Vignette Title

Aleksandra Dąbrowska, Alicja Gosiewska

2017-04-06

Contents

Package

Package . . . the detection of differentially expressed and methylated genes.

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 between methylation and expression for patients with different subtypes of BRCA cancer.

Function test_diff

The main function of the package is `test_diff`. It allows to find differences between genes methylation or expression, taking into account additional information about samples.

Methylation

Info about methylation. . .

In this section, we will work with the methylation level data from TCGA database.
Data set `BRCA_methylation_chr17` contains information about methylation of CpG islands located on 17th chromosome for patienst with breast cancer.

```
load("BRCA_methylation_chr17.rda")
head(BRCA_methylation_chr17)[1:5,1:4]
```

```
##                              SUBTYPE cg00021527 cg00031162  cg00032227
## TCGA-A1-A0SD-01A-11D-A112-05    LumA 0.03781858  0.7910348 0.006391233
## TCGA-A2-A04N-01A-11D-A112-05    LumA 0.01437552  0.7359370 0.008752293
```

```
## TCGA-A2-A04P-01A-31D-A032-05    Basal 0.01360124   0.6967802 0.009442039
## TCGA-A2-A04Q-01A-21D-A032-05    Basal 0.01525656   0.5341244 0.014674247
## TCGA-A2-A04T-01A-21D-A032-05    Basal 0.01167384   0.7378100 0.012251559
```

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

```
library(MetExpR)
BRCA_methylation_chr17_gen <- map_to_gene(BRCA_methylation_chr17[,-1])
head(BRCA_methylation_chr17_gen)[1:5,1:4]
```

```
##                                 AANAT    AARSD1      AATF      AATK
## TCGA-A1-A0SD-01A-11D-A112-05 0.7148533 0.8625816 0.24294092 0.7835302
## TCGA-A2-A04N-01A-11D-A112-05 0.5850106 0.8355825 0.21367129 0.8466190
## TCGA-A2-A04P-01A-31D-A032-05 0.4495537 0.8786166 0.03277413 0.3417919
## TCGA-A2-A04Q-01A-21D-A032-05 0.7120650 0.8819490 0.03460160 0.7264985
## TCGA-A2-A04T-01A-21D-A032-05 0.6010397 0.7739978 0.02501599 0.6276399
```

Function `test_diff` allows us to test for differences between the base means for two or more conditions.

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 vector, which each element corresponds to a sample.

```
condition <- ifelse(BRCA_methylation_chr17$SUBTYPE=="LumA","LumA", "other")
head(condition,8)
```

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

T-test

One of the tools for testing differences between values is t-test. To use it in `test_diff` function, we set value of parameter `test` on "ttest".

```
test.mety <- test_diff(BRCA_methylation_chr17_gen, condition, test="ttest")
```

As a result we obtain a data frame with columns corresponds to: id of gene, mean, logarithm of fold change, p-value for t-test, adjusted p-value (BH method). For more information about customizing this function see the help page for `test_diff`.

```
head(test.mety)
```

```
##                 id      mean    log.fold          pval         padj
## ICAM2        ICAM2 0.3330801 -0.15151320 3.754116e-17 3.063359e-14
## RILP          RILP 0.3341447 -0.05073691 2.575168e-13 1.050668e-10
## PIPOX        PIPOX 0.3647812  0.11505558 5.360053e-12 1.196885e-09
## TNFSF12     TNFSF12 0.2485025 -0.13412855 5.867083e-12 1.196885e-09
## CD7            CD7 0.8127112  0.09822690 1.641919e-11 2.679612e-09
## KSR1          KSR1 0.5549808  0.19973400 2.054467e-11 2.794075e-09
```

Expression

Info about expression. . .

