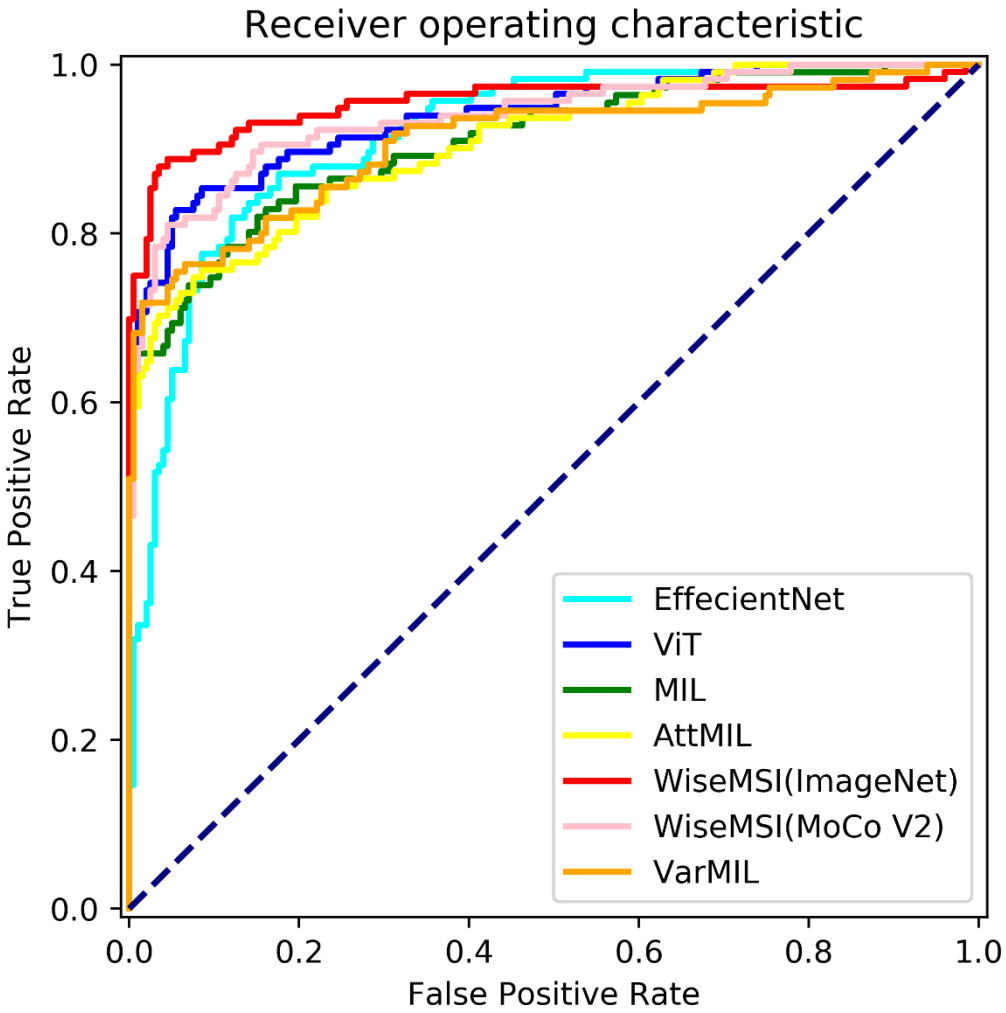
Cell Reports Medicine, Volume *4*

Supplemental information

Predicting colorectal cancer microsatellite instability with a self-attention-enabled convolutional neural network

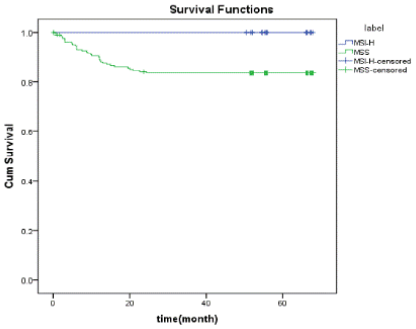
Xiaona Chang, Jianchao Wang, Guanjun Zhang, Ming Yang, Yanfeng Xi, Chenghang Xi, Gang Chen, Xiu Nie, Bin Meng, and Xueping Quan



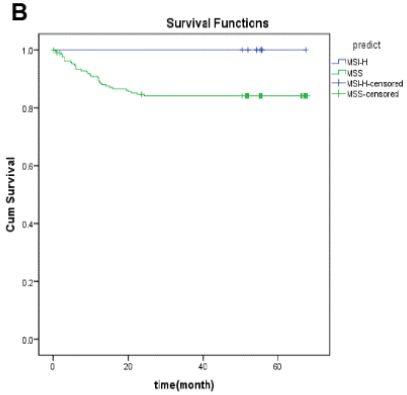
Supplemental Figure 1. The ROC curves of seven models. EfficientNet (cyan), Vit (blue), MIL (green), AttMIL (yellow), VarMIL (Orange), WiseMSI (red), that used

pretrained ImageNet feature extractor, and the WiseMSI (pink) using self-trained MocoV2 feature extractor. Related to Table 3.

