Differential expression in colorectal cancer progression

Aline Duarte Gomes, Paulo Thiago Santos, Nicole Scherer, Mariana Boroni

Laboratório de Bioinformática e Biologia Computacional - Instituto Nacional de Câncer (INCA), Rio de Janeiro, RJ, Brazil

Colorectal cancer (CRC) is the third worldwide most commonly diagnosed cancer in men and the second one in women, with more than 1 million cases per year. Deaths from CRC are relatively lower (8.5% of the total) with more deaths (52%) in the less developed regions of the world, reflecting a poorer survival in these regions. Estimates for Brazil indicate that CRC will affect more than 34,000 people in 2016. Furthermore, CRC provides a good model for the study of morphological and genetic stages in cancer progression, since its tumor progression process is very well described. Large-scale studies have pointed out genomic, epigenomic and gene expression alterations that contribute to this tumor development. However, each of these studies provides an one-dimensional and limited view of this whole system. We aim to study the differential gene expression in CRC during its progression and investigate important pathways affected in different stages. In order to identify differentially expressed genes, we have analyzed RNA-seq data from the TCGA database. Our data set was composed by 287 colon cancer and 41 normal tissue samples (14 of which are paired), totalizing 328 samples. For the differential expression analysis, we employed DESeq2 package from Bioconductor for R, with a cutoff values set to $P < 0.001$ and log2FoldChange >= 1.58. Then, using the package clusterProfiler from Bioconductor repository, we identified the enriched pathways in the Gene Ontology (GO) database in differentially expressed genes with a cutoff set to 0.05 for both pvalue and qvalue. A total of 2,823 differentially expressed genes were found, of which 1,158 are up-regulated and 1,594 are down-regulated. We found that the differentially expressed genes are enriched in high activity channels, transmembrane transporters, receptors and growth factor in molecular function categorie; for biological process, we see an impact on the muscular system, on the cellular homeostasis processes, in regulating hormone levels and blood circulation; and the terms enriched in cellular component indicate alterations in the extracellular matrix, cellular transport, membrane components and components of the muscular system. These data shows important processes that are affected in CRC, as higher expression of receptors, growth factors and hormones that could be related to proliferative signaling, and also alterations in the extracellular matrix that could be important to the activation of invasion and metastasis, important hallmarks of cancer.

Keywords: colorectal cancer, gene expression, RNA-seq, TCGA