Journal of Engineering and Applied Science

REVIEWS

Open Access

Harmony in transcripts: a systematic literature review of transcriptome-wide association studies

Mahinaz A. Mashhour[1*], Ahmed Hisham Kandil[2], Manal AbdElwahed[2] and Mai S. Mabrouk[3]

*Correspondence:
Mahinaz.mashhour@must.edu.eg

[1] Biomedical Engineering Department, Misr University for Science and Technology, Giza, Egypt
[2] Systems and Biomedical Engineering Department, Cairo University, Giza, Egypt
[3] School of Information Technology and Computer Science, Nile University, Sheikh Zayed City, Egypt

Abstract

Transcriptome-wide association studies (TWAS) goal is to better understand the etiology of diseases and develop preventative and therapeutic approaches by examining the connections between genetic variants and phenotypes while overcoming the limitations of the genome-wide association study (GWAS). It is a valuable complement to GWAS, reducing the negative effects of multiple tests and enabling a more thorough investigation of gene expression patterns in various tissues. A systematic review is presented in this paper to identify articles that utilize TWAS to understand the genetic factors behind complex diseases. A detailed selection process was carried out using standard PRISMA criteria to select relevant articles for the review. Twenty-five articles passed the inclusion criteria and were selected for additional review. The studies cover a diverse range of disorders, including Tourette's syndrome, Alzheimer's disease, rheumatoid arthritis, and major depression. Leveraging gene expression data from different tissues and populations, these investigations successfully identified novel genes and pathways associated with the studied conditions. The collective findings highlight the transformative impact of integrative genomics in advancing our understanding of complex diseases, providing insights into potential therapeutic targets, and laying the foundation for precision medicine approaches.

Keywords: Transcriptome-wide association studies, Genomic wide association study, Genetics, Expression quantitative trait loci, Gene expression

Introduction

Transcriptome-wide association study (TWAS) is a cutting-edge genetic approach that uncovers the relationships between genes and certain traits such as complex diseases that aid in the understanding of how changes in the amounts of gene expression may be linked to various traits [1]. By analyzing the RNA in particular tissues, TWAS can identify which genes are active and their corresponding expression levels. TWAS offers significant insight into gene-trait interactions in a variety of complex traits since expression patterns vary across tissue types [2]. TWAS offers a framework for discovering and ranking candidate genes that may be involved in complex traits or disorders. This is

accomplished by combining genome-wide association study (GWAS) data and tissue-specific gene expression profiles [3].

GWAS is a research method used to examine the associations between genetic variants and phenotypes across different populations. The primary objective of GWAS is to enhance the understanding of the etiology of diseases to improve strategies for prevention and treatment [4]. Through conducting an analysis of polymorphisms in two distinct groups, namely a group of healthy controls and a group with the disease under investigation, it is possible to establish connections between single nucleotide polymorphisms (SNPs) and the likelihood of developing those diseases [5]. GWAS offers an objective approach to exploring the genetic foundations of phenotypes by identifying disease-associated SNPs. GWAS data can be utilized to forecast how susceptible a person is to both physical and mental ailments, based on their genotype [6]. Unfortunately, GWAS have encountered constraints in yielding therapeutic insights due to barriers to interpreting their findings, mostly because the majority of GWAS variations reside in non-coding areas of the genome, hence rendering their direct influence on gene coding sequences questionable [7].

Investigating the correlation between a trait and gene expression is an alternate strategy for deciphering the molecular basis of complicated traits. Using this strategy, we can find genes whose expression in disease-related cell types differs significantly between patients and controls using RNA sequencing. Nevertheless, performing such a study is currently not feasible because it would involve gene expression profiling on a massive scale across multiple tissues and a large number of samples in both the case and control groups.

Instead of expensive RNA sequencing, genotypes can impute cell type-specific gene expression profiles. TWAS leverages data from GWAS and a reference panel such as expression quantitative trait loci (eQTL) catalogs to directly predict gene expression in cases and controls. An eQTL is a specific location in the genome that accounts for a portion of the genetic variation in gene expression. This reference panel enables the development of a predictive model capable of imputing gene expression variation. Imputation is the statistical estimation of gene expression levels in a target population using genetic variants and a reference panel [8]. Standard eQTL analysis is conducting a direct association test between genetic variations and gene expression levels [9]. This approach eliminates the need to personally measure gene expression in each sample participating in the GWAS. This imputation is plausible because gene expression is strongly heritable. An individual's genotype is used to predict their transcriptome levels using TWAS, which trains predictors using tissue-specific eQTL maps as reference datasets. By prioritizing the heritable component of gene expression, this prediction approach enables the direct association between a disease and the expression of each gene. The prediction model is then applied to the genotyping data obtained from a GWAS. This allows for the imputation of gene expression values that are directly associated with statistical SNPs discovered during the GWAS. Once gene expression levels have been estimated, gene-trait association analyses are carried out to investigate the correlations between expected expression levels, genotypes, and observed traits among individuals in the study [10].

TWAS offers an extra benefit by reducing the problem of multiple testing penalties in GWAS during statistical inference. This is achieved by testing the imputed expression

of hundreds of genes instead of millions of SNPs in GWAS [11]. The Genotype-Tissue Expression (GTEx) project is widely recognized as the most prominent eQTL investigation, in which multiple tissues from hundreds of individuals were examined to uncover eQTLs specific to each tissue [12]. Version 8 of the GTEx has examined a total of 15,201 RNA-sequencing samples obtained from 838 postmortem donors across 49 different tissues. As a step in TWAS, GWAS examines the correlation between genetic variations and phenotype, as mentioned earlier. This can be accomplished by starting from the beginning as a stage in TWAS utilizing individual-based genetic data, or by gathering previously conducted GWAS-summary statistics. GWAS summary statistics refer to the combined p values and association data for each variant examined in a GWAS [13]. GWAS summary statistics offer advantages over individual phenotype and genotype data, such as being openly accessible, originating from meta-analyses, and bypassing challenges at the sample level. They are often derived from numerous studies, a larger cohort than individual samples, and can help identify non-normal distributions, confounding covariates, or outliers [14]. A flowchart summarizing the process of a TWAS is shown in Fig. 1.

