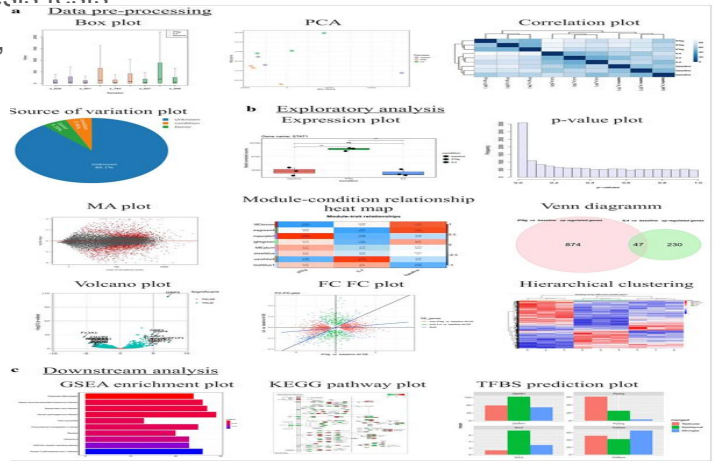
DSeq, pca and volcano plot

Abstract for Dseq

1. DESeq method detects and corrects dispersion estimates that are too low through modeling of the dependence of the dispersion on the average expression strength over all samples the comparative analysis of transcriptomics data
2. logarithmic fold change (LFC) between treatment and control for a gene's expression is exactly zero, p -value
3. the number of genes called significantly differentially expressed depends as much on the sample size
4. DESeq2 integrates methodological advances with several novel features to facilitate a more quantitative analysis of comparative RNA-seq data using shrinkage estimators for dispersion and fold change.

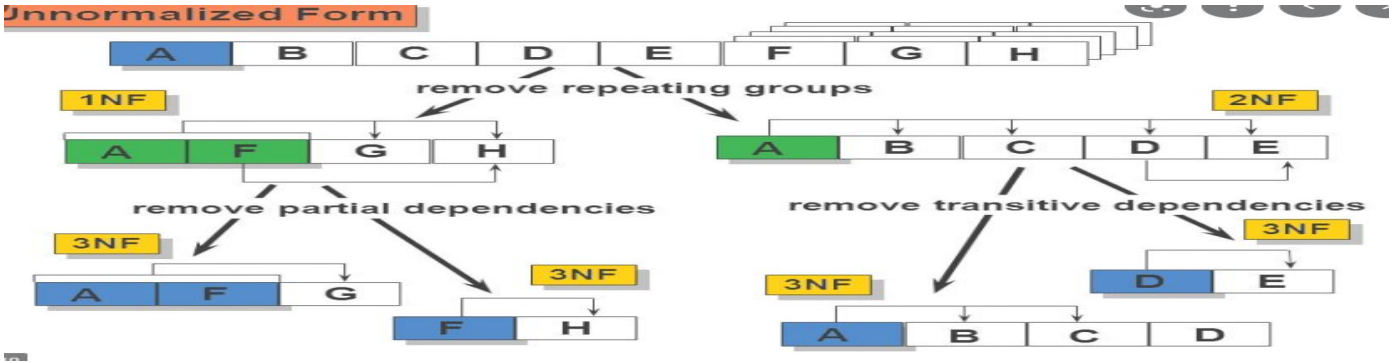
Abstract Cont

1. DESeq2’s new features by describing a number of applications possible with shrunken fold changes and their estimates of standard error
2. improved gene ranking and visualization, hypothesis tests
3. DESeq2’s statistical power with existing tools, y and precision



Matrix operation

1. matrix Kij (itself are negative binomial)
2. indicate the number of sequencing reads that have been unambiguously mapped to a gene in a sample
3. Normalization using μij = sijqij



