Mutational landscape in the precancerous stages of sporadic colorectal cancer

ANNA VALICKOVA[1,2], MARKETA URBANOVA[1,2], JOSEF HORAK[1], JIRI JUNGWIRTH[3], JAN KRAL[4], TOMAS HUCL[5], VERONIKA MAKAJEVOVA[6], SANDRA SUMMEROVA[7], PAVEL KOHOUT[7], RADOSLAV MATEJ[8], PAVEL VODICKA[1,2,9] and VERONIKA VYMETALKOVA[1,2]

[1]Department of Molecular Biology of Cancer, Institute of Experimental Medicine of The Czech Academy of Sciences, 142 00 Prague, Czech Republic; [2]Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University, 128 00 Prague, Czech Republic; [3]Institute of Physiology, First Faculty of Medicine, Charles University, 128 00 Prague, Czech Republic; [4]Department of Internal Medicine, Second Faculty of Medicine, Charles University, 150 06 Prague, Czech Republic; [5]Department of Hepatogastroenterology, Institute for Clinical and Experimental Medicine, 142 00 Prague, Czech Republic; [6]Department of Surgery, Thomayer University Hospital, 142 00 Prague, Czech Republic; [7]Department of Internal Medicine, Third Faculty of Medicine, Charles University and Thomayer University Hospital, 142 00 Prague, Czech Republic; [8]Department of Pathology and Molecular Medicine, Third Faculty of Medicine, Charles University and Thomayer University Hospital, 142 00 Prague, Czech Republic; [9]Biomedical Centre, Faculty of Medicine in Pilsen, Charles University, 323 00 Pilsen, Czech Republic

Received September 16, 2025; Accepted March 30, 2026

DOI: 10.3892/ol.2026.15634

Abstract. Colorectal adenoma (CRA) is a precancerous lesion that can progress to colorectal carcinoma (CRC); however, its malignant potential varies considerably. The present study aimed to characterize the putatively pathogenic variants (PPVs) of CRA and to assess their potential clinical relevance in identifying lesions with an increased risk of progression at precancerous stages. PPVs in a panel of 176 cancer-associated genes were analyzed in 67 CRA samples and matched adjacent normal mucosa using next-generation sequencing. The panel included genes involved in DNA repair pathways, cell cycle regulation and the genes directly associated with CRC development. PPVs in CRA tissue were identified in 44 patients. The most frequently mutated genes were found to be *APC*, *KRAS*, *FBXW7*, *BRAF* and *MAP2K4*, with *APC* mutations being the most prevalent. A higher frequency of PPVs was observed in CRA with high-grade dysplasia and tubulo-villous features. Notably, among the entire sample set, there were 10 hyperplastic polyps, which are generally considered low risk; however, three carried PPVs, specifically one polyp carried *APC*, *MAP2K4* and *FEN1* mutations, another polyp carried *BRAF* and *POLQ* mutation, and a third carried a *MAP2K4* mutation. These findings suggested that mutational profiling may provide additional molecular information beyond histopathological assessment and could support improved risk stratification of colorectal lesions in early, precancerous stages. Such molecular characterization may be of value for identifying subsets of patients who could benefit from closer clinical surveillance.

Correspondence to: Dr Veronika Vymetalkova, Department of Molecular Biology of Cancer, Institute of Experimental Medicine of The Czech Academy of Sciences, 1083 Videnska, 142 00 Prague, Czech Republic
E-mail: veronika.vymetalkova@iem.cas.cz

Key words: colorectal adenoma, mutation, next-generation sequencing

Introduction

Colorectal carcinoma (CRC) is one of the leading global cancers , causing ~1.9 million new cases and >900,000 deaths annually, and ranking third in incidence and second in cancer-related mortality worldwide (1). The majority of CRC cases arise from a precursor lesions known as a colorectal adenoma (CRA) (2), which undergoes a series of genetic and epigenetic alterations in a process described by the Vogelstein model of adenoma-carcinoma progression (3). This stepwise accumulation of mutations enables the transformation of normal colonic epithelium into invasive carcinoma (4).

