Original Article

COX-2/C-MET/KRAS status-based prognostic nomogram for colorectal cancer: A multicenter cohort study

Jianhua Liu[#], Chengzhi Huang[1#], Junjiang Wang[1], Ling Huang, Shaojie Chen[2]

Departments of Oncology and [1]Gastrointestinal Surgery, Cancer Center, Guangdong General Hospital, Guangdong Academy of Medical Sciences, [2]Department of Gastroenterology, Sun Yat-Sen Memorial Hospital, Guangzhou, China

[#]These authors contributed to the work equally.

Abstract

Background/Aim: To construct quantitative prognostic models for colorectal cancer (CRC) based on COX-2/C-MET/KRAS expression status in clinical practice.

Patients and Methods: Clinical factors and COX-2/C-MET/KRAS expression status of 578 eligible patients from two Chinese hospitals were included. The patients were randomly allocated into training and validation datasets. We created several models using Cox proportional hazard models: Signature^{C} contained clinical factors, Signature^{G} contained COX-2/C-MET/KRAS expression status, and Signature^{CG} contained both. After comparing their accuracy, nomograms for progression-free survival (PFS) and overall survival (OS) were built for the best signatures, with their concordance index and calibration tested. Further, patients were subgrouped by the median of the best signatures, and survival differences between the subgroups were compared.

Results: For PFS, among the three signatures, Signature$^{PFS\text{-}CG}$ had the best area under the curve (AUC), with the 1-, 2- and 3-year AUCs being 0.70, 0.73 and 0.89 in the training dataset, respectively and 0.67, 0.73 and 0.87 in the validation dataset, respectively. For OS, the AUCs of Signature$^{OS\text{-}CG}$ for 1-, 2- and 3-years were 0.63, 0.71 and 0.81 in the training dataset, respectively and 0.68, 0.71 and 0.76 in validation dataset, respectively. The nomograms based on Signature$^{PFS\text{-}CG}$ and Signature$^{OS\text{-}CG}$ had good calibrations. Subsequent stratification analysis demonstrated that the subgroups were significantly different for both PFS (training: $P < 0.001$; validation: $P < 0.001$) and OS (training: $P < 0.001$; validation: $P < 0.001$).

Conclusions: Combining clinical factors and COX-2/C-MET/KRAS expression status, our models provided accurate prognostic information in CRC. They can be used to aid treatment decisions in clinical practice.

Keywords: C-MET, colorectal cancer, COX-2, KRAS, prognosis

Address for correspondence: Dr. Jianhua Liu, Department of Oncology, Cancer Center, Guangdong General Hospital, Guangdong Academy of Medical Sciences, 123 Huifu Road West, Guangzhou 510180, China.
E-mail: liujianh80@126.com

INTRODUCTION

Colorectal cancer (CRC) is one of the most common cancers worldwide, with an incidence ranking 3rd and 2nd in males and females, respectively, and a mortality rate ranking 4th and 3rd, respectively.[1] To provide reasonable and effective treatment for patients, we need to have accurate prognoses for patients. Many studies have focused

Access this article online	
Quick Response Code:	Website: www.saudijgastro.com
	DOI: 10.4103/sjg.SJG_502_18

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

How to cite this article: Liu J, Huang C, Wang J, Huang L, Chen S. COX-2/C-MET/KRAS status-based prognostic nomogram for colorectal cancer: A multicenter cohort study. Saudi J Gastroenterol 2019;25:293-301.

on this issue.[2-4] However, most of them were qualitative rather than quantitative studies, meaning patients were divided into subgroups with high or low risks. Although this approach aided the identification of patients at risks for a specific patient, it provided limited information for their survival rate at a fixed time. Thus, these studies have not been able to aid doctors in making reasonable choices between aggressive and conservative treatment.

Apart from traditional clinical factors, genomic data may also be informative, especially when several genetic mutations are assessed in combination.[5] To date, mutations resulting in the overexpression of *COX-2*, *C-MET* and *KRAS*, the key gene abnormalities in CRC, have been shown to have utility in CRC prognosis. According to a variety of studies, COX-2 overexpression is an adverse predictor for disease-specific survival in CRC.[6-8] Mesenchymal-epithelial transition factor (C-MET) overexpression has also been shown to be closely correlated with CRC progression and metastasis and may interfere with anti-epidermal growth factor receptor (*EGFR)* strategies.[9,10] Moreover, as a driver mutation and the principle aspect of somatic changes in CRC, KRAS mutations contribute to cell invasion and apoptosis suppression during tumor progression, along with resistance to anti-*EGFR* therapy.[11-13] Therefore, a prognostic model combining mutations in these three genes status and clinical factors may be helpful not only for the current therapeutic regimens but also for future applications of targeted therapies.

Thus, we conducted this study in a multi-center cohort to develop a COX-2/C-MET/KRAS abnormalities expression-based prognostic nomogram for survival in CRC.

PATIENTS AND METHODS

Patients

This multi-center study initially included 683 patients diagnosed with CRC at Guangdong General Hospital and Sun Yat-sen Memorial Hospital, between November 2010 and October 2014. After screening, 578 eligible patients were randomized using the R sample function with a 3:2 ratio to create a training dataset (346 cases) and a validation dataset (232 cases). The inclusion criteria were that the patient had (1) baseline characteristics recorded, including radiological examination and necessary laboratory tests; (2) at least one radiological progressive disease (PD) (confirmed by computed tomography [CT]/magnetic resonance imaging [MRI]) or death confirmed, or had a follow-up >36 months before the cut-off date; and (3) the status of COX-2/C-MET/KRAS was confirmed at diagnosis. The exclusion criteria were (1) irregular follow-up and (2) the presence of other cancers. The inclusion and exclusion flowchart is shown in Figure 1.

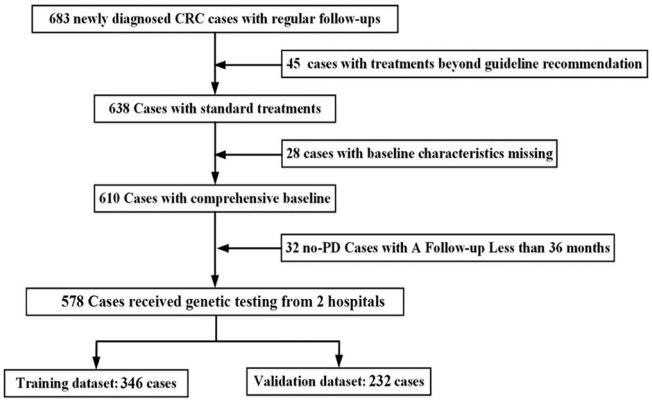

Figure 1: Flowchart of patients' inclusion and exclusion

This study was approved by the Ethics Committee of Guangdong General Hospital and Sun Yat-sen Memorial Hospital. Informed consent was waived in lieu of the retrospective design of the study. In addition, before the analyses, all patients' information was anonymized and de-identified.

Treatments and follow-ups

Patient treatments were decided according to established guidelines.[14,15] In general, radical surgery was prioritized for participants with stage I disease, and complete surgical excision with adjuvant 5-fluorouracil-based chemotherapy was used for stage II and III cases. Regarding patients with advanced CRC, chemotherapy alone or chemotherapy combined with targeted therapy [anti-vascular endothelial growth factor monoclonal antibody for all patients and anti-EGFR antibodies for patients with the absence of *KRAS* or *BRAF* mutations] were alternatives. The follow-up interval was 4–6 weeks, with chest X-rays, abdominal CT/MRI[14,15] and necessary laboratory tests. In addition, all suspected new lesions were confirmed by CT/MR.

