

RTeQTL Database Manual (version 1.0)

June 13, 2014

Contact:

Baoshan Ma (mabaoshan@dlmu.edu.cn), Jinyan HUANG (hiekeen@gmail.com) and Liming Liang (lliang@hsph.harvard.edu)

Web address:

<http://eqtl.rc.fas.harvard.edu>

Citation:

Baoshan Ma, Jinyan Huang and Liming Liang.(2014) RTeQTL: Real-time online engine for expression quantitative trait loci analyses. database: the journal of biological databases and curation.

I. Datasets and background

Gene expression in lymphoblastoid cell lines was characterized in two independent datasets, one sample of 405 siblings using Affymetrix HG U133 Plus 2.0 chips (more than 54,000 transcription probesets, MRCA dataset) and the other sample of 550 siblings using Illumina Human6 V1 array (more than 47,000 transcription probes, MRCE dataset). All samples are Caucasian of British descendant. Among these individuals, 928 were also genotyped at >300,000 SNPs using the Illumina HumanHap300 arrays, with additional genotypes for 2 million SNPs in the HapMap Project filled in using imputation. These two datasets together identified genome-wide significant *cis* and *trans* eQTLs for 14,177 genes (Liang et al, 2013 Genome Research 23(4):716-26).

II. What can be retrieved from this database browser?

A. Single SNP eQTL analysis

User can query the association result between any one gene and any one SNP available on the Illumina or Affymetrix expression platforms and HapMap 2 SNP panel. For example, a gene might be a candidate in the pathway of a disease associated SNP. It would be interesting to see if the disease associated SNP might be associated with the expression of the candidate gene. The statistical model for this analysis is:

$$\text{expression} = \alpha + \text{SNP1} \cdot \beta_1 + Z + \varepsilon$$

where β_1 is the fixed effect for SNP1, Z is random effect for family relatedness and ε is residual error.

B. Two SNPs conditional eQTL analysis

User can query the association result between any one gene and any SNP adjusting for any other SNP chosen by the user. For example, a disease associated SNP might be an eQTL of the candidate gene but it might not be the peak eQTL showing the strongest effect on the gene expression. It would be interesting to examine whether the disease associated SNP shows independent effect on gene expression while adjusting for the peak eQTL of the same gene. The statistical model for this analysis

is:

$$\text{expression} = \alpha + \text{SNP1} \cdot \beta_1 + \text{SNP2} \cdot \beta_2 + Z + \varepsilon$$

where β_1 is the fixed effect for SNP1 and β_2 is the fixed effect for SNP2, Z is random effect for family relatedness and ε is residual error.

III. User input

A. For single SNP eQTL analysis

Step 1. Type SNP1 rs name in the “SNP1” column and “-” (short dash) in the “SNP2” column.

Step 2.

- (a) If you know the probe ID, you may type probe ID in the “Probe for analysis” column.
- (b) If you do not know the probe ID but the gene of interest, you may select/type the gene name in the “Gene” column and then select the corresponding probe ID in the box below. The probe ID will be added to the “Probe for analysis” column after you click the “add” button.
- (c) You can also browse and select probe ID from the “All Probesets” column and click the “add” button to include in the “Probe for analysis” column.

Step 3. Choose whether the analysis should be stratified by sex from the drop-down menu “Stratify by gender”

- (a) When “Male & Female” is chosen, result based on the full samples will be returned.
- (b) When “Male” is chosen only results based on male samples will be return. Similarly for the “Female” option.
- (c) When “Gender Specific” is chosen, results based on male samples and results based on female samples will be returned.

Step 4. Click the “Submit” button to get result table.

B. For two SNPs conditional eQTL analysis

Step 1. Type SNP1 rs name in the “SNP1” column and SNP2 rs name in the “SNP2” column.

Step 2.

- (a) If you know the probe ID, you may type probe ID in the “Probe for analysis” column.
- (b) If you do not know the probe ID but the gene of interest, you may select/type the gene name in the “Gene” column and then select the corresponding probe ID in the box below. The probe ID will be added to the “Probe for analysis” column after you click the “add” button.
- (c) You can also browse and select probe ID from the “All Probesets” column and click the “add” button to include in the “Probe for analysis” column.

Step 3. Choose whether the analysis should be stratified by sex from the drop-down menu “Stratify by gender”

- (a) When “Male & Female” is chosen, result based on the full samples will be returned.
- (b) When “Male” is chosen only results based on male samples will be return. Similarly for the “Female” option.
- (c) When “Gender Specific” is chosen, results based on male samples and results based on female samples will be returned.

Step 4. Click the “Submit” button to get result table.

