Exercices in linear models and experimental design, Lima

Dima

Introduction

These exercises describe experimental situations that have been studied using some omics technology like microarrays.

Comparison between three types of breast cancer

This case study is based on a paer published in http://www.ncbi.nlm.nih.gov/pubmed/15897907 whose data are available in GEO as series GSE1561 series on the following link http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE1561

The researchers investigated three types of breast cancer tumors: apocrine (APO), basal (BAS) and luminal (LUMI). The classification is based on the resistance of tumors to estrogen and androgen receptors.

- Tumors classified as "APO" are negative for estrogen receptor (ER-) and positive for the androgen receptor (AR +).
- Those classified as "LUMI" are ER + and AR + and
- Those classified as "BAS" are ER- and AR.

The assignment of each sample to an experimental group can be obtained from this link: http://www.ncbi.nlm.nih.gov/geo/gds/profileGraph.cgi?gds=1329

Obviously this is an observational study but its analysis can be done using a linear model approach as well.

1. Identify the experimental factors and their levels.
2. Write down the design matrix associated with this study design.
3. Build the contrast matrix needed to compare each tumor type with the oher two, that is:
    1. "APO" vs "LUMI"
    2. "APO" vs "BAS"
    3. "LUMI" vas "BAS"

```
##    SampleName          grupos ShortNAme         Colors
## 1  GSM3143969 Infection_total      IT69     firebrick3
## 2  GSM3143970 Infection_total      IT70     firebrick3
## 3  GSM3143971 Infection_total      IT71     firebrick3
## 4  GSM3143972   Control_total      CT72          cyan3
## 5  GSM3143973   Control_total      CT73          cyan3
## 6  GSM3143974   Control_total      CT74          cyan3
## 7  GSM3143975 Infection_polyA      IP75      darkgreen
## 8  GSM3143976 Infection_polyA      IP76      darkgreen
## 9  GSM3143977 Infection_polyA      IP77      darkgreen
## 10 GSM3143978   Control_polyA      CP78 darkgoldenrod2
## 11 GSM3143979   Control_polyA      CP79 darkgoldenrod2
## 12 GSM3143980   Control_polyA      CP80 darkgoldenrod2
```

```
##  [1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

```
##  [1] "1"  "2"  "3"  "4"  "5"  "6"  "7"  "8"  "9"  "10" "11" "12"
```

```
##    it ct ip cp
## 1   0  0  0  1
## 2   0  0  0  1
## 3   0  0  0  1
## 4   0  1  0  0
## 5   0  1  0  0
## 6   0  1  0  0
## 7   0  0  1  0
## 8   0  0  1  0
## 9   0  0  1  0
## 10  1  0  0  0
## 11  1  0  0  0
## 12  1  0  0  0
## attr(,"assign")
## [1] 1 1 1 1
## attr(,"contrasts")
## attr(,"contrasts")$`targets$grupos`
## [1] "contr.treatment"
```

```
##       Contrasts
## Levels itvsct itvsip itvscp ctvsip ctvscp ipvscp
##     it      1      1      1      0      0      0
##     ct     -1      0      0      1      1      0
##     ip      0     -1      0     -1      0      1
##     cp      0      0     -1      0     -1     -1
```

```
##               logFC   AveExpr        t      P.Value    adj.P.Val        B
## 6709_at   0.9119978 1.1537637 13.12875 4.075478e-08 0.0005323923 8.861119
## 113829_at 0.7288723 0.3016420 13.06037 4.305117e-08 0.0005323923 8.815866
## 8342_at   1.1239047 0.2388142 11.52929 1.577365e-07 0.0011209665 7.714932
## 60509_at  0.7144823 0.5796208 11.37021 1.820506e-07 0.0011209665 7.590171
## 1173_at   1.1896924 1.3959312 10.88434 2.851421e-07 0.0011209665 7.195845
## 64324_at  0.7270113 0.9900555 10.76102 3.203991e-07 0.0011209665 7.092471
```

```
##               logFC  AveExpr         t      P.Value    adj.P.Val         B
## 23586_at -0.9781839 1.260912 -16.17287 4.461663e-09 8.719844e-05 10.322442
## 1031_at   1.1221075 1.002875  15.22218 8.520042e-09 8.719844e-05  9.864326
## 3655_at  -1.1941241 1.804888 -14.62052 1.308195e-08 8.719844e-05  9.550727
## 28984_at  1.4201543 1.636372  14.51739 1.410236e-08 8.719844e-05  9.495004
## 4907_at  -1.5640048 1.043499 -14.20263 1.778969e-08 8.799846e-05  9.321206
## 57476_at -1.0076289 1.032540 -13.34356 3.436443e-08 1.274756e-04  8.816762
```

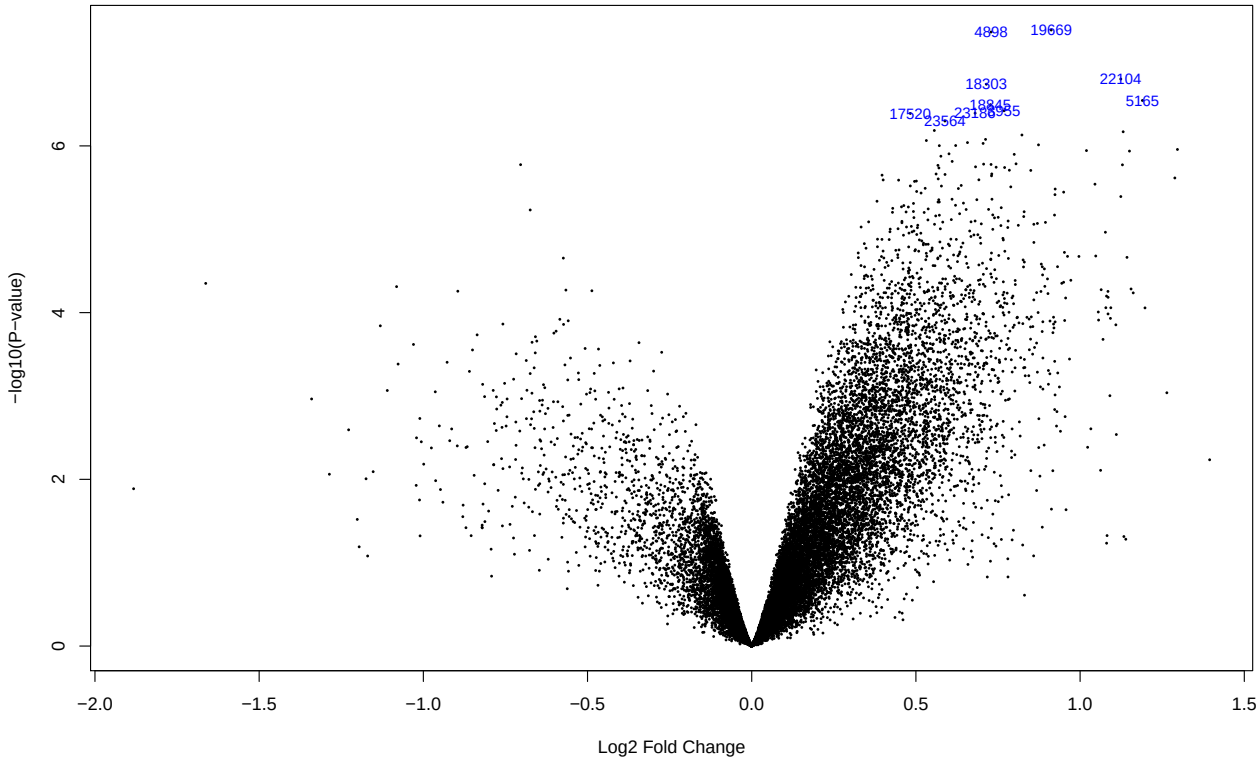
```
