

74288_Auto_Edited.docx

Name of Journal: *World Journal of Gastroenterology*

Manuscript NO: 74288

Manuscript Type: ORIGINAL ARTICLE

Basic Study

Fecal Gene Detection Based on Next Generation Sequencing for Colorectal Cancer Diagnosis

Fecal Gene Detection for CRC Diagnosis

Si-Yu He, Ying-Chun Li, Yong Wang, Hai-Lin Peng, Cheng-Lin Zhou, Chuan-Meng Zhang, Sheng-Lan Chen, Jian-Feng Yin, Mei Lin

Abstract

BACKGROUND

Colorectal cancer (CRC) is one of the most common malignancies worldwide. Given its insidious onset, the condition often already progresses to advanced stage when symptoms occur. Thus, early diagnosis is of great significance for timely clinical intervention, efficacy enhancement and prognostic improvement. Featuring high throughput, fastness and rich information, Next Generation Sequencing (NGS) can greatly shorten the detection time, which is a widely used detection technique at present.

AIM

To screen specific genes or gene combinations in fecal DNA that are suitable for diagnosis and prognostic prediction of CRC; and to establish a technological platform for CRC screening, diagnosis and efficacy monitoring through fecal DNA detection.

METHODS

NGS was used to sequence the stools of patients with CRC, which were then compared with the stool samples of normal control and benign intestinal disease groups, as well as the tumor tissues of CRC patients. Specific genes or gene combinations in fecal DNA suitable for diagnosis and prognostic prediction of CRC were screened, and their significances in diagnosing CRC and predicting patients' prognosis were comprehensively evaluated.

RESULTS

High mutation frequencies of TP53, APC and KRAS were detected in the stools and tumor tissues of CRC patients prior to surgery. Contrastively, no pathogenic mutations of the above 3 genes were noted in the postoperative stools, the normal controls, or in the benign intestinal disease group. This indicates that the tumor-specific DNA was detectable in the preoperative stools of CRC patients. The preoperative fecal expressions

of tumor-associated genes can reflect the gene mutations in tumor tissues to some extents. Compared to the postoperative stools and the stools in the two control groups, the pathogenic mutation frequencies of TP53 and KRAS were significantly higher for the preoperative stools ($\chi^2 = 7.328, P < 0.05$; $\chi^2 = 4.219, P < 0.05$), suggesting that fecal TP53 and KRAS genes can be used for CRC screening, diagnosis and prognostic prediction. No significant difference in the pathogenic mutation frequency of APC gene was found from the postoperative stools or the two control groups ($\chi^2 = 0.878, P > 0.05$), so further analysis with larger sample size is required. Among CRC patients, the pathogenic mutation sites of TP53 occurred in 16 of 27 preoperative stools, with a true positive rate of 59.26%, while the pathogenic mutation sites of KRAS occurred in 10 stools, with a true positive rate of 37.04%. The sensitivity and negative predictive value for the combined genetic testing of TP53 and KRAS were 66.67% (18/27) and 68.97%, respectively, both of which were higher than the TP53 or KRAS mutation detection alone, suggesting that the combined genetic testing can improve the CRC detection rate. The mutation sites TP53 exon4 A84G and EGFR exon20 I821T (mutation start and stop positions were both 7579436 for the former, while 55249164 for the latter) were found in the preoperative stools and tumor tissues. These "undetected" mutation sites may be new types of mutations occurred during the CRC carcinogenesis and progression, which needs to be confirmed through further research. Some mutations of "unknown clinical significance" were found in such genes as TP53, PTEN, KRAS, BRAF, AKT1 and PIK3CA, whose clinical values are worthy of further exploration.

CONCLUSION

NGS-based fecal genetic testing can be used as a complementary technique for the CRC diagnosis. Fecal TP53 and KRAS can be used as specific genes for the screening, diagnosis, prognostic prediction and recurrence monitoring of CRC. Moreover, the combined testing of TP53 and KRAS genes can improve the CRC detection rate.

Key Words: Colorectal cancer; Feces; Next Generation Sequencing; Diagnosis

He SY, Li YC, Wang Y, Peng HL, Zhou CL, Zhang CM, Chen SL, Yin JF, Lin M. Fecal Gene Detection Based on Next Generation Sequencing for Colorectal Cancer Diagnosis. *World J Gastroenterol* 2022; In press

Core Tip: Colorectal Cancer (CRC) is characterized by high morbidity and mortality, as well as low early diagnosis rate. The development of current gold standard for clinical diagnosis of CRC is restricted due to its invasiveness. In this study, Next Generation Sequencing (NGS) was used to detect genes in the tumor tissues and preoperative/postoperative stools of patients with CRC. The genetic testing results were compared with the fecal genes in the normal control and benign intestinal disease groups. It is found that the NGS-based fecal gene detection can serve as a complementary technique for the CRC diagnosis, and that the TP53 and KRAS in stools can be used as specific genes for the screening, diagnosis, prognostic prediction and recurrence monitoring of CRC. Moreover, the combined testing of TP53 and KRAS genes can improve the CRC detection rate. Additionally, we found "undetected" mutation sites and mutations of "unknown clinical significance" in the preoperative stools and tumor tissues of CRC patients. These mutation sites require further exploration in the future.

