Bonferroni correction method was employed to correct the *P* values.

## Pathway-based enrichment analysis

To explain the biological function of the prioritized insomnia-related risk genes from *Sherlock* Bayesian-based analysis, we utilized an easy use plug-in of Cytoscape platform [35] called ClueGO [36] to generate a functional organized pathway-term network. First, we performed a pathway-based analysis depending on a popular public source of the Kyoto Encyclopedia of Genes and Genomes (KEGG). By using over-representation analysis, we could identify and prioritize functional associations between chosen genes and biological pathways. Furthermore, we performed a Gene Ontology (GO) analysis including three categories of GO terms: molecular function, cellular component, and biological process. The method of “*GO Term Fusion*” was used to reduce the redundancies among GO terms. The hypergeometric test was used to calculate *P* value and Bonferroni step down correction was used for multiple testing.

## Phenotype- and drug-related gene set enrichment analysis

To explore whether identified genes were significantly enriched in gene sets related to phenotypes or drugs, we utilized the web-based tool of WebGestalt [37] to perform functional enrichment analysis based on the resources of GLAD4U [38], DrugBank [39], and the Human Phenotype Ontology [37]. The web-access tool has three main functions of over-representation analysis, network topology-based analysis, and gene set enrichment analysis. Here we used the function of overrepresentation analysis to analyze the submitted genes identified from the *Sherlock* integrative analysis. Current enrichment analysis utilized all genome protein-coding genes as background genes. We only selected the gene size of each gene set ranging from 5 to 2000 for the current analysis. The method of the Benjamini–Hochberg false discovery rate was employed for adjustment.

## Validation eQTL datasets using Sherlock integrative analysis

To further replicate the authenticity of these identified insomnia-associated genes, we re-performed the *Sherlock* integrative analysis with the use of an independent eQTL dataset (i.e. 136 brain cortex samples) from GTEx portal (data release v7) [40]. RNA-sequencing was used for quantifying the level of RNA expression, and the Illumina OMNI 5M SNP Array was used for SNP genotyping. Based on the additive genetic model, eQTL analysis was conducted through the tool of Matrix eQTL [41] using linear regression analysis. For *Sherlock* Bayesian-based analysis, all the parameters were set to be the same with those in discovery eQTL data. The *P* values calculated from *Sherlock* were corrected with the application of the Bonferroni correction.

## MAGMA gene-based analysis

Furthermore, we intended to perform a gene-based enrichment analysis of the above-used GWAS summary dataset on insomnia [13] by adopting an independent approach of the Multi-marker Analysis of GenoMic Annotation (MAGMA). The IDs and *P* values of all SNPs were utilized as submitted information for the MAGMA tool. For revealing the multi-variant convergent genetic effects, the multiple regression model was utilized to integrate the linkage disequilibrium among SNPs within a specific genomic region. The definition that an SNP belongs to a specific gene depended on the location of the SNP. Whether it's mapped into the gene body or a region extended  $\pm 20$  kb downstream or upstream of the gene [42]. More detailed illustrations of the MAGMA are demonstrated in the official website of <https://ctg.cncr.nl/software/magma>. The SNP–SNP linkage disequilibrium information was computed as reference for the 1000 Genome European Panel, and the location of each SNP in the present study was referred to as the Human Genome Build 37.

## PPI network-based analysis

Numerous studies have been published to show that susceptibility genes for complex diseases are predisposed to be collectively interacted [20,43,44]. Furthermore, network-based analytic approaches have been widely used to search for functional patterns of identified genes associated with traits of interest [45,46]. Therefore, we performed a protein–protein interaction (PPI) network-based analysis of these identified insomnia-associated risk genes by using the GeneMANIA software [47], which is a user-friendly tool for speculating the functions of inputted genes and prioritizing the promising genes for further molecular experiments. This tool could extend the identified genes with functionally similar genes by integrating available proteomics and genomics data.

## Replication of candidate gene expression in brain tissue of insomnia patients

Under the assumption that aberrant expression of genes may convey risk to complex diseases, *Sherlock* Bayesian-based analysis is used to identify disease-associated risk genes. To determine whether these five identified susceptibility genes' expression show a significant difference between insomnia and control brain samples, we downloaded one available RNA expression dataset by using Affymetrix Human Genome U133+ 2.0 chip from the NCBI's GEO database (accession number: GSE40562) and performed a differential gene expression (DGE) analysis. For this dataset [19], there were three insomnia patients enrolled in the present study, and total RNA of the thalamus and the parietal cortex of insomnia patients were extracted using RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol. The Ethical Committee of National Institute for Viral Disease Prevention and Control, China CDC, approved for those human brain samples used in investigation. More detailed information on these samples including their genetic, pathogenic, and neuropathological features was reported in previous studies [48,49]. The web-based tool of GEO2R [50] was used to calculate the expression difference between insomnia and control. *P*-value < 0.05 was considered to be significant. The co-expression patterns of five identified genes were analyzed by the Pearson correlation analysis, and the *Corrplot* R package was used for visualization. The R script used for this analysis is shown in the github website (<https://github.com/mayunlong89/insomnia/blob/master/coexpression.r>).

## Temporal changes in the expression of candidate genes in mice brain cortex

To explore whether these identified candidate genes have significantly temporal changes in expression patterns in brain cortex between sleep and wake (sleep deprivation), we performed a differential expression pattern analysis at different time points by downloading the RNA expression data from NCBI GEO database (accession number: GSE6514). For this dataset, the male mice (C57BL/6J) at age of  $10 \pm 1$  week were used in the experiments. With a dark–light cycle of 12 h, chosen mice were housed in a pathogen-free, humidity- and temperature-controlled room. These male mice were subjected to 14 days of acclimatization for establishing a nighttime feeding pattern. The detailed experimental information is recorded in a previous article [51]. The sleep deprivation procedure was initiated at lights-on with gentle handling. Five sleep-deprived mice were killed at each time point of 3, 6, 9, and 12 h of total sleep deprivation. Similarly, five undisturbed sleeping mice were killed at the same diurnal time points (i.e. 3, 6, 9, and 12 h) as sleep-deprived mice. Additionally, at the time of lights-on at 7:00 AM called as anchor time, five control mice were killed. The Affymetrix GeneChip Mouse Genome 430 2.0 array with >45,000 probe sets and ~34,000 well-annotated genes. The intensity data of probes were analyzed with the application of the affy package of the R software, which is used to evaluate the quality of expression and generate summary measures of expression. Student's *t*-test was used to calculate the expression difference between sleep and wake at each time point.

