

Selecting DE Genes Continued

...

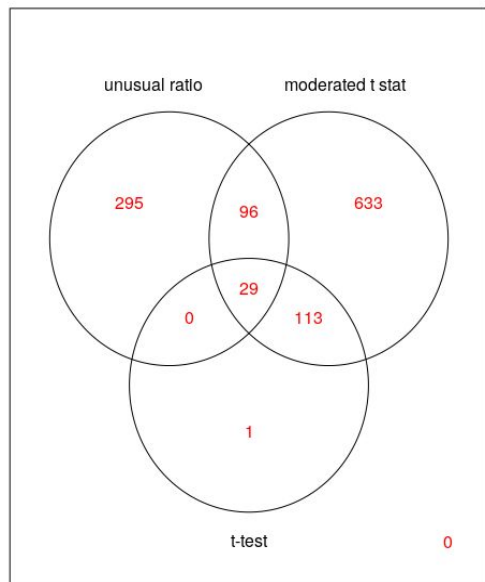
by Jake Sauter

Covered This Week

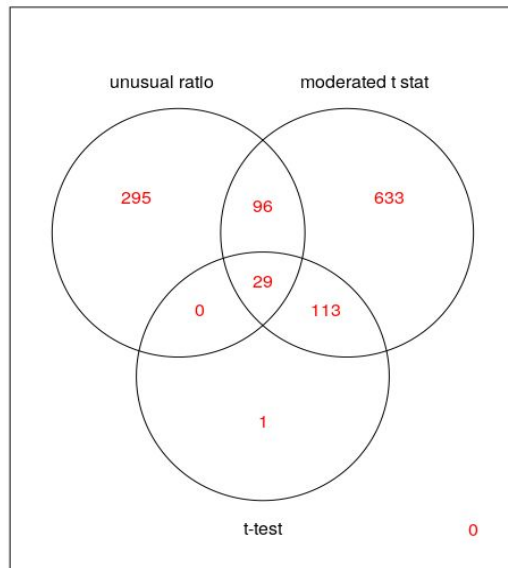
- Implemented Holms and FDR Family Wise Error Rate corrections for t-test
- Use of SAM in R
- Implemented permutation testing for t-test