In this section we will use data set BRCA_mRNAseq_chr17, which contains information about gene expression. This data set contains per-gene read counts computed for genes from 17th chromosome for 100 patients with breast cancer.

```r
load("BRCA_mRNAseq_chr17.rda")
```

```r
BRCA_mRNAseq_chr17[1:5,1:5]
```

```
##                              SUBTYPE AANAT AARSD1 AATF AATK
## TCGA-A1-A0SB-01A-11R-A144-07  Normal     9   2354 2870  317
## TCGA-A1-A0SD-01A-11R-A115-07    LumA     2   1846 5656  312
## TCGA-A1-A0SE-01A-11R-A084-07    LumA    11   3391 9522  736
## TCGA-A1-A0SF-01A-11R-A144-07    LumA     0   2169 4625  169
## TCGA-A1-A0SG-01A-11R-A144-07    LumA     1   2273 3473   92
```

Nbinom test

Negative binomial test is an another tool for finding differences between the base means of data having two or more conditions.

As in the t-test w need also a description of the samples, which we keep in a vector, whose elements correspond to different gorups.

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.

```r
condition<-ifelse(BRCA_mRNAseq_chr17$SUBTYPE=="LumA","LumA","other")
head(condition,8)
```

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

For using negative binomial test, in function `test_diff` we set value "nbinom" for parameter `test`. Evaluation for nbinorm may take a few minutes

```r
test.expr <- test_diff(BRCA_mRNAseq_chr17[,-1], condition, test="nbinom")
```

As a result we obtain the following data frame:

```r
head(test.expr)
```

```
##       id        mean     log.fold      pval      padj
## 1  AANAT    3.455436 -0.40216187 0.9453634 0.9920531
## 2 AARSD1 2779.448414  0.07427334 0.6627296 0.8992428
## 3   AATF 6750.269650 -0.13313947 0.6864598 0.9111193
## 4   AATK  352.805108 -0.02272566 0.8051408 0.9585881
## 5  ABCA5 1933.257431  0.07031881 0.5837682 0.8489059
## 6  ABCA6  689.547294  0.44680041 0.1289441 0.3857753
```

