Khan Inan hw 4

2023-04-07

```
#STEP 1
GSE124548.raw.fixed <- read.delim("C:/Users/khani/Downloads/GSE124548.raw.fixed.txt", header=FALSE)
#Specifying columns with HC or CF base
readcount <- GSE124548.raw.fixed[, c(1,4,5,7,8,10,11,13,14,15,16,18,19,21,22,24,26,27,28,29,30,
                                     32,33,35,36,37,38,39,40,41,42,43,45,47,49)]
#Setting row names and column names
rownames(readcount) <- readcount[,1]
readcount <- readcount[,-1]
colnames(readcount) = readcount[1,]
readcount <- readcount[-1,]
readcount = data.matrix(readcount)
head(readcount)
```

```
##        Raw_10_HC_Auto_066_237 Raw_11_Orkambi_006_Base Raw_13_HC_Auto_068_239
## 1                        3482                    2013                   2281
## 503538                   2712                    1720                   1468
## 144571                   3348                     114                   2310
## 8086                      767                    3146                   3230
## 65985                    1149                     352                    238
## 195827                   1018                     164                   1137
##        Raw_14_Orkambi_007_Base Raw_16_HC_Auto_072_243 Raw_17_Orkambi_009_Base
## 1                         3105                    405                     702
## 503538                    2266                   3923                    4164
## 144571                    2404                     92                    3120
## 8086                       238                   1686                     637
## 65985                      787                   1452                    1840
## 195827                    1063                   1666                    2174
##        Raw_19_HC_Auto_074_245 Raw_1_Orkambi_001_Base Raw_20_Orkambi_010_Base
## 1                          71                   3242                      94
## 503538                   2908                   2105                    3143
## 144571                   1167                    625                     693
## 8086                      630                   3225                     522
## 65985                     670                    381                    1276
## 195827                   1582                    254                    1535
##        Raw_21_HC_Immune_004 Raw_23_HC_Auto_076_247 Raw_24_Orkambi_012_Base
## 1                       275                   3450                      60
## 503538                 3799                   3474                    3236
## 144571                 4051                   3814                     513
## 8086                   1220                   1349                     465
## 65985                  1329                   1010                    1454
## 195827                 1982                   1757                    1790
##        Raw_26_HC_Auto_078_249 Raw_27_Orkambi_013_Base Raw_29_HC_Auto_080_251
## 1                        3039                    3620                    667
## 503538                   2548                    2655                   3931
## 144571                   3364                    2774                    163
## 8086                      304                    1039                   1479
## 65985                     581                     897                   1252
## 195827                    697                    1287                   1773
##         Raw_30_Orkambi_014_Base Raw_3_HC_Auto_062_233 Raw_42_HC_Auto_089_261
## 1                          2955                  3198                   3549
## 