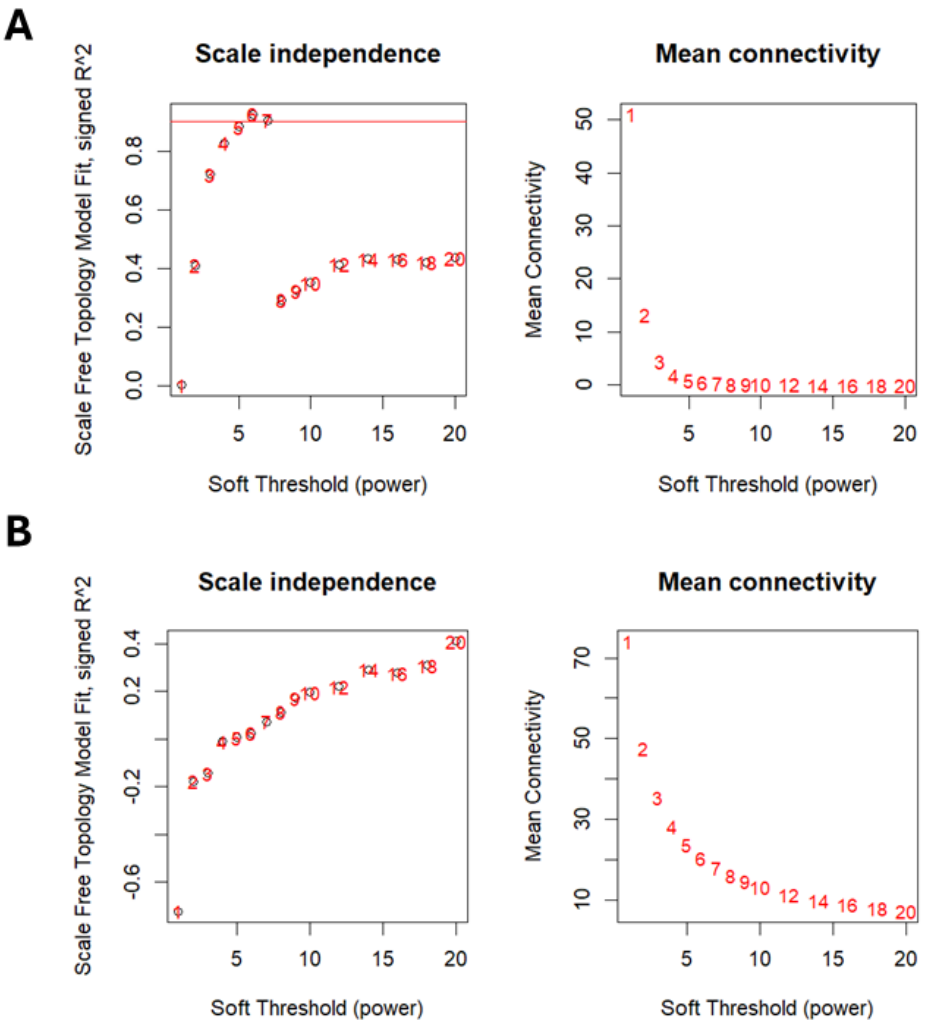
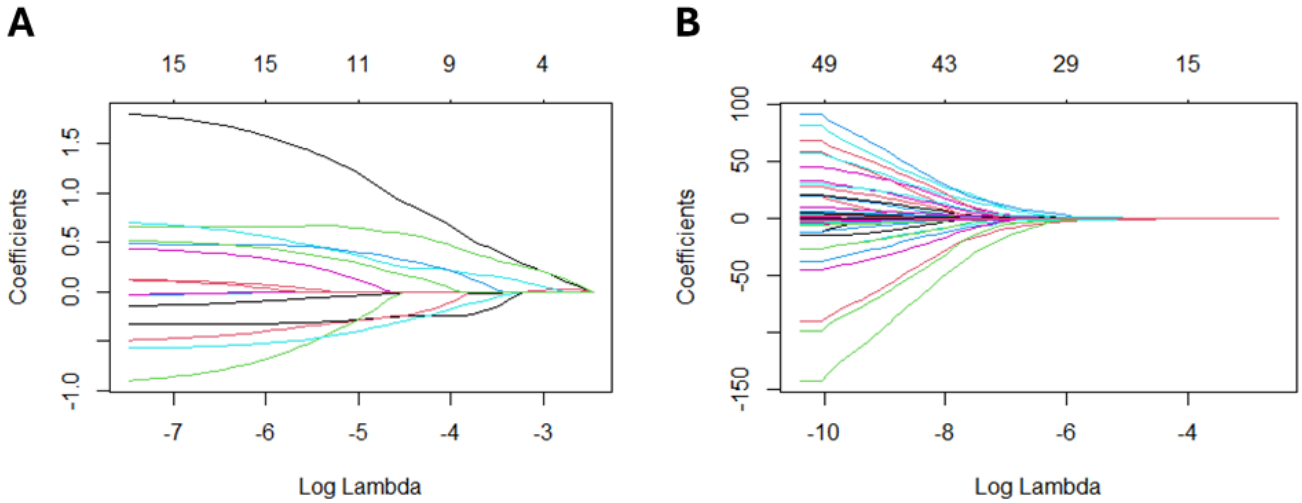
