

Network-based identification of subtype-specific candidate genes and associated drugs for new therapies in colorectal cancer

Nicole de Miranda Scherer, Luís Felipe Ribeiro Pinto, Mariana Boroni, Cristóvão Antunes de Lanna

INCA - Instituto Nacional de Câncer, Instituto Nacional de Câncer

Abstract

Colorectal cancer (CRC) is the fourth most incident carcinoma worldwide, being the second in Brazil. Incidence is related to hereditary factors, eating habits, overweight and obesity, and physical inactivity. The variety of etiologic factors results in highly heterogeneous tumors with distinct prognosis and response to treatment. Different classification strategies have been proposed to characterize tumors more efficiently. The Colorectal Cancer Subtyping Consortium (CRCSC) recently identified four consensus molecular subtypes (CMS1-4) from primary CRC transcriptomic data. Identification of disease-related genes with high potential for drug interactions may assist in discovering new targets and more effective therapeutic strategies. This enables repositioning of previously approved drugs to treat other diseases and may reduce the time required to approve new treatments. However, few works have assertively proposed new treatments based on this classification system. For this reason, the aim of this work is to identify candidate genes and associated drugs for the development of new therapies for different molecular subtypes of colorectal cancer from large-scale genomic and transcriptomic data. Gene expression data from 623 patients generated by The Cancer Genome Atlas (TCGA) were used, totaling 623 samples from primary tumor tissue and 51 from tumor-adjacent tissue. Tumor samples were classified into 4 groups using the CMSClassifier package, with posterior subdivision of CMS4 samples into epithelial and stromal. Unique differentially expressed genes (DEGs) in each CMS subtype were identified with DESeq2 and InteractiVenn. Co-expression modules were constructed using weighted gene correlation network analysis (WGCNA), correlated with subtypes and normal samples, and used in the construction of protein-protein interaction networks using the STRING base. Interactions with low confidence were filtered out and subgraphs were identified within each module based on modularity using the igraph package. Candidate target genes were selected based on degree, closeness betweenness, and pagerank centralities. Each of these centralities was measured, converted to z-score, and combined with log2 fold change. For each measure, the top 20 genes with the highest score were selected for drug-gene interaction identification in the DGIdb database. Drug-gene interactions were validated using tumor-derived cell line sensitivity data from Genomics of Drug Sensitivity in Cancer (GDSC), classified into CMS subtypes from expression data available from the Gene Expression Omnibus (GEO) database using CMScaller, a cell line-specific classifier. Subtype-specific drug-gene interactions as well as interaction overlaps were evaluated in order to propose candidate drug combinations for further testing. These results demonstrate the potential for the evaluation and implementation of new therapeutic strategies in CRC and the possibility of implementing these analyses in other tumor types.

Funding: Capes, Ministério da Saúde
Link to Video: