Genes & Diseases (2026) 13, 101850

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.keaipublishing.com/en/journals/genes-diseases

REVIEW ARTICLE

eQTL analysis: A bridge from genome to mechanism

Zhe Jia [a,c], Jing Xu [a,c], Yingnan Ma [a,c], Siyu Wei [a,c], Chen Sun [a], Xingyu Chen [a], Jingxuan Kang [a], Haiyan Chen [a], Chen Zhang [a], Yu Dong [a], Junxian Tao [a], Xuying Guo [a], Hongchao Lv [a], Guoping Tang [b,***], Yongshuai Jiang [a,**], Mingming Zhang [a,*]

[a] *College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang 150086, China*
[b] *Department of Medical Engineering, The Fourth Affiliated Hospital of School of Medicine, and International School of Medicine, International Institutes of Medicine, Zhejiang University, Yiwu, Zhejiang 322000, China*

Received 27 December 2023; received in revised form 6 May 2025; accepted 18 June 2025
Available online 17 September 2025

KEYWORDS
Expression quantitative trait locus;
Gene expression;
Genetic variation;
Genome-wide association study;
Single nucleotide polymorphism

Abstract Expression quantitative trait locus (eQTL) refers to a genetic variation associated with the expression of specific genes. It has been widely applied to explain the regulatory mechanisms linking genetic variations to complex traits or diseases. Several eQTLs have been identified from tissues and single cells in individuals. Furthermore, the integration of eQTL and other omics data can be used to detect novel susceptibility genes and consequently understand the dynamic regulation of trait-associated genetic variations at the system level. Here, we review the identification methods, analysis tools, research progress, and common data resources of eQTLs, as well as their role in four typical diseases. Finally, we discussed the application fields, challenges, and future development perspectives of eQTL.
© 2025 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

* Corresponding author.
** Corresponding author.
*** Corresponding author.
E-mail addresses: tgp@zju.edu.cn (G. Tang), jiangyongshuai@hrbmu.edu.cn (Y. Jiang), zhangmingming@hrbmu.edu.cn (M. Zhang).
Peer review under the responsibility of Chongqing Medical University.
[c] These authors contributed equally to this work.

https://doi.org/10.1016/j.gendis.2025.101850
2352-3042/© 2025 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Introduction

One of the challenges faced by researchers in the post-genomic era is understanding the genetic mechanism between genomic variations (such as single-nucleotide polymorphisms or SNPs) and the phenotype. Quantitative trait loci (QTLs) are genomic loci that exert a quantitative effect on certain traits.[1] Expression quantitative trait loci (eQTL), the most common QTLs, are regulatory genetic variation loci that govern the expression of specific genes. The standard method of identifying eQTLs is a linear regression model in which gene expression is considered a quantitative trait to quantify the effect of genetic variation on gene expression.[2]

Numerous studies have implicated eQTLs in multiple ways. For example, the eQTLs located within the gene transcription initiation regions have been demonstrated to affect the affinity of transcription factor binding sites or alter chromatin accessibility, thereby altering gene expression and functions. Similarly, the eQTLs located in the coding region of the gene directly influence the activity of the encoded protein.[3] Another mode of action of eQTLs is to regulate the expression of target genes through regulatory proteins or microRNAs.[4] However, the most important function of eQTLs is the explanation of the mechanism between trait-associated genetic variants in non-coding regions and complex traits after genome-wide association studies (GWAS). Several methods have been developed using this concept, such as the co-localization analysis. In addition, significantly associated SNPs in GWAS that have been detected as eQTLs are speculated to be functional genetic variations, but they need to be further verified. eQTLs can also be used to study causal inference. For example, transcriptome-wide association analysis (TWAS) uses eQTL data to predict gene expression and identify the association between expression and the associated trait.[5] Mendelian randomization can infer causal genes regulated by eQTLs for a complex trait by integrating GWAS and the eQTL summary statistics.[6] Initially, eQTL analysis was conducted using the gene microarray data. The development of sequencing technology has allowed eQTL researchers to use both bulk sequencing data and single-cell RNA sequencing (scRNA-seq) data. For instance, Will et al. (2013) demonstrated the impact of single-cell analysis on genetic variants.[7] Several tissue-specific eQTLs and cell-type-specific eQTLs have been discovered. Numerous studies have reported that single-cell eQTL studies can be used to understand the complex genetic mechanisms of several diseases.[8,9] For instance, Yazar et al identified 26,597 independent cis-eQTLs in 14 immune cells by integrating the scRNA-seq and genotype data from 1.27 million peripheral blood mononuclear cells from 982 healthy samples, and found that approximately 305 eQTLs were associated with autoimmune diseases because of altered gene expression in specific cell types.[10] Nathan et al reported enriched autoimmune-related genetic variants in cell state-dependent eQTLs and demonstrated that cell state was crucial to explain the function of eQTLs.[11] At present, molecular eQTL annotation has emerged as a powerful strategy to explain the genetic mechanisms of diseases.

Identification of eQTLs and elucidation of their functions in several diseases will assist us in better understanding the disease etiology and facilitate disease diagnosis or treatment. Here, we review the relevant knowledge of eQTLs, including their identification and analytical methods, development across different physiological periods, existing data resources, and their application in four typical diseases.

Identification and analysis of eQTLs

eQTLs play an irreplaceable role in the pathogenesis of several diseases. It is indispensable to be familiar with eQTL data analyses and identification methods.[12,13] Here, we systematically used the data for analysis, identification, analysis tools, and subsequent analysis for researchers.

Data

The eQTL analysis requires multiple data during calculation, including SNP genotype matrix and gene expression matrix from the same samples, physical location data, and optional covariates matrix.

Raw genotype data are required to perform eQTL analysis. Briefly, genotype data are usually labeled as 0, 1, and 2, representing common allele homozygous, heterozygous genotype, and minor allele homozygous, respectively. In the case of insufficient genotype density in certain samples, genotype imputation is performed using HapMap[14] (http://www.hapmap.org/) or 1000 Genomes Project[15](http://www.1000genomes.org/) as the reference data using the IMPUTE2 software.[16] The criteria for selecting SNPs after imputation are: i) minor allele frequency (MAF) $\geq$ 5%, ii) SNP missing rate <5%, and iii) Hardy–Weinberg equilibrium test p-value $>1 \times 10^{-6}$.[17,18]

Gene expression data can be microarray or sequencing data. These data are finally converted into a gene expression matrix. The R package "affy" can read raw microarray data and perform standardized processing workflows for it.[19,20] Raw sequencing data need to be processed using multiple software tools for quality control, data filtering, normalization, and other steps to be converted into a matrix format.

Covariate data are processed into a matrix for better analysis, with columns representing samples and rows representing different covariates. Usually, apart from SNPs, other factors affecting gene expression are selected as covariates, and biologically relevant factors such as age, gender, and race should be prioritized. Other potential confounding factors need to be evaluated for their impact on gene expression to ensure the selection of the appropriate covariates.

Location files for genes and SNPs are required input for eQTL mapping so that the software knows which gene-SNP pairs to include for cis or trans eQTL testing. Gene positions are obtained from the National Center for Biotechnology Information's (NCBI) genome database (www.ncbi.nlm.nih.gov/genome/), whereas SNP positions are obtained from the dbSNP database (https://www.ncbi.nlm.nih.gov/SNP/).

Identification of eQTLs

eQTL analysis, a linear regression analysis, includes the SNP genotype as the independent variable and gene expression as the dependent variable (Fig. 1A).[21] For example, a SNP has two alleles (A and a), and g_i is the expression value of the ith gene. The linear regression model of detecting eQTL is $g_i = \alpha_i + \beta_i x + \epsilon_i$, where x represents the genotype, and AA, Aa, and aa can be coded as 2, 1, and 0, respectively (Fig. 1B). Multiple regression is used in cases including covariates. The linear regression model is fitted to detect eQTLs, and statistical data such as SNP—gene pairs, beta values, p-values, and distances between SNPs and genes can be obtained. Type I errors are controlled using commonly applied multiple testing correction methods, such as Bonferroni and false discovery rate. When the adjusted p-value is smaller than the usually set α (*e.g.*, 0.05), a SNP is identified as an eQTL (Fig. 1C).

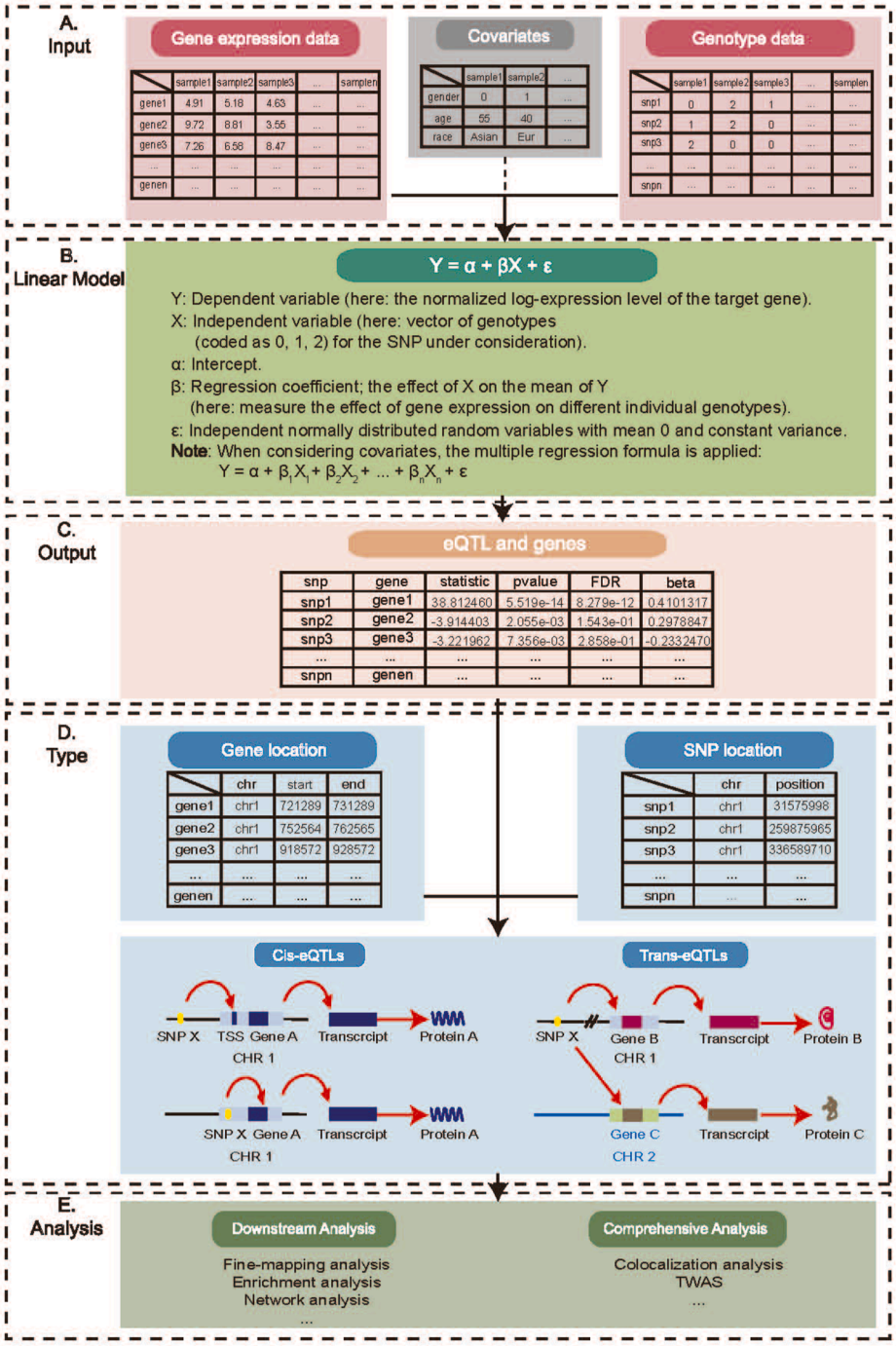

Figure 1 Workflow for eQTL identification and analysis. **(A)** Input data preparation. **(B)** Linear regression for eQTL identification. **(C)** Results of eQTLs. **(D)** Different types of eQTLs. **(E)** Subsequent analysis of eQTLs.

eQTLs are classified into cis-eQTLs and trans-eQTLs based on the distance between eQTLs and transcription start sites of regulated genes. cis-eQTLs refer to genetic variations located within a genomic window of the target gene, such as regions upstream and downstream of 1 Mb region.[22,23] They are known to affect the structure of gene regulatory regions or chromatin and alter gene expression by influencing transcripts in the coding regions.[24] In contrast, trans-eQTLs constitute genetic variations located outside this window region of the target gene, including those on different chromosomes, and can modify gene expression by interacting with regulatory proteins, such as transcription factors.[17] For instance, the eQTL "SNP X" is shown in Figure 1D. The left image illustrates that SNP X is a cis-eQTL for gene *A* located on the same chromosome as the regulated gene *A*. Conversely, the right image displays that, being far from the regulatory gene, SNP X is a trans-eQTL for target genes *B* and *C*. Although both eQTL types can explain the regulatory relationship between genetic variations and gene expression, most present studies focus on the regulatory function of cis-eQTLs.[18] Because cis-eQTLs are located near the target gene and typically exert stronger effects on gene expression, they are more likely to influence gene expression through direct regulation. cis-eQTLs simplify causal inference from genetic variation to phenotype, facilitating subsequent functional studies and drug target discovery. In contrast, trans-eQTLs often involve indirect regulatory networks, making them more challenging to study, and yet remain indispensable for comprehensively understanding genetic regulation.

We conducted a comprehensive eQTL analysis on simulation data using the MatrixeQTL software; specific details are provided in the supplementary material.

Common eQTL identification tools

The increasing use of eQTL analysis has led to the application of numerous tools in this field.

MatrixeQTL (www.bios.unc.edu/research/genomic_software/Matrix_eQTL/) is a classic and common eQTL analysis software used for identifying cis-eQTLs and trans-eQTLs.[21] It can efficiently analyze large-scale data, support linear model analysis of genotypes and expression levels, support the adjustment of multiple covariates (such as age and gender), and provide an R language interface. Its drawback is that the correction of multiple tests relies on traditional methods (such as Bonferroni), which may be overly strict. Additionally, it is limited in the analysis of single-cell data rand complex cellular heterogeneity.

FastQTL (http://fastqtl.sourceforge.net/) is another commonly used software, specifically designed for cis-eQTL analysis and widely applied in the analysis of projects such as Genotype-Tissue Expression (GTEx).[25] Its advantages are fast running speed, support for covariate adjustment and normalization of gene expression data, optimized multiple corrections, and suitability for parallel analysis of large-scale single-cell datasets. Its disadvantage is that the ability to analyze trans-eQTL is relatively weak.

QTLtools (https://qtltools.github.io/qtltools/) applies to a variety of molecular QTL analyses, including eQTL, splicing QTL, and chromatin interaction QTL analysis. It supports multiple types of QTLs and phenotypic data (*e.g.*, RNA-seq, ATAC-seq), but the installation is relatively complex, has high requirements for computing resources, and requires a high-performance computing environment.[26]

In recent years, the rapid development of single-cell sequencing (scRNA-seq) technology has provided us with an opportunity to uncover associations at the single-cell level between genetic variations and gene expression. Many eQTL tools at the cellular level that have emerged. SCeQTL[27] is an R package (https://github.com/XuegongLab/SCeQTL/), that can not only perform eQTL analysis on single-cell data, but also detect variations in gene expression associated with other grouping factors, such as cell lineages or cell types. It can capture cellular heterogeneity and reveal the regulatory effects masked by traditional batch analysis. However, currently, due to high data sparsity and low statistical power, a large sample size is required for its support. The analysis process is not standardized yet, and the computational complexity is high.

Single-cell eQTL analysis usually uses genetic variation and expression data from the same cells. However, the current technology of such paired data in single cells is still immature. The eQTLsingle method (https://github.com/horsedayday/eQTLsingle) addresses the limitation, which can detect mutations in coding regions using scRNA-seq data to capture both mutation and expression profiles from individual cells.[28] Therefore, eQTLsingle makes it possible by identifying eQTLs only with scRNA-seq data or snRNA-seq data. Nevertheless, this approach has limitations in identifying potential eQTLs in non-coding regions and genes with low expression levels.

SURGE (https://github.com/BennyStrobes/surge) is a method to identify context-specific eQTLs from single-cell transcriptomic data, which can apply to any eQTL data.[29] SURGE is particularly effective at automatically detecting latent contexts and demonstrates high statistical power. Nonetheless, it might face challenges in interpreting latent contexts, detecting small-effect eQTLs, and scaling to very large datasets.

Cell Regulatory Map (CellRegMap, https://github.com/limix/CellRegMap) is a statistical framework to perform multi-context eQTL mapping using scRNA-seq data based on a linear mixed model.[30] CellRegMap can account for sample structure, including population structure and repeated observations from the same samples. Its key strengths include effective integration of diverse data types and the capability to uncover cell type-specific regulatory mechanisms. However, it necessitates high-quality multimodal data and may encounter challenges in interpreting complex regulatory networks.

Presently, there is a rising trend in extensive scRNA-seq and cell type-specific eQTL (ct-eQTL) investigations.[31] It is worth noting the web tool developed by Igor Mandric et al (https://mandricigor.github.io/ct-eqtl-design/) that effectively aids scientists in the optimization of ct-eQTL research experiment design.[32] By entering desired cell types, budget constraints, and sequencing expenses, the optimal sample size and number of cells are automatically determined by this calculator, leading to a significant enhancement in statistical power.

Additionally, single-nucleotide variants can function as eQTLs. The RBP-Var tool (www.rbp-var.biols.ac.cn) is employed to investigate the relationship between single-

nucleotide variants and gene expression, as well as provide annotations for functional variants involved in post-transcriptional regulation. The details of the aforementioned tools are available on our website (http://www.onethird-lab.com/eQTLbrowser).

Analysis of eQTLs

The identified eQTLs can be mapped to the reference genome, with a few located within the genes and others located between genes. These eQTLs offer diverse analytical applications, including downstream analysis of regulatory relationships between the target gene and other genomic elements.[24] For instance, enrichment analysis can deduce the biological functions associated with eQTLs, whereas network analysis investigates their impacts within a multi-omics framework.[33] In addition to downstream analysis, integrated analysis is widely applied. GWAS, developed over a decade ago, is a well-established technique for identifying associations between SNPs and complex traits.[34] Data sources such as UKbiobank[35](https://www.ukbiobank.ac.uk/), GWAS catalog[36–38] (http://www.ebi.ac.uk/gwas/), and GWAS ATLAS[39] (https://bigd.big.ac.cn/gwas/) provide successful GWAS outcomes. Although hundreds of risk SNPs have been identified using GWAS,[40] the mechanism between these SNPs and the complex traits remains elusive. In this regard, the eQTL analysis can detect the regulatory relationship between SNPs and gene expression and explain the regulation route from genetic variations to disease risk (SNP → gene expression → disease).[41,42] Therefore, causal relationships can be studied by integrating eQTL and GWAS analyses.

