

Mutational Landscape and Clinicopathological Correlates of 553 Chinese Colorectal Cancer Cases Assessed by 116-Gene Targeted Sequencing

Yitong Liu†, Jianglin Li†, Mingfeng Wu, Zidan Luo, Yiyun Sun, Xinyu Chen, Enle Zhou*

Department of Pathology, Chongqing University Cancer Hospital, Chongqing 400030, China

† These authors contributed equally to this work and share the first authorship.

*Corresponding author: Enle Zhou, zhouenle@163.com

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: *Objective:* Comprehensive molecular profiling of colorectal cancer (CRC) can inform precision oncology, yet real-world data from Chinese cohorts based on clinical targeted panels remain limited. This study aimed to characterize the mutational landscape of Chinese CRC using a 116-gene targeted next-generation sequencing (NGS) panel and to evaluate associations between key genomic alterations and clinicopathological features. *Method:* We retrospectively enrolled 553 patients with pathologically confirmed colorectal adenocarcinoma or its histological subtypes who underwent 116-gene targeted sequencing. Single-nucleotide variants (SNV)/short insertions and deletions (InDel), copy number amplifications, and gene fusions were identified using a standardized bioinformatics pipeline. Somatic variants were annotated using both AMP/ASCO/CAP oncogenicity labels and laboratory ACMG-style pathogenicity labels. Gene-level analyses were based on a nonsynonymous variant set. Associations with clinicopathological features were assessed using Fisher's exact test with Benjamini-Hochberg correction. Potentially actionable molecular events were stratified into three evidence-based tiers. *Results:* Among 553 CRC patients, 343 (62.0%) were male and 210 (38.0%) were female, with a median age of 62 years (range, 17-94 years). The rectum was the most common tumor site (53.3%), followed by the left colon (24.8%) and right colon (21.9%). A total of 4,725 SNV/InDel variants were identified, of which 3,107 were nonsynonymous (involving 102 genes); the median number of nonsynonymous mutations per patient was 4 (range, 1-60). The five most frequently mutated genes were TP53 (78.1%), KRAS (53.0%), APC (49.4%), PIK3CA (22.4%), and FBXW7 (20.6%). Copy number amplification was detected in 12.7% of patients (70/553), most commonly involving MYC (2.9%) and FGFR3 (2.5%). Gene fusions were identified in 7 patients (1.3%). Co-mutation analysis revealed significant co-occurrence of TP53 and APC (OR = 2.07, 95% CI: 1.36-3.14, adjusted p = 0.005) and mutual exclusivity between KRAS and BRAF (OR = 0.10, 95% CI: 0.04-0.25, adjusted p < 0.001). By evidence-based tiering, 19.9% (110/553) of patients harbored at least one Tier 1 molecular event, and 66.7% (369/553) harbored at least one Tier 1 or Tier 2 event. *Conclusion:* This study delineates the 116-gene mutational landscape of 553 Chinese CRC patients, characterizing the TP53-KRAS-APC core mutation pattern, hotspot distributions, low-frequency copy number amplifications and fusion events, and a tiered map of potentially actionable molecular events, providing real-world

regional data to inform precision oncology in CRC.

Keywords: Colorectal neoplasms; Targeted sequencing; Mutational landscape; Copy number amplification; Gene fusion; Molecular stratification

Online publication: October 20, 2026

1. Introduction

Colorectal cancer (CRC) is the third most commonly diagnosed malignancy and the second leading cause of cancer-related mortality worldwide ^[1]. Despite advances in surgical resection, chemoradiotherapy, and targeted therapy, the 5-year survival rate for advanced CRC remains suboptimal, underscoring the need for deeper molecular characterization to guide individualized treatment strategies.

The Cancer Genome Atlas (TCGA) performed whole-exome sequencing on 276 CRC cases, revealing high-frequency mutated genes and core signaling pathways ^[2]. However, TCGA primarily represents Western populations, whereas China bears a substantial and rising CRC burden ^[3], with Chinese patients exhibiting distinct tumor site distributions and molecular features compared with Western cohorts ^[4-6]. Although several large-scale studies have characterized the genomic landscape of Chinese CRC ^[4-6], systematic mutational profiling using clinical targeted panels with standardized somatic variant interpretation remains limited. In this study, we analyzed 116-gene targeted sequencing data from 553 Chinese CRC patients to characterize mutation frequencies, copy number amplifications, gene fusions, and variant classification patterns, and to explore associations between key genomic alterations and clinicopathological features, thereby providing real-world regional data to inform precision oncology in CRC.

2. Materials and methods

2.1. Study population

A total of 561 patients with CRC who underwent tumor tissue NGS testing at Chongqing University Cancer Hospital between August 2024 and July 2026 (last case tested on July 9, 2026) were retrospectively enrolled.

The inclusion criteria were:

- (1) pathologically confirmed colorectal adenocarcinoma or its histological subtypes (including mucinous adenocarcinoma and signet ring cell carcinoma components);
- (2) successful completion of 116-gene targeted NGS;
- (3) availability of core clinicopathological data.

The exclusion criteria were:

- (1) NGS failure or substandard sequencing quality;
- (2) non-colorectal primary tumors;
- (3) histological types not meeting adenocarcinoma definitions.

Based on criterion (3), 8 patients were excluded (2 neuroendocrine carcinomas, 5 adenomas with focal malignant transformation, and 1 mixed carcinoma with a neuroendocrine component), yielding a final cohort of 553 patients. This study was approved by the Ethics Committee of Chongqing University Cancer Hospital (Approval No.: CZLL2026-278-002), with the requirement for informed consent waived.

2.2. Clinicopathological data collection

Clinical data collected included sex, age, tumor site (right colon, left colon, or rectum), histological type, differentiation grade, specimen type (surgical or biopsy), neoadjuvant treatment status, and TNM stage. Tumor site classification followed Chinese CRC diagnostic guidelines: the right colon comprises the cecum, ascending colon, hepatic flexure, and transverse colon; the left colon comprises the splenic flexure, descending colon, sigmoid colon, and rectosigmoid junction; and the rectum refers to the segment below the peritoneal reflection.

