Genomics, Proteomics & Bioinformatics, 2024, **22(5)**, qzae077
https://doi.org/10.1093/gpbjnl/qzae077
Advance access publication: 29 October 2024
Review

Multiome-wide Association Studies: Novel Approaches for Understanding Diseases

Mengting Shao [1], Kaiyang Chen [1], Shuting Zhang [1], Min Tian [1], Yan Shen [1], Chen Cao [1,*], Ning Gu [1,2,*]

[1]Key Laboratory for Bio-Electromagnetic Environment and Advanced Medical Theranostics, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing 211166, China
[2]Nanjing Key Laboratory for Cardiovascular Information and Health Engineering Medicine, Institute of Clinical Medicine, Nanjing Drum Tower Hospital, Medical School, Nanjing University, Nanjing 210093, China

*Corresponding authors: caochen@njmu.edu.cn (Cao C), guning@nju.edu.cn (Gu N).

Handling Editor: Zhongming Zhao

Abstract

The rapid development of multiome (transcriptome, proteome, cistrome, imaging, and regulome)-wide association study methods have opened new avenues for biologists to understand the susceptibility genes underlying complex diseases. Thorough comparisons of these methods are essential for selecting the most appropriate tool for a given research objective. This review provides a detailed categorization and summary of the statistical models, use cases, and advantages of recent multiome-wide association studies. In addition, to illustrate gene–disease association studies based on transcriptome-wide association study (TWAS), we collected 478 disease entries across 22 categories from 235 manually reviewed publications. Our analysis reveals that mental disorders are the most frequently studied diseases by TWAS, indicating its potential to deepen our understanding of the genetic architecture of complex diseases. In summary, this review underscores the importance of multiome-wide association studies in elucidating complex diseases and highlights the significance of selecting the appropriate method for each study.

Key words: Genome-wide association study; Transcriptome-wide association study; Multiome; Gene-based association study; Complex disease.

Introduction

The genome-wide association study (GWAS) was first proposed in 2005 [1] and aim to discover statistical associations between millions of genomic variants and phenotypes, particularly for complex traits in humans [2–6]. GWAS screens a large number of whole genomes to find variants that appear more frequently in individuals with a specific trait than those without it. Over the past two decades, research into GWAS methods has expanded rapidly, and GWAS has been successfully applied in many datasets to identify numerous genetic variants related to multiple complex traits and diseases [7–9]. However, GWAS still has significant limitations such as poor interpretability and insufficient statistical power [10], due to its use of statistical associations to estimate the correlation, not causation, between variants and traits. Furthermore, many false-positive associations have been identified by GWAS as a result of linkage disequilibrium (LD) [11–13].

With the development of high-throughput sequencing technologies and the increasing number of multiomic datasets, such as transcriptomes, proteomes, and epigenomes, many multiome-wide association study methods have emerged to address the limitations of GWAS. These innovative methods include transcriptome-wide association study (TWAS) [14,15], proteome-wide association study (PWAS) [16], cistrome-wide association study (CWAS) [17], imaging-wide association study (IWAS) [18–20], and regulome-wide association study (RWAS) [21,22] (**Figure 1**). These approaches are collectively termed "multiome-mediated" methods due to their use of an additional data source as an intermediate step between genotype and phenotype association (**Figure 2A**). Generally, multiome-mediated methods aim to identify biomarkers related to both genetic variants and phenotypes, thereby improving statistical power and enabling the interpretation of genetic variants through their associated biomarkers (Figure 2A). The integrated analysis of these multiomic datasets refines our understanding of human traits by providing insights into the genetic determinants underlying complex diseases.

TWAS was the first framework established to integrate multiomic data following the development of GWAS (Figures 1 and 2B). The two earliest TWAS methodologies PrediXcan [14] and TWAS-FUSION [15] have been widely adopted, allowing biologists to better identify and interpret susceptibility genes in various diseases [23–27], such as calcific aortic valve stenosis [28], macular degeneration [29], schizophrenia [30,31], and also applied into other domains, including drug repurposing [32,33].

Following the development of TWAS, the proteome was next integrated with GWAS, based on the hypothesis that genetic variants in coding regions influence phenotypes by affecting the biochemical function of protein products. PWAS, introduced in 2020 [16] (Figures 1 and 2B), is a novel framework for identifying gene–trait associations mediated by functional changes in proteins. PWAS estimates the effect of variants on protein function using FIRM [34], a machine-learning model that considers the proteomic context of each variant. This approach addresses the limitations of GWAS, particularly the poor interpretability and limited statistical power due to numerous variant loci.

Received: 16 August 2023; **Revised:** 6 October 2024; **Accepted:** 23 October 2024.

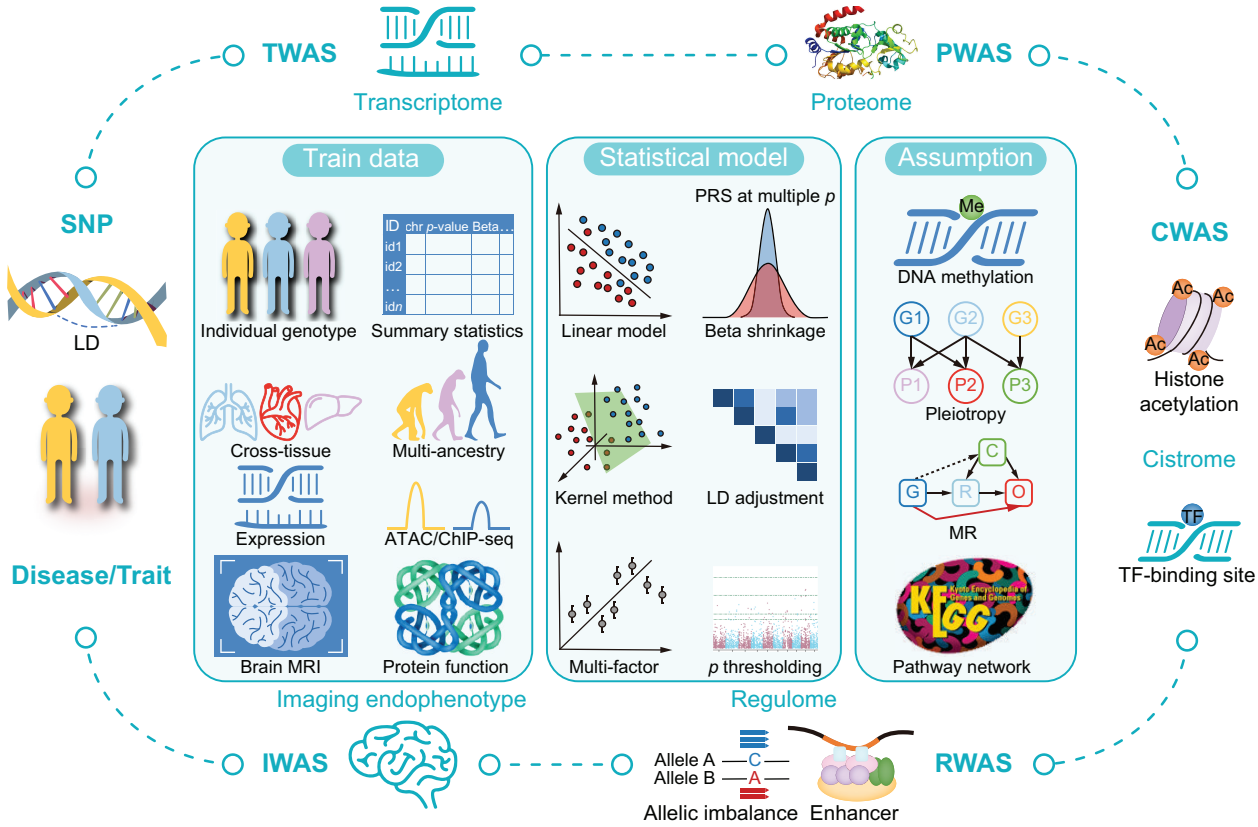

Figure 1 Overview of multiome-wide association study methods

This figure depicts the datasets, models, and biological assumptions employed in the development of multiome-wide association study methods. Outer ring: path indicating how multiome-wide association study methods link SNPs to diseases or traits, thereby increasing the interpretation and power of traditional GWAS methods. Left: from top to bottom, three parts show primary data sources used in multiome-wide association study methods. Starting with individual-level genotype data and GWAS summary statistics. Then some extensions to TWAS methods distinguish between various tissues (cross-tissue) or ancestries (multi-ancestry) to search for tissue- or ancestry-dependent variants and mechanisms. Finally, data sources for five multiome-mediated methods are also shown, which are expression data for TWAS, ATAC-seq data for RWAS, ChIP-seq data for CWAS, imaging data for IWAS, and protein function data for PWAS. Middle: statistical models used in multiome-mediated methods. Linear models such as LASSO and BSLMM are widely used in the traditional pipeline, and the kernel machine method is a non-linear model for capturing the interaction of variants. The multi-factor model is used to investigate the effects of multiple genes on traits. Beta shrinkage, LD adjustment, and multiple *p*-value thresholding are widely applied in polygenic risk score methods, which extend TWAS methods to enable the training of models on summary statistics. Right: biological assumptions used in multiome-mediated methods. Epigenetic data, such as DNA methylation, are based on the assumption that integrated *cis*-regulatory elements can help improve the accuracy of expression prediction. Pleiotropy, MR, and pathway networks are other common tools used to enhance the power of multiome-mediated methods. P, phenotype; C, confounding factor; O, outcome; R, risk factor; G, genetic variant; SNP, single-nucleotide polymorphism; GWAS, genome-wide association study; TWAS, transcriptome-wide association study; PWAS, proteome-wide association study; CWAS, cistrome-wide association study; RWAS, regulome-wide association study; IWAS, imaging-wide association study; LD, linkage disequilibrium; ATAC-seq, assay for transposase-accessible chromatin using sequencing; ChIP-seq, chromatin immunoprecipitation followed by sequencing; MRI, magnetic resonance imaging; PRS, polygenic risk score; TF, transcription factor; Me, methylation; Ac, acetylation; LASSO, least absolute shrinkage and selection operator; BSLMM, Bayesian sparse linear mixed model; MR, Mendelian randomization.

Similar to PWAS, IWAS was proposed in 2017 [18] (Figures 1 and 2B). This approach is a potent method that integrates GWAS with imaging endophenotypes. It has been demonstrated that IWAS enhances statistical power and improves the biological interpretability of GWAS findings [35]. Additionally, IWAS has shown that the TWAS methodology can be extended to a larger variety of endophenotypes beyond gene expression, with many potential applications.

Recently, CWAS [17] was introduced to integrate GWAS with epigenomics (Figures 1 and 2B). The main goal of CWAS is to discover chromatin quantitative trait loci (cQTLs) and discern allele-specific effects on chromatin states associated with traits. The efficacy of CWAS has been demonstrated through its application to epigenetic and GWAS summary statistic data on prostate cancer. In contrast to utilizing expression quantitative trait loci (eQTLs) in TWAS, CWAS can detect and prioritize more variants relative to transcriptome factors, making it a promising approach for future genetic investigations.

Additionally, two independent methods termed RWAS have recently been published [21,22] (Figures 1 and 2B). The first method employs stratAS [36] for allelic imbalance analysis using cancer assay for transposase-accessible chromatin using sequencing (ATAC-seq) data. It then identifies associated regulatory elements, their impact mechanism, and potential upstream regulators through allele-specific accessibility QTLs (as-aQTLs). Another pipeline, facilitated by the MAGMA software package [37], aims to link enhancers with diseases and has shown success applied to schizophrenia data. Essentially, these two RWAS pipelines represent innovative approaches that integrate regulome data with GWAS data. They hold promise for improving the power and interpretability of GWAS findings, thereby aiding in discovering novel genetic factors underlying complex traits.