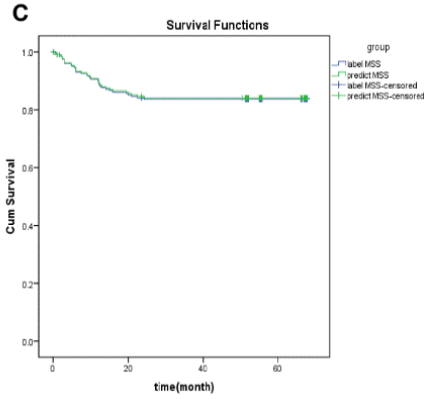
A


Survival Functions
label
MSI-H
MSS
MSI-H-censored
MSS-censored
Cum Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
time(month)

B


Survival Functions
predict
MSI-H
MSS
MSI-H-censored
MSS-censored
Cum Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
time(month)

C

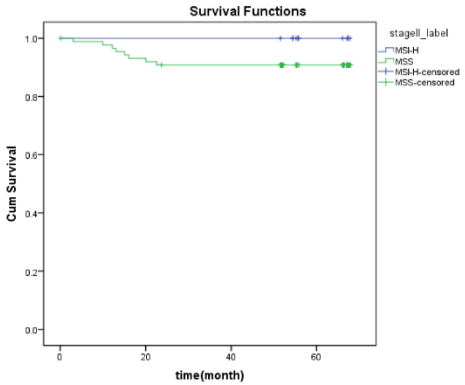

Survival Functions
group
label MSS
predict MSS
label MSS-censored
predict MSS-censored
Cum Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
time(month)

Supplemental Figure 2 Kaplan–Meier-estimated disease-free survival (DFS) curves of CRC patients from the TSMCC Cohort.

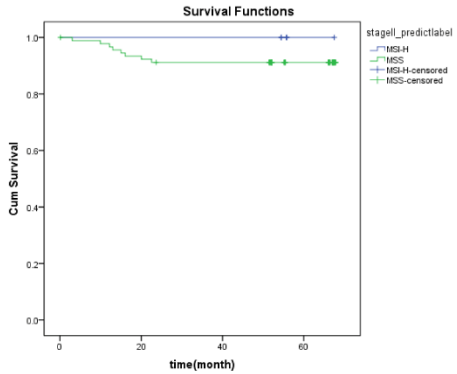
(A) DFS curves stratified by labeled MSI-H and MSS. (B) DFS curves stratified by MSI-H and MSS predicted by the CNN Wise MSI model combining self-attention mechanism. (C) The overlay of DFS curves of CRC patients with labeled and predicted MSS.

Related to “QUANTIFICATION AND STATISTICAL ANALYSIS” section of STAR Methods.

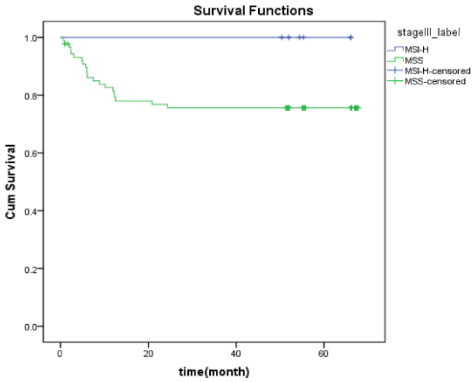
A


Survival Functions
stagell_label
MSI-H
MSS
MSI-H-censored
MSS-censored
Cum Survival
time(month)

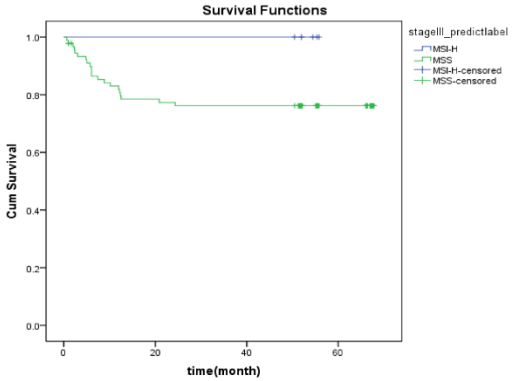
B


Survival Functions
stagell_predictlabel
MSI-H
MSS
MSI-H-censored
MSS-censored
Cum Survival
time(month)

