

Gene pair based prognostic signature for colorectal colon cancer

Peng Shu, MD^a, Jianping Wu, MD^b, Yao Tong, BS^c, Chunxia Xu, MD^a, Xingguo Zhang, MD^{a,*}

Abstract

Background: The identification of high-risk colorectal cancer (CRC) patient is key to individualized treatment after surgery and reliable prognostic biomarkers are needed identifying high-risk CRC patients.

Methods: We developed a gene pair based prognostic signature that could can the prognosis risk in patients with CRC. This study retrospectively analyzed 4 public CRC datasets, and 1123 patients with CRC were divided into a training cohort (n=300) and 3 independent validation cohorts (n=507, 226, and 90 patients).

Results: A signature of 9 prognosis-related gene pairs (PRGPs) consisting of 17 unique genes was constructed. Then, a PRGP index (PRGPI) was constructed and divided patients into high- and low-risk groups according to the signature score. Patients in the high-risk group showed a poorer relapse-free survival than the low-risk group in both the training cohort [hazard ratio (HR) range, 4.6, 95% confidence interval (95% CI), 2.55–8.32; $P < .0001$] and meta-validation set (hazard ratio range, 4.09, 95% CI, 1.99–8.39; $P < .0001$). The PRGPI signature achieved a higher accuracy [mean concordance index (C-index): 0.6~0.74] than a commercialized molecular signature (mean C-index, 0.48~0.56) for estimation of relapse-free survival in comparable validation sets.

Conclusion: The gene pair based prognostic signature is a promising biomarker for estimating relapse-free survival of CRC.

Abbreviations: CRC = colorectal cancer, EMT = epithelial-mesenchymal transition, GSEA = gene set enrichment analysis, PCPI = prognosis-clinical prognostic index, PRGPs = prognostic-related gene pairs, PRGs = prognosis-related genes, RMS = restricted mean survival.

Keywords: colorectal cancer, prognosis, prognosis-related gene pairs, relapse-free survival

1. Introduction

Colorectal cancer (CRC) is the third most diagnosed cancer in the world.^[1] Nearly 1.4 million people are diagnosed as new cases of CRC every year.^[2] Although new tests and treatments have been achieved for the management of CRC, the 5-year survival rate is only approximately 55%.^[3] Surgery remains the first priority means of curative treatment. However, a proportion of patients will suffer local recurrences and remote metastases after surgery. Meanwhile, patients with equal clinical or pathological conditions show unpredictable clinical outcomes, even when treated similarly.^[4] The patients' genetic heterogeneity contributes most to the inherent clinical diversity.^[5]

Biomarkers that can estimate the genetic diversity of CRC and accurately evaluate patient survival can guide new and more effective clinical management of CRC.^[6] For example, considering the high risk (50–60%) of recurrence, stage III patients are routinely treated with adjuvant therapy despite potentially curative surgery.^[7] However, stage II patients are not recommended to undergo adjuvant therapy under the current guidelines.^[8] Studies have reported a 20% to 30% relapse rate for stage II CRC patients, and the clinical popular staging system remains ineffective in distinguishing this subgroup,^[9] for whom the toxic side effects may outweigh the benefit of adjuvant therapy. Therefore, stratification of the subgroup of CRC patients with a high risk of recurrence and death who have the greatest requirement for treatment adjustment is needed beyond the clinical pathological factors.

Regarding prognosis biomarkers, researchers have investigated the possibility of stratifying patients with CRC based on gene expression signatures and they have built multigene-expression signatures that can be used to stratify high-risk subgroups.^[10–13] Although these survival-related signatures hold promise, they do not perform well when validated in independent cohorts due to the diversity of data. Gene expression levels sequenced by traditional approaches require suitable normalization, which is difficult considering the biological heterogeneity and technical biases across sequencing platforms.^[14] Instead, researchers have proposed new methods to eliminate the limitations in data processing, such as normalization and scaling based on relative ranking of gene expression levels, and have produced robust outcomes in various studies.^[15–17]

In this study, we constructed prognosis-related gene pairs (PRGSs) to develop and validate an individualized prognostic signature for CRC.

Editor: Jianxun Ding.

Funding/support: This work was supported by the Ningbo Social Development Major Project (2017C510009).

The authors report no conflicts of interest.

Supplemental Digital Content is available for this article.

^a Beilun People's Hospital, Ningbo, ^b Department of Laboratory Medicine, The First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, China, ^c College of Agriculture and Life Sciences, Cornell University, Ithaca, NY.

* Correspondence: Xingguo Zhang, Beilun People's Hospital, No.1288 of East Lushan Road, Beilun District, Ningbo 315800, China (e-mail: 995936007@qq.com).