Since 2015, various methodologies have been developed to conduct tissue-specific and multiple-tissue TWAS. Single-tissue TWAS examines the relationship between gene expression patterns in a specific tissue or cell type and complex traits or diseases, providing tissue-specific insights. Multiple-tissue TWAS analyzes the association between gene expression patterns across various tissues or cell types and the studied traits or diseases, allowing for the identification of shared and tissue-specific associations.

Single-tissue models

PrediXcan [15] uses Elastic NET regression and transcriptome data from reference panels to predict gene expression levels in specific tissues based on genotype data. FUnctional Summary-based ImputatiON (FUSION) [1] was the initial attempt to overcome PrediXcan's issue that large-scale GWAS data are only publically available at the summary association statistic level. They used the Bayesian sparse linear mixed model

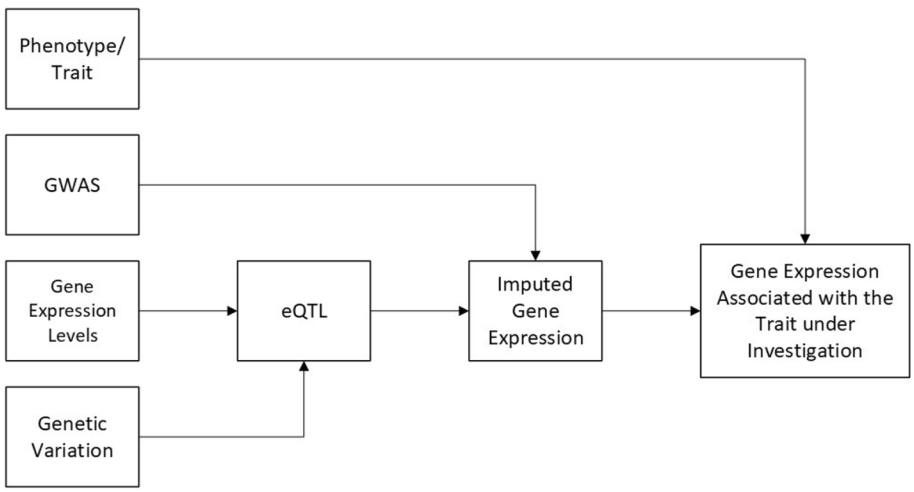

Fig. 1 Flowchart summarizing the process of a TWAS study

(BSLLM) to develop the prediction model and impute expression-trait association statistics directly from GWAS summary statistics. S-PrediXcan [16] was then introduced to extend the PrediXcan by employing GWAS summary statistics instead of genotype data to facilitate gene expression-trait associations without genetic data. Previous presented methods relied on parametric imputation models; however, they cannot model the complex genomic architecture of transcriptomic data. Transcriptome-Integrated Genetic Association Resource (TIGAR) [17] has been developed to specifically address these limitations, by employing a nonparametric Bayesian method that was originally proposed for the genetic prediction of complex traits, known as Dirichlet process regression (DPR) model. DPR is a more generalized model that uses PrediXcan's Elastic-Net and FUSION's BSLMM as special cases. Then, kernel-based transcriptome-wide association study (kTWAS) [18] was introduced, focusing on a kernel-based approach, using genomic data to construct kernels that capture genetic relationships and employing a regression framework to predict gene expression and assess associations with traits. Summary-level Unified Method for Modeling Integrated Transcriptome (SUMMIT) [19] and Omnibus Transcriptome Test using Expression Reference Summary data (OTTERS) were introduced to improve the accuracy of the expression prediction model and the power of TWAS by overcoming the limitation of small-expression reference panel sample sizes by using summary-level expression panels utilizing larger samples and allowing for more accurate expression prediction models and ultimately strengthening the power of TWAS.

Multi-tissue models

Through tissue integration, multiple tissue TWAS reveals shared and tissue-specific gene-trait associations. MultiXcan [20] extended PrediXcan by merging tissue data to create a meta-model that predicts gene expressions across tissues. Also, S-MultiXcan [20], building upon MultiXcan, predicts multi-tissue gene expressions using GWAS summary data. S-MultiXcan facilitates association testing across several tissues without requiring individual genetic information using summary statistics instead of genotype data. Hu et al. 2019 [21] addressed the limitations of previous methods, stating that previous methodologies often train separate imputation models for different tissues, neglecting transcriptional regulation similarities. They introduced the Unified Test for MOlecular SignaTures (UTMOST) framework that involves training cross-tissue expression imputation, assessing single-tissue associations, and using a generalized BerkJones test for each gene to summarize single-tissue association statistics into a powerful metric that quantifies the gene-trait association. Finally, the joint-tissue imputation (JTI) [22] approach was developed as an extension to improve target tissue prediction accuracy by integrating all tissues using a weighted square error loss function, preferring comparable tissues over dissimilar ones.

In conclusion, multiple approaches have been proposed, each one achieving a balance between specificity and breadth in association testing by utilizing different prediction models.

This article examines the various ways in which TWAS techniques can be used to uncover the complicated genetic foundations of traits and disorders. By combining gene expression data with genetic information, TWAS offers a potent approach to uncover

the regulatory mechanisms that control variations in traits. One of its key benefits is its ability to provide significant scientific knowledge by explaining how genetic variations affect gene expression and, therefore, the characteristics of various tissues. Using TWAS, tissue-specific effects can be identified, shedding light on the complex functions of genes in many biological settings. Finally, TWAS is a remarkable tool that will change the face of precision medicine and therapies by opening the door to a detailed understanding of the genetic architecture of complex disorders.

