Cancer Management and Research

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

A novel 4-gene prognostic signature for hypermutated colorectal cancer

This article was published in the following Dove Medical Press journal:
Cancer Management and Research

Weiting Ge[1]
Wen Cai[1,2]
Rui Bai[1]
Wangxiong Hu[1]
Dehao Wu[1,2]
Shu Zheng[1]
Hanguang Hu[1,2]

[1]Cancer Institute (Key Laboratory of Cancer Prevention and Intervention, China National Ministry of Education), The Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, China; [2]Department of Medical Oncology, The Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, China

Correspondence: Hanguang Hu
Department of Medical Oncology, The Second Affiliated Hospital, School of Medicine, Zhejiang University, 88 Jiefang Road, Hangzhou, Zhejiang, 310009, China
Tel +86 0571 8778 4795
Fax +86 0571 8776 7088
Email huhanguang@zju.edu.cn

Background: Hypermutated colorectal cancer (CRC) reportedly accounts for 15%–17% of all cases of CRC. However, the proportion and number of patients with hypermutated CRC cannot be unappreciated. Additionally, therapy options for these patients differ from those for CRC patients, with a greater potential benefit from immunotherapy.

Materials and methods: We sequenced the tumor mucosa of CRC patients with >24 months of follow-up data at our center and identified mutation profiles of hypermutated CRC as a training data set (Zhejiang University [ZJU]); we then collected patients from The Cancer Genome Atlas (TCGA) as a validation data set. Recurrently mutated genes were combined to calculate a compound score via Cox proportional hazards model. Patients with higher-than-median scores were segregated as the high-risk group. Outcomes were analyzed by Kaplan–Meier and Cox regression analyses using Python (3.6.0) and R (3.4.0).

Results: We constructed a 4-gene signature (*ACVR2A, APC, DOCK2*, and *POLE*), with training in 45 hypermutated patients at ZJU and validation in 24 hypermutated patients from TCGA. Patients in the high-risk group showed poor survival (adjusted HR =9.85, 95% CI: 2.07–46.81, *P*=0.004). Further subgroup analysis was performed for stage II and III colon cancer (HR =10.91, 95% CI: 1.36–87.5, *P*=0.005) and high microsatellite instability (MSI-H) CRC (HR =12.57, 95% CI: 1.57–100.69, *P*=0.002) subgroups, which verified that our signature is universal. We then compared our prognostic signature with other risk factors (including MSI status, POLE driver mutation, BRAF-p.V600E, tumor mutational burden, and TNM staging). The results proved that our 4-gene signature is better than the other risk factor for prognosis in hypermutated CRC.

Conclusion: Our 4-gene signature is a good predictor of survival for hypermutated CRC, and this signature is powerful in stage II and III colon cancer and MSI-H CRC. Future prospective studies are needed to confirm the power of the 4-gene signature in patients receiving immunotherapy.

Keywords: colorectal cancer, hypermutation, gene signature, prognosis

Introduction

Colorectal cancer (CRC) is one of the most commonly diagnosed cancers worldwide. Indeed, most recent data show that CRC is the second leading cause of cancer death, with over 500,000 deaths annually.[1]

Hypermutation is characterized by mismatch repair (MMR) deficiency or POLE/POLD1 driver mutations, indicating a high mutational rate. Accumulating evidence confirms that hypermutation occurs in many cancer types, such as melanoma,[2] lung cancer,[3] and bladder cancer.[4]

Acquisition of genomic instability is a crucial feature of CRC development, and the microsatellite instability (MSI) pathway, which in involved in 15%[5,6] of CRC

submit your manuscript | www.dovepress.com
Dovepress
http://dx.doi.org/10.2147/CMAR.S190963

pathogeneses, is known to arise from inactivation of one of four MMR genes: *MSH2, MLH1, MSH6*, and *PMS2*.[7–10] An aberrant MMR process leads to additive mutations throughout the genome and ultimately a "hypermutator" phenotype.[10,11] The results of a recent large-scale sequencing analysis also suggest that hypermutation in CRC should not be underestimated.[12] Moreover, hypermutation is associated with a predicted high neo-epitope load, which can be exploited for immunotherapy in selected patients who experience conventional therapy failure.[13]

Because hypermutated CRC differs from CRC with regard to tumorigenesis, prognosis, and treatment, it is unreasonable to use the same prognostic biomarkers as employed for patients without hypermutations. Here, we propose a prognostic prediction signature for patients with hypermutated CRC. This signature can determine the hypermutated status of tumors and distinguish between patients who have a high or low risk of disease-related death and predict those who are likely to benefit from immunotherapy.

Materials and methods

Sample collection and DNA extraction

Fresh tissues were flash frozen in liquid nitrogen after surgery and stored at −80°C until genomic DNA isolation. Only samples from patients with a pathologic diagnosis of CRC were evaluated, but samples were excluded if they contained <40% tumor cells. Tumor- and matched normal mucosa-derived DNA was purified using the QIAamp DNA mini kit (Qiagen NV, Venlo, The Netherlands).

All patients signed a consent form when the frozen tissue was obtained. This consent form authorized us to conduct scientific research and publish the results anonymously with the approval of the Institutional Review Board (IRB). All protocols and procedures of this study were approved under The Second Affiliated Hospital, School of Medicine of Zhejiang University IRB protocol 2013-042. This study was conducted in accordance with the Declaration of Helsinki.

DNA sequencing and MSI status detection

Panel sequencing was performed using a custom-designed panel utilizing Agilent SureSelect capture kit technology (Agilent Technologies, Santa Clara, CA, USA), targeting the exonic region of 524 genes (Table S1), with HiSeq 2000 (Illumina, San Diego, CA, USA). Sequence reads were aligned to the human genome (GRCh37/hg19), and unique pairs were used for variant calling. Candidate variants and indels were detected using GATK. Somatic mutations and indels were then identified using MuTect and Strelka, respectively. These variants were annotated with ANNOVAR (Openbioinformatics, USA).[14] MSI status was detected using mSINGS.[15]

Detection of hypermutated samples and candidate gene screening

The threshold of hypermutation was higher than 10 Mut/Mb, which was determined according to a recently published large-scale analysis.[12] As there are hundreds of mutated genes in a hypermutated tumor, it is impractical to use a mutation frequency ≥5% as a filtering criterion, which will include most of the genes sequenced. Therefore, we employed multiple strategies to screen for candidate prognostic genes in hypermutated CRC. First, we employed a univariate Cox proportional hazard (PH) model to evaluate the association between gene mutations and overall survival (OS) for each gene. Dedicator of cytokinesis 2 (*DOCK2*) was the only gene significantly associated with OS. Second, *POLE* driver mutation and MSI statuses were selected as candidates because replication repair mutations and MSI are observed in most types of hypermutated cancer. We identified *POLE* driver mutations based on the results of Campbell et al.[12] However, POLD1 was not included because no driver mutations were found in the present study. Third, activin A receptor type 2A (*ACVR2A*) and *BRAF* were included according to previous studies. *APC, TP53*, and *KRAS*, which have an overall higher mutation frequency in CRC, were also included. Thus, we obtained eight candidate factors (*DOCK2, ACVR2A, BRAF, APC, TP53, KRAS, POLE,* and MSI status) to construct a prognostic model. We summarize the selection process in Figure 1.

