Open Access Full Text Article

ORIGINAL RESEARCH

A novel 4-gene prognostic signature for hypermutated colorectal cancer

Weiting Ge¹ Wen Cai^{1,2} Rui Bai¹ Wangxiong Hu¹ Dehao Wu^{1,2} Shu Zheng¹ Hanguang Hu^{1,2}

¹Cancer 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; ²Department 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

1985

(where the second secon

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 mucosaderived 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 $\geq 5\%$ 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 highrisk 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

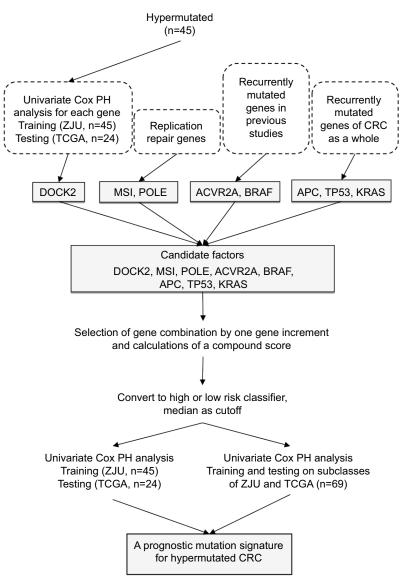


Figure I Flowchart of prognostic signatures generation and validation procedures. Abbreviations: TCGA, The Cancer Genome Atlas; ZJU, Zhejiang University; PH, proportional hazard; CRC, colorectal cancer.

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 (<u>https://www.python.org/</u>) and R 3.4.0 (<u>https://www.r-project.org/</u>).

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

	ZJU center	TCGA	P-value
	(n=45)	(n=24)	
Age, years, mean \pm 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	73.62±12.08	43.01±7.67	0.344
± SD			
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)	(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

Table I Characteristics of included patients

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

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).

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-thanmedian 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 highrisk patients shared a higher death risk, the results suggested

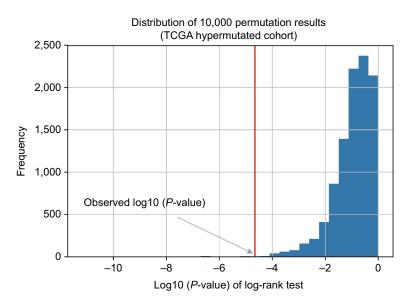


Figure 2 Distribution of 10,000 permutation results. Abbreviation: TCGA, The Cancer Genome Atlas.

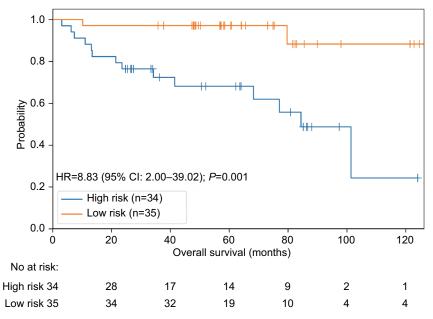


Figure 3 The pool data demonstrating effectiveness of the 4-gene signature.

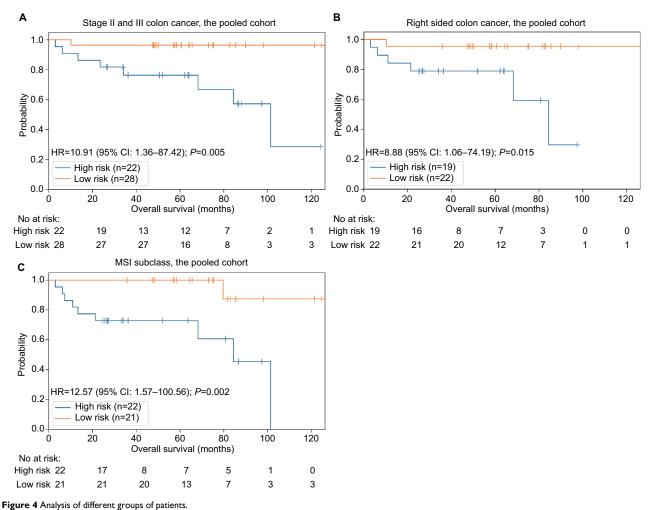
Table 2 Multivariate Co	ox regression anal	lysis of 4-gene s	ignature
-------------------------	--------------------	-------------------	----------

	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



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 al¹² 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

- 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.
- Akbani R, Akdemir KC, Aksoy BA, et al. Genomic classification of cutaneous melanoma. *Cell*. 2015;161(7):1681–1696.
- 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.
- Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature*. 2014;507(7492):315–322.
- Rehder DS, Nelson RW, Borges CR. Glycosylation status of vitamin D binding protein in cancer patients. *Protein Sci.* 2009;18(10):2036–2042.
- de La Chapelle A, Hampel H. Clinical relevance of microsatellite instability in colorectal cancer. J Clin Oncol. 2010;28(20):3380–3387.
- 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.
- Aaltonen LA, Peltomäki P, Leach FS, et al. Clues to the pathogenesis of familial colorectal cancer. *Science*. 1993;260(5109):812–816.
- Jiricny J. The multifaceted mismatch-repair system. Nat Rev Mol Cell Biol. 2006;7(5):335–346.
- Lee V, Murphy A, Le DT, Diaz LA. Mismatch repair deficiency and response to immune checkpoint blockade. *Oncologist*. 2016;21(10):1200–1211.
- 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.
- 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.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164
- Salipante SJ, Scroggins SM, Hampel HL, Turner EH, Pritchard CC. Microsatellite instability detection by next generation sequencing. *Clin Chem.* 2014;60(9):1192–1199.
- Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012;487(7407):330–337.
- 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.
- 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.
- 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.
- La D. PD-1 blockade in tumors with mismatch-repair deficiency. NEngl J Med. 2015;373(20):1979.
- 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.
- Dobbs K, Dominguez 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.
- 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.