Determination of COX-2/C-MET/KRAS expression status

The 578 formalin-fixed paraffin-embedded (FFPE) primary CRC samples, which were obtained from surgical/endoscopic biopsies were cut into 4-μm thick sections longitudinally, and the COX-2 and C-MET proteins were detected using a standard immunohistochemistry protocol.[16] Briefly, after dewaxing, re-hydrating and antigen-retrieving, the sections were inhibited for endogenous peroxidase activity and for blocked nonspecific antibody binding, followed by incubation with primary antibody against COX-2 (BD, NJ, USA; lyophilized, 1:200 dilution) or C-MET (BD; lyophilized, 1:100 dilution)

overnight at 4°C. Subsequently, slides were incubated with secondary antibodies for 30 min; then, the immunoreaction was visualized with the streptavidin-biotin peroxidase complex method. Finally, the sections were visualized using an optical microscope (CX31; Olympus Corporation, Tokyo, Japan). The findings were judged by two independent pathologists, and COX-2 or C-MET staining was analyzed by multiplying the percentage of positive cells (P) by staining intensity (I). Briefly, *P* was defined as 0, 1, 2 or 3 score when *P* accounted for <5%, between 5% and 25%, between 26% and 50%, or >50% of total cells, respectively. I was scored 0 for absent/weak staining, 1 for moderate staining, and 2 for strong staining. Thus, the total scores ranged from 0 to 6, and overexpression was defined as a score of 6.

KRAS mutations were analyzed for each FFPE specimen using polymerase chain reaction (PCR)-based direct gene sequencing. Genomic deoxyribonucleic acid (DNA) was extracted with QIAamp DNA FFPE Tissue Kit Qiagen (Hilden, Germany) in accordance with the manufacturer's protocol, and cancer cell rich regions were ascertained beforehand using the application of hematoxylin and eosin (H and E) staining to ensure that all cases tested for enrichment of ≥70% tumor cells. Subsequently, DNA concentration was measured using an ND-1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA), and *KRAS* mutations (including exons 2, 3, and 4) of all tumor samples were determined. Primer pairs for gene amplifications were designed with AmpliSeq Designer software, version 1.2.6 software (Life Technologies).[17] DNA amplification was carried out with GoTaq® Hot Start Polymerase (Promega, Madison, WI, USA) and 0.2 μM of each primer using the GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA, USA). Cycling conditions were as described previously.[18] After identification by agarose gel electrophoresis, the PCR products were purified with the DNA Clean/Extraction Kit (GeneMark), then submitted to direct sequencing on an ABI 3730xl genetic analyzer (Invitrogen Life Technologies, Carlsbad, CA, USA), and detailed steps were consistent with those reported earlier.[19] The resulting reads were aligned using Chromas software, and all mutations were confirmed in both the sense and anti-sense directions.

Clinical factors

Clinical factors included sex (classified as male or female), age (classified as ≤60 or >60 years), tumor location (classified as right colon or left colon/rectum), differentiation (classified as moderate/well or poor), TNM stage (classified as I, II, III, or IV), T stage (classified as T1/T2 or T3/T4), N stage (classified as N0, N1 or N2), M stage (classified as M0 or M1), carcinoembryonic antigen (CEA) (classified as <20 ng/mL or >20 ng/mL), KRAS (classified as mutation and wild-type), COX-2 (classified as low expression or overexpression), and C-MET (classified as low expression or overexpression). We assumed that T, N and M had different prognostic values, so we included both TNM stage as a comprehensive factor and T, N and M as separate factors. Because treatments were decided according to TNM stage, treatment was not included in the model construction to avoid the influence of collinearity.

Outcomes

Compared with progression-free survival (PFS), overall survival (OS) may be influenced more by confounding factors, so we used PFS as our primary outcome along with OS as our secondary outcome. PFS and OS were determined by the time span from enrollment start time to PD/censoring and death/censoring, respectively. PD was defined in accordance with Response Evaluation Criteria in Solid Tumors (RECIST 1.1).[20]

Statistical analysis

Categorical variables expressed as percentages were compared by the Pearson's χ^2 test or the Fisher's exact test (when the expected frequency was less than 5) as appropriate. PFS and OS were compared using the logrank test.

For the general process of model construction, to test the prognostic value of COX-2/C-MET/KRAS for PFS and OS, we first randomly separated the cohort into training and validation sets. Second, we performed the analysis for PFS and OS seperately. In general, signatures were built using clinical factors ($Signature^{PFS-C}$ and $Signature^{OS-C}$), gene expression/mutation abnormalities ($Signature^{PFS-G}$ and $Signature^{OS-G}$) and both ($Signature^{PFS-CG}$ and $Signature^{OS-CG}$). All candidate predictors were included in the development of the prediction models. The two types of variables (clinical factors and DNA mutations) separately entered $Signature^{C}$ and $Signature^{G}$. All the factors identified above were then introduced into $Signature^{CG}$ using the backward stepwise COX-2 proportional hazard model by Akaike information criterion. Third, for 1-, 2- and 3-year PFS and OS, likelihood ratio tests were used to compare the different models. In addition, for 1-, 2- and 3-year PFS and OS, we estimated the area under curve (AUC) of the three signatures using the approach of Heagerty *et al*.,[21] and a nomogram was constructed for the optimal signatures with their calibration tested to compare the expected and observed survival probabilities. Finally, split-sample validation of the prediction model

was conducted to assess the stability of the model. The patients were subgrouped by the median of the optimal signatures, the Kaplan-Meier method was used to estimate the survival function, and the logrank test was used to compare survival.

All statistical tests performed were two-sided, and *P* values less than 0.05 were considered as statistically significant. Data was analyzed by the R statistical package (http://www.r-project.org/).

RESULTS

Patients' baseline characteristics

After random allocation into the two groups, the number of patients in the training and validation datasets were 346 and 232, respectively. In the total study population, 258 cases had cancer located in the right colon (training: 149; validation: 109), and 320 had cancer located in the left colon or rectum (training: 197; validation: 123). By the end date of follow-up, 411 PD occurred (training: 248; validation: 163), and 337 cases died (training: 200; validation: 137). Their characteristics are shown in Tables 1 and 2.

