# HW4 - Gene expression - Clustering and Enrichment

```
install.packages('tinytex')
tinytex::install_tinytex()
tinytex::pdflatex('HW4_2019_template.tex')
```

Submitted by:

ID1: 320983216 Name1: Maxim Kolchinsky ID2: 321340580 Name2: Anna Romanov

## The Biology

The data for this lesson comes from:

> Saigi et al. "MET-Oncogenic and JAK2-Inactivating Alterations Are Independent Factors That Affect Regulation of PD-L1 Expression in Lung Cancer" PLoS ONE. 2018 Jun 13;9(6):e99625. PMID: 24926665.

Purpose: The blockade of immune checkpoints such as PDL1 and PD-1 is being exploited therapeutically in several types of malignancies. Here, we aimed to understand the contribution of the genetics of lung cancer to the ability of tumor cells to escape immunosurveillance checkpoints. Experimental Design: More than 150 primary non-small cell lung cancers, including pulmonary sarcomatoid carcinomas, were tested for levels of the HLA-I complex, PD-L1, tumor-infiltrating CD8b lymphocytes, and alterations in main lung cancer genes. Correlations were validated in cancer cell lines using appropriate treatments to activate or inhibit selected pathways. We also performed RNA sequencing to assess changes in gene expression after these treatments. Results: MET-oncogenic activation tended to associate with positive PD-L1 immunostaining, whereas STK11 mutations were correlated with negative immunostaining. In MET-altered cancer cells, MET triggered a transcriptional increase of PD-L1 that was independent of the IFNgmediated JAK/STAT pathway. The activation of MET also upregulated other immunosuppressive genes (PDCD1LG2 and SOCS1) and transcripts involved in angiogenesis (VEGFA and NRP1) and in cell proliferation. We also report recurrent inactivating mutations in JAK2 that co-occur with alterations in MET and STK11, which prevented the induction of immunoresponse-related genes following treatment with IFNg. Conclusions: We show that MET activation promotes the expression of several negative checkpoint regulators of the immunoresponse, including PD-L1. In addition, we report inactivation of JAK2 in lung cancer cells that prevented the response to IFNg. These alterations are likely to facilitate tumor growth by enabling immune tolerance and may affect the response to immune checkpoint inhibitors

### Data

This data was downloaded from GEO (GSE:GSE109720)

### Import count data and metadata

```
library(readr)
library(dplyr)
library(ggplot2)

rawcounts <- read_csv("data/lung_counts.csv")
metadata <- read_csv("data/lung_metadata.csv")

rawcounts</pre>
```

```
## # A tibble: 58,347 x 34
##
      ensgene AE1148 AE1149 AE1150 AE1151 AE1152 AE1153 AE1154 AE1155 AE1156
                     <dbl>
##
               <dbl>
                              <dbl>
                                     <dbl> <dbl> <dbl>
                                                           <dbl>
  1 ENSGOO~
##
                   0
                          0
                                  5
                                         0
                                                7
                                                        0
                                                               0
                                                                      0
                                                                              5
##
    2 ENSGOO~
                 374
                         529
                                304
                                       366
                                               466
                                                      519
                                                             319
                                                                    379
                                                                            326
## 3 ENSG00~
                                                        0
                                                               0
                   0
                          0
                                  0
                                         0
                                                0
                                                                      0
                                                                              0
## 4 ENSG00~
                   0
                           0
                                  0
                                         0
                                                0
                                                        0
                                                               0
                                                                       0
## 5 ENSG00~
                   0
                           0
                                  0
                                         0
                                                0
                                                        0
                                                               0
                                                                       0
                                                                              0
## 6 ENSG00~
                   0
                           0
                                  0
                                         0
                                                0
                                                        0
                                                               0
                                                                       0
                                                                              0
## 7 ENSGOO~
                                                                              0
                   0
                           0
                                  0
                                         0
                                                0
                                                        0
                                                               0
                                                                       0
## 8 ENSG00~
                   0
                           0
                                  0
                                         0
                                                0
                                                        0
                                                               0
                                                                       0
                                                                              0
## 9 ENSG00~
                           0
                                  0
                                         0
                                                0
                                                        0
                                                               0
                                                                       0
                                                                              0
                   0
## 10 ENSGOO~
                  11
                           1
                                  8
                                        24
                                                18
                                                       21
                                                              25
                                                                     17
                                                                             17
## # ... with 58,337 more rows, and 24 more variables: AE1157 <dbl>,
       AE1158 <dbl>, AE1159 <dbl>, AE1160 <dbl>, AE1161 <dbl>, AE1162 <dbl>,
## #
       AE1163 <dbl>, AE1164 <dbl>, AE1165 <dbl>, AE1166 <dbl>, AE1167 <dbl>,
       AE1168 <dbl>, AE1169 <dbl>, AE1170 <dbl>, AE1171 <dbl>, AE1172 <dbl>,
## #
       AE1173 <dbl>, AE1174 <dbl>, AE1175 <dbl>, AE1176 <dbl>, AE1177 <dbl>,
## #
       AE1178 <dbl>, AE1179 <dbl>, AE1180 <dbl>
```