INTRODUCTION

CRC, as a health issue of great concern worldwide, poses a serious threat to human health. According to the latest data² released by the International Agency for Research on Cancer (IARC), there were approximately 19.3 million new cancer cases and nearly 10 million cancer deaths globally in 2020. Among them, new CRC cases amounted to about 1.93 million (10%), ranking third, while CRC deaths amounted to about 0.94 million (9.4%), ranking second. For female patients, there were approximately 860,000 new cases of CRC (9.4%), second only to breast cancer, and nearly 870,000 CRC deaths (9.5%), ranking third. As for male population, the numbers of new cases and deaths

both ranked third for CRC, which were 930,000 (10.6%) and 50,000 (9.3%), respectively^[1].

⁵ In recent years, the morbidity and mortality of CRC have been on the rise in China, and the onset age of patients has gradually become younger. Early detection of CRC is difficult due to the insidious onset, long progressive course, and unobvious symptoms and signs of patients at the onset stage. As the tumor grows and gradually compresses the intestinal cavity, such symptoms as altered bowel habit and abnormal stool color or traits may occur. At this time, the condition often already progresses to advanced stage.¹ Studies have shown that the 5-year survival of patients with early CRC can reach 90%, while is only about 12.5% for advanced stage patients^[2]. Thus, early diagnosis is conducive to early intervention and treatment of CRC, which is of great significance for improving the efficacy and prognosis^[3,4]. In response, the American Cancer Society (ACS) recommended the CRC screening since the age of 45 for adults in 2018, which is 5 years earlier than the previously recommended age^[5].

¹ The occurrence and progression are a multi-factor, multi-stage complex process^[6-8]. It is generally believed that CRC is caused by the mutations and malignant proliferation of colorectal mucosal epithelial cells. Other studies have shown that CRC can develop from inflammatory bowel disease (IBD). IBD is mainly characterized by intestinal mucosal barrier dysfunction, because the changes of cytokines aggravate tissue damage and eventually lead to tumorigenesis. It is reported that patients with ulcerative colitis have a higher risk of developing CRC than the general population^[9]. In addition, microRNA and Tumor microenvironment (TME) also have potential pathogenic effects on CRC^[10,11]. Currently, the clinically recognized gold standard for CRC detection remains endoscopy combined with histopathological analysis. Given the heterogeneity of tumor tissues and the invasiveness of endoscopy, the same patients cannot be sampled multiple times, so that the genetic testing results cannot fully display the mutation spectrum of tumor genes^[12]. According to the Chinese *Expert Consensus on Early Diagnosis and Screening Strategies for Colorectal Tumors* published in October 2018, fecal detection is a promising method of early diagnosis^[13]. Its theoretical basis is that

the tumor cells in the intestinal cavity of CRC patients adhere less to the base and can be continuously exfoliated from the colorectal mucosa. Studies have shown that tumor-associated DNA can stably exist in the stools, which have high specificity for the detection of CRC and are the most suitable biomarkers in fecal detection^[14-16]. By extracting the DNA of tumor shed cells in stools and analyzing their mutations, the traits of intestinal tumor cells can be reflected. Fecal DNA detection, as an emerging, noninvasive technique for screening CRC, has attracted great attention in recent years, which has the advantages of noninvasiveness, convenient access to materials, good patient compliance and continuous dynamic monitoring. It is in line with the concept and conforms to the trend of "precision medicine", which is an exceptionally promising detection method.

With the continuous development of molecular biotechnology, NGS has emerged to gradually become a hot research topic worldwide. It is also known as high-throughput sequencing, which operates on the following principles: 4 different dNTPs are labeled with different fluorescent colors. Conforming to the complementary base pairing rule, the bound base releases the corresponding fluorescence by the action of DNA polymerase every time a dNTP is added. Finally, the base is identified according to the fluorescent signal. Compared to the first-generation Sanger sequencing, NGS is a set of technologies that completely change the standard concept of nucleic acid sequencing. Capable of detecting millions of DNA molecules simultaneously, it achieves concurrent synthesis and sequencing, which can greatly shorten the detection time, and has the advantages of high throughput, high efficiency and rich information^[17-18]. NGS can more comprehensively display the full spectrum of tumor-associated genes, which can reflect the changes in tumor-specific genes, expand the understanding of tumor-specific genes, and improve the predictability of targeted therapy, suggesting its significance in achieving personalized tumor treatment and developing new therapeutic strategies^[19-21]. Currently, NGS has been widely applied in the research of various solid tumors, including CRC. Kraus and his colleagues performed NGS analysis on the tissue paraffin sections and blood samples of 152 CRC

patients based on 18 CRC-related genes. Their results showed that NGS had a high detection rate and could effectively identify the high-risk patients with hereditary CRC^[22]. In a study by Li *et al*, the NGS (Ion Torrent PGM platform) technology was applied to analyze 22 tumor-associated mutation hotspots in 526 CRC patients, finding that TP53 and KRAS mutations were the most common in tumor tissues, and confirming that NGS was a highly sensitive detection technique^[23].