Copyright © 2018 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

Medicine (2018) 97:42(e12788)

Received: 6 June 2018 / Accepted: 18 September 2018

<http://dx.doi.org/10.1097/MD.00000000000012788>

2. Materials and methods

2.1. Ethical approval

The researchers were granted approval to conduct the research by their Departmental Research Ethics Committee at the Beilun People's Hospital, Ningbo, China. The study protocol was approved by the institutional review board of Beilun People's Hospital. All the procedures were performed in accordance with the Declaration of Helsinki and relevant policies in China.

2.2. Data collection

We retrospectively analyzed gene expression profiles from public CRC cohorts, including microarray datasets and RNA-seq data for The Cancer Genome Atlas (TCGA) of CRC. Normalized CRC gene expression data sets were downloaded from the Gene Expression Omnibus (GEO) database (<http://www.ncbi.nlm.nih.gov/geo/>) by GEOquery (version: v2.46.15) package.^[18] Four datasets, that is, CIT cohort (GSE39582,^[19] n=566) used as a training set, TCGA^[20] (n=624), Jorissen cohort (GSE14333,^[21] n=226), and De Sousa cohort (GSE33113,^[22] n=90) were included in this study. Only CRC patients with relapse-free survival (RFS) were included. Patients in the CIT cohort without adjuvant chemotherapy (GSE39582,^[19] n=300) were used as the training set, while TCGA^[20] (n=507), Jorissen cohort (GSE14333,^[21] n=226), and De Sousa cohort (GSE33113,^[22] n=90) were used as validation cohorts. Overall, we included 1123 CRC patients in our study (Supplemental Table 1, <http://links.lww.com/MD/C546>).

2.3. Data preprocessing

All Affymetrix microarrays were normalized with the MAS5.0 method using the *affy*^[23] package (version: v1.56.0). The probe IDs were annotated with gene symbols using the Affymetrix annotation file (<http://www.affymetrix.com>). When multiple probe IDs matched to an identical gene symbol, the probe ID with the largest mean of expression values among all probe IDs was selected to represent the gene.^[24] Normalized TCGA RNA-Seq data and clinical information were downloaded from FireBrowse (<http://firebrowse.org>).

2.4. Construction of CRC-specific PRGPs for prognosis prediction

CIT (GSE39582, n=300) was used as the discovery data-set and to train the model. Of 21,049 genes, 4746 were used as initial candidates and were selected on the basis of mean absolute deviation (MAD) > 0.5 and the average expression level must be beyond the median values of all genes in this data set. Then, we identified 296 prognosis-related genes (PRGs) using the proportional hazards regression model (FDR $P < .01$). Among them, 267 genes were measured by all platforms in this study. From these 267 PRGs, we constructed 35,511 prognostic-related gene pairs (PRGPs) and then filtered out PRGPs with relatively minor variation (MAD=0) and 62 PRGPs were kept. Then, we constructed a PRGP index (PRGPI) using Lasso Cox proportional hazards regression on the training set (CIT, n=300) and 9 gene pairs were used to define the final model.

2.5. Validation of the PRGPI

The PRGPI prognostic value was evaluated in all stages of CRC patients and early stages group (stages I and II) in the training,

meta-validation, and independent validation cohorts using the log-rank test. We then combined the PRGPI with other clinical factors in multivariate analyses. Age and stage were treated as continuous variables. Stage I was transformed into 1; Stage II was transformed into 2; Stage III was transformed into 3; and Stage IV was transformed into 4. The prognostic accuracy of the PRGPI was estimated using the concordance index (C-index), which ranges from 0 to 1.0. The prognostic efficiency of the PRGPI was compared with the existing multigene signature Oncotype Dx Colon Cancer^[25] with C-index.

2.6. Gene set enrichment analysis

We performed gene set enrichment analysis^[26] using GSEA software (<http://www.broadinstitute.org/gsea>) with 1000 permutations. For the CIT data-set, we took as phenotype the log₂ fold change between the gene expression profile of high- and low-risk groups. Gene sets used in this study were downloaded from MSigDB (C2 and C5 databases, version 6) (<http://www.broadinstitute.org/gsea/msigdb/collections.jsp>). FDR-adjusted $P < .05$ or nominal (NOM) $P < .05$ was used to select statistically significant gene sets.

2.7. Construction and validation of a composite prognosis-clinical prognostic index

We integrated age, stage, sex, and the PRGPI risk score into an entirety prognosis-clinical prognostic index (PCPI) using Cox proportional hazards regression in the training cohort. The prognostic efficiency of the PCPI was compared with the C-index of the PRGPI and depicted by the restricted mean survival (RMS) curve.^[27] A high RMS time ratio represents a large prognostic difference.

2.8. Statistical analysis

All statistical analyses were performed using R (version: 3.4.2). Univariate and multivariate Cox proportional hazard analyses of the PRGPI and other clinical factors with RFS were evaluated using the log-rank test. The C-index was calculated by *survcomp* package^[28] (version: 1.28.4). The RMS curve and time ratio were calculated by *survival* package (version: 2.41.3). Differences were considered as statistically significant when $P < .05$.

