

Supplementary information

**Improved prediction of immune
checkpoint blockade efficacy across
multiple cancer types**

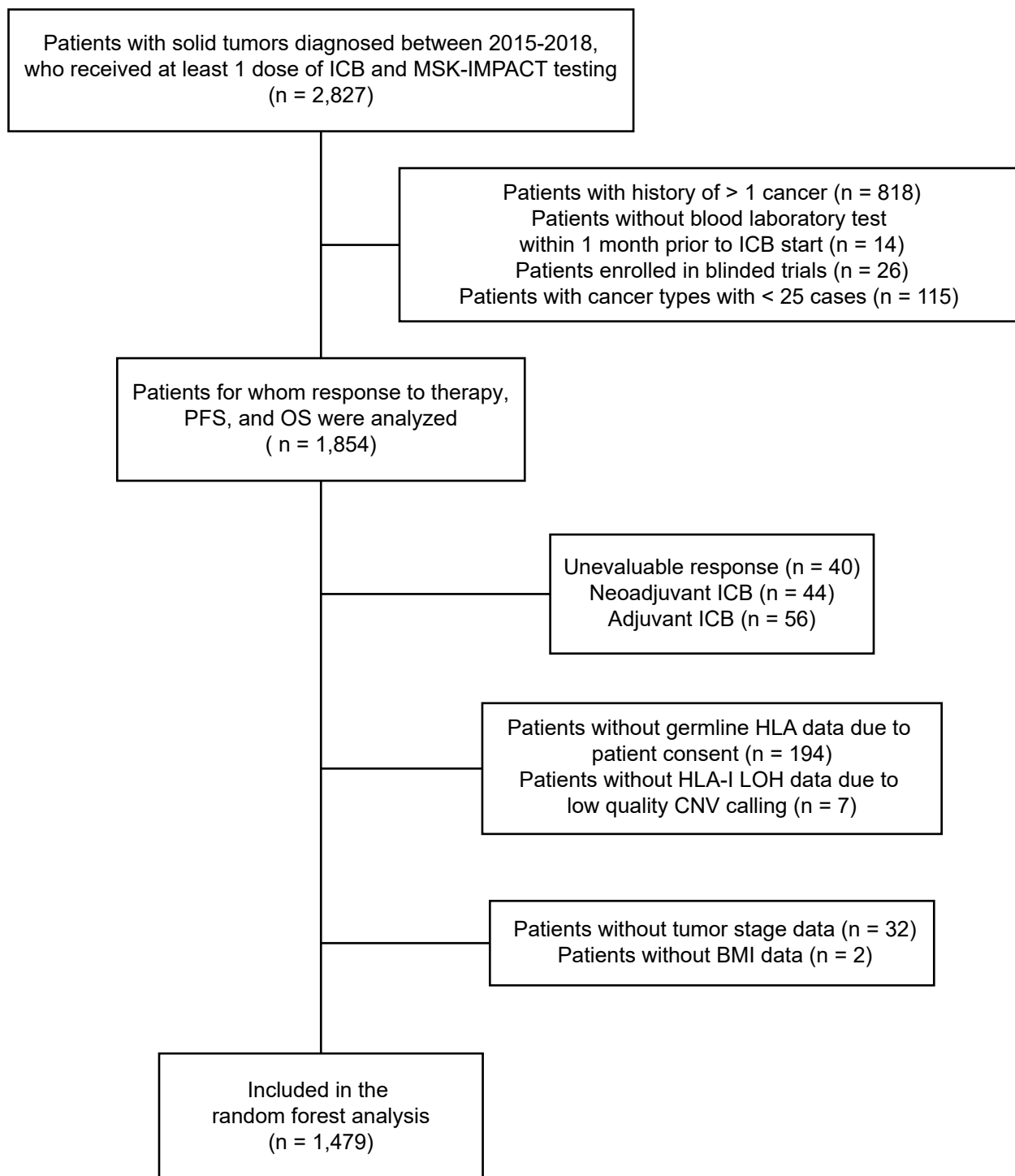
In the format provided by the
authors and unedited

Supplementary Information for

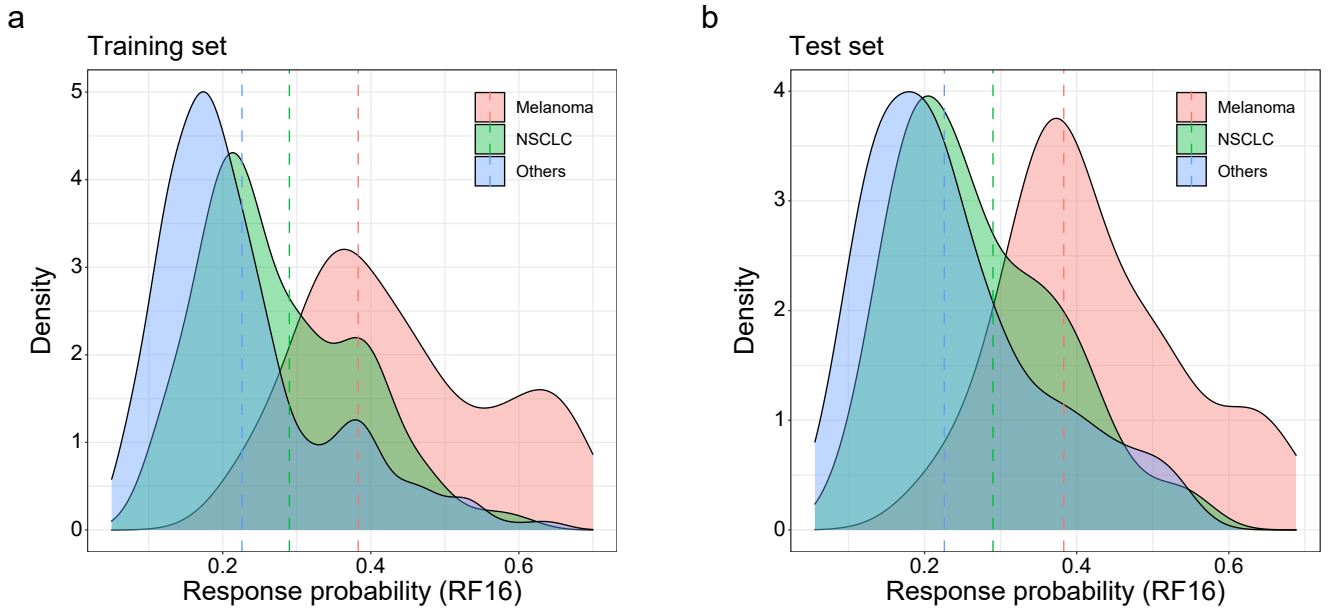
**Improved prediction of immune checkpoint blockade efficacy
across multiple cancer types**

Diego Chowell, Seong-Keun Yoo, Cristina Valero, Alessandro Pastore, Chirag Krishna, Mark Lee, Douglas Hoen, Hongyu Shi, Daniel W. Kelly, Neal Patel, Vladimir Makarov, Xiaoxiao Ma, Lynda Vuong, Erich Y. Sabio, Kate Weiss, Fengshen Kuo, Tobias L. Lenz, Robert M. Samstein, Nadeem Riaz, Prasad S. Adusumilli, Vinod P. Balachandran, George Plitas, A. Ari Hakimi, Omar Abdel-Wahab, Alexander N. Shoushtari, Michael A. Postow, Robert J. Motzer, Marc Ladanyi, Ahmet Zehir, Michael F. Berger, Mithat Gönen, Luc G.T. Morris, Nils Weinhold, and Timothy A. Chan

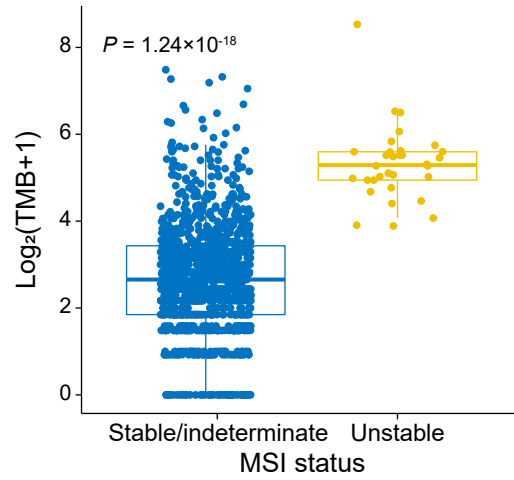
Correspondence to
Luc G.T. Morris (morrisl@mskcc.org)
Nils Weinhold (weinholn@mskcc.org)
Timothy A. Chan (CHANT2@ccf.org)



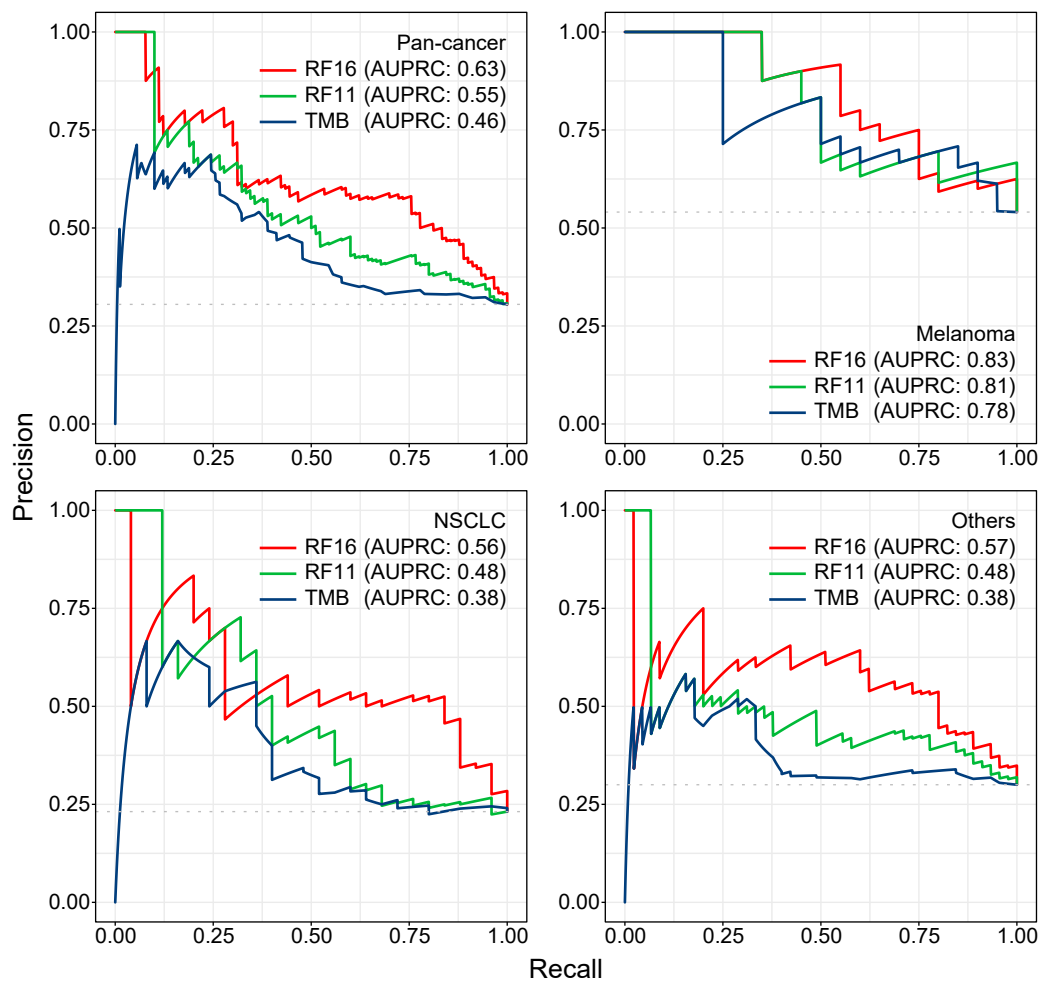
Supplementary Fig 1. Flow diagram showing sample collection and filtering process.



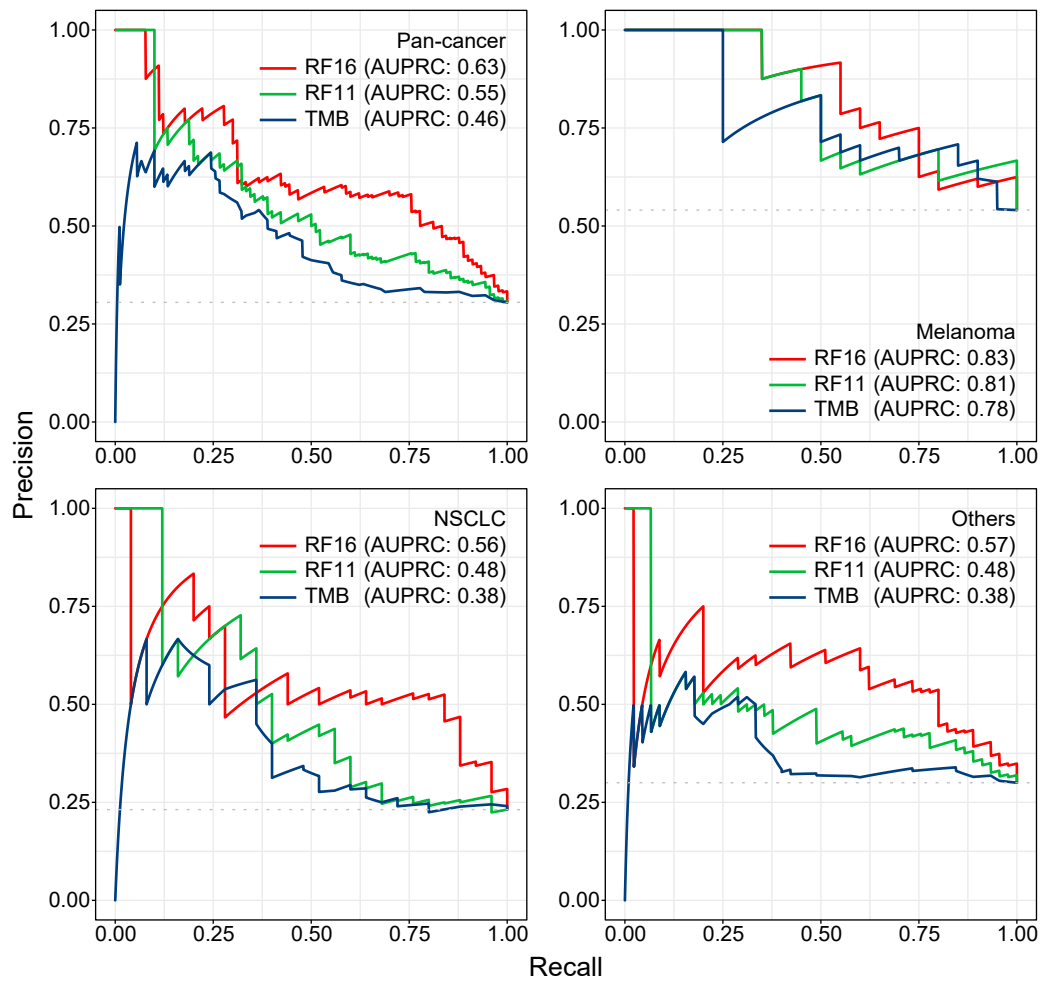
Supplementary Fig 2. Distribution of response probabilities calculated by the integrated model in each cancer group. Density plots showing response probability values (a) in the training set and (b) in the test set. The vertical lines denote the cancer-specific group optimal cut-points identified in the training set.



Supplementary Fig 3. Association between TMB and MSI status. Two-sided Mann-Whitney U test P -value is presented. Center bar, median; box, interquartile range; whiskers, first and third quartiles $\pm 1.5 \times$ interquartile range.



Supplementary Fig 4. Model performance illustrated by the precision-recall curve across multiple cancer types in the training set. The corresponding AUPRC values of RF16, RF11, and TMB are shown in pan-cancer and each cancer group [melanoma, NSCLC, and others (not melanoma/NSCLC)].




Supplementary Fig 5. Model performance illustrated by the precision-recall curve across multiple cancer types in the test set. The corresponding AUPRC values of RF16, RF11, and TMB are shown in pan-cancer and each cancer group [melanoma, NSCLC, and others (not melanoma/NSCLC)].

a

Training set


Feature	AUC				AUPRC			
	Pan-cancer	Melanoma	NSCLC	Others	Pan-cancer	Melanoma	NSCLC	Others
Age	0.53	0.61	0.46	0.53	0.31	0.52	0.28	0.25
Albumin	0.61	0.62	0.56	0.62	0.35	0.57	0.31	0.30
BMI	0.55	0.50	0.51	0.57	0.30	0.46	0.29	0.25
Chemo prior ICB	0.63	0.53	0.57	0.62	0.37	0.46	0.33	0.35
Drug	0.53	0.53	0.52	0.52	0.30	0.46	0.30	0.23
FCNA	0.53	0.51	0.55	0.54	0.29	0.43	0.33	0.24
HED	0.52	0.50	0.49	0.55	0.29	0.46	0.29	0.26
HGB	0.59	0.53	0.56	0.58	0.34	0.46	0.31	0.30
LOH in HLA-I	0.51	0.52	0.53	0.50	0.27	0.47	0.30	0.22
MSI	0.52	0.50	0.50	0.55	0.30	0.44	0.27	0.30
NLR	0.55	0.47	0.57	0.54	0.28	0.39	0.32	0.22
Platelets	0.51	0.49	0.55	0.50	0.27	0.43	0.32	0.21
RF16	0.85	0.82	0.83	0.89	0.67	0.81	0.66	0.66
Sex	0.52	0.54	0.49	0.53	0.28	0.46	0.27	0.23
Stage	0.49	0.47	0.50	0.50	0.26	0.43	0.28	0.22
TMB	0.62	0.68	0.57	0.60	0.43	0.65	0.39	0.34

