Characterization and clinical evaluation of microsatellite instability and loss of heterozygosity within tumor-related genes in colorectal cancer

Xueyun Huo (huoxueyun@126.com)
Capital Medical University https://orcid.org/0000-0002-1646-0963

Dandan Feng
Capital Medical University

Shuangyue Zhang
Chinese Academy of Medical Sciences Institute of Laboratory Animal Sciences

Zhenkun Li
Capital Medical University

Xiaohong Li
Capital Medical University

Changlong Li
Capital Medical University

Meng Guo
Capital Medical University

Jin Wang
Capital Medical University Affiliated Beijing Friendship Hospital

Zhongtao Zhang
Capital Medical University Affiliated Beijing Friendship Hospital

Qingxian Lu
University of Louisville

Xiaoyan Du
Capital Medical University

Zhigang Bai
Capital Medical University Affiliated Beijing Friendship Hospital

Zhenwen Chen
Capital Medical University

Research article

Keywords: Microsatellite instability, Loss of heterozygosity, Colorectal cancer, Prognosis, Tumor- related genes

Posted Date: October 18th, 2019

DOI: https://doi.org/10.21203/rs.2.16206/v1

Version of Record: A version of this preprint was published at BMC Medical Genomics on September 25th, 2021. See the published version at https://doi.org/10.1186/s12920-021-01051-5.

Abstract

Background: Microsatellite instability (MSI) has been known as a biomarker for better outcome in colorectal cancer (CRC). However, the conclusion is controversy. In addition, MS can also be a useful marker for loss of heterozygosity (LOH) of genes but it has not been well studied yet. Here, aimed to clarify the predictive value of MSI/LOH within tumor-related genes in CRCs, we detected MSI/LOH of MSs in tumor-related genes and the Bethesda (B5) panel and further analyzed the relationship between MSI/LOH and clinical features or outcomes. Results: As expected, the MSI rate of B5 loci were all very high, suggesting that B5 panel criterion is powerful for MSI status determining of CRCs. Interestingly, MSI/LOH of 2 loci in the B5 panel and 12 loci in tumor-related genes were associated with poorer outcome while MSI/LOH of B5 panel was failed to predict outcomes of CRCs. MSI of BAT25, MSI/LOH of BAT26 and MSI of B5 panel showed closer relationship with mucinous carcinoma. In addition, LOH-H of B5 panel associated with more lymphatic metastasis. Conclusions: In summary, MSI/LOH of certain loci or whole panel of B5 were related to the clinical features, and several loci within tumor-related genes showed a prognostic value in outcomes of CRCs.

Background

Colorectal cancer (CRC) is one of the most common cancers in the world [1]. In China, CRC is one of the five leading causes of the cancer-related death, and one of the most two common cancers in both men and women according to the data from the National Central Cancer Registry of China (NCCR) [2]. CRC often shows a significant heterogeneity in both prognosis and chemotherapeutic response, despite similar histologic features and tumor stage [3].

The carcinogenesis of colorectal cancer may follow various pathways: chromosomal instability (CIN) and microsatellite instability (MSI). CIN is detected in up to 80% of CRCs and may be accompanied by the loss of heterozygosity (LOH) and chromosomal rearrangement [4]. MSI is known as a hypermutable phenotype resulted from the loss or dysfunction of mismatch repair (MMR) system which detects and repairs mismatches that occur during DNA replication [5]. It has been reported that approximately 15% of CRCs carry MSI in the Western countries [6], whereas about 14.3% of CRCs in China were identified as MSI-positive [7]. Thus, it is well accepted that MSI status is associated with CRC. The Bethesda panel (B5 panel) has been recommended by American National Cancer Institute for the MSI test [8,9]. According to MSI determination by B5 panel, CRCs with MSI exhibit distinctive features, including a tendency to arise in the proximal colon, lymphocytic infiltrate, and a poorly differentiated, mucinous or signet ring appearance, and they have a better prognosis than those without MSI, since their different chemotherapeutics effects [6,10]. A recent study showed that MSI detected by the B5 panel seemed to be a good biomarker for better outcome and longer disease-free survival [11]. However, the conclusion is controversial [12,13]. On the other hand, LOH analysis identifies allelic imbalances, which reflect gains and losses of chromosomal regions. It is known that severity of LOH differs among tumors. Some tumors have LOH at many loci in various chromosomes, whereas others have less frequent LOH [14]. A few studies indicated that the different LOH mutation frequency of loci might be related to the biologic

behavior of CRC. But the conclusion is still disputable. We analyzed the data of 440 CRC patients in China by the B5 panel, as expected, they are quite sensitive to MSI, but the MSI and LOH status of B5 panel have no prognosis and predicting significance to the CRC. Noteworthy, MSI/LOH mutations of the loci or panels recombination by loci in B5 panel and tumor-related genes correlated with the clinical features of CRC, and could be used for the treatment of individual patients. Thus, development of novel robust biomarker for CRC population may be benefit for prognosis and prediction of the chemotherapeutic responses.

More new panels of MSI test have been developed recently. Ronald J Hause *et al.* showed that MSI status was scattered across the human genome in 18 cancer types including CRC, and revealed that MSI showed prognosis significance [15]. But the amount of MS loci was too large to be used in practice. It is reported that several critical genes, such as *TP53*, *APC* inactivation, *KRAS* and *BRAF* mutation, *MYC* amplification, and other tumor related genes, were altered in CRC. These molecular events lead to dysregulation of cell growth, proliferation, survival, apoptosis, and invasion, which are involved in tumorigenesis and development [16]. The prevalence and clinical significances of *KRAS*, *BRAF*, *NRAS*, and *PIK3CA* mutations have been documented in the Chinese CRC population [17]. However the data are quite limited, and the MSI status within these tumor-related genes has not been fully explored yet. Therefore, we hypothesize that MSI/LOH within these genes would be appropriate markers for the clinical pathological staging, prognosis and predicting response to chemotherapy in CRC patients. In the present study, we investigated the MSI/LOH profile in the tumor-relate genes, as well as to illuminate the relationship between MSI/LOH status and the clinicopathological characteristics in Chinese CRC patients.

Results

The MSI/LOH status in tumor-related genes

In order to determine the MSI/LOH status in tumor-related genes, we selected 61 MS loci (Table S2) based on the optimization of the appropriate PCR conditions. Among the 61 MS loci, 53 were located in introns, 1 found in an exon region, 5 loci located in the non-coding regions (referred to the area except for 3’-untranslated regions (UTR), 5’-UTR, exon and intron of gene), and the rest of 2 loci in the 3'-UTR. Based on the STR scanning, 217 MSI events in 18 genes were detected in 48 tumor specimens, representing approximately 4.52 mutations per tumor. In addition, 909 LOH events in 18 genes were detected in 147 tumors representing approximately 6.18 mutations per tumor. Of the 256 cases, 18.8% harbored one or more than one MSI events, 57.4% had one or more than one LOH events and no mutation (MSI and LOH) were found in 23.8% cases. For the 61 loci, 70.49 % (43/61) contained at least one MSI event and 83.61% (51/61) contained at least one LOH event. MSI occurrence in 61 loci varied widely. BRAF-9 was most frequently affected with significant higher occurrence rate than other loci (5.08%, 13/256) (Fig. 1A). The MSI frequency of the top 3 frequent MS loci in the tumor-related genes (range from 4.30% to 5.08%) was lower than that of B5 loci (range from 7.42% to 9.38%). The results showed that the MSI mutation percentages of BAT25, BAT26, D5S346, D2S123, and D17S250 were very high. Remarkably, loci TP53-1 had the highest LOH occurrence rate (26.95%, 69/256). LOH occurrence in 61 loci were also ranged widely

(Fig. 1B). The LOH frequency of the top 3 frequent loci in the tumor-related genes (range from 14.45% to 26.95%) was similar to that of dinucleotide loci in B5 (range from 10.16% to 21.88%). Furthermore, the *P21* was the most commonly affected gene with MSI frequency (4.69%, 12/256), and gene *TP53* had much higher LOH frequency (26.95%, 69/256) than other genes (Fig. 1C-D and Table S3-S4).