Compound score calculation and high-risk patient detection

The final follow-up date was October 1, 2016 for patients from The Second Affiliated Hospital, School of Medicine of Zhejiang University (the ZJU cohort) and August 20, 2015 for patients from The Cancer Genome Atlas (TCGA) cohort. Only patients with >24 months of follow-up survival data were used for the prognosis-related analysis. OS was measured in months from the date of surgery to the date that the patient died. Stage was assessed according to the seventh version of the American Joint Commission on Cancer guidelines.

The compound score was calculated based on the sum of candidate gene mutations for each patient using a Cox PH model. Patients with higher-than-median scores were segregated as the high-risk group; others were defined as the low-risk group. The Cox PH model was employed to evaluate

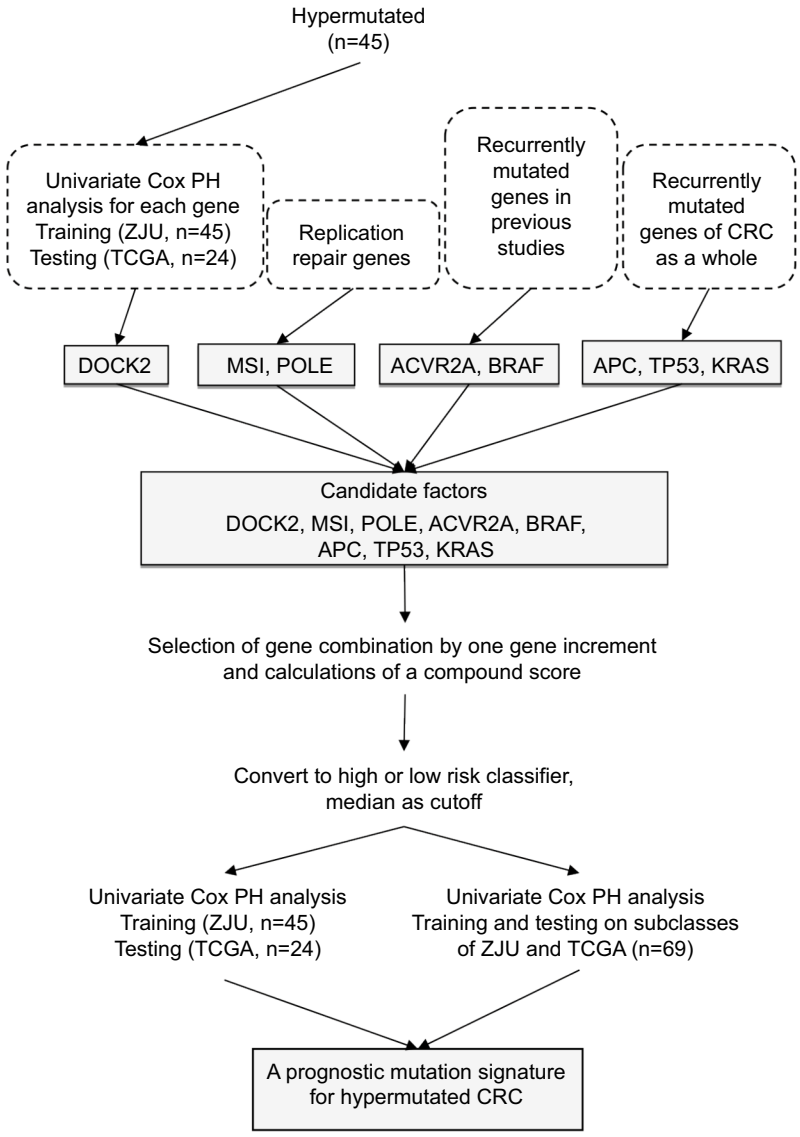

Figure 1 Flowchart of prognostic signatures generation and validation procedures.
Abbreviations: TCGA, The Cancer Genome Atlas; ZJU, Zhejiang University; PH, proportional hazard; CRC, colorectal cancer.

the association between the mutation signature and the clinical endpoint. To rule out over-fitting of the model, MSI data, somatic mutation data, and clinicopathological information for the TCGA cohort were obtained from the TCGA project data portal (http://www.cbioportal.org) on July 3, 2017.[16] These patients were used as an independent cohort for verification of the significance between the panel and OS.

Statistical analyses

Kaplan–Meier survival curve analysis with the log-rank test was applied to estimate the mutation signature in relation to OS. Fisher's exact test, Student's *t*-test, and the Mann–Whitney *U*-test were used to determine differences in clinicopathological variables between subgroups (high-risk vs low-risk/mutant vs wild-type). Multivariate Cox regression was performed to determine the contribution of the mutation signature to survival, adjusting for age, sex, and stage; the Wald test was employed in this analysis. Clinical subgroups of training and testing cohorts were also used to verify the prognostic prediction of the panel. Harrell's concordance index (C-index) was used to quantify predictive accuracies. In addition, multiple permutation tests were performed on both cohorts. To obtain a sufficient sample size for clinical subclass analysis, we pooled the two cohorts and utilized 5-fold

cross-validation. All statistical analyses were two-sided. A *P*-value of <0.05 was considered statistically significant. All analyses were performed using Python 3.6.0 (https://www.python.org/) and R 3.4.0 (https://www.r-project.org/).

Results

Patients' basic characteristics

We sequenced 338 samples in our center, from which we identified 45 (13.30%) hypermutated patients as the training data set to construct the prognostic mutation signature. The proportion of hypermutation obtained in this study coincides with that of a previous report. Samples that met the criteria (63 patients) of hypermutation were selected from the TCGA project data portal for an independent testing data set. The demographics of the two initial cohorts, with a total of 108 patients, are shown in Table S2. Due to the lack of information on specific staging, location, and survival data in the TCGA data set, 24 patients from TGCA were ultimately selected as the testing group for prognosis analysis. The baseline characteristics of the two groups are presented in Table 1.

