

Differential Gene Expression: Understanding Gene Activity Across Conditions

Idowa Tiwao^{1*} | Oluymi Akinlo¹

¹Center for Genomics of Noncommunicable Diseases and Personalized Health Care, University of Lagos, Akoka, Lagos, Nigeria

***Correspondence:** IdowaTiwao, Center for Genomics of Noncommunicable Diseases and Personalized Health Care, University of Lagos, Akoka, Lagos, Nigeria, E-mail: idowa@tiwao.edu.ng

Citation: Tiwao I, Oluymi A (2024) Differential Gene Expression: Understanding Gene Activity Across Conditions. *Int. J. Health Sci. Biomed.* 1: 1-3. DOI: 10.5678/IJHSB.2024.417

Received Date: 2024-07-02, **Accepted Date:** 2024-07-22, **Published Date:** 2024-07-30

Keywords: Differential gene expression; RNA-seq; Transcriptomics; Gene regulation; DESeq2; edgeR; Biomarker discovery; Gene expression profiling

Abstract

Differential gene expression (DGE) analysis is a fundamental method in genomics that allows researchers to identify genes whose expression levels vary significantly across different biological conditions, such as diseased vs. healthy states, treated vs. untreated samples, or developmental stages. With the advent of high-throughput RNA sequencing (RNA-seq), DGE analysis has become more precise, scalable, and widely used across biomedical research. This article provides an overview of the principles, methodologies, applications, and challenges of DGE analysis, highlighting its critical role in understanding biological processes and disease mechanisms.

Introduction

Gene expression is the process by which information from a gene is used to synthesize functional gene products, primarily proteins or functional RNAs. The level of gene expression can vary depending on the cell type, developmental stage [1], or environmental condition. Differential gene expression (DGE) analysis seeks to identify these variations by comparing gene expression profiles between different sample groups.

For decades, researchers have studied gene expression using technologies like microarrays. However, the advent of RNA sequencing (RNA-seq) has revolutionized transcriptomic studies by allowing unbiased, high-resolution, and quantitative analysis of the transcriptome. RNA-seq data enables the detection of both known and novel transcripts and provides greater sensitivity for low-abundance genes.

DGE analysis has wide-ranging applications [2], including identifying disease biomarkers, understanding developmental biology, analyzing drug responses, and revealing gene regulatory mechanisms.

Overview of DGE Analysis

Differential gene expression analysis involves several key steps:

Sample Preparation and RNA Sequencing

RNA is extracted from biological samples and converted into cDNA libraries for sequencing using platforms such as Illumina [3].

Read Alignment or Quantification

Sequenced reads are either aligned to a reference genome/transcriptome using tools like STAR or pseudo-aligned using methods like Kallisto and Salmon.

Expression Quantification

Reads are counted for each gene to produce a raw count matrix, where rows represent genes and columns represent samples [4].

Normalization

Raw counts are normalized to account for sequencing depth and gene length, often using methods like TPM (Transcripts Per Million), FPKM, or statistical models in tools like DESeq2 and edgeR.

Statistical Testing

Statistical models are applied to identify genes that show significant changes in expression between groups. Multiple testing correction (e.g., Benjamini-Hochberg FDR) controls the false discovery rate.

Interpretation

Significantly differentially expressed genes are interpreted in the context of biological functions, often using pathway or gene ontology (GO) analysis.

Common Tools and Methods

Several bioinformatics tools are widely used for DGE analysis:

DESeq2: A robust R package that uses negative binomial distributions for count data and includes methods for normalization and variance estimation [5].

edgeR: Suitable for small sample sizes, edgeR also models count data using negative binomial distributions and offers powerful statistical tests.

limma-voom: Originally developed for microarray data, it can be used with RNA-seq after transforming count data using the voom method.

Sleuth: Works with pseudo-alignment tools like Kallisto and provides interactive DGE analysis.

Each tool has unique strengths and is chosen based on data characteristics, experimental design, and user familiarity.

Applications of DGE Analysis

Disease Research and Biomarker Discovery

DGE analysis helps identify genes that are upregulated or downregulated in diseases such as cancer, diabetes, or neurodegenerative disorders. These genes may serve as biomarkers or therapeutic targets.

Drug Response and Toxicogenomics

By comparing gene expression before and after treatment, researchers can uncover molecular responses to drugs and identify

genes involved in resistance or toxicity.

Developmental Biology and Differentiation

DGE analysis is used to study gene regulation during embryonic development, stem cell differentiation, and organogenesis, revealing how genes control complex biological transitions.

Environmental and Evolutionary Studies

Organisms exposed to different environmental conditions show changes in gene expression, which can be analyzed to understand adaptation, stress response, and evolutionary divergence.

Challenges in DGE Analysis

Despite the power of DGE, several challenges exist:

Biological and Technical Variability: Gene expression is influenced by batch effects, sample quality, and biological heterogeneity, which must be accounted for during analysis.

Multiple Testing: Thousands of genes are tested simultaneously, increasing the risk of false positives; hence, stringent corrections are necessary.

Low-Count Genes: Lowly expressed genes may yield unreliable estimates and often require filtering before analysis.

Interpretation Complexity: While statistical methods can identify differentially expressed genes, understanding their roles often requires additional biological validation.

Interpreting DGE Results

After identifying differentially expressed genes, researchers typically perform:

Gene Ontology (GO) Enrichment Analysis: Categorizes genes into biological processes, molecular functions, and cellular components.

Pathway Analysis: Tools like KEGG, Reactome, or GSEA identify enriched pathways or gene sets.

Visualization: Heatmaps, volcano plots, MA plots, and PCA plots help in interpreting the data and spotting patterns.

These analyses help contextualize DGE results and prioritize genes for experimental validation.

Case Example: Cancer Transcriptomics

A common application of DGE analysis is comparing tumor samples with adjacent normal tissues to identify oncogenes and tumor suppressor genes. For instance, in breast cancer, genes

like HER2 and BRCA1/2 may be differentially expressed, offering diagnostic and therapeutic insights. Large-scale projects like The Cancer Genome Atlas (TCGA) provide public datasets that are frequently analyzed using DGE methods.

Conclusion

Differential gene expression analysis is an essential tool in modern molecular biology, enabling researchers to uncover the dynamic nature of gene activity in various biological contexts. With RNA-seq technologies and powerful statistical tools, DGE analysis has become more accurate, accessible, and versatile. Its applications span from basic research to clinical diagnostics and drug development. While challenges remain—such as dealing with variability and interpreting high-dimensional data—ongoing methodological advances promise to enhance the precision and impact of DGE studies.

References

1. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15: 550.
2. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140.
3. Law CW (2014) voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology* 15: R29.
4. Trapnell C (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* 7: 562–578.
5. Subramanian A (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *PNAS* 102: 15545-15550.