Table 1: Baseline demographics and characteristics of training and validation dataset (I)

Variable	Training dataset (*n*=346)	Validation dataset (*n*=232)	*P*
Sex			0.780
Male	192 (55.5%)	126 (54.3%)	
Female	154 (44.5%)	106 (45.7%)	
Age			0.397
≤60	175 (50.6%)	109 (47.0%)	
>60	171 (49.4%)	123 (53.0%)	
Location			0.353
Right colon	149 (43.1%)	109 (47.0%)	
Left colon and rectum	197 (56.9%)	123 (53.0%)	
Differentiation			0.965
Moderate and well	244 (70.5%)	164 (70.7%)	
Low	102 (29.5%)	68 (29.3%)	
T stage			0.069
T1 and T2	25 (7.2%)	27 (11.6%)	
T3 and T4	321 (92.8%)	205 (88.4%)	
N stage			0.012
N0	92 (26.6%)	73 (31.5%)	
N1	152 (43.9%)	116 (50.0%)	
N2	102 (29.5%)	43 (18.5%)	
M stage			0.939
M0	275 (79.5%)	185 (79.7%)	
M1	71 (20.5%)	47 (20.3%)	

Table 2: Baseline demographics and characteristics of training and validation dataset (II)

Variable	Training dataset (*n*=346)	Validation dataset (*n*=232)	*P*
TNM stage			0.139
Ⅰ	12 (3.5%)	18 (7.8%)	
Ⅱ	62 (17.9%)	43 (18.5%)	
Ⅲ	201 (58.1%)	124 (53.4%)	
Ⅳ	71 (20.5%)	47 (20.3%)	
CEA level			0.507
<20 ng/ml	84 (24.3%)	62 (26.7%)	
≥20 ng/ml	262 (75.7%)	170 (73.3%)	
KRAS status			0.269
Mutation	123 (35.5%)	93 (40.1%)	
Wild-type	223 (64.5%)	139 (59.9%)	
COX-2			0.147
Low expression	80 (23.1%)	42 (18.1%)	
Overexpression	266 (76.9%)	190 (81.9%)	
C-MET			0.050
Low expression	107 (30.9%)	90 (38.8%)	
Overexpression	239 (69.1%)	142 (61.2%)	

Gene expression status

For all participants, 456 cases COX-2 overexpression (training: 266; validation: 190), 381 was C-MET overexpression (training: 239; validation: 142) [Figure S1], and 216 cases exhibited *KRAS* mutations (training: 123; validation: 93). A total of 326 patients had overexpression for both COX-2+C-MET (training: 212; validation: 114), 176 had both COX-2+KRAS abnormalities (training: 92; validation: 86), 180 had abnormalities of both C-MET+KRAS (training: 114; validation: 66), and 132 had abnormalities of all three (training: 86; validation: 46). However, there was no statistical difference between the training and validation datasets.

Model construction

Results of univariate and multivariate analyses are shown in Tables S1 and S2. Considering the prognostic effects of COX-2 and C-MET,[22-24] although they were not statistically different in the multivariate analysis of PFS [Tables S1 and S3], we still tested whether the integration of COX-2 and C-MET improved the models. The results showed that for Signature$^{PFS\text{-}CG}$, adding information of COX-2 and C-MET expression increased the AUC of 3-year PFS in the validation dataset (0.87 vs. 0.84, Figures 2 and S2). Thus, although calibration between the nomograms were similar [Figure S3], we believed it was reasonable to add COX-2 and C-MET expression into the nomogram for Signature$^{PFS\text{-}CG}$. Finally, the formulas were as follows: PFS (with COX-2 and C-MET): points = 23 × (age > 60) +26 × (tumor location in the left colon and rectum) +37 × (N stage = N1) +78 × (N stage = N2) + 100 × (M stage = M1) + 29 × (CEA ≥ 20 ng/mL) + 49 × (KRAS mutation) + 14 × (COX-2 overexpression) + 2 × (C-MET overexpression); PFS (without COX-2 and C-MET): points = 25 × (age > 60) + 27 × (tumor location in the left colon and rectum) + 36 × (N stage = N1) + 75 × (N stage = N2) + 100 × (M stage = M1) + 30 × (CEA ≥ 20 ng/mL) + 40 × (KRAS mutation); OS: points = 41 × (age > 60) + 35 × (N stage = N1) + 64 × (N stage = N2) + 100 × (M stage = M1) + 39 × (CEA ≥ 20 ng/mL) + 55 × (KRAS mutation).

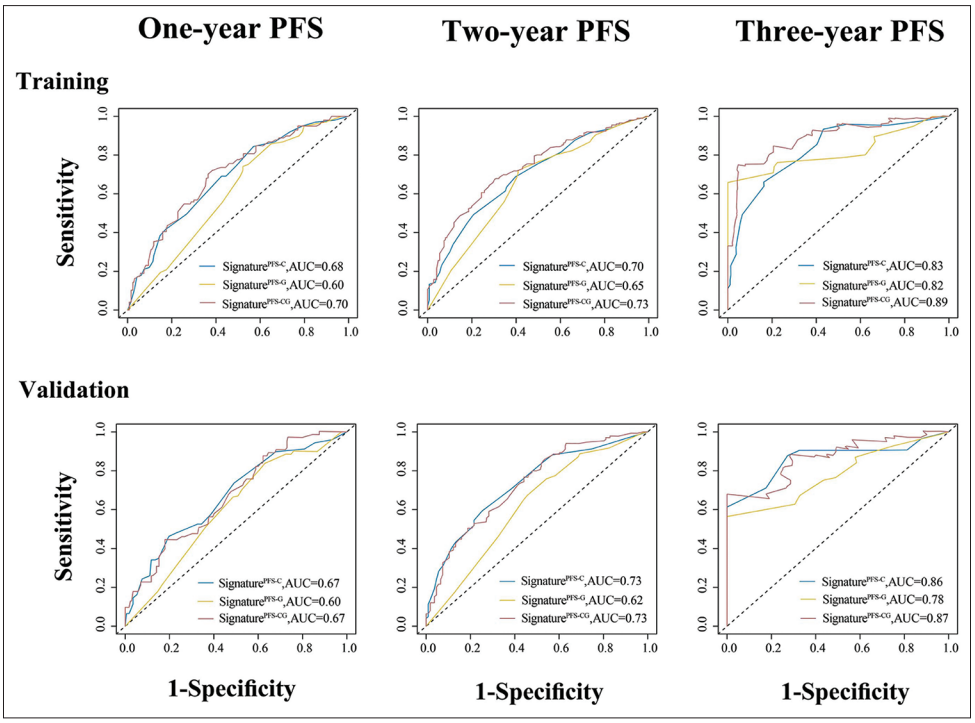

Figure 2: Comparing the accuracy of progressioin-free survival (PFS) by the three signatures. For the 1-, 2-, and 3-year PFS, the AUC of Signature^{PFS-C}, Signature^{PFS-G}, and Signature^{PFS-CG} in the training dataset were 0.68, 0.60 and 0.70; 0.70, 0.65 and 0.73; and 0.83, 0.82 and 0.89. Those in the validation dataset were 0.67, 0.60 and 0.67; 0.73, 0.62 and 0.73; and 0.86, 0.78 and 0.87. In addition, Signature^{PFS-CG} has the highest accuracy in both datasets. AUC: Area under the curve; C: Clinical factors; G: Gene abnormalities; CG: Clinical factors plus gene abnormalities

For PFS, among the three signatures (Signature^{PFS-C}, Signature^{PFS-G}, and Signature^{PFS-CG}), the AUCs in training dataset were 0.68, 0.60 and 0.70 for 1 year; 0.70, 0.65 and 0.73 for 2 years; and 0.83, 0.82 and 0.89 for 3 years [Figure 2], respectively, with Signature^{PFS-CG} having the best performance [Table 3]. In the validation dataset, the AUCs were 0.67, 0.60 and 0.67 for 1 year; 0.73, 0.62 and 0.73 for 2 years; and 0.86, 0.78 and 0.87 for 3 years [Figure 2], respectively, and Signature^{PFS-CG} also had the best performance [Table 3].