Colorectal carcinomas with high-frequency microsatellite instability (MSI-H) account for 15% of all colorectal cancers, including 12% of sporadic cases and 3% of cancers associated with Lynch syndrome. Using the B5 panel, we classified the tumor as MSI-H, MSI-L and MSS. Colorectal cancers with MSI-H accounted for 10.94% of all cases, which were comparable with the data in literature. Moreover, MSI-H tumors defined by B5 panel were more prone to mutate in MS loci of tumor-related genes (Fig. S2).

The prognostic value and predicting the response to chemotherapy of MSI/LOH in tumor-related genes

Microsatellite instability can provide rich information for prognosis and evaluation of the chemotherapy response in the cancer patients [18,19]. The overall survival (OS) in patients with MSI-H were also longer than those with MSS/MSI-L (63.5 months versus 60.0 months, p=0.013) [20]. In the present study, we explored the relationship of MSI/LOH of 32 sensitive loci and the outcomes of CRCs only in training group (n=256) due to the lack of survival information of the second batch of samples.

However, according to MSI status of the B5 panel, the outcome were not significantly different CRC patients which in group of all stage (Fig. 2A-B), stage Ⅱ (Fig. 2C-D), stage Ⅲ (Fig. 2E-F)adjuvant chemotherapy (Fig. 2G-H). (at least two of the B5 loci show LOH) status of the B5 panel analysis also failed to indicate the outcome of CRC patients (Fig. 3) Fortunately, we found the MSI/LOH status of BAT25 and BAT26 in B5 panel and 12 loci in tumor-related genes could be sensitive markers for the outcome prediction of CRC patients (Table 1 and Fig. 4).

In group of entire patients (n=256), the MSI in D17S250 (p=0.001), MSH2-15 (p=0.001), pinch5 (p=0.03) and MCC-10 (p=0.001) loci demonstrated a poor prognosis in 5-year OS (Fig. 4A); patients with MSI in D17S250 (p=0.02), MSH2-15 (p=0.006), MCC-25 (p=0.048) and MCC-10 (p=0.001) loci showed significantly poorer outcome in 5-year progression free survival (PFS) (Table 1 and Fig. 4B).

In group of stage Ⅱ patients (n=127), the MSI in D17S250 (p=0.006), Pinch-5 (p=0.001), MSH2-15 (p=0.001), MCC-25 (p=0.024) and MCC-10 (p=0.001) MCC-3 (p=0.036), MCC-26 (p=0.049), MGMT-10 (p=0.04), APC-6 (p=0.049) loci showed a bad outcome in 5-year OS in the stage Ⅱ CRC patients (Fig. 4C). Patients with MSI in Pinch-5 (p=0.001), MSH2-15 (p=0.001), MCC-25 (p=0.024), MCC-10 (p=0.001), MGMT-10 (p=0.04) and BRAF-9 (p=0.001), showed significantly poorer outcome in 5-year PFS (Table 1 and Fig. 4D).

In group of stage Ⅲ patients (n=93), MSI in D17S250 (p=0.002) and MCC-10 (p=0.003), and LOH in loci P21 (p=0.009) and MLH1-2 (p=0.006) was related to a bad outcome in 5-year OS (Fig. 4E); MSI in D17S250 (p=0.01) and MCC-10 (p=0.001), and LOH in BRAF-9 (p=0.001), P21 (p=0.021) MLH1-2

(p=0.004) and Pinch-13 (p=0.035) showed significantly poorer outcome in 5-year PFS (Table 1 and Fig. 4F).

We also examined the association of MSI/LOH in the tumor-related genes with the response to adjuvant chemotherapy. In adjuvant chemotherapy group (n=132), the patients with MSI in D17S250 (p=0.01) and MCC-10 (p=0.001), and LOH in BAT-25 (p=0.048) presented a poorer outcome in 5-year OS (Fig. 4G); meanwhile, patients with MSI in MCC-10 (p=0.001) loci presented a poorer outcome in 5-year PFS (Table 1 and Fig. 4H).

The association of MSI/LOH profile with CRC clinical features

The clinical features such as the TNM (tumor-node-metastasis) stage and pathological type are usually important prognostic factors for patients with colorectal cancer [19]. The analysis of the association of MSI/LOH profile with CRC clinical features was set up in the training cohort (n=256) and were clarified in the validation cohort (n=440). Here, we showed that the numbers of patients with mucinous carcinoma who have MSI in BAT25 (p=0.005), MSI/LOH in BAT26 (p=0.004) or MSI-H in B5 panel (p=0.012) were significantly higher than that in the adenocarcinoma in (Table S5-S6). These results illustrated that, compared to loci in tumor-related genes, MSI/LOH of certain loci or whole panel of B5 have closer relation to the pathological type of CRCs. Next, we explored the MSI/LOH profile and its association with other clinicopathological features. Although MSI/LOH of several loci were remarkably related to TNM stage, lymphatic metastasis, infiltration depth, differentiation degree and recurrence in training group, but they all failed to be confirmed in the validation group (Table S7-S14). In regard to B5 panel, LOH-H patients showed more lymphatic metastasis than LOH-L+non-LOH CRCs in training (p=0.05) and validation (p=0.04) set (Table S9-S10).

The characteristics of the MSI/LOH within the tumor-related genes

Among MSI/LOH events, 46% MSI (100/217) and 56% LOH (511/909) were found in tumor suppressor (TS) genes. Specifically, we found that the MSI frequency in TS genes (1.50%, 100/26*256) and DNA repairing (DNAR) genes (1.50%, 23/6*256) were higher than that in oncogenes (1.25%, 64/20*256) and MMR (1.30%, 30/9*256), but the difference had no statistically significance . However, the LOH frequency in TS genes (7.68%, 511/26*256) was remarkably higher than that in DNAR genes (5.79%, 89/6*256), MMR (4.69%, 108/9*256) and oncogenes (3.93%, 201/20*256) (p=0.011; p<0.001; p<0.001, respectively). In addition, significant difference in the LOH frequency was detected between the DNAR and oncogenes (p=0.002) (Fig. 5A-B).

Regarding the location of MS in the tumor-related genes, the MSI frequency within introns was 1.5% (203/53×256), which was higher than that located in the non-coding (1.02%, 13/5×256) and exon (0.39%, 1/1×256) regions, but they did not differ significantly between each other (Fig. 5C). On the other hand, the LOH frequency within 3'UTR 7.03% (36/2×256) and intron 6.31% (856/53×256) were significantly higher than that located in the non-coding region 1.33% (17/5×256) (p<0.001; p<0.001, respectively) (Fig. 5D).

These results suggested that the MSs were rich in the introns and were prone to mutate than other regions.

Most MSIs (75.1%, 163/217) and LOHs (63.3%, 575/909) were characterized by the dinucleotide repeats existed within the tumor-related genes. The frequency of MSI with dinucleotide repeats (1.68%, 163/38×256) was remarkably higher than that with the tetranucleotide repeats (0.78%, 26/13×256) (p<0.001), and also showed distinct difference as compared with the trinucleotide repeats (0.93%, 19/8×256) (p=0.013) (Fig. 5E). The frequency of LOH with dinucleotide repeats (5.91%, 575/38×256) was higher than that with the tetranucleotide repeats (4.72%, 157/13×256) (p=0.010), but showed no significant difference with the trinucleotide repeats (5.27%, 108/8×256) (p=0.262) (Fig. 5F). These data indicated that most of MS loci were characterized with dinucleotide repeats which were more prone to mutate than other types.

