Nomogram Integrating Genomics with Clinicopathological Features Improves Prognosis Prediction for Colorectal Cancer

Yongfu Xiong†, Wenxian You†, Min Hou‡, Linglong Peng†, He Zhou†, Zhongxue Fu*†

†Department of Gastrointestinal Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China.
‡Department of Oncology, Affiliated Hospital of North Sichuan Medical College, Nanchong, China.
*†Equal contributor
†Correspondence: Zhongxue Fu, Ph.D. Department of Gastrointestinal Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing 400016, China.
E-mail: fzx19990521@126.com

Running title
A Nomogram to Predict Colorectal Cancer Prognosis

Keywords: CRC, genomic signature, nomogram, clinicopathological features

Disclosure of potential conflicts of interest
The authors have no potential conflicts of interest to disclose.

Financial support
This work was financially supported by a grant from the National Natural Science foundation of China (grant no. 81572319; project recipient: ZhongXue Fu).
Abstract
The current tumor staging system is insufficient for predicting the outcomes of patients with colorectal cancer (CRC) because of its phenotypic and genomic heterogeneity. Integrating gene expression signatures with clinicopathological factors may yield a predictive accuracy exceeding that of the currently available system. Twenty-seven signatures that used gene expression data to predict CRC prognosis were identified and re-analyzed using bioinformatic methods. Next, clinically annotated CRC samples (n=1710) with the corresponding expression profiles, that predicted a patient's probability of cancer recurrence, were pooled to evaluate their prognostic values and establish a clinicopathologic-genomic nomogram. Only 2 of the 27 signatures evaluated showed a significant association with prognosis and provided a reasonable prediction accuracy in the pooled cohort (HR = 2.46, 95% CI = 1.183-5.132, p < 0.001, AUC = 60.83; HR =2.33, 95% CI = 1.218-4.453, p < 0.001, AUC = 71.34). By integrating the above signatures with prognostic clinicopathological features, a clinicopathologic-genomic nomogram was cautiously constructed. The nomogram successfully stratified CRC patients into three risk groups with remarkably different DFS rates and further stratified stage II and III patients into distinct risk subgroups. Importantly, among patients receiving chemotherapy, the nomogram determined that those in the intermediate- (HR = 0.98, 95% CI = 0.255-0.679, p < 0.001) and high-risk (HR = 0.67, 95% CI = 0.469-0.957, p = 0.028) groups had favorable responses.

Implications: These findings offer evidence that genomic data provides independent and complementary prognostic information, and incorporation of this information refines the prognosis of CRC.
**Introduction**

Colorectal carcinoma (CRC) is the third most commonly diagnosed malignant disease and the second leading cause of cancer death worldwide (1). Despite advances in CRC screening, diagnosis, and curative resection, its prognosis is not entirely satisfactory because optimal management and individual therapy strategies still present great challenges, as CRC is a well-recognized heterogeneous disease. Currently, treatment decisions and prognoses for patients with CRC are primarily driven by assessment of the tumor-node-metastasis (TNM) staging system, which is based merely on anatomical information (2). Previous studies revealed that patients with stage I CRC have a 5-year survival rate of approximately 93%, which decreases to approximately 80% for patients with stage II disease and to 60% for patients with stage III disease (3). However, discrepancies in the survival outcomes of patients at the same stage and receiving similar treatments is commonly observed.

More importantly, according to the current TNM stage, adjuvant therapy is generally recommended for all patients with stage III disease (4). However, patients with T1-2N1M0 tumors (stage IIIA) have significantly higher survival rates than those with stage IIB tumors (3), suggesting that stratifying patients at high risk of recurrence, who are most likely to benefit from adjuvant therapy, is critical. Therefore, substantial effort has been made to discover new clinicopathological indicators to optimize the current staging system. To date, some clinicopathological features, such as emergency presentation, adjacent organ involvement (T4), intestinal perforation or obstruction, high tumor grade, inadequate sampling of lymph nodes and lymphatic/vascular invasion, with prognostic value to classify CRC patients who are at a high risk of recurrence have been applied in clinical practice. However, these factors are all relatively weak and insufficient to identify CRC patients who may benefit from adjuvant therapy, which leads to potential under- or over-treatment (5).

CRC biological behavior (e.g., recurrence, metastasis, drug resistance) is a tightly regulated process that requires the aberrant expression of related genes to empower carcinoma cells with their corresponding abilities. Accordingly, before clinicopathologically detectable changes occur, the underlying alterations necessary
for recurrence have already occurred in the primary CRC, providing the possibility to
develop robust prognostic tools by using multiple genes in combination (6). During
the last decade, gene signatures have shown great promise in predicting the long-term
outcomes and treatment responses of individual patients (7). Of note, because of the
outstanding ability to predict the prognoses of patients with breast cancer, multigene
assays, such as Oncotype DX and MammaPrint, have been successfully approved by
the US Food and Drug Administration and are available in routine clinical practice (8).
Actually, genomic signatures to predict the prognosis of CRC have also been
continually developed in the past 10 years, but none are commercially used in the
clinic (9). For example, Agesen et al. established a 13-gene expression classifier,
ColoGuideEx, for prognosis predictions specific to patients with stage II CRC (10).
Based on the essential role of lipid metabolism in carcinogenesis, Teodoro et al.
identified a group of 4 genes that predict survival in intermediate-stage colon cancer
patients, allowing the delineation of a high-risk group that may benefit from adjuvant
therapy. In addition, Smith (11), Chen (12), Oh (13) and Popovici et al. (14) also
published their own signatures based on different mechanisms. Despite a large
number of studies, no signatures have remained credible in either meta-analyses or
prospective trials (15). However, these published signatures clearly show low
prediction accuracies but moderate clinical usefulness (15). Moreover, when confined
to a specific CRC stage, promising results regarding risk stratification have also been
reported. Therefore, investigating the predictive ability of these published signatures
on comprehensive large-scale datasets and identifying whether any can be used to
clinically guide treatment decisions is necessary.