- Hampton T. Gut microbes may shape response to cancer immunotherapy. JAMA. 2018;319(5):430–431.
- 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.
- 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.
- 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

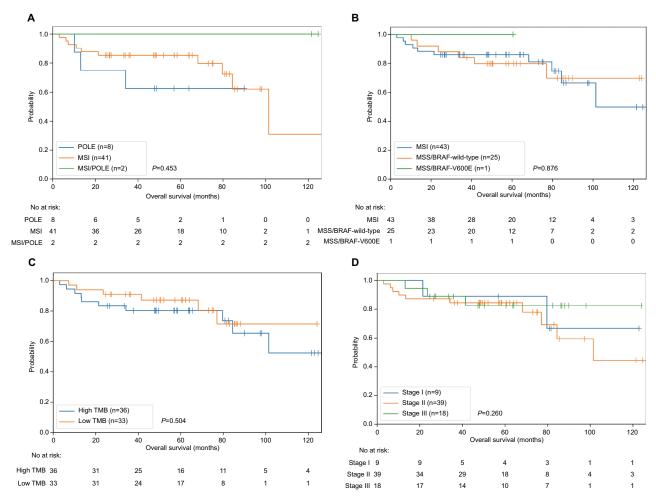


Figure SI Power of different prognosis biomarkers.

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.

