MLExpResso

Cheat Sheet

MI DATA LAB

Introduction

MLExpResso is an R package for integrative analyses and visualization of gene expression and DNA methylation data.

Key functions of this package are:

- identification of DMR - differentially methylated regions,
- identification of genes with affected expression,
- identification regions with changes in expression and methylation,
- visualization of identified regions.

The joint modeling and visualization of genes expression and methylation improve interpretability of identified signals.

The methodology is supplemented with example applications to The Cancer Genome Atlas data.

Data

Expression

Datasets for testing expression differences must contain per gene read counts. Columns should correspond to genes, rows to samples. As input for tests corresponds to expression, the **calculate_test()** function expects count data in the form of a matrix of integer values. The value in the i-th row and the j-th column tells how many reads can be assigned to gene j in sample i.

Note that for some tests the values in the matrix should be un-normalized counts, so transformed or normalized values such as counts scaled by library size should not be used as input.

Methylation

Datasets for testing methylation differences must contain a percentage of methylation for each CpG probe. Columns should correspond to probes, rows to samples. As input for tests corresponds to methylation, the **calculate_test()** function expects values aggregated to genes. For aggregation of probes for genes is used the **aggregate_probes()** function. The value in the i-th row and the j-th column tells how much gene j is methylated in sample i.

Expression

MLExpResso::calculate_test(data, condition, test)

Function **calculate_test()** computes log folds, p-values and means for choosen test for data from methylation or expression.

Value	Test
'ttest'	student's t-tets
'nbinom2'	negative binomial test
'lrt'	likelihood-ratio test
'qlf'	quasi-likelihood F-test

Example

```
library(MLExpResso)
library(MLExpRessodata)
exp <- BRCA_mRNAseq_chr17[ ,-1]
gr_exp <- BRCA_mRNAseq_chr17[ ,1]
gr_exp <- ifelse(gr_exp=='LumA', 'LumA', 'other')
res_exp <- calculate_test(exp, gr_exp, 'lrt')
```

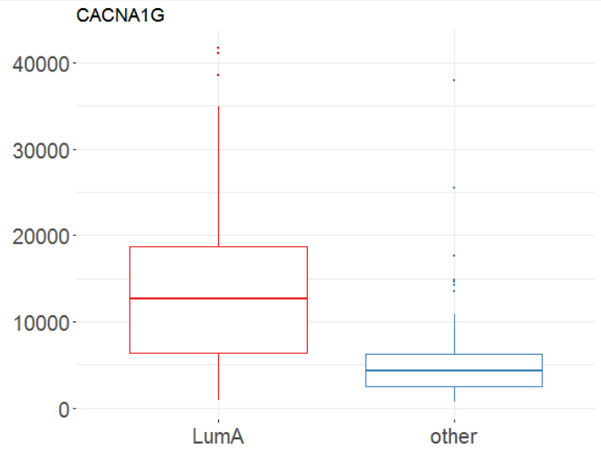
	id	log2.fold	pval	mean_LumA	mean_other	mean
1	AURKB	2.3399204	3.191000e-32	539.0426	2323.8868	1485.01
2	CBX2	2.8950625	2.834335e-26	632.5106	4296.6038	2574.48
3	KPNA2	1.4472883	8.551812e-24	11547.36	26427.38	19433.77
4	PRR11	3.8221475	2.286874e-22	396.383	3479.981	2030.69
5	BIRC5	1.9889977	1.953941e-21	1957.085	6658.358	4448.76

MLExpResso::plot_diff_boxplot(data, condition, gene)

Function plot_diff_boxplot() generates a boxplot of values from choosen data frame column with division in groups (two or more).

Example

```
plot_diff_boxplot(exp, gr_exp, 'CACNA1G')
```



Methylation

MLExpResso::calculate_test(data, condition, test)

MLExpResso::aggregate_probes(data)

Function **aggregate_probes()** aggregates CpG probes to corresponding genes using, by default, the IlluminaHumanMethylation data.

Value	Test
'ttest'	student's t-test
'methyanalysis'	quasi-likelihood F-test

Example

```
library(MLExpResso)
library(MLExpRessodata)
met <- aggregate_probes(BRCA_methylation_chr17)
gr_met <- BRCA_methylation_chr17[ ,1]
gr_met <- ifelse(gr_met=='LumA', 'LumA', 'other')
res_met <- calculate_test(met, gr_met, 'ttest')
```

