D1122–D1128 *Nucleic Acids Research, 2023, Vol. 51, Database issue* *Published online 4 November 2022*
https://doi.org/10.1093/nar/gkac1020

QTLbase2: an enhanced catalog of human quantitative trait loci on extensive molecular phenotypes

Dandan Huang[1,†], Xiangling Feng[2,†], Hongxi Yang[2], Jianhua Wang[2], Wenwen Zhang[2], Xutong Fan[2], Xiaobao Dong[3], Kexin Chen[3], Ying Yu[2], Xin Ma[1,*], Xianfu Yi[3,*] and Mulin Jun Li[2,3,*]

[1]Wuxi School of Medicine, Jiangnan University, Wuxi, China, [2]Department of Pharmacology, The Province and Ministry Co-sponsored Collaborative Innovation Center for Medical Epigenetics, Tianjin Key Laboratory of Inflammation Biology, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China and [3]Department of Bioinformatics, Tianjin Key Laboratory of Molecular Cancer Epidemiology, Tianjin Medical University Cancer Institute and Hospital, Tianjin Medical University, Tianjin, China

Received September 09, 2022; Revised October 17, 2022; Editorial Decision October 18, 2022; Accepted October 21, 2022

ABSTRACT

Deciphering the fine-scale molecular mechanisms that shape the genetic effects at disease-associated loci from genome-wide association studies (GWAS) remains challenging. The key avenue is to identify the essential molecular phenotypes that mediate the causal variant and disease under particular biological conditions. Therefore, integrating GWAS signals with context-specific quantitative trait loci (QTLs) (such as different tissue/cell types, disease states, and perturbations) from extensive molecular phenotypes would present important strategies for full understanding of disease genetics. Via persistent curation and systematic data processing of large-scale human molecular trait QTLs (xQTLs), we updated our previous QTLbase database (now QTLbase2, http://mulinlab.org/qtlbase) to comprehensively analyze and visualize context-specific QTLs across 22 molecular phenotypes and over 95 tissue/cell types. Overall, the resource features the following major updates and novel functions: (i) 960 more genome-wide QTL summary statistics from 146 independent studies; (ii) new data for 10 previously uncompiled QTL types; (iii) variant query scope expanded to fit 195 QTL datasets based on whole-genome sequencing; (iv) supports filtering and comparison of QTLs for different biological conditions, such as stimulation types and disease states; (v) a new linkage disequilibrium viewer to facilitate variant prioritization across tissue/cell types and QTL types.

INTRODUCTION

Exploring the molecular basis underlying genotype–phenotype association is critical to fully understand disease pathogenesis and trait evolution. Many large-scale quantitative trait locus (QTL) studies have been performed in the past decade to interrogate the genetic effect on diverse molecular phenotypes (1). To date, >20 fine-grained molecular traits have been profiled at cohort level and numerous molecular trait QTLs (xQTLs) were identified in different tissue/cell types (2,3). These intensive efforts have substantially facilitated the interpretation of disease/trait-associated variants revealed by genome-wide association studies (GWAS) and guided functional follow-ups for dissecting the molecular mechanisms of disease/trait development (4).

Despite the achievements in genome-wide human xQTL profiling, there are several persistent challenges regarding xQTL discovery and interpretation. First, most molecular traits and their regulation are context-specific; therefore, the transmission of genetic effects varies in different biological conditions. In recent years, researchers have applied multiple biological models (such as specific disease status, cell development process, drug treatment, or cytokine stimulation) to investigate dynamic QTLs and their allele-specific effects (5). However, no resource specifically integrates such context-specific information. Second, both empirical and experimental evidence demonstrated that multiple *cis*-QTLs or numerous genetic variants commonly regulate gene expression (6,7), suggesting the great difficulty of causal xQTL fine-mapping in high linkage disequilibrium (LD). Incorporating whole-genome sequencing (WGS)-based xQTL mapping and multi-dimensional functional annotations would benefit causal variant identifica-

*To whom correspondence should be addressed. Tel: +86 22 83336668; Fax: +86 22 83336668; Email: mulin0424.li@gmail.com
Correspondence may also be addressed to Xianfu Yi. Email: yixianfu@tmu.edu.cn
Correspondence may also be addressed to Xin Ma. Email: maxin@jiangnan.edu.cn
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

tion. Third, emerging xQTL studies indicated that novel omics traits can mediate the effect from QTL to conventional traits, for example, gene expression QTLs (eQTLs) have been frequently reported to exert their genetic effects by first affecting chromatin accessibility (8), histone modification (9), DNA methylation (10), alternative polyadenylation (11) or N6-methyladenosine (12,13). This emphasizes the complexity of genetic effects on cooperative molecular phenotypes and the necessity of studying their genetic pleiotropy. Altogether, there is an urgent need to develop more comprehensive resources for in-depth xQTL characterization, comparison, and interpretation.

