Final_PS3_Output

Dane Dewees

Wed Nov 29 18:07:05 2017

Part 1 (Experimental Design)

Question 1:

```
The are no response variables in metadata file, the only response variables ass
ociated to this deisgn is genecounts     (in seperate file).
```

Question 2:

```
Explanatory vairables: cell passage, MBNL_induced, RBFOX1_induced.
```

Question 3:

```
The design is factorial (multiple factorial design) due to the levels of each f
actor in     the design. Since there are three facotrs (explained above) that s
hows a full factorial design.
```

Question 4:

```
It is not balanced due to the variety in total count of levels in each factor.
```

Question 5:

```
No because we have different combinations within one factor compared to another
.
```

Part 2 (DESeq2 object construction and initial visualization)

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

After trying to create a matrix, an error occurs due to the model matrix not being in full rank (the column of cell_plate_sample_ID have duplicates which shows that there is a nesting factor in the matrix which prevents it to not be fully ranked). After observing the files, I determined that the technical replication is the sample ID b/c (you test the technical replicates for each sample and since you test it multiple times you have to get rid of the replicate) and the batch variable is student ID (needed later on to correct for student variation with regards to differences in prep).

Part 3 - Construct DESeqDataset, get ridof genes < 10, normalize values, and produce PCA plot for both Technical and Batch effect variables

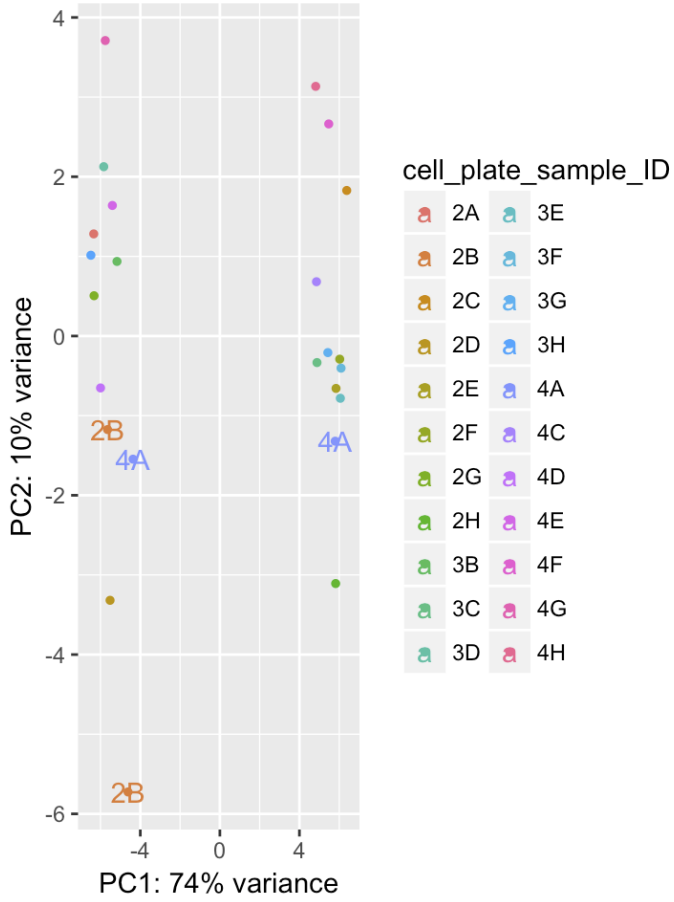
```
                        *Technical Replication Variance*
```

```
## class: DESeqDataSet
## dim: 52636 24
## metadata(1): version
## assays(1): counts
## rownames(52636): ENSMUSG00000000001 ENSMUSG00000000003 ...
##   ENSMUSG00000114967 ENSMUSG00000114968
## rowData names(0):
## colnames(24): 1_2A_control 2_2B_control ... 32_4G_both 34_4H_both
## colData names(5): cell_plate_sample_ID cell_passage MBNL_induced
##   RBFOX1_induced student
```

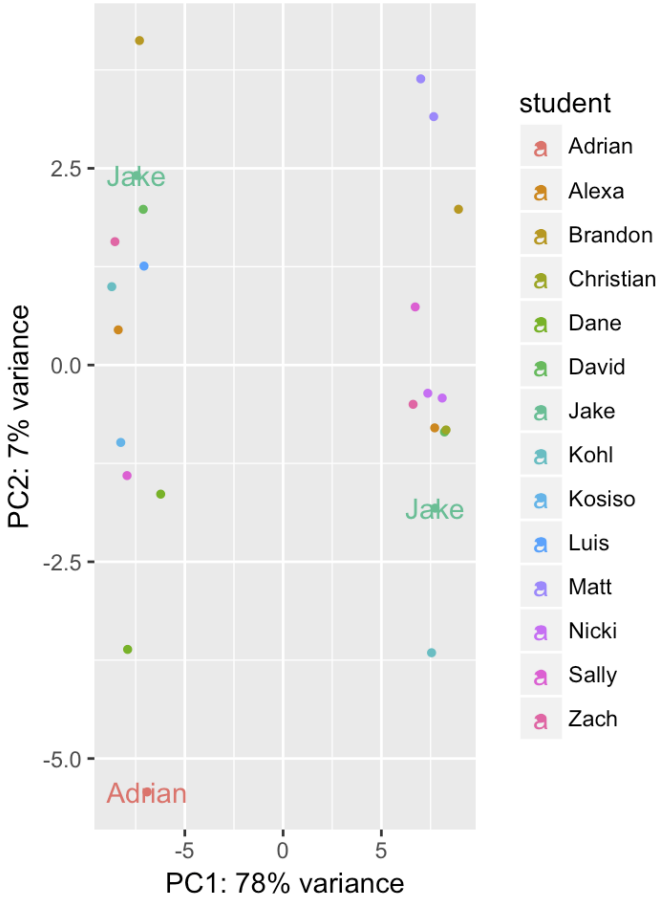
```
                          *Batch Effect Variance*
```

```
## class: DESeqDataSet
## dim: 52636 24
## metadata(1): version
## assays(1): counts
## rownames(52636): ENSMUSG00000000001 ENSMUSG00000000003 ...
##   ENSMUSG00000114967 ENSMUSG00000114968
## rowData names(0):
## colnames(24): 1_2A_control 2_2B_control ... 32_4G_both 34_4H_both
## colData names(5): cell_plate_sample_ID cell_passage MBNL_induced
##   RBFOX1_induced student
```

PCA plot for Technical replication



PCA plot for Batch effect variable



There is not a clear batch effect b/c there is no clear clustering shown that indicates batch effect (Nicky and Matt show some signs of batch effect) when looking at the tehnical replication plots. There is some slight composition similarirty with regards to technical replicates (reference cluster above 4A on +4 value of x-axis).

Because only 2 libraries were technically duplicated, drop the duplicate with fewer total reads in both cases from the original counts matrix and metadata data frame for subsequent analyses.

Part 3 (Main effect tests)

Construct a new DESeqDataSet object, specifying an additive model with all terms except the technical replicate and batch effect terms (we decided to exclude those).

```
## estimating size factors
```

```
## estimating dispersions
```

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

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

```
## final dispersion estimates
```

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

```
## log2 fold change (MAP): RBFOX1_induced Yes vs No
## Wald test p-value: RBFOX1_induced Yes vs No
## DataFrame with 18185 rows and 6 columns
##                      baseMean log2FoldChange      lfcSE        stat
##                     <numeric>      <numeric>  <numeric>   <numeric>
## ENSMUSG00000000001 2201.692028    -0.08989237 0.05719619 -1.57164966
## ENSMUSG00000000028  948.247532     0.03427040 0.04342429  0.78919884
## ENSMUSG00000000049    1.605198     0.09475364 0.04516143  2.09810991
## ENSMUSG00000000056  215.479566    -0.00135589 0.05911234 -0.02293751
## ENSMUSG00000000058  655.420873    -0.03157591 0.07088280 -0.44546650
## ...                        ...            ...        ...         ...
## ENSMUSG00000114929   0.5650606    0.008531043 0.02412713   0.3535871
## ENSMUSG00000114931   1.1498626   -0.006570610 0.03182535  -0.2064584
## ENSMUSG00000114937   0.8829727   -0.042705017 0.03468622  -1.2311810
## ENSMUSG00000114940  24.1065909    0.050522447 0.08660659   0.5833557
## ENSMUSG00000114951  14.7417609   -0.125955369 0.08563932  -1.4707656
##                        pvalue      padj
##                     <numeric> <numeric>
## ENSMUSG00000000001 0.11603182 0.9078815
## ENSMUSG00000000028 0.42999580 0.9490895
## ENSMUSG00000000049 0.03589544 0.9078815
## ENSMUSG00000000056 0.98170012 0.9998762
## ENSMUSG00000000058 0.65598267 0.9819760
## ...                       ...       ...
## ENSMUSG00000114929  0.7236483 0.9876225
## ENSMUSG00000114931  0.8364329 0.9958447
## ENSMUSG00000114937  0.2182552 0.9140505
## ENSMUSG00000114940  0.5596539 0.9712613
## ENSMUSG00000114951  0.1413545 0.9078815
```

