Progress in Neuropsychopharmacology & Biological Psychiatry 86 (2018) 180–185

Contents lists available at ScienceDirect

Progress in Neuropsychopharmacology & Biological Psychiatry

journal homepage: www.elsevier.com/locate/pnp

Integrative analysis of genome-wide association study and brain region related enhancer maps identifies biological pathways for insomnia

Miao Ding[1], Ping Li[1], Yan Wen, Yan Zhao, Bolun Cheng, Lu Zhang, Mei Ma, Shiqiang Cheng, Li Liu, Yanan Du, Xiao Liang, Awen He, Xiong Guo, Feng Zhang*

Key Laboratory of Trace Elements and Endemic Diseases of National Health and Family Planning Commission, School of Public Health, Xi'an Jiaotong University Health Science Center, Xi'an, PR China

ARTICLE INFO

Keywords:
Insomnia
Enhancer
Genome-wide association study
Enrichment analysis

ABSTRACT

Insomnia is a common sleep disorder whose genetic mechanism remains unknown. The aim of this study is to identify novel genes, gene enrichment sets and enriched tissue/cell types for insomnia considering the differences across different brain regions. We conducted an integrative analysis of genome-wide association study (GWAS) and brain region related enhancer maps. Summary data was derived from a large-scale GWAS of insomnia, involving 113,006 unrelated individuals. The chromosomal enhancer maps of 6 brain regions were then aligned with the GWAS summary data to obtain the association testing results of enhancer regions for insomnia. Gene prioritization, tissue/cell and pathway enrichment analysis were implemented by Data-driven Expression Prioritized Integration for Complex Traits (DEPICT) tool. We identified multiple cross-brain regions or brain-region specific prioritized genes for insomnia, such as MADD (P = .0013 in angular gyrus), PPP2R3C (P = .0319 in cingulate gyrus), CASP9 (P = .0066 in angular gyrus and P = .0278 in hippocampus middle), PLEKHM2 (P = .0032 in angular gyrus, P = .0052 in anterior caudate, P = .0385 in cingulate gyrus and P = .0011 in inferior temporal lobe). This study also detected a group of insomnia associated biological pathways within multiple or specific brain regions, such as REACTOME_SIGNALING_BY_NOTCH and KEGG_GLYCEROPHOSPHOLIPID_METABOLISM. Our results showed that insomnia associated genes were significantly enriched in neural stem cells. Our results highlight a set of potential points, particularly neural stem cells, for subsequent biological studies for insomnia.

1. Introduction

Insomnia is the most prevalent sleep disorder, characterized by having difficulties in initiating sleep or keeping sleep, or waking up too early and being unable to restart sleep (Alsaadi et al., 2013). The reported prevalence of chronic insomnia in general populations varied approximately from 10% to 15% (Budhiraja et al., 2011). With the increasing pressure of modern life and work, the prevalence of insomnia tends to continuously rise. Insomnia may lead to daytime functional impairment, such as fatigue, mood decline, decreased concentration, or worse psychomotor performance (Alsaadi et al., 2013; Budhiraja et al., 2011). Moreover, it has been reported to be related to increased risk for various human disorders, such as major psychiatric illnesses(Baglioni et al., 2011), type 2 diabetes(Cappuccio et al., 2010), and cardiovascular diseases(Mallon et al., 2002). Furthermore, chronic insomnia can cause the decreased productivity, the increased traffic accidents, the increased cost of health care, and the worse health-related quality of life, resulting in heavy personal and social burden (Kessler et al., 2011; Léger et al., 2001; Ozminkowski et al., 2007).

Previous studies have demonstrated that genetic factors contributed greatly to the development of insomnia (Wing et al., 2012). The estimated heritability of insomnia is about 25–45% for the whole population (Wing et al., 2012), 59% for the females and 38% for the males (Lind et al., 2015). Multiple susceptibility genes have been reported for insomnia. For instance, genome-wide association study (GWAS) identified multiple loci (near MEIS1, TMEM132E, CYCL1 and TGFBI in females and WDR27 in males) associated with insomnia complaint (Lane et al., 2017). A recent genome-wide gene-based association study reported that HHEX, RHCG, IPO7 and TSNARE1 genes involved in the genetic basis of insomnia symptoms (Hammerschlag et al., 2017). Although multiple candidate genes have been identified for insomnia, the genetic basis of insomnia remains elusive now. Further studies are needed to reveal the complicated genetic mechanism of insomnia.