The first version of QTLbase provided a unique resource for biologists and geneticists to search and visualize tissue/cell type-specific xQTLs in a tissue-, phenome-, and variant-specific manner. Researchers have leveraged the database to investigate putative molecular connections with many human complex traits, such as common disease causality (14,15), COVID-19 severity (16) and trait evolution (17). Here, we present an updated QTLbase2 database that incorporates more extensive QTL types and advanced functions. QTLbase2 incorporates 1681 new genome-wide QTL summary statistics from 377 independent studies and 22 molecular phenotypes. It is also equipped with many highly interactive web functions to ease xQTL filtering, comparison, and visualization. The QTLbase2 is free and open to all users without login requirement at http://mulinlab.org/qtlbase.

MATERIALS AND METHODS

Summary of data curation, processing and normalization

Consistent with the previous version, i.e. QTLbase, we manually curated and filtered human molecular phenotype QTL studies by keyword-matching from Google Scholar. We discarded QTL studies that did not report genome-wide significant variants and associated summary information. QTL studies containing multiple independent QTL mapping results were split into separate datasets with unique IDs. As many QTLs were identified in different biological conditions, we labelled the context-specific QTL datasets with different condition types. We assigned the QTL datasets to 95 standardized tissue/cell types and five human superpopulations. The previous versions of some QTL datasets were upgraded to the most recent version, for example, Genotype-Tissue Expression (GTEx) V7 was replaced with GTEx V8 (18). Variants and molecular traits were normalized as described previously (2). For new traits, such as alternative polyadenylation and N6-methyladenosine, the trait name was standardized across different QTL studies. Furthermore, the genomic positions of both variants and molecular traits were uniformly mapped to the GRCh38/hg38 human genome assembly in this update using LiftOver (19).

Compilation of WGS QTLs

Given the high volume of significant variants and associated summary statistics from WGS-based QTL studies, QTLbase only retained the significant variants tagged at the Infinium HumanOmni2.5 BeadChip for those WGS-based QTLs. Here, by optimizing the data indexing structure of QTL variants and associated summary information using VarNote (20) and NoSQL, we were able to fully accommodate WGS-based xQTL datasets, such as GTEx (18), Geuvadis (21) and BLUEPRINT (22), for more efficient data retrieval.

Integration of condition-specific QTLs

Recent QTL studies have extensively observed the context-specific genetic effects on different molecular phenotypes. In QTLbase2, we comprehensively curated QTL mapping results under various biological conditions and uniformly categorized these conditions into four major types: disease, stimulation, differentiation and drug treatment. The original condition description was extracted from the publication and then mapped to corresponding categories. We also labelled the QTL datasets in each condition group for intuitive visualization.

Calculation of LD across populations

To benefit the identification of likely causal variants and colocalized signals among correlated alleles in the LD region of specific human populations, we leveraged the 1000 Genomes Project phase 3 genotypes (23) in five human super-populations (AFR, AMR, EAS, EUR, SAS) to estimate LD information. Implementing an LD calculation program using the VarNote index structure and bit-encoding compression (http://www.mulinlab.org/varnote/utilsq.html?f=ld) resulted in greatly increased computational efficiency for real-time queries.

Linking variant to its associated genes

All candidate genes associated with each query variant are reported in four ways: (i) the host gene(s) whose genomic regions overlay the query variant, (ii) the gene(s) closest to the query variant within 1 Mb, (iii) the QTL gene(s) associated with the query variant, (iv) the interacted gene(s) supported by 5 kb Hi-C interactions of 60 tissues/cell types in GWAS4D (24).

RESULTS

New data types and features

Given the rapid progress in the QTL field regarding novel molecular phenotype characterization and context-specific genetic effects in different biological conditions, QTLbase2 presents a timely introduction of five features to enhance the database content and retrieval function (Figure 1). First, it includes the genetic associations for 10 new molecular phenotypes: alternative polyadenylation (apaQTL), mRNA N6-methyladenosine (m6AQTL), circular RNA expression (circQTL), transcription factor (TF) binding (bQTL), promoter usage (puQTL), transcript usage (tuQTL), enhancer activity (eaQTL), mRNA stability (stQTL), promoter interaction expression (pieQTL), and gene expression variance (vQTL). These expanded xQTLs will facilitate investigations of the potential cascade effect of functional variants.