This project performs NGS-based genetic testing of 50 tumor-associated mutation hotspots in the stools of CRC patients, and compare them with the genetic testing results of corresponding tumor tissues, with a view to finding the fecal genes or gene combinations with high specificity and sensitivity, and to establishing a technological platform for CRC screening, diagnosis and efficacy monitoring through fecal DNA detection. The findings herein provide support and basis for personalized diagnosis and treatment of CRC.

MATERIALS AND METHODS

Subjects

Experimental group: 50 tumor tissues, 27 preoperative stools and 19 postoperative stools (stools formed over 7 days postoperatively) were collected from patients with CRC.

Normal control group: 20 stool samples were collected from healthy subjects.

Benign control group: 17 stool samples of patients with benign intestinal polyps confirmed by digestive endoscopy were collected, as well as 3 stool samples from patients with ulcerative colitis, totaling 20 samples. (Postoperative pathological reports of polyp patients revealed inflammatory or adenomatous polyps, none of which had progressed to the CRC stage)

This project has been approved by the Ethics Committee of Taizhou People's Hospital in Jiangsu, and the patients signed informed consents for sample collection.

Main Reagents

Fecal DNA sample preservation tube, tissue nucleic acid preservation tube, fecal DNA extraction kit (immunomagnetic bead method), universal columnar genomic nucleic acid extraction kit, fast DNA library preparation kit, *etc.* were all provided by Cowin Bio., Jiangsu, while high-throughput kit (MGISEQ-2000RS) was provided by MGI, Shenzhen.

Samples sources

The subjects of this project are patients who were diagnosed with CRC for the first time and received surgery at Taizhou People's Hospital from January 2019 to January 2021. A total of 54 tumor tissue samples, 45 preoperative stool samples and 41 postoperative stool samples were collected. The normal control group consisted of healthy physical examination takers without malignancies (including CRC), intestinal polyps or inflammatory bowel disease, from whom 20 stool samples were collected. The benign control group comprised 20 patients who were diagnosed with intestinal polyps or inflammatory bowel disease by digestive endoscopy, and admitted to the Department of Gastroenterology, Taizhou People's Hospital for polypectomy or anti-inflammatory treatment. 20 formed stool samples were collected from patients who had not taken laxatives preoperatively or from patients before receiving anti-inflammatory therapies. The above samples were tested by NGS, unreliable results were filtered during analysis of genetic testing results based on biological information, and poor quality samples were eliminated (heterogeneous tumor tissues, atypical fecal specimens, little or degraded DNA in stools). Finally, effectively detected samples included 50 tumor tissues, 27 preoperative stools, 19 postoperative stools, 20 stools in the normal control group, and 20 stools in the benign control group.

Sample collection and processing

Stool samples: Patients were instructed to use the fecal DNA preservation tube correctly (tubes contained protective solution to prevent the degradation of nucleic acid, while steel beads played a mixing role). They were asked to take samples immediately

after natural defecation. Using sampling spoon located on the tube cap, typical formed stool about the size of soybean was excavated while avoiding spillover of protective solution and steel beads in the tube. After tightening the tube cap, the patients were asked to shake the tubes for approximately 30 s for uniform mixing. The samples were stored in a -80°C refrigerator for subsequent testing.

Tissue samples: The freshly isolated tumor tissues were quickly cut into pieces (< 0.5 cm in thickness) using sterile tissue scissors, and then the tissue pieces were soaked in the protective solution of preservation tubes (if the volume was excessively large, the protective solution would not easily penetrate the tissues, which could not protect the nucleic acid from degradation). After tightening the tube caps, the samples were stored in a 4–8°C refrigerator for subsequent testing.

DNA extraction from stool samples

The DNA in stool samples should be extracted strictly in accordance with the kit instructions (immunomagnetic bead method). Initially, the samples were homogenized, and then the precipitates (fecal particles) were removed, and the supernatants were retained for subsequent use. Different reagents were added to the corresponding wells of deep 96-well plates as per the instruction procedures, and then the deep well plates and magnetic sleeves containing the reagents were placed at the corresponding positions of nucleic acid extractor, followed by running of the stool extraction program. Under high salinity condition, DNA can bind to the surfaces of silica-coated Magbeads. After rinsing, high-purity DNA was eluted and stored in Buffer GE. The DNA-containing Buffer GE was transferred to an EP tube and stored under refrigeration for subsequent testing.