C. Batch model

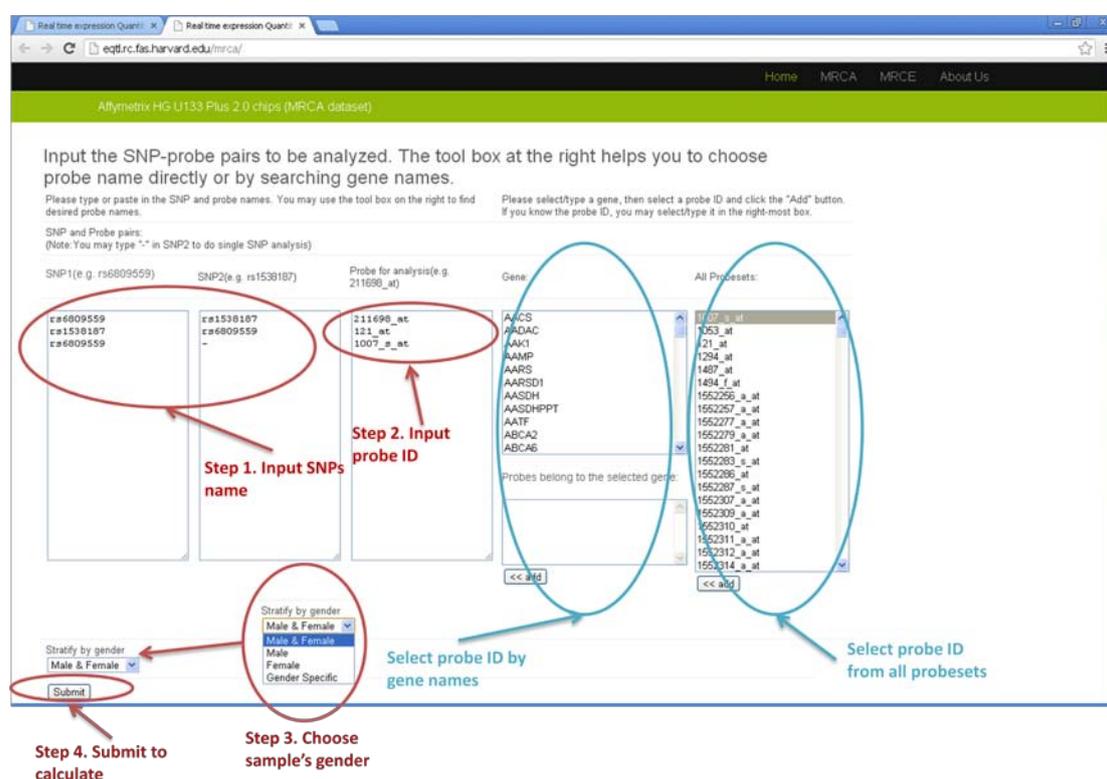
For analyses involving multiple pairs of gene and SNPs (batch mode), the list of SNP rs names and probe names can be copied and pasted into the corresponding columns (SNP1, SNP2, Probe). For example, the input setting shown in Figure 1 will return eQTL results for the following three models:

Expression of 211698_at = rs6809559 + rs1538187 (Two-SNP analysis)

Expression of 121_at = rs1538187 + rs6809559 (Two-SNP analysis)

Expression of 1007_s_at = rs6809559 (Single-SNP analysis)

Figure 1. Input interface webpage (MRCA database as an example)



IV. Results output

A. How to display or download the result

Results table will be displayed on the web page and can also be downloaded to desktop computer by clicking the link "[Download the table as csv file](#)" on top of the result table.

B. Content of output file

Each row corresponds to result for each pair of gene and SNPs. Full detail for regression results are available, including effect size, standard error, test statistics, pvalue as well as gene annotation of the expression probe and SNPs (chromosomal position, allele label, allele frequency, MaCH imputation quality score, Rsq- see Table 1), and the "Gender" column indicates the samples used for analysis (male, female or both). For single SNP eQTL result, the corresponding outputs of the

SNP2 are masked as “NA”. The last row is used to display any error message, for example, if the input SNP or probe name could not be found in our data files. See Figure 2 for an output example.