3. Results

3.1. Construction and definition of the PRGPs signature

A total of 1123 CRC patients were included in this study (Supplemental Table 1, <http://links.lww.com/MD/C546>). The CIT dataset (n=300) was selected as the training cohort. Within this dataset, 296 PRGs were identified and 267 overlapped genes were measured by all platforms in this study. On the basis of the 267 genes, we constructed 35,511 PRGPs. We removed PRGPs with relative minor variation (MAD=0) and 62 PRGPs were kept. We then constructed a PRGP index (PRGPI) using L1-penalized Cox proportional hazards regression on the training data set and 9 gene pairs (Supplemental Table 2, <http://links.lww.com/MD/C546>) were used to define the final model. The PRGP signature consisted of 17 unique prognostic genes (*RPS6KA5*, *ITGA5*, *FSTL3*, *S100A2*, *PEX6*, *KRT17*, *GZMA*, *MBTD1*, *ZNF468*, *KLK10*, *MINPP1*, *BRIP1*, *ZNF184*, *RPS6KC1*, *GLMN*, *WWC3*, and *C5orf30*) and the corresponding coefficient are shown in (Supplemental Table 3, <http://links.lww.com/MD/C546>).

lww.com/MD/C546). The median value of the risk score was used as the cutoff to further stratify patients into high and low-risk groups.

3.2. Validation of the PRGPI as a prognostic factor

The PRGPI significantly classified CRC patients into high- and low-risk groups in terms of RFS in the training cohort (Supplemental Table 3, <http://links.lww.com/MD/C546> and Fig. 1A). A multivariate analysis showed that PRGPI high tumors were associated with a higher risk of recurrence than PRGPI low ones [hazard ratio (HR) range, 6.86, 95% confidence interval (95% CI), 3.89–12.12; $P < .0001$]; the HR associated with the PRGPI status was higher than clinical

and pathologic factors such as age, sex, tumor stage, and tumor location (Supplemental Table 4 and 5, <http://links.lww.com/MD/C546>). Furthermore, the PRGPI significantly classified early stage CRC patients (I and II) into high- and low-risk groups (HR range, 4.09, 95% CI, 1.99–8.39; $P < .0001$) (Fig. 1B). Similarly, a higher PRGPI was significantly associated with a worse prognosis at all stages (HR range, 1.76, 95% CI, 1.31–2.36; $P = 1.46E-4$) and early-stage (HR range, 2.03, 95% CI, 1.26–3.29; $P = 3.17E-3$) CRC patients in meta-validation (Fig. 1C, D). When testing in independent validation cohorts, the PRGPI remained highly prognostic for all and early-stage CRC patients (Supplemental Fig. 1, <http://links.lww.com/MD/C546>). In summary, the PRGPI appears to estimate RFS for CRC.