The table below details the reaction system. Relevant reagents were added to the corresponding wells of deep 96-well plates (wells 1–6 corresponded to one sample, wells 7–12 corresponded to another sample, and so forth)(see Table 1 for details).

The deep 96-well plates and corresponding magnetic sleeves were placed into the corresponding positions of automatic nucleic acid extractor, and then the fecal DNA extraction program was initiated. After running for about 20 minutes, the program was paused, and each 220 μ l of pre-prepared mixture (Magbeads PN 20 μ l, isopropanol 200 μ l) was added to the first well of each sample, followed by continuation of the program. The product in the last well of each sample was transferred to the centrifuge tube and stored at low temperature for subsequent testing.

DNA extraction from tissue samples

DNA was extracted from tissue samples strictly following the instructions of universal columnar genomic nucleic acid extraction kit. Initially, fresh tumor tissues were ground to small pieces and added with lysis buffer and proteinase K for complete digestion and lysis. Then, RNase was added to remove RNA. Since DNA was insoluble in absolute ethanol, the samples added with absolute ethanol were centrifuged to adsorb the DNA on the column. The residual impurities dissolved in absolute ethanol were discarded along with the liquid waste in the collection tube. Next, the collection tube was replaced with a new one, and the centrifugation was repeated several times. The DNA-containing adsorption column was placed at room temperature, and let stand to dry the absolute ethanol. The elution buffer was suspended and added to the adsorption column with a new collection tube, centrifuged, and then the DNA-containing eluate in the collection tube was collected and refrigerated for subsequent testing.

The nucleic acid extraction was carried out in Cowin Bio., Jiangsu.

Library construction and NGS detection

In accordance with the instructions of fast DNA library preparation kit, the hotspot mutation panels of 50 tumor-associated genes designed for specific genomic regions or targets were added initially based on the multiplex PCR library construction technology, and then multiplex PCR enrichment was performed on specific gene region

fragments. Next, the obtained target gene fragments were subjected to end repair, barcode adapter ligation and PCR amplification (Eastwin Scientific Equipment, Suzhou) to prepare a library. Afterwards, the library was quantified with Qubit 3.0 fluorescence quantifier (Thermo Fisher Scientific, Shanghai), and the fragment size of the library was analyzed with Bioptic Qsep 100 automatic nucleic acid analyzer (Bioptic Inc., Jiangsu), thereby evaluating the library quality. Finally, a high-throughput sequencer (MGI, Shenzhen) was used for sequencing, and the data splitting, denoising and comparison were accomplished in the Linux system to obtain the gene mutation results of various samples.

The library construction and NGS were carried out in Cowin Bio., Jiangsu.

Statistical analysis

Data were analyzed using SPSS 26.0 software. The gene mutation frequencies were expressed as the numbers of cases or percentages. Inter-group differences were examined by χ^2 and *Fisher exact tests*. $P < 0.05$ were considered statistically significant.

RESULTS

Clinical data analysis of CRC patients

Among the CRC patients corresponding to 50 tumor tissue samples, 24 were females (48%) and 26 were males (52%). The age span was large (35–84 years), and the median age was 59.5 years. According to the histopathological results of tumor tissues, there were 29 cases of rectal cancer (58%), 11 cases of left colon cancer + sigmoid colon cancer (22%), and 10 cases of right colon cancer (20%). Twelve cases (24%) were well-differentiated + well-to-moderately differentiated, 25 cases (50%) were moderately differentiated, and 12 cases (24%) were moderately-to-poorly differentiated + poorly differentiated. Among 19 cases (38%) of lymph node metastasis, only 1 case had distant metastasis (metastasis to the liver). Regarding pathological staging (according to eighth edition of AJCC TNM staging system), 33 cases (66%) were at stages I + II, and 17 cases (34%) were at stages III + IV (see Table 2 for details).

Gene mutations in various samples

As is clear from the NGS testing results of CRC specimens in Table 3, genes like TP53, APC, KRAS, PTEN, MET, HRAS and PDGFRA had high incidences of mutations in tumor tissues and preoperative stools, while in postoperative stools, the mutation frequencies of TP53, HRAS and PDGFRA were high. Despite high mutation frequencies of HRAS and PDGFRA in the three groups of samples, their mutations were mostly same sense mutations, meaning there were only base mutations, without changes in amino acids, which had no clinical significance.

As shown in Table 4, the PDGFRA, HRAS, KIT mutations occurred in only 3 of 20 stool samples in the normal control group. In the benign control group, only 4 stool samples had PDGFRA, HRAS, KIT and STK11 mutations.