Figure 2. The output webpage (MRCA database as an example)

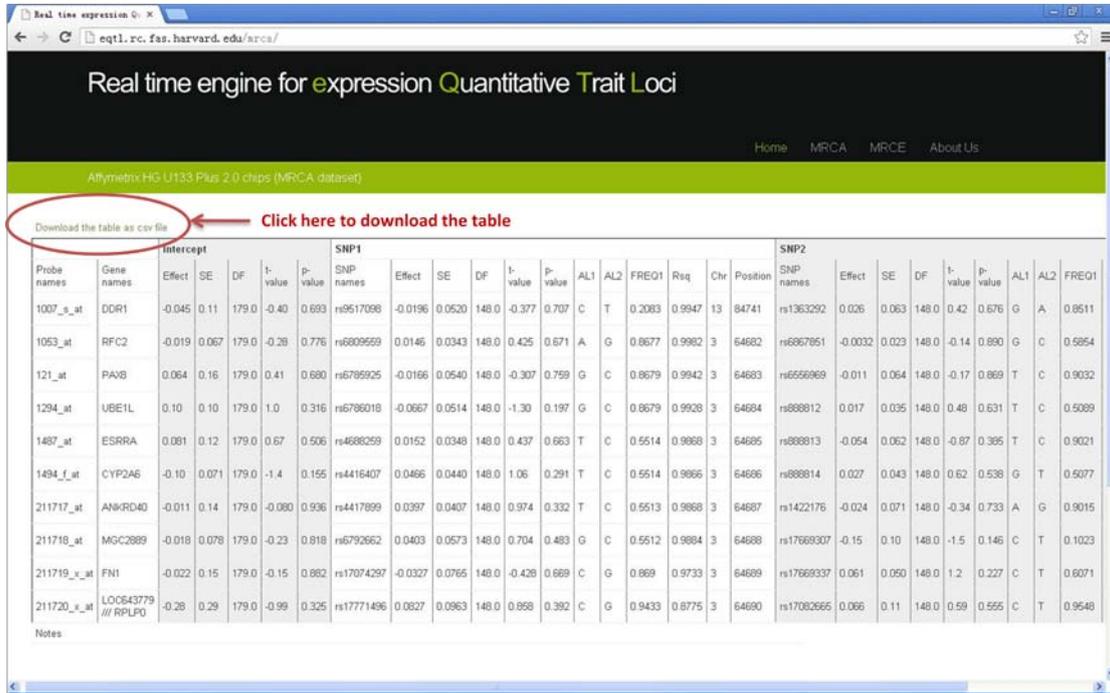


Table 1. Columns in output table

Column Order	Name	Notes
1	Effect	effect size $\hat{\beta}$ from linear mixed model. The amount of increase/decrease expression by one copy of the Allele 1 in the unit of one standard deviation
2	SE	standard error of $\hat{\beta}$
3	DF	degrees of freedom of the test
4	t-value	$\hat{\beta} / \text{sd}(\hat{\beta})$
5	p-value	the probability of $P\{t > T \}$
6	AL1/2	the allele1/2 label
7	FREQ1	frequency for Allele1
8	Chr	chromosome
9	Position	position on chromosome (NCBI 36)
10	Rsq	MaCH imputation quality score, which estimates the squared correlation between imputed and true allele counts
11	Gender	Sample's gender when analysis stratified by gender

FAQ

Q: How do I contact you?

A: By e-mail: mabaoshan@dlmu.edu.cn, hiekeen@gmail.com, lliang@hsph.harvard.edu

If you get any question when using the RTeQTL database and need our help, please contact us. We will reply you as soon as possible.

Q: Which kinds of internet browsers are needed?

A: Popular browsers such as Google Chrome, Firefox, Safari and IE are good to use.

Q: What is the purpose of this database?

A: The database can efficiently provide eQTL association results that are not available in existing eQTL databases browsers: (1) single SNP and (2) two-SNP conditional eQTL effects on gene expression regardless magnitude of pvalues. This is the only tool that can assess the independent effect of disease or trait associated SNP on gene expression conditioning on other eQTL of the same gene or other SNP of interest.

Q: How do I submit a query to the database?

A: Type the SNP1, SNP2 rs names and Probe ID in the corresponding columns, and click the “Submit” button. You may find detailed operation steps in the manual.

Q: Do I need to register on your website?

A: No, the database is free for all users.

Q: Do I have to fill in SNP name for both SNP1 and SNP2 columns?

A: No, you can query eQTL result for single SNP by simply put “-” (short dash) in the SNP2 column.

Q: What is the best way to do query in batch mode?

A: To query a large number of gene and SNP pairs. You may copy and paste the probe names, SNP names in corresponding columns. Results will be displayed in a table where each row is for one gene-SNP pair.

Q: Can we download the output table?

A: Yes. You can download it to desktop computer by clicking the link "[Download the table as csv file](#)" on top of the result table.

Q: How many pairs of probe and SNP can be submitted in one time?

A: We did not set the maximum number of probe-SNP pairs. The analysis for 100 probe-SNP pairs takes only 20 seconds. We recommend an initial try to be no more than 1000 probe-SNP pairs in one query.

Q: Why some genes cannot be found in your database?

A: Because some genes might not be covered by the Affymetrix U133 Plus 2 or Illumina Human6 V1 platform. Also try to use other alias of the gene name.