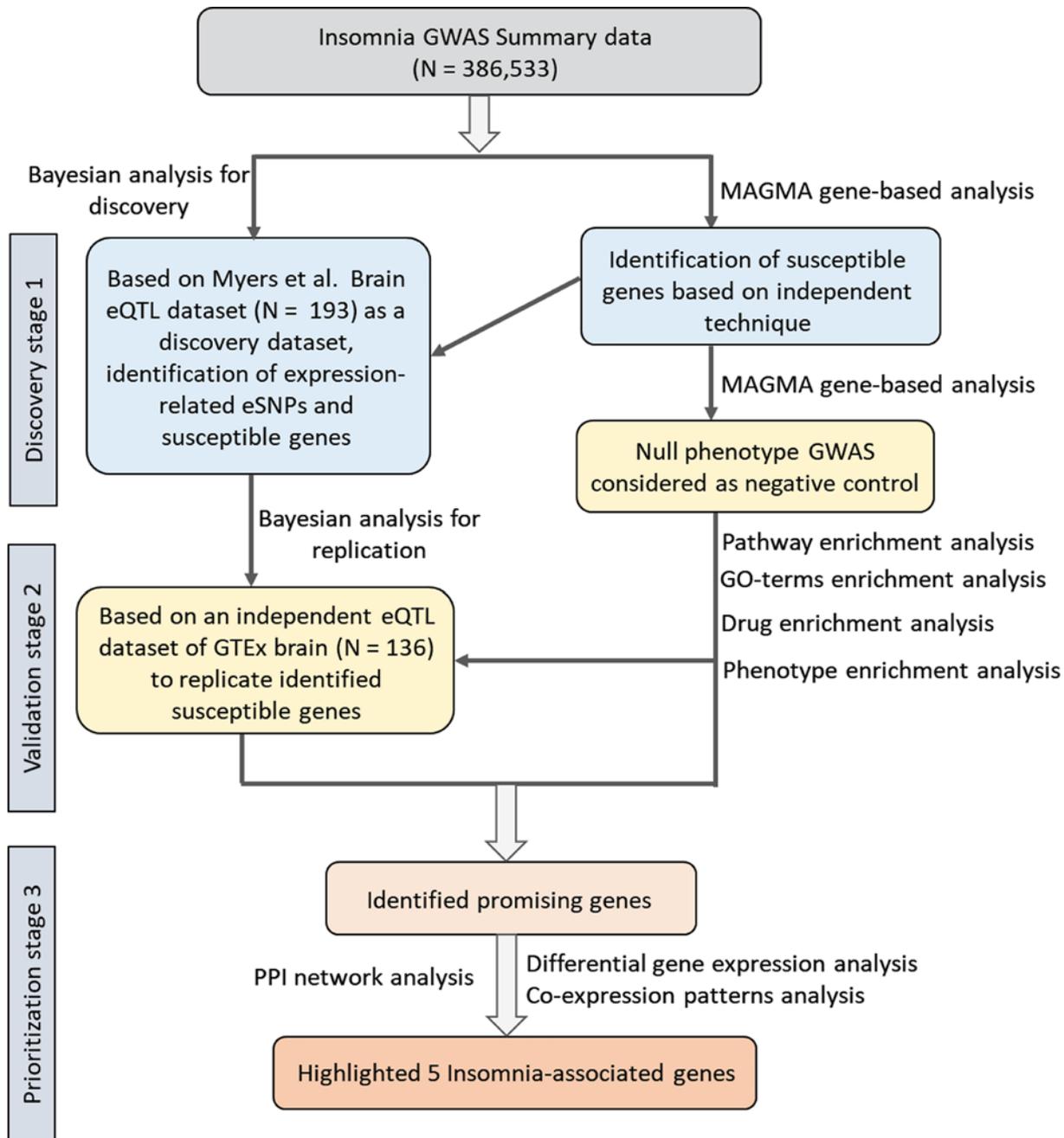


Figure 1. The workflow of identifying and prioritizing the insomnia risk genes

## Results

### Sherlock Bayesian-based analysis prioritizes insomnia-associated risk genes

The workflow of our current study is shown in Figure 1. First, we employed *Sherlock* analysis to explore the association between SNP and expression by integrating insomnia-related GWAS summary statistics based on 386,533 samples and brain eQTL data based on 193 samples. Based on this Bayesian method of *Sherlock*, a total of 449 genes were found to be nominally significantly associated with insomnia risk by alteration in its expression ( $P$ -value  $< 0.05$ , Table 1, and Supplementary Table S1). For example, the top-ranked genes of *FOXF2* (simulated  $P=8.89 \times 10^{-6}$ ), *FAM193A* (simulated  $P=1.22 \times 10^{-4}$ ), *PAIP1* (simulated  $P=3.33 \times 10^{-4}$ ), *MPG* (simulated  $P=3.67 \times 10^{-4}$ ),

**Table 1** Top-ranked 20 insomnia-associated genes identified by *Sherlock* integrative analysis

Gene symbol	Supporting SNP <sup>1</sup>	eSNP-based <i>P</i> value <sup>2</sup>	GWAS-based <i>P</i> value <sup>3</sup>	<i>Sherlock</i> -based <i>P</i> value <sup>4</sup>
<i>FOXF2</i>	rs4836269	$9.96 \times 10^{-6}$	0.067	$8.89 \times 10^{-6}$
<i>FAM193A</i>	rs7109421	$9.92 \times 10^{-6}$	0.010	$1.22 \times 10^{-4}$
<i>PAIP1</i>	rs11173336	$9.87 \times 10^{-6}$	0.018	$3.33 \times 10^{-4}$
<i>MPG</i>	rs1867116	$9.96 \times 10^{-6}$	0.076	$3.67 \times 10^{-4}$
<i>INO80</i>	rs4477668	$9.72 \times 10^{-6}$	0.49	$5.00 \times 10^{-4}$
<i>VPS13B</i>	rs6017342	$9.74 \times 10^{-6}$	0.61	$5.98 \times 10^{-4}$
<i>TGFB3</i>	rs2372321	$9.99 \times 10^{-6}$	0.029	$7.78 \times 10^{-4}$
<i>LOC283537</i>	rs9908305	$9.83 \times 10^{-6}$	0.076	$8.00 \times 10^{-4}$
<i>LOC155340</i>	rs810517	$9.54 \times 10^{-6}$	0.29	$9.16 \times 10^{-4}$
<i>GLUL</i>	rs1866877	$9.70 \times 10^{-6}$	0.034	$1.05 \times 10^{-3}$
<i>RPS17</i>	rs7810180	$9.65 \times 10^{-6}$	0.0028	$1.22 \times 10^{-3}$
<i>ZNF621</i>	rs11135930	$9.96 \times 10^{-6}$	0.040	$1.33 \times 10^{-3}$
<i>SMYD5</i>	rs951821	$9.80 \times 10^{-6}$	0.056	$1.34 \times 10^{-3}$
<i>PLLP</i>	rs418682	$9.55 \times 10^{-6}$	0.82	$1.46 \times 10^{-3}$
<i>RWDD2B</i>	rs8098365	$9.67 \times 10^{-6}$	0.57	$1.60 \times 10^{-3}$
<i>CAT</i>	rs2172962	$9.42 \times 10^{-6}$	0.59	$1.74 \times 10^{-3}$
<i>SMEK1</i>	rs286451	$9.80 \times 10^{-6}$	0.97	$1.89 \times 10^{-3}$
<i>NDUFS6</i>	rs1218887	$9.36 \times 10^{-6}$	0.16	$1.90 \times 10^{-3}$
<i>LYL1</i>	rs4694022	$9.28 \times 10^{-6}$	$4.19 \times 10^{-6}$	$2.00 \times 10^{-3}$
<i>FAF1</i>	rs4775953	$9.30 \times 10^{-6}$	0.0023	$2.11 \times 10^{-3}$

<sup>1</sup>SNP influences the expression level of risk gene.

<sup>2</sup>*P*-value from expression quantitative trait analysis of Myers et al.

<sup>3</sup>*P*-value from GWAS on insomnia of Jansen et al.

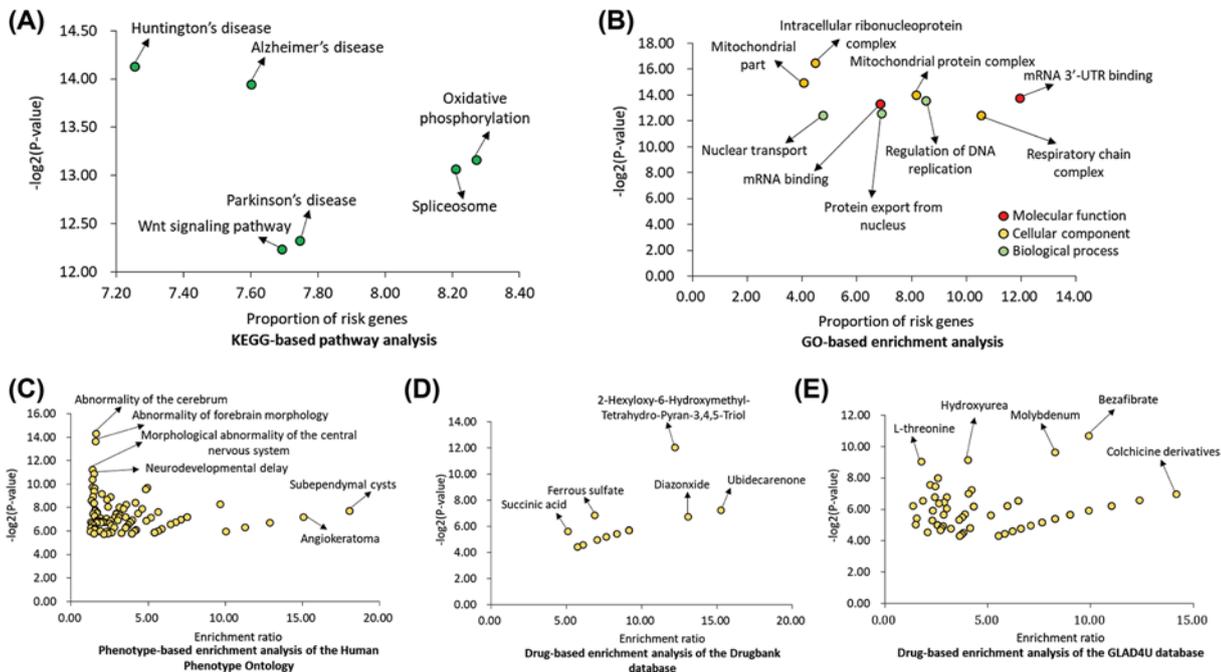
<sup>4</sup>*P*-value from calculation based on the *Sherlock* Bayesian integrative analysis. In light of *Sherlock* tool uses finite times of permutation test to calculate the *P* value for each gene, some of these top-ranked genes have distinct logarithm of the Bayes factor values but obtain the same rankings (namely, their *P* values are same).