Methods

To ensure transparency, a comprehensive analysis of existing research was conducted following the well-established Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [23].

Search strategy and inclusion criteria

To assess the latest research in the context of TWAS, a comprehensive search for relevant studies was conducted. The search was conducted electronically to select papers published in the previous 10 years in the PubMed database. The keywords and search algorithms employed to refine the selection of articles that are significant to this study are presented below.

"Transcriptome-Wide Association Study" AND ((y_10[Filter]) AND (ffrft[Filter]) AND (excludepreprints[Filter]) AND (humans[Filter]) AND (data[Filter]) AND (english[Filter]))".

Studies were selected based on three inclusion requirements: (1) focusing primarily on human traits (2) employing TWAS, and (3) should be an original article. Figure 2 demonstrates the PRISMA flowchart that outlines the criteria for selecting studies and the grounds for their exclusion.

Data collection

The relevant data were extracted from the articles after performing a qualitative screening of publications and acquisition of related research that satisfied the inclusion requirements. The subsequent information was collected from every article: year of publication, investigated trait, type of data used, TWAS approach, and their findings.

Results and discussion

Three hundred and sixty-one references were retrieved from the PubMed database. After performing a preliminary assessment of each publication, 341 articles were eliminated because they failed to fulfill the inclusion requirements. After evaluating 47 suitable full-text references, 22 proved irrelevant and were eliminated. Eventually, 25 articles were chosen for final evaluation based on the previously demonstrated eligibility criteria. Figure 2 illustrates the criteria used for research inclusion.

Researchers conducted different TWAS approaches to gain an understanding of the intricate relationship between genetic variations and complex traits or diseases. Table 1 presents the findings and methods employed in the chosen studies. Figure 3 analyzes the distribution of the TWAS algorithms employed in the chosen research.

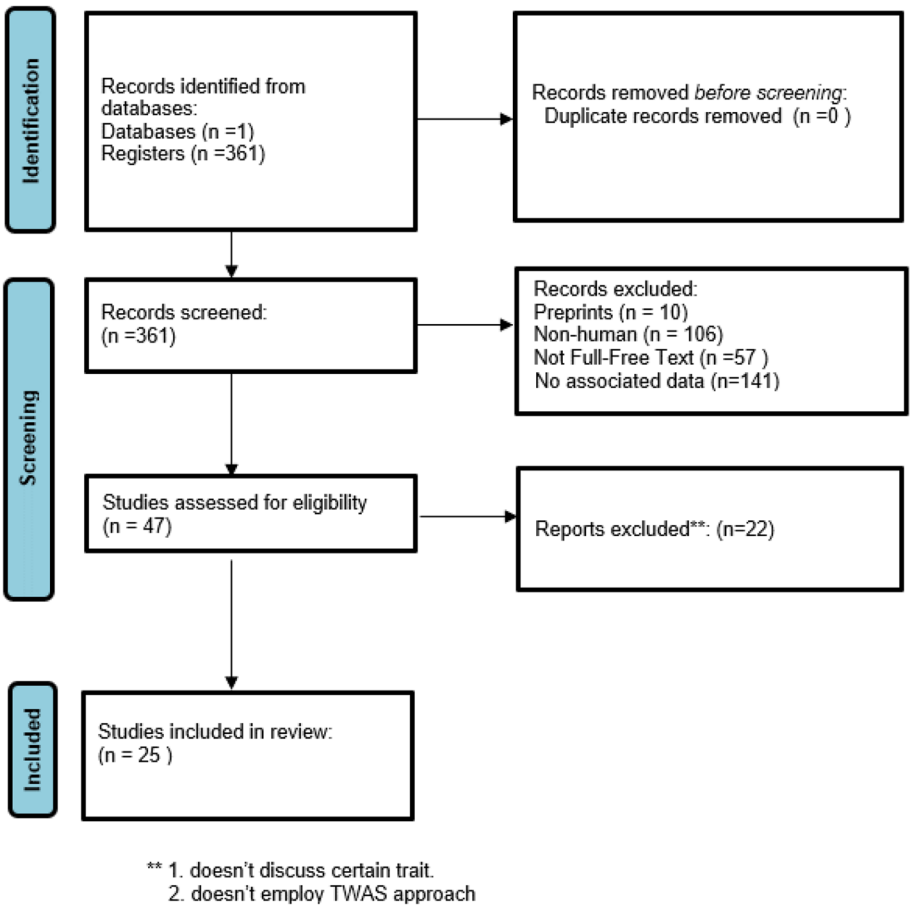

Fig. 2 PRISMA flowchart

Fifteen of the presented studies used FUSION as their approach to further understand complex diseases. Liao et al. [24] investigated the biological significance of GWAS signals of Tourette's syndrome and determined gene targets for further functional analysis by performing a TWAS utilizing summary statistics from a recent GWAS involving more than 14,000 participants. They successfully Provided evidence that elevated FLT3 expression in the dorsolateral prefrontal cortex is linked to Tourette syndrome. Li et al. [26] attempted to convert the GWAS findings of Depression into risk genes by combining GWAS summary statistics from 807,553 individuals with summary-level gene-expression data from the dorsolateral prefrontal cortex of the samples. They successfully identified fifty-three risk genes associated with depression, 23 of which were not included in the initial GWAS, and 7 were found to be associated with depression in the two separate brain eQTL datasets. Gockley et al. [27] adjusted the FUSION TWAS pipeline to incorporate gene expression data from various neocortical regions by conducting a TWAS analysis on Alzheimer's disease, using weights that were trained based on RNA-Seq expression values obtained from six different cortical regions. Consequently, they presented proof of genetic variations that contribute to the risk of Alzheimer's disease through 8 genes located in six different genomic regions. Park et al. [28] conducted a TWAS to uncover genes associated with