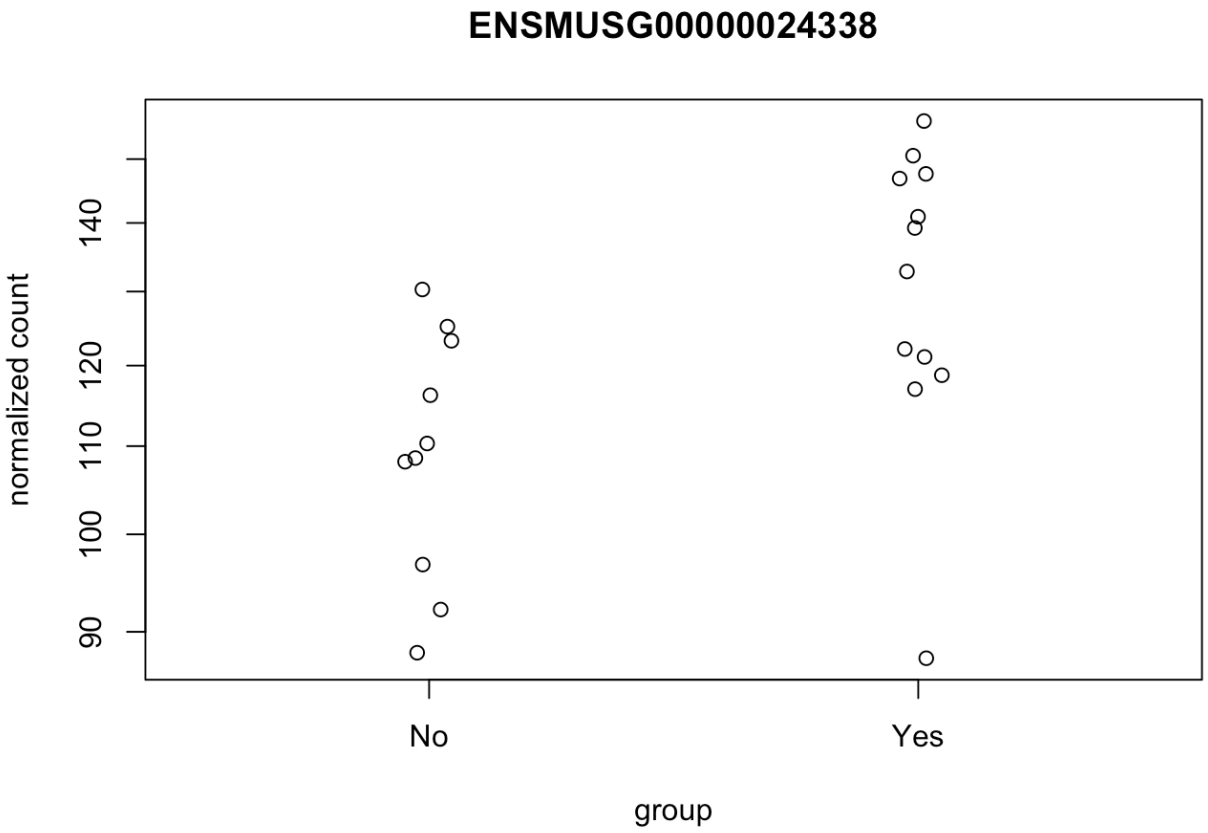
Test for any genes differentially expressed by MBNL induction, controlling for effects of cell passage and RBFOX1 induction.

```
##
## out of 18185 with nonzero total read count
## adjusted p-value < 0.1
## LFC > 0 (up)     : 0, 0%
## LFC < 0 (down)   : 0, 0%
## outliers [1]     : 0, 0%
## low counts [2]   : 0, 0%
## (mean count < 0)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results
```

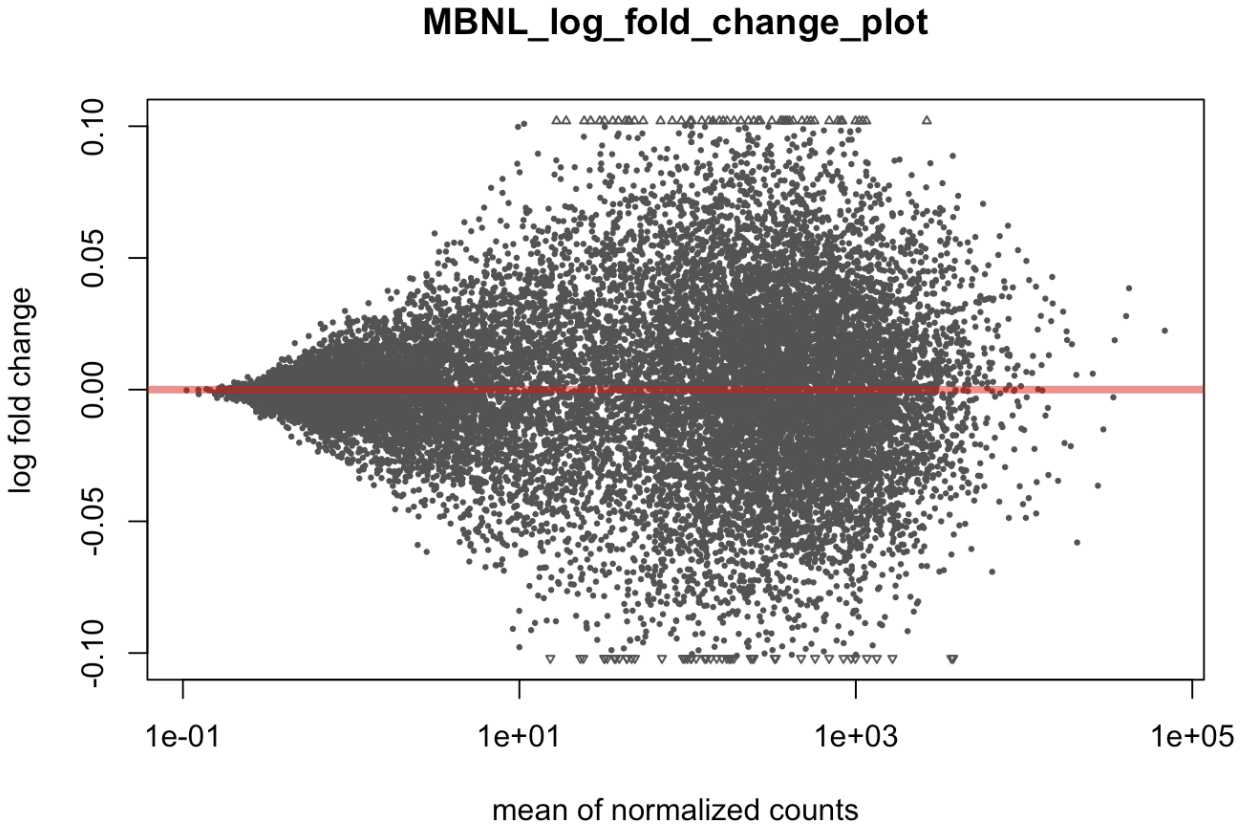
```
## log2 fold change (MAP): MBNL_induced Yes vs No
## Wald test p-value: MBNL_induced Yes vs No
## DataFrame with 18185 rows and 6 columns
##                        baseMean log2FoldChange       lfcSE          stat
##                       <numeric>      <numeric>   <numeric>     <numeric>
## ENSMUSG00000006273    831.49726      0.1106502  0.03365866      3.287422
## ENSMUSG00000026956    539.89646      0.1515608  0.04644197      3.263445
## ENSMUSG00000100158     15.27673     -0.1216126  0.03735461     -3.255627
## ENSMUSG00000024593     18.96491      0.1233709  0.03903559      3.160471
## ENSMUSG00000015468     33.28489     -0.1544724  0.04949740     -3.120818
## ...                         ...            ...         ...           ...
## ENSMUSG00000063275   73.3974206  -1.784464e-05 0.051759723 -3.447592e-04
## ENSMUSG00000113510   96.1791199   1.541744e-05 0.051334561  3.003325e-04
## ENSMUSG00000039200    0.3569061   1.304981e-06 0.008888581  1.468155e-04
## ENSMUSG00000015120 1856.5509577   3.587740e-06 0.045243514  7.929843e-05
## ENSMUSG00000003604  144.2270733  -7.158209e-07 0.051375417 -1.393314e-05
##                         pvalue      padj
##                      <numeric> <numeric>
## ENSMUSG00000006273 0.001011093 0.9999253
## ENSMUSG00000026956 0.001100667 0.9999253
## ENSMUSG00000100158 0.001131424 0.9999253
## ENSMUSG00000024593 0.001575143 0.9999253
## ENSMUSG00000015468 0.001803495 0.9999253
## ...                       ...       ...
## ENSMUSG00000063275   0.9997249 0.9999253
## ENSMUSG00000113510   0.9997604 0.9999253
## ENSMUSG00000039200   0.9998829 0.9999889
## ENSMUSG00000015120   0.9999367 0.9999889
## ENSMUSG00000003604   0.9999889 0.9999889
```

There are zero genes that are differentially expression when looking at summary of MBNL.

Plot the expression values (compare MBNL Yes vs. MBNL No) of the gene with the highest fold change estimate



Produce an "MA plot" to show the distribution of log2 fold changes and expression levels among genes for the MBNL test.

