Cancer Management and Research

Dovepress

open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Mutation spectrum of germline cancer susceptibility genes among unselected Chinese colorectal cancer patients

Rui Gong[1,2,*]
Yuan He[1,2,*]
Xiao-Yun Liu[1,2]
Hai-Yun Wang[1,2]
Li-Yue Sun[1,2]
Xin-Hua Yang[1,2]
Bin Li[3]
Xin-Kai Cao[3]
Zu-Lu Ye[1,2]
Ling-Heng Kong[1,4]
Da-Dong Zhang[3]
Yu-Hong Li[1,5]
Rui-Hua Xu[1,5]
Jian-Yong Shao[1,2]

[1]State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou 510060, People's Republic of China; [2]Department of Molecular Diagnostics, Sun Yat-sen University Cancer Center, Guangzhou 510060, People's Republic of China; [3]Research and Development Institute of Precision Medicine, 3D Medicine Inc., Shanghai, 201114, People's Republic of China; [4]Department of Colorectal Surgery, Sun Yat-sen University Cancer Center, Guangzhou, People's Republic of China; [5]Department of Medical Oncology, Sun Yat-sen University Cancer Center, Guangzhou, People's Republic of China

*These authors contributed equally to this work

Correspondence: Jian-Yong Shao
State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou 510060, People's Republic of China
+86 20 8734 5599
Email shaojy@sysucc.org.cn

Rui-Hua Xu
State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou 510060, People's Republic of China
Tel +86 20 8734 5689
Email xurh@sysucc.org.cn

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

Background: Genetic factors play an important role in colorectal cancer (CRC) risk, yet the prevalence and spectrum of germline cancer susceptibility gene mutations among unselected Chinese CRC patients is largely undetermined.

Methods: We performed next-generation sequencing with a 73-genes panel and analyzed the prevalence and spectrum of germline mutations in 618 unselected Chinese CRC patients. We classified all identified germline alterations for pathogenicity and calculated the frequencies of pathogenic mutations. Clinical characteristics were assessed by age and mutation status. Protein expressions and interactions of MLH1 missense variants were evaluated by western blot and co- immunoprecipitation.

Results: Overall, 112 (18.1%) of 618 unselected Chinese CRC patients were found to carry at least one pathogenic or likely pathogenic variant (totaling 97 variants), including 70 (11.3%) Lynch syndrome (LS) mutation carriers and 42 (6.8%) non-LS mutation carriers. LS mutation carriers were significantly younger at CRC diagnosis and were more likely to have right-sided, poorly differentiated, early stage, high-frequency microsatellite instability (MSI-H) or dMMR CRC and a family history of cancer compared with noncarriers. Non-LS mutation carriers were more likely to be proficient mismatch repair (pMMR) than noncarriers (p=0.039). We found four clinical variables (gender, tumor histological stage, cancer stage and mutation status) that showed significant differences between patients younger and older than 50 years old. Higher mutation rates were found in patients under 50 years old (p=0.017). Thirty-three novel variants were discovered and evaluated as pathogenic mutations by our study.

Conclusion: Given the high frequency and wide spectrum of mutations, genetic testing with a multigene panel should be considered for all Chinese CRC patients under 50 years old and is also needed to determine whether a gene is associated with CRC susceptibility and to promote clinical translation.

Keywords: genetic factor, germline mutations, Lynch syndrome, next-generation sequencing

Introduction

Colorectal cancer (CRC) is the third most common malignancy and the fourth leading cause of cancer-related mortality worldwide.[1] The majority of CRC cases are sporadic, but inherited factors contribute to approximately 30–35% of CRC cases.[2] Between 5–10% of CRCs are associated with high-risk mutations in known CRC susceptibility genes, predominantly the mismatch repair (MMR) genes (Lynch Syndrome, LS), *APC* (Familial Adenomatous Polyposis, FAP) and *MUTYH* (*MUTYH* - associated Polyposis, MAP).[3–5] Overall, mutation carriers have an increased risk of CRC (lifetime risk

30–70%).[6,7] Genetic factors play an important role in CRC risk and predisposition.[8,9] The identification of individuals at high risk for CRC, especially those who carry mutations in a CRC susceptibility gene, is important to provide various options for risk management and targeted screening for cancer prevention.[10]

While hereditary CRC has been a common indicator for germline genetic evaluation since MMR genes were identified, many additional genes have subsequently been implicated in CRC.[11] Patients who performed germline genetic testing for CRC were typically tested for a limited number of genes that were strongly associated with CRC, such as MMR genes, *APC* and *MUTYH*. With the advent of next-generation sequencing (NGS), germline genetic testing for CRC has shifted from a limited number of phenotype-specific gene assessments to broad panels with multiple genes indicating various hereditary cancer syndromes. Since NGS reduces costs and increases the capacity to analyze multiple genes in parallel, there is an opportunity to provide more information on a large number of genes, allowing for more accurate cancer surveillance and tailored prevention options.[12,13] The genes included in different panels vary, ranging from well-established cancer susceptible genes with quantifiable risk levels to less well-defined genes that are not traditionally associated with CRC.[12,13]

Herein we used NGS with a 73-genes panel associated with various hereditary cancer syndromes to determine the prevalence and spectrum of germline mutations in a consecutive series of 618 CRC patients. This may help to shape a more comprehensive understanding of genetic structure of CRC and generate accurate individualized risk management strategies for mutation carriers.[14]

Materials and methods

Patients and methods

We reviewed the genetic test results and clinical data from a consecutive series of 618 CRC patients evaluated by an NGS hereditary cancer panel between September 2014 and September 2017 at the Sun Yat-sen University Cancer Center (SYSUCC) (Guangzhou, China). All patients submitted a peripheral blood sample, from which germline DNA was isolated for clinical sequencing. Germline DNA was tested for mutations in 73 cancer susceptibility genes using the NGS method (gene list is detailed in Table 1). Patient demographics, medical history, family history (three generation), colonoscopy or resection findings, tumor location, tumor histology and phenotype, MMR status and CRC stage were obtained through review of medical records and through detailed patient inquiries. The SYSUCC ethical review board approved this study, and written informed consents were obtained.

Clinical genetic testing

NGS-based clinical sequencing of germline DNA for mutations was performed at Clinical Laboratory Improvement Amendments (CLIA)-certified commercial genetic testing laboratory (The Beijing Genomics Institute, BGI, Shenzhen, China) using sequencing by synthesis (SBS) in accordance with current practice standards. NGS was performed to detect single nucleotide variations (SNVs), copy number variants (CNV) and insert/deletion mutations (indels) within 20 bp for all exon regions as well as a portion of intronic regions (±10 bp) for all of the genes in the panel. The average depth was more than 300× in all samples and the coverage of the target regions was approximately 99.5%.

IHC and MSI analysis

MMR status were determined by immunohistochemistry (IHC) testing for *MLH1*, *MSH2*, *MSH6* and *PMS2* protein (antibody: Roche, 07862237001, 078622530, 07862245001, 07862261001) expression in tumor tissue. MSI status were tested using the Bethesda consensus panel (mononucleotide repeats BAT25 and BAT26, and dinucleotide repeats D2S123, D5S346, and D17S250) by multiplex fluorescent PCR and capillary electrophoresis. Tumors showing MSI at 0 marker were classified as microsatellite stable (MSS). Tumors showing MSI at 1 marker were classified as low-frequency microsatellite instability (MSI-L). Tumors showing MSI at 2 or more markers were classified as high-frequency microsatellite instability (MSI-H).

Data analysis

The outcomes of clinical genetic testing were analyzed by us with the assistance of 3D Medicines (Shanghai, China) according to the 2015 American College of Medical Genetics and Genomics (ACMG) standards and guidelines[15] for variant classification. Genetic variants were classified as pathogenic, likely pathogenic, variants of uncertain significance (VUS), likely benign or benign. Every variant was analyzed by at least a master- or PhD-level analyst and a PhD-level Fellow clinical molecular geneticist. We also referred to specific tools and resources include the Exome Sequencing Project (ESP),[16] The 1,000 Genomes Projects, The Exome Aggregation Consortium (Ex AC), The Single Nucleotide Polymorphism Database (dbSNP), ClinVar, SIFT[17] and Polyphen.[18]