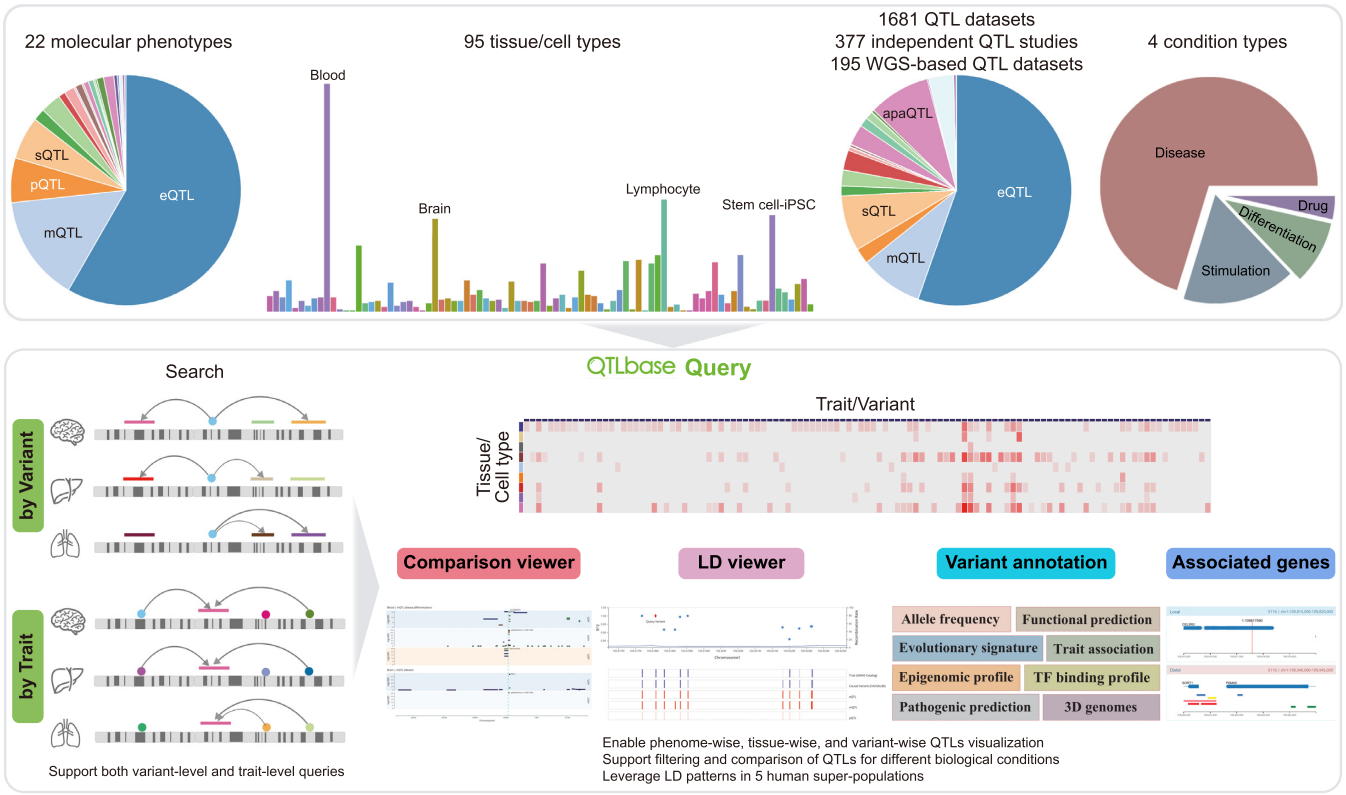

Figure 1. The database structure and newly added functions in QTLbase2.

Second, QTL mapping based on WGS variants can maximize the fine-mapping power of causal variants, but the high volume of tested variants across the whole genome limited the query efficacy and display burden in QTLbase. By optimizing a backend database storage structure and frontend web components, QTLbase2 incorporates 195 WGS-based xQTL datasets (67.7% from eQTL studies), such as that from GTEx (18), Geuvadis (21), BLUEPRINT (22) and PCAWG (Pan-Cancer Analysis of Whole Genomes) (25). Third, to assist the interrogation of context-specific genetic effects in dynamic biological processes, QTLbase2 comprehensively integrates 434 xQTL datasets with different conditions, including 305 on disease states, 73 with stimulations, 41 under differentiation processes and 15 through drug treatment (Supplementary Figure S1). It also allows users to filter and compare these context-specific QTLs in a highly interactive webpage. Fourth, by leveraging LD patterns in different human super-populations, QTLbase2 supports the investigation of likely causal QTL variants within LD across various molecular phenotypes and different biological conditions. Finally, to systematically annotate and evaluate the potential biological functions and target genes of each query variant, QTLbase2 incorporates 40 genome-wide variant annotations in VannoPortal (26) and Hi-C interactions of 60 tissue/cell types in GWAS4D (24).

