

# Integrative analyses of single-cell and bulk transcriptomics to study the tumor microenvironment in high-grade serous ovarian cancer

Mariana Boroni, MARCO ANTONIO MARQUES PRETTI, Gabriela Rapozo Guimarães, Palloma Porto Almeida, Giovanna Resk Maklouf

*INCA - Instituto Nacional de Câncer, Instituto Nacional de Câncer*

## Abstract

Ovarian cancer (OC) is the major cause of death related to gynecological tumors, with high-grade serous ovarian cancer (HGSOC) being the most common subtype of this disease accounting for the majority of deaths. The elevated mortality rates of HGSOC have been reported to be related to the interconnected signaling networks, cellular composition, and spatio-temporal localization of cells within the tumor microenvironment (TME). In this way, the aim of this project is to characterize the tumor microenvironment, integrating bulk and single-cell transcriptomics of HGSOC and investigate the impact of different subpopulations in patients' outcome. In order to accomplish that, single-cell RNA sequencing data (scRNA-seq) from public datasets were analyzed using the Seurat R package. Quality control, normalization, clustering, and differential gene expression analysis were performed to define the resulting clusters into cell types. Also, a thorough review of gene markers was made to support the annotation of the subpopulations. For a comprehensive assessment of the subpopulation's roles in a larger cohort, we used CIBERSORTx, a deconvolution tool that employs machine-learning to detect the abundance of subpopulations in bulk RNA-seq data based on a single-cell reference signature matrix. Subsequently, Cox Proportional Hazard Models were used with Survival and Survminer R packages to determine which cell type significantly influenced the patients' risk of death. Using the scRNA-seq data of 26,786 cells derived from 5 patients, we have identified 24 clusters with distinct profiles within malignant, immune, and stromal major populations relevant to HGSOC. The cell types identified in the analysis comprised not only commonly observed subpopulations related to cancer, e.g. T CD4 and T CD8, but also more rare ones, e.g. pericytes and adipogenic fibroblasts. Finally, deconvolution with bulk RNA-Seq data of 454 patients from The Cancer Genome Atlas and International Cancer Genome Consortium databases revealed that cancer-associated fibroblasts, endothelial cells, and malignant cells were the most abundant subpopulations in HGSOC patients. However, regarding the hazard ratio (HR) related to the overall survival of the patients, the enrichment in T CD8 (HR = 0.82,  $p = 0.002$ ) and T CD4 (HR = 0.85,  $p = 0.012$ ) signatures were associated with a good prognosis, which corroborates with the high lymphocyte density as a common indicator of better outcomes at different stages of disease in many malignancies, including HGSOC. Our approach to explore the HGSOC TME will provide insights into how intratumoral content besides cancer cells can operate as a prevalent factor in the patient's prognosis, as well as offer potential diagnostic biomarkers and therapeutic targets in future clinical practice.

Funding: MS/INCA, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) e Chan Zuckerberg Initiative (CZI)

Link to Video: