# The G-S-M, Grouping, Scoring and Modeling Approach. Application of Biological Domain Knowledge for Groups Selection on Gene Expression Data

## Malik Yousef, PhD
## Zefat Academic College, Israel

In the last two decades, there have been massive advancements in the high throughput technologies, which resulted in exponential growth of public repositories of gene expression datasets for various phenotypes. It is possible to unravel biomarkers by comparing the gene expression profiles in different conditions (e.g. cases vs. controls, drug treatments after different time points, different tissues). This problem refers to a well-studied problem in the machine learning domain, i.e., the feature selection problem. In biological data analysis, most of the computational feature selection methodologies were taken from other fields, without considering the nature of the biological data. Thus integrative approaches are necessary for this kind of data. The main aim of integrative gene selection is to generate a ranked list considering both statistical metrics applied on the gene expression data and the biological background information provided as external datasets or as omics data sets.

During the last years we have developed an integrative approach that perform groups selections rather than feature selections. The generic approach G-S-M (Grouping, Scoring, and Modeling). I will present our related works in this topic such as SVM-RCE, maTE, CogNet, miRcorrNet, 3Mint, GediNET, PriPath, TextNetTopics GeNetOntology MicroBiomeNet, DisMir-Net and other tools under development.

**Thursday August 25, 2022, 3:30-4:30 PM Eastern**

**133 Rosenau Hall**

Virtual using link and info below.

Link: Link: https://unc.zoom.us/j/92602267820?pwd=YW1wN1pjdUNVd1A4TTl2OStmVHBjQT09
Meeting ID: 926 0226 7820  Passcode: 533114

BIOSTATISTICS SEMINAR