

## Data Elements for the QTL Viewer

The QTL Viewer utilizes R and several different libraries in order to calculate the data for various types of QTL projects. The following sections will explain each element in detail.

*Please note that some data element must be pre-computed.*

### RData Environment Overview

The following elements should be contained within the RData file.

*ensembl.version* – the numerical version of Ensembl

*genoprobs* – the genotype probabilities

*K* – the kinship matrix

*map* – list of one element per chromosome, with the genomic position of each marker

*markers* – marker names and positions

The following element is a special element and there must be at least one per RData file.

*dataset.\** - where \* should be a very short, unique and informative name. This element will contain most of the data and will be detailed in the section below.

*Exact case of element and variable names is very important.*

*Other meta data can be included in the RData file as long as there are no conflicting names.*

### Elements

#### ensembl.version

This specifies the genome release version for the genomic marker positions and for annotations attached to molecular phenotypes IF any, i.e. mRNA. Please see the documentation at Ensembl (<http://www.ensembl.org>) for build and version information.

**R data type:** numeric

#### genoprobs

This is the genotype probabilities and must be supplied by the user. This is a list with one element per chromosome of  $\mathbf{N} * \mathbf{K} * \mathbf{M}_j$  arrays, where:

**N** represents the number of mice  
**K** represents the number of founders  
**M<sub>j</sub>** represents the number of markers on chromosome **j**

`rownames(genoprobs)` are the same value of the `mouse.id` column in the `samples` element

`colnames(genoprobs)` are the founder strain symbol (A,B,C,D,E,F,G,H)

`dimnames(genoprobs[[j]])` are marker names on chromosome **j**

**R data type:** list, `calc_genoprobs`

*May be produced by `qtl2convert::probs_doqtl_to_qtl2`. Please see the documentation of `R/qtl2geno`.*

K

A list of kinship matrices, with one element per chromosome of **N \* N** matrices, where:

**N** represents the number of mice

`rownames(K)` are the same value of the `mouse.id` column in the `samples` element

`colnames(K)` are the same value of the `mouse.id` column in the `samples` element

**R data type:** list

*May be produced by `qtl2geno::calc_kinship(genoprobs, type="Loco")`. Please see the documentation of `R/qtl2geno`.*

map

This is a list with one element per chromosome of named numeric vector. Elements of the vector are positions along the chromosome in Mb units. Element names are marker names and must match the `dimnames` of `genoprobs`.

Users can download maps for MUGA platforms or for 69k pseudomarker grid.

**R data type:** list

May be produced by `qtl2convert::map_df_to_list`. Please see the documentation of `R/qtl2geno`.

## markers

Marker information containing the following information:

`marker.id` – character string, unique name of the marker

`chr` – character string, the chromosome

`pos` – numeric, position in Mbp

**R data type:** `tibble`

## The dataset.\* Element

The environment must contain at least one object of this type, multiple are allowed. The \* should be a very short, unique and informative name. It is for internal use only and will not appear in the QTL Viewer interface.

The main purpose of the `dataset.*` element is to store multiple datasets per RData file with informative information regarding the data.

The `dataset.*` element is a list that should contain the following named elements:

*annot.datatype* – annotations, where datatype is one of **mrna**, **protein**, or **phenotype**

*annot.samples* – annotation data for the samples

*covar.matrix* – a matrix of covariate data, samples (rows) x covariates (columns)

*covar.info* – information describing the covariates

*data* – either a matrix containing data or a list containing several kinds of data

*datatype* – one of **mrna**, **protein**, or **phenotype**

*display.name* – name of the dataset, for QTL Viewer display purposes

*lod.peaks* – a list of LOD peaks over a certain threshold

## annot.datatype

The `annot.datatype` element will have different data and column names depending on whether this is a **mrna**, **protein**, or **phenotype** dataset.

