Prognostic and biological relevance of collagen-related genes in prostate cancer

Methods, figures and tables for the transcriptome part

2025-02-21

# Figures

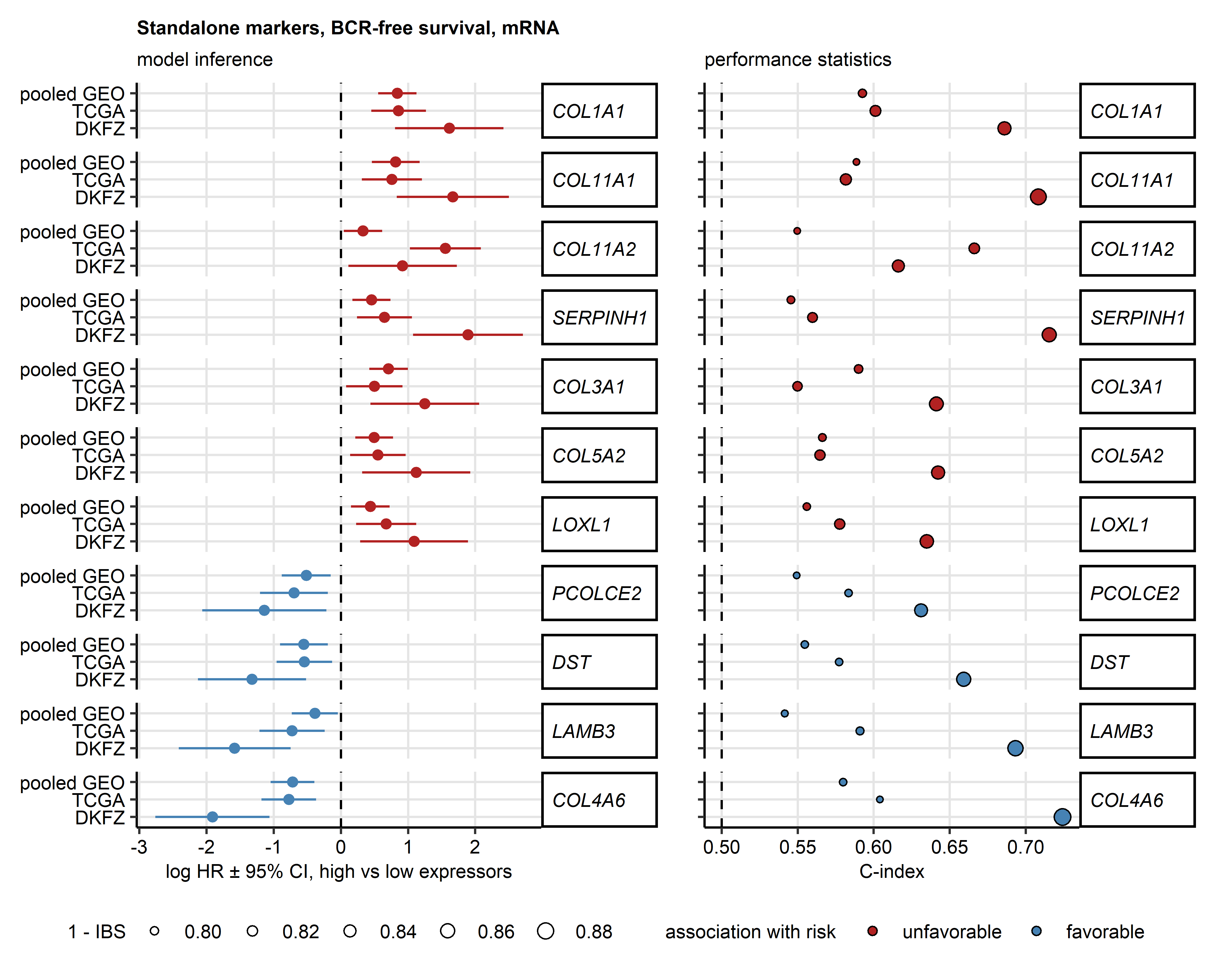


Figure 1: Univariable analysis of BCR-free survival with collagen-related transcripts as candidate standalone prognostic factors.

**Figure 1. Univariable analysis of BCR-free survival with collagen-related transcripts as candidate standalone prognostic factors.**

*Prostate cancer (PCA) patients in the pooled GEO (total: n = 485, biochemical relapse [BCR]: n = 194), TCGA (total: n = 493, BCR: n = 93), and DKFZ cohort (total: n = 105, BCR: n = 24) were classified as high and low expressors for each of the rlength(globals$genes)` collagen-related transcripts with the expression cutoffs corresponding to the largest differences in survival between the high and low expressors.* *BCR risk was compared between the high and low expressors for each gene by univariable Cox proportional hazard regression. P values were corrected for multiple testing with the false discovery rate (FDR) method.* *Standalone BCR risk markers were defined as transcripts associated with unfavorable prognosis (high versus low expressors, hazard ratio [HR] > 1, pFDR < 0.05) or favorable prognosis (HR < 1, pFDR < 0.05) in all three cohorts.*

*HR with 95% confidence intervals for BCR risk in the high versus low expression strata are presented in a Forest plot (left). The model performance statistics, Harrell’s concordance index (C-index, measure of model accuracy), and Integrated Brier Scores (IBS, measure of overall model calibration, low values indicative of good calibration) are presented in a dot plot (right).*