It is noteworthy to mention TWAS, a method that integrates eQTL and GWAS to unravel the genetic regulatory mechanisms underlying complex traits or diseases.[5] Its core principle involves constructing predictive models of gene expression to translate associations between genetic variants and phenotypes into associations between gene expression levels and phenotypes, thereby identifying potential causal genes. A variety of methods or software (*e.g.*, PrediXcan, S-prediXcan, FUSION, UTMOST) have been developed for TWAS analysis.[43] TWAS, as a gene-based analysis method, presents a reduced multiple testing burden compared with GWAS. By presenting results at the level of specific genes, it facilitates subsequent functional analyses and enables more direct interpretation of biological significance. The GTEx project provides an exceptionally rich resource of genomic and transcriptomic data for TWAS analysis, allowing researchers to construct gene expression prediction models using this reference panel. TWAS is particularly well-suited for identifying susceptibility genes for complex diseases, as well as applications in early diagnosis, disease prediction, and risk assessment research.[44] TWAS integrated with eQTL analysis has become a pivotal tool for dissecting the genetic architecture of complex traits by bridging genetic variation and functional genes. With continuous optimization of multi-tissue modeling, fine-mapping, and cross-omics integration methods, its applications have expanded from disease mechanisms to agricultural breeding. It holds great promise for future roles in precision medicine and molecular design breeding. However, challenges, such as cross-population data compatibility, dynamic regulatory mechanism resolution, and causal validation, remain critical bottlenecks requiring breakthroughs.

Common eQTL analysis tools

Recently, several analytical tools have been applied for integrating eQTL and GWAS, facilitating further screening of disease-related functions of SNPs. Here, we introduced certain efficient and typical tools such as eCAVIAR (http://zarlab.cs.ucla.edu/tag/ecaviar/),[42] SMR (https://yanglab.westlake.edu.cn/software/smr),[45] XGR (https://xgr.r-forge.r-project.org/),[46] and OmicKriging (http://www.scandb.org/newinterface/tools/OmicKriging.html).[47] eCAVIAR is a Bayesian probabilistic model used to identify causal variants in both GWAS and eQTL studies based on the co-localization method. Its operating speed is fast. Summary-based Mendelian randomization (SMR) is a linear regression-based eQTL analysis tool known for its efficient processing speed. It facilitates simultaneous analysis of associations between gene expression and genotype data. XGR is a comprehensive bioinformatics toolkit offering several analysis modules, including eQTL analysis. The actual running speed could vary depending on the selected module and dataset size. OmicKriging is an eQTL analysis tool that utilizes Kriging interpolation; it has a slightly slower running speed, particularly when dealing with large-scale datasets. eCAVIAR and SMR can accept genotype and expression data in the summary statistics format as inputs, whereas OmicKriging specifically accepts data in the PLINK format. XGR, in contrast, supports data analysis in both VCF and PLINK formats. SMR provides exceptionally beautiful visualizations of the output data. Similarly, XGR can visualize the results. The performance of eCAVIAR is assessed from an algorithmic standpoint by comparing it against simulated data or known results. Key evaluation metrics include accuracy, recall, and F1 score. The efficacy of SMR is measured by comparing it with the established results. Relevant evaluation metrics include the precision of association statistics and *p*-values. Similarly, the performance of XGR is assessed by comparing it with known results. Common evaluation metrics include the accuracy of association statistics, *p*-values, and annotation information accuracy. The performance of OmicKriging is assessed by comparing it with known results. Common evaluation metrics include the accuracy of eQTL mapping and gene expression prediction. The details of all tools are available at http://www.onethird-lab.com/eQTLbrowser.

eQTL in different development periods

The eQTL mapping analysis was first implemented in 2001, associating the SNPs with the target transcript abundance.[48] Over the years, eQTL studies have evolved and progressed significantly (Fig. 2), driven by advancements in technology and methodologies. We will discuss the major developments in eQTL analysis across four periods: microarray period, high-throughput sequencing period, post-genome period, and single-cell sequencing period.

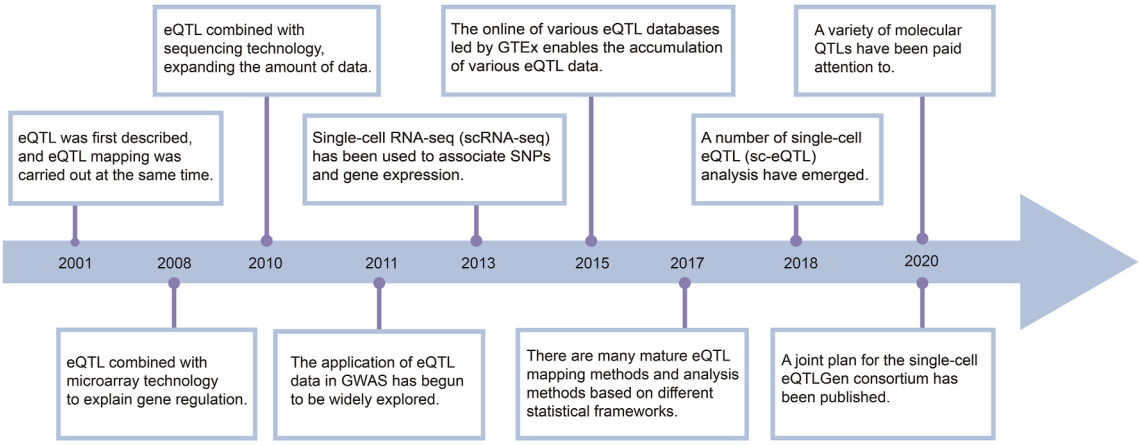

Figure 2 eQTL development.

Microarray period

In the 1990s, high-throughput DNA chips or microarrays were applied to different kinds of gene expression scenarios.[49] These have emerged as efficient methods that can simultaneously perform gene expression analysis and genotyping.[50] This became a prerequisite for the development of eQTL research. After obtaining gene expression and the SNP genotype of the same sample from the microarray, an eQTL is identified by regression analysis. Chip technology was used in the early study of eQTLs in several diseases, such as rheumatoid arthritis (RA) and ischemic stroke. eQTL analysis using Illumina BeadChip microarrays from 344 patients with RA revealed that *METL21B*, *JAZF1*, *IKZF3*, and *PADI4* were regulated by cis-eQTLs in $CD4^+$ lymphocyte subsets, and alterations of their expression suggested a role of $CD4^+$ lymphocyte subsets in an RA-specific effect.[51] Four SNP-gene pairs were identified by microarray eQTL analysis in ischemic stroke, which were implicated in immune responses and contributed to the development of personalized treatment.[52]

Altogether, chip technology contributes to the development of early eQTL research and prompts researchers to understand the regulation of several genes from the perspective of genetic variations.[53]

High-throughput sequencing period

The continuous advancement and evolution of experimental technology have led to the emergence of large-scale and high-resolution RNA-seq, greatly promoting the development of eQTLs and enhancing the understanding of eQTLs' potential regulatory mechanisms. Compared with the microarray technology, RNA-seq technology can more accurately explain the relationship between genetic variations and phenotypes. Therefore, RNA-seq technology has been widely used for eQTL research since 2010.[54]

In the era of high-throughput sequencing, eQTL analysis has greatly promoted the study of the pathogenesis of several diseases.[55] For instance, an eQTL analysis using sequencing technology confirmed certain novel loci related to obesity, such as *Lpl*, *Lactb*, and *Ppm1l*.[56] Another eQTL study of the prostate tissue revealed that rs13215402 regulated the expression of *RGS17*, which was finally confirmed as a susceptibility gene for prostate cancer. Specifically, the higher expression of *RGS17* is associated with the severity of prostate cancer.[57]

Several databases have emerged during the high-throughput sequencing period to store the eQTL results. The largest one is the GTEx database, which is a landmark and provides researchers with rich eQTL data to explore gene expression patterns in different tissues.[58] Several eQTL analysis tools based on sequencing data have been developed, such as FastQTL[25] and QTLtools.[26]

Although hundreds of eQTLs associated with complex traits have been identified, it is challenging to integrate eQTLs with other omics data for a comprehensive understanding of human diseases.

