Bioinformatics Research Starter Program by Omics Lite

Week 2 Assignment: *Differentially Expressed Genes* (DEG) *Analysis Using* GEO2R

by 257_Gadis

I. Background

Cervical cancer is one of the most common gynecological malignancies worldwide and remains a major global health concern, with approximately 570,000 new cases and 311,000 deaths reported annually [1]. Despite screening and vaccination programs, incidence and mortality rates remain high. Advances in high-throughput technologies such as microarray and RNA sequencing enable comprehensive gene expression profiling, allowing the identification of differentially expressed genes (DEGs) associated with cervical cancer progression [2]. This study aims to analyze differences in gene expression between cervical cancer and non-tumor tissues to identify potential biomarkers and therapeutic targets.

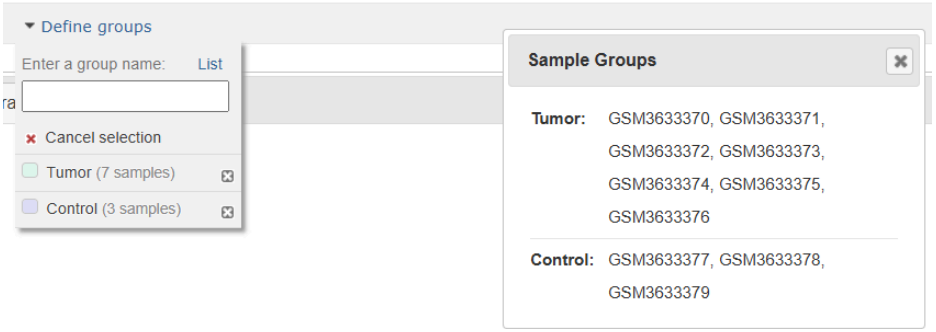
II. Method

A. Dataset

The dataset used in this analysis was GSE127265 which was obtained from Gene Expression Omnibus (GEO). The dataset contains gene expression/transcriptomics data of tumour and non-tumour cervical tissue samples analyzed using the Clariom_D_Human array platform.

B. Grouping

This dataset contains 10 samples. The samples were divided by 2 groups



- Group I → Tumor/Disease: 7 cervical cancer tissue samples
- Group II → Normal/Control: 3 normal cervical tissue samples

C. Analysis

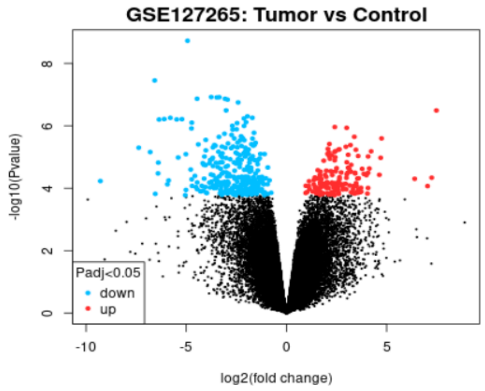
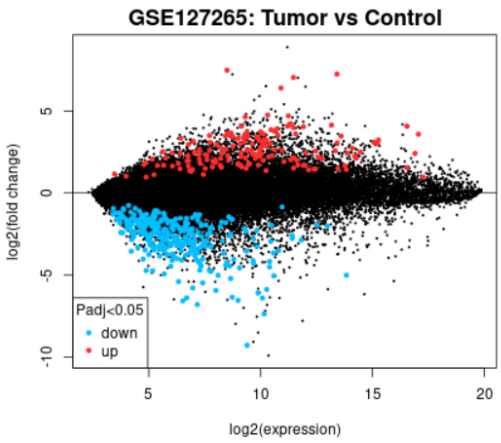
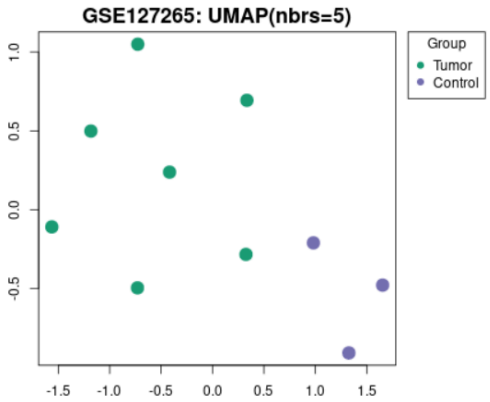
Grouping → limma → FDR → output table → filtering → Up/downregulated classification