Table 1 Genes analyzed by a multigene hereditary cancer panel

Syndrome	Associated gene(s)	Associated cancer spectrum
Colon cancer susceptibility genes		
Lynch syndrome (LS)	*MLH1, MSH2, MSH6, PMS2, EPCAM*	Colorectal, endometrial, ovarian, gastric, urothelial, pancreaticobiliary, cutaneous sebaceous neoplasms, brain
Familial adenomatous polyposis (FAP)	*APC*	Colorectal, small intestine, ampullary, gastric, desmoid, thyroid
MUTYH-associated polyposis (MAP)	*MUTYH*	Colorectal, duodenal
Juvenile polyposis syndrome (JPS)	*SMAD4, BMPR1A*	Colorectal, Gastric
Peutz–Jeghers syndrome (PJS)	*STK11*	Colorectal, breast, pancreatic, gastric, small intestine, cervical, ovarian
PTEN hamartoma tumor syndrome, Cowden syndrome	*PTEN*	Colorectal, breast, endometrial, thyroid, renal
Oligodontia-colorectal Cancer syndrome	*AXIN2*	Colorectal, breast cancer, neuroblastoma
Other cancer susceptibility genes		
Li-Fraumeni syndrome (LFS)	*TP53*	Breast, sarcoma, brain, adrenocortical, leukemia, gastric
Hereditary breast/ovarian cancer	*BRCA1, BRCA2*	Breast, ovarian, pancreatic, prostate, melanoma
Hereditary diffuse gastric cancer syndrome	*CDH1*	Gastric, breast
Familial atypical multiple-mole melanoma syndrome	*CDKN2A*	Melanoma, pancreatic
Gorlin syndrome	*PTCH1*	Skin, brain, breast
Bloom syndrome	*BLM*	Leukemia, lymphoma
Other genes	*ATM*	Breast, pancreatic
	CHEK2	Breast, prostate, lung, kidney, thyroid
	BARD1, BRIP1, MRE11A, NBN, PALB2, RAD50, RAD51C, RAD51, FANCA, FANCC, FANCD2, FANCE, FANCF, FANCG, ALK, AR, CDC73, CDK4, CDKN1B, CYLD, DICER1, EGFR, EXT1, EXT2, FH, FLCN, HNF1A, HRAS, HSD3B1, KIT, LMO1, MAX, MEN1, MET, MLH3, MPL, NF1, NF2, NTRK1, PDGFRA, PMS1, PRKAR1A, RB1, RET, SDHAF2, SDHB, SDHC, SDHD, SMARCB1, SUFU, TMEM127, TSC1, TSC2	

PCR amplification and sanger validation

To validate the results of NGS, sequence-specific PCR was used to amplify the target fragment with specific primers designed with PelPrimer. Total PCR volume was 25 μl, including12.5 μl 2×Taq PCR MasterMix (KT201, Tiangen Biotech, Beijing), 1 μl template DNA, 9.5 μl ddH2O, 1 μl PCR Primer F and 1 μl PCR Primer R. The first step in the PCR was performed at 94 °C for 3 min, followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s, extension at 72 °C for 1 min and a final extension at 72 °C for 5 min. Agarose gel electrophoresis and a DNA purification kit were used to isolate and purify the PCR products. Sanger sequencing with an ABI 3500 (Applied Biosystems 3500) genetic analyzer was used to validate all of the pathogenic and likely pathogenic variants. Finally, the sequences were aligned to the Nucleotide Database with NCBI Blast online software.

Statistical analysis

The primary outcome of this study was the detection of cancer susceptibility gene mutations. Information related to patient numbers and demographics were presented using descriptive statistics. Data for qualitative variables were reported as percentages. The association of mutation status with clinical characteristics was analyzed using the chi-squared test or Fisher's exact test (when cells have an expected count of less than 5) to determine *p*-values (qualitative variables) and level of significance was set at 0.05. Patient age at CRC diagnosis was analyzed as a continuous variable and assessed by the two-sample *t*-test, reported as median and range. Wilson score intervals with continuity correction were used to compute confidence intervals. Statistical analysis was performed using SPSS 20.0.

Site-directed mutagenesis and expression plasmid construction

pcDNA3.1B was kindly provided by Dr Xiao-Feng Zhu (State Key Laboratory of Oncology in Southern China, Cancer Center, Sun Yat-sen University). To analyze protein expression in human cells, human h*MLH1/MSH2/MSH6* cDNA was synthesized by PCR and then cloned into the plasmid pcDNA3.1B by double enzyme digestion (EcoR-I/BamH-I restriction enzyme from NEB) and homologous recombination (CloneExpress II One Step Cloning Kit, C112, Vazyme Biotech, Nanjing). Selected *MLH1* variants were constructed by site-directed mutagenesis using a PCR-based protocol. In addition, eight plasmids were constructed and tagged with Flag protein (N-DYKDDDDK-C): *MLH1*-WT, *MLH1* c.1153C>T (p.R385C), *MLH1* c.1230_1232delinsTG (p.I411Vfs*80), *MLH1* c.1573T>G (p.L525V), *MLH1* c.1713delT (p.F571Lfs*2), *MLH1* c.1866delT (p.A623Qfs*14), *MLH1* c.278delG (p.S93Ifs*15) and *MLH1* c.522delG (p.I176Ffs*26). pcDNA3.1B (empty vector) used as a negative control and *MLH1*-WT was used as a positive control.

HEK293T cell culture and transfection

HEK293T cell line was purchased from the Stem Cell Bank, Chinese Academy of Sciences (Shanghai, China). HEK293T cells were cultured in DMEM with 10% fetal bovine serum (FBS) at 37 °C in a humidified 5% CO_2 atmosphere. HEK293T cells were transfected at 60–70% confluence with expression plasmids using Lipofectamine 3,000 reagent (L3000015, Invitrogen, Shanghai) according to the manufacturer's instructions. After 48 h, the cells from each group were harvested for western blot.

Western blot analysis and co-immunoprecipitation

Cells were harvested and washed with phosphate-buffered saline (PBS). The lysates were obtained with RIPA lysis buffer (containing 1 mM PMSF) followed by centrifugation (4 °C, 12,000 g, 15 min). Total protein concentrations in the supernatant were examined using an Enhanced BCA Protein Assay Kit (Beyotime Biotechnology, P0010, China). For co-immunoprecipitation, 1 mg of lysate was incubated with anti-Flag Ab or control mouse IgG overnight at 4 °C with rotation and then for 2 h at 4 °C with 20 µl Protein A/G PLUS-Agarose (Santa Cruz Biotechnology). Immunoprecipitates were collected by centrifugation and washed five times with ice-cold PBS. After the final wash, the supernatant was discarded and the pellet was resuspended in SDS lysis buffer, then boiled in 5× SDS loading dye for 5 min. Western blot was performed according to the standard procedure. Briefly, proteins were normalized to 25 µg/lane and loaded on a 10% SDS-polyacrylamide gel for electrophoresis and then transferred to PVDF membranes. The membranes were blocked with 5% BSA in TBST (10 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.1% Tween-20) for 1 h. The membranes were incubated with primary antibodies (anti-*MLH1*, CST3515T, mouse, 1:1,000; anti-*PMS2*, Abcam, ab110638, rabbit, 1:1,000 and anti-β-actin, 60008-1-Ig, Proteintech, mouse, 1:1,000) overnight at 4 °C, then washed three times with TBST for 10 min each time and incubated with horseradish peroxidase (HRP)-conjugated secondary antibody (Santa Cruz Biotech) at room temperature for 1 h. After washing the secondary antibody, the bands in the membrane were detected using chemiluminescence ECL kit. Finally, the membranes were processed using an enhanced chemiluminescence detection system.

Results

Clinical characteristics

A total of 77.8% (481/618) CRC patients were from south China in the Pearl River Valley and 12.5% (n=77) of them were from Yangtze River Valley and the rest were from other catchment areas. Patient characteristics including gender, age at CRC diagnosis, family history of cancer, primary tumor site, tumor histological stage, cancer stage,

MMR and MSI status are summarized in Table 2. Men accounted for 59.2% (n=366) of patients and 44.7% (n=276) had a family history of cancer. The mean age at CRC diagnosis was 49.8 years and 48.7% (n=301) patients were diagnosed before 50 years old.

A total of 82.7% (n=511) CRC patients had tumors with MMR IHC data available, of which 27% (n=167) were MMR deficient (dMMR). The most common pattern of MMR protein loss at all sites was *MLH1±PMS2*-. 45% (n=278) patients had tumors with MSI data available, of which 10.7% (n=66) were MSI-H. Most tumor screening results (MSI and IHC) were consistent for each patient, except for 12 patients (2 MSI-H, pMMR and 10 MSS/MSI-L, dMMR).

According to the clinical characteristics analysis by age (Table 2), we found 4 clinical variables (gender, tumor histological stage, cancer stage, mutation status) that showed significant differences between patients younger and older than 50 years old. There were more male patients than females in both groups, and the difference was more significant in patients over 50 years old. 20.3% (n=61) of patients under 50 and 13.3% (n=42) of patients over 50 were diagnosed with CRC at stage IV. Patients under 50 tended to have more poorly differentiated tumor than patients over 50 years. Patients under 50 years old were significantly more likely to carry pathogenic or likely pathogenic mutations than patients older than 50 (21.9% vs 14.5%, respectively). LS mutation carriers account for 14.3% (n=43) and non-LS mutation carriers account for 7.6% (n=23) of 301 patients under 50 years old.

