# Qtlizer: comprehensive QTL annotation of GWAS results

Matthias Munz[1-5,*], Inken Wohlers[3], Eric Simon[6], Hauke Busch[3], Arne S. Schaefer[4,¶], Jeanette Erdmann[1,4-5,¶,*]

[¶]These authors contributed equally

[1]Institute for Cardiogenetics, University of Lübeck, 23562 Lübeck, Germany
[2]Charité – University Medicine Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Institute for Dental and Craniofacial Sciences, Department of Periodontology and Synoptic Dentistry, Berlin, Germany
[3]Medical Systems Biology Group, Institute of Experimental Dermatology, Institute for Cardiogenetics, University of Lübeck, 23562 Lübeck, Germany
[4]DZHK (German Research Centre for Cardiovascular Research), partner site Hamburg/Lübeck/Kiel, 23562 Lübeck, Germany
[5]University Heart Center Lübeck, 23562 Lübeck, Germany
[6]Target Discovery Research, Boehringer Ingelheim Pharma GmbH & Co KG, Biberach 88397, Germany

[*] To whom correspondence should be addressed.
Email: m.munz@uni-luebeck.de
Telephone: +49 (0) 451 500 50860
Address:
Medical Systems Biology Group
Institute of Experimental Dermatology
Institute for Cardiogenetics
University of Lübeck
Ratzeburger Allee 160
23562 Lübeck, Germany

Correspondence may also be addressed to
Email: jeanette.erdmann@uni-luebeck.de
Telephone: +49 (0) 451 3101 8300
Address:
Institute for Cardiogenetics
University of Lübeck
Ratzeburger Allee 160
23562 Lübeck, Germany

**ABSTRACT**

Exploration of genetic variant-to-gene relationships by quantitative trait loci (QTLs) helps to identify candidate causal variants and genes in post genome-wide association study analyses. However, the wide range of public QTL databases and the lack of batch annotation features make it cumbersome to investigate these relationships in a comprehensive manner. In this work, we introduce the tool 'Qtlizer' to annotate lists of common variants and genes in humans with associated changes in gene expression and protein abundance using the, to-date, most comprehensive database of published QTLs. The features include incorporation of LD variants, sophisticated prioritization functionality and linking to other resources. The web application of Qtlizer is available at http:// genehopper.de/qtlizer, a guide on how to use the REST API is available at http://genehopper.de/rest.

## INTRODUCTION

In the past decade, genome-wide association studies (GWAS) led to the discovery of thousands of genetic loci that are associated with variation in traits and diseases. However, according to the NHGRI-EBI Catalog of published GWAS (GWAS Catalog), only a small fraction of 5-6% of the associated sentinel variants are located in protein-coding regions (**Supplementary Figure 1**). For the majority of variants possible functional consequences are still unclear, resulting in a lack of mechanistic understanding of these associations. Elucidating the relationships between risk variants and genes by quantitative trait loci (QTLs) such as expression QTLs (eQTLs) can help to identify candidate variants and genes for functional studies.

Consequently, several public databases (DBs) have evolved: the GTEx Portal (1), Haploreg (2), GRASP (3), GEUVADIS (4), SCAN (5), seeQTL (6), Blood eQTL Browser (7), pGWAS (8), ExSNP (9) and BRAINEAC (10). Some of them, but in particular the GTEx Portal, provide specific tools and visualizations to analyze their QTL data. However, most of the existing DBs only provide unique QTL datasets resulting in a poor overlap between the DBs. In addition, the databases are very limited regarding the used search terms and are lacking the feature to search for multiple variants or genes in parallel. Hence, it is very cumbersome to comprehensively annotate lists of variants or genes with immediate results. Another aspect, that the current databases do not sufficiently consider, is that for both GWAS and QTL studies, true association signals are very often accompanied by multiple other variants, which can be attributed to the human linkage disequilibrium (LD) structure. Therefore, it is not trivial to break down a signal to one or more causal variants just as it is to find true signal overlaps between a GWAS and a QTL signal.

To address these shortcomings, we have developed the web-based application ‘Qtlizer’, which facilitates and improves QTL annotation. Qtlizer allows exploration of QTL data for a given list of common variants and/or genes in a fast and efficient manner by integrating a large number of QTL datasets. All data is extensively annotated in the web application and direct links to the source platform are provided for further analyses. To enhance the interpretability regarding signal overlaps, Qtlizer allows for the inclusion of variants in LD with a query variant, as well as flag them which helps to dissect the QTL data into independent association signals. In line with the release of Qtlizer, we provide two unique eQTL datasets from the Cardiogenics Consortium (11, 12).

## MATERIALS AND METHODS

### Data sources

Qtlizer currently integrates 166 tissue-specific QTL datasets of which two are response expression QTLs (reQTLs) and protein abundance QTLs (pQTLs), while the remainder are eQTLs. Further details are provided in **Supplementary Table 1**. eQTL data was obtained from 13 sources that included datasets from 53 publications and cover 114 different tissues or cell types. For 34 of these tissues more than one eQTL data set is available, with most included studies performed for lymphoblastoid cell lines (n=9), liver (n=8), cerebellum (n=4) and monocytes (n=4). The number of variants per study varies considerably and reflects the diversity of the included data; the median is approx. 3,700, but the number of variants ranges up to 2,017,804 for a thyroid dataset from the GTEx Portal. Some of the included datasets cover only variants that are statistically significant after multiple testing in the original study, while others also cover suggestively associated SNPs.
Moreover, we included several other datasets. Published variant phenotype associations identified in GWAS were taken from GWAS Catalog (13, 14). We used genotype data from 1000 Genomes Project Phase 3 (1000GP3) (15). A dataset of topological associated domain (TAD) boundaries was taken from Dixon et al. (16). It contains boundaries of 3,054 TADs with a mean length of 853 kilo bases (kb; maximum length = 4.44 mega bases, minimum length = 0.8 kb). TADs are spatially close genetic regions in the genome and have been defined based on Hi-C chromatin interaction data, i.e. physical interactions occur more frequently within such a compartment. All other data such as genetic variant and gene information was taken from the Genehopper DB (http://genehopper.de) (17). Genehopper DB is an integrated DB that mainly relies on data collected from Ensembl (18, 19).

### Implementation

We used Perl to implement an extraction, transformation and loading (ETL) process in which QTL summary statistics and other aforementioned data was downloaded from the respective websites (extraction), manipulated to fit the database schema of the host database (transformation) and the resulting tables were loaded into the host database. As a host database we used the Genehopper DB which is based on the relational database management system MySQL (https://www.mysql.com). On top of the DB we built the web application of Qtlizer by using the Play! Framework (https://www.playframework.com). We used Java for implementing the backend and Javascript for implementing the frontend of the application. Lastly, we used the front-end component library Bootstrap (https://getbootstrap.com/docs/3.3/getting-started/) to give Qtlizer a responsive and appealing design.

## RESULTS

### Data transformation

After selecting and downloading relevant datasets, we applied a transformation procedure as part of the ETL process. This procedure consisted of several crucial steps. Most importantly, the source variant and gene identifiers were mapped to internal identifiers of Genehopper DB. To make tissue names are consistent across all datasets, we manually created a mapping file to convert tissue names to a normalized form. If available, we took QTL significance information from the source, if not, the significance was determined by adjusting for multiple testing using a family-wise error rate (FWER) of 5%, resulting in an adjusted significance level of $P = 10^{-12}$ which corresponds to 1000,000 variants and 50,000 genes. eQTL entities that passed the study-wide significance threshold were flagged as "is_sw_significant". Here and in the following we define a QTL entity as a quadruple of variant, gene, tissue and source study. We utilized the TAD boundaries to categorize QTLs into *cis* or *trans* to tell whether a variant and gene of the same QTL entity were located in the same TAD or not. All variants and genes, that were included in the QTL data, were mapped to GWAS Catalog for possible variant-phenotype or locus-phenotype associations. Variant-variant correlations were calculated as LD based on the 1000GP3 genotypes **(Supplementary Figure 2)**. For each QTL entity we used this LD information to check whether the underlying variant had the lowest P-value among all variants within its LD group in a specific tissue of a QTL study, and flagged the QTL entity accordingly with "is_best_in_ld_group". We defined a LD group of a variant as the variant and all other variants with $r^2 > 0.2$. Moreover, QTL entities were flagged as "is_best" if it has the overall lowest P-value in a specific tissue of a QTL study. Lastly, we calculated count characteristics for each QTL entity. For each gene in a specific tissue of a QTL study, we calculated the number of putative QTLs ("n_qtls"), the number of tissues and studies in which the putative QTL is the best regarding the P-value ("n_best"), the number of tissues and studies in which the putative QTL is study-wide significant and the best ("n_sw_significant") and the number of tissues and studies in which the putative QTL occurs ("n_occ").

40,883,209 (37,014,094 study-wide significant) QTLs from 3,856,968 variants and 32,987 genes were finally added to the Genehopper DB (**Figure 1a, b**). Of these, 25,432,954 QTLs were annotated as being *cis* QTLs and 26,449 variants and 11,511 genes could be linked to an entry in the GWAS Catalog.

### User interface workflow

The user interface of Qtlizer takes a list of query terms as input (**Figure 1c**). A single query term can either identify a genetic variant or a gene. Accepted variant identifiers are reference SNP identifiers (rsIds) and chromosomal positions of the reference builds hg19/GRCh37 or hg38/GRCh38. Genes can be specified using a variety of database identifiers (e.g. Ensembl Gene Id, EntrezGene Id) or gene symbols consisting of upper-case letters and Arabic numerals. Gene terms are mapped to an internal gene identifier by using the full text term-to-gene index and a prioritization algorithm provided by Genehopper to account for ambiguous identifiers or symbols (17). By using of regular expressions,

Qtlizer requires no special input format and is able to interpret all standard text formats (e.g. tab-, comma-, space-separated, or a mixture of all). The regular expressions are listed in **Supplementary Figure 3**.

Optionally, queried variants can be enriched for proxy variants using the European reference population of 1000GP3 by setting a LD threshold for $r^2$ or D'.

The query result is displayed in a joint table view, with one row for every QTL entity (**Figure 1d**). This means that each row represents a putative association of a variant with the expression of a gene in a given tissue and within a QTL study that has been obtained from the given source.

Initially, only basic information is displayed: the variant rsId and hyperlinks to dbSNP and Ensembl; the rsId of the proxy variant, if input variants were enriched for proxies; a gene symbol with a hyperlink to Ensembl; the chromosomal distance in kilo bases between variant and gene; the normalized tissue name; an association P-value; a beta value representing the effect size; effect and non-effect alleles (EA, NEA); the source name with hyperlinks to the source database and the original publication in which the dataset was published; the flags "is_significant", "is_best", "is_best_in_ld_group" which are grouped in a single column; hyperlinks that connecting the variant, proxy variant and/or the gene with entries in the GWAS Catalog.

Additionally, there are a number of optional columns that can be interactively displayed. The search type can be a "Variant" if the query term was mapped to a variant, or "Gene" if the query term was mapped to a gene; the original query term that was entered by the user; the type of QTL, e.g. eQTL or pQTL; Ensembl Gene Id; co-localization (*cis* versus *trans*); a significance information column providing details about the multiple testing correction and the applied significance threshold; the Pubmed Id of the original publication and a counts column summarizing the data for several eQTL studies.

The resulting table can be interactively sorted and filtered column-wise, either by selecting an arbitrary subset of column values (e.g. a selection of SNPs or genes or tissues) or by specifying numerical thresholds (e.g. P-value < 0.001 or beta > 0). The current filter state of the table can be downloaded as text file for further use.

Qtlizer allows to summarize the results such that for each index variant and gene all available tissues and the best P-value are merged into a single row. QTLs of proxy variants are merged with the index variant to allow for a haplotype block-centric view.

**Example use case**

We use Qtlizer to annotate a coronary artery disease (CAD) GWAS locus at chromosome 3q22.3, whose index SNP is rs2306374 (20) (**Supplementary Figure 4**). After specifying the input SNP and clicking the button "Annotate", Qtlizer lists 17 QTL entities (specified as "#Results: 17") for this SNP in the Qtlizer database. After enabling the column "QTL type" using the button "Select columns", it turns out that all results are eQTLs. By clicking the button "Aggregate", we see that there are eight eQTL genes and that SNP rs2306374 is associated with the expression of *MRAS* (Muscle RAS oncogene homolog) in multiple tissues. Most of these tissues are relevant for coronary artery disease, e.g. "Artery – Aorta", "Artery – Tibial", "Heart - Atrial appendage" and "Heart". By clicking on the SNP in the

“GWAS Catalog” column, we learn from the GWAS Catalog website that the SNP’s C allele is associated with CAD according to a GWAS performed in altogether more than 100,000 Europeans (21). After shifting back to Qtlizer’s detailed table view by using again the “Aggregate” button, we investigate the *MRAS* gene closer by selecting it exclusively in the “Gene” column header. Nine associations remain, all have the flag “is_sw_significant” and are thus study-wide significant eQTLs. By inspecting the allele columns “EA” and “NEA”, we observe that the eQTL effect allele (EA) is C and from the “Beta” column we see that in four CAD-relevant tissues *MRAS* expression is increased for the EA, whereas it is decreased in three other tissues. The “Distance” column specifies a distance of 0 which denotes that SNP rs2306374 is located within gene boundaries. The absence of the “is_best” and “is_best_in_ld_group” flags denotes that the query SNP is not the eQTL sentinel variant.

In a next step we start a new Qtlizer search which includes proxy variants. We choose a conservative LD threshold of $r^2 = 0.8$, resulting in 315 associations. The aggregate view confirms that *MRAS* is still the eQTL gene observed in most relevant tissues, and additional relevant tissues for proxies are “Atherosclerotic aortic root”, “Artery - Coronary”, “Artery – Mammary” and “Heart - Left ventricle”. We again select only *MRAS* eQTLs and filter further to keep those associations in heart, artery/aortic tissues tagged with flags “is_sw_significant” and one or both of “is_best” and “is_best_in_ld_group”. We obtain eight associations for three proxy variants with $r^2 = 0.94$ (provided in brackets behind the respective rsIds). Two sources, GTEx and Haploreg, concordantly report the eQTL signal as being best for the LD block in three heart tissues, and two sources, GTEx and Franzén et al., as being best for the LD block in five artery tissues.

Using effect predictions for all study-wide significant variants by also displaying eQTLs with “is_sw_significant” only, we prioritize SNP rs9851766 ($r^2 = 0.98$, beta = 1.05) and rs13324341 ($r^2 = 0.94$; beta = 1.03) for functional studies. SNP rs9851766 is located within the 3’ UTR of *MRAS* and 26 base pairs downstream of a hsa-miR-135-5p binding site which was predicted by TargetScanHuman release 7.2 (22). hsa-miR-135-5p is a microRNA with relevance in CAD (23) whose post-transcriptional regulation of MRAS may be affected by this variant. SNP rs13324341 is intronic of *MRAS* and lies in a DNase I hypersensitivity site according to the ENCODE track in the UCSC Browser (24, 25). Interestingly, according to the “Count” column, this SNP is the best QTL for *MRAS* in six studies in heart and artery tissues.

## DISCUSSION

With Qtlizer we introduce a web-based tool to annotate lists of common small variants and genes with QTL data in a comprehensive manner. Qtlizer has an easy-to-use graphical interface for jointly reviewing many QTL studies on various tissues, covering a large portion of humane common small variants and genes. This is possible because we integrated QTL data from many different sources, connected them with various other data types and calculated meaningful characteristics. To-date, Qtlizer contains the most diverse and comprehensive set of QTL data. Moreover, it provides unique functionality to analyze multiple inputs in parallel.

