**Data Processing**

All analyses of the RNA-Seq data from the TCGA database were performed using the R language (version 4.3.3) or public databases. Prior to differential analysis, the scale function was applied to transform gene expression levels into z-scores by performing the (x-μ)/σ transformation. To verify the findings from TCGA, external validation was conducted using the GEO database or other relevant databases. Initially, we converted the data into unitless Z-Score values by calculating (x-μ)/σ to eliminate dimensional discrepancies among various samples. Outliers with Z-Score values below -3 or above 3 were subsequently removed. Following the removal of outliers, the tumor was included in the analysis only if there were at least three normal samples available. In the case of proteomic data from the CPTAC database, missing values in the expression data were eliminated during the preprocessing phase to guarantee the accuracy and reliability of subsequent analyses. Subsequently, samples were distinctly classified into tumor and normal groups according to their clinical information.

**Statistical Analysis**

The Wilcoxon rank-sum test, a non-parametric method, is especially suitable for data that do not conform to a normal distribution and can effectively assess expression differences between two groups. Ultimately, box plots were employed to visually illustrate the mRNA and protein expression levels in both tumor and normal groups, clearly highlighting the expression differences of the target genes. Box plots not only depict the median and interquartile range but also help identify potential outliers, offering a more comprehensive view for subsequent analysis. ROC curve analysis was conducted using the "pROC" package to evaluate the diagnostic performance of genes. All statistical tests were two-tailed, with p-values less than 0.05 deemed statistically significant and p-values less than 0.0001 regarded as highly significant. Statistical significance was reported as follows: \* P<0.05, \*\*P<0.01, \*\*\*P<0.001, and \*\*\*\*P<0.0001. Kaplan-Meier survival analysis was employed in this study, with data analysis performed using the survival package in R. Initially, the optimal cutoff value for MPZL2 gene expression levels was established using the R package survminer to differentiate between high and low expression groups, ensuring that the sample proportion in each group was at least 0.3. Subsequently, the survfit function was utilized to conduct the Log-rank test, evaluating whether the difference in survival curves between the two patient groups was statistically significant. In addition, this study utilized the survival package in R to conduct univariate and multivariate Cox survival analysis on gene expression and traditional clinical variables, evaluating their influence on patient survival time. The Cox proportional hazards model, a semi-parametric regression model, is well-suited for analyzing survival time data and can evaluate the impact of multiple variables on survival time. During the analysis, we calculated the hazard ratio (HR) and its corresponding 95% confidence interval (CI) for each variable to characterize the relative risk. To visualize the results, we employed the forestploter package to create forest plots, a widely utilized graphical tool for presenting effect sizes and their confidence intervals from various studies.