Currently, while doubts about the predictive value of clinicopathological features
are increasing, they still provide the most reliable guidelines for the prognostication
and treatment of CRC (16). Thus, we hypothesized that integrating genomic
signatures with clinicopathological features in a model would yield a predictive
accuracy exceeding that of the currently available prognostic system. Nomogram is a
statistical prediction model that combines multiple prognostic factors to make
intuitive graphical and individualized predictions (17). Here, we aimed to apply a
systematic approach to evaluate the clinical usefulness of CRC-related signatures and then construct a composite clinicopathologic-genomic nomogram by integrating factors with potential prognostic value in a training set. Moreover, using another independent set, the capacity of the nomogram to stratify CRC patients most likely to benefit from chemotherapy was further validated.

Materials and Methods

Patients and prognostic signatures

To identify gene expression data arrayed using the Affymetrix platform with clinically annotated data, we systematically searched Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/), ArrayExpress (http://www.ebi.ac.uk/arrayexpress/) and related literature with the keywords “colorectal cancer”, “CRC”, “colon cancer”, “survival”, “relapse”, “recurrence”, “prognostic”, “prognosis” and “outcomes” published through Aug 1, 2017. For some datasets whose clinical data did not accompany their gene expression profiles, we either searched the supplements or contacted one or more of the investigators to obtain the missing information. Moreover, datasets with small sample sizes and duplications were excluded. Raw microarray data and the corresponding clinical data of these datasets were retrieved and manually organized when available. Only patients diagnosed with colorectal cancer having clinicopathological and survival information available were included. Patients with follow-up or survival times of less than 1 month as well as patients with missing or insufficient data on age, local invasion, lymph node metastasis, distant metastasis and TNM stage were excluded from subsequent processing. Eventually, all patients satisfying the inclusion criteria were combined and summarized in Supplementary Table S1 and S2.

Expression data processing

For raw CEL files available from Affymetrix microarrays, the data were normalized and annotated using a MAS5 algorithm and the corresponding annotation files from R Bioconductor to obtain summarized values for each probeset; otherwise,
pre-processed data as provided by the contributors were used. For each sample in every data set, measurements without a gene annotation were excluded, and multiple probesets corresponding to a single gene were summarized into a gene symbol by taking the most variable probeset measured by the interquartile range (IQR).

**Identification and analysis of gene signatures potentially related to CRC prognosis**

Gene signatures potentially related to CRC prognosis were systematically retrieved from PubMed. The search was restricted to recent papers to increase validity (from January 2004 to August 2017). The selection criteria are detailed in Supplementary Figure S1. Articles that provided a list of differentially expressed genes in primary tumor samples associated with CRC prognosis were included in our study. Studies based on tissue microarray and those that were exclusively focused on differences between stages or between primary tumors and metastases were excluded. The signatures finally included in our analysis are described in Table 1 (detailed descriptions provided in Supplementary Table S3). In addition, the probesets or genes of those signatures were re-annotated using the SOURCE web tool to address the retired gene symbols and their differences in the tested platforms. The re-annotated genes were then subjected to biological function enrichment analysis, and the online analytical tool DAVID (Database for Annotation, Visualization and Integrated Discovery) (18) was used to enrich gene ontology (GO) functions and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. GO terms and KEGG pathways with significant enrichment false discovery rate (FDR) values less than 0.05 were selected for further analysis. In addition, genes from the abovementioned signatures were mapped and imported into the Retrieval of Interacting Genes/Proteins (STRING) 9.1 database, which queried the human protein-protein interaction network for interactions between effective linkers and seeds to construct a functional subnetwork. Cluster analyses were performed using correlation distance metrics and the average linkage agglomeration algorithm (R package nclust version 1.9.0; ref. 34). Non-negative matrix factorization (NMF) was done with the NMF package (version 0.20.5) and standard strategies.
**Subclass prediction of CRC patients in the training set**

The re-preprocessed microarray dataset, which represents the genomic features of the individual, were classified with the prognostic signatures identified above using the nearest template prediction (NTP) method (19) as implemented in Gene Pattern software (Broad Institute of Harvard and MIT, Boston, MA) (20). NTP requires only a list of pre-specified template signature genes and a dataset to be tested and not a corresponding training dataset to capture good and poor gene expression patterns in each sample. Briefly, a template containing representative expression patterns of the signature genes was defined based on published gene signatures from their respective studies. The proximity of the signature gene expression patterns of the sample to the template was evaluated by calculating the cosine distance. The FDR was used to correct the p-values for multiple hypothesis testing, and prediction analysis was performed separately for each dataset. A prediction of good or poor for the related signature was determined based on FDR predictions <0.05, and the remaining samples with intermediate expression levels of the correlated genes in the signature were classified as uncertain.

Concordance among these predictions was evaluated using unsupervised clustering according to the Cramer’s V coefficient of the paired prediction overlap as previously described (21). Cramer’s V statistic values range from 0 to 1, with values between 0.36 and 0.49 indicating substantial correlation, and values higher than 0.5 indicating strong correlation.

**Development, comparison, and validation of the prognostic models**

All the patients included in this study were randomly separated into training and validation sets. Disease-free survival (DFS) was defined as the interval between the day that surgery was performed and the day that recurrence was first detected. If recurrence was not diagnosed, the date of death due to CRC or last follow-up was used. Overall survival (OS) was calculated from the date of diagnosis or surgery to the date of death or last follow-up. Continuous variables were expressed as median (IQR) or median (range) values, and group comparisons were performed by the Student’s t test or the Wilcoxon rank sum test. Categorical variables were expressed as
percentages, and group comparisons were performed by Pearson’s χ² test or the Fisher’s exact test. Median follow-up was calculated using the reverse Kaplan-Meier method (22). In the training set, survival curves for different variable values were generated using Kaplan-Meier estimates and compared using the log-rank test. Variables that achieved significance at p values less than 0.05 were subjected to multivariate analyses via the Cox regression model. Then, statistical analyses to identify independent prognostic factors were conducted. Based on the multivariate analysis results, a prognostic nomogram was established using a backward stepwise Cox proportional hazard model. The entire population was divided into three risk groups (high, intermediate, low) according to the tertiles of the total scores given by the established nomogram in the training set. Concordance index (c-index) values were used to measure the discrimination performance. Finally, for external validation, the total scores of each patient in the validation set were calculated according to the proposed nomogram to verify its generalization. Three risk groups were determined by the tertiles defined in the training set, and the respective Kaplan-Meier survival curves were delineated. All statistical tests performed were two-sided, and p values less than 0.05 were considered statistically significant.