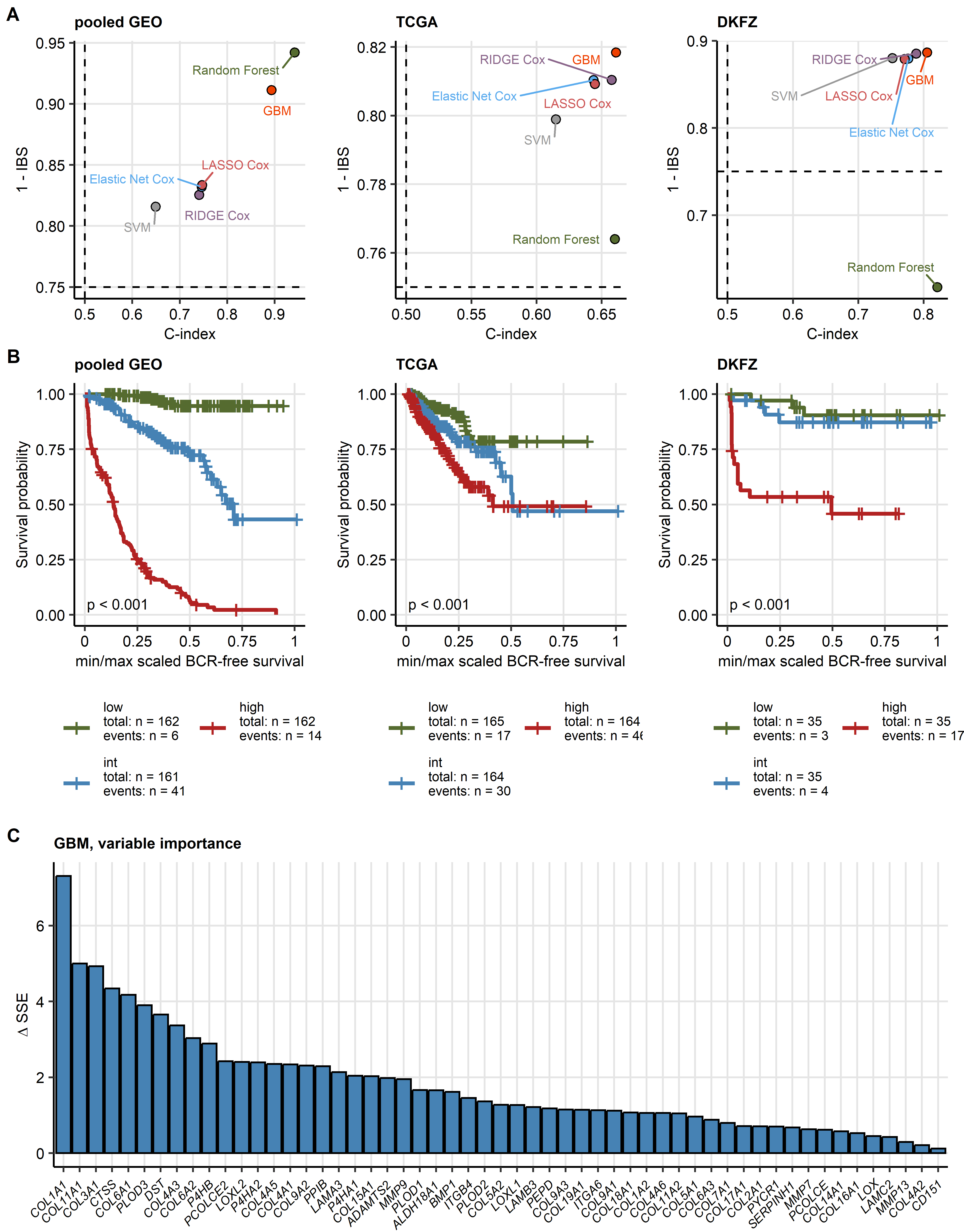


Figure 2: Machine learning modeling of BCR-free survival with expression levels of the collagen-related transcripts as explanatory factors.

**Figure 2. Machine learning modeling of BCR-free survival with expression levels of the collagen-related transcripts as explanatory factors.**

*Risk of biochemical relapse (BCR) in prostate cancer (PCA) was modeled by multi-parameter machine learning models with ComBat-processed -transformed expression levels of 55 collagen-related transcripts as explanatory factors.* *The models were trained in the pooled GEO cohort with six machine learning algorithms (RIDGE Cox regression, Elastic Net Cox regression, LASSO Cox regression, survival Support Vector Machines [SVM], survival Random Forest, and survival Gradient Boosted Machines [GBM]).* *Performance of the models was evaluated in the pooled GEO training cohort (total: n = 485, BCR: n = 194), and the TCGA (total: n = 493, BCR: n = 93) and DKFZ (total: n = 105, BCR: n = 24) validation collectives.*

*(A) Metrics of performance of the survival models, Harrell’s concordance index (C-index, measure of model accuracy) and Integrated Brier Score (IBS, measure of overall model calibration, low values are characteristic for good calibration), are presented in dot plots. Dashed lines represent C-index and IBS values expected for random BCR risk prediction. Note the superior accuracy and calibration of the GBM model in the validation collectives.*

*(B) BCR-free survival in PCA patients stratified by tertiles of the predictor scores of the best performing GBM model. Statistical significance of differences between the predictor score tertiles was determined by Peto-Peto test corrected for multiple testing wit the false discovery rate method. Fractions of BCR-free patients and minimum/maximum scaled time after diagnosis are visualized in Kaplan-Meier plots. Numbers of patients and BCR in the predictor score tertiles are indicated in the legends. P values are displayed in the plots.*

*(C) Importance of the explanatory variables for BCR-risk prediction by the GBM model was gauged by difference in sum of squared error () attributed to inclusion of a particular collagen-related gene in the GBM learner ensemble. High values are characteristic for highly influential variables. values for the variables that contributed substantially to the BCR risk prediction ( > 0) are presented in a bar plot.*

# References