



ABSTRACT

Investigating differentially expressed genes in colorectal cancer datasets using a bioinformatics approach.

Sama Milhem, Tasnim Nour.

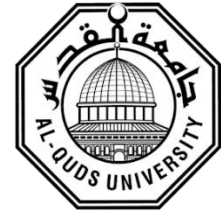
Supervisor: Haneen Nur.

Hebron University

Background: This study aimed to analyze GEO and TCGA data to enhance understanding of prognosis in early-stage colorectal cancer patients. Therefore, SLC26A3 and CLCA4 are supposed to be a risk classification markers for early-stage colorectal cancer and were discovered through bioinformatics tools.

Objectives: To investigate the prognostic significance of CLCA4 and SLC26A3 in early-stage colorectal cancer (CRC) by analyzing data from GEO and TCGA databases. Through the use of bioinformatics tools, the study aims to validate whether these genes can serve as reliable risk classification markers, thereby contributing to improved early prognosis and personalized treatment strategies for CRC patients.

Methods: Microarray data for 2 datasets, GSE32323 and GSE110224, from colorectal cancer patients from GEO and RNA-sequencing datasets from TCGA were downloaded. Using R language for overlapping DEGs and creating a volcano plot for the most up- and down-regulated genes. performed functional enrichment analyses of pathways by ShinyGo KEGG and Go for up and down-regulated genes separately, shown as a bar graph. From TCGA data, Cbioportal then



used STRING and Cytoscape to visualize genes and identify hub genes, then investigate colorectal adenocarcinoma (COAD) immune infiltration in relation to SLC26A3, CLCA4 expression.

Results: By R language produced a volcano plot and DEGs. Down-regulated genes have more confidence and interaction, in KEGG pathway affects and GO enriched analysis have clarified the effect of the gene of high enrichment in cytokine receptor interaction in up-regulated genes, and the nitrogen metabolism pathway is the most highly enriched entity among down-regulated genes. explores the relationship between SLC26A3 and CLCA4 expression and immune infiltration in colorectal adenocarcinoma (COAD). SLC26A3 shows a weak correlation with immune cells, while CLCA4 is significantly associated with macrophages, CD8+ T cells, and NK cells, negatively correlating with tumor purity. These findings suggest a potential role of CLCA4 in immune modulation, warranting further investigation. Then visualized and performed candidate genes via PPI and Cytoscape genes consisting annotated network of 15 nodes and 42 edges for down-regulated genes and 15 nodes and 17 edges between up-regulated genes.

Conclusion: The results from this study could guide the importance of adopting the integration of gene biomarkers into routine clinical practice to predict the diagnosis or treatment of CRC. It also helps in knowing the essential and early biomarkers for further detection of this cancer to ensure a healthy quality of life and treatment without resorting to surgery and chemotherapy for patient satisfaction.