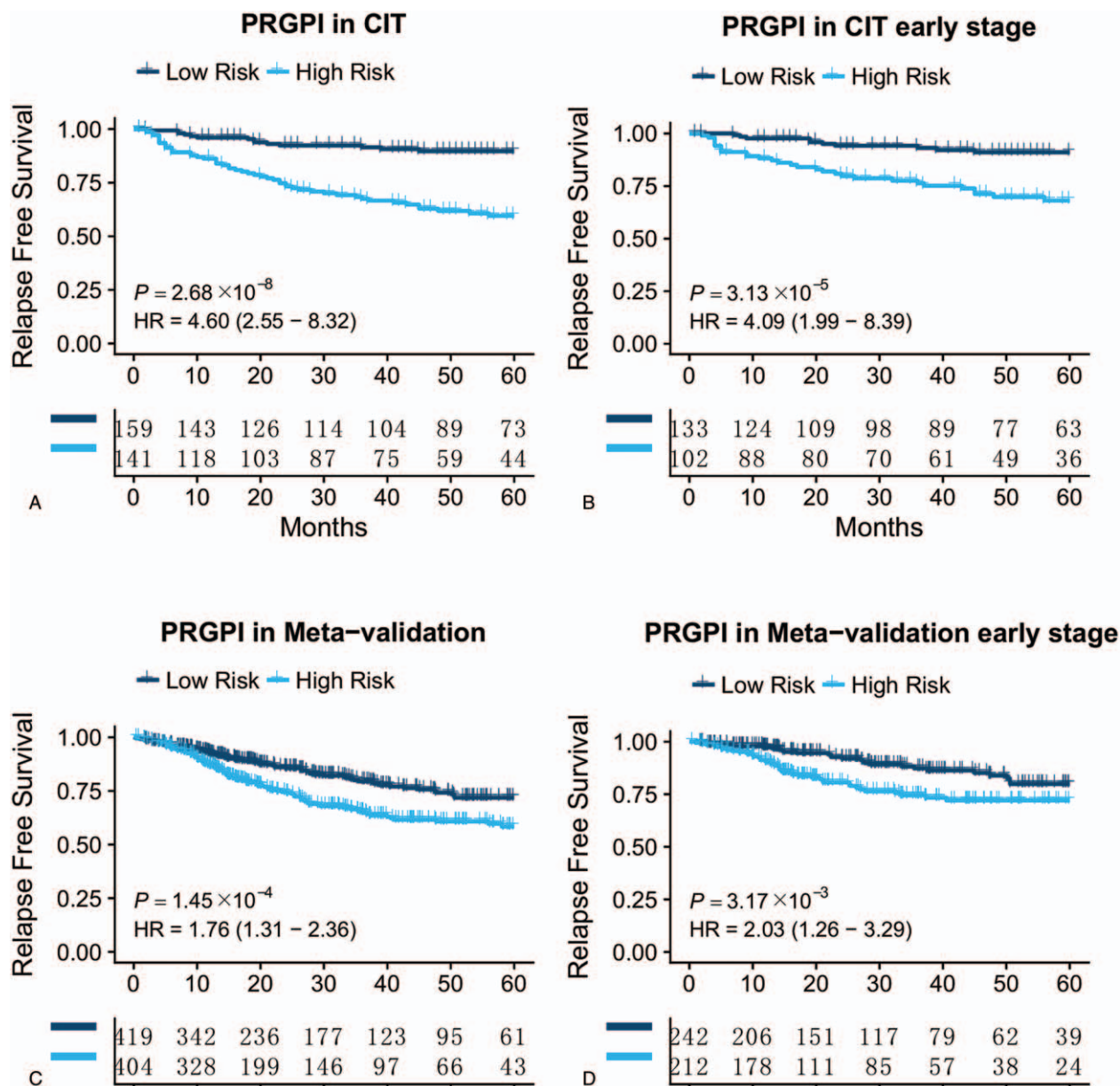


Figure 1. Kaplan-Meier curves of relapse-free survival among colorectal cancer (CRC) patients. Patients are stratified by the prognostic-related gene pair index (PRGPI) (low and high risk). (A) and (C) Relapse-free survival for patients in the training and meta-validation cohorts. (B) and (D) Relapse-free survival among patients with stages I and II CRC in the training and meta-validation cohorts. Hazard ratios (HRs) and 95% CIs are for high and low immune risk. P values are based on log-rank tests.

3.3. Biological processes associated with the signature

Patients were stratified into high- and low-risk groups according to the PRGPI signature, and gene set enrichment analysis (GSEA) was performed on the CIT dataset. Indeed, genes comprising the signatures of collagen binding, extracellular matrix, epithelial-mesenchymal transition (EMT), and focal adhesion—4 programs widely accepted for their important contribution in a mesenchymal phenotype—were highly enriched for the group with a high PRGPI signature (Fig. 2).

3.4. Comparison with oncotype Dx colon cancer

We also compared the PRGPI signature with Oncotype Dx colon cancer,^[25] which consisted of a 12-gene signature for stage II and

III CRC. We calculated Oncotype Dx risk scores for both training and validation cohorts. For the CIT data sets, the PRGPI achieved a higher C-index (mean C-index, 0.74) compared with the 12-gene signature (mean C-index, 0.56) for estimation of RFS. The PRGPI signature also achieved a higher accuracy [mean concordance index (C-index): 0.6~0.62] than Oncotype Dx (mean C-index, 0.48~0.53) in comparable validation sets (Fig. 3).

3.5. Integrated prognostic index by combining the PRGPI

In multivariate analysis, clinical factors (age, stage, and sex) and the PRGPI were independent prognostic factors in the CIT dataset, suggesting their complementary value. To improve the