##                logFC  AveExpr         t      P.Value    adj.P.Val        B
## 23586_at  -1.3087528 1.260912 -21.63835 1.921848e-10 4.753307e-06 13.54747
## 1031_at    1.3652297 1.002875  18.52030 1.038782e-09 1.284610e-05 12.27687
## 6709_at    1.2367743 1.153764  17.80411 1.589594e-09 1.310514e-05 11.93857
## 28984_at   1.6947637 1.636372  17.32456 2.132732e-09 1.318722e-05 11.70090
## 79622_at   1.7625903 1.505011  15.57842 6.657694e-09 3.182065e-05 10.75164
## 113829_at  0.8574222 0.301642  15.36380 7.719399e-09 3.182065e-05 10.62508
```

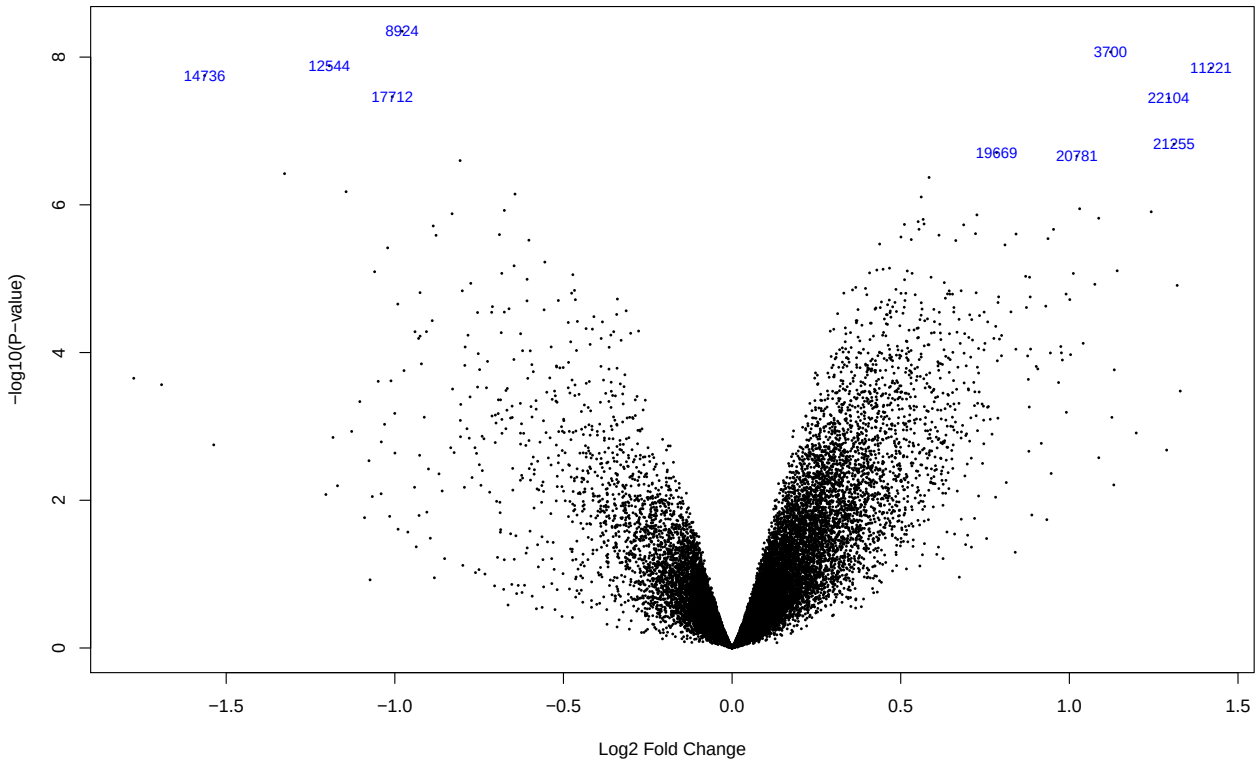
```