To investigate the mutation patterns of the tumor-related genes in human CRCs, we divided mutations into two patterns: MSI and LOH. Among 1126 mutation events, the rates of MSI and LOH were 19.27% (n=217) and 80.73% (n=909), respectively (Fig. S3A). We found LOH was the more common mutation type in the tumor-related genes (Fig. S3B). Of the 61 MS loci, we found mutations in 54 MS loci and most (40 loci) of them exhibited both MSI and LOH patterns (Fig. S3C). There were 11 loci exhibiting LOH pattern alone and 3 loci only showing MSI pattern. Statistical analysis results indicated that the MSI frequency was similar among the four types of genes. MSI in TS genes (1.50%, 100/26×256) was similar to the DNAR genes (1.50%, 23/6×263), MMR genes (1.30%, 30/9×263), and oncogenes (1.25%, 64/20×263). But the proportion of MSI in TS genes was much lower than that in the oncogenes (p<0.01) (Fig. S3D). When focus on the locations of MS, we found introns and non-coding regions harbored two mutation types, while 3'UTR only had LOH mutation and exon had MSI mutation. The proportion of LOH pattern in introns represented 80.83% (856/1059) of all mutation events. The proportion of MSI and LOH in non-coding regions were semblable (Fig. S3E).

We further analyzed the mutation patterns based on the number of repeat unit, especially in introns, the types of repeat unit, and the length of repeat unit. It showed no correlation among these subgroups (Fig. S4), which indicated that mutation patterns have not been interfered by repeat units.

Mutational profile of MS in human CRCs

Given that the B5 panel has been frequently applied in clinical practice, and the MMR system is of pivotal importance for the occurrence of MSI, we analyzed if the MSI of tumor-related genes we studied was relevant to the status of B5 or MMR. The samples of CRC were divided into B5-MSI and B5-MSS or MMR-deficient (MMR-d) and MMR-proficient (MMR-p) groups according to the MSI status of B5 or MMR.

The data showed that the MSI frequency of 16 tumor-related genes (84.2%, 16/19) we detected was significantly higher in the B5-MSI group than that in the B5-MSS group (Table S15). It indicated that B5 panel is a high-efficiency criterion on assessing the integral MSI status of the genome.

Similarly, except four MMR genes including MSH2, MLH1, MSH6 and PMS2, MSI frequency of majority tumor-related genes (80%, 12/15) we detected was remarkably higher in MMR-d tumors than that in MMR-p tumors (Table S16). In accordance with the statement that MMR system plays an vital role in the occurrence of MSI.

The MSI/LOH spectrum in CRC patients

Increased number of mutations was detected in CRCs [21], suggesting that the mutation spectrum in CRCs was very complicated. In our study, 54.17% (26/48) of CRCs harbored MSI events within one gene, while 6.25% (3/48) in two genes simultaneously. Of 48 MSI patients, the number of MSI events detected in each individual patient was ranged from 1 (52.08%, 25/48) to 18 (2.08%, 1/48) with a mean value of 4.52 MSI events (217/48) per individual (Table S17-18). Furthermore, 22.22% (42/189) of CRCs harbored LOH events within one gene, while 17.46% (33/189) in two genes. Of 189 LOH patients, the number of LOH events detected in each individual patient was ranged from 1 (22.22%, 42/189) to 18 (0.53%, 1/189) with a mean value of 4.81 MSI events (909/189) per individual (Table S19-20). These results suggested a complicated mutation spectrum in the CRC patients.

We further found that both the gene numbers and the MSI loci numbers in the non-adenocarcinoma patients were higher than that in the adenocarcinoma patients (p=0.002; p=0.002, respectively) (Fig. S5A-B). Moreover, both the MSI genes number and the MSI loci number in the colon patients were higher than in the rectal patients (p=0.006; p=0.007, respectively) (Fig. S5C-D). However, no significant differences of the LOH genes number and the LOH loci number were found between each pair of groups. These findings suggested that the MSI frequency of tumor-related genes in colorectal cancer was associated with pathological type and tumor location.

Discussion

MSI is an important feature observed in many tumors types, especially for sporadic CRC patients with prognosis and treatment value, and has been used in clinic [11, 22]. It is associated with pathological characteristics and cancer outcomes, and used to predict response to adjuvant chemotherapy [23]. In addition, the evolution of genetic instability in colon cancer may involve chromosomal instability (CIN) pathway, which may be accompanied by a loss of heterozygosity (LOH). CIN-high CRC showed significantly poorer outcome than CIN-low CRC. Therefore, MSI and LOH status have been considered to be valuable and independent prognostic markers in CRC patients [24]. Although the risk scores based on clinical and pathological parameters have been developed to predict outcomes, the existing prognostic markers are not likely to be enough for clinical decision, and not interpreted well across institutions [25]. In our study, as evaluated with the B5 panel, the MSI and MSS patients, showed similar outcomes in 5-year OS and 5-year PFS (Fig. 2G-L), and there were no difference between the LOH and non-LOH patients in outcomes in 5-year OS and 5-year PFS as well. It suggested that, occasionally, the MSI/LOH detected by the B5 panel may not be a sufficient biomarker for predicting the outcomes in Chinese patients. Therefore, it is urgent to screen more practicable markers for colorectal cancer.

Recently, Jun Yu *et al* have shown that seven significantly mutated genes in Asian CRC and a mutation signature can predict survival outcomes [21]. Our hypothesis is that the MSI/LOH in tumor-related genes may serve as complementary markers to predict the outcome of CRC. Fortunately, several loci in B5 panel and tumor-related genes we detected showed remarkable prognostic value for all CRC patients, stage Ⅱ, stage Ⅲ group CRC patients respectively.

Although the prediction of the chemotherapy response by MSI remains controversial [12, 13], some studies have shown that MSI CRCs are particularly responsive to immunotherapy, such as anti-PD-1 blockage [26]. In the present study, we found that the patients with LOH in BAT25 or MSI in MCC-10 would not be benefit from adjuvant chemotherapy (Table 1). Therefore, the MSI/LOH status of these 2 loci may be useful, convenient, and applicable for predicting the response to chemotherapy in CRC patients.

Notably, the MS loci exhibited prognostic value were located in *MCC*, *MSH2*, *Pinch5*, *Mgmt*, *MLH1*, *APC*, *BRAF* and *P21* genes, which were reported to be involved in CRC progression [27,28]. This study may also provide a foundation for further investigation of the mechanisms underlying the functional involvement of these MSI loci in the development of CRCs.

It is suggested that there is strong correlation between CRC clinicopathological features and the MSI status. For example, the prevalence of CRCs with microsatellite is different among disease stage with 15% in stage Ⅱ and Ⅲ and more common in stage Ⅰ [12]. The MSI events may help to determine the degree of the tumor malignancy. Moreover, MSI tumors shared similar histomorphology regardless of their respective pathogenesis and frequently had a mucinous phenotype [29]. Uniformly, our data also showed a higher frequency of MSI in mucinous carcinoma compared to adenocarcinoma in certain loci or whole panel of B5. In the present study, the MSI-H status of B5 panel is related to mucinous carcinoma (p=0.012). Surprisingly, MSI of BAT25 or MSI/LOH of BAT26 which belongs to B5 panel also showed sensitive correlation to mucinous carcinoma in both training and validation sets of CRC, respectively. These indicated that patients with MSI in certain loci or whole panel of B5 tends to develop mucinous carcinoma rather than adenocarcinoma. Besides, the result that LOH-H of B5 panel was related to more lymphatic metastasis indicated that, except MSI-H, LOH-H status was also a potential marker for CRCs features.

For MSI/LOH profile, the results showed that the MSI mutation percentages of B5 were very high. These findings indicated that B5 panel was meaningful for the study of colorectal cancer. The LOH frequency of the top 3 frequent MS loci (P53, APC-6 and Nup88-3) in the tumor-related genes was similar to the dinucleotide loci (D5S346 and D17S250) of B5. Therefore, MS loci in tumor-related genes may play an important role in the study of CRC.