0.2  0.8

b

Test set

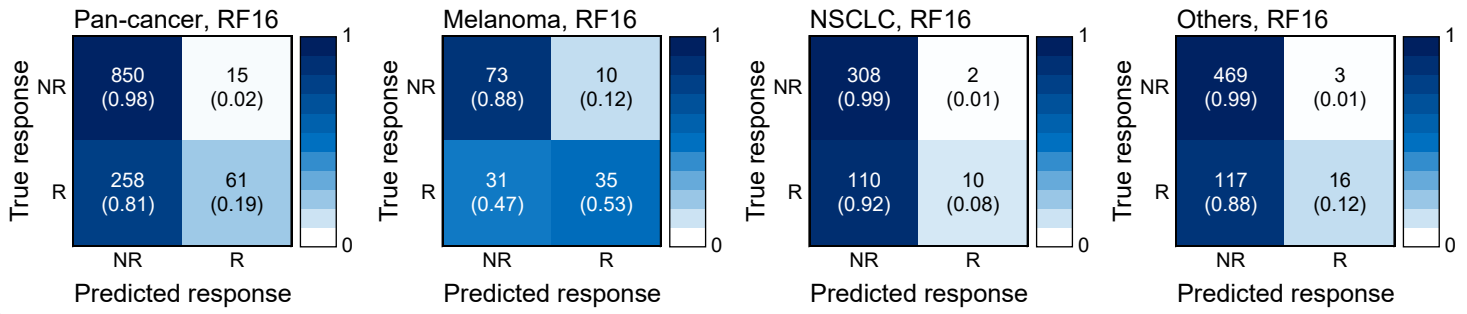
Feature	AUC				AUPRC			
	Pan-cancer	Melanoma	NSCLC	Others	Pan-cancer	Melanoma	NSCLC	Others
Age	0.49	0.54	0.51	0.48	0.31	0.57	0.26	0.28
Albumin	0.63	0.53	0.66	0.60	0.38	0.51	0.32	0.38
BMI	0.56	0.65	0.52	0.52	0.35	0.64	0.27	0.32
Chemo prior ICB	0.70	0.56	0.71	0.66	0.50	0.57	0.41	0.51
Drug	0.54	0.43	0.53	0.52	0.35	0.50	0.26	0.33
FCNA	0.55	0.81	0.49	0.57	0.32	0.72	0.22	0.35
HED	0.53	0.49	0.61	0.47	0.32	0.55	0.33	0.27
HGB	0.68	0.54	0.65	0.69	0.45	0.56	0.37	0.46
LOH in HLA-I	0.48	0.44	0.46	0.52	0.29	0.48	0.21	0.32
MSI	0.52	0.50	0.50	0.55	0.35	0.54	0.23	0.37
NLR	0.65	0.60	0.62	0.65	0.38	0.56	0.30	0.37
Platelets	0.52	0.44	0.48	0.58	0.31	0.50	0.25	0.34
RF16	0.79	0.78	0.82	0.79	0.63	0.83	0.56	0.57
Sex	0.50	0.46	0.49	0.50	0.30	0.51	0.23	0.30
Stage	0.49	0.48	0.48	0.51	0.30	0.53	0.23	0.30
TMB	0.63	0.74	0.61	0.58	0.46	0.78	0.38	0.38

0.2  0.8

Supplementary Fig 6. Comparison of each model feature performance with that from the integrated RF16 model across multiple cancer types. AUC and AUPRC of each feature are shown in pan-cancer and each cancer group [melanoma, NSCLC, and others (not melanoma/NSCLC)] in the **(a)** training set and in the **(b)** test set. None of RF16's features alone could achieve the level of performance achieved by RF16.