Table 1 Findings and methods employed in the chosen studies

Authors	Trait	Data		Methods	Findings
		GWAS	Tissue		
Liao et al. 2022 [24]	Tourette's syndrome	GWAS Summary statistics	14 tissue types	Fusion	Increase in expression of the *FLT3* gene across many brain tissues
Levey et al. 2021 [25]	Major depressive disorder	Individual-based	13 brain and whole-blood tissues	MetaXcan	Links were observed with the expression of the *DRD2* gene in the nucleus accumbens and the *NEGR1* gene in the hypothalamus
Li et al. 2021 [26]	Depression	GWAS summary statistics	Tissue-type-specific 54 human tissues	FUSION	Of the 53 genes associated with Depression, 23 were not included in the initial GWAS and 7 were found to be associated with depression in the 2 separate brain eQTL datasets
Gockley et al. 2021 [27]	Alzheimer's disease	GWAS summary statistics	6 distinct cortical regions	FUSION	They identified an association between 8 different genes and Alzheimer's disease
Park et al. 2021 [28]	Amyotrophic lateral sclerosis	GWAS summary statistics	19 tissue reference panels	FUSION	They identified an association between 7 novel genes and Amyotrophic lateral sclerosis
Traylor et al. 2021 [29]	Lacunar stroke	GWAS summary statistics + additional cases	Multi-tissue	FUSION	They identified the association between 6 novel genes and lacunar stroke
Yao et al. 2021 [30]	Bipolar disorder	GWAS summary statistics	13 brain tissues	FUSION	The risk of bipolar disorder is associated with 44 genes whose expression levels can be predicted genetically. Additionally, 11 novel genes were found in the cerebellar hemisphere, 1 of which is ASB16
Wang et al. 2021 [31]	Schizophrenia	GWAS summary statistics	Peripheral blood and brain tissues	FUSION	The expression of TMEM180 mRNA was found strongly linked to an increased risk of developing schizophrenia
Bhat et al. 2021 [32]	Mismatch negativity	Individual level	Cortex and frontal cortex	PrediXcan	They identified the association between 2 novel genes and Mismatch Negativity

Table 1 (continued)

Authors	Trait	Data		Methods	Findings
		GWAS	Tissue		
Xu et al. 2021 [33]	Hand osteoarthritis	GWAS summary statistics	Skeletal muscle and blood	FUSION	They identified the association between 5 novel genes and Hand Osteoarthritis
Reus et al. 2021 [34]	Frontotemporal dementia	GWAS summary statistics	53 tissue types	FUSION	They identified 73 significant gene-tissue associations, involving 44 distinct genes across 34 different types of tissues
Liu et al. 2021 [35]	Alzheimer's disease	GWAS summary statistics	Hippocampal tissue	S-PrediXcan	They identified the association between 24 novel genes and Alzheimer's disease in hippocampal tissue
Bruinooge et al. 2021 [36]	Inflammatory bowel disease	Individual-level	44 non-diseased human tissue	PrediXcan	They discovered that different genetically regulated genes in different tissues, including skeletal muscle, the cerebellar hemisphere of the brain, and the frontal cortex of the brain, are associated with inflammatory bowel disease
Huang et al. 2021 [37]	Autism spectrum disorder	GWAS summary statistics	10 brain tissues	UTMOST	31 genes were discovered to be associated with Autism, including the *POU3F2* gene
Kia et al. 2021 [38]	Parkinson's disease	GWAS summary statistics	10 brain regions	FUSION	They identified the association between 11 novel genes with Parkinson's disease
Wu et al. 2021 [39]	Rheumatoid arthritis	GWAS summary statistics	4 different tissues	FUSION	They identified the association between 692 novel genes with rheumatoid arthritis, four of which were associated and the four tissues
Dall'Aglio et al. 2021 [40]	Major depression	GWAS summary statistics	21 tissue datasets	FUSION	They linked 94 novel genes to major depression, half of which were novel

Table 1 (continued)

Authors	Trait	Data		Methods	Findings
		GWAS	Tissue		
Guo et al. 2021 [41]	Colorectal cancer risk	GWAS summary statistics	Different colon tissues, including carcinoma and adenoma tissues	MetaXcan	They linked 25 unique genes to colorectal cancer, including 4 novel loci. Furthermore, in 9 known GWAS loci, they discovered nine novel genes
Lu et al. 2018 [42]	Epithelial ovarian cancer risk	GWAS summary statistics	53 different tissues	MetaXcan	They discovered 35 genes, which include FZD4, a possible new epithelial ovarian cancer risk
Wu et al. 2018 [43]	Breast cancer	Individual-based	Breast tissue	PrediXcan	They linked 48 genes to breast cancer, including 14 novel genes
Shi et al. 2019 [44]	Age at natural menopause	Individual-based	Normal hypothalamus and ovarian tissues	PrediXcan	They revealed 34 genes strongly linked with natural age Menopause, including 4 entirely novel genes, located over 1 Mb away from any previously identified genetic variations linked to menopause through GWAS, 24 genes found inside known GWAS regions but not previously associated with menopause, and 6 previously discovered genes
Gusev et al. 2018 [45]	Schizophrenia	GWAS summary statistics	Brain, blood, and adipose tissues	FUSION	157 unique genes were linked to schizophrenia, 35 of which did not match an existing GWAS locus
Lamontagne et al. 2018 [46]	Chronic obstructive pulmonary disease	GWAS summary statistics	Lung tissues	S-PrediXcan	They identified the association between 12 genes/loci and the disease
Thériault et al. 2018 [47]	Calcific aortic valve stenosis	GWAS summary statistics	Aortic valve tissues	FUSION	The study identified PALMD as a susceptibility gene