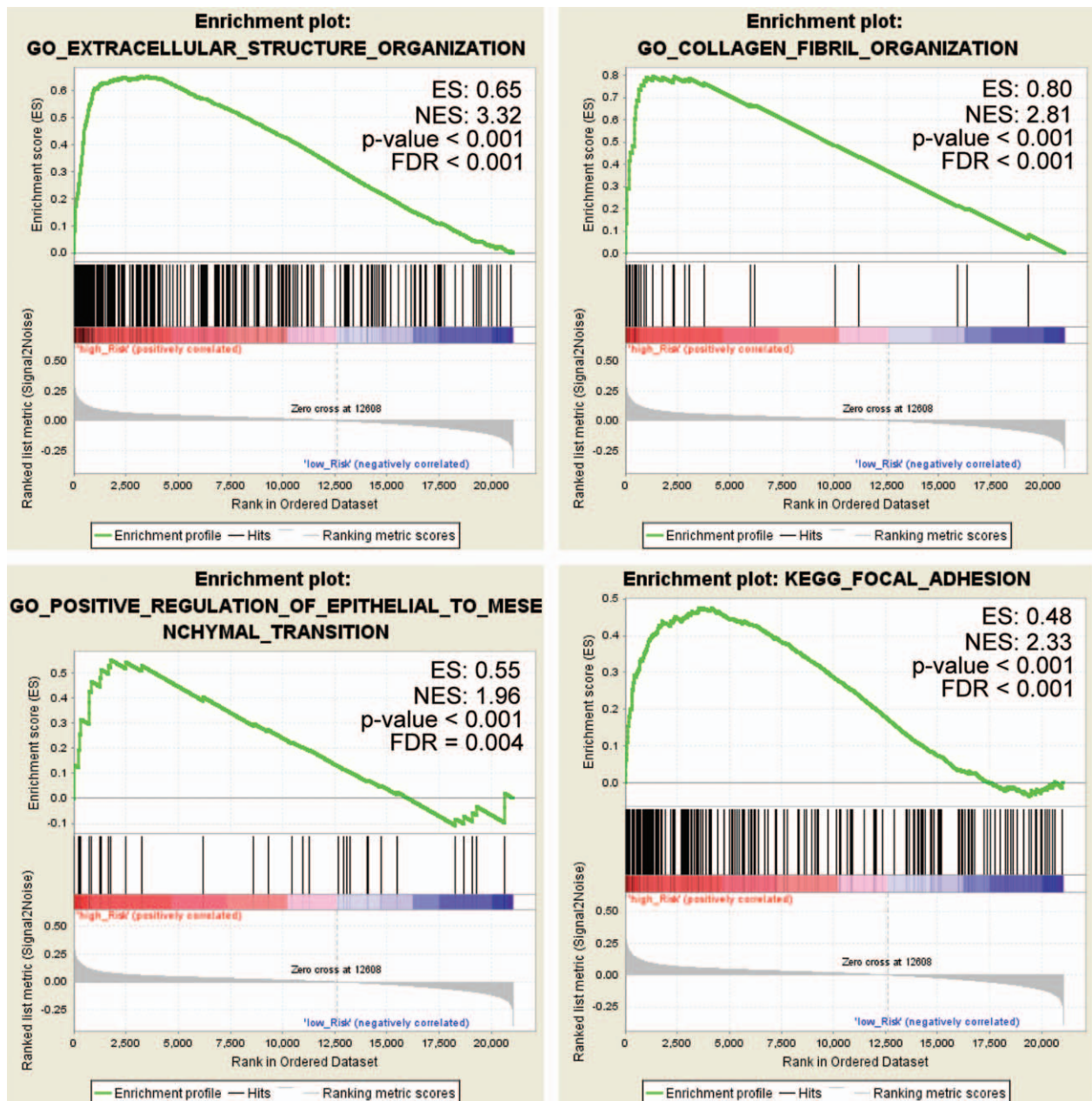


Figure 2. Gene set enrichment analysis (GSEA). Gene set enrichment analysis confirmed that EMT-related programs were upregulated in the high-risk group in the CIT data set. P values were calculated by GSEA software.

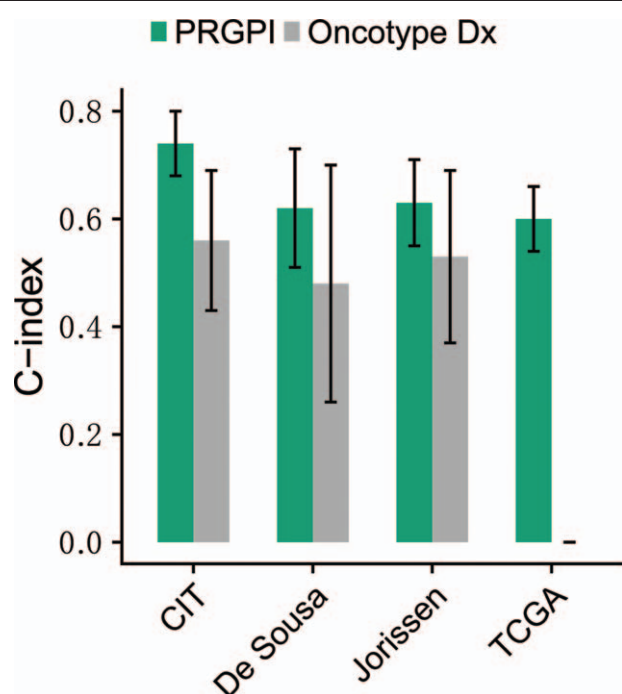


Figure 3. C-index comparison between PRGPI and oncotype Dx colon cancer. Comparison of C-index between oncotype Dx colon cancer signature and the PRGPI in the training and independent validation cohorts.

prognostic efficiency, the PRGPI signature was combined with age, stage, and sex to fit a Cox proportional hazards regression model using the CIT data set and formed a PCPI as $(1.834 \times \text{PRGPI}) + (0.999 \times \text{Sex}) + (0.022 \times \text{Age}) + (0.845 \times \text{Stage})$. On account of time-dependent ROC curve analysis, the optimal cutoff for the PCPI signature was chosen at 0.77 to classify patients into high- and low-risk groups in the meta-training data set (Fig. 4A). Significantly improved prognostic power was achieved by the PCPI compared with the PRGPI in the meta-validation cohorts (Fig. 4B).

