# Deep Transcriptome Profiling - A comparison of gene expression analysis programs for RNAseq

***Willian A. da Silveira[1], E. Starr Hazard[1,2], Dongjun Chung[3] ,Gary Hardiman[1,3]***

*[1]MUSC Bioinformatics, Center for Genomics Medicine, [2]Library Science and Informatics, [3]Department of Public Health Sciences, Medical University of South Carolina (MUSC), Charleston SC*

## Background

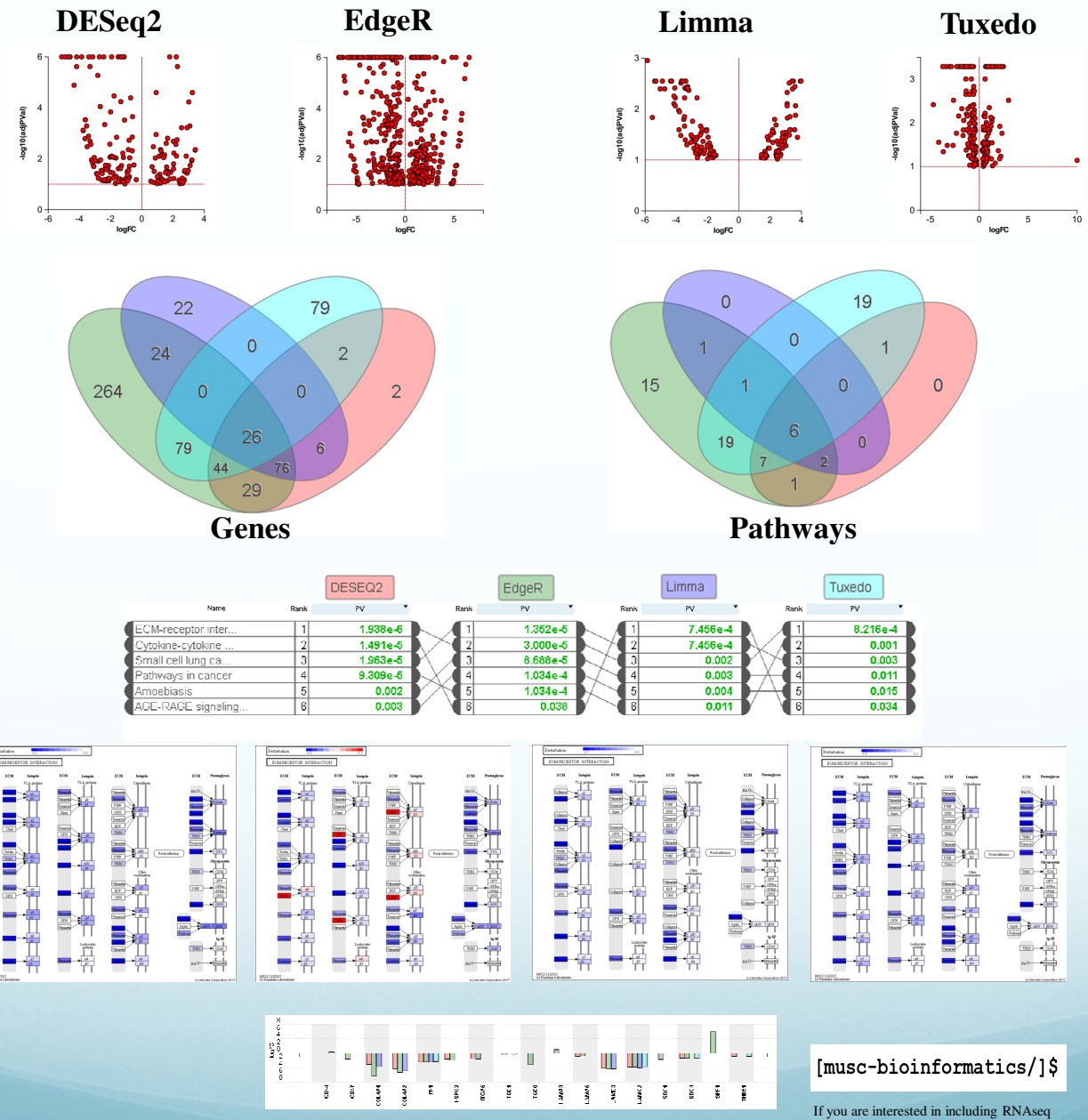
- Different methods have evolved over the past five years to significantly determine differentially expressed genes from a comparison of the transcriptome under different physiological conditions (e.g. healthy vs diseased tissue).
- Selecting the best method is guided by the experimental design.

## Methodology

- Next-generation RNA sequencing experiments were performed in HCC827 human lung adenocarcinoma cells expressing, ZEB1(a well known Epithelial-to-Mesenchymal Transition stimulator) or empty vector and extracted from GEO (accession number GSE81167).
- The FASTQ files were downloaded and the quality of the FASTQ files were assessed using the FAST**QC** and preprocessing using Cutadapt and Trimmomatic.
- These processes increase the quality and reliability of the analyses and diminish the computational cost and execution time. We used the reference GRCh37/hg19 human genome and the Tophat aligner. We then performed differential expression (DE) analysis using the TUXEDO pipeline, and after defining mRNA level counts with HTSeq, the DESeq2 , edgeR and Limma/voom programs. q≤0.1
- Systems level analysis was performed using Advaita Bio's iPathwayGuide.

## Figure

- **1st row** – Volcano Plots – Gene Expression Data.
- **2nd row** – Venn Diagrams
- **3rd row** – Pathways in common across the 4 analytical approaches.
- **4th row** – Perturbation in the ECM Receptor pathway in the 4 analyses
- **5th row** – Gene Expression of the ECM Receptor pathway in the 4 analyses




## Results

- Using a Bonferroni-Hochberg adjusted *p*-value ≤ 0.1, TUXEDO uncovered 476 DE genes, DESeq2 found 186, edgeR 540 and Limma/voom 155.
- Considering only the lowest 10 adj *p*-values, no mRNAs were shared across all methods.
- For the complete list, DESeq2 provided the most overlap, i.e. 43.0% with TUXEDO, 58.0% with Limma/voom and 94.6% with edgeR.
- edgeR yielded the greatest list of DE genes, Tuxedo has a 59.6% of overlap with edgeR, DESeq2 has 94.6%, and Limma 81.2%, but only 48.0% of edgeR DE genes are unique compared to only 0.01% with DESeq2.
- Only 19.4% of the mRNAs from Limma/voom overlap the TUXEDO results, the lowest agreement.
- Even with this gene level discordance, the systems level analysis identified the same pattern of affected pathways.

## Conclusion

- Between this methods, gene-level analysis of significance show a variability, but system level analysis revealed consistency.

## Bibliography

1. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
2. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr 20;43(7):e47.
3. Robinson MD, McCarthy DJ, Smyth GK.: edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–140.
4. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc. 2012 Mar 1;7(3):562-78.
5. Zhang T, Guo L, Creighton CJ, Lu Q, Gibbons DL, Yi ES, Deng B, Molina JR, Sun Z, Yang P, Yang Y. A genetic cell context-dependent role for ZEB1 in lung cancer. Nat Commun. 2016 Jul 26;7:12231.
6. Draghici S, Khatri P, Tarca AL, Amin K, Done A, Voichita C, Georgescu C, Romero R. A systems biology approach for pathway level analysis. Genome Res. 2007 Oct;17(10):1537-45.

[musc-bioinformatics/]$

If you are interested in including RNAseq analysis in your work, please, contact us:
hardiman@musc.edu