Post-genome period

In the post-genome period, several methods have been applied to integrate the eQTL analysis with GWAS for the causal analysis of human diseases,[59] such as SMR and colocalization. SMR-based integrative analysis has identified numerous risk genes in several diseases, such as *TRAF1* and *ANKRD55* in RA and *SNX19* and *NMRAL* in schizophrenia (SZ).[45] For instance, Huang et al identified that inflammatory bowel disease-related variants had significant colocalization signals with eQTLs in $CD4^+$ T lymphocytes.[60] Similarly, Franceschini et al identified two candidate genes, namely, *CCDC71L* and *PRKAR2B*, by using GWAS of cardiovascular signatures to colocalize with eQTLs in the vascular tissue, leading to atherosclerosis in cardiovascular disease.[61]

The integration of eQTL and survival analyses for identifying survival-related eQTL has also been attempted in the study of different cancers. For example, Gong et al identified 22,212 survival-related eQTLs to explain the relationship between genetic variations and survival.[62] Network analysis of integrating eQTL and other omics data is a current research hotspot.[63] Mo et al constructed a genome-wide long non-coding RNA (lncRNA)-eQTL network and identified eQTL as an intermediary factor in mRNA—lncRNA functional modules.[64]

Single-cell sequencing period

Traditional studies involving bulk RNA-seq data often do not capture information regarding allele-specific expression that pertains to specific cell types. The introduction of scRNA-seq around the mid-2010s has brought about new possibilities for analyzing eQTLs at the individual cell level, providing valuable insights into the diversity of cells, differentiation among cell types, and disease mechanisms.[65] Early approaches for linking SNPs with gene expression in scRNA-seq data began to surface in 2013, establishing the groundwork for studying single-cell eQTLs.[7] In 2018, the first genuine single-cell eQTL (sc-eQTL) analysis was conducted by Kang et al using droplet-based single-cell sequencing technology. They examined 78,000 peripheral blood mononuclear cells from eight major immune cell populations from 23 donors and identified 32 cis-eQTLs, 22 of which were specific to particular cell types.[66] In the same year, Van Derwijst et al analyzed scRNA-seq data from 25,000 peripheral blood mononuclear cells of 45 Dutch donors and identified cell type-specific cis-eQTLs and co-expression QTLs.[67]

The rapid advancement of scRNA-seq technology has enabled numerous research groups to explore the intricate regulatory landscape in various biological systems, including peripheral blood mononuclear cells, T cells, induced pluripotent stem cells, brain cortex, and dopaminergic neuron differentiation.[8] These investigations have yielded significant findings regarding cell-type-specific eQTLs associated with immune response genes,[10,68,69] the genetic influence on gene expression during cellular reprogramming and differentiation processes,[70] genetic regulatory mechanisms governing gene expression in various cell types like neurons, astrocytes, and oligodendrocytes, the genetic foundation of cellular diversity and functional specialization in the brain,[71,72] and genetic factors that impact gene expression patterns in dopaminergic neurons.[73] Thus, these discoveries enhance our comprehension of complex diseases (*e.g.*, autoimmune diseases and neurodevelopmental disorders), cellular reprogramming, and brain development, laying the groundwork for precision medicine and therapeutic interventions. Alongside sc-eQTL studies, several tools and approaches have been developed to optimize and design sc-eQTL analysis, including SCeQTL[27] and scPower.[74] Additionally, various resources are accessible, such as the single-cell eQTLGen consortium[75] and the OneK1K cohort.[10]

Despite significant progress, several challenges and issues need to be addressed in sc-eQTL studies.[76] The current analysis methods and tools require improvements to handle the complexity and unique features of single-cell data. Breakthroughs in sc-eQTL research are expected with further technological advancements. These breakthroughs will help unravel gene regulatory networks, uncover disease mechanisms, and support personalized medicine and therapeutic interventions.

Common eQTL data resources

The innovation of sequencing technology and the wide application of eQTL identification methods have resulted in the development of certain large eQTL databases, providing great convenience for users to understand the functions of genetic variations in the genome.

GTEx (V8) (https://gtexportal.org/home/) is a comprehensive public resource, containing the eQTL data of 49 non-disease tissues from nearly 1,000 individuals.[77] Data can be downloaded by tissue and eQTL type (cis-eQTL or trans-eQTL) from the GTEx database.[78] GTEx contributes to the identification of disease-related risk genes. For example, Yang et al confirmed that *GLT8D1* and *CSNK2B* were risk genes for SZ using brain tissue data of 92 subjects from the GTEx database,[79] thus providing novel ideas for the prevention and clinical treatment of SZ. Kundu et al identified 340 pathogenic variant loci with a high posterior probability for multiple immune-mediated diseases using the GTEx database.[80]

eQTLGen (www.eqtlgen.org/) is a free eQTL database that stores cis- and trans-eQTLs from 31,684 whole blood samples.[81] In addition, the database stores single-cell eQTLs from more than 30,000 blood samples, contributing to differences in eQTLs' replication rates between single-cell data and mixed cell-type data.[18] These data assist in understanding the tissue-specific expression of eQTLs in blood. A systematic study of pancreatic cancer identified rs3802266 as an important cis-eQTL for *ZHX2* using eQTLGen and found that the lower *ZHX2* expression was related to a higher risk of pancreatic cancer in individuals.[82] In 2021, Yang et al identified SNPs, namely, rs13428812, rs116555563, and rs9977672, in multiple sclerosis and regulatory pathways shared between different diseases through a meta-analysis combined with eQTLGen.[83] Multiple sclerosis and inflammatory bowel disease share the same tissue-specific heritability enrichment pattern for eQTLs in whole blood tissue, suggesting a biological basis for co-morbidity between the two.

eQTL Catalogue (www.ebi.ac.uk/eqtl) is an invaluable resource providing uniformly processed gene expression cis-eQTLs and splicing QTLs from 32 different studies, where summary statistics are available for several cell types and tissues.[8] It provides a comprehensive collection of genetic variants associated with altered gene expression and splicing, and can systematically explain human GWAS associations across a wide range of cell types and tissues.[84] Koel et al performed the GWAS meta-analysis, shedding light on the genetics of cervical cancer phenotypes and providing valuable information for risk stratification of cervical cancer.[85] Chun et al utilized the data from the eQTL Catalog to identify pleiotropic variants impacting both gene expression and obstructive sleep apnea, thereby providing valuable insights into the underlying mechanisms of complex human diseases.[86] eQTL Catalogue was updated in 2023, which increased the number of uniformly processed studies from 21 to 31 and added X chromosome QTLs for 19 compatible studies.[87]

OneK1K (www.onek1k.org/) is a large cohort containing scRNA-seq data from 1.27 million peripheral blood mononuclear cells collected from 982 donors.[10] An eQTL analysis was performed across 14 immune cell types that identified 26,597 cis-eQTLs. The majority of these eQTLs exhibited cell type-specific effects.[8] These findings contribute to our understanding of how genetic variation impacts gene regulation and signaling pathways in a cell type-specific manner. Yazar et al used the OneK1K to identify single-cell eQTLs in

nine immune-mediated diseases and found that most eQTLs affected T-cell activation and signaling.[88] Another study conducted using OneK1K confirmed the dynamic nature of HLA eQTL effects across different immune cells. This highlights the significance of including cell types when investigating the genetic mechanisms underlying HLA expression.[89]

scQTLbase (http://bioinfo.szbl.ac.cn/scQTLbase) is the first integrated comprehensive database designed to explore human sc-eQTLs, which unravel the genetic basis of cell type-specific gene regulation.[90] By consolidating 304 sc-eQTLs datasets from diverse sources, scQTLbase spans 57 cell types and 95 cell states. It includes approximately 16 million SNPs significantly linked to cell type/state-dependent gene expression, along with ~690,000 disease-associated sc-eQTLs derived from 3333 traits/diseases. Beyond sc-eQTL search and browsing functionalities, scQTLbase enables gene expression visualization in UMAP plots and colocalization visualization with GWAS datasets. scQTLbase will help in the discovery of disease susceptibility genes.

SingleQ (http://www.sqraolab.com/scqtl) is a single-cell eQTL interactive database, which collects sc-eQTL datasets and provides online visualization of sc-eQTLs across different cell types in a user-friendly manner.[91] SingleQ sc-eQTL database curated up to 77467 eQTL summary statistics from 273 unique cell types covering different developmental stages of diverse tissues or cell states. SingleQ is the most comprehensive summary statistics by far.

PancanQTL (http://gong_lab.hzau.edu.cn/PancanQTL/) is the first cancer-related database that stores eQTLs of 9196 tumor samples from 33 cancer types in the Cancer Genome Atlas (TCGA). It contains 5,606,570 cis-eQTLs, 231,210 trans-eQTLs, 22,212 survival-related eQTLs, and 337,131 GWAS-related eQTLs, and supports querying, browsing, and downloading data.[62] Carrot-Zhang et al[92] identified tens of thousands of cancer-type-specific eQTLs using PancanQTL and explained the impact of reproductive system genetic variations on cancer gene expression. PancanQTLv2.0 (https://hanlaboratory.com/PancanQTLv2/) has updated the fine-mapping of causal variants for eQTLs, identifying eQTLs associated with drug response and immune infiltration. The database finalized 58,747 fine-mapped eQTL credible sets and identified 84,592,135 linkage associations between eQTLs and existing GWAS loci, marking a 50-fold increase compared with PancanQTLv1.0. Additionally, 659,516 associations between eQTLs and drug responses and 146,948 associations between eQTLs and immune cell abundance were contained in PancanQTLv2.0, which provides a foundation for the potential clinical application of eQTLs in cancer therapy.[93]

Several other eQTL databases, such as Blood eQTL Browser,[94] EyeGEx,[95] and ImmuNexUT,[96] have been widely used. More details of these important eQTL databases are shown in Tables S1—S3.

The role of eQTL in clinical diseases

eQTLs have several unique functions in disease research; they act as a bridge from genetic variations to functions and can cover the limitations of GWAS. Numerous disease-associated eQTL genes have been discovered, thereby elucidating the regulatory mechanism of risk SNPs and diseases from a functional perspective. We will focus on the contribution of eQTL analysis in four typical diseases: RA, type 2 diabetes (T2D), breast cancer, and SZ. Certain eQTL-gene pairs of four diseases in the whole blood are shown in Table S4. More details are available at www.onethird-lab.com/eQTLbrowser.

Rheumatoid arthritis

RA is a chronic and inflammatory autoimmune disease that could lead to joint damage and systemic complications.[97] It is known to affect approximately 0.5%—1% of the world's population.[98] The etiology of RA is complex, and genetic factors can explain approximately 60% of RA susceptibility.[99] The eQTL analysis has provided a boost to understanding the pathogenic mechanisms underlying RA.

