Data Challenge

Clinical context

Multiple Myeloma (MM) is a type of bone marrow cancer. Treatment for MM involves combinations of drugs over multiple cycles. There is huge heterogeneity in treatment response with some individuals not responding to treatment and some patients responding well to treatment for some time before a relapse. A better characterization of patients who relapse early can influence the treatment options and combinations.

In this challenge, we propose to develop a model for predicting the risk of dying or relapsing of newly diagnosed multiple myeloma patients from baseline clinical and expression data.

Data

The data for this test are extracted from an old Synapse Dream Challenge (https://www.synapse.org/#!Synapse:syn6187098/wiki/401884).

It consists of clinical data, gene expression data and follow-up for newly diagnosed Multiple Myeloma patients extracted from the MMRF CoMMpass IA9 study. In the data, newly diagnosed MM patients are classified as High Risk (HR) when they relapse or die before 18 months.

To access the data, you first need to create an account on the synapse platform, and then to download the following files:

- Expression data:
  MMRF_CoMMpass_IA9_E74GTF_Salmon_entrezID_TPM_hg19.csv
  (https://www.synapse.org/#!Synapse:syn10573789)
  [notice that the first column gives *Entrez IDs* for gene]

- Clinical data and labels:
  sc3_Training_ClinAnnotations.csv
  (https://www.synapse.org/#!Synapse:syn9926878)

- Explanation of the clinical and label annotations:
  Harmonized_Clinical_Dictionary
  (https://www.synapse.org/#!Synapse:syn9744732)

Aim

The purpose of this data challenge is to develop a machine learning model for predicting the risk of fast dying or relapsing of newly diagnosed MM patients. The label for high risk death or relapse is provided under the label HR_FLAG in the clinical data file.

Notice that observations undergo [censoring](#) and that the HR_FLAG can be either "True", "False" or "Censored". As a first approach, you can restrict yourself to uncensored patients, leading to a binary classification problem.

.