Germline findings

Overall, 460 patients carried at least one pathogenic or likely pathogenic or VUS mutation (Figure 1A and C). A total of 596 VUS germline variants were detected among 419 patients (67.8%, 95% CI, 63.9–71.4%). The genes most likely to have a VUS variant discovered included *ATM* (n=42), *FANCA* (n=37), and *BRCA2* (n=31) (Figure 1B). The highest VUS frequency was observed in the *ATM* gene. A total of 97 pathogenic and likely pathogenic germline variants were detected among 112 (18.1%; 95% CI, 15.2–21.4%) of the 618 patients, including 80 (12.9%; 95% CI, 10.5–15.9%) with high-penetrance mutations (one with a concurrent moderate-penetrance mutation) and 19 (3.1%; 95% CI, 1.9–4.8%) with only moderate-penetrance mutations. Additionally, 33 pathogenic or likely pathogenic mutations were newly discovered by our study which were unreported in public data bases (Table 3). They were not listed in ClinVar 20150330, dbSNP 138 databases or other literatures and their frequency in the databases of 1,000 Genomes Projects 2015 Aug, ESP 6500 and Ex AC is 0. The rest of the pathogenic and likely pathogenic germline mutations were detailed in Table S1. Evidence of pathogenicity were according to the 2015 American College of Medical Genetics and Genomics (ACMG) standards and guidelines.

Seventy patients (11.3%, 95%CI 9% to 14.2%) carried LS mutations (36 *MLH1*, 23 *MSH2,* 7 *MSH6*, and 3 *PMS2*, 2 deletions of the 3′-end of *EPCAM*, including one patient with both *MSH2* and *EPCAM* mutations) and 42 (6.8%, 95% CI, 5–9.15%) carried non-LS mutations (including one patient with both a LS and a non-LS mutations, *MLH1*/*BLM*). Thirteen patients (2.1%; 95% CI, 1.2–3.7%) carried high penetrance non-LS mutations (7 *APC*, 1 *PTCH1*, 1 *PTEN*, 1 *TP53*, 1 *BRCA1*, 1 *BRCA2* and 1 biallelic *MUTYH*). Twenty patients (3.2%; 95% CI, 2–5.1%) carried mutations in moderate-penetrance genes (5 *ATM*, 3 *BARD1*, 4 *BLM*, 2 *BRIP1*, 1 *CDK4*, 4 *CHEK2* and 1 Monoallelic *MUTYH*). The rest carried mutations in less well-defined genes. The spectrum of pathogenic and likely pathogenic germline mutations is shown in (Figure 2). 61 (87.1%; 95% CI, 76.5–93.6%) of 70 LS mutation carriers demonstrated MSI-H and/or dMMR (2 pMMR or MSS; 6 had missing MSI/MMR data). All the pathogenic and likely pathogenic mutations detected by NGS were validated by Sanger sequencing or qPCR.

In the clinical characteristics analysis of mutation carriers compared with noncarriers, LS mutation carriers were significantly younger at CRC diagnosis and more likely to have right-sided, poorly differentiated, early stage, MSI-H, dMMR CRC and family history of cancer compared with noncarriers (Table 4). Non-LS mutation carriers were more likely to be pMMR than noncarriers. Age at CRC diagnosis, gender, primary tumor site, tumor histological stage, cancer stage and family history of cancer was not significantly associated with the presence of a non-LS mutation. However 23 (54.8%) of 42 non-LS mutation carriers and 40 (59.7%) of LS mutation carriers were given a diagnosis of CRC under 50 years old. Eleven (26.2%) of 42 non-LS mutation carriers were given a diagnosis of CRC at age ≥50 years old and lacked a family history of cancer.

Table 2 Clinical characteristics of 618 CRC patients

Characteristics		Total evaluable cohort No. (%)	Age at diagnosis		*p*-value
			< 50 (%)	≥50 (%)	
No. of patients		618	301	317	
Gender					0.005*
	Male	366 (59.2)	161 (53.5)	205 (64.7)	
	Female	252 (40.8)	140 (46.5)	112 (35.3)	
Age at CRC diagnosis, years					
	Mean (SD)	49.8 (12.2)	39.5(6.9)	59.6(6.8)	
Family history of cancer					0.595
	No	323 (52.3)	159 (52.8)	165 (52.0)	
	Yes	276 (44.7)	128 (42.5)	147 (46.4)	
	Unkonwn	19 (3.0)	14 (4.7)	5 (1.6)	
Primary tumor site					0.647
	Right sided colon cancer	184 (29.8)	96 (31.9)	88 (27.8)	
	Left sided colon cancer	191 (30.9)	93 (30.9)	98 (30.9)	
	Rectal cancer	193 (31.2)	86 (28.6)	107 (33.7)	
	Multiple primary colorectal cancer	41 (6.6)	21 (7.0)	20 (6.3)	
	Unknown	9 (1.5)	5 (1.6)	4 (1.3)	
Tumor histological stage					0.012*
	Well differentiated adenocarcinoma	21 (3.4)	14 (4.7)	7 (2.2)	
	Moderately differentiated adenocarcinoma	495 (80.1)	224 (74.4)	271 (85.5)	
	Poorly differentiated adenocarcinoma	71 (11.5)	43 (14.3)	28 (8.8)	
	Unknown	31 (5.0)	20 (6.6)	11 (3.5)	
Cancer stage					0.02*
	0	19 (3.1)	12 (4.0)	7 (2.2)	
	I	51 (8.2)	18 (6.0)	33 (10.4)	
	II	192 (31.1)	84 (27.9)	108 (34.1)	
	III	228 (36.9)	110 (36.5)	118 (37.2)	
	IV	103 (16.7)	61 (20.3)	42 (13.3)	
	Unknown	25 (4.0)	16 (5.3)	9 (2.8)	
MMR status					0.062
	dMMR	167 (27.0)	88 (29.2)	79 (24.9)	
	pMMR	344 (55.7)	151 (50.2)	193 (60.9)	
	Unknown	107 (17.3)	62 (20.6)	45 (14.2)	
MSI					0.173
	MSS	206 (33.3)	97 (32.2)	109 (34.4)	
	MSI-L	6 (1.0)	5 (1.7)	1 (0.3)	
	MSI-H	66 (10.7)	35 (11.6)	31 (9.8)	
	Unknown	340 (55)	164 (54.5)	176 (55.5)	

(*Continued*)

Table 2 (Continued).

Characteristics		Total evaluable cohort No. (%)	Age at diagnosis		*p*-value
			< 50 (%)	≥50 (%)	
Mutation status					0.017*
	Yes	112	66 (21.9)	46 (14.5)	
	No	506	235 (78.1)	271 (85.5)	

Notes: Unknown data were excluded from analysis in each group. The mutation status refers to carrying pathogenic or likely pathogenic mutations. * Statistical significance.
Abbreviations: CRC, colorectal cancer; MMR, mismatch repair; dMMR, MMR deficient; MSI, microsatellite instability; MSI-L, low-frequency microsatellite instability; MSI-H, high-frequency microsatellite instability; pMMR, proficient mismatch repair.