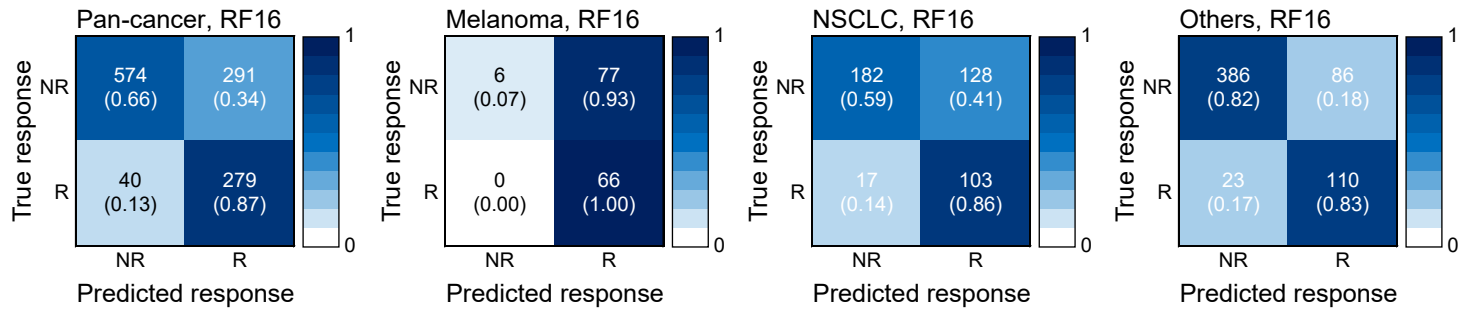
a

RF16 with 0.5 cutpoint



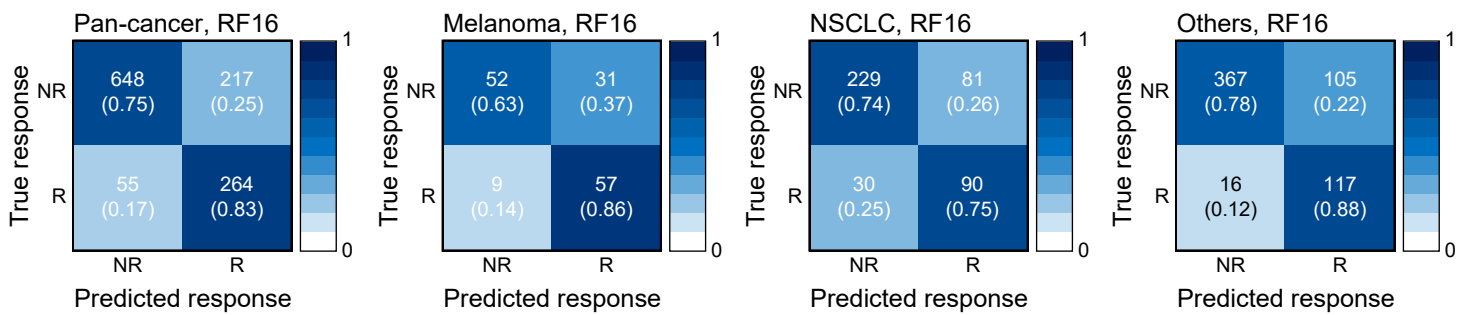
b

RF16 with pan-cancer cutpoint

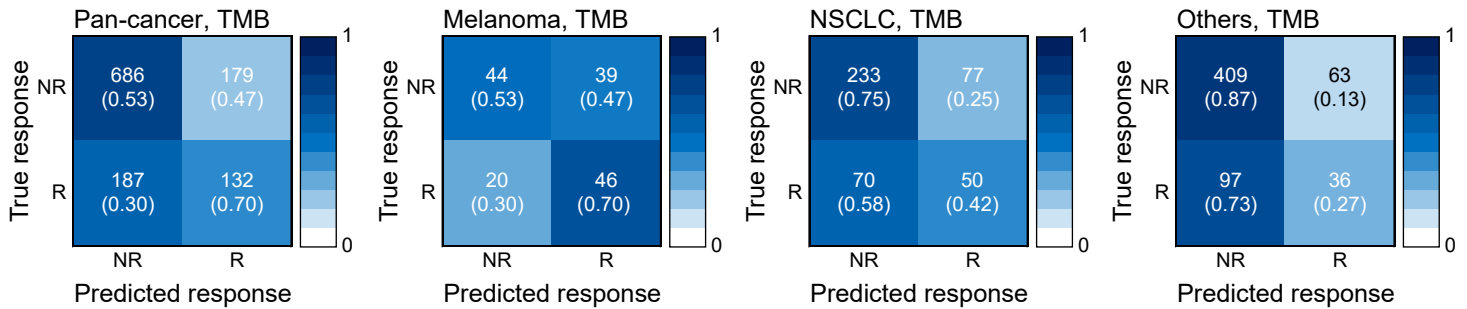


c

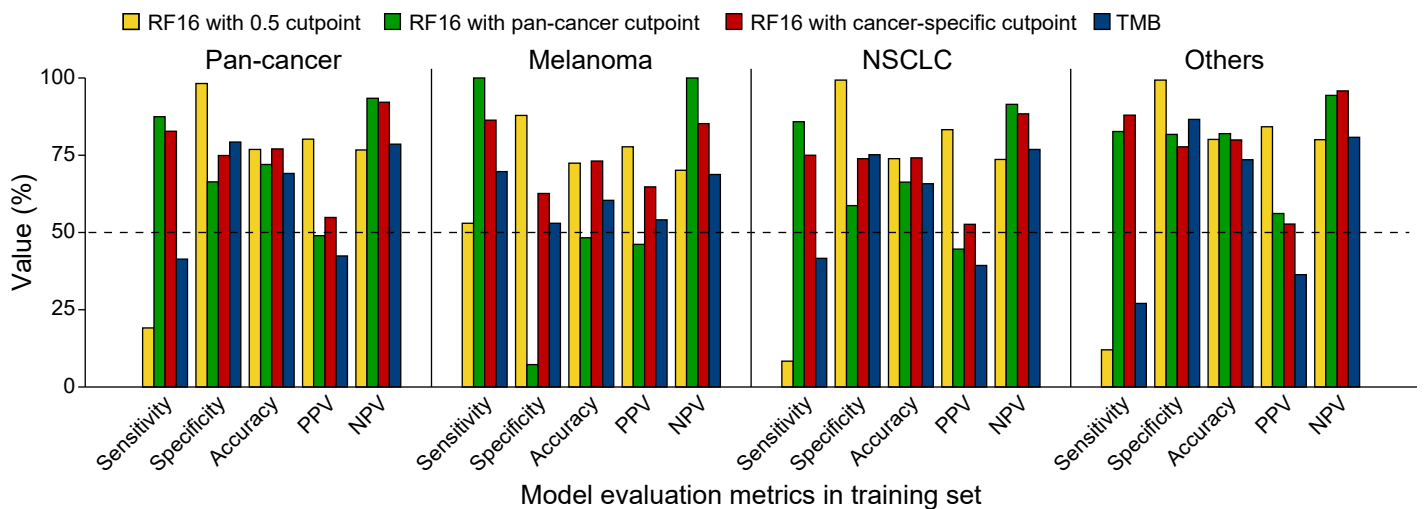
RF16 with cancer-specific cutpoint



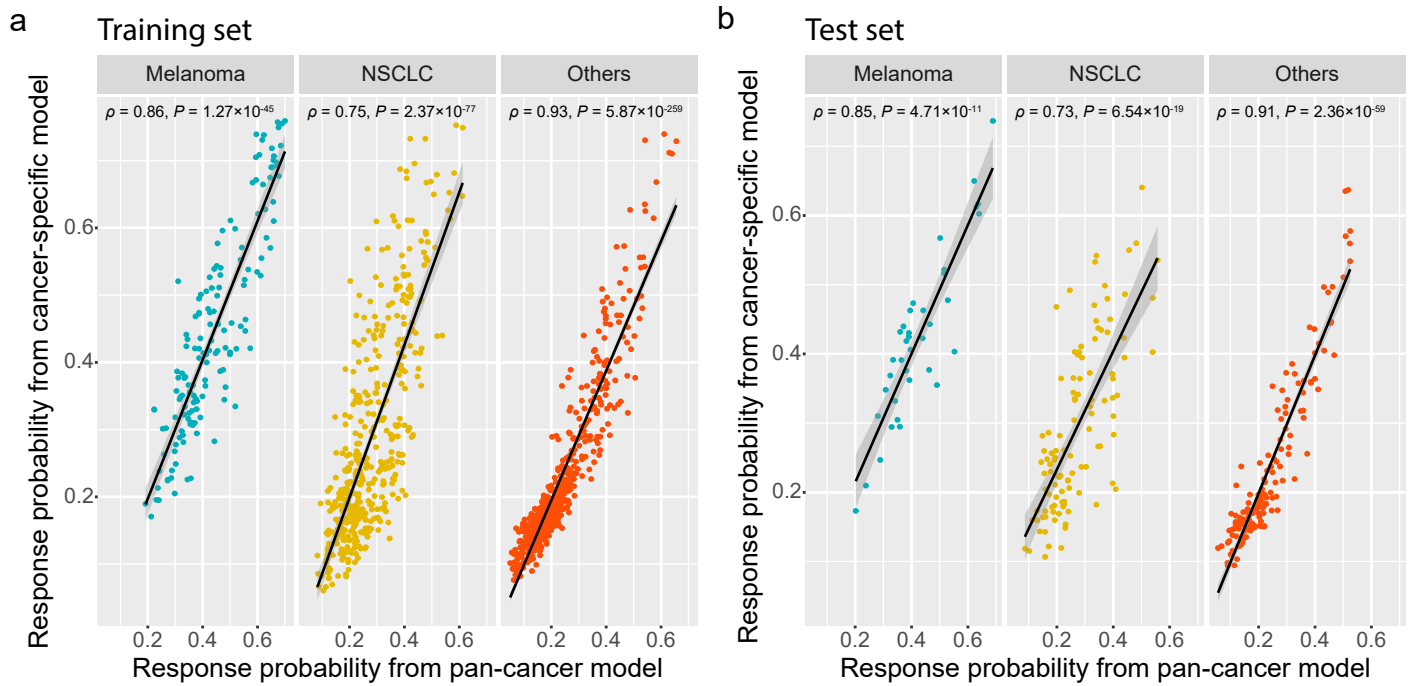
d



e



Supplementary Fig 7. Model development in the training set. (a) Confusion matrices showing suboptimal response outcomes predicted by RF16 in pan-cancer and each cancer group [melanoma, NSCLC, and others (not melanoma/NSCLC)] using the default cutoff of 0.5 which was used during cross-validation to obtain the best model based on the highest average accuracy. **(b)** Confusion matrices showing suboptimal response outcomes predicted by RF16 in pan-cancer and each cancer group [melanoma, NSCLC, and others (not melanoma/NSCLC)] using a single pan-cancer optimal threshold based on the ROC curve. **(c)** Confusion matrices showing improved response predictions using RF16 in each cancer group using three different optimal thresholds derived from each cancer-group-specific ROC curve (Fig.1d). **(d)** Confusion matrices showing response outcomes predicted by TMB in each cancer group. To define high TMB tumors, we applied the threshold of ≥ 10 mut/Mb5. **(e)** Performance measurements of the model and TMB illustrated by sensitivity, specificity, accuracy, PPV, and NPV.



Supplementary Fig 8. Comparison of the response probabilities computed from RF16 model trained on pan-cancer data with those from separate models trained on cancer-specific data [melanoma, NSCLC, and others (not melanoma/NSCLC)]. Correlation plots of response probabilities calculated by the pan-cancer RF16 model and those calculated by separate cancer-specific models in the **(a)** training set and in the **(b)** test set. The others group is composed of 14 different cancer types (Table 1). Spearman's ρ and two-sided P -values are presented.

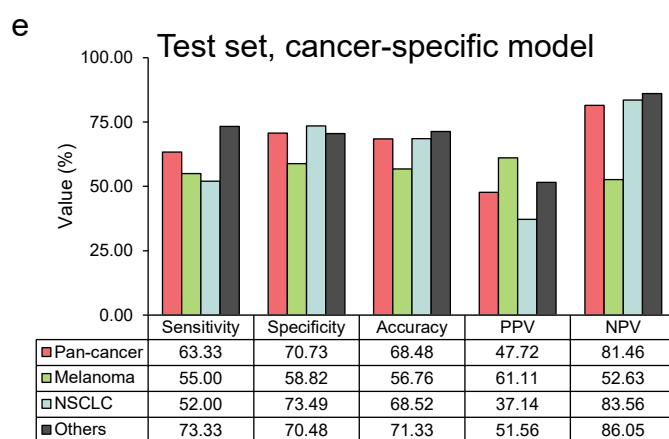
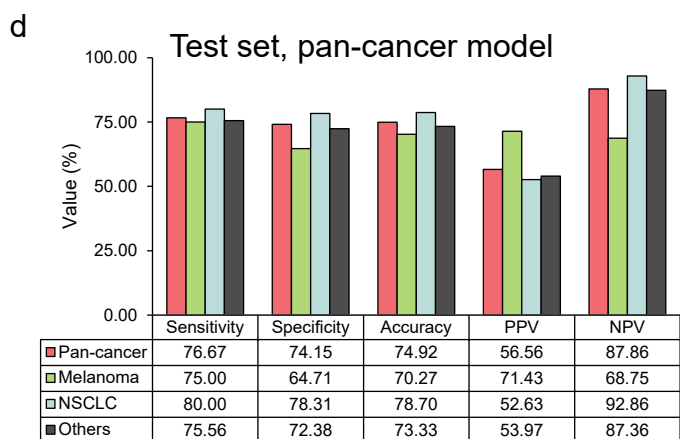
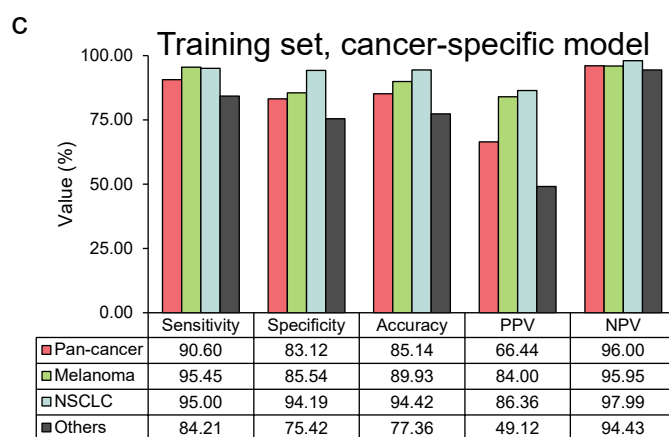
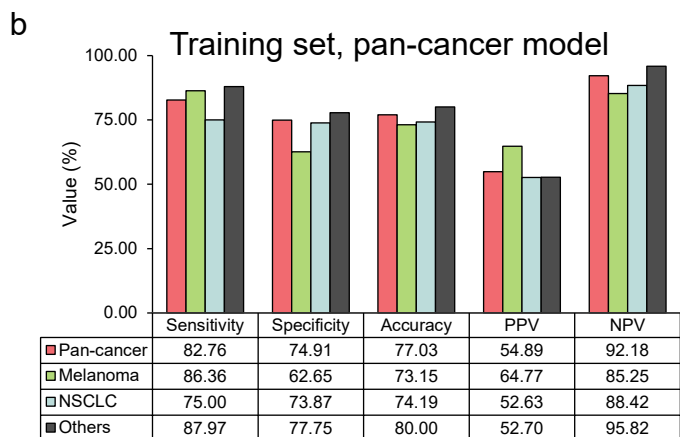
a

Pan-cancer model

Method	Training			Test		
	Melanoma	NSCLC	Others	Melanoma	NSCLC	Others
AUC	0.82	0.83	0.89	0.78	0.82	0.79
AUPRC	0.81	0.66	0.66	0.83	0.56	0.57

Cancer-specific model

Method	Training			Test		
	Melanoma	NSCLC	Others	Melanoma	NSCLC	Others
AUC	0.96	0.98	0.88	0.66	0.72	0.77
AUPRC	0.95	0.96	0.71	0.75	0.53	0.58



Supplementary Fig 9. Comparison of the predictive performance of the RF16 model trained on pan-cancer with those from separate models trained on cancer-specific data [melanoma, NSCLC, and others (not melanoma/NSCLC)]. (a) Comparison of the AUC and AUPRC values for RF16 trained on pan-cancer data with those from the separate cancer-specific models in the training and test set. Performance of the (b) pan-cancer RF16 model and the (c) cancer-specific model in the training set. Performance of the (d) pan-cancer RF16 model and the (e) cancer-specific model in the test set. In (c) and (e), the pan-cancer row represents the pan-cancer performances, which were calculated by aggregating the prediction results of melanoma, NSCLC, and others. To calculate the different performance metrics of the cancer-specific models presented in (c) and (e), we found the optimal cutoff on the ROC curve in the training set for melanoma, NSCLC, and Others, separately. The cutoffs are: 0.420 for melanoma; 0.341 for NSCLC; and 0.197 for Others. The others group is composed of 14 different cancer types (Table 1).