the amount of shrinkage

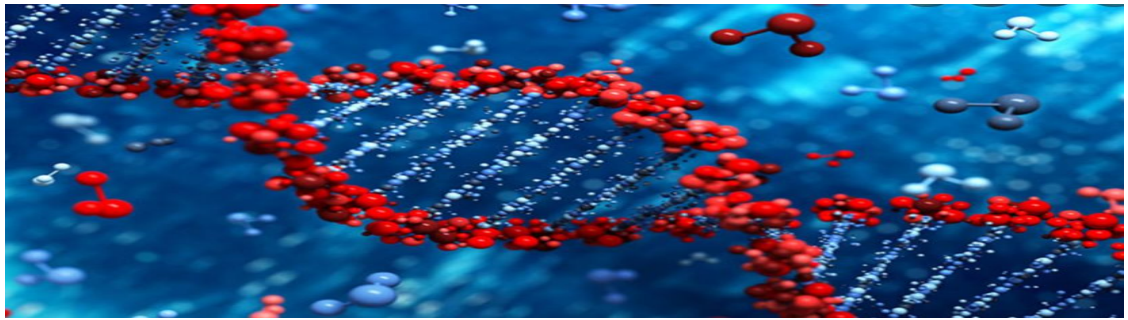
1. DESeq2 estimates the width of the prior distribution from the data and therefore automatically controls the amount of shrinkage based on the observed properties of the data.
2. DESeq2 handles these cases by using the gene-wise estimate instead of the shrunken estimate when the former is more than 2 residual standard deviations above the curve.
3. the reason for extraordinarily high dispersion of a gene is that it does not obey our modeling assumptions
4. some genes may show much higher variability than others for biological or technical reasons

Method

1. Empirical Bayes shrinkage for fold-change estimation
2. overcomes this issue by shrinking LFC estimates toward zero in a manner such that shrinkage is stronger when the available information for a gene is low
3. DESeq2 reports the standard error for each shrunken LFC estimate, obtained from the curvature of the coefficient’s posterior
4. DESeq2 uses a Wald test: the shrunken estimate of LFC is divided by its standard error, resulting in a z-statistic, which is compared to a standard normal distribution also the Max likelihood ratio

Method

1. average expression strength of each gene, across all samples, as its filter criterion, and it omits all genes with mean normalized counts below a filtering threshold from multiple testing adjustment
2. choose a threshold that maximizes the number of genes found at a user-specified target FDR
3. Hypothesis tests with thresholds on effect size Specifying minimum effect size

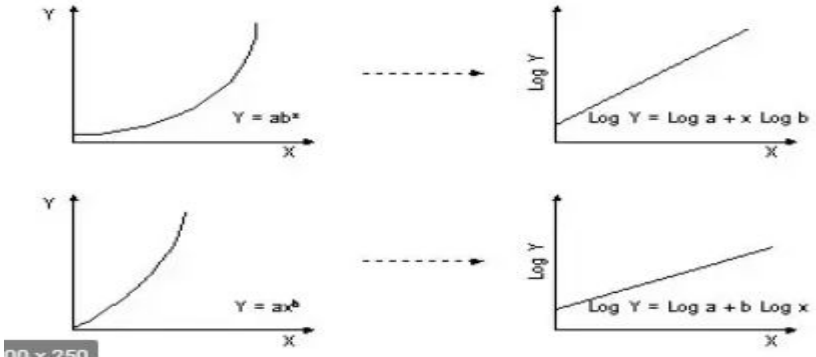


P-values, outliers, and hypothesis

1. for composite null hypotheses of the form $|\beta ir| \leq \theta$, where βir is the shrunken LFC from the estimation procedure
2. Two method for flag outlier
    a. <=6 replicates cause the whole gene to be flagged and removed from subsequent analysis, including P value adjustment for multiple testing
    b. >7 replaces the outlier counts with an imputed value, namely the trimmed mean over all samples, scaled by the size factor, and then re-estimates the dispersion, LFCs and P values for these genes.

Transformation and Gene-level analysis and precision

1. Regularized logarithm transformation(a log2 transformation for genes with high counts, while shrinking together the values for different samples for genes with low counts)
2. Gene-level analysis DESeq2 for the analysis of per-gene counts, i.e., the total number of reads that can be uniquely assigned to a gene DESeq2 framework as gene- and sample-specific normalization factors. In addition, the approach used in DESeq2 can be extended to isoformspecific analysis



Sensitivity compare to other tool and precision

1. DESeq2 had comparable sensitivity to edgeR and voom though less than DSS.
2. Precision of fold change estimates DESeq2 had consistently low rootmean-square error and mean absolute error across a range of sample sizes and models for a distribution of true LFCs.

Conclusion

- Deseq2 is the best for finding the RNA sequencing and calculation
- Also a efficient way to do the further result