3) CADRAN CUEMA EFFA FINIA FIFTURE LERA MAY AND CITCH PLAY OPERIT SCANA STRAID FINIAL CONCLARGE ECAS FTAGI TRIVING CONCLARGE EFFAA FUNDI FETU HAVET FINIAL CONCLARGE EFFAA FUNDI FETU HAVET FINIAL CAS CANAID STRAID FETU HAVET HAVET FINIAL FINIAL CAS CANAID STRAID FETU HAVET HAVET FINIAL FINIAL CAS CANAID STRAID FETU HAVET HAVET FINIAL FINIAL CAS CANAID STRAID FETU HAVET HAVET HAVET FINIAL CAS CANAID STRAID FETU HAVET HAVET HAVET FINIAL FINIAL CAS CANAID STRAID FETU HAVET HAVET HAVET HAVET FINIAL CAS CAS CANAID STRAID FETU HAVET HAVET HAVET CAS CAS CANAID STRAID FETU HAVET HAVET HAVET HAVET CAS CAS CANAID STRAID FETU HAVET HAVET HAVET HAVET CAS CAS CANAID STRAID FETU HAVET HAVET HAVET HAVET CAS CAS CANAID STRAID STRAID FETU HAVET HAVET HAVET CAS CAS CANAID STRAID STATES AND				0				0,1004		Ĺ	- 4100			0110	
ATMOS Confide ENAD FOLLING HANT FEAD	ASPM	+	CULEAS	EF3E	FA14		LLUKL	MSH2	NULLI	rrus	QSEKI	SCN9A	5 I 85 A6	I KAK2	ZFP36L2
ATM CS GRait Errat REXVIX HISTHABRI LGS MISHE NOTCH2 PMS1 ADD31 STMA10 TTGA10 ATPIA2 CACNAIB CSTAA EMOD2 FCUB HUA5 UG1 MTV31 NRA31 SMD31 SFRAD1 STM311 TTGA11 ATP1A2 CACNAIB CSTAA EMOD2 FCUB HUA5 UG1 MTV31 NRA32 POLD1 RAD31 STG712 TSPA01 ATTRA2 CACNAIB CSTAA EMOD2 FL MUC1 NTTA1 POLD1 RAD31 STC11 UNC31 STC31 STC91 UNC4 NTA1 STC13 STC91 UNC4 NTA1 STC13 STC91 UNC4 NTA1 STC14 UNC4 NTA1 STC14 UNC4 NTA1 STC14 UNC4 NTA1 UNC4 <td>ATAD5</td> <td></td> <td>COL6A6</td> <td>EIF4A2</td> <td>FBN2</td> <td>_</td> <td>LEKRI</td> <td>MSH3</td> <td>NOTCHI</td> <td>PLXNCI</td> <td>RABGGTB</td> <td>SEC63</td> <td>STAGI</td> <td>TRIM49B</td> <td>ZFPMI</td>	ATAD5		COL6A6	EIF4A2	FBN2	_	LEKRI	MSH3	NOTCHI	PLXNCI	RABGGTB	SEC63	STAGI	TRIM49B	ZFPMI
ATPLOA CATC1 EMO1 FCR1 HVER2 LIFR MTOR PMDS PMDS FTOR1 TSA10 TSA10 ATPLOA CACNUIB GE7RA EMMO2 FTU0A HUC3 MATPLOA CACNUIB GE7RA1 TSA11 TS450 ATPLOA CACNUIB GE7RA EUMO2 FTU0A HUC3 NUTA1 POLD1 RADD TSA11 TS450 TS470 TS47	ATM	ů	CRBI	ELF3	FBXW7		LGR5	MSH6	NOTCH2	PMSI	RAD21	SEMA3D	STKII	TRNTI	ZFPM2
ATPLA2 CACMAIL CSTAA EMOD2 FTUB LIG MTUA CACMAIL CSTAAL EMOD2 FTUB CTFAIL TFAIL ATPLA2 CACMAIL ESPR01 FEG HMCNL LGG HMCNL LGG HMCNL LGG FTPAIL TFPAIL ATPRA2 CAR EVERD1 FLG HMCNL LGG MTUD71 POLE MASA1 SICTAAL SUFU TCPAIL ATNU CCDC16 CTNNID BBB2 PCXA1 HSM LMFL NUC17	ATP10		CRTCI	ELMOI	FCRLI	HIVEP2	LIFR	MTOR	NPIPB5	PMS2	RAD50	SERPINFI	STK31	TSGA10	ZNF208
ATPR82 CAFPR8 CAFPR SCHOL FEG2 HHCNL UNA MUC12 NUTLL PCD2 ALGAPAI S1 STFG2 TFG23 ATPR02 CAFPR43 CBCH CTNNB1 REB3 FLI HNRNL UNA MUC12 NUDT11 POLE1 RASAI SLC1AAI SUFU UAC3 ATYNL CCDC146 CTNNB1 REB3 FLI MUC15 NUDT11 POLE1 RASAI SLC3AI SUFU UAC3 AXINU CCDC46 CTX3 REC1 FXE1 MUC4 ORD1 NUS1 SUC440 REP1 UAC4 AXINU CCDC64 CUX1 REC1 RVE1 MUC4 ORD1 REM1 FTG1 UAC4 AXINU CCDC64 CUX1 REC1 RVE1 MUC4 ORD1 REM1 UC172 UUC4 REM1 UC172 UUC172 <td></td> <td></td> <td></td> <td>ELMOD2</td> <td>FETUB</td> <td>HLA-B</td> <td>LIGI</td> <td>MTUSI</td> <td>NRAS</td> <td>POLDI</td> <td>RAD5 I AP2</td> <td>SHKBPI</td> <td>STONI- GTF2AIL</td> <td>TSHZ3</td> <td>ZNF233</td>				ELMOD2	FETUB	HLA-B	LIGI	MTUSI	NRAS	POLDI	RAD5 I AP2	SHKBPI	STONI- GTF2AIL	TSHZ3	ZNF233
ATPRA2 CBL CSM03 FECAM FLG2 HNRNLL UNMA MUC1 NUTS1 POLE RALT SLC1AI STCPI ULGA ATXNL CCDCI64 TTNND RBR2 FUL HNRNL UFH7 NUC1 NUD17 POLE RAST SLC1AI STVEL UBFA ATXNL CCDC164 TTNND RBR2 FLU HNRNL UCT NUD17 POLE RAST SLC1AI STVEL UBFA ATXNL CCDC164 TTNND RECC6 LUT RHRNL UND17 POLE RBR17 UBFA RUT RUSC REV UD17 RUSC REV UD17 RUSC REV UD17 RUSC			CSMDI	EP300	FLG	HMCNI	LIG4	MTUS2	NTHLI	POLD3	RALGAPAI	SI	STPG2	TSPO2	ZNF267