In the present study, we generated MSI/LOH profiles in 19 tumor-related genes which were prone to alter and might be involved in CRC tumorigenesis and progression [30-34]. Selected as one hot locus in *BRAF* gene, the BRAF-9 was the most frequent locus that was prone to mutate in the CRC patients (5.08%, 13/256), and the TP53-1 was the most frequently mutated gene in the CRC patients (26.95%, 69/256). In our previous study, we examined the MS (TP53ALU) status in the intron 1 and mutation in all exon of

TP53 gene. Also, we studied the association between *TP53*-exon mutation and TP53ALU alterations. It showed that the prevalence of *TP53*-exon mutations was significantly higher in TP53ALU-LOH tumors than that in TP53ALU-non-LOH tumors (p=0.003) (Fig. S6), suggesting that the *TP53*-exon are more likely to mutate when the MS of *TP53*-intron is in the status of LOH. However, there was no corelaiton was found between the *TP53*-exon mutation and TP53ALU-MSI status (Fig. S6) (have not been published). This indicated that, the LOH of MS in *TP53*-intron seems to be a sensitive marker for the mutaion status of *TP53*-exon which always play a crucial role in CRC tumorigenesis.

In addition, our study showed that MS in TS genes were more prone to mutate than MS located in MMR genes or oncogenes, suggesting the inactivation of tumor suppressor genes in CRC as previous reports [35]. We also found less frequency of MSI events in certain genes, such as *MYC*, *MDM2*, *BBC3*, and *KRAS*. It is notable that lower occurrence of MSI in *KRAS* genes, which are often mutated in CRC. A larger cohort of CRC patients may validate the phenomenon.

MSs are abundant in both non-coding and coding regions in mammalian genomes [36]. MS mutations occurring in the coding regions, introns, or untranslated regions may positively or negatively influence gene expression or protein function by interrupting gene transcription or splicing [37]. We observed that the MSI/LOH frequency in introns was higher than that in other locations. The prognostic and predicted panels of MSI/LOH were mostly located in introns, which suggested that MSs in introns may be prone to alter and relevant with clinicopathological features of CRC. Moreover, we also found one MSI event in exon 2 of *MYC* with $(CAG)_5$ repeats, and a new alleles emerged (174/174 to 165/174). Although only one MSI event found in the exon of *MYC* gene in one sample, this MSI may play a pivotal role in CRC as Jason B *et al*. reported [38]. In addition, we also found two loci with LOH in the 3'UTR of *MDM2* gene. Mutations within the 3'UTR might be contributable to alteration in the recognition sites of microRNAs or RNA-binding proteins, thus affecting gene expression. Importantly, 88.46% (69/78) mutation events at TP53-1 $(AAAAT)_8$ were LOH, which may be a key event in the pathogenesis of CRCs involved in “second hit” (mutation and subsequent LOH) process [22].

The MSI status of B5 panel and the expression of MMR genes are frequently-used criterion to determine the MSI status of a CRC. In the present study, the MSI frequency of the tumor-related genes (except *BBC3* and *MYC*) were higher in B5-MSI tumors than those in B5-MSS tumors, suggesting that B5 panel is a powerful tool in defining the MSI status of CRC. Similarly, the MSI frequency of 80% tumor-related genes we detected were significantly higher in MMR-MSI tumors than those in MMR-MSS tumors. These results are in agreement with previous report that higher mutation loads were frequently found in the tumors with mismatch repair deficiency [5].

There are several limitations in our study. As a retrospective study, drawing more convincing conclusions was unavoidably limited. Due to the restrictions of medical records and short follow-up time, we insufficiently collected the data regarding the treatment and survival information from the patients we recruited in this study. Moreover, to identify potential risk factors for prognosis of CRC patients, multivariate cox regression analyses were conducted. The analyses showed that tumor recurrence was a

high risk for prognosis of CRC patients (RR =9.379, 95% CI: 4.522–19.453, p<0.001). In the experimental group, several loci were related to recurrence. However, the validation group did not confirm these loci. In addition, the 61 MS loci we selected from 19 genes were predetermined based on the PCR amplification efficacy, not all MS loci in these tumor related genes were included, potentially missing other important loci in these genes.

In conclusion, we described the MSI/LOH profile in 19 tumor-related genes and performed analysis to identify the clinical correlations and significance. Most important of all, we found several loci in B5 panel and tumor-related genes for prognosis and predicting the response to the chemotherapy in Chinese CRC patients. Two loci of MSI/LOH were associated with pathological type of CRC. Our study offered a landscape of MS in the 19 tumor-related genes in Chinese CRC patients and provide significant implications for the clinical application.

Methods

Cohort Selection and DNA Extraction

This study was performed on 440 pairs of CRC and adjacent normal tissues collected of the local Institutional Review Broad of Beijing Friendship Hospital. Patients were divided into training (n=256; 2006 to 2014) and validation sets (n=440; 2005 to 2014). The patients with colorectal cancers were at the stage -, according to the TNM system classification of the American Joint Committee on Cancer. Informed consent was obtained from all individuals, and the principal inclusion criteria were as follows: histologically confirmed papillary/tubular adenocarcinoma, signet ring carcinoma and mucinous carcinoma of the colon or rectum. The patients were followed until their last contact or death. Vital status and cause of death were obtained from medical records, tumor registry correspondence, or death confirmation.

Clinicopathologic data were obtained from medical record archive. The clinical and histopathological information of 256 patients was shown in Supplementary Table S1. In brief, the mean age was 67.39 years (range=49-86 years.), while 57.48% were male and 42.52% were female, but the gender information of two patients were missing. Moreover, 34.51% (n=88) and 22.35% (n=57) of the patients showed the history of smoking and drinking. Overall, 131 patients (51.17%) received adjuvant treatment; Stage and tumors represented 57.73% and 42.27% of the cases, respectively. Regarding the location of tumors, 54.69% (n=140) and 45.31% (n=116) of tumors were located within the colon and rectum, respectively. The overall survival rate and the disease-free survival rate were 81.64% (n=209) and 76.17% (n=195), respectively. Among the patients studied, forty-seven died during the data collection.

Genomic DNA was extracted from total 880 samples (440 pairs) by a standard phenol-chloroform method [39]. The DNA quality was analyzed by a micro-volume spectrophotometer (Thermo Scientific NanoDrop 2000, Waltham, MA, USA) and agarose gel electrophoresis.

MS in Tumor-related Genes

Based on previous research about the development of CRC, we firsty selected 19 genes which are closely related to the tumorigenesis of CRC. We determined 145 microsatellite loci in 4 MMR genes (*MLH1*, *MSH6*, *PMS2* and *MSH2*), 7 TS genes (*Tp53*, *CDKN1A*, *ATM*, *APC*, *MCC*, *BBC3*, and *PTEN*), 7 oncogenes (*KRAS*, *NUP88*, *BRAF*, *LIMS1*, *MDM2*, *MYC* and *TMEM97*), and 1 DNAR (*MGMT*) using SSRHunter software. We designed and synthesized primers to amplify these loci. The PCR conditions for these 145 MS loci were optimized by PCR amplification in a gradient thermal cycler (BIO-RAD Inc. ALS1296, Hercules, CA, USA) using the same protocol as we previously reported [39]. The PCR products were evaluated on 2% agarose gels and visualized using a UV transilluminator (BIO-RAD Inc. Gel Doc^{TM} XR+), through which 61 Microsatellite loci were successfully amplified (Table S2).

Microsatellite Instability and Loss of Heterozygosity

Microsatellite status in CRC was determined by PCR amplification using the primer pairs for 61 microsatellite loci. The 5'-end of the forward primer for each locus was tagged with a FAM, HEX, or TAMRA fluorescent marker. The PCR amplification was performed using the optimized annealing temperature for each pair of primers. The PCR products were evaluated on 2% agarose gels prior to STR scanning.