As this brief history of multiome-mediated methods demonstrates, there are many different ways to improve and modify existing GWAS methods that have been successfully

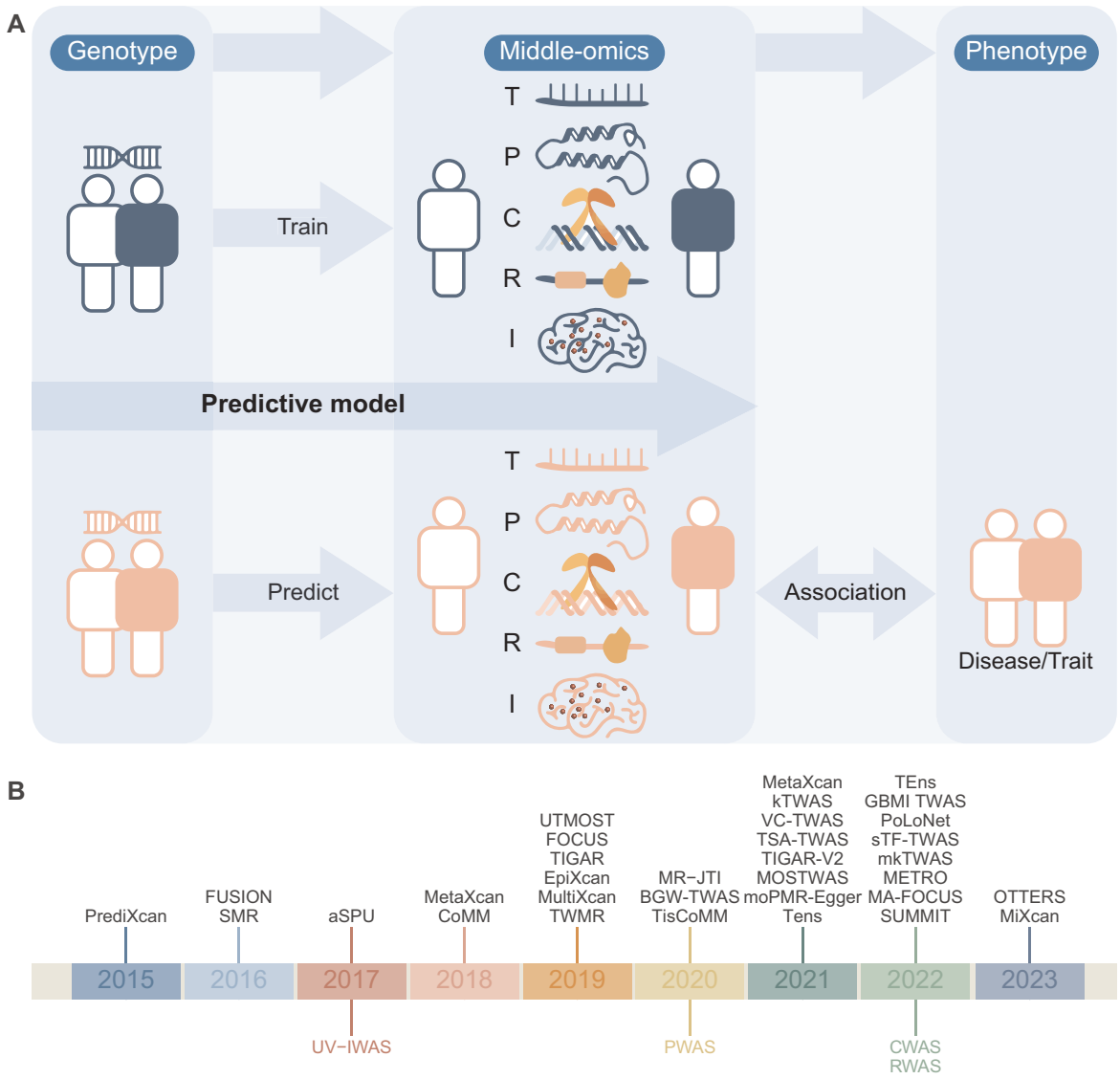

Figure 2 General framework and landmark publications of multiome-wide association study methods
A. The general workflow of multiome-wide association study methods. First (top), different statistical models are utilized to estimate how genotype determines multiome-mediated features such as the transcriptome (T), proteome (P), cistrome (C), regulome (R), or images (I). Then, the trained model is used to predict multiome-mediated features from test data. Finally, associations between the predicted multiome-mediated features and complex human traits or diseases are identified (bottom). **B**. Timeline for the development of key multiome-wide association study methods over the past few years. The upper black font represents the TWAS methods, while non-TWAS methods are indicated by colored font below the year.

used to identify disease-associated genes. To aid researchers navigate the large number of methods that have been developed, our review provides a comprehensive summary and analysis of the statistical models, use cases, and advantages of all currently available multiome-wide association study methods. Furthermore, using Alzheimer's disease (AD) as an example, we discussed multiple studies employing various TWAS methods to discover susceptibility genes associated with AD. We also explored the limitations of current studies as well as the factors impacting the association results.

Classification of multiome-wide association study methods

TWAS

TWAS, a gene-based approach, identifies associations between genetically regulated expression (GReX) components and complex diseases or traits. The biological hypothesis underlying GReX is that variations in complex diseases and traits often result from the cumulative effect of genetic variations across many genes, with single genetic variants typically exerting small effects on phenotypic variation [38,39]. Thus, TWAS considers the collective contribution of single-nucleotide polymorphism (SNPs) within a single gene or genetic region to phenotype rather than focusing on a single locus of genetic variation. Notably, eQTL represents a crucial class of regulatory variations elucidating the relationship between genetic variants and the regulation of transcription and translation [40–42].

The traditional TWAS framework comprises three steps (Figure 2A). First, an imputation model, such as a linear model, is trained on a reference panel (*e.g.*, Genotype-Tissue Expression Consortium [43]) containing matched genotype and transcriptome data. This model utilizes individual-level

genotypes to predict gene expression associated with GReX [14]. Subsequently, a single-tissue or cross-tissue model is applied to a GWAS cohort lacking transcriptome data to estimate gene expression from the genetic variants within the cohort [44,45]. Finally, gene–trait association tests are conducted on the predicted expression, either within a specific tissue or across multiple tissues.

TWAS stands out as the most widely used multiome-wide association study. Since its inception in 2015 [14], various improvements have been made to the traditional TWAS approach to boost accuracy and applicability. Here, we review representative extension methods, categorizing them into five distinct groups based on the differences in algorithmic construction and statistical models employed in various TWASs (**Table 1**).

Extension in improving GReX accuracy

The first category focuses on improving the accuracy of GReX predictions. GReX, initially proposed in 2015 with the development of PrediXcan [14], focuses on the mechanism of gene expression regulation, excluding the effect of environmental factors and traits. PrediXcan employs the elastic net [46] to train weights between *cis*-SNPs and gene expression based on a reference transcriptome dataset. These weights, along with individual-level genotype data, are utilized to predict GReX, which is then assessed for association with phenotypes of interest.

Subsequently, a Bayesian sparse linear mixed model (BSLMM)-based method, FUSION [15] was presented, which achieves the best power in the BSLMM model among five statistical models also including best linear unbiased predictor (BLUP), least absolute shrinkage and selection operator (LASSO) regression model [47], elastic net [46], and the most significantly associated SNP model. FUSION uses BSLMM to capture weights between *cis*-SNPs and gene expression, and this model [48] combines the Bayesian variable selection regression (BVSR) model and the linear mixed model (LMM), offering a higher degree of accuracy. The traditional TWAS framework is derived from PrediXcan and FUSION. Although the two methods differ in their training models and input data types, they form the basis of the standard TWAS approach. Researchers have continued to refine each step of the traditional framework, resulting in extensive research [49].

Compared with parametric imputation methods adopted by both PrediXcan and FUSION, the nonparametric Bayesian model is also suitable for estimating the complex effect of genetics on the transcriptome because it takes parametric models as special cases. Thus, a nonparametric Bayesian method was proposed to estimate the effects of *cis*-eQTLs through data-driven approaches [50]. Comparative analysis indicated that the nonparametric Bayesian method provided a better fit for transcriptomic imputation models under specific settings than the parametric Bayesian method. The authors also introduced TIGAR [50], a software tool that allows for the use of both parametric and nonparametric methods.

To increase efficiency and reduce computational costs, TIGAR-V2 [51] was developed. It directly accepts variant call format (VCF) files of individual-level and summary-level GWAS data as input, and enables parallel computation. TIGAR-V2 offers nonparametric Bayesian Dirichlet process regression (DPR) [52] and elastic net for training imputation models. TIGAR-V2 reduces training computation time by 90% and memory usage by 50% compared with TIGAR [50].

Beyond the model construction step of TWAS, confounding factors such as LD and co-regulation in the genome and transcriptome may reduce the accuracy of TWAS [41,53–57]. FOCUS [58] was proposed to address this. It utilizes a standard Bayesian approach [48] to compute the posterior inclusion probability of each gene associated with the TWAS signal, effectively controlling the effects of LD, predictive weights, and pleiotropy on association outcomes. To avoid overfitting, FOCUS employs a multivariate Gaussian prior, which leads to strong agreement between simulation results and real data. To improve the identification power of susceptibility genes, the multi-ancestry FOCUS model (MA-FOCUS) [59] was developed in 2022, assuming that disease or complex trait-related genes are shared across ancestries [60]. MA-FOCUS employs a Bayesian approach for posterior computation, and simulation results demonstrate its robustness across various datasets and its efficacy in identifying susceptibility genes across ancestries.

While the aforementioned TWAS methods were designed under distinct biological assumptions regarding gene contribution to complex traits or *cis*-SNP effects on gene expression, their efficacy under different biological scenarios remains unclear. Thus, a novel TWAS method, tissue specificity-aware TWAS (TSA-TWAS) [44], was developed to capture the effect of tissue specificities on TWAS power. In simulation studies, TSA-TWAS utilizes two representative prediction models, elastic net and LASSO, and two association approaches, principal component (PC) regression [61] and generalized Berk-Jones (GBJ) test [45], to show that different TWAS protocols have different power based on different biological questions. TSA-TWAS integrates and maximizes the power of multiple methods while controlling type-I error rates based on the single tissue matched to the trait-related tissue. Applying TSA-TWAS to acquired immune deficiency syndrome (AIDS) Clinical Trial Group (ACTG) clinical trial data highlighted its ability to identify both statistically significant and novel associations.

Extension in adopting cross-tissue datasets

To overcome the limitation of single-tissue sample size, UTMOST [45] was proposed to perform cross-tissue gene expression imputation using multivariate regression. It improves the prediction accuracy across all available tissues by combining the results of testing used in each tissue with the cross-tissue z-score accuracy of expression values. Another method, MultiXcan [61], also utilizes multivariate regression for gene–trait association testing across multiple tissues but differs from UTMOST by utilizing the first k PCs of the predicted gene expression. Real data results applied on 222 traits in UK Biobank (UKBB) demonstrate that MultiXcan can discover more associations for some traits, while simulation results reveal that MultiXcan performs poorly under the assumption that causal expression in a specific tissue is known.