ATR CCDC141 CTNNB1 BBB2 FUI HRNR UPPI07 NUDT(1) POLE2 RASA1 STCP1 USCP1 MSA1 ATXN1 CCDC168 CTNUD3 BRB4 HYDN1 NUP107 NUP107 NUP107 S1C3A63 STT61 UB64 ATXN1 CCDC166 CUX1 RCC1 FACA1 RCC1 FACA1 S1C3A63 STT61 UB87 AXIN1 CCDC66 CUX1 RCC5 FC3 UC47 NUC4 OPC1 RAS71 S1C3A63 STT61 UB843 AXIN2 CCDC66 CUX1 RCC5 FC4 NUC4 OPC1 RAS71 UC732 BA201 CDC46 DXA1 FS17 IUR17 MC87 PAC4182 PUC1 RAS71 UC732 UA731 UC732 B3G716 DXA1 FS17 ITLN1 LRR3 MC16 PAC41 PUC6 RAS11 IUC32 UC732 UC732 UC732 UC732 UC732 UC732 UC732			CSMD3	EPCAM	FLG2	HNRNPL	LMNA	MUCI	NTSR2	POLE	RALY	SLC12A1	SUFU	TUSC3	ZNF285
ATXN1 CCDC168 CTNND2 RBM4 LKIF1 MUC16 NUP107 POLE4 RASSP2 SLC30A6 STTIE UBR5 ATTXN1 CCDC64 CTSA REXC1 FOXI LKIP MUC17 NUP30 PDL7 REM1 SLC30A6 STTIE UBR5 A XIN2 CCDC63 CTSA REXC1 FACI LG533 LRXC4 MUC4 PRD1 POLO REM1 SLC3016 TEC107 UBR5 A XIN2 CCDC73 DCAF6 RCC6 FKC1 LLXB1 MUC4 PAPR1 POTE REM12 SLC310 TEC17 UBR5 A XIN2 CCDC73 DCAF6 RCG1 FS17 LLXD1 LKRC4 MUC4 PAPR1 POTE SLC310 TEC17 UBR5 B XIN1 CCDC43 DCAF6 RCG1 FS17 LLXD3 MCC1 SLC411 PCF72 UNC52 US174 TEC712 UNC52 B XIN1 CD7 DNAH1 F18 CT12 RC117<		CCDC141	CTNNBI	ERBB2	FLI	HRNR	LPHN3	MUC12	NUDTII	POLE2	RASAI	SLCI3AI	SYCPI	UACA	ZNF334
AUTS2 CCDC64 EFACI EXCI1 <t< td=""><td></td><td></td><td></td><td>ERBB4</td><td>FOXAI</td><td>HSPA8</td><td>LRIFI</td><td>MUC16</td><td>NUP107</td><td>POLE4</td><td>RASSF2</td><td>SLC25A51</td><td>SYNEI</td><td>UBE4A</td><td>ZNF382</td></t<>				ERBB4	FOXAI	HSPA8	LRIFI	MUC16	NUP107	POLE4	RASSF2	SLC25A51	SYNEI	UBE4A	ZNF382
XINI CCDC66 CUXI RCC5 RCH2 IDO2 RP2 MUC4 ORD RP112 SLC4AIO TBEL1XI UGT322 AXIN2 CCDC637 DCAF6 RCC5 RG1 IGS17 LRC1 RVC1 SLC0B1 TRC113 RV SLC0B1 TBP UGT322 JZ ZCD DCA6 RCC1 FHR ILS1 LRR10 MCC6 RC1112 DCO16 TEC120 UMC132 S130T5 CCD04G DFB106 FT3 LRR10 MCC6 RR11 PADF1 PCTER RF712 UMC132 S130T5 CCM411 F136 GABG3 ITL10 LRR10 MC712 SLC06 SL1 SLC16 UMC33 BAG7 DVAH F138 GABG3 ITL1 LRR19 MC03 FRA1 FRA1 FEAD2 UNC34 BAR12 CDT1 DVAH FAM138 GC11 KND3 FRA1 RC632 SLA4 FEAD2 UNC34 BAR12			CTSA	ERCCI	FOXEI	HYDIN	LRPIB	MUCI7	NUP50	POLN	RBI	SLC30A8	SYT16	UBNI	ZNF417
AXIND CCDC73 DC466 RFCG RFC1 DC1FL RCACIO Table LXRI UGTA2 AXGND CCDC73 DC463 RFC1 LIS RACID PTC4 RCACI RCFL RCACI S1C RCACIB DC401 FRF71 UNC32 11 ZKGN CCMEI DC441 ERRO1 FRT RT UNC3 S1C <td< td=""><td>_</td><td>CCDC66</td><td>CUXI</td><td>ERCC5</td><td>FREM2</td><td>ID02</td><td>LRP2</td><td>MUC4</td><td>OPRDI</td><td>POLQ</td><td>RBM12</td><td>SLC4A I0</td><td>TBCID7</td><td>UBR5</td><td>ZNF469</td></td<>	_	CCDC66	CUXI	ERCC5	FREM2	ID02	LRP2	MUC4	OPRDI	POLQ	RBM12	SLC4A I0	TBCID7	UBR5	ZNF469
AZGF1 CCDG88 DCH3 FR/CH3 LII5 LR/D1 MUTH1 PABPC1 POTER RMX13 SLCOIBI TBP UGT32 21 B2M CONEI DDX6 EXO1 FSH2 ILR3 ITR1 UTC32 UGT32 21 B2M COMEI DDX6 EXO1 FSH2 ILN3 ITR1 IRR01 NUCC31 UGT32 UGT33 UGT32 UGT33 UGT3 UGT33 UGT3 UGT32 UGT33 UGT3 UGT32 UGT33			DCAF6	ERCC6	FRGI	IGSF3	LRRC4C	MUC6	ORI IHI 2	POSTN	RBMX	SLC5A10	Table I XRI	UGT2A2	ZNF479
11 B2M COLEI DDX6 EVOI FHR IRFS IRRVI MCGAS1 SLC06AII TCF12 UNCI32 20 B3GNT6 CO40LG DFEH108 FST ITN2 IRRV2 MCVI3 PARTI PTC1 SLTKKI TDRD3 UNC33 20 B3GNT6 CO40LG DFH10 F10 F17.2 IRRV2 MCV3 SLTKKI TDRD3 UNC34 TDRD3 UNC34 8431 CD7 DNH14 F138 GARG3 IV IRRV3 MCV3 SMTA TDRD3 UNC34 TFDD3 UNC34 8A11 CD7 DNH1 F138 GAT12 MCO3 MCO35 SMAD FRD3 KTP3 FRD3 KTP3 SMTA TDRD3 UNC33 UNC33 8A11 CD7 DNAH FAM138 GUT KCN13 MATC4 MATC3 FRD3 FRC3 UTC33 UTC33 UTC33 UTC33 UTC33 UTC33 UTC33 UTC33 UTC33				ERICH3	FRY	ILI5	LRRDI	МИТҮН	PABPCI	POTEC	RBMXL3	SLCOIBI	TBP	UGT3A2	ZNF521
