OXFORD

A multi-omics study of brain tissue transcription and DNA methylation revealing the genetic pathogenesis of ADHD

Jingkai Wang (ID), Qiu-Wen Zhu (ID), Jia-Hao Mai (ID), Shun Zhang (ID), Yuqing Wang (ID), Jiatong Liang (ID), Ji-Yuan Zhou (ID)*

Department of Biostatistics, State Key Laboratory of Organ Failure Research, Ministry of Education, and Guangdong Provincial Key Laboratory of Tropical Disease Research, School of Public Health, Southern Medical University, Guangzhou 510515, China

*Corresponding author. Department of Biostatistics, State Key Laboratory of Organ Failure Research, Ministry of Education, and Guangdong Provincial Key Laboratory of Tropical Disease Research, School of Public Health, Southern Medical University, Guangzhou 510515, China. E-mail: zhoujy@smu.edu.cn

Abstract

Attention-deficit/hyperactivity disorder (ADHD) is a chronic psychiatric disease that often affects a patient's whole life. Research has found that genetics plays an important role in the development of ADHD. However, there is still a lack of knowledge about the tissue-specific causal effects of biological processes beyond gene expression, such as alternative splicing (AS) and DNA methylation (DNAm), on ADHD. In this paper, a multi-omics study was conducted to investigate the causal effects of the transcription and the DNAm on ADHD, by integrating ADHD genome-wide association data with quantitative trait loci data of gene expression, AS, and DNAm across 14 different brain tissues. The causal effects were estimated using four different two-sample Mendelian randomization methods. Finally, we also prioritized the expression of 866 genes showing significant causal effects, including *COMMD5*, *ENSG00000271904*, *HYAL3*, etc., within at least one brain tissue. We prioritized 966 unique genes that have statistically significant causal AS events, within at least one of the 14 different brain tissues. These genes include *PPP1R16A*, *GGT7*, *TREM2*, etc. Furthermore, through mediation analysis, 106 regulatory pathways were inferred where DNAm influences ADHD through gene expression or AS processes. Our research findings provide guidance for future experimental studies on the molecular mechanisms of ADHD development, and also put forward valuable knowledge for the prevention, diagnosis, and treatment of ADHD.

Keywords: attention-deficit/hyperactivity disorder; multi-omics; Mendelian randomization; transcriptome-wide association study; alternative splicing; DNA methylation

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a chronic neurodevelopmental condition impacting approximately 5% of children and adolescents as well as 2.5% of adults globally [1]. It increases the risks of various challenges, including psychiatric disorders, educational and occupational difficulties, accidents, criminal involvement, social impairments, and addictive behaviors [1]. Recently, there has been a growing focus on preventing and treating ADHD. Studies have shown that pharmaceuticals like d-amphetamine, dl-threo-methylphenidate, and atomoxetine are effective in managing ADHD symptoms [2–6]. Also, it was reported that environmental influences and proper physical exercise can improve neural growth and development, potentially leading to sustained, long-term effects on the course of ADHD [7]. Therefore, ADHD is not uncontrollable. However, effective prevention and treatment of ADHD require a thorough understanding of its causative factors, enabling targeted interventions for individuals with ADHD.

Previously, research indicated that genetics plays an important role in ADHD. Specific genes associated with dopamine regulation and other neurotransmitters have been identified as the contributors to the development of ADHD [8]. To reveal potential genetic factors of ADHD, genome-wide association studies (GWAS)

have been employed to identify single nucleotide polymorphisms (SNPs), such as rs11420276, rs1222063, and rs1427829, associated with ADHD [9]. However, the limitation of GWAS lies in the incapacity to ascertain the pathways through which SNPs influence traits [10]. To solve this problem, one approach is to map SNPs to multi-omics profiles (exposures), such as gene transcription events, based on the quantitative trait loci (QTL). For example, transcriptome-wide association studies (TWAS) integrate the data of GWAS and expression QTL (eQTL) to search trait-associated genes and have found multiple genes that are associated with ADHD, such as *HYAL3*, *NUP43*, *PIDD*, and *PNPL2* [11]. Nevertheless, several limitations persist in existing research. First, tissue-specific genes causing ADHD have received limited investigation. Given the spatial specificity of transcription, the expression of the same gene can significantly vary across different tissues (Figure 1a). ADHD, which is a complex psychiatric disorder, is influenced by gene activities in multiple tissues, especially brain tissues [12]. Although the expression patterns across different brain regions exhibit high correlations [13], it remains unclear whether the causal genes of ADHD are the same across these tissues. Second, alternative splicing (AS) has not yet been explored for its causal effects on ADHD. AS is the process of splicing the precursor messenger RNA (pre-mRNA), which is an important
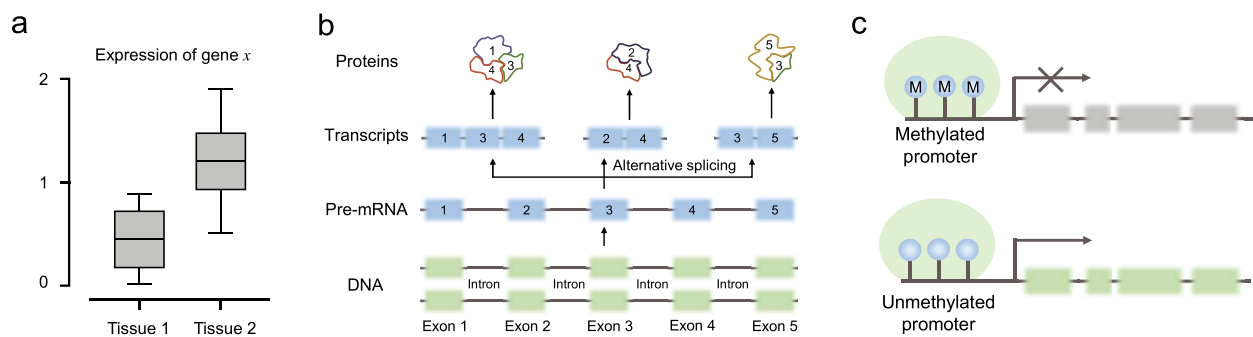
Figure 1. The biological hypotheses of this study. (a) The expression of gene *x* varies between tissues. (b) Pre-mRNA of a gene undergoes AS to form different transcripts, which are then translated into different protein isoforms, thereby influencing traits. (c) The principle diagram of how DNAm controls gene expression. DNAm and demethylation to some extent determine the transcriptional activity of promoters.

mechanism for increasing the diversity of transcripts (Figure 1b). The research found that AS is controlled by SNPs, leading to trait variation through complex mechanisms [14]. Following the structure of TWAS, with the help of splicing QTL (sQTL), GWAS have revealed associations between AS and multiple traits, such as schizophrenia [15] and Alzheimer's disease [16]. Third, it remains unclear whether the ADHD causal genes are regulated by specific mechanisms. Research indicated that DNA methylation (DNAm) can regulate transcription, subsequently influencing psychiatric disorders [17] (Figure 1c). Epigenome-wide association studies, which integrate methylation QTL (mQTL) and GWAS data, have mainly searched for ADHD-associated DNAm sites and linked the sites to genes relying on physical proximity or annotation databases [18, 19]. However, these studies lack a quantitative analysis of the mechanism by which transcription-mediated DNAm influences ADHD. In summary, understanding the genetic etiology of ADHD merely relying on genomics is inadequate. It is essential to integrate multiple tissue-specific omics data, such as transcriptomics, spliceomics, and methylomics, and analyze the causal effects between transcription, methylation, and ADHD to uncover the pathogenesis comprehensively.

To infer the causal associations between the multi-omics exposures and ADHD, we prioritize using the methods based on Mendelian randomization (MR). Following Mendel's Laws of Inheritance, alleles segregate during gamete formation, leading to naturally randomized distributions of SNPs in a population [20, 21]. MR methods leverage SNPs as instrumental variables (IVs) to substitute exposures in an association study, thereby avoiding confounding effects and enabling the interpretation of causal relationships between exposures and outcomes.