2.3. NGS testing

All tumor tissue specimens were formalin-fixed and paraffin-embedded (FFPE), and genomic DNA was extracted. A custom 116-gene targeted capture panel (Pan116, capture region 0.25 Mb) covering genes commonly mutated in CRC, DNA damage repair genes, targeted therapy-related genes, and core signaling pathway genes was used. Paired-end sequencing was performed on the Illumina NextSeq platform with an average sequencing depth exceeding 500x. The bioinformatics pipeline included BWA alignment, GATK-based preprocessing, Mutect2 for SNV/InDel calling, ANNOVAR annotation, CNVkit for copy number analysis, and FACTERA for fusion detection. Microsatellite instability (MSI) status was determined by the NGS panel based on microsatellite locus instability counting, with at least 8 unstable loci classified as MSI-H (laboratory-established threshold). Fusion events were confirmed by manual review using the Integrative Genomics Viewer (IGV). Amplification was determined by CNVkit-estimated copy number with a laboratory-established threshold of 3.5 or greater. Nineteen genes were included in the copy number analysis module: AKT2, AURKA, CCND1, CCNE1, CD274, CDK6, EGFR, ERBB2, FGF19, FGF3, FGFR1, FGFR3, IGF1R, MYC, PGR, PSMD4, RET, RSF1, and TOP2A.

2.4. Variant classification and analysis set definition

Somatic variants retained two reporting label systems:

- (1) AMP/ASCO/CAP oncogenicity labels (Oncogenic/Likely oncogenic) per the 2017 joint guidelines^[7];
- (2) laboratory-issued ACMG-style pathogenicity labels^[8] (Pathogenic/Likely pathogenic/Uncertain/Likely benign/Benign).

These two systems differ in scope and were used solely to describe reporting distributions, not for cross-conversion or combined interpretation. It should be noted that the ACMG framework was originally designed for germline variants; this study employed a tumor-only somatic workflow without paired normal controls, and these labels reflect laboratory reporting conventions only. Variant functional annotation referenced the COSMIC database^[9] and cBioPortal^[10].

Two variant sets were defined:

- (1) All detected variants: all SNV/InDel variants reported by the panel, used only for describing variant type composition and overall pathogenicity distribution;
- (2) Nonsynonymous analysis set: after excluding synonymous mutations, intronic variants (except splice sites), UTR, and flanking region variants, missense, nonsense, frameshift, in-frame insertion/deletion, and splice site variants were included.

All gene-level analyses (mutation frequency, hotspot mutations, co-mutation, clinicopathological associations, and potentially actionable molecular events) were based on the nonsynonymous analysis set.

During quality control, 5 samples were found to have duplicate bioinformatics analysis records; 22 fully duplicate variant entries were removed by sample-gene-transcript-HGVS deduplication, and all statistics were based on deduplicated, sample-level variant events. Additionally, as this study employed a tumor-only clinical somatic testing workflow without paired normal controls, germline-origin variants cannot be fully excluded; therefore, clinical interpretation of DNA damage repair gene mutations (e.g., BRCA1/2, PALB2) requires caution.

2.5. Statistical methods

Categorical variables were described as frequencies (percentages), and continuous variables as medians (ranges). Associations between gene mutations and clinicopathological features were analyzed using Fisher's exact test: tumor site was compared as colon (right and left combined) versus rectum in 2 x 2 tests; differentiation was compared as poorly differentiated (including poor-to-moderate, 48 cases) versus moderately differentiated (including moderate-to-well, 334 cases) in 2 x 2 tests, with well-differentiated (6 cases) and missing records (165 cases) excluded; and sex was compared as male versus female. Copy number amplification associations were analyzed using the same methods. Multiple comparisons were corrected using the Benjamini-Hochberg method (BH-FDR) per clinical variable across 30 genes (30 tests per variable), with adjusted $p < 0.05$ considered statistically significant. The association between BRAF mutation and tumor site was prespecified based on prior literature as a right colon versus rectum pairwise comparison, with uncorrected p-values reported. Co-mutation analysis was performed using pairwise Fisher's exact tests on the 15 most frequently mutated genes (105 pairs), reporting odds ratios (OR), 95% confidence intervals (CI), and BH-FDR-adjusted p -values. Statistical analyses were performed using Python 3.13.

2.6. Potentially actionable molecular event definition and evidence tiering

Referencing the NCCN Colon Cancer Guidelines (Version 3.2025) ^[11], potentially actionable molecular events were stratified into three evidence-based tiers as a research classification (not ESCAT/OncoKB standard):

- (1) Tier 1 (guideline-recommended or relatively high-level evidence molecular events): Tier 1A includes CRC guideline-recommended events: MSI-H/dMMR status (an independent biomarker phenotype guiding immunotherapy), BRAF p.V600E mutation, KRAS p.G12C mutation, ERBB2 (HER2) amplification, and NTRK1 gene fusion; Tier 1B includes pan-cancer approved or rare high-level evidence events in CRC: ERBB2 mutation, ALK fusion, and RET fusion, where ERBB2 mutation classification is based on anti-HER2 response evidence from pan-cancer basket trials (evidence level below that of ERBB2 amplification in CRC guidelines) and ALK/RET fusion evidence is mostly from case reports or small series;
- (2) Tier 2 (predictive/resistance markers): markers that are not direct therapeutic targets but guide treatment decisions, including KRAS non-G12C mutations and NRAS mutations (anti-EGFR resistance prediction);
- (3) Tier 3 (exploratory molecular events): events with no approved drugs in CRC but with clinical trial or preclinical evidence, including PIK3CA, MET, BRCA1/2, and BRAF non-V600E mutations, and FGFR1/FGFR3, MYC, and CCND1 amplifications. EGFR amplification was not included in the tiering analysis owing to insufficient evidence.

3. Results

3.1 Patient characteristics

Among the 553 CRC patients, 343 (62.0%) were male and 210 (38.0%) were female, with a median age of 62 years (range, 17-94 years). The rectum was the most common tumor site (295 cases, 53.3%), followed by the left colon (137 cases, 24.8%) and right colon (121 cases, 21.9%). Adenocarcinoma was the predominant histological type (509 cases, 92.0%), with 34 cases (6.1%) containing mucinous adenocarcinoma components and 5 cases (0.9%) containing signet ring cell carcinoma components; histological type was not recorded in 5 cases (0.9%). Surgical specimens accounted for 363 cases (65.6%) and biopsy specimens for 157 cases (28.4%), with specimen type not recorded in 33 cases (6.0%). Fifty-seven cases (10.3%) were post-neoadjuvant treatment. NGS testing revealed MSI-H in 35 cases (6.3%) and MSS in 518 cases (93.7%) (**Table 1**). Pathological TNM staging was available for only 144 cases (26.0%) and was therefore not included in subsequent association analyses.

