

# The CD90/Thy1 in Triple Negative Breast Cancer: associations by bioinformatics between dysregulated genes and Signaling Pathways

Marco Lázaro de Sousa Batista, Aline Ramos Maia Lobba, Mari Cleide Sogayar,  
Ana Claudia Oliveira Carreira, Milton Yutaka Nishiyama Junior

*NUCEL (Cell and Molecular Therapy Center), Biochemistry Department, Chemistry  
Institute, University of São Paulo, Brazil.*

## Abstract

Breast cancer is the most frequently diagnosed type of cancer among women in the World, with the ductal-invasive triple-negative (TNBC) type being the most aggressive and lethal. We previously demonstrated that the CD90 is a promising diagnostic and therapeutic target for TNBC patients, in the TNBC-derived cell line non-tumorigenic MCF10A overexpressing CD90 (MCF10A/CD90+) and the tumorigenic Hs578T knockdown CD90 (Hs578T/shCD90). The approaches for high-throughput gene expression analysis, such as DEGs identification, and Pathway Enrichment Analysis are well established. Therefore, the aim of this work is to elucidate the possible crosstalk between CD90 and signaling pathways in CD90 TNBC-derived cell line models, through the RNA-seq, applying and developing new System Biology approaches integrating the in house curated TNBC canonical pathways and genes. The edgeR approach has been used for DEGs identification, and EnrichR software, for the identification and comparison of enriched pathways based on KEGG and Reactome databases. A pipeline has been developed to integrate the TNBC biological knowledge, with the DEGs, the Enriched pathways and with the gene set enrichment analysis. A new proposed strategy improved the integration between the DEGs with the Enriched pathways, separating into three DEGs sets: All, Up- and Down-regulated genes. The enrichment analyses was conducted with the sets in MCF10A/CD90+ (718 DEGs) and Hs578T/shCD90 (16 DEGs). Therefore, a new strategy for RNA-seq analysis has been proposed based on the cell models assays, and the new curated Gene sets from Molecular Signature Databases (MsigDB), integrated with the DEGs and enriched pathways. Each assay was conducted using all gene expression profiles, in each curated gene sets on GSEA (Gene Set Enrichment Analysis) method, with fGSEA software. The PCA analysis was adopted to compare the gene and pathways merged between them, using a strategy to calculate the average of each gene in a specific pathway. The weighted gene co-expression network, using the WGCNA software has been used to integrate the enriched pathways and PCA analysis. . Based on the literature, we have identified the TNBC canonical genes (151) and pathways (65), which have been used as a validation set for the identified DEGs and enriched pathways in the first approach. From the MCF10A/CD90+ DEGs, has been found 14 canonical genes, while for Hs578T/shCD90 we have found 3. The PCA analysis has shown two groups of pathways set, segregating MCF10A/CD90+ and Hs578T/shCD90, showing 5 important pathways involved in opposite biological process in these lineages and they will be studied more carefully.

Funding: Support: FAPESP, PIBITI/CNPq, CAPES, BNDES, FINEP, MS-DECIT, MCTI.