Results

Screening and analysis of eligible samples and published signatures

To comprehensively evaluate their clinical usefulness, we retrieved all of the published signatures. Accordingly, 81 signatures published as valid prognostic tools in CRC were obtained from 507 potentially related articles. Among these signatures, those with genes not clearly described in their respective studies or investigating the prognostic value base on tissue array, immune cells, circulation blood or RT-PCR results were further excluded (see flowchart in Supplementary Figure S1 for details). Finally, a total of 27 signatures from 26 studies met all of the inclusion criteria and were retained for subsequent analysis (Supplementary Table S3). Among these signatures, almost all were obtained from experiments based on CRC tissues except for one, which was derived from an animal experiment (23). The 27 signatures
contained 1274 total genes, among which 1041 were unique, and the signature sizes ranged from 4 to 264 (Figure 1A). The top overlapping genes in the above signatures were FN1 (ten times), CYP1B1 and POSTN (six times), 7 genes (five times), 7 genes (four times), 29 genes (three times), and 107 genes (twice). As is clearly shown in Figure 1A, none of these genes appeared in all signatures. In addition, some repeated genes in different signatures even presented opposite prognostic values. This lower-than-expected overlap, at least in part, accounted for the low reproducibility of the signatures and the failure in clinical practice.

Accumulating evidence strongly suggests that the prognosis of CRC patients, such as recurrence and metastasis, is tightly regulated by aberrantly expressed genes via specific biological processes (16). To gain insight into the underlying biological meanings of the gene set obtained from the above signatures, enrichment (GO and KEGG) and protein-protein interaction (PPI) analyses were conducted (Supplementary Table S4). Figure 1B-D shows that specific GO categories closely related to CRC prognosis, such as cell cycle (FDR < 0.0001), regulation of cell proliferation (FDR < 0.0001), extracellular matrix (FDR < 0.0001), and chemokine activity (FDR < 0.0001), were significantly enriched. Additionally, some KEGG pathways well known to be involved in CRC metastasis and recurrence were enriched, including colorectal cancer metastasis signaling (FDR < 0.0001), wnt-β catenin signaling (FDR = 0.0002), VEGF signaling (FDR = 0.0008), and chemokine signaling (FDR = 0.0021) (Figure 1E). More importantly, a visualized network of these gene interactions revealed that the abovementioned pathways not only form highly integrated modules but also play essential roles in the entire network (Figure 1F). These findings suggested that notwithstanding the low reproducibility in the system test and clinical validation, genes obtained from CRC signatures actually have potential prognostic value.

Furthermore, to assess the global performance of these signature and construct a composite nomogram, publicly available datasets with whole-genome profiles and corresponding clinical information in CRC patients were screened and downloaded (see flowchart in Supplementary Figure S1 for details). Ultimately, 1710 CRC
patients pertaining to 9 cohorts (Supplementary Table S5) met the set criteria and were retained. Although the profiles were obtained using different technologies and based on fresh or formalin-fixed and paraffin-embedded (FFPE) specimens as source material, they showed no evidence of cohort-bias clustering by principal component analysis (PCA, Supplementary Figure S2A-C) or hierarchical clustering (Supplementary Figure S2D), suggesting that patients from different cohorts in the present study could be blended together. Moreover, to balance the baseline characteristics, all eligible patients were randomly divided into a training set (n = 855) and a validation set (n = 855). The demographic and major clinicopathological characteristics of the patients are summarized in Supplementary Table S2. As we expected, no significant differences were observed between the statistical properties of the training and validation sets.

Global prognostic performance of the published signatures

To determine the correlations of signatures and CRC patient outcomes, the prognostic performances of the 27 signatures were evaluated in the training set using a modified NTP method as previously described (19). Among the 27 signatures evaluated, all except one (Teodoro’s (24) signature had the only poor prognostic gene) were able to confidently stratify patients (FDR <0.05) into good and poor subgroups. Table 1 and Figure 2A summarize the prediction results obtained for each of the 855 patients. Popovici’s signature [34] was the most prevalent prediction in the training set (83.6%; 615 of 855), whereas Agesen’s signature [42] was securely identified in only 19.9% (167 of 855) of CRC patients. Of note, conflicting prognostic outcomes were commonly observed among the training set, and only 34% patients had consistent results (good or poor prognosis) in more than 5 signatures. However, as demonstrated in Figure 2A, the overall tendency of the NTP analysis was consistent and roughly consistent with previous studies (25). In addition, global Cramer’s V coefficients were calculated for each signature to gain insight into their concordances. Unsupervised clustering based on these coefficients indicated a substantial association among these signatures (Figure 2B). Obviously, 5 signatures obtained from the studies of Oh (13), Minh (11), Chang (9), Jiang (26), and Fritzmann et al. (27) were at least moderately
correlated with each other. Moreover, another 4 signatures (Toshiaki’s, Aziz’s, Chen’s, Xu’s) were also clustered with similar coefficients. These findings are largely consistent with the initial purpose of each signature, which further confirms the potential prognostic value of these signatures.

We then applied Cox regression analysis to assess whether any signatures were statistically significant for CRC-related recurrence/progression and independent of clinical features in the training set. The univariate and clinical factor-adjusted multivariate analysis results are displayed in Supplementary Table S6 and Figure 2C. Univariate analyses revealed that age (HR = 1.12, 95% CI = 1.021-1.269, \( p = 0.048 \)), local invasion (T stage; HR = 2.21, 95% CI = 1.139-7.029, \( p = 0.027 \)), lymph node metastasis (N stage; HR = 2.28, 95% CI = 1.189-3.784, \( p = 0.021 \)), distant metastasis (M stage; HR = 2.46, 95% CI = 1.363-4.453, \( p = 0.003 \)), and TNM stage (HR = 2.15, 95% CI = 1.165-3.971, \( p = 0.014 \)) were confidently associated with poor prognosis. However, with respect to genomic factors, only 8 signatures were validated to have significant effects on DFS (Figure 2C). Next, we used multivariate Cox analysis with the covariates, including age, local invasion, lymph node metastasis, distant metastasis and TNM stage, to evaluate the predictive values of the 8 signatures. As shown in Figure 4C, after each covariate was adjusted, Popovici’s signature (HR = 1.92, 95% CI = 1.298-2.841, \( p = 0.001 \)) and Fritzmann’s signature (HR = 2.16, 95% CI = 1.144-4.069, \( p < 0.0001 \)) remained powerful enough to indicate prognosis in the training set for DFS, which revealed that these two signatures are independent prognostic factors.