Table 1. Clinicopathological and molecular characteristics of 553 colorectal cancer patients

Characteristic	Value
Total cases	553
Sex (Male/Female)	343 (62.0%) / 210 (38.0%)
Age, median (range)	62 (17-94)
Tumor site (Rectum/Left colon/Right colon)	295 (53.3%) / 137 (24.8%) / 121 (21.9%)
Specimen type (Surgical/Biopsy/Not recorded)	363 / 157 / 33
Post-neoadjuvant cases	57 (10.3%)
MSI status (MSI-H/MSS)	35 (6.3%) / 518 (93.7%)
Total SNV/InDel variants detected	4725
Nonsynonymous analysis set	3107 (excluding synonymous/intronic/UTR/flanking: 1618)
Genes with variants (all/nonsynonymous)	105 / 102
Nonsynonymous mutations/case, median (range)	4 (1-60)
Cases with copy number amplification	70 (12.7%)
Cases with fusion events	7 (1.3%)
Cases with ≥ 1 Tier 1 / Tier 1+2 / any tier event	110 (19.9%) / 369 (66.7%) / 435 (78.7%)
AMP/ASCO/CAP Oncogenic/Likely oncogenic	608 (12.9%)
ACMG-style P/LP/Uncertain/LB/B	404 (8.6%) / 675 (14.3%) / 2015 (42.6%) / 937 (19.8%) / 86 (1.8%)

Note: Pathogenicity and tier rows merged for brevity; full classification detail in Section 3.6.

3.2. Gene mutation profile

The 116-gene panel detected 4,725 SNV/InDel variants (involving 105 genes) across the 553 CRC cases, of which 3,107 were nonsynonymous variants (involving 102 genes), with 1,618 synonymous, intronic, UTR, and flanking region variants excluded. The median number of nonsynonymous mutations per patient was 4 (range, 1-60), with a median of 4 mutated genes per patient (range, 1-38).

Based on the nonsynonymous analysis set, the 10 most frequently mutated genes were TP53 (78.1%, 432/553), KRAS (53.0%, 293/553), APC (49.4%, 273/553), PIK3CA (22.4%, 124/553), FBXW7 (20.6%, 114/553), SMAD4 (17.2%, 95/553), ARID1A (17.0%, 94/553), BRCA2 (13.2%, 73/553), TSC2 (11.6%, 64/553), and MSH6 (11.2%, 62/553) (**Figure 1**). The patient-level mutational landscape of the 15 most

frequently mutated genes is shown in **Figure 2**.

By gene function, tumor suppressor genes (TP53, APC, SMAD4, FBXW7, ARID1A) showed high mutation frequencies; KRAS, PIK3CA, and BRAF represented oncogenes; and DNA damage repair genes (BRCA1, BRCA2, ATM) had relatively lower mutation frequencies with potential relevance for molecular stratification or exploratory investigation in selected contexts. Notably, a substantial proportion of variants in certain genes, particularly TSC2, were classified as VUS and are further addressed in the Discussion.

Codon-level variant analysis was performed for 6 high-frequency genes (KRAS, TP53, BRAF, PIK3CA, APC, and NRAS). KRAS mutations predominantly affected codon 12, with p.G12D (83 cases, 15.0%), p.G12V (66 cases, 11.9%), and p.G13D (55 cases, 9.9%) being the most common hotspots, collectively accounting for 69.6% of KRAS mutations; p.G12C (17 cases, 3.1%) is noteworthy as the target of KRAS G12C inhibitors. TP53 mutations were relatively dispersed, with p.R175H (50 cases, 9.0%) being the most common hotspot, followed by p.R282W (21 cases, 3.8%), p.R273C (20 cases, 3.6%), and p.R273H (18 cases, 3.3%), concentrated in the DNA-binding domain (exons 5-8). Among BRAF mutations, p.V600E accounted for 70.5% (31/44), while non-V600E variants accounted for 29.5% (13/44). PIK3CA hotspot mutations clustered at p.E545K (22 cases, 4.0%), p.H1047R (17 cases, 3.1%), and p.E542K (16 cases, 2.9%). APC mutations were predominantly truncating (nonsense and frameshift), with p.R1450* (29 cases, 5.2%) being the most common truncation site, concentrated in the exon 16 mutation cluster region. NRAS mutations (19 cases, 3.4%) were predominantly p.G12D (6 cases) and p.Q61K/R (8 cases combined).