Identified potential genes for constructing the signature

According to the three procedures described above, *DOCK2* was found to be the only gene significantly associated with OS in both the ZJU and TCGA cohorts (HR =1.73, *P*=0.007 and HR =2.08, *P*=0.001, respectively). Combining the remaining selected genes, *DOCK2, ACVR2A, BRAF, APC, TP53, KRAS*, and *POLE* as well as MSI status were chosen as candidate factors for constructing the prognostic signature.

To determine the minimum number of genes able to discriminate outcomes, we began training in the ZJU cohort and stopped increasing the size of the signature when we obtained the maximal C-index. We next constructed a risk classifier based on four genes (*DOCK2, ACVR2A, APC*, and *POLE*), with good concordance (C-index =0.748). We further used multiple permutation testing to confirm the robustness of this signature, which indicated that our signature was not generated by coincidence (Figure 2).

Table 1 Characteristics of included patients

	ZJU center (n=45)	TCGA (n=24)	*P*-value
Age, years, mean ± SD	60.00±2.10	70.04±2.87	0.168
Median OS, months, mean ± SD	67.97±4.30	42.32±5.76	0.410
Sex, female, n (%)	16 (35.56)	11 (45.83)	0.405
Mutation burden, mb, mean ± SD	73.62±12.08	43.01±7.67	0.344
Stage, n (%)			0.411
I	4 (8.89)	5 (20.83)	
II	27 (60.00)	12 (50.00)	
III	13 (28.89)	5 (20.83)	
IV	1 (2.22)	0 (0.00)	
Location, n (%)			0.025
Right-side colon	22 (48.89)	19 (79.17)	
Left-side colon	14 (31.11)	3 (12.50)	
Rectum	9 (20.00)	1 (4.17)	
MSI-H, n (%)	22 (48.89)	21 (87.50)	0.002
POLE driver mutant, n (%)	9 (20.00)	1 (4.17)	0.075
ACVR2A mutant, n (%)	34 (75.55)	7 (29.17)	0.002
APC mutant, n (%)	28 (62.22)	7 (29.17)	0.312
DOCK2 mutant, n (%)	20 (44.44)	7 (29.17)	0.278

Abbreviations: OS, overall survival; MSI-H, high microsatellite instability; TCGA, The Cancer Genome Atlas; ZJU, Zhejiang University.

Construction of a 4-gene signature for high-risk hypermutated CRC patients

We divided patients into two groups by calculating the compound score for four genes. The compound score was generated from the ZJU cohort and further calculated for each patient in the TCGA cohort. Patients with higher-than-median scores were segregated as the high-risk group, and others were defined as the low-risk group. Patients with a high-risk classifier in the ZJU (HR =10.19, 95% CI: 1.25–83.23, *P*=0.007) and TCGA (HR =8.62, 95% CI: 1.87–39.67, *P*=0.001) cohorts had significantly worse survival.

Due to the limited number of patients exhibiting hypermutation (n=45 in ZJU and n=24 in TCGA), we pooled the two cohorts to display the results and perform further subgroup analysis. A high-risk classifier was significantly associated with poor survival (HR =8.83, 95% CI: 2.00–39.04, *P*=0.001; Figure 3).

We also employed a multivariate Cox regression analysis including our 4-gene risk classifier, age, sex, and stage to reveal that the 4-gene risk classifier can serve as an independent determinant of OS in patients with hypermutated CRC (adjusted HR =9.85, 95% CI: 2.07–46.81, *P*=0.004) (Table 2). There was no statistically significant difference in the proportion of patients receiving adjuvant chemotherapy between the high-risk and low-risk groups (*P*=0.457, Table S3).

The 4-gene signature is powerful in specific subgroups

We first employed the 4-gene signature for subgroup analysis of stage II and III colon cancer patients, accounting for 74% (78/105) of all included patients. Twenty-two patients were in the high-risk group and twenty-eight in the low-risk group according to the compound score. Because the high-risk patients shared a higher death risk, the results suggested

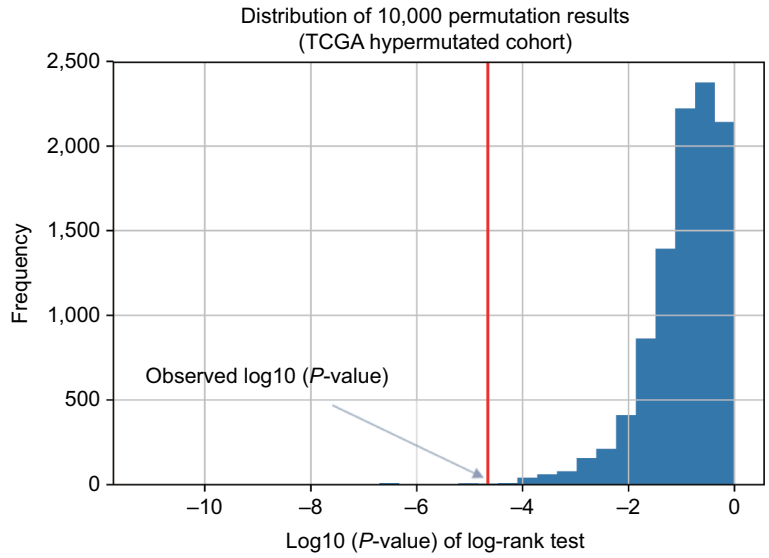

Figure 2 Distribution of 10,000 permutation results.
Abbreviation: TCGA, The Cancer Genome Atlas.

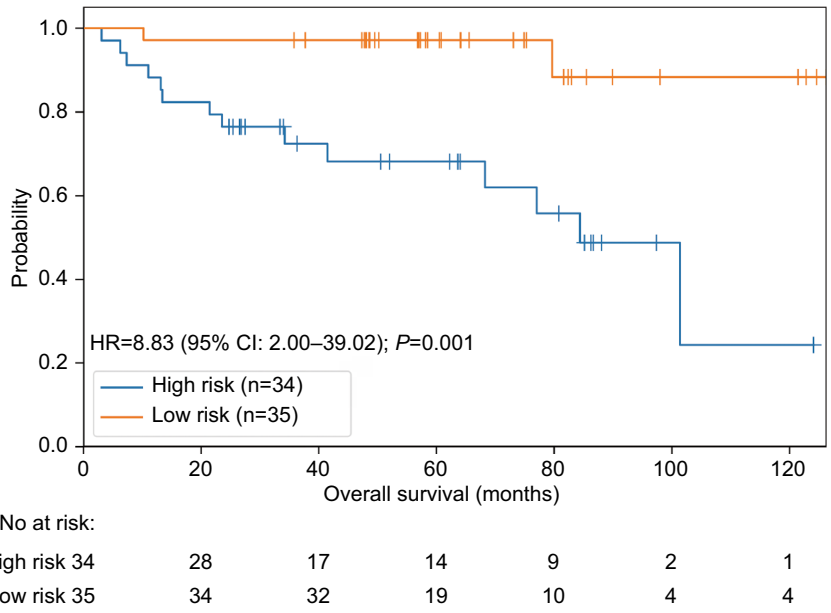

Figure 3 The pool data demonstrating effectiveness of the 4-gene signature.