While screening programs have reduced CRC incidence and mortality through early CRA detection and removal, a subset of CRA still carries higher malignant potential based on molecular characteristics, such as dysplasia grade, histological subtype and specific gene mutations (5). Thus, characterizing the mutational landscape of the CRA can provide insight into early oncogenic events and help identify lesions with a greater risk of progression, potentially guiding personalized surveillance strategies. Advances in next-generation sequencing (NGS) have enabled the comprehensive characterization of these mutational landscapes, identifying key driver genes that may distinguish high-risk adenoma cases from those less likely to progress (6).

In addition to classical CRA, other types of colorectal polyps have gained further attention, particularly hyperplastic

polyps, which are traditionally regarded as low-risk or non-neoplastic (7). While the majority of hyperplastic polyps are considered to result from inflammatory or reactive processes, occasional findings of somatic mutations in oncogenes raise the possibility that a subset may harbor early neoplastic potential (8,9), especially in the context of the serrated neoplasia pathway (10,11). Therefore, comprehensive genetic profiling of both adenomas and hyperplastic polyps could improve risk stratification and challenge current clinical assumptions.

Although previous studies (12-15), including an earlier study by Jungwirth *et al* (6), have focused on the frequency of mutations in well-known cancer-related genes, there remains a need to explore these alterations in larger gene panels, higher sample sizes and with improved technical sensitivity. The present study builds upon this by analyzing the mutational profile of CRA tissues and their adjacent mucosa using NGS across an expanded gene panel. This approach not only re-evaluated the prevalence of key driver mutations in adenomas of varying dysplasia grades, but also examined potential pathogenic alterations in hyperplastic polyps, thereby offering new insights into early CRC development.

The present study aimed to further explore the profile of the putatively pathogenic variants (PPVs) of CRA using a targeted panel of 176 cancer-associated genes in a cohort of CRA samples, with the goal of identifying recurrent PPVs and evaluating their association with dysplasia grade and histological features. Ultimately, the present study may contribute to the development of a clinically useful oncogene panel for risk assessment and early detection of CRC.

Materials and methods

Sample collection. Workflow order is clarified in Fig. 1. The present study included 67 consecutive analyzed patients with CRA (Table I). The cohort consisted of 44 men (66%) and 23 women (34%), with a median age of 65 years (range: 43-77 years). Patients were eligible for inclusion if they were diagnosed with sporadic CRA detected during routine screening or diagnostic colonoscopy. Only patients without a personal history of CRC, and without clinical, endoscopic or histopathological features suggestive of hereditary CRC syndromes, including familial adenomatous polyposis or Lynch syndrome, were included. Patients with a known hereditary cancer syndrome, a personal history of any malignant disease, synchronous CRC or inflammatory bowel disease were excluded. In addition, patients with incomplete clinical data or insufficient tissue material for molecular analysis were not included in the present study. From each patient, a single CRA and the corresponding adjacent normal mucosa were obtained. Patients were recruited in the Czech Republic and thus represented a Central European population.

The CRA tissue samples and adjacent mucosa were collected during routine preventive colonoscopic examinations performed after the age of 50 years or earlier upon a physician's recommendation, between May 2017 and November 2021, at the Department of Gastroenterology, Thomayer University Hospital (Prague, Czech Republic) and the Department of Hepatogastroenterology, Institute for Clinical and Experimental Medicine (Prague, Czech Republic). The present study was reviewed and approved by the Institute for Clinical and Experimental Medicine and Thomayer University Hospital, Institute of Experimental Medicine Ethics Committee (Prague, Czech Republic; approval no. G-17-03-02). All patients provided written informed consent, prepared in accordance with the guidelines of The Declaration of Helsinki.

The collected biopsies were immediately snap-frozen and stored at -80˚C. The characteristics of the adenomas were subsequently determined by pathologists. The demographic and clinicopathological characteristics of the patient cohort, including age, sex, adenoma histology, grade of dysplasia and colonic location, are summarized in Table I.

Nucleic acid isolation. Total genomic DNA was isolated from all samples using a DNA Mini Kit (Qiagen GmbH), according to the manufacturer's instructions. The DNA concentration was measured using a Qubit™ 3.0 fluorometer with Qubit dsDNA High Sensitivity and Broad Range Assay Kits (Invitrogen; Thermo Fisher Scientific, Inc.), and DNA purity was assessed by OD260/280 and OD260/230 ratios using a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Inc.).

Gene panel design and selection criteria. A custom 176-gene panel was designed in-house to target genes with established relevance to CRC development and genomic instability. The gene selection strategy was based on an extensive review of the literature, publicly available cancer genomics resources, including cBioPortal for Cancer Genomics (https://www.cbioportal.org/) and the NCI Genomic Data Commons (https://portal.gdc.cancer.gov/), and previously published and clinically used targeted sequencing panels, including those developed for CRC (16-23). The panel included key CRC driver genes and the components of major oncogenic signaling pathways frequently altered in CRC, such as the Wnt/β-catenin, MAPK, PI3K/Akt and TGF-β pathways, as well as a comprehensive set of genes involved in DNA damage response and DNA repair mechanisms. Probes were designed by Altium International s.r.o. according to our specifications, targeting the exonic regions defined by NM [NCBI Reference Sequence (RefSeq) mRNA accession number] reference transcripts with maximal coverage of alternative transcripts for each gene. Gene reference sequences (NM accession numbers) were obtained from the NCBI RefSeq database (https://www.ncbi.nlm.nih.gov/refseq/). The regions of interest included approximately ±10 bp of exon-intron boundaries, both 5' and 3' untranslated regions and promoter regions (-500 bp upstream of the transcription start site). This approach ensured comprehensive coverage of the coding and regulatory regions relevant to CRC-associated genetic alterations.

Library preparation and DNA sequencing of a panel of selected genes. A total of 250-300 ng DNA, isolated from CRA and adjacent mucosa, was used for library preparation according to the manufacturer's instructions of the SureSelect XT HS Target Enrichment System for Illumina Paired-End Multiplexed Sequencing [SureSelect XT Low Input Reagent Kit (1-96) cat. no. G9703A, SureSelect Enzymatic Fragmentation Kit cat. no. 5191-4080; Agilent Technologies, Inc.]. DNA concentration, quality and purity of the samples were assessed using a Qubit™ 3.0 fluorometer (Invitrogen; Thermo Fisher