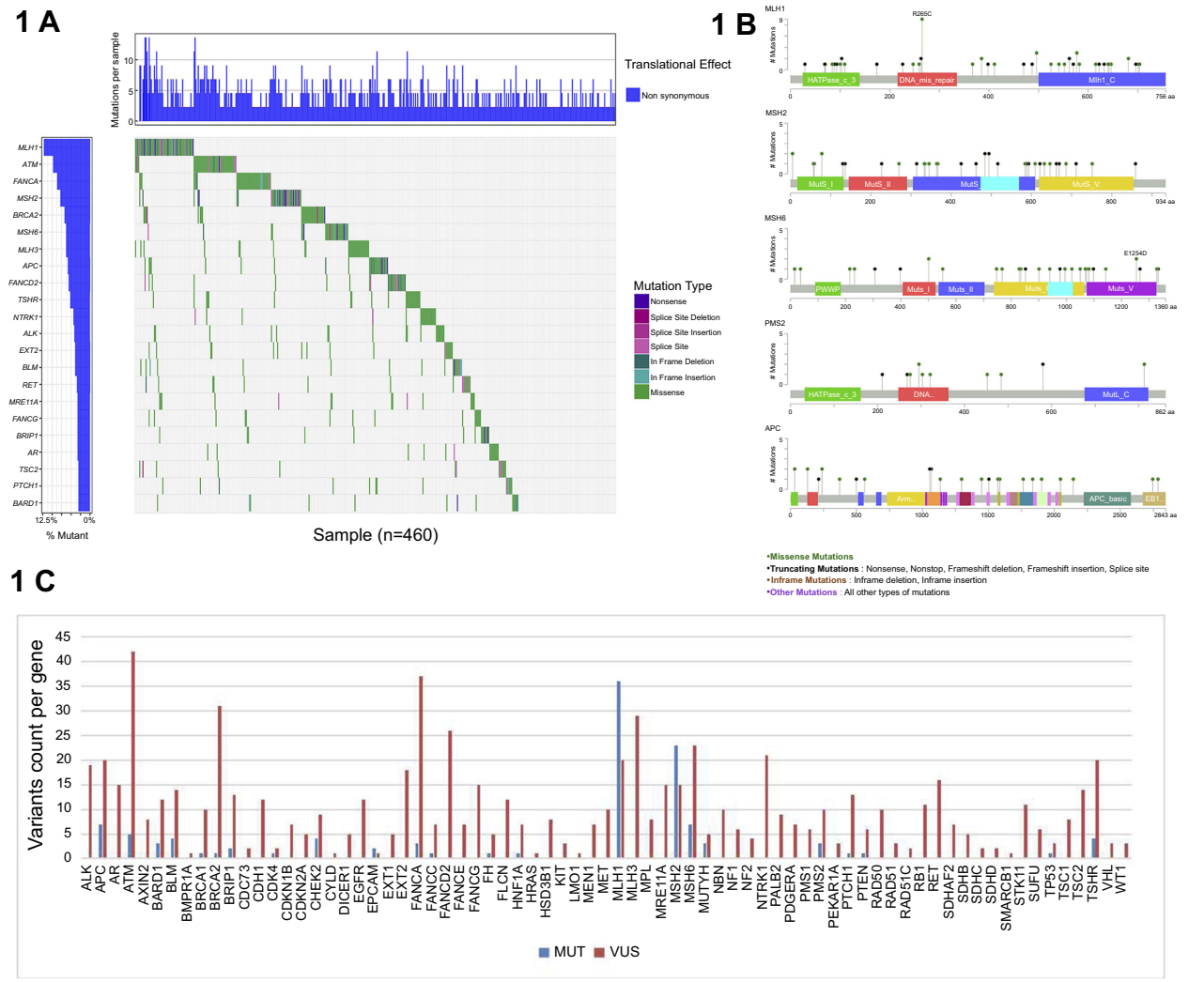

Figure 1 The mutations found in 618 CRC patients. (**A**) Heatmap of genes (mutation frequency >3%) with germline non-benign variants identified among 618 unselected colorectal cancer patients. Each column represents a patient, and each row represents a gene with multiple germline variants. (**B**) A lollipop diagram of germline non-benign variants of *MLH1*, *MSH2*, *MSH6*, *PMS2* and *APC* genes identified among 618 unselected colorectal cancer patients. (**C**) The number of MUT and VUS variants per gene detected with a multigene panel among 618 unselected colorectal cancer patients. The MUT group includes both pathogenic and likely pathogenic variants. Abbreviations: MUT, mutation; CRC, colorectal cancer ; VUS, variants of uncertain significance.

Pedigree analysis

Two families that had a typical family cancer history and novel mutations were taken into the study after informed consent. Analysis of the candidate variant was performed in additional family members by using NGS or Sanger sequencing.

Table 3 Summary of novel pathogenic germline mutations

Gene	Transcript_Ref	Chr	Chr_start	Chr_end	C_dot	P_dot	Mutation_type	Evidence	Count
APC	NM_000038	chr5	112174454	112174455	c.3164_3165del	p.I1055Nfs*4	Frameshift	1 PVS1+1 PM2+1 PM6	1
APC	NM_000038	chr5	112174456	112174456	c.3166del	p.I1056*	Frameshift	1 PVS1+1 PM2+1 PM6	1
ATM	NM_000051	chr11	108216477	108216478	c.8431_8432del	p.K2811Vfs*3	Frameshift	1 PVS1+1 PM2+1 PM6	1
BARD1	NM_000465	chr2	215593664	215593664	c.2069dup	p.N690Kfs*4	Frameshift	1 PVS1+1 PM2+1 PM6	1
BLM	NM_000057	chr15	91308540	91308540	c.2093_2094dup	p.Y699Vfs*19	Frameshift	1 PVS1+1 PM2+1 PM6	1
BLM	NM_000057	chr15	91347516	91347516	c.3678C>A	p.C1226*	Nonsense	1 PVS1+1 PM2+1 PM6	1
BRCA1	NM_007294	chr17	41246594	41246595	c.953delinsTGT	p.H318Lfs*24	Frameshift	1 PVS1+1 PM2+1 PM6	1
FANCA	NM_000135	chr16	89806404	89806405	c.3931_3932del	p.S1311*	Frameshift	1 PVS1+1 PM2+1 PM6	1
FANCA	NM_000135	chr16	89806417	89806417	c.3918dup	p.Q1307Sfs*6	Frameshift	1 PVS1+1 PM2+1 PM6	2
FANCC	NM_000136	chr9	97869503	97869504	c.1377_1378del	p.S459Rfs*58	Frameshift	1 PVS1+1 PM2+1 PM6	1
MLH1	NM_000249	chr3	37035125	37035131	c.87_93del	p.N30Sfs*4	Frameshift	1 PVS1+1 PM2+1 PM6	1
MLH1	NM_000249	chr3	37042500	37042501	c.263_264del	p.F88*	Frameshift	1 PVS1+1 PM2+1 PM6	1
MLH1	NM_000249	chr3	37042516	37042516	c.278del	p.S93Ifs*15	Frameshift	1 PVS1+1 PM2+1 PM6	1
MLH1	NM_000249	chr3	37045892	37045901	c.309_318del	p.L104*	Frameshift	1 PVS1+1 PM2+1 PM6	1
MLH1	NM_000249	chr3	37045894	37045894	c.311del	p.L104Wfs*4	Frameshift	1 PVS1+1 PM2+1 PM6	1
MLH1	NM_000249	chr3	37050371	37050371	c.522del	p.I176Ffs*26	Frameshift	1 PVS1+1 PM2+1 PM6	1
MLH1	NM_000249	chr3	37067318	37067321	c.1230_1232delinsTG	p.I411Vfs*80	Frameshift	1 PVS1+1 PM2+1 PM6	1
MLH1	NM_000249	chr3	37083801	37083801	c.1713del	p.F571Lfs*20	Frameshift	1 PVS1+1 PM2+1 PM6	1
MLH1	NM_000249	chr3	37090036	37090036	c.1926dup	p.I643Dfs*2	Frameshift	1 PVS1+1 PM2+1 PM6	1
MSH2	NM_000251	chr2	47630505	47630505	c.176del	p.K59Rfs*5	Frameshift	1 PVS1+1 PM2+1 PM6	1
MSH2	NM_000251	chr2	47693833	47693833	c.1547del	p.S516Mfs*10	Frameshift	1 PVS1+1 PM2+1 PM6	1
MSH2	NM_000251	chr2	47698192	47698194	c.1751_1752delinsA	p.I584Nfs*6	Frameshift	1 PVS1+1 PM2+1 PM6	1
MSH2	NM_000251	chr2	47702168	47702168	c.1764T>A	p.Y588*	Nonsense	1 PVS1+1 PS1+1 PM2+1 PM6	1
MSH2	NM_000251	chr2	47702409	47702409	c.2005del	p.G669Afs*16	Frameshift	1 PVS1+1 PM2+1 PM6	1
MSH2	NM_000251	chr2	47703557	47703557	c.2058dup	p.L687Tfs*12	Frameshift	1 PVS1+1 PM2+1 PM6	1
MSH2	NM_000251	chr2	47707952	47707982	c.2577_2606delinsT	p.E859Dfs*13	Frameshift	1 PVS1+1 PM2+1 PM6	1
MSH6	NM_000179	chr2	48025816	48025816	c.695dup	p.P233Afs*2	Frameshift	1 PVS1+1 PM2+1 PM6	1
MSH6	NM_000179	chr2	48026038	48026038	c.916G>T	p.G306*	Nonsense	1 PVS1+1 PM2+1 PM6	1
MSH6	NM_000179	chr2	48026314	48026314	c.1192del	p.V398Cfs*13	Frameshift	1 PVS1+1 PM2+1 PM6	1
MSH6	NM_000179	chr2	48027609	48027609	c.2488dup	p.S830Ffs*17	Frameshift	1 PVS1+1 PM2+1 PM6	1
MSH6	NM_000179	chr2	48027676	48027677	c.2554_2555del	p.K852Efs*5	Frameshift	1 PVS1+1 PM2+1 PM6	1
MSH6	NM_000179	chr2	48030680	48030680	c.3294C>A	p.C1098*	Nonsense	1 PVS1+1 PM2+1 PM6	1
PTCH1	NM_001083603	chr9	98279069	98279069	c.33dup	p.V12Cfs*27	Frameshift	1 PVS1+1 PM2+1 PM6	1

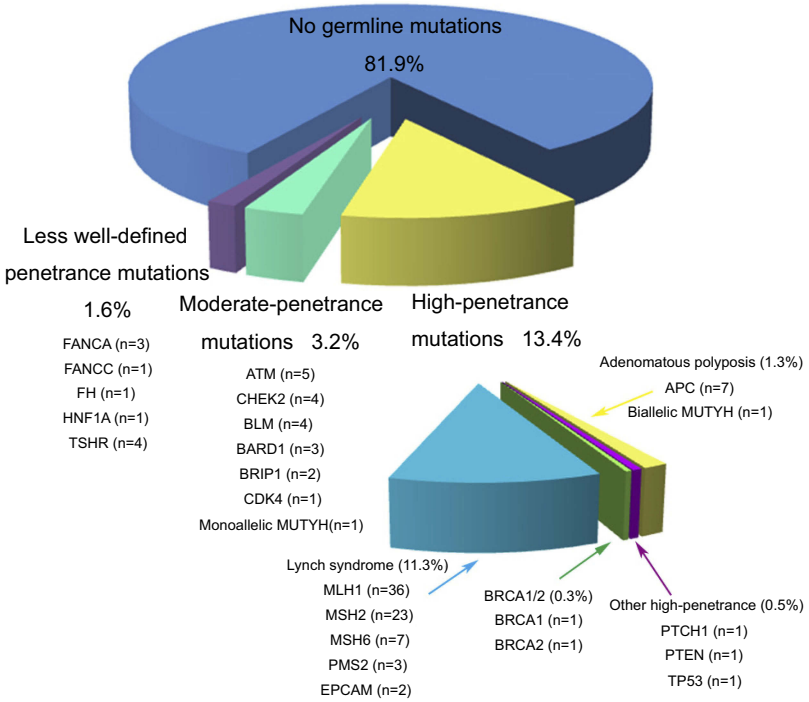

Figure 2 The identified pathogenic and likely pathogenic germline mutations. There was one patient with both *MSH2* and *EPCAM* mutations and one patient with both *MLH1* and *BLM* mutations.

In family W, the proband was diagnosed with sigmoid colon cancer at 48 years old (pMMR/MSI unknown) and ovarian cancer at 61 years old. The patient carried 3 mutations of uncertain significance (*MUTYH* c.924 +7C>T, *MLH1* c.T1573G, *EGFR* c.A1703G). The patient's father, mother and 4 siblings were diagnosed with different types of cancer (Figure 3A).

In family T, the proband was diagnosed with ascending colon cancer at 43 years old. Immunohistochemistry showed a lack of staining of *MLH1* and *PMS2* proteins. The MSI status was unknown. The patient carried one novel pathogenic mutation (*MLH1* c.278delG). Eleven family members in 3-generations were diagnosed with CRC (Figure 3B).

The mutation status of the pedigrees suggests that these novel mutations may be pathogenic (supporting evidence, PP4).

In vitro protein analysis of *MLH1* missense variants

Expression was determined by transient transfection of mutated plasmid (Characterization of constructed *MLH1* mutated plasmids were showed in Figure 3D) into HEK293T cells, which do not express the endogenous *MLH1* gene due to promoter hyper-methylation.[19] The c.1230_1232delinsTG (p.I411Vfs*80), c.1713delT (p.F571Lfs*2), , c.278delG (p.S93Ifs*15) and c.522delG (p.I176Ffs*26) mutations were novel *MLH1* variants that were unreported in public databases and were classified as pathogenic mutations by our study. The MLH1 c.1866delT (p.A623Qfs*14) mutation were pathogenic MLH1 variant that has been previously reported. The c.1153C>T (p.R385C) and c.1573T>G (p.L525V) mutations were VUS *MLH1* variants that have been previously reported. The c.1230_1232delinsTG, c.278delG and c.522delG variants showed a strong reduction of *MLH1* expression, and the c.1713delT and c.1866delT variants showed protein truncation at approximately 65 kd and 70 kd. PMS2 was not properly co-expressed in the truncated c.1713delT, c.1866delT and c.522delG *MLH1* variants since *PMS2* was destabilized on the protein level in the absence of proper dimerization with the *MLH1* C-terminus, which was absent in these variants. Immunoprecipitation experiment showed that the truncating variants *MLH1* c.1713delT, c.1866delT, c.278delG and c.522delG inhibited the interaction between *MLH1* and *PMS2*. In contrast, we found that the c.1153C>T and c.1573T>G variants had no effect on protein expression and interactions (Figure 3C).

The function analysis suggests that *MLH1* c.1230_1232delinsTG (p.I411Vfs*80), c.1713delT

Table 4 Clinical characteristics analysis of mutation carriers compared with noncarriers

Characteristic		Noncarriers	LS mutation carriers		Non-LS mutation carriers	
		No. (%)	No. (%)	*p*-value	No. (%)	*p*-value
No. of patients		506	70		42	
Gender				0.487		0.666
	Male Female	296(58.5) 210(41.5)	44(62.9) 26(37.1)		26(61.9) 16(38.1)	
Age at diagnosis						
	Mean ± SD <50 y ≥50 y	50.5±12.0 235 (46.4） 271 (53.6）	45.3±10.4 43(61.4) 27(38.6)	0.001* 0.019*	49.4±15.4 23 (54.8） 19 (45.2）	0.666 0.299
Family history of cancer				<0.001*		0.238
	Yes No Unknown	211(41.7) 281(55.5) 14(2.8)	44(62.9) 23(32.9) 3(4.3)		21 (50.0) 19 (45.2) 2 (4.8)	
Primary tumor site				<0.001*		0.122
	Right sided Left sided Rectal Multiple Unknown	136(26.9) 166(32.8) 176(34.8) 22(4.3) 6(1.2)	37(52.9) 10(14.3) 7(10.0) 14(20.0) 2(2.9)		11 (26.2) 15 (35.7) 10 (23.8) 5 (11.9) 1(2.4)	
Tumor histological stage				0.02*		0.375
	Well differentiated Moderately differentiated Poorly differentiated Unknown	17(3.4) 412(81.4) 52(10.3) 25(4.9)	1(1.4) 51(72.9) 15(21.4) 3(4.3)		3(7.1) 32(76.2) 4(9.5) 3(7.1)	
Cancer stage				0.023*		0.39
	0 I II III IV Unknown	15(3.0) 42(8.3) 145(28.7) 198(39.1) 86(17.0) 20(3.9)	1(1.4) 5(7.1) 34(48.6) 18(25.7) 9(12.9) 3(4.3)		3 (7.1) 4 (9.5) 13 (31.0) 12 (28.6) 8 (19.0) 2 (4.8)	
MMR status				<0.001*		0.039*
	dMMR pMMR Unknown	103(20.4) 309(61.1) 94(18.6)	61(87.1) 5(7.1) 4(5.7)		3(7.1) 30 (71.4) 9 (21.5)	
MSI				<0.001*		0.242
	MSS MSI-L MSI-H Unknown	183(36.2) 6(1.2) 40(7.9) 277(54.7)	2(2.9) 0(0.0) 25(35.7) 43(61.4)		21 (50.0) 0 (0.0) 1(2.4) 20(47.6)	

Notes: *p*-values were used for comparisons of mutation carriers to noncarriers. One patient with both an LS and a non-LS mutation (*MLH1*/*BLM*) was included in the LS carriers group. Unknown data were excluded from analysis in each group. *Statistical significance.

Abbreviations: LS, Lynch syndrome; MMR, mismatch repair; dMMR, MMR deficient; MSI, microsatellite instability; MSI-L, low-frequency microsatellite instability; MSI-H, high-frequency microsatellite instability.