Table 1 (continued)

Authors	Trait	Data		Methods	Findings
		GWAS	Tissue		
Mancuso et al. 2017 [48]	30 different complex traits	GWAS summary statistics	Multi-tissue	FUSION	They identified 1196 complex trait-associated genes, including 168 unique genes. Furthermore, 43 trait pairs had a substantially high association with estimated expression

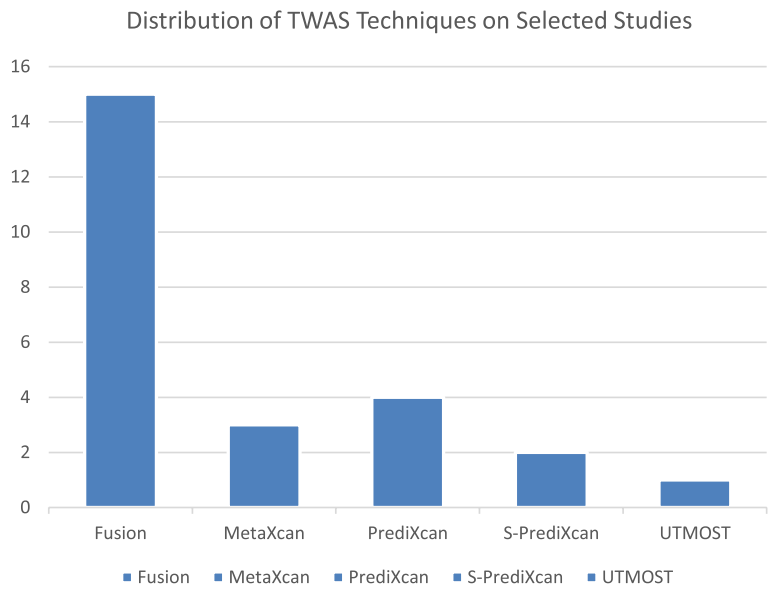

Fig. 3 Distribution of TWAS techniques in the selected studies

amyotrophic lateral sclerosis. They successfully discovered seven novel genes in amyotrophic lateral sclerosis using the greatest GWAS summary statistic ($n = 80{,}610$) and 19 tissue reference panels. Traylor et al. [29] combined data from recently recruited lacunar stroke patients and previous GWAS to implement a TWAS to identify genes associated with lacunar stroke and successfully found links between six genes and lacunar stroke. Yao et al. 2021 [30] employed TWAS to reveal novel bipolar disorder risk genes and causative genes at GWAS-previously identified loci. They discovered 14 conditionally independent genes and 11 novel genes. They also showed that the Bipolar Disorder GWAS is influenced by genetically regulated expression, resulting in many genome-wide meaningful signals. Wang et al. [31] performed integrated analysis using blood eQTL data and GWAS data to investigate schizophrenia in East Asian populations and demonstrated a significant association between reduced TMEM180 mRNA expression and the risk of schizophrenia. Xu et al. 2021 [33] utilized the GWAS summary of hand osteoarthritis to conduct a TWAS while employing skeletal muscle

and blood as a reference for gene expression. As a result, they successfully identified 177 genes linked with skeletal muscle and 423 genes associated with blood. Reus et al. [34] conducted a TWAS to discover genes with anticipated expression levels linked to frontotemporal dementia. This was achieved by integrating GWAS summary statistics with reference gene expression data. A total of 73 gene-tissue associations were discovered for frontotemporal dementia, encompassing 44 distinct genes across 34 different tissue types. Kia et al. 2021 [38] attempted to enhance our comprehension of the fundamental genes and mechanisms at the earlier discovered GWAS loci to gain insight into the development of Parkinson's disease by employing TWAS and successfully identified the association between 11 novel genes with Parkinson's disease. Wu et al. [39] attempted to find genetic factors associated with rheumatoid arthritis by applying TWAS considering four distinct tissue summary data from a GWAS involving 5539 patients and 20,169 controls. They successfully discovered a total of 692 genes, with four of them being linked to the four used tissues. Dall'Aglio et al. [40] conducted a TWAS to investigate the genetic factors of major depression. The analysis relied on summary statistics obtained from the largest genome-wide association study of major depression, which included a sample size of 135,458 cases and 344,901 controls. Additionally, gene expression levels from 21 tissue datasets were included. They linked 94 novel genes to major depression, half of which were novel. Although GWAS have shown a large number of genetic regions associated with an increased risk of schizophrenia, the specific mechanisms responsible for this link are still largely unclear. Gusev et al. [45] conducted a TWAS by combining a schizophrenia GWAS involving 79,845 individuals with expression data obtained from 3693 control individuals. They successfully discovered 157 genes, 35 of which were not associated with any previously reported GWAS location. By integrating GWAS and eQTL data, Thériault et al. [47]were able to determine the underlying molecular factors responsible for calcific aortic valve stenosis. Through TWAS, they discovered that the *PALMD* gene is strongly linked to calcific aortic valve stenosis. Finally, Mancuso et al. [48] utilized gene expression data from 45 panels and combined it with summary GWAS data to conduct 30 TWASs, which involved analyzing gene expression across many tissues. Of the 1196 genes related to these phenotypes, 168 are more than 0.5 Mb from any previously published GWAS significant variant.