The PCR product of microsatellite was visualized through capillary electrophoresis on an ABI-3730XL DNA Analyzer system (PE Biosystems, Carlsbad, CA, USA). The peak height of the wave for each specimen was determined using GeneMarker version 1.75. Microsatellite instability was also accessed by 5 Bethesda loci including BAT25, BAT26, D2S123, D5S346, and D17S250. Using capillary array electrophoresis, MSI may be demonstrated using two main features: de novo alleles that appear as new peaks (i.e., peaks that did not exist in the normal tissue genotype) and slipped pre-existing alleles for the few base pairs [40, 41], the samples do not exhibit MSI are defined as MSS. Except MSI, we also analyzed the LOH mutation, which is another mutant phenomenon different from MSI and a partial (>35%) to complete signal loss of one heterozygote allele is an indicator of LOH [42, 43], the samples do not exhibit LOH are defined as non-LOH. Exemplary images of MSI and LOH for BAT-25/P53-1 loci were shown in the Fig. S1.

Statistical Analysis

The statistical analysis was performed using the IBM SPSS® Statistics 16.0 package software (SPSS Inc). Pearson's Chi-square or Fisher exact test was performed to compare the proportion of MSI/LOH within tumor-related genes among groups with different clinicopathological variables as appropriate in this study. The Kaplan-Meier method was used to estimate the survival outcomes. A p value<0.05 was considered statistically significant. The symbol* indicates p-value<0.05, ** indicates p-value<0.01, *** indicates p-value<0.001.

Abbreviations

MSI: Microsatellite instability; CRC: Colorectal cancer; LOH: Loss of heterozygosity; B5: Bethesda; CIN: Chromosomal instability; MMR: Mismatch repair; TS: Tumor suppressor; DNAR: DNA repair; OS: Overall survival; PFS: progression free survival; UTR: untranslated regions; TNM: tumor-node-metastasis.

Declarations

Acknowledgement:

We thank the local Institutional Review Broad of Beijing Friendship Hospital and medical record archive for providing samples and clinicopathological data.

Authors' contributions

Zhenwen Chen and Zhigang Bai designed the study. Dandan Feng and Shuangyue Zhang performed the experiments. Xueyun Huo analysed the data and wrote the manuscript with the help of Dandan Feng and Qingxian Lu. Xueyun Huo complete the submission and revision. Zhenkun Li, Xiaohong Li, Changlong Li, Meng Guo, Jin Wang participated in the data interpretation and Zhongtao Zhang and Xiaoyan Du contributedstudy. All co-authors commented on the manuscript. All authors read and approved to designing the the final manuscript.

Funding

This work was supported by the National Science Foundation of China (Nos. 31772545) and Support Project of High-level Teachers in Beijing Municipal Universities in the Period of 13th five-year Plan (IDHT 20170516).

Availability of data and materials

The data that support the findings of this study are available on request from the corresponding author.

Ethics approval and consent to participate

The local Institutional Review Broad of Beijing Friendship Hospital approved the study proposal and all patients involved in the research have signed written informed consent.

Consent for publication

Not applicable.

Competing interests

No competing interests declared.

References

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics. *CA Cancer J Clin.* 2017;67:7-30.
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China. *CA Cancer J Clin*. 2016;66:115-32.
3. Gelsomino F, Barbolini M, Spallanzani A, Pugliese G, Cascinu S. The evolving role of microsatellite instability in colorectal cancer: A review. *Cancer Treat Rev*. 2016;51:19-26.
4. Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. *Histopathology*. 2007;50: 113-130.
5. Lin EI, Tseng LH, Gocke CD, Reil S1, Le DT3, Azad NS, et al. Mutational profiling of colorectal cancers with microsatellite instability. *Oncotarget.* 2015;6:42334-44.
6. Boland CR, Goel A. Microsatellite instability in colorectal cancer. *Gastroenterology*. 2010;138:2073-2087.e3.
7. Yan WY, Hu J, Xie L, Cheng L, Yang M, Li L, et al. Prediction of biological behavior and prognosis of colorectal cancer patients by tumor MSI/MMR in the Chinese population. *Onco Targets Ther.* 2016;9:7415-7424.
8. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, et al. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.* 1998;58:5248-57.
9. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Rüschoff J, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst.* 2004;96:261-8.
10. Kim H, Jen J, Vogelstein B, Hamilton SR. Clinical and pathological characteristics of sporadic colorectal carcinomas with DNA replication errors in microsatellite sequences. *Am J Pathol.* 1994;145:148-56.
11. Kim CG, Ahn JB, Jung M, Beom SH, Kim C, Kim JH, et al. Effects of microsatellite instability on recurrence patterns and outcomes in colorectal cancers. *Br J Cancer.* 2016;115:25-33.
12. Ribic CM, Sargent DJ, Moore MJ, Thibodeau SN, French AJ, Goldberg RM, et al. Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. *N Engl J Med.* 2003;349:247-57.
13. Lamberti C, Lundin S, Bogdanow M, Pagenstecher C, Friedrichs N, Büttner R, et al. Microsatellite instability did not predict individual survival of unselected patients with colorectal cancer. *Int J Colorectal Dis.* 2007;22:145-52.
14. Watanabe T, Wu TT, Catalano PJ, Ueki T, Satriano R, Haller DG, et al. Molecular predictors of survival after adjuvant chemotherapy for colon cancer. *N Engl J Med.* 2001;344:1196-1206.
15. Hause RJ, Pritchard CC, Shendure J, Salipante SJ. Classification and characterization of microsatellite instability across 18 cancer types. *Nat Med.* 2016;22:1342-1350.

16. Rodriguez-Salas N, Dominguez G, Barderas R, Mendiola M4, García-Albéniz X5, Maurel J, et al. Clinical relevance of colorectal cancer molecular subtypes. *Crit Rev Oncol Hematol.* 2017;109:9-19.
17. Ma BB, Mo F, Tong JH, Wong A, Wong SC, Ho WM, et al. Elucidating the prognostic significance of KRAS, NRAS, BRAF and PIK3CA mutations in Chinese patients with metastatic colorectal cancer. *Asia Pac J Clin Oncol.* 2015;11:160-9.
18. Pritchard CC, Grady WM. Colorectal cancer molecular biology moves into clinical practice. *Gut.* 2011;60:116-29.
19. Brenner H, Kloor M, Pox CP. Colorectal cancer. *Lancet.* 2014;383:1490-502.
20. Sung PH, Byung SM, Tae IK, Cheon JH, Kim NK, Kim H, et al. The differential impact of microsatellite instability as a marker of prognosis and tumor response between colon cancer and rectal cancer. *EJC.* 2012;48:1235-1243.
21. Yu J, Wu WK, Li X, He J, Li XX, Ng SS, et al. Novel recurrently mutated genes and a prognostic mutation signature in colorectal cancer. *Gut.* 2015;64:636-45.
22. Carethers JM, Jung BH. Genetics and Genetic Biomarkers in Sporadic Colorectal Cancer. *Gastroenterology.* 2015;149:1177-1190.e3.
23. Hemminki A, Mecklin JP, Jarvinen H, Aaltonen LA, Joensuu H. Microsatellite instability is a favorable prognostic indicator in patients with colorectal cancer receiving chemotherapy. *Gastroenterology.* 2000;119:921-8.
24. Chang SC, Lin JK, Lin TC, Liang WY. Loss of heterozygosity: an independent prognostic factor of colorectal cancer. *World J Gastroenterol.* 2005;11: 778-784.
25. Zakaria S, Donohue JH, Que FG, Farnell MB, Schleck CD, Ilstrup DM, et al. Hepatic resection for colorectal metastases: value for risk scoring systems? *Ann Surg.* 2007;246:183-91.
26. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med.* 2015;372:2509-20.
27. Fukuyama R, Niculaita R, Ng KP, Obusez E, Sanchez J, Kalady M, et al. Mutated in colorectal cancer, a putative tumor suppressor for serrated colorectal cancer, selectively represses beta-catenin-dependent transcription. *Oncogene.* 2008;27:6044-55.
28. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature.* 2012;487:330-7.
29. Wei C, Benjamin JS, Wendy LF. Molecular genetics of microsatellite unstable colorectal cancer for pathologists. *Diagnostic Pathology.* 2017;12:24.
30. Zhao ZR, Zhang LJ, Wang YY, Li F, Wang MW, Sun XF. Increased serum level of Nup88 protein is associated with the development of colorectal cancer. *Med Oncol.* 2012;29:1789-95.
31. Moparthi SB, Arbman G, Wallin A, Kayed H, Kleeff J, Zentgraf H, et al. Expression of MAC30 protein is related to survival and biological variables in primary and metastatic colorectal cancers. *Int J Oncol.* 2007;30:91-5.