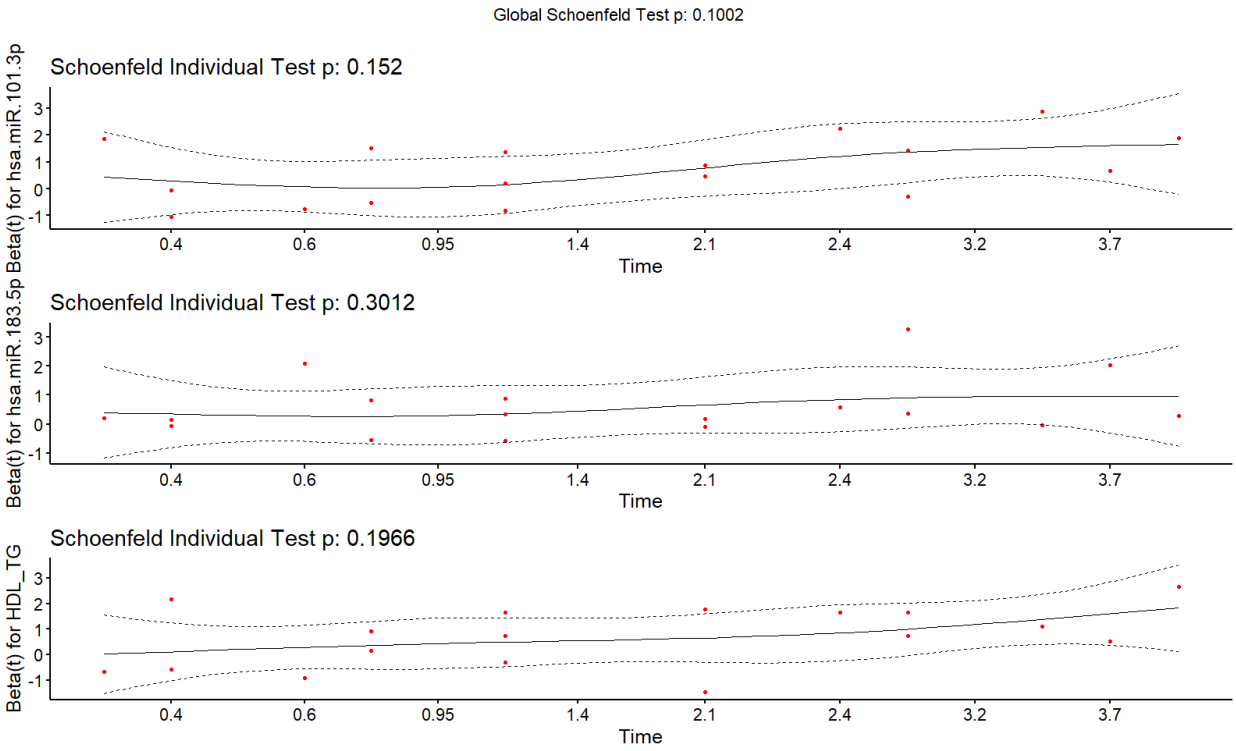
Supplementary Figures