Since every gene has numerous mutation sites, different site mutations have different meanings. We classified and summarized the results of gene sequencing, and queried them in the <https://www.ncbi.nlm.nih.gov/clinvar/>. After combining the start and stop positions of each gene mutation site, various gene mutation sites were classified into four types depending on their pathogenicity: "pathogenic", "benign", "unknown clinical significance" and "undetected in the system" (hereinafter referred to as "undetected"). "Benign" gene mutation sites imply a large number of mutations that can occur multiple times in the normal population, albeit less than 1% incidences. Mutation sites of "unknown clinical significance" indicate that such mutations are neither pathogenic nor benign, on whom the experimental conclusions in the literature reports are inconsistent. There is insufficient evidence to clearly classify them, and further exploration is required. "Undetected" mutation sites are identified by comprehensively checking multiple items such as the exons, the types of amino acid mutations, the start and end positions of mutations. In the case of non-conformity of any item, the corresponding mutation site is regarded as "undetected". In Table 3, the mutations of PTEN gene in tumor tissues were quite frequent, all of which were mutation sites of "unknown clinical significance". Among the 27 cases of preoperative

stools, 1 case had pathogenic mutation. Given the small case number, the possible cause of tumor heterogeneity was not ruled out. The MET gene exhibited high mutation frequencies in both tumor tissues and preoperative stools, all of which were "benign" mutations upon query. The four MET mutations in the postoperative stools were all "undetected" mutations. In Table 4, the mutation sites of KIT and STK11 in the benign control group were benign mutation sites, while the rest were all same sense mutations and had no clinical significance.

This project focuses on analyzing the **pathogenic mutation sites** of various genes.

Comparison between preoperative stool and tumor tissue results in CRC patients

As shown in Table 5, the TP53, APC, and KRAS genes had high incidences of pathogenic mutations in tumor tissues and preoperative stools among CRC patients. Despite high mutation frequency of PIK3CA in tumor tissues, its pathogenic mutation frequency in preoperative stools was only 3.7% (1/27). The possible causes of tumor heterogeneity or individual disparity could not be ruled out. Although other genes had pathogenic mutations, their mutation frequencies were rather low.

Comparison between preoperative/postoperative stool results of CRC patients and control stool results

For CRC patients, the pathogenic mutation incidences of APC, TP53 and KRAS in the preoperative stools were 11.1% (3/27), 37.4% (10/27) and 25.9% (7/27), respectively. There were mutations of PIK3CA, BRAF and MPL genes in the postoperative stools. Although they were pathogenic, only one case was found for each gene. The possible cause of tumor heterogeneity could not be ruled out. No pathogenic gene mutation sites were detected in stool samples from the normal and benign control groups (see Tables 6–8 for details).

According to a combination of Tables 5–8, the TP53, APC and KRAS genes all had high mutation frequencies in the preoperative stools and tumor tissues of CRC patients,

while exhibiting no pathogenic mutations in the postoperative stools, or in the stool samples of normal or benign control group. This indicates that tumor-specific DNA can be detected in the preoperative stools of CRC patients, and that the preoperative fecal expressions of tumor-associated genes can reflect the gene mutations in tumor tissue to some extents. Compared to the postoperative stools and the stool samples of the two control groups, the pathogenic mutation frequencies of TP53 and KRAS were significantly higher in the preoperative stools ($P < 0.05$), suggesting that fecal TP53 and KRAS genes can be used for the screening, diagnosis and prognostic prediction of CRC. Contrastively, the pathogenic mutation frequency of APC gene in the preoperative stools differed insignificantly from that in the postoperative stool or the two control groups ($P > 0.05$), which was probably associated with the excessively small sample size.

Mutation site analysis for TP53, KRAS and APC genes

As shown in Table 9, TP53 mutations occurred in 31 of 50 tumor tissues, and the total mutation frequency was 40 times, of which pathogenic mutation sites accounted for 60% (24/40), with exon5 R175H, exon7 R248Q/W and exon8 R273H/C being the most common. The proportion of pathogenic mutation sites in 29 tumor tissues with KRAS mutations was 90% (27/30), most of which were located in exon2, with G12D/V/S being predominant, followed by G13D/C. Additionally, there was 1 case of exon 3 Q61H mutation. As for the APC gene, all its pathogenic mutation sites were located in exon17.

12 of 27 TP53 mutations in preoperative stools were pathogenic mutations, which were mostly exon7 R248Q and exon8 R273H/C. The pathogenic mutation sites of KRAS gene were primarily exon2 G12D/V. As for APC gene, all its pathogenic mutation sites were located in exon17 (see Table 10 for details).

As are clear from Tables 9 and 10, the aforementioned pathogenic mutation sites occurred in both preoperative stools and tumor tissues, suggesting the possible correlation between preoperative stool and tumor tissue in CRC patients.

Combined TP53–KRAS detection

Among the 27 preoperative stools of CRC patients, 16 had pathogenic mutations of the TP53 gene, with a true positive rate of 59.26%, and 10 had pathogenic mutations of the KRAS gene, with a true positive rate of 37.04%. Pathogenic mutations of TP53 or KRAS occurred in 18 preoperative stools, showing a true positive rate of 66.67% (18/27). According to Table 11, the sensitivity and negative predictive value of the combined TP53–KRAS detection were 66.67% and 68.97%, respectively, which were higher than the TP53 or KRAS mutation detection alone. This suggests that the combined testing of TP53 and KRAS genes can improve the detection rate.