For **mrna**, the following fields are required:

`gene.id` – character string, Ensembl gene id

symbol – character string, Symbol of the gene  
chr – character string, chromosome  
start – numeric, position in Mbp  
end – numeric, position in Mbp

For **protein**, all **mrna** fields *PLUS* the following field:

protein.id – character string, Ensembl protein id

For **phenotype**, the following fields are required:

data.name – character string, phenotype id  
short.name – character string, short descriptive name  
R.name – character string, name used by R  
description – character string, phenotype description  
units – character string, measuring units  
category – character string, category if any  
R.category – character string, category used by R  
is.id – logical, should only be 1 TRUE  
is.numeric – logical, is this a numeric field  
is.date – logical, does this contain a date  
is.factor – logical, is this a factor  
factor.levels – character string, “:” separated values  
is.covar – logical, is this a covariate  
is.pheno – logical, is this an actual phenotype  
is.derived – logical, is this phenotype derived  
omit – logical, T to omit, F to include  
use.covar – character string, Ensembl gene id

**R data type:** tibble

*Extra information in the tibble will be ignored by the QTL Viewer.*

annot.samples

Annotations for the samples in this dataset. The unique identifying column is **mouse.id**. There should be a unique value for **mouse.id** in every row.

For the purpose of doing certain scans, there will need to be other columns that match the information stored in the covar.info element.

**R data type:** tibble

## covar.info

This element controls how we scan and interact with the RData object. The following columns must be present:

`sample.column`

name of the column in the `annot.sample` element

`display.name`

QTL Viewer uses this to display a nice name

`interactive`

TRUE for an interactive covariate, must also set `lod.peaks` if TRUE. If FALSE, `lod.peaks` value should be NA. This controls whether or not interactive scans are performed for a particular covariate.

`primary`

which covariate to display preselected in the Effect Plot

`lod.peaks`

named tibble in the `lod.peaks` element

**R data type:** tibble

## covar.matrix

Covariates data, samples (rows) x covariates (columns) as produced by `model.matrix`.

**R data type:** matrix

## data

This element is either a matrix or a list.

If it is a matrix, there is one and only set of data for this dataset.

If it is a `list`, each named element in the list should be a matrix with the following controlled vocabulary for the names:

**rz**  
**norm**  
**raw**  
**log**  
**transformed**

Each matrix will contain numerical data with samples (rows) by annotations (columns).

**R data type:** `matrix` or `list`

`datatype`

This will be used to identify the type of dataset. This is a controlled vocabulary consisting of the following values:

**mrna**  
**protein**  
**phenotype**

Based upon the value of this element, the QTL Viewer will treat the data as accordingly.

**R data type:** `character`

`display.name`

This will be used to display the name of the dataset to the user in the QTL Viewer. This will be used in a dropdown menu to switch among the datasets.

**R data type:** `character`

`lod.peaks`

This is a list with each value in the list being either **additive** (the default) or one of the interactive covariates (if set in `covar.info`). The **additive** values should always be present.

The `covar.info` element should have values with `interactive` set to `TRUE` and `lod.peaks` set to the name of the element in this list.

Depending on the value of `datatype` (**mrna**, **protein**, **phenotype**), the annotation column identifier will match to the appropriate column in the `annot.datatype` element.

The following shows the required fields in each tibble.

If `datatype` is **mrna**, the following fields are required:

`gene.id`  
the Ensembl gene identifier in the `annot.mrna` element

`marker.id`  
the marker identifier in the `markers` element

`lod`  
the lod score

If `datatype` is **protein**, the following fields are required:

`protein.id`  
the Ensembl protein id in the `annot.protein` element

`marker.id`  
the marker identifier in the `markers` element

`lod`  
the lod score

If `datatype` is **phenotype**, the following fields are required:

`data.name`  
the unique identifier in the `annot.pheno` element

`marker.id`  
the marker identifier in the `markers` element

`lod`  
the lod score

**R data type:** list