SFig. 1. Soft-thresholding power selection for WGCNA. A) Presents soft-thresholding results for the cmiRs. The left panel shows the scale-free topology model fit (signed R^2) as a function of the soft-thresholding power, with a threshold (0.9) indicated for scale-free topology. The right panel presents mean connectivity as a function of soft-thresholding power. B) Soft-thresholding results of cMets, with similar plots for scale-free topology fit and mean connectivity. A higher soft-thresholding power achieves the desired scale-free network topology while balancing mean connectivity.

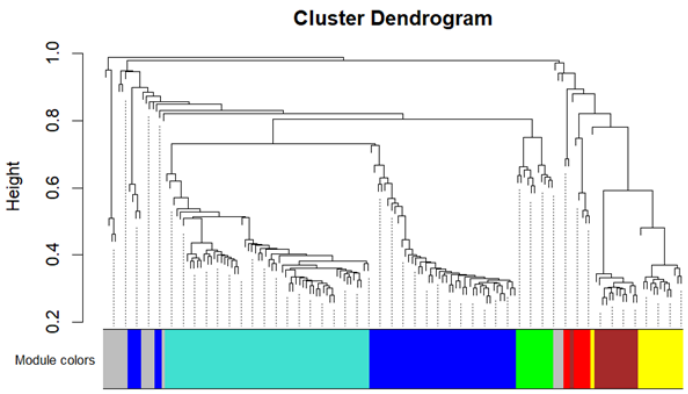


SFig. 2. Coefficients from Lasso Cox regression for selected features. (A) Coefficients for cmiRs as the tuning parameter lambda (λ) varies. (B) Coefficients for cMets as λ varies. The x-axis shows the log-transformed values of λ, with smaller values representing less regularization. The y-axis represents the coefficient values. As λ increases, coefficients move towards zero, leaving only cmiRs and cMets above zero in the predictive model.



SFig. 3. Schoenfeld residuals for testing the proportional hazards assumption in the Cox regression model. Each plot shows the time-dependent residuals for individual predictors. The y-axis represents the beta coefficient for each predictor over time, while the x-axis shows time. Solid lines represent the estimated coefficient trend over time, and dashed lines indicate confidence intervals. The p-values from individual Schoenfeld tests are provided for each predictor, with the global Schoenfeld test result displayed at the top, assessing overall proportionality in the model.

A

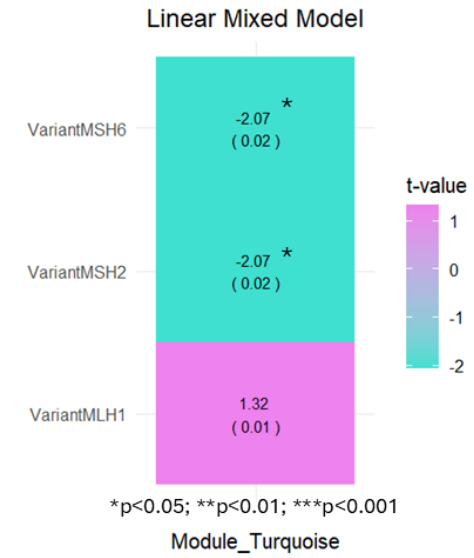


B

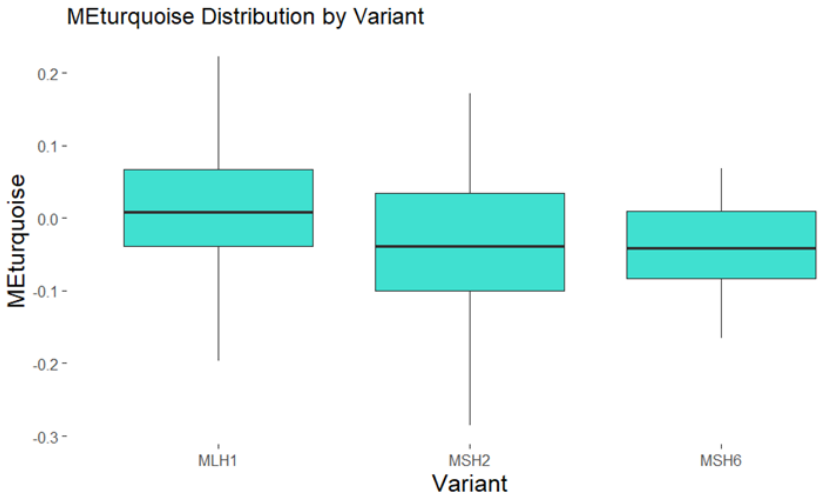
Module	t	df	p value
MEgreen	-1.3006	21.468	0.2072
MEblue	-1.4191	21.131	0.1704
Meturquoise	-0.001115	21.234	0.9991
MEred	1.369	21.48	0.1851
MEbrown	0.65644	23.994	0.5178
MEyellow	-0.64313	21.696	0.5269

SFig. 4. Weighted Correlation Network Analysis (WGCNA) on cMets levels. A) The cluster dendrogram depicting gene modules generated via hierarchical clustering of the expression similarity matrix. B) Using the T-test, the table shows the mean difference between the cMet modules and status groups (Future cancer-Healthy).