Distribution of mutation sites of "unknown clinical significance"

Mutation sites of "unknown clinical significance" occurred for such genes as TP53, PTEN, KRAS, BRAF and AKT1 in the tumor tissues of CRC patients. The five mutation sites of "unknown clinical significance" for TP53 were respectively exon5 A159V/V172G/G154D/P153S and exon7 S241T. In the case of PTEN, 9 of 11 mutations were of "unknown clinical significance", all of which were exon6 V133I. For the KRAS gene, the mutation sites of "unknown clinical significance" accounted for 10% (3/30), all of which were exon4 A146T. In preoperative stools, the mutation sites of "unknown clinical significance" were scattered in TP53, KRAS, AKT1 and STK11. In the case of TP53, such mutation sites were exon5 V172G and exon7 I251T/S241T, respectively, which were identical to the mutation locations in tumor tissues. For the KRAS gene, the mutation site of "unknown clinical significance" was exon4 A146T. As for AKT1 and STK11, such mutation sites were located separately in exon3 E17K and exon4 E199K.

Distribution of "undetected" mutation sites

As shown in Table 12, genes like TP53, APC, EGFR, FBXW7 and NOTCH1 in tumor tissues all had "undetected" mutation sites, among which exon4 A84G of TP53 occurred more frequently (4/40). For the APC gene, "undetected" mutation sites were all found on exon17. The pathogenic mutation sites of APC in Tables 9 and 10 were also located in exon17, suggesting that the "undetected" mutation sites of APC in Table 12 may be unreported new sites. They are probably associated with the carcinogenesis and progression of CRC, which require further exploration. Table 13 shows that 6 of 27 TP53 mutation sites in the preoperative stools were "undetected" sites. They were distributed in exon4, exon5, exon7 and exon8, among which there were three exon7 N247D (mutation start and stop positions were both 7577542). Additionally, genes like CDKN2A, HRAS, EGFR, IDH1 and PDGFRA also had "undetected" mutations sites.

DISCUSSION

Studies have shown that the genetic factors coupled with prolonged unfavorable external factors induce the disturbance of homeostasis, which leads to a series of changes including angiogenesis, cell hyperproliferation, apoptosis evasion and enhanced invasion capacity, ultimately resulting in the occurrence of malignancies. It is generally believed that CRC often develops slowly from adenomas, during which multiple steps are involved, ⁶ including the activation of proto-oncogenes and the inactivation of tumor suppressor genes. The mutation statuses of genes like TP53, APC and KRAS have been reported to be closely associated with the carcinogenesis and progression of CRC^[24,25]. With the introduction of "precision medicine" concept, molecular targeted therapy has become a hot research topic in recent years. This NGS-based research project investigates the value of fecal genetic testing in the CRC diagnosis.

As a tumor suppressor gene, TP53 regulates downstream genes when cells are under stress, which induces apoptosis and cell cycle arrest *via* a series of processes, thereby preventing the growth of tumor cells^[26,27]. Its mutation is a key step in facilitating the adenoma–adenocarcinoma transition. In malignancies, the p53 protein

encoded by TP53 is mutant. The mutant p53 protein can lose its tumor suppressor effect and acquire new oncogenic functions to promote the invasion and metastasis of tumor cells^[28]. In this study, TP53 had 62% (24/40) and 59.3% (16/27) mutations in the tumor tissues and preoperative stools of CRC patients, respectively, among which pathogenic mutation sites accounted for 60% (24/40) and 40.7% (11/27) of total mutation sites. Such pathogenic mutation sites as exon5 C176Y, exon6 R196*, exon7 R248Q, exon7 G245C/S and exon8 R273H/C existed in both types of samples, showing consistency. Such mutation hotspots as R273H, R248Q and R282W have been reported as typical "hotspots" in the DNA binding domain. Over 95% of missense TP53 mutations occur in the DNA binding domain, which can affect the ability of proteins to bind to their target DNA sequences^[29-31]. Lo *et al* screened a patient with CRC lymphatic metastasis for tumor infiltrating lymphocytes. Since some tumor patients had TP53 R175H mutations during genetic testing, and HLA-A*0201 could limit the recognition of mutant TP53 R175H, they identified the T cell receptors capable of recognizing the TP53 R175H mutation hotspots, thereby finding patients with different tumor types who shared the same mutation hotspots. In their opinion, the mutant TP53 could serve as an ideal common antigen target, and these specific hotspot mutations had potential as valuable therapeutic targets for a variety of tumor cells^[32]. A study by Olszewski suggested that changes in the oncogenic activity of TP53 Led to acquisition of new functions for the mutated tumor suppressor p53 protein, and that the overexpression of such mutation hotspots as codons 175, 248 or 273 might be associated with the chemotherapeutic resistance or invasive changes. Their analysis of TP53 mutation database in human somatic cells revealed that patients carrying TP53 R248Q allele had earlier tumor onsets, but lower frequencies of metastasis^[33].