T-test Results

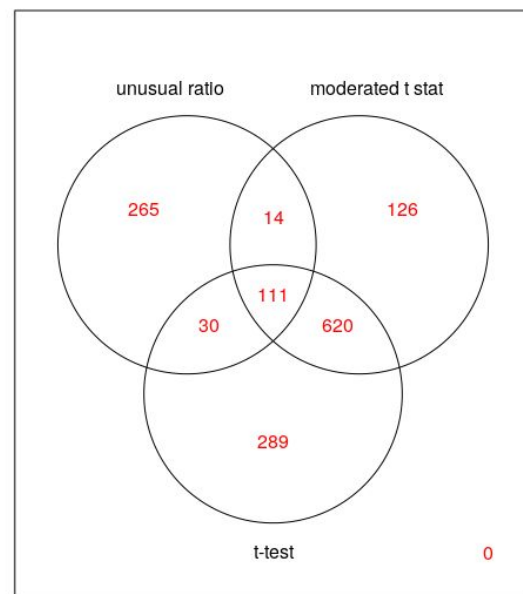
Bonferroni



Holms



FDR

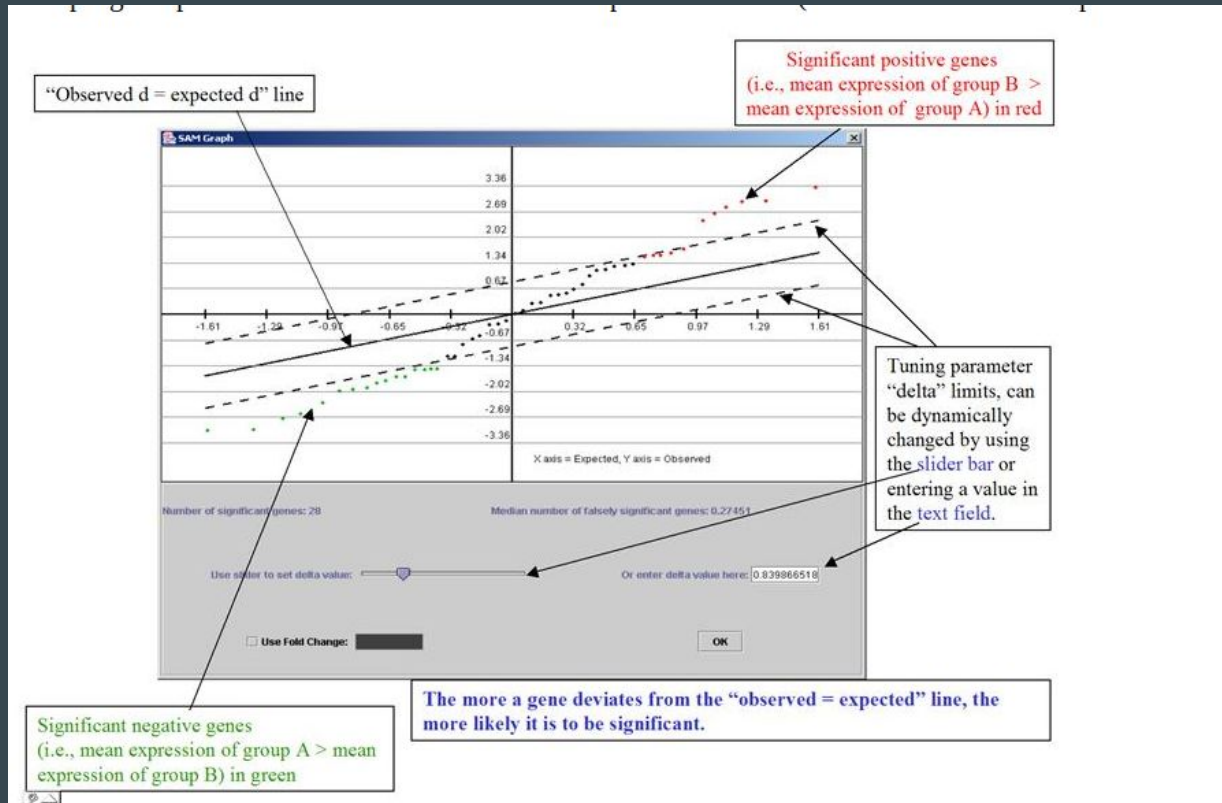


SAM Results

- The way to control for FDR in SAM is through the **delta** parameter
- SAM results allow for a table to be printed that shows how many genes are selected, and various FDR statistics associated with a particular delta value

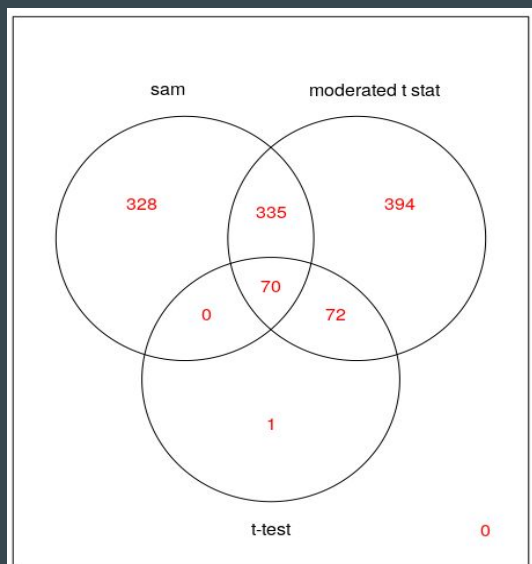
```
> delta.table[, c(1,4,5,6)]
      delta # called median FDR 90th perc FDR
[1,] 0.000000000    2127 0.504348388 0.527150150
[2,] 0.001772127    2126 0.503690508 0.527142359
[3,] 0.007088508    2118 0.502512485 0.526771709
[4,] 0.015949144    2110 0.499779422 0.525728268
[5,] 0.028354033    2091 0.496259888 0.524004677
[6,] 0.044303177    2061 0.491611968 0.522425104
[7,] 0.063796575    2048 0.484511353 0.520192628
[8,] 0.086834227    2024 0.478973931 0.516179702
[9,] 0.113416133    1994 0.472546391 0.512766006
[10,] 0.143542293    1972 0.463618762 0.508257404
[11,] 0.177212707    1945 0.450486464 0.501000211
[12,] 0.214427376    1861 0.428748663 0.492703063
[13,] 0.255186298    1774 0.401349687 0.478891857
[14,] 0.299489475    1693 0.375429673 0.466830259
[15,] 0.347336906    1595 0.345829794 0.455152233
[16,] 0.398728591    1485 0.305725077 0.435594161
[17,] 0.453664530    1372 0.269479568 0.417217779
[18,] 0.512144724    1248 0.231122297 0.399203319
[19,] 0.574169171    1179 0.199223692 0.370823543
[20,] 0.639737873    1058 0.158026708 0.349200482
[21,] 0.708850829    1004 0.129700857 0.313664871
[22,] 0.781508039     937 0.106189736 0.281083651
[23,] 0.857709503     850 0.082516695 0.248509582
[24,] 0.937455221     780 0.060993632 0.206541863
[25,] 1.020745193     719 0.044994468 0.177104274
[26,] 1.107579420     656 0.035225416 0.151013431
[27,] 1.197957901     586 0.023659938 0.120712077
[28,] 1.291880635     545 0.016461086 0.086794817
[29,] 1.389347624     488 0.010027524 0.064844658
[30,] 1.490358867     438 0.004965431 0.047668134
[31,] 1.594914365     385 0.004236738 0.026973896
```