In the second stage, PrediXcan was utilized in 4 studies, where Bhat et al. 2021 [32] conducted a TWAS on a sample of 728 individuals to examine the genetic factors underlying Mismatch negativity, an electrophysiological response that measures the cortical's ability to adapt to unexpected stimulation. This study identified two genes, FAM89A and ENGASE, whose expression in cortical tissues is linked to mismatch negativity. Bruinooge et al. [36] utilized TWAS to examine the genetic factors that underlie Inflammatory bowel disease utilizing genetically regulated gene expression patterns that were inferred from the genetic profiles of 240 individuals with inflammatory bowel disease and 44 non-diseased human tissue-specific reference models obtained from the GTEx. They discovered that different genetically regulated genes in different tissues, including skeletal muscle, the cerebellar hemisphere of the brain, and the frontal cortex of the brain, are associated with Inflammatory bowel disease. In order to find new risk locations and genes suspected to cause breast cancer, Wu et al. [43] conducted a TWAS study that

analyzed the relationships between genetically predicted gene expression and breast cancer risk. The study included 122,977 cases and 105,974 controls of European descent and linked 48 genes to breast cancer, including 14 novel genes. Finally, Shi et al. [44] attempted to discover new genes that make individuals more susceptible to experiencing natural menopause at a certain age. They revealed 34 genes strongly linked with natural age menopause, including 4 entirely novel genes, located over 1 Mb away from any previously identified genetic variations linked to menopause through GWAS, 24 genes found inside known GWAS regions but not previously associated with menopause, and six previously discovered genes.

MetaXcan was employed in 3 studies, where Levey et al. [25] performed a comprehensive meta-analysis of depression using TWAS, and observed links between the major depressive disorder and the expression of the *DRD2* gene in the nucleus accumbens and the *NEGR1* gene in the hypothalamus. Guo et al. [41] performed a TWAS to discover potential genes linked to colorectal cancer. They linked 25 unique genes to colorectal cancer, including 4 novel loci. Furthermore, in 9 known GWAS loci, they discovered nine new novel genes. Lu et al. [42] conducted a TWAS in order to identify new genomic regions and potential causative genes at previously identified GWAS regions. They successfully discovered 35 genes, including FZD4, a possible new epithelial ovarian cancer risk factor.

S-PrediXcan was utilized in two studies. Liu et al. [35] explored the relationship between gene expression in the hippocampus and Alzheimer's disease using TWAS and identified the association between 24 novel genes and Alzheimer's disease in hippocampal tissue. Lamontagne et al. [46] attempted to identify genes that may cause chronic obstructive pulmonary disease and provide valuable biological insights into the recently identified chronic obstructive pulmonary disease susceptibility loci. They identified an association between 12 genes/loci and chronic obstructive pulmonary disease. Finally, Huang et al. [37] utilized UMOST to perform a TWAS to better understand the genetic factors behind autism spectrum disorder. As a result, 31 genes were discovered to be associated with autism, including the *POU3F2* gene.

Our main goal was to act as a reference for future TWAS investigations. The framework describes various computational models that are used at each computational stage and highlights the significance of choosing models that are in line with SNP regulatory effects on target genes and relevant tissues related to the trait under study. Subsequently, case studies of TWAS implementations are demonstrated, including case studies. After a comprehensive examination of 15 studies that employed the FUSION approach and further studies using PrediXcan, MetaXcan, S-PrediXcan, and UMOST, an intriguing pattern was revealed that highlights the critical function of TWAS in interpreting complex genetic factors of a range of complex diseases. The FUSION studies demonstrate the diversity of disorders examined, from Alzheimer's disease to Tourette's syndrome, and the effectiveness of TWAS in identifying new genes and pathways linked to these disorders. By integrating gene expression data from various tissues and populations with GWAS summary statistics, scientists have been able to understand previously unknown genetic variations, which has led to important new understandings of the molecular mechanisms underlying disease. Also, the effective utilization of PrediXcan, MetaXcan, S-PrediXcan, and UMOST in various settings highlights the adaptability of these

techniques in determining the genetic components of disorders such as autism spectrum disorder, Alzheimer's disease, breast cancer, inflammatory bowel disease, and mismatch negativity. When taken as a whole, these studies demonstrate how genomic research is changing and how it may change how we understand complicated diseases by opening up new possibilities for tailored medicine and more focused therapeutic interventions.

Conclusion

Aspects of expanded TWAS applications were examined in this review article, which also sheds light on the significance of gene-trait associations for complex diseases and traits. Providing an all-encompassing examination of recent developments, methodologies, and practical implementations in the field of complex trait analysis. The presented array of studies employing TWAS and related methodologies shed light on the pivotal role of integrative genomics in advancing our understanding of complex diseases. These investigations not only unravel the complex genetic landscapes associated with various disorders but also showcase the adaptability of TWAS methodologies across different types of conditions. The findings presented in these studies not only contribute to our understanding of the genetic underpinnings of diseases such as Tourette's syndrome, Alzheimer's, and rheumatoid arthritis but also unveil novel genes and pathways that may serve as potential therapeutic targets. Furthermore, the application of advanced methodologies of TWAS in subsequent stages of research emphasizes the need for comprehensive and multidimensional approaches in deciphering the genetic architecture of complex traits.

Although this review comprehensively summarizes the applications of TWAS, it is important to acknowledge certain inherent limitations. As TWAS is a rapidly evolving field, methodologies and tools are constantly changing, making it difficult to directly compare studies. Additionally, the heterogeneity observed in the types of diseases, tissues, and populations studied can make it challenging to draw generalized conclusions. Lastly, statistical complexities and challenges in interpreting biological mechanisms further necessitate cautious interpretation of results. Despite these limitations, TWAS's transformative impact in revealing the genetic foundations of complex disorders is clear, and future research in this field promises to advance our understanding of disease genesis while establishing the path for novel personalized disease prevention, diagnosis, and treatment, ultimately fostering a new era in precision medicine.