**Construction and comparison analysis of the composite nomogram**
Accurately predicting DFS in patients with CRC is important for counseling, follow-up planning, and selection of appropriate adjuvant therapy. Thus, to provide the clinician with a quantitative and user-friendly method to generate individualized predictions, we constructed a nomogram that integrated the genomic signatures identified above with clinicopathological features (Figure 3A). Calibration plots revealed excellent agreement between the nomogram-predicted probabilities and the actual observations of 1-, 3- and 5-year DFS (Figure 3B). The nomogram
demonstrated that distant metastasis made the largest contribution to prognosis, followed by the TNM stage and lymph node metastasis. Age and Popovici’s signature had moderate effects on survival rate. Each category within these variables was assigned a point on the top scale based on the coefficients from Cox regression. By summing all of the assigned points for the seven variables and drawing a vertical line between total points and the survival probability axis, we easily obtained the estimated probabilities of 1-, 3- and 5-year DFS. The risk score cutoff values (≤10.6, 10.6-21.2 and ≥21.2) were selected in terms of total points to stratify patients into roughly equal tertiles in the training set, which accurately divided patients into low-(reference), intermediate-(HR = 2.227, 95% CI = 1.507-3.291, p < 0.001), and high-risk (HR = 6.787, 95% CI = 4.786-9.624, p < 0.001) subgroups with significantly different DFS rates (Figure 3C, D).

Although the TNM stage in combination with other clinical factors is a well-recognized prediction system for CRC prognosis (28), its effectiveness urgently needs to be increased. To determine whether the genomic signatures added additional prognostic value to the current system, time-dependent receiver operating characteristic (ROC) analysis was applied to compare the performances between the nomogram, clinicopathological factors and genomic signatures (Figure 3E). Expectedly, the nomogram achieved the greatest area under the ROC curve (AUCs at 5-year DFS = 78.07, 71.40, 71.34, and 60.83, p < 0.05), suggesting that the integrated clinicopathological genomic nomogram had a prognostic performance superior to those of clinicopathological and genomic information by themselves.

**Validating the nomogram for stratifying CRC patient risks**

To validate the correlation between the nomogram score (total points) and DFS statuses of patients with CRC, Kaplan-Meier analysis and log-rank tests based on the same cutoff value were conducted on the validation set. As shown in Figure 4A and B, applying the clinicopathologic-genomic nomogram stratified patients into three distinct risk subgroups with significantly different DFS rates. In addition, the prognostic accuracy of the nomogram was remarkably better than those of both clinicopathological and genomic information by themselves (Figure 4C), which was
consistent with the results obtained from the training set. Disagreements on potentially appropriate candidates for adjuvant therapy prevail, especially for patients belonging to AJCC stages II and III. In the present study, the survival rates predicted by the nomogram were significantly distinct between the Kaplan-Meier curves (Figure 4D, E) in patients categorized by the abovementioned major clinicopathological features (AJCC stages II and III).

Moreover, to ensure the effectiveness of statistical power, patients with available chemotherapy information were integrated together regardless of the set to which they belonged (training set or validation set). Intriguingly, we noted that adjuvant chemotherapy did not enhance OS in all 869 patients (HR = 0.98, 95% CI = 0.769-1.258, \( p = 0.873 \); Figure 5A) or in patients who were defined as low risk by the nomogram (HR = 0.93, 95% CI = 0.515-1.695, \( p = 0.824 \); Figure 8B). However, patients in the nomogram defined as intermediate (HR = 0.98, 95% CI = 0.255-0.679, \( p < 0.001 \); Figure 5C) or high risk (HR = 0.67, 95% CI = 0.469-0.957, \( p = 0.028 \); Figure 5D) had favorable responses to adjuvant chemotherapy. Taken together, these results suggested that the clinicopathologic-genomic nomogram has the discriminatory power to identify CRC patients who are suitable candidates for adjuvant chemotherapy.

**Discussion**

Although surgery remains the mainstay of curative treatment, the prognosis of patients with CRC, especially for long-term outcome, depends more on pre- or postoperative individualized therapies (29), including various strategies of chemotherapy and chemoradiation. However, as a well-recognized heterogeneous cancer, its prognosis varies significantly, and optimal management presents challenges. Currently, adjuvant chemotherapy is regarded as standard treatment for patients with stage III colon cancer (4). However, patients with T1-2N1M0 tumors (stage IIIA) have significantly higher survival rates than those with stage IIB tumors (3). Moreover, for some patients, long-term survival continues to be jeopardized by the high risk of recurrence even after chemotherapy, and personalized therapy needs to
further optimized. With respect to stage II patients, disagreement regarding which patients are potentially appropriate candidates for adjuvant therapy prevails. The MOSAIC trial (30) has shown only a 20% benefit from adjuvant therapy at 3 years for patients with stage II disease and a 28% benefit for patients with recognized clinical risk factors. Thus, for approximately 75% of stage II patients who receive curative surgery, the administration of adjuvant chemotherapy is unnecessary and harmful. Unfortunately, low-risk stage II patients have not been accurately defined. In addition, patients with microsatellite instability (MSI CRC) have a better prognosis (31) and do not benefit from 5-fluorouracil (5-FU)-based adjuvant chemotherapy (32), similar to patients with CpG island methylation (33). All the abovementioned studies highlight that the rational application of therapeutic strategies requires accurate prognosis prediction and risk stratification for CRC patients.