*INO80* (simulated  $P = 5.00 \times 10^{-4}$ ), *VPS13B* (simulated  $P = 8.89 \times 10^{-6}$ ), and *TGFB3* (simulated  $P = 8.89 \times 10^{-6}$ ) with supportive eSNPs conveying risk to insomnia (Table 1). Among them, 20 genes that are associated with insomnia or sleep-related phenotypes have been documented in the GWAS catalog database (Supplementary Table S1). For example, Spada and co-workers reported the rs62388641 in the *FOXF2* gene ( $P=1.0 \times 10^{-6}$ ) is suggestively associated with daytime sleep phenotypes [52].

## Identification of significantly enriched pathways

Subsequently, we conducted a pathway analysis based on the KEGG source for these identified 449 insomnia-associated genes. There were six biological pathways significantly enriched by these inputted genes (Figure 2A and Supplementary Table S2; Corrected *P*-value < 0.05). Interestingly, these overrepresented pathways have been well-documented to be implicated in psychiatric disorders or neurodegenerative diseases [53–59]. For example, the pathways of Huntington's disease ( $P=5.58 \times 10^{-5}$ ), Alzheimer's disease ( $P=5.58 \times 10^{-5}$ ), Parkinson's disease ( $P=6.34 \times 10^{-5}$ ), spliceosome ( $P=1.17 \times 10^{-4}$ ), oxidative phosphorylation ( $P=1.09 \times 10^{-4}$ ), and wnt signaling pathway ( $P=2.07 \times 10^{-4}$ ). Furthermore, we carried out a GO enrichment analysis according to three categories of GO terms. With respect to the category of molecular function (Figure 2B and Supplementary Table S3), we found that these identified insomnia-associated genes were significantly enriched in mRNA 3'-UTR binding ( $P=7.39 \times 10^{-5}$ ) and mRNA binding ( $P=1.01 \times 10^{-4}$ ). With regard to the category of cellular component (Figure 2B), five terms were significantly enriched; for example, mitochondrial part ( $P=3.22 \times 10^{-5}$ ), mitochondrial protein complex ( $P=6.08 \times 10^{-5}$ ), and mitochondrial membrane part ( $P=1.01 \times 10^{-4}$ ). For the category of biological process (Figure 2B), we observed three significantly enriched terms: regulation of DNA replication ( $P=8.28 \times 10^{-5}$ ), protein export from nucleus ( $P=1.65 \times 10^{-4}$ ), and nuclear transport ( $P=1.82 \times 10^{-4}$ ).

In addition, we utilized the WebGestalt software to conduct phenotype- and drug-based enrichment analysis. Interestingly, with regard to phenotype-focused enrichment analysis, we found a number of enriched gene sets related to several defined phenotypes (Figure 2C and Supplementary Table S4). For example, the gene sets were related to abnormality of the cerebrum ( $P=4.88 \times 10^{-5}$ ), abnormality of forebrain morphology ( $P=7.66 \times 10^{-5}$ ), morphological abnormality of the central nervous system ( $P=4.17 \times 10^{-4}$ ), and neurodevelopmental delay ( $P=5.19 \times 10^{-4}$ ). For



**Figure 2. Functional enrichment analysis of insomnia-associated genes identified by Sherlock Bayesian analysis**

(A) Pathway enrichment analysis based on the KEGG database. (B) GO-terms enrichment analysis based on three wide-used terms of molecular function, cellular component, and biological process. (C) Phenotype-based enrichment analysis of the Human Phenotype Ontology based on the WebGestalt tool. (D) Drug-based enrichment analysis of the Drugbank database based on the WebGestalt. (E) Drug-based enrichment analysis of the GLAD4U database based on the WebGestalt.

drug-focused enrichment analysis, 83 gene sets related to drugs were significantly enriched based on two widely-used databases of DrugBank (Figure 2D and Supplementary Table S5) and GLAD4U (Figure 2E and Supplementary Table S6).

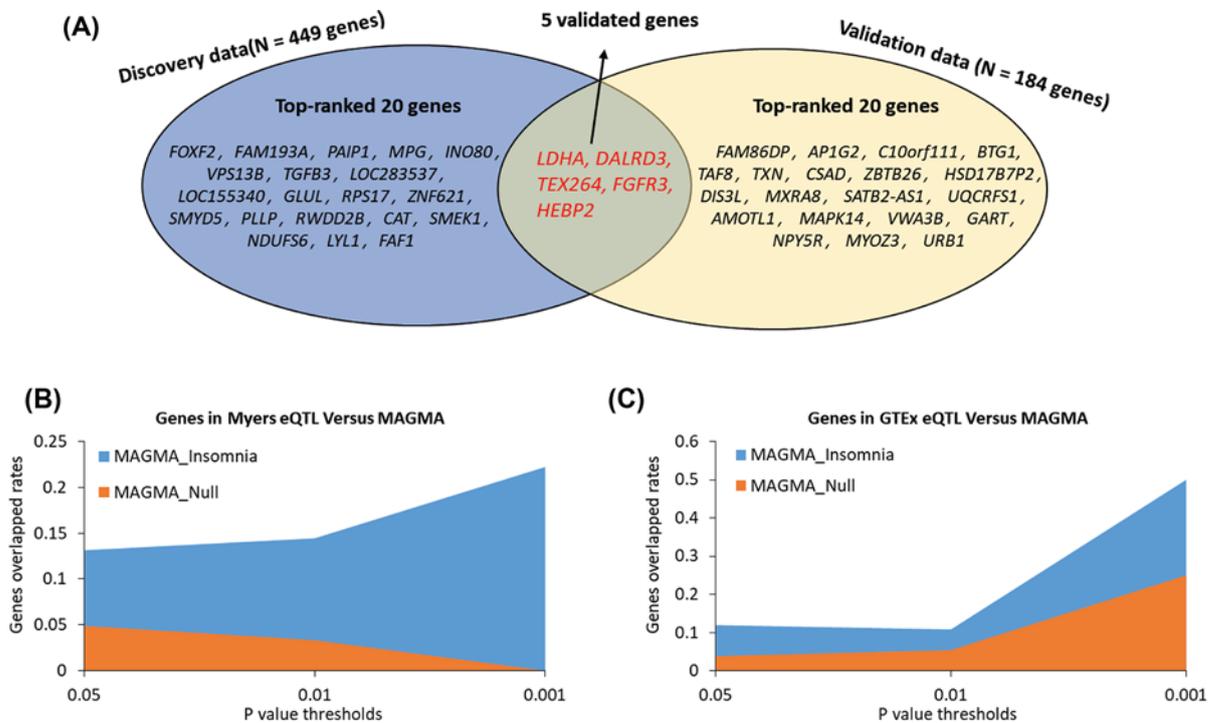
## Replication of identified risk genes using an independent eQTL dataset

For validation of above identified genes, we then reconducted the *Sherlock* Bayesian-based integrative analysis with the same parameter settings using independent brain cortex eQTL data ( $N=136$ ). *Sherlock* Bayesian-based analysis identified 184 significant insomnia-associated genes (Supplementary Table S7). There were two genes of *C6orf201* and *AK5* reported to be associated with sleep-related traits [52,60]. By compared with genes identified from the discovery stage, we found that five significant genes were overlapped between discovery and replication stage; namely, *HEBP2* (simulated  $P=0.015$ ), *LDHA* (simulated  $P=0.018$ ), *TEX264* (simulated  $P=0.02$ ), *FGFR3* (simulated  $P=0.023$ ), and *DALRD3* (simulated  $P=0.029$ ) (Figure 3A and Supplementary Table S8).

To further ensure the reliability of current investigation, we conducted MAGMA gene analyses for GWAS on insomnia and GWAS on Null phenotype (referred to as negative control). For comparing the findings between real and fake data, we employed three distinct thresholds of  $P$  values: 0.05, 0.01, and 0.001. At each threshold, we observed that *Sherlock*-identified genes (Myers eQTL and GTEx eQTL) were obviously higher overlapped with MAGMA-identified genes from GWAS on insomnia than those from GWAS on Null phenotype (Figure 3B,C).

