



**Biomarkers for Diagnosis, Therapy and Prognosis in  
Colorectal Cancer: a study from databases, machine learning  
predictions to laboratory confirmations**

av

**Xueli Zhang**

**Akademisk avhandling**

Avhandling för medicine doktorsexamen i  
medicinsk vetenskap med inriktning mot biomedicin,  
som kommer att försvaras offentligt  
torsdag den 11 juni 2020 kl. 09.00,  
hörsal C1, Örebro universitet

Opponent:  
Professor Mauno Vihinen  
Lunds universitet  
Lund, Sverige

Örebro universitet  
Institutionen för medicinska vetenskaper  
701 82 ÖREBRO

## Abstract

Xueli Zhang (2020): Biomarkers for Diagnosis, Therapy and Prognosis in Colorectal Cancer: a study from databases, machine learning predictions to laboratory confirmations. Örebro Studies in Medicine 214.

Colorectal cancer (CRC) is one of the leading causes of cancer death worldwide. Early diagnosis and better therapy response have been believed to be associated with better prognosis. CRC biomarkers are considered as precise indicators for the early diagnosis and better therapy response. It is, therefore, of importance to find out, analyze and evaluate the CRC biomarkers to further provide the more precise evidence for predicting novel potential biomarkers and eventually to improve early diagnosis, personalized therapy and prognosis for CRC.

In this study, we started with creating and establishing a CRC biomarker database. (CBD: <http://sysbio.suda.edu.cn/CBD/index.html>) In the CBD database, there were 870 reported CRC biomarkers collected from the published articles in PubMed. In this version of the CBD, CRC biomarker data was carefully collected, sorted, displayed, and analyzed. The major applications of the CBD are to provide 1) the records of CRC biomarkers (DNA, RNA, protein and others) concerning diagnosis, treatment and prognosis; 2) the basic and clinical research information concerning the CRC biomarkers; 3) the primary results for bioinformatics and biostatistics analysis of the CRC biomarkers; 4) downloading/uploading the biomedicine information for CRC biomarkers.

Based on our CBD and other public databases, we further analyzed the presented CRC biomarkers (DNAs, RNAs, proteins) and predicted novel potential multiple biomarkers (the combination of single biomarkers) with biological networks and pathways analysis for diagnosis, therapy response and prognosis in CRC. We found several hub biomarkers and key pathways for the diagnosis, treatment and prognosis in CRC. Receiver operating characteristic (ROC) test and survival analysis by microarray data revealed that multiple biomarkers could be better biomarkers than the single biomarkers for the diagnosis and prognosis of CRC.

There are 62 diagnosis biomarkers for colon cancer in our CBD. In the previous studies, we found these present biomarkers were not enough to improve significantly the diagnosis of colon cancer. In order to find out novel biomarkers for the colon cancer diagnosis, we have performed /machine learning (ML) techniques such as support vector machine (SVM) and regression tree to predict candidate to discover diagnostic biomarkers for colon cancer. Based on the protein-protein interaction (PPI) network topology features of the identified biomarkers, we found 12 protein biomarkers which were considered as the candidate colon cancer diagnosis biomarkers. Among these protein biomarkers Chromogranin-A (CHGA) was the most powerful biomarker, which showed good performance in bioinformatics test and Immunohistochemistry (IHC). We are now expanding this study to CRC.

Expression of CHGA protein in colon cancer was further verified with a novel logistic regression based meta-analysis, and convinced as a valuable diagnostic biomarker as compared with the typical diagnostic biomarkers, such as TP53, KRAS and MKI67.

microRNAs (miRNAs/miRs) have been considered as potential biomarkers. A novel miRNA-mRNA interaction network-based model was used to predict miRNA biomarkers for CRC and found that miRNA-186-5p, miRNA-10b-5p and miRNA-30e-5p might be the novel biomarkers for CRC diagnosis. In conclusion, we have created a useful CBD database for CRC biomarkers and provided detailed information for how to use the CBD in CRC biomarker investigations. Our studies have been focusing on the biomarkers in diagnosis, therapy and prognosis. Based on our CBD and other powerful cancer associated databases, ML has been used to analyze the characteristics of the CRC biomarkers and predict novel potential CRC biomarkers. The predicted potential biomarkers were further confirmed at biomedical laboratory.

**Keywords:** biomarkers, diagnosis, therapy response, prognosis, database, machine learning, CRC

Xueli Zhang, School of Medical Sciences

Örebro University, SE-701 82 Örebro, Sweden, [zhang.xueli@oru.se](mailto:zhang.xueli@oru.se)