Table 2 Multivariate Cox regression analysis of 4-gene signature

	HR (95% CI)	*P*-value
4-Gene risk model	9.85 (2.07–46.81)	0.004
Age	1.02 (0.98–1.06)	0.442
Sex, female vs male	2.49 (0.82–7.56)	0.108
Stage	0.65 (0.28–1.49)	0.307

Notes: Multivariate Cox regression analysis was performed on 67 of 69 pooled patients who met the criteria of having >24 months follow-up survival data, as well as age, sex, and stage information. *P*-value was calculated by Wald test.

that our 4-gene signature is a powerful prognostic tool for this subgroup (HR =10.91, 95% CI: 1.36–87.50, *P*=0.005; Figure 4A).

As hypermutations are more likely to occur on the right side of colon cancer,[14] we selected right-side colon cancer patients and employed the 4-gene signature to predict outcomes. The results showed markedly worse survival

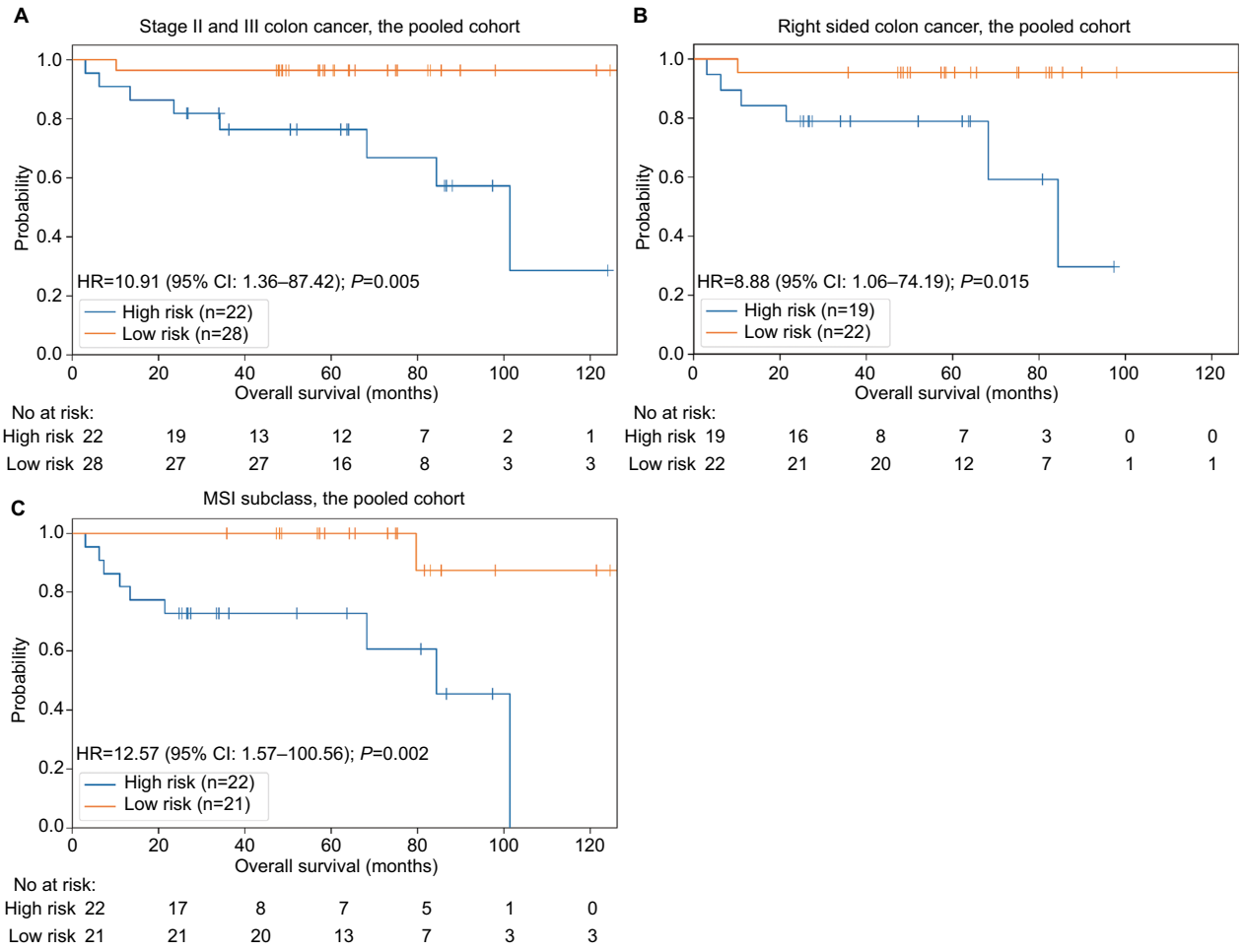

Figure 4 Analysis of different groups of patients.
Abbreviation: MSI, microsatellite instability.

among patients in the high-risk group (HR =8.88, 95% CI: 1.06–74.50, *P*=0.015; Figure 4B).

MSI status is important for both prognosis and therapy choice among patients with hypermutated tumors. Thus, we segregated patients into MSI and microsatellite stable (MSS) subgroups, and the results showed that our signature can well predict outcomes in the MSI group (HR =12.57, 95% CI: 1.57–100.69, *P*=0.002; Figure 4C), although the difference was not significant in the MSS group (*P*=0.082).

Comparison of prognostic biomarkers in hypermutated CRC

We established a prognostic predictor based on four genes (*DOCK2, ACVR2A, APC,* and *POLE*) for hypermutated CRC. Previous studies have also proposed prognostic risk stratifications based on MSI/POLE or MSI/BRAF. Hypermutation, MSI status, and tumor mutational burden (TMB) are correlated with the response to immune checkpoint inhibitors, and we therefore conducted a performance comparison for these risk factors/groups using the pooled cohort.