Recently, an eQTL analysis identified certain novel susceptibility genes and revealed a new genetic mechanism for RA. Thalayasingam et al identified novel eQTLs regulating *METL21B*, *JAZF1*, *IKZF3*, *PTPN22*, and *PADI4* in $CD4^+$ T cells for the first time.[51,100] Among these, SNP rs142845557 is located in a non-coding region, which is a synovial tissue-specific eQTL. It regulates chromatin accessibility in synovial tissue (such as the open state of enhancers or promoters), thereby influencing the expression of *JAZF1*. *JAZF1* plays a critical role in the activation of synovial fibroblasts and may be involved in joint inflammation and tissue destruction processes.[101] *PTPN22* expression showed significant associations with RA risk across 44 tissues based on an integrative analysis of cross-tissue TWAS and eQTL. The mechanistic investigation revealed that the rs2476601 modulated *PTPN22* expression and regulated T-cell receptor signaling pathway activity. Dysfunctional *PTPN22* led to T-cell overactivation, disrupted immune tolerance, and thereby promoted autoantibody production in RA pathogenesis.[102] An example in *PADI4* demonstrates that eQTL analysis can help us understand disease risk from genotype to RA. The expression level of *PADI4* is associated with rs2240335. The A allele at rs2240335 is associated with an elevated risk of RA. rs2240335-A is associated with increased expression of *PADI4* in neutrophils but reduced expression in monocytes.[103] Ishigaki et al identified that an eQTL rs7616215 regulated *CCR2* and *CCR3* simultaneously in monocytes, thereby activating immune conditions and increasing the risk of RA.[104] These results highlight how eQTLs serve as functional bridges connecting genetic variants to biological mechanisms.

Single-cell eQTL modeling is an increasingly popular research direction, and environment-dependent effects in RA pathogenesis have been comprehensively studied using single-cell analyses. In total, 6511 *cis*-eQTLs were identified by Nathan et al, among which RA risk variants near *ORMDL3* and *CTLA4* were enriched in cell-state-dependent eQTLs.[11] Regulatory regions of immune cells in RA tissues overlap with several GWAS gene loci, especially $CD4^+$ T cells and $CD8^+$ T cells, which is consistent with earlier findings.[105]

Type 2 diabetes

T2D is a common metabolic disease affecting more than 300 million people worldwide.[106] Although GWAS has identified

several susceptibility SNPs in T2D, it does not completely explain how risk loci lead to T2D, and eQTLs fills this gap well. Here, we explored the relationship between genetic variants and risk genes influencing T2D susceptibility and associated glycemic characteristics.

Certain effects on islet cell function have been identified through various studies involving eQTLs. For example, the typical eQTL rs4547172, as identified by Keildson et al, leads to the overexpression of the *PFKM* gene, which is associated with fasting plasma insulin, reduced metabolic flexibility, and the promotion of T2D.[107] The presence of the cis-eQTL rs75409866 of the *KCNA6* gene in pancreatic islets, as confirmed by Varshney et al, impacts glucose-stimulated insulin secretion (*GSIS*) and contributes to the development of T2D.[108] Khamis et al reported that eQTLs regulated *TTLL6* (rs11657371), *MLX* (rs646123), and *KIF9* (rs2276854) through co-localization with T2D GWAS, even though these were not considered risk loci in GWAS, they still exhibited significant associations with T2D risk genes.[109] Mulder reported that the cis-eQTL rs950994-A of *TFB1M* enhances TFB1M expression, thus inhibiting insulin secretion in human islets, making individuals more susceptible to T2D.[110,111] Additionally, co-localization analyses have yielded new insights into the mechanisms underlying T2D. Broadaway et al identified 36 insulin-promoting genes that co-localize with pancreatic islet eQTLs, potentially identifying candidate genes for the disease.[112] A co-analysis with depression revealed a significant causal effect on T2D. CorrespondingeQTLs for *CCND2* and *IRS1* were found in pancreatic islets, brain, and fat, indicating the multi-tissue and pleiotropic nature of T2D pathogenesis. Furthermore, Mendelian randomization demonstrated causal effects, emphasizing the importance of preventing T2D when depressive symptoms first appear.[113] These results not only elucidate the genetic architecture underlying T2D but also provide valuable insights into the molecular mechanisms driving disease progression, emphasizing the importance of eQTLs as a bridge from genomic variation to functional consequences in complex metabolic diseases.

Breast cancer

Breast cancer is the most common cancer in females worldwide,[114] which presents familial aggregation, suggesting that genetics is an important factor. GWAS and meta-analysis have identified several risk loci, such as *BRCA1*, *BRCA2*, and *PALB2*; however, these loci can only explain about 30% of the familial risk.[115] The eQTL analysis can directly or indirectly probe the expression patterns affecting target genes and thus influence breast cancer, largely compensating for heritability analysis methods.

Li et al performed a trans-eQTL analysis using breast cancer data from TCGA and identified three new risk loci, namely, rs2046210, rs418269, and rs471467, regulating *ESR1*, *MYC*, and *KLF4*,[116] which can promote the occurrence and development of breast cancer from multiple signaling pathways and transcriptional regulatory mechanisms.[115,117,118] In addition, regulatory elements of breast cancer target genes are affected by eQTLs. Guo et al found that cis-eQTLs, including rs11552449 (*DCLRE1B*), rs7257932 (*SSBP4*), rs2236007 (*PAX9*), and rs73134739 (*ATG10*), significantly altered the promoter activity of these breast cancer-related target genes.[119] Among them, rs11552449 can affect the expression of *DCLRE1B* by enhancing its promoter activity, thereby affecting DNA stability and increasing the probability of cancer.[119] Similarly, rs2236007 can promote the occurrence and development of breast cancer by regulating *PAX9*, changing the transcriptional activity of downstream genes, and participating in signaling pathways.[120] Rezaie et al identified 1030 ncRNAs significantly mutated in breast cancer. They were significantly enriched for eQTL polymorphisms, and their genome-wide association SNP levels were unprecedentedly higher than genome-wide levels.[121] Recently, Shidal et al conducted a cis-eQTL analysis on the classical breast cancer risk SNP, rs35418111, to re-explore the causative mechanism. The results demonstrated that rs35418111 was associated with reduced expression of the *YBEY* gene, which is involved in a network of genes associated with inflammation and metabolism. Thus, the finding provides an important basis for exploring the association with breast cancer risk.[122]

Schizophrenia

SZ is a serious mental illness characterized by frequent hallucinations and is caused by changes in brain function. Genetic factors constitute the major factors affecting SZ. eQTL facilitates the identification of genetic markers and the explanation of the complex genetic structure in SZ.[123] The combination of eQTLs and single-cell RNA sequencing analysis has allowed researchers to deepen their comprehension of SZ further.

In particular, eQTL analysis identified several novel SZ-related loci in the hippocampus and and in cerebrospinal fluid metabolizing species. Yang et al first identified *ALMS1*, *GLT8D1*, and *CSNK2B* as pathophysiologically relevant risk genes in SZ by comprehensive eQTL analysis in the brain tissue.[79] Another study further confirmed *ALMS1* as a causal gene for SZ.[124] In addition, Walker et al used a combination of cis-eQTLs and GWAS to identify dozens of novel candidate risk genes, including *SNX19*, *VSP29*, and *XRCC3*.[125] Schulz et al identified 288 eQTLs enriched in the hippocampus, suggesting the role of the hippocampus in the mechanism of SZ disease.[126] A cerebrospinal fluid analysis of 414 subjects conducted by Luykx et al revealed that MM-associated eQTL rs11628551 regulates the expression of *PDE9A*, involved in the transmission of monoaminergic signals, as well as major depressive disorder and antidepressant responses.[127]

A combined analysis of the Bayesian statistical framework and Mendelian randomization identified *TYW5* as an SZ risk gene and demonstrated that rs203772 was associated with a high expression of *TYW5* in the prefrontal cortex in the eQTL analysis.[128] A novel risk missense variant, rs1801311, was identified in 2021 to potentially contribute to SZ risk by regulating *NAGA* expression, causing the up-regulation of *NAGA* expression, and affecting the proliferation and differentiation of neural stem cells.[129] Furthermore, rs1801311 contributed to SZ by preventing the binding of *YY1*, *TAF1*, and *POLR2A*, requiring further in-

depth studies to explore the mechanism. In addition to tissue-derived eQTL, integrated cell-type-based analysis better elucidates the SZ risk genes. Wu et al identified 6108 cis-eQTL genes screened from eight types of brain cells and revealed nine SZ risk genes by analyzing eQTLs in inhibitory neurons.[130]

Discussion

eQTL as a bridge from the genome to the mechanism, allows investigation of the potential effects between genetic variations and complex traits. In this review, we provided an overview of eQTLs, including their development process, identification and analysis methods, data resources, and their contributions to four typical diseases.

The development of sequencing technology has generated abundant eQTL data, such that the research focus has changed from eQTL identification to its application. eQTL analysis can comprehensively explain the regulatory mechanisms involved in genomic markers, and therefore, it can be performed in multiple ways. For example, eQTL analysis can study the causal association between risk SNPs and phenotypes after GWAS. In addition, it can compare the regulatory differences in eQTLs in case—control samples. eQTL analysis can be combined with other omics data to perform network analysis and infer the regulatory relationships of multiple molecular biomarkers. The integrated analysis of eQTL and single-cell data can provide new clues to uncover the regulatory patterns underlying genetic variations in specific cell types.