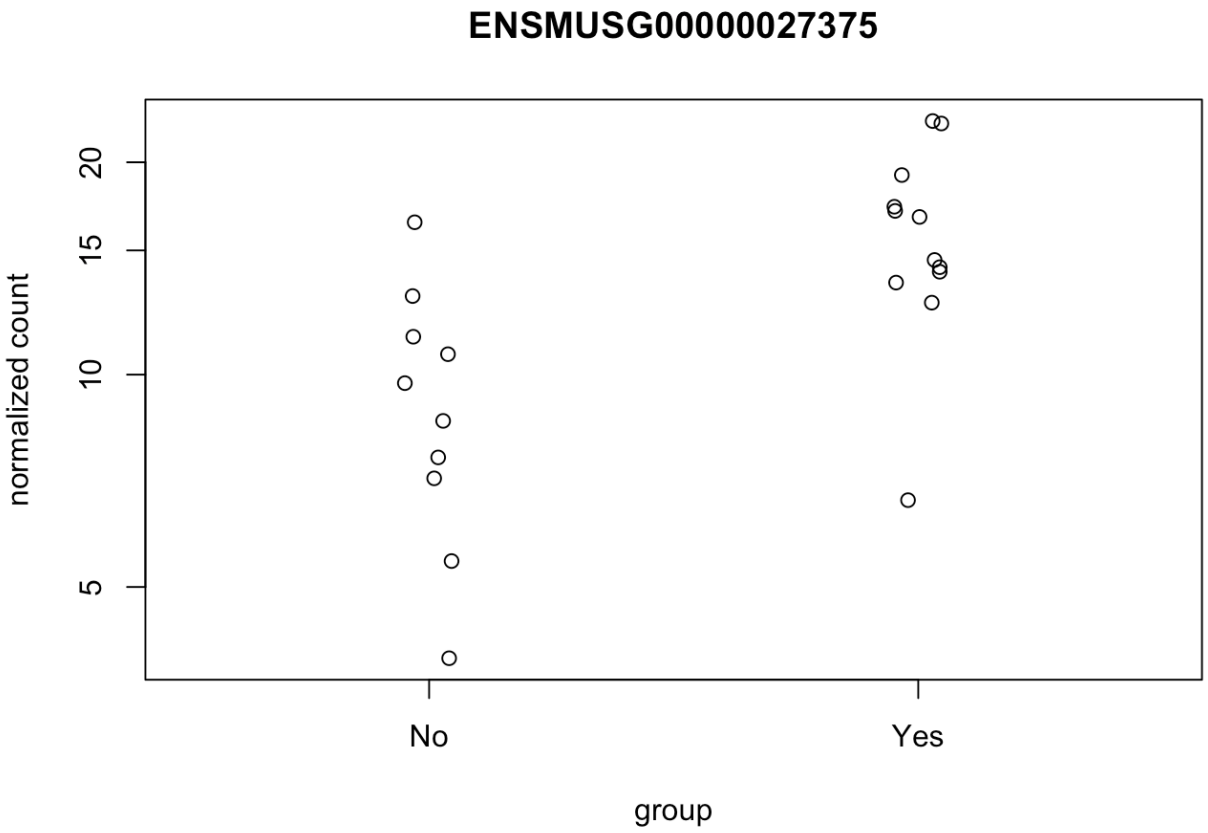


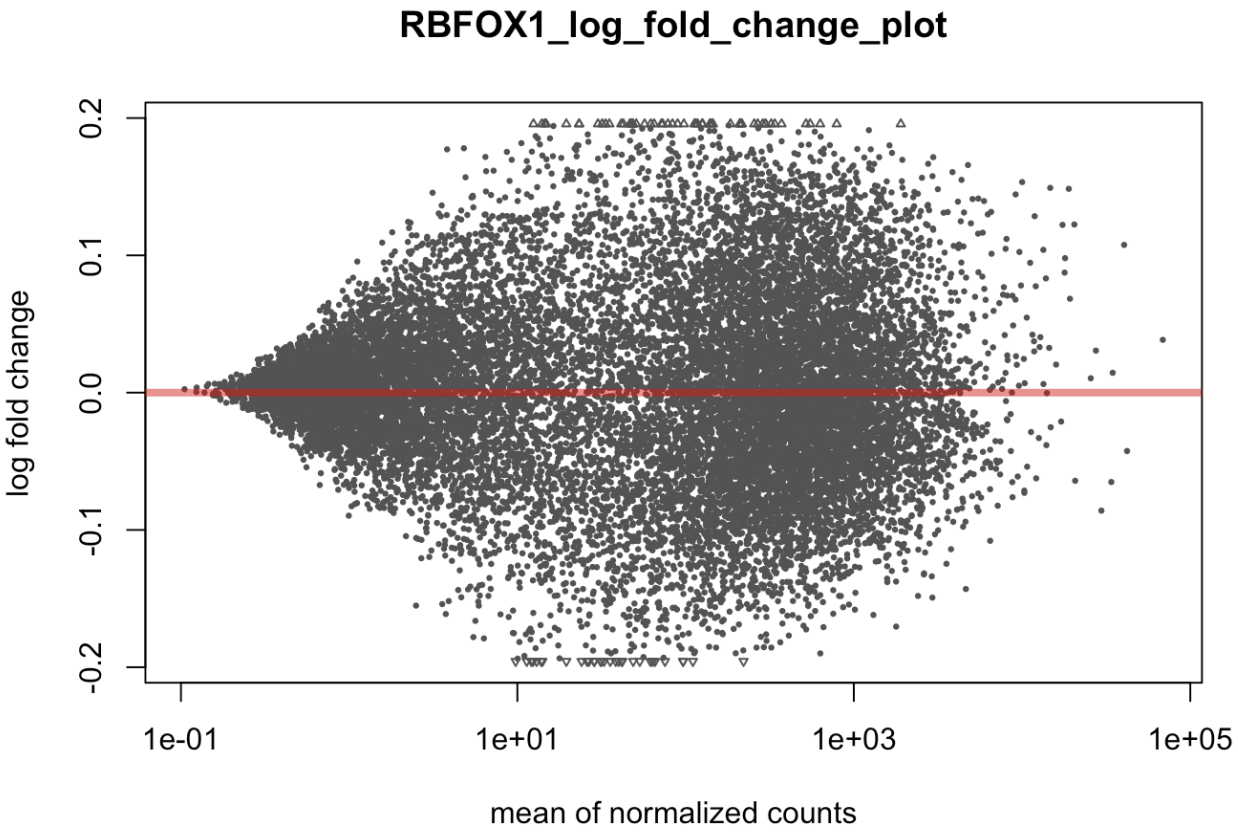
Repeat all steps from above but for the other categorical variable; RBFOX1

```
##
## out of 18185 with nonzero total read count
## adjusted p-value < 0.1
## LFC > 0 (up)     : 0, 0%
## LFC < 0 (down)   : 0, 0%
## outliers [1]     : 0, 0%
## low counts [2]   : 0, 0%
## (mean count < 0)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results
```

```
## log2 fold change (MAP): RBFOX1_induced Yes vs No
## Wald test p-value: RBFOX1_induced Yes vs No
## DataFrame with 18185 rows and 6 columns
##                     baseMean log2FoldChange      lfcSE          stat
##                    <numeric>      <numeric>  <numeric>     <numeric>
## ENSMUSG00000098009   53.38676     -0.3032441 0.08339899     -3.636064
## ENSMUSG00000032733  255.73233      0.2329245 0.06540788      3.561108
## ENSMUSG00000027375   12.46480      0.2916713 0.08352922      3.491847
## ENSMUSG00000038080  220.91149     -0.2487006 0.07334744     -3.390719
## ENSMUSG00000016526   78.51405      0.2628818 0.07822387      3.360633
## ...                       ...            ...        ...           ...
## ENSMUSG00000020486   1.208891   1.472804e-05 0.03935701  3.742164e-04
## ENSMUSG00000052763 282.857739   1.816731e-05 0.05505670  3.299747e-04
## ENSMUSG00000059022  14.107892   1.874500e-05 0.08488216  2.208356e-04
## ENSMUSG00000039478 120.654399  -1.555672e-05 0.07466331 -2.083583e-04
## ENSMUSG00000025353 285.329216  -1.106155e-06 0.05722957 -1.932839e-05
##                         pvalue      padj
##                      <numeric> <numeric>
## ENSMUSG00000098009 0.0002768356 0.9078815
## ENSMUSG00000032733 0.0003692938 0.9078815
## ENSMUSG00000027375 0.0004796923 0.9078815
## ENSMUSG00000038080 0.0006970946 0.9078815
## ENSMUSG00000016526 0.0007776394 0.9078815
## ...                         ...       ...
## ENSMUSG00000020486    0.9997014 0.9998887
## ENSMUSG00000052763    0.9997367 0.9998887
## ENSMUSG00000059022    0.9998238 0.9998887
## ENSMUSG00000039478    0.9998338 0.9998887
## ENSMUSG00000025353    0.9999846 0.9999846
```

There are zero genes that are differentially expression when looking at summary of RBFOX1.


