EBADIMEX

Tobias Madsen

2018-09-05

Data

First load the `ebadimex`-package

```
library(ebadimex)
```

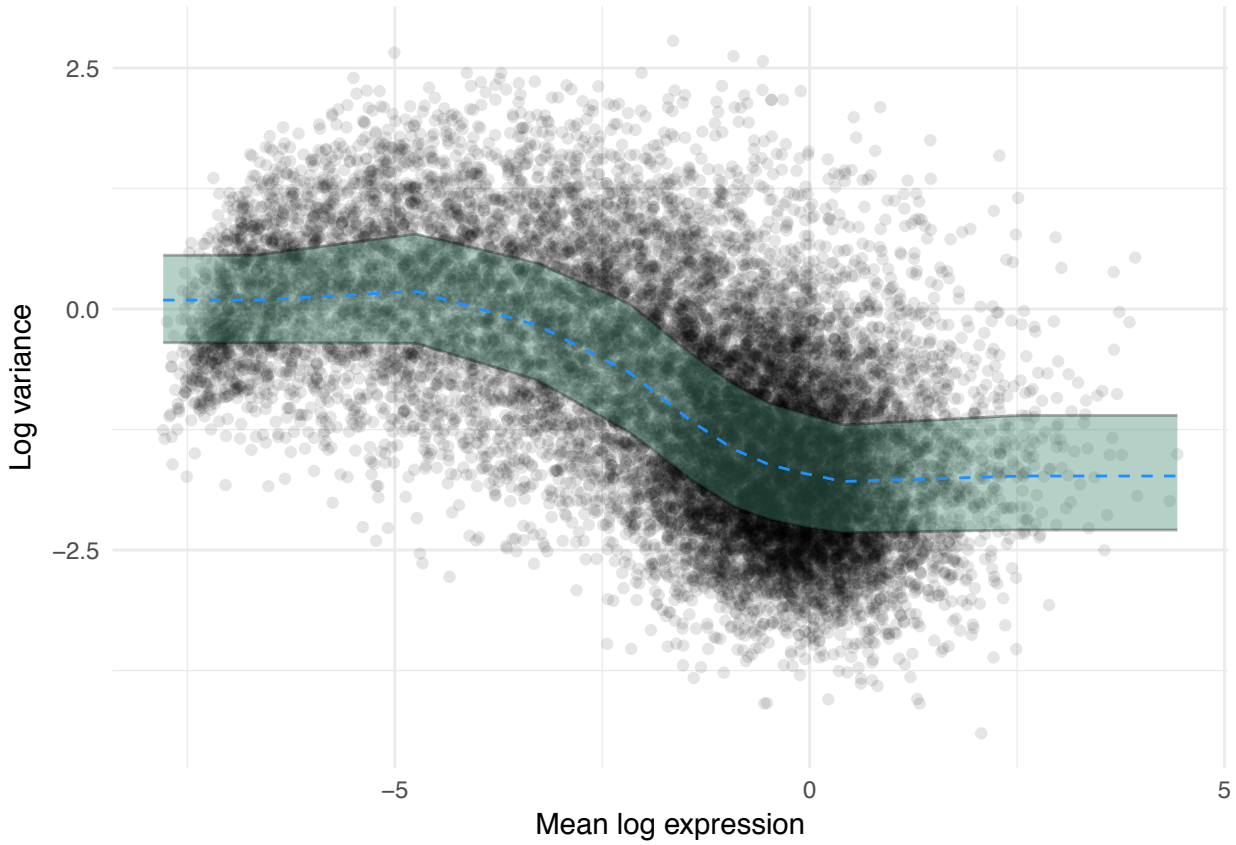
In the package we have included a subset of the data analyzed in our paper, with 5 normals and 10 tumours. The `brca` data set is a list with tree elements; `is_tumour` containing a vector indicating whether the i'th sample a tumour or normal, `sample_names` containing the name of the TCGA sample and finally `expr_meth` which is a list with expression (expr), promotor methylation (pr) and genebody methylation (gb) for a set of 16218 genes.

```
data(brca)
brca$expr_meth$PTEN
#>               expr        pr        gb
#>  [1,] 1.343331602 -4.917518 2.9680207
#>  [2,] 0.911211558 -5.143823 2.9627090
#>  [3,] 2.045277544 -4.920027 2.2493503
#>  [4,] 0.729825456 -5.311021 2.8071660
#>  [5,] 1.454789844 -2.537185 3.3612236
#>  [6,] 1.060067964 -4.513959 2.7685198
#>  [7,] 0.410336988 -4.575666 2.2727798
#>  [8,] 0.433309651 -4.756105 3.2010611
#>  [9,] 0.300567229 -4.508408 3.3166604
#> [10,] 0.449045962 -4.637717 3.3856938
#> [11,] 0.575343085 -4.154861 2.8651715
#> [12,] 0.350725670 -4.811216 3.3404441
#> [13,] 0.534079843 -4.859806 0.3677625
#> [14,] 0.007347179 -4.733323 3.4843213
#> [15,] 0.437210427 -4.715468 4.0996422
```

Priors

Before analyzing any individual gene, we train the empirical priors with the two methods `exprPrior` and `methPrior`. The expression prior is a function of the expression level and the figure shows the 25th and 75th percentile and the median of the prior distribution overlayed over point estimates of the expression variance for each gene.

```
ep <- exprPrior(brca$expr_meth)
```



The methylation prior consists of the scale matrix and degrees of freedom for an inverse Wishart distribution

```
mp <- methPrior(brca$expr_meth, brca$is_tumour)
mp
#> $Lambda
#>            pr        gb
#> pr 0.3375337 0.1356111
#> gb 0.1356111 0.2536949
#>
#> $nu0
#> [1] 6.127854
```

Differential Regulation

To test for differential expression use the `testDE` function.

```
pten_expr <- brca$expr_meth$PTEN[,'expr']

testDE(pten_expr, grouping = brca$is_tumour, prior = ep)
#> $p_location
#> [1] -7.773667
#>
#> $p_location_welch
#> [1] -6.017436
#>
#> $p_scale
#> [1] -2.097696
```

```
#>
#> $kl_12
#> [1] 1.928384
#>
#> $kl_21
#> [1] 4.393217
```

The output shows the log p-values for the moderated t-test (p_location), the moderated Welch t-test (p_location_welch) and the moderated f-test for equal variance (p_scale).

Test for differential methylation use `testDM` function.

```
pklr_expr <- brca$expr_meth$PKLR[,'expr']
pklr_meth <- brca$expr_meth$PKLR[,c('pr','gb')]

testDM(meth = pklr_meth, expr = pklr_expr, grouping = brca$is_tumour, prior = mp)
#> $p_location
#> [1] -6.411875
#>
#> $f_test_statistic
#> [1] 9.542486
#>
#> $p_bartlett
#> [1] -0.557031
#>
#> $kl_12
#> [1] 2.483216
#>
#> $kl_21
#> [1] 3.190577
```

The output shows the log p-value for the moderated F-test for shift in location (p_location) and the corresponding f-test statistic. The test for differential methylation corrects for the expression level. Thus the two p-values from testing differential expression and differential methylation can be combined using Fisher's method.

Classification

For classification pick a candidate gene and fit both the expression and the methylation model using `fitExpr` and `fitMeth`

```
pten_expr <- brca$expr_meth$PTEN[,'expr']
pten_meth <- brca$expr_meth$PTEN[,c('pr','gb')]
exprFit <- fitExpr(pten_expr, grouping = brca$is_tumour, prior = ep)
methFit <- fitMeth(meth = pten_meth, expr = pten_expr, grouping = brca$is_tumour, prior = mp)
```

A new samples can then be classified be evaluating the probability of the data under the model fit for the tumours and normals respectively and then Bayes factor.

```
new_sample_expr_meth <- c(1.4148645950, -4.622975, 2.75108451)

exprLogLikRatio <- exprLogLik(expr = new_sample_expr_meth[1],
                              param = exprFit,
                              same_sd = F)
methLogLikRatio <- methLogLik(expr = new_sample_expr_meth[1],
                              meth = new_sample_expr_meth[2:3],
```

```
                                param = methFit$param)

methLogLikRatio + exprLogLikRatio
#> [1] -4.391265
```

The reported log likelihood ratios are the `is_tumour`-true group (tumours) against `is_tumour`-false (normals). The above negative log lik ratio suggest that the new sample is more likely to belong to the normal group (ignoring imbalance in the prior distribution).