a

Training set

Cancer type	RF16		Logistic regression	
	AUC	AUPRC	AUC	AUPRC
Pan-cancer	0.85	0.67	0.72	0.51
Melanoma	0.82	0.81	0.74	0.71
NSCLC	0.83	0.66	0.66	0.45
Others	0.89	0.66	0.72	0.43

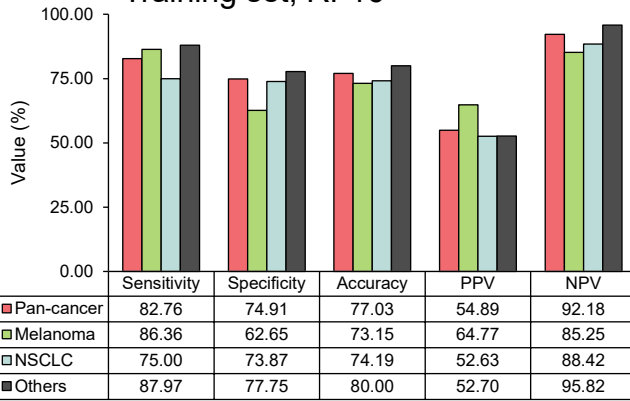
b

Test set

Cancer type	RF16		Logistic regression	
	AUC	AUPRC	AUC	AUPRC
Pan-cancer	0.79	0.63	0.78	0.61
Melanoma	0.78	0.83	0.81	0.87
NSCLC	0.82	0.56	0.77	0.47
Others	0.79	0.57	0.78	0.55

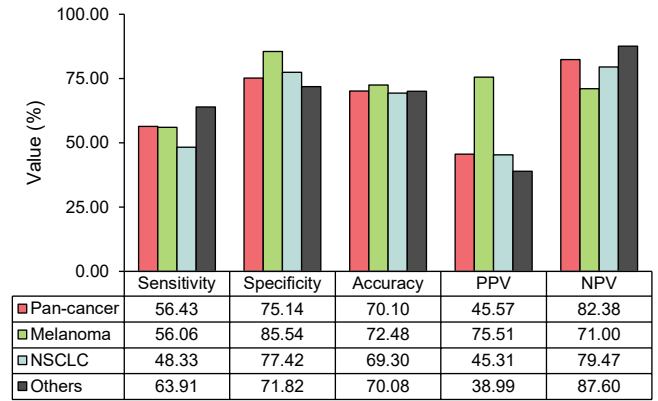
c

Training set, RF16



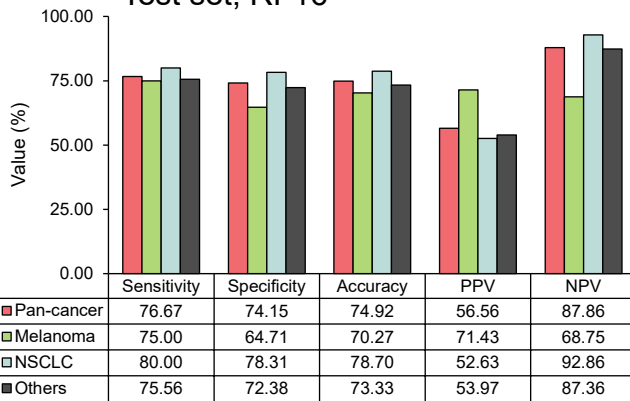
d

Training set, Logistic regression



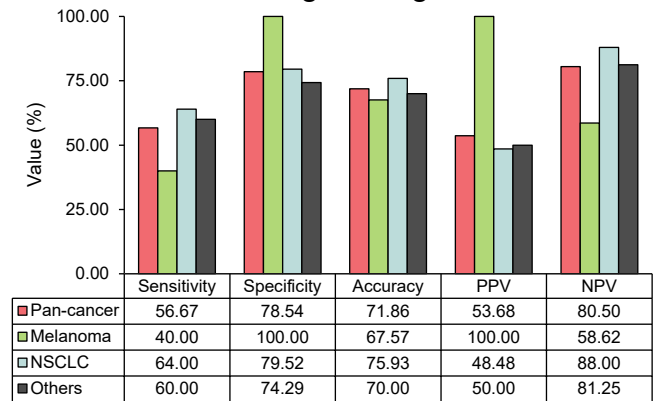
e

Test set, RF16

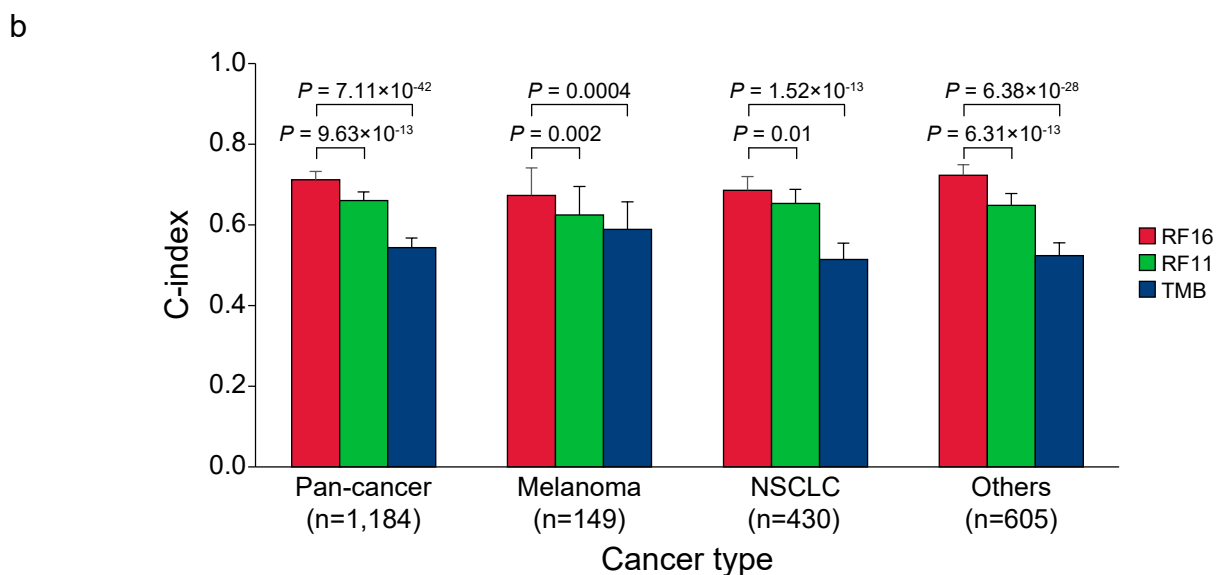
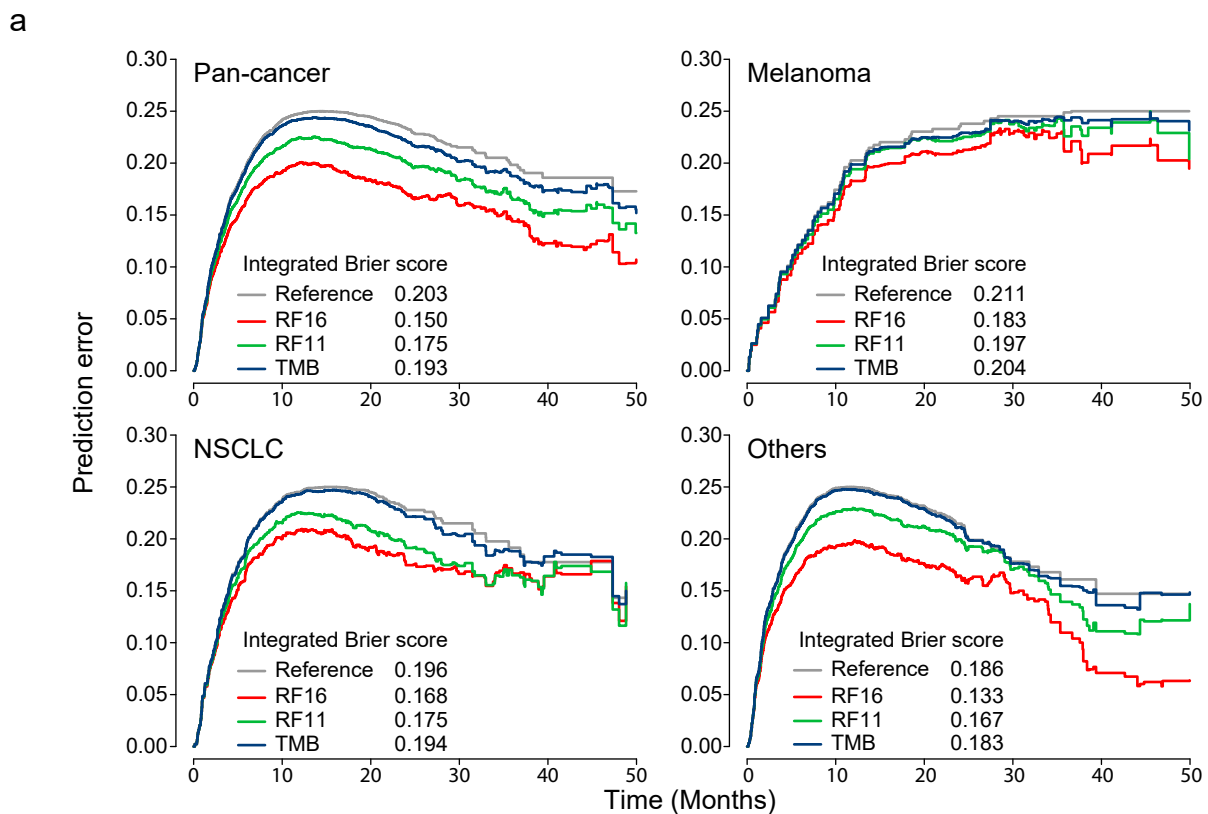