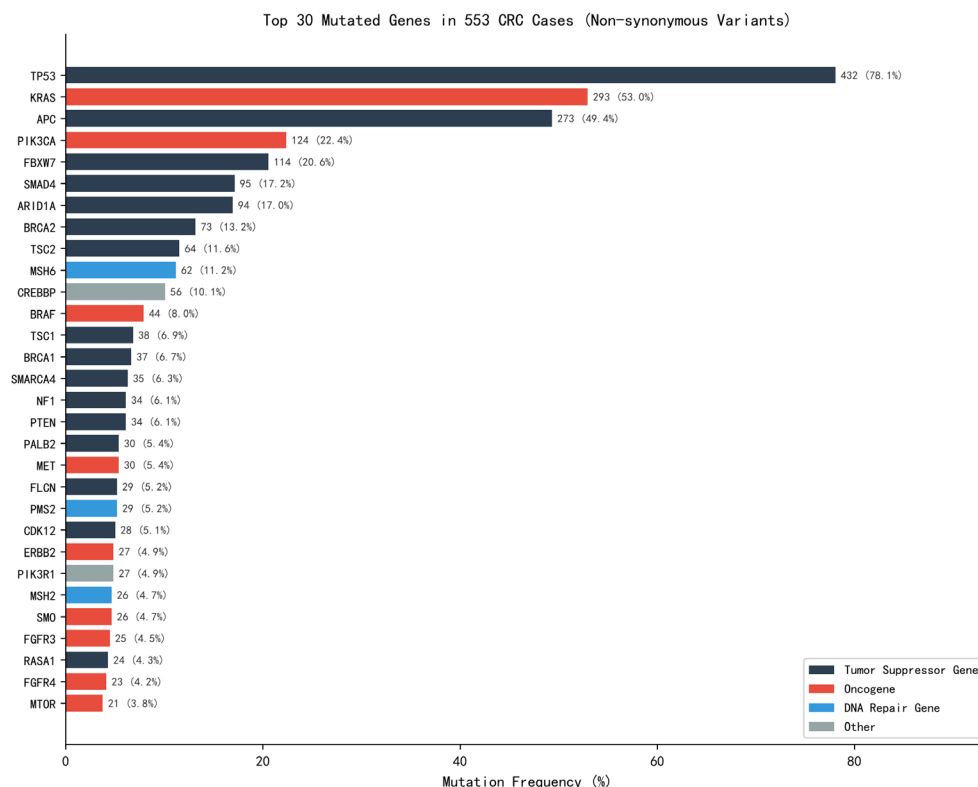


Figure 1. Mutation frequencies of the 30 most frequently mutated genes in 553 colorectal cancer cases (nonsynonymous variant analysis set; descriptive mutational profile, not driver gene ranking). Bars show mutation frequency (%); bar-end labels indicate case count and frequency; colors distinguish gene functional types (tumor suppressor/oncogene/DNA repair/other, classified by dominant biological function; categories are not mutually exclusive).

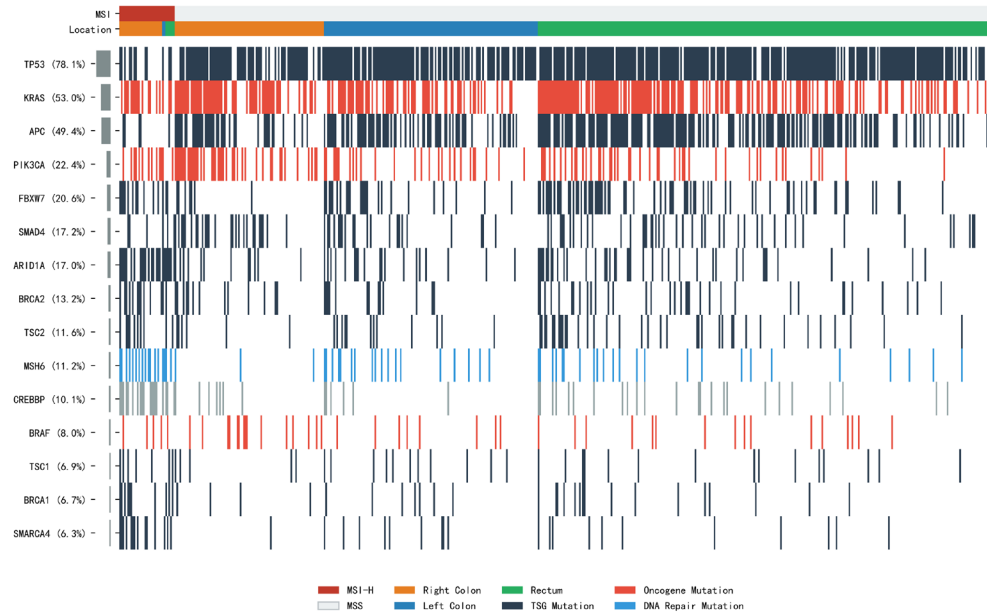


Figure 2. Patient-level mutational landscape of the 15 most frequently mutated genes in 553 colorectal cancer cases. Each column represents a single patient; multiple variant events in the same gene are recorded as a single cell (presence/absence). Patients are sorted by MSI status (top track, red = MSI-H) and tumor site (bottom track, orange = right colon, blue = left colon, green = rectum); genes are colored by functional type with mutation frequencies on the left.

3.3. Copy number amplification

Among the 553 CRC cases, 70 (12.7%) harbored at least one gene copy number amplification (the CNV analysis assessed amplification events only; homozygous deletions were not evaluated). Among the 19 CNV detection genes (see Section 2.3), the five most frequently amplified genes were MYC (2.9%, 16/553), FGFR3 (2.5%, 14/553), FGFR1 (2.0%, 11/553), ERBB2 (1.6%, 9/553), and AURKA (1.4%, 8/553). CCND1, FGF19, and FGF3 amplification frequencies were each 1.3% (7/553), and CD274 and EGFR amplification frequencies were each 0.9% (5/553), with the remaining genes below 0.5%.

Association analysis showed that ERBB2 amplification frequencies were similar across the right colon (1.7%), left colon (1.5%), and rectum (1.7%) ($p = 1.000$). MYC amplification was more frequent in the rectum (3.7%) and left colon (2.9%) than in the right colon (0.8%), although the difference was not statistically significant ($p = 0.309$).

3.4. Gene fusions

Seven patients (1.3%) harbored fusion events: 2 ALK fusions (EML4-ALK), 1 RET fusion (NCOA4-RET), and 4 NTRK1-related fusions. The 4 NTRK1-related fusions involved 3 distinct structures: TPM3-NTRK1 (2 cases, breakpoints at ex9-ex9 and ex8-ex9), TPR-NTRK1 (1 case, ex9-ex9), and NTRK1-TPR (1 case, ex10-ex21). In the first three, NTRK1 was the 3' gene with an intact kinase domain, representing canonical oncogenic NTRK fusions; in NTRK1-TPR, NTRK1 was the 5' gene with the kinase domain not retained, of uncertain oncogenic significance. Among fusion-positive patients, 4 were female and 3 were male, with a median age of 68 years (range, 54-80 years); by original anatomical site, 5 were located in the ileocecal

region, 1 in the ascending colon, and 1 in the rectum; by the study's site grouping, the first 6 were classified as right colon. Given the small sample size, site distribution is presented descriptively. Four fusion-positive patients were MSI-H (3 kinase domain-intact NTRK1 fusions and 1 EML4-ALK fusion).

