The increase or increment of a signaling pathway in a biological sample may indicate which biological processes are getting activated in the sample analyzed. In the case of tumor RNA-Seq, as is the case of the original dataset downloaded for the Machine Learning study, we may understand which phenotypes characterize a sample by some quantification of the phenotype of interest.

Previously, we determined hypoxia, a condition of limited oxygen availability for cancer cells, as a factor that modulates prognosis in neuroblastoma. Hypothetically, we can quantify how much hypoxia a tumor has, by the quantification of the expression of genes or gene sets related to hypoxia.

Among a variety of computational methods available to quantify phenotypes in tumor samples, we chose the GSVA – Gene Set Enrichment Analysis, to quantify signaling pathways or phenotypes. GSVA allows power to detect activity of a signaling pathway, or phenotype, in a sample or a population of samples.

In the data frame containing 22 phenotypic scores, 6 phenotypes are simple quantification of 6 gene expressions: MYCN, STC1, P4HA1, BHLHE40, HIF1A and ST8SIA1 measured in in log2RPM. The other 16 phenotypic scores are GSVA phenotypic scores: the GSVA estimation of the expression of that gene set (column names) in the tumor sample (row labels). The ML method allowed us to classify the tumor risk based on the expression of these 22 phenotypic scores.