## PPI network-based analysis of five insomnia-risk genes

To determine whether identified insomnia-risk genes functionally interacted together, we performed a PPI network-based analysis with the application of interactions of physical interactions, co-expression, predictions, pathways, and shared protein domains based on the well-documented database of GeneMANIA [47]. Figure 4 demonstrated that these identified insomnia-risk genes is generated a biological network, suggesting that there exist highly functional links among these identified risk genes. For example, two hub genes of *LDHA* and *FGFR3* have the most interactions with other genes (Figure 4). Additionally, the gene of *TEX264* shows evidence of shared protein domains with the insomnia-risk gene of *HEBP2*, and has co-expression evidence with *DALRD3* (Figure 4)



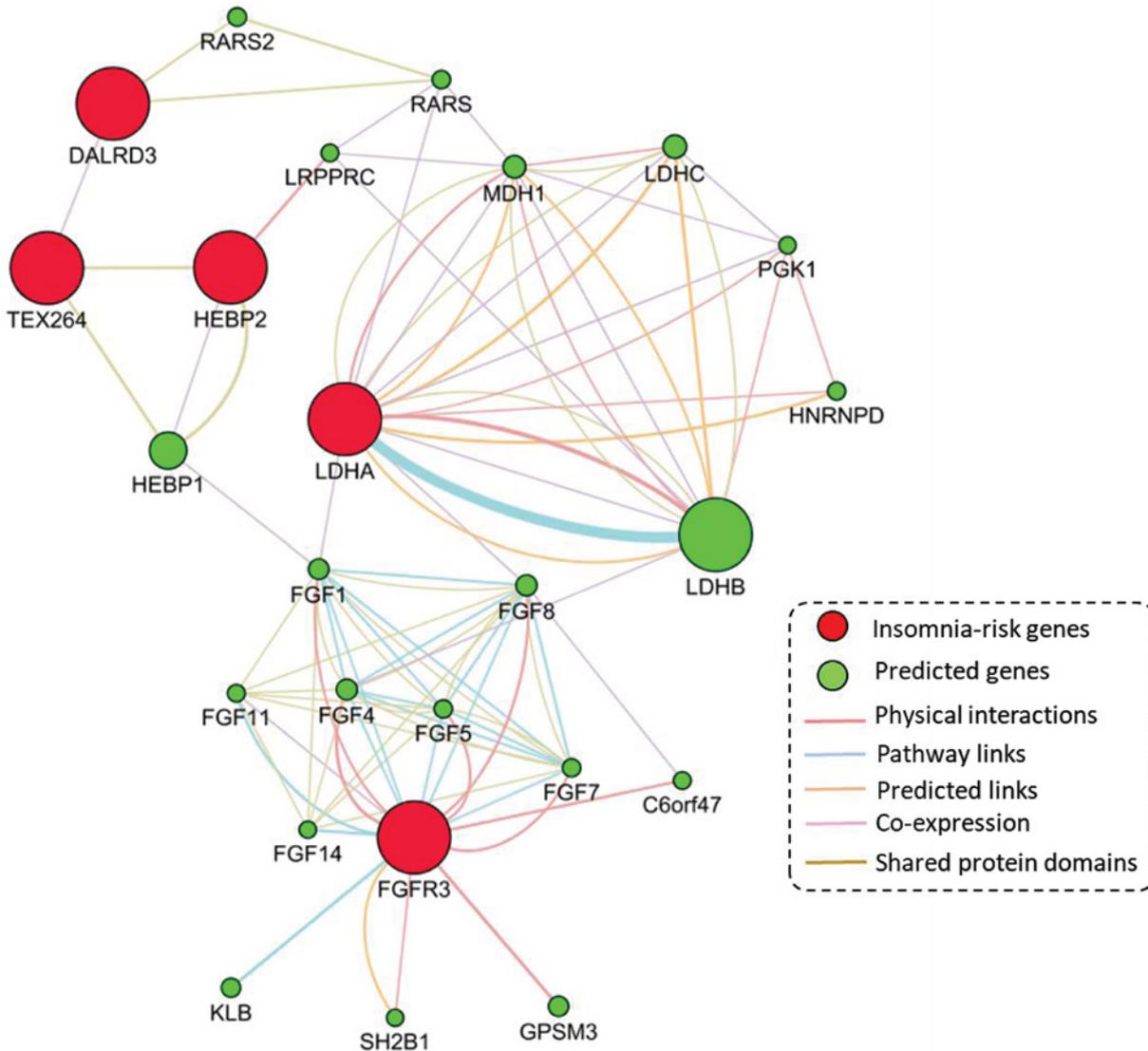
**Figure 3. Identification of insomnia risk genes based on two independent eQTL datasets**

(A) Venn plot demonstrated the overlapped genes between discovery (Myers eQTL) and replication stage (GTEx eQTL). (B) Sherlock-identified genes from Myers eQTL data (the discovery stage) were obviously higher overlapped with genes identified from MAGMA analysis of GWAS on insomnia than those from MAGMA analysis of GWAS on null phenotype. (C) Sherlock-identified genes from GTEx eQTL data (the replication stage) were obviously higher overlapped with genes identified from MAGMA analysis of GWAS on insomnia than those identified from MAGMA analysis of GWAS on null phenotype.

## Differential expression of identified risk genes between insomnia and control brain samples

To determine whether the co-expression patterns among five identified genes were altered by disease status, we carried out a Pearson correlation analysis with the use of *corrplot* package for a visualization based on the RNA expression data from GSE40562 from the NCBI GEO database. Interestingly, we found prominent alterations of the co-expression relationships among five genes categorized by insomnia status (Figure 5A,B and Supplementary Tables S9–S10). For example, the positive correlation coefficient of *LDHA* with *DALRD3* was 0.62 in all samples, but it was largely reduced to 0.06 in insomniac patients. The negative correlation score of *LDHA* with *TEX264* was reduced from  $-0.40$  in all samples to  $-0.18$  in insomniac patients. Similarly, the negative correlation score between *DALRD3* and *TEX264* was decreased from  $-0.48$  in all samples to  $-0.31$  in insomniac patients. In addition, the co-expression correlation between *DALRD3* and *HEBP2* fundamentally changed from 0.54 in all samples to  $-0.24$  in insomniac patients.

Furthermore, by conducting a DGE analysis, we found that three genes of *DALRD3* ( $P=5.0 \times 10^{-5}$ ), *LDHA* ( $P=0.044$ ), and *HEBP2* ( $P=0.032$ ) showed significantly down-regulated expression in insomnia brain samples compared with controls (Figure 5C–E and Supplementary Table S8). Unfortunately, there was no significant evidence for both *FGFR3* and *TEX264* between insomnia patients and controls (Supplementary Table S8). In addition, based on the dataset of GSE6514, we found that the expression patterns of these five genes between sleep and sleep deprivation remarkably changed across different time points in mice brain cortex (Figure 6A–E). For example, the gene of *DALRD3* has similar decreased expression patterns of sleep and wake states before the time point of 6 h, but subsequently the gene expression is prominently increased in wake state and still decreased in sleep state (Figure 6A).