Visualization

log-log p-value

```r
p_values_plot(test.expr, test.mety)
```
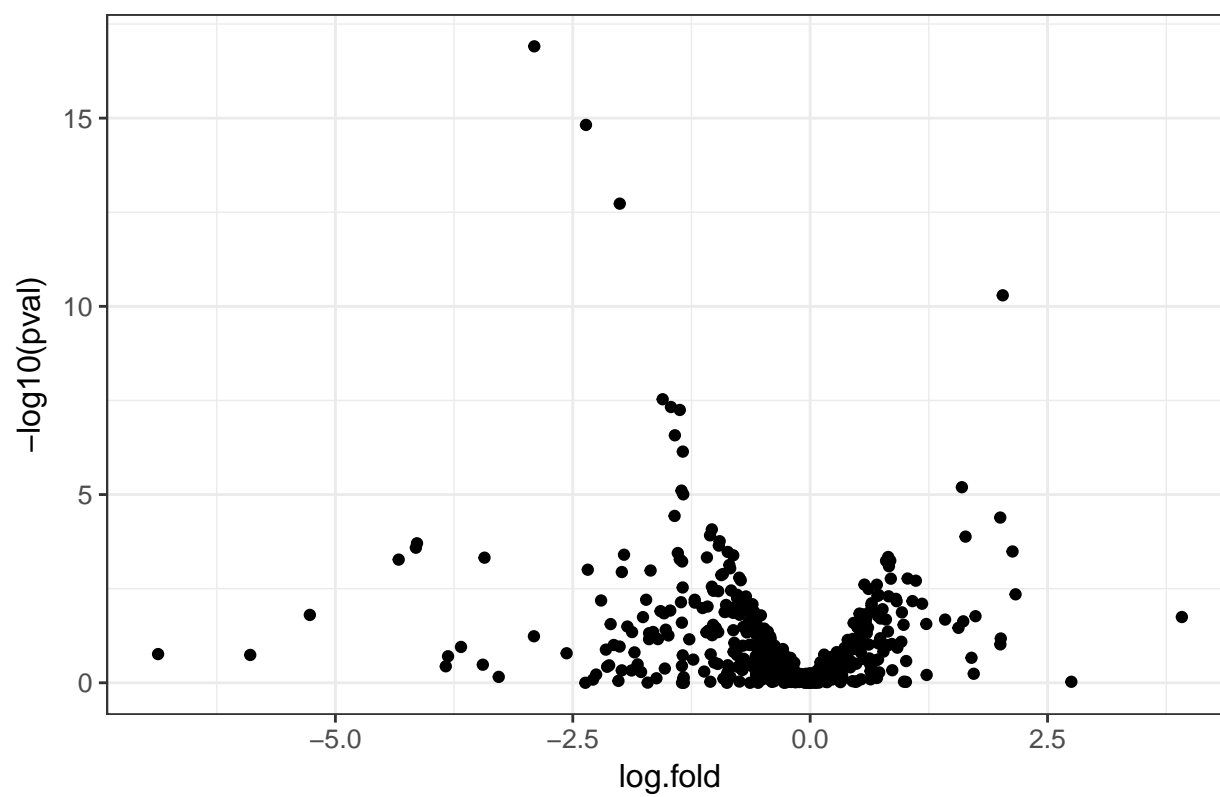