3.5. Co-mutation analysis

Pairwise Fisher's exact tests with BH-FDR correction on the 15 most frequently mutated genes revealed 21 significantly co-mutated or mutually exclusive gene pairs (adjusted $p < 0.05$). Key findings included: significant co-occurrence of TP53 and APC (OR = 2.07, 95% CI: 1.36-3.14, $p = 6.61 \times 10^{-4}$, adjusted $p = 0.005$), consistent with their synergistic driver roles in CRC tumorigenesis; significant mutual exclusivity between KRAS and BRAF (OR = 0.10, 95% CI: 0.04-0.25, adjusted $p < 0.001$) and between APC and BRAF (OR = 0.20, 95% CI: 0.09-0.45, adjusted $p < 0.001$), suggesting alternative activation patterns within the MAPK pathway; and significant co-occurrence of KRAS and PIK3CA (OR = 2.32, 95% CI: 1.52-3.54, adjusted $p = 0.001$). Additionally, co-mutation clustering of MSH6 with CREBBP (OR = 5.42, 95% CI: 2.87-10.24), SMARCA4 (OR = 4.20), BRCA1 (OR = 4.47), and ARID1A (OR = 2.45) primarily reflects the high mutational burden phenotype of MSI-H tumors (**Figure 3**).

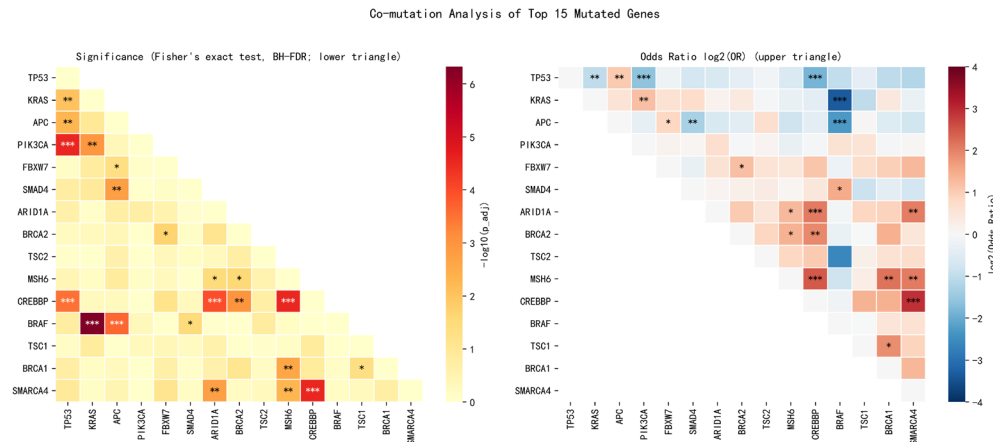


Figure 3. Co-mutation analysis of the 15 most frequently mutated genes. Left panel (lower triangle): Fisher's exact test significance after BH-FDR correction ($-\log_{10}$ adjusted p); right panel (upper triangle): $\log_2(\text{OR})$, red for co-mutation ($\text{OR} > 1$), blue for mutual exclusivity ($\text{OR} < 1$). Significance symbols shown only for adjusted $p < 0.05$ (* adjusted $p < 0.05$, ** adjusted $p < 0.01$, *** adjusted $p < 0.001$); cells without asterisks indicate non-significance after BH-FDR correction, not missing values. OR estimates for low-frequency events (fewer than 5 co-mutated cases) are unstable and should be interpreted as association signals only. Note: MSH6-related co-mutation clustering primarily reflects the high mutational burden background of MSI-H tumors.

3.6. Variant classification

All 4,725 detected SNV/InDel variants were described using two label systems (see Section 2.4). In ACMG-style pathogenicity labels, Uncertain was the most prevalent category (42.6%, 2,015), followed by Likely benign (19.8%, 937), Likely pathogenic (14.3%, 675), Pathogenic (8.6%, 404), and Benign (1.8%, 86), with Pathogenic/Likely pathogenic totaling 22.8% (1,079). In AMP/ASCO/CAP oncogenicity labels, Likely oncogenic accounted for 9.4% (443) and Oncogenic for 3.5% (165), totaling 12.9% (608).

The variant type distribution (all detected variants) showed missense mutations as the most common (2,197, 46.5%), followed by intronic variants (896, 19.0%), synonymous mutations (680, 14.4%), nonsense

mutations (424, 9.0%), and frameshift deletions (239, 5.1%). Synonymous, intronic, and UTR/flanking region variants were used for descriptive statistics only and were not included in gene-level analyses.

3.7. Potentially actionable molecular event tiering

Per the evidence tiering criteria in Section 2.6 (Figure 4), patient-level prevalence was: Tier 1A, 15.9% (88/553); Tier 1B, 5.4% (30/553); and their union, 19.9% (110/553, with 8 patients harboring both Tier 1A and 1B events). In Tier 1A, MSI-H was the most common event (6.3%), followed by BRAF p.V600E (5.6%), KRAS p.G12C (3.1%), ERBB2 amplification (1.6%), and NTRK1 fusion (0.5%, kinase domain-intact); Tier 1B included ERBB2 mutation (4.9%), ALK fusion (0.4%), and RET fusion (0.2%).

Tier 2: KRAS non-G12C mutations (49.9%) and NRAS mutations (3.4%), with 369 patients (66.7%) harboring at least one Tier 1 or Tier 2 event.

Tier 3: PIK3CA mutations (22.4%), BRCA2 mutations (13.2%), BRCA1 mutations (6.7%), MET mutations (5.4%), BRAF non-V600E mutations (2.4%), and amplifications of MYC (2.9%), FGFR3 (2.5%), FGFR1 (2.0%), and CCND1 (1.3%). Including Tier 3, 435 patients (78.7%) harbored at least one event of any tier. It should be noted that BRCA1/2 mutations are mostly passenger events and are not equivalent to homologous recombination deficiency; MET mutations differ fundamentally from MET amplification or exon 14 skipping in targetability; and these Tier 3 events should be regarded as exploratory.