For OS, among the three signatures (Signature^{OS-C}, Signature^{OS-G} and Signature^{OS-CG}) the AUCs in the training dataset were 0.61, 0.57 and 0.63 for 1 year; 0.67, 0.61 and 0.71 for 2 years; and 0.72, 0.76 and 0.81 for 3 years [Figure 3], respectively, with Signature^{OS-CG} achieving the best performance [Table 3]. In the validation dataset, the AUCs were 0.64, 0.62 and 0.68 for 1 year; 0.70, 0.63 and 0.71 for 2 years; and 0.77, 0.61 and 0.76 for 3 years [Figure 3], respectively. Although Signature^{OS-CG} had the best performance for 1 and 2-year OS, it was inferior to Signature^{OS-C} for 3-year OS [Table 3].

From these results, nomograms based on Signature^{PFS-CG} were constructed for PFS [Figure 4-a] and achieved good calibration in both the training [Figure 4-b] and validation [Figure 4-c] datasets. For OS, a nomogram was also constructed [Figure 4-d], with good calibration in both datasets [Figure 4-e and f].

Kaplan-Meier survival analysis

When divided by the median of Signature^{PFS-CG} (threshold: 0.0640), the two groups had a statistically significant difference for PFS, both in the training ($P < 0.001$, Table 4 and Figure 5a) and validation ($P < 0.001$, Table 4 and Figure 5b) datasets. Similar results were observed for OS when patients were separated by the median of Signature^{OS-CG} (threshold: 0.1969; training: $P < 0.001$; and validation: $P < 0.001$; Table 4 and Figure 5c, 5d).

DISCUSSION

Due to the nature of its clinical heterogeneity and aggressiveness, CRC presents with a high incidence of distant metastasis and accompanying poor prognosis. During the progression from uncontrolled cell growth to invasive cancer, a series of clinical and genetic abnormalities as well as deregulation in cellular homeostasis are involved. At present, many genes have been studied in CRC, with *COX-2/C-MET/KRAS* being the most representative. Up-regulation of COX-2 and C-MET as well as KRAS mutations play pivotal roles in accelerating cancer-promoting activity and worsening prognosis. To select more individualized treatments for patients, accurate

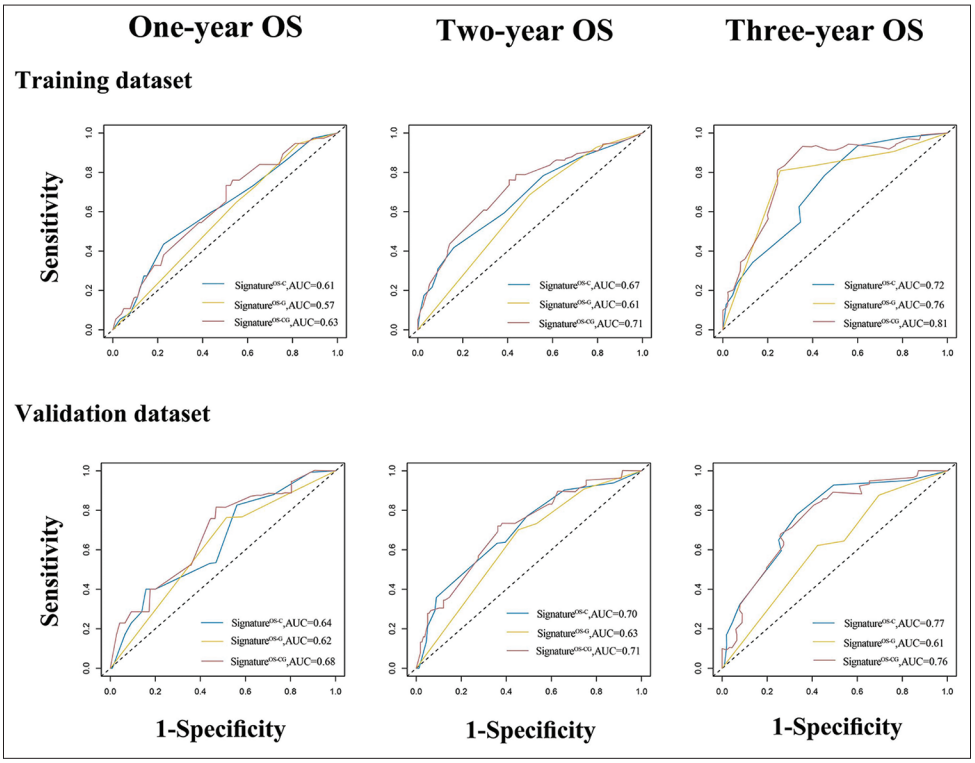

Figure 3: Comparing the accuracy of overall survival (OS) by the three signatures. For the 1-, 2- and 3-year OS, the AUC of Signature^{OS-C}, Signature^{OS-G}, and Signature^{OS-CG} in the training dataset were 0.61, 0.57 and 0.63; 0.67, 0.61 and 0.71; and 0.72, 0.76 and 0.81. Signature^{OS-CG} has the highest accuracy. Those in the validation dataset were 0.64, 0.62 and 0.68; 0.70, 0.63 and 0.71; and 0.77, 0.61 and 0.76. Signature^{PFS-CG} has the highest accuracy in the 1- and 2-year, but for 3-year, Signature^{PFS-C} was a little better. AUC: Area under the curve; PFS: Progression-free survival; C: Clinical factors; G: Gene abnormalities; CG: Clinical factors and gene abnormalities

Table 3: Pairwise comparison among signatures by Likelihood ratio tests (*P*-value)

	Training dataset	Validation dataset
PFS		
Signature^{PFS-C} vs. Signature^{PFS-G}	<0.001	<0.001
Signature^{PFS-C} vs. Signature^{PFS-CG}	0.005	<0.001
Signature^{PFS-G} vs. Signature^{PFS-CG}	<0.001	<0.001
OS		
Signature^{OS-C} vs. Signature^{OS-G}	<0.001	<0.001
Signature^{OS-C} vs. Signature^{OS-CG}	<0.001	<0.001
Signature^{OS-G} vs. Signature^{OS-CG}	<0.001	<0.001

PFS: Progression-free survival; OS: Overall survival; C: Clinical factors; G: Gene abnormalities; CG: Clinical factors and gene abnormalities

prognostic models should be constructed including both clinical factors and genetic information. In this study, we constructed COX-2/C-MET/KRAS expression-based prognostic models in a multicenter cohort of patients. The signatures were accurate in both the training and validation datasets, especially for long-term survival (3-year PFS and OS). In addition, the nomograms based on them also showed good calibration for 1-, 2- and 3-year PFS and OS. Furthermore, when divided by the median of the signatures, patients with low and high risk for CRC could be identified.