QTLbase2 data statistics

Based on the August 2022 version, 960 new genome-wide xQTL summary statistics from 146 independent studies were collected and curated in QTLbase2. Therefore, QTLbase2 now includes a total of 377 independent QTL studies involving 1681 QTL summary statistics across 22 molecular phenotypes (Supplementary Figures S2 and S3), 95 tissue/cell types (Supplementary Figure S4) and 178 biological conditions. Compared to QTLbase, QTLbase2 has compiled nearly 60% more QTL studies and an additional 10 molecular phenotypes, where most were derived from publications in the past three years (Table 1). This indicates an accelerated course of profiling the genetic effects on various molecular phenotypes to aid the interpretation of disease/trait-causal variants in the post-GWAS era. Moreover, QTLbase2 involves approximately 18.1% or 7.7% of QTL studies performed on disease-associated or -perturbed tissue/cell types, respectively, suggesting that investigating context-specific and dynamic biological processes may boost the QTL discovery power. Notably, most of the 10 newly added QTL types are associated with RNA phenotypes in transcriptional and post-transcriptional processes, which greatly expands the scope of deciphering genetic effects on the fine-grained molecular phenotypes of gene expression. Similar to the previous version, most xQTLs were identified in blood- and brain-derived tissue/cell types (20.6% and 18.1%, respectively) and conducted on European populations (47.1%) (Supplementary Figure S5).

Illustration of use cases and new functions

QTL investigations across different biological conditions. We used several reported cases to illustrate the advances of

Table 1. Feature comparisons between the two major versions of QTLbase (up to the August 2022)

Features	QTLbase	QTLbase2
Number of QTL types	13 molecular phenotypes	22 molecular phenotypes
Number of independent QTL studies	233	377
Number of QTL datasets	712	1681
Number of WGS-based QTL datasets	66 (with restriction)	195
Number of tissue/cell types	78 tissue/cell types	95 tissue/cell types
Biological conditions	N/A	305 on disease state, 41 under differentiation, 73 with stimulations, and 15 through drug treatment
LD support	N/A	LD for 5 human super-populations
Functional annotations	20 internal annotations	40 annotations in VannoPortal (26)
Interactive graphical visualizations	Phenome-wise, tissue-wise and variant-wise visualizations	Phenome-wise, tissue-wise and variant-wise visualizations; xQTL comparison viewer; LD viewer; associated gene viewer

QTLbase2 for identifying the context-specific genetic effect on molecular traits. First, the variant rs2275888 represents a stimulus-specific *cis*-eQTL for *IFNB1* in monocytes after 2-h lipopolysaccharide (LPS) treatment (5). In QTLbase2, selecting the 'Filter by condition' option enables specific inspection of the QTL evidence under different conditions. Consistent with the literature, rs2275888 receives a context-specific eQTL associated with *IFNB1* expression in monocytes under 2-h LPS exposure, but not at naïve status or after 24-h stimulation (Supplementary Figure S6). Moreover, users can investigate and compare the detailed QTL information and affected molecular traits in QTLbase2 'Comparison Viewer' (Figure 2A). Another example involves genetic variants that demonstrate stimulus-specific genetic effects on gene expression under statin treatment (27). In alignment with the reported data, QTLbase2 demonstrates that rs61396151 is a weak simvastatin treatment-responsive eQTL associated with *SNX14* expression in lymphocytes from the AFR population (Figure 2B and Supplementary Figure S7).

Visualization and comparison of context-specific QTLs in LD. We applied a previously reported QTL variant to demonstrate the usability of QTLbase2 LD Viewer for investigating the likely causal and multi-trait effects of xQTL in its LD region. The variant rs3763469 affected its enclosed chromatin accessibility and regulated *COL1A2* gene expression (8). Searching rs3763469 in QTLbase2 revealed that it could be associated with other molecular traits in lymphocytes, including TF binding, DNA methylation, and histone modifications, yet the causality of this variant on these molecular phenotypes is unknown. LD Viewer revealed that rs3763469 demonstrates a moderate correlation with other variants in LD ($r^2 < 0.8$ on five human super-populations) (Figure 3A). The xQTL tracks also indicate that the variants linked with rs3763469 obtain less genetic evidence, suggesting a higher casual probability of rs3763469. The intuitive visualization through QTLbase2 Comparison Viewer further demonstrates that this variant is likely to affect the fine-grained molecular phenotypes in lymphocytes, including the DNA methylation CpG sites at the *COL1A2* promoter (28), histone modifications on H3K4me1 (29), and TF binding on *POU2F1* (30) (Figure 3B).

Annotation of common and rare functional variants. QTLbase2 has improved variant interpretation ability and annotation scope over QTLbase. For example, rs12041331 was associated with platelet reactivity and cardiovascular disease by affecting the allele-specific effect on DNA methylation of the *PEAR1* gene (31), which was supported by many pieces of QTLbase-based evidence on chromatin-associated QTLs and TF binding. By incorporating large-scale functional annotations in VannoPortal (26) and integrative target gene information, QTLbase2 provides more biological insights into the rs12041331 regulatory mechanism. This intronic variant demonstrates significant allelic imbalance of open chromatin in CD14+ monocytes (32) and obtains a high genome-scale pathogenicity prediction score [e.g. Eigen-PC Phred score (33) = 15], suggesting its potential chromatin-modulating function in addition to affecting DNA methylation. By linking rs12041331 to candidate genes based on different strategies, QTLbase2 reports many associated genes by QTL trait analysis and several potential target genes by Hi-C interaction evidence in blood cells, such as *PEAR1* and *PRCC* (Supplementary Figure S8). Furthermore, QTLbase2 facilitates the functional interpretation of rare variants revealed by WGS-based association analysis. For example, by applying GTEx V8 RNA-seq and WGS data, the analysis of genetically driven transcriptome abnormalities revealed many functional large-effect rare variants (34). A rare 5'UTR exonic variant rs200388505 in the *FAAH* gene previously associated with pain sensitivity (35) contributes to extreme gene expression patterns in two individuals. By querying QTLbase2, rs200388505 was documented to significantly affect surrounding chromatin accessibility in lymphocytes from a QTL fine-mapping study (36) (Supplementary Figure S9) and was annotated with many active transcriptional signals in VannoPortal, which reveals its strong regulatory potential and possibly causal effect on complex traits.

CONCLUSIONS

QTLbase provided scientists with a comprehensive, high-quality, fast-response and interactive biological database for studying genetic effects on widespread molecular phenotypes in recent years. Given the increasing number of QTL studies on fine-grained molecular traits and various biological contexts, QTLbase2 involves the timely curation

Figure 2. Investigating eQTLs across different biological conditions in QTLbase2 'Comparison Viewer'. (**A**) rs2275888 displays a stimulus-specific genetic effect on *IFNB1* expression in monocytes only after 2-h LPS treatment; (**B**) rs61396151 exhibits a weak context-specific genetic effect associated with *SNX14* expression in lymphocytes after simvastatin treatment.

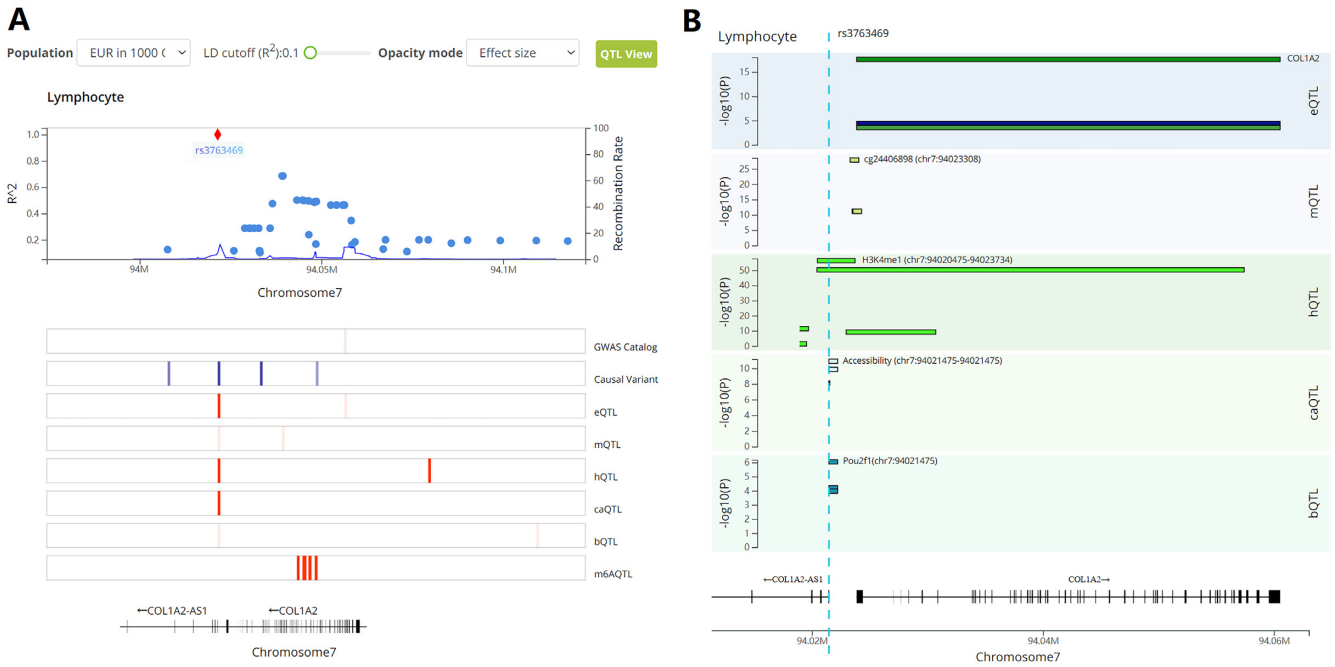

Figure 3. Comparing and inspecting highly linked xQTLs in QTLbase2 'LD Viewer' and 'Comparison Viewer'. (**A**) rs3763469 obtains xQTLs evidence on multiple molecular phenotypes and shows moderate correlation with other variants in its LD region; (**B**) rs3763469 is associated with several regulatory molecular traits in its proximal region.

and integration of these fresh data on a one-stop platform. The QTLbase2 database also introduces several advanced functions to compare and screen QTLs according to trait and tissue/cell types and biological conditions, which significantly facilitates the discovery of true causal variants for functional follow-up (37–39). As xQTL studies at the single-cell level (40), somatic mutation angle (25) and translational medicine (41,42) are evolving quickly, QTLbase2 will be constantly updated to benefit research interrogation of the context-specific genetic mechanism and create more significant effects in the era of human genetics and functional genomics.

DATA AVAILABILITY

QTLbase2 is free and open to all users without login requirement at http://mulinlab.org/qtlbase. No new data were generated or analysed in support of this research.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

Chinese National Key Research and Development Project [2021YFC2500403]; National Natural Science Foundation of China [32270717 to M.J.L. and 81622007 to X.M.]; Natural Science Foundation of Tianjin [19JCJQJC63600 to M.J.L. and 19JCQNJC09000 to X.Y.]. Funding for open access charge: National Natural Science Foundation of China [32270717 to M.J.L.].

REFERENCES

1. Ye,Y., Zhang,Z., Liu,Y., Diao,L. and Han,L. (2020) A multi-omics perspective of quantitative trait loci in precision medicine. *Trends Genet.*, **36**, 318–336.
2. Zheng,Z., Huang,D., Wang,J., Zhao,K., Zhou,Y., Guo,Z., Zhai,S., Xu,H., Cui,H., Yao,H. *et al.* (2020) QTLbase: an integrative resource for quantitative trait loci across multiple human molecular phenotypes. *Nucleic Acids Res.*, **48**, D983–D991.
3. Vandiedonck,C. (2018) Genetic association of molecular traits: a help to identify causative variants in complex diseases. *Clin. Genet.*, **93**, 520–532.
4. Neumeyer,S., Hemani,G. and Zeggini,E. (2020) Strengthening causal inference for complex disease using molecular quantitative trait loci. *Trends Mol. Med.*, **26**, 232–241.
5. Fairfax,B.P., Humburg,P., Makino,S., Naranbhai,V., Wong,D., Lau,E., Jostins,L., Plant,K., Andrews,R., McGee,C. *et al.* (2014) Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science*, **343**, 1246949.
6. Zeng,B., Lloyd-Jones,L.R., Holloway,A., Marigorta,U.M., Metspalu,A., Montgomery,G.W., Esko,T., Brigham,K.L., Quyyumi,A.A., Idaghdour,Y. *et al.* (2017) Constraints on eQTL fine mapping in the presence of multisite local regulation of gene expression. *G3 (Bethesda)*, **7**, 2533–2544.
7. Abell,N.S., DeGorter,M.K., Gloudemans,M.J., Greenwald,E., Smith,K.S., He,Z. and Montgomery,S.B. (2022) Multiple causal variants underlie genetic associations in humans. *Science*, **375**, 1247–1254.
8. Kumasaka,N., Knights,A.J. and Gaffney,D.J. (2016) Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nat. Genet.*, **48**, 206–213.
9. Ng,B., White,C.C., Klein,H.U., Sieberts,S.K., McCabe,C., Patrick,E., Xu,J., Yu,L., Gaiteri,C., Bennett,D.A. *et al.* (2017) An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.*, **20**, 1418–1426.
10. Taylor,D.L., Jackson,A.U., Narisu,N., Hemani,G., Erdos,M.R., Chines,P.S., Swift,A., Idol,J., Didion,J.P., Welch,R.P. *et al.* (2019) Integrative analysis of gene expression, DNA methylation, physiological traits, and genetic variation in human skeletal muscle. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 10883–10888.