In addition to cross-tissue approaches, the Mendelian randomization (MR) method helps avoid reverse causation and detect confounding factors in observational studies by examining the causal effect of exposure variables on phenotypes. To further improve TWAS accuracy, MR-JTI (joint-tissue

Table 1 Classification and detailed information of the multiome-wide association study methods

Classification		Method	Website	Description	Input	Output	Data	Ref.
TWAS	Improving GReX accuracy	PrediXcan	https://github.com/hakyimlab/PrediXcan	Correlate estimated genetically regulated gene expression with the phenotype	IGD	GTAs	GTEx; GEUVADIS; DGN; WTCCC; BioVU	[14]
		FUSION	http://gusevlab.org/projects/fusion/	Train weights between *cis*-SNPs and gene expression based on BSLMM	IGD; GSS	GTAs	METSIM; YFS; NTR; lipids2010; MuTHER	[15]
		TIGAR	https://github.com/yanglab-emory/TIGAR	Obtain the weights between genes based on a nonparametric Bayesian model	IGD; GSS	GTAs	ROS/MAP; 1000 Genomes Project; IGAP	[50]
		TIGAR-V2	https://github.com/yanglab-emory/TIGAR	Take VCF files as input and provide both DPR and elastic net to train imputation models	IGD; GSS	GTAs	GTEx; BCAC; OCAC	[51]
		FOCUS	https://github.com/bogdanlab/focus/	Calculate the probability of each gene associated with the TWAS signal based on Bayesian model	GSS	GTAs	1000 Genomes Project; lipids2010	[58]
		MA-FOCUS	https://github.com/mancusolab/ma-focus	Identify causal genes by considering which are shared across ancestries related to diseases or complex traits	GSS	GTAs	GENOA; GEUVADIS; EAS; HapMap; DisGeNET	[59]
		TSA-TWAS	https://github.com/RitchieLab/multi_tissue_twas_sim	Identify tissue-specific causal genes	IGD; GSS	GTAs	ACTG; GTEx	[44]
	Adopting cross-tissue datasets	UTMOST	https://github.com/Joker-Jerome/UTMOST/	Apply cross-tissue weights on expression imputation	IGD; GSS	GTAs	GERA; IGAP; 1000 Genomes Project; GTEx; GEUVADIS	[45]
		MultiXcan/ S-MultiXcan	https://github.com/hakyimlab/MetaXcan	Utilize the first *k* principal components of the estimated gene expression in multiple tissues	IGD; GSS	GTAs	GTEx; UKBB; EGA	[61]
		MR-JTI	https://github.com/gamazonlab/MR-JTI	Combine MR and TWAS	IGD; GSS	GTAs	GTEx; PsychENCODE; ENCODE; Roadmap; UKBB; GEUVADIS	[62]
		TisCoMM/ TisCoMM-S	https://github.com/XingjieShi/TisCoMM	Identify the tissue-specific gene–trait associations	IGD; GSS	GTAs	GTEx; NFBC1966; UKBB; 1000 Genomes Project	[63]
	Separating feature selection and aggregation	kTWAS	https://github.com/theLongLab/kTWAS	Integrate kernel-machine with TWAS	IGD	GTAs	1000 Genomes Project; GTEx; WTCCC; MSSNG	[65]
		VC-TWAS	https://github.com/yanglab-emory/VC_TWAS	Assume the effects of *cis*-eQTL SNPs on phenotype are random and use an equivalent kernel test for aggregation	IGD; GSS	GTAs	ROS/MAP; 1000 Genomes Project; MCADGS; IGAP; Synapse	[66]
		mkTWAS	https://github.com/theLongLab/mkTWAS	Replace GReX with marginal effect-based feature selection and kernel-based feature aggregation	IGD	GTAs	1000 Genomes Project; GTEx; WTCCC; DisGeNET	[71]
	Integrating other biological information	EpiXcan	https://bitbucket.org/roussoslab/epixcan	Integrate epigenome to estimate the functional importance of genetic variations on gene expression	GSS	GTAs	CMC; GTEx; STARNET; REMC; DLPFC; MSigDB; Harmonizome	[72]
		aSPU	https://github.com/ChongWu-Biostat/aSPU2	Conduct association test by integrating single set or multiple sets of weights	IGD; GSS	GTAs	WTCCC; lipids2010; lipids2013; HapMap; DGN; 1000 Genomes Project; NTR; YFS; METSIM	[74]
		BGW-TWAS	https://github.com/yanglab-emory/BGW-TWAS	Integrate both *cis*- and *trans*-eQTL weights to predict expression	IGD; IGWD	GTAs	ROS/MAP; MCADGS; RADC; 1000 Genomes Project; IGAP	[67]

(continued)

Table 1 (continued)

Classification		Method	Website	Description	Input	Output	Data	Ref.
		MOSTWAS	https://github.com/bhattacharya-a-bt/MOSTWAS	Prioritize distal-SNPs contributing to genes	IGD; IGWD	GTAs	ROS/MAP; TCGA; PsychENCODE; IGAP; PGC; UKBB; iCOGs	[76]
		METRO	https://github.com/zhengli09/METRO	Utilize expression data from multiple ancestries	IGD; GSS	GTAs	GENOA; GEUVADIS; UKBB; GLGC; AAAGC; MEDIA; 1000 Genomes project; GENCODE	[81]
		GBMI-TWAS	https://github.com/bhattacharya-a-bt/gbmi_twas	Conduct inverse-variance weighted meta-analysis on multi-ancestry datasets from GBMI	GSS	GTAs	1000 Genomes project; GBMI; UKBB	[85]
		moPMR-Egger	https://github.com/yuanzhongshang/PMR	Investigate causal influence on multiple traits based on LD between *cis*-SNPs of one gene	IGD; GSS	GTAs	GEUVADIS; GERA; UKBB	[95]
		TEns	https://zenodo.org/record/6845955	Identify eQTLs of transcribed enhancers	GSS	GTAs	Synapse; SuperAgerEpiMap; GTEx	[98]
		sTF-TWAS	https://github.com/theLongLab/TF-TWAS	Integrate prior knowledge of susceptible TF occupied elements to TWAS	IGD; GSS	GTAs	GTEx; BCAC; PGC; PRACTICAL; TRICL-ILCCO; LC3	[101]
		PoLoNet	https://github.com/Liye222/PoLoNet	Integrate biological network regression and proportional odds logistic model in TWAS	IGD	GETAs	GEUVADIS; UKBB; KEGG; GCTA; GENCODE	[103]
	Training with summary statistics	OTTERS	https://github.com/daiqile96/OTTERS	Adopt multiple PRS methods to estimate eQTL weights	GSS	GTAs	UKBB; GTEx; eQTLGen; 1000 Genomes Project; ROS/MAP; MSBB; UKBB	[106]
		Multi-variant TWAS	/	Utilize multi-variant TWAS models and larger eQTL summary statistic datasets to identify causal genes	ESS	GTAs	GTEx; eQTLGen; MetaBrain; YFS; GTEx; HapMap3	[110]
PWAS		/	https://github.com/nadavbra/pwas	Aggregate the signal of all variants jointly affecting a protein-coding gene and assess their overall impact on the protein's function	IGD	PGTAs	UKBB	[16]
CWAS		/	https://github.com/scbaca/cwas	Identify chromatin (cistrome) features that are genetically associated with a trait of interest	GSS	CTAs	GSE205885; GTEx; 1000 Genomes Project	[17]
RWAS		/	https://zenodo.org/record/6371678	Consider as-aQTLs' influence on disease risk	GSS	RTAs	TCGA; GTEx; ENCODE	[21]
			https://github.com/casalex/RWAS	Identify the characteristics of enhancers that contribute to genetic risk for disease	GSS	ETAs	GTEx; Roadmap; Hi-C data; PGC; UCSC	[22]
IWAS		UV-IWAS	https://github.com/kathalexknuts/MVIWAS	Exchange transcriptome by using imaging endophenotypes	IGWD; GSS	GTAs	ADNI; GTEx; IGAP; lipids2013	[18]
		MV-IWAS	https://github.com/kathalexknuts/MVIWAS	Exchange transcriptome by using imaging and other endophenotypes	IGWD; GSS	GTAs	ADNI; UKBB; ENIGMA; IGAP	[19]
		BrainXcan	https://github.com/hakyimlab/brainxcan	Leverage reference brain MRI data	IGD; GSS	BTAs	UKBB; HapMap3; PGC	[20]

Note: GWAS, genome-wide association study; TWAS, transcriptome-wide association study; PWAS, proteome-wide association study; CWAS, cistrome-wide association study; RWAS, regulome-wide association study; IWAS, imaging-wide association study; SNP, single-nucleotide polymorphism; BSLMM, Bayesian sparse linear mixed model; VCF, variant call format; DPR, Dirichlet process regression; GReX, genetically regulated expression; PRS, polygenic risk score; MRI, magnetic resonance imaging; IGD, individual-level genotype data; GSS, GWAS summary statistics; IGWD, individual-level GWAS data; eQTL, expression quantitative trait locus; as-aQTL, allele-specific accessibility quantitative trait locus; MR, Mendelian randomization; LD, linkage disequilibrium; TF, transcription factor; ESS, eQTL summary statistics; GTA, gene–trait association; GETA, gene/edge–trait association; PGTA, protein-coding gene–trait association; CTA, cistrome–trait association; RTA, regulome–trait association; ETA, enhancer–trait association; BTA, brain feature–trait association. The abbreviations of all databases are detailed in Table S3.

imputation) method [62] was proposed, utilizing tissue similarities in gene transcriptome and epigenome data as weights between different tissues. On the other hand, to ensure tissue-specific gene–trait associations, TisCoMM [63] was proposed to demonstrate the co-regulation of genetic variations across different tissues using a unified probabilistic model, achieving robust results with a low false-positive rate.

These improvements have significant implications for identifying the associations between genetic variants and diseases, overcoming dataset sample size limitations and furthering our understanding of underlying mechanisms and potential treatment targets.

Extension in separating feature selection and aggregation

Traditional two-stage TWAS methods first build an expression imputation model using genotype and reference transcriptomic data, and then conduct an association test between GReX and the phenotype. These tools assume that the effect of genetic variants on reference transcriptomic data and phenotype is linear. Consequently, GReX is derived as a linear combination of genetic variants, and the same linear model links GReX and phenotype. However, various factors can influence the power of TWAS methods, such as population differences between modeling and test datasets and the non-linear effect of variants on phenotype [64].

Recently, several new kernel-based methods have been developed to improve TWAS power. One such protocol is kernel-based TWAS (kTWAS) [65], which integrates the TWAS-like method and the sequence kernel association test (SKAT)-like method for feature selection and aggregation, separately. Extensive simulations have shown that kTWAS performs robustly and is more resistant to non-linear effects by combining the advantages of both methods. Especially, it can also detect interactions between numerous variants.

Another method to improve TWAS power is the variance-component TWAS (VC-TWAS) method [66], which uses a random effect model suitable for both individual-level and summary-level GWAS data, making it effective for analyzing continuous and dichotomous phenotypes. VC-TWAS estimates eQTL effect sizes using various methods like elastic net [14], nonparametric Bayesian DPR [50], and BVSR [67]. It then adopts a powerful framework analogous to SKAT [68–70] method to factorize the kernel matrix. Simulation tests have shown that VC-TWAS achieves greater power when the linearity assumption is relaxed.

Another promising extension of TWAS is the marginal + kernel TWAS (mkTWAS) [71], which utilizes marginal effect-based and kernel-based methods for feature selection and aggregation to replace GReX. Analyses of real and simulated data have demonstrated that mkTWAS significantly increases the power of identifying susceptibility genes by applying feature selection and aggregation into different statistical models.

These novel TWAS protocols set themselves apart from traditional two-stage methods by decoupling feature selection and aggregation, resulting in improved power. They offer a viable alternative to GReX in most scenarios.

Extension in integrating other biological information

In addition to three modifications mentioned above to the traditional TWAS approaches, integrating biological information such as pathway networks and epigenetic data is also a promising strategy for improving gene expression imputation accuracy.

EpiXcan [72], based on the assumption that SNPs within *cis*-regulatory elements (CREs) are more likely to influence gene expression regulation and exhibit functional relevance [73], has improved the accuracy of transcriptome imputation. By incorporating epigenetic data, including DNA methylation, histone modification, and chromatin accessibility, EpiXcan improves the prediction of the functional effect of genetic variations on gene expression levels.

Alternatively, another way to extend the traditional TWAS frameworks is to improve the association testing methodology between genes and traits. To achieve this goal, an adaptive sum of powered score (aSPU) [74] test was proposed based on generalized linear models (GLMs). This approach integrates multiple sets of weights obtained from diverse datasets to conduct association tests, thus it is capable of uncovering more novel gene–trait associations.