It is important to note that we explicitly did not aim to statistically combine the QTL data over studies or tissues. Such efforts are difficult because studies differ in various aspects such as sample sizes, disease status and genetic background of genotyped individuals as well as techniques used for genotyping and expression quantification (e.g. array versus sequencing). Besides these differences, there are also other restrictions regarding the public availability. For example, some studies provide only significant associations after multiple testing correction, others use thresholds on nominal association P-values. Another challenge is missing information. For example, beta and effect alleles are occasionally not reported and as a result the effect direction cannot be determined. Due to QTL data heterogeneity, Qtlizer is designed to provide comprehensive access to as many available QTL datasets as possible in order to provide a starting point for in-depth literature review and further analysis in native QTL browsers, if available.

Also, we refrained from filtering variants or genes to be included in Qtlizer to enable the user to review all available information. However, we do allow interactive prioritization of variant-gene relationships.

In genetic association studies such as GWAS and QTL studies, association signals are strongly influenced by the LD structure. This means that a true association signal usually does not only consist of the causal variant(s), but is supported by variants which are correlated with the causal variant(s), producing an inherent ambiguity in interpreting association study results. This phenomenon is visualized as coherent elevations in regional association plots for which we provide an example in **Supplementary Figure 4**. Consequently, the problem of correlated variants makes it also difficult to determine whether two association signals overlap and point to the same causal variant(s). A typical case use for Qtlizer is to compare a GWAS signal, which is represented by an index variant given as input by the user, with that of a QTL signal in Qtlizer. To assist the user with investigating the signal overlap between the query and the QTL data in Qtlizer, we implemented the option to include LD variants and to prioritize the QTLs using informative flags and summary counts. Flags that we want to highlight in this context are “is_best” and “is_best_in_ld_group”, informing the user whether a QTL entity is overall the best or the best among all correlated QTLs ($r^2 > 2$) in a tissue and study regarding the association P-value. If sufficient data is available for the tissue of interest, the user can also export selected QTL data for testing the significance of overlap of association signals using dedicated tools such as Coloc (26).

It is common practice to divide between *cis* eQTLs and *trans* eQTLs. Usually, the functional mechanism is not known and eQTL studies use the term *cis*, if the variant and the associated gene are close to each other (e.g. within 1 MB) and *trans* otherwise. In Qtlizer we improve on this crude definition by using TADs boundaries. Although multiple publications have suggested that chromosomal TAD boundaries are largely conserved across cell-types (16, 27, 28), recent studies indicate that TAD boundaries might be more unstable than previously assumed (29). However, in this work, we assumed the former.

The frontend functionality of Qtlizer was implemented in Javascript which means, that the performance of rendering and manipulating the result table is highly dependent on the hardware resources of the user. For this reason, we restricted queries to a maximum of 50 query terms and a maximum table size of 5,000 entries. For queries exceeding these thresholds, we provide a web service, allowing programmatic access to the QTL data using any programming language. Example scripts for Perl and R are shown in **Supplementary Figure 5, 6**.

Keeping the database of Qtlizer up-to-date is especially important when the source databases release new versions or new datasets are being published. Therefore, we implemented a semi-automated ETL process allowing to update Qtlizer in a timely and simple manner.

## CONCLUSION AND OUTLOOK

In summary, Qtlizer provides a high performant web-based solution for annotating lists of genetic variants and genes with QTLs, facilitating the exploration of the consequential genetic associations in non-coding regions on the molecular mechanisms of a phenotype. Qtlizer outperforms each of the existing databases regarding the number of QTL datasets and the variety of tissues that can be explored. Since QTLs can be tissue-specific, this is particularly helpful to select appropriate tissues for a phenotype of interest.

In addition to incorporating new QTL datasets, we also intend to map QTLs to functional chromosomal sites such as enhancers by integrating data from the Encyclopedia of DNA Elements (ENCODE) (24). This would help to discriminate causal from non-causal QTL variants. Another promising new function of Qtlizer would be a pathway (or ontology, transcription factor, etc.) enrichment analysis. It could be realized by integrating GWAS data, that would ideally be inputted by the user, and QTL data of Qtlizer, that would ideally be filtered by the user prior to the analysis to match the phenotype. If effect directions are available for both entities, the function would even be able to distinguish between up- and downregulated pathways. Thus, this approach will heavily rely on the availability of complete QTL summary statistics, including effect size and effect allele.

## AVAILABILITY

The web application of Qtlizer is available at http://www.genehopper.de/qtlizer, a guide on how to use the REST API is available at http://www.genehopper.de/rest (**Supplementary Figure 4, 5**). The website is free and open to all and there is no login requirement.


## ACKNOWLEDGMENTS

We acknowledge computational support from the OMICS compute cluster at the University of Lübeck. Special thanks to Dr. Carsten Kemena for extensive testing of the web application and Dr. Loreto Munoz Venegas for revising the manuscript. We cordially thank the Cardiogenics Consortium for data related to the macrophage and monocyte eQTLs. Members of the Cardiogenics Consortium are listed in **Supplementary Table 2**.

## FUNDING

This work was supported by a research grant of the German Research Foundation DFG (Deutsche Forschungsgemeinschaft; GZ: SCHA 1582/3-1) as well as the Land Schleswig-Holstein within the funding program “Open Access Publikationsfonds” and the clusters of excellence “Inflammation at Interfaces” (IaI) and “Precision Medicine in Chronic Inflammation” (PMI). I.W. was supported by the Peter und Traudl Engelhorn Foundation. H.B. and J.E. are supported by the German Research Foundation DFG (Deutsche Forschungsgemeinschaft) under Germany’s Excellence Strategy – EXC 22167-390884018.


## CONFLICT OF INTEREST

None declared.