The 21 kDa protein encoded by KRAS, a member of the ras proto-oncogene family, can participate in the G protein signaling pathway and regulate the proliferation and differentiation of cells. KRAS mutations can continuously activate the RAS-RAF-MAPK pathway, causing uncontrolled cellular proliferation and differentiation, thereby inducing CRC^[14]. Research has demonstrated that the patients' resistance to epidermal

growth factor receptor (EGFR)-targeted drugs was closely associated with the activation of KRAS gene^[34]. In the present study, the incidences of KRAS mutations in tumor tissue and preoperative stools of CRC patients were 58.0% (29/50) and 37.0% (10/27), respectively. Among them, pathogenic mutation sites existed in 27 tissues and 7 preoperative stools, with the most common ones being G12D/V and G13D located in exon2. Armengol *et al* explored the gene mutations in stool samples of 52 CRC patients. They found 12 cases of KRAS gene mutations, of which 58.3% (7/12) were located in codons 12 and 13 of exon 2^[35]. Knight argued that when there was a KRAS G12D carcinogenic mutation in the body, the formation or growth of intestinal adenomas was not inhibited by rapamycin^[36]. Compared to KRAS G12D, patients with KRAS A146T mutation could have stronger resistance to EGFR GFR tolerance, and better survival rates^[37]. Mulla *et al* analyzed the RAS mutations in 51 CRC patients, finding that 43% of them had RAS mutations, 91% of which were KRAS mutations. The majority of KRAS mutations were located in G12D, followed by G13D. They also found that the incidence of KRAS mutations was slightly higher among individuals under 50 years of age than that of the elderly (>50 years)^[38]. Ottaiano *et al* reported that mutant KRAS was a negative influencing factor of prognosis in CRC patients. Among these patients, those with G12C/S mutation showed the shortest survival and the worst prognosis^[39]. Varshavi *et al* investigated how KRAS mutations in different codons affected the metabolic pathways. According to their findings, metabolic differences existed between various KRAS mutation sites, which might be one of the reasons for the varying efficacies of anti-cancer therapies across patients. Hence, more effective, individualized molecular targeted drugs can be searched based on such metabolic differences^[40].

Mutations of APC were linked to familial adenomatous polyposis, and its inactivation and the activation of Wnt signaling pathway also played pivotal roles in the occurrence of CRC, which could affect the adhesion between tumor cells to accelerate the cellular migration^[41,42]. It has been reported that in sporadic CRC, most of the APC mutations occurred around the mutation cluster region, especially between codons 1286 and 1513^[43]. In this study, the incidences of APC mutations in tumor

tissues and preoperative stools were 24% (12/50) and 11.1% (3/27), respectively. As shown in Tables 5 and 6, the pathogenic mutation frequency of APC gene in preoperative stools differed insignificantly from that in control stools ($P=0.370$), which thus cannot serve as a specific gene for CRC screening and diagnosis. This may be attributed to the small sample size of this project, which remains further exploration. We found that all the pathogenic mutation sites of APC were located in exon 17, most of which used the mutations of certain amino acids as stop codons, thus allowing premature termination of mRNA translation (Tables 8, 9). Luo's team performed NGS-based gene sequencing on the blood and tissue samples of 22 CRC patients and 21 patients with intestinal polyps, in order to understand gene mutations. Through screening, they found that the APC mutations were most common in both types of samples, which were mostly pathogenic mutation sites. These pathogenic mutation sites of the APC gene accounted for 77.3% (17/22) of the total pathogenic mutation sites in all samples^[44].

In addition to the pathogenic mutation sites of various samples, this study also found multiple mutation sites of "unknown clinical significance", which were scattered in such genes as TP53, PTEN, KRAS, BRAF, AKT1 and PIK3CA. There were also some gene mutation sites that had not been found in the Clinvar database. TP53 exon4 A84G and EGFR exon20 I821T were present in both the preoperative stools and tumor tissues of CRC patients. The mutation start and stop positions were both 7579436 for the former, while 55249164 for the latter.

CONCLUSION

Conclusively, NGS-based fecal DNA detection can be used as a complementary technique for the CRC diagnosis. Fecal TP53 and KRAS can serve as a specific gene combination for the screening, diagnosis, prognostic prediction and recurrence monitoring of CRC. Moreover, the combined TP53-KRAS testing can improve the CRC detection rate, which provides a molecular reference for developing personalized treatment programs. The mutation sites of "unknown clinical significance" and the

"undetected" mutation sites may be unreported new mutation sites associated with CRC. Whether these genes are linked to the CRC diagnosis and prognosis remains to be further explored and researched.

