MExpResso: differential expression and methylation analysis

Case study using RTCGA data

Aleksandra DÄ. . . browska, Alicja Gosiewska

Contents

1 Package (Abstract?)

It is considered that the result of increased methylation is decreased gene expression. While, recent studies suggest that the relationship between methylation and expression is more complex than was previously thought. The package `MExpResso` provides methods to test for differential expression and methylation by use of the negative binonial distribution and t-test. Additionaly 'MExpResso allows to visualize results in a simple way.

2 Standard Workflow

In this vignette we will work with the data sets containing information about gene expression and methylation for patients with breast cancer. We will analyze differences in methylation and expression for patients with different subtypes of BRCA cancer.

2.0.0.1 `BRCA_methylation_chr17` data set

In this section, we will work with the methylation level data from TCGA database. Package contains `BRCA_methylation_all` dataset. `BRCA_methylation_all` contains information about methylation of CpG islands for patients with breast cancer. Rows of this data set correspond to patients, more precisely, to samples taken from patients. First column `SUBTYPE`corresponds to a subtype of BRCA cancer, next columns correspond to CpG islands. Values inside the table indicate the methylation level of CpG island for specified sample.

```
library(MetExpR)
```

```
## Setting options('download.file.method.GEOquery'='auto')
```

```
## Setting options('GEOquery.inmemory.gpl'=FALSE)
```

```
## No methods found in "genoset" for requests: toGenomeOrder
```

```
##
```

```
head(BRCA_methylation_all)[1:5,1:4]
```

```
##                               SUBTYPE cg00000292 cg00002426 cg00003994
## TCGA-A1-A0SD-01A-11D-A112-05    LumA  0.6055526 0.06197412 0.33345006
## TCGA-A2-A04N-01A-11D-A112-05    LumA  0.7433957 0.07044132 0.32317983
## TCGA-A2-A04P-01A-31D-A032-05   Basal  0.2897206 0.25927969 0.02402149
## TCGA-A2-A04Q-01A-21D-A032-05   Basal  0.7898920 0.63619354 0.10885097
## TCGA-A2-A04T-01A-21D-A032-05   Basal  0.6512270 0.27268734 0.03413620
```

In this analysis we would like to find genes with different methylation and expression. At first we need to use function `map_to_gene`, which generates new data frame with CpG islands mapped to genes.

```
BRCA_methylation_gen <- map_to_gene(BRCA_methylation_all[,-1])
head(BRCA_methylation_gen)[1:5,1:4]
```

```
##                               c.0.627525304291071..0.61889803781234..0.561973209745455..0.64687254168
## TCGA-A1-A0SD-01A-11D-A112-05                                                                      0.6
## TCGA-A2-A04N-01A-11D-A112-05                                                                      0.6
## TCGA-A2-A04P-01A-31D-A032-05                                                                      0.5
## TCGA-A2-A04Q-01A-21D-A032-05                                                                      0.6
## TCGA-A2-A04T-01A-21D-A032-05                                                                      0.6
##                                     X7A5      A1BG      A2BP1
## TCGA-A1-A0SD-01A-11D-A112-05 0.13199966 0.9686056 0.03378443
## TCGA-A2-A04N-01A-11D-A112-05 0.11862215 0.9785676 0.06679088
## TCGA-A2-A04P-01A-31D-A032-05 0.08032758 0.9793897 0.29396794
## TCGA-A2-A04Q-01A-21D-A032-05 0.08958826 0.9718291 0.21287231
## TCGA-A2-A04T-01A-21D-A032-05 0.13135664 0.9801575 0.21864058
```

In this case we have two conditions, connected with subtypes of breast cancer.

Before we go to the testing, we need to define condition values for each sample. We would like to test for differences between `LumA` subtype and `other` subtypes of breast cancer, so we create a vector, which each element corresponds to a sample. Our division into this two groups relies on numbers of occurences of each subtype. The `LumA` subtype is the most common, in case of breast cancer.

```
condition_met <- ifelse(BRCA_methylation_all$SUBTYPE=="LumA","LumA", "other")
head(condition_met, 8)
```

```
## [1] "LumA"  "LumA"  "other" "other" "other" "other" "LumA"  "other"
```

2.0.0.2 `BRCA_mRNAseq_all` data set

Data set `BRCA_mRNAseq_all` contains information about gene expression. This data set contains per-gene read counts computed for genes for 736 patients with breast cancer. Rows of this data set correspond to samples taken from patients. First column `SUBTYPE`corresponds to a subtype of BRCA cancer, next columns correspond to genes.

```
BRCA_mRNAseq_all[1:5,1:5]
```

```
##                               SUBTYPE A1BG A1CF A2BP1 A2LD1
## TCGA-A1-A0SB-01A-11R-A144-07  Normal  164    0    22   127
## TCGA-A1-A0SD-01A-11R-A115-07    LumA  546    0     1   331
## TCGA-A1-A0SE-01A-11R-A084-07    LumA 1341    0     2   498
## TCGA-A1-A0SF-01A-11R-A144-07    LumA  836    1     0   526
## TCGA-A1-A0SG-01A-11R-A144-07    LumA  512    3    25   451
```