210 B3GNT6 CP440.C DEFB1088 EYS FITU2 URK2 MYCN PARF1 FOTEH REX1 TDRD3 UNC5CL B3GNT7 CD58 DNAH11 F10 F5TL3 ITRN1 LRN3 MYO16 PARF4 FOTE SLTRK4 TDRD3 UNC5CL B3GNT7 CD58 DNAH11 F10 F5TL3 ITRN3 MYO36 PBYN1 PRM9 RG23 SLTR44 TDRD3 UNC52L BA1P212 CD72 DNAH3 FAN1338 GC11 KCN13 MA12 MYO95 PBN1 PRM9 RG23 SNAD3 TG20 UT733 BA1P212 CDF412 DNAH7 FAN1338 G17P23 MA22 MA12 PRM9 RM1 SG23 SNAD4 TGFB2 VN182 BA1 CDH1 DNAH7 FAN1388 G17P23 MA22 MA22 RM183 RN13 SNTG2 TGFB2 VN182 BA1 EC19 DNAH7 FAN138 G17P2 RCN12		CCNEI	9XQQ	EXOI	FSHR	IRF5	LRRIQI	MYCBP2	PAFAH I B2	POTEE	REV3L	SLCO6AI	TCF7L2	UNCI3C	ZNF569
B3GNT9 CD44 DNAH11 F10 FTL5 ITPR1 LRRN3 MYO16 PARP4 POTE SLITK4 TDRD6 UNC93A BAGE3 CD7 DNAH14 F13B GABRG3 IVL LRRTN3 MYO3B PXX1 PPMIE RG52 SLX4 TED2 UP12A BAGE3 CD7 DNAH14 F13B GABRC3 IVL LRRTN3 MYO3B PXX1 PPMIE RG522 SLX4 TED2 UP12A BAJ7 CD7 DNAH16 FAN133B GLT9D2 KON3 MAC PKR1 RND3 SNTG2 TED2 UT12 BAX CDKNDA DNAH7 FAN133B GLT9D2 KCN2 MAALAD PFN16 RG52 SNTG3 TGFI VT133 B BAX CDKNDA DNAH7 FAN133 GLT9D2 KCN12 MAALAD PCHBIS PKR1 R1M3 SNTG3 TGFI VT143 B ACI CHEC3 DNA/7 MAPX1 NBEA PCD			DEFB108B	EYS	FSIP2	ITLN2	LRRK2	MYCN	PARPI	POTEH	RFCI	SLITRKI	TDRD3	UNC5CL	ZNF600
BAGE3CD58DNAH14F13BCABRG3N/LLRTM3MYO3BPAX1PPMIERGS22SLX4TED22USH2ABAJ32CD7DNAH3F5GAUNT12JAX2MAGEB3MYO9APBRMIPRDM9RHCGSMAD2TELO2UTP23BAAP212CD7DNAH6FAM13BGCTIKCNI3MAL2DPYPEB1PRKA1RHN2SMAD3TGUT122BAP1CDN12DNAH6FAM13BGLT92KCN13MAL2DPYPE16PCDH9PRKA1RM32TGFR2VN1721BAP1CDN12DNAH6FAM13BGLT92KCN13MAL2DPCDH9PKKA1RM32TGFR2VN1822BC19CFM23DNAH8FAM13BGLT92KCN12MAD2K1NALCNPCGF6PSS3RM1SOR82TLR4VT11A2BC19LCHK2DNAH2FAM18BGLT2KON12MAP2K1NBF15PCS7RNF3SO22TLR7V8SC118BUY1CHC11DOCZFAM162GN12KNA04MAP2K1NBF15PCS7RNF3SO22TLR7V8SC118BUY1CHU1DOCZFAM162GN12KNA04NBF15PCS7RNF3SO22TLR7V8SC118BUY1CHU1DOCZFAN162GN12NBF15PCS7RNF3PCS7FNM13KN958BUY1CHU1DOCZFAN16NBF15PCS7RNF3PCS7 </td <td>~</td> <td></td> <td>DNAHII</td> <td>FI0</td> <td>FSTL5</td> <td>ITPRI</td> <td>LRRN3</td> <td>MYO16</td> <td>PARP4</td> <td>POTEJ</td> <td>RGSI7</td> <td>SLITRK4</td> <td>TDRD6</td> <td>UNC93A</td> <td>ZNF665</td>	~		DNAHII	FI0	FSTL5	ITPRI	LRRN3	MYO16	PARP4	POTEJ	RGSI7	SLITRK4	TDRD6	UNC93A	ZNF665
Bal3 CD7 DNH3 F5 GALNT12 JAC2 MAGE3 MYO9A BRM1 PCDM9 RHCG SMAD2 TELO2 UT733 BAP712 CC742FI DNH4 FAM133B GT1 KCNJ5 MAGE12 MYOCD PCBH9 PRKAR RIM22 RHD3 TG UT733 BAP712 CC742FI DNAH6 FAM133B G13B2 KCNN3 MAL2 MYOCD PCBH9 PRKAR RIM32 R11 SPR33 TGG V733 1 BXX CDKN2A DNAH8 GN12 KCN13 MAD24 NAL20 PCDH9 PRKAR R1N3 SNTG2 TGGF V718 2 BC19 CFFA58 DNAH8 GN12 KN66A MAD24 NAL20 PCDH8 PKRAR R1N3 SNTG2 TGGF V7162 2 BC14 CHE73 KN66A MAD24 NBF15 PC10 PSC33 RM1 SNTG2 TLR1 V718 V7163 Y7183 Y7183 <td></td> <td></td> <td>DNAH14</td> <td>FI3B</td> <td>GABRG3</td> <td>IVL</td> <td>LRRTM3</td> <td>MYO3B</td> <td>PAXI</td> <td>PPMIE</td> <td>RGS22</td> <td>SLX4</td> <td>TEAD2</td> <td>NSH2A</td> <td>ZNF671</td>			DNAH14	FI3B	GABRG3	IVL	LRRTM3	MYO3B	PAXI	PPMIE	RGS22	SLX4	TEAD2	NSH2A	ZNF671
BAIAPL2CDC42EP1DNAH5FAM123BGGT1KCNU3MAGEL2MYCDCPCBH9PRKA1RMD3TGUT32BBAP1CDH1DNAH6FAN133BGU33KCNU3MAL2MYPOPPCDH9PKAN1RMD3TGFB72VN1221BAXCDKN2ADNAH6FAN133BGU33KCNU3MAL2MAD2PCDH91PKAN1RM32SMD4TGFB72VN1232BCL91CFM73DNAH6FAN134BGLTP2KCN12MAP2K1NBEAPCCH3PKAN1RN13SOR22TL44VT132BCL91CHE72DNAG24FAN194BGNT72KDM6AMAP2K1NBEAPCC10PCG5SN32RN11SOR52TL44VT132BLMCLUL1DOCK1FAN192GN12KIN43MAP2K1NBF13PC10PK72PK871VB87172BLMCLUL1DOCK2FAN194GN172KN72BMAP2K1NBF13PH72PT01PR671PK87172BLMCUUL1DOCK2FAN192GOLGB1KIRELMAP2K1NBF13PH72PT01PA023PM8713XR723BMF1CUUL1DOCK2FAN19GOLGB1KIRELMAP2K1NBF13PH72PT01PM133XR723BMF1CMV3DP10FAN10GOLGB1KIRT2BMAP2K1NBF13PH72PT01PT012PM6713XR793BMF2CMD1DC		CD7	DNAH3	F5	GALNT12	JAK2	MAGEB3	МҮО9А	PBRMI	PRDM9	RHCG	SMAD2	TELO2	UTP23	ZNF679