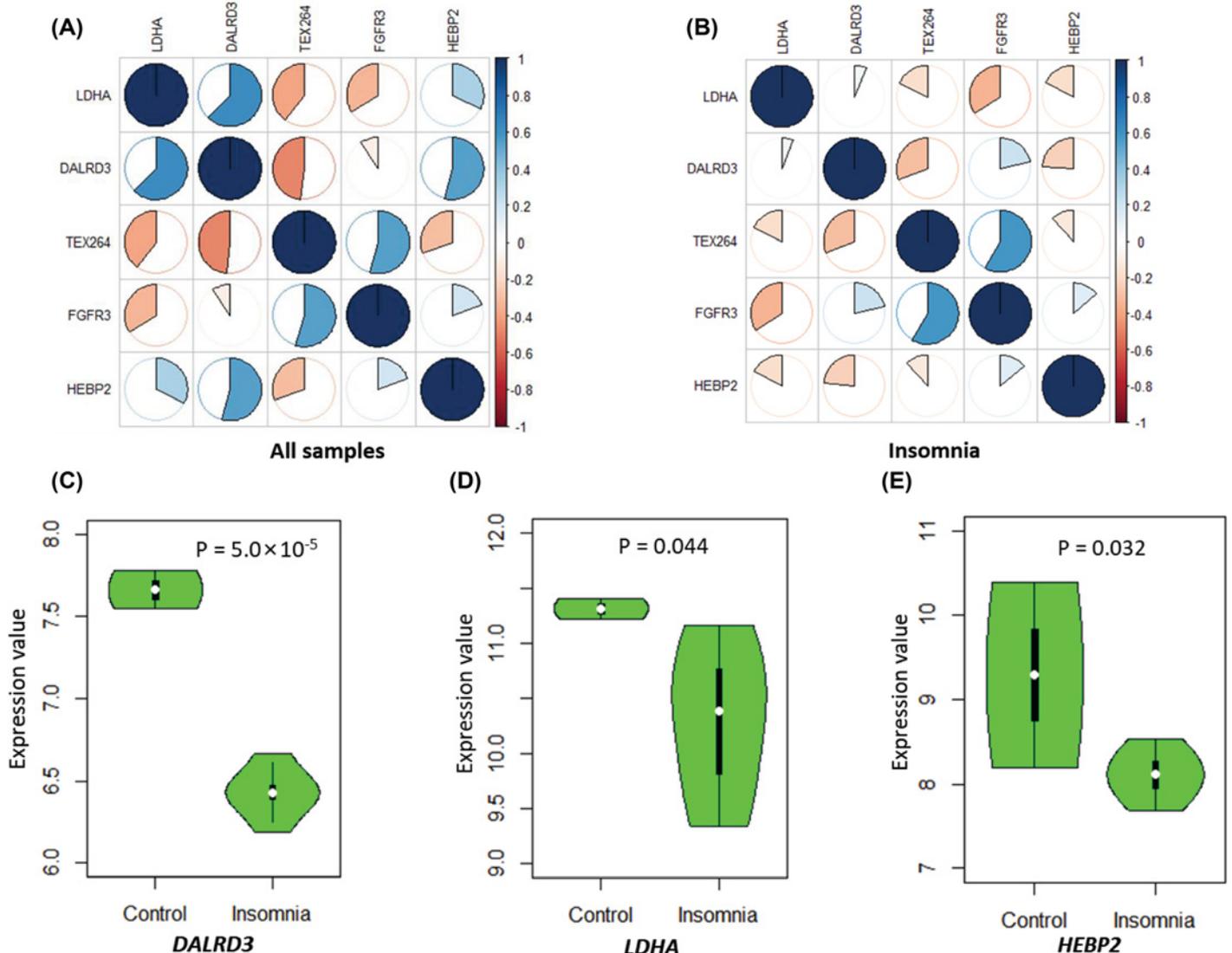


**Figure 4. Protein-protein interactions network constructed by using five identified insomnia risk genes**

The five insomnia-associated risk genes were provided as query (red nodes) and a number of additional genes were predicted to be linked (green nodes). The interactions evidence was based on physical interactions, pathway links, predicted links, co-expression, and shared protein domains.

## Discussion

Insomnia disorder is one of the most prevalent mental disorders worldwide [61]. Multiple lines of evidence from family and twin studies have been reported to suggest the involvement of genetic components in the pathogenesis of insomnia [9–11]. To date, a few insomnia-related linkages and candidate genetic association studies with relatively small sample sizes have been carried out. With the development of high-throughput genotyping technologies and the increase in sample size, numerous highly significant genetic variants among more than 200 genomic loci have been identified to be associated with insomnia by using the GWAS approach [13–17]. However, it remains equivocal how these identified genetic variants convey susceptibility to insomnia. In view of suffering the influence of linkage disequilibrium, the vast majority of reported susceptibility loci contain many highly linked genetic variants with similar significant association signals, enhancing the difficulty to find the authentic causal variants. As we all know, most of the GWAS-identified genetic variants are mapped into non-coding regions of the genome [62,63]. Thus, there is a tough conundrum from GWAS left which is how a genetic variant change in the non-coding region could

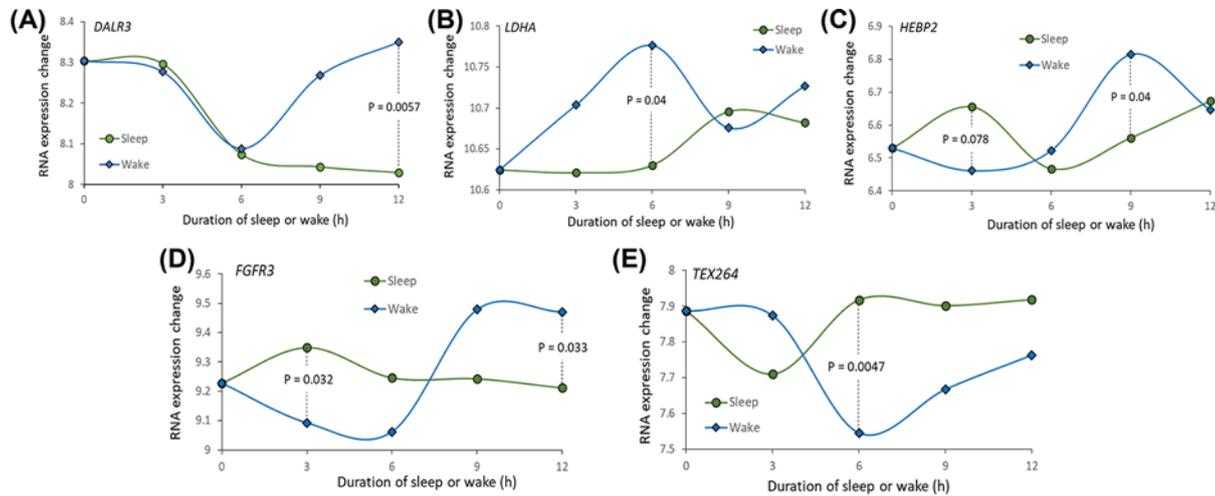


**Figure 5. The differential expression patterns in the brain between insomnic patients and controls**

(A) Co-expression patterns of five susceptible genes in all samples based on the Pearson correlation analysis. (B) Co-expression patterns of five susceptible genes in insomnic patients based on the Pearson correlation analysis. (C and D) Boxplots show differential expression signatures of identified genes in the brain between insomnic patients and controls. (C) for *DALRD3*, (D) for *LDHA*, and (E) for *HEBP2*.

contribute to increased risk to a specific phenotype, such as insomnia. One possible explanation is that these identified genetic variants in non-coding regions are more likely to result in alterations of gene expression levels rather than in alterations of protein functions [34,64].

To address this issue, many researchers have developed various bioinformatics statistical approaches for data integration of the genetic variants information from GWAS summary statistics and eQTL datasets [22,64,65]. In the present study, we performed a *Sherlock* Bayesian-based integrative analysis of two independent brain eQTL datasets ( $N=329$ ) and GWAS summary data on insomnia ( $N=386,533$ ) to identify candidate gene conferring susceptibility to insomnia. The approach of GWAS scanning tens of millions of genetic variants is extensively applied to identify genomic regions harboring common SNPs that are significantly associated with traits of interest, including insomnia. Nevertheless, the burden of correction for multiple-testing from millions of SNPs remarkably limits the power of GWAS to confirm the associated risk SNPs. Many SNPs with small-to-modest effects that do not reach a genome-wide significance but is still pivotal to insomnia may be ignored by GWAS. In addition, GWAS alone could not infer whether



**Figure 6. The temporal changes in the expression of five identified risk genes in the brain between behavioral states of sleep and sleep deprivation**

(A) for *DALRD3*, (B) for *LDHA*, (C) for *HEBP2*, (D) for *FGFR3*, (E) for *TEX264*.

these identified disease-associated SNPs have functional consequences. Thus, *Sherlock* Bayesian analysis used in the current investigation is an effective approach for identifying novel risk genes based on genetic and expression information, and has been widely employed in numerous diseases [22,26–29].

In the discovery stage, we observed 499 genes whose alterations in expression may be implicated in the etiology of insomnia by using *Sherlock* analysis. Numerous identified genes have been demonstrated to be associated with sleep-related traits or insomnia in previous studies. For example, the genes of *FOXF2* [52], *PLLP* [66], and *WWC1* [52]. More interestingly, these identified genes were over-represented in six biological pathways, which have been widely implicated in the etiology of neuropsychiatric or neurological disorders [53–59,67], including insomnia-related disease [68,69]. Of these pathways, three pathways of Huntington's disease, Alzheimer's disease, and Parkinson's disease were derived from neurodegenerative disorders, which is the umbrella term for the progressive loss of structure or function of neurons. For example, Huntington's disease is a neurodegenerative genetic disorder that leads to mental decline and behavioral symptoms [70]. Also, the other three significantly enriched pathways of spliceosome, oxidative phosphorylation, and wnt signaling pathway have been extensively reported to be involved in neurodegenerative disorders, such as autism [71–73], Alzheimer's disease [74,75], Parkinson's disease [76–78], and Huntington's disease [79,80]. Consistently, previous studies [60] have reported that there existed shared genetics between sleep disturbance traits and neuropsychiatric disorders. Consistently, our phenotype-focused enrichment analysis also found these *Sherlock*-identified genes were significantly enriched in several phenotypes relevant to neurodevelopment or brain morphological abnormality, including abnormality of the cerebrum, abnormality of forebrain morphology, and morphological abnormality of the central nervous system. Thus, our results provided supportive evidence that insomnia has an impact on the quality of life in patients and may involve in the pathogenesis of various neurodegenerative disorders.