Our 4-gene risk classifier stratifies patients with hypermutated CRC into high-risk and low-risk groups. Worse outcomes were significantly associated with the high-risk group. MSI/POLE status divided patients into three groups: POLE, MSI, and POLE/MSI. Although the MSI group accounted for the majority of patients, with better survival than those of the POLE group (75% vs 40%, 5-year survival rate), the difference was not significant (Figure S1A). No death occurred in two patients in the POLE/MSI group. MSI/BRAF status also classified patients into three groups, MSI, MSS/BRAF-wild-type, and MSS/BRAF-p.V600E, and there was no prognostic difference between the first two groups, with only one patient in the last group (Figure S1B). To investigate the association between TMB and prognosis, we divided patients into high-TMB and low-TMB groups based on the median (40 mut/Mb). High-TMB patients showed a worse,

but not significant, outcome (Figure S1C). All patients with a POLE driver mutation were sorted into the high-TMB group, which was confirmed by the poor prognosis of the POLE group. Lastly, consistent with the results of the multivariate Cox regression analysis, TNM staging confirmed that the prognosis of hypermutated CRC does not comply with the TNM staging system (Figure S1D).

Discussion

In our study, patients with hypermutated CRC comprised 13.30%, which agrees with previous TCGA studies (in which hypermutated CRC accounted for 15.6%,[16] 16.9%,[17] and 15%).[18] To predict the prognosis of hypermutated CRC, we constructed a prognostic mutation signature of four genes (*DOCK2, ACVR2A, APC,* and *POLE*) that separated patients into two risk groups. The high-risk group was significantly associated with worse survival (HR =8.62, *P*=0.001). Overall, the proportion of hypermutated CRC suggested that hypermutation should not be underestimated in CRC. Furthermore, this special subgroup of CRC patients deserves greater attention because there are currently many novel therapeutic options. For example, Campbell et al[12] recently reported dramatically different survival rates among 217 patients with hypermutated cancer. In addition, the abundance of immune checkpoint inhibitors that target programmed cell death 1 (PD-1) or their ligands (PD-L1) brings new hope to hypermutated tumor patients. Many studies have proven that hypermutation and MSI status predict the clinical benefit of immune checkpoint inhibitors, which can lead to durable remission in some patients with conventional therapy failure.[19,20] Thus, establishing a prognostic biomarker will help physicians screen patients who have poor prognoses but good responses to immunotherapy. Regardless, not only are the clinicopathological features of hypermutated tumors poorly understood, but prognostic tools are also lacking due to a lack of understanding of this type of disease. Recently, sequencing of individual cancer genomes has prompted scientists to search for biomarkers based on gene mutation signatures. However, most previous efforts treated hypermutation and non-hypermutation as a whole or focused on non-hypermutated CRC.[21] To the best of our knowledge, no well-known study has been conducted on hypermutated CRC, highlighting the value of our work.

The genes in our signature include *DOCK2, ACVR2A, APC,* and *POLE*. *APC* and *POLE* play important roles in the development and progression of CRC, and *DOCK2* and *ACVR2A* are worth discussing further. *DOCK2* is a gene frequently mutated in CRC and esophageal cancer,[22] and data for the pooled cohort showed that *DOCK2* is more frequently mutated in hypermutated CRC (38.0%) than in non-hypermutated CRC (3.9%). Germline deficiency of *DOCK2* leads to life threatening, invasive bacterial and viral infections,[23,24] and transgenic mice experiments confirmed that *DOCK2* is required for recruitment and infiltration of immune cells into the colon mucosa during bacterial infection.[24,25] Several recent studies have reported that the gut microbiome influences the efficacy of anti-PD-1/L1 immunotherapy.[26,27] Accordingly, *DOCK2* may participate in the immune response initiated by gut microbes, and its deficiency is most likely involved in an immune evasion mechanism of high-risk hypermutated CRC. *ACVR2A* has two 8-bp polyadenine tracts, a hot spot for mutation in MSI CRC.[28,29] *ACVR2A* is a member of the TGF-β superfamily, which plays a key role during CRC progression, and a previous study showed that *ACVR2A* is the most frequently mutated gene of hypermutated CRC. Further study of *ACVR2A* is urgently needed.

We performed subgroup analysis to verify the universality of our 4-gene signature. Due to the different treatment strategies and clinical outcomes in colon and rectal cancer, stage II and III colon cancer patients were selected as the first subgroup, and the results showed that our signature is sufficiently powerful in this subgroup. Nonetheless, the effectiveness of our 4-gene signature in rectal cancer needs to be further investigated. We then focused on the subgroup of high microsatellite instability (MSI-H) patients because these CRC patients are considered to have different clinical outcomes and therapeutic choices; for example, patients with MSI-H CRC usually do not respond to 5-fluorouracil chemotherapy, unlike non-hypermutated CRC patients. We further classified MSI-H patients into two risk groups using our 4-gene signature and observed that the low-risk group showed markedly good survival. Although recent studies have reported that MSI-H patients have a good response to immune checkpoint inhibitors, there are still some patients for whom benefit may not actually be gained or for whom the response may not be translated to OS. Using our 4-gene signature in MSI patients to identify different risk groups is promising in clinical practice that selecting patients who will actually gain benefits from immunotherapy, with an effect on OS. The final subgroup analysis focused on the primary locations, as hypermutation is more common in right-side colon cancer.[16] This subgroup analysis further verified that our 4-gene signature is a powerful tool for patients with hypermutated CRC.

Using gene mutation status to construct a prognostic signature is currently a hot topic in cancer research. A powerful

prognostic signature will identify different risk group of patients to receive different intensity of therapies. Our 4-gene signature classifies hypermutation CRC patients into two groups and we suggest that high-risk group patients should receive more aggressive treatments than the low-risk group patients. A previous study stratified patients into three groups based on the MSI/BRAF status: MSI/BRAF-wild-type or mutant (best prognosis), MSS/BRAF-wild-type (intermediate prognosis), and MSS/BRAF mutant (worst prognosis).[30] Although the MSI/BRAF status can as a whole serve as a prognostic biomarker for CRC, it was not suitable for hypermutated CRC according to our study. TMB has also been recommended as a critical criterion correlated with the objective response rate of anti-PD-1/L1 immunotherapy. However, the threshold in different kinds of tumor remains controversial, and the power for predicting patient prognosis needs to be developed. Compared with the above biomarkers, our 4-gene risk classifier exhibits good performance for predicting clinical outcome in patients with hypermutated CRC.

However, there are some limitations to our study; the most important one is the limited number of patients in our study group due to the small percentage of hypermutated CRC. Fortunately, gene sequencing technology is gradually maturing and becoming faster and less expensive. We will continue to collect cases of hypermutated CRC to further verify our signature. Furthermore, although we believe that the 4-gene signature is promising in selecting patients who will gain benefits from immunotherapy, its significant value still needs to be verified in prospective studies that we are devoting ourselves to.