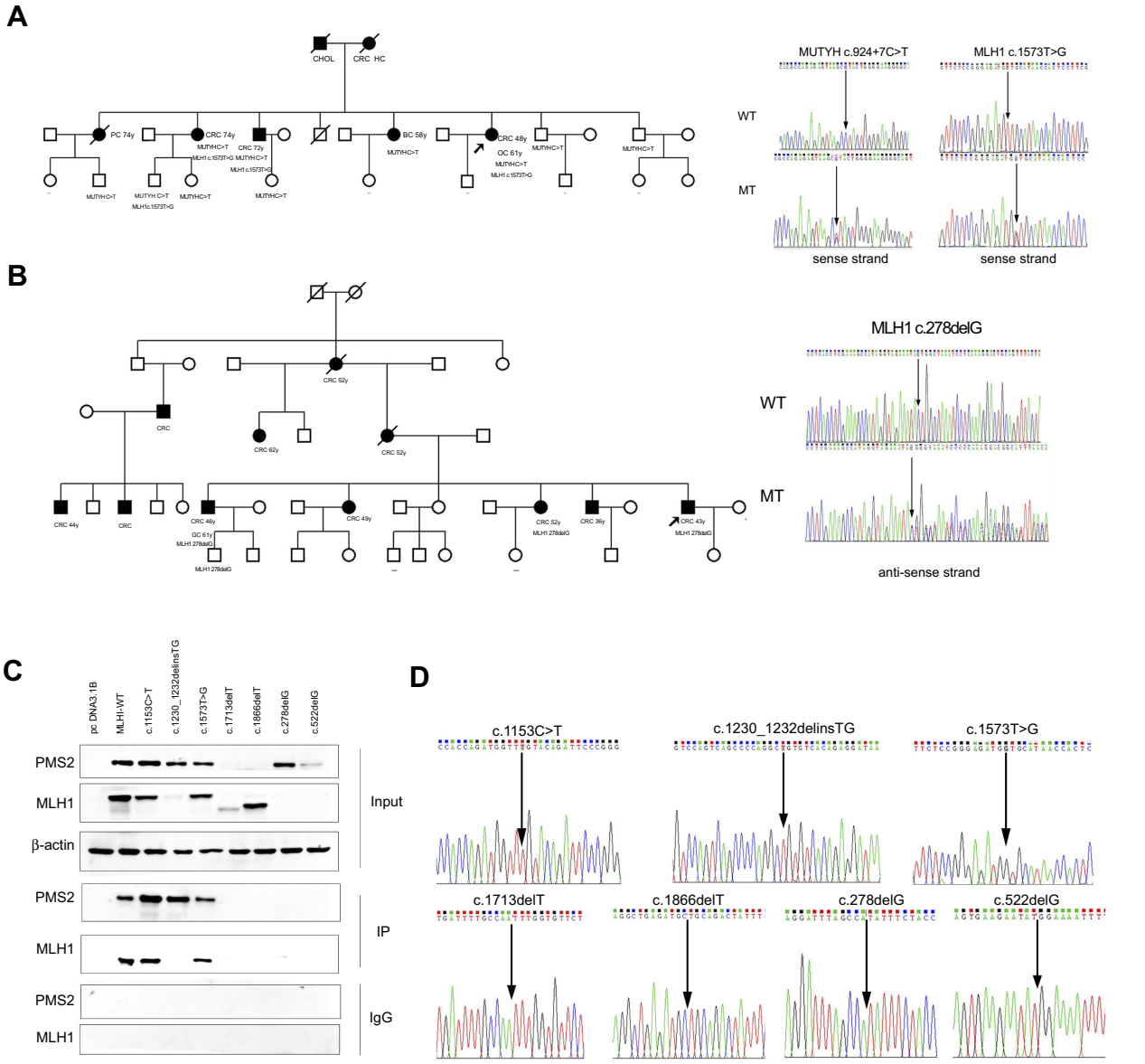

Figure 3 Evidence to identify germline variants for pathogenicity. (**A**) Pedigree of family W. (**B**) Pedigree of family T. (**C**) Western-blot analysis of *MLH1* and *PMS2* proteins. (**D**) Characterization of constructed *MLH1* mutated plasmids. Numbers represent age at diagnosis. Minus signs indicate that the individual was confirmed not to carry the specific mutation. Shading indicates that the individual was affected with cancer. The arrow heads indicate the proband for that family.

Abbreviations: CRC, colorectal cancer; CHOL, cholangiocarcinoma; HC, hepatic cancer; BC, breast cancer; PC, pancreatic carcinoma; OC, ovarian cancer; GC, gastric cancer.

(p.F571Lfs*2), c.1866delT (p.A623Qfs*14), c.278delG (p.S93Ifs*15) and c.522delG (p.I176Ffs*26) may be pathogenic (moderate evidence).

Discussion

Using multigene panel testing, we determined the prevalence and spectrum of germline mutations in 73 genes associated with various hereditary cancer syndromes in 618 unselected Chinese CRC patients. One hundred twelve patients (18.1%) carried at least a pathogenic or likely pathogenic germline mutation, most of which were in high-penetrance cancer susceptibility genes. One of every 5 patients with CRC diagnosed younger than 50 years had at least 1 pathogenic or likely pathogenic germline mutation (21.9%). The mutation rate was between Pearlman' (16%)[9] and Stoffel's studies (25.1%).[20] This differences may due to different testing panels or populations. Therefore, we recommend colon cancer screening in the general population starting at age 50.

The prevalence of LS mutation carriers reported in this study (11.3%) was slightly higher than previous publications.[21,22] All the CRC patients in our cancer center were recommended for immunohistochemistry and genetic

testing. However practical challenges in implementation and concerns regarding cost-effectiveness may result in certain selection bias, which means the patients who participated in the NGS are not completely random. After all, 44.7% of CRC patients in the cohort had a family cancer history and 46.4% of them were under 50, which was also higher than the previous publications. Therefore the prevalence we found for pathogenic and likely pathogenic mutations probably represents the maximum prevalence. Consistent with prior studies that performed germline LS testing after preselection with MSI/MMR tumor testing,[23–25] 36.5% of dMMR CRC patients (61/167) in the current study had LS mutations and 87.1% of LS mutation carriers (61/70) demonstrated MSI-H and/or dMMR. These results support the current practice of performing systematic MSI/MMR immunohistochemistry for all CRCs to screen for LS.[26,27] For screening, immunohistochemistry is almost equally sensitive as MSI. However limiting tumor analysis to patients who fulfill the Bethesda criteria or the Amsterdam II criteria would fail to identify 20% (14/70) of LS cases.

In addition to LS, our study also determined the prevalence and spectrum of other hereditary cancer syndromes found in 618 unselected CRC patients. Among 618 CRC patients, 6.8% (42/618) carried non-LS mutations (one with both LS and non-LS mutations), accounting for 37.5% (42/112) of all positive variants identified in our overall testing population. The large number of non-LS mutations found in this cohort suggests that MMR/MSI testing alone is insufficient to identify cancer risk in unselected CRC patients. Furthermore, 26.2% (n=11) of non-LS mutation carriers in this study also lacked traditional phenotypic characteristics of hereditary CRC risk (age under 50 years or family history of cancer), which made it more difficult to identify non-LS mutations. ATM mutations were found in 0.8% (n=5) of CRC patients in our study which was higher than the estimated 0.37% (455/123136) of general population prevalence in a previous publication[28] and raises questions about whether such mutations predispose to CRC. Other moderate genes that were not traditionally associated with CRC, such as CHEK2 and BLM (0.6%, 4/618), also showed a possible link between CHEK2/BLM mutations and CRC risk. These findings further support the hypothesis that the analysis of genes currently excluded from routine molecular diagnostic screens may be predisposed to a wider range of cancers, potentially including CRC.[28] While studies support an association with cancer, the magnitude of the risk and complete cancer spectrum for variants in these genes is unclear.[29]

In our study, pedigree analysis and western blot analysis also provided some moderate or supporting evidence to identify germline variants for pathogenicity. Families W and T had phenotypes and family histories that were highly specific for CRC with a single genetic etiology. The protein expression and interaction of *MLH1* and *PMS2* were affected by some inframe variants especially those that led to protein truncation or had an impact on the interacting domains which is consistent with previous studies.[30,31]

Our study has certain limitations. We could not detect a large deletion (>20 bp), methylation of *MLH1* or the *BRAF* V600E mutation to confirm that there were no such mutations in the patients who were non-carriers especially non-LS mutation carriers with dMMR. We failed to obtain a comprehensive gene mutation status of the pedigrees due to patient compliance.

In conclusion, the advancements in NGS have led to a refined understanding of the genomics of colorectal cancer. The prevalence and spectrum of germline cancer susceptibility gene mutations have been investigated in previous studies among unselected or high-risk Caucasian, American and Asian populations, yet these conditions among unselected Chinese CRC patients are largely undetermined. Our study is the first to our knowledge to determine the prevalence and spectrum of germline cancer susceptibility gene mutations in unselected CRC patients in Chinese population using an NGS panel of 73 genes. Multigene panel testing facilitated the identification of germline mutations in patients who may have otherwise been missed. Only through extensive testing and the accumulation of large international datasets will sufficient information be generated to provide overwhelming evidence to determine whether a gene is associated with CRC susceptibility. Despite these improvements, further studies are needed to determine the function of various mutations in each gene which will enable us to promote clinical translation.

Ethics approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki. The ethical review board of Sun Yat-sen University Cancer Center approved this study (grant No. C2018-013-01).

Patient consent

Written informed consent were obtained.

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments

The authors are grateful to the family members who participated in this study and we would like to thank Dr. Xiang-Feng Zhu for the gift of pcDNA3.1B plasmids and the assistance from their laboratory. This work was supported by the Guangzhou Science and Technology Plan Projects (Health Medical Collaborative Innovation Program of Guangzhou) (grant No. 201803040019) and the National Natural Science Foundation of China (No. 81472522).

Author contributions

RG and YH contributed equally to this study. JYS and RHX conceived and designed the experiments; RG and YH performed the experiments; RG, XHY, XYL XKC and BL performed the bioinformatics analysis; LYS, LHK and ZLY collected the samples; RG, YH, YHL and DDZ analyzed the data. HYW and RG wrote the manuscript. All authors contributed to data analysis, drafting and revising the article, gave final approval of the version to be published, and agreed to be accountable for all aspects of the work

