



Summer Institute
in Statistical Genetics 2018

Integrative Genomics

3a. Normalization Tutorial



ggibson.gt@gmail.com
<http://www.cig.gatech.edu>

Georgia
Tech



Center for
Integrative Genomics

Steps in the R Normalization Tutorial

1. Overview of the "Eight Adults" RNASeq Samples
2. edgeR evaluation of DE genes by Weight class
3. qValue FDR computation
4. Saving and Visualizing normalized logcpm
5. Principal Variance Components Analysis (PVCA)
6. Surrogate Variable Analysis (SVA)
7. Supervised Normalization of Microarrays (SNM)

Overview of the “Eight Adults” RNASeq Samples

The dataset consists of 32 RNASeq samples each with 1M to 3M reads

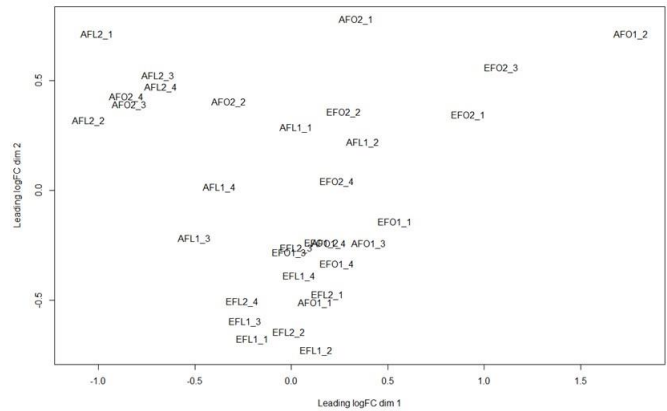
- There are:
- 2 Obese African American Women (AFO1 and AFO2)
 - 2 Normal Weight African American Women (AFL1 and AFL2)
 - 2 Obese European American Women (EFO1 and EFO2)
 - 2 Normal Weight European American Women (EFL1 and EFL2)