To validate and prioritize these identified genes, we reconducted the *Sherlock* Bayesian analysis with the use of an independent brain eQTL dataset. There were five promising genes of *DALRD3*, *LDHA*, *HEBP2*, *TEX264*, and *FGFR3* replicated. Meanwhile, we detected that the *Sherlock*-identified genes from both discovery and replication stage were prominently higher overlapped with MAGMA-discovered genes from GWAS on insomnia than that from GWAS on Null phenotype, indicating that these *Sherlock*-identified genes associated insomnia risk are probably due to genetic components rather than random chances. Through conducting the PPI network analysis, we noticed that these five genes collectively interacted with each other. The insomnia-associated gene of *HEBP2* has shared protein domains with the insomnia-associated gene of *TEX264* [81], and *TEX264* show evidence of co-expression with *DALRD3* [82]. The insomnia-associated gene of *LDHA* has co-expression links with predicted genes of *PGK1* and *RARS* [83]. By using co-expression analysis, we found that the gene–gene co-expression patterns were significantly different between insomniac patients and controls. Furthermore, we observed *DALRD3*, *LDHA*, and *HEBP2* are significantly lower expressed in insomniac patients than those in controls. Consistently, these five identified genes showed

differential expression patterns between sleeping duration and sleep deprivation at different time points. Together, these findings provided consistent evidence to support that these newly identified genes have roles in the etiology of insomnia risk and may represent therapeutic targets for treating insomnia.

In the current investigation, by using numerous bioinformatics analyses including *Sherlock* Bayesian analysis, MAGMA gene enrichment analysis, pathway enrichment analysis, drug-based enrichment analysis, PPI network-based analysis, co-expression analysis, DGE-based analysis, and differential time point-based expression analysis based on multilayer data from various sources, we identified five novel susceptible genes associated with insomnia risk. With regard to five identified genes, there existed a bulk of evidence to demonstrate their molecular functions. The protein, encoded by *LDHA* gene, catalyzes the conversion of L-lactate and NAD to pyruvate and NADH in the final step of anaerobic glycolysis. The *LDHA* gene has been reported to be implicated in various neurodegenerative disorders [84,85]. For example, Newington and co-workers [84] have shown that the overexpression of *LDHA* in a rat B12 cell line conveys resistance to amyloid  $\beta$  and other neurotoxins, which may elucidate why some people tolerate high levels of amyloid  $\beta$  deposition without the development of Alzheimer's disease. Furthermore, a single conserved exon 5 haplotype in *LDHA* is remarkably associated with the risk of panic disorder, which is a type of anxiety disorder [86]. The mRNA expression levels of *LDHA* gene increased in major depressive disorder patients in both depressive state and remissive state in comparison with healthy control subjects [87]. For the gene of *DALRD3*, it encodes a protein with a DALR anticodon binding domain similar to that of class Ia aminoacyl tRNA synthetases. The abnormal expression of *DALRD3* has been reported to be significantly associated with sleep-related phenotypes ( $P=0.033$ ) [88]. The *FGFR3* gene, encoding a member of the fibroblast growth factor receptor family, has been reported to be exclusively expressed in the locus coeruleus in patients with major depressive disorder [89]. Previous studies have demonstrated that skeletal dysplasia patients with Asn540Lys mutation in the *FGFR3* gene have been documented to accompany with medial temporal lobe dysgenesis and epilepsy [90–92].

In conclusion, the current comprehensive study provides multiple lines of evidence for supporting *DALRD3*, *LDHA*, *HEBP2*, *TEX264*, and *FGFR3* as insomnia-associated genes whose abnormal expression level may convey risk to insomnia. Our findings indicate that individuals suffering from insomnia may be more vulnerable to various neurodegenerative disorders. Furthermore, we link insomnia risk variants to susceptible genes and biological pathways, offering a possible explanation of biological mechanism between genetic variation and insomnia risk. Further molecular experiments are warranted to investigate the molecular functions of identified genes and risk variants.

### Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

### Funding

This work was supported by Zhejiang Medical and Health Science and Technology Plan [grant number 2017KY355]. The funders had no role in the designing and conducting of this study and collection, analysis, and interpretation of data and in writing the manuscript.

### Author Contribution

H.S. and J.Z. managed the literature search, data collection, and bioinformatics analysis. H.S. wrote the first draft of the manuscript. Y.M. and J.L. conceived the study and wrote the manuscript. All the authors read and approved the final manuscript.

### Ethics Approval

This study does not contain any studies with human subjects or animals performed by any of the authors.

### Consent for Publication

The authors declare that we gave consent for publication.

### Abbreviations

DGE, differential gene expression; eQTL, expression quantitative trait loci; LBF, logarithm of the Bayes factor; MAGMA, Multi-marker Analysis of GenoMic Annotation; PPI, protein–protein interaction.

### References

- 1 Chung, K.F. et al. (2015) Cross-cultural and comparative epidemiology of insomnia: the Diagnostic and statistical manual (DSM), International classification of diseases (ICD) and International classification of sleep disorders (ICSD). *Sleep Med.* **16**, 477–482, <https://doi.org/10.1016/j.sleep.2014.10.018>