#### metadata

```
## # A tibble: 33 x 4
##
     id
            celltype dex
                                  geo_id
##
     <chr> <chr>
                     <chr>
                                  <chr>>
##
  1 AE1160 EBC1
                     No treatment GSM2949380
                     No treatment GSM2949381
##
  2 AE1163 EBC1
   3 AE1166 EBC1
                     No treatment GSM2949382
## 4 AE1162 EBC1
                     Crizotinib GSM2949383
## 5 AE1165 EBC1
                     Crizotinib GSM2949384
## 6 AE1168 EBC1
                     Crizotinib GSM2949385
   7 AE1161 EBC1
                     Interferon <U+03B3> GSM2949386
## 8 AE1164 EBC1
                     Interferon <U+03B3> GSM2949387
## 9 AE1167 EBC1
                     Interferon <U+03B3> GSM2949388
                     No treatment GSM2949389
## 10 AE1169 H1573
## # ... with 23 more rows
```

### Clustering - k-means

```
}
# cluster the genes
rawcounts_RN<- data.frame(rawcounts, row.names=1)
km <- kmeans(rawcounts RN,5)
km$centers
##
          AE1148
                      AE1149
                                  AE1150
                                             AE1151
                                                         AE1152
                                                                      AE1153
## 1 261075.5000 355397.5000 249188.5000 298081.500 312029.0000 330232.5000
                    260.8659
                                184.8346
                                            225.175
                                                       226.5053
        204.3794
                                                                    263.4797
                             89713.5238 117618.429 120754.0000 151602.5238
## 3 101059.4762 147734.7143
       6589.0833
                   8841.2332
                               6127.4839
                                          7358.684
                                                      7346.2171
                                                                   8890.0477
    31052.4260
                 42212.6686
                              27571.1893 34696.260 32887.3373 41325.6095
##
          AE1154
                      AE1155
                                  AE1156
                                              AE1157
                                                          AE1158
                                                                      AE1159
## 1 222399.5000 302143.5000 258949.5000 295158.0000 257717.0000 284918.500
## 2
        177.6818
                    222.9686
                                196.5011
                                            232.7849
                                                        189.8281
                                                                    212.521
## 3 90998.6190 124462.2857 104368.4286 134149.9048 106395.0952 117189.476
      5868.3233
                  7272.5940
                               6356.2941
                                          8001.1913
                                                       6352.8116
                                                                   7040.299
## 5
     26505.6568
                 33457.2663 29596.4379
                                         37330.6331
                                                     30042.5799
                                                                  32418.935
##
         AE1160
                     AE1161
                                 AE1162
                                             AE1163
                                                         AE1164
## 1 413308.500 281465.5000 419312.0000 335008.5000 264320.5000 344465.0000
## 2
        242.584
                   180.0566
                               253.0504
                                           219.7236
                                                       191.4948
## 3 143324.762 104374.0952 132841.0000 122184.8571 107568.4286 112637.7143
       7666.075
                  5913.3574
                             7290.1723
                                          6847.2006
                                                      6203.9289
                                                                   6185.2936
## 5
     35868.917 27201.2367
                             34315.4142 31814.4260 29037.9941
                                                                 29120.4734
          AE1166
                      AE1167
                                  AE1168
                                              AE1169
                                                          AE1170
                                                                      AE1171
## 1 384117.5000 304171.0000 421639.5000 605629.0000 430845.0000 507135.5000
        241.4646
                    198.3505
                                238.9257
                                            253.7209
                                                        189.6234
## 3 135000.0476 111633.0476 126291.2857 169666.4762 124785.2381 141406.8095
       7557.6928
                   6488.0501
                               6934.8676
                                           7469.2434
                                                       5595.4722
                                                                    6493.8734
## 5
    34757.7219
                  29575.0533
                             32625.8402
                                          36791.6982 27468.7456
                                                                  31881.7219
                      AE1173
                                  AE1174
                                             AE1175
          AE1172
                                                         AE1176
                                                                     AE1177
## 1 422003.5000 530275.0000 492796.5000 382136.500 308577.5000 364062.5000
                    226.5967
                               210.5955
                                            222.858
                                                       183.9239
        185.7504
                                                                    211.8739
## 3 117254.7619 146959.6667 137927.1429 127871.333 104154.3333 123308.0952
       5550.7936
                   6760.0141
                               6311.3476
                                           6946.255
                                                      5771.8062
                                                                   6654.9367
## 5
     27203.1006
                  33370.0118
                              31380.9112
                                          29166.260 24118.6923
                                                                 28170.8698
##
          AE1178
                      AE1179
                                  AE1180
## 1 280669.0000 308508.0000 400801.0000
        165.6252
                    178.0214
                                225.1644
    94938.3333 104582.6190 133835.4286
## 4
       5181.0136
                   5586.3374
                               7055.9640
## 5 21687.6036 23821.4024
                              29880.8935
num genes <- table(km$cluster)</pre>
num_genes
##
```