* Corresponding author at: Key Laboratory of Trace Elements and Endemic Diseases of National Health and Family Planning Commission, School of Public Health, Xi'an Jiaotong University Health Science Center, No. 76 Yan Ta West Road, Xi'an 710061, PR China.
E-mail address: fzhxjtu@mail.xjtu.edu.cn (F. Zhang).
[1] The two authors contributed equally to this work.

https://doi.org/10.1016/j.pnpbp.2018.05.026
Received 6 February 2018; Received in revised form 30 May 2018; Accepted 30 May 2018
Available online 06 June 2018
0278-5846/

Given that the growth and development of human brain comes down to the complicated interactions among different brain regions, psychiatric disorders usually involve the dysfunction of various brain regions with different molecular pathogenesis(Bagley et al., 2017). For example, Lo et al. suggested different roles of various brain regions for Alzheimer's disease (Lo et al., 2010). Previous studies observed that diverse brain regions were involved in the biological process of sleep. For instance, it was found that hippocampus and neocortex contributed to the regulation of sleep-waking states, sleep structure, state transition timing, and proportion of time in sleep states (Emrick et al., 2016). Functional neuroimaging of sleep found that the decreased activity in basal ganglia, prefrontal cortex and anterior cingulate cortex might contribute to regulating slow-wave sleep (SWS), and enhanced activity of anterior cingulate cortex involved in the regulation of REM sleep (Maquet, 2000). Therefore, it is reasonable to infer that multiple brain regions contribute to the development of insomnia. However, the biological mechanism of different brain regions implicated in the development of insomnia are largely unknown now. Few genetic studies have been conducted to explore the different roles of various brain regions in the development of insomnia. Genetic studies considering potentially different roles of multiple brain regions may provide insights into the biological mechanism of insomnia.

It is well-known that GWAS was an effective tool for scanning the genetic variations affecting complicated traits. But, the genetic variants with modest phenotypic effect may be missed by GWAS, arising from stringent genome-wide statistical significance threshold (Serretti et al., 2003). Pathway-based association studies were proposed and succeeded in the genetic mechanism studies of complex diseases. It is capable of capturing the genetic information of GWAS and known functional information of gene sets (Wang et al., 2007). Identifying insomnia-related biological pathways may enhance our comprehension of the genetic basis of insomnia.

Recent studies observed that the significant single nucleotide polymorphisms (SNPs) identified by GWAS were enriched in chromosomal non-coding regulatory regions (Nicolae et al., 2010). Integrating GWAS and regulatory loci information can provide novel clues for pathogenetic studies of complex traits (Zhu et al., 2016). Enhancers are genomic regulatory regions that can recruit and be bound by transcription factors, and modulate the gene transcription by interacting with the promoters of target genes (Ren et al., 2017). In this study, we conducted an integrative analysis of GWAS and 6 brain region related enhancer maps to identify insomnia associated genes and pathways considering the differences across different brain regions.

2. Materials and methods

2.1. GWAS summary dataset of insomnia

The summary data of a recent large-scale genome-wide association analysis for insomnia was applied here (Hammerschlag et al., 2017). This GWAS dataset included 113,006 individuals from UK Biobank and 7565 participants from deCODE. For UK samples, Genotyping was performed on the UK BiLEVE custom array and the UK Biobank Axiom array. Genotype imputation was carried out using the UK10K haplotype panel and the 1000 Genomes Project Phase 3 panel as a reference panel. Genome-wide association analysis was performed by SNPTEST using logistic regression model. Finally, 12,444,916 SNPs with imputation quality > 0.8 and minor allele frequency (MAF) > 0.001 were retained. For deCODE sample, the GWAS data were generated on the basis of whole-genome sequencing, genotype imputation and long-range haplotype phasing of Icelandic samples. The association test was performed by logistic regression using the tool proposed by deCODE Genetics. The meta-analysis of SNPs in UK Biobank and deCODE samples was conducted using METAL. The detailed description of cohorts, genotyping, imputation, quality control and analytic procedure was available in the previously published studies (Hammerschlag et al., 2017).