Considering the inherent deficiency of TNM stages, substantial effort has been placed on their optimization (2), and they have been continuously developed over the past decade. Substantial changes in CRC classifications, based mainly on the analysis of surveillance, epidemiology, and end results (SEER) data, were made in the seventh edition (34). However, a comparative study demonstrated that the seventh TNM edition did not provide a greater accuracy for predicting CRC patient prognoses but rather resulted in a more complex classification for daily clinical use (35). The eighth edition of the CRC staging system, containing major advances, including the introduction of molecular markers, such as MSI, KRAS and BRAF mutations, to strengthen its discriminatory power, will be implemented worldwide on January 1, 2018 (2). Nonetheless, the forthcoming edition may not provide substantial improvement because the prognostic value of clinicopathological factors has peaked. Moreover, the current models still lack genomic information, which may directly reflect the heterogeneity of CRC and present substantial potential prognostic value.

Genomic signatures predicting the prognosis of CRC have been continually developed in the past 10 years. In the present study, we retrieved 81 articles related to the development of signatures for predicting CRC prognosis, including OS, DFS, disease-specific survival (DSS) and various types of recurrence. For instance, based...
on the association between the aberrantly expressed gene and the duration of individual DFS, Zhang et al. (36) constructed a six RNA-based classifier, which was further validated to have reliable predictive power for detecting recurrence in patients with stage II colon cancer. Similarly, Anna et al. (23) developed a genomic profile specific for adult intestinal stem cells (ISC), which was highly enriched in genes that positively predicted the risk of recurrence. Patients bearing primary CRCs with high average expression levels of ISC genes had a relative risk of relapse that was 10-fold higher than that for patients with low levels (23). Moreover, recent evidence (37) suggest that gene signatures that closely reflect specific biological processes or oncogenic pathway statuses also have potential prognostic values for stratifying CRC patients. Even though almost all of these signatures were proven to have prognostic value in their respective publications, none are routinely used in clinical practice.

Potential explanations for the unsatisfactory results must be considered. Thus, in the current study, we exhaustively reviewed and analyzed the published multi-gene signatures in CRC. To eliminate the potential confounding effects, only data derived from the Affymetrix human genome platform having clearly described genes were included. Of the 81 published signatures, 27 signatures from 26 publications met all of the inclusion criteria and were retained (Table 1 and Supplementary Table S3). The 27 signatures contained 1274 total genes, among which 1041 were unique, and the signature sizes ranged from 4 to 264 (Figure 1A). The top overlapping genes in the above signatures were FN1 (ten times), CYP1B1 and POSTN (six times), 7 genes (five times), 7 genes (four times), 29 genes (three times), and 107 genes (twice). As is clearly shown in Figure 2A, none of these genes appeared in all signatures. These results were expected since they were previously reported (15) and in agreement with the systematic analyses of lung cancer and breast cancer prognostic signatures (38). Obviously, the slight overlap of these signatures in gene identity was perplexing but sufficiently explained the low reproducibility. Moreover, genetic heterogeneity was recently shown to exist not only between but also within tumors, meaning that biological statuses, including oncogenic and carcinostatic states, might regularly vary in the same patient (39). Our NTP analysis revealed that some patients concurrently
harbored more than one signature, which further confirmed the abovementioned results and indicated that their tumors were genomic instability at the individual level. Taken together, the different signatures might reflect diverse biological processes and be active in different or identical individual tumors, which partially accounts for the wide nonoverlapping among signatures constructed in different CRC studies.

In terms of global performance, in the training set, 8 of 27 signatures showed a significant association with prognosis and could reasonably predict the outcome. However, after multivariate adjustment, only 2 signatures remained powerful enough to indicate prognosis for DFS, which directly demonstrated the low reproducibility of CRC genomic signatures. More importantly, Popovici’s signature, Fritzmann’s signature and TNM stage had remarkable prognostic abilities and were independent of each other, indicating that these signatures may be used to refine the current prognostic model and facilitate further stratification of patients with CRC at the same TNM stage.

Our primary goal was to construct a nomogram that could integrate genomic signatures with clinicopathological variables in a large cohort of CRC patients to add additional prognostic value to the current staging system. Our multivariate analysis in the training set revealed that age, local invasion, lymph node metastasis, distant metastasis, TNM stage, Popovici’s signature and Fritzmann’s signature were independent prognostic factors (Supplementary Table S6), which was highly consistent with previous studies on CRC risk factors (40,41). Accordingly, we cautiously built a clinicopathologic-genomic nomogram using the abovementioned significant factors. The AUC of the nomogram showed a superior prediction accuracy compared to that of the combined clinical factors in the training and validation sets (Figure 3E, 4C). Additionally, the calibration plots showed optimal agreement between the expected and observed survival probabilities (Figure 3B), ensuring the reliability of our nomogram. Moreover, the nomogram successfully stratified CRC patients into three risk groups with remarkably different DFS values based on total points, which was further confirmed using the same cutoff value in the validation set.

Intriguingly, the TNM staging system is the most important tool for guiding
treatment in clinical practice. However, disagreement prevails on potentially appropriate candidates for adjuvant therapy, especially for stage II and III CRC patients. Therefore, the identification of a high-risk subgroup among these patients is of urgent clinical needed. Our findings suggest that the clinicopathologic-genomic nomogram could be a promising tool to stratify stage II and III patients into distinct risk subgroup and guide individualized therapy. Moreover, in the current study, we clearly revealed that chemotherapy provides a survival benefit to patients classified as intermediate- and high-risk by the nomogram (Figure 5).

Despite the promising results, there were several shortcomings in this work. First, the NTP method uses only a list of signature genes to make class predictions in each patient’s expression data, allowing the method to be less sensitive to differences in experimental and analytical conditions and applicable to every patient without optimization of the analysis parameters. In the current study, we applied the NTP method for all prediction models, which generally consisted of two major components: a gene signature and a prediction algorithm. Obviously, the intrinsic limitation of the NTP method is that it inevitably causes the algorithm lose its prognostic value. Second, the forthcoming 8th edition of the CRC TNM staging system may better predict prognosis, but data used in the present study was from the current 7th edition or the 6th edition. Third, the prognostic signatures of CRC included in our study were not exhaustive, and some promising signatures were excluded because of insufficient information. In addition, some clinicopathological factors, even those well recognized to have prognostic value, were not included in the composite nomogram because of the low availability of information.