Numerous TWAS frameworks are limited to considering only *cis*-eQTLs, neglecting *trans*-eQTLs due to the computational burden. However, emerging research indicates that *trans*-eQTLs are crucial in explaining the expression of most genes [75]. To address this, Bayesian genome-wide TWAS (BGW-TWAS) [67] was proposed to enable the accounting of both *cis*-eQTL and *trans*-eQTL weights, utilizing summary statistics from standard eQTL analysis, allowing for effective computation through Bayesian variable selection regression.

To account for the effect of distal SNPs and other biomarkers underlying the SNP–gene associations, the Multi-Omic Strategies for TWAS (MOSTWAS) [76] method was developed. MOSTWAS integrates distal-SNPs and multiomic data, including epigenomes, microRNAs, and transcription factors (TFs), for transcriptome imputation. This approach enables more robust testing of gene–trait associations, achieving greater power in the process.

Allele frequency and LD patterns differ across populations with diverse genetic backgrounds, which can significantly impact eQTL analysis results and the identification of disease susceptibility genes [77–80]. To address this, METRO [81] was developed, it utilizes expression data from multiple ancestries to improve statistical power through a joint likelihood-based inference model. Additionally, METRO enables the inference of how distinct genetic ancestries' expression prediction models contribute to explaining gene–trait associations. Probing the ancestry-dependent transcriptome mechanisms driving gene–trait associations enables gaining deeper insights into the genetic underpinnings of complex traits.

Biobanks have been crucial in identifying relationships between genetic variants and human traits, but they have historically comprised primarily individuals of European descent [82,83]. This lack of diversity can influence genetic discoveries and limit our understanding of complex diseases in non-European populations. Different ancestry groups may have varying frequencies of disease-associated genetic variants, highlighting the need for more diverse biobanks to improve our understanding of complex diseases across all populations.

The Global Biobank Meta-analysis Initiative (GBMI) [84] aims to address this issue by fostering collaboration among over 20 biobank resources worldwide. Recently, a pipeline called GBMI-TWAS [85] has been introduced. This pipeline utilizes inverse-variance weighted meta-analysis on multi-

ancestry datasets derived from GBMI. It describes practical considerations related to ancestry and tissue specificities, meta-analytic techniques, and challenges encountered at each stage of the TWAS framework. For instance, a comparison between ancestry-aligned and ancestry-misaligned models revealed that utilizing training and testing samples of the same ancestry enhances the predictive capability of the model during expression prediction model construction. The effectiveness of TWAS is influenced by both the expression prediction model and expression heritability [10,86]. Consequently, this finding holds significance in selecting a reference expression panel for actual data analysis. Moreover, another crucial result highlights the importance of employing a suitable ancestry LD reference panel in the meta-analysis of GWAS summary statistics to ensure the accuracy of TWAS associations. By implementing ancestry-specific expression models that adhere to the ancestry-stratified TWAS framework, GBMI-TWAS demonstrates minimal test statistic inflation. These findings underscore the impact of ancestry-specific genetic architecture on gene–trait associations in disease studies and offer insights into the influence of genetic variants on disease susceptibility in populations of diverse ancestries.

Integrating MR to identify gene–trait associations has emerged as a growing trend. In contrast to the traditional two-stage TWAS approach, SMR method [41] utilizes the two-step least-squares (2SLS) method to detect gene–trait associations based on GWAS summary and eQTL data. Additionally, the method incorporates heterogeneity in dependent instruments (HEIDI) approach to identify potential pleiotropic effects of genes. Building upon this framework, TWMR approach [87] combines MR with TWAS to pinpoint causal gene–trait associations. This method utilizes multiple SNPs and gene expression data as instrumental variables and exposures, respectively, thereby mitigating biases arising from pleiotropic effects.

Due to the limited availability of paired genotype and single-cell transcriptome data, most TWAS approaches identify gene–trait associations using bulk transcriptome data. In contrast, MiXcan [88] has been developed as a framework for conducting cell type-level TWAS, enabling the identification of disease-associated cell types and genes. By leveraging prior information on scaled xCell [89] cell type enrichment scores from the training data, MiXcan enhances the estimation of cell type proportions through fitting mixture models for the expression levels of cell type signature genes. However, the current MiXcan framework only decomposes a tissue into two cell type components, necessitating an extension to create a comprehensive TWAS model encompassing all cell types.

SUMMIT [90] is a TWAS method designed to predict gene expression using eQTL summary-level data. Subsequently, associations between the predicted gene expression levels and traits are tested for the selected fitted expression prediction model. The combined p-value is then aggregated using the Cauchy combination test [91,92] to quantify the overall gene–trait associations. When applied to COVID-19 GWAS summary data, SUMMIT identifies risk genes for COVID-19 severity. However, the predominance of European ancestry in most eQTL data necessitates the corresponding eQTL summary data for extension to other ancestral populations.

Existing TWAS methods focus primarily on examining individual outcome traits separately in a univariate manner. However, given the shared genetic basis among numerous human complex traits, the potential for increased effectiveness of TWAS methods through multivariate analysis cannot be ignored [93,94].

To address this issue, moPMR-Egger [95] defines instrumental variables as possible LD between *cis*-SNPs of a gene and investigates its causal effect on multiple traits simultaneously. The significant aspect of moPMR-Egger is its capacity to evaluate and control for horizontal pleiotropic effects resulting from instruments, thereby increasing power while reducing false associations for TWAS methods.

Enhancers are crucial DNA regulatory elements that regulate the expression of target genes, and can be transcribed with expression levels as a signal for their activation [96,97]. Consequently, previous eQTL analysis may have missed essential information about enhancer-mediated genetic mechanisms. A recent population-scale analysis of enhancer expression in the human cerebral cortex [98], based on the integration of cell type-specific transcriptome and regulome data, revealed that enhancer eQTLs and genes account for a significant percentage of the heritability of neuropsychiatric traits. Enhancer eQTLs improved the functional interpretation of trait–locus associations in TWAS.

In addition to enhancers, TFs are essential regulatory elements for gene expression [99,100]. To fill this gap, sTF-TWAS [101,102] was developed to integrate prior knowledge of susceptible TF (sTF)-occupied *cis*-regulatory elements (STFCREs) with TWAS. This approach predicts gene expression using putative genetic variants located in STFCREs and provides evidence that genetic variants mediate the binding affinity of susceptible TFs to impact disease risk.

Current TWAS methods mainly focus on one gene at a time, while complex diseases typically result from biological networks involving multiple genes. From this perspective, PoLoNet [103] was developed to identify the association between specific networks and binary or ordinal categorical traits. PoLoNet is applied in the traditional two-stage TWAS framework. First, it obtains the SNP effect on genes using a distribution-robust nonparametric DPR model. Then, pointwise mutual information (PMI) is used to calculate all the between-node or edge correlations of predicted gene expression, followed by a proportional odds logistic model to perform association analysis on all nodes and edges. The biological network is selected from pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG), ultimately identifying the trait-related network nodes or edges.

Extension in training with summary statistics

In the first step of traditional TWAS frameworks, numerous methods estimate the effect size of SNPs using individual-level reference expression and genotype data. However, utilizing individual-level data in TWAS is more restricted compared to GWAS summary statistics, which have a broader range of applications. This is because GWAS summary statistics typically have larger sample size and are more readily accessible. TWAS is analogous to polygenic risk score (PRS) [104,105], which includes several methods using GWAS summary statistics instead of individual-level data. Therefore, it is feasible for TWAS to use summary-level data to indicate molecular mechanisms of genetic variations by integrating eQTL data and GWAS summary statistics.

For instance, OTTERS [106] is a TWAS framework that synthesizes various PRS methods [107–109] to estimate eQTL effect sizes from eQTL summary statistics, similar to

GWAS summary statistics. OTTERS can successfully identify potential risk genes that may be missed using limited individual-level datasets.

A previous study compared schizophrenia TWAS results based on datasets at two different levels [110]. The results showed that the model using summary statistics outperformed the model using individual-level data for many genes. For schizophrenia, several new susceptibility genes were identified that were not noticeable in other models. Thus, using GWAS summary statistics in TWAS reveals an effective way to elucidate the mechanisms of genetic variations and expands the utility of TWAS. Particularly, integrating GWAS summary statistics with eQTL data provides a feasible approach to understanding complex phenotypes.

Identification of susceptibility genes based on TWAS

Recently, a comprehensive compilation of TWAS discoveries named TWAS Atlas [111] was presented, offering high-quality human gene–trait associations gleaned from 200 peer-reviewed papers. Here we manually curated the publications of gene–disease association studies based on TWAS methods from the TWAS Atlas database, as well as the papers that were omitted in the TWAS Atlas.

First, we conducted a targeted exploration of the disease types and collected 347 entries from 168 papers by using the "browse" page of the TWAS Atlas website (https://ngdc.cncb.ac.cn/twas/browse). Subsequently, we performed an exhaustive literature search on the PubMed database using the keywords "transcriptome-wide association study" and "disease", meticulously evaluating results to supplement any gaps in the TWAS Atlas. In doing so, we procured 131 entries from 67 distinguished papers. For better classification of diseases, we converted all annotation properties of each disease from the Experimental Factor Ontology (EFO) [112] to MeSH (https://www.ncbi.nlm.nih.gov/mesh) terms (Table S1).

Various pie charts were created to demonstrate the distribution of different disease categories among the 478 entries (**Figure 3A**). The R package "wordcloud2" was utilized to construct wordcloud images, which were used for analyzing disease frequency (Figure 3B). The bar chart (Figure 3C) illustrates the number of publications per year, with 235 publications in total.

Out of the 22 disease categories identified, "Mental Disorders" was the most prevalent, accounting for 29% (137/478) of all cases. The top 3 diseases among the total 149 diseases were "Alzheimer's Disease" (35/235), "Schizophrenia" (31/235), and "Depressive Disorder" (19/235) (Figure 3B). TWAS has also been employed to investigate complex traits in numerous species including plants, animals, and humans, which highlights its powerful statistical performance in fully identifying the genetic mechanisms underlying these traits.

AD is a prevalent form of dementia in older individuals, initially manifesting as memory loss and progressing to severe executive and cognitive impairment [113]. Current therapies

