

SCeQTL is an R package that uses zero-inflated negative binomial regression to do eQTL analysis on single-cell data. It can distinguish two type of gene-expression differences among different genotype groups. It's more suitable to use SCeQTL to identify eQTLs from single-cell data. It can also be used for finding gene expression variations associated with other grouping factors like cell lineages. Following is the detail information and usage of this program. You can also found them in [README](#).

[R Packages](#) is a book based around this workflow.

Authors

Hu Yue y-hu10@qq.com

Installation

To install the developmental version from [GitHub](#):

```
if(!require(devtools)) install.packages("devtools")
devtools::install_github("XuegongLab/SCeQTL")
```

To install from source package, you could download the R package from [Mac and Linux version](#) or [Windows version](#) and run the command in R:

```
install.packages(<source_code_file>, repos=NULL, type="source")
```

Requirement

Prerequisite(R-packages):

```
qvalue
ggplot2
doParallel
pscl
```

Input:

SCeQTL takes two inputs: `gene` and `snp`.

The input `gene` is a scRNA-seq **raw read counts matrix**. The rows of the matrix are genes and columns are cells.

The other input `snp` is a **genotype matrix**, where each element should be 0/1/2 indicating three types of genotypes. The order of columns should be the same with `gene`.

TestData:

Users can load the test data in `SceQTL` by

```
library(SceQTL)
data(test)
```

The toy data `gene` in `test` is a scRNA-seq read counts matrix which has 151 genes (rows) and 1529 cells (columns).

```
dim(gene)
gene[1:6, 1:6]
```

The object `snp` in `test` is a matrix of genotype which has three levels and equal length to the column number of `snp`.

```
dim(snp)
snp[1:6, 1:6]
```

Usage

Here is an example to run `SceQTL` with read counts matrix input:

```
# Load library and the test data for SceQTL
library(SceQTL)
data(test)

# check whether the non-zero part could fit negative binomial well
# This function may fail since It's possible that the random picked gene can't
# be fit to negative binomial distribution, all zero value for example
checkdist(gene)

# normalize gene matrix using DEseq method
normalize(gene)

# Detecting the SceQTL
result <- cal.pvalue(gene, snp)

# Picking one sample to visualize
check.sample(gene[1,], snp[1,])
```

Output

SceQTL will output the p-value of each gene-snp pair. You could also get q-value by running `cal.qvalue(gene, snp)` instead of `cal.pvalue(gene, snp)`.

```
SceQTL::cal.qvalue(gene[1:10,], snp)
```

The result is like the picture below.

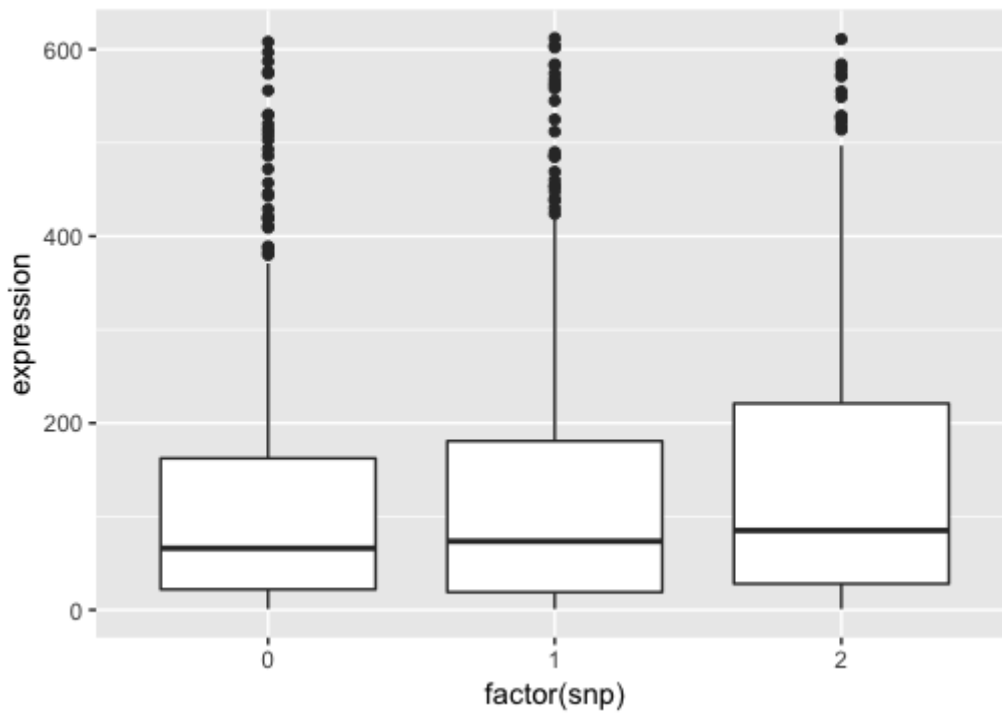
| | gene | snp | pvalue | qvalue |
|----|----------|-----|-------------|-----------|
| 91 | ARID3A | 10 | 0.001585913 | 0.1045663 |
| 27 | HIPK3 | 3 | 0.006616460 | 0.1454175 |
| 30 | HIPK3 | 3 | 0.004698696 | 0.1454175 |
| 7 | CDC5L | 1 | 0.039188731 | 0.3061283 |
| 15 | TPP2 | 2 | 0.022037359 | 0.3061283 |
| 23 | HIPK3 | 3 | 0.029422254 | 0.3061283 |
| 24 | HIPK3 | 3 | 0.041786275 | 0.3061283 |
| 60 | S100A13 | 6 | 0.035170908 | 0.3061283 |
| 87 | SLC39A10 | 9 | 0.031555068 | 0.3061283 |
| 10 | CDC5L | 1 | 0.064629195 | 0.3200852 |
| 14 | TPP2 | 2 | 0.063997435 | 0.3200852 |