4. Discussion and conclusion

Prognostic-related biomarkers are key to the risk stratification of patients with CRC and the decision regarding treatment. Reliable prognostic biomarkers are urgently needed to screen patients who are at highest risk of recurrence and who might require for additional systemic therapy. Currently, stage, grade, and microsatellite instability remain the most prevalent ways of assessing risk for CRC patients. A handful of multigene prognostic signatures^[10-13] has been developed in regard to CRC, but their accuracy of prognosis estimation remains uncertain. In this study, we established a prognostic signature based on 9 PRGPs for CRC and validated it in 3 independent cohorts. The PRGPI can classify CRC patients into groups with different clinical and biological outcomes. The PRGPI achieved higher accuracy than a commercialized molecular biomarker. We further combined the PRGPI signature and clinical factors and showed a higher accuracy estimation of RFS in CRC.

Considering the potential heterogeneity of tumors and the technical bias caused by the sequencing or microarray platforms, traditional prognostic risk models require appropriate normali-

zation of gene expression profiles, which is a difficulty of data analysis. To identify a robust signature for CRC prognosis, we used a method that can perform robustly regardless of the technical biases across different platforms.^[29] Our proposed signature has no need for data preprocessing, such as scaling and normalization, which is accomplished by relative ranking of gene expression values and selecting pairwise comparison. This approach has been reported to generate robust outcomes in various studies, including cancer subtyping.^[15] The risk score calculated by this gene pair based signature was wholly based on the gene expression values of single-sample, individualized CRC patients. This showed the robustness of the PRGPs signature.

The 9 PRGPs contained 17 unique genes. Within these 17 signature genes, only KLK10 has been reported to be an independent unfavorable predictor of DFS and OS in CRC patients.^[30,31] Overexpression of ITGA5 and S100A2 has previously been reported to be associated with poor outcome in non-small-cell lung cancer (NSCLC).^[32,33] WWC3 downregulation is correlated with malignant phenotype and poor prognosis in human gastric cancer (GC).^[34] KRT17 has been shown to be a possible biomarker in GC promoting tumor growth, motility, and invasion.^[35] Growing evidence suggest that BRIP1 may have an antioncogenic role, and downregulation of BRIP1 has been observed in multiple types of cancer.^[36] On account of the unbalanced expression in specific gene pairs, which might play a more important role than individual genes, the remaining 11 genes might also play a role in CRC prognosis prediction.

The limitations of our study should also be confronted for our study. First, the PRGPI was built on gene expression profiles generated by RNA-seq or microarray platforms, which are difficult to promote in routine clinical applications due to their exorbitant price, long conversion cycle, and requirement of bioinformatics expertise. Several alternative options may be worth exploring, such as IHC-based assays derived from optimized signature genes filtered from the full list of PRGPs. Second, the training cohort used for constructing the PRGPI was derived from a retrospective study and contained fresh frozen samples; hence, the robustness and efficiency of the PRGPI on FFPE samples are still questionable. More data sets with different sample properties need to be integrated for extensive validation. Third, we have only included the microarray data based on Affymetrix, which might produce selection bias. More sequencing platforms should be considered to test the robustness of the PRGPI, such as Illumina and Agilent.

In conclusion, our study developed a novel gene pairs signature for CRC prognosis estimation.

Author contributions

Conceptualization: Xingguo Zhang.

Data curation: Xingguo Zhang, Peng Shu, Yao Tong.

Formal analysis: Peng Shu.

Funding acquisition: Xingguo Zhang, Chunxia Xu.

Investigation: Xingguo Zhang, Chunxia Xu.

Methodology: Xingguo Zhang, Jianping Wu.

Project administration: Xingguo Zhang.

Resources: Jianping Wu.

Supervision: Xingguo Zhang.

Validation: Peng Shu.

Visualization: Peng Shu.

Writing – original draft: Xingguo Zhang, Peng Shu, Jianping Wu.

Writing – review & editing: Xingguo Zhang, Chunxia Xu.