Hundreds of eQTLs have been identified in several complex diseases to address the limitations of GWAS and allow causal inference between risk loci and diseases. However, eQTL analysis has certain limitations. Firstly, it is difficult to obtain SNP genotype and gene expression data from the same samples to conduct an eQTL analysis in public databases. Secondly, eQTLs identified in public databases could have regulatory differences in different populations (*e.g.*, cases and controls), which need to be further detected. Thirdly, more efficient tools are required to integrate eQTL and other omics data effectively.

The development of eQTL research has led to the application of its related theories and methods in the research between genetic variations and other biological elements. At least 30 other types of molecular QTLs have emerged, such as protein QTLs influencing protein abundance, methylation QTLs affecting DNA methylation patterns, splicing QTLs regulating alternative transcript isoforms, and metabolic QTLs modulating metabolite levels. While eQTLs primarily capture transcriptional regulation, these complementary QTL classes often map to distinct genomic loci and reveal post-transcriptional, epigenetic, or translational regulatory mechanisms. Notably, partial overlaps between eQTLs and other QTL types may occur when genetic variants exert pleiotropic effects across molecular layers. Integration of multi-omic QTL data enhances our capacity to construct causal regulatory networks and decipher the hierarchical relationships between DNA variation, molecular phenotypes, and organismal traits.[131,132] This systematic approach helps disentangle whether the observed expression changes drive downstream biological effects or merely correlate with other molecular events.

Spatial transcriptomics has revolutionized our ability to map gene expression patterns within intact tissue architectures, revealing how cellular microenvironments and spatial organization influence transcriptional regulation.[133] In this framework, eQTLs, genetic variants associated with gene expression variation, may exhibit spatial context-dependent effects that are obscured in bulk or single-cell RNA-seq analyses. For instance, spatially resolved eQTLs could reflect cell-type-specific regulatory mechanisms localized to distinct tissue regions (*e.g.*, tumor margins *vs.* core) or interactions with spatially varying factors such as hypoxia gradients or immune cell infiltration. Recent studies[134–136] have identified spatial eQTLs, where genetic effects on gene expression are modulated by positional information, suggesting that regulatory genetic variants may drive expression heterogeneity across tissue compartments. Integrating spatial transcriptomics with eQTL mapping could thus disentangle whether genetic regulation operates uniformly or is constrained by niche-specific molecular or cellular interactions. While technical challenges, including limited spatial resolution and statistical power, remain, advances in spatial omics and multivariate analytical frameworks (*e.g.*, tensor-based QTL mapping) promise to uncover how genetic variation shapes spatially organized gene networks.

eQTL studies have emerged as a cornerstone in deciphering the genetic architecture of gene regulation, bridging the gap between genomic variation and phenotypic diversity. Integrating eQTLs with GWAS data enables biomarker discovery and polygenic risk scoring, accelerating precision medicine. The comprehensive integration framework of eQTL and other omics data will bring a new perspective to the understanding of functional genomics in the future. sc-eQTL mapping resolves cellular heterogeneity, capturing transient or context-dependent regulatory effects obscured in bulk analyses. Innovations in scRNA-seq design, such as enriched rare-cell sequencing, enhance the detection of cell-type-specific eQTLs.[65,90] These advancements are pivotal for unraveling dynamic processes like immune responses and neuronal activation. Systematic analyses of lncRNA eQTLs will help uncover the potential of non-coding genomics in disease etiology. To sum up, the integration of cutting-edge technologies, diverse population datasets, and multi-omics approaches is revolutionizing eQTL research. These developments not only refine our understanding of genetic regulation but also pave the way for actionable insights into complex diseases, from cancer to neuropsychiatric disorders. As methodologies evolve and resources expand, eQTL studies will play a key role in unlocking the functional genome and delivering tailored therapeutic interventions.

CRediT authorship contribution statement

Zhe Jia: Writing — original draft. **Jing Xu:** Writing — original draft. **Yingnan Ma:** Writing — review & editing. **Siyu Wei:** Writing — review & editing. **Chen Sun:** Software, Resources. **Xingyu Chen:** Software, Conceptualization. **Jingxuan Kang:** Visualization, Conceptualization. **Haiyan Chen:**

Methodology, Data curation. **Chen Zhang:** Conceptualization, Formal analysis. **Yu Dong:** Software, Resources. **Junxian Tao:** Resources. **Xuying Guo:** Data curation. **Hongchao Lv:** Supervision, Data curation. **Guoping Tang:** Supervision, Writing — review & editing. **Yongshuai Jiang:** Writing — review & editing, Supervision, Funding acquisition. **Mingming Zhang:** Methodology, Writing — review & editing, Supervision, Funding acquisition.

Conflict of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work was supported by the Program for Young Talents of Basic Research in Universities of Heilongjiang Province, China (No. YQJH2023036), the National Natural Science Foundation of China (No. 31970651, 92046018, 81601422, 81600403), the Mathematical Tianyuan Fund of the National Natural Science Foundation of China (No. 12026414), the Joint Funds of the Zhejiang Provincial Natural Science Foundation of China (No. LBY24H170001), and the Marshal Initiative Funding of Harbin Medical University, Heilongjiang, China (No. HMUMIF-22010).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gendis.2025.101850.

References

1. Tanksley S. Mapping polygenes. *Annu Rev Genet*. 1993;27:205—233.
2. Farrall M. Quantitative genetic variation: a post-modern view. *Hum Mol Genet*. 2004;13(Spec No 1):R1—R7.
3. Zhang H, Wang Z, Qiao X, Wu J, Cheng C. Investigating potential drug targets for the treatment of glioblastoma: a Mendelian randomization study. *BMC Cancer*. 2025;25(1):654.
4. Wittkopp PJ. Genomic sources of regulatory variation in *cis* and in *trans*. *Cell Mol Life Sci*. 2005;62(16):1779—1783.
5. Gusev A, Ko A, Shi H, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet*. 2016;48(3):245—252.
6. Morrison J, Knoblauch N, Marcus JH, Stephens M, He X. Mendelian randomization accounting for correlated and uncorrelated pleiotropic effects using genome-wide summary statistics. *Nat Genet*. 2020;52(7):740—747.
7. Wills QF, Livak KJ, Tipping AJ, et al. Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments. *Nat Biotechnol*. 2013;31(8):748—752.
8. Zhang J, Zhao H. eQTL studies: from bulk tissues to single cells. *J Genet Genomics*. 2023;50(12):925—933.
9. Maria M, Pouyanfar N, Örd T, Kaikkonen MU. The power of single-cell RNA sequencing in eQTL discovery. *Genes*. 2022;13(3):502.
10. Yazar S, Alquicira-Hernandez J, Wing K, et al. Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science*. 2022;376(6589):eabf3041.
11. Nathan A, Asgari S, Ishigaki K, et al. Single-cell eQTL models reveal dynamic T cell state dependence of disease loci. *Nature*. 2022;606(7912):120—128.
12. Schadt EE, Monks SA, Drake TA, et al. Genetics of gene expression surveyed in maize, mouse and man. *Nature*. 2003;422(6929):297—302.
13. Rockman MV, Kruglyak L. Genetics of global gene expression. *Nat Rev Genet*. 2006;7(11):862—872.
14. International HapMap Consortium. The international HapMap project. *Nature*. 2003;426(6968):789—796.
15. 1000 Genomes Project Consortium, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68—74.
16. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3 (Bethesda)*. 2011;1(6):457—470.
17. Zou Y, Su Z, Huang W, Gu X. Histone modification pattern evolution after yeast gene duplication. *BMC Evol Biol*. 2012;12:111.
18. Võsa U, Claringbould A, Westra HJ, et al. Large-scale *cis*- and *trans*-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet*. 2021;53(9):1300—1310.
19. Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy: analysis of *Affymetrix GeneChip* data at the probe level. *Bioinformatics*. 2004;20(3):307—315.
20. a Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003;4(2):249—264;
b Irizarry RA. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003;4(2):249—264.
21. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics*. 2012;28(10):1353—1358.
22. Steinberg J, Southam L, Roumeliotis TI, et al. A molecular quantitative trait locus map for osteoarthritis. *Nat Commun*. 2021;12(1):1309.
23. Viñuela A, Varshney A, van de Bunt M, et al. Genetic variant effects on gene expression in human pancreatic islets and their implications for T2D. *Nat Commun*. 2020;11(1):4912.
24. Doss S, Schadt EE, Drake TA, Lusis AJ. *Cis*-acting expression quantitative trait loci in mice. *Genome Res*. 2005;15(5):681—691.
25. Ongen H, Buil A, Brown AA, Dermitzakis ET, Delaneau O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics*. 2016;32(10):1479—1485.
26. Delaneau O, Ongen H, Brown AA, Fort A, Panousis NI, Dermitzakis ET. A complete tool set for molecular QTL discovery and analysis. *Nat Commun*. 2017;8:15452.
27. Hu Y, Xi X, Yang Q, Zhang X. SCeQTL: an R package for identifying eQTL from single-cell parallel sequencing data. *BMC Bioinf*. 2020;21(1):184.
28. Ma T, Li H, Zhang X. Discovering single-cell eQTLs from scRNA-seq data only. *Gene*. 2022;829:146520.
29. Strober BJ, Tayeb K, Popp J, et al. SURGE uncovering context-specific genetic-regulation of gene expression from single-cell RNA sequencing using latent-factor models. *Genome Biol*. 2024;25(1):28.
30. Cuomo ASE, Heinen T, Vagiaki D, Horta D, Marioni JC, Stegle O. CellRegMap: a statistical framework for mapping context-specific regulatory variants using scRNA-seq. *Mol Syst Biol*. 2022;18(8):e10663.
31. Zhang T, Choi J, Kovacs MA, et al. Cell-type-specific eQTL of primary melanocytes facilitates identification of melanoma susceptibility genes. *Genome Res*. 2018;28(11):1621—1635.
32. Mandric I, Schwarz T, Majumdar A, et al. Optimized design of single-cell RNA sequencing experiments for cell-type-specific eQTL analysis. *Nat Commun*. 2020;11(1):5504.