CRC progression is a multi-step process that includes various genetic events involving dysregulation of metastasis-associated genes and inactivation of tumor repressor genes.[25] Recently, a variety of studies have contributed to the development of CRC prognostic models, demonstrating that the abnormalities of clinical and genetic features might be significant predictors for poor clinical outome.[26-28] However, there are limitations to the application of these conclusions. First, fragmented processes identifying informative factors can only qualitatively identify patients at risk but are without comprehensive integration of prognostic factors, so quantitative survival prediction remains a challenge. Second, models should be validated in independent datasets to show that the conclusions could be applied beyond the original cohorts. Third, to facilitate clinical application, the factors included in the models should be easily collected during diagnosis and treatment. With similar accuracy, the fewer additional examinations that need to be performed, the more convenient the models might be. From these issues, we chose the three most widely tested genes in

Table 4: Kaplan-Meier analysis based on subgrouping by the signatures

	Training dataset			Validation dataset		
	Median (months)	95%CI	*P*	Median (months)	95%CI	*P*
PFS			<0.001			<0.001
Signature^{PFS-CG} ≤0.0640	32	30-39		29	24-36	
Signature^{PFS-CG} >0.0640	16	14-18		18	15-21	
OS			<0.001			<0.001
Signature^{OS-CG} ≤0.1969	45	40-50		42	34-48	
Signature^{OS-CG} >0.1969	25	24-27		26	21-29	

OS: Overall survival; PFS: Progression-free survival; C: Clinical factors; G: Gene abnormalities; CG: Clinical factors plus gene abnormalities; CI: Confidence interval

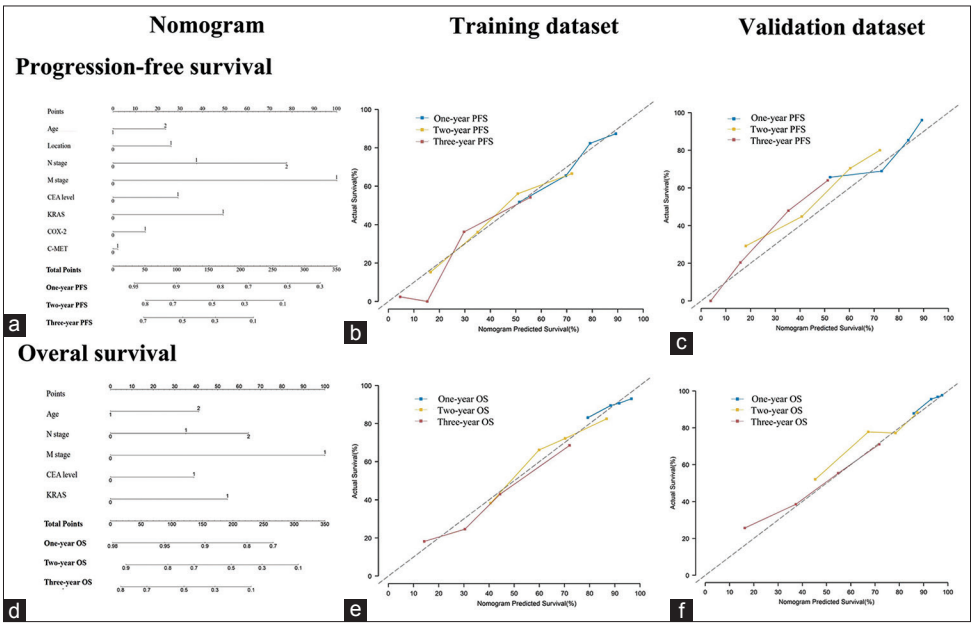

Figure 4: Nomogram and calibration based on Signature^{PFS-CG} and Signature^{OS-CG}. For PFS, the nomogram based on Signature^{PFS-CG} (a) has good calibration, both in training (b) and validation (c) datasets. For OS, the nomogram based on Signature^{OS-CG} (d) also has good calibration, both in training (e) and validation (f) datasets. AUC: Area under the curve; PFS: Progression-free survival; OS: Overall survival; C: Clinical factors; G: Gene abnormalities; CG: Clinical factors and gene abnormalities

clinical practice (expression of *COX-2, C-MET* and *KRAS* mutation status) and combined them into an integral signature with testing in both training and validation datasets. The accuracy of our models was comparable to a previous study,[25] especially for the 3-year PFS. However, we used fewer genes (3 vs. 13), which are widely tested in CRC management, so our model may be more convenient for clinical practice. In addition, we constructed models not only for PFS but also for OS.

In addition, our results indicated that some prognostic factors were common for both PFS and OS, such as age, N stage, M stage, CEA level and KRAS status. However, tumor location and the expression levels of COX-2 and C-MET were only informative for PFS. These data may have suggested that N and M stages were more important than T stage because cases with metastases in lymph nodes or beyond the primary site were more aggressive and more difficult to control by current therapeutic approaches.[28] In addition, tumor location might influence the short-term clinical response to anti-tumor therapies, especially for chemotherapy.[29] Third, a high expression of both COX-2 and C-MET had synergistic effects in regulating the malignant behavior of CRC, and these proteins are important markers for invasion and metastasis as well as potential molecular targets for specific treatments.[30,31] In addition, OS is the consequence of multiple effects, not only including therapeutic options but also the natural behaviors of the tumor. Thus, KRAS status was informative for both PFS and OS. Notably, *KRAS* mutation has been shown to be associated with resistance to anti-EGFR therapies, and influences the response of some CRC patients to targeted medicine.

However, there were some limitations in our study. First, because we did not have PD or death information within 2-years of follow-up for many patients, the AUCs of 1- and 2-year PFS and OS were less accurate compared with those for 3-years. Second, because of the retrospective nature of our study, multiple confounding factors likely existed after

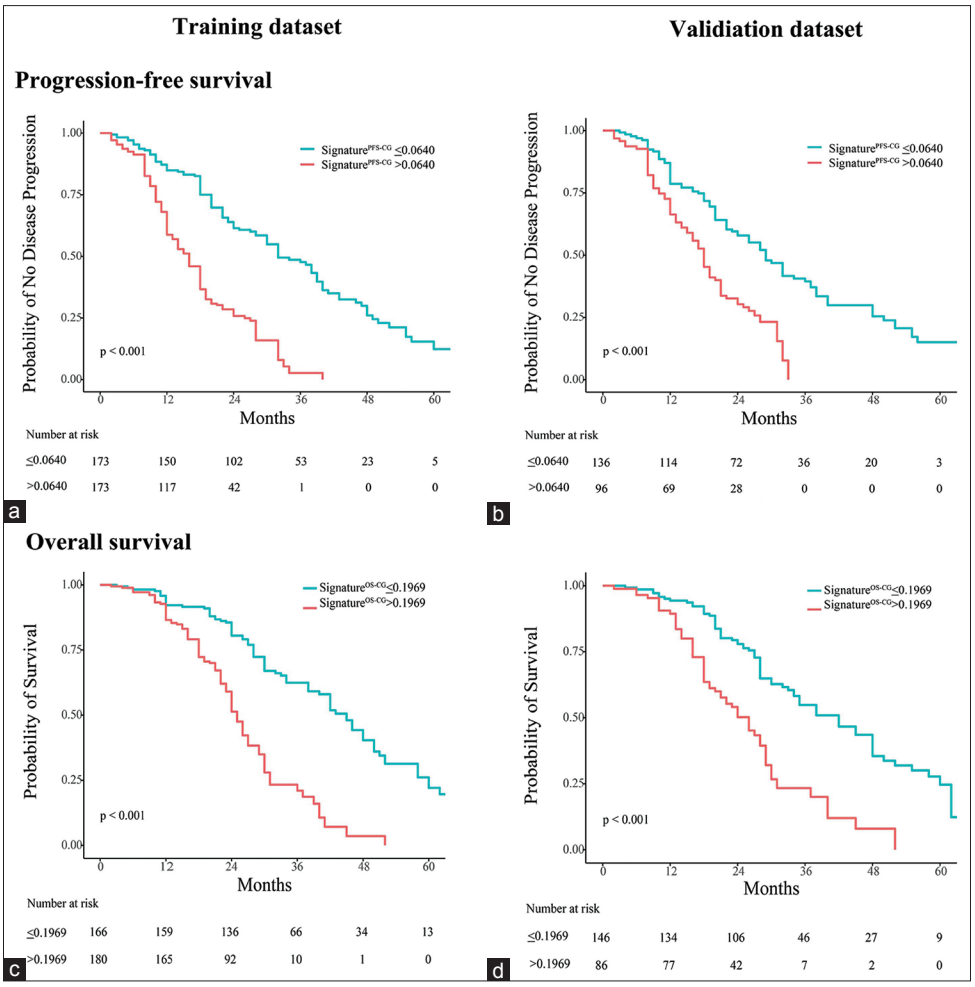

Figure 5: Signatures based subgroup Kaplan-Meier analysis. Separated by the median of Signature^{PFS-CG} at 0.0640, the two subgroups have significant statistical difference in PFS, both for training (a) and validation (b) datasets. Separated by the median of Signature^{OS-CG} at 0.1969, similar results were observed in training (c) and validation (d) datasets. PFS: Progression-free survival; OS: Overall survival; C: Clinical factors; G: Gene abnormalities; CG: Clinical factors and gene abnormalities