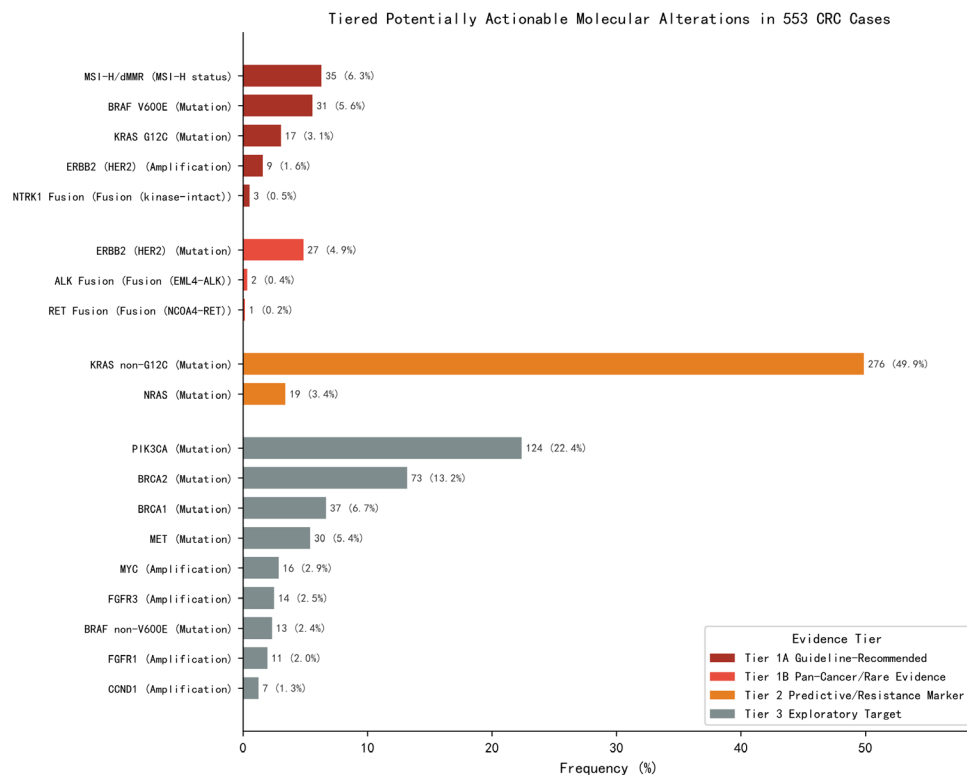


Figure 4. Tiered potentially actionable molecular events in 553 colorectal cancer cases. Classification is a research framework based on NCCN guidelines and literature evidence (not ESCAT/OncoKB standard): Tier 1A (dark red) = CRC guideline-recommended molecular events (88 cases, 15.9%), Tier 1B (light red) = pan-cancer approved or rare events with relatively high-level evidence (30 cases, 5.4%), Tier 2 (orange) = predictive/resistance markers, Tier 3 (gray) = exploratory molecular events; patient-level deduplication across tiers: Tier 1A+1B = 19.9% (110/553, with 8 patients harboring both Tier 1A and 1B events), Tier 1+2 = 66.7% (369/553), all tiers = 78.7% (435/553). Note: BRCA1/2, MET, and other Tier 3 events are exploratory and not equivalent to proven CRC therapeutic targets.

3.8. Associations between gene mutations and clinicopathological features

Association analysis results for the 30 most frequently mutated genes with tumor site, differentiation, and sex. Tumor site *p*-values correspond to 2 x 2 Fisher's exact tests comparing colon (right and left combined) versus rectum (BH-FDR corrected, 30 tests per variable), with subgroup frequencies shown as descriptive breakdowns. PIK3CA mutations were significantly enriched in colon (right colon 43.0%, left colon 19.7% vs. rectum 15.3%, adjusted *p* < 0.001); TP53 was enriched in rectum (rectum 83.1% vs. colon combined 72.5%, adjusted *p* = 0.022; right colon 64.5%, left colon 79.6%); APC was enriched in rectum (rectum 55.6% vs. colon combined 43.0%, adjusted *p* = 0.022; right colon 35.5%, left colon 48.2%); and CDK12 (right colon 12.4%, left colon 4.4% vs. rectum 2.4%, adjusted *p* = 0.022) and SMO (right colon 13.2%, left colon 2.2% vs. rectum 2.4%, adjusted *p* = 0.047) mutations were enriched in colon. A prespecified pairwise comparison (not additionally corrected) showed BRAF mutation enrichment in the right colon (15.7% vs. rectum 5.8%, uncorrected *p* = 0.002), consistent with prior literature. In poorly differentiated tumors, SMAD4 (31.2% vs. moderately differentiated 12.3%, adjusted *p* = 0.012), ARID1A (37.5% vs. 15.6%, adjusted *p* = 0.012), BRAF (22.9% vs. 4.8%, adjusted *p* = 0.003), CDK12 (16.7% vs. 3.6%, adjusted *p* = 0.012), and PALB2 (14.6% vs. 3.3%, adjusted *p* = 0.021) mutations were significantly enriched. APC mutations were more common in males (54.8% vs. females 40.5%, adjusted *p* = 0.036). Given the high proportion of missing differentiation records (29.8%, 165/553), differentiation-related results should be interpreted as exploratory. Additionally, CDK12, SMO, and PALB2 mutations are predominantly variants of uncertain significance (VUS), and their associations with clinicopathological features may be influenced by VUS inclusion, requiring cautious interpretation in conjunction with pathogenicity classification.

4. Discussion

This study systematically analyzed 116-gene targeted sequencing data from 553 Chinese CRC patients, delineating the mutational landscape across SNV/InDel, copy number amplification, gene fusion, and variant classification dimensions based on a well-defined nonsynonymous variant analysis set, providing real-world regional data for CRC precision oncology.

4.1. Comparison of high-frequency mutated genes with TCGA and Chinese studies

The mutation frequencies of TP53 (78.1%), KRAS (53.0%), and APC (49.4%) in this study differ from TCGA reports (TP53, 50%-60%; KRAS, 40%-45%; APC, 70%-80%)^[2]. The higher TP53 frequency may reflect:

- (1) population differences, as this cohort has a higher proportion of rectal cancer (53.3%), in which both TP53 and APC were enriched;
- (2) the high panel sequencing depth (>500x) increasing sensitivity for low-abundance TP53 variants, whereas APC large fragment deletions may be underestimated;
- (3) the 28.4% biopsy specimens potentially affecting tumor purity. Additionally, TP53 detection included splice site variants, which may contribute to the elevated apparent frequency.

Compared with other Chinese large-scale studies, Guo et al.^[5] (887-gene panel, 869 cases) reported TP53 (72%), APC (63%), and KRAS (42%) as the most frequent mutations, consistent with our ranking; Cheng et al.^[6] (73-gene panel, 414 cases) reported lower mutation frequencies overall, likely reflecting the smaller panel size. Zhao et al.^[4] (whole-exome sequencing, 1,015 cases) similarly confirmed TP53-KRAS-

APC as the core driver combination. The TSC2 mutation frequency of 11.6% in this study exceeds TCGA's approximately 5%-6%^[2]; however, 94.0% (63/67) of TSC2 variants were VUS missense mutations, with only 4 patients (0.7%) harboring pathogenic or likely pathogenic TSC2 variants, consistent with TCGA's truncation-dominated inactivation pattern, indicating that the difference primarily stems from variant inclusion criteria.