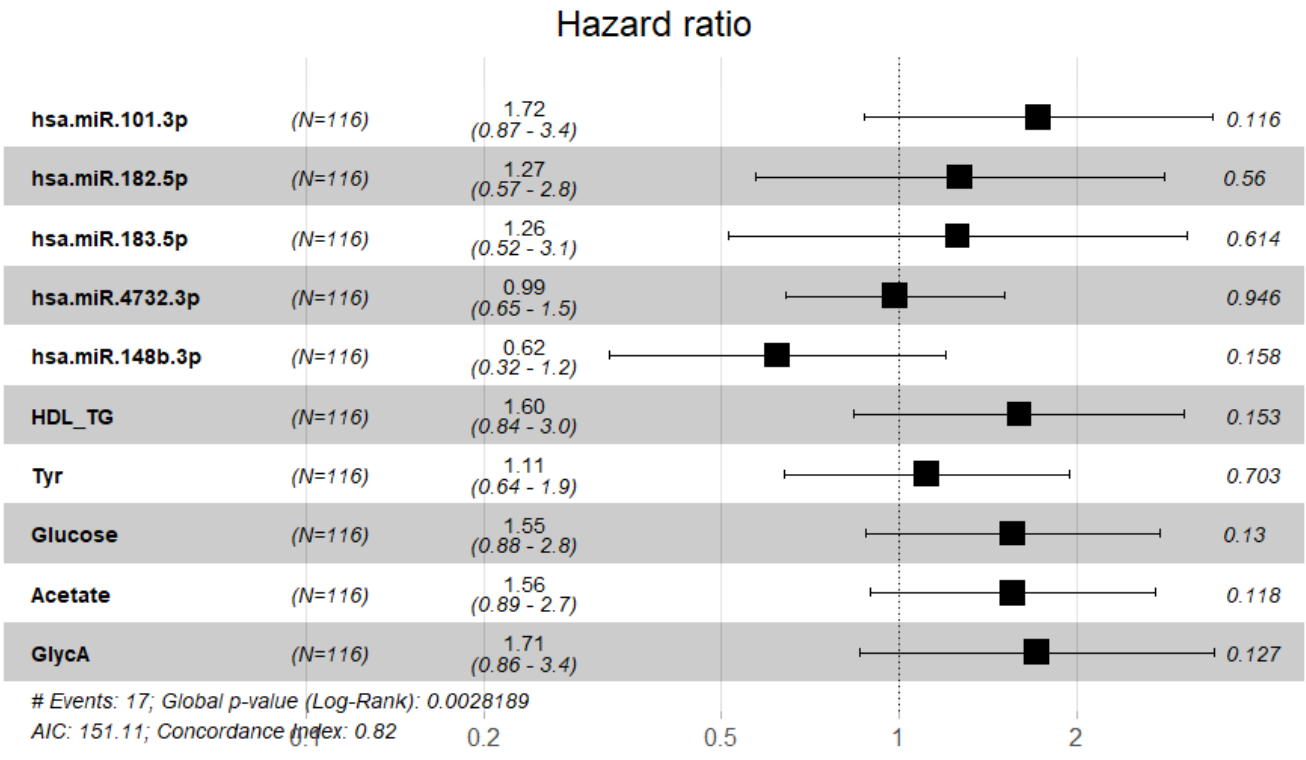
A



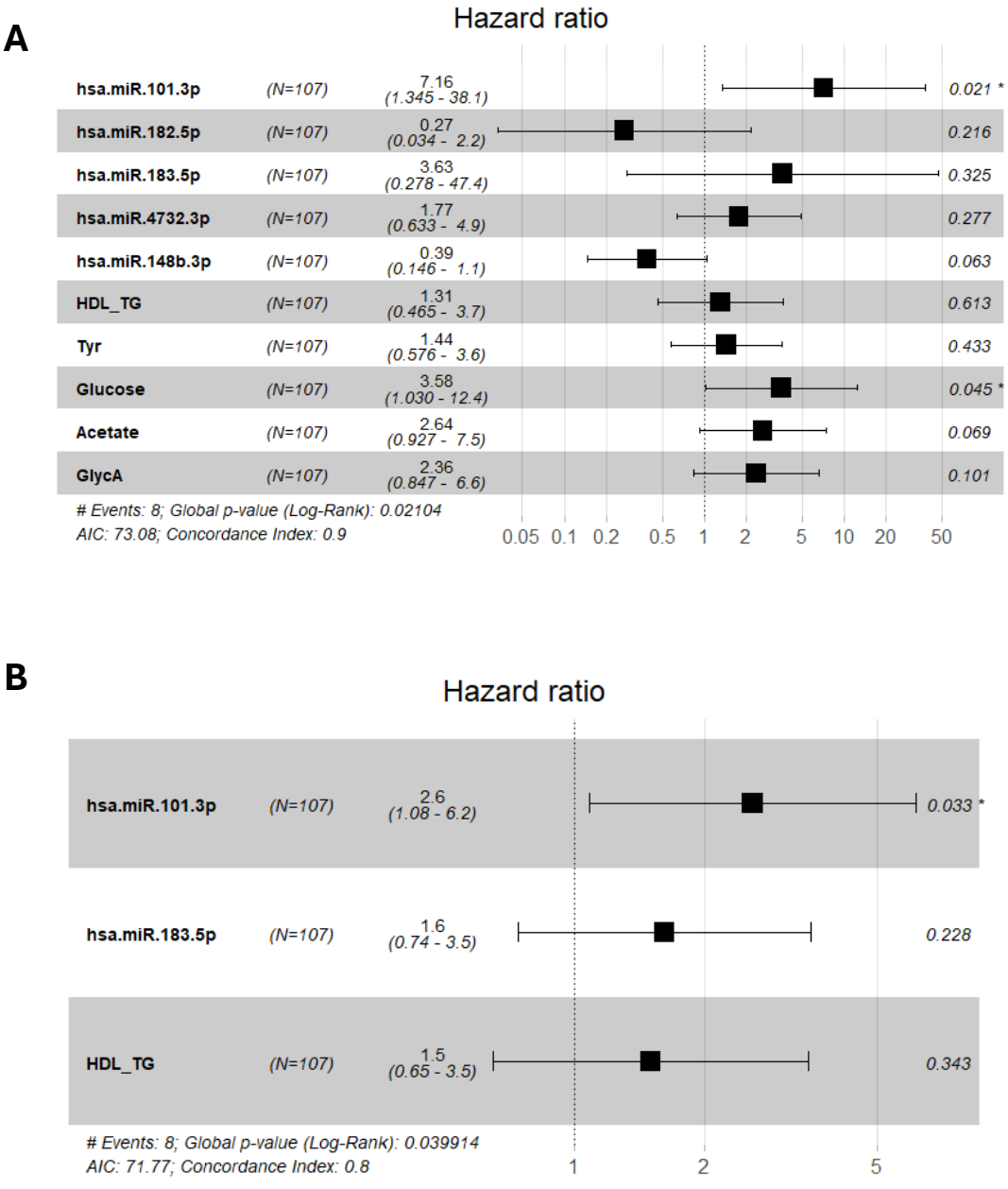
B



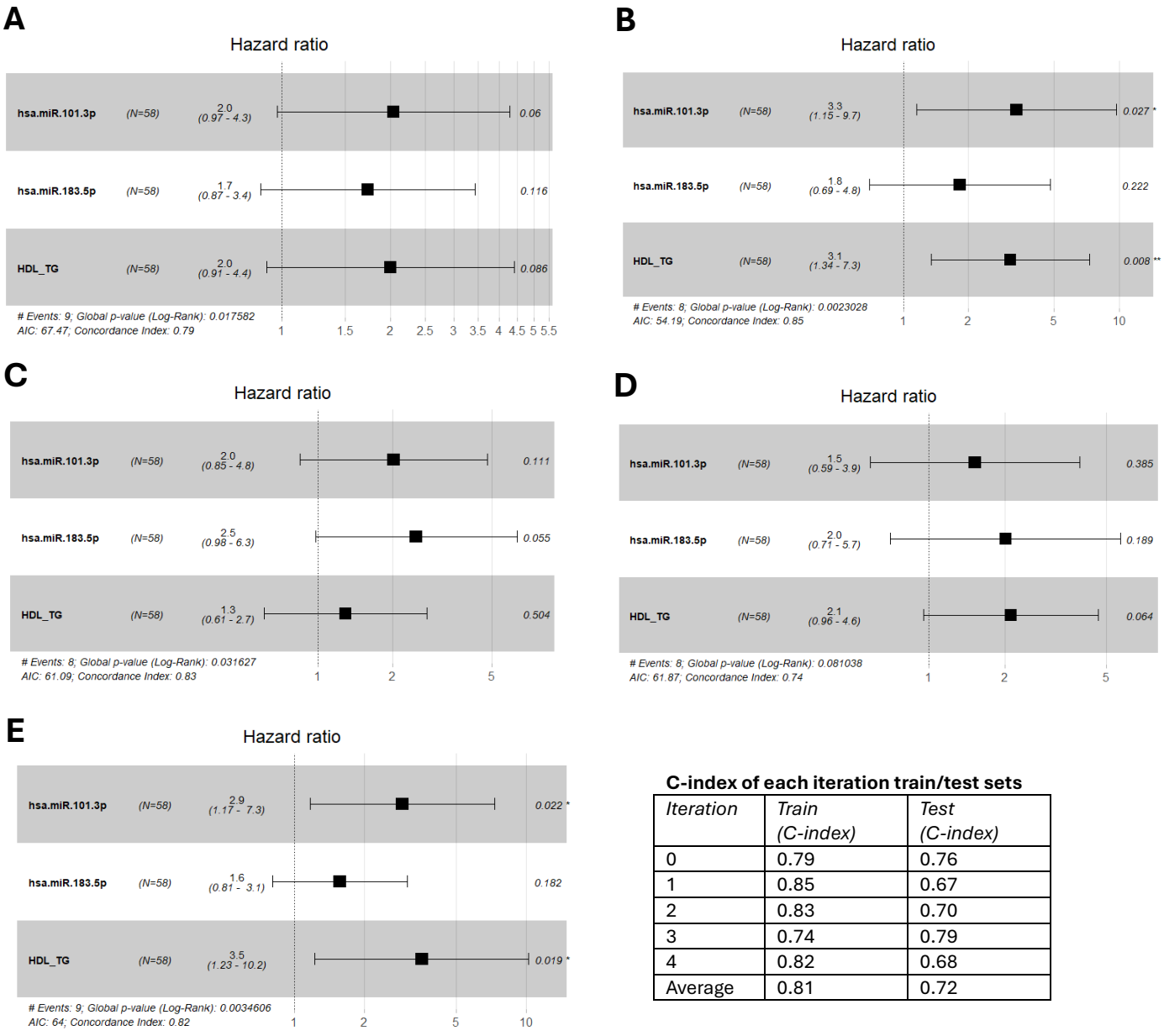
SFig. 5. A) Linear mixed model of the average effect of predictor variables *(path_MMR* variant) on the response variable (MEturquoise) across all levels of the random effects (Age & BMI). Trait-module associations are shown as t-values with (Std. Error), and significance indicators *p<0.05; **p<0.01; ***p<0.001. B) Distribution of the MEturquoise among *path_MMR* variant carriers.



SFig. 6. A forest plot of the Cox regression model fit with 10 predictive biomarkers obtained from Lasso Cox feature selection. The plot shows biomarkers coefficients (estimated HRs), confidence intervals, C-index, and significance. The coefficients above 1 are considered to increase the cancer risk.



SFig. 7. Cox regression model performance on CRC cases only using A) 10 predictive biomarkers and B) 3 predictive biomarkers.



C-index of each iteration train/test sets

Iteration	*Train (C-index)*	*Test (C-index)*
0	0.79	0.76
1	0.85	0.67
2	0.83	0.70
3	0.74	0.79
4	0.82	0.68
Average	0.81	0.72

SFig. 8. Forest plots of Cox Proportional Hazards Model from 5 iterations and model's predictive accuracy of A) 0, B) 1st, C) 2nd, D) 3rd, and E) 4th iteration presenting the model (trained with corresponding train data) coefficients, CIs, C-index, and significance. The table presents each iteration's predictive accuracy (C-index) on corresponding test datasets.