## REFERENCES


1. GTEx Consortium,Gte. (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, **348**, 648–60.
https://doi.org/10.1126/science.1262110
http://www.ncbi.nlm.nih.gov/pubmed/25954001

2. Ward,L.D. and Kellis,M. (2016) HaploReg v4: Systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.*, **44**, D877–D881.
https://doi.org/10.1093/nar/gkv1340
http://www.ncbi.nlm.nih.gov/pubmed/26657631

3. Eicher,J.D., Landowski,C., Stackhouse,B., Sloan,A., Chen,W., Jensen,N., Lien,J.-P., Leslie,R. and Johnson,A.D. (2015) GRASP v2.0: an update on the Genome-Wide Repository of Associations between SNPs and phenotypes. *Nucleic Acids Res.*, **43**, D799-804.
https://doi.org/10.1093/nar/gku1202
http://www.ncbi.nlm.nih.gov/pubmed/25428361

4. Lappalainen,T., Sammeth,M., Friedländer,M.R., 't Hoen,P.A.C., Monlong,J., Rivas,M.A., Gonzàlez-Porta,M., Kurbatova,N., Griebel,T., Ferreira,P.G., *et al.* (2013) Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*, **501**, 506–11.
https://doi.org/10.1038/nature12531
http://www.ncbi.nlm.nih.gov/pubmed/24037378

5. Gamazon,E.R., Zhang,W., Konkashbaev,A., Duan,S., Kistner,E.O., Nicolae,D.L., Dolan,M.E. and Cox,N.J. (2010) SCAN: SNP and copy number annotation. *Bioinformatics*, **26**, 259–62.
https://doi.org/10.1093/bioinformatics/btp644
http://www.ncbi.nlm.nih.gov/pubmed/19933162

6. Xia,K., Shabalin,A.A., Huang,S., Madar,V., Zhou,Y.-H., Wang,W., Zou,F., Sun,W., Sullivan,P.F. and Wright,F.A. (2012) seeQTL: a searchable database for human eQTLs. *Bioinformatics*, **28**, 451–2.
https://doi.org/10.1093/bioinformatics/btr678
http://www.ncbi.nlm.nih.gov/pubmed/22171328

7. Westra,H.-J., Peters,M.J., Esko,T., Yaghootkar,H., Schurmann,C., Kettunen,J., Christiansen,M.W., Fairfax,B.P., Schramm,K., Powell,J.E., *et al.* (2013) Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.*, **45**, 1238–43.
https://doi.org/10.1038/ng.2756
http://www.ncbi.nlm.nih.gov/pubmed/24013639

8. Suhre,K., Arnold,M., Bhagwat,A.M., Cotton,R.J., Engelke,R., Raffler,J., Sarwath,H., Thareja,G., Wahl,A., DeLisle,R.K., *et al.* (2017) Connecting genetic risk to disease end points through the human blood plasma proteome. *Nat. Commun.*, **8**, 14357.
https://doi.org/10.1038/ncomms14357
http://www.ncbi.nlm.nih.gov/pubmed/28240269

9. Yu,C.-H., Pal,L.R. and Moult,J. (2016) Consensus Genome-Wide Expression Quantitative Trait Loci and Their Relationship with Human Complex Trait Disease. *OMICS*, **20**, 400–14.
https://doi.org/10.1089/omi.2016.0063
http://www.ncbi.nlm.nih.gov/pubmed/27428252

10. Ramasamy,A., Trabzuni,D., Guelfi,S., Varghese,V., Smith,C., Walker,R., De,T., UK Brain Expression Consortium,U.B.E., North American Brain Expression Consortium,N.A.B.E., Coin,L., *et al.* (2014) Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat. Neurosci.*, **17**, 1418–1428.
https://doi.org/10.1038/nn.3801
http://www.ncbi.nlm.nih.gov/pubmed/25174004

11. Garnier,S., Truong,V., Brocheton,J., Zeller,T., Rovital,M., Wild,P.S., Ziegler,A., Cardiogenics Consortium,T.C., Munzel,T., Tiret,L., *et al.* (2013) Genome-wide haplotype analysis of cis expression quantitative trait loci in monocytes. *PLoS Genet.*, **9**, e1003240.
https://doi.org/10.1371/journal.pgen.1003240
http://www.ncbi.nlm.nih.gov/pubmed/23382694

12. Codoni,V., Blum,Y., Civelek,M., Proust,C., Franzén,O., Cardiogenics Consortium,C., IDEM Leducq Consortium CADGenomics,I.L.C., Björkegren,J.L.M., Le Goff,W., Cambien,F., *et al.* (2016) Preservation Analysis of Macrophage Gene Coexpression Between Human and Mouse Identifies PARK2 as a Genetically Controlled Master Regulator of Oxidative Phosphorylation in Humans. *G3 (Bethesda).*, **6**, 3361–3371.
https://doi.org/10.1534/g3.116.033894
http://www.ncbi.nlm.nih.gov/pubmed/27558669

13. Welter,D., MacArthur,J., Morales,J., Burdett,T., Hall,P., Junkins,H., Klemm,A., Flicek,P., Manolio,T., Hindorff,L., *et al.* (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.*, **42**, D1001-6.
https://doi.org/10.1093/nar/gkt1229
http://www.ncbi.nlm.nih.gov/pubmed/24316577

14. MacArthur,J., Bowler,E., Cerezo,M., Gil,L., Hall,P., Hastings,E., Junkins,H., McMahon,A., Milano,A., Morales,J., *et al.* (2017) The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.*, **45**, D896–D901.
https://doi.org/10.1093/nar/gkw1133
http://www.ncbi.nlm.nih.gov/pubmed/27899670