This study will integrate genomics, transcriptomics, spliceomics, and methylomics data to identify causal genes and causal AS events leading to ADHD. It will also investigate the genetic pathogenic mechanisms of ADHD and the regulatory mechanisms of transcription mediated by DNAm.

Methods
GWAS summary statistics dataset
The summary statistics for the ADHD GWAS were made available from the Psychiatric Genomics Consortium. A meta-analysis [22] incorporated extensive data from various sources, including iPSYCH, deCODE genetics, and aggregated information from 10 ADHD cohorts with European ancestry, which was curated by the Psychiatric Genomics Consortium. The combined dataset encompassed 38 691 individuals diagnosed with ADHD and 186 843 controlled individuals. The specific variable information for

all the data used for the analyses is detailed in Supplementary Note 1.

QTL summary statistics datasets
For multi-tissue TWAS analysis, we acquired the eQTL and sQTL summary statistics for 14 human brain tissues from GTEx project version 8 [23, 24] (European-American samples). The 14 human brain tissues include the amygdala ($n = 119$), anterior cingulate cortex ($n = 135$), caudate (basal ganglia, $n = 172$), cerebellar hemisphere ($n = 157$), cerebellum ($n = 188$), cortex ($n = 183$), frontal cortex ($n = 157$), hippocampus ($n = 150$), hypothalamus ($n = 156$), nucleus accumbens (basal ganglia, $n = 181$), pituitary ($n = 219$), putamen (basal ganglia, $n = 153$), spinal cord (cervical c-1, $n = 115$), and substantia nigra ($n = 100$). For mediation analysis, we obtained Qi *et al.*'s eQTL and sQTL meta-analysis summary statistics of brain cortex samples ($n = 2,865$) [14] as well as Qi *et al.*'s mQTL meta-analysis summary statistics ($n = 1,160$) [25] of brain samples from YangLab (https://yanglab.westlake.edu.cn/). In this paper, these two data were referred to as the 'BrainMeta' dataset. For quality control, we deleted the QTL with a *P*-value > 0.05 or minor allele frequency < 0.1.

Multi-tissue TWAS of ADHD
To produce the association analyses, first, for preparation, eQTL and sQTL data were collected from 14 different GTEx brain tissues to separately colocalize with the ADHD GWAS data based on the common SNPs. Then, four MR methods were employed to estimate and test the causal effect of each exposure on ADHD. In strict terms, both of the above two colocalization studies fall under TWAS. However, for the sake of clarity in this paper, we refer to the TWAS focusing on gene expression as eTWAS, and the TWAS focusing on AS as sTWAS.

Statistical methods for estimating and testing causal effects of exposures on ADHD
The causal structure of TWAS is depicted in Figure 2, where *Y* represents the outcome, which is the occurrence of ADHD in our study. *X* denotes the exposure, which corresponds to the phenotype in multi-omics QTL data. Z_1, Z_2,..., and Z_p are *p* IVs, i.e. SNPs. *U* refers to unknown confounding factors affecting both the outcome and the exposure. The $X \rightarrow Y$ represents the causal effect of the exposure on ADHD, which is the focus of our study. The effect sizes of the association between *Z* and *X* and the effect sizes of the direct association between *Z* and *Y* can be obtained from the QTL and GWAS summary data, respectively.

In this study, we used four two-sample MR methods to estimate and test the causal effect between the exposures and ADHD:

Table 1. Summary of four two-sample MR methods

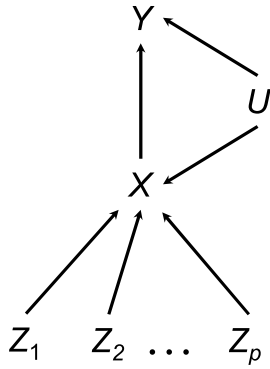| Method	Causal inference method	SNP correlation	Pleiotropy assumption
IVW	Meta-analysis	Independent	No
MR-Egger	Egger regression	Independent	Yes
LDA MR-Egger	Egger regression	Correlated	Yes
PMR-Egger	Likelihood-based method	Correlated	Yes

Figure 2. Directed acyclic graph of causal structure that TWAS assume. U refers to unknown confounding factors affecting both the outcome and the exposure. $X \rightarrow Y$ represents the causal effect of the exposure on the outcome.

inverse-variance weighted (IVW) method [26], MR-Egger [27], LDA MR-Egger [28], and PMR-Egger [29]. Table 1 provides a summary of these four methods together with the causal inference methods, the assumptions of SNP correlation (independent or correlated) and horizontal pleiotropy. Since the assumptions of these methods differ, synthesizing their analysis results will better accommodate diverse datasets and adjust the potential factors influencing the estimation of causal effects, such as horizontal pleiotropy and linkage disequilibrium (LD) between SNPs, thereby enhancing the reliability of the results. The details of all the methods are described in Supplementary Note 2.

In statistical testing, significant exposures are detected based on Bonferroni correction, which means that the P-value threshold is calculated as 0.05 divided by the number of exposures in certain omics. Furthermore, we performed an additional prioritization of the detected causal exposures based on criteria that are commonly used when implementing multiple methods. First, the corresponding causal effects are statistically significant for at least two methods [30, 31]. Second, the statistically significant causal effects should be in the same effect direction.

For the LD information between SNPs, we constructed a reference panel based on the 1000 Genome Project phase 3 datasets and employed a pruning method for SNP selection with PLINK 1.90 beta software [32, 33] (https://www.cog-genomics.org/plink/1.9/).

Bioinformatics analysis

To reveal the potential signaling pathways or processes, gene ontology (GO) enrichment analyses were conducted within each brain tissue with the combined gene list of causal genes identified by eTWAS and the genes mapped by the causal AS events detected by sTWAS. The enrichment analysis was conducted using the PANTHER Classification System (https://pantherdb.org/) based on the background databases including traditional GO and PANTHER GO-Slim v19.0 annotation datasets. GO is divided into three categories: biological process (BP), molecular function (MF), and

cellular component (CC). Additionally, the enrichment analysis also incorporated the PANTHER Protein Class (PC) database, aiming to determine if the products of the causal genes share common characteristics. The enrichment P-values were adjusted using false discovery rate (FDR).

To further understand the dynamic processes that the causal genes of ADHD involve, we analyzed the interactions between the products of the causal genes through protein-protein interaction (PPI) network analyses. The PPI networks were constructed in each brain tissue based on the STRING v12.0 database [34] (https://cn.string-db.org/). In addition to tissue-specific PPI networks, a cross-tissue PPI network was also established to reveal the interactions between different brain tissues. The interactions between the proteins are assessed using a combined score based on eight indicators: experimentally determined interaction, database annotated interaction, neighborhood on the chromosome, gene fusion, phylogenetic co-occurrence, homology, co-expression, and automated text mining. To ensure the reliability of the PPI analysis results, we removed interactions with combined scores below 700.

Mediation mechanism analysis

To explore whether the causal effects of transcription on traits originate from the regulation of DNAm, Wu *et al*. [35] proposed a hierarchical mediation mechanism framework involving methylomics, transcriptomics, and genomics data (Figure 3). In our study, we applied this framework to the ADHD GWAS data, mQTL, eQTL, and sQTL BrainMeta summary data. The mediation analysis can be summarized as the following processes: first, map the DNAm sites to their neighboring genes within a *cis*-region of ±2 Mb. Then, test the causal effects of the DNAm sites on the transcription of neighboring genes. Next, test the causal effects of the transcription on ADHD and the causal effects of DNAm on ADHD separately. Finally, if a DNAm site exhibits a significant causal effect on the transcription of a single gene, and both the DNAm site and the gene transcription show significant causal effects on ADHD, then it is inferred that this DNAm site influences the trait by regulating gene transcription.