f

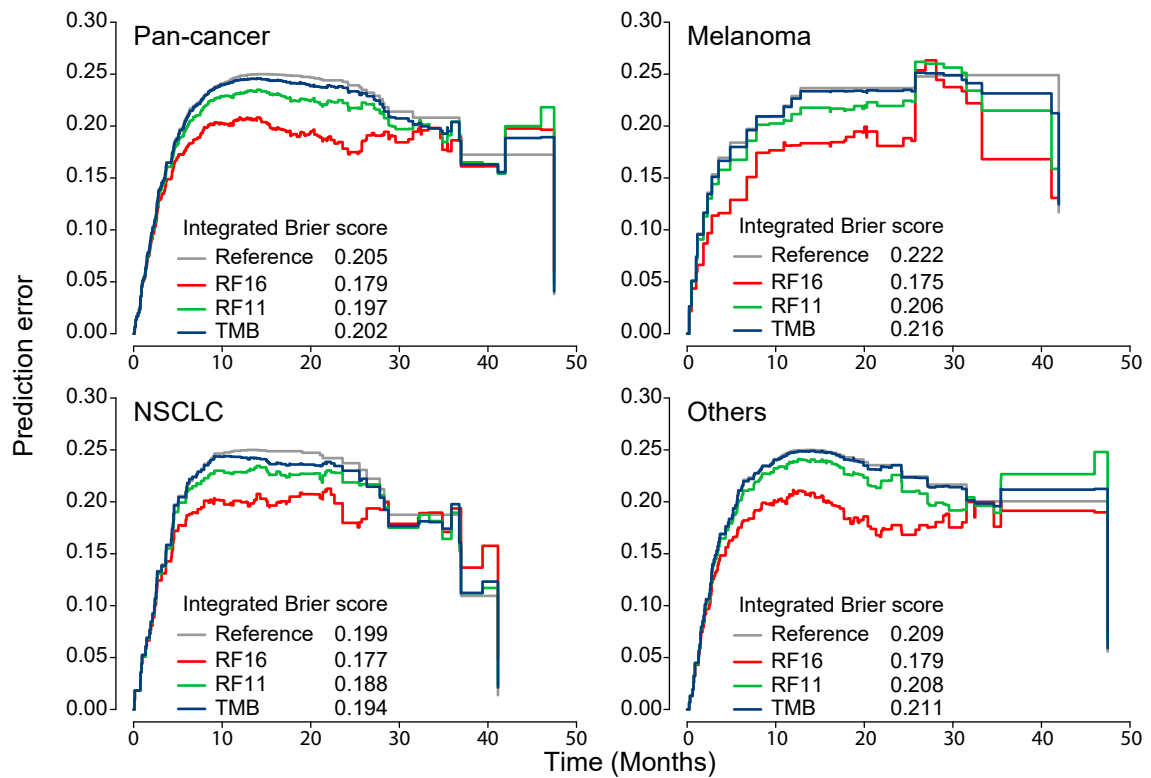
Test set, Logistic regression



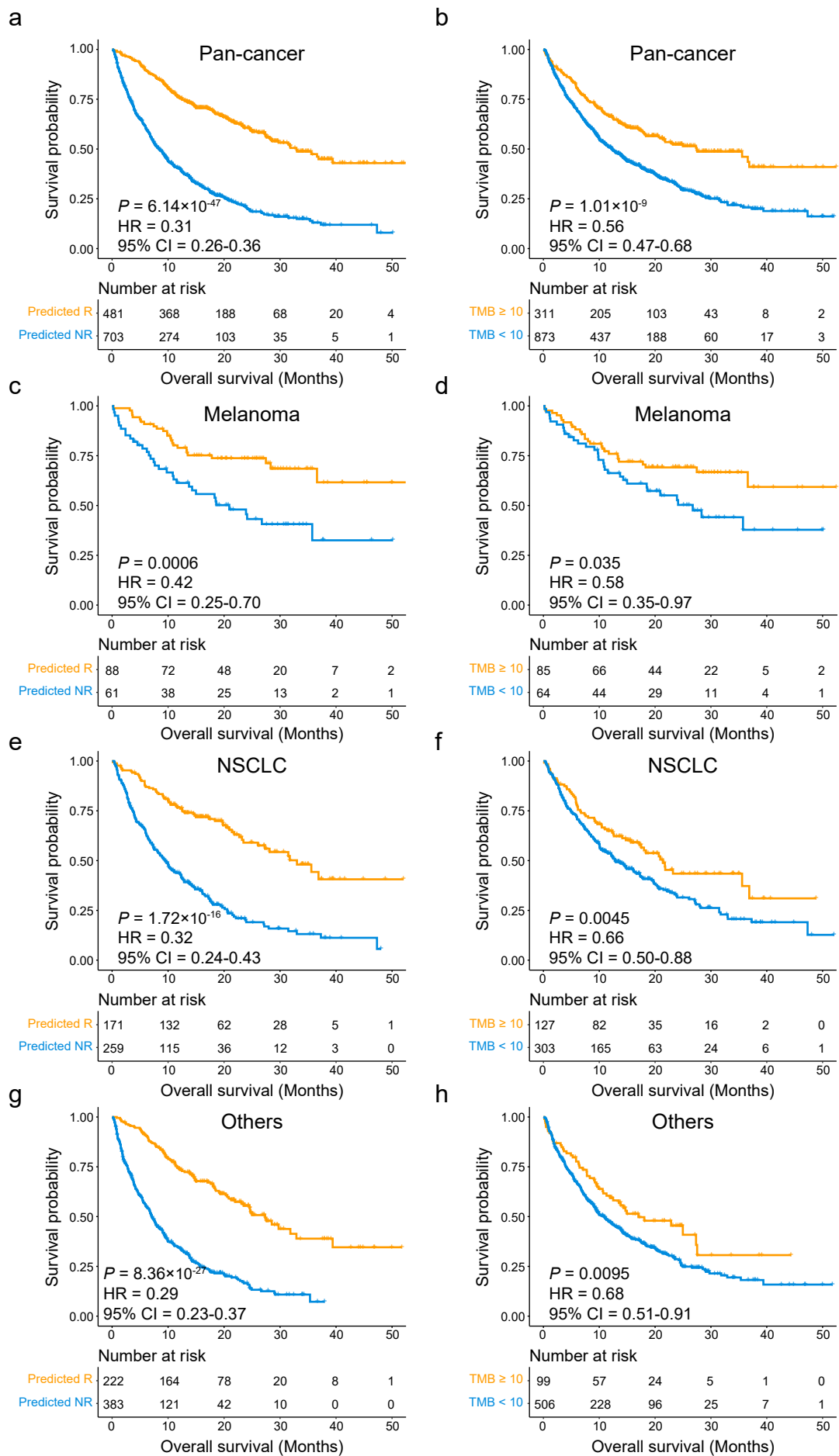
Supplementary Fig 10. Comparison of the predictive performance of RF16 with a logistic regression model. AUC and AUPRC values of RF16 and the logistic regression in the (a) training and (b) test set. Performance of the (c) RF16 model and the (d) logistic regression in the training set. Performance of the (e) RF16 model and the (f) logistic regression in the test set. In (d) and (f), the pan-cancer row represents the pan-cancer performances, which were calculated by aggregating the prediction results of melanoma, NSCLC, and others.