##               logFC   AveExpr         t      P.Value    adj.P.Val
## 57476_at -1.3108320 1.0325396 -17.35874 2.087996e-09 4.918734e-05
## 3655_at  -1.3351590 1.8048882 -16.34731 3.977466e-09 4.918734e-05
## 28984_at  1.4459254 1.6363721  14.78084 1.165136e-08 9.605766e-05
## 4907_at  -1.5616960 1.0434987 -14.18167 1.807004e-08 9.655450e-05
```

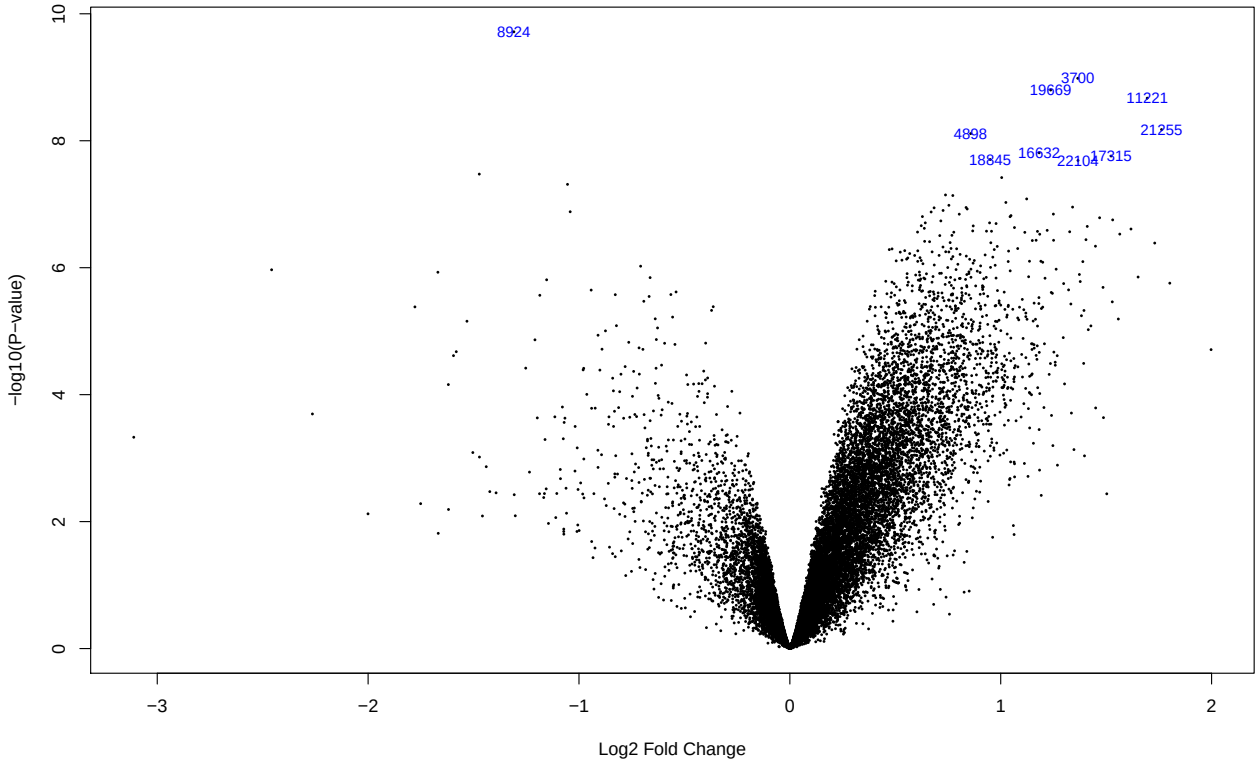
```
## 23586_at -0.8515180 1.2609119 -14.07863 1.951937e-08 9.655450e-05
## 2591_at  -0.8391549 0.6248026 -12.11778 9.415714e-08 3.102326e-04
##                  B
## 57476_at 11.216059
## 3655_at  10.743770
## 28984_at  9.918234
## 4907_at   9.568305
## 23586_at  9.506048
## 2591_at   8.191415
```

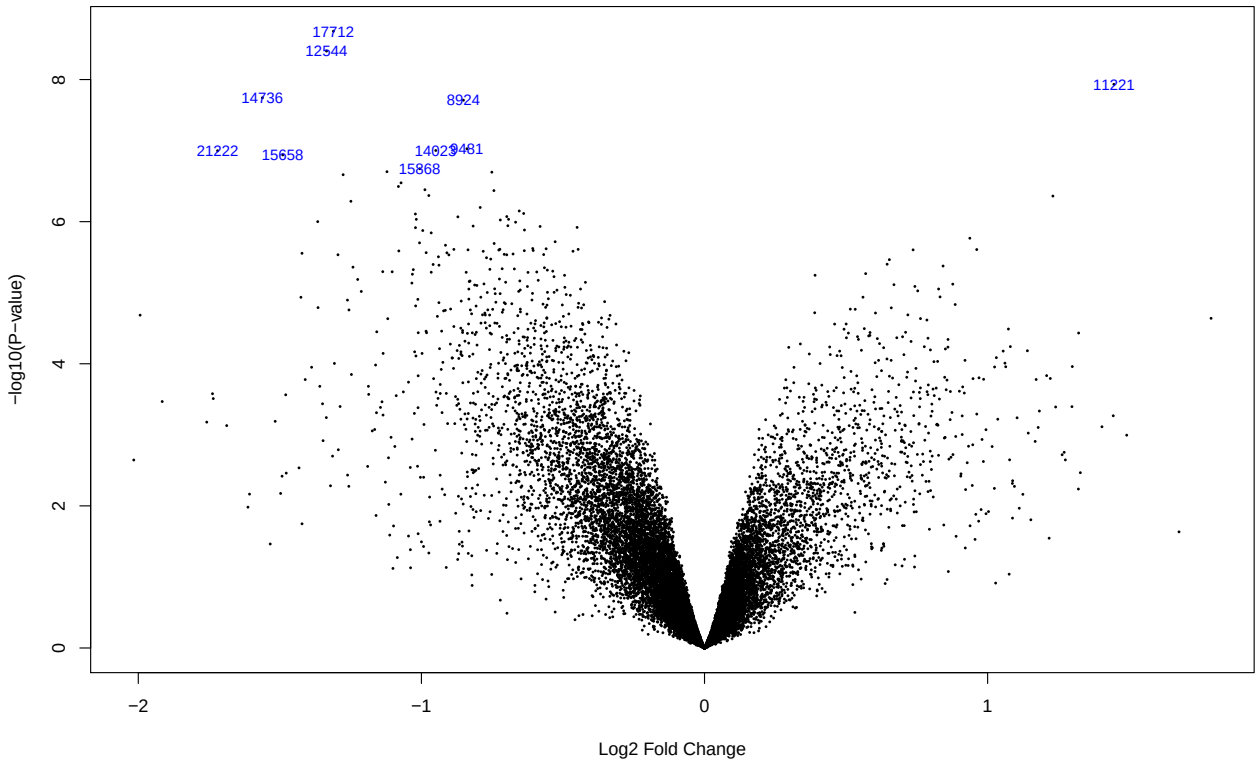
```
##              logFC  AveExpr         t      P.Value    adj.P.Val         B
## 23586_at -1.182087 1.260912 -19.54411 5.804606e-10 1.435653e-05 11.167452
## 28984_at  1.720535 1.636372  17.58800 1.813039e-09 2.242095e-05 10.519045
## 3655_at  -1.195389 1.804888 -14.63600 1.293579e-08 1.066470e-04  9.255488
## 4907_at  -1.470451 1.043499 -13.35308 3.410773e-08 2.108966e-04  8.567680
## 5121_at   1.394139 1.323829  11.82812 1.210431e-07 5.987520e-04  7.610009
## 57476_at -0.848099 1.032540 -11.23098 2.066802e-07 8.519703e-04  7.186756
```