4.2. Clinical significance of copy number amplifications and gene fusions

Copy number amplification events were relatively infrequent (12.7%), but some carry therapeutic significance: ERBB2 amplification (1.6%) is a potential target for anti-HER2 therapy^[12], MYC amplification (2.9%) suggests aberrant WNT pathway activation^[13], and EGFR amplification (0.9%) has unclear treatment guidance value and was not included in the tiering analysis.

Gene fusions were rare (1.3%), but some fusion-positive patients may benefit from targeted therapy: kinase domain-intact NTRK1 fusions (0.5%, 3 cases) represent established targets for larotrectinib^[14], and ALK fusions (0.4%) and RET fusions (0.2%) also have corresponding targeted agents^[15-16]. Breakpoint structure analysis revealed that the same "NTRK1 fusion" designation can correspond to different molecular structures (3 kinase domain-intact versus 1 kinase domain-not-retained), underscoring the importance of evaluating fusion direction and kinase domain integrity in clinical report interpretation. All 3 kinase domain-intact NTRK1 fusions were found in MSI-H patients, but the small sample size precludes further inference.

4.3. Challenges in somatic variant classification

This study retained both AMP/ASCO/CAP oncogenicity labels^[7] and laboratory ACMG-style pathogenicity labels^[8] for somatic variants (see Section 2.4). The high proportion of variants classified as Uncertain (42.6%) reflects the practical difficulty of somatic variant interpretation in solid tumor panel testing: unlike germline variants, clinical interpretation evidence for somatic variants is more limited, particularly for missense mutations lacking functional experimental validation. Pathogenic/likely pathogenic variants accounted for 22.8% under the ACMG framework. In contrast, oncogenic/likely oncogenic variants accounted for 12.9% under the AMP/ASCO/CAP framework (the two systems partially overlap and cannot be simply summed), indicating that a considerable proportion of variants receive some interpretive label in clinical testing reports. However, their clinical significance requires context-specific interpretation. For genes dominated by VUS such as TSC2, apparent mutation frequency alone may overestimate clinical significance, and pathogenicity-stratified interpretation is essential.

4.4. Co-mutation patterns and pathway interactions

TP53-APC co-mutation (OR = 2.07) is one of the canonical synergistic driver combinations in CRC, with both genes functioning as tumor suppressors in the p53 and WNT pathways, respectively, and their co-mutation accelerating adenoma-to-carcinoma progression^[17]. The significant mutual exclusivity of KRAS-BRAF (OR = 0.10) and APC-BRAF (OR = 0.20) is consistent with alternative activation of MAPK pathway driver events: BRAF-mutant CRC constitutes a molecularly distinct subgroup (often associated with MSI-H and serrated pathway origin) that rarely overlaps with classical KRAS or APC driver events^[18]. KRAS-PIK3CA co-mutation (OR = 2.32) suggests that synergistic activation of the MAPK and PI3K pathways confers a selective advantage in some CRC cases, consistent with prior reports^[18]. The co-mutation clustering of MSH6 with CREBBP, SMARCA4, BRCA1, and ARID1A primarily reflects the high mutational burden

background of MSI-H tumors rather than specific pathway interactions, and interpretation should account for MSI status as a confounder.

4.5. Potentially actionable molecular events and precision treatment

This study stratified potentially actionable molecular events into three evidence-based tiers to avoid overestimating clinical actionability. At the Tier 1 level, 19.9% of patients harbored at least one Tier 1 molecular event. MSI-H patients are a well-defined population that benefits from immune checkpoint inhibition: pembrolizumab is established as first-line standard therapy for MSI-H advanced CRC ^[19], nivolumab plus ipilimumab has shown durable benefit ^[20], and PD-1 blockade has demonstrated remarkably high complete response rates in dMMR locally advanced rectal cancer ^[21]. Patients with BRAF p.V600E mutations can benefit from encorafenib plus cetuximab ^[22]; those with KRAS p.G12C (3.1%) from sotorasib or adagrasib combined with anti-EGFR therapy ^[23]; and those with ERBB2 amplification from tucatinib plus trastuzumab ^[24]. ERBB2 mutation evidence is primarily derived from basket trials and does not have the same guideline-level support as ERBB2 amplification in CRC. NTRK1/ALK/RET fusions, although rare, have clearly effective targeted agents, highlighting the value of multigene panel testing.

At the Tier 2 level, KRAS non-G12C mutations (49.9%) and NRAS mutations (3.4%) are negative predictive markers for anti-EGFR therapy, with decisive influence on first-line treatment decisions. At the Tier 3 level, PIK3CA, MET, BRCA1/2, and other variants have no approved drugs in CRC and represent exploratory directions for later-line treatment. It should be specifically noted that BRAF non-V600E mutations (29.5% of BRAF mutations) are mostly class II or III variants that are insensitive to BRAF monomer inhibitors, whereas class III mutations may instead benefit from MEK inhibitors or combined anti-EGFR strategies ^[25], and should not be equated with V600E in clinical interpretation.

4.6. Limitations

This study has several limitations:

- (1) the retrospective single-center design with low TNM staging completeness (26.0%) precluded analysis of gene-stage associations;
- (2) the limited panel coverage (0.25 Mb) is unsuitable for standard tumor mutational burden calculation and may miss off-panel variants (e.g., POLE proofreading domain ^[26]);
- (3) the lack of paired normal controls means that some variants may be of germline origin, and certain genes have high VUS proportions (e.g., TSC2), requiring functional validation;
- (4) the amplification threshold (3.5 or greater copies) is lower than high-level amplification criteria used in some studies (e.g., 6 or greater copies), necessitating caution in cross-study comparisons, and the low amplification and fusion detection rates limit statistical power;
- (5) the lack of follow-up data and consensus molecular subtype analysis precludes assessment of gene-prognosis and gene-transcriptomic subtype associations. Nevertheless, this study reflects real-world clinical sequencing practice at a provincial cancer center and remains representative of routine molecular diagnostic workflows.