Disclosure

The authors report no conflicts of interest in this work.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. *CA Cancer J Clin*. 2016;66(1):7–30. doi:10.3322/caac.21332
2. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer–analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med*. 2000;343(2):78–85. doi:10.1056/NEJM200007133430201
3. Valle L. Recent discoveries in the genetics of familial colorectal cancer and polyposis. *Clin Gastroenterol Hepatol*. 2017;15(6):809–819. doi:10.1016/j.cgh.2016.09.148
4. Lorans M, Dow E, Macrae FA, Winship IM, Buchanan DD. Update on hereditary colorectal cancer: improving the clinical utility of multigene panel testing. *Clin Colorectal Cancer*. 2018;17(2):e293–e305. doi:10.1016/j.clcc.2018.01.001
5. Stoffel EM, Mangu PB, Gruber SB, et al. Hereditary colorectal cancer syndromes: American Society of Clinical Oncology Clinical Practice Guideline endorsement of the familial risk-colorectal cancer: European Society for Medical Oncology Clinical Practice Guidelines. *J Clin Oncol*. 2015;33(2):209–217. doi:10.1200/JCO.2014.58.1322
6. Balmana J, Balaguer F, Cervantes A, Arnold D. Familial risk-colorectal cancer: ESMO Clinical Practice Guidelines. *Ann Oncol*. 2013;24(Suppl 6):vi73–80. doi:10.1093/annonc/mdt209
7. Lynch HT, de la Chapelle A. Hereditary colorectal cancer. *N Engl J Med*. 2003;348(10):919–932. doi:10.1056/NEJMra012242
8. Obuch JC, Ahnen DJ. Colorectal cancer: genetics is changing everything. *Gastroenterol Clin North Am*. 2016;45(3):459–476. doi:10.1016/j.gtc.2016.04.005
9. Pearlman, R, Frankel WL, Swanson B, et al. Prevalence and spectrum of germline cancer susceptibility gene mutations among patients with early-onset colorectal cancer. *JAMA Oncol*. 2017;3(4):464–471. doi:10.1001/jamaoncol.2016.5194
10. Yurgelun MB, Kulke MH, Fuchs CS, et al. Cancer susceptibility gene mutations in individuals with colorectal cancer. *J Clin Oncol*. 2017;35(10):1086–1095. doi:10.1200/JCO.2016.71.0012
11. Yurgelun MB, Allen B, Kaldate RR, et al. Identification of a variety of mutations in cancer predisposition genes in patients with suspected Lynch syndrome. *Gastroenterology*. 2015;149(3):604–13.e20. doi:10.1053/j.gastro.2015.05.006
12. Gallego CJ, Shirts BH, Bennette CS, et al. Next-generation sequencing panels for the diagnosis of colorectal cancer and polyposis syndromes: a cost-effectiveness analysis. *J Clin Oncol*. 2015;33(18):2084–2091. doi:10.1200/JCO.2014.59.3665
13. Susswein LR, Marshall ML, Nusbaum R, et al. Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. *Genet Med*. 2016;18(8):823–832. doi:10.1038/gim.2015.166
14. Hahn MM, de Voer RM, Hoogerbrugge N, Ligtenberg MJL, Kuiper RP, van Kessel AG. The genetic heterogeneity of colorectal cancer predisposition - guidelines for gene discovery. *Cell Oncol (Dordr)*. 2016;39(6):491–510. doi:10.1007/s13402-016-0284-6
15. Sue Richards NAPS. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;5(17):405–423. doi:10.1038/gim.2015.30
16. Casper J, Zweig AS, Villarreal C, et al. The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res*. 2018;46(D1):D762–D769. doi:10.1093/nar/gkx1020
17. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4(7):1073–1081. doi:10.1038/nprot.2009.86
18. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248–249. doi:10.1038/nmeth0410-248
19. Trojan J, Zeuzem S, Randolph A, et al. Functional analysis of hMLH1 variants and HNPCC-related mutations using a human expression system. *Gastroenterology*. 2002;122(1):211–219.
20. Stoffel EM, Koeppe E, Everett J, et al. Germline genetic features of young individuals with colorectal cancer. *Gastroenterology*. 2018;154(4):897–905.e1. doi:10.1053/j.gastro.2017.11.004
21. Barnetson RA, Tenesa A, Farrington SM, et al. Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer. *N Engl J Med*. 2006;354(26):2751–2763. doi:10.1056/NEJMoa053493
22. Hampel H, Frankel WL, Martin E, et al. Feasibility of screening for Lynch syndrome among patients with colorectal cancer. *J Clin Oncol*. 2008;26(35):5783–5788. doi:10.1200/JCO.2008.17.5950
23. Ward RL, Hicks S, Hawkins NJ. Population-based molecular screening for Lynch syndrome: implications for personalized medicine. *J Clin Oncol*. 2013;31(20):2554–2562. doi:10.1200/JCO.2012.46.8454
24. Moreira L, Balaguer F, Lindor N, et al. Identification of Lynch syndrome among patients with colorectal cancer. *JAMA*. 2012;308(15):1555–1565. doi:10.1001/jama.2012.13088
25. Kastrinos F, Ojha RP, Leenen C, et al. Comparison of prediction models for Lynch syndrome among individuals with colorectal cancer. *J Natl Cancer Inst*. 2016;108(2):djv308. doi:10.1093/jnci/djv308
26. Syngal S, Brand RE, Church JM, Giardiello FM, Hampel HL, Burt RW. ACG clinical guideline: genetic testing and management of hereditary gastrointestinal cancer syndromes. *Am J Gastroenterol*. 2015;110(2):223–262. quiz 263. DOI:10.1038/ajg.2014.435.

27. Giardiello FM, Allen JI, Axilbund JE, et al. Guidelines on genetic evaluation and management of Lynch syndrome: a consensus statement by the US Multi-society Task Force on colorectal cancer. *Am J Gastroenterol*. 2014;109(8):1159–1179. doi:10.1038/ajg.2014.186
28. Hu C, Hart SN, Polley EC, et al. Association between inherited germline mutations in cancer predisposition genes and risk of pancreatic cancer. *JAMA*. 2018;319(23):2401–2409. doi:10.1001/jama.2018.6228
29. Ma X, Zhang B, Zheng W. Genetic variants associated with colorectal cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. *Gut*. 2014;63(2):326–336. doi:10.1136/gutjnl-2012-304121
30. Kondo E, Horii A, Fukushige S. The interacting domains of three MutL heterodimers in man: hMLH1 interacts with 36 homologous amino acid residues within hMLH3, hPMS1 and hPMS2. *Nucleic Acids Res*. 2001;29(8):1695–1702.
31. Guerrette S, Acharya S, Fishel R. The interaction of the human MutL homologues in hereditary nonpolyposis colon cancer. *J Biol Chem*. 1999;274(10):6336–6341.

Supplementary material

Table S1 Summary of other detected pathogenic or likely pathogenic germline mutations

Gene	Transcript_Ref	Chr	Chr_start	Chr_end	C_dot	P_dot	Mutation_type	Clinvar annotation	Evidence	Mutation annotation	Recurrency	ExAC_ ALL	ExAC_EAS	1000G_ALL	1000G_EAS
APC	NM_000038	chr5	112175799	112175802	c.4510_4513del	p.S1504Afs*2	Frameshift	.	1 PVS1+1 PM2+1 PM6	MUT	1	.	.	.	.
APC	NM_000038	chr5	112174490	112174493	c.3202_3205del	p.S1068Gfs*57	Frameshift	.	1 PVS1+1 PM2+1 PM6	MUT	2	.	.	.	.
APC	NM_000038	chr5	112162891	112162891	c.1495C>T	p.R499*	Nonsense	MUT	1 PM2	MUT	1	.	.	.	.
APC	NM_000038	chr5	112116592	112116592	c.637C>T	p.R213*	Nonsense	MUT	1 PM2	MUT	1	.	.	.	.
ATM	NM_000051	chr11	108141977	108141977	c.2922-1G>T	NA	Splicing	LP	1 PVS1+1 PM2+1 PM6	MUT	1	.	.	.	.
ATM	NM_000051	chr11	108236087	108236087	c.9023G>A	p.R3008H	Missense	LP	1 PM2+1 PP3	LP	1	8.24E-06	0	.	.
ATM	NM_000051	chr11	108165729	108165729	c.4852C>T	p.R1618*	Nonsense	MUT	1 PM2	MUT	1	8.24E-06	0	.	.
ATM	NM_000051	chr11	108205807	108205807	c.8122G>A	p.D2708N	Missense	LP	1 PM2+1 PP3	LP	1	.	.	.	.
BARD1	NM_000465	chr2	215645386	215645386	c.1211dup	p.Y404*	Frameshift	MUT	1 PVS1+1 PM2+1 PM6	MUT	1	.	.	.	.
BARD1	NM_000465	chr2	215634013	215634013	c.1338C>A	p.Y446*	Nonsense	.	1 PVS1+1 PM2+1 PM6	MUT	1	8.27E-06	0.0001	.	.
BLM	NM_000057	chr15	91292816	91292816	c.320dup	p.L107Ffs*36	Frameshift	.	1 PVS1+1 PM2+1 PM6	MUT	2	8.25E-06	0.0001	.	.
BRCA2	NM_000059	chr13	32911601	32911601	c.3109C>T	p.Q1037*	Nonsense	MUT	1 PM2	MUT	1	.	.	.	.
BRIP1	NM_032043	chr17	59938887	59938887	c.14G>A	p.W5*	Nonsense	LP	1 PVS1+1 PM2+1 PM6	MUT	1	.	.	.	.
BRIP1	NM_032043	chr17	59871088	59871088	c.1343G>A	p.W448*	Nonsense	MUT	1 PM2	MUT	1	3.70E-05	0.0005	.	.

(*Continued*)

Table S1 (Continued).

