

Title: Adaptive prediction model in prospective molecular-signature-based clinical studies

Authors: Guanghua Xiao¹, Shuangge Ma², John Minna^{4,5,6}, Yang Xie^{1,3}

Supplementary Table 1.

	X ₁ , ..., X ₁₀	X ₁₁ , ..., X ₂₀	X ₂₁ , ..., X ₃₀	X ₃₁ , ..., X ₅₀
Training set	√	√	-	-
Testing set	√	-	√	-

Table S1. Simulation settings. The variable with a check mark is a predictor, otherwise is not. It indicates the X₁,...,X₁₀ are good predictors and X₁₁,...,X₂₀ are unstable predictors.

Supplementary Table 2.

AffyID	Accession	Symbol	Cancer
219049_at	NM_018371	ChGn	Prostate cancer
204995_at	NM_003885	CDK5R1	Gangliogliomas, Prostate cancer
205668_at	NM_002349	LY75(CD205)	Lymphoma, thymoma
201109_s_at	NM_003246	THBS1	non-small cell lung cancer, carcinoma, neuron blastoma, colorectal cancer, melanoma
213156_at	BC035518		
227345_at	NM_003840	TNFRSF10D	Breast cancer, Acutemyeloid leukemia, Neuro blastoma, Prostate cancer
205808_at	NM_004318	ASPH	Hepatocellular carcinoma, Bileduct cancer
216379_x_at	NM_013230	CD24	Non-small cell lung cancer, Ovarian cancer, Colorectal cancer, Carcinoma, Breast cancer
203603_s_at	NM_014795	ZFHX1B (SIP1,ZEB2)	Breast cancer
226142_at	AC022507		
209309_at	NM_001185	AZGP1	Prostate cancer
221911_at	NM_004956	ETV1	Prostate cancer
224973_at	AF350451	FAM46A	Colorectal cancer

Table S2. Biological verification of genes with high adaptive scores. Among the top 13 genes with adaptive scores higher than 30, eleven of them have been shown to be important cancer related genes in other studies.

Supplementary Figure 1.

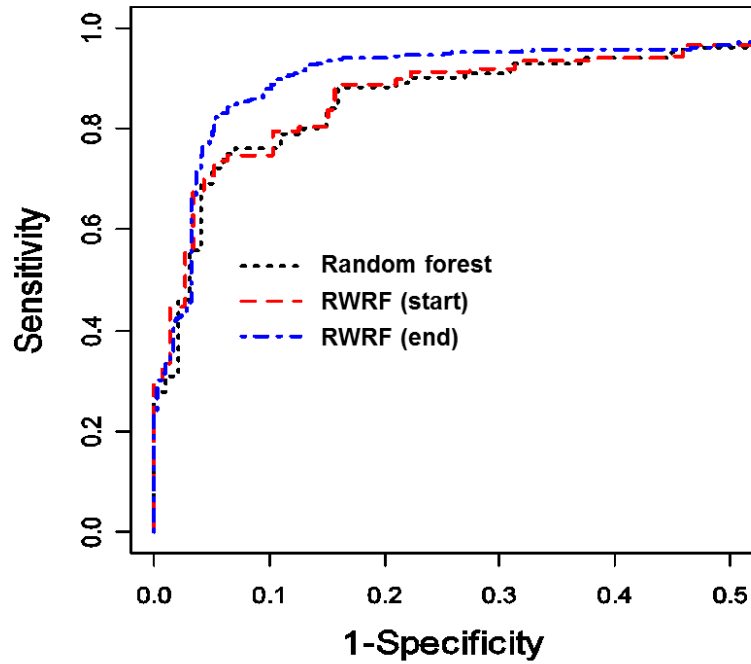


Figure S1. ROC curves in the simulation study. The black, red and blue lines represent the standard RF, the starting performance of re-weighted RF and ending performance of re-weighted RF. As expected, the starting performance of re-weighted RF is similar as the standard RF and prediction performance has improved significantly at the end of the study.