##

##

2

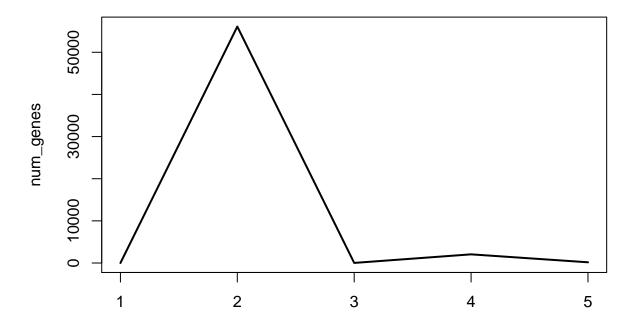
2 56101

3

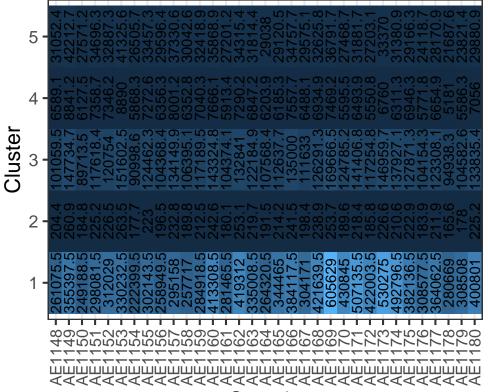
21 2054

5

169



plot\_heatmap(km)