ARTICLE HIGHLIGHTS

Research background

Colorectal cancer (CRC) is currently a health problem of global concern. In recent years, the incidence of CRC a trend of gradual increase. Most patients have unobvious early symptoms, and they are commonly in mid and advanced stages when the symptoms become evident, with rather high mortalities. Featuring high throughput, fastness and rich information, Next Generation Sequencing (NGS) can greatly shorten the detection time, which is a research hotspot at home and abroad at present.

Research motivation

As we all know, histopathological examination is the gold standard of diagnosis, but its invasiveness limits its development. Therefore, it is imperative to explore the screening, diagnosis and prognosis of CRC with strong specificity, high sensitivity and non-invasive methods.

Research objectives

In this project, NGS technology was used to conduct genetic testing on stool samples of CRC patients, and the results were compared with the corresponding tumor tissue genetic testing results. In order to find genes or gene combinations with high specificity and sensitivity in the stool. To establish a technical platform for CRC screening and diagnosis and curative effect monitoring through fecal DNA detection, providing a strong basis and support for personalized diagnosis and treatment of CRC.

Research methods

NGS was used to sequence the stools of patients with CRC, which were then compared with the stool samples of normal control and benign intestinal disease groups, as well as the tumor tissues of CRC patients. Specific genes or gene combinations in fecal DNA suitable for diagnosis and prognostic prediction of CRC were screened, and their significances in diagnosing CRC and predicting patients' prognosis were comprehensively evaluated.

Research results

High mutation frequencies of TP53, APC and KRAS were detected in the stools and tumor tissues of CRC patients prior to surgery. Contrastively, no pathogenic mutations of the above 3 genes were noted in the postoperative stools, or two control groups. This indicates that the tumor-specific DNA was detectable in the preoperative stools of CRC patients. Compared to the postoperative stools and the stools in the two control groups, the pathogenic mutation frequencies of TP53 and KRAS were significantly higher for the preoperative stools ($P < 0.05$), suggesting that fecal TP53 and KRAS genes can be used for CRC screening, diagnosis and prognostic prediction. No significant difference in the pathogenic mutation frequency of APC gene was found from the postoperative stools or the two control groups ($P > 0.05$), so further analysis with larger sample size is required. In 27 preoperative stools of CRC patients, the sensitivity negative predictive value of TP53、KRAS gene combination detection was higher than that of TP53 mutation alone or KRAS mutation, suggesting that TP53、KRAS gene combination detection can improve the detection rate of CRC. The "undetected" mutation sites found in preoperative stool and tumor tissue may be new mutation types in the occurrence and development of CRC, which need to be further studied. In addition, some mutations of "unknown clinical significance" were found, and their clinical value is worth further study.

Research conclusions

NGS-based fecal genetic testing can be used as a complementary technique for the CRC diagnosis. Fecal TP53 and KRAS can be used as specific genes for the screening, diagnosis, prognostic prediction and recurrence monitoring of CRC. Moreover, the combined testing of TP53 and KRAS genes can improve the CRC detection rate.

Research perspectives

Fecal genetic detection is a new method for CRC diagnosis, which has the advantages of non-invasive, convenient sampling and dynamic monitoring. Although the sensitivity of fecal genetic test in CRC screening is low, it is certain that it has great potential and broad prospects in the diagnosis and prognosis assessment of CRC. In addition, the "undetected" mutation sites in preoperative stool of CRC and the "unknown clinical significance" mutation sites are related to the occurrence and development of CRC, which requires further research and exploration.

2%

SIMILARITY INDEX

PRIMARY SOURCES

- 1

Siyu He, Chenglin Zhou, Hailin Peng, Mei Lin. "Recent advances in fecal gene detection for colorectal cancer diagnosis", Biomarkers in Medicine, 2021

Crossref

45 words — 1%
- 2

Qianqian Zhao, Luyu Zhang, Yingying Wang, Ye Sun et al. "A Bioinformatic Analysis: The Overexpression and Prognostic Potential of GPX7 in Lower-Grade Glioma", International Journal of General Medicine, 2022

Crossref

24 words — < 1%
- 3

www.x-mol.com

Internet

16 words — < 1%
- 4

id.nii.ac.jp

Internet

15 words — < 1%
- 5

Zhenjie Yin, Lifeng Zhu, Bingyong You, Yingming Sun et al. "What's The Next Era Of Immune Checkpoint Inhibitors Of Renal Cell Carcinoma In Clinical Practice: An Evidence From Extended Bibliometric Analysis", Research Square Platform LLC, 2022

Crossref Posted Content

12 words — < 1%
- 6

bgee.org

Internet

12 words — < 1%
- 7

www.ncbi.nlm.nih.gov

Internet

12 words — < 1%

8 www.spandidos-publications.com
Internet

12 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	< 12 WORDS
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 12 WORDS