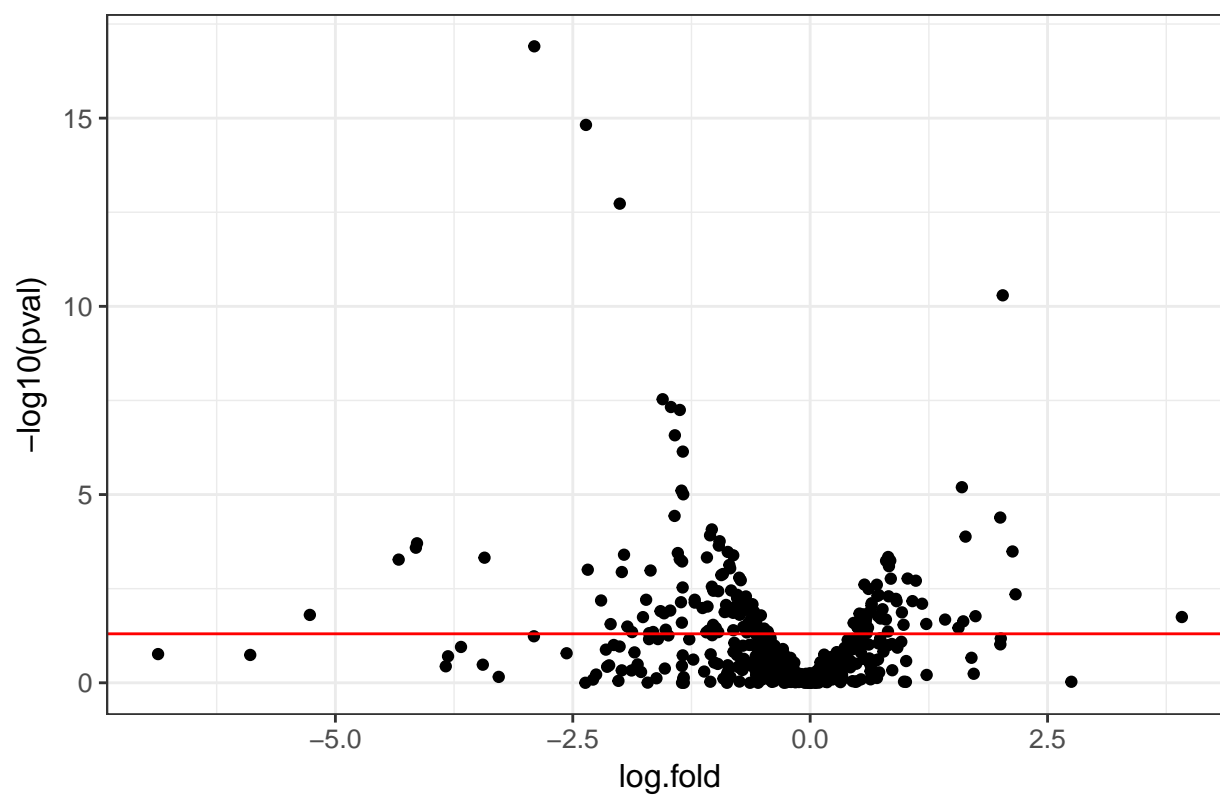## P−values comparison
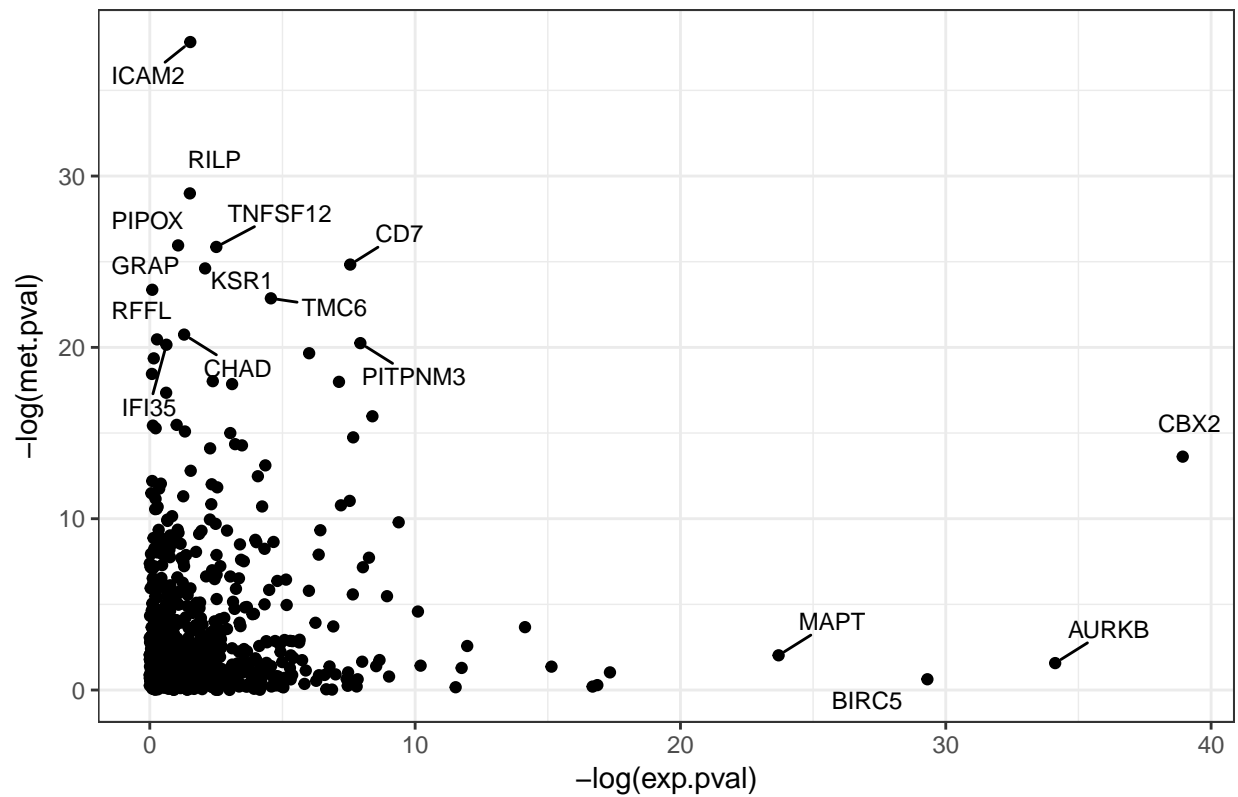
volcano

