RNA-seq Lab Answer Key

1) When we multiply all counts by 10, the p-value for ENSG00000160202 becomes 1.18e-4.
Example code:

```
countMat2 <- countMat*10
y2   <- edgeR::DGEList(counts=countMat2, group=group)
y2   <- edgeR::calcNormFactors(y2        )
y2   <- edgeR::estimateGLMTrendedDisp(y2, design)
y2   <- edgeR::estimateGLMTagwiseDisp(y2, design)
fit2 <- edgeR::glmFit(y2, design)
lrt2 <- edgeR::glmLRT(fit2, coef=2)
topTags(lrt2)
```

2) When we just multiply all counts in "mtBC06" by 100, the raw p-value for ENSG00000160202 becomes 1.19e-4 and the new most-significant P-value is for the gene ENSG00000183486.
Note: The per-gene dispersion estimate makes use of the more general estimate, so you need to run common dispersion OR tagwise dispersion first.
Example code:

```
countMat3     <- countMat
countMat3[,2] <- countMat3[,2] * 100
summary(countMat3)
y3   <- edgeR::DGEList(counts=countMat3, group=group)
y3   <- edgeR::calcNormFactors(y3        )
y3   <- edgeR::estimateGLMTrendedDisp(y3, design)
y3   <- edgeR::estimateGLMTagwiseDisp(y3, design)
fit3 <- edgeR::glmFit(y3, design)
lrt3 <- edgeR::glmLRT(fit3, coef=2)
topTags(lrt3)
```

3) The maximum number of (valid) permutations for this data (the "B" parameter in MTP) is 20. This is the value 6 choose 3. The reason that permutation works poorly on this data is that we have too few replicates to estimate p-values with enough precision.