An overview of study design

Our study design is illustrated in Figure 4. First, we acquired ADHD GWAS and GTEx multi-tissue QTL summary data and performed an integration by colocalization. We then estimated the LD reference panel using 1000 Genomes Project data. With these datasets, we applied four MR methods to conduct eTWAS and sTWAS, identifying causal genes and causal AS events for ADHD. Second, we performed GO enrichment analysis and PPI network analysis on the causal genes and the genes mapped by causal AS events. Third, we employed mediation analysis to identify pathways in which DNAm regulates transcription, ultimately leading to ADHD. Details on data preprocessing, model implementation, bioinformatics analysis, and mediation analysis are provided in Supplementary Notes 3–6.
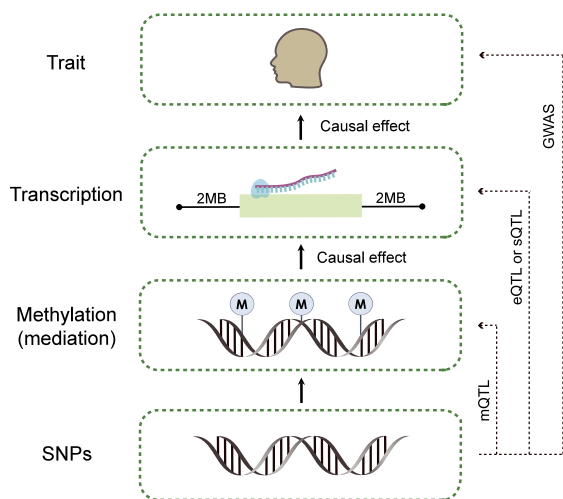
Figure 3. Framework of mediation analysis. In the mediation mechanism analysis framework, DNAm sites have causal effects on gene expression, subsequently impacting the onset of ADHD.

Results

Identified multi-tissue ADHD causal exposures

We conducted eTWAS and sTWAS, and then, compared the significant genes identified in both of the two studies across different brain tissue samples (Figure 5). Although there are overlapping significant genes between the two studies, this overlap comprises only a small fraction. Among all tissues, the pituitary exhibits the highest number of significant genes. Additionally, we provided Manhattan plots showcasing the results of eTWAS and sTWAS across the four MR methods. (Supplementary Note 7 and Supplementary Figs 1–8). All results generated by the programs are available for download at https://console.cloud.google.com/storage/browser/output20240516.

Identification results of eTWAS

We prioritized 866 genes that were identified as significant by at least two of the four MR methods within at least one brain tissue (Supplementary Table 1 and the corresponding variables used in it are listed in Supplementary Note 8). The expression of these genes may potentially contribute to ADHD. Supplementary Figure 9 illustrates the Venn diagram showing the number of causal genes and the overlapping genes identified by the four methods. Notably, we identified one causal gene in the amygdala and one in the pituitary (Fig. 6a and 6b), each found by the four methods: ENSG00000271904 in the amygdala (Figure 6c) and COMMD5 in the pituitary (Figure 6d). We generated a heatmap to display the numbers of overlapping causal genes between each pair of the 14 brain tissues (Figure 6e). The heatmap shows that the cortex and the cerebellum have the highest number of overlapping causal genes, with a total of 26, indicating that these two tissues may share common biological mechanisms or pathways involved in the pathophysiology of ADHD. Following this, the cerebellum and the cerebellar hemisphere share 21 causal genes, which is possibly due to their close physical proximity.

Identification results of sTWAS

In the sTWAS, AS events are marked by LeafCutter [36], an approach that only marks the introns of AS events of the pre-mRNA. Across the 14 brain tissues, we prioritized a total of 2,653 different AS events, which can be mapped to 966 unique genes (Supplementary Table 2). Supplementary Figure 10 provides

the Venn diagrams summarizing the number of significant AS events tested by the four MR methods across the 14 brain tissues. Notably, four unique AS events were identified as significant by all the methods. In the pituitary, two AS events were identified as significant by all the methods (Fig. 7a–7c) and can be mapped to PPP1R16A gene (Figure 7d). Notably, PPP1R16A was also prioritized by eTWAS. This further supports the evidence that PPP1R16A is a causal gene for ADHD. In the cerebellar hemisphere and the putamen, one AS event in each tissue was identified as significant by all the methods, mapping to the GGT7 and TREM2 genes, respectively (Fig. 7e–7h).

Additionally, we presented a schematic diagram showing the causal AS events of the causal genes that were found in eTWAS and sTWAS (Fig. 8a–8f). Due to the limited space, we only showed the genes that were significant in at least three methods in the eTWAS here while others were shown in Supplementary Figs 11–91. We found that HYAL3 was identified in the cerebellar hemisphere, cerebellum, and pituitary in the eTWAS, and seven causal AS events of HYAL3 were also found in the sTWAS within these tissues. We depicted the schematic of the excised introns of HYAL3 to explain the causal mechanism at the sequence level (Figure 8a). It can be observed that in each of the three tissues, there is an AS event with the splicing region from 50 295 619 to 50 299 213 on Chromosome 3, indicating that these tissues may produce similar upstream products of the HYAL3 gene, which could influence the onset of ADHD. Another gene, MROH8, was prioritized by the eTWAS and was found to have two causal AS events in the cortex (Figure 8b). These two AS events share a common feature: their introns both terminate at 37,143,717 bp on Chromosome 20. NAA80 is also the overlapping gene of eTWAS and sTWAS within the cortex and is mapped by two causal AS events, which both ended up at 50,299,213 bp on Chromosome 3 (Figure 8c). Besides the aforementioned genes, ELOVL1, USP4, and ZFTRAF1 are the genes prioritized by both eTWAS and sTWAS in the frontal cortex, cerebellum, and nucleus accumbens, respectively, each with only one ADHD causal AS event within the corresponding tissue (Fig. 8d–8f).

In addition, we used a heatmap to display the number of common AS events between each pair of the 14 brain tissues. The nucleus accumbens and the cerebellar hemisphere, as well as the pituitary and the cerebral cortex, have 15 overlapping AS events. A higher number of overlapping AS events suggests that the genes influencing ADHD may exhibit similar AS patterns.

Multi-tissue enrichment analysis

We constructed a bubble plot of the results of GO enrichment analyses (Figure 9). We removed the CC terms from the chart because they provide little help in understanding the pathogenesis of ADHD. Finally, GO terms were significantly enriched in only six tissues among the 14 brain tissues.

Among all the tissues, the hippocampus showed the highest number of enriched terms, with a total of 15. In addition to common processes such as the 'Primary metabolic process' ($FDR = 4.39 \times 10^{-6}$), the term 'DNA repair' ($FDR = 4.49 \times 10^{-2}$) and 'DNA damage response' ($FDR = 3.25 \times 10^{-3}$) were uniquely enriched in the hippocampus. This indicates that the influence of the hippocampus on ADHD may be related to DNA damage or repair processes. Furthermore, the ERBB signaling pathway was also enriched in the hippocampus ($FDR = 4.23 \times 10^{-2}$). This enrichment suggests that the genes involved in the ERBB signaling pathway might be critical in the hippocampus, potentially affecting processes such as cell proliferation, differentiation, migration, and survival [37, 38], thereby influencing the pathogenesis of
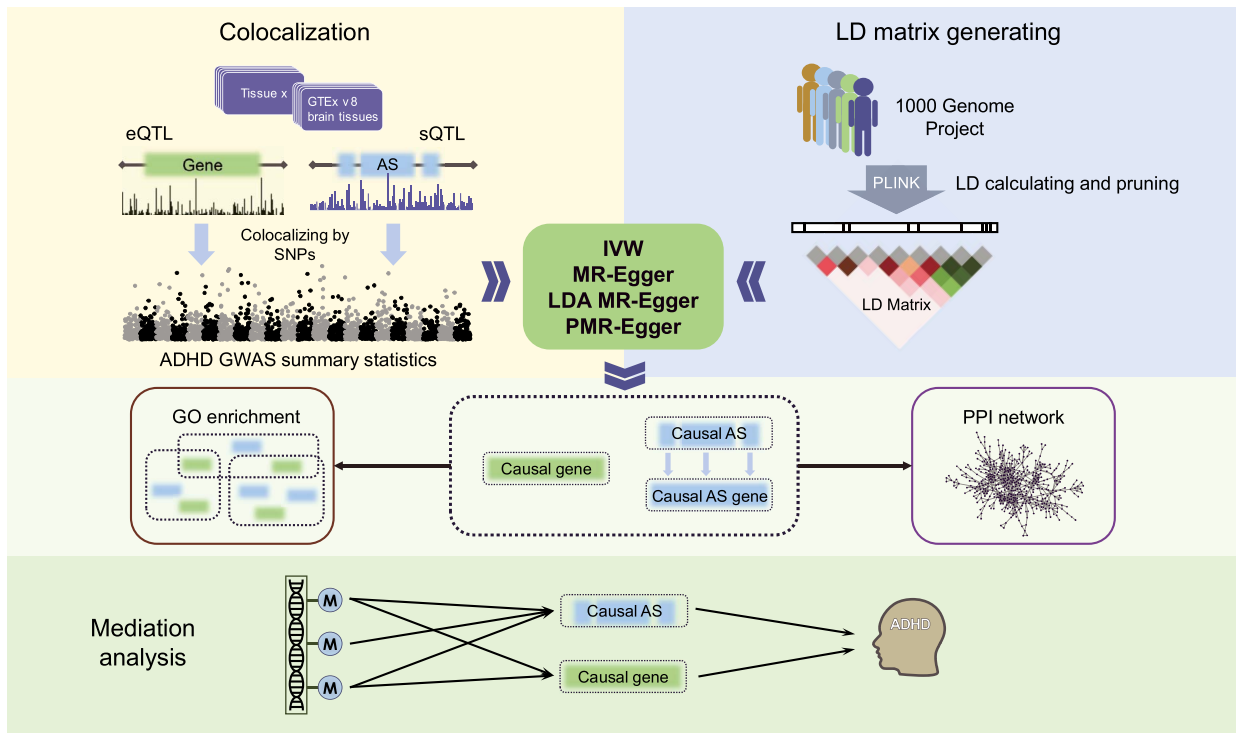
Figure 4. The study design in this paper. The workflow delineates the data, statistical methods, and bioinformatics analyses utilized in our study. Genes or AS events outlined with dashed borders in the figure indicate their significant causal effects on ADHD.
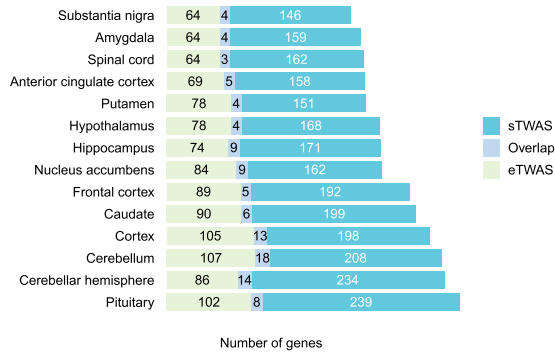

Figure 5. Bar chart showing the summary of the numbers of genes and the overlapping genes in eTWAS and sTWAS results. The x-axis represents the number of genes, while the y-axis represents different brain tissues.

ADHD. The term 'Nitrogen compound metabolic process' was enriched in the cortex ($FDR = 5.50 \times 10^{-4}$), hippocampus ($FDR = 1.15 \times 10^{-5}$), and hypothalamus ($FDR = 1.49 \times 10^{-2}$), indicating that the metabolic activities involving nitrogen compounds are particularly active in these tissues and may be related to the development and progression of ADHD. The significant enrichment of non-membrane spanning protein tyrosine kinase (NM-PTK) activity in the anterior cingulate cortex ($FDR = 3.00 \times 10^{-2}$) suggests that the signaling mediated by these kinases could be crucial for cellular functions and behavioral regulation in the anterior cingulate cortex, thereby influencing the development and progression of ADHD.

Moreover, we found that three PANTHER PC terms were significantly enriched: 'Microtubule (binding) cytoskeletal protein' ($FDR = 2.16 \times 10^{-2}$) in the hypothalamus, 'Microtubule binding motor protein' ($FDR = 4.06 \times 10^{-2}$), and 'Exoribonuclease' ($FDR = 2.36 \times 10^{-2}$) in the cortex. These PC terms suggest that the classification of proteins expressed by the causal genes exhibits consistency within the hypothalamus and the cortex, indicating the genetic features that may influence the development of ADHD. Results of the enrichment analysis are detailed in Supplementary Table 3.

PPI network analysis

We constructed tissue-specific PPI networks for each of the 14 brain tissues inputting the union set of the genes prioritized in eTWAS as well as sTWAS. The 14 tissue-specific PPI networks are detailed in Supplementary Figure 92 and Supplementary Table 4.

Furthermore, we constructed a cross-tissue PPI network analysis, which revealed 6736 interactions across the 14 brain tissues (Supplementary Table 5). For better visualization, we depicted only the subnetwork with genes that were identified by at least three MR methods in either eTWAS or sTWAS (Figure 10). In the cross-tissue PPI network, the interaction between *USP19* in the hypothalamus and *RNF123* in the cerebellum exhibits the strongest evidence, with a combined score of 994 and a co-expression score of 430. This suggests that the co-expression of these two proteins in their respective tissues may be an important process in the pathogenesis of ADHD. Besides, the protein of *PPP1R16A* in the pituitary interacts with the protein encoded by the prioritized gene *GPT* in the cortex and the hippocampus, with a combined score of 731 and a gene fusion score of 693. This suggests that a new fusion protein containing functional domains from both *PPP1R16A* and *GPT* might be produced, playing a crucial role in ADHD pathogenesis by affecting neurodevelopment or neurotransmission. In addition to these interactions, proteins with more complex interaction networks also warrant attention.
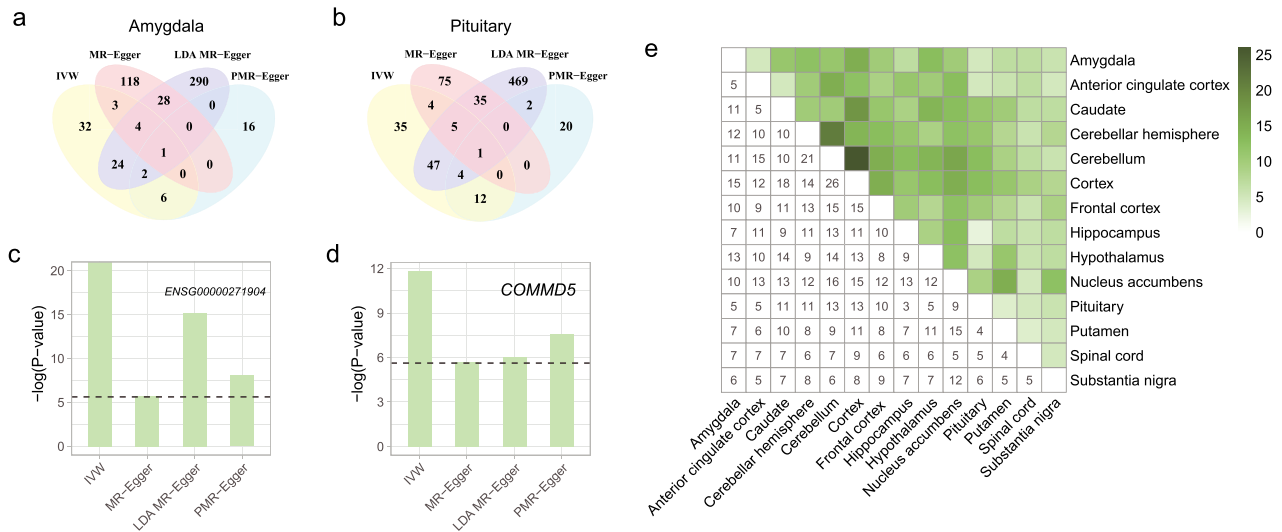
Figure 6. The genes identified by the MR methods in eTWAS. (a) A Venn diagram illustrating the number of causal genes identified by the four MR methods in the amygdala. (b) A Venn diagram illustrating the number of causal genes identified by the four MR methods in the pituitary. (c) The -log10 transformed causal P-values of *ENSG00000271904* in the amygdala, as determined by the four MR methods. (d) The -log10 transformed causal P-values of the *COMMD5* gene in the pituitary. (e) A heatmap showing the number of overlapping causal genes among the 14 brain tissues.
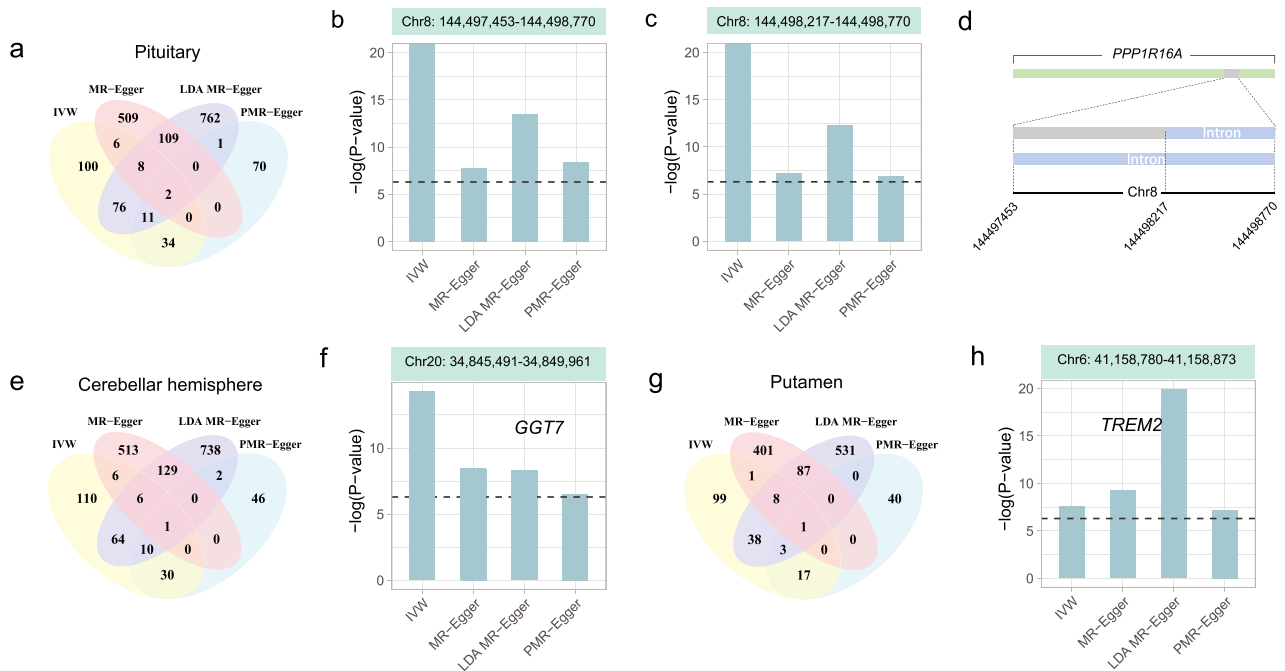