Conclusion

We constructed a 4-gene signature for patients with hypermutated CRC, which can classify patients into different risk groups and predict prognosis. This signature is a powerful tool in stage II and III colon cancer and MSI-H CRC. Future prospective studies are needed to confirm the power of the 4-gene signature in patients receiving immunotherapy.

Acknowledgments

This study was supported by the Fundamental Research Funds for the Central Universities, National Human Genetic Resources Sharing Service Platform (2005DKA21300), and National High Technology Research and Development Program of China (2012AA02A204). The abstract of this paper was presented at the MAP Conference named "A novel 4-gene prognostic signature for hypermutated colorectal cancer (CRC)" as a poster presentation with interim findings. The poster's abstract was published in "Poster Abstracts" in ESMO's official journal Annals of Oncology.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Wolf AMD, Fontham ETH, Church TR, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American cancer Society. *CA Cancer J Clin*. 2018;68(4):250–281.
2. Akbani R, Akdemir KC, Aksoy BA, et al. Genomic classification of cutaneous melanoma. *Cell*. 2015;161(7):1681–1696.
3. Govindan R, Ding L, Griffith M, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell*. 2012;150(6):1121–1134.
4. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature*. 2014;507(7492):315–322.
5. Rehder DS, Nelson RW, Borges CR. Glycosylation status of vitamin D binding protein in cancer patients. *Protein Sci*. 2009;18(10):2036–2042.
6. de La Chapelle A, Hampel H. Clinical relevance of microsatellite instability in colorectal cancer. *J Clin Oncol*. 2010;28(20):3380–3387.
7. Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature*. 1993;363(6429):558–561.
8. Aaltonen LA, Peltomäki P, Leach FS, et al. Clues to the pathogenesis of familial colorectal cancer. *Science*. 1993;260(5109):812–816.
9. Jiricny J. The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol*. 2006;7(5):335–346.
10. Lee V, Murphy A, Le DT, Diaz LA. Mismatch repair deficiency and response to immune checkpoint blockade. *Oncologist*. 2016;21(10):1200–1211.
11. Lee SY, Chung H, Devaraj B, et al. Microsatellite alterations at selected tetranucleotide repeats are associated with morphologies of colorectal neoplasias. *Gastroenterology*. 2010;139(5):1519–1525.
12. Campbell BB, Light N, Fabrizio D, et al. Comprehensive analysis of hypermutation in human cancer. *Cell*. 2017;171(5):1042–1056.
13. Jover R, Nguyen TP, Pérez-Carbonell L, et al. 5-fluorouracil adjuvant chemotherapy does not increase survival in patients with CpG island methylator phenotype colorectal cancer. *Gastroenterology*. 2011;140(4):1174–1181.
14. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164
15. Salipante SJ, Scroggins SM, Hampel HL, Turner EH, Pritchard CC. Microsatellite instability detection by next generation sequencing. *Clin Chem*. 2014;60(9):1192–1199.
16. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012;487(7407):330–337.
17. Tian S, Roepman P, Popovici V, et al. A robust genomic signature for the detection of colorectal cancer patients with microsatellite instability phenotype and high mutation frequency. *J Pathol*. 2012;228(4):586–595.
18. Arora S, Yan H, Cho I, et al. Genetic variants that predispose to DNA double-strand breaks in lymphocytes from a subset of patients with familial colorectal carcinomas. *Gastroenterology*. 2015;149(7):1872–1883.
19. Dudley JC, Lin MT, Le DT, Eshleman JR. Microsatellite instability as a biomarker for PD-1 blockade. *Clin Cancer Res*. 2016;22(4):813–820.
20. La D. PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl J Med*. 2015;373(20):1979.
21. Sho S, Court CM, Winograd P, Russell MM, Tomlinson JS. A prognostic mutation panel for predicting cancer recurrence in stages II and III colorectal cancer. *J Surg Oncol*. 2017;116(8):996–1004.

22. Dulak AM, Stojanov P, Peng S, et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat Genet*. 2013;45(5):478–486.
23. Dobbs K, Domínguez Conde C, Zhang SY, et al. Inherited DOCK2 deficiency in patients with early-onset invasive infections. *N Engl J Med*. 2015;372(25):2409–2422.
24. Liu Z, Man SM, Zhu Q, et al. DOCK2 confers immunity and intestinal colonization resistance to *Citrobacter rodentium* infection. *Sci Rep*. 2016;6(1):27814.
25. Miao S, Zhang RY, Wang W, et al. Overexpression of dedicator of cytokinesis 2 correlates with good prognosis in colorectal cancer associated with more prominent CD8+ lymphocytes infiltration: a colorectal cancer analysis. *J Cell Biochem*. 2018;119(11):8962–8970.
26. Routy B, Le Chatelier E, Derosa L, et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. *Science*. 2018;359(6371):91–97.
27. Hampton T. Gut microbes may shape response to cancer immunotherapy. *JAMA*. 2018;319(5):430–431.
28. Hempen PM, Zhang L, Bansal RK, et al. Evidence of selection for clones having genetic inactivation of the activin A type II receptor (ACVR2) gene in gastrointestinal cancers. *Cancer Res*. 2003;63(5):994–999.
29. Jung B, Doctolero RT, Tajima A, et al. Loss of activin receptor type 2 protein expression in microsatellite unstable colon cancers. *Gastroenterology*. 2004;126(3):654–659.
30. Lochhead P, Kuchiba A, Imamura Y, et al. Microsatellite instability and BRAF mutation testing in colorectal cancer prognostication. *J Natl Cancer Inst*. 2013;105(15):1151–1156.