SAM Results

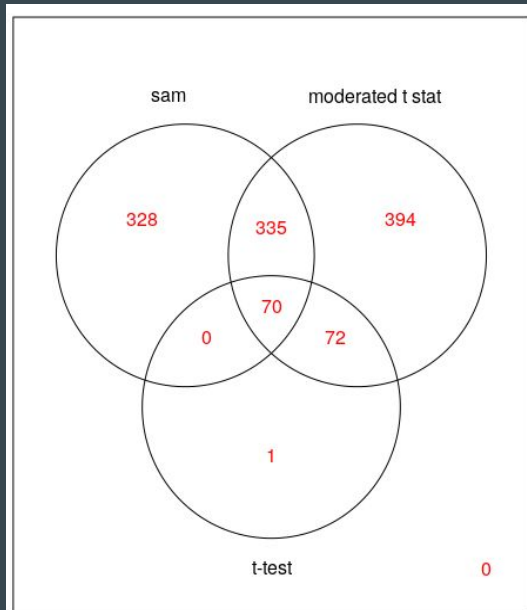


SAM Results

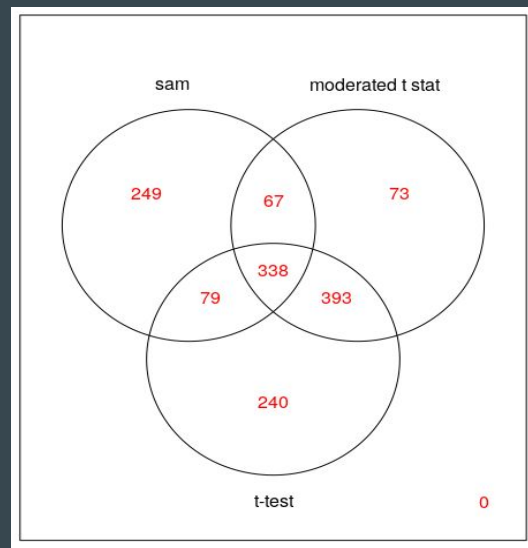
Bonferroni



Holms



FDR

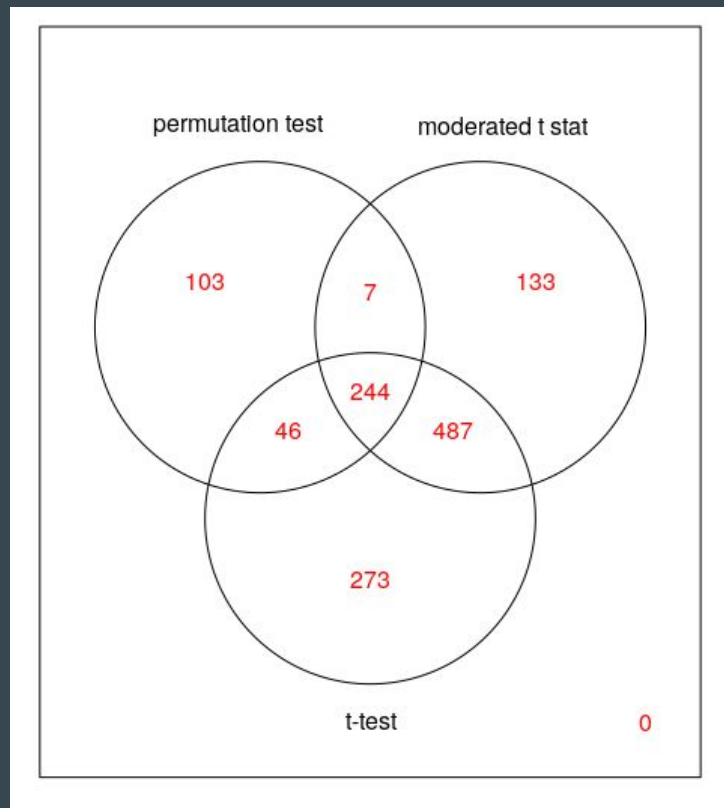


Permutation Testing

- For permutation testing, our goal is to determine the distribution of our test statistic by computing the test statistic for all or some amount of permutations of the data
 - With the distribution of our test-statistic, we can produce a p-value with our actual test statistic, counting how many statistics in our distribution are greater than or equal to our actual test statistic