15. Auton,A., Abecasis,G.R., Altshuler,D.M., Durbin,R.M., Abecasis,G.R., Bentley,D.R., Chakravarti,A., Clark,A.G., Donnelly,P., Eichler,E.E., *et al.* (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.
https://doi.org/10.1038/nature15393

16. Dixon,J.R., Selvaraj,S., Yue,F., Kim,A., Li,Y., Shen,Y., Hu,M., Liu,J.S. and Ren,B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–80.
https://doi.org/10.1038/nature11082
http://www.ncbi.nlm.nih.gov/pubmed/22495300

17. Munz,M., Tönnies,S., Balke,W.-T. and Simon,E. (2015) Multidimensional gene search with Genehopper. *Nucleic Acids Res.*, **43**, 98–103.
https://doi.org/10.1093/nar/gkv511
http://www.ncbi.nlm.nih.gov/pubmed/25990726

18. Chen,Y., Cunningham,F., Rios,D., McLaren,W.M., Smith,J., Pritchard,B., Spudich,G.M., Brent,S., Kulesha,E., Marin-Garcia,P., *et al.* (2010) Ensembl variation resources. *BMC Genomics*, **11**, 293.

https://doi.org/10.1186/1471-2164-11-293

19. Yates,A., Akanni,W., Amode,M.R., Barrell,D., Billis,K., Carvalho-Silva,D., Cummins,C., Clapham,P., Fitzgerald,S., Gil,L., *et al.* (2016) Ensembl 2016. *Nucleic Acids Res.*, **44**, D710–D716.
https://doi.org/10.1093/nar/gkv1157

20. Erdmann,J., Kessler,T., Munoz Venegas,L. and Schunkert,H. (2018) A decade of genome-wide association studies for coronary artery disease: the challenges ahead. *Cardiovasc. Res.*, 10.1093/cvr/cvy084.
https://doi.org/10.1093/cvr/cvy084
http://www.ncbi.nlm.nih.gov/pubmed/29617720

21. Schunkert,H., König,I.R., Kathiresan,S., Reilly,M.P., Assimes,T.L., Holm,H., Preuss,M., Stewart,A.F.R., Barbalic,M., Gieger,C., *et al.* (2011) Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.*, **43**, 333–338.
https://doi.org/10.1038/ng.784

22. Agarwal,V., Bell,G.W., Nam,J.-W. and Bartel,D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, **4**.
https://doi.org/10.7554/eLife.05005
http://www.ncbi.nlm.nih.gov/pubmed/26267216

23. Xu,Z., Han,Y., Liu,J., Jiang,F., Hu,H., Wang,Y., Liu,Q., Gong,Y. and Li,X. (2015) MiR-135b-5p and MiR-499a-3p Promote Cell Proliferation and Migration in Atherosclerosis by Directly Targeting MEF2C. *Sci. Rep.*, **5**, 12276.
https://doi.org/10.1038/srep12276

24. Consortium,T.E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
https://doi.org/10.1038/nature11247

25. Kent,W.J., Sugnet,C.W., Furey,T.S., Roskin,K.M., Pringle,T.H., Zahler,A.M. and Haussler,D. (2002) The human genome browser at UCSC. *Genome Res.*, **12**, 996–1006.
https://doi.org/10.1101/gr.229102
http://www.ncbi.nlm.nih.gov/pubmed/12045153

26. Giambartolomei,C., Vukcevic,D., Schadt,E.E., Franke,L., Hingorani,A.D., Wallace,C. and Plagnol,V. (2014) Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLoS Genet.*, **10**, e1004383.
https://doi.org/10.1371/journal.pgen.1004383
http://www.ncbi.nlm.nih.gov/pubmed/24830394

27. Pope,B.D., Ryba,T., Dileep,V., Yue,F., Wu,W., Denas,O., Vera,D.L., Wang,Y., Hansen,R.S., Canfield,T.K., *et al.* (2014) Topologically associating domains are stable units of replication-timing regulation. *Nature*, **515**, 402–405.
https://doi.org/10.1038/nature13986

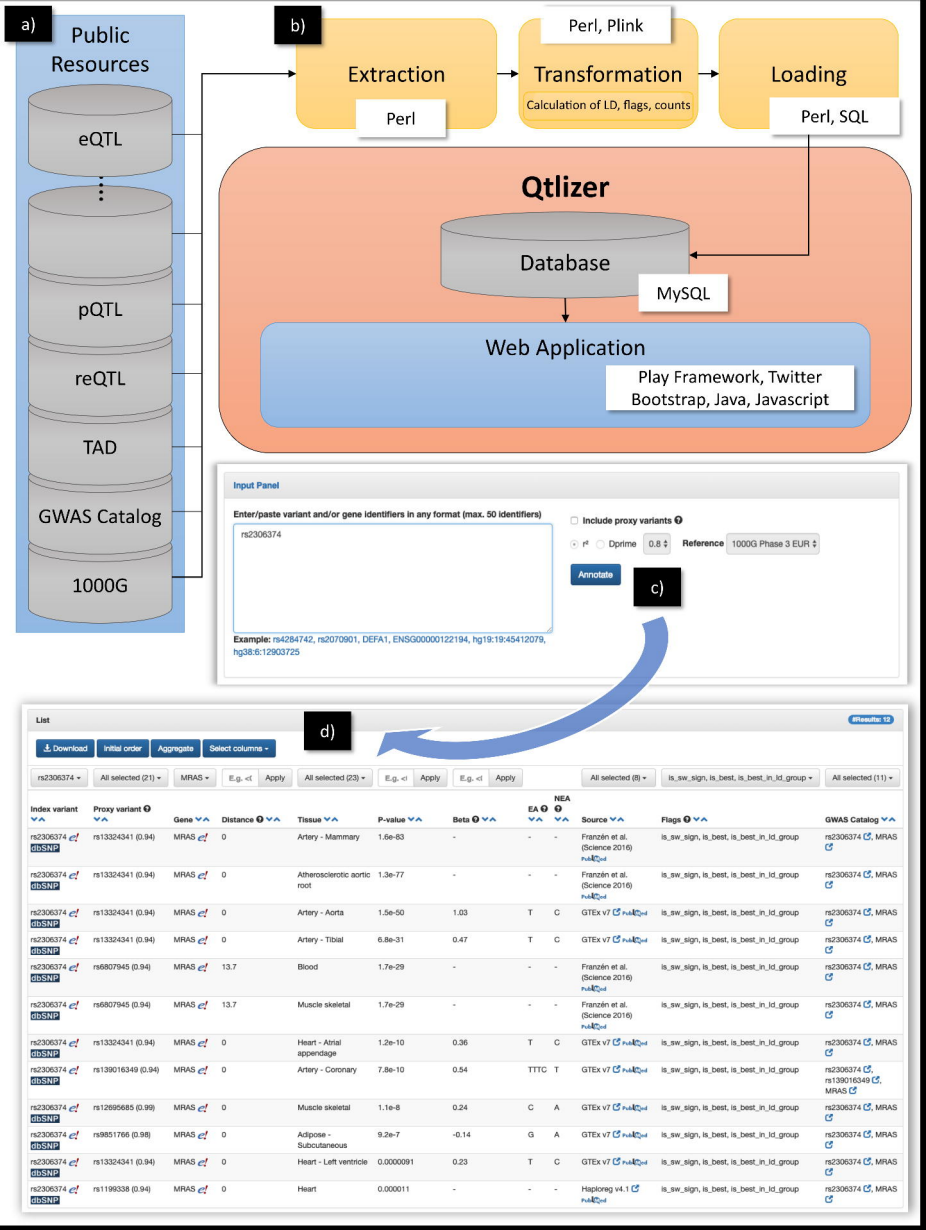
28. Dekker,J., Meisig,J., Heard,E., Barillot,E., Sedat,J., Okamoto,I., van Berkum,N.L., Blüthgen,N., Servant,N., Nora,E.P., *et al.* (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, **485**, 381–385.

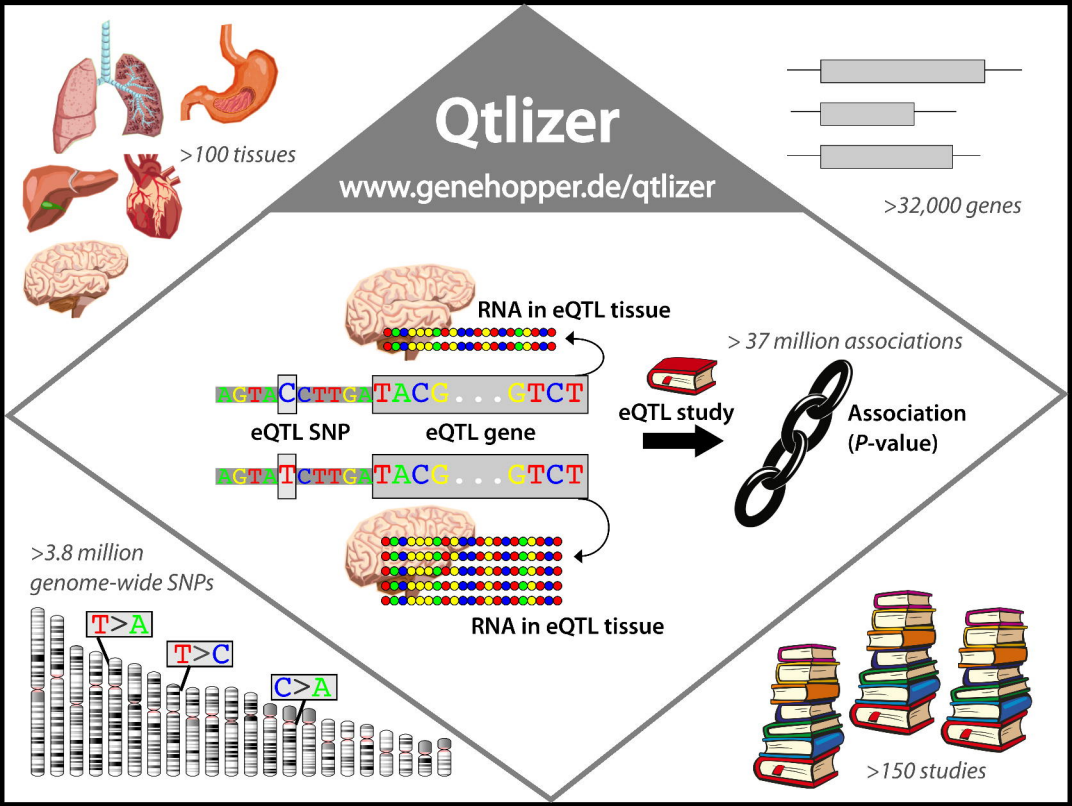
https://doi.org/10.1038/nature11049

29. Sauerwald,N., Singhal,A. and Kingsford,C. (2018) Analysis of the structural variability of topologically associated domains as revealed by Hi-C. *bioRxiv*, 10.1101/498972. https://doi.org/10.1101/498972