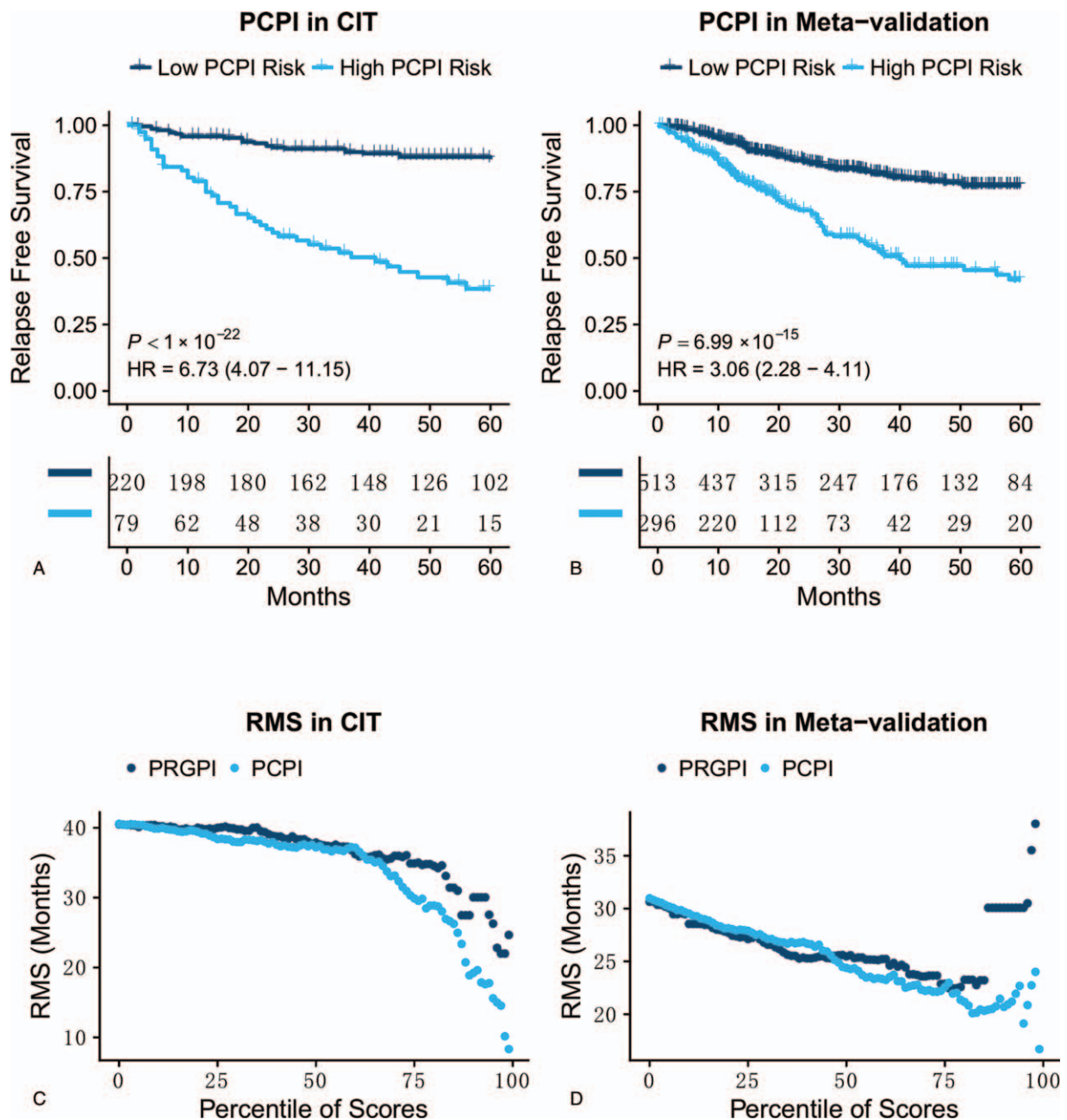


Figure 4. Kaplan–Meier curves and restricted mean survival (RMS) curves for prognostic-clinical prognostic index (PCPI) prediction. Kaplan–Meier curves for relapse-free survival of all patients stratified by the PCPI in the training (A) and meta-validation cohorts (B). The RMS curve for prognostic-related gene pair index (PRGPI) and PCPI scores in the training (C) and meta-validation cohorts (D).

References

- [1] Siegel RL, Miller KD, Fedewa SA, et al. Colorectal cancer statistics, 2017. *CA Cancer J Clin* 2017;67:177–93.
- [2] Verma M, Kumar V. Epigenetic biomarkers in colorectal cancer. *Mol Diagn Ther* 2016;21:153–65.
- [3] Frampton M, Houlston RS. Modeling the prevention of colorectal cancer from the combined impact of host and behavioral risk factors. *Genet Med* 2017;19:314–21.
- [4] Inadomi JM. Screening for colorectal neoplasia. *N Engl J Med* 2017;376:149–56.
- [5] Chang W, Gao X, Han Y, et al. Gene expression profiling-derived immunohistochemistry signature with high prognostic value in colorectal carcinoma. *Gut* 2014;63:1457–67.
- [6] Dalerba P, Sahoo D, Paik S, et al. CDX2 as a prognostic biomarker in stage II and stage III colon cancer. *N Engl J Med* 2016;374:211–22.
- [7] QUASAR Collaborative Group, QUASAR Collaborative Group. Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomised study. *Lancet* 2007;370:2020–9.
- [8] Figueredo A, Coombes ME, Mukherjee S. Adjuvant therapy for completely resected stage II colon cancer. *Cochrane Database Syst Rev* 2008;CD005390.
- [9] Dotan E, Cohen SJ. Challenges in the management of stage II colon cancer. *Semin Oncol* 2011;38:511–20.
- [10] Barrier A, Boelle PY, Roser F, et al. Stage II colon cancer prognosis prediction by tumor gene expression profiling. *J Clin Oncol* 2006;24:4685–91.
- [11] Wang Y, Jatkoe T, Zhang Y, et al. Gene expression profiles and molecular markers to predict recurrence of Dukes’ B colon cancer. *J Clin Oncol* 2004;22:1564–71.
- [12] Yothers G, O’Connell MJ, Lee M, et al. Validation of the 12-gene colon cancer recurrence score in NSABP C-07 as a predictor of recurrence in patients with stage II and III colon cancer treated with fluorouracil and