value

6e+05

5e+05 4e+05 3e+05 2e+05 1e+05

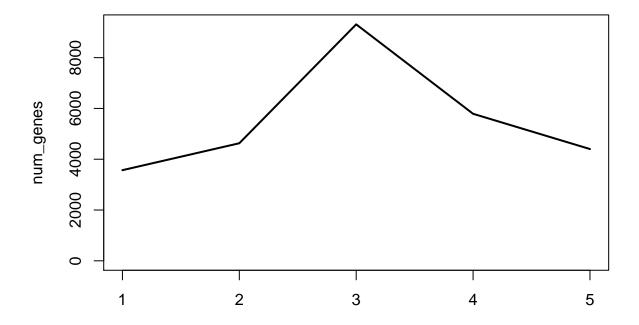
# Sample

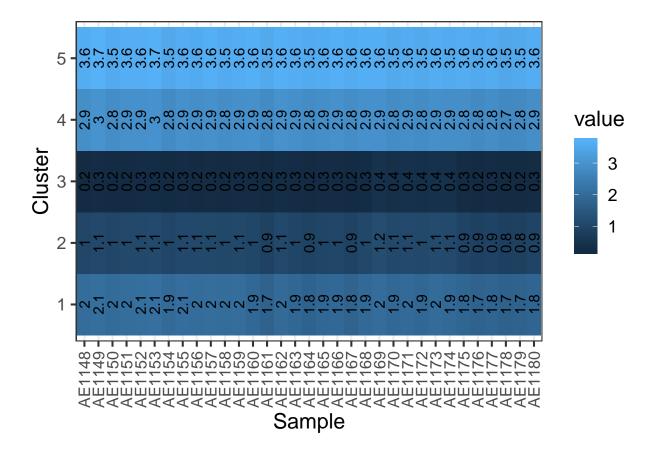
```
# remove rows of zeroes
rawcounts_RN\$sum <- rowSums(rawcounts_RN)
rawcounts_clean <- rawcounts_RN[ rawcounts_RN\$sum >= 10, ]
rawcounts_clean\$sum <- NULL

# apply log scaling
rawcounts_log = log10(1+rawcounts_clean)
km <- kmeans(rawcounts_log,5)
km\$centers</pre>
```

```
##
        AE1148
                  AE1149
                            AE1150
                                      AE1151
                                                 AE1152
                                                           AE1153
                                                                     AE1154
## 1 2.0053516 2.0718609 1.9683162 2.0457680 2.0753388 2.0995963 1.9408546
## 2 0.9926849 1.0658002 0.9937652 1.0406866 1.0859241 1.1021633 0.9782394
## 3 0.2224750 0.2595178 0.2365299 0.2465905 0.2664494 0.2856338 0.2270980
## 4 2.8768341 2.9629047 2.8159122 2.9162445 2.9151470 2.9548990 2.7748572
## 5 3.5738212 3.6867595 3.5233650 3.6183988 3.6159370 3.6748730 3.4829931
##
        AE1155
                  AE1156
                            AE1157
                                      AE1158
                                                 AE1159
                                                           AE1160
## 1 2.0608478 2.0370740 2.0484974 1.9979362 2.0467039 1.8938259 1.7459993
## 2 1.0667283 1.0603057 1.0567061 1.0293423 1.0579908 1.0154241 0.8954655
## 3 0.2611225 0.2488747 0.2554619 0.2473489 0.2562278 0.2735666 0.2295744
## 4 2.9035514 2.8555548 2.9088671 2.8171388 2.8838187 2.9247148 2.7762690
## 5 3.6039529 3.5517147 3.6395846 3.5207057 3.5946933 3.6473609 3.5152520
##
        AE1162
                  AE1163
                            AE1164
                                      AE1165
                                                 AE1166
                                                           AE1167
                                                                     AE1168
## 1 1.9523912 1.8613406 1.7745325 1.8706251 1.8907524 1.7858088 1.9209397
## 2 1.0746557 0.9825569 0.9117918 0.9844687 1.0037558 0.9144569 1.0278350
## 3 0.3069437 0.2660096 0.2308454 0.2634969 0.2683428 0.2351569 0.2865346
```

```
## 4 2.9481320 2.8845318 2.8080246 2.8748433 2.9221639 2.8221034 2.9211485
## 5 3.6335486 3.5997041 3.5385613 3.5622239 3.6428058 3.5588146 3.6114502
                 AE1170
                           AE1171
                                      AE1172
## 1 2.0208283 1.892123 1.9587703 1.8819145 1.968288 1.9379473 1.8364030
## 2 1.1728400 1.052466 1.1166752 1.0400485 1.123027 1.1024374 0.9364739
## 3 0.4410393 0.373576 0.4068394 0.3665689 0.409207 0.3943318 0.2684219
## 4 2.9333166 2.806446 2.8682297 2.7963220 2.882997 2.8517416 2.8476059
## 5 3.6191436 3.491511 3.5584802 3.4891624 3.575307 3.5434462 3.5860752
##
        AE1176
                  AE1177
                            AE1178
                                       AE1179
                                                 AE1180
## 1 1.7454409 1.8129884 1.7102739 1.7326465 1.8332838
## 2 0.8611594 0.9207595 0.8351714 0.8435144 0.9420726
## 3 0.2289570 0.2646176 0.2237919 0.2291777 0.2697421
## 4 2.7616280 2.8245047 2.7176708 2.7502298 2.8514261
## 5 3.5038501 3.5662006 3.4583779 3.4908875 3.5926027
num_genes <- table(km$cluster)</pre>
num_genes
##
##
           2
                3
## 3569 4629 9311 5789 4401
plot(num_genes, type="l", main='num of genes in each cluster')
```