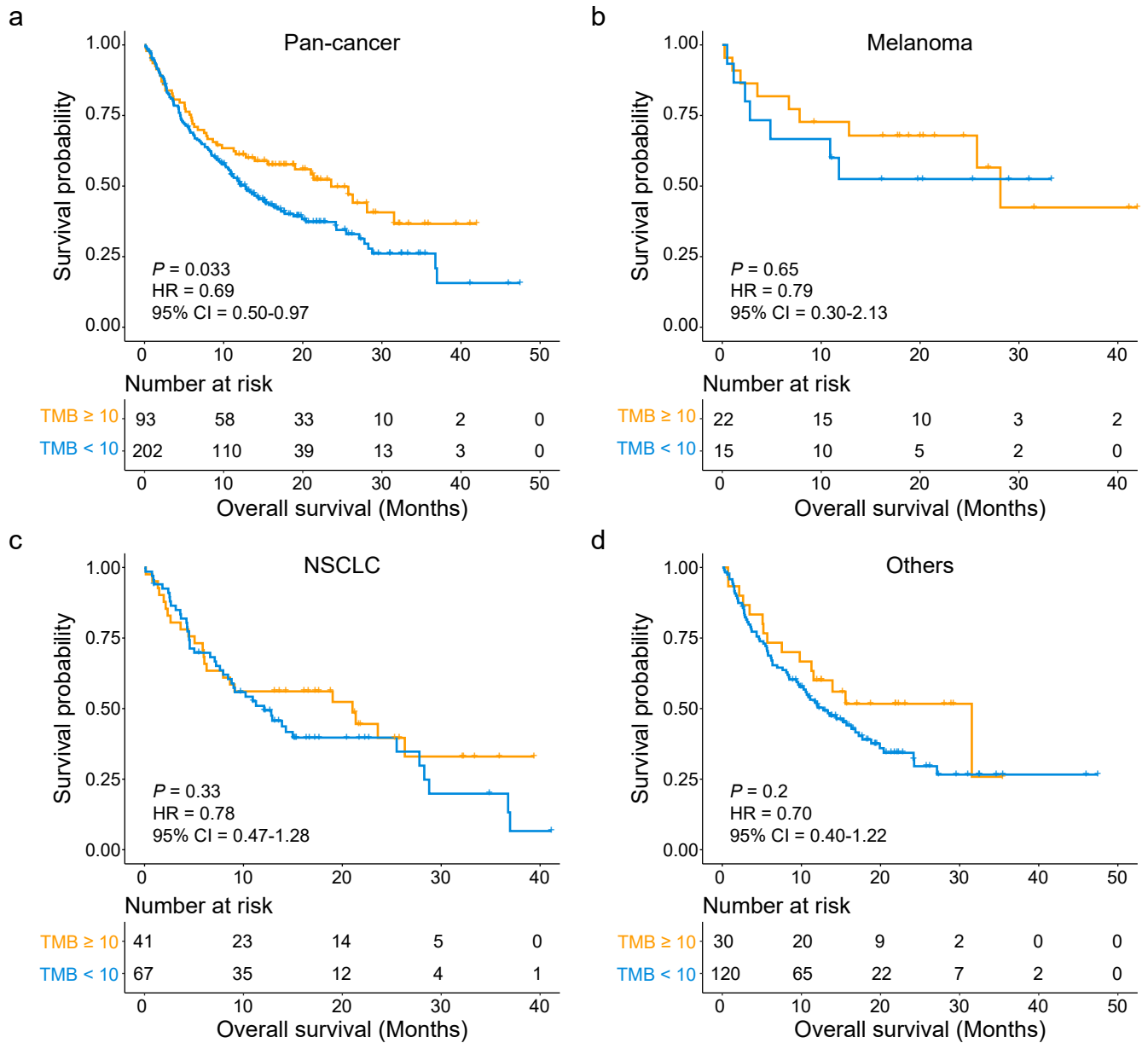
Supplementary Fig 11. Model performance illustrated by the brier score and C-index in the training set. (a) Brier score showing that RF16 for predicting OS has a smaller error compared to a reference model, based on RF11, or based on TMB alone in pan-cancer and each cancer group [melanoma, NSCLC, and others (not melanoma/NSCLC)]. Reference model denotes a random model. **(b)** Comparison of C-index and 95% CI for predicting OS among RF16, RF11, and TMB. Two-sided *P*-values were computed using the paired Student t-test.



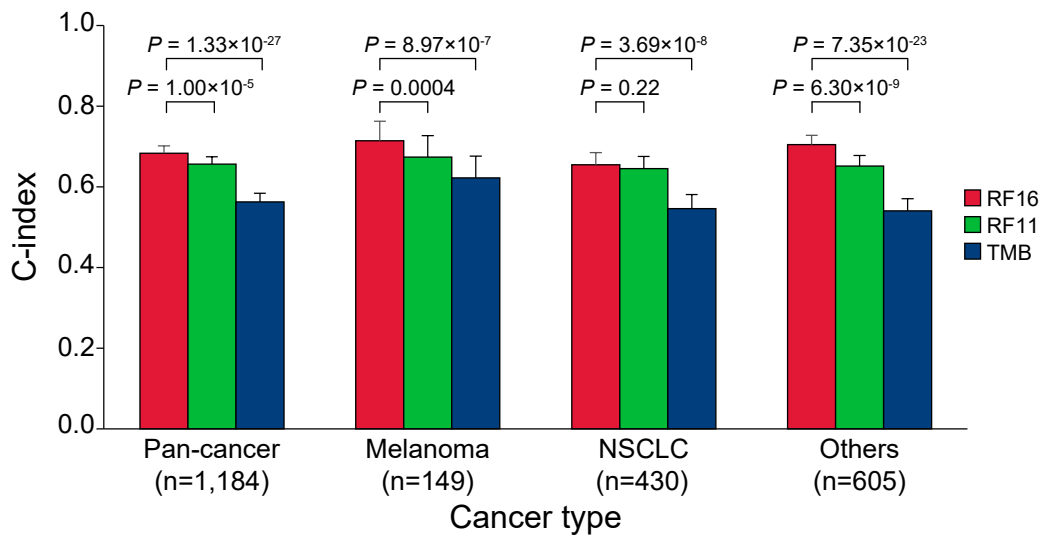
Supplementary Fig 12. Model performance illustrated by the brier score in the test set. Brier score showing that RF16 for predicting OS has a smaller error compared to a reference model, based on RF11, or based on TMB alone in pan-cancer and each cancer group [melanoma, NSCLC, and others (not melanoma/NSCLC)]. Reference model denotes a random model.



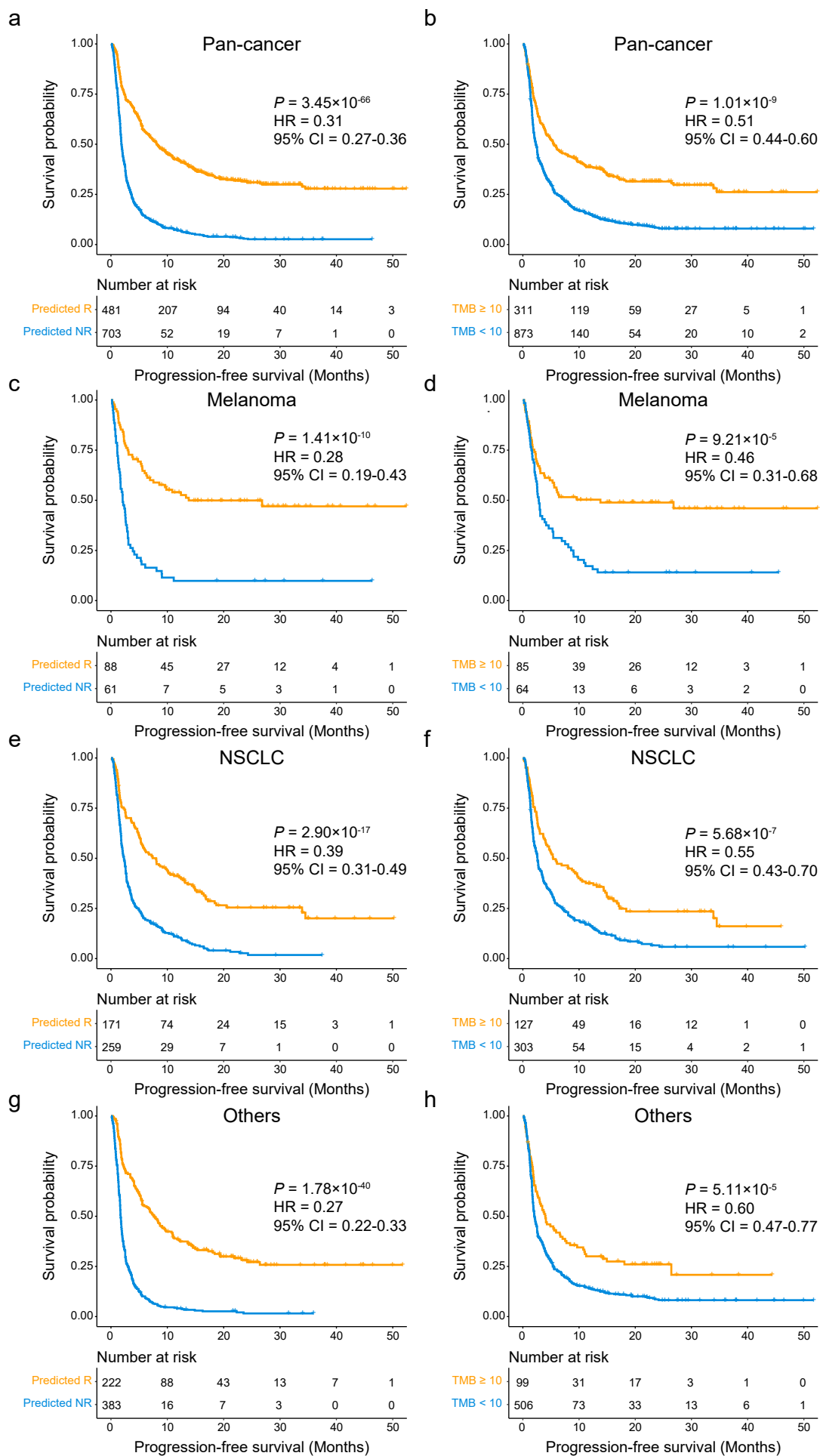
Supplementary Fig 13. Model predicts overall survival across multiple cancer types in the training set. (a)(c)(e)(g) Association between ICB response predicted by RF16 and OS in pan-cancer and each cancer group. (b)(d)(f)(h) Association between high TMB and OS in pan-cancer and each cancer group. High TMB was defined as any tumor with ≥ 10 mut/Mb. *P*-values were computed using the two-sided log-rank test.