- 2 Fernandez-Mendoza, J. and Vgontzas, A.N. (2013) Insomnia and its impact on physical and mental health. *Curr. Psychiatry Rep.* **15**, 418, <https://doi.org/10.1007/s11920-013-0418-8>
- 3 Luyster, F.S. et al. (2012) Sleep: a health imperative. *Sleep* **35**, 727–734, <https://doi.org/10.5665/sleep.1846>
- 4 Baglioni, C. et al. (2011) Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. *J. Affect. Disord.* **135**, 10–19, <https://doi.org/10.1016/j.jad.2011.01.011>
- 5 Yehuda, R. et al. (2015) Post-traumatic stress disorder. *Nat. Rev. Dis. Primers* **1**, 15057, <https://doi.org/10.1038/nrdp.2015.57>
- 6 Morin, C.M. and Jarrin, D.C. (2013) Insomnia and healthcare-seeking behaviors: impact of case definitions, comorbidity, sociodemographic, and cultural factors. *Sleep Med.* **14**, 808–809, <https://doi.org/10.1016/j.sleep.2013.05.003>
- 7 Roth, T. et al. (2011) Prevalence and perceived health associated with insomnia based on DSM-IV-TR; International Statistical Classification of Diseases and Related Health Problems, Tenth Revision; and Research Diagnostic Criteria/International Classification of Sleep Disorders, Second Edition criteria: results from the America Insomnia Survey. *Biol. Psychiatry* **69**, 592–600
- 8 Cunnington, D., Junge, M.F. and Fernando, A.T. (2013) Insomnia: prevalence, consequences and effective treatment. *Med. J. Aust.* **199**, S36–40, <https://doi.org/10.5694/mja13.10718>
- 9 Bastien, C.H. and Morin, C.M. (2000) Familial incidence of insomnia. *J. Sleep Res.* **9**, 49–54, <https://doi.org/10.1046/j.1365-2869.2000.00182.x>
- 10 Dauvilliers, Y. et al. (2005) Family studies in insomnia. *J. Psychosom. Res.* **58**, 271–278, <https://doi.org/10.1016/j.jpsychores.2004.08.012>
- 11 Wing, Y.K. et al. (2012) Familial aggregation and heritability of insomnia in a community-based study. *Sleep Med.* **13**, 985–990, <https://doi.org/10.1016/j.sleep.2012.04.013>
- 12 Lind, M.J. et al. (2015) A Longitudinal Twin Study of Insomnia Symptoms in Adults. *Sleep* **38**, 1423–1430, <https://doi.org/10.5665/sleep.4982>
- 13 Jansen, P.R. et al. (2019) Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat. Genet.* **51**, 394–403, <https://doi.org/10.1038/s41588-018-0333-3>
- 14 Lane, J.M. et al. (2017) Genome-wide association analyses of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits. *Nat. Genet.* **49**, 274–281, <https://doi.org/10.1038/ng.3749>
- 15 Hammerschlag, A.R. et al. (2017) Genome-wide association analysis of insomnia complaints identifies risk genes and genetic overlap with psychiatric and metabolic traits. *Nat. Genet.* **49**, 1584–1592, <https://doi.org/10.1038/ng.3888>
- 16 Stein, M.B. et al. (2018) Genome-wide analysis of insomnia disorder. *Mol. Psychiatry* **23**, 2238–2250, <https://doi.org/10.1038/s41380-018-0033-5>
- 17 Byrne, E.M. et al. (2013) A genome-wide association study of sleep habits and insomnia. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **162B**, 439–451, <https://doi.org/10.1002/ajmg.b.32168>
- 18 Choi, Y.J. et al. (2017) Transcriptional profiling of human femoral mesenchymal stem cells in osteoporosis and its association with adipogenesis. *Gene* **632**, 7–15, <https://doi.org/10.1016/j.gene.2017.08.015>
- 19 Wang, W. et al. (2017) Integrative analysis of GWAS, eQTLs and meQTLs data suggests that multiple gene sets are associated with bone mineral density. *Bone Joint Res.* **6**, 572–576, <https://doi.org/10.1302/2046-3758.610.BJR-2017-0113.R1>
- 20 Calabrese, G.M. et al. (2017) Integrating GWAS and Co-expression Network Data Identifies Bone Mineral Density Genes SPTBN1 and MARK3 and an Osteoblast Functional Module. *Cell Syst.* **4**, 46e4–59e4
- 21 Uyhelji, H.A. et al. (2018) Exploring gene expression biomarker candidates for neurobehavioral impairment from total sleep deprivation. *BMC Genomics* **19**, 341, <https://doi.org/10.1186/s12864-018-4664-3>
- 22 He, X. et al. (2013) Sherlock: detecting gene-disease associations by matching patterns of expression QTL and GWAS. *Am. J. Hum. Genet.* **92**, 667–680, <https://doi.org/10.1016/j.ajhg.2013.03.022>
- 23 Ayalew, M. et al. (2012) Convergent functional genomics of schizophrenia: from comprehensive understanding to genetic risk prediction. *Mol. Psychiatry* **17**, 887–905, <https://doi.org/10.1038/mp.2012.37>
- 24 Schadt, E.E. et al. (2005) An integrative genomics approach to infer causal associations between gene expression and disease. *Nat. Genet.* **37**, 710–717, <https://doi.org/10.1038/ng1589>
- 25 Zhu, Z. et al. (2016) Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487, <https://doi.org/10.1038/ng.3538>
- 26 Huo, Y.X. et al. (2016) Identification of SLC25A37 as a major depressive disorder risk gene. *J. Psychiatr. Res.* **83**, 168–175, <https://doi.org/10.1016/j.jpsychores.2016.09.011>
- 27 Zhong, J. et al. (2019) Integration of GWAS and brain eQTL identifies FLOT1 as a risk gene for major depressive disorder. *Neuropsychopharmacology* **44**, 1542–1551
- 28 Lee, M.G. et al. (2019) Integrative Genome-Wide Association Studies of eQTL and GWAS Data for Gout Disease Susceptibility. *Sci. Rep.* **9**, 4981, <https://doi.org/10.1038/s41598-019-41434-4>
- 29 Yang, C.P. et al. (2018) Comprehensive integrative analyses identify GLT8D1 and CSNK2B as schizophrenia risk genes. *Nat. Commun.* **9**, 838, <https://doi.org/10.1038/s41467-018-03247-3>
- 30 Visscher, P.M. (2008) Sizing up human height variation. *Nat. Genet.* **40**, 489–490, <https://doi.org/10.1038/ng0508-489>
- 31 Landi, M.T. et al. (2009) A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am. J. Human Genet.* **85**, 679–691, <https://doi.org/10.1016/j.ajhg.2009.09.012>
- 32 Ferreira, M.A.R. et al. (2019) Genetic Architectures of Childhood- and Adult-Onset Asthma Are Partly Distinct. *Am. J. Hum. Genet.* **104**, 665–684, <https://doi.org/10.1016/j.ajhg.2019.02.022>
- 33 Myers, A.J. et al. (2007) A survey of genetic human cortical gene expression. *Nat. Genet.* **39**, 1494–1499, <https://doi.org/10.1038/ng.2007.16>
- 34 Hindorff, L.A. et al. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9362–9367, <https://doi.org/10.1073/pnas.0903103106>

- 35 Shannon, P. et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504, <https://doi.org/10.1101/gr.1239303>
- 36 Bindea, G. et al. (2009) ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093, <https://doi.org/10.1093/bioinformatics/btp101>
- 37 Zhang, B., Kirov, S. and Snoddy, J. (2005) WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res.* **33**, W741–W748, <https://doi.org/10.1093/nar/gki475>
- 38 Jourquin, J. et al. (2012) GLAD4U: deriving and prioritizing gene lists from PubMed literature. *BMC Genomics* **13**, S20, <https://doi.org/10.1186/1471-2164-13-S8-S20>
- 39 Wishart, D.S. et al. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **36**, D901–D906, <https://doi.org/10.1093/nar/gkm958>
- 40 Consortium, G.T. (2013) The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585
- 41 Shabalin, A.A. (2012) Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353–1358, <https://doi.org/10.1093/bioinformatics/bts163>
- 42 Network and C (2015) Pathway Analysis Subgroup of Psychiatric Genomics, Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat. Neurosci.* **18**, 199–209, <https://doi.org/10.1038/nn.3922>
- 43 Ma, Y. and Li, M.D. (2017) Establishment of a Strong Link Between Smoking and Cancer Pathogenesis through DNA Methylation Analysis. *Sci. Rep.* **7**, 1811, <https://doi.org/10.1038/s41598-017-01856-4>
- 44 Gulsuner, S. et al. (2013) Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154**, 518–529, <https://doi.org/10.1016/j.cell.2013.06.049>
- 45 Zhao, X. et al. (2019) Identification of LIFR, PIK3R1, and MMP12 as Novel Prognostic Signatures in Gallbladder Cancer Using Network-Based Module Analysis. *Front. Oncol.* **9**, 325, <https://doi.org/10.3389/fonc.2019.00325>
- 46 Taylor, I.W. et al. (2009) Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat. Biotechnol.* **27**, 199–204, <https://doi.org/10.1038/nbt.1522>
- 47 Warde-Farley, D. et al. (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* **38**, W214–W220, <https://doi.org/10.1093/nar/gkq537>
- 48 Shi, X.H. et al. (2010) Clinical, histopathological and genetic studies in a family with fatal familial insomnia. *Infect. Genet. Evol.* **10**, 292–297, <https://doi.org/10.1016/j.meegid.2010.01.007>
- 49 Xie, W.L. et al. (2013) Comparison of the pathological and pathogenic features in six different regions of postmortem brains of three patients with fatal familial insomnia. *Int. J. Mol. Med.* **31**, 81–90, <https://doi.org/10.3892/ijmm.2012.1194>
- 50 Davis, S. and Meltzer, P.S. (2007) GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846–1847, <https://doi.org/10.1093/bioinformatics/btm254>
- 51 Mackiewicz, M. et al. (2007) Macromolecule biosynthesis: a key function of sleep. *Physiol. Genomics* **31**, 441–457, <https://doi.org/10.1152/physiolgenomics.00275.2006>
- 52 Spada, J. et al. (2016) Genome-wide association analysis of actigraphic sleep phenotypes in the LIFE Adult Study. *J. Sleep Res.* **25**, 690–701, <https://doi.org/10.1111/jsr.12421>
- 53 Ross, C.A. and Tabrizi, S.J. (2011) Huntington's disease: from molecular pathogenesis to clinical treatment. *Lancet Neurol.* **10**, 83–98, [https://doi.org/10.1016/S1474-4422\(10\)70245-3](https://doi.org/10.1016/S1474-4422(10)70245-3)
- 54 Abeliovich, A. and Gitler, A.D. (2016) Defects in trafficking bridge Parkinson's disease pathology and genetics. *Nature* **539**, 207–216, <https://doi.org/10.1038/nature20414>
- 55 McNaught, K.S. and Olanow, C.W. (2003) Proteolytic stress: a unifying concept for the etiopathogenesis of Parkinson's disease. *Ann. Neurol.* **53**, S73–S84, discussion S84–6, <https://doi.org/10.1002/ana.10512>
- 56 Rios, J.A. et al. (2014) Is Alzheimer's disease related to metabolic syndrome? A Wnt signaling conundrum. *Prog. Neurobiol.* **121**, 125–146, <https://doi.org/10.1016/j.pneurobio.2014.07.004>
- 57 Bergman, O. and Ben-Shachar, D. (2016) Mitochondrial Oxidative Phosphorylation System (OXPHOS) Deficits in Schizophrenia: Possible Interactions with Cellular Processes. *Can. J. Psychiatry* **61**, 457–469, <https://doi.org/10.1177/0706743716648290>
- 58 Morris, G. et al. (2017) A model of the mitochondrial basis of bipolar disorder. *Neurosci. Biobehav. Rev.* **74**, 1–20, <https://doi.org/10.1016/j.neubiorev.2017.01.014>
- 59 Kobayashi, M. et al. (2017) Diabetic polyneuropathy, sensory neurons, nuclear structure and spliceosome alterations: a role for CWC22. *Dis. Model. Mech.* **10**, 215–224, <https://doi.org/10.1242/dmm.028225>
- 60 Lane, J.M. et al. (2017) Genome-wide association analyses of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits. *Nat. Genet.* **49**, 274–281, <https://doi.org/10.1038/ng.3749>
- 61 Wittchen, H.U. et al. (2011) The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur. Neuropsychopharmacol.* **21**, 655–679, <https://doi.org/10.1016/j.euroneuro.2011.07.018>
- 62 Freedman, M.L. et al. (2011) Principles for the post-GWAS functional characterization of cancer risk loci. *Nat. Genet.* **43**, 513–518, <https://doi.org/10.1038/ng.840>
- 63 Blattler, A. et al. (2014) Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. *Genome Biol.* **15**, 469, <https://doi.org/10.1186/s13059-014-0469-0>
- 64 Tak, Y.G. and Farnham, P.J. (2015) Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin* **8**, 57, <https://doi.org/10.1186/s13072-015-0050-4>
- 65 Wu, Y., Yao, Y.G. and Luo, X.J. (2017) SZDB: A Database for Schizophrenia Genetic Research. *Schizophr. Bull.* **43**, 459–471