Figure 7. The AS events identified by the four MR methods in sTWAS. (a) A Venn diagram illustrating the number of identified causal AS events in the pituitary. (b) Significance of the first causal AS event of *PPP1R16A* in the pituitary. The text in the gray area above the bar chart indicates the splicing region of the AS event. (c) Significance of the second causal AS event of *PPP1R16A* in the pituitary. The text in the gray area above the bar chart indicates the splicing region of the AS event. (d) A schematic of the splicing regions for the two AS events of *PPP1R16A*. (e) A Venn diagram illustrating the number of identified causal AS events in the cerebellar hemisphere. (f) Causal significance of the AS event of *GGT7* in the cerebellar hemisphere. (g) A Venn diagram illustrating the number of identified causal AS events in the putamen. (h) Causal significance of the AS event of *GGT7* in the putamen.

Mediation analysis

We conducted mediation analysis by integrating the mQTL, eQTL, and GWAS data and revealed a total of 26 mediated regulatory relationships from DNAm sites to genes, which are visualized in a network (Figure 11). The network involves 11 genes and 15 DNAm sites. Based on the results, it can be inferred that the DNAm sites in the network exert causal effects on ADHD by mediating the expression of the connected genes. It is important to note that the samples used for mediation analysis were not specific to individual brain tissues, and thus only represent the overall epigenetic profile of the brain. However, there are four overlapping genes between the results of mediation analysis and those of eTWAS: *RHEBL1*, *LSMEM2*, *HYAL3*, and *ICA1L*. These tissue-specific causal genes for ADHD help us to understand the DNAm regulatory patterns specific to each tissue. For instance, both *RHEBL1* and *ICA1L* are the prioritized causal genes in the cortex, therefore the regulatory relationships cg15878670-*RHEBL1* and cg13521797-*ICA1L* are likely to occur in
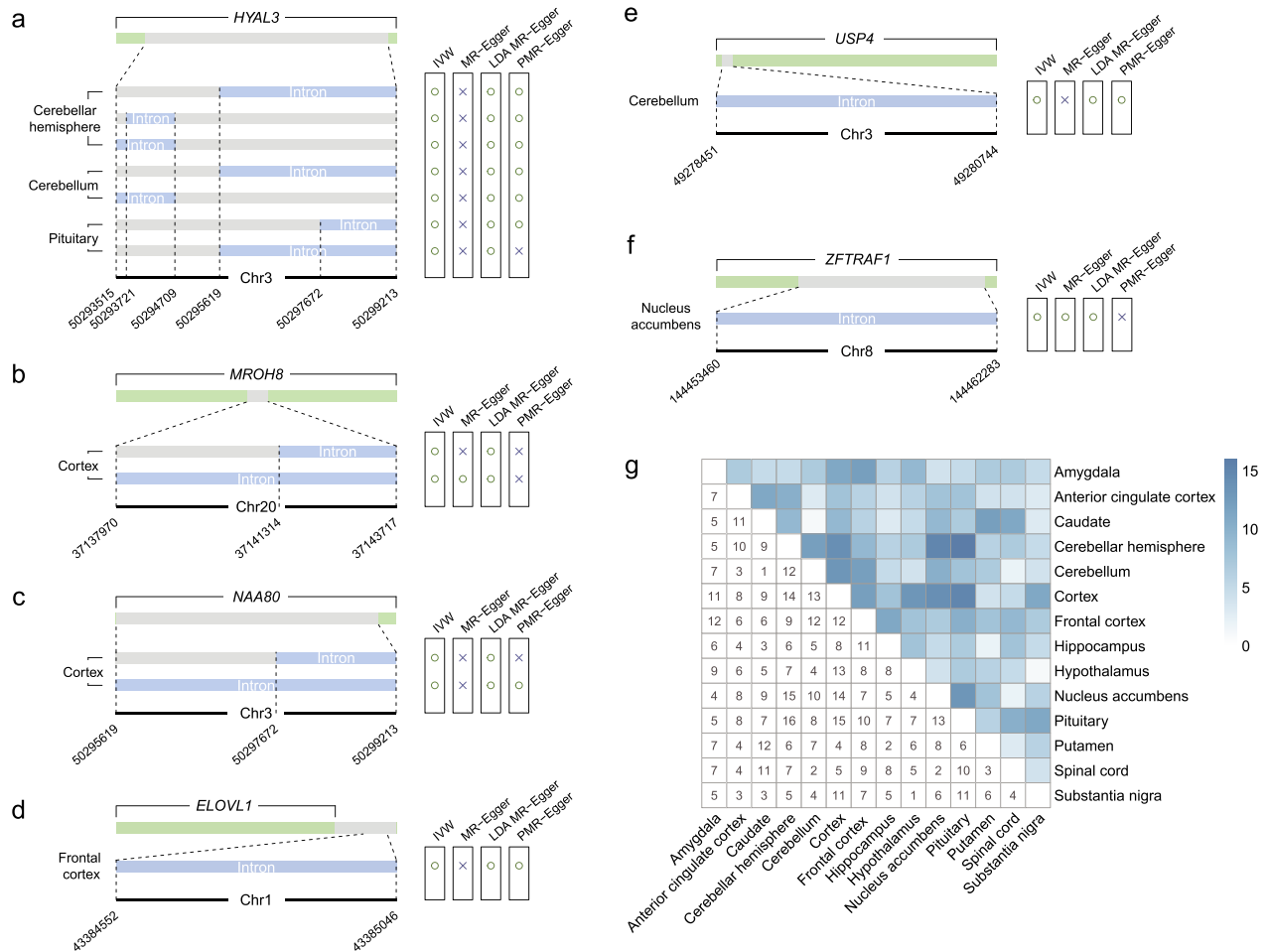
Figure 8. The schematic diagrams of AS for the overlapping genes of eTWAS and sTWAS and the overlap of AS events in different tissues. (a) The splicing diagram of *HYAL3*. The blue segments indicate the excised introns of significant causal AS events detected in different tissues. On the right side of the graph, the test results of the causal effect from the four MR methods are displayed. A green circle indicates a significant result, while a blue cross represents a non-significant result. (b) The splicing diagram of *MROH8*. (c) The splicing diagram of *NAA80*. (d) The splicing diagram of *ELOVL1*. (e) The splicing diagram of *USP4*. (f) The splicing diagram of *ZFTRAF1*. (g) A heatmap showing the numbers of overlapping genes across the 14 brain tissues.