In summary, based on analyzing a large-scale CRC cohort and published multi-gene prognostic signatures, we confirmed herein for the first time that genomic information in combination with clinical data can improve patient prognostic evaluations and should be considered as an independent and complementary approach to the current clinicopathological prognostic model. Furthermore, considering the wide promise of our nomogram, which integrates genomic signatures with clinicopathological features to improve the prognosis prediction of colorectal cancer,
further systematic research is needed.

Acknowledgments

Conception and design: Z. Fu and Y. Xiong. Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): Y. Xiong, W. You, and M. Hou. Writing, review, and/or revision of the manuscript: L. Peng, H. Zhou. Administrative, technical, and/or material support (i.e., reporting or organizing data): Y. Xiong, W. You, M. Hou, L. Peng, and H. Zhou. Study supervision: Z. Fu.
References


21 / 26


Table 1 Genomic Signatures Included in the Study

<table>
<thead>
<tr>
<th>First author</th>
<th>Prognosis</th>
<th>Number of genes</th>
<th>Patients with the signature classified as (%)*</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aziz</td>
<td>OS</td>
<td>19</td>
<td>0.78 (78.9%)</td>
<td>201 (23.5%)</td>
</tr>
<tr>
<td>Chang</td>
<td>DFS</td>
<td>95</td>
<td>0.5 (100.0%)</td>
<td>341 (39.9%)</td>
</tr>
<tr>
<td>Chen</td>
<td>OS</td>
<td>7</td>
<td>0.6 (85.7%)</td>
<td>293 (34.3%)</td>
</tr>
<tr>
<td>Shi</td>
<td>DFS</td>
<td>9</td>
<td>0.1 (100.0%)</td>
<td>396 (46.3%)</td>
</tr>
<tr>
<td>Xu</td>
<td>OS</td>
<td>15</td>
<td>0.1 (86.7%)</td>
<td>299 (35%)</td>
</tr>
<tr>
<td>Minh</td>
<td>DFS</td>
<td>77</td>
<td>0.74 (96.5%)</td>
<td>358 (41.9%)</td>
</tr>
<tr>
<td>Chen</td>
<td>DFS</td>
<td>4</td>
<td>0.1 (100.0%)</td>
<td>0 (0.0%)</td>
</tr>
<tr>
<td>Agesen</td>
<td>DFS</td>
<td>13</td>
<td>0.1 (100.0%)</td>
<td>74 (8.7%)</td>
</tr>
<tr>
<td>Popovici</td>
<td>DFS</td>
<td>64</td>
<td>0.56 (87.5%)</td>
<td>333 (38.9%)</td>
</tr>
<tr>
<td>Seven</td>
<td>DFS</td>
<td>7</td>
<td>0.1 (100.0%)</td>
<td>140 (16.4%)</td>
</tr>
<tr>
<td>Merlos</td>
<td>DFS</td>
<td>264</td>
<td>0.247 (93.6%)</td>
<td>156 (18.2%)</td>
</tr>
<tr>
<td>Oh</td>
<td>DFS, OS</td>
<td>87</td>
<td>0.81 (93.1%)</td>
<td>278 (32.5%)</td>
</tr>
<tr>
<td>Salazar</td>
<td>DFS</td>
<td>18</td>
<td>0.168 (88.9%)</td>
<td>192 (22.5%)</td>
</tr>
<tr>
<td>Kalady</td>
<td>DFS, OS</td>
<td>36</td>
<td>0.246 (66.7%)</td>
<td>245 (28.7%)</td>
</tr>
<tr>
<td>Peng</td>
<td>DFS</td>
<td>8</td>
<td>0.07 (87.5%)</td>
<td>120 (14%)</td>
</tr>
<tr>
<td>Toshiaki</td>
<td>DFS</td>
<td>10</td>
<td>0.09 (90.0%)</td>
<td>323 (37.8%)</td>
</tr>
<tr>
<td>Fritzmann</td>
<td>DFS</td>
<td>113</td>
<td>0.100 (88.5%)</td>
<td>568 (66.4%)</td>
</tr>
<tr>
<td>Hao</td>
<td>DFS</td>
<td>4</td>
<td>0.1 (100.0%)</td>
<td>188 (22%)</td>
</tr>
<tr>
<td>Jorissen</td>
<td>DFS</td>
<td>122</td>
<td>0.102 (83.6%)</td>
<td>370 (43.3%)</td>
</tr>
<tr>
<td>Smith</td>
<td>DFS, OS</td>
<td>34</td>
<td>0.298 (85.3%)</td>
<td>337 (39.4%)</td>
</tr>
<tr>
<td>Watanabe</td>
<td>DFS</td>
<td>31</td>
<td>0.227 (81.0%)</td>
<td>156 (18.2%)</td>
</tr>
<tr>
<td>Garman</td>
<td>DFS, OS</td>
<td>46</td>
<td>0.378 (80.4%)</td>
<td>193 (22.6%)</td>
</tr>
<tr>
<td>Jiang</td>
<td>DFS</td>
<td>7</td>
<td>0.06 (85.7%)</td>
<td>216 (25.3%)</td>
</tr>
<tr>
<td>Lin.a</td>
<td>DFS</td>
<td>16</td>
<td>0.148 (87.5%)</td>
<td>141 (16.5%)</td>
</tr>
<tr>
<td>Lin.b</td>
<td>DFS</td>
<td>22</td>
<td>0.18 (81.8%)</td>
<td>319 (37.3%)</td>
</tr>
<tr>
<td>Wang</td>
<td>DFS</td>
<td>19</td>
<td>0.19 (100.0%)</td>
<td>226 (26.4%)</td>
</tr>
</tbody>
</table>

*Patients were classified as good, poor or uncertain by the respective published genomic signatures based on prediction results (false discover rate [FDR] <0.05) of the nearest template prediction (NTP).
Figure legends

Figure 1. Bioinformatic analysis of gene lists included in this study. A. Landscape distribution characteristics of the gene lists as reported in 27 prognostic signatures. Genes with good and poor prognostic values in each signature are shown as short green and red vertical bars, respectively. The number of gene occurrences is graphically depicted in the upper panel. Each row represents one signature annotated by the first author of the corresponding article. All the unique genes were sorted alphabetically. B-E. GO and KEGG pathway analyses of genes in the 27 signatures. The vertical axis represents Go or KEGG pathway annotations. The horizontal axis represents the number of genes assigned to the corresponding annotation. (B) Biological process; (C) cellular components; (D) molecular function; and (E) KEGG pathway. F. Protein-protein interaction (PPI) networks. Nodes represent the proteins, and edges represent the physical interaction. The black and gray nodes represent genes with good and poor prognostic values, respectively. CRC recurrence-related pathways that were significantly enriched by KEGG analysis form some closely connected regions and are highlighted in the colored box.