- 66 Byrne, E.M. et al. (2013) A genome-wide association study of sleep habits and insomnia. *Am. J. Med. Genet. Part B, Neuropsychiatric Genet.: Off. Publ. Int. Soc. Psychiatric Genet.* **162B**, 439–451, <https://doi.org/10.1002/ajmg.b.32168>
- 67 Hoseth, E.Z. et al. (2018) Exploring the Wnt signaling pathway in schizophrenia and bipolar disorder. *Transl Psychiatry* **8**, 55, <https://doi.org/10.1038/s41398-018-0102-1>
- 68 Tian, C. et al. (2014) Analyses of the similarity and difference of global gene expression profiles in cortex regions of three neurodegenerative diseases: sporadic Creutzfeldt-Jakob disease (sCJD), fatal familial insomnia (FFI), and Alzheimer's disease (AD). *Mol. Neurobiol.* **50**, 473–481, <https://doi.org/10.1007/s12035-014-8758-x>
- 69 Tian, C. et al. (2013) Comparative analysis of gene expression profiles between cortex and thalamus in Chinese fatal familial insomnia patients. *Mol. Neurobiol.* **48**, 36–48, <https://doi.org/10.1007/s12035-013-8426-6>
- 70 van Bruggen-Rufi, M. et al. (2016) Music therapy in Huntington's disease: a protocol for a multi-center randomized controlled trial. *BMC Psychol.* **4**, 38–38, <https://doi.org/10.1186/s40359-016-0146-z>
- 71 Gonatopoulos-Pournatzis, T. et al. (2018) Genome-wide CRISPR-Cas9 Interrogation of Splicing Networks Reveals a Mechanism for Recognition of Autism-Misregulated Neuronal Microexons. *Mol. Cell* **72**, 510e12–524e12, <https://doi.org/10.1016/j.molcel.2018.10.008>
- 72 Packer, A. (2018) Enrichment of factors regulating canonical Wnt signaling among autism risk genes. *Mol. Psychiatry* **23**, 492–493, <https://doi.org/10.1038/mp.2016.228>
- 73 Durak, O. et al. (2016) Chd8 mediates cortical neurogenesis via transcriptional regulation of cell cycle and Wnt signaling. *Nat. Neurosci.* **19**, 1477–1488, <https://doi.org/10.1038/nn.4400>
- 74 Bai, B. et al. (2013) U1 small nuclear ribonucleoprotein complex and RNA splicing alterations in Alzheimer's disease. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 16562–16567, <https://doi.org/10.1073/pnas.1310249110>
- 75 Sorrentino, V. et al. (2017) Enhancing mitochondrial proteostasis reduces amyloid-beta proteotoxicity. *Nature* **552**, 187–193, <https://doi.org/10.1038/nature25143>
- 76 Liu, Y. et al. (2017) Wnt/beta-catenin signaling plays an essential role in alpha7 nicotinic receptor-mediated neuroprotection of dopaminergic neurons in a mouse Parkinson's disease model. *Biochem. Pharmacol.* **140**, 115–123, <https://doi.org/10.1016/j.bcp.2017.05.017>
- 77 Salasova, A. et al. (2017) A proteomic analysis of LRRK2 binding partners reveals interactions with multiple signaling components of the WNT/PCP pathway. *Mol Neurodegener* **12**, 54, <https://doi.org/10.1186/s13024-017-0193-9>
- 78 Berger, B.S. et al. (2017) Parkinson's disease-associated receptor GPR37 is an ER chaperone for LRP6. *EMBO Rep.* **18**, 712–725, <https://doi.org/10.15252/embr.201643585>
- 79 Lim, R.G. et al. (2017) Huntington's Disease iPSC-Derived Brain Microvascular Endothelial Cells Reveal WNT-Mediated Angiogenic and Blood-Brain Barrier Deficits. *Cell Rep.* **19**, 1365–1377, <https://doi.org/10.1016/j.celrep.2017.04.021>
- 80 Johri, A., Chandra, A. and Flint, M. (2013) Beal, PGC-1alpha, mitochondrial dysfunction, and Huntington's disease. *Free Radic. Biol. Med.* **62**, 37–46, <https://doi.org/10.1016/j.freeradbiomed.2013.04.016>
- 81 Mitchell, A.L. et al. (2019) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **47**, D351–D360, <https://doi.org/10.1093/nar/gky1100>
- 82 Bahr, T.M. et al. (2013) Peripheral blood mononuclear cell gene expression in chronic obstructive pulmonary disease. *Am. J. Respiratory Cell Mol. Biol.* **49**, 316–323, <https://doi.org/10.1165/rcmb.2012-02300C>
- 83 Roth, R.B. et al. (2006) Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. *Neurogenetics* **7**, 67–80, <https://doi.org/10.1007/s10048-006-0032-6>
- 84 Newington, J.T. et al. (2012) Overexpression of pyruvate dehydrogenase kinase 1 and lactate dehydrogenase A in nerve cells confers resistance to amyloid beta and other toxins by decreasing mitochondrial respiration and reactive oxygen species production. *J. Biol. Chem.* **287**, 37245–37258, <https://doi.org/10.1074/jbc.M112.366195>
- 85 Krzysztan-Russjan, J. et al. (2013) A study of molecular changes relating to energy metabolism and cellular stress in people with Huntington's disease: looking for biomarkers. *J. Bioenerg. Biomembr.* **45**, 71–85, <https://doi.org/10.1007/s10863-012-9479-3>
- 86 Philibert, R.A. et al. (2003) Association of an exonic LDHA polymorphism with altered respiratory response in probands at high risk for panic disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **117B**, 11–17, <https://doi.org/10.1002/ajmg.b.10015>
- 87 Shibata, T. et al. (2013) The alteration of hypoxia inducible factor-1 (HIF-1) and its target genes in mood disorder patients. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **43**, 222–229, <https://doi.org/10.1016/j.pnpbp.2013.01.003>
- 88 Baier, P.C. et al. (2014) Mice lacking the circadian modulators SHARP1 and SHARP2 display altered sleep and mixed state endophenotypes of psychiatric disorders. *PLoS One* **9**, e110310, <https://doi.org/10.1371/journal.pone.0110310>
- 89 Bernard, R. et al. (2011) Altered expression of glutamate signaling, growth factor, and glia genes in the locus coeruleus of patients with major depression. *Mol. Psychiatry* **16**, 634–646, <https://doi.org/10.1038/mp.2010.44>
- 90 Okazaki, T. et al. (2017) Epileptic phenotype of FGFR3-related bilateral medial temporal lobe dysgenesis. *Brain Dev.* **39**, 67–71, <https://doi.org/10.1016/j.braindev.2016.07.004>
- 91 Garcia-Vargas, A. et al. (2008) An epidermal nevus syndrome with cerebral involvement caused by a mosaic FGFR3 mutation. *Am. J. Med. Genet. A* **146A**, 2275–2279, <https://doi.org/10.1002/ajmg.a.32429>
- 92 Manikkam, S.A. et al. (2018) Temporal Lobe Malformations in Achondroplasia: Expanding the Brain Imaging Phenotype Associated with FGFR3-Related Skeletal Dysplasias. *AJNR Am. J. Neuroradiol.* **39**, 380–384, <https://doi.org/10.3174/ajnr.A5468>