503538                     1967                  2484                   2953
## 144571                     2010                  2103                   3757
## 8086                        161                   655                    601
## 65985                       311                   357                    778
## 195827                      760                   792                   1173
##         Raw_4_HC_Immune_002 Raw_5_Orkambi_002_Base Raw_7_HC_Auto_064_235
## 1                      2560                   2163                  3626
## 503538                 2166                   2006                  3421
## 144571                 3247                   3217                  3544
## 8086                    339                    424                  1064
## 65985                   493                    902                   813
## 195827                  600                    891                  1107
##         Raw_8_Orkambi_004_Base Raw_HC_Auto_082_253 Raw_HC_Auto_084_255
## 1                         2867                 452                2981
## 503538                    2548                3191                2047
## 144571                     520                3324                2276
## 8086                      3108                 939                  34
## 65985                     3364                1063                 518
## 195827                       3                1857                 558
##         Raw_HC_Auto_088_259 Raw_HC_Auto_091_263 Raw_HC_Auto_093_265
## 1                      3699                   2                 177
## 503538                 3080                3059                4457
## 144571                  713                3790                1008
## 8086                     42                3755                1388
## 65985                  1058                 926                2094
## 195827                 1401                1967                2779
##         Raw_HC_Auto_095_267 Raw_HC_Immune_006 Raw_HC_Immune_008
## 1                      3886              3786               430
## 503538                 2704              2728              2816
## 144571                  282              3583              2466
## 8086                    256               186              3665
## 65985                  1115              1320               778
## 195827                 1642              1571              1794
##         Raw_Orkambi_015_Base Raw_Orkambi_016_Base Raw_Orkambi_017_Base
## 1                       3899                 1854                  262
## 503538                  3498                 1970                 3153
## 144571                  3104                  966                 3946
## 8086                    1255                 2518                 4142
## 65985                   1377                  458                 1606
## 195827                  1357                  499                 1799
##         Raw_Orkambi_018_Base
## 1                        244
## 503538                  2981
## 144571                   133
## 8086                       2
## 65985                   1705
## 195827                  2058
```

```
#STEP 2
#Creating data frame with condition column
expgroup <- data.frame(condition = c("HC", "CF", "HC", "CF", "HC", "CF", "HC", "CF", "CF", "HC",
                                     "HC", "CF", "HC", "CF", "HC", "CF", "HC", "HC", "HC", "CF",
                                     "HC", "CF", "HC", "HC", "HC", "HC", "HC", "HC", "HC", "HC",