- leucovorin (FU/LV) and FU/LV plus oxaliplatin. *J Clin Oncol* 2013;31:4512–9.
- [13] Salazar R, Roepman P, Capella G, et al. Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer. *J Clin Oncol* 2011;29:17–24.
- [14] Leek JT, Scharpf RB, Bravo HC, et al. Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat Rev Genet* 2010;11:733–9.
- [15] Popovici V, Budinska E, Tejpar S, et al. Identification of a poor-prognosis BRAF-mutant-like population of patients with colon cancer. *J Clin Oncol* 2012;30:1288–95.
- [16] Heinäniemi M, Nykter M, Kramer R, et al. Gene-pair expression signatures reveal lineage control. *Nat Methods* 2013;10:577–83.
- [17] Li B, Cui Y, Diehn M, et al. Development and validation of an individualized immune prognostic signature in early-stage nonsquamous non-small cell lung cancer. *JAMA Oncol* 2017;3:1529–37.
- [18] Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 2007;23:1846–7.
- [19] Marisa L, de Reyniès A, Duval A, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. *PLoS Med* 2013;10:e1001453.
- [20] Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012;487:330–7.
- [21] Jorissen RN, Gibbs P, Christie M, et al. Metastasis-associated gene expression changes predict poor outcomes in patients with Dukes Stage B and C colorectal cancer. *Clin Cancer Res* 2009;15:7642–51.
- [22] de Sousa E, Melo F, Colak S, et al. Methylation of cancer-stem-cell-associated Wnt target genes predicts poor prognosis in colorectal cancer patients. *Cell Stem Cell* 2011;9:476–85.
- [23] Gautier L, Cope L, Bolstad BM, et al. affy: analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 2004;20:307–15.
- [24] Miller JA, Cai C, Langfelder P, et al. Strategies for aggregating gene expression data: the collapseRows R function. *BMC Bioinformatics* 2011;12:322.
- [25] O'Connell MJ, Lavery I, Yothers G, et al. Relationship between tumor gene expression and recurrence in four independent studies of patients with stage II/III colon cancer treated with surgery alone or surgery plus adjuvant fluorouracil plus leucovorin. *J Clin Oncol* 2010;28:3937–44.
- [26] Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–50.
- [27] Eng KH, Schiller E, Morrell K. On representing the prognostic value of continuous gene expression biomarkers with the restricted mean survival curve. *Oncotarget* 2015;6:36308–18.
- [28] Schröder MS, Culhane AC, Quackenbush J, et al. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics* 2011;27:3206–8.
- [29] Eddy JA, Sung J, Geman D, et al. Relative expression analysis for molecular cancer diagnosis and prognosis. *Technol Cancer Res Treat* 2010;9:149–59.
- [30] Alexopoulou DK, Papadopoulos IN, Scorilas A. Clinical significance of kallikrein-related peptidase (KLK10) mRNA expression in colorectal cancer. *Clin Biochem* 2013;46:1453–61.
- [31] Talieri M, Alexopoulou DK, Scorilas A, et al. Expression analysis and clinical evaluation of kallikrein-related peptidase 10 (KLK10) in colorectal cancer. *Tumour Biol* 2011;32:737–44.
- [32] Dingemans A-MC, van den Boogaart V, Vosse BA, van Suylen R-J, Griffioen AW, Thijssen VL. Integrin expression profiling identifies integrin alpha5 and beta1 as prognostic factors in early stage non-small cell lung cancer. *Mol Cancer* 2010;9:152.
- [33] Wang H, Zhang Z, Li R, et al. Overexpression of S100A2 protein as a prognostic marker for patients with stage I non small cell lung cancer. *Int J Cancer* 2005;1162:285–90.
- [34] Hou J, Zhou J. WWC3 downregulation correlates with poor prognosis and inhibition of Hippo signaling in human gastric cancer. *Onco Targets Ther* 2017;10:2931–42.
- [35] Chivu-Economescu M, Dragu DL, Necula LG, et al. Knockdown of KRT17 by siRNA induces antitumoral effects on gastric cancer cells. *Gastric Cancer* 2017;206:948–59.
- [36] Rafnar T, Gudbjartsson DF, Sulem P, et al. Mutations in BRIP1 confer high risk of ovarian cancer. *Nat Genet* 2011;4311:1104–7.