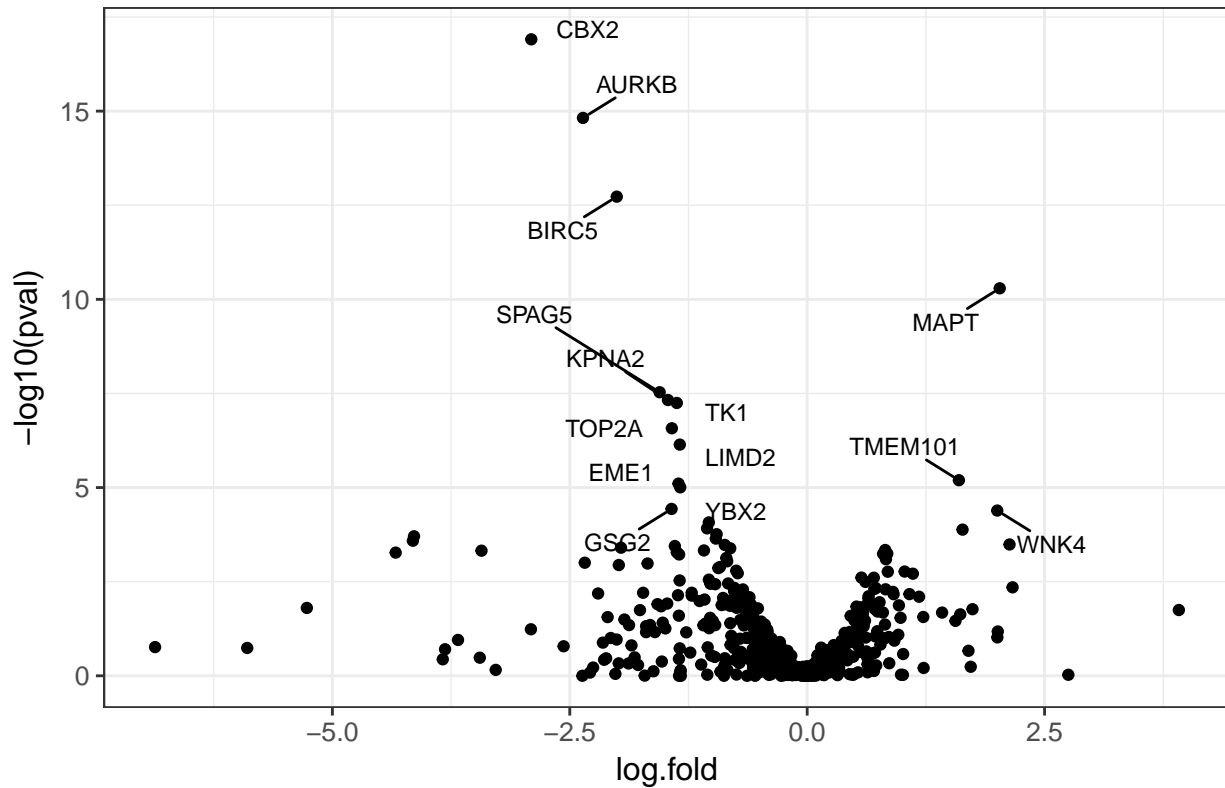Volcan plot for expression

```
volcano_plot(test.expr)
```

Volcano plot

```r
volcano_plot(test.expr, line = 0.05)
```

Volcano plot


```r
volcano_plot(test.expr, names = 0.00005)
```

Volcano plot

methyl "road" + boxplots for expr

```
BRCA1_gene <- CpG_mean(BRCA_methylation_chr17, "BRCA1")
head(BRCA1_gene)
```

```
##              Name Symbol CPG_ISLAND_LOCATIONS       mean
## 7047   cg07054526  BRCA1 17:38525979-38526990 0.93138624
## 4712   cg04658354  BRCA1 17:38530194-38531162 0.05802403
## 8929   cg08993267  BRCA1 17:38530194-38531162 0.06280625
## 19075  cg19088651  BRCA1 17:38530194-38531162 0.11090400
## 19527  cg19531713  BRCA1 17:38530194-38531162 0.10361921
## 6961   cg06973652  BRCA1 17:38531525-38532730 0.83772148
```

```
genereg_vs_met(BRCA_methylation_chr17, "BRCA1")
```