Figure 2. Global prognostic performance of the published CRC signatures. A. Concordance of signature-based prediction results in the training set. Each column represents the prediction of each individual sample. The orange, blue and white bars indicate good, poor and uncertainty prognoses of the corresponding signature, respectively. B. Heatmap of Cramer’s V coefficients showing correlation between these published signatures. C. Analysis of prognostic performance using Cox regression; left: univariate analysis; right: clinical factors adjusted for multivariate analysis.

Figure 3. Development of the composite clinicopathologic-genomic nomogram. A. Composite nomogram to predict disease-free survival for patients with CRC. Each category within these variables was assigned to a point on the top scale based on the coefficients from Cox regression. By summing all of the assigned points for the seven
variables and drawing a vertical line between the total points and survival probability axis, we easily obtained the estimated probabilities of 1-, 3- and 5-year disease-free survival (DFS). **B.** Calibration curves for predicting patient survival at 1, 3 and 5 years. The actual DFS is plotted on the y-axis, and the nomogram-predicted DFS is plotted on the x-axis; a plot along the 45-degree line would indicate a perfect calibration model in which the predicted probabilities are identical to the actual outcomes. **C.** The total point distribution of each CRC patient and their disease-free survival time and status. The blue dotted line represents the cutoff value dividing patients into low-, intermediate- and high-risk groups. **D.** Kaplan-Meier survival curves of DFS for different risk groups using the nomogram in the training set. **E.** Time-dependent ROC curves comparing the prognostic accuracies of 5-year disease-free survival among the prognostic signatures combined with clinicopathological features and the nomogram in the training set. The AUC 95% CIs were calculated from 1000 bootstraps of the survival data. ROC, receiver operator characteristic. AUC, area under the curve.

**Figure 4. Verifying the prognostic value of the nomogram in the validation set. A.** The total point distribution of each CRC patient and their disease-free survival time and status. The blue dotted line represents the cutoff value dividing patients into low-, intermediate- and high-risk groups. **B.** Kaplan-Meier survival curves of DFS for different risk groups using the nomogram in the validation set. **C.** Time-dependent ROC curves comparing the prognostic accuracies of 5-year disease-free survival among the prognostic signatures in combination with clinicopathological features and the nomogram in the validation set. **D, E.** Kaplan-Meier survival curves for patients in different risk subgroups, which were stratified by TNM stages II (D) and III (E).

**Figure 5. Effect of chemotherapy on overall survival in patients stratified by the nomogram. A.** Kaplan-Meier survival curves for all patients who received chemotherapy. **B-D.** Effects of adjuvant chemotherapy on overall survival in the different subgroups. Low-risk group (B); intermediate risk group (C); high-risk group...
(D).
Figure 2

A

B

C

Univariate analyses

Multivariate analyses

<table>
<thead>
<tr>
<th>Study</th>
<th>Hazard Ratio</th>
<th>HR</th>
<th>95%–CI</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agesen</td>
<td>1.08</td>
<td>[0.99; 1.18]</td>
<td>0.0818</td>
<td></td>
</tr>
<tr>
<td>Aziz</td>
<td>1.11</td>
<td>[0.91; 1.35]</td>
<td>0.3251</td>
<td></td>
</tr>
<tr>
<td>Chang</td>
<td>1.08</td>
<td>[0.98; 1.20]</td>
<td>0.1189</td>
<td></td>
</tr>
<tr>
<td>Chen</td>
<td>1.00</td>
<td>[0.93; 1.08]</td>
<td>0.9424</td>
<td></td>
</tr>
<tr>
<td>Watanabe</td>
<td>1.11</td>
<td>[0.91; 1.35]</td>
<td>0.3251</td>
<td></td>
</tr>
<tr>
<td>Garman</td>
<td>1.01</td>
<td>[1.00; 1.02]</td>
<td>0.0981</td>
<td></td>
</tr>
<tr>
<td>Hao</td>
<td>1.04</td>
<td>[0.79; 1.36]</td>
<td>0.7804</td>
<td></td>
</tr>
<tr>
<td>Jiang</td>
<td>1.08</td>
<td>[0.99; 1.18]</td>
<td>0.1011</td>
<td></td>
</tr>
<tr>
<td>Jorissen</td>
<td>1.42</td>
<td>[1.12; 1.80]</td>
<td>0.0033</td>
<td></td>
</tr>
<tr>
<td>Kalady</td>
<td>1.05</td>
<td>[0.95; 1.15]</td>
<td>0.3372</td>
<td></td>
</tr>
<tr>
<td>Lin.a</td>
<td>0.91</td>
<td>[0.62; 1.00]</td>
<td>0.0413</td>
<td></td>
</tr>
<tr>
<td>Lin.b</td>
<td>1.07</td>
<td>[0.96; 1.19]</td>
<td>0.2105</td>
<td></td>
</tr>
<tr>
<td>Merlos</td>
<td>0.96</td>
<td>[0.87; 1.04]</td>
<td>0.3133</td>
<td></td>
</tr>
<tr>
<td>Minh</td>
<td>2.60</td>
<td>[1.47; 4.61]</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Oh</td>
<td>1.27</td>
<td>[1.06; 1.53]</td>
<td>0.0089</td>
<td></td>
</tr>
<tr>
<td>Peng</td>
<td>0.58</td>
<td>[0.34; 0.97]</td>
<td>0.0399</td>
<td></td>
</tr>
<tr>
<td>Salazar</td>
<td>1.16</td>
<td>[1.04; 1.30]</td>
<td>0.0071</td>
<td></td>
</tr>
<tr>
<td>Popovici</td>
<td>2.46</td>
<td>[1.18; 5.13]</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Seven</td>
<td>1.07</td>
<td>[0.98; 1.18]</td>
<td>0.1252</td>
<td></td>
</tr>
<tr>
<td>Shen</td>
<td>1.10</td>
<td>[0.98; 1.24]</td>
<td>0.1032</td>
<td></td>
</tr>
<tr>
<td>Shi</td>
<td>1.41</td>
<td>[0.74; 2.67]</td>
<td>0.2949</td>
<td></td>
</tr>
<tr>
<td>Smith</td>
<td>1.07</td>
<td>[0.96; 1.19]</td>
<td>0.2105</td>
<td></td>
</tr>
<tr>
<td>Toshiaki</td>
<td>1.10</td>
<td>[0.99; 1.22]</td>
<td>0.0861</td>
<td></td>
</tr>
<tr>
<td>Wang</td>
<td>1.00</td>
<td>[0.90; 1.11]</td>
<td>0.9948</td>
<td></td>
</tr>
<tr>
<td>Fritzmann</td>
<td>2.33</td>
<td>[1.22; 4.45]</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Xu</td>
<td>1.16</td>
<td>[0.94; 1.43]</td>
<td>0.1759</td>
<td></td>
</tr>
</tbody>
</table>