32. Loof J, Rosell J, Bratthall C, Doré S, Starkhammar H, Zhang H, et al. Impact of PINCH expression on survival in colorectal cancer patients. *BMC Cancer.* 2011;11:103.
33. Helwa R, Gansmo LB, Romundstad P, Hveem K, Vatten L, Ryan BM, et al. MDM2 promoter SNP55 (rs2870820) affects risk of colon cancer but not breast-, lung-, or prostate cancer. *Sci Rep.* 2016;6:33153.
34. Chen D, Wei L, Yu J, Zhang L. Regorafenib inhibits colorectal tumor growth through PUMA-mediated apoptosis. *Clin Cancer Res.* 2014;20:3472-84.
35. Sheaffer KL, Elliott EN, Kaestner KH. DNA Hypomethylation Contributes to Genomic Instability and Intestinal Cancer Initiation. *Cancer Prev Res (Phila).* 2016;9:534-46.
36. Weber JL. Informativeness of human (dC-dA)n.(dG-dT)n polymorphisms. *Genomics.* 1990;7:524-30.
37. Gymrek M, Willems T, Guilmatre A, Zeng H, Markus B, Georgiev S, et al. Abundant contribution of short tandem repeats to gene expression variation in humans. *Nat Genet.* 2016;48:22-9.
38. Wright JB, Brown SJ, Cole MD. Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. *Mol Cell Biol.* 2010;30:1411-20.
39. Zhang S, Huo X, Li Z, Li X, Tang W, Li C, et al. Microsatellite instability detected in tumor-related genes in C57BL/6J mice with thymic lymphoma induced by N-methyl-N-nitrosourea. *Mutat Res.* 2015;782:7-16.
40. Dietmaier W, Wallinger S, Bocker T, Kullmann F, Fishel R and Rüschoff J. Diagnostic microsatellite instability: defnition and correlation with mismatch repair protein expression. *Cancer Res.* 1997;57:4749-4756.
41. Castagnaro A, Marangio E, Verduri A, Chetta A, D'Ippolito R, Del Donno M, et al. Microsatellite analysis of induced sputum DNA in patients with lung cancer in heavy smokers and in healthy subjects. *Exp Lung Res.* 2007;33:289-301.
42. Green MR, Jardine P, Wood P, Wellwood J, Lea RA, Marlton P, et al. A new method to detect loss of heterozygosity using cohort heterozygosity comparisons. *BMC Cancer.* 2010;10:195.
43. Powierska-Czarny J, Miścicka-Sliwka D, Czarny J, Grzybowski T, Wozniak M, Drewa G, et al. Analysis of microsatellite instability and loss of heterozygosity in breast cancer with the use of a well characterized multiplex system. *Acta Biochim Pol.* 2003;50: 1195-1203.

Figures

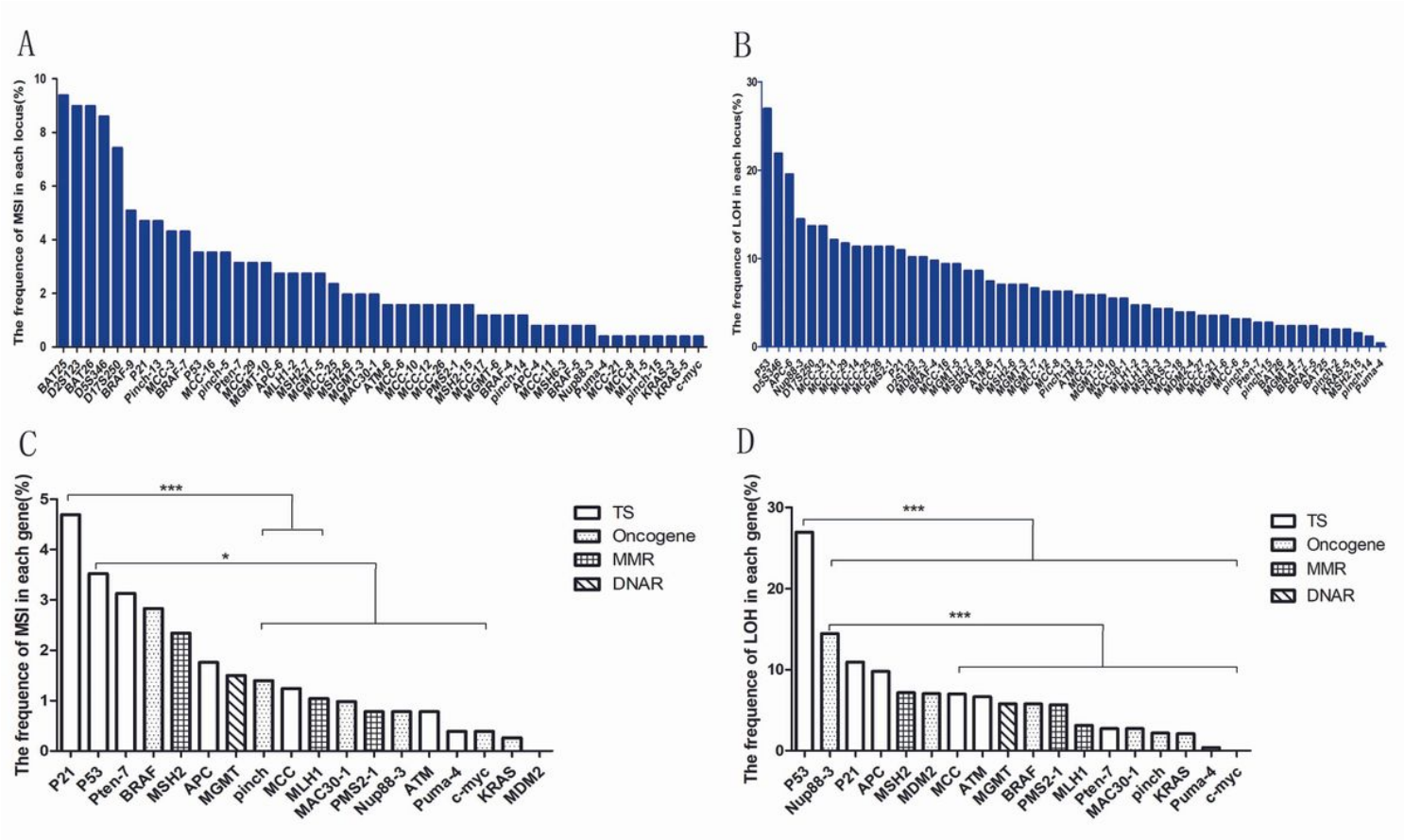

Figure 1

MSI/LOH frequency of loci in B5 and tumor-related genes in CRCs. (A) MSI occurrence in 43 of the 61 loci, expressed as percentage of patients carrying MSI on the indicated locus per total 256 patients. (B) LOH occurrence in 51 of the 61 loci, expressed as percentage of patients carrying LOH on the indicated locus per total 256 patients. (C-D) Frequency of MSI/LOH in each of 18 tumor-related genes, expressed as percentage of patients carrying MSI/LOH on the indicated gene per total 256 patients. p values were obtained from $\chi 2$-test. *$p<0.05$; ***$p<0.001$.