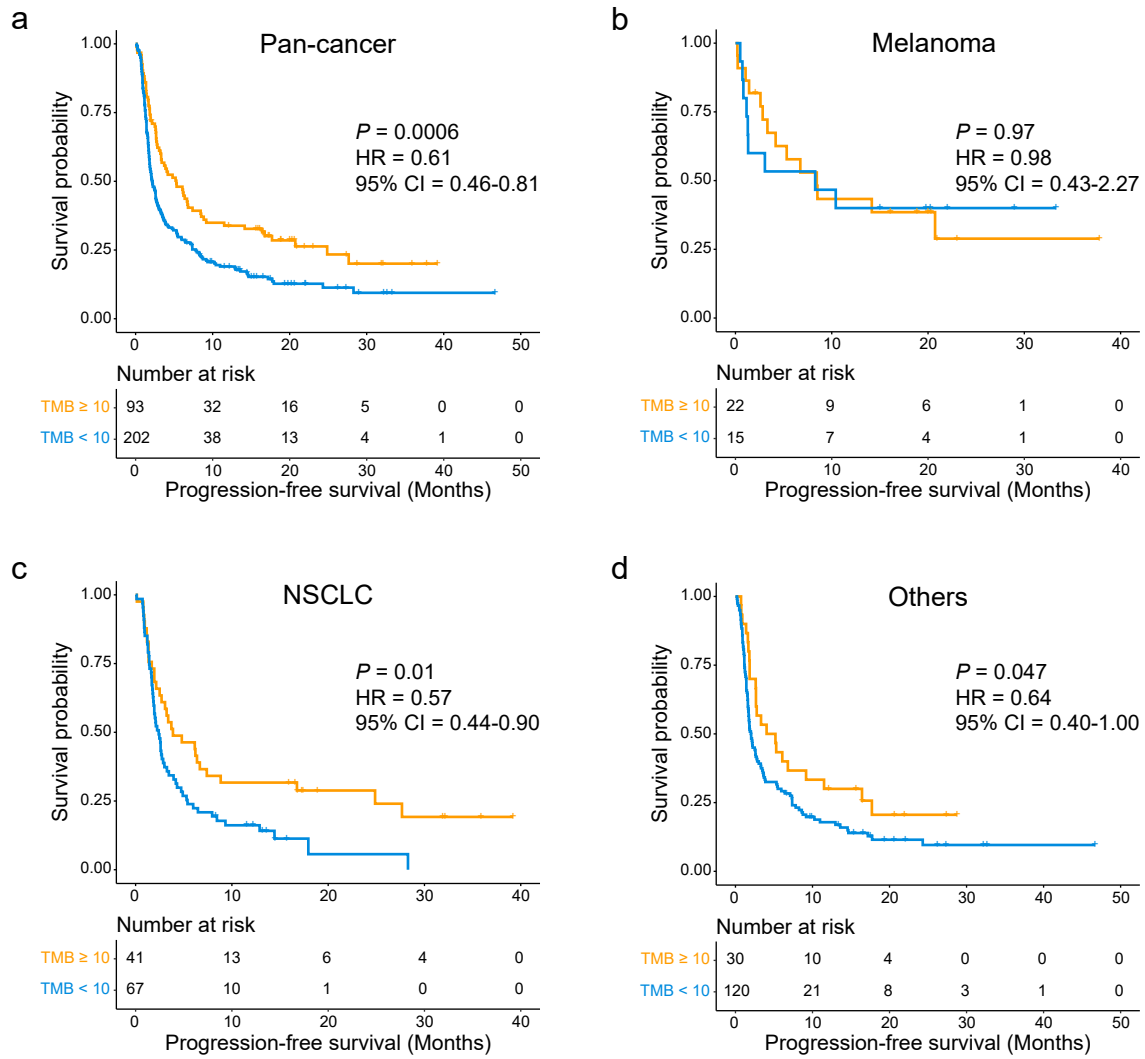
Supplementary Fig 14. Association of high TMB and overall survival in pan-cancer and each cancer group in the test set. *P*-values were computed using the two-sided log-rank test. High TMB was defined as any tumor with ≥ 10 mut/Mb.



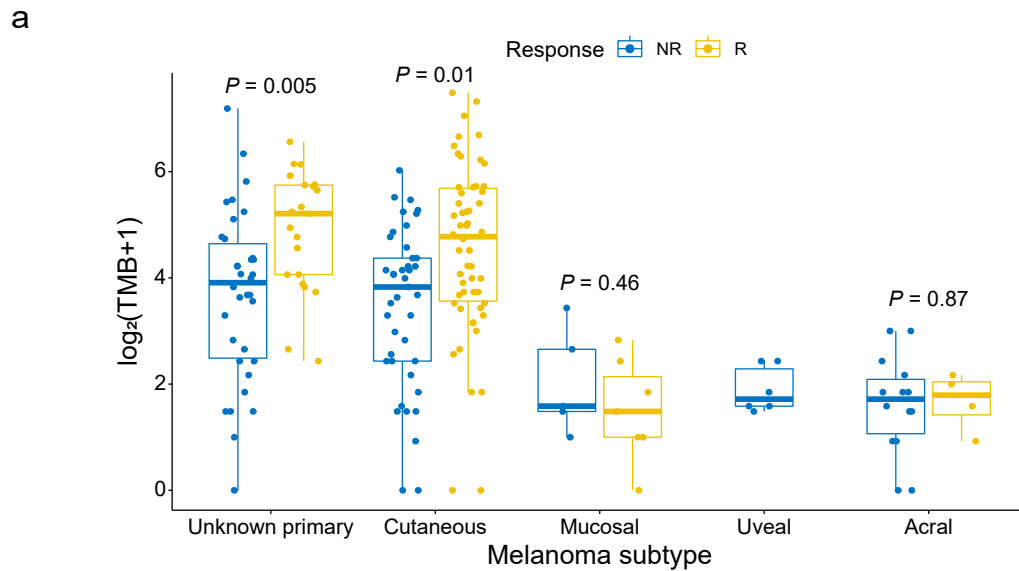
Supplementary Fig 15. Model predicts progression-free survival across multiple cancer types in the training set. Comparison of C-index and 95% CI for predicting PFS among RF16, RF11, and TMB in pan-cancer and each cancer group. Two-sided *P*-values were computed using the paired Student t-test.



Supplementary Fig 16. Model predicts progression-free survival across multiple cancer types in the training set. (a)(c)(e)(g) Association between ICB response predicted by the integrated RF16 model and PFS in pan-cancer and each cancer group. (b)(d)(f)(h) Association between high TMB and PFS in pan-cancer and each cancer group. High TMB was defined as any tumor with ≥ 10 mut/Mb. P -values were computed using the two-sided log-rank test.



Supplementary Fig 17. Association of high TMB and progression-free survival in pan-cancer and each cancer group in the test set. P -values were computed using the two-sided log-rank test. High TMB was defined as any tumor with ≥ 10 mut/Mb.



b

	Standardized coefficient	Standard error	z-value	<i>P</i> -value
(Intercept)	-1.33	0.57	-2.34	0.02
TMB	0.03	0.01	3.48	0.0005
Acral as reference group on melanoma subtype				
Cutaneous	0.89	0.63	1.42	0.16
Unknown primary	0.04	0.67	0.05	0.96
Mucosal	1.58	0.82	1.93	0.05
Uveal	-15.32	979.49	-0.02	0.99

Supplementary Fig 18. Association of TMB with immunotherapy response is not confounded by melanoma subtype. (a) Differences in TMB between responders and non-responders across different melanoma subtypes. Two-sided Mann-Whitney U test *P*-values are presented. Center bar, median; box, interquartile range; whiskers, first and third quartiles $\pm 1.5 \times$ interquartile range. (b) Logistic regression for immunotherapy response where TMB and melanoma subtype are included the model. Results show that the effect of TMB is not affected by the melanoma subtype.

Supplementary Table 1. Comparison of the AUC values for RF16 trained on pan-cancer data with those from separate cancer-specific models in the training and test set for all cancer types within the others group.

Cancer type	Pan-cancer model		Cancer-specific model	
	Training set AUC	Test set AUC	Training set AUC	Test set AUC
Bladder	0.85	0.70	0.98	0.40
Breast	0.98	0.25	1.00	0.50
Colorectal	0.92	1.00	0.99	0.93
Endometrial	0.86	0.86	0.95	0.67
Esophageal	0.88	0.95	0.94	1.00
Gastric	0.92	0.68	0.97	0.83
Head & Neck	0.91	0.45	0.97	0.52
Hepatobiliary	0.92	0.94	1.00	0.69
Mesothelioma	0.84	0.83	1.00	0.17
Ovarian	0.95	1.00	1.00	0.50
Pancreatic	0.95	1.00	NA ^a	0.50
Renal	0.86	0.81	0.90	0.86
Sarcoma	0.88	0.83	1.00	0.83
SCLC	0.92	0.78	0.99	0.78

^a'NA' is due to the absence of responders in the training set; the only one responder was present in the test set.

Supplementary Table 2. Effects of the model features and their corresponding P-values estimated from the logistic regression.

Feature	Standardized coefficient	Standard error	z-value	P-value
(Intercept)	-1.13	0.07	-15.16	6.49×10 ⁻⁵²
TMB	0.52	0.11	4.59	4.52×10 ⁻⁶
Albumin	0.44	0.10	4.54	5.60×10 ⁻⁶
Chemo prior ICB	-0.39	0.08	-4.67	3.05×10 ⁻⁶
NLR	-0.19	0.11	-1.69	0.09
FCNA	-0.19	0.08	-2.44	0.01
Age	0.17	0.08	2.14	0.03
Platelets	0.16	0.08	2.00	0.05
LOH in HLA-I	0.11	0.07	1.58	0.11
Cancer type ^a : NSCLC	0.10	0.12	0.85	0.40
Drug: Combo	0.08	0.07	1.09	0.27
BMI	0.07	0.07	0.96	0.33
Cancer type ^a : Others	0.07	0.13	0.50	0.62
MSI: Unstable	0.07	0.07	0.90	0.37
HED	0.06	0.07	0.86	0.39
HGB	-0.04	0.09	-0.39	0.69
Sex: Male	0.03	0.08	0.35	0.72
Tumor stage: IV	0.00	0.07	0.03	0.98

^aMelanoma as reference group

Supplementary Table 3. De-identified data needed to replicate all analyses is provided as a separate Excel file.