33. Lee S, Xing EP. Leveraging input and output structures for joint mapping of epistatic and marginal eQTLs. *Bioinformatics*. 2012;28(12):i137—i146.
34. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet*. 2012;90(1):7—24.
35. Bycroft C, Freeman C, Petkova D, et al. The UK biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203—209.
36. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res*. 2014;42(D1):D1001—D1006.
37. MacArthur J, Bowler E, Cerezo M, et al. The new NHGRI-EBI catalog of published genome-wide association studies (GWAS catalog). *Nucleic Acids Res*. 2017;45(D1):D896—D901.
38. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019;47(D1):D1005—D1012.
39. Tian D, Wang P, Tang B, et al. GWAS atlas: a curated resource of genome-wide variant-trait associations in plants and animals. *Nucleic Acids Res*. 2020;48(D1):D927—D932.
40. Foley CN, Staley JR, Breen PG, et al. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat Commun*. 2021;12(1):764.
41. Cano-Gamez E, Trynka G. From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. *Front Genet*. 2020;11:424.
42. Hormozdiari F, van de Bunt M, Segrè AV, et al. Colocalization of GWAS and eQTL signals detects target genes. *Am J Hum Genet*. 2016;99(6):1245—1260.
43. Evans P, Nagai T, Konkashbaev A, Zhou D, Knapik EW, Gamazon ER. Transcriptome-wide association studies (TWAS): methodologies, applications, and challenges. *Curr Protoc*. 2024;4(2):e981.
44. Wainberg M, Sinnott-Armstrong N, Mancuso N, et al. Opportunities and challenges for transcriptome-wide association studies. *Nat Genet*. 2019;51(4):592—599.
45. Zhu Z, Zhang F, Hu H, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016;48(5):481—487.
46. Fang H, Knezevic B, Burnham KL, Knight JC. XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits. *Genome Med*. 2016;8(1):129.
47. Wheeler HE, Aquino-Michaels K, Gamazon ER, et al. Poly-omic prediction of complex traits: omickriging. *Genet Epidemiol*. 2014;38(5):402—415.
48. Jansen R. Genetical genomics: the added value from segregation. *Trends Genet*. 2001;17(7):388—391.
49. Kurian KM, Watson CJ, Wyllie AH. DNA chip technology. *J Pathol*. 1999;187(3):267—271.
50. Wei CW. Using a microfluidic device for 1 l DNA microarray hybridization in 500 s. *Nucleic Acids Res*. 2005;33(8):e78.
51. Thalayasingam N, Massey J, Anderson AE, et al. THU0003 $CD4^+$ and B lymphocyte expression quantitative traits at rheumatoid arthritis risk loci in untreated early arthritis: implications for causal gene identification? *Ann Rheum Dis*. 2017;76:200—201.
52. Amini H, Shroff N, Stamova B, et al. Genetic variation contributes to gene expression response in ischemic stroke: an eQTL study. *Ann Clin Transl Neurol*. 2020;7(9):1648—1660.
53. Nica AC, Dermitzakis ET. Expression quantitative trait loci: present and future. *Philos Trans R Soc B*. 2013;368(1620):20120362.
54. Pickrell JK, Marioni JC, Pai AA, et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature*. 2010;464(7289):768—772.
55. Majewski J, Pastinen T. The study of eQTL variations by RNA-seq: from SNPs to phenotypes. *Trends Genet*. 2011;27(2):72—79.
56. Chen Y, Zhu J, Lum PY, et al. Variations in DNA elucidate molecular networks that cause disease. *Nature*. 2008;452(7186):429—435.
57. Zhang P, Xia JH, Zhu J, et al. High-throughput screening of prostate cancer risk loci by single nucleotide polymorphisms sequencing. *Nat Commun*. 2018;9(1):2022.
58. Team TAR, Sanchez-Roige S, Fontanillas P, et al. Genome-wide association study of delay discounting in 23,217 adult research participants of European ancestry. *Nat Neurosci*. 2018;21(1):16—18.
59. Liu C. Brain expression quantitative trait locus mapping informs genetic studies of psychiatric diseases. *Neurosci Bull*. 2011;27(2):123—133.
60. Huang H, Fang M, Jostins L, et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature*. 2017;547(7662):173—178.
61. Franceschini N, Giambartolomei C, de Vries PS, et al. GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. *Nat Commun*. 2018;9(1):5141.
62. Gong J, Mei S, Liu C, et al. PancanQTL: systematic identification of *cis*-eQTLs and trans-eQTLs in 33 cancer types. *Nucleic Acids Res*. 2018;46(D1):D971—D976.
63. Li L, Zhang X, Zhao H. eQTL. *Methods Mol Biol*. 2012;871:265—279.
64. Mo XB, Wu LF, Lu X, et al. Detection of lncRNA-mRNA interaction modules by integrating eQTL with weighted gene co-expression network analysis. *Funct Integr Genom*. 2019;19(2):217—225.
65. Kang JB, Raveane A, Nathan A, Soranzo N, Raychaudhuri S. Methods and insights from single-cell expression quantitative trait loci. *Annu Rev Genom Hum Genet*. 2023;24:277—303.
66. Kang HM, Subramaniam M, Targ S, et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat Biotechnol*. 2018;36(1):89—94.
67. Study LC, van der Wijst MGP, Consortium B, et al. Single-cell RNA sequencing identifies celltype-specific *cis*-eQTLs and co-expression QTLs. *Nat Genet*. 2018;50(4):493—497.
68. Onuora S. Single-cell RNA sequencing sheds light on cell-type specific gene expression in immune cells. *Nat Rev Rheumatol*. 2022;18(7):363.
69. Perez RK, Gordon MG, Subramaniam M, et al. Single-cell RNA-seq reveals cell type-specific molecular and genetic associations to lupus. *Science*. 2022;376(6589):eabf1970.
70. Neavin D, Nguyen Q, Daniszewski MS, et al. Single cell eQTL analysis identifies cell type-specific genetic control of gene expression in fibroblasts and reprogrammed induced pluripotent stem cells. *Genome Biol*. 2021;22(1):76.
71. Lake BB, Chen S, Sos BC, et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol*. 2018;36(1):70—80.
72. Bryois J, Calini D, MacNair W, et al. Cell-type-specific *cis*-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. *Nat Neurosci*. 2022;25(8):1104—1112.
73. Jerber J, Seaton DD, Cuomo ASE, et al. Population-scale single-cell RNA-seq profiling across dopaminergic neuron differentiation. *Nat Genet*. 2021;53(3):304—312.
74. Schmid KT, Höllbacher B, Cruceanu C, et al. scPower accelerates and optimizes the design of multi-sample single cell transcriptomic studies. *Nat Commun*. 2021;12(1):6625.
75. van der Wijst M, de Vries DH, Groot HE, et al. The single-cell eQTLGen consortium. *eLife*. 2020;9:e52155.