Gene	Transcript_Ref	Chr	Chr_start	Chr_end	C_dot	P_dot	Mutation_type	Clinvar annotation	Evidence	Mutation annotation	Recurrency	ExAC_ ALL	ExAC_EAS	1000G_ALL	1000G_EAS
CDK4	NM_000075	chr12	58145317	58145317	c.184C>T	p.R62*	Nonsense	.	1 PVS1+1 PM2+1 PM6	MUT	1	8.24E-06	0	.	.
CHEK2	NM_007194	chr22	29121015	29121015	c.542G>A	p.R181H	Missense	MUT(1); VUS(5)	1 PM2	LP	1	0.0001	0.0012	.	.
CHEK2	NM_007194	chr22	29091175	29091175	c.1315C>T	p.Q439*	Nonsense	MUT/LP	1 PM2	LP	1	.	.	.	.
CHEK2	NM_007194	chr22	29095917	29095917	c.917G>C	p.G306A	Missense	LP(2);VUS (1)	1 PM2+1 PP3	LP	1	5.77E-05	0	.	.
CHEK2	NM_007194	chr22	29095926	29095926	c.909-1G>A	NA	Splicing	LP	1 PVS1+1 PM2+1 PM6	MUT	1	.	.	.	.
EPCAM	NM_002354	chr2	.	.	exon1_9 del	.	Deletion	.	1 PVS1+1 PM2	LP	1	.	.	.	.
EPCAM	NM_002354	chr2	.	.	exon8_9 del	.	Deletion.	.	1 PVS1+1 PM2+PP5	MUT	1	.	.	.	.
FH	NM_000143	chr1	241667402	241667402	c.1048C>T	p.R350W	Missense	LP	1 PM2+1 PP3	LP	1	2.48E-05	0	.	.
HNF1A	NM_001306179	chr12	121437263	121437263	c.1624-2A>T	NA	Splicing	.	1 PVS1+1 PM2+1 PM6	MUT	1	8.35E-06	0.0001	.	.
MLH1	NM_000249	chr3	37055984	37055984	c.739T>C	p.S247P	Missense	MUT	1 PM2+1 PP3	MUT	1	.	.	.	.
MLH1	NM_000249	chr3	37070324	37070324	c.1459C>T	p.R487*	Nonsense	MUT	1 PM2	MUT	1	.	.	.	.
MLH1	NM_000249	chr3	37067242	37067242	c.1153C>T	p.R385C	Missense	LP(1);VUS (6)	1 PM2+1 PP3	LP	1	3.30E-05	0.0003	.	.
MLH1	NM_000249	chr3	37058999	37058999	c.793C>T	p.R265C	Missense	MUT	1 PM2+1 PP3	MUT	9	.	.	.	.
MLH1	NM_000249	chr3	37053589	37053589	c.676C>T	p.R226*	Nonsense	MUT	1 PM2	MUT	1	8.27E-06	0	.	.
MLH1	NM_000249	chr3	37090508	37090508	c.2103G>C	p.Q701H	Missense	MUT	1 PM2	MUT	1	.	.	.	.
MLH1	NM_000249	chr3	37090492	37090492	c.2089del	p.L697Sfs*86	Frameshift	MUT	1 PVS1+1 PM2+1 PM6	MUT	1	.	.	.	.
MLH1	NM_000249	chr3	37083775	37083775	c.1684C>T	p.Q562*	Nonsense	MUT	1 PM2	MUT	1	.	.	.	.

(*Continued*)

Table S1 (Continued).

Gene	Transcript_Ref	Chr	Chr_start	Chr_end	C_dot	P_dot	Mutation_type	Clinvar annotation	Evidence	Mutation annotation	Recurrency	ExAC_ ALL	ExAC_EAS	1000G_ALL	1000G_EAS
MLH1	NM_000249	chr3	37067281	37067281	c.1192C>T	p.Q398*	Nonsense	MUT	1 PM2	MUT	1	.	.	.	.
MLH1	NM_000249	chr3	37042488	37042488	c.250A>G	p.K84E	Missense	LP	1 PM2+1 PP3	LP	1	.	.	.	.
MLH1	NM_000249	chr3	37089123	37089125	c.1852_1854del	p.K618del	Inframe_deletion	MUT	1 PM2+1 PM4	MUT	1	.	.	.	.
MLH1	NM_000249	chr3	37089143	37089143	c.1866del	p.A623Qfs*14	Frameshift	MUT	1 PVS1+1 PM2+1 PM6	MUT	1	.	.	.	.
MLH1	NM_000249	chr3	37090446	37090446	c.2041G>A	p.A681T	Missense	MUT	1 PM2+1 PP3	MUT	2	.	.	.	.
MLH1	NM_000249	chr3	37070273	37070276	c.1410-2_1411del	NA	Frameshift	.	1 PVS1+1 PM2+1 PM6	MUT	1	.	.	.	.
MLH1	NM_000249	chr3	37083823	37083823	c.1731+1G>A	NA	Splicing	MUT	1 PVS1+1 PM2	MUT	1	.	.	.	.
MLH1	NM_000249	chr3	37042445	37042445	c.208-1G>A	NA	Splicing	MUT	1 PVS1+1 PM2	MUT	1	.	.	.	.
MLH1	NM_000249	chr3	37056036	37056036	c.790+1G>A	NA	Splicing	MUT	1 PVS1+1 PM2	MUT	2	.	.	0.0002	0.001
MSH2	NM_000251	chr2	47690264	47690264	c.1481C>G	p.S494*	Nonsense	LP	1 PM2	LP	1	.	.	.	.
MSH2	NM_000251	chr2	47703631	47703631	c.2131C>T	p.R711*	Nonsense	MUT	1 PM2	MUT	1	.	.	.	.
MSH2	NM_000251	chr2	47702265	47702265	c.1861C>T	p.R621*	Nonsense	MUT	1 PM2	MUT	1	.	.	.	.
MSH2	NM_000251	chr2	47702388	47702388	c.1984C>T	p.Q662*	Nonsense	MUT	1 PM2	MUT	1	.	.	.	.
MSH2	NM_000251	chr2	47702181	47702181	c.1777C>T	p.Q593*	Nonsense	MUT	1 PM2	MUT	1	.	.	.	.
MSH2	NM_000251	chr2	47637254	47637255	c.388_389del	p.Q130Vfs*2	Frameshift	MUT	1 PM2	MUT	1	.	.	.	.
MSH2	NM_000251	chr2	47690235	47690238	c.1457_1460del	p.N486Tfs*10	Frameshift	MUT	1 PM2	MUT	2	.	.	.	.
MSH2	NM_000251	chr2	47705451	47705451	c.2251G>A	p.G751R	Missense	MUT	1 PM2+1 PP3	MUT	1	.	.	.	.

(*Continued*)

Table S1 (Continued).

Gene	Transcript_Ref	Chr	Chr_start	Chr_end	C_dot	P_dot	Mutation_type	Clinvar annotation	Evidence	Mutation annotation	Recurrency	ExAC_ ALL	ExAC_EAS	1000G_ALL	1000G_EAS
MSH2	NM_000251	chr2	47637272	47637272	c.408del	p.F136Lfs*38	Frameshift	.	1 PVS1+1 PM2+1 PM6	MUT	1	8.24E-06	0	.	.
MSH2	NM_000251	chr2	47643489	47643489	c.997T>C	p.C333R	Missense	LP(2);VUS(1)	1 PM2+1 PP3	LP	1	.	.	.	.
MSH2	NM_000251	chr2	47672797	47672797	c.1386+1G>T	NA	Splicing	LP	1 PVS1+1 PM2	LP	1	.	.	.	.
MSH2	NM_000251	chr2	47641560	47641560	c.942+3A>T	NA	Splicing	MUT	1 PM2	MUT	1	.	.	.	.
MSH2	NM_000251	chr2	47639588	47639588	c.687del	p.A230Lfs*16	Frameshift	MUT	1 PVS1+1 PM2+1 PM6	MUT	1	.	.	.	.
MSH2	NM_000251	chr2	.	.	exon1_6 del	.	Deletion	MUT	1 PVS1+1 PM2+PP5	MUT	1	.	.	.	.
MSH2	NM_000251	chr2	.	.	exon7 del	.	Deletion.	.	1 PVS1+1 PM2+PP5	MUT	1	.	.	.	.
MSH6	NM_000179	chr2	48028053	48028053	c.2931C>G	p.Y977*	Nonsense	MUT	1 PM2	MUT	1	.	.	.	.
MUTYH	NM_012222	chr1	45798627	45798627	c.458G>A	p.W153*	Nonsense	MUT/LP	1 PM2	LP	1	4.96E-05	0.0006	.	.
MUTYH	NM_012222	chr1	45798467	45798467	c.535C>T	p.R179C	Missense	MUT/LP	1 PM2+1 PP3	LP	1	8.24E-06	0	.	.
MUTYH	NM_012222	chr1	45797914	45797914	c.848G>A	p.G283E	Missense	LP(2);MUT(2);VUS(1)	1 PM2+1 PP3	LP	1	2.50E-05	0.0001	.	.
PMS2	NM_000535	chr7	6038813	6038813	c.631C>T	p.R211*	Nonsense	MUT	1 PM2	MUT	1	8.24E-06	0	.	.
PMS2	NM_000535	chr7	6026658	6026658	c.1738A>T	p.K580*	Nonsense	MUT	1 PM2	MUT	2	.	.	.	.
PTEN	NM_000314	chr10	89692905	89692905	c.389G>A	p.R130Q	Missense	MUT	1 PM2+1 PP3	MUT	1	.	.	.	.
TP53	NM_000546	chr17	7577547	7577547	c.734G>T	p.G245V	Missense	LP	1 PM2+1 PP3	LP	1	8.24E-06	0.0001	.	.
TSHR	NM_000369	chr14	81609751	81609751	c.1349G>A	p.R450H	Missense	MUT	1 PM2+1 PP3	MUT	4	0.0003	0.0034	0.0002	0.001

MUT: pathogenic
LP: likely pathogenic
VUS: uncertain significance