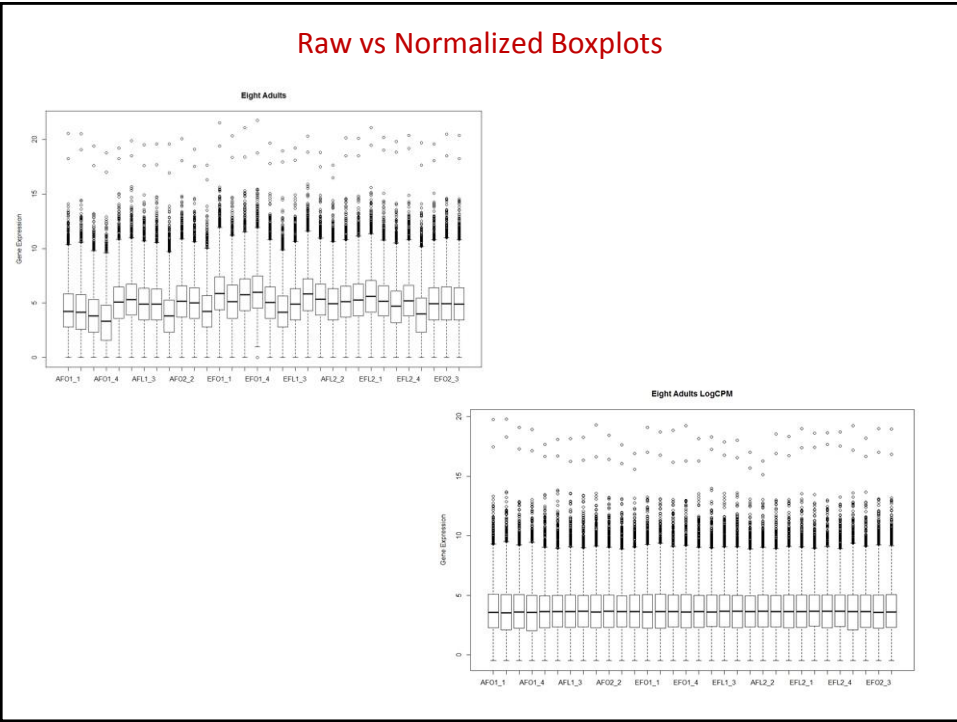
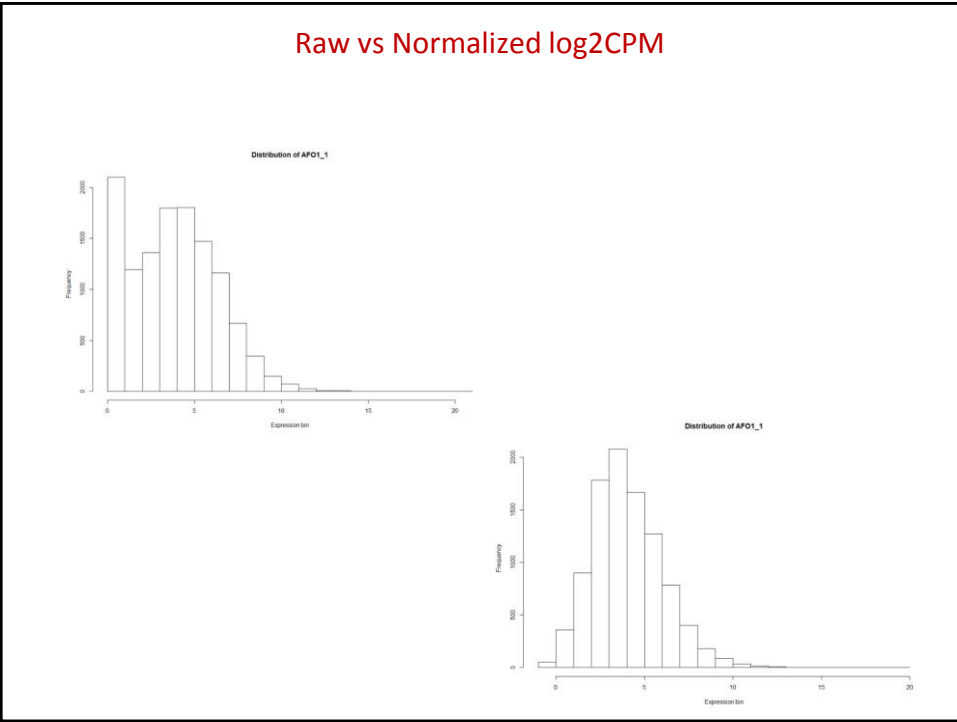
Each Woman was a healthy adult volunteer sampled at 4 ~12-month intervals

```

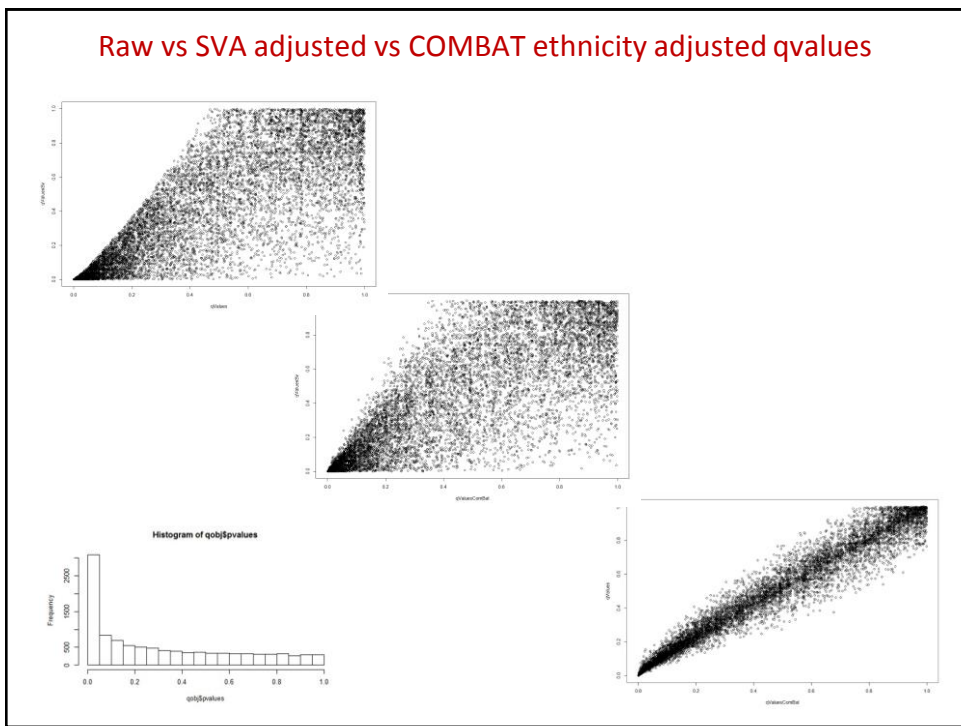
group      lib.size      norm.factors
AFR.LOW:8  Min.    :1020150  Min.    :1
AFR.OB :8   1st Qu.:2103568  1st Qu.:1
EUR.LOW:8  Median :2627652  Median :1
EUR.OB :8   Mean    :2866461  Mean    :1
          3rd Qu.:3171455  3rd Qu.:1
          Max.    :6866271  Max.    :1
> y <- calcNormFactors(y)
> head(y$samples)
      group lib.size norm.factors
AFO1_1 AFR.OB  2778253  0.6197938
AFO1_2 AFR.OB  3068801  0.5472522
AFO1_3 AFR.OB  1538792  0.8050460
AFO1_4 AFR.OB  1020150  0.8779080
AFL1_1 AFR.LOW 2355800  1.2489012
AFL1_2 AFR.LOW 3080954  1.1263825
    
```