PD that affected OS, and as such subtle inconsistencies were found regarding the AUC comparison between the training and validation models, in which Signature^{OS-CG} was the most accurate in the training dataset but was slightly inferior in the validation dataset. Third, considering racial and regional differences, whether our results could be applied for Western patients requires further exploration. Fourth, other factors such as Plastin3, SPC18 and PTEN[32-34] may also have contributed to CRC survival, and with improvements in testing methods and interpretion of these genes, we might be able to use them to improve the performance of our models. All of these aspects necessitate future study with a larger cohort, especially with prospective data.

In conclusion, we developed abnormal COX-2/C-MET/KRAS expression-based prognostic models for both PFS and OS and verified their accuracy in both training and validation datasets. We concluded that with proper integration, combining these three genes with clinical factors provided more prognostic information for CRC and may facilitate individualized treatment decision-making in the future.

Acknowledgments

This work was supported by a grant from the Science and Technology Program of Guangzhou, China (grant no. 201804010426).

Financial support and sponsorship

This work was supported by a grant from the Science and Technology Program of Guangzhou, China (grant no. 201804010426).

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global

cancer statistics, 2012. CA Cancer J Clin 2015;65:87-108.
2. Franko J, Shi Q, Meyers JP, Maughan TS, Adams RA, Seymour MT, *et al*. Prognosis of patients with peritoneal metastatic colorectal cancer given systemic therapy: An analysis of individual patient data from prospective randomised trials from the analysis and research in cancers of the digestive system (ARCAD) database. Lancet Oncol 2016;17:1709-19.
3. Tabernero J, Lenz HJ, Siena S, Sobrero A, Falcone A, Ychou M, *et al*. Analysis of circulating DNA and protein biomarkers to predict the clinical activity of regorafenib and assess prognosis in patients with C-METastatic colorectal cancer: A retrospective, exploratory analysis of the CORRECT trial. Lancet Oncol 2015;16:937-48.
4. Salazar R, Roepman P, Capella G, Moreno V, Simon I, Dreezen C, *et al*. Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer. J Clin Oncol 2011;29:17-24.
5. Willis JA, Vilar E. Refining prognosis in early-stage colorectal cancer: One or multiple genes at a time? Ann Oncol 2017;28:1686-8.
6. Peng L, Zhou Y, Wang Y, Mou H, Zhao Q. Prognostic significance of COX-2 immunohistochemical expression in colorectal cancer: A C-METa-analysis of the literature. Plos One 2013;8:e58891.
7. Wang JY, Sun J, Huang MY, Wang YS, Hou MF, Sun Y, *et al*. STIM1 overexpression promotes colorectal cancer progression, cell motility and COX-2 expression. Oncogene 2015;34:4358-67.
8. Shin IY, Sung NY, Lee YS, Kwon TS, Si Y, Lee YS, *et al*. The expression of multiple proteins as prognostic factors in colorectal cancer: Cathepsin D, p53, COX-2, epidermal growth factor receptor, C-erbB-2, and Ki-67. Gut Liver 2014;8:13-23.
9. Takahashi N, Iwasa S, Taniguchi H, Sasaki Y, Shoji H, Honma Y, *et al*. Prognostic role of ERBB2, MET and VEGFA expression in metastatic colorectal cancer patients treated with anti-EGFR antibodies. Br J Cancer 2016;114:1003-11.
10. Pietrantonio F, Oddo D, Gloghini A, Valtorta E, Berenato R, Barault L, *et al*. MET-driven resistance to dual EGFR and BRAF blockade may be overcome by switching from EGFR to C-MET Inhibition in BRAF-mutated colorectal cancer. Cancer Discov 2016;6:963-71.
11. McCubrey JA, Steelman LS, Abrams SL, Abrams SL, Wong EW, Chang F, *et al*. Roles of the RAF/MEK/ERK and PI3K/PTEN/AKT pathways in malignant transformation and drug resistance. Adv Enzyme Regul 2006;46:249-79.
12. Mendelsohn J, Baselga J. Epidermal growth factor receptor targeting in cancer. Semin Oncol 2006;33:369-85.
13. De Roock W, De Vriendt V, Normanno N, Ciardiello F, Tejpar S. KRAS, BRAF, PIK3CA, and PTEN mutations: Implications for targeted therapies in metastatic colorectal cancer. Lancet Oncol 2011;12:594-603.
14. National Comprehensive Cancer Network. Rectal Cancer (Version 3.2017), 2017. Available from: https://www.nccn.org/professionals/physician_gls/pdf/rectal.pdf. [Last accessed on 2016 Oct 13].
15. National Comprehensive Cancer Network. Colon Cancer (Version 2.2017), 2017. Available from: https://www.nccn.org/professionals/physician_gls/pdf/colon.pdf. [Last accessed on 2016 Oct 13].
16. Bottaro DP, Rubin JS, Faletto DL, Chan AM, Kmiecik TE, Vande Woude GF, *et al*. Identification of the hepatocyte growth factor receptor as the c-met proto-oncogene product. Science 1991;251:802-4.
17. Ion Ampliseq Designer, [https://www.ampliseq.com/browse.action].
18. Xu Y, Bi R, Xiao Y, Tu X, Li M, Li A, *et al*. Low frequency of BRAF and KRAS mutations in Chinese patients with low-grade serous carcinoma of the ovary. Diagn Pathol 2017;12:87.
19. Gao J, Wu H, Wang L, Zhang H, Duan H, Lu J, *et al*. Validation of targeted next-generation sequencing for RAS mutation detection in FFPE colorectal cancer tissues: Comparison with Sanger sequencing and ARMS-Scorpion real-time PCR. BMJ Open 2016;6:e009532.
20. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, *et al*. New response evaluation criteria in solid tumours: Revised RECIST guideline (Version 1.1). Eur J Cancer 2009;45:228-47.
21. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics 2000;56:337-44.
22. Cooke VG, LeBleu VS, Keskin D, Khan Z, O'Connell JT, Teng Y, *et al*. Pericyte depletion results in hypoxia-associated epithelial-to-mesenchymal transition and metastasis mediated by met signaling pathway. Cancer Cell 2012;21:66-81.
23. Tuynman JB, Vermeulen L, Boon EM, Kemper K, Zwinderman AH, Peppelenbosch MP, *et al*. Cyclooxygenase-2 inhibition inhibits c-Met kinase activity and Wnt activity in colon cancer. Cancer Res 2008;68:1213-20.
24. Wu QB, Sun GP. Expression of COX-2 and HER-2 in colorectal cancer and their correlation. World J Gastroenterol 2015;21:6206-14.
25. Tian X, Zhu X, Yan T, Yu C, Shen C, Hu Y, *et al*. Recurrence-associated gene signature optimizes recurrence-free survival prediction of colorectal cancer. Mol Oncol 2017;11:1544-60.
26. Summers MG, Smith CG, Maughan TS, Kaplan R, Escott-Price V, Cheadle JP. BRAF and NRAS locus-specific variants have different outcomes on survival to colorectal cancer. Clin Cancer Res 2017;23:2742-9.
27. Chu QD, Zhou M, Medeiros K, Peddi P. Positive surgical margins contribute to the survival paradox between patients with stage IIB/C (T4N0) and stage IIIA (T1-2N1, T1N2a) colon cancer. Surgery 2016;160:1333-43.
28. Kong X, Li J, Cai Y, Tian Y, Chi S, Tong D, *et al*. A modified TNM staging system for non-metastatic colorectal cancer based on nomogram analysis of SEER database. BMC Cancer 2018;18:50.
29. Patel M, McSorley ST, Park JH, Roxburgh CSD, Edwards J, Horgan PG, *et al*. The relationship between right-sided tumour location, tumour microenvironment, systemic inflammation, adjuvant therapy and survival in patients undergoing surgery for colon and rectal cancer. Br J Cancer 2018;118:705-12.
30. Sui H, Zhou S, Wang Y, Liu X, Zhou L, Yin P, *et al*. COX-2 contributes to P-glycoprotein-mediated multidrug resistance via phosphorylation of c-Jun at Ser63/73 in colorectal cancer. Carcinogenesis 2011;32:667-75.
31. Li F, Zhu YT. HGF-activated colonic fibroblasts mediates carcinogenesis of colonic epithelial cancer cells via PKC-cMET-ERK1/2-COX-2 signaling. Cell Signal 2015;27:860-6.
32. Yokobori T, Iinuma H, Shimamura T, Imoto S, Sugimachi K, Ishii H, *et al*. Plastin3 is a novel marker for circulating tumor cells undergoing the epithelial-mesenchymal transition and is associated with colorectal cancer prognosis. Cancer Res 2013;73:2059-69.
33. Hattori T, Sentani K, Naohide O, Sakamoto N, Yasui W. Clinicopathological significance of SPC18 in colorectal cancer: SPC18 participates in tumor progression. Cancer Sci 2017;108:143-50.
34. Trotman LC, Wang X, Alimonti A, Chen Z, Teruya-Feldstein J, Yang H, *et al*. Ubiquitination regulates PTEN nuclear import and tumor suppression. Cell 2007;128:141-56.