76. Hicks SC, Townes FW, Teng M, Irizarry RA. Missing data and technical variability in single-cell RNA-sequencing experiments. *Biostatistics*. 2018;19(4):562—578.
77. Consortium GTEx. The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369 (6509):1318—1330.
78. Burgess DJ. Reaching completion for GTEx. *Nat Rev Genet*. 2020;21(12):717.
79. Yang CP, Li X, Wu Y, et al. Comprehensive integrative analyses identify *GLT8D1* and *CSNK2B* as schizophrenia risk genes. *Nat Commun*. 2018;9(1):838.
80. Kundu K, Tardaguila M, Mann AL, et al. Genetic associations at regulatory phenotypes improve fine-mapping of causal variants for 12 immune-mediated diseases. *Nat Genet*. 2022;54 (3):251—262.
81. Liang L, Morar N, Dixon AL, et al. A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. *Genome Res*. 2013;23(4):716—726.
82. Ke J, Peng X, Mei S, et al. Evaluation of polymorphisms in microRNA-binding sites and pancreatic cancer risk in Chinese population. *J Cell Mol Med*. 2020;24(3):2252—2259.
83. Yang Y, Musco H, Simpson-Yap S, et al. Investigating the shared genetic architecture between multiple sclerosis and inflammatory bowel diseases. *Nat Commun*. 2021;12(1):5641.
84. Kerimov N, Hayhurst JD, Peikova K, et al. A compendium of uniformly processed human gene expression and splicing quantitative trait loci. *Nat Genet*. 2021;53(9):1290—1299.
85. Koel M, Võsa U, Jõeloo M, et al. GWAS meta-analyses clarify the genetics of cervical phenotypes and inform risk stratification for cervical cancer. *Hum Mol Genet*. 2023;32(12): 2103—2116.
86. Chun S, Akle S, Teodosiadis A, et al. Leveraging pleiotropy to discover and interpret GWAS results for sleep-associated traits. *PLoS Genet*. 2022;18(12):e1010557.
87. Kerimov N, Tambets R, Hayhurst JD, et al. eQTL catalogue 2023: new datasets, X chromosome QTLs, and improved detection and visualisation of transcript-level QTLs. *PLoS Genet*. 2023;19(9):e1010932.
88. Demela P, Pirastu N, Soskic B. Cross-disorder genetic analysis of immune diseases reveals distinct gene associations that converge on common pathways. *Nat Commun*. 2023;14(1):2743.
89. Kang JB, Shen AZ, Sakaue S, et al. Mapping the dynamic genetic regulatory architecture of HLA genes at single-cell resolution. *medRxiv*. 2023;2023, 03.14.23287257.
90. Ding R, Wang Q, Gong L, et al. scQTLbase: an integrated human single-cell eQTL database. *Nucleic Acids Res*. 2024;52 (D1):D1010—D1017.
91. Zhou Z, Du J, Wang J, et al. SingleQ: a comprehensive database of single-cell expression quantitative trait loci (sc-eQTLs) cross human tissues. *Database*. 2024;2024:baae010.
92. Carrot-Zhang J, Chambwe N, Damrauer JS, et al. Comprehensive analysis of genetic ancestry and its molecular correlates in cancer. *Cancer Cell*. 2020;37(5):639—654.e6.
93. Chen C, Liu Y, Luo M, et al. PancanQTLv2.0: a comprehensive resource for expression quantitative trait loci across human cancers. *Nucleic Acids Res*. 2024;52(D1):D1400—D1406.
94. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet*. 2013;45(10):1238—1243.
95. Ratnapriya R, Sosina OA, Starostik MR, et al. Retinal transcriptome and eQTL analyses identify genes associated with age-related macular degeneration. *Nat Genet*. 2019;51(4): 606—610.
96. Ota M, Nagafuchi Y, Hatano H, et al. Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell*. 2021;184(11):3006—3021.e17.
97. Di Matteo A, Bathon JM, Emery P. Rheumatoid arthritis. *Lancet*. 2023;402(10416):2019—2033.
98. Alamanos Y, Drosos AA. Epidemiology of adult rheumatoid arthritis. *Autoimmun Rev*. 2005;4(3):130—136.
99. MacGregor AJ, Snieder H, Rigby AS, et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum*. 2000;43(1):30—37.
100. Suzuki A, Yamada R, Chang X, et al. Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat Genet*. 2003;34(4):395—402.
101. Jiang F, Hu SY, Tian W, et al. A landscape of gene expression regulation for synovium in arthritis. *Nat Commun*. 2024;15(1): 1409.
102. Ni J, Wang P, Yin KJ, et al. Novel insight into the aetiology of rheumatoid arthritis gained by a cross-tissue transcriptome-wide association study. *RMD Open*. 2022;8(2):e002529.
103. Naranbhai V, Fairfax BP, Makino S, et al. Genomic modulators of gene expression in human neutrophils. *Nat Commun*. 2015; 6:7545.
104. Ishigaki K, Kochi Y, Suzuki A, et al. Polygenic burdens on cell-specific pathways underlie the risk of rheumatoid arthritis. *Nat Genet*. 2017;49(7):1120—1125.
105. Mu Z, Wei W, Fair B, Miao J, Zhu P, Li YI. The impact of cell type and context-dependent regulatory variants on human immune traits. *Genome Biol*. 2021;22(1):122.
106. Zhang T, Chen S, Syed I, et al. A LC-MS-based workflow for measurement of branched fatty acid esters of hydroxy fatty acids. *Nat Protoc*. 2016;11(4):747—763.
107. Keildson S, Fadista J, Ladenvall C, et al. Expression of phosphofructokinase in skeletal muscle is influenced by genetic variation and associated with insulin sensitivity. *Diabetes*. 2014;63(3):1154—1165.
108. Varshney A, Scott LJ, Welch RP, et al. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proc Natl Acad Sci U S A*. 2017;114(9):2301—2306.
109. Khamis A, Canouil M, Siddiq A, et al. Laser capture microdissection of human pancreatic islets reveals novel eQTLs associated with type 2 diabetes. *Mol Metabol*. 2019;24: 98—107.
110. Mulder H. Transcribing β-cell mitochondria in health and disease. *Mol Metabol*. 2017;6(9):1040—1051.
111. Koeck T, Olsson AH, Nitert MD, et al. A common variant in TFB1M is associated with reduced insulin secretion and increased future risk of type 2 diabetes. *Cell Metab*. 2011;13 (1):80—91.
112. Broadaway KA, Yin X, Williamson A, et al. Loci for insulin processing and secretion provide insight into type 2 diabetes risk. *Am J Hum Genet*. 2023;110(2):284—299.
113. Maina JG, Balkhiyarova Z, Nouwen A, et al. Bidirectional mendelian randomization and multiphenotype GWAS show causality and shared pathophysiology between depression and type 2 diabetes. *Diabetes Care*. 2023;46(9):1707—1714.
114. Tang X, Shi L, Xie N, et al. SIRT7 antagonizes TGF-β signaling and inhibits breast cancer metastasis. *Nat Commun*. 2017;8 (1):318.
115. Escala-Garcia M, Abraham J, Andrulis IL, et al. A network analysis to identify mediators of germline-driven differences in breast cancer prognosis. *Nat Commun*. 2020;11(1):312.
116. Li Q, Seo JH, Stranger B, et al. Integrative eQTL-based analyses reveal the biology of breast cancer risk loci. *Cell*. 2013; 152(3):633—641.
117. Chen Y, Olopade OI. MYC in breast tumor progression. *Expert Rev Anticancer Ther*. 2008;8(10):1689—1698.
118. Yu F, Li J, Chen H, et al. Kruppel-like factor 4 (KLF4) is required for maintenance of breast cancer stem cells and for cell migration and invasion. *Oncogene*. 2011;30(18): 2161—2172.
119. Guo X, Lin W, Bao J, et al. A comprehensive *cis*-eQTL analysis revealed target genes in breast cancer susceptibility loci

identified in genome-wide association studies. *Am J Hum Genet*. 2018;102(5):890—903.
120. Ren N, Li Y, Xiong Y, Li P, Ren Y, Huang Q. Functional screenings identify regulatory variants associated with breast cancer susceptibility. *Curr Issues Mol Biol*. 2021;43(3):1756—1777.
121. Rezaie N, Bayati M, Hamidi M, et al. Somatic point mutations are enriched in non-coding RNAs with possible regulatory function in breast cancer. *Commun Biol*. 2022;5(1):556.
122. Shidal C, Shu X, Wu J, et al. Functional genomic analyses of the 21q22.3 locus identifying functional variants and candidate gene *YBEY* for breast cancer risk. *Cancers*. 2021;13(9):2037.
123. Sullivan PF, Daly MJ, O'Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat Rev Genet*. 2012;13(8):537—551.
124. Ma C, Gu C, Huo Y, Li X, Luo XJ. The integrated landscape of causal genes and pathways in schizophrenia. *Transl Psychiatry*. 2018;8(1):67.
125. Walker RL, Ramaswami G, Hartl C, et al. Genetic control of expression and splicing in developing human brain informs disease mechanisms. *Cell*. 2019;179(3):750—771.e22.
126. Schulz H, Ruppert AK, Herms S, et al. Genome-wide mapping of genetic determinants influencing DNA methylation and gene expression in human hippocampus. *Nat Commun*. 2017;8(1):1511.
127. Luykx JJ, Bakker SC, Lentjes E, et al. Genome-wide association study of monoamine metabolite levels in human cerebrospinal fluid. *Mol Psychiatr*. 2014;19(2):228—234.
128. Zhang C, Li X, Zhao L, et al. Comprehensive and integrative analyses identify *TYW5* as a schizophrenia risk gene. *BMC Med*. 2022;20(1):169.
129. Li Y, Ma C, Li W, et al. A missense variant in NDUFA6 confers schizophrenia risk by affecting YY1 binding and NAGA expression. *Mol Psychiatr*. 2021;26(11):6896—6911.
130. Wu Y, Zhang CY, Wang L, Li Y, Xiao X. Genetic insights of schizophrenia via single cell RNA-sequencing analyses. *Schizophr Bull*. 2023;49(4):914—922.
131. Zhang Z, Fang T, Chen L, Ji F, Chen J. Multi-omics Mendelian randomization integrating GWAS eQTL, and mQTL data identified genes associated with breast cancer. *Am J Cancer Res*. 2024;14(3):1433—1445.
132. Chen J, Ruan X, Sun Y, et al. Multi-omic insight into the molecular networks of mitochondrial dysfunction in the pathogenesis of inflammatory bowel disease. *EBioMedicine*. 2024;99:104934.
133. Chen C, Cai Y, Hu W, et al. Single-cell eQTL mapping reveals cell subtype-specific genetic control and mechanism in malignant transformation of colorectal cancer. *Cancer Discov*. 2025:752153.
134. Pudjihartono N, Ho D, Golovina E, Fadason T, Kempa-Liehr AW, O'Sullivan JM. Juvenile idiopathic arthritis-associated genetic loci exhibit spatially constrained gene regulatory effects across multiple tissues and immune cell types. *J Autoimmun*. 2023;138:103046.
135. Zaied RE, Gokuladhas S, Walker C, O'Sullivan JM. Unspecified asthma, childhood-onset, and adult-onset asthma have different causal genes: a Mendelian randomization analysis. *Front Immunol*. 2024;15:1412032.
136. Pudjihartono N, Ho D, O'Sullivan JM. Integrative analysis reveals novel insights into juvenile idiopathic arthritis pathogenesis and shared molecular pathways with associated traits. *Front Genet*. 2024;15:1448363.