Multi-Dimensional Scaling of the Dataset

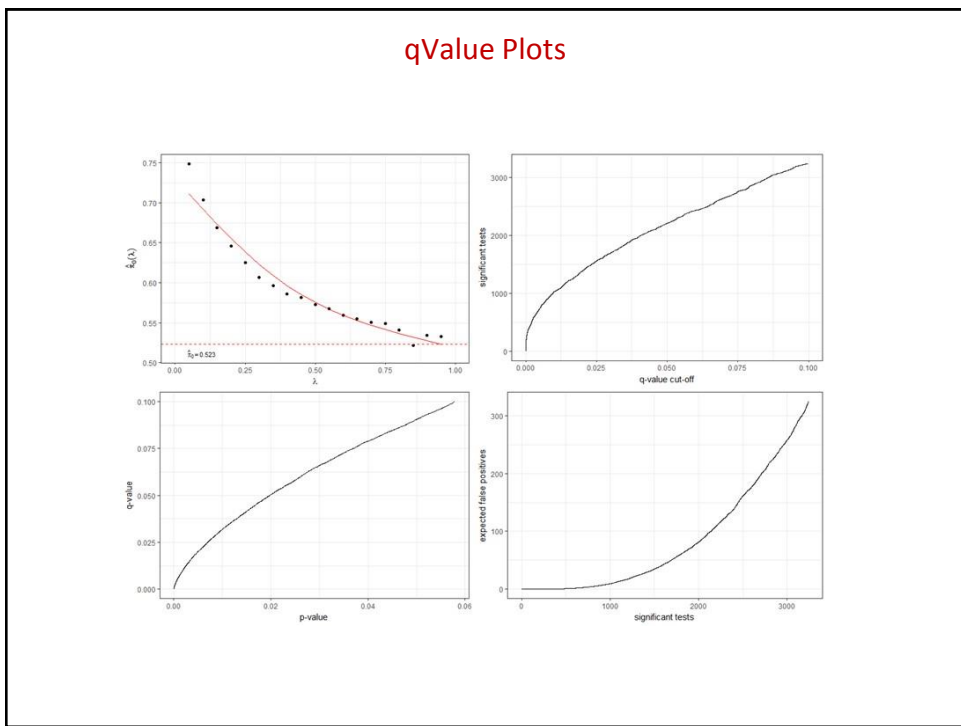


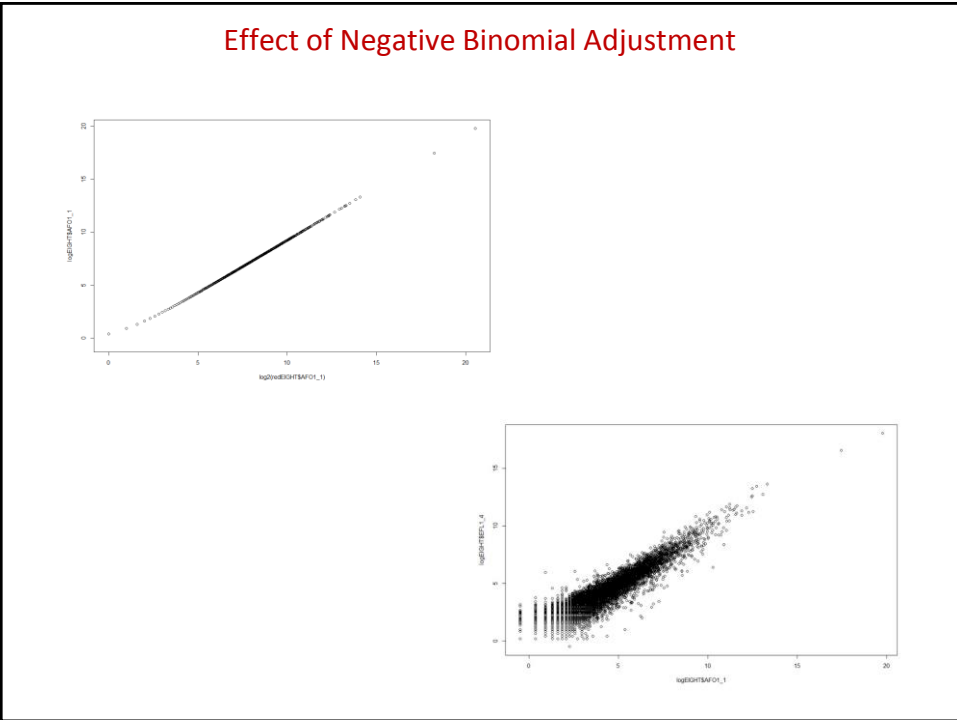
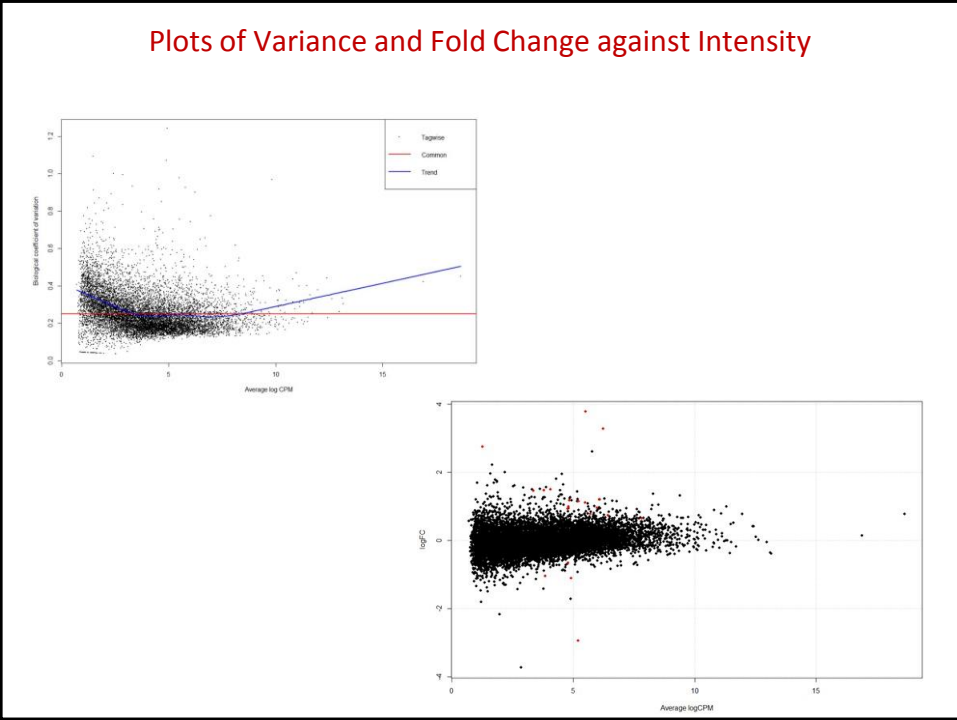