SUPPLEMENTARY APPENDIX

Table S1: Univariate and multivariate analysis of the three signatures for progression-free survival (PFS) (including COX-2 and C-MET)

Factor	Univariate analysis			Multivariate analysis		
	HR	95%CI	*P*	HR	95%CI	*P*
Signature^{PFS-C}						
Age	1.082	0.841-1.392	0.541	1.269	0.979-1.646	0.072
Location	1.282	0.995-1.653	0.055	1.355	1.047-1.754	0.021
N stage						
N1 stage	1.935	1.374-2.724	<.001	1.849	1.315-2.601	<.001
N2 stage	2.883	2.022-4.109	<.001	2.614	1.834-3.725	<.001
M stage	2.672	2.001-3.569	<.001	2.689	1.996-3.623	<.001
CEA level	1.779	1.311-2.415	<.001	1.45	1.044-2.013	0.027
Signature^{PFS-G}						
KRAS	1.86	1.392-2.485	<.001	1.814	1.273-2.585	0.001
COX-2	0.895	0.695-1.153	0.392	1.191	0.886-1.601	0.248
C-MET	1.283	0.975-1.687	0.075	1.153	0.872-1.526	0.318
Signature^{PFS-CG}						
Age	1.082	0.841-1.392	0.541	1.282	0.985-1.669	0.064
Location	1.282	0.995-1.653	0.055	1.308	1.010-1.694	0.042
N stage						
N1 stage	1.935	1.374-2.724	<.001	1.489	1.030-2.154	0.034
N2 stage	2.883	2.022-4.109	<.001	2.269	1.567-3.286	<.001
M stage	2.672	2.001-3.569	<.001	2.798	2.072-3.779	<.001
CEA level	1.779	1.311-2.415	<.001	1.349	0.966-1.884	0.079
KRAS	1.86	1.392-2.485	<.001	1.655	1.135-2.414	0.009
COX-2	0.895	0.695-1.153	0.392	1.155	0.854-1.563	0.350
C-MET	1.283	0.975-1.687	0.075	1.029	0.774-1.369	0.842

HR: Hazard ratio; CI: Confidence interval; C: Clinical factors; G: Gene abnormalities; CG: Clinical factors and gene abnormalities

Table S2: Univariate and multivariate analysis of the three signatures for overall survival (OS)

Factor	Univariate analysis			Multivariable analysis		
	HR	95%CI	*P*	HR	95%CI	*P*
Signature^{OS-C}						
Age	1.263	0.955~1.670	0.102	1.954	1.102~1.954	0.009
N stage						
N1 stage	1.993	1.378~2.884	<.001	2.747	1.313~2.747	0.001
N2 stage	2.602	1.777~3.809	<.001	3.38	1.573~3.380	<.001
M stage	2.75	2.015~3.752	<.001	3.712	1.959~3.712	<.001
CEA level	1.936	1.375~2.726	<.001	2.331	1.141~2.331	0.007
Signature^{OS-G}						
KRAS	2.147	1.535~3.005	<.001	2.631	1.311~2.631	<.001
Signature^{OS-CG}						
Age	1.263	0.955~1.670	0.102	2.074	1.163~2.074	0.003
N stage						
N1 stage	1.993	1.378~2.884	<.001	2.177	0.977~2.177	0.065
N2 stage	2.602	1.777~3.809	<.001	2.954	1.339~2.954	0.001
M stage	2.75	2.015~3.752	<.001	4.039	2.124~4.039	<.001
CEA level	1.936	1.375~2.726	<.001	2.189	1.059~2.189	0.023
KRAS	2.147	1.535~3.005	<.001	2.63	1.232~2.630	0.002

HR: Hazard ratio; CI: Confidence interval; C: Clinical factors; G: Gene abnormalities; CG: Clinical factors plus gene abnormalities

Table S3: Univariate and multivariate analysis of the three signatures for PFS (exclusive of COX-2 and C-MET)

Factor	Univariate analysis			Multivariate analysis		
	HR	95%CI		95%CI	HR	95%CI
Signature^{PFS-C}						
Age	1.082	0.841~1.392	0.541	1.269	0.979~1.646	0.072
Location	1.282	0.995~1.653	0.055	1.355	1.047~1.754	0.021
N stage						
N1 stage	1.935	1.374~2.724	<.001	1.849	1.315~2.601	<.001
N2 stage	2.883	2.022~4.109	<.001	2.614	1.834~3.725	<.001
M stage	2.672	2.001~3.569	<.001	2.689	1.996~3.623	<.001
Signature^{PFS-G}						
CEA level	1.779	1.311~2.415	<.001	1.522	1.103~2.099	0.011
KRAS	1.86	1.392~2.485	<.001	1.634	1.206~2.214	0.002
Signature^{PFS-CG}						
Age	1.082	0.841~1.392	0.541	1.303	1.003~1.692	0.047
Location	1.282	0.995~1.653	0.055	1.324	1.024~1.712	0.033
N stage						
N1 stage	1.935	1.374~2.724	<.001	1.478	1.021~2.140	0.038
N2 stage	2.883	2.022~4.109	<.001	2.243	1.552~3.242	<.001
M stage	2.672	2.001~3.569	<.001	2.839	2.104~3.831	<.001
CEA level	1.779	1.311~2.415	<.001	1.375	0.990~1.909	0.057
KRAS	1.86	1.392~2.485	<.001	1.523	1.090~2.127	0.014

OS: Overall survival; PFS: Progression-free survival; HR: Hazard ratio; CI: Confidence interval; C: Clinical factors; G: Gene abnormalities; CG: Clinical factors and gene abnormalities.