                                     "CF", "CF", "CF", "CF"), row.names = colnames(readcount))
head(expgroup)
```

```
##                          condition
## Raw_10_HC_Auto_066_237         HC
## Raw_11_Orkambi_006_Base        CF
## Raw_13_HC_Auto_068_239         HC
## Raw_14_Orkambi_007_Base        CF
## Raw_16_HC_Auto_072_243         HC
## Raw_17_Orkambi_009_Base        CF
```

```
#STEP 3
#creating the matrix for the counts dataset
cds <- DESeqDataSetFromMatrix(countData = readcount, colData = expgroup, design = ~ condition)
```

```
## Warning in DESeqDataSet(se, design = design, ignoreRank): some variables in
## design formula are characters, converting to factors
```

```
head(cds)
```

```
## class: DESeqDataSet
## dim: 6 34
## metadata(1): version
## assays(1): counts
## rownames(6): 1 503538 ... 65985 195827
## rowData names(0):
## colnames(34): Raw_10_HC_Auto_066_237 Raw_11_Orkambi_006_Base ...
##   Raw_Orkambi_017_Base Raw_Orkambi_018_Base
## colData names(1): condition
```

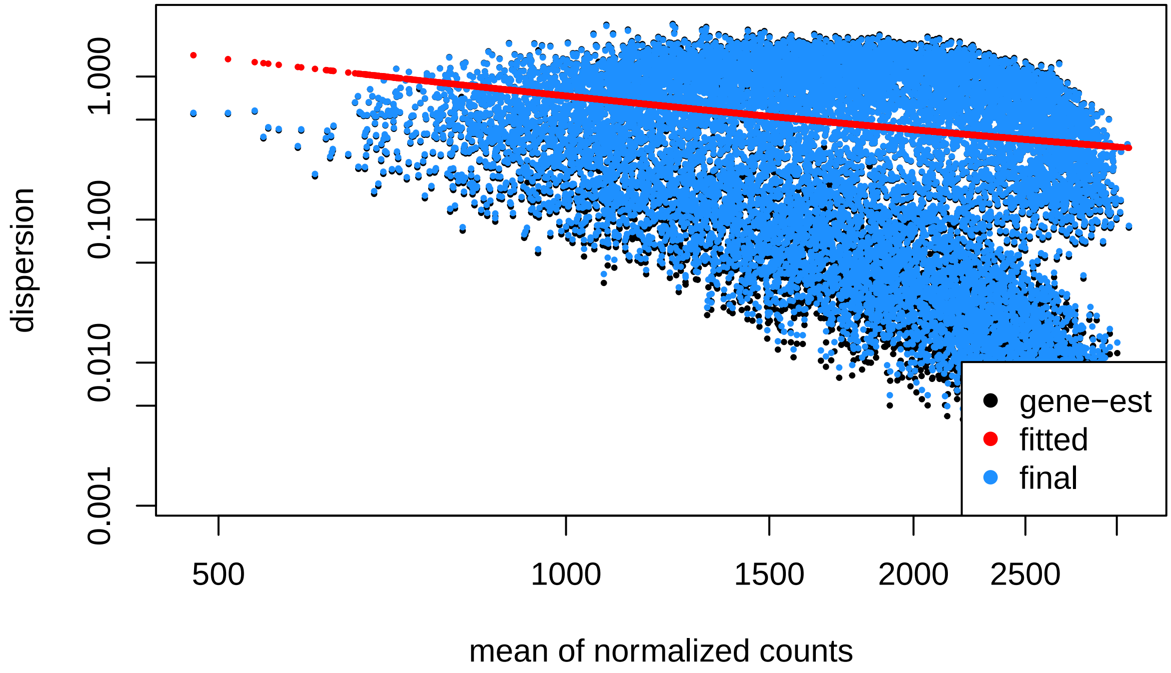
```
#STEP 4
#correcting the size
cds <- estimateSizeFactors(cds)
cds <- estimateDispersions(cds)
```

```
## gene-wise dispersion estimates
```

```
## mean-dispersion relationship
```

```
## final dispersion estimates
```

```
#plotting
plotDispEsts(cds)
```



The graph essentially tells me the about how the dispersion related to the normalized counts. The red dotted line respresents the fitted line for the estimates of the dispersion. The grap itself shows me that the dispersion level increased along with the means. The dispersion cluster does not and should not show any kind of bending or skew

```
#STEP 5
#doing the differential analysis
cds <- DESeq(cds)
```

```
## using pre-existing size factors

## estimating dispersions

## found already estimated dispersions, replacing these

## gene-wise dispersion estimates

## mean-dispersion relationship

## final dispersion estimates

## fitting model and testing

## -- replacing outliers and refitting for 18 genes
## -- DESeq argument 'minReplicatesForReplace' = 7
## -- original counts are preserved in counts(dds)
## estimating dispersions
```

```
## fitting model and testing
```

```r
#getting the results
results <- results(cds)
```

```r
#STEP 6
#subsetting the genes with specific parameters such as p-value and log2foldchange values
diffexpgenes <- subset(results, padj < 0.05 & abs(log2FoldChange) > 1)
#counting the genes in this subset
nrow(diffexpgenes)
```

```
## [1] 78
```

```r
#STEP 7
#normalizing the data from the counts
normvalues <- counts(cds, normalized = TRUE)
head(normvalues)
```

```
##        Raw_10_HC_Auto_066_237 Raw_11_Orkambi_006_Base Raw_13_HC_Auto_068_239
## 1                   3190.7881               2602.4327              2955.8028
## 503538              2485.1859               2223.6385              1902.2878
## 144571              3067.9950                147.3807              2993.3820
## 8086                 702.8531               4067.1899              4185.5515
## 65985               1052.9051                455.0702               308.4091
## 195827               932.8611                212.0213              1473.3660
##        Raw_14_Orkambi_007_Base Raw_16_HC_Auto_072_243 Raw_17_Orkambi_009_Base
## 1                    3153.7934              289.74110                450.5253
## 503538               2301.6090             2806.55390               2672.3467
## 144571               2441.7776               65.81773               2002.3347
## 8086                  241.7400             1206.18146                408.8100
## 65985                 799.3673             1038.77549               1180.8641
## 195827               1079.7045             1191.87326               1395.2166
##        Raw_19_HC_Auto_074_245 Raw_1_Orkambi_001_Base Raw_20_Orkambi_010_Base
## 1                     65.8643              3780.5536                 77.8705
## 503538              2697.6534              2454.6778               2603.6914
## 144571              1082.5865               728.8236                574.0879
## 8086                 584.4297              3760.7296                432.4298
## 65985                621.5364               444.2908               1057.0507
## 195827              1467.5680               296.1939               1271.6088
##        Raw_21_HC_Immune_004 Raw_23_HC_Auto_076_247 Raw_24_Orkambi_012_Base
## 1                  209.8622              2673.7707                 51.7377
## 503538            2899.1506              2692.3709               2790.3868
## 144571            3091.4607              2955.8729                442.3574
## 8086               931.0249              1045.4831                400.9672
## 65985             1014.2067               782.7561               1253.7770
## 195827            1512.5340              1361.6856               1543.5081
##        Raw_26_HC_Auto_078_249 Raw_27_Orkambi_013_Base Raw_29_HC_Auto_080_251
## 1                   3642.2150               3040.6218               504.9728
## 503538              3053.7558               2230.0693              2976.0840
## 144571              4031.7247               2330.0235               123.4041
## 8086                 364.3414                872.7089              1119.7223
## 65985                696.3234                753.4359               947.8650
## 195827               835.3484               1081.0167              1342.3040
##        Raw_30_Orkambi_014_Base Raw_3_HC_Auto_062_233 Raw_42_HC_Auto_089_261
## 1                    3568.6726             3452.0078              2893.6901
## 503538               2375.4920             2681.2968              2407.7394
## 144571               2427.4220             2270.0351              3063.2837
## 8086                  194.4353              707.0247               490.0275
## 65985                 375.5862              385.3555               634.3451
## 195827                917.8312              854.9062               956.4098
##        Raw_4_HC_Immune_002 Raw_5_Orkambi_002_Base Raw_7_HC_Auto_064_235
## 1                2947.2962              2170.0541             3277.2420
## 503538           2493.6889              2012.5421             3091.9595
## 144571           3738.2308              3227.4915             3203.1290
## 8086              390.2865               425.3828              961.6618
## 65985             567.5848               904.9417              734.8036
## 195827            690.7726               893.9058             1000.5259
##        Raw_8_Orkambi_004_Base Raw_HC_Auto_082_253 Raw_HC_Auto_084_255
## 1                 3053.687271            391.7258          3400.75818
## 503538            2713.915301           2765.4804          2335.24052
## 144571             553.860265           2880.7449          2596.48629
## 8086              3310.380202            813.7844            38.78758
## 65985             3583.049871            921.2490           590.94020
## 195827               3.195348           1609.3692           636.57265
##        Raw_HC_Auto_088_259 Raw_HC_Auto_091_263 Raw_HC_Auto_093_265
## 1               2815.73516            