Supplementary materials

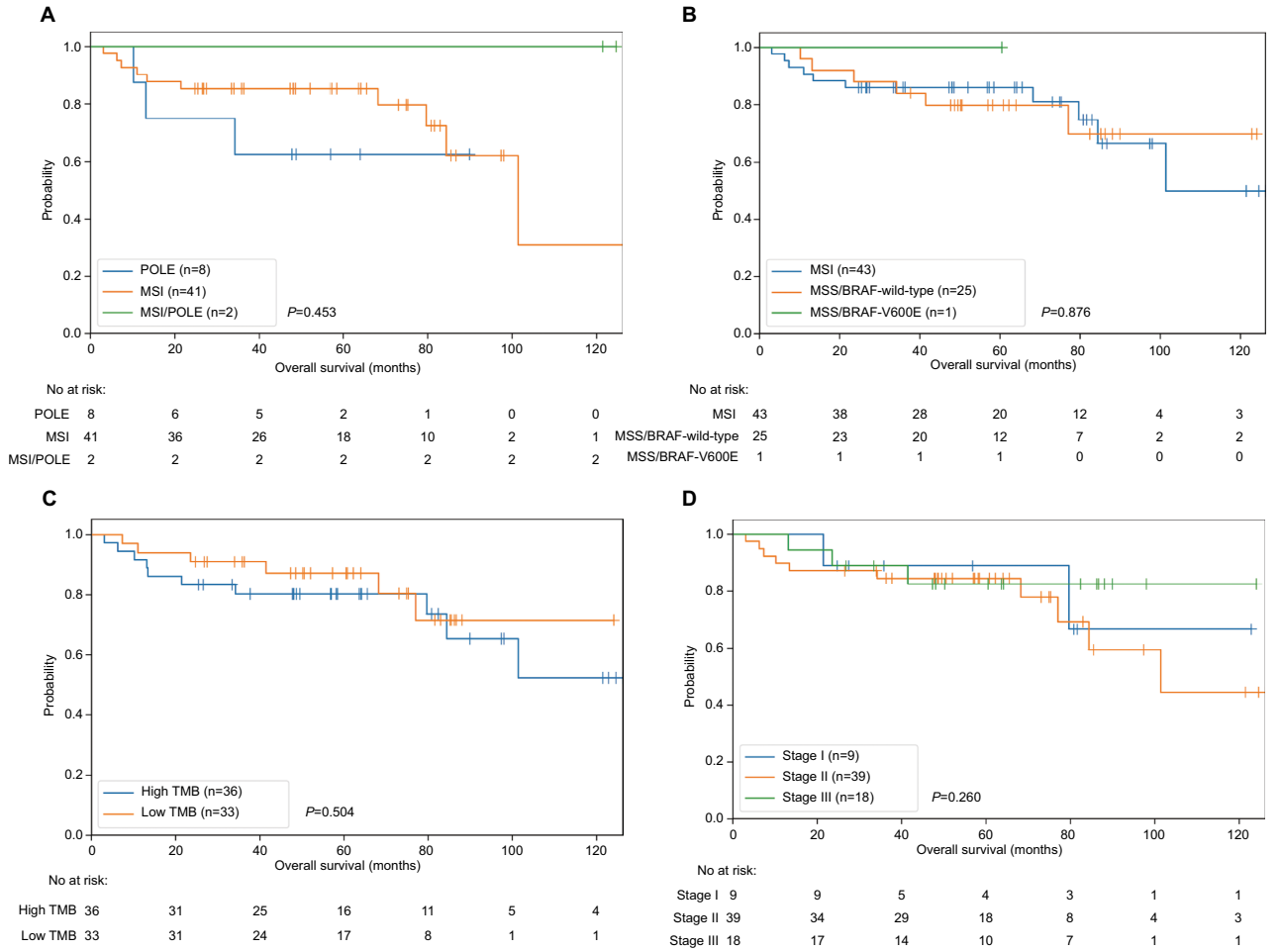

Figure S1 Power of different prognosis biomarkers.

Notes: (**A**) MSI/POLE status divides patients into three groups: POLE, MSI, and POLE/MSI. The difference of patients' survival is not significant. (**B**) MSI/BRAF status divides patients into three groups: MSI, MSS/BRAF-wild-type, and MSS/BRAF-V600E. The difference of patients' survival is not significant. (**C**) Patients were divided into high TMB and low TMB groups by median(as 40 mut/Mb). The difference of patients' survival is not significant. (**D**) Hypermutated patients were not comply the traditional TNM stages.

Abbreviations: MSI, microsatellite instability; MSS, microsatellite stable; TMB, tumor mutational burden.

