**Integrative analysis identified genes pleiotropically**

**associated with the risk of insomnia**

# Introduction

Insomnia is the second most prevalent mental disorder featured as difficulties in falling asleep, maintaining sleep, and early morning awakening, and ineffective sleep.(Jansen, et al. 2019; Riemann, et al. 2015) Insomnia is caused by combinations of genetic and environmental factors, and the genetic factors could explain a substantial proportion of the variance for insomnia. Its meta-analytic estimate of heritability was estimated at 40%. (Barclay, et al. 2021; Kocevska, et al. 2021; Madrid-Valero, et al. 2021) Genome-wide association studies (GWAS) have identified 57 loci for self-reported insomnia symptoms in the UK Biobank (n = 453,379) and confirmed their effects on self-reported insomnia symptoms in the HUNT Study (n = 14,923 cases and 47,610 controls), physician-diagnosed insomnia in the Partners Biobank (n = 2,217 cases and 14,240 controls). This may provide evidence of shared genetic factors was found between frequent insomnia symptoms and restless legs syndrome, aging, and cardiometabolic, behavioral, psychiatric, and reproductive traits. (Lane, et al. 2019) But the genetic underpinnings and pathophysiology of insomnia is still the unclear, and most of GWAS loc are in noncoding regions, making functional interpretation difficult.

These risk SNPs identified by GWAS may contribute to the risk of insomnia through modulating the expression level of nearby genes in different tissues rather than disturb the structure of proteins. DNA methylation is one of the most studied epigenetic modifications. And the best-known function of DNA methylation is to regulate nearby gene expression. We can integrate gene expression level data (e.g., expression quantitative trait loci, eQTL or mQTL) into GWAS data of disease to identify functional variants from GWAS results(Nicolae, et al. 2010). Summary data-based Mendelian Randomization(SMR) and heterogeneity in dependent instruments (HEIDI) is a mendelian randomization (MR) method that uses summary-level data to examine whether the expression level of a gene and a complex phenotype caused by pleiotropy and discern pleiotropy from linkage. Through the SMR analysis, several novel candidate genes underlying GWAS hits of complex diseases or traits were prioritized for follow-up functional studies. Strikingly, through integrating different omics data, we could gain further insights into the underlying genetic mechanisms of GWAS hits and disease.

To investigate insomnia risk genes roles in pathogenesis, we first combined the insomnia GWAS data and eQTL using SMR test. Next, integrated the mQTL data into insomnia GWAS data. Then, we made a comprehensive analysis of the results.

# MATERIALS AND METHODS

## GWAS data

We obtained complete summary-level of AD GWAS from UK Biobank GWAS result**s (**<http://fastgwa.info/ukbwes/phenotypes>). This insomnia GWAS summary statistics consisted of 455744 samples of European ancestry. The individuals with one or two parents diagnosed with insomnia in UKB were defined as proxy cases, and patients with two parents were upweighted. Meanwhile, participants with two parents without AD were defined as proxy controls, and older cognitively normal

parents were also upweighted.

## eQTL summary data

The blood eQTL data was obtained from the Westra eQTL summary data, which consisted of non-transformed peripheral blood samples from 5,311 individuals with replication in 2,775 individuals.(Westra, et al. 2013) We also used CAGE eQTL summary data which involved mRNA levels for 36,778 transcript expression traits (probes) from 2,765 individuals.(Lloyd-Jones, et al. 2017)

## mQTL summary data

In the SMR analysis, we integrated the insomnia GWAS data with blood mQTL data and brain mQTL data, respectively. For blood mQTL data, we obtained mQTL summary data from McRae et al. mQTL summary data which replicated 52,916 cis and 2,025 trans DNA methylation quantitative trait loci (mQTL) using methylation from whole blood measured on Illumina HumanMethylation450 arrays in the Brisbane Systems Genetics Study (n = 614 from 177 families) and the Lothian Birth Cohorts of 1921 and 1936 (combined n = 1366).(McRae, et al. 2018) For brain mQTL data, we used Brain-mMeta mQTL data which consisted of sample sizes (n = 1980 to 14,115).(Qi, et al. 2018)

## SMR analysis

To prioritize candidate causal genes of insomnia, we integrated GWAS and eQTL data, and GWAS and mQTL data through SMR method respectively, which examine the putative pleiotropic relationships between insomnia and eQTL. The SMR method mainly comprises of two steps. First, genetic variations are used as instrumental variables to examine for causative effect of gene expression on insomnia. Second, we applied the heterogeneity in dependent instruments (HEIDI) test implemented in SMR software to distinguish the causality and pleiotropy model from the linkage model. If the HEIDI test is significant (PHEIDI< 0.05), the identified genes by SMR can be a result of linkage. To account for multiple testing, we adjusted PSMR values using the Bonferroni approach. Those probes with PSMR <0.05/n and little evidence of heterogeneity (PHEIDI > 0.05) were used to highlight genome-wide significant genes. The SMR software was downloaded from https://cnsgenomics.com/software/smr.

# RESULTS

## eQTL data for SMR analysis

After we performed the HEIDI analysis for the identified genes to reduce the effect of potential linkage, we found 4 genes identified in the SMR analysis.

## mQTL data for SMR analysis

We integrated insomnia GWAS with mQTL data from the blood and brain, respectively. We found the 22 genes and 19 genes respectively identified in the SMR analysis after the HEIDI test.

# Discussion

The current study represented an effort to identify associated genes by integrating GWAS data with QTL data such as eQTL and mQTL data using the SMR method and further comprehensive analysis to find their correlation. And we found 4 genes in SMR analysis with eQTL and 23 genes with mQTL.

GWAS confirmed a number of susceptibility sites. However, it is difficult to explain why. Two-sample MR methods such as SMR used QTL as instrumental variable to detect whether gene expression level had a causal effect. MR analysis has several advantages over traditional observational epidemiological studies, including the ability to control for environmental confounders and to assess the impact of exposures on outcomes without having to measure exposures and outcomes in the same group of individuals. This is a feasible method to integrate multiple omics data, and this integration method can help to discover the causative factors behind GWAS recognition sites.

In conclusion, by integrating genomic, transcriptomic, and epigenomic data, we prioritized some functional gene which may cause insomnia, and found some genetic variations which have significant associations.

# References

Barclay, N. L., et al.

2021 The heritability of insomnia: A meta-analysis of twin studies. Genes Brain Behav 20(4):e12717.

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(3):394-403.

Kocevska, D., et al.

2021 Heritability of sleep duration and quality: A systematic review and meta-analysis. Sleep Med Rev 59:101448.

Lane, J. M., et al.

2019 Biological and clinical insights from genetics of insomnia symptoms. Nat Genet 51(3):387-393.

Lloyd-Jones, L. R., et al.

2017 The Genetic Architecture of Gene Expression in Peripheral Blood. Am J Hum Genet 100(2):228-237.

Madrid-Valero, J. J., et al.

2021 The heritability of insomnia: Systematic review and meta-analysis of twin studies. Sleep Med Rev 58:101437.

McRae, A. F., et al.

2018 Identification of 55,000 Replicated DNA Methylation QTL. Sci Rep 8(1):17605.

Nicolae, D. L., et al.

2010 Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet 6(4):e1000888.

Qi, T., et al.

2018 Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. Nat Commun 9(1):2282.

Riemann, D., et al.

2015 The neurobiology, investigation, and treatment of chronic insomnia. Lancet Neurol 14(5):547-58.

Westra, H. J., et al.

2013 Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet 45(10):1238-1243.