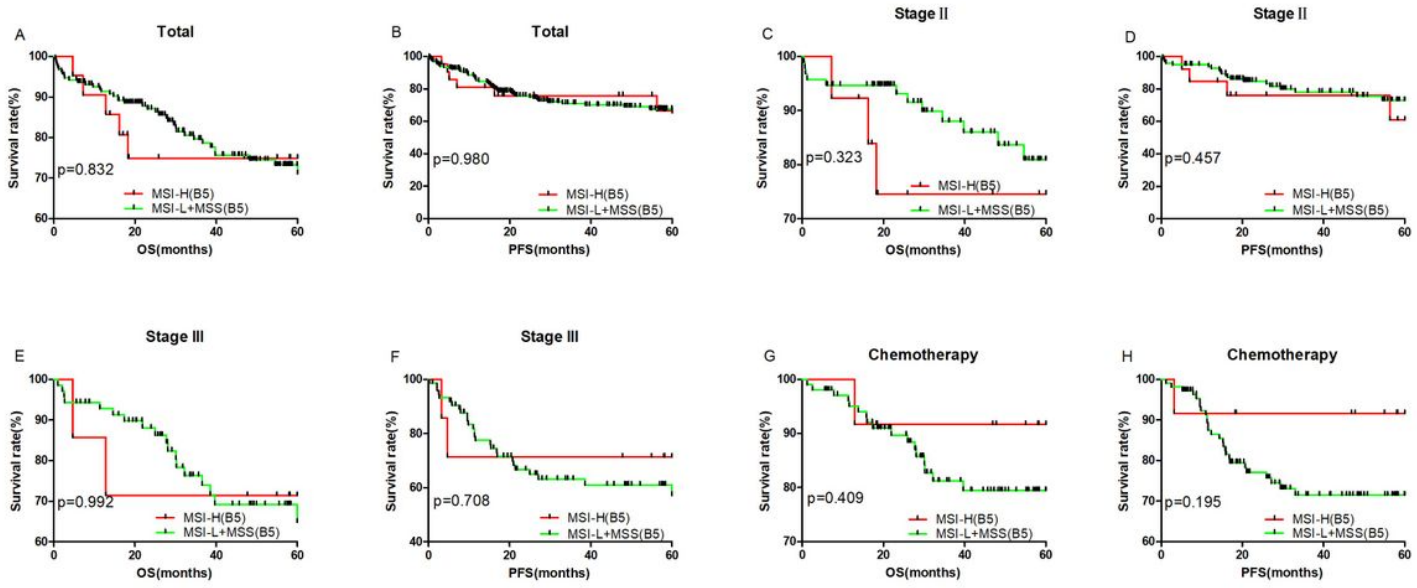

Figure 2

Survival analysis in group of 256 CRC patients, stage Ⅱ, stage Ⅲ and chemotherapy patients according MSI status detected by B5 panel. MSI presenting at least one loci unstable among B5 panel. All p values were obtained by log-rank test. Kaplan-Meier analysis for OS and PFS of MSI patients in (A-B) total 256, (C-D) stage Ⅱ, (E-F) stage Ⅲ and (G-H) chemotherapy treatment according to B5 panel.

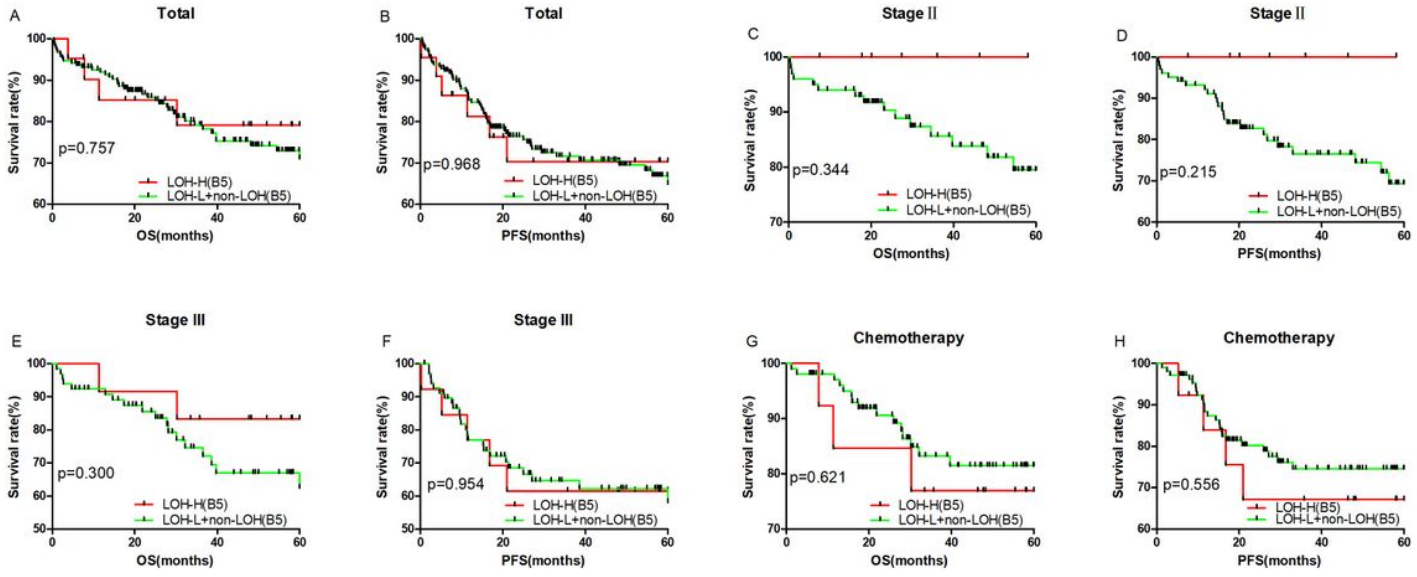

Figure 3

Survival analysis in group of 256 CRC patients, stage Ⅱ, stage Ⅲ and chemotherapy patients according LOH status detected by B5 panel. LOH presenting at least one loci unstable among B5 panel. All p values were obtained by log-rank test. Kaplan-Meier analysis for OS and PFS of LOH patients in (A-B) total 256, (C-D) stage Ⅱ, (E-F) stage Ⅲ and (G-H) chemotherapy treatment according to B5 panel.

