Transcriptomics approach to identify subtype-specific candidate genes and associated drugs for new therapies in colorectal cancer

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

Instituto Nacional de Câncer

Abstract

Colorectal cancer (CRC) is the fourth most prevalent carcinoma worldwide, being the third most common in men and the second in women in Brazil. The incidence is related to hereditary factors, eating habits, overweight and physical inactivity. The variety of combinations of these factors results in highly heterogeneous tumors reflecting different 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 the discovery of 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. 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 tissue adjacent to the tumor. 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 subgroups were identified within each module using the igraph package. Hub genes were selected by the subgroups' connectivity degrees and used to search for drug-gene interactions in the DGIdb database. Molecular-type Drug propositioning were validated using sensitivity data in cell lines 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. Thirty drugs for CMS1, 33 for CMS2 and 33 for CMS4e were identified, 16 of which have already been tested in cell lines. These 16 drugs were evaluated for repositioning in repoDB. Of these, four have not yet undergone cancer clinical trials, and among the others, only two have been tested for CRC, one of them being approved. Seven hubs genes identified within the criteria defined in this work do not have known interactions with drugs. 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, INCA, Ministério da Saúde