ENSMUSG00000027375
normalized count
5
10
15
20
No
Yes
group


RBFOX1_log_fold_change_plot
log fold change
-0.2
-0.1
0.0
0.1
0.2
1e-01
1e+01
1e+03
1e+05
mean of normalized counts

Part 4 (Interaction effects tests using contrasts)

```
## 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
```

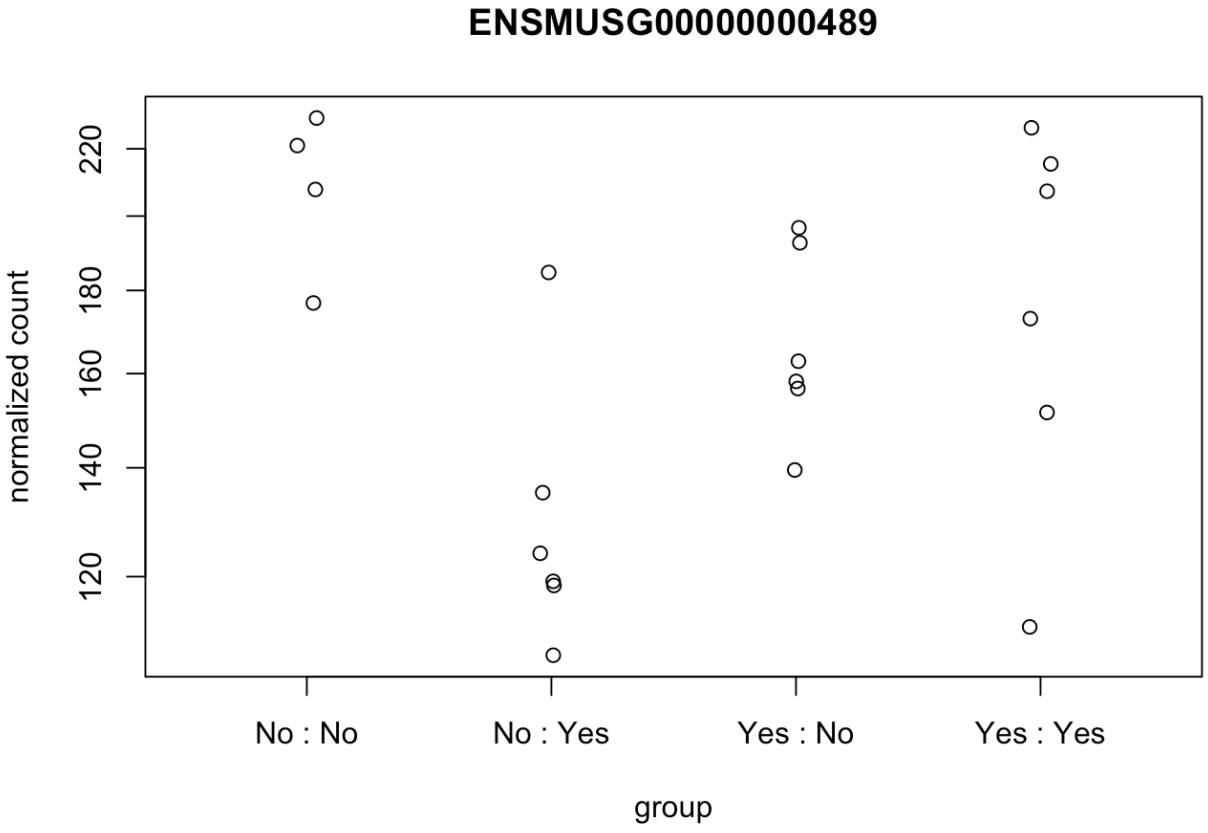
```
## [1] "Intercept"                          "MBNL_induced_Yes_vs_No"
## [3] "RBFOX1_induced_Yes_vs_No"           "MBNL_inducedYes.RBFOX1_inducedYes"
```

```
##
## out of 18185 with nonzero total read count
## adjusted p-value < 0.1
## LFC > 0 (up)     : 0, 0%
## LFC < 0 (down)   : 0, 0%
## outliers [1]     : 0, 0%
## low counts [2]   : 0, 0%
## (mean count < 0)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results
```

```
##
## out of 18185 with nonzero total read count
## adjusted p-value < 0.1
## LFC > 0 (up)     : 759, 4.2%
## LFC < 0 (down)   : 309, 1.7%
## outliers [1]     : 0, 0%
## low counts [2]   : 5994, 33%
## (mean count < 6)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results
```

```
##
## out of 18185 with nonzero total read count
## adjusted p-value < 0.1
## LFC > 0 (up)     : 0, 0%
## LFC < 0 (down)   : 0, 0%
## outliers [1]     : 0, 0%
## low counts [2]   : 0, 0%
## (mean count < 0)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results
```

```
##
## out of 18185 with nonzero total read count
## adjusted p-value < 0.1
## LFC > 0 (up)     : 0, 0%
## LFC < 0 (down)   : 0, 0%
## outliers [1]     : 0, 0%
## low counts [2]   : 0, 0%
## (mean count < 0)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results
```



The statistical power to detect interactions compared to the power to detect main effects shows a

clear differentiation with regards to output. Power of statistical tests increases when treatments and samples are reduced. Main effects model will have a higher power due to less variations within samples and treatment groups. As mentioned from the Logan *Biostatistical Text Book*: .."tests of interactions are typically more powerful than main effects (due to greater available degrees of freedom) and for fixed models, efforts to improve the power of any of the main effects will also benefit the corresponding interactions. Power analyses for mixed and random factorial designs should reflect the appropriate residuals".

Part 5 (Individual Candidate Genes)

```
## using pre-existing size factors
```

```
## estimating dispersions
```

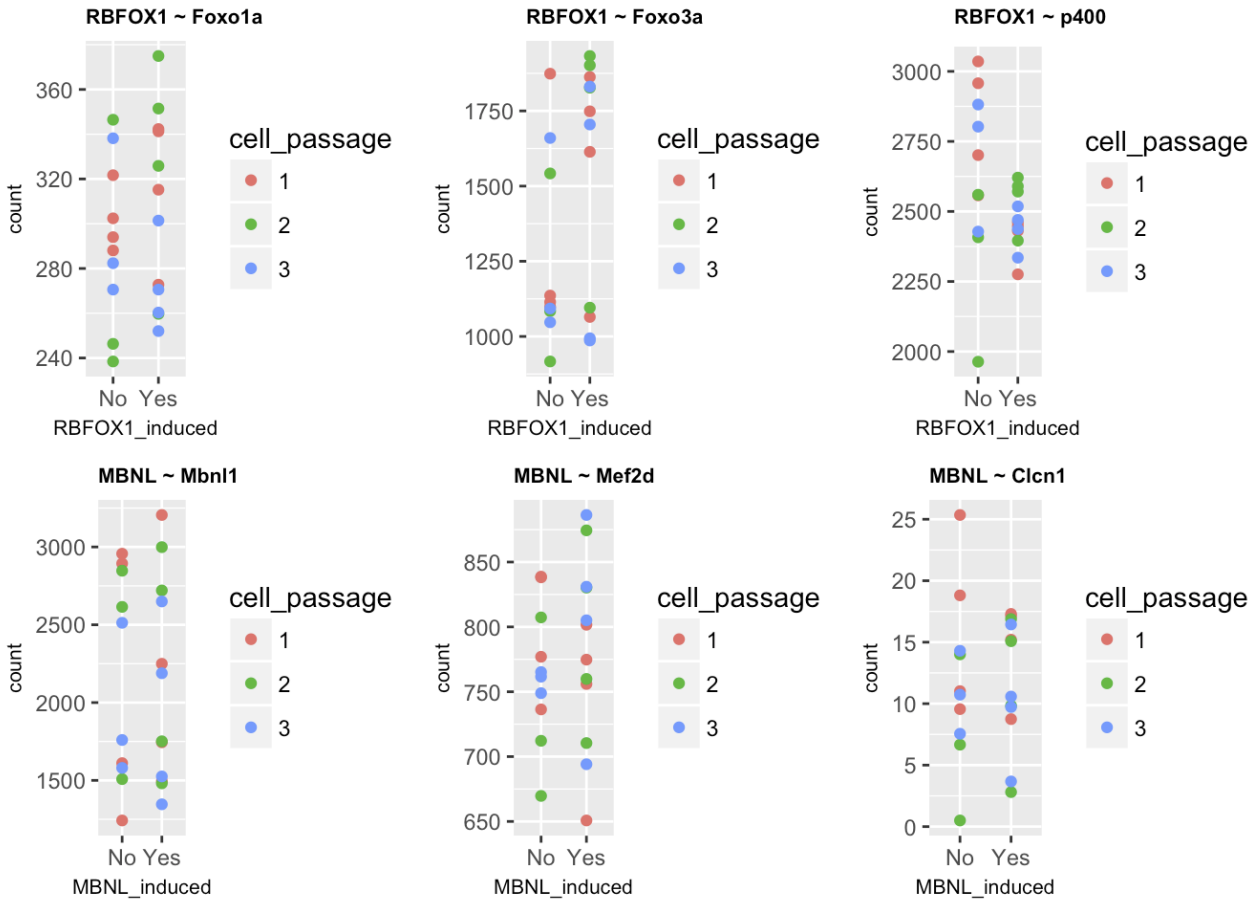
```
## found already estimated dispersions, replacing these
```

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

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

```
## final dispersion estimates
```

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



As shown above, despite picking 6 different genes associated to either the RBFOX1 pathway or the MBNL pathway, no real influence was made with regards to cell passge and gene abundancy. No distribution patterns were made, thus concluding that cell passge had no clear effect on gene counts.

Repeat all steps from above but on new gene counts file

As shown below, there were slight differences with regards to the output from running DESeq2. There was a difference in batch effect spread when plotting the PCA plots (Dane & David seemed to show the largest spread). In particular, there was a difference in the upregulation and downregulation abundancy when looking at the MBNL and RBFOX1 (seperately) & it's effect it has on gene abundancy. For my file, there were no genes that were differentially expressed until I looked at the interaction between the two treatements, whereas in my collegues file, I found that when referencing RBFOX1, it showed a 4.4% upregulation and a 1.8% downregulation (where in my RBFOX1 there was 0% for both pathways). This may be due to the treatment and overall prep of ones library compared to others. Overall plots were relatively the same (of course except the RBFOX1 MA & counts plots).

Part 1 (Experimental Design)

Answered above ^^

Part 2 (DESeq2 object construction and initial visualization)

After trying to creating a matrix, an error occurs due to the model matrix not being in full rank (the column of cell_plate_sample_ID have duplicates which shows that there is a nesting factor in the matrix which prevents it no be fully ranked). After observing the files, I determined that the technical replication is the sample ID b/c (you test the technical replicates for each sample and since you test it multiple times you have to get rid of the replicate) and the batch variable is student ID (needed later on to correct for student variation with regards to differences in prep).

Part 3 - Construct DESeqDataset, get ridof genes < 10, normalize values, and produce PCA plot for both Technical and Batch effect variables

```
                        *Technical Replication Variance*
```

```
## class: DESeqDataSet
## dim: 51826 24
## metadata(1): version
## assays(1): counts
## rownames(51826): ENSMUSG00000000001 ENSMUSG00000000003 ...
##   ENSMUSG00000114194 ENSMUSG00000114195
## rowData names(0):
## colnames(24): 1_2A_control 2_2B_control ... 32_4G_both 34_4H_both
## colData names(5): cell_plate_sample_ID cell_passage MBNL_induced
##   RBFOX1_induced student
```

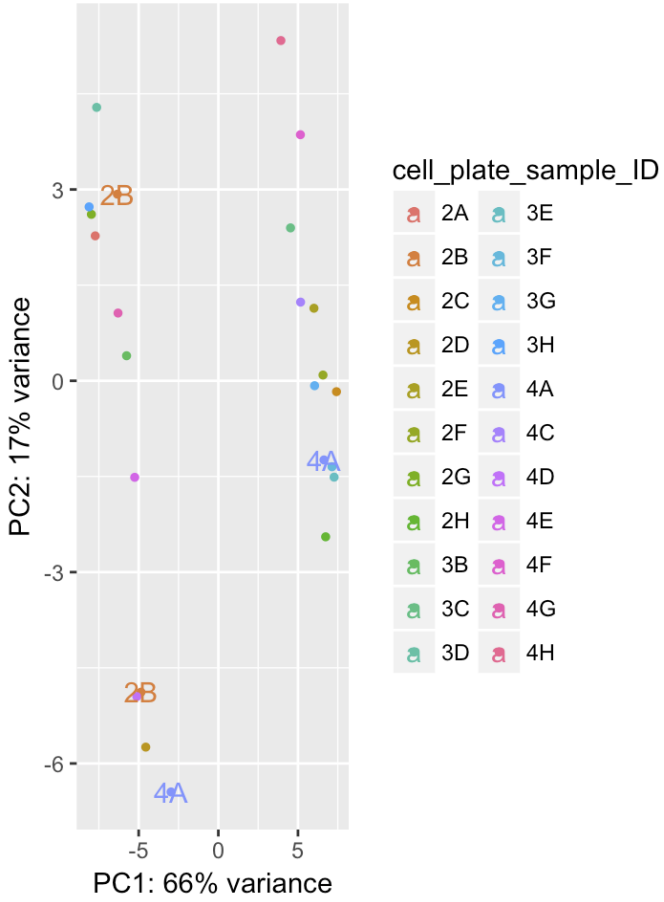
```
## class: DESeqDataSet
## dim: 18850 24
## metadata(1): version
## assays(1): counts
## rownames(18850): ENSMUSG00000000001 ENSMUSG00000000028 ...
##   ENSMUSG00000114184 ENSMUSG00000114193
## rowData names(0):
## colnames(24): 1_2A_control 2_2B_control ... 32_4G_both 34_4H_both
## colData names(5): cell_plate_sample_ID cell_passage MBNL_induced
##   RBFOX1_induced student
```

```
                        *Batch Effect Variance*
```

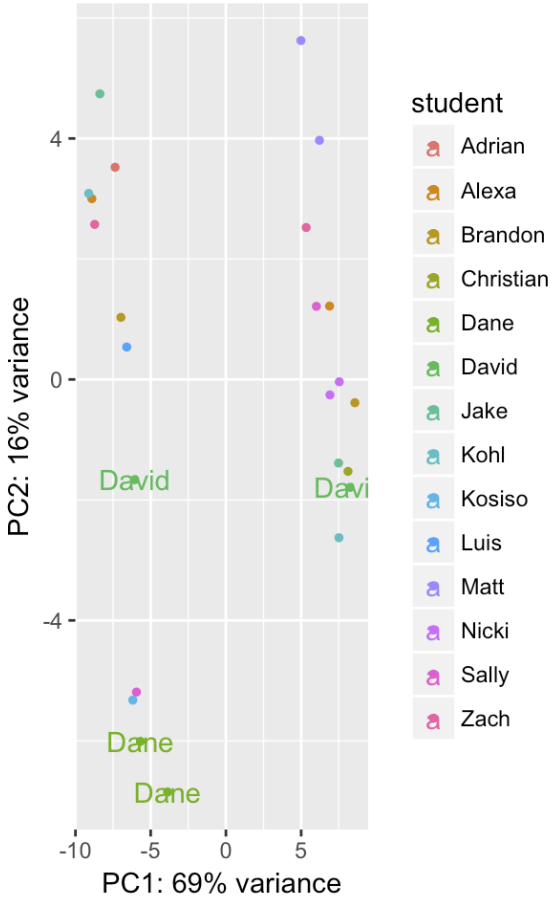
```
## class: DESeqDataSet
## dim: 51826 24
## metadata(1): version
## assays(1): counts
## rownames(51826): ENSMUSG00000000001 ENSMUSG00000000003 ...
##   ENSMUSG00000114194 ENSMUSG00000114195
## rowData names(0):
## colnames(24): 1_2A_control 2_2B_control ... 32_4G_both 34_4H_both
## colData names(5): cell_plate_sample_ID cell_passage MBNL_induced
##   RBFOX1_induced student
```

```
## class: DESeqDataSet
## dim: 18850 24
## metadata(1): version
## assays(1): counts
## rownames(18850): ENSMUSG00000000001 ENSMUSG00000000028 ...
##   ENSMUSG00000114184 ENSMUSG00000114193
## rowData names(0):
## colnames(24): 1_2A_control 2_2B_control ... 32_4G_both 34_4H_both
## colData names(5): cell_plate_sample_ID cell_passage MBNL_induced
##   RBFOX1_induced student
```

```
  *PCA plot for Technical replication*
```




```
*PCA plot for Batch effect variable*
```



There is not a clear batch effect b/c there is no clear clustering shown that indicates batch effect (Nicky and Matt show some signs of batch effect) when looking at the tehnical replication plots. There is some slight composition similarirty with regards to technical replicates (reference cluster above 4A on +4 value of x-axis).

Because only 2 libraries were technically duplicated, drop the duplicate with fewer total reads in both cases from the original counts matrix and metadata data frame for subsequent analyses.

Part 3 (Main effect tests)

Construct a new DESeqDataSet object, specifying an additive model with all terms except the technical replicate and batch effect terms (we decided to exclude those).

```
## estimating size factors
```

```
## estimating dispersions
```

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

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

```
## final dispersion estimates
```

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

```
## log2 fold change (MAP): RBFOX1_induced Yes vs No
## Wald test p-value: RBFOX1_induced Yes vs No
## DataFrame with 18714 rows and 6 columns
##                      baseMean log2FoldChange      lfcSE        stat
##                     <numeric>      <numeric>  <numeric>   <numeric>
## ENSMUSG00000000001 2200.650677  -0.0909728500 0.05719552 -1.59055911
## ENSMUSG00000000028  951.695039   0.0332518533 0.04339497  0.76626055
## ENSMUSG00000000049    1.604615   0.0967761746 0.04615253  2.09687698
## ENSMUSG00000000056  262.533372   0.0007072742 0.05470456  0.01292898
## ENSMUSG00000000058  658.841669  -0.0306275589 0.07073985 -0.43296047
## ...                        ...            ...        ...         ...
## ENSMUSG00000114169   0.5284424    -0.02527088 0.02866975  -0.8814475
## ENSMUSG00000114172  12.6101561     0.10506494 0.08119900   1.2939192
## ENSMUSG00000114179   2.2169903     0.01427454 0.04888209   0.2920197
## ENSMUSG00000114184   9.6896311    -0.03013902 0.08126382  -0.3708787
## ENSMUSG00000114193   3.2008226    -0.08531662 0.05283155  -1.6148802
##                       pvalue      padj
##                     <numeric> <numeric>
## ENSMUSG00000000001 0.11170883 0.9007979
## ENSMUSG00000000028 0.44352128 0.9547628
## ENSMUSG00000000049 0.03600447 0.9007979
## ENSMUSG00000000056 0.98968445 0.9999285
## ENSMUSG00000000058 0.66504349 0.9839573
## ...                       ...       ...
## ENSMUSG00000114169  0.3780757 0.9414748
## ENSMUSG00000114172  0.1956933 0.9043132
## ENSMUSG00000114179  0.7702715 0.9951910
## ENSMUSG00000114184  0.7107279 0.9880835
## ENSMUSG00000114193  0.1063366 0.9007979
```

Test for any genes differentially expressed by MBNL induction, controlling for effects of cell passage and RBFOX1 induction.