1.660216            110.2262
## 503538          2344.54293         2539.300047           2775.5831
## 144571           542.74646         3146.108917            627.7289
## 8086              31.97104         3117.055141            864.3728
## 65985            805.36572          768.679910           1304.0321
## 195827          1066.46255         1632.822227           1730.6138
##        Raw_HC_Auto_095_267 Raw_HC_Immune_006 Raw_HC_Immune_008
## 1                3514.3460         3126.4690          399.8607
## 503538           2445.3916         2252.7754         2618.6226
## 144571            255.0297         2958.8322         2293.1545
## 8086              231.5164          153.5983         3408.1149
## 65985            1008.3623         1090.0526          723.4689
## 195827           1484.9604         1297.3278         1668.2560
##        Raw_Orkambi_015_Base Raw_Orkambi_016_Base Raw_Orkambi_017_Base
## 1                 2897.0026            2366.9243             201.9968
## 503538            2599.0549            2515.0167            2430.9002
## 144571            2306.3083            1233.2518            3042.2874
## 8086               932.4797            3214.6254            3193.3995
## 65985             1023.1271             584.7095            1238.1940
## 195827            1008.2669             637.0524            1386.9932
##        Raw_Orkambi_018_Base
## 1                186.850617
## 503538          2282.793810
## 144571           101.848902
## 8086               1.531562
## 65985           1305.656976
## 195827          1575.977746
```

```
#STEP 8
#creating the matric for the genes that are differentially expressed
diffexpvalues <- normvalues[rownames(diffexpgenes), ]
```

```
#STEP 9
#doing the hierarchical clustering
hc <- hclust(dist(diffexpvalues), method = "complete")
#splitting the tree into 8 groups
groups <- cutree(hc, k = 8)
#counting the number of genes in each of the groups
table(groups)
```

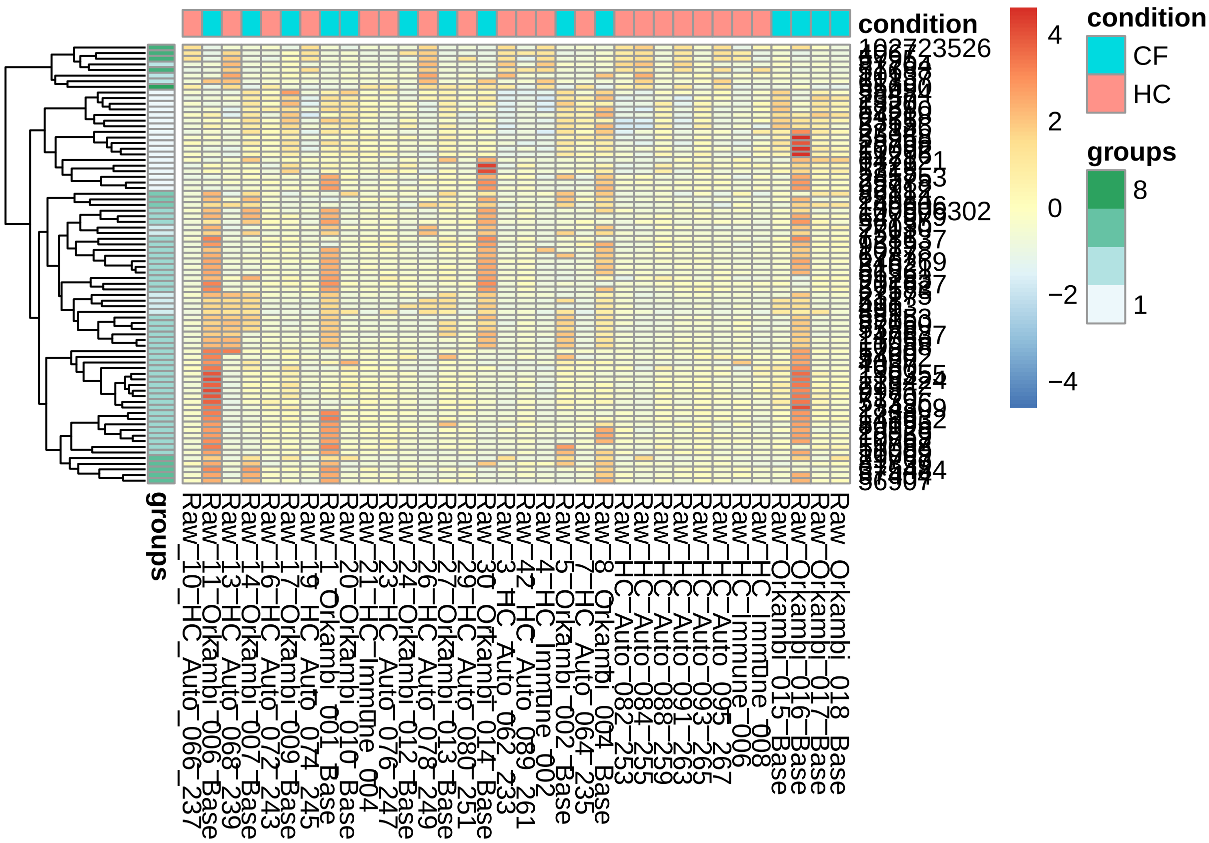
```
## groups
##  1  2  3  4  5  6  7  8
## 18  6  3 38  3  5  4  1
```

The number of genes in each group are shown in the table above

```
#STEP 10
#creating the heatmap. I had some difficulty incorporating annotation_row at first
library(pheatmap)
```

```
## Warning: package 'pheatmap' was built under R version 4.2.3
```

```
groups <- data.frame(groups)
pheatmap(diffexpvalues, scale = "row", cluster_cols = F, shown_rownames = F,
         annotation_col = expgroup, annotation_row = groups)
```



```
#STEP 11
params <- new("GOHyperGParams", geneIds = rownames(diffexpgenes),
              universeGeneIds = rownames(readcount),
              annotation = "org.Hs.eg.db", ontology = "BP",
              pvalueCutoff = 0.001, testDirection = "over")
```

#STEP 12

Based on my analysis throughout this process I have found genes that are differentially expressed between the CF patients before the treatment as well as the healthy patients. In reference to GO terms and GO stats, the genes do show signs of being impacted in CF. The heatmap shows patterns of differential expression as well because some genes show higher expression than others. The hiarchical clustering was useful for grouping the genes based on the patterns they portray during expression. In conclusion, I would say that there is definitely a difference between healthy and CF patients. The next steps would be to find out what are the roles of these genes in terms of vital biological functions, and then we can better understand the effects of CF and what it can target