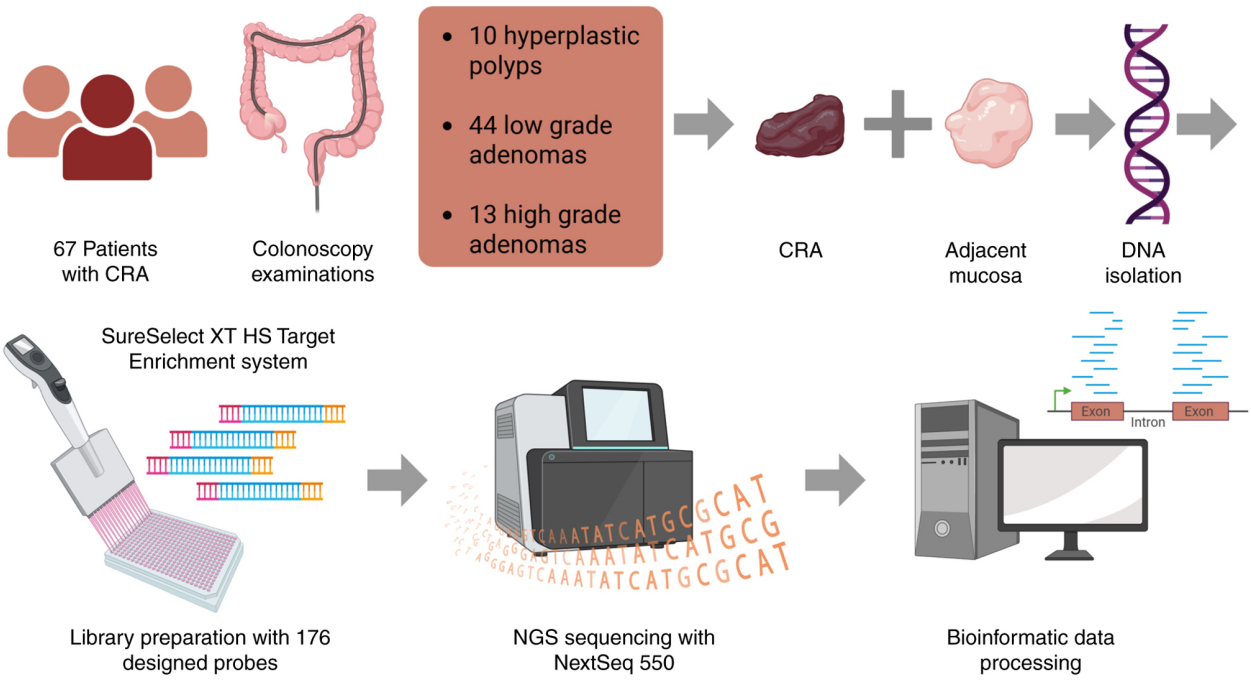

Figure 1. Methodology diagram of the present study. Created using BioRender.com. CRA, colorectal adenoma; HS, high sensitivity; NGS, next-generation sequencing.

Scientific, Inc.) and NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Inc.). The library preparation process included enzymatic DNA fragmentation, end-repair (dA-tailing), adapter ligation with molecular barcodes and amplification using SureSelect XT HS Index Primer adapters (cat. no. 5190-6444; Agilent Technologies, Inc.). Samples were then pooled (48 samples per run) and hybridization with custom-designed probes targeting 176 genes (Table SI) was performed, followed by amplification of the targeted regions.

Each step was accompanied by purification and size selection of fragments using Agencourt® AMPure® XP Magnetic Beads (New England Biolabs) and Dynabeads™ MyOne™ Streptavidin T1 (Invitrogen; Thermo Fisher Scientific, Inc.). Library quality and fragment size were verified using an Agilent High Sensitivity DNA chip on a 2100 Bioanalyzer System (Agilent Technologies, Inc.).

The resulting library, with an average fragment size of 300 bp, was sequenced on the NextSeq 550 Sequencing System (Illumina, Inc.) with NextSeq 500/550 Mid Output Kit v2.5 (150 cycles) (cat. no. 20024904; Illumina, Inc.) using paired-end sequencing (2x125 bp) with a target of ≥10 million reads per sample. The final library was loaded at a concentration of 1.5 pM in 1.3 ml standard loading volume.

Data processing. Standard bioinformatics tools were used for data processing. Quality control was conducted using FastQC (v0.11.9, https://www.ncbi.nlm.nih.gov/sra/PRJNA1441829) and MultiQC (v1.21) tools (24), while Trimmomatic (v0.39) (25) was employed for preprocessing sequencing data, including adapter removal, trimming of low-quality sequences and filtering short reads. Sequencing reads were subjected to quality trimming, with 95% of reads passing the applied filters (minimum read length of 36 bp, sliding window of four bases with a minimum average quality score of 15 and retention of properly paired reads). After trimming, the mean sequencing coverage for a sample was 105x and the alignment rate was >99%. The Burrows-Wheeler Aligner (version 0.7.17-r1188) (26) was used to map the reads to the human GENCODE genome reference (version 38). The mean sequencing coverage across samples was 105.2x (range: 14.9-279.2x). Duplicate reads were removed with PicardTools (version 3.3.0.). The Genome Analysis Toolkit (GATK; version 4.6.0.0.) (27) pipeline was applied to identify somatic variants using the 'Mutect2' tool. The default filtering parameters of GATK were used to exclude sequencing artefacts and low-confidence calls. Furthermore, for each patient, sequencing data from CRA were paired with data from adjacent mucosa, and the Broad Institute Panel of Normals (https://console.cloud.google.com/storage/browser/gatk-best-practices/