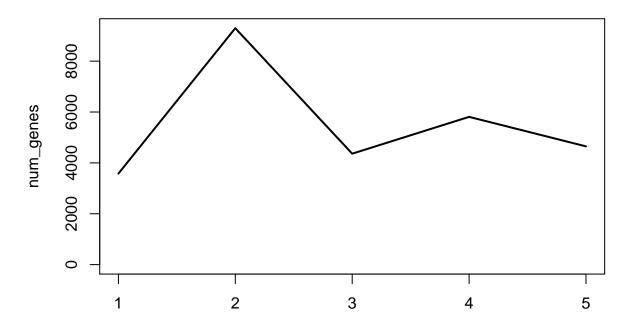




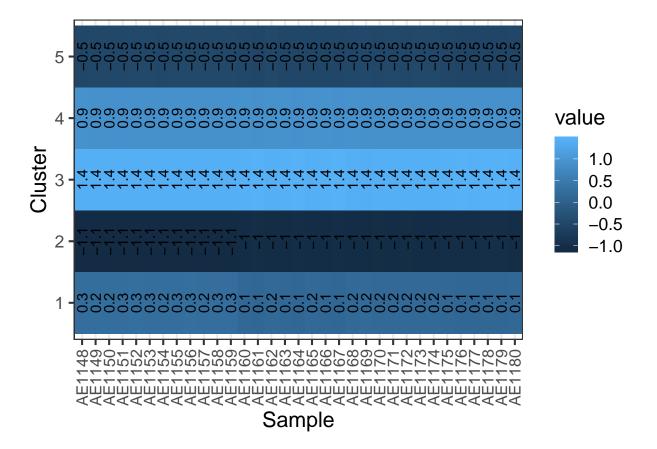
```
# Apply standardization
rawcounts_scaled <- scale(rawcounts_log)
km <- kmeans(rawcounts_scaled,5)
km$centers</pre>
```

```
AE1151
                                                               AE1153
##
        AE1148
                   AE1149
                              AE1150
                                                    AE1152
## 1 0.2528079 0.2451968 0.2458377
                                     0.2538863 0.2649167
                                                           0.2567951
## 2 -1.0693269 -1.0688768 -1.0633402 -1.0722110 -1.0782449 -1.0693484
     1.4114811 1.4119526
                          1.4156985
                                      1.4092971
                                                1.4048429
                                                           1.4039390
## 4 0.8964867 0.8893151 0.8828349 0.8935212 0.8863283
                                                          0.8799729
## 5 -0.5008578 -0.4873795 -0.4943596 -0.4901676 -0.4734303 -0.4761762
                                                               AE1159
##
        AE1154
                   AE1155
                              AE1156
                                         AE1157
                                                    AE1158
## 1 0.2447623 0.2625378 0.2683054 0.2498533 0.2583316
                                                           0.2608463
  2 -1.0586070 -1.0739779 -1.0785432 -1.0690393 -1.0662110 -1.0741021
     1.4127032
               1.4041744
                          1.4048258
                                      1.4157064
                                                 1.4051733
                                                           1.4108421
     0.8760129
                0.8860388 0.8821716
                                      0.8804559
                                                 0.8744648
                                                           0.8826296
## 5 -0.4916615 -0.4791412 -0.4702332 -0.4830918 -0.4779076 -0.4795858
##
        AE1160
                   AE1161
                              AE1162
                                         AE1163
                                                    AE1164
                                                               AE1165
## 1 0.1468150 0.1198996 0.1673943 0.1452124 0.1274776
                                                          0.1580055
## 2 -1.0365985 -1.0180156 -1.0427133 -1.0321912 -1.0220277 -1.0340571
## 3 1.4240540 1.4423877 1.4017358
                                      1.4245303 1.4359796
                                                           1.4101721
## 4 0.8973837 0.8894981 0.8979515 0.8978315 0.8936499 0.9007449
## 5 -0.4976051 -0.5213826 -0.4809970 -0.5061895 -0.5183713 -0.5024793
```

```
##
        AE1166
                   AE1167
                              AE1168
                                        AE1169
                                                   AE1170
## 1 0.1481989 0.1284032 0.1652151 0.1754403 0.1611897 0.1696153
## 2 -1.0346146 -1.0215992 -1.0380963 -1.0333268 -1.0232746 -1.0306172
## 3 1.4222051 1.4381401 1.4073590 1.3956382 1.4062621 1.4039338
## 4 0.8979212 0.8933030 0.8994738 0.8715029 0.8722969 0.8709739
## 5 -0.5015744 -0.5215339 -0.4957272 -0.4671921 -0.4872764 -0.4752458
        AE1172
                   AE1173
                              AE1174
                                        AE1175
                                                   AE1176
## 1 0.1597162 0.1691050 0.1655199 0.1384030 0.1238105 0.1326525
## 2 -1.0237152 -1.0311316 -1.0302534 -1.0126447 -1.0052797 -1.0095451
## 3 1.4123471 1.4040635 1.4059962 1.4240011 1.4340972 1.4275570
## 4 0.8720849 0.8718143 0.8711739 0.8850181 0.8846325 0.8835753
## 5 -0.4907044 -0.4749960 -0.4750045 -0.5235211 -0.5360008 -0.5268240
        AE1178
                   AE1179
                              AE1180
## 1 0.1194956 0.1214050 0.1338500
## 2 -1.0017273 -1.0019488 -1.0131401
## 3 1.4386298 1.4360721 1.4257434
## 4 0.8831780 0.8858309 0.8852047
## 5 -0.5422155 -0.5441582 -0.5208928
num_genes <- table(km$cluster)</pre>
num_genes
##
##
          2
      1
               3
## 3580 9297 4362 5810 4650
plot(num_genes, type="l", main='num of genes in each cluster')
```



plot\_heatmap(km)

