Identifying Potential Biomarkers for Colorectal Cancer Diagnosis Using An RNA-seqAnalysis Workflow

[1] **Zubayr Kader** and Mandeep Kaur

[1] School of Molecular and Cell Biology, University of the Witwatersrand, Private Bag 3, WITS – 2050, Johannesburg, South Africa Presenter's Email: 1384707@students.wits.ac.za
Supervisor: mandeep.kaur@wits.ac.za

Introduction: In South Africa, colorectal cancer (CRC) is the second most common cancer inmen and the third most common cancer in women, located at the lower end of the digestive system, in the colon and rectum, and is most accurately diagnosed through colonoscopy procedures. The prevalence of CRC is on the rise globally as well as locally, yet participation rates for screening tools remain low due to the invasive nature of the screening process, i.e. colonoscopy. In the present study, bioinformatics methods were employed to explore the molecular biology of patient CRC data as well as to identify dysregulated genes as potential biomarkers to be used as an alternative screening tool.

Methods: The analysis was performed by creating an RNA-Seq analysis workflow, that identified differentially expressed genes that were further used in functional analysis to identifybiological processes and pathways relating to CRC onset and progression in patients. The RNA-Seq workflow used tools from Bioconductor in the programming language R. The geneswere validated as biomarkers using existing CRC gene databases and tested in silico using statistical tests for specificity and sensitivity, as well as publicly-available blood expression data.

Results: Using the workflow, the genes identified included COL11A1, INHBA, CLDN1, ETV4 and FOXQ1 as potential tissue biomarkers, and MMP1, CTHRC1, KRT17 and IGFBP1as potential blood biomarkers.

Discussion: The identified biomarkers require future wet lab validation, however this study illustrates the potential for developing a novel CRC screening test to reduce the dependency ontraditional tools that are ineffective due to poor patient participation and associated cost. We hypothesise that acquiring South African RNA-Seq data will help establish a foundation fromwhich to identify biomarkers to improve local healthcare that focuses on early detection.

Acknowledgements: This research was financially supported by National Research Foundation (NRF), South Africa (grant number: 129356) to MK, and NRF Doctoral Scholarship (MND200503519532) to ZK.

Keywords: colorectal cancer, bioinformatics, RNA-Seq