Abbreviations

TWAS	Transcriptome-wide association studies
GWAS	Genome-wide association study
SNP	Single nucleotide polymorphism
eQTL	Expression Quantitative Trait Loci
GTEx	Genotype-tissue expression
FUSION	FUnctional Summary-based ImputatiON
BSLMM	Bayesian sparse linear mixed model
TIGAR	Transcriptome-Integrated Genetic Association Resource
DPR	Dirichlet process regression
kTWAS	Kernel-based transcriptome-wide association study
SUMMIT	Summary-level Unified Method for Modeling Integrated Transcriptome
OTTERS	Omnibus Transcriptome Test using Expression Reference Summary data
UTMOST	Unified Test for MOlecular SignaTures
JTI	Joint-tissue imputation
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Acknowledgements
Not applicable.

Authors' contributions
MM contributed to data processing and analysis, manuscript drafting and revision, and figure design. AK identified the research focus, conceptualized the analysis, and contributed to data processing and analysis, figures design, and manuscript revision. MA identified the research focus, conceptualized the analysis, verified analytical methods, and drafted and revised the manuscript. MS verified analytical methods, supervised the research findings, and contributed to writing and revising the manuscript. All authors have reviewed and approved the submitted version of this manuscript and any subsequent substantially modified versions that incorporate their individual contributions. Each author agrees to be held accountable for their own contributions to this work and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and documented in the literature.

Funding
No funding was received to assist with the preparation of this manuscript.

Availability of data and materials
Not applicable.

Declarations

Competing interests
The authors declare that they have no competing interests.

Received: 17 April 2024 Accepted: 24 July 2024
Published online: 02 August 2024

References

1. Gusev A et al (2016) Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet 48(3):245–252. https://doi.org/10.1038/ng.3506
2. Wainberg M et al (2019) Opportunities and challenges for transcriptome-wide association studies. Nat Genet 51(4):592–599. https://doi.org/10.1038/s41588-019-0385-z
3. Li B, Ritchie MD. From GWAS to gene: transcriptome-wide association studies and other methods to functionally understand GWAS discoveries. Front Genet 2021;12. https://doi.org/10.3389/fgene.2021.713230.
4. Visscher PM et al (2017) 10 years of GWAS discovery: biology, function, and translation. Am J Human Genet 101(1):5–22. https://doi.org/10.1016/j.ajhg.2017.06.005. Cell Press.
5. Jeck WR, Siebold AP, Sharpless NE (2012) Review: a meta-analysis of GWAS and age-associated diseases. Aging Cell 11(5):727–731. https://doi.org/10.1111/j.1474-9726.2012.00871.x
6. Uffelmann E et al (2021) Genome-wide association studies. Nat Rev Methods Primers 1(1):59. https://doi.org/10.1038/s43586-021-00056-9
7. Cano-Gamez E, Trynka G. From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. Front Genet 2020;11. https://doi.org/10.3389/fgene.2020.00424.
8. Cao C, Ding B, Li Q, Kwok D, Wu J, Long Q (2021) Power analysis of transcriptome-wide association study: Implications for practical protocol choice. PLoS Genet 17(2):e1009405. https://doi.org/10.1371/journal.pgen.1009405
9. Nica AC, Dermitzakis ET (2013) Expression quantitative trait loci: present and future. Philos Transact Royal Soc B: Biol Sci 368(1620):20120362. https://doi.org/10.1098/rstb.2012.0362
10. Xie Y, Shan N, Zhao H, Hou L (2021) Transcriptome wide association studies: general framework and methods. Quant Biol 9(2):141–150. https://doi.org/10.15302/J-QB-020-0228
11. Kho PF et al (2021) Multi-tissue transcriptome-wide association study identifies eight candidate genes and tissue-specific gene expression underlying endometrial cancer susceptibility. Commun Biol 4(1):1211. https://doi.org/10.1038/s42003-021-02745-3
12. Aguet F et al (2020) The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science (1979) 369(6509):1318–1330. https://doi.org/10.1126/science.aaz1776
13. MacArthur J et al (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res 45(D1):D896–D901. https://doi.org/10.1093/nar/gkw1133
14. Svishcheva GR, Belonogova NM, Zorkoltseva IV, Kirichenko AV, Axenovich TI (2019) Gene-based association tests using GWAS summary statistics. Bioinformatics 35(19):3701–3708. https://doi.org/10.1093/bioinformatics/btz172
15. Gamazon ER et al (2015) A gene-based association method for mapping traits using reference transcriptome data. Nat Genet 47(9):1091–1098. https://doi.org/10.1038/ng.3367
16. Barbeira AN et al (2018) Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun 9(1):1825. https://doi.org/10.1038/s41467-018-03621-1
17. Nagpal S et al (2019) TIGAR: an improved Bayesian tool for transcriptomic data imputation enhances gene mapping of complex traits. Am J Human Genet 105(2):258–266. https://doi.org/10.1016/j.ajhg.2019.05.018
18. Cao C. et al. kTWAS: integrating kernel machine with transcriptome-wide association studies improves statistical power and reveals novel genes. Brief Bioinform 2021;22(4). https://doi.org/10.1093/bib/bbaa270.
19. Zhang Z, Bae YE, Bradley JR, Wu L, Wu C (2022) SUMMIT: an integrative approach for better transcriptomic data imputation improves causal gene identification. Nat Commun 13(1):6336. https://doi.org/10.1038/s41467-022-34016-y