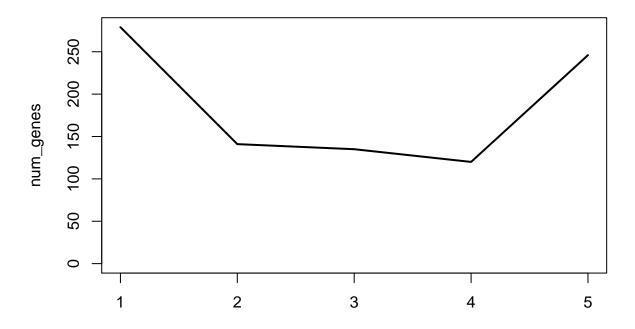


```
# load DE gene list
DE_res <- read_csv("data/sigresults.csv")
DE_genes <- DE_res$row

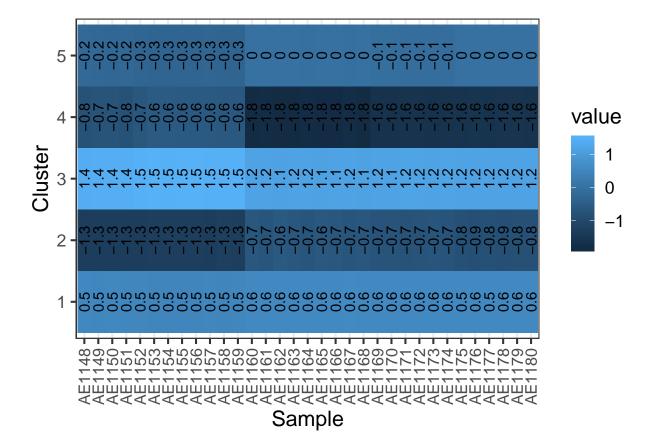
# example of filtering a data frame
dat_filtered = rawcounts_clean[DE_genes,]
rawcounts_log = log10(1+dat_filtered)
rawcounts_scaled <- scale(rawcounts_log)
km <- kmeans(rawcounts_scaled,5)
km$centers</pre>
```

```
##
        AE1148
                   AE1149
                              AE1150
                                         AE1151
                                                    AE1152
                                                               AE1153
## 1 0.5337270 0.5208088 0.5288254 0.5269107 0.5070860 0.4624104
## 2 -1.3451860 -1.3400504 -1.3413247 -1.3383443 -1.3333927 -1.2999979
## 3 1.3979433 1.4178354 1.4232187
                                     1.4188640 1.4592332 1.4795627
## 4 -0.7921002 -0.7437571 -0.7423865 -0.7738489 -0.6628963 -0.5678233
## 5 -0.2150769 -0.2378679 -0.2498523 -0.2316517 -0.2882827 -0.3142885
##
        AE1154
                   AE1155
                              AE1156
                                         AE1157
                                                    AE1158
                                                               AE1159
## 1 0.4685541 0.4791618 0.4817043 0.4907729 0.4707111
                                                            0.4894650
## 2 -1.2963713 -1.3125244 -1.3146074 -1.3443903 -1.2590951 -1.3127958
## 3 1.4683110 1.4749882 1.4717207 1.4606436 1.4621541
## 4 -0.5718176 -0.6045295 -0.6022802 -0.6102236 -0.6377484 -0.6261328
## 5 -0.3152119 -0.3056913 -0.3066851 -0.2899458 -0.3034836 -0.2986412
##
           AE1160
                      AE1161
                                   AE1162
                                                AE1163
                                                            AE1164
## 1 0.560050266 0.56716303 0.556002973 0.559808462 0.56777725
## 2 -0.659407594 -0.70395081 -0.603571663 -0.656599303 -0.69711689
```

```
## 3 1.159134035 1.16559390 1.128134362 1.158098245 1.17289646
## 4 -1.811127781 -1.77119449 -1.836104417 -1.817277603 -1.77209577
## 5 -0.009859003 -0.01541975 -0.008076561 -0.007626057 -0.02360122
##
          AE1165
                       AE1166
                                   AE1167
                                                AE1168
                                                            AE1169
## 1 0.563552918 0.570237586 0.57621058 0.570855527 0.56843546
## 2 -0.637641321 -0.663459154 -0.71511153 -0.657881684 -0.72536901
## 3 1.131615781 1.144621835 1.16222723 1.131294763 1.15077538
## 4 -1.823423184 -1.825492717 -1.76059351 -1.815757369 -1.63119344
## 5 -0.005208074 -0.004119381 -0.02260765 -0.005452371 -0.06474766
##
                                                       AE1174
          AE1170
                    AE1171
                               AE1172
                                           AE1173
                                                                    AE1175
## 1 0.57096347 0.5685684 0.5708907 0.56730121 0.56847642 0.548673130
## 2 -0.72956508 -0.7469951 -0.7377530 -0.71086443 -0.72282577 -0.822227473
## 3 1.14679012 1.1754322 1.1751532 1.16706575 1.17839994 1.155013738
## 4 -1.62581136 -1.6047913 -1.6153435 -1.62795150 -1.61537253 -1.602217409
## 5 -0.06564811 -0.0789133 -0.0815438 -0.08229614 -0.08912917 -0.003278435
##
           AE1176
                      AE1177
                                   AE1178
                                                AE1179
                                                             AE1180
## 1 0.554284664 0.54980088 0.553864011 0.551009763 0.551675756
## 2 -0.856315033 -0.82038617 -0.852661005 -0.844702189 -0.823259963
## 3 1.158201568 1.15543006 1.156852192 1.163838498 1.151822431
## 4 -1.569161687 -1.59562654 -1.576154352 -1.583907268 -1.602058266
## 5 -0.007978905 -0.00905638 -0.005444638 -0.006821303 -0.004418364
num_genes <- table(km$cluster)</pre>
num_genes
##
##
        2
           3 4
     1
                    5
## 279 141 135 120 246
plot(num_genes, type="l", main='num of genes in each cluster')
```



