

# Use of data mining for Onco-targets to analyze Breast Cancer through the construction of Ontology Networks

Edgar Lacerda de Aguiar<sup>1</sup>, Lissur Azevedo Orsine<sup>2</sup>, José Miguel Ortega<sup>3</sup>,

*1 Cefet-MG*

*2 UFMG*

*3 Universidade Federal de Minas Gerais, Laboratório de Biodados*

## Abstract

Studies indicate that by the end of 2017 more than 23 million patients will develop some type of cancer. Among the various types of cancer, breast cancer has the most impact among women and one of the highest mortality rate. Breast cancer has high biological heterogeneity, which implies a high diversity of molecular forms which are associated with distinct subtypes and distinct drug targets. This high range of variations in the biological entities involved in disease pathology impacts directly on diagnosis and treatment. Because of these facts this work aims to mine the possible genes related to breast cancer, from different databases (DBs), Cancer Genome Atlas (TCGA), COSMIC, KEGG and to relate the genes to database Gene Ontology (GO) with its molecular functions, biological processes, cellular components, thus inferring the main ontologies associated with breast cancer. Initially there was an interpolation of genes between BDs CGA and COSMIC after cured genes were mined and crossed with the top mutated genes of breast cancer. The genes were cured with the BDs UniProt, NCBI and KEGG, focusing on DB KEGG Pathway, which was used to search for the genes of the various types of cancer. With the UniProt values Id of cured Genes there was a search for Ontologies in the GO database. The initial results weren't very satisfactory due to high specificity and high granularity of ontological terms. Better treatment of the data and a new methodological approach to the Ontological terms was necessary. The ontological terms of GoSlim, which are terms and healed with a median specificity were used. Several enrichment analyzes were performed comparing breast cancer genes with genes responsible for breast development. Through these analyzes it was possible to note that approximately 10% of the breast cancer genes were found in the group of genes responsible for breast development. In the analysis of biological processes of the genes were associated 42% in metabolic process, 39% in response to stimulus and 33% in developmental process. In the Molecular function 33% in binding, catalytic activity 34% and 10% receptor activity. The analyzes show a significant correlation between many biological processes and molecular functions encountered. This work is important for a better understanding of breast cancer and the genes responsible for breast development, through better analysis of biological processes, molecular functions using data mining and ontology network, to aid in the search for Onco-targets.

Funding: Cefet-Mg, UfmG