11. Li,L., Huang,K.L., Gao,Y., Cui,Y., Wang,G., Elrod,N.D., Li,Y., Chen,Y.E., Ji,P., Peng,F. *et al.* (2021) An atlas of alternative polyadenylation quantitative trait loci contributing to complex trait and disease heritability. *Nat. Genet.*, **53**, 994–1005.
12. Zhang,Z., Luo,K., Zou,Z., Qiu,M., Tian,J., Sieh,L., Shi,H., Zou,Y., Wang,G., Morrison,J. *et al.* (2020) Genetic analyses support the contribution of mRNA N(6)-methyladenosine (m(6)A) modification to human disease heritability. *Nat. Genet.*, **52**, 939–949.
13. GTEx Consortium, Xiong,X., Hou,L., Park,Y.P., Molinie,B., Gregory,R.I. and Kellis,M. (2021) Genetic drivers of m(6)A methylation in human brain, lung, heart and muscle. *Nat. Genet.*, **53**, 1156–1165.
14. Stein,M.B., Levey,D.F., Cheng,Z., Wendt,F.R., Harrington,K., Pathak,G.A., Cho,K., Quaden,R., Radhakrishnan,K., Girgenti,M.J. *et al.* (2021) Genome-wide association analyses of post-traumatic stress disorder and its symptom subdomains in the million veteran program. *Nat. Genet.*, **53**, 174–184.
15. Galata,G., Garcia-Montero,A.C., Kristensen,T., Dawoud,A.A.Z., Munoz-Gonzalez,J.I., Meggendorfer,M., Guglielmelli,P., Hoade,Y., Alvarez-Twose,I., Gieger,C. *et al.* (2021) Genome-wide association study identifies novel susceptibility loci for KIT D816V positive mastocytosis. *Am. J. Hum. Genet.*, **108**, 284–294.
16. Li,Y., Ke,Y., Xia,X., Wang,Y., Cheng,F., Liu,X., Jin,X., Li,B., Xie,C., Liu,S. *et al.* (2021) Genome-wide association study of COVID-19 severity among the chinese population. *Cell Discov*, **7**, 76.
17. Quan,C., Li,Y., Liu,X., Wang,Y., Ping,J., Lu,Y. and Zhou,G. (2021) Characterization of structural variation in tibetans reveals new evidence of high-altitude adaptation and introgression. *Genome Biol.*, **22**, 159.
18. GTEx Consortium (2020) The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science*, **369**, 1318–1330.
19. Navarro Gonzalez,J., Zweig,A.S., Speir,M.L., Schmelter,D., Rosenbloom,K.R., Raney,B.J., Powell,C.C., Nassar,L.R., Maulding,N.D., Lee,C.M. *et al.* (2021) The UCSC genome browser database: 2021 update. *Nucleic Acids Res.*, **49**, D1046–D1057.
20. Huang,D., Yi,X., Zhou,Y., Yao,H., Xu,H., Wang,J., Zhang,S., Nong,W., Wang,P., Shi,L. *et al.* (2020) Ultrafast and scalable variant annotation and prioritization with big functional genomics data. *Genome Res.*, **30**, 1789–1801.
21. Lappalainen,T., Sammeth,M., Friedlander,M.R., t Hoen,P.A., Monlong,J., Rivas,M.A., Gonzalez-Porta,M., Kurbatova,N., Griebel,T., Ferreira,P.G. *et al.* (2013) Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*, **501**, 506–511.
22. Chen,L., Ge,B., Casale,F.P., Vasquez,L., Kwan,T., Garrido-Martin,D., Watt,S., Yan,Y., Kundu,K., Ecker,S. *et al.* (2016) Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell*, **167**, 1398–1414.
23. Genomes Project,C., Auton,A., Brooks,L.D., Durbin,R.M., Garrison,E.P., Kang,H.M., Korbel,J.O., Marchini,J.L., McCarthy,S., McVean,G.A. *et al.* (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.
24. Huang,D., Yi,X., Zhang,S., Zheng,Z., Wang,P., Xuan,C., Sham,P.C., Wang,J. and Li,M.J. (2018) GWAS4D: multidimensional analysis of context-specific regulatory variant for human complex diseases and traits. *Nucleic Acids Res.*, **46**, W114–W120.
25. Group,P.T.C., Calabrese,C., Davidson,N.R., Demircioglu,D., Fonseca,N.A., He,Y., Kahles,A., Lehmann,K.V., Liu,F., Shiraishi,Y. *et al.* (2020) Genomic basis for RNA alterations in cancer. *Nature*, **578**, 129–136.