Fixed effect model: 1.02 [1.01; 1.03], p = 0.0037
Random effects model: 1.06 [1.03; 1.10], p < 0.001

Heterogeneity: I² = 64%,
Figure 4

A. Disease-free survival (blue) and overall survival (red) for patients with low, intermediate, and high-risk using the nomogram. The Kaplan-Meier curves show a clear separation between the risk groups.

B. Validation set, total. Risk stratification using the nomogram with disease-free survival. The p-value is less than 0.0001.

C. Validation set, total. Time-dependent ROC curves for disease-free survival. The ROC curves indicate good discrimination between the risk groups.

D. Validation set, TNM stage II. Risk stratification using the nomogram with disease-free survival. The p-value is 0.00035.

E. Validation set, TNM stage III. Risk stratification using the nomogram with disease-free survival. The p-value is 0.00067.

Legend:
- Low risk (≤10.6 points)
- Intermediate risk (10.6-21.2 points)
- High risk (≥21.2 points)

Number at risk by time:
- Low risk
  - 370 at 0 months,
  - 250 at 50 months,
  - 73 at 100 months,
  - 17 at 150 months,
  - 3 at 200 months,
- Intermediate risk
  - 208 at 0 months,
  - 125 at 50 months,
  - 31 at 100 months,
  - 1 at 150 months,
  - 0 at 200 months,
- High risk
  - 275 at 0 months,
  - 87 at 50 months,
  - 20 at 100 months,
  - 3 at 150 months,
  - 0 at 200 months.

<table>
<thead>
<tr>
<th>Risk Group</th>
<th>AUC</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nomogram</td>
<td>0.773</td>
<td>0.680-0.817</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Fritsman’s signature</td>
<td>0.713</td>
<td>0.622-0.777</td>
<td>0.029</td>
</tr>
<tr>
<td>Popovski’s signature</td>
<td>0.601</td>
<td>0.506-0.691</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Clinical pathological factors</td>
<td>0.736</td>
<td>0.640-0.830</td>
<td>0.047</td>
</tr>
</tbody>
</table>
Figure 5

A  All patients

Survival probability

Number at risk by time

<table>
<thead>
<tr>
<th>Time (0-200)</th>
<th>478</th>
<th>270</th>
<th>57</th>
<th>16</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>No chemotherapy</td>
<td>Chemotherapy</td>
<td>No chemotherapy</td>
<td>Chemotherapy</td>
<td>No chemotherapy</td>
<td>Chemotherapy</td>
</tr>
</tbody>
</table>

B  Low risk group

Survival probability

Number at risk by time

<table>
<thead>
<tr>
<th>Time (0-200)</th>
<th>225</th>
<th>153</th>
<th>33</th>
<th>7</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>No chemotherapy</td>
<td>Chemotherapy</td>
<td>No chemotherapy</td>
<td>Chemotherapy</td>
<td>No chemotherapy</td>
<td>Chemotherapy</td>
</tr>
</tbody>
</table>

C  Intermediate risk group

Survival probability

Number at risk by time

<table>
<thead>
<tr>
<th>Time (0-200)</th>
<th>178</th>
<th>94</th>
<th>17</th>
<th>9</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>No chemotherapy</td>
<td>Chemotherapy</td>
<td>No chemotherapy</td>
<td>Chemotherapy</td>
<td>No chemotherapy</td>
<td>Chemotherapy</td>
</tr>
</tbody>
</table>

D  High risk group

Survival probability

Number at risk by time

<table>
<thead>
<tr>
<th>Time (0-100)</th>
<th>75</th>
<th>23</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>No chemotherapy</td>
<td>Chemotherapy</td>
<td>No chemotherapy</td>
<td>Chemotherapy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time (0-100)</th>
<th>180</th>
<th>64</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>No chemotherapy</td>
<td>Chemotherapy</td>
<td>No chemotherapy</td>
<td>Chemotherapy</td>
</tr>
</tbody>
</table>
Molecular Cancer Research

Nomogram Integrating Genomics with Clinicopathological Features Improves Prognosis Prediction for Colorectal Cancer

Yongfu Xiong, Wenxian You, Min Hou, et al.

Mol Cancer Res  Published OnlineFirst May 21, 2018.

Updated version  Access the most recent version of this article at: doi:10.1158/1541-7786.MCR-18-0063

Supplementary Material  Access the most recent supplemental material at: http://mcr.aacrjournals.org/content/suppl/2018/05/19/1541-7786.MCR-18-0063.DC1

Author Manuscript  Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, use this link http://mcr.aacrjournals.org/content/early/2018/05/19/1541-7786.MCR-18-0063. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.