20. Barbeira AN, Pividori M, Zheng J, Wheeler HE, Nicolae DL, Im HK (2019) Integrating predicted transcriptome from multiple tissues improves association detection. PLoS Genet 15(1):e1007889. https://doi.org/10.1371/journal.pgen.1007889
21. Hu Y et al (2019) A statistical framework for cross-tissue transcriptome-wide association analysis. Nat Genet 51(3):568–576. https://doi.org/10.1038/s41588-019-0345-7
22. Zhou D, Jiang Y, Zhong X, Cox NJ, Liu C, Gamazon ER (2020) A unified framework for joint-tissue transcriptome-wide association and Mendelian randomization analysis. Nat Genet 52(11):1239–1246. https://doi.org/10.1038/s41588-020-0706-2
23. Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 6(7):e1000097. https://doi.org/10.1371/journal.pmed.1000097
24. Liao C et al (2022) Transcriptome-wide association study reveals increased neuronal FLT3 expression is associated with Tourette's syndrome. Commun Biol 5(1):289. https://doi.org/10.1038/s42003-022-03231-0
25. Levey DF et al (2021) Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. Nat Neurosci 24(7):954–963. https://doi.org/10.1038/s41593-021-00860-2
26. Li X et al (2021) Transcriptome-wide association study identifies new susceptibility genes and pathways for depression. Transl Psychiatry 11(1):306. https://doi.org/10.1038/s41398-021-01411-w
27. Gockley J et al (2021) Multi-tissue neocortical transcriptome-wide association study implicates 8 genes across 6 genomic loci in Alzheimer's disease. Genome Med 13(1):76. https://doi.org/10.1186/s13073-021-00890-2
28. Park S, Kim D, Song J, Joo JWJ (2021) An integrative transcriptome-wide analysis of amyotrophic lateral sclerosis for the identification of potential genetic markers and drug candidates. Int J Mol Sci 22(6):3216. https://doi.org/10.3390/ijms22063216
29. Traylor M et al (2021) Genetic basis of lacunar stroke: a pooled analysis of individual patient data and genome-wide association studies. Lancet Neurol 20(5):351–361. https://doi.org/10.1016/S1474-4422(21)00031-4
30. Yao S et al (2021) Epigenetic element-based transcriptome-wide association study identifies novel genes for bipolar disorder. Schizophr Bull 47(6):1642–1652. https://doi.org/10.1093/schbul/sbab023
31. Wang J-Y et al (2021) Integrative analyses followed by functional characterization reveal *TMEM180* as a Schizophrenia risk gene. Schizophr Bull 47(5):1364–1374. https://doi.org/10.1093/schbul/sbab032
32. Bhat A et al (2021) Transcriptome-wide association study reveals two genes that influence mismatch negativity. Cell Rep 34(11):108868. https://doi.org/10.1016/j.celrep.2021.108868
33. Xu J et al (2021) Integrating transcriptome-wide association study and mRNA expression profile identified candidate genes related to hand osteoarthritis. Arthritis Res Ther 23(1):81. https://doi.org/10.1186/s13075-021-02458-2
34. Reus LM et al (2021) Gene expression imputation across multiple tissue types provides insight into the genetic architecture of frontotemporal dementia and its clinical subtypes. Biol Psychiatry 89(8):825–835. https://doi.org/10.1016/j.biopsych.2020.12.023
35. Liu N et al (2021) Hippocampal transcriptome-wide association study and neurobiological pathway analysis for Alzheimer's disease. PLoS Genet 17(2):e1009363. https://doi.org/10.1371/journal.pgen.1009363
36. Bruinooge A et al (2021) Genetic predictors of gene expression associated with psychiatric comorbidity in patients with inflammatory bowel disease – a pilot study. Genomics 113(3):919–932. https://doi.org/10.1016/j.ygeno.2021.02.001
37. Huang K et al (2021) Transcriptome-wide transmission disequilibrium analysis identifies novel risk genes for autism spectrum disorder. PLoS Genet 17(2):e1009309. https://doi.org/10.1371/journal.pgen.1009309
38. Kia DA et al (2021) Identification of candidate parkinson disease genes by integrating genome-wide association study, expression, and epigenetic data sets. JAMA Neurol 78(4):464. https://doi.org/10.1001/jamaneurol.2020.5257
39. Wu C et al (2021) Transcriptome-wide association study identifies susceptibility genes for rheumatoid arthritis. Arthritis Res Ther 23(1):38. https://doi.org/10.1186/s13075-021-02419-9
40. Dall'Aglio L, Lewis CM, Pain O (2021) delineating the genetic component of gene expression in major depression. Biol Psychiatry 89(6):627–636. https://doi.org/10.1016/j.biopsych.2020.09.0105
41. Guo X et al (2021) Identifying novel susceptibility genes for colorectal cancer risk from a transcriptome-wide association study of 125,478 subjects. Gastroenterology 160(4):1164-1178.e6. https://doi.org/10.1053/j.gastro.2020.08.062
42. Lu Y et al (2018) A transcriptome-wide association study among 97,898 women to identify candidate susceptibility genes for epithelial ovarian cancer risk. Cancer Res 78(18):5419–5430. https://doi.org/10.1158/0008-5472.CAN-18-0951
43. Wu L et al (2018) A transcriptome-wide association study of 229,000 women identifies new candidate susceptibility genes for breast cancer. Nat Genet 50(7):968–978. https://doi.org/10.1038/s41588-018-0132-x
44. Shi J et al (2019) Transcriptome-wide association study identifies susceptibility loci and genes for age at natural menopause. Reprod Sci 26(4):496–502. https://doi.org/10.1177/1933719118776788
45. Gusev A et al (2018) Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. Nat Genet 50(4):538–548. https://doi.org/10.1038/s41588-018-0092-1
46. Lamontagne M et al (2018) Leveraging lung tissue transcriptome to uncover candidate causal genes in COPD genetic associations. Hum Mol Genet 27(10):1819–1829. https://doi.org/10.1093/hmg/ddy091
47. Thériault S et al (2018) A transcriptome-wide association study identifies PALMD as a susceptibility gene for calcific aortic valve stenosis. Nat Commun 9(1):988. https://doi.org/10.1038/s41467-018-03260-6
48. Mancuso N, Shi H, Goddard P, Kichaev G, Gusev A, Pasaniuc B (2017) Integrating gene expression with summary association statistics to identify genes associated with 30 complex traits. Am J Human Genet 100(3):473–487. https://doi.org/10.1016/j.ajhg.2017.01.031

Publisher's Note