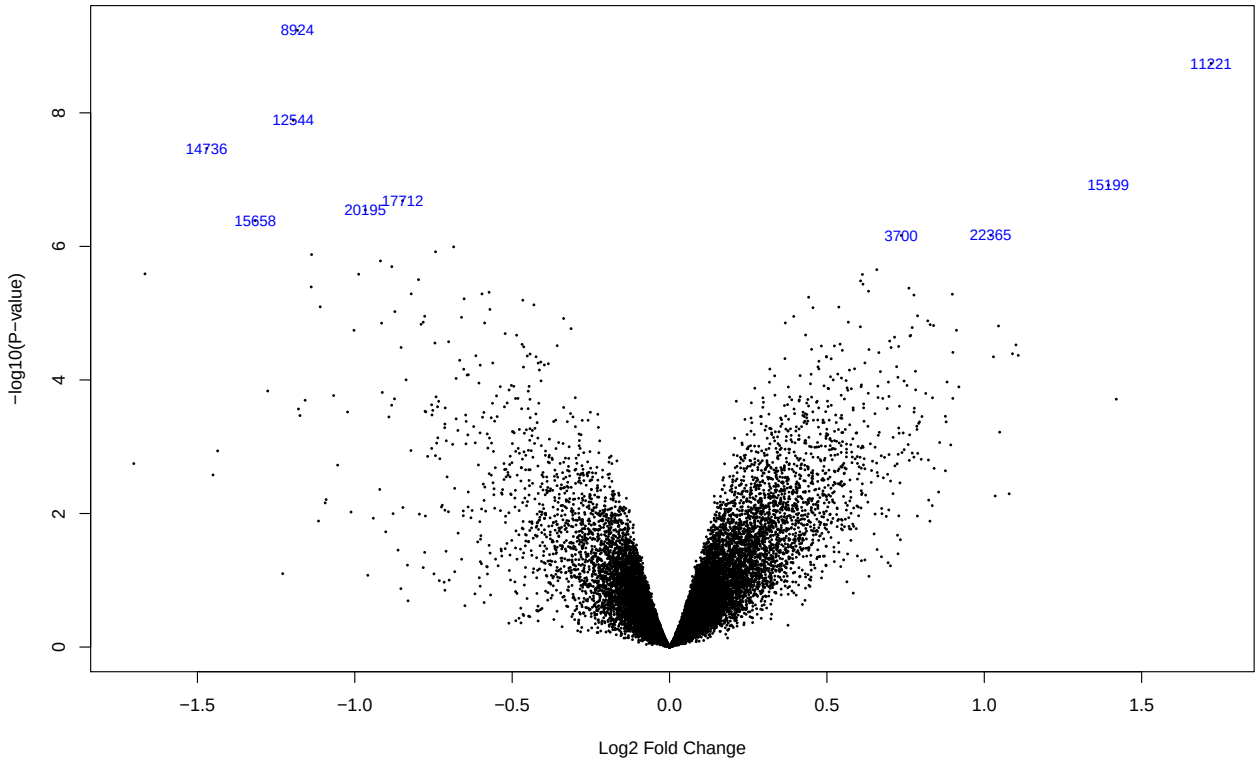
```
##               logFC   AveExpr          t      P.Value  adj.P.Val        B
## 6023_at  -2.5841171 0.7615244 -10.035440 6.512732e-07 0.01416185 6.054565
## 1729_at   0.6534534 1.1558589   9.487205 1.145179e-06 0.01416185 5.601223
## 2040_at   0.6958301 0.9499314   8.839623 2.308763e-06 0.01658099 5.022881
## 9877_at   0.8080422 1.7098592   8.050260 5.736753e-06 0.01658099 4.249122
## 26801_at -1.6929885 1.4887458  -7.927394 6.648995e-06 0.01658099 4.121400
## 22853_at  0.4012584 0.7817825   7.842311 7.371591e-06 0.01658099 4.031752
```

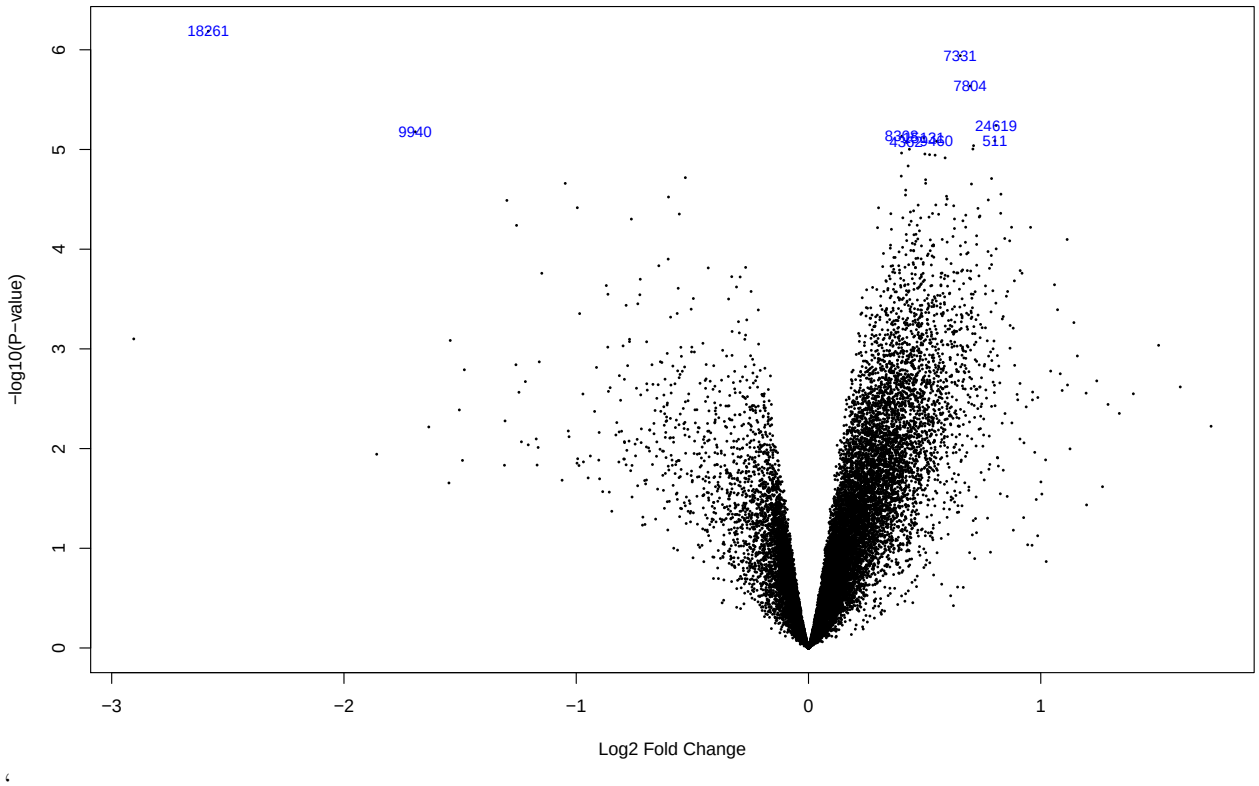



8924
14736
12544
17712
3700
11221
22104
21255
19669
20781
−log10(P−value)
Log2 Fold Change


8924
3700
19669
11221
21255
4998
18845
16632
22104
17315
−log10(P−value)
Log2 Fold Change


17712
12544
14736
8924
11221
21222
15658
15868
−log10(P−value)
0
2
4
6
8
−2
−1
0
1
Log2 Fold Change


8924
11221
12544
14736
15199
20195
17712
15658
3700
22365
−log10(P−value)
0
2
4
6
8
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
Log2 Fold Change


18261
7331
7804
9940
24619
511
−log10(P−value)
0
1
2
3
4
5
6
−3
−2
−1
0
1
Log2 Fold Change