Then you may interested in the significant gene-snp pairs. You could stat and visualize it by `check.sample`. Here take the most significant pair for example.

```
SCEQTl::check.sample(gene[rownames(gene)=='ARID3A'], snp[10,])
```

Then you could get the statistic about the pair and a boxplot showing the relationship.

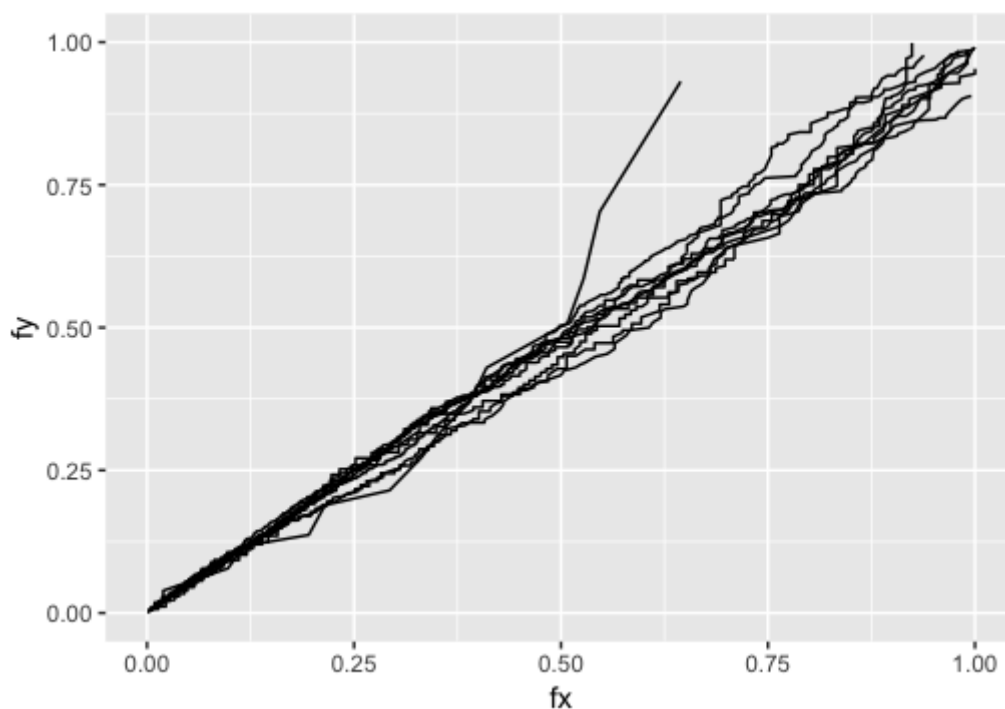
```
zero ratio of group 0 0.08074534
zero ratio of group 1 0.05660377
zero ratio of group 2 0.07674419
mean of group 0: 105.6087
mean of group 1: 120.1486
mean of group 2: 134.3442
non-zero part mean of group 0: 114.8851
non-zero part mean of group 1: 127.3575
non-zero part mean of group 2: 145.5113
non-zero part standard error of group 0: 129.1127
non-zero part standard error of group 1: 144.3739
non-zero part standard error of group 2: 150.4358
```



Before run `cal.qvalue`, you may want to check whether the distribution fit the binomial negative assumption well by `checkdist`

```
checkdist(gene)
```

A Q-Q plot will be shown. It will randomly pick some genes so you may get different kind of result.



Help

Use the following code in R to get access to the help documentation for `DEsingle`:

```
# Documentation for checking and preparing gene expression dataset
?checkdist
?normalize
```

```
# Documentation for finding single cell eQTL
```

```
?cal.pvalue
```

```
?cal.qvalue
```

```
# Documentation for checking single gene-snp pair
```

```
?check.sample
```

```
# Documentation for TestData
```

```
?test
```

```
?snp
```

```
?group
```

You are also welcome to contact the author by email for help.

Code of conduct

Please note that the SCellQTL project is released with a [Contributor Code of Conduct](#). By contributing to this project, you agree to abide by its terms.