Comparative analysis of gene expression data between colorectal cancer cell lines with wild-type and silenced MMR genes

Cristóvão Antunes de Lanna, Nicole Scherer, Mariana Boroni

Laboratório de Bioinformática e Biologia Computacional (LBBC), Instituto Nacional de Câncer (INCA)

Colorectal cancer (CRC) is a high incidence carcinoma in Brazil and in the world, with high mortality rates in less developed countries. 34,280 new cases are estimated for 2016 in Brazil. Since the obtention of CRC samples at different stages of development is relatively easy, it allowed the in-depth study of CRC progression, as well as its molecular basis and the characterization of CRC subtypes. One of the known pathways of CRC progression is the microsatellite instability pathway (MSI), caused by mutations or epigenetic silencing in genes involved in mismatch repair mechanisms (MMR). As cell lines can be used to represent different stages of tumors according to their origin, this study aims to characterize the expression profiles of CRC cell lines with altered MMR pathway (MMR-) compared with control ones (MMR+). For this, RNA-seq data was obtained for nine CRC cell lines from project SRP052201 at the SRA database: MMR+ : SW480, HT29, COLO205, Caco2 and MMR- : LS174T, LoVo, HCT116, HCT15, RKO. After analyzing the quality of the reads and removing low-quality bases using FastQC and Trimmomatic, respectively, these sequences were aligned to the human genome (GRCh37 version) and sorted by coordinate, using the STAR algorithm. Then, the quality of the alignments was analyzed using RSeQC. After that, the mapped reads for each gene were quantified using HTSeq-count and the expression levels of genes were estimated through the edgeR tool, written in R, from the Bioconductor repository. A total of 245 differentially expressed genes were obtained, of which 229 were down-regulated and 16 were up-regulated in the MMR- cell lines. Differentially expressed genes were classified according to the Gene Ontology (GO) database categories, using the clusterProfile Bioconductor package. Enriched biological function categories include processes related to angiogenesis, response to wounding, and lipid and steroids metabolism. Membrane and extracellular matrix components represented the majority of cellular components enriched in this analysis. Enriched categories for molecular function included sulfur compound, heparin, and glycosaminoglycan binding.