**VALIDATION ANALYSIS**

Following the screening of external proteins using the machine learning model, validation analyses were conducted to compare the mean genomic alterations between breast cancer (BC)-related proteins (prediction 1) and BC non-related proteins (prediction 0). For this purpose, we selected the study *‘Breast Invasive Carcinoma (TCGA, PanCancer Atlas)’* from cBioPortal (<https://www.cbioportal.org/>). Subsequently, genomic alterations (CNA amplifications, CNA deep deletions, in-frame driver mutations, truncating driver mutations, missense driver mutations, mRNA overexpression, mRNA underexpression, protein overexpression, and protein underexpression) associated with the 4,504 queried proteins were downloaded and analyzed across a cohort of 1,066 individuals, as per the Pan-Cancer Atlas.

The results were compiled in an Excel document titled *‘Validation\_BC\_proteins’*, which includes the following sheets:

**Sheet 1: SUMMARY**

This sheet provides a table summarizing the mean genomic alterations among CIPs (cancer immunotherapy proteins) related to BC vs. non-related to BC, MDPs (metastasis driver proteins) related to BC vs. non-related to BC, and RBPs (RNA-binding proteins) related to BC vs. non-related to BC. Additionally, rankings of BC-related proteins associated with cancer immunotherapy, metastasis, and RNA binding from the original study, undersampling, and SMOTE analyses are included.

**Sheet 2: DATA\_PAPER**

Contains genomic alterations published in the original study.

**Sheet 3: UNDERSAMPLING**

Includes genomic alterations of BC-related proteins (three datasets) analyzed using the prediction model with undersampling.

**Sheet 4: SMOTE**

Presents genomic alterations of BC-related proteins (three datasets) analyzed using the prediction model with SMOTE.

**Sheet 5: UNDER\_MDP**

Details genomic alterations of metastasis driver proteins analyzed using undersampling.

**Sheet 6: SMOTE\_MDP**

Contains genomic alterations of metastasis driver proteins analyzed using SMOTE.

**Sheet 7: UNDER\_CIP**

Includes genomic alterations of cancer immunotherapy proteins analyzed using undersampling.

**Sheet 8: SMOTE\_CIP**

Details genomic alterations of cancer immunotherapy proteins analyzed using SMOTE.

**Sheet 9: UNDER\_RBP**

Contains genomic alterations of RNA-binding proteins analyzed using undersampling.

**Sheet 10: SMOTE\_RBP**

Details genomic alterations of RNA-binding proteins analyzed using SMOTE.