plot\_heatmap(km)



# try again with less clusters
km <- kmeans(rawcounts\_scaled,3)
km\$centers</pre>

```
AE1149
##
        AE1148
                              AE1150
                                         AE1151
                                                    AE1152
                                                              AE1153
## 1 -0.2359628 -0.2569866 -0.2628749 -0.2498311 -0.3033400 -0.3263202
## 2 -1.0897526 -1.0578491 -1.0588992 -1.0754732 -0.9956799 -0.9289602
     0.9421614 0.9443945 0.9513452 0.9479200 0.9544944
                                                           0.9367624
##
        AE1154
                              AE1156
                   AE1155
                                         AE1157
                                                    AE1158
                                                               AE1159
## 1 -0.3285240 -0.3204519 -0.3247792 -0.3044573 -0.3181097 -0.3168620
## 2 -0.9224395 -0.9499582 -0.9432554 -0.9797709 -0.9325391 -0.9582931
## 3 0.9349843 0.9437989 0.9441724 0.9456099 0.9302650 0.9452465
##
         AE1160
                    AE1161
                                AE1162
                                            AE1163
                                                        AE1164
                                                                   AE1165
## 1 0.03374135 0.0278140 0.04422371 0.03676334 0.02076732 0.04338717
## 2 -1.36041448 -1.3633053 -1.35171193 -1.36449421 -1.35355507 -1.35932245
## 3
     0.82571167
                 0.8338699
                            0.80900981 0.82507007 0.83521581
                                                               0.81472352
##
         AE1166
                     AE1167
                                 AE1168
                                             AE1169
                                                         AE1170
                                                                     AE1171
     0.04070543 \quad 0.02474592 \quad 0.04333567 \quad -0.01050916 \quad -0.01174189 \quad -0.02645446
## 1
  2 -1.37172915 -1.36239212 -1.36760089 -1.31968529 -1.31536911 -1.30887236
## 3 0.82544492 0.83656652 0.82002231 0.84714767 0.84572955
                                                               0.85731922
##
         AE1172
                     AE1173
                                 AE1174
                                             AE1175
                                                         AE1176
                                                                     AE1177
## 1 -0.02671463 -0.02737011 -0.03305205 0.02437834 0.02164494 0.01947869
## 2 -1.31136582 -1.30026943 -1.29824918 -1.35161032 -1.35404301 -1.34482053
                            0.85284727
## 3
     0.85917638
##
         AE1178
                     AE1179
                                AE1180
## 1 0.02218627 0.02069012 0.0230893
```

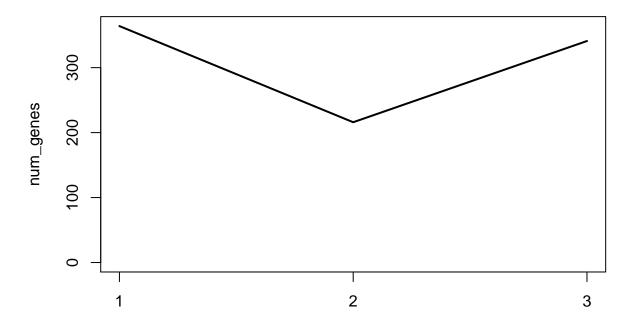
```
## 2 -1.35443662 -1.35491449 -1.3511697
## 3 0.83425955 0.83615931 0.8312262

num_genes <- table(km$cluster)

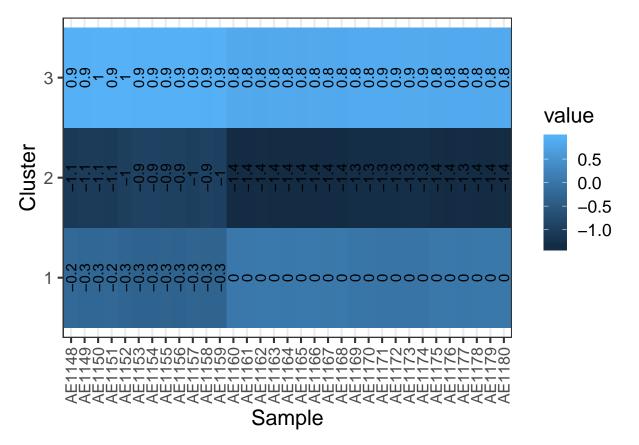
num_genes

##
## 1 2 3
## 364 216 341

plot(num_genes, type="l", main='num of genes in each cluster')</pre>
```