- If there are too many possible permutations to account for, we can choose permutations randomly (Monte Carlo Sampling) [2]
 - The more random permutations we perform, the closer our analysis become to an exact test

Permutation Testing

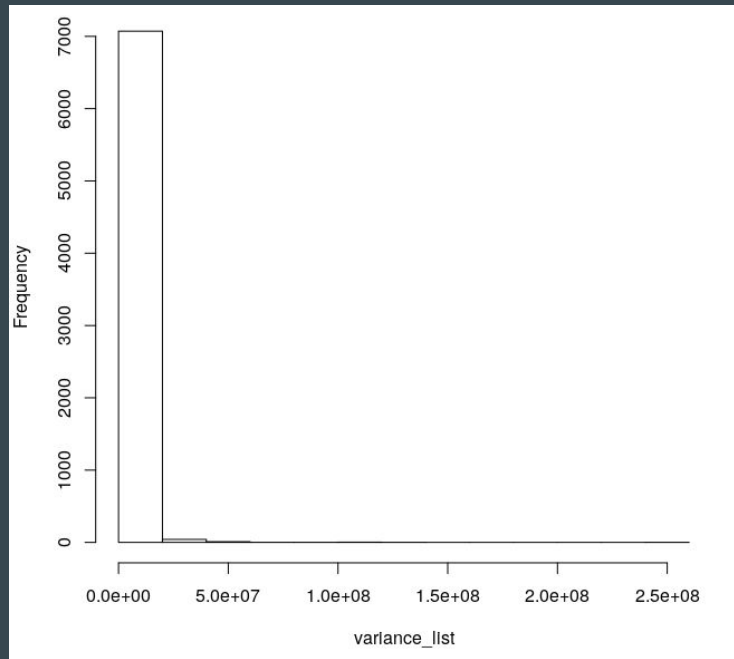
- 10,000 tests were performed for each gene
- Only 1 value was below the .05 threshold
- The 10 lowest p-values were 0.0324, 0.0694, 0.0741, 0.0775, 0.1013, 0.1033, 0.1103, 0.1186, 0.1205, 0.1284



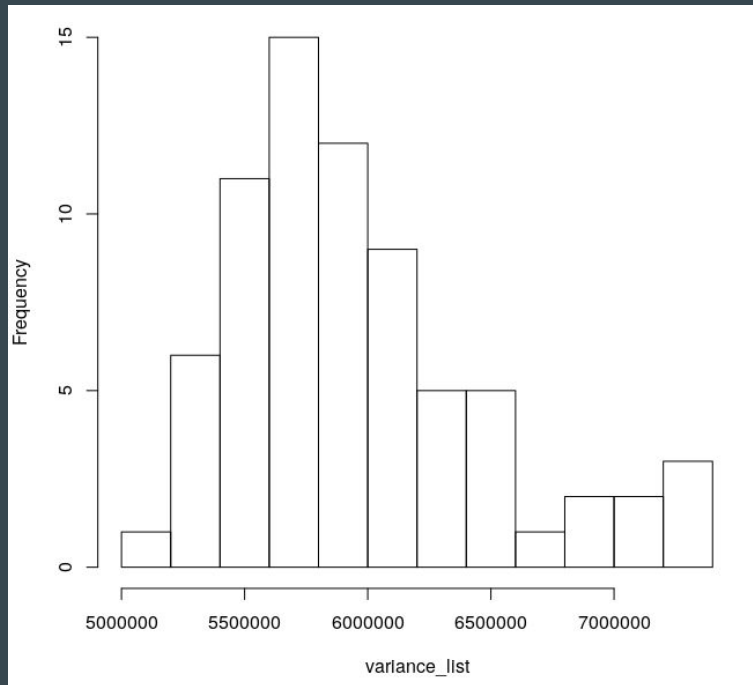
Selecting the 400 lowest p-values of permutation testing

