

Title: Integrated Bioinformatics and Machine Learning Analysis Reveals Shared Key Candidate Biomarkers and Therapeutic Targets in Ulcerative Colitis and Colorectal Cancer

Author: Sakib Sarker, Md. Faruk Hosen, Md. Abul Basar, Emon Ahammed

Abstract: The interplay between ulcerative colitis (UC) and colorectal cancer (CRC) has garnered significant research interest due to their potential shared molecular mechanisms. This study aims to identify common significant biomarkers and potential therapeutic targets for UC and CRC. We utilized two microarray datasets to perform differential expression analysis, identifying DEGs for both conditions. Subsequent ML-based gene selection was conducted using SHapley Additive exPlanations (SHAP) algorithm models on the respective datasets. Common ML-based DEGs were then identified and a protein-protein interaction (PPI) network was constructed using the STRING database. The PPI network was visualized and analyzed in Cytoscape, with the top ten hub genes identified using the Degree method in the cytoHubba plugin. The hub genes identified were CDC20, ANLN, HMMR, CCNB1, CDK1, KIF20A, ECT2, KIF11, NUF2, and CCNA2. These genes were further validated through survival analysis, establishing their significance in patient outcomes. Finally, we explored the drug-gene interaction network to identify potential therapeutic drugs targeting these hub genes. This comprehensive bioinformatics approach provides insights into the shared molecular pathways in UC and CRC and highlights potential therapeutic targets for future research and drug development.

Keywords: Biomarkers, Colorectal cancer, Protein-Protein Interaction (PPI), Hub Genes, Survival Analysis

DOI: <https://doi.org/10.1109/ICICT64387.2024.10839675>