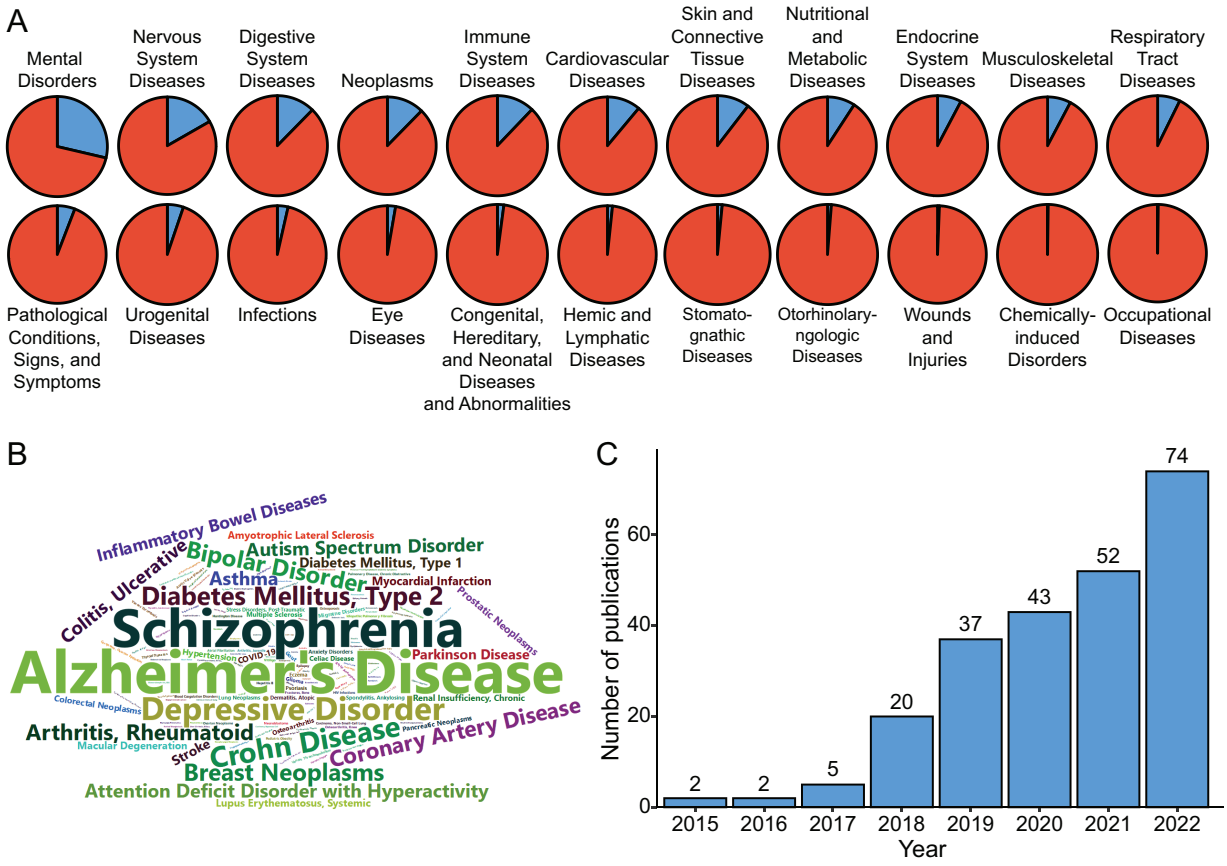

Figure 3 Statistical results of disease types and publications of identifying susceptibility genes based on TWAS methods

A. Pie chart showing the proportion of each MeSH disease category in the gene–disease association publications based on TWAS. Blue denotes the portion attributed to each disease, while red represents the all disease background. **B**. A word cloud graph showing all disease terms. Font size is determined by the frequency of each disease. **C**. Bar chart showing the number of publications that conduct gene–disease association analyses based on TWAS each year.

only manage symptoms without curing the disease [114], resulting in significant economic and societal burdens. Consequently, effective risk assessment procedures for AD are essential. Despite the development of numerous risk assessment strategies, inconsistencies across studies have been noted, possibly due to limitations like small sample size, selection bias, potential confusion, and reverse causality.

Various TWAS methods have been utilized in AD research. For instance, a previous study using FUSION identified four novel AD-associated genes (*MLH3*, *FNBP4*, *CEACAM19*, and *CLPTM1*) in brain tissue, adipose tissue, and whole blood [115]. Another study integrated expression and intron usage into the FUSION framework to investigate the association of alternative splicing with AD, uncovering 21 genes significantly linked to AD at the gene expression or intronic excision level [116]. Through the TWAS statistical framework UTMOST, 69 AD-associated genes were pinpointed by analyzing 17,008 cases and 37,154 controls [45]. Additionally, research employing PrediXcan to identify susceptibility genes associated with AD discovered 50 distinct genes and 126 tissue-specific correlations [117].

While numerous TWAS studies have explored AD, results often exhibit variability due to lack of a reliable gold standard to evaluate results, and not all identified genes and associations are necessarily causal with disease. Furthermore, heterogeneity among sample batches and factors like sample size, tissue origin, and ancestor population may influence final results. Hence, selecting appropriate analysis tools based on distinct sample data and analytical requirements is particularly crucial (Table S2).

EWAS

The concept of epigenome-wide association study (EWAS) was first introduced in 2011 [118] to explore epigenetic features linked to the susceptibility and progression of complex human diseases, similar to GWAS which connects the epigenome to phenotype through investigating modifications such as methylation across various disease states. Currently, EWAS analysis primarily centers on methylomes, often utilizing microarray-based methylation data due to the cost of whole-genome bisulfite sequencing (WGBS) [119].

EWAS analysis can be categorized into two main approaches [119]: differentially methylated positions (DMPs) and differentially methylated regions (DMRs). DMP analysis correlates methylation levels of individual CpG sites (using beta- or M values) with the phenotype, while DMR analysis links genomic regions containing multiple adjacent DMPs to the phenotype. Various tools exist for DMP analysis. For example, in case-control studies, the R package CHAMP [120] identifies significant differential CpG sites based on mean methylation levels from a beta-matrix after preprocessing raw microarray data through probe filtering, imputation, and normalization.

Additionally, the EWAS2.0 software [121] was developed for EWAS analysis within the "population epigenetic framework". Like GWAS analysis, EWAS2.0 can conduct epigenome-wide single-marker association studies, methylation haplotype association studies, and association meta-analyses using chi-square tests, *t*-tests, linear regression, and logistic regression. Before analysis, methylation data inputted into EWAS2.0 should undergo preprocessing and normalization similar to those in CHAMP [119].

In response to the improvement of EWAS and the growing availability of microarray-based data, EWASdb [122] curated 1319 EWASs associated with 302 phenotypes, including EWASs for single markers, KEGG pathways, and GO (Gene Ontology) categories. Recently, the EWAS Open Platform [123] was developed and comprises the EWAS Atlas [124], EWAS Data Hub [125], and EWAS Toolkit [123]. The EWAS Atlas [124] has curated 675,261 epigenome–phenotype associations from 1079 publications related to 798 traits. The EWAS Data Hub [125] provides normalized DNA methylation array data from 143,678 samples across 1210 tissues/cells associated with 658 diseases. The EWAS Open Platform [122] offers comprehensive resources, data, and tools for EWAS enrichment, annotation, and visualization.

PWAS

Following TWAS, PWAS [16] was introduced in 2020, based on the assumption that genetic variants within coding regions can influence phenotypes by affecting the biochemical functions of protein products. This novel framework identifies gene–trait associations mediated by protein functional changes. It estimates protein damage based on an individual's genotype to capture the impacts of variations that affect genes' coding regions, such as missense, nonsense, and frameshift. PWAS initially computes effect scores for non-synonymous variants within each gene's coding region, indicating their potential to disrupt the gene's protein product. These scores range from 0 (indicating complete loss of function) to 1 (suggesting no functional impact). Missense, nonsense, frameshift, in-frame insertion and deletion (indel), and canonical splice-site variants are recognized as influencing protein function. FIRM [34], a machine learning model that considers the rich proteomic context of each impacting variant, is employed to calculate the effect scores for missense variants, while nonsense, frameshift, and canonical splice-site variants are categorized as loss-of-function and assigned a score of 0. Regarding in-frame indels, the effect score is determined based on the number of altered amino acids.

By combining genotyping information from the cohort with these predicted effects, PWAS generates functional prediction of each gene, representing each protein-coding gene's functional effect score. Statistical analysis is then performed to determine whether a gene's effect score is associated with phenotype. A significant association between the effect scores of cases and controls, in the case of a binary phenotype, would indicate that the protein is more (or less) damaged in affected individuals. Simulation results show that PWAS performs exceptionally well in cases of recessive inheritance, and its discovery power is robust when applied to real datasets. Additionally, PWAS identified numerous known associations for most phenotypes. For instance, when the PWAS was applied to the UKBB cohort [126,127] to assess its applicability to different phenotypes, 5249 of the 500,000 participants contained a GWAS-significant non-synonymous variant in the gene's coding region.

One of PWAS's distinctive characteristics is its ability to model both dominant and recessive inheritance. While substantial evidence suggests that the commonly used additive model can capture most of human trait heritability, non-additive and epistatic effects are crucial in many phenotypes [128]. However, PWAS's reliance on complete individual-level data, including genotype and phenotype, could be a

drawback, preventing it from examining GWAS summary statistics alone, unlike other approaches. This dependency on raw data is due to the non-linear aggregation algorithm used to obtain gene impact scores from variation effect scores.

CWAS

GWAS and TWAS have identified numerous risk-associated genetic variants, but the mechanisms of non-coding genetic variants that influence complex traits and diseases remain to be fully explored. Many studies have shown that eQTLs influence gene expression by altering chromatin activity, which has led to increased research attention on the impact of risk-associated genetic variants on chromatin [129–135].

Analogous to eQTL, cQTL is a SNP associated with chromatin states, such as histone modifications, TF binding, and chromatin accessibility. Moreover, allelic imbalance in epigenomic data, defined as differences in the representation of heterozygous SNP alleles in sequencing reads, can be employed to identify variations that influence chromatin states [135–138].

Nevertheless, the application of cQTL and allelic imbalance to understand trait heritability is limited by two main factors: first, the lack of substantial reference epigenomes from relevant organs, and second, the absence of a uniform methodology for incorporating these data into GWAS. To overcome these limitations, a novel approach named CWAS [17] has been introduced. CWAS aims to identify genetic determinants of TF binding and histone modifications, and associates genetically predicted chromatin signals with traits based on GWAS summary statistics.

To implement CWAS, the growing number of chromatin immunoprecipitation followed by sequencing (ChIP-seq) datasets was leveraged to create and benchmark an approach that imputes genotypes from ChIP-seq data with high accuracy. As an extension of the conventional TWAS pipeline, CWAS takes into account allele-specific information and chromatin phenotype.

CWAS findings demonstrate the efficacy of an integrative cistrome approach in identifying genetic determinants of gene regulation. Furthermore, CWAS is shown to complement other methods, such as TWAS/eQTL-based methods, which may fail to detect associations involving genes with intricate regulation and context-dependent expression. In the context of prostate cancer, CWAS identifies key regulatory elements and androgen receptor-binding sites, explaining the associations at 52 out of 98 known prostate cancer risk loci. Additionally, CWAS discovered 17 novel risk loci, emphasizing the power of this approach in uncovering previously unidentified genetic determinants of disease risk.

RWAS

Researchers have developed an innovative work, known as RWAS [21], which utilizes cancer ATAC-seq datasets from The Cancer Genome Atlas (TCGA) to identify germline as-aQTLs and links them to potential risk mechanisms. ATAC-seq is a sequencing-based technique that enables the identification of accessible regions within the genome, often correlated with TF-binding sites [135,139]. RWAS has successfully been applied to seven cancer GWAS datasets, revealing numerous cancer-specific as-aQTLs that demonstrate a higher enrichment for cancer risk heritability compared to other functional annotations. Moreover, the majority of these cancer-specific as-aQTLs have been observed to modulate TF patterns, thereby influencing differential TF binding and gene expression. RWAS has identified genetically linked accessible peaks in over 70% of recognized breast and prostate loci and has unearthed novel risk loci across all cancer types analyzed. Through the integration of as-aQTL discovery, motif analysis, and RWAS, potential susceptibility regulatory elements and their likely upstream regulators have been identified.

Another concurrent study also referred to as RWAS [22], aims to pinpoint particular enhancer attributes potentially contributing to genetic disease risk. The RWAS methodology comprises three core stages. First, genotyped SNPs are mapped to regulatory features like enhancers specific to distinct cell types or tissues. Subsequently, the association between each regulatory feature and a trait of interest is examined. Finally, enhancer-set enrichment analyses are performed to disclose quantitative or categorical features of regulatory elements associated with the trait. These procedures denote a novel application of MAGMA [140], originally designed for gene-centric GWAS analysis.

IWAS

IWAS was initially proposed in 2017 [18] to investigate the relationship between genes and complex diseases via imaging endophenotypes. The protocol of IWAS is analogous to TWAS. The difference is that IWAS substitutes gene expression with imaging endophenotypes in the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Analogous to TWAS, the elastic net was employed to train the weights of genetic variations for each imaging endophenotype and subsequently compared with gene expression-based weights from PrediXcan. Distinct tests were conducted utilizing sum of powered score (SPU) (1), SPU (2), and aSPU for each set of weights, respectively, with numerous significant associations using imaging-based weights while non-significant associations using expression-based weights. Moreover, a test integrating multiple weights named doubly aSPU (daSPU) test was applied to imaging endophenotypes, unveiling a substantial number of significant genes. Conversely, a small number of significant genes overlapped with GWAS were detected when applied to the International Genomics of Alzheimer's Project (IGAP) GWAS summary statistics. These results demonstrate that univariate IWAS (UV-IWAS) has higher power in specific complex diseases (*e.g.*, AD) and daSPU-based IWAS has a better interpretation through the imaging endophenotypes.