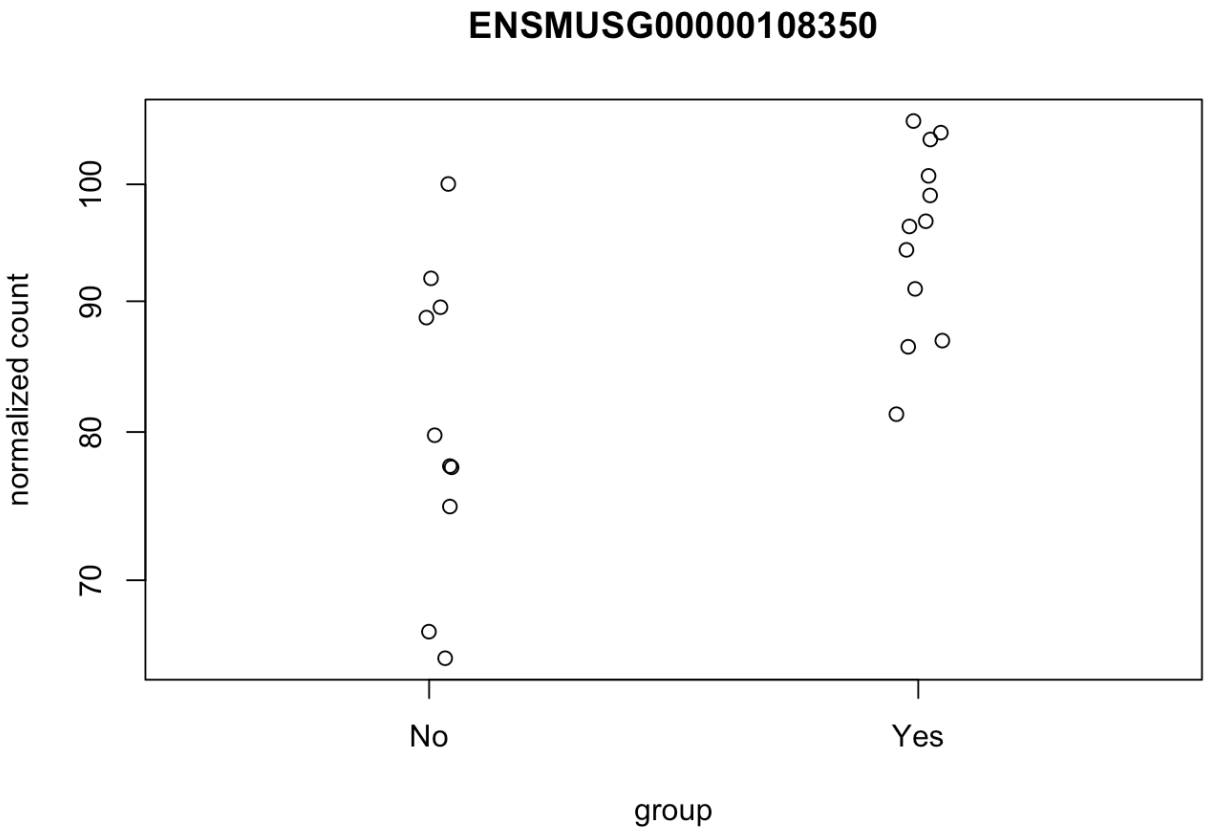
```
## log2 fold change (MAP): MBNL_induced Yes vs No
## Wald test p-value: MBNL_induced Yes vs No
## DataFrame with 6 rows and 6 columns
##                       baseMean log2FoldChange      lfcSE        stat
##                      <numeric>      <numeric>  <numeric>   <numeric>
## ENSMUSG00000000001 2200.650677    0.038461514 0.04650503  0.82703979
## ENSMUSG00000000028  951.695039    0.002432953 0.03863734  0.06296897
## ENSMUSG00000000049    1.604615   -0.019930129 0.01681539 -1.18523109
## ENSMUSG00000000056  262.533372   -0.039069389 0.04526855 -0.86305805
## ENSMUSG00000000058  658.841669    0.029162011 0.05087527  0.57320603
## ENSMUSG00000000078 1434.933233   -0.029744545 0.05035228 -0.59072881
##                       pvalue      padj
##                    <numeric> <numeric>
## ENSMUSG00000000001 0.4082145 0.9998939
## ENSMUSG00000000028 0.9497912 0.9998939
## ENSMUSG00000000049 0.2359261 0.9998939
## ENSMUSG00000000056 0.3881056 0.9998939
## ENSMUSG00000000058 0.5665052 0.9998939
## ENSMUSG00000000078 0.5547021 0.9998939
```

```
##
## out of 18714 with nonzero total read count
## adjusted p-value < 0.1
## LFC > 0 (up)     : 0, 0%
## LFC < 0 (down)   : 0, 0%
## outliers [1]     : 0, 0%
## low counts [2]   : 0, 0%
## (mean count < 0)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results
```

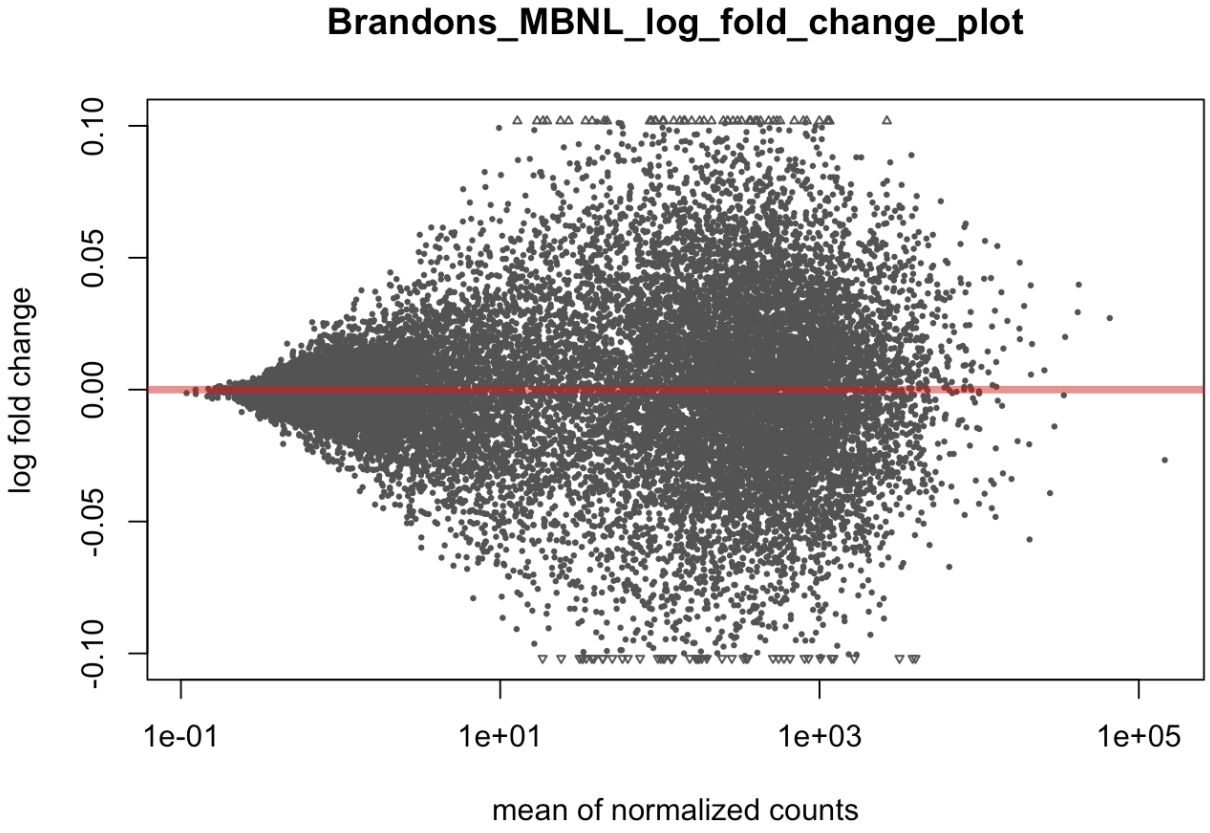
```
## log2 fold change (MAP): MBNL_induced Yes vs No
## Wald test p-value: MBNL_induced Yes vs No
## DataFrame with 18714 rows and 6 columns
##                      baseMean log2FoldChange      lfcSE          stat
##                     <numeric>      <numeric>  <numeric>     <numeric>
## ENSMUSG00000006273   833.21059      0.1117299 0.03367714      3.317679
## ENSMUSG00000100158    18.43454     -0.1301851 0.03950923     -3.295056
## ENSMUSG00000026956   541.69963      0.1504737 0.04650278      3.235800
## ENSMUSG00000108350    88.55361      0.1611581 0.05067616      3.180156
## ENSMUSG00000024338   122.08000      0.1570363 0.05102031      3.077918
## ...                        ...            ...        ...           ...
## ENSMUSG00000018412  790.211641  -2.058793e-05 0.05136335 -4.008292e-04
## ENSMUSG00000028330 1255.947786  -5.492361e-06 0.04264413 -1.287952e-04
## ENSMUSG00000075581    1.484269   1.804744e-06 0.01421285  1.269798e-04
## ENSMUSG00000022601  304.971254   4.020161e-06 0.04546277  8.842755e-05
## ENSMUSG00000038045    1.692337   1.167220e-07 0.01628138  7.169051e-06
##                         pvalue      padj
##                      <numeric> <numeric>
## ENSMUSG00000006273 0.0009076884 0.9998939
## ENSMUSG00000100158 0.0009840205 0.9998939
## ENSMUSG00000026956 0.0012130221 0.9998939
## ENSMUSG00000108350 0.0014719568 0.9998939
## ENSMUSG00000024338 0.0020845203 0.9998939
## ...                         ...       ...
## ENSMUSG00000018412    0.9996802 0.9998939
## ENSMUSG00000028330    0.9998972 0.9999829
## ENSMUSG00000075581    0.9998987 0.9999829
## ENSMUSG00000022601    0.9999294 0.9999829
## ENSMUSG00000038045    0.9999943 0.9999943
```

There are zero genes that are differentially expression when looking at summary of MBNL.