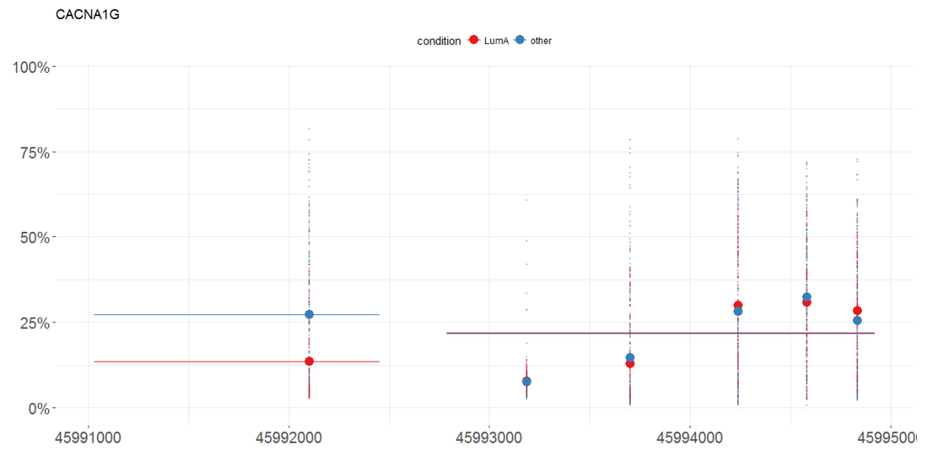
	id	log2.fold	pval	mean_LumA	mean_other	mean
1	ICAM2	-0.1515131975	3.754116e-17	0.2547275	0.4062407	0.33308008
2	RILP	-0.0507369105	2.575168e-13	0.3079069	0.3586438	0.33414469
3	PIPOX	0.1150555802	5.360053e-12	0.4242804	0.3092248	0.36478123
4	TNFSF12	-0.1341285454	5.867083e-12	0.1791401	0.3132686	0.24850250
5	CD7	0.0982268951	1.641919e-11	0.8635077	0.7652808	0.81271124

MLExpResso::plot_methylation_path(data, condition, gene)

Function **plot_methylation_path()** visualises a chosen gene with marked CpG probes. It shows the mean methylation level for each probe in group. Also we can exact the line corresonding to gene. In this case we see what are the locations of probes on gene in HG18 coordinates. We can as well draw a locations of CpG islands.

Example

```
plot_methylation_path(BRCA_methylation_chr17 ,gr_met,  'CACNA1G')
```



Visualizations

Plot_volcanoes()

MLExpResso::plot_volcanoes(data.m, data.e, condition.m, condition.e, gene, test.m, test.e)

Function **plot_volcanoes()** generates a dashboard with volcano plots for expression and methylation. Also it adds a tables with basic statistics for chosen gene.

Example

```
plot_volcanoes(met, exp, gr_met, gr_exp, 'CACNA1G', res_met, res_exp)
```

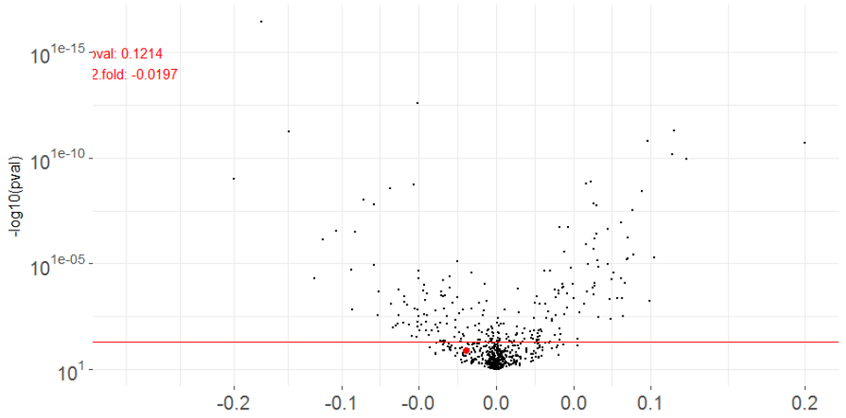
CACNA1G

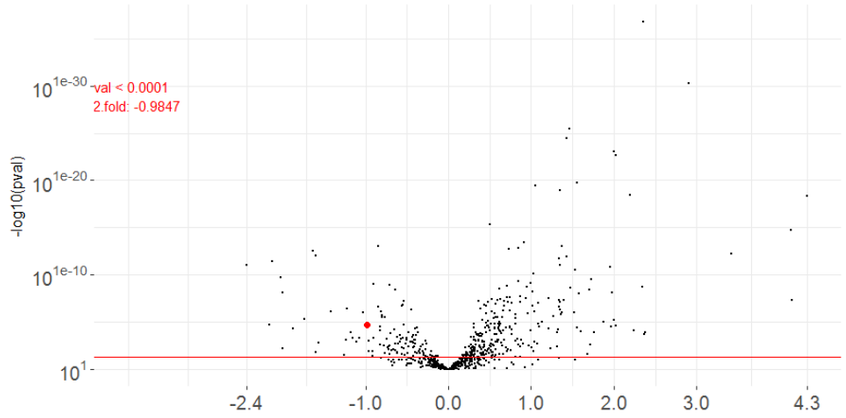
Methylation

	min	1st Q	med	mean	3rd Q	max	count
LumA	0.05	0.14	0.19	0.21	0.27	0.49	155
other	0.04	0.13	0.2	0.23	0.31	0.63	166

Expression (cpm)

	min	1st Q	med	mean	3rd Q	max	count
LumA	977	6351	12628	14302	18716	41717	47
other	752	2556	4360	6185	6314	37883	53





Plot_gene()

MLExpResso::plot_genes(data.m, data.e, condition.m, condition.e, gene)

Function **plot_gene()** generates a dashboard with methylation path for methylation and boxplots for groups for chosen gene.

Example

```
plot_volcanoes(met, exp, gr_met, gr_exp, 'CACNA1G', res_met, res_exp)
```