**BINF6310 Public Databases, Container, and Pipelines - Module 08 Activity**

**Aim:** This assignment aims to familiarize students with analyzing mutations and gene expression data in specific cancer types, and performing a survival analysis.

**Objective:** To explore the relationship between gene mutations, gene expression, and patient survival in a specific cancer type using cBioPortal.

---

## Part 1: Data Exploration and Gene Mutations

1. Go to cBioPortal: cBioPortal
2. Choose a Cancer Study
    - Click on "Query" in the top menu.
    - In the search bar, type **"Breast Invasive Carcinoma (TCGA, PanCancer Atlas)"** and select it. The dataset has 1,084 samples.
    - Click "**Explore Selected Studies**."
3. Gene Query
    - In the gene query box on the top right, enter **TP53** (a commonly mutated gene in cancer).
    - Click "**Submit**."
4. Analyze Mutations
    - You will be presented with an overview of TP53 mutations in breast cancer on the **OncoPrint** menu.
    - Note the mutation frequency, the types of mutations (e.g., missense, nonsense), and the affected regions of the gene.

**Questions**:

- What percentage of patients have TP53 mutations in this dataset?
- What are the most common types of TP53 mutations in breast cancer?

## Part 2: Gene Expression Analysis

1. Scroll to Plots, and click the **"Plots"** menu tab.
2. On the left-hand side, Set the Horizontal Axis (x-axis)
    - In the **Horizontal Axis** section, change the **Data Type** to **Mutation**.
    - Under **Gene**, select **TP53** (or keep it if it's already selected).
    - This will set **TP53 mutation status** as the x-axis to differentiate between mutant and wild-type samples.
3. Set the Vertical Axis (y-axis)
    - In the **Vertical Axis** section, under **Data Type**, select **mRNA**.
    - For the **mRNA Profile**, select **mRNA Expression, RSEM (Batch normalized)** (as visible in your screenshot).
    - Ensure that **TP53** is selected as the gene to analyze its expression.
4. Generate the Plot
    - Once these options are selected, you should generate a plot showing the relationship between **TP53 mutation status** and **TP53 mRNA expression**.

- Deselect **"Copy Number" and "Structural Variant"**
- The plot will now display the mRNA expression levels on the vertical axis and TP53 mutation status (mutant vs. non-mutant) on the horizontal axis.

**Questions**:

- Would you hypothesize that certain TP53 mutation types (e.g., truncating vs. missense) are more likely to be associated with cancer progression?

- Does the presence of multiple mutations or splice variants seem to have a distinct impact on TP53 expression compared to a single missense mutation?

## Part 3: Perform a Survival Analysis

1. Click on the **"Survival"** tab.
2. From the dropdown menu, choose **Overall Survival** and compare patients with and without TP53 mutations.
3. Generate survival plots to see if there is a significant difference in survival between these groups.

**Questions**:

- Do patients with TP53 mutations have a significantly different survival rate compared to those without?
- How could TP53 mutations impact patient prognosis?