Plot the expression values (compare MBNL Yes vs. MBNL No) of the gene with the highest fold change estimate



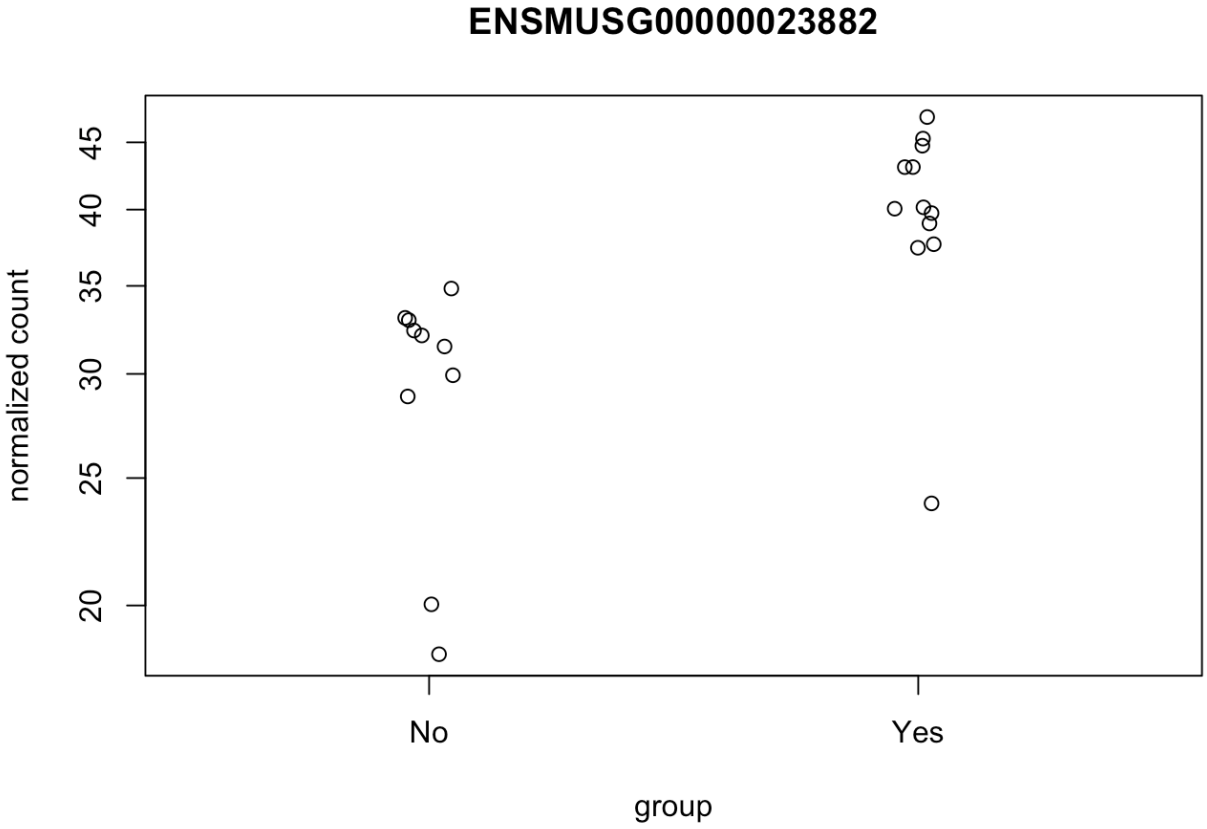
Produce an "MA plot" to show the distribution of log2 fold changes and expression levels among genes for the MBNL test.

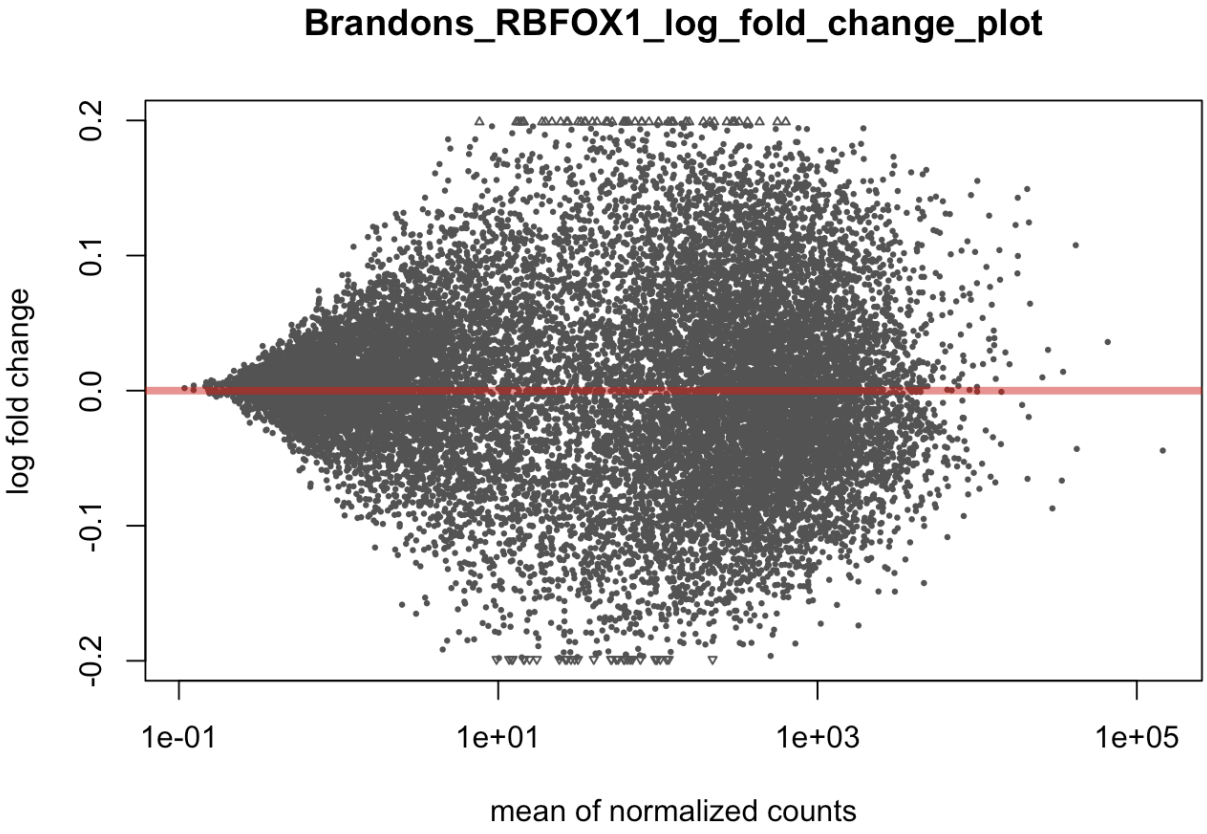


Repeat all steps from above but for the other categorical variable; RBFOX1

```
##
## out of 18714 with nonzero total read count
## adjusted p-value < 0.1
## LFC > 0 (up)     : 0, 0%
## LFC < 0 (down)   : 0, 0%
## outliers [1]     : 0, 0%
## low counts [2]   : 0, 0%
## (mean count < 0)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results
```

```
## log2 fold change (MAP): RBFOX1_induced Yes vs No
## Wald test p-value: RBFOX1_induced Yes vs No
## DataFrame with 18714 rows and 6 columns
##                      baseMean log2FoldChange      lfcSE          stat
##                     <numeric>      <numeric>  <numeric>     <numeric>
## ENSMUSG00000032733   304.05617      0.2226263 0.06263739      3.554208
## ENSMUSG00000028420    96.26967     -0.2351700 0.06770165     -3.473623
## ENSMUSG00000023882    34.73895      0.2883685 0.08366685      3.446627
## ENSMUSG00000038080   220.23374     -0.2527015 0.07382731     -3.422873
## ENSMUSG00000063698    32.96440      0.2872201 0.08764575      3.277056
## ...                        ...            ...        ...           ...
## ENSMUSG00000104378   1.2839129   1.760469e-06 0.04326860  4.068700e-05
## ENSMUSG00000081071   0.6265668   1.084210e-06 0.03041769  3.564405e-05
## ENSMUSG00000047370   0.9257389  -1.254566e-06 0.03593443 -3.491264e-05
## ENSMUSG00000090500   0.7901651  -2.416991e-07 0.02932847 -8.241109e-06
## ENSMUSG00000039740 235.7991658  -1.067411e-07 0.05489723 -1.944380e-06
##                         pvalue      padj
##                      <numeric> <numeric>
## ENSMUSG00000032733 0.0003791191 0.9007979
## ENSMUSG00000028420 0.0005134813 0.9007979
## ENSMUSG00000023882 0.0005676307 0.9007979
## ENSMUSG00000038080 0.0006196309 0.9007979
## ENSMUSG00000063698 0.0010489544 0.9007979
## ...                         ...       ...
## ENSMUSG00000104378    0.9999675 0.9999984
## ENSMUSG00000081071    0.9999716 0.9999984
## ENSMUSG00000047370    0.9999721 0.9999984
## ENSMUSG00000090500    0.9999934 0.9999984
## ENSMUSG00000039740    0.9999984 0.9999984
```


ENSMUSG00000023882
normalized count
20
25
30
35
40
45
No
Yes
group


Brandons_RBFOX1_log_fold_change_plot
log fold change
-0.2
-0.1
0.0
0.1
0.2
1e-01
1e+01
1e+03
1e+05
mean of normalized counts

Part 4 (Interaction effects tests using contrasts)

```
## 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
```