26. Huang,D., Zhou,Y., Yi,X., Fan,X., Wang,J., Yao,H., Sham,P.C., Hao,J., Chen,K. and Li,M.J. (2022) VannoPortal: multiscale functional annotation of human genetic variants for interrogating molecular mechanism of traits and diseases. *Nucleic Acids Res.*, **50**, D1408–D1416.
27. Theusch,E., Chen,Y.I., Rotter,J.I., Krauss,R.M. and Medina,M.W. (2020) Genetic variants modulate gene expression statin response in human lymphoblastoid cell lines. *BMC Genomics*, **21**, 555.
28. Gutierrez-Arcelus,M., Lappalainen,T., Montgomery,S.B., Buil,A., Ongen,H., Yurovsky,A., Bryois,J., Giger,T., Romano,L., Planchon,A. *et al.* (2013) Passive and active DNA methylation and the interplay with genetic variation in gene regulation. *Elife*, **2**, e00523.
29. Delaneau,O., Zazhytska,M., Borel,C., Giannuzzi,G., Rey,G., Howald,C., Kumar,S., Ongen,H., Popadin,K., Marbach,D. *et al.* (2019) Chromatin three-dimensional interactions mediate genetic effects on gene expression. *Science*, **364**, eaat8266.
30. Tehranchi,A.K., Myrthil,M., Martin,T., Hie,B.L., Golan,D. and Fraser,H.B. (2016) Pooled chip-Seq links variation in transcription factor binding to complex disease risk. *Cell*, **165**, 730–741.
31. Izzi,B., Pistoni,M., Cludts,K., Akkor,P., Lambrechts,D., Verfaillie,C., Verhamme,P., Freson,K. and Hoylaerts,M.F. (2016) Allele-specific DNA methylation reinforces PEAR1 enhancer activity. *Blood*, **128**, 1003–1012.
32. Vierstra,J., Lazar,J., Sandstrom,R., Halow,J., Lee,K., Bates,D., Diegel,M., Dunn,D., Neri,F., Haugen,E. *et al.* (2020) Global reference mapping of human transcription factor footprints. *Nature*, **583**, 729–736.
33. Ionita-Laza,I., McCallum,K., Xu,B. and Buxbaum,J.D. (2016) A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat. Genet.*, **48**, 214–220.
34. Ferraro,N.M., Strober,B.J., Einson,J., Abell,N.S., Aguet,F., Barbeira,A.N., Brandt,M., Bucan,M., Castel,S.E., Davis,J.R. *et al.* (2020) Transcriptomic signatures across human tissues identify functional rare genetic variation. *Science*, **369**, eaaz5900.
35. Habib,A.M., Okorokov,A.L., Hill,M.N., Bras,J.T., Lee,M.C., Li,S., Gossage,S.J., van Drimmelen,M., Morena,M., Houlden,H. *et al.* (2019) Microdeletion in a FAAH pseudogene identified in a patient with high anandamide concentrations and pain insensitivity. *Br. J. Anaesth.*, **123**, e249–e253.
36. Tehranchi,A., Hie,B., Dacre,M., Kaplow,I., Pettie,K., Combs,P. and Fraser,H.B. (2019) Fine-mapping cis-regulatory variants in diverse human populations. *Elife*, **8**, e39595.
37. Umans,B.D., Battle,A. and Gilad,Y. (2021) Where are the disease-associated eQTLs? *Trends Genet.*, **37**, 109–124.
38. Cano-Gamez,E. and Trynka,G. (2020) From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. *Front. Genet.*, **11**, 424.
39. Chen,L. and Li,M.J. (2021) Editorial: deciphering non-coding regulatory variants: computational and functional validation. *Front. Bioeng. Biotechnol.*, **9**, 769614.
40. van der Wijst,M., de Vries,D.H., Groot,H.E., Trynka,G., Hon,C.C., Bonder,M.J., Stegle,O., Nawijn,M.C., Idaghdour,Y., van der Harst,P. *et al.* (2020) The single-cell eQTLGen consortium. *Elife*, **9**, e52155.
41. Cui,H., Zuo,S., Liu,Z., Liu,H., Wang,J., You,T., Zheng,Z., Zhou,Y., Qian,X., Yao,H. *et al.* (2020) The support of genetic evidence for cardiovascular risk induced by antineoplastic drugs. *Sci. Adv.*, **6**, eabb8543.
42. ULTRA-DD Consortium, Fang,H., De Wolf,H., Knezevic,B., Burnham,K.L., Osgood,J., Sanniti,A., Lledo Lara,A., Kasela,S., De Cesco,S. *et al.* (2019) A genetics-led approach defines the drug target landscape of 30 immune-related traits. *Nat. Genet.*, **51**, 1082–1091.