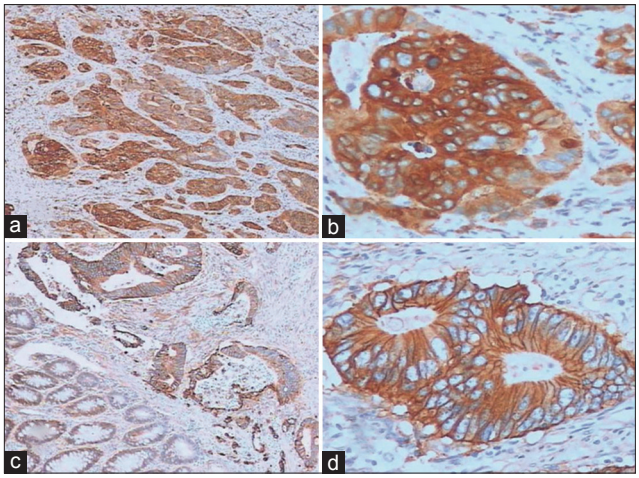

Figure S1: Analysis of COX-2 and C-MET expression by immunohistochemistry in colorectal carcinomas. (a) Strong COX-2 staining in tumor nests (40×); (b) Positive COX-2 staining in the cytoplasm (200×); (c) Strong C-MET staining in tumor nests (40×); (d) Positive membrane staining of C-MET in tumor cells (200×)

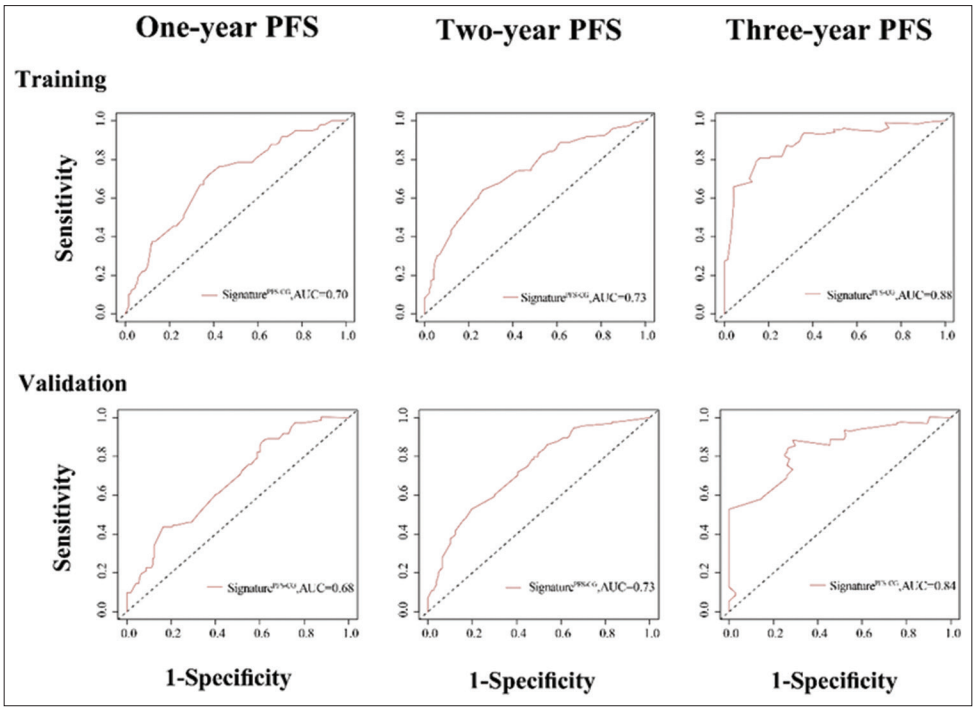

Figure S2: Comparing the accuracy of PFS by the three signatures without COX-2 and C-MET. For the 1-, 2- and 3-year PFS, the AUC of Signature^{PFS-CG} in the training dataset were 0.70, 0.73 and 0.88. Those in the validation dataset were 0.67, 0.73 and 0.84. PFS: Progression-free survival; C: Clinical factors; G: Gene abnormalities; CG: Clinical factors and gene abnormalities

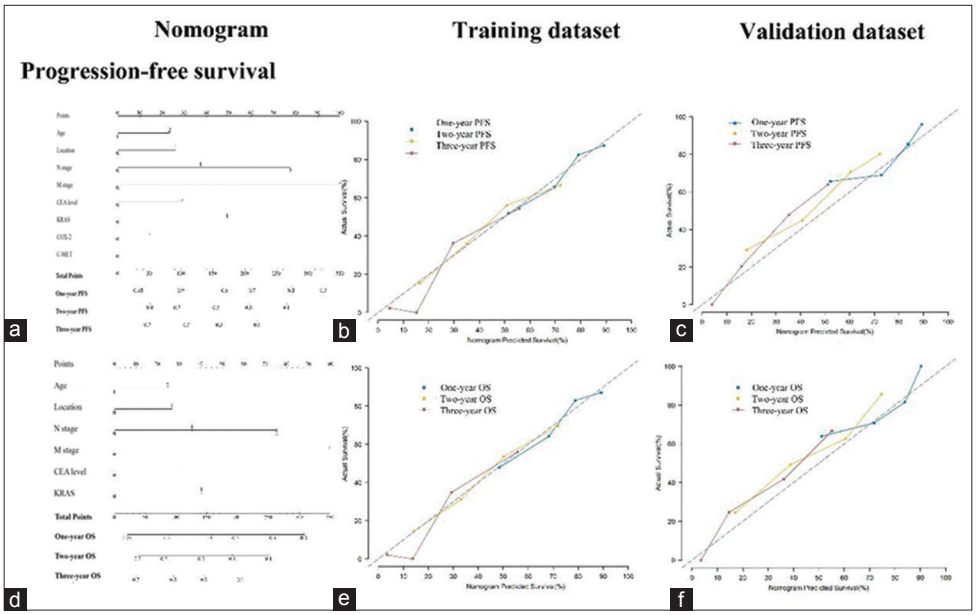

Figure S3: Nomogram and calibration based on Signature^{PFS-CG} with and without COX-2 and C-MET. For PFS, the nomogram based on Signature^{PFS-CG} with COX-2 and C-MET (a) has good calibration, both in training (b) and validation (c) datasets. The nomogram based on Signature^{PFS-CG} without COX-2 and C-MET (d) also has good calibration, both in training (e) and validation (f) datasets. PFS: Progression-free survival; C: Clinical factors; G: Gene abnormalities; CG: Clinical factors and gene abnormalities