BeP1CDH1DNAH6FAM135BGL13KCNN3MAL2MYPOPPCDH9PKKAA1RIM32SMD4TGFBR2VNIR21BXXCDKN2ADNAH7FAM133BGL7BD2KCN22MANSC4NALAD2PCDHB15PKKIRRIN3SNTG2TGF11VF337B2BCL9CFAP58DNAH8FAM144GL7D2KCN12MAP2K1NBEAPCLOPSCARNH3SON22TLR4VT1IA2BCL9LCHEK2DNAJC4FAM194BGNT2KDM6AMAP2K1NBEAPCLOPSCARNH3SON2TLR7WBSC1112BUM-CLUL1DOCK10FAM1912GOLG41K1F4BMAP3K1NBF15PCC5PTCH1RP1RPM173XRP2112BUMCUUL1DOCK2FAN1GOLG41K1F4BMAP3K1NBF12PTCH1RP1RP173ZRP24MN173ZRP2413BIM1CUUL1DOCK2FAN1GOLG41K1F4BMAP3K1NBF2P1CH1RP1RP173ZRP24MN173ZRP2414CUUL1DOCK2FAN1GOLG41K1F4BMAP281NBF12P1CH1RP1RP173ZRP24ZRP3121TM6113ZRP23ZRP3121ZMP173ZRP24ZRP3121ZMP173ZRP24ZRP3121ZMP173ZRP24ZRP3121ZMP173ZRP24ZRP3121ZMP173ZRP24ZMP173ZRP24ZMP173ZRP24ZMP173ZRP24ZMP173ZRP23	BAIAP2		DNAH5	FAMI 23B	GGTI	KCNJ5	MAGEL2	MYOCD	PCBPI	PREX2	RHOA	SMAD3	TG	UTS2B	ZNF727
1BXXCDKNJADNH7FAMI53BGLTBD2KCNS2MANSC4NALAD2PCDHB15PKRIRR1N3SNTG2TGF1VP37B2BC19CFAP58DNH8FAM184GLTPD2KCNT2MAPXK1NBEAPCLOPSCARNH3SORES2TLR4VTI1A2BC19LCHER2DNJG24FAM194BGNT2KDM6AMAPXK1NBEAPCLOPSCARNH3SOX22TLR4VT11A8BUM-CLEC18BDOCK10FAM71E2GNG12K1Aa0408MAPXK1NBF15PCLOPSCARNH3SOX22TLR4VT11A12BUM-CLUL1DOCK2FAM1GNG12K1RALMAPXK1NBF15PF17PF17NB571VBSC1112BUM1CUUL1DOCK2FAM1GOLG4K1RALMAPXK1NBF13PHF2PF17RP143SOX2TLR4VD11A13BMR1ACMT05DPP101FAM12GNG13K1RALMAP311NB72PHF2PF201RP113RP113Z1731214CUUL1DOCK2FAN1GOLG4K1RALMAPX19NBF13NF22PHF2RP123SP766TNN136Z1773Z177315BMR12CNN10DEV102FAN1KRTAP45M167NK221PK333PHF2RP213RP123ZP766TNN136Z1773115BMR22CNN10DV14FAN1NK221NK221RP133PT761RP223 <td< td=""><td></td><td>CDHI</td><td>DNAH6</td><td>FAMI 35B</td><td>GLI3</td><td>KCNN3</td><td>MAL2</td><td>МҮРОР</td><td>PCDH9</td><td>PRKAAI</td><td>RIMS2</td><td>SMAD4</td><td>TGFBR2</td><td>VNIR2</td><td>ZNF728</td></td<>		CDHI	DNAH6	FAMI 35B	GLI3	KCNN3	MAL2	МҮРОР	PCDH9	PRKAAI	RIMS2	SMAD4	TGFBR2	VNIR2	ZNF728
2 BCJ9 CFAP58 DNAH8 FAMI946 GTFD2 KCNT2 MAP18 NALCN PCGF6 PSS33 RMI SORB22 TLR4 VTIA 8 BCJ9L CHEK2 DNAJC34 FAMI948 GNAT2 KDM6A MAP2K1 NBF15 PCLO PSC4 RMF3 SOX2 TLR7 WBSCR17 8 BIVM CLEC188 DOCK10 FAM1E2 GNG12 KIA0408 MAP2K1 NBF15 PSC4 PSC4 SOX2 TLR7 WBSCR17 8 BIVM CLEC188 DOCK10 FAM1E2 GNG12 KIA0408 MAP2K1 NBF15 PSC4 PSC4 PSC4 PSC4 PSC4 PSC4 PSC11 PSC4 PSC11 PSC11 PSC11 PSC11 PSC11 PSC4 PSC4 PSC4 PSC4 PSC4 PSC4 PSC11 PS	=	CDKN2A	DNAH7	FAMI 53B	22	KCNS2	MANSC4	NAALAD2		PRKRIR	RIN3	SNTG2	TGIFI	VPS37B	ZNF729
BCL9LCHEK2DNAJC24FAM194BGNAT2KDM6AMAP2K1NBF15PCL0PSCARNF3SOX2TLR7WBSCR17BVMCLEC18BDOCK10FAM71E2GNG12KIA0408MAP2K4NBP15PCSK7PSG5ROBO2SOX9TM651WDR11ERCC5EMCLULIDOCK10FAM71E2GNG12KIA0408MAP2K4NBP15PF15PF12SOX9TM651WDR1112BMCLULIDOCK2FAN1GOLGA4KIFBLMAP71NBP73PH22PTCH1RP12SPAC11TMEM173XIRP213BMPR.2CNB1DPY19L2FANC3GP10KMT2BMAPK1NEBPF202PTPLARP44SPAT31D1TMEM184XR9914CNTNIDSELFANC3GP112KMT2BMAPK1NEBPF202PTPLARP122SPD766TMN136XR02115BMPR.2CNTNIDSELFANC3GP112KR72P16MN12RF212SPD766TMN136XR02116BMPR.2CNTNIBDSELFANC3GP112KR72P16MN12RF222SPD766TMN136XR02116BRA7CNTNIBDSELFANC3GP112KR72P16MN12RF222SPD766TNN136ZEC13ZEC1317BRA7CNTNIBDSELFANC3KTAP16MR27NK271PK2212PTP12SPD766TNN136ZEC2318RCA1CNTN6D		CFAP58	DNAH8	FAM184A	GLTPD2	KCNT2	MAPIB	NALCN	PCGF6	PRSS35	RMII	SORBS2	TLR4	VTIIA	ZNF735
BIVM-CLECIBBDOCK10FAM7IE2GNG12KIAA0408MAP2K4NBPF15PCK7PGG5ROBO2SOX9TM65F1WDR11ERCC5ERCC4KTCMAP1MAP3K19NBPF3PHF2PTCH1RP1SPARCL1TMEM173XIRP22BLMCLULIDOCK2FANIGOLGA4KIF4BMAP3K19NBPF3PHF2PTCH1RP1SPARCL1TMEM173XIRP230BMPR1ACMY35DPP10FANC2GOLGB1KIRELMAP81P1NEHPHF2PTCH1RP1SPARCL1TMEM173XIRP231BMPR2CNBD1DPY19L2FANC2GPN12KRT2BMAP88IP1NEF12PHF3PTCA1RP1RP2SPD766TMP133XPC31BRAFCNTNIDSELFANC3GPN12KRT2P16MF72PK331PTP14RP203SPD166TMP133XPC31BRAFCNTNADSELFANC3GPN13KRT4P4.5MI67NIX31PK231PTP14RP203SP194TOP3BZB67331BRACCNTNAPSDYNC2H1FANC4GR112KRT4P4.5MI67NIX31PK232PTP14RSP03SP194ZD673ZB67331BRP1CUL1AIEBNI1FATGR12MH28MI67NIX31PLC1PTP14RSP03SP194ZD673ZB67332BRP1CUL1AIEBNI1FATGR12MH28MI673MK167NIX421 </td <td></td> <td>CHEK2</td> <td>DNAJC24</td> <td>FAMI 94B</td> <td>GNAT2</td> <td>KDM6A</td> <td>MAP2K I</td> <td>NBEA</td> <td>PCLO</td> <td>PSCA</td> <td>RNF43</td> <td>SOX2</td> <td>TLR7</td> <td>WBSCR17</td> <td>ZNF750</td>		CHEK2	DNAJC24	FAMI 94B	GNAT2	KDM6A	MAP2K I	NBEA	PCLO	PSCA	RNF43	SOX2	TLR7	WBSCR17	ZNF750
ERCC5ERCC5MMM<		-	DOCK10	FAM71E2		KIAA0408	MAP2K4	NBPFI 5	PCSK7	PSG5	ROBO2	6XOS	TM6SFI	WDRII	ZNF761
12BLMCLULIDOCK2FANIGOLGA4KIF4BMAP3K19NBF3PHF2PTCH1RP1SPACL1TMEM173XIRP280ABMPR1ACMYA5DPP10FANCAGOLGB1KIRELMAPK1NEBPEZO2PTENRP4SPATA31D1TMEM184AXK9980ABMPR2CNBD1DPY19L2FANCD2GPN1KMT2BMAPK8IP1NEHPK3CAPTPLARP122SPDYE6TMPN3KXFC2138BRAFCNTN1DSELFANC1GPR12KR3MPR2NINPIK3R3PTPK1RP220SPDYE6TNN3KXFC2137BRCA1CNTN6DUX4FANC1GPR12KR3MFPNINPIK3R3PTPK1RS200SPC30TNN3KXFC2131BRCA1CNTN6DUX4FANC1GPR14KR7P10.6MFPNINPIK3R3PTPK1RS202SPI1TOP2BZBE7931BRP1CU11AIEBLNIFAT1GR12KR7P4.5ML61NLX2-1PLC11PTPK1RSP03SP11TOP2BZBE7232BRP1CU11AIEBLNIFAT2GRN2AKTAP5.5ML11NLR914PLC11PTPK1RSP03SP11TOP2BZBE7332BRP1CU11AIEBLNIFAT3GRN2AKTAP5.5ML11NLR914PLC11PTPK1RSP03SP11TOP3BZG31432C17orf70CU12AIEDNRBFAT3GRN2AMN11 <td>ERCC5</td> <td></td>	ERCC5														
80ABMPR.IACMYASDPP10FANCAGOLGBIKIRELMAPKNEBPIEZO2PTENRPA4SPATA3ID1TMENI84AXKR938BMPR.2CNBD1DPY1912FANCD2GPN1KMT2BMAPK8IPINEHPK3CAPTPLARPL22SPDY66TMPRSS13XPC38BRAFCNTNIDSELFANCDGPN12KRA7MDM2NFE212PK3R1PTPLARPL22SPDY66TNN13KXRCC137BRCA1CNTN6DUX4FANC1GPR149KRTAPI0-6MFRNINPK3R3PTPKRSP02SPDRTOP2BZBE0931BRC1CNTN65DYNC2HIFANC1GRP149KRTAP10-6MFRNINPK3R3PTPKRSP02SP1RTOP2BZBE0932BRP1COL1AIEBLNIFAT1GRA2KTAP5-5MLHINLRP14PLCE1PTPK1RSP03SP11TOP2BZDB7232BRP1COL1AIEBLNIFAT2GRN2AKTAP5-5MLHINLRP14PLCE1PTPK1RSP03SP11TOP2BZD67332BLP1COL1AIEBLNIFAT3GRN2AKTAP5-5MLH1NLRP14PLCE1PTHNISC03SP11TOP2BZFC3H112C176rf70COL12AIEDNRBFAT3GRN2AKTAP5-5MLH1NLRP14PLCE1PTHNISC33SP11TOP3BZFC3H112C176rf70COL12AIEDNRBFAT3G		CLULI	DOCK2	FANI	A4	KIF4B	MAP3K19	NBPF3	PHF2	РТСНІ	RPI	SPARCLI	TMEM173	XIRP2	ZNF787
BMPR2 CNBD1 DPY19L2 FANCD2 GPN1 KMT2B MAPK8IPI NEH PIF3CA FPLA RPL22 SPDYE6 TMPRSS13 XPC 3B BRAF CNTNI DSEL FANCE GPR12 KR3 MDM2 NFE212 PIK3R1 FPLA SPDYE6 TNNI3K XRCc1 3F BRCAI CNTN6 DUX4 FANCI GPR149 KRTAPI0-6 MFP NIN PIF3 PTPK RSP02 SPDR TOP2B ZBE09 715 BRCA2 CNTNAP5 DYNC2HI FANCI GRM14 NLA2-1 PKD32 PTPK RSP02 SPDR TOP2B ZBE09 715 BRCA2 CNTNAP5 DYNC2HI FANCM GRM14 NLC91 PKD212 PTPR7 RSP03 SPDR TOP2B ZDB72 715 FOL1AI EBLNI FAT1 GRM2 MLH3 NLC11 PLC11 PTPR1 RSP03 SPTAI ZDA3 ZDA4 712 CT3rdf70 <td></td> <td></td> <td>DPP10</td> <td>FANCA</td> <td>ВІ</td> <td>KIRREL</td> <td>MAPKI</td> <td>NEB</td> <td>PIEZO2</td> <td>PTEN</td> <td>RPA4</td> <td>SPATA31D1</td> <td></td> <td>XKR9</td> <td>ZNF804A</td>			DPP10	FANCA	ВІ	KIRREL	MAPKI	NEB	PIEZO2	PTEN	RPA4	SPATA31D1		XKR9	ZNF804A
