IdMarker

2023-05-14

Introduction

`IdMarker` was a comprehensive framework to identify multigenic prognostic signature for cancer outcome prediction and patients stratification. First, IdMarker will randomly divide the samples into train group and test group according to the `ratio` specified by the user. In train data, we will apply univariate Cox proportional hazards regression analysis to filter candidate genes. With these candidate genes, multivariate Cox regression model to construct the risk model. IdMarker will calculate the value of `Overall Risk Score` by calling the model. According to Overall Risk Score, The patients will classified into high-risk group and low-risk group. Later, the performance of the model will be validated by performing the Kaplan-Meier (KM) survival analysis and the log-rank test.

Model

The formula to calcuate the `Overall Risk Score` was listed as follows:

0.141 * IGF2BP3 + 0.009 * SMC1B + 0.053 * IGF2BP1 + 0.034 * CLDN6 + 0.069 * LY6K

Results

Univariate Cox proportional hazards regression analysis

After applying univariate Cox proportional hazards regression analysis (p value < 0.05), the candidate genes were narrowed to genes as follows:

```
##           beta HR (95% CI for HR) wald.test  p.value
## IGF2BP3  0.209    1.23 (1.1-1.38)      13.2 0.000285
## SMC1B    0.111   1.12 (1.02-1.22)      6.33   0.0119
## IGF2BP1  0.127   1.14 (1.06-1.22)      12.7 0.000368
## CLDN6   0.0952     1.1 (1.03-1.18)      7.68  0.00558
## LY6K     0.133   1.14 (1.05-1.24)        10  0.00154
```

Coefficients from multivariate Cox regression analysis

Applying multivariate Cox regression analysis, the coefficients for genes are listed as follows:

```
##                 coef exp(coef)   se(coef)         z   Pr(>|z|)
## IGF2BP3 0.141297334  1.151767 0.06564668 2.1523912 0.03136656
## SMC1B   0.008531092  1.008568 0.05306976 0.1607524 0.87228840
## IGF2BP1 0.052971622  1.054400 0.04070752 1.3012737 0.19316478
## CLDN6   0.034097883  1.034686 0.03758046 0.9073302 0.36423223
## LY6K    0.068645013  1.071056 0.04956054 1.3850739 0.16602989
```