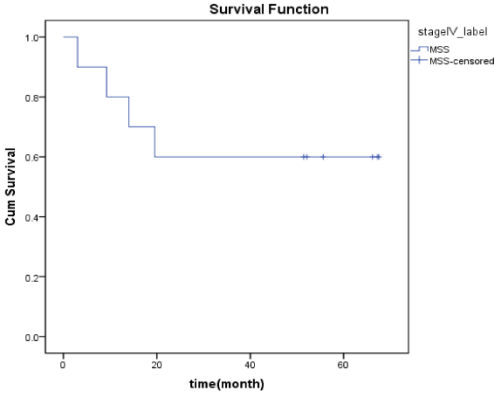
C


Survival Functions
stagelll_label
MSI-H
MSS
MSI-H-censored
MSS-censored
Cum Survival
time(month)

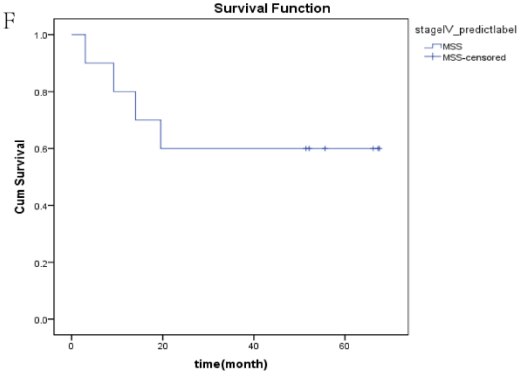
D


Survival Functions
stagelll_predictlabel
MSI-H
MSS
MSI-H-censored
MSS-censored
Cum Survival
time(month)

E


Survival Function
stageIV_label
MSS
MSS-censored
Cum Survival
time(month)

F


Survival Function
stageIV_predictlabel
MSS
MSS-censored
Cum Survival
time(month)

Supplemental Figure 3. Kaplan–Meier-estimated disease-free survival (DFS) curves of subgroups of CRC patients in Stage II - IV from the TSMCC Cohort. (

A) DFS curves of stage II patients with MSI status detected by PCR. (B) DFS curves of stage II patients with MSI status predicted by AI. (C) DFS curves of stage III patients with MSI status detected by PCR. (D) DFS curves of stage III patients with MSI status predicted by AI. (E) DFS curves of stage IV patients with MSI status detected by PCR. (D) DFS curves of stage IV patients with MSI status predicted by AI.

Related to “QUANTIFICATION AND STATISTICAL ANALYSIS” section of STAR Methods.

Supplemental Table 2. Comparison of running time on five model performance for MSI prediction.

Related to Table 3.

Model	Step	Avg time(s)
WiseMSI	Normal/Tumor Patch classifier extract feature MSI classifier	807.797
ViT	100 patches per slide MSI classifier	3.298
EfficientNet	100 patches per slide MSI classifier	4.003
MIL	extract feature MSI classifier	825.540
AttMIL	extract feature MSI classifier	825.955

Supplemental Table 3. The performance for MSI prediction in multi ethnics.

Related to Table 2.

Race	MSS	Specificity	MSI-H	Sensitivity	Accuracy
American Indian or Alaska native	0/1	0	-	-	0
Asian	11/11	1	1/1	1	1
Black or African American	32/59	0.542	7/9	0.778	0.574
White	105/240	0.438	37/43	0.860	0.502
Not reported	109/212	0.514	23/33	0.697	0.539

Supplemental Table 4. WiseMSI model performance tested with TCGA-STAD data.

Related to "WiseMSI on Other Cancer Type" section of STAR Methods.

Model	Specificity (95% CI)	Sensitivity (95% CI)	AUC (95% CI)
Self-attention	0.963 (0.936, 0.991)	0.185 (0.062, 0.307)	0.750 (0.699, 0.802)
ViT	0.972 (0.948,0.997)	0.131 (0.030, 0.231)	0.682 (0.635, 0.729)

Supplemental Table 5. Patient demographic and baseline characteristics stratified by MSI status by molecular typing and Wise MSI-based prediction.

Related to Figure 4.

Variables	N	label		predict label		Fisher Exact sig(2-sided) P value
		MSI-H	MSS	MSI-H	MSS	
Gender						
Male	139	10	129	6	133	0.441
Female	87	11	76	8	79	0.628
Age, years						
mean±SD	60.56±11.27	56.67±11.11	60.96±11.24	55.29±6.81	60.91±11.43	
>=65	90	4	86	2	88	0.682
<65	136	17	119	12	124	0.432
Stage						
I	20	2	18	1	19	1
II	99	11	88	7	92	0.459

III	97	8	89	6	91	0.783
IV	10	0	10	0	10	1
Tumor differentiation						
Low	202	4	20	3	21	1
Moderate	24	17	185	11	191	0.327
Tumor location						
Colon	216	13	203	6	210	0.158
Rectum	10	8	2	8	2	1
Laterality						
Left	178	8	170	4	174	0.379
Right	38	5	33	2	36	0.43
Lynch syndrome						
Yes	142	9	133	3	139	0.137
No	84	12	72	11	73	1