Raw vs SVA adjusted vs COMBAT ethnicity adjusted qvalues

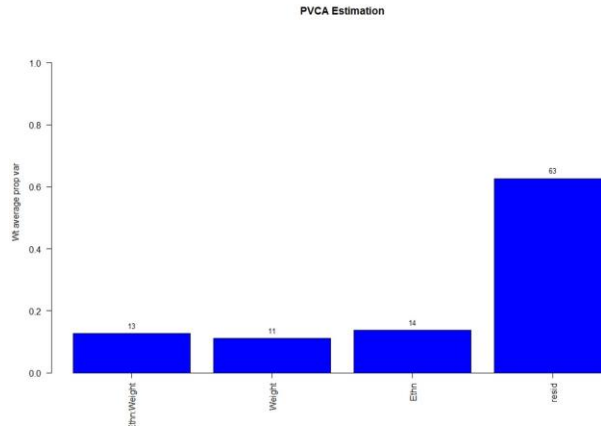


qValue Plots





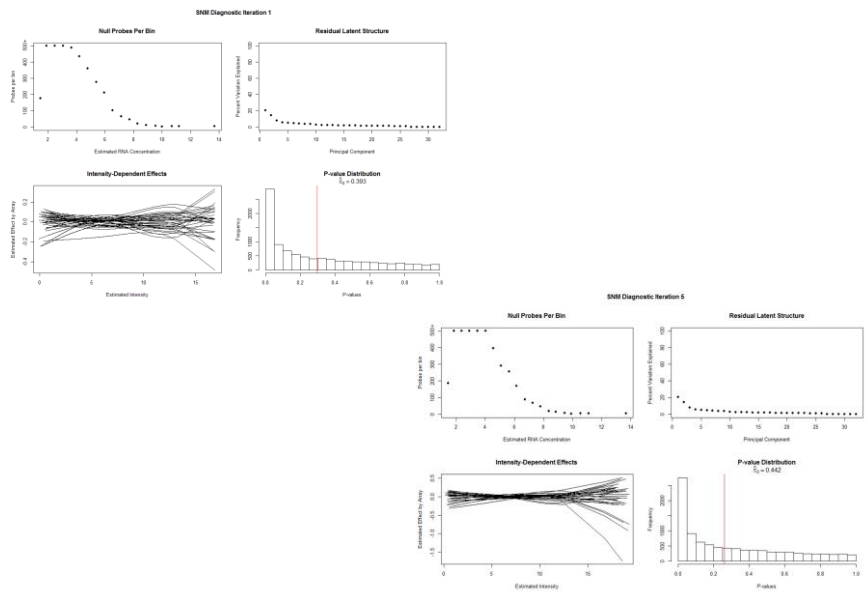
Weighted Principal Variance Components Analysis



ANOVA of Surrogate Variables

```
> summary(fit2)
      Df Sum Sq Mean Sq F value    Pr(>F)
ethn   1  0.0070  0.00699    0.708   0.409
weight 1  0.0015  0.00153    0.155   0.698
visit  3  0.0175  0.00585    0.593   0.626
person 5  0.7668  0.15335   15.545 1.92e-06 ***
Residuals 21  0.2072  0.00987
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> fit3 <- aov(SV3 ~ ethn + weight + visit + person, data=phEIGHT)
> summary(fit3)
      Df Sum Sq Mean Sq F value    Pr(>F)
ethn   1  0.5516  0.5516   74.451 2.4e-08 ***
weight 1  0.0005  0.0005    0.061 0.806947
visit  3  0.0188  0.0063    0.846 0.483938
person 5  0.2735  0.0547    7.383 0.000392 ***
Residuals 21  0.1556  0.0074
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

SNM Iterations 1 and 5



Comparison of SNM/SAS vs edgeR for Weight