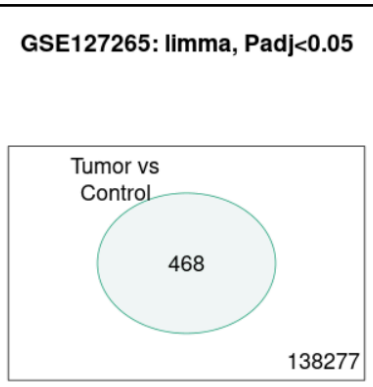
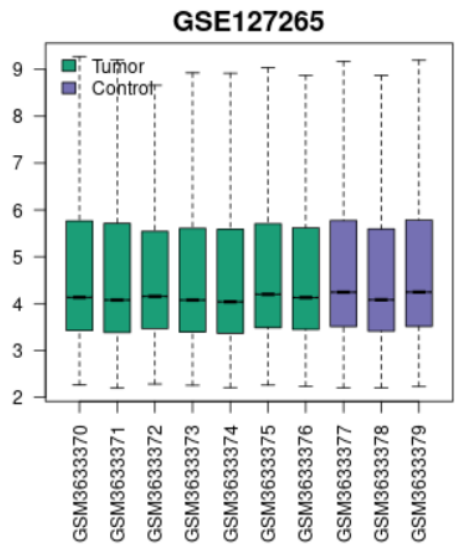
1. After grouping the samples, differential expression analysis was performed using GEO2R with its default settings. GEO2R automatically applies the limma (Linear Models for Microarray Data) statistical method to compare gene expression levels between groups.
2. Multiple testing correction is performed using the Benjamini–Hochberg False Discovery Rate (FDR) method to control for false positives.
3. The analysis generates an output table containing the following parameters for each gene
    a) Log2 fold change (log2FC)
    b) p-value
    c) Adjusted p-value (Benjamini-Hochberg)
    d) Average expression values per group
4. Genes were considered significantly differentially expressed if they met the following criteria
    a) Adjusted p-value < 0.05
    b) |log2 Fold Change| > 1 (if applied for biological relevance)
    c) Based on the log2FC value, genes were classified as either upregulated (positive log2FC) or downregulated (negative log2FC).

III. Results and Interpretation

From the analysis using the limma method 138.745 genes were analysed. Based on adjusted p-value<0.05, 468 genes show significant different expression between control and tumor group. Among these, 159 genes were up-regulated (Log2FC > 0) and 309 genes were down-regulated (log2FC < 0) in tumor groups compared to control groups.

Volcano Plot	The volcano plot illustrates the

	distribution of differentially expressed genes between the tumor and control groups. The majority of significant genes were down-regulated in the tumor group.
Mean Difference Plot 	The Mean Difference (MD) plot illustrates the relationship between average gene expression (log2 expression) and log2 fold change between tumor and control groups. Most genes cluster around log2FC = 0, indicating no substantial change in expression. However, several genes exhibit significant differential expression across various expression levels, with down-regulated genes being more predominant.
UMAP Plot 	The UMAP plot visualizes the similarity between samples. Tumor samples tend to cluster on the left side of the plot, while control samples cluster on the right side. There is a clear separation between the two groups, indicating that the global gene expression profiles differ between the tumor and control groups.
Venn Diagram	Venn diagram shows the number of significantly differentially expressed genes based on the comparison between tumor and control groups using the limma method (adjusted p-value < 0.05). From the figure, it can be observed that

	468 genes are significantly involved in the comparison between tumor and control groups, while 138,277 genes are not significantly differentially expressed in this comparison.
Boxplot 	The boxplot shows the distribution of gene expression values for each sample in the tumor and control groups. It shows comparable median expression levels and distribution ranges across all samples, suggesting good normalization and absence of major technical variation. So that indicates that the data are appropriate for downstream differential expression analysis.

IV. Conclusion

- Data were properly normalized, as indicated by comparable median expression levels across samples in the boxplot.
- A total of 138,745 genes were analyzed, and 468 genes were identified as significantly differentially expressed (adjusted p-value < 0.05).
- Among the significant genes, 159 were up-regulated and 309 were down-regulated in tumor tissues compared to non-tumor tissues.
- **That indicates, in cervical tumor tissues, gene expression changes are predominantly characterized by down-regulation rather than up-regulation.**
- Volcano and MD plots confirmed statistically significant expression changes across multiple genes.

- UMAP analysis demonstrated clear separation between tumor and non-tumor samples, indicating distinct global transcriptomic profiles.
- In general, **cervical tumor tissues exhibit substantial and biologically meaningful alterations in gene expression compared to non-tumor tissues.**

V. References

[1] S. D. A, Pasumarthi D, Pasha A, Doneti R, B. S, Botlagunta M, et al. Identification of Differentially Expressed Genes in Cervical Cancer Patients by Comparative Transcriptome Analysis. Pandi G, editor. BioMed Research International. 2021 Mar 19;2021:1–13.

[2] Exploring Differentially Expressed Genes to Identify Biomarkers of Cervical Cancer: A Bioinformatics Approach. Indonesian Journal of Medical Chemistry and Bioinformatics. 2025 Jun 9;4(1).