Data- Independent Analysis of EPS-urine coupled to Machine Learning: a predictive model for prostate cancer

Licia PRESTAGIACOMO - *Magna Graecia University of Catanzaro*

Prostate cancer (PCa) is the most frequently diagnosed cancer in the male population with an estimated 268,490 new cases and 34,500 estimated deaths in the 2022 in the United States. To date, PCa detection is based on the digital rectal exam (DRE) and on the serum prostate specific antigen (PSA) assay. Although PSA-based screening is used for the detection of PCa, it exhibits low specificity, sensitivity and the inability to discriminate between aggressive and indolent tumours. For this reason, it is necessary to detect a new biomarker or a biomarkers panel with better diagnostic power than PSA.

In this work, the proteomic profile of EPS-urine sample from PCa patients (n= 73) and benign prostatic hyperplasia (BPH; n= 60) was analyzed by mass spectrometry analysis in order to identify differentially expressed proteins between the two groups. To achieve deep proteome coverage, EPS-urine samples were analyzed by data-independent analysis (DIA) a sensitive method suitable for identifying even low abundance proteins[3].

This approach allowed us to identify a total of 2615 proteins in our sample set (n= 133); the matrix with the quantified proteins was merged with clinical parameters and the hybrid data was analyzed by Machine Learning algorithms in order to draw up a predictive model.

Our predictive model consisting of two proteins and two clinical variables, included: semaphorin-7A (sama7A), secreted protein acid and rich in cysteine (SPARC), FT-ratio and prostate gland size.