the cortex. Besides, *HYAL3*, which is a causal gene identified in eTWAS across all the 14 brain tissues excluding the putamen, is regulated by cg03655330, cg22973319, cg02490920, cg16913124, and cg06980053 according to the network, suggesting that these regulatory relationships may be widely present throughout the brain.

We also analyzed the regulation of DNAm mediating AS that leads to ADHD by integrating the data of mQTL, sQTL, and GWAS. The findings revealed 80 regulatory relationships of DNAm sites and AS events, involving 15 genes. Among these 15 genes, five genes overlap with the AS genes identified in sTWAS, including *SLC25A22*, *PRKAG1*, *LMBR1L*, *UQCC2*, and *RPLP2*. These tissue-specific causal genes for ADHD help us understand the DNAm regulatory patterns specific to each tissue. For example, *SLC25A22* has six different AS products regulated by six different DNAm sites in the mediation analysis. In the sTWAS of the hypothalamus and nucleus accumbens, *SLC25A22* was found to have alternatively spliced products that contribute to ADHD. Therefore, it can be inferred that the causal effect of *SLC25A22* on ADHD in the hypothalamus and nucleus accumbens may result from the regulatory influence of multiple DNAm sites. Overall, we identified a total of 106 regulatory relationships, which are listed in Supplementary Table 6.

Discussion

In our study, we investigated the pathogenic mechanisms of ADHD at three levels of omics: gene expression, AS, and DNAm. To start with, we prioritized 866 genes by eTWAS and identified 966 genes with causal AS events by sTWAS across 14 brain tissues, based on the results of four MR methods, to determine their causal effects on ADHD. Using these causal genes, we conducted an enrichment analysis and identified 30 GO enriched terms in six brain tissues. Furthermore, we constructed tissue-specific and cross-tissue PPI networks using the causal genes. Finally, by integrating methylation and transcriptome data, we identified 106 processes where DNAm regulates transcription, leading to ADHD. All steps of this study are summarized in a workflow (Figure 4).

Given the results of previous research, our study primarily addressed two key issues. First, previous studies on ADHD-associated genes have relied solely on gene expression data without delving into the level of AS. For example, Demontis *et al.* conducted TWAS and identified 76 risk genes for ADHD with PrediXcan [22]. Cabana-Domínguez *et al.* found 56 genes associated with ADHD using S-PrediXcan in 15 tissues [39]. Fahira *et al.* utilized multiple datasets to investigate the causal gene expression, and prioritized 47 genes [11]. These studies identified
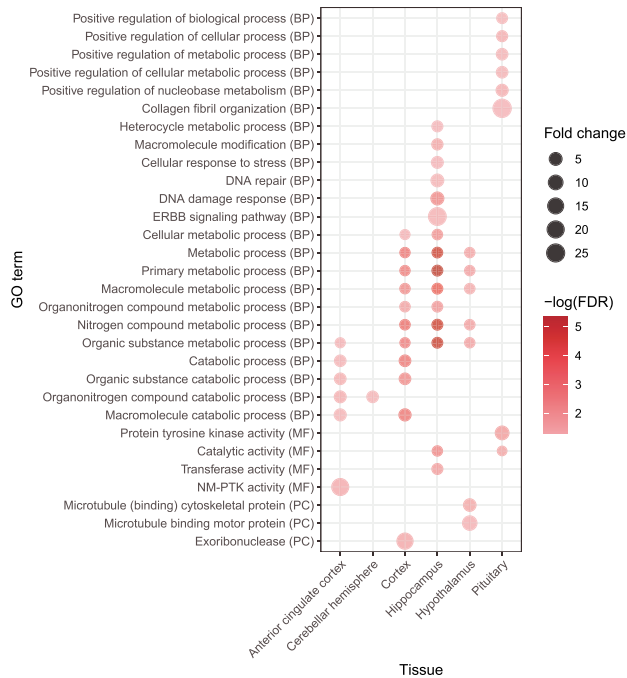
Figure 9. Bubble plot of multi-tissue GO enrichment analysis. Each enriched term is followed by a corresponding enrichment type indicated in parenthesis, where 'BP' means biological process, 'MF' stands for molecular function, and 'PC' represents PANTHER protein class. *P*-values were adjusted for FDR. The size and color of the bubbles represent the fold change and the negative logarithm (base 10) of the FDR of the term in the corresponding tissue, respectively.
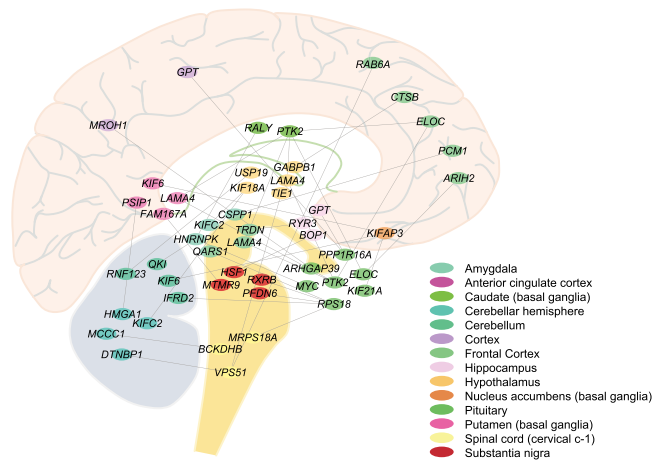

Figure 10. Cross-tissue PPI network among the 14 GTEx brain tissues. In the PPI network, each node represents a protein encoded by a gene, and different colored nodes represent different tissues.

ADHD-associated genes, such as *NUP43* and *PNPL2*. Notably, 330 of these genes are consistent with the conclusions of our research, including *LSG1*, *HYAL3*, and *PIDD*. From the above research, it is evident that the causal genes revealed by expression data are limited. By incorporating sQTL data, more causal genes for ADHD were identified in our study. Second, previous studies have lacked quantitative investigations on how methylation regulates transcription and subsequently affects ADHD. In some studies, DNAm sites are mapped to genes based on the physical locations. For instance, Neumann *et al.* identified nine DNAm sites associated with ADHD via an epigenome-wide association study,

such as cg01271805 (mapping to *ERC2*) and cg25520701 (mapping to *CREB5*) [18]. While other studies mapped the DNAm sites to the genes based on prior knowledge of annotation. For instance, Ehlinger *et al.* identified ADHD-associated DNAm sites including cg02280912, cg12603272 and cg22601108, and mapped them to gene *ZNF814*, *ELF4* and *OR6K6*, respectively [19]. In our mediation analysis, we modeled the regulatory effects of methylation on transcription and constructed a DNAm regulation network based on the analyses. This allowed us to gain deeper and more intuitive insights into the pathways through which methylation influences ADHD, such as cg15878670-*RHEBL1* and cg13521797-*ICA1L*.

Notably, we performed functional annotation for all ADHD causal genes prioritized in our study (Supplementary Table 7) using Metascape [40] (https://metascape.org/) and found that 562 of these genes have annotations related to the nervous system or psychiatric disorders (Supplementary Table 8). The evidence linking these genes to the nervous system aids in further understanding their impact on ADHD. We selected several key examples as elucidated below. For instance, *TREM2*, which was identified by sTWAS, has been reported to regulate microglial activity by enhancing their survival, proliferation, and phagocytic ability, and by modulating inflammatory responses through signaling pathways, which is crucial for clearing amyloid-beta plaques and maintaining neuronal health [41]. Additionally, *ARHGAP39*, which is involved in the PPI network, also elucidates some potential pathogenic mechanisms of ADHD based on the current scientific evidence. *ARHGAP39*, a Rho GTPase activating protein, has been proven to play a crucial role in neuronal development by regulating dendritic spine morphology and plasticity, which are essential for synaptic function, learning, and memory [42]. *MROH1* gene product in the cortex interacts with the *ARHGAP39* protein in the pituitary, with a combined score of 724. This suggests that the pituitary may secrete *ARHGAP39*, which then acts on *MROH1* in the cortex, leading to neuronal morphological changes and potentially contributing to ADHD. However, this hypothesis is based on data inference and requires further experimental validation. Another gene, *SLC25A22* which was also an overlapping gene in eTWAS and sTWAS and was involved in the DNAm regulatory network in our study, is known to disrupt mitochondrial glutamate transport, leading to severe neurological conditions such as neonatal epileptic encephalopathy [43]. These pieces of scientific evidence not only validate the reasonableness of our results but also reveal the underlying mechanisms behind these data-driven findings.

Our study has several innovative aspects. First, for the sTWAS, our study is the first to explore the causal effects of AS events on ADHD. By mapping causal AS events to their corresponding genes, our study provides an alternative perspective for identifying causal genes for ADHD. This approach yielded findings that differ from those of eTWAS, such as *PPP1R16A*. Second, we constructed a cross-tissue PPI network to understand the interactions between the products from different tissues that potentially contribute to ADHD. Third, we integrated mQTL, eQTL (or sQTL), and ADHD GWAS data to identify transcription-mediated DNAm pathogenic pathways, allowing us to gain a deeper understanding of ADHD causal genes.

The value of the study lies in different aspects. In terms of experimental research, our findings offered guidance and validation for future investigations in molecular biology and biochemical studies of ADHD. From a public health perspective, our research unveiled novel genetic and epigenetic characteristics of ADHD, which could be further explored for early prediction of
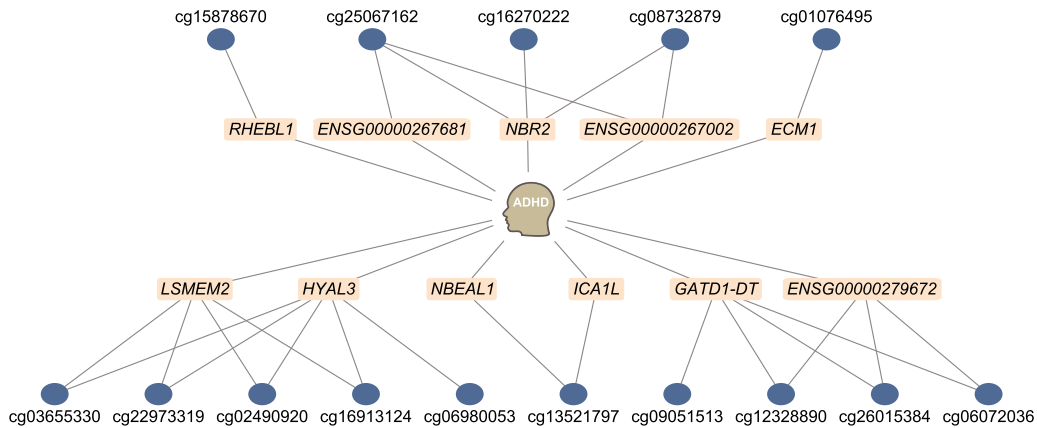
Figure 11. The DNAm regulation network constructed based on the mediation analysis. The network illustrates the complex regulatory mechanisms of 15 DNAm sites on the expression of 11 genes. In the network, rectangles represent genes, while nodes represent DNAm sites.

ADHD occurrence during fetal development. In terms of clinical implications, our findings provided potential biomarkers for diagnosis and presented opportunities for the development of targeted drug therapies, thereby expanding the repertoire of treatment modalities available for ADHD.

However, our study still has limitations. First, the sample sizes of multi-tissue eQTLs and sQTL data are limited, probably resulting in low power of statistical testing, and possibly yielding fewer causal genes. It would be beneficial to involve a larger volume of transcriptomic data. Second, considering the tissue specificity in our study, it was challenging to find suitable multi-omics data while maintaining the focus on specific tissues. Given that ADHD is a complex psychiatric disorder, its etiology cannot be solely explained by transcription and DNAm. In the future, as biological data continue evolving, additional omics, such as proteomics or metabolomics, can be introduced while maintaining tissue consistency.

Conclusion

In this study, TWAS were performed by integrating GWAS summary data with sQTL and eQTL summary data across 14 GTEx brain tissues. Four MR methods were employed to estimate and test the causal effects of AS events and gene expressions on ADHD. We identified 966 genes with significant causal AS events and 866 genes with significant causal expressions affecting ADHD. Enrichment analysis of these causal genes revealed 30 GO-enriched terms in six brain tissues. Subsequently, we constructed tissue-specific and cross-tissue PPI networks using the identified causal genes. Additionally, by integrating DNAm and transcriptome data, we identified 106 regulatory processes where DNAm influences transcription, contributing to ADHD.

Key Points
- TWAS were conducted by integrating the summary data of GWAS with the summary data of eQTL and sQTL, within 14 GTEx brain tissues.
- Four different Mendelian randomization methods were employed to estimate and test the causal effects of the alternative splicing events and gene expressions on ADHD.

- The expression of 866 genes and the alternative splicing events of 966 genes were identified to have significant causal effects on ADHD.
- Both the tissue-specific and cross-tissue PPI networks were constructed aiming to reveal the interaction and regulation of the upstream products of the causal genes identified.
- 106 regulatory pathways were inferred where DNA methylation influences ADHD through some transcription processes.

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data availability statement

GWAS summary data analysed in our study are accessible at https://pgc.unc.edu/. Access for GTEx brain eQTL and sQTL data can be found at https://www.gtexportal.org/. BrainMeta data can be downloaded at https://yanglab.westlake.edu.cn/. All the codes for this study are available for download at https://github.com/SMUJK/BIB-24-0876.

Author contributions statement

J.W., Q.W.Z., and J.H.M. were responsible for conceptualization, formal analysis, methodology, and writing original draft. S.Z., Y.W., and J.L. were responsible for deploying the codes, visualization, and formatting the paper. J.Y.Z. was responsible for project design, administration, and supervision.

References

1. Faraone SV, Asherson P, Banaschewski T. *et al.* Attention-deficit/hyperactivity disorder. *Nat Rev Dis Primers* 2015;**1**:15020. https://doi.org/10.1038/nrdp.2015.20.

2. Wilens TE, Biederman J. The stimulants. *Psychiatr Clin North Am* 1992;**15**:191–222. https://doi.org/10.1016/S0193-953X(18)30265-X.

3. Spencer T, Biederman J, Wilens T. *et al.* Pharmacotherapy of attention-deficit hyperactivity disorder across the life cycle. *J Am Acad Child Psy* 1996;**35**:409–32. https://doi.org/10.1097/00004583-199604000-00008.

4. Elia J, Ambrosini PJ, Rapoport JL. Treatment of attention-deficit-hyperactivity disorder. *N Engl J Med* 1999;**340**:780–8. https://doi.org/10.1056/NEJM199903113401007.

5. Michelson D, Buitelaar JK, Danckaerts M. *et al.* Relapse prevention in pediatric patients with ADHD treated with atomoxetine: a randomized, double-blind, placebo-controlled study. *J Am Acad Child Psy* 2004;**43**:896–904. https://doi.org/10.1097/01.chi.0000125089.35109.81.

6. Heal DJ, Cheetham SC, Smith SL. The neuropharmacology of ADHD drugs in vivo: insights on efficacy and safety. *Neuropharmacology* 2009;**57**:608–18. https://doi.org/10.1016/j.neuropharm.2009.08.020.

7. Halperin JM, Bédard AC, Curchack-Lichtin JT. Preventive interventions for ADHD: a neurodevelopmental perspective. *Neurotherapeutics* 2012;**9**:531–41. https://doi.org/10.1007/s13311-012-0123-z.

8. Faraone SV, Larsson H. Genetics of attention deficit hyperactivity disorder. *Mol Psychiatry* 2019;**24**:562–75. https://doi.org/10.1038/s41380-018-0070-0.

9. Ridler C. First genetic risk loci for ADHD identified. *Nat Rev Neurol* 2019;**15**:4. https://doi.org/10.1038/s41582-018-0117-5.

10. Zhu Z, Zhang F, Hu H. *et al.* Integration of summary data from GWAS and EQTL studies predicts complex trait gene targets. *Nat Genet* 2016;**48**:481–7. https://doi.org/10.1038/ng.3538.

11. Fahira A, Li Z, Liu N. *et al.* Prediction of causal genes and gene expression analysis of attention-deficit hyperactivity disorder in the different brain region, a comprehensive integrative analysis of ADHD. *Behav Brain Res* 2019;**364**:183–92. https://doi.org/10.1016/j.bbr.2019.02.010.

12. Klein M, Onnink M, van Donkelaar M. *et al.* Brain imaging genetics in ADHD and beyond - mapping pathways from gene to disorder at different levels of complexity. *Neurosci Biobehav R* 2017;**80**:115–55.

13. Nadler JJ, Zou F, Huang H. *et al.* Large-scale gene expression differences across brain regions and inbred strains correlate with a behavioral phenotype. *Genetics* 2006;**174**:1229–36. https://doi.org/10.1534/genetics.106.061481.

14. Qi T, Wu Y, Fang H. *et al.* Genetic control of RNA splicing and its distinct role in complex trait variation. *Nat Genet* 2022;**54**:1355–63.

15. Takata A, Matsumoto N, Kato T. Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. *Nat Commun* 2017;**8**:14519. https://doi.org/10.1038/ncomms14519.

16. Raj T, Li YI, Wong G. *et al.* Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. *Nat Genet* 2018;**50**:1584–92. https://doi.org/10.1038/s41588-018-0238-1.

17. Starnawska A, Demontis D. Role of DNA methylation in mediating genetic risk of psychiatric disorders. *Front Psych* 2021;**12**:596821. https://doi.org/10.3389/fpsyt.2021.596821.

18. Neumann A, Walton E, Alemany S. *et al.* Association between DNA methylation and ADHD symptoms from birth to school age: a prospective meta-analysis. *Transl Psychiatry* 2020;**10**:398.

19. Ehlinger JV, Goodrich JM, Dolinoy DC. *et al.* Associations between blood leukocyte DNA methylation and sustained attention in mid-to-late childhood. *Epigenomics* 2023;**15**:965–81. https://doi.org/10.2217/epi-2023-0169.

20. Beatty RA. The genetics of the mammalian gamete. *Biol Rev Camb Philos Soc* 1970;**45**:73–117. https://doi.org/10.1111/j.1469-185X.1970.tb01076.x.

21. Emdin CA, Khera AV, Kathiresan S. Mendelian randomization. *JAMA* 2017;**318**:1925–6. https://doi.org/10.1001/jama.2017.17219.

22. Demontis D, Walters GB, Athanasiadis G. *et al.* Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic architecture and implicate several cognitive domains. *Nat Genet* 2023;**55**:198–208. https://doi.org/10.1038/s41588-022-01285-8.

23. GTEx Consortium. The genotype-tissue expression (GTEx) project. *Nat Genet* 2013;**45**:580–5.

24. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 2020;**369**:1318–30. https://doi.org/10.1126/science.aaz1776.

25. Qi T, Wu Y, Zeng J. *et al.* Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat Commun* 2018;**9**:2282. https://doi.org/10.1038/s41467-018-04558-1.

26. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* 2013;**37**:658–65. https://doi.org/10.1002/gepi.21758.

27. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol* 2015;**44**:512–25. https://doi.org/10.1093/ije/dyv080.

28. Barfield R, Feng H, Gusev A. *et al.* Transcriptome-wide association studies accounting for colocalization using Egger regression. *Genet Epidemiol* 2018;**42**:418–33. https://doi.org/10.1002/gepi.22131.

29. Yuan Z, Zhu H, Zeng P. *et al.* Testing and controlling for horizontal pleiotropy with probabilistic mendelian randomization in transcriptome-wide association studies. *Nat Commun* 2020;**11**:3861. https://doi.org/10.1038/s41467-020-17668-6.

30. Moussa AA, Mandozai A, Jin Y. *et al.* Genome-wide association screening and verification of potential genes associated with root architectural traits in maize (Zea mays l.) at multiple seedling stages. *BMC Genomics* 2021;**22**:558. https://doi.org/10.1186/s12864-021-07874-x.

31. Vlietstra WJ, Vos R, van Mulligen EM. *et al.* Identifying genes targeted by disease-associated non-coding SNPs with a protein knowledge graph. *PloS One* 2022;**17**:e0271395. https://doi.org/10.1371/journal.pone.0271395.

32. 1000 Genomes Project Consortium, Auton A, Brooks LD. *et al.* A global reference for human genetic variation. *Nature* 2015;**526**:68–74. https://doi.org/10.1038/nature15393.

33. Purcell S, Neale B, Todd-Brown K. *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;**81**:559–75. https://doi.org/10.1086/519795.

34. von Mering C, Jensen LJ, Snel B. *et al*. String: Known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res* 2005;**33**:D433–7.

35. Wu Y, Zeng J, Zhang F. *et al*. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat Commun* 2018;**9**:918. https://doi.org/10.1038/s41467-018-03371-0.

36. Li YI, Knowles DA, Humphrey J. *et al*. Annotation-free quantification of RNA splicing using leafcutter. *Nat Genet* 2018;**50**:151–8. https://doi.org/10.1038/s41588-017-0004-9.

37. Yarden Y, Sliwkowski MX. Untangling the ERBB signalling network. *Nat Rev Mol Cell Biol* 2001;**2**:127–37. https://doi.org/10.1038/35052073.

38. Hynes NE, Lane HA. ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat Rev Cancer* 2005;**5**:341–54. https://doi.org/10.1038/nrc1609.

39. Cabana-Dominguez J, Llonga N, Arribas L. *et al*. Transcriptomic risk scores for attention deficit/hyperactivity disorder. *Mol Psychiatry* 2023;**28**:3493–502. https://doi.org/10.1038/s41380-023-02200-1.

40. Zhou Y, Zhou B, Pache L. *et al*. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 2019;**10**:1523. https://doi.org/10.1038/s41467-019-09234-6.

41. Rayaprolu S, Mullen B, Baker M. *et al*. TREM2 in neurodegeneration: Evidence for association of the p.R47H variant with frontotemporal dementia and parkinson's disease. *Mol Neurodegener* 2013;**8**:19. https://doi.org/10.1186/1750-1326-8-19.

42. Govek EE, Newey SE, Van Aelst L. The role of the RHO GTPases in neuronal development. *Genes Dev* 2005;**19**:1–49. https://doi.org/10.1101/gad.1256405.

43. Molinari F, Kaminska A, Fiermonte G. *et al*. Mutations in the mitochondrial glutamate carrier slc25a22 in neonatal epileptic encephalopathy with suppression bursts. *Clin Genet* 2009;**76**:188–94. https://doi.org/10.1111/j.1399-0004.2009.01236.x.