In our example we will test for differential expression between groups with LumA breast cancer subtype and other subtypes of that cancer. Again we will use vector `conditions`, which consist of two values corresponds

to subtype of breast cancer: `LumA` and `other`.

```
condition_exp <- ifelse(BRCA_mRNAseq_all$SUBTYPE=="LumA","LumA","other")
head(condition_exp, 8)
```

```
## [1] "other" "LumA"  "LumA"  "LumA"  "LumA"  "LumA"  "other" "LumA"
```

3 Testing

```
genes_comparison <- comparison_table(BRCA_mRNAseq_all[ ,-1], BRCA_methylation_gen, condition_exp, condi
head(genes_comparison)
```

```
##       id nbinom2.log2.fold  nbinom2.pval ttest.log2.fold   ttest.pval
## 1   A1BG        -0.4855862  7.330514e-13      0.01318900 1.119637e-01
## 2  A2BP1        -0.8461944  4.315880e-07     -0.05677500 1.019935e-02
## 3    A2M         0.2352576  1.675724e-03      0.03310777 1.686741e-01
## 4  A2ML1         4.1919667 8.164823e-201      0.17016379 1.019044e-12
## 5 A4GALT        -0.1927530  9.741784e-03     -0.00841392 5.203636e-01
## 6  A4GNT         0.2208814  8.021080e-02      0.06387711 1.957709e-05
```

3.1 Choosing most different genes

Sorting ...

```
genes_comparison_sorted <- genes_comparison[order(genes_comparison$ttest.pval), ]
head(genes_comparison_sorted)
```

```
##              id nbinom2.log2.fold  nbinom2.pval ttest.log2.fold
## 10601 TNFRSF10A         0.2038466  2.272347e-04       0.1401746
## 10218     TBX19         1.0212930  3.226882e-49      -0.3479977
## 4958     IGFALS        -1.9559531  1.892160e-54      -0.1581887
## 6214      MGST1         0.4688093  8.122084e-08       0.2583506
## 5895      LUC7L        -0.1772576  5.466990e-05       0.1414415
## 8870       RRM2         1.4786220 5.630314e-103       0.1051548
##         ttest.pval
## 10601 2.568571e-22
## 10218 2.252784e-21
## 4958  6.144850e-21
## 6214  7.121403e-20
## 5895  6.298008e-18
## 8870  7.790260e-18
```

IGFALS!

4 Visualization

Visualizing chosen gene - IGFALS.

```
test_exp <- genes_comparison[ ,c(1,2,3)]
test_met <- genes_comparison[ ,c(1,4,5)]
```

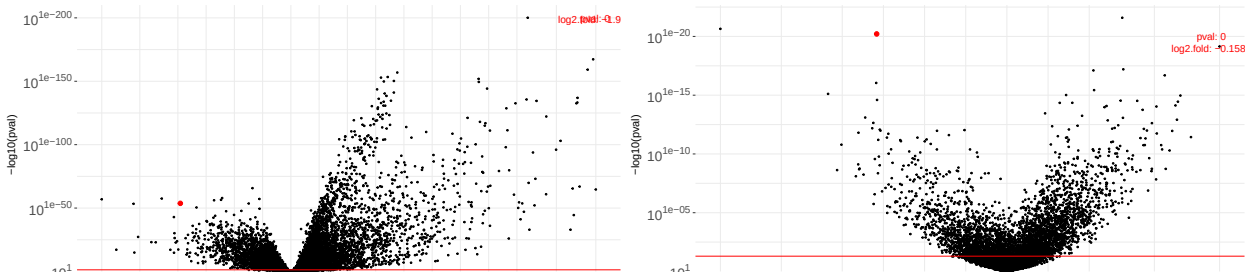
```
visual_volcano(condition_exp, condition_met,
               BRCA_mRNAseq_all[,-1], BRCA_methylation_all[,-1],
               "IGFALS",
               list(test_exp), list(test_met),
               values=TRUE)
```

Methylation

	min	1st Q	med	mean	3rd Q	max	count
LumA	0.15	0.47	0.54	0.54	0.63	0.89	139
other	0.34	0.59	0.73	0.7	0.82	0.94	153

Expression (cpm)

	min	1st Q	med	mean	3rd Q	max	count
LumA	29	407	994	2214	2444	30771	359
other	0	74	160	544	503	21792	377



Note that `visual_gene` methylation require data frame with cpg islands, not genes.

```
visual_gene(condition_exp, condition_met,
            BRCA_mRNAseq_all[,-1], BRCA_methylation_all[,-1],
            "IGFALS")
```