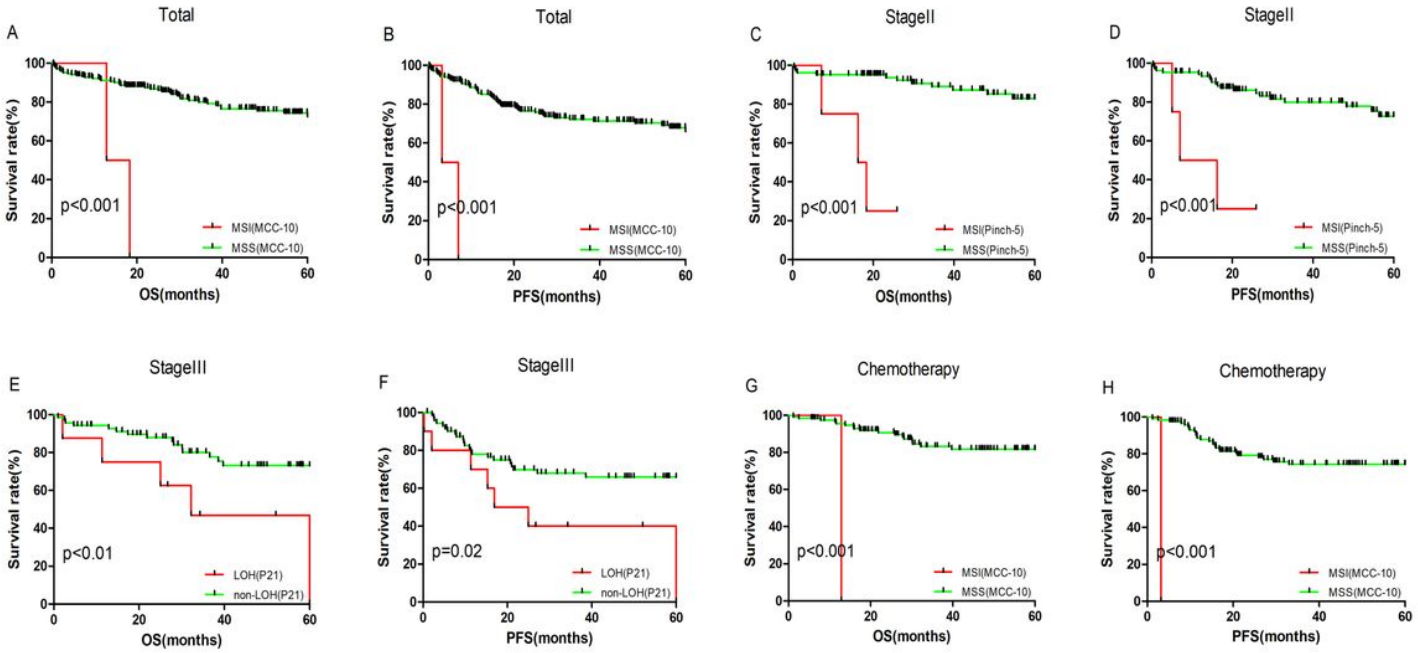

Figure 4

Survival analysis in group of 256 CRC patients, stage Ⅱ, stage Ⅲ and chemotherapy patients according MSI/LOH status detected by loci in B5 panel and tumor-related genes. All p values were obtained by log-rank test. Kaplan-Meier analysis for OS and PFS of MSI/LOH patients in (A-B) total 256, (C-D) stage Ⅱ, (E-F) stage Ⅲ and (G-H) chemotherapy treatment according to loci in B5 panel and tumor-related genes.

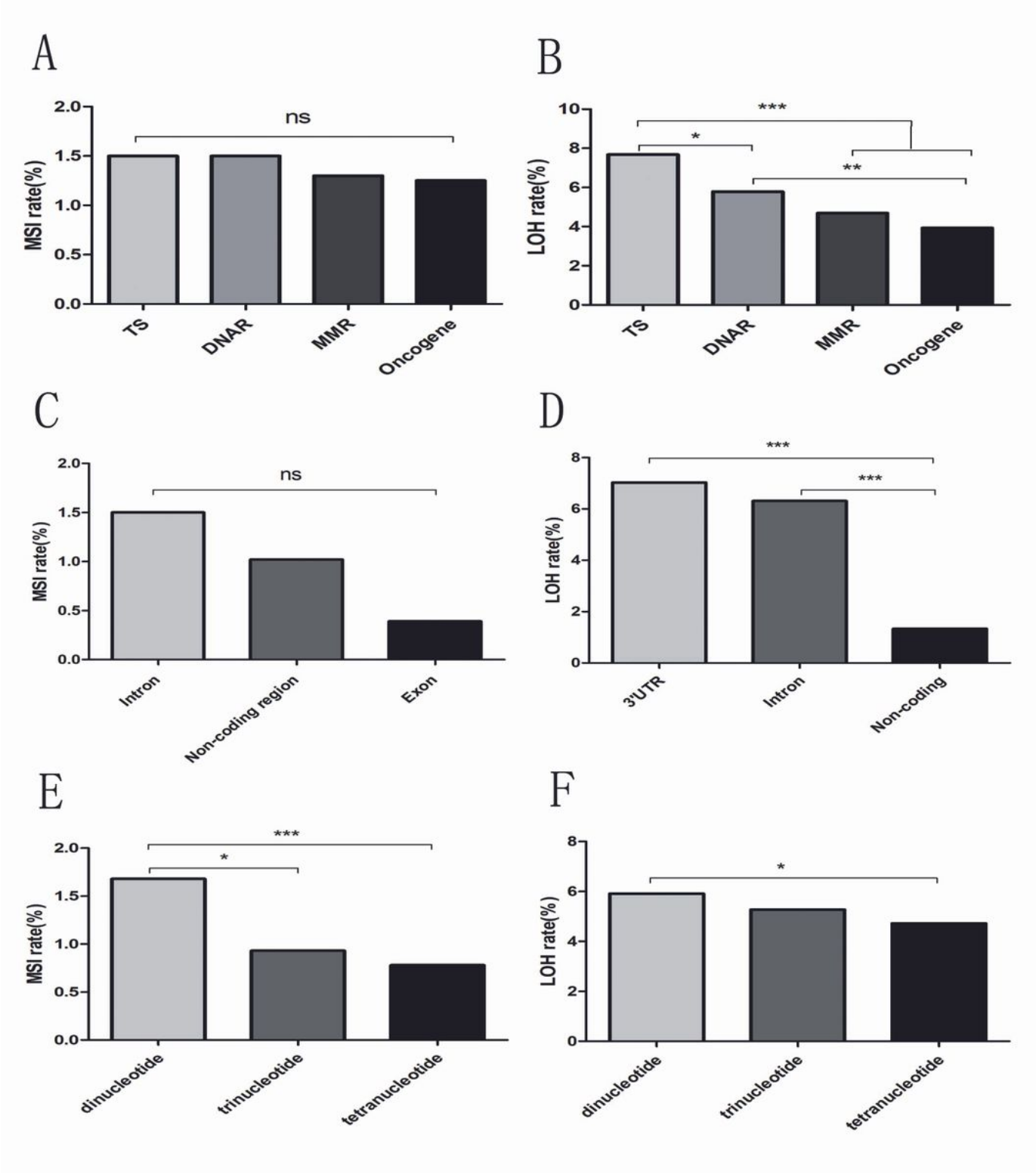

Figure 5

Gene types, location, repeat units, and patterns of MSI/LOH in the tumor-related genes of CRCs.(A-B) The frequency of the MSI/LOH events occurred within each type of the TS, DNAR, MMR, and oncogene groups. The frequency= (the number of MSI/LOH events)/(the number of loci affected within the gene type×the number of tumor samples). (C) The ratios of the MSI events appeared in intron, non-coding region and exon. The ratio= (the number of MSI events within the indicated region)/ (the number of loci

affected within the region×the number of tumor samples). (D) The ratios of the LOH events appeared in intron, non-coding region and 3'UTR. The ratio= (the number of MSI events within the indicated region)/(the number of loci affected within the region×the number of tumor samples). (E) The frequency of MSI events grouped by the number of nucleotides repeats, which includes dinucleotide, trinucleotide, and tetranucleotide repeats. The rate= (the number of MSI events carrying the indicated repeats)/(the number of loci affected×the number of tumor samples). (F) The frequency of LOH events grouped by the number of nucleotides repeats, which includes dinucleotide, trinucleotide, and tetranucleotide repeats. The rate= (the number of MSI events carrying the indicated repeats)/(the number of loci affected×the number of tumor samples). Note, one locus with pentanucleotide repeat occurred MSI and was excluded from this analysis. p values were obtained from $\chi 2$-test. *$p<0.05$; ***$p<0.001$.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- figureS3.tif
- figureS2.tif
- figureS5.tif
- figureS6.tif
- figureS4.tif
- supplementtables2019.5.21.doc
- figureS1.tif