Table S1 List of 524 genes for panel sequencing

ABCA13	ASPM	C20orf54	COL6A3	EIF3E	FAT4	HEMGN	LCORL	MSH2	NOLC1	PLD5	QSER1	SCN9A	ST8SIA6	TRAK2	ZFP36L2
ABCA6	ATAD5	C2orf44	COL6A6	EIF4A2	FBN2	HIST1H2BC	LEKR1	MSH3	NOTCH1	PLXNC1	RABGGTB	SEC63	STAG1	TRIM49B	ZFPM1
ABCA8	ATM	C9	CRB1	ELF3	FBXW7	HIST1H2BM	LGR5	MSH6	NOTCH2	PMS1	RAD21	SEMA3D	STK11	TRNT1	ZFPM2
ABCA9	ATP10A	CA4	CRTC1	ELMO1	FCRL1	HIVEP2	LIFR	MTOR	NPIPB5	PMS2	RAD50	SERPINF1	STK31	TSGA10	ZNF208
ABCB1	ATP1A2	CACNA1B	CSF2RA	ELMOD2	FETUB	HLA-B	LIG1	MTUS1	NRAS	POLD1	RAD51AP2	SHKBP1	STON1-GTF2A1L	TSHZ3	ZNF233
ACOX2	ATP1B2	CASP8	CSMD1	EP300	FLG	HMCN1	LIG4	MTUS2	NTHL1	POLD3	RALGAPA1	SI	STPG2	TSPO2	ZNF267
ACSL6	ATP8A2	CBL	CSMD3	EPCAM	FLG2	HNRNPL	LMNA	MUC1	NTSR2	POLE	RALY	SLC12A1	SUFU	TUSC3	ZNF285
ACTC1	ATR	CCDC141	CTNNB1	ERBB2	FLI1	HRNR	LPHN3	MUC12	NUDT11	POLE2	RASA1	SLC13A1	SYCP1	UACA	ZNF334
ACVR1B	ATXN1	CCDC168	CTNND2	ERBB4	FOXA1	HSPA8	LRIF1	MUC16	NUP107	POLE4	RASSF2	SLC25A51	SYNE1	UBE4A	ZNF382
ACVR2A	AUTS2	CCDC62	CTSA	ERCC1	FOXE1	HYDIN	LRP1B	MUC17	NUP50	POLN	RB1	SLC30A8	SYT16	UBN1	ZNF417
ADAM21	AXIN1	CCDC66	CUX1	ERCC5	FREM2	IDO2	LRP2	MUC4	OPRD1	POLQ	RBM12	SLC4A10	TBC1D7	UBR5	ZNF469
ADAM29	AXIN2	CCDC73	DCAF6	ERCC6	FRG1	IGSF3	LRRC4C	MUC6	OR11H12	POSTN	RBMX	SLC5A10	Table 1 XR1	UGT2A2	ZNF479
ADAM7	AZGP1	CCDC88A	DCHS2	ERICH3	FRY	IL15	LRRD1	MUTYH	PABPC1	POTEC	RBMXL3	SLCO1B1	TBP	UGT3A2	ZNF521
ADAMTS12	B2M	CCNE1	DDX6	EXO1	FSHR	IRF5	LRRIQ1	MYCBP2	PAFAH1B2	POTEE	REV3L	SLCO6A1	TCF7L2	UNC13C	ZNF569
ADAMTS20	B3GNT6	CD40LG	DEFB108B	EYS	FSIP2	ITLN2	LRRK2	MYCN	PARP1	POTEH	RFC1	SLITRK1	TDRD3	UNC5CL	ZNF600
ADCY2	B3GNT9	CD44	DNAH11	F10	FSTL5	ITPR1	LRRN3	MYO16	PARP4	POTEJ	RGS17	SLITRK4	TDRD6	UNC93A	ZNF665
ADM2	BAGE3	CD58	DNAH14	F13B	GABRG3	IVL	LRRTM3	MYO3B	PAX1	PPM1E	RGS22	SLX4	TEAD2	USH2A	ZNF671
AGBL2	BAI3	CD7	DNAH3	F5	GALNT12	JAK2	MAGEB3	MYO9A	PBRM1	PRDM9	RHCG	SMAD2	TELO2	UTP23	ZNF679
AKT1	BAIAP2L2	CDC42EP1	DNAH5	FAM123B	GGT1	KCNJ5	MAGEL2	MYOCD	PCBP1	PREX2	RHOA	SMAD3	TG	UTS2B	ZNF727
ALPK2	BAP1	CDH1	DNAH6	FAM135B	GLI3	KCNN3	MAL2	MYPOP	PCDH9	PRKAA1	RIMS2	SMAD4	TGFBR2	VN1R2	ZNF728
ALS2CR11	BAX	CDKN2A	DNAH7	FAM153B	GLT8D2	KCNS2	MANSC4	NAALAD2	PCDHB15	PRKRIR	RIN3	SNTG2	TGIF1	VPS37B	ZNF729
ALS2CR12	BCL9	CFAP58	DNAH8	FAM184A	GLTPD2	KCNT2	MAP1B	NALCN	PCGF6	PRSS35	RMI1	SORBS2	TLR4	VTI1A	ZNF735
ALX4	BCL9L	CHEK2	DNAJC24	FAM194B	GNAT2	KDM6A	MAP2K1	NBEA	PCLO	PSCA	RNF43	SOX2	TLR7	WBSCR17	ZNF750
AMPD1	BIVM-ERCC5	CLEC18B	DOCK10	FAM71E2	GNG12	KIAA0408	MAP2K4	NBPF15	PCSK7	PSG5	ROBO2	SOX9	TM6SF1	WDR11	ZNF761
ANKRD12	BLM	CLUL1	DOCK2	FAN1	GOLGA4	KIF4B	MAP3K19	NBPF3	PHF2	PTCH1	RP1	SPARCL1	TMEM173	XIRP2	ZNF787
ANKRD30A	BMPR1A	CMYA5	DPP10	FANCA	GOLGB1	KIRREL	MAPK1	NEB	PIEZO2	PTEN	RPA4	SPATA31D1	TMEM184A	XKR9	ZNF804A
APC	BMPR2	CNBD1	DPY19L2	FANCD2	GPN1	KMT2B	MAPK8IP1	NEFH	PIK3CA	PTPLA	RPL22	SPDYE6	TMPRSS13	XPC	ZNF804B
APOBEC3B	BRAF	CNTN1	DSEL	FANCE	GPR112	KRAS	MDM2	NFE2L2	PIK3R1	PTPN11	RPS20	SPG20	TNNI3K	XRCC1	ZNF831
APOBEC3F	BRCA1	CNTN6	DUX4	FANCI	GPR149	KRTAP10-6	MFRP	NIN	PIK3R3	PTPRK	RSPO2	SPIDR	TOP2B	ZBED9	ZNF850
ARHGAP15	BRCA2	CNTNAP5	DYNC2H1	FANCM	GREM1	KRTAP4-5	MKI67	NKX2-1	PKD2L2	PTPRT	RSPO3	SPP1	TOP3B	ZDBF2	
ARHGAP32	BRIP1	COL11A1	EBLN1	FAT1	GRIA2	KRTAP5-5	MLH1	NLRP14	PLCE1	PTPRZ1	RYR2	SPTA1	TOPAZ1	ZFC3H1	
ARHGEF12	C17orf70	COL12A1	EDNRB	FAT2	GRIN2A	KTN1	MLH3	NME8	PLCL1	PYHIN1	SACS	SPZ1	TP53	ZFHX4	
ARID1A	C1ORF74	COL18A1	EFHB	FAT3	HDGFRP2	LCA5	MNX1	NMI	PLCZ1	QRFPR	SCN1A	SSBP1	TPM3	ZFP28	

Table S2 Baseline values of all initial selected patients

	ZJU center (n=45)	TCGA (n=63)	*P*-value
Age, years, mean ± SD	60.00±2.10	66.02±1.94	0.459
Median OS, months, mean ± SD	67.97±4.30	40.73±4.73	0.413
Sex, female, n (%)	16 (35.56)	25 (39.68)	0.663
Mutation burden, mb, mean ± SD	73.62±12.08	46.40±5.14	0.000
Stage, n (%)			0.519
I	4 (8.89)	11 (17.46)	
II	27 (60.00)	35 (55.55)	
III	13 (28.89)	13 (20.63)	
IV	1 (2.22)	2 (3.17)	
Location, n (%)			0.003
Right-side colon	22 (48.89)	50 (79.37)	
Left-side colon	14 (31.11)	7 (11.11)	
Rectum	9 (20.00)	5 (7.94)	
MSI-H, n (%)	22 (48.89)	39 (61.90)	0.179
POLE driver mutant, n (%)	9 (20.00)	5 (7.94)	0.066
ACVR2A mutant, n (%)	34 (75.55)	18 (23.57)	0.000
APC mutant, n (%)	28 (62.22)	27 (42.86)	0.152
DOCK2 mutant, n (%)	20 (44.44)	42 (33.33)	0.284

Abbreviations: OS, overall survival; MSI-H, high microsatellite instability; TCGA, The Cancer Genome Atlas; ZJU, Zhejiang University.

Table S3 Chemotherapy of patients in the final analysis

	High-risk group (n=33)	Low-risk group (n=35)	*P*-value
Chemotherapy, n (%)	18 (54.55)	14 (40.00)	0.457