2.2. Brain enhancer-gene dataset

Enhancers are regulatory DNA elements, which are bound by transcription factors and cofactors, and activate gene transcription by interacting with the promoters of target genes (Ren et al., 2017). Recent studies have demonstrated the key roles of chromosomal enhancer regions in the development of complex diseases (Das et al., 2017). It was also found that enhancers played different roles in gene regulation across different brain regions (Cheng et al., 2015; Leung et al., 2015). For instance, Cheng et al. observed different effects of enhancers on the gene expression of different brain tissues in mice (Cheng et al., 2015). In this study, the published enhancer-gene networks of various brain regions were utilized in our analysis (Cao et al., 2017). In short, DNase–seq data, RNA-seq data, and ChIP–seq data of H3K27ac, H3K27me3, and H3K4me1 for 127 samples of human cell/tissue types and cell lines were obtained from the ENCODE and Roadmap Epigenomics. Moreover, ChromHMM-predicted active enhancers from states 6, 7 and 12 in 15-state model for 127 samples were also obtained. The residual overlapped enhancers were merged after excluding the enhancers larger than 2500 bp. A total of 489,581 enhancers were retained finally. Enhancer targets were identified by joint effect of multiple enhancers (JEME), which mainly contained investigating candidate modulating enhancers of each transcription start site (TSS) and then identifying the enhancers modulating every TSS in a given sample. Specific for this study, we used 15,624, 18,569, 18,768, 21,629, 18,455 and 22,596 enhancer-gene pairs for brain angular gyrus, anterior caudate, cingulate gyrus, hippocampus middle, inferior temporal lobe, and substantia nigra, respectively. Detailed information about sample characteristics, experiment design, data analysis, quality control and reconstruction approaches of enhancer-target could be found in the published study (Cao et al., 2017).

2.3. Statistical analysis

In this study, the GWAS summary data of insomnia was first combined with the enhancer–gene networks of 8 brain regions, through aligning the SNP positions of GWAS summary data with the chromosomal ranges of enhancers. If a SNP was located between the start and end positions of an enhancer, then this SNP was mapped to this enhancer. Cao et al. originally constructed the enhancer–gene networks for 8 brain regions, including angular gyrus, anterior caudate, cingulate gyrus, hippocampus middle, inferior temporal lobe, substantia nigra, germinal matrix and dorsolateral prefrontal cortex (Cao et al., 2017). The brain regions were divided by the types of biological material and anatomical location (Kundaje et al., 2015). The profiled tissues and cell types in the 8 brain regions were representative of major lineages in brain. After integrating the GWAS data with the enhancer-gene map data, only 6 brain regions met the analysis requirement of DEPICT (Pers et al., 2015). 51383, 62231, 62350, 70754, 59946 and 69042 enhancer SNP-gene pairs were generated for angular gyrus, anterior caudate, cingulate gyrus, hippocampus middle, inferior temporal lobe, and substantia nigra, respectively. For each of the 6 brain regions, the enhancer SNPs with GWAS *P value* $< .001$ were further input into DEPICT for gene prioritization, and tissue/cell and pathway enrichment analysis. DEPICT is an integrated tool that prioritizes the potential casual genes at given loci, and identifies enriched pathways, tissues and cell types (Pers et al., 2015; Sapkota et al., 2017). Briefly, enriched tissue/cell types are identified by testing whether the genes in given loci are highly expressed in 209 Medical Subject Heading (MeSH) tissue/cell types using 37,427 human Affymetrix HGU133a2.0 microarrays data. Gene set enrichment analysis is performed by assessing whether genes in associated loci are enriched for 14,461 reconstituted gene sets, including phenotypic gene sets obtained from phenotype–-gene pairs from the Mouse Genetics Initiative, Gene ontology(GO) terms, Kyoto Encyclopedia of genes and genomes (KEGG) and Reactome canonical pathways, and molecular pathways from high-