## TABLE AND FIGURE LEGENDS

**Figure 1.** (a) Data from various publicly available resources was integrated (b) applying an Extraction, Transformation and Loading (ETL) process. (c) Qtlizer can be queried for QTL data via the web-based user interface or programmatically by inputting lists of genetic variants and genes using a REST API. (d) Annotation results are displayed in a table view.


a)
Public Resources
eQTL
pQTL
reQTL
TAD
GWAS Catalog
1000G
b)
Extraction
Perl
Perl, Plink
Transformation
Calculation of LD, flags, counts
Loading
Perl, SQL
Qtlizer
Database
MySQL
Web Application
Play Framework, Twitter Bootstrap, Java, Javascript
Input Panel
Enter/paste variant and/or gene identifiers in any format (max. 50 identifiers)
rs2306374
Include proxy variants
r²
Dprime
0.8
Reference
1000G Phase 3 EUR
Annotate
Example: rs4284742, rs2070901, DEFA1, ENSG00000122194, hg19:19:45412079, hg38:6:12903725
c)
d)
List
#Results: 12
Download
Initial order
Aggregate
Select columns
rs2306374
All selected (21)
MRAS
E.g. <!
Apply
All selected (23)
All selected (8)
is_sw_sign, is_best, is_best_in_ld_group
All selected (11)
Index variant
Proxy variant
Gene
Distance
Tissue
P-value
Beta
EA
NEA
Source
Flags
GWAS Catalog
rs2306374 dbSNP | rs13324341 (0.94) | MRAS | 0 | Artery - Mammary | 1.6e-83 | - | - | - | Franzén et al. (Science 2016) PubMed | is_sw_sign, is_best, is_best_in_ld_group | rs2306374, MRAS
rs2306374 dbSNP | rs13324341 (0.94) | MRAS | 0 | Atherosclerotic aortic root | 1.3e-77 | - | - | - | Franzén et al. (Science 2016) PubMed | is_sw_sign, is_best, is_best_in_ld_group | rs2306374, MRAS
rs2306374 dbSNP | rs13324341 (0.94) | MRAS | 0 | Artery - Aorta | 1.5e-50 | 1.03 | T | C | GTEx v7 PubMed | is_sw_sign, is_best, is_best_in_ld_group | rs2306374, MRAS
rs2306374 dbSNP | rs13324341 (0.94) | MRAS | 0 | Artery - Tibial | 6.8e-31 | 0.47 | T | C | GTEx v7 PubMed | is_sw_sign, is_best, is_best_in_ld_group | rs2306374, MRAS
rs2306374 dbSNP | rs6807945 (0.94) | MRAS | 13.7 | Blood | 1.7e-29 | - | - | - | Franzén et al. (Science 2016) PubMed | is_sw_sign, is_best, is_best_in_ld_group | rs2306374, MRAS
rs2306374 dbSNP | rs6807945 (0.94) | MRAS | 13.7 | Muscle skeletal | 1.7e-29 | - | - | - | Franzén et al. (Science 2016) PubMed | is_sw_sign, is_best, is_best_in_ld_group | rs2306374, MRAS
rs2306374 dbSNP | rs13324341 (0.94) | MRAS | 0 | Heart - Atrial appendage | 1.2e-10 | 0.36 | T | C | GTEx v7 PubMed | is_sw_sign, is_best, is_best_in_ld_group | rs2306374, MRAS
rs2306374 dbSNP | rs139016349 (0.94) | MRAS | 0 | Artery - Coronary | 7.8e-10 | 0.54 | TTTC | T | GTEx v7 PubMed | is_sw_sign, is_best, is_best_in_ld_group | rs2306374, rs139016349, MRAS
rs2306374 dbSNP | rs12695685 (0.99) | MRAS | 0 | Muscle skeletal | 1.1e-8 | 0.24 | C | A | GTEx v7 PubMed | is_sw_sign, is_best, is_best_in_ld_group | rs2306374, MRAS
rs2306374 dbSNP | rs9851766 (0.98) | MRAS | 0 | Adipose - Subcutaneous | 9.2e-7 | -0.14 | G | A | GTEx v7 PubMed | is_sw_sign, is_best, is_best_in_ld_group | rs2306374, MRAS
rs2306374 dbSNP | rs13324341 (0.94) | MRAS | 0 | Heart - Left ventricle | 0.0000091 | 0.23 | T | C | GTEx v7 PubMed | is_sw_sign, is_best, is_best_in_ld_group | rs2306374, MRAS
rs2306374 dbSNP | rs1199338 (0.94) | MRAS | 0 | Heart | 0.000011 | - | - | - | Haploreg v4.1 PubMed | is_sw_sign, is_best, is_best_in_ld_group | rs2306374, MRAS


Qtlizer
www.genehopper.de/qtlizer
>100 tissues
>32,000 genes
RNA in eQTL tissue
> 37 million associations
AGTACCTTGATACG...GTCT
eQTL SNP
eQTL gene
eQTL study
Association
(P-value)
AGTATCTTGATACG...GTCT
RNA in eQTL tissue
>3.8 million
genome-wide SNPs
T>A
T>C
C>A
>150 studies