5. Conclusion

This study delineates the 116-gene mutational landscape of 553 Chinese CRC patients, characterizing

the TP53-KRAS-APC core mutation pattern, classic hotspot distributions, low-frequency copy number amplification and fusion events, and a tiered map of potentially actionable molecular events (19.9% of patients harboring Tier 1 events), providing real-world regional data to inform precision oncology in CRC.

Disclosure statement

The author declares no conflict of interest.

References

- [1] Sung H, Ferlay J, Siegel RL, et al., 2021, Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 71(3): 209–249.
- [2] Cancer Genome Atlas Network, 2012, Comprehensive Molecular Characterization of Human Colon and Rectal Cancer. *Nature*, 487(7407): 330–337.
- [3] Chen W, Zheng R, Baade PD, et al., 2016, Cancer Statistics in China, 2015. *CA: A Cancer Journal for Clinicians*, 66(2): 115–132.
- [4] Zhao Q, Wang F, Chen YX, et al., 2022, Comprehensive Profiling of 1015 Patients' Exomes Reveals Genomic-Clinical Associations in Colorectal Cancer. *Nature Communications*, 13: 2342.
- [5] Guo L, Wang Y, Yang W, et al., 2023, Molecular Profiling Provides Clinical Insights Into Targeted and Immunotherapies as Well as Colorectal Cancer Prognosis. *Gastroenterology*, 165(2): 414–428.
- [6] Cheng B, Xu L, Zhang Y, et al., 2024, Correlation Between NGS Panel-Based Mutation Results and Clinical Information in Colorectal Cancer Patients. *Heliyon*, 10(7): e29299.
- [7] Li MM, Datto M, Duncavage EJ, et al., 2017, Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer: A Joint Consensus Recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *Journal of Molecular Diagnostics*, 19(1): 4–23.
- [8] Richards S, Aziz N, Bale S, et al., 2015, 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. *Genetics in Medicine*, 17(5): 405–424.
- [9] Tate JG, Bamford S, Jubb HC, et al., 2019, COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research*, 47(D1): D941–D947.
- [10] Cerami E, Gao J, Dogrusoz U, et al., 2012, The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery*, 2(5): 401–404.
- [11] National Comprehensive Cancer Network, 2025, NCCN Clinical Practice Guidelines in Oncology: Colon Cancer (Version 3.2025), visited on July 20, 2026, <https://www.nccn.org>.
- [12] Sartore-Bianchi A, Trusolino L, Martino C, et al., 2016, Dual-Targeted Therapy With Trastuzumab and Lapatinib in Treatment-Refractory, KRAS Codon 12/13 Wild-Type, HER2-Positive Metastatic Colorectal Cancer (HERACLES): A Proof-of-Concept, Multicentre, Open-Label, Phase 2 Trial. *The Lancet Oncology*, 17(6): 738–746.
- [13] Blank A, Schenk L, Moch H, 2021, The Impact of MYC Alterations in Colorectal Cancer. *Der Pathologe*, 42(3): 255–261.
- [14] Drilon A, Laetsch TW, Kummar S, et al., 2018, Efficacy of Larotrectinib in TRK Fusion-Positive Cancers in

- Adults and Children. *New England Journal of Medicine*, 378(8): 731–739.
- [15] Lin JJ, Riely GJ, Shaw AT, 2019, Targeting ALK: Precision Medicine Takes on Drug Resistance. *Cancer Discovery*, 9(7): 899–915.
 - [16] Subbiah V, Gainor JF, Rahal R, et al., 2018, Precision Targeted Therapy With BLU-667 for RET Fusion-Positive Solid Tumors. *Cancer Discovery*, 8(7): 836–849.
 - [17] Fearon ER, Vogelstein B, 1990, A Genetic Model for Colorectal Tumorigenesis. *Cell*, 61(5): 759–767.
 - [18] Yaeger R, Chatila WK, Lipsyc MD, et al., 2018, Clinical Sequencing Defines the Genomic Landscape of Metastatic Colorectal Cancer. *Cancer Cell*, 33(1): 125–136.
 - [19] Andre T, Shiu KK, Kim TW, et al., 2020, Pembrolizumab in Microsatellite-Instability-High Advanced Colorectal Cancer. *New England Journal of Medicine*, 383(23): 2207–2218.
 - [20] Overman MJ, Lonardi S, Wong KYM, et al., 2018, Durable Clinical Benefit With Nivolumab Plus Ipilimumab in DNA Mismatch Repair-Deficient/Microsatellite Instability-High Metastatic Colorectal Cancer. *Journal of Clinical Oncology*, 36(8): 773–779.
 - [21] Cercek A, Lumish M, Sinopoli J, et al., 2022, PD-1 Blockade in Mismatch Repair-Deficient, Locally Advanced Rectal Cancer. *New England Journal of Medicine*, 386(25): 2363–2376.
 - [22] Kopetz S, Grothey A, Yaeger R, et al., 2019, Encorafenib, Binimetinib, and Cetuximab in BRAF V600E-Mutated Colorectal Cancer. *New England Journal of Medicine*, 381(17): 1632–1643.
 - [23] Fakih MG, Kopetz S, Kuboki Y, et al., 2022, Sotorasib for Previously Treated Colorectal Cancers With KRAS G12C Mutation (CodeBreak100): A Prespecified Analysis of a Single-Arm, Phase 2 Trial. *The Lancet Oncology*, 23(1): 115–124.
 - [24] Strickler JH, Cercek A, Siena S, et al., 2023, Tucatinib Plus Trastuzumab for Chemotherapy-Refractory, HER2-Positive, RAS Wild-Type Unresectable or Metastatic Colorectal Cancer (MOUNTAINEER): A Multicentre, Open-Label, Phase 2 Study. *The Lancet Oncology*, 24(5): 496–508.
 - [25] Yaeger R, Corcoran RB, 2019, Targeting Alterations in the RAF-MEK Pathway. *Cancer Discovery*, 9(3): 329–341.
 - [26] Palles C, Cazier JB, Howarth KM, et al., 2013, Germline Mutations Affecting the Proofreading Domains of POLE and POLD1 Predispose to Colorectal Adenomas and Carcinomas. *Nature Genetics*, 45(2): 136–144.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.