Variance Observations

Genes

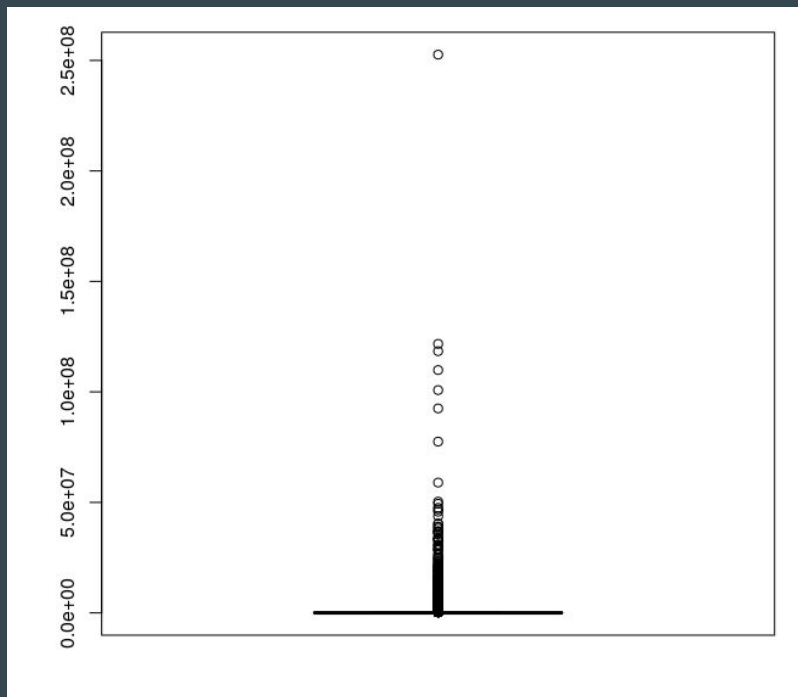


People

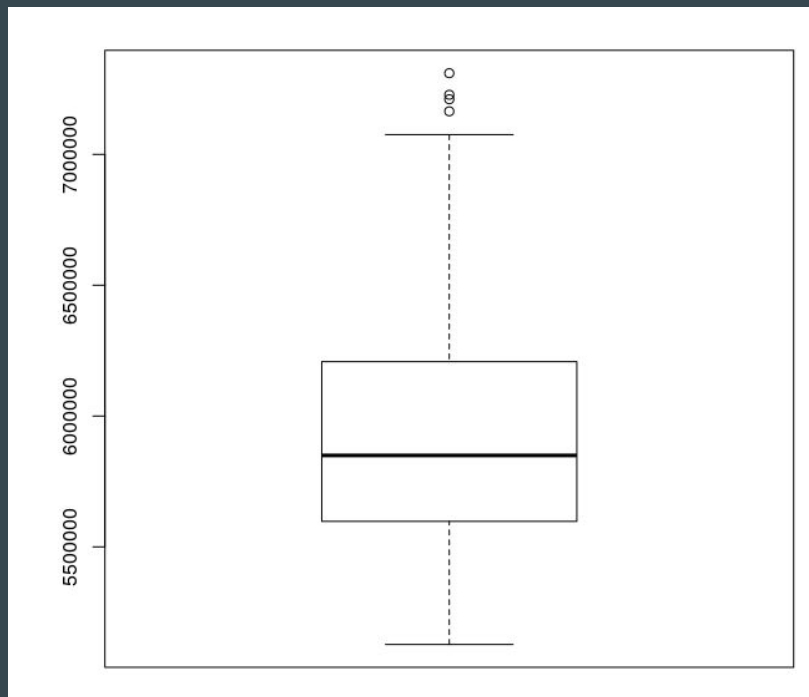


Variance Observations

Genes



People



References

- [1] Cui, Yan. “Data File: Stanford_Large.Txt.” Microarray Data Analysis II, compbio.uthsc.edu/microarray/lecture2.htm.
- [2] Jianqiang, MA. “Permutation Test & Monte Carlo Sampling.” Permutation Test & Monte Carlo Sampling, 18 Mar. 2009, www.let.rug.nl/nerbonne/teach/rema-stats-meth-seminar/presentations/Permutation-Monte-Carlo-Jianqiang-2009.pdf.
- [3] Drăghici Sorin. Statistics and Data Analysis for Microarrays: Using R and Bioconductor. Chapman and Hall, 2012.