3B BRAF CNTNI DSEL FANCE GPRI12 KRAS MDM2 NFE212 PIK3R1 PTPNI1 RPS20 SPG20 TNN13K XRCC1 3F BRCAI CNTN6 DUX4 FANCI GPR149 KRTAPI0-6 MFR NIN PK3R3 PTPRK RS202 SPDA TOP2B ZBED9 7I BRCA2 CNTNAF5 DYNC2HI FANCM GREMI KRTAPI-5 MKI67 NKX2-1 PKD212 PTPRK RSP03 SPPI TOP2B ZBED9 312 BRIPI COLI1AI EBLNI FATI GRM3 KTAP5-5 MLHI NLRP14 PLCEI PTPRZ RYR2 SPPI TOP3B ZPC3HI 12 C176rf70 COL12AI EDNB FAT2 GRN2A KTAP5-5 MLH3 NME8 PLCL1 PTPRZ RYR2 SPTA ZPC3HI ZFC3HI 12 C176rf70 COL12AI EDNB FAT3 GRN2A KTAP5-5 MNL4 NLC1 <t< td=""><td>BMPR2</td><td></td><td>DPY 19L2</td><td>FANCD2</td><td>GPNI</td><td>KMT2B</td><td>MAPK8IP1</td><td>NEFH</td><td>PIK3CA</td><td>PTPLA</td><td>RPL22</td><td>SPDYE6</td><td>TMPRSS13</td><td>XPC</td><td>ZNF804B</td></t<>	BMPR2		DPY 19L2	FANCD2	GPNI	KMT2B	MAPK8IP1	NEFH	PIK3CA	PTPLA	RPL22	SPDYE6	TMPRSS13	XPC	ZNF804B
3F BRCAI CNTN6 DUX4 FANCI GPR149 KRTAPI0-6 MERP NIN PIK33 PTPRK RSPO2 SPIDR TOP2B ZBED9 15 BRCA2 CNTNAP5 DYNC2HI FANCM GRM1 KRTAP4-5 MKl67 NX2-1 PKD212 PTPRT RSPO3 SPI1 TOP3B ZDBF2 32 BRIPI COLI1AI EBLNI FAT1 GRIA2 KRTAP5-5 MLH1 NLRP14 PLCE1 PTPRZ1 RYR2 SPTA1 TOP3B ZFC3HI 12 C170-rf70 COL12AI EDNB FAT2 GRIN2A KTNI MH3 NME8 PLCLI PYHINI SAC3 SPTA1 ZFC3HI ZFC3HI 12 C170-rf70 COL12AI EDNB FAT3 GRIN2A KTNI MLH3 NME8 PLCLI PYHINI SAC3 SPT1 TP33 ZFHX4 12 C170-rf70 COL12AI ETHB FAT3 GRIN2A MN1 NM1 PLCZ1 <t< td=""><td></td><td>CNTNI</td><td>DSEL</td><td>FANCE</td><td>GPR112</td><td>KRAS</td><td>MDM2</td><td>NFE2L2</td><td>PIK3R I</td><td>PTPNII</td><td>RPS20</td><td>SPG20</td><td>TNNI3K</td><td>XRCCI</td><td>ZNF831</td></t<>		CNTNI	DSEL	FANCE	GPR112	KRAS	MDM2	NFE2L2	PIK3R I	PTPNII	RPS20	SPG20	TNNI3K	XRCCI	ZNF831
¹⁵ BRCA2 CNTNAP5 DYNC2HI FANCM GREMI KRTAP4-5 MKI67 NKX2-1 PKD2L2 PTPRT RSPO3 SPPI TOP3B ³² BRIPI COLI1AI EBLNI FATI GRA2 KRTAP5-5 MLHI NLRP14 PLCEI PTPRZI RYR2 SPTAI TOPAZI ¹² C17orf70 COL12AI EDNRB FAT2 GRIN2A KTNI MLH3 NME8 PLCLI PYHINI SACS SPZI TP53 ¹³ CIORF74 COL18AI EFHB FAT3 HDGFRP2 LCA5 MNXI NMI PLCZI QRFPR SCNIA SSBPI TPM3			DUX4	FANCI	GPR149		MFRP	NIN	PIK3R3	PTPRK	RSPO2	SPIDR	TOP2B	6D38Z	ZNF850
32 BRIPI COLIIAI EBLNI FATI GRIA2 KRTAP5-5 MLHI NLRPI4 PLCEI PTPRZI RYR2 SPTAI TOPAZI 12 CI7orf70 COL12AI EDNRB FAT2 GRIN2A KTNI MLH3 NME8 PLCLI PYHINI SACS SPZI TP53 CIORF74 COLI8AI EFHB FAT3 HDGFRP2 LCA5 MNXI NMI PLCZI QRFPR SCNIA SSBPI TPM3				FANCM	GREMI	KRTAP4-5	MK167	NKX2-I	PKD2L2	PTPRT	RSPO3	IddS	TOP3B	ZDBF2	
12 C170-F70 COL12AI EDNRB FAT2 GRIN2A KTNI MLH3 NME8 PLCLI PYHINI SACS SPZI TP53 C10RF74 COL18AI EFHB FAT3 HDGFRP2 LCA5 MNXI NMI PLCZI QRFPR SCNIA SSBPI TPM3		COLIIAI	EBLNI	FATI	GRIA2	KRTAP5-5	MLHI	NLRP14	PLCEI	PTPRZI	RYR2	SPTAI	TOPAZI	ZFC3H1	
CIORF74 COLI8AI EFHB FAT3 HDGFRP2 LCA5 MNXI NMI PLCZI QRFPR SCNIA SSBPI TPM3	12		EDNRB	FAT2	GRIN2A	KTNI	MLH3	NME8	PLCLI	ΡΥΗΙΝΙ	SACS	SPZI	TP53	ZFHX4	
			EFHB	FAT3		LCA5	MNXI	IΨN	PLCZI	QRFPR	SCNIA	SSBPI	TPM3	ZFP28	

Table S2 Baseline values of all initial selected patient
--

	ZJU center (n=45)	TCGA (n=63)	P-value
Age, years, mean \pm 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
l	4 (8.89)	11 (17.46)	
II	27 (60.00)	35 (55.55)	
III	13 (28.89)	13 (20.63)	
IV	I (2.22)	2 (3.17)	
Location, n (%)			0.003
Right-side colon	22 (48.89)	50 (79.37)	
Left-side colon	4 (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 (%	6) 18 (54.55)	14 (40.00)	0.457

Cancer Management and Research

Publish your work in this journal

Cancer Management and Research is an international, peer-reviewed open access journal focusing on cancer research and the optimal use of preventative and integrated treatment interventions to achieve improved outcomes, enhanced survival and quality of life for the cancer patient. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit http://www.dovepress.com/testimonials.php to read real quotes from published authors.

Submit your manuscript here: https://www.dovepress.com/cancer-management-and-research-journal

Dovepress