Applying the model to train group and KM analysis to evaluate its preformance

The data of training group was applied to the model. According to the calculated `Overall Risk Score`, the threshold was 2.124. With the threshold, the patients in the train group was classified to high-risk group and low-risk group. KM analysis was applied to evaluate the performance and the survival plot was as follows.
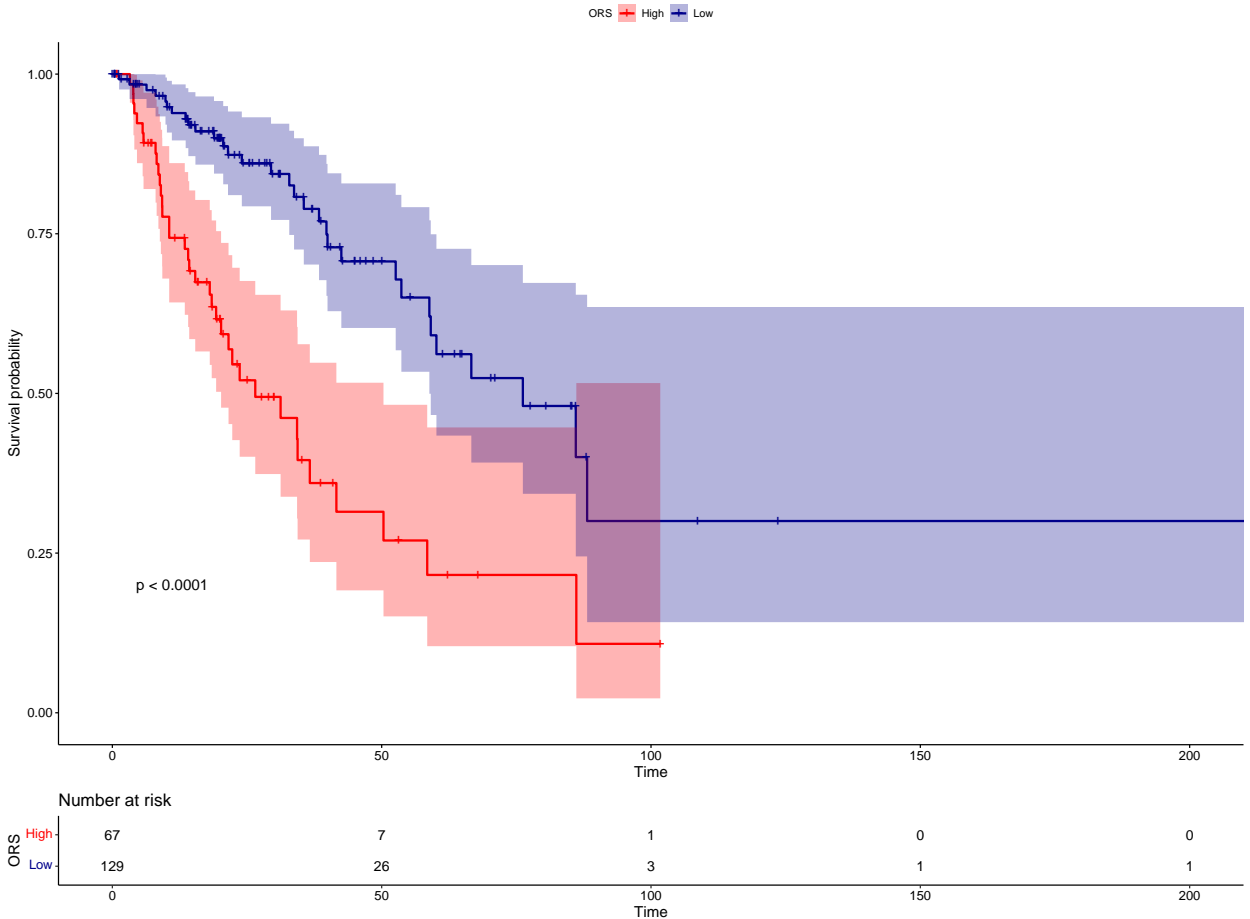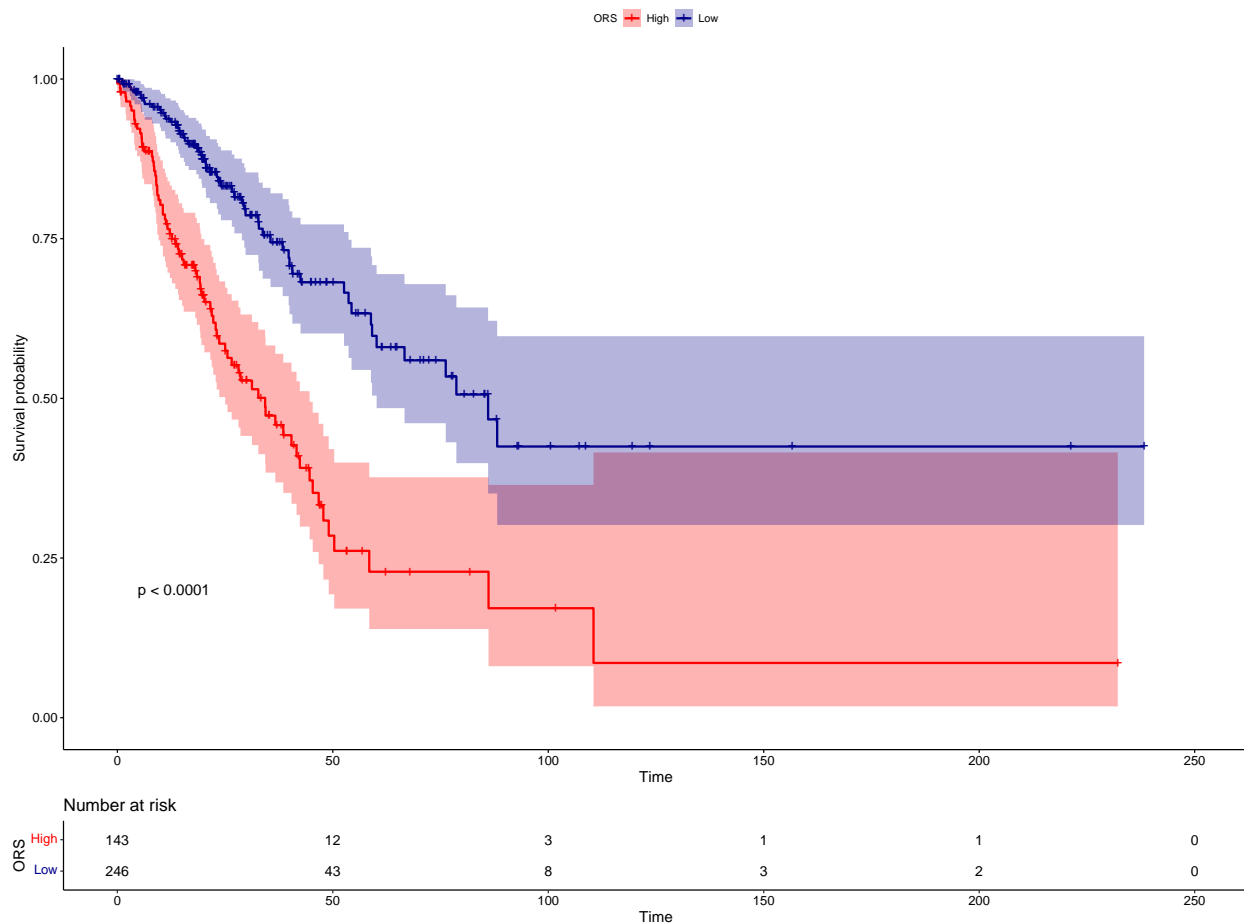

KM analysis for test group

The threshold for test group was 2.138. The survival plot was as follows.
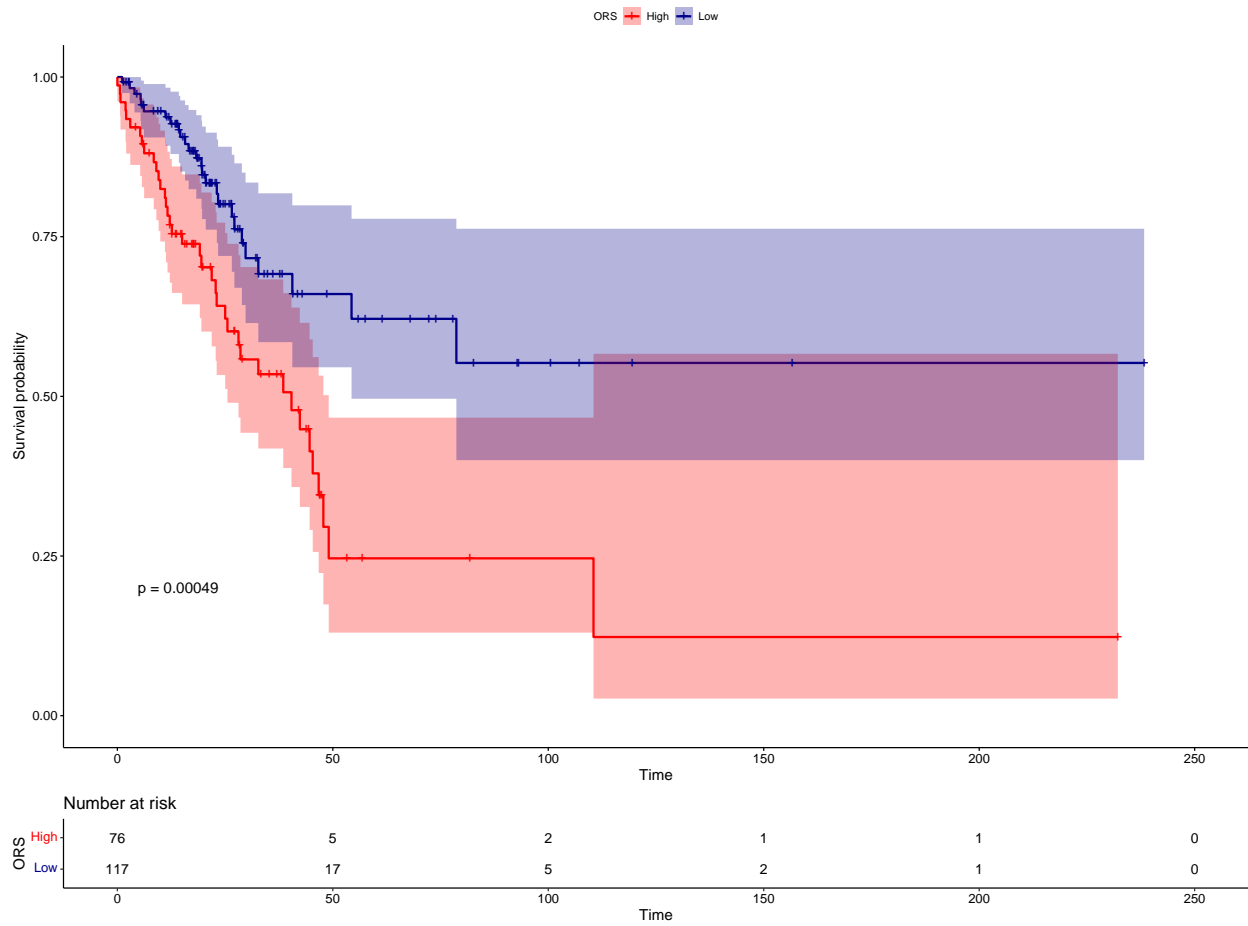
KM analysis for total patients

The threshold for total patients was 2.138. The survival plot was as follows.

Results Files

Patients' information of train group was saved in "`Results/train_group_patients.txt`".

Patients' information of test group was saved in "`Results/test_group_patients.txt`".

Result of patients' univariate Cox regression was saved in "`Results/uni_candidate.txt`".

Result of patients' multivariate Cox regression was saved in "`Results/multi_candidate.txt`"

Result of KM Analysis for train group was saved in "`Results/train_survival_plot.pdf`"

Result of KM Analysis for test group was saved in "`Results/test_survival_plot.pdf`"

Result of KM Analysis for all the patients were saved in "`Results/total_survival_plot.pdf`"