Subsequently, multivariate IWAS (MV-IWAS) [19] was introduced as an extension of UV-IWAS to reduce type-I errors originating from horizontal pleiotropy, with the residual pleiotropic effects addressed via Egger estimators denoted as MV-IWAS-Egger. Under the scenario of directional pleiotropy, MV-IWAS-Egger controls the type-I error well and has maximum power across all multivariate models. Notably, causal associations for the left hippocampus and right inferior temporal cortex volumes concerning AD were pinpointed when both tests were applied to the ADNI endophenotypes, a conclusion supported by extensive literature. Furthermore, MV-IWAS identified numerous new causal brain phenotypes based on UKBB data which were missed by UV-IWAS, and also implicated many potential false-positive UV-IWAS results.

Recently, a preprint named BrainXcan [20] was published, which utilizes GWAS data to test the associations between genetic predictors of brain magnetic resonance imaging

(MRI)-derived features and complex traits based on three modules as an extension of PrediXcan. First, the "Prediction weight training" module computes image-derived phenotype (IDP) prediction weights, IDP QTLs, and "reference LD" information across the UKBB and Psychiatric Genomics Consortium (PGC) data. Then, the "Association" module will generate regression estimated coefficients between predicted IDPs and traits. Last, the "Mendelian randomization" module examines the direction of causal flow, *i.e.*, whether phenotype influences disease or disease affects phenotype. BrainXcan identified many risk phenotypes supporting the disconnectivity hypothesis of schizophrenia.

In conclusion, IWAS achieved enhanced power and interpretability through the integration of imaging datasets. However, limitations persist, including the size of the training cohort constraining the predictor's efficacy. Further investigation is required to validate the causal relationship, and the enhancements to the linear regression model are necessary.

mGWAS and MWAS

Metabolomics-based genome-wide association study (mGWAS) links thousands of metabolites and millions of genetic variants to understand the genetic regulation of metabolites in complex phenotypes, crucial for unraveling their genetic underpinnings [141,142]. mGWAS involves examining genetic variants across the entire genome to identify associations with metabolite-level variations, typically utilizing large-scale datasets containing genotypic and metabolomic data. Predominantly conducted in populations of European ancestry, these studies predominantly focus on blood and urine metabolites, employing proton nuclear magnetic resonance (NMR) spectroscopy or mass spectrometry (MS) for quantification [143]. By elucidating the genetic basis of metabolite traits, mGWAS offers insights into metabolic pathways and their regulation, aiding in identifying disease biomarkers, understanding metabolic dysregulation, and discovering therapeutic targets [144].

Metagenome-wide association studies (MWAS) are primarily modeled on the GWAS framework, aiming to identify genes in the human metagenome associated with phenotypes, often diseases. In MWAS, gene relative abundance in a metagenome is correlated with a disease of interest, typically after grouping genes into strain-level clusters termed metagenomic linkage groups, metagenomic clusters, or metagenomic species to reduce data dimensionality [145]. MWAS entails analyzing the genetic composition of microbial communities (the metagenome) through sequencing microbial DNA from samples like fecal, oral, or skin swabs, followed by bioinformatic analysis to characterize microbial taxa or functional genes [146]. By identifying microbial taxa or functional genes associated with specific phenotypes or diseases, MWAS sheds light on host-microbiome interactions, microbial community dynamics, and the impact of environmental factors on the microbiota [147,148].

In summary, mGWAS and MWAS are robust methodologies enabling researchers to investigate the genetic basis of metabolite concentrations and microbial community composition, respectively. These approaches offer crucial insights into the intricate interplay between genetic and environmental factors in health and disease, with implications for personalized medicine, biomarker discovery, and therapeutic interventions.

Discussion

Multiome-wide association study methods offer an effective strategy for integrating diverse multiomic data, addressing the challenge of interpreting GWAS significance signals within non-coding regions to some extent and enhancing the statistical power of genome–phenotype associations. These studies aim to explore the impact of various factors, including genomes, epigenomes, transcriptomes, proteomes, metabolome, and microbiomes, on phenotypes, and their interactions within biological systems. Compared to single-omics research, multiomics integration provides a more comprehensive approach, considering the complexity and diversity of human traits and diseases.

Recent research shows a tendency toward joint analysis by integrating GWAS with multiome-mediated methods. For instance, an integrated analysis of GWAS and TWAS on patients with drug-induced liver injury successfully identified significant clinical risk predictors, aiding in the susceptibility assessment to liver injury due to amoxicillin-clavulanate [149]. Another study conducted an integrated analysis of GWAS and TWAS on type 2 diabetes mellitus and schizophrenia to explore shared pathways between the two diseases [150]. Similarly, an integrated analysis of TWAS, PWAS, Bayesian colocalization, and MR on Parkinson's disease prioritized two susceptibility genes based on their effects on brain proteins and transcriptome levels [151]. These studies underscore the potential of integrated analyses of multiome-mediated methods in understanding human complex diseases.

The findings of research through multiome-mediated methods have proven to be significant reference values in investigating complex traits and diseases. However, several limitations persist in multiome-mediated methods. Firstly, while many significant signals identified by these approaches are associated with diseases, they may not necessarily be causal. Moreover, these methods typically rely on paired public genotypic and multiomic data as reference panels, which are still scarce in large population cohorts, often restricted by privacy concerns, and particularly lacking in most tissues, especially among populations of non-European ancestries. Additionally, comprehensive data resources for multiome-mediated methods remain relatively limited. Currently, only three databases for TWAS (TWAS-hub [152], webTWAS [86], and TWAS Atlas [111]), and four databases for EWAS (EWASdb [122], EWAS Atlas [124], EWAS Data Hub [125], and EWAS Open Platform [123]) are available. Lastly, the rapid advancement of single-cell sequencing has underscored the importance of analyzing disease mechanisms from a single-cell perspective, emphasizing the need to integrate single-cell data into multiome-wide association study methods.

In summary, future studies will be required to distinguish causal signals of complex human diseases or traits and elucidate single cell-specific genetic mechanisms. With the release of large datasets such as UKBB, multiome-wide association study methods are poised to experience an influx of new method extensions.

CRediT author statement

Mengting Shao: Data curation, Visualization, Writing – original draft, Writing – review & editing. **Kaiyang Chen:** Writing

– original draft. **Shuting Zhang:** Writing – review & editing. **Min Tian:** Writing – original draft. **Yan Shen:** Writing – review & editing. **Chen Cao:** Conceptualization, Supervision, Writing – review & editing, Project administration. **Ning Gu:** Conceptualization, Supervision. All authors have read and approved the final manuscript.

Supplementary material

Supplementary material is available at *Genomics, Proteomics & Bioinformatics* online (https://doi.org/10.1093/gpbjnl/qzae077).

Competing interests

The authors have declared no competing interests.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 62102068, 62231013, 61821002, 81971750, and 81701833) and the National Key R&D Program of China (Grant No. 2017YFA0104300). We also thank Dr. Quan Long (University of Calgary, Canada) for putting forward the name of the multiome-wide association study.

ORCID

0009-0001-7443-7542 (Mengting Shao)
0009-0001-9110-7414 (Kaiyang Chen)
0009-0001-2515-4677 (Shuting Zhang)
0009-0008-3332-2723 (Min Tian)
0009-0006-4350-5932 (Yan Shen)
0000-0001-5343-808X (Chen Cao)
0000-0003-0047-337X (Ning Gu)

References

[1] Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. Science 2005;308:385–9.
[2] Sud A, Kinnersley B, Houlston RS. Genome-wide association studies of cancer: current insights and future perspectives. Nat Rev Cancer 2017;17:692–704.
[3] Loos RJF. 15 years of genome-wide association studies and no signs of slowing down. Nat Commun 2020;11:5900.
[4] Zhang H, Ahearn TU, Lecarpentier J, Barnes D, Beesley J, Qi G, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. Nat Genet 2020;52:572–81.
[5] Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nat Rev Genet 2019;20:467–84.
[6] Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, et al. Association analysis identifies 65 new breast cancer risk loci. Nature 2017;551:92–4.
[7] Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS discovery: biology, function, and translation. Am J Hum Genet 2017;101:5–22.
[8] Conti DV, Darst BF, Moss LC, Saunders EJ, Sheng X, Chou A, et al. *Trans*-ancestry genome-wide association meta-analysis of prostate cancer identifies new susceptibility loci and informs genetic risk prediction. Nat Genet 2021;53:65–75.
[9] McKay JD, Hung RJ, Han Y, Zong X, Carreras-Torres R, Christiani DC, et al. Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. Nat Genet 2017;49:1126–32.
[10] Cao C, Ding B, Li Q, Kwok D, Wu J, Long Q. Power analysis of transcriptome-wide association study: implications for practical protocol choice. PLoS Genet 2021;17:e1009405.
[11] Gallagher MD, Chen-Plotkin AS. The Post-GWAS Era: from association to function. Am J Hum Genet 2018;102:717–30.
[12] Christoforou A, Dondrup M, Mattingsdal M, Mattheisen M, Giddaluru S, Nöthen MM, et al. Linkage-disequilibrium-based binning affects the interpretation of GWASs. Am J Hum Genet 2012;90:727–33.
[13] Fachal L, Aschard H, Beesley J, Barnes DR, Allen J, Kar S, et al. Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes. Nat Genet 2020;52:56–73.
[14] Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. Nat Genet 2015;47:1091–8.
[15] Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet 2016;48:245–52.
[16] Brandes N, Linial N, Linial M. PWAS: proteome-wide association study-linking genes and phenotypes by functional variation in proteins. Genome Biol 2020;21:173.
[17] Baca SC, Singler C, Zacharia S, Seo JH, Morova T, Hach F, et al. Genetic determinants of chromatin reveal prostate cancer risk mediated by context-dependent gene regulation. Nat Genet 2022;54:1364–75.
[18] Xu Z, Wu C, Pan W. Imaging-wide association study: integrating imaging endophenotypes in GWAS. Neuroimage 2017;159:159–69.
[19] Knutson KA, Deng Y, Pan W. Implicating causal brain imaging endophenotypes in Alzheimer's disease using multivariable IWAS and GWAS summary data. Neuroimage 2020;223:117347.
[20] Liang Y, Melia O, Caroll TJ, Brettin T, Brown A, Im HK. BrainXcan identifies brain features associated with behavioral and psychiatric traits using large scale genetic and imaging data. medRxiv 2022;21258159.
[21] Grishin D, Gusev A. Allelic imbalance of chromatin accessibility in cancer identifies candidate causal risk variants and their mechanisms. Nat Genet 2022;54:837–49.
[22] Casella AM, Colantuoni C, Ament SA. Identifying enhancer properties associated with genetic risk for complex traits using regulome-wide association studies. PLoS Comput Biol 2022;18:e1010430.
[23] Zhong J, Jermusyk A, Wu L, Hoskins JW, Collins I, Mocci E, et al. A transcriptome-wide association study identifies novel candidate susceptibility genes for pancreatic cancer. J Natl Cancer Inst 2020;112:1003–12.
[24] Dall'Aglio L, Lewis CM, Pain O. Delineating the genetic component of gene expression in major depression. Biol Psychiatry 2021;89:627–36.
[25] Pain O, Pocklington AJ, Holmans PA, Bray NJ, O'Brien HE, Hall LS, et al. Novel insight into the etiology of autism spectrum disorder gained by integrating expression data with genome-wide association statistics. Biol Psychiatry 2019;86:265–73.
[26] Pividori M, Schoettler N, Nicolae DL, Ober C, Im HK. Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. Lancet Respir Med 2019;7:509–22.
[27] Liu D, Zhu J, Zhou D, Nikas EG, Mitanis NT, Sun Y, et al. A transcriptome-wide association study identifies novel candidate susceptibility genes for prostate cancer risk. Int J Cancer 2022;150:80–90.

[28] Thériault S, Gaudreault N, Lamontagne M, Rosa M, Boulanger MC, Messika-Zeitoun D, et al. A transcriptome-wide association study identifies *PALMD* as a susceptibility gene for calcific aortic valve stenosis. Nat Commun 2018;9:988.
[29] Ratnapriya R, Sosina OA, Starostik MR, Kwicklis M, Kapphahn RJ, Fritsche LG, et al. Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. Nat Genet 2019;51:606–10.
[30] Gusev A, Mancuso N, Won H, Kousi M, Finucane HK, Reshef Y, et al. Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. Nat Genet 2018;50:538–48.
[31] Jaffe AE, Hoeppner DJ, Saito T, Blanpain L, Ukaigwe J, Burke EE, et al. Profiling gene expression in the human dentate gyrus granule cell layer reveals insights into schizophrenia and its genetic risk. Nat Neurosci 2020;23:510–9.
[32] Fabbri C, Pain O, Hagenaars SP, Lewis CM, Serretti A. Transcriptome-wide association study of treatment-resistant depression and depression subtypes for drug repurposing. Neuropsychopharmacology 2021;46:1821–9.
[33] Gaspar HA, Gerring Z, Hübel C, Middeldorp CM, Derks EM, Breen G. Using genetic drug-target networks to develop new drug hypotheses for major depressive disorder. Transl Psychiatry 2019;9:117.
[34] Brandes N, Linial N, Linial M. Quantifying gene selection in cancer through protein functional alteration bias. Nucleic Acids Res 2019;47:6642–55.
[35] Shen L, Thompson PM. Brain imaging genomics: integrated analysis and machine learning. Proc IEEE Inst Electr Electron Eng 2020;108:125–62.
[36] Gusev A, Spisak S, Fay AP, Carol H, Vavra KC, Signoretti S, et al. Allelic imbalance reveals widespread germline-somatic regulatory differences and prioritizes risk loci in renal cell carcinoma. bioRxiv 2019;631150.
[37] de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput Biol 2015;11:e1004219.
[38] Hamza TH, Zabetian CP, Tenesa A, Laederach A, Montimurro J, Yearout D, et al. Common genetic variation in the *HLA* region is associated with late-onset sporadic Parkinson's disease. Nat Genet 2010;42:781–5.
[39] Delaneau O, Zazhytska M, Borel C, Giannuzzi G, Rey G, Howald C, et al. Chromatin three-dimensional interactions mediate genetic effects on gene expression. Science 2019;364:eaat8266.
[40] Ota M, Nagafuchi Y, Hatano H, Ishigaki K, Terao C, Takeshima Y, et al. Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. Cell 2021;184:3006–21.e17.
[41] Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet 2016;48:481–7.
[42] Wang T, Qiao J, Zhang S, Wei Y, Zeng P. Simultaneous test and estimation of total genetic effect in eQTL integrative analysis through mixed models. Brief Bioinform 2022;23:bbac038.
[43] Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013;45:580–5.
[44] Li B, Veturi Y, Verma A, Bradford Y, Daar ES, Gulick RM, et al. Tissue specificity-aware TWAS (TSA-TWAS) framework identifies novel associations with metabolic, immunologic, and virologic traits in HIV-positive adults. PLoS Genet 2021;17:e1009464.
[45] Hu Y, Li M, Lu Q, Weng H, Wang J, Zekavat SM, et al. A statistical framework for cross-tissue transcriptome-wide association analysis. Nat Genet 2019;51:568–76.
[46] Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Series B Stat Methodol 2005;67:301–20.
[47] Tibshirani R. Regression shrinkage and selection via the Lasso. J R Stat Soc Series B Stat Methodol 1996;58:267–88.
[48] Zhou X, Carbonetto P, Stephens M. Polygenic modeling with bayesian sparse linear mixed models. PLoS Genet 2013;9:e1003264.
[49] Zeng P, Dai J, Jin S, Zhou X. Aggregating multiple expression prediction models improves the power of transcriptome-wide association studies. Hum Mol Genet 2021;30:939–51.
[50] Nagpal S, Meng X, Epstein MP, Tsoi LC, Patrick M, Gibson G, et al. TIGAR: an improved Bayesian tool for transcriptomic data imputation enhances gene mapping of complex traits. Am J Hum Genet 2019;105:258–66.
[51] Parrish RL, Gibson GC, Epstein MP, Yang J. TIGAR-V2: efficient TWAS tool with nonparametric Bayesian eQTL weights of 49 tissue types from GTEx V8. HGG Adv 2022;3:100068.
[52] Zeng P, Zhou X. Non-parametric genetic prediction of complex traits with latent Dirichlet process regression models. Nat Commun 2017;8:456.
[53] Pierce BL, Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. Am J Epidemiol 2013;178:1177–84.
[54] 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.
[55] Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, et al. Opportunities and challenges for transcriptome-wide association studies. Nat Genet 2019;51:592–9.
[56] Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Hum Mol Genet 2014;23:R89–98.
[57] Yuan Z, Zhu H, Zeng P, Yang S, Sun S, Yang C, et al. Testing and controlling for horizontal pleiotropy with probabilistic Mendelian randomization in transcriptome-wide association studies. Nat Commun 2020;11:3861.
[58] Mancuso N, Freund MK, Johnson R, Shi H, Kichaev G, Gusev A, et al. Probabilistic fine-mapping of transcriptome-wide association studies. Nat Genet 2019;51:675–82.
[59] Lu Z, Gopalan S, Yuan D, Conti DV, Pasaniuc B, Gusev A, et al. Multi-ancestry fine-mapping improves precision to identify causal genes in transcriptome-wide association studies. Am J Hum Genet 2022;109:1388–404.
[60] Shi H, Burch KS, Johnson R, Freund MK, Kichaev G, Mancuso N, et al. Localizing components of shared transethnic genetic architecture of complex traits from GWAS summary data. Am J Hum Genet 2020;106:805–17.
[61] Barbeira AN, Pividori M, Zheng J, Wheeler HE, Nicolae DL, Im HK. Integrating predicted transcriptome from multiple tissues improves association detection. PLoS Genet 2019;15:e1007889.
[62] Zhou D, Jiang Y, Zhong X, Cox NJ, Liu C, Gamazon ER. A unified framework for joint-tissue transcriptome-wide association and Mendelian randomization analysis. Nat Genet 2020;52:1239–46.
[63] Shi X, Chai X, Yang Y, Cheng Q, Jiao Y, Chen H, et al. A tissue-specific collaborative mixed model for jointly analyzing multiple tissues in transcriptome-wide association studies. Nucleic Acids Res 2020;48:e109.
[64] Keys KL, Mak ACY, White MJ, Eckalbar WL, Dahl AW, Mefford J, et al. On the cross-population generalizability of gene expression prediction models. PLoS Genet 2020;16:e1008927.
[65] Cao C, Kwok D, Edie S, Li Q, Ding B, Kossinna P, et al. kTWAS: integrating kernel machine with transcriptome-wide association studies improves statistical power and reveals novel genes. Brief Bioinform 2021;22:bbaa270.
[66] Tang S, Buchman AS, De Jager PL, Bennett DA, Epstein MP, Yang J. Novel Variance-Component TWAS method for studying complex human diseases with applications to Alzheimer's dementia. PLoS Genet 2021;17:e1009482.

[67] Luningham JM, Chen J, Tang S, De Jager PL, Bennett DA, Buchman AS, et al. Bayesian genome-wide TWAS method to leverage both *cis*- and *trans*-eQTL information through summary statistics. Am J Hum Genet 2020;107:714–26.
[68] Wu MC, Kraft P, Epstein MP, Taylor DM, Chanock SJ, Hunter DJ, et al. Powerful SNP-set analysis for case-control genome-wide association studies. Am J Hum Genet 2010;86:929–42.
[69] Li Q, Perera D, Cao C, He J, Bian J, Chen X, et al. Interaction-integrated linear mixed model reveals 3D-genetic basis underlying Autism. Genomics 2023;115:110575.
[70] Lee S, Teslovich TM, Boehnke M, Lin X. General framework for meta-analysis of rare variants in sequencing association studies. Am J Hum Genet 2013;93:42–53.
[71] Cao C, Kossinna P, Kwok D, Li Q, He J, Su L, et al. Disentangling genetic feature selection and aggregation in transcriptome-wide association studies. Genetics 2022; 220:iyab216.
[72] Zhang W, Voloudakis G, Rajagopal VM, Readhead B, Dudley JT, Schadt EE, et al. Integrative transcriptome imputation reveals tissue-specific and shared biological mechanisms mediating susceptibility to complex traits. Nat Commun 2019; 10:3834.
[73] Gaffney DJ, Veyrieras JB, Degner JF, Pique-Regi R, Pai AA, Crawford GE, et al. Dissecting the regulatory architecture of gene expression QTLs. Genome Biol 2012;13:R7.
[74] Xu Z, Wu C, Wei P, Pan W. A powerful framework for integrating eQTL and GWAS summary data. Genetics 2017;207:893–902.
[75] Lloyd-Jones LR, Holloway A, McRae A, Yang J, Small K, Zhao J, et al. The genetic architecture of gene expression in peripheral blood. Am J Hum Genet 2017;100:371.
[76] Bhattacharya A, Li Y, Love MI. MOSTWAS: multi-omic strategies for transcriptome-wide association studies. PLoS Genet 2021;17:e1009398.
[77] Lappalainen T, Sammeth M, Friedländer MR, 't Hoen PAC, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature 2013;501:506–11.
[78] Mogil LS, Andaleon A, Badalamenti A, Dickinson SP, Guo X, Rotter JI, et al. Genetic architecture of gene expression traits across diverse populations. PLoS Genet 2018;14:e1007586.
[79] Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR, et al. Genetic analyses of diverse populations improves discovery for complex traits. Nature 2019;570:514–8.
[80] Shang L, Smith JA, Zhao W, Kho M, Turner ST, Mosley TH, et al. Genetic architecture of gene expression in European and African Americans: an eQTL mapping study in GENOA. Am J Hum Genet 2020;106:496–512.
[81] Li Z, Zhao W, Shang L, Mosley TH, Kardia SLR, Smith JA, et al. METR–O: multi-ancestry transcriptome-wide association studies for powerful gene–trait association detection. Am J Hum Genet 2022;109:783–801.
[82] Swede H, Stone CL, Norwood AR. National population-based biobanks for genetic research. Genet Med 2007;9:141–9.
[83] Shifman S, Kuypers J, Kokoris M, Yakir B, Darvasi A. Linkage disequilibrium patterns of the human genome across populations. Hum Mol Genet 2003;12:771–6.
[84] Zhou W, Kanai M, Wu KH, Rasheed H, Tsuo K, Hirbo JB, et al. Global biobank meta-analysis initiative: powering genetic discovery across human disease. Cell Genom 2022;2:100192.
[85] Bhattacharya A, Hirbo JB, Zhou D, Zhou W, Zheng J, Kanai M, et al. Best practices for multi-ancestry, meta-analytic transcriptome-wide association studies: lessons from the global biobank meta-analysis initiative. Cell Genom 2022;2:100180.
[86] Cao C, Wang J, Kwok D, Cui F, Zhang Z, Zhao D, et al. webTWAS: a resource for disease candidate susceptibility genes identified by transcriptome-wide association study. Nucleic Acids Res 2022;50:D1123–30.
[87] Porcu E, Rüeger S, Lepik K, Santoni FA, Reymond A, Kutalik Z. Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. Nat Commun 2019;10:3300.
[88] Song X, Ji J, Rothstein JH, Alexeeff SE, Sakoda LC, Sistig A, et al. MiXcan: a framework for cell-type-aware transcriptome-wide association studies with an application to breast cancer. Nat Commun 2023;14:377.
[89] Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol 2017;18:220.
[90] Zhang Z, Bae YE, Bradley JR, Wu L, Wu C. SUMMIT: an integrative approach for better transcriptomic data imputation improves causal gene identification. Nat Commun 2022; 13:6336.
[91] Liu Y, Chen S, Li Z, Morrison AC, Boerwinkle E, Lin X. ACAT: a fast and powerful p value combination method for rare-variant analysis in sequencing studies. Am J Hum Genet 2019; 104:410–21.
[92] Liu Y, Xie J. Cauchy combination test: a powerful test with analytic *p*-value calculation under arbitrary dependency structures. J Am Stat Assoc 2020;115:393–402.
[93] Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. Nat Rev Genet 2013;14:483–95.
[94] Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet 2015;47:1236–41.
[95] Liu L, Zeng P, Xue F, Yuan Z, Zhou X. Multi-trait transcriptome-wide association studies with probabilistic Mendelian randomization. Am J Hum Genet 2021;108:240–56.
[96] Schoenfelder S, Fraser P. Long-range enhancer-promoter contacts in gene expression control. Nat Rev Genet 2019; 20:437–55.
[97] Long HK, Prescott SL, Wysocka J. Ever-changing landscapes: transcriptional enhancers in development and evolution. Cell 2016;167:1170–87.
[98] Dong P, Hoffman GE, Apontes P, Bendl J, Rahman S, Fernando MB, et al. Population-level variation in enhancer expression identifies disease mechanisms in the human brain. Nat Genet 2022;54:1493–503.
[99] Deplancke B, Alpern D, Gardeux V. The genetics of transcription factor DNA binding variation. Cell 2016;166:538–54.
[100] Choudhuri A, Trompouki E, Abraham BJ, Colli LM, Kock KH, Mallard W, et al. Common variants in signaling transcription-factor-binding sites drive phenotypic variability in red blood cell traits. Nat Genet 2020;52:1333–45.
[101] He J, Wen W, Beeghly A, Chen Z, Cao C, Shu XO, et al. Integrating transcription factor occupancy with transcriptome-wide association analysis identifies susceptibility genes in human cancers. Nat Commun 2022;13:7118.
[102] Guo X, He J, Beeghly-Fadiel A, Chen Z, Cao C, Shu X, et al. Integrating prior knowledge of transcription factor occupied elements with transcriptome-wide association analysis identifies 153 breast cancer susceptibility genes. Cancer Res 2022; 82:5900.
[103] Zhang L, Ju T, Jin X, Ji J, Han J, Zhou X, et al. Network regression analysis for binary and ordinal categorical phenotypes in transcriptome-wide association studies. Genetics 2022; 222:iyac153.
[104] Yang S, Zhou X. Accurate and scalable construction of polygenic scores in large biobank data sets. Am J Hum Genet 2020; 106:679–93.
[105] Yang S, Zhou X. PGS-server: accuracy, robustness and transferability of polygenic score methods for biobank scale studies. Brief Bioinform 2022;23:bbac039.
[106] Dai Q, Zhou G, Zhao H, Võsa U, Franke L, Battle A, et al. OTTERS: a powerful TWAS framework leveraging summary-level reference data. Nat Commun 2023;14:1271.
[107] Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to

risk of schizophrenia and bipolar disorder. Nature 2009; 460:748–52.
[108] Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat Commun 2019;10:1776.
[109] Zhou G, Zhao H. A fast and robust Bayesian nonparametric method for prediction of complex traits using summary statistics. PLoS Genet 2021;17:e1009697.
[110] Pain O, Gerring Z, Derks E, Wray NR, Gusev A, Al-Chalabi A. Polygenic prediction of molecular traits using large-scale meta-analysis summary statistics. bioRxiv 2022;517213.
[111] Lu M, Zhang Y, Yang F, Mai J, Gao Q, Xu X, et al. TWAS Atlas: a curated knowledgebase of transcriptome-wide association studies. Nucleic Acids Res 2023;51:D1179–87.
[112] Malone J, Holloway E, Adamusiak T, Kapushesky M, Zheng J, Kolesnikov N, et al. Modeling sample variables with an Experimental Factor Ontology. Bioinformatics 2010; 26:1112–8.
[113] Kukull WA, Bowen JD. Dementia epidemiology. Med Clin North Am 2002;86:573–90.
[114] Yiannopoulou KG, Papageorgiou SG. Current and future treatments in Alzheimer disease: an update. J Cent Nerv Syst Dis 2020;12:1179573520907397.
[115] Hao S, Wang R, Zhang Y, Zhan H. Prediction of Alzheimer's disease-associated genes by integration of GWAS summary data and expression data. Front Genet 2018;9:653.
[116] Raj T, Li YI, Wong G, Humphrey J, Wang M, Ramdhani S, et al. Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. Nat Genet 2018;50:1584–92.
[117] Gerring ZF, Lupton MK, Edey D, Gamazon ER, Derks EM. An analysis of genetically regulated gene expression across multiple tissues implicates novel gene candidates in Alzheimer's disease. Alzheimers Res Ther 2020;12:43.
[118] Wei S, Tao J, Xu J, Chen X, Wang Z, Zhang N, et al. Ten years of EWAS. Adv Sci (Weinh) 2021;8:e2100727.
[119] Campagna MP, Xavier A, Lechner-Scott J, Maltby V, Scott RJ, Butzkueven H, et al. Epigenome-wide association studies: current knowledge, strategies and recommendations. Clin Epigenetics 2021;13:214.
[120] Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, et al. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. Bioinformatics 2017;33:3982–4.
[121] Xu J, Zhao L, Liu D, Hu S, Song X, Li J, et al. EWAS: epigenome-wide association study software 2.0. Bioinformatics 2018;34:2657–8.
[122] Liu D, Zhao L, Wang Z, Zhou X, Fan X, Li Y, et al. EWASdb: epigenome-wide association study database. Nucleic Acids Res 2019;47:D989–93.
[123] Xiong Z, Yang F, Li M, Ma Y, Zhao W, Wang G, et al. EWAS Open Platform: integrated data, knowledge and toolkit for epigenome-wide association study. Nucleic Acids Res 2022; 50:D1004–9.
[124] Li M, Zou D, Li Z, Gao R, Sang J, Zhang Y, et al. EWAS Atlas: a curated knowledgebase of epigenome-wide association studies. Nucleic Acids Res 2019;47:D983–8.
[125] Xiong Z, Li M, Yang F, Ma Y, Sang J, Li R, et al. EWAS Data Hub: a resource of DNA methylation array data and metadata. Nucleic Acids Res 2020;48:D890–5.
[126] Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015;12:e1001779.
[127] Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. Genome-wide genetic data on ~ 500,000 UK Biobank participants. bioRxiv 2017;166298.
[128] Moore JH, Williams SM. Epistasis and its implications for personal genetics. Am J Hum Genet 2009;85:309–20.
[129] Li YI, van de Geijn B, Raj A, Knowles DA, Petti AA, Golan D, et al. RNA splicing is a primary link between genetic variation and disease. Science 2016;352:600–4.
[130] McVicker G, van de Geijn B, Degner JF, Cain CE, Banovich NE, Raj A, et al. Identification of genetic variants that affect histone modifications in human cells. Science 2013;342:747–9.
[131] Chen L, Ge B, Casale FP, Vasquez L, Kwan T, Garrido-Martín D, et al. Genetic drivers of epigenetic and transcriptional variation in human immune cells. Cell 2016;167:1398–414.e24.
[132] Waszak SM, Delaneau O, Gschwind AR, Kilpinen H, Raghav SK, Witwicki RM, et al. Population variation and genetic control of modular chromatin architecture in humans. Cell 2015;162:1039–50.
[133] del Rosario RC, Poschmann J, Rouam SL, Png E, Khor CC, Hibberd ML, et al. Sensitive detection of chromatin-altering polymorphisms reveals autoimmune disease mechanisms. Nat Methods 2015;12:458–64.
[134] Grubert F, Zaugg JB, Kasowski M, Ursu O, Spacek DV, Martin AR, et al. Genetic control of chromatin states in humans involves local and distal chromosomal interactions. Cell 2015;162:1051–65.
[135] Gate RE, Cheng CS, Aiden AP, Siba A, Tabaka M, Lituiev D, et al. Genetic determinants of co-accessible chromatin regions in activated T cells across humans. Nat Genet 2018;50:1140–50.
[136] Kumasaka N, Knights AJ, Gaffney DJ. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. Nat Genet 2016; 48:206–13.
[137] Wang AT, Shetty A, O'Connor E, Bell C, Pomerantz MM, Freedman ML, et al. Allele-specific QTL fine mapping with PLASMA. Am J Hum Genet 2020;106:170–87.
[138] Benaglio P, D'Antonio-Chronowska A, Ma W, Yang F, Young Greenwald WW, Donovan MKR, et al. Allele-specific NKX2-5 binding underlies multiple genetic associations with human electrocardiographic traits. Nat Genet 2019;51:1506–17.
[139] Liang D, Elwell AL, Aygün N, Krupa O, Wolter JM, Kyere FA, et al. Cell-type-specific effects of genetic variation on chromatin accessibility during human neuronal differentiation. Nat Neurosci 2021;24:941–53.
[140] Sey NYA, Hu B, Mah W, Fauni H, McAfee JC, Rajarajan P, et al. A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. Nat Neurosci 2020;23:583–93.
[141] Lotta LA, Pietzner M, Stewart ID, Wittemans LBL, Li C, Bonelli R, et al. A cross-platform approach identifies genetic regulators of human metabolism and health. Nat Genet 2021;53:54–64.
[142] Surendran P, Stewart ID, Au Yeung VPW, Pietzner M, Raffler J, Wörheide MA, et al. Rare and common genetic determinants of metabolic individuality and their effects on human health. Nat Med 2022;28:2321–32.
[143] Köttgen A, Raffler J, Sekula P, Kastenmüller G. Genome-wide association studies of metabolite concentrations (mGWAS): relevance for nephrology. Semin Nephrol 2018;38:151–74.
[144] Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, et al. An atlas of genetic influences on human blood metabolites. Nat Genet 2014;46:543–50.
[145] Wang J, Jia H. Metagenome-wide association studies: fine-mining the microbiome. Nat Rev Microbiol 2016;14:508–22.
[146] Oh J, Byrd AL, Deming C, Conlan S, Kong HH, Segre JA. Biogeography and individuality shape function in the human skin metagenome. Nature 2014;514:59–64.
[147] Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, et al. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 2010;464:59–65.
[148] Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 2012;490:55–60.
[149] Nicoletti P, Dellinger A, Li YJ, Barnhart HX, Chalasani N, Fontana RJ, et al. Identification of reduced ERAP2 expression and a novel *HLA* allele as components of a risk score for susceptibility to liver injury due to amoxicillin-clavulanate. Gastroenterology 2023;164:454–66.

[150] Cai L, Sun Y, Liu Y, Chen W, He L, Wei DQ. Evidence that the pituitary gland connects type 2 diabetes mellitus and schizophrenia based on large-scale trans-ethnic genetic analyses. J Transl Med 2022;20:501.
[151] Zhou S, Tian Y, Song X, Xiong J, Cheng G. Brain proteome-wide and transcriptome-wide association studies, Bayesian colocalization and Mendelian randomization analyses revealed causal genes of Parkinson's disease. J Gerontol A Biol Sci Med Sci 2023;78:563–8.
[152] Mancuso N, Shi H, Goddard P, Kichaev G, Gusev A, Pasaniuc B. Integrating gene expression with summary association statistics to identify genes associated with 30 complex traits. Am J Hum Genet 2017;100:473–87.

Genomics, Proteomics & Bioinformatics, 2024, 22, 1–17
https://doi.org/10.1093/gpbjnl/qzae077
Review