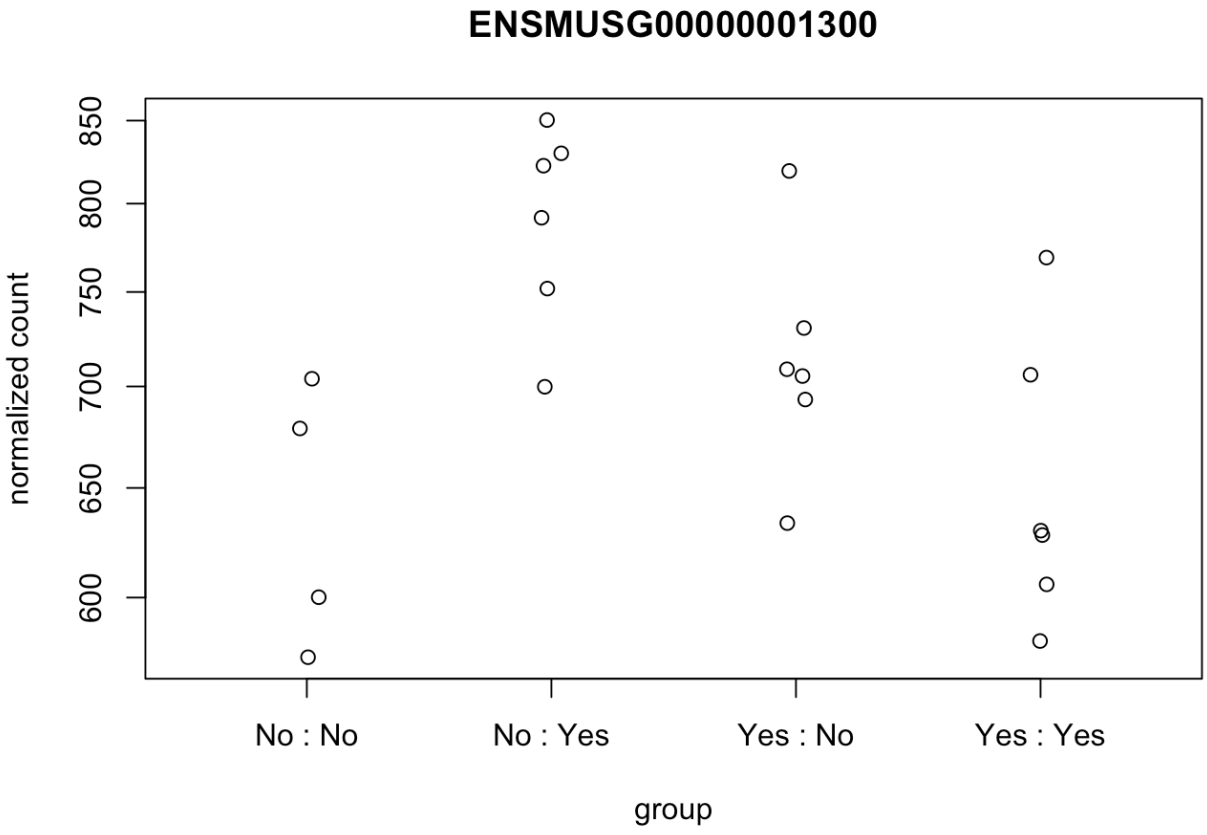
```
## [1] "Intercept"                          "MBNL_induced_Yes_vs_No"
## [3] "RBFOX1_induced_Yes_vs_No"           "MBNL_inducedYes.RBFOX1_inducedYes"
```

```
##
## out of 18714 with nonzero total read count
## adjusted p-value < 0.1
## LFC > 0 (up)     : 0, 0%
## LFC < 0 (down)   : 0, 0%
## outliers [1]     : 0, 0%
## low counts [2]   : 0, 0%
## (mean count < 0)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results
```

```
##
## out of 18714 with nonzero total read count
## adjusted p-value < 0.1
## LFC > 0 (up)     : 824, 4.4%
## LFC < 0 (down)   : 341, 1.8%
## outliers [1]     : 0, 0%
## low counts [2]   : 6531, 35%
## (mean count < 7)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results
```

```
##
## out of 18714 with nonzero total read count
## adjusted p-value < 0.1
## LFC > 0 (up)     : 0, 0%
## LFC < 0 (down)   : 0, 0%
## outliers [1]     : 0, 0%
## low counts [2]   : 0, 0%
## (mean count < 0)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results
```

```
##
## out of 18714 with nonzero total read count
## adjusted p-value < 0.1
## LFC > 0 (up)     : 0, 0%
## LFC < 0 (down)   : 0, 0%
## outliers [1]     : 0, 0%
## low counts [2]   : 0, 0%
## (mean count < 0)
## [1] see 'cooksCutoff' argument of ?results
## [2] see 'independentFiltering' argument of ?results
```



The statistical power to detect interactions compared to the power to detect main effects shows a

clear differentiation with regards to output. Power of statistical tests increases when treatments and samples are reduced. Main effects model will have a higher power due to less variations within samples and treatment groups. As mentioned from the Logan *Biostatistical Text Book*: .."tests of interactions are typically more powerful than main effects (due to greater available degrees of freedom) and for fixed models, efforts to improve the power of any of the main effects will also benefit the corresponding interactions. Power analyses for mixed and random factorial designs should reflect the appropriate residuals".

Part 5 (Individual Candidate Genes)

```
## using pre-existing size factors
```

```
## estimating dispersions
```

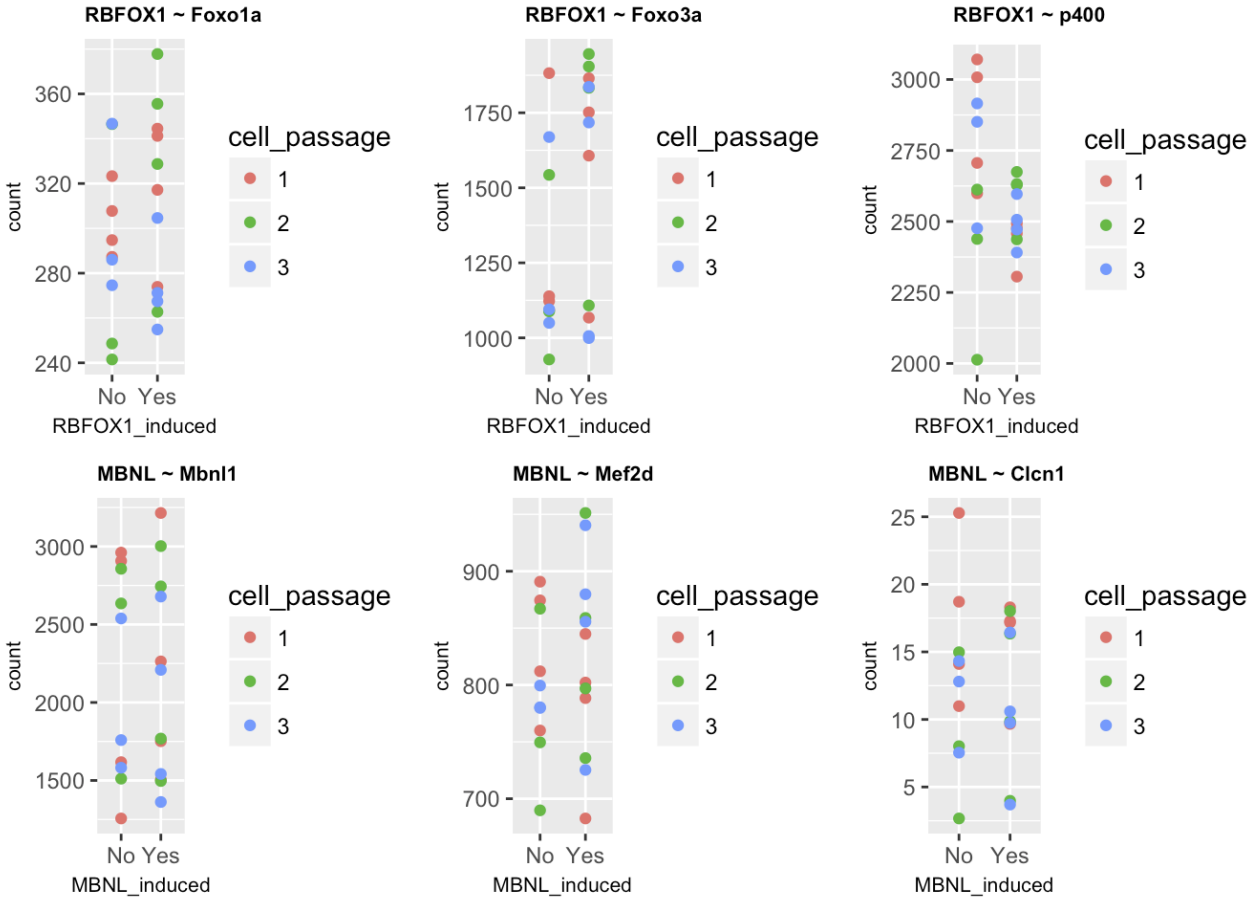
```
## found already estimated dispersions, replacing these
```

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

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

```
## final dispersion estimates
```

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


RBFOX1 ~ Foxo1a
count
360
320
280
240
No Yes
RBFOX1_induced
cell_passage
1
2
3
RBFOX1 ~ Foxo3a
count
1750
1500
1250
1000
No Yes
RBFOX1_induced
cell_passage
1
2
3
RBFOX1 ~ p400
count
3000
2750
2500
2250
2000
No Yes
RBFOX1_induced
cell_passage
1
2
3
MBNL ~ Mbnl1
count
3000
2500
2000
1500
No Yes
MBNL_induced
cell_passage
1
2
3
MBNL ~ Mef2d
count
900
800
700
No Yes
MBNL_induced
cell_passage
1
2
3
MBNL ~ Clcn1
count
25
20
15
10
5
No Yes
MBNL_induced
cell_passage
1
2
3