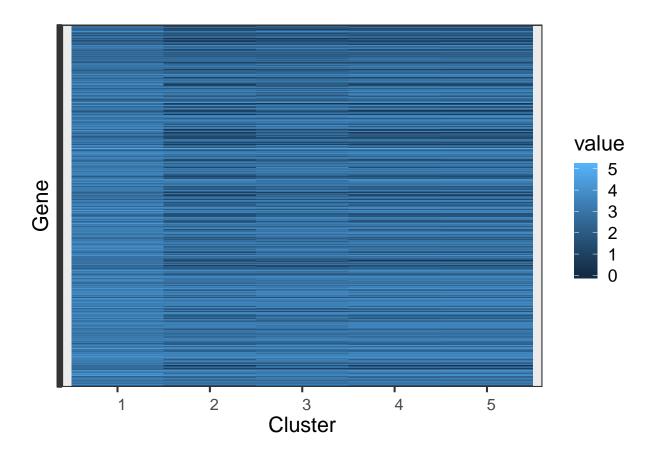
```
plot_heatmap(km)
```



```
# a plotting function
plot_heatmap <- function(km) {</pre>
  centers <- km$centers %>%
      tbl_df() %>%
      rownames_to_column('Cluster') %>%
      gather(Gene, value, -Cluster) %>%
      mutate(
          Cluster = factor(Cluster),
          Gene = factor(Gene)
   ggplot(centers,aes(Cluster,Gene)) + geom_tile(aes(fill=value)) + theme(axis.text.x=element_text(angl.))
}
# cluster the samples
rawcounts_clean <- rawcounts_RN[ rawcounts_RN$sum >= 10, ]
rawcounts_log = log10(1+rawcounts_clean)
DE_res <- read_csv("data/sigresults.csv")</pre>
DE_genes <- DE_res$row</pre>
dat_filtered = rawcounts_log[DE_genes,]
km <- kmeans(t(dat_filtered),5)</pre>
#km$centers
table(km$cluster)
##
## 1 2 3 4 5
```

## 13 3 12 3 3

### plot\_heatmap(km)



### Record sessionInfo()

The sessionInfo() prints version information about R and any attached packages. It's a good practice to always run this command at the end of your R session and record it for the sake of reproducibility in the future.

### sessionInfo()

```
## R version 3.5.1 (2018-07-02)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 17134)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_Israel.1252 LC_CTYPE=English_Israel.1252
## [3] LC_MONETARY=English_Israel.1252 LC_NUMERIC=C
## [5] LC_TIME=English_Israel.1252
##
## attached base packages:
##
## attached base packages:
## [1] stats graphics grDevices utils datasets methods base
```

```
##
## other attached packages:
## [1] bindrcpp_0.2.2 tidyr_0.8.2
                                    tibble_1.4.2
                                                   dplyr_0.7.8
## [5] readr_1.2.1
                     ggplot2_3.1.0 knitr_1.20
## loaded via a namespace (and not attached):
## [1] Rcpp_1.0.0
                        pillar_1.3.0
                                         compiler_3.5.1
                                                          plyr_1.8.4
## [5] bindr_0.1.1
                        tools_3.5.1
                                         digest_0.6.18
                                                          evaluate_0.12
## [9] gtable_0.2.0
                        pkgconfig_2.0.2 rlang_0.3.0.1
                                                          cli_1.0.1
                        withr_2.1.2
## [13] yaml_2.2.0
                                         stringr_1.3.1
                                                          hms_0.4.2
## [17] grid_3.5.1
                        tidyselect_0.2.5 glue_1.3.0
                                                          R6_2.3.0
## [21] fansi_0.4.0
                        rmarkdown_1.11
                                         purrr_0.2.5
                                                          magrittr_1.5
## [25] scales_1.0.0
                        codetools_0.2-15 htmltools_0.3.6
                                                          assertthat_0.2.0
## [29] colorspace_1.3-2 labeling_0.3
                                         utf8_1.1.4
                                                          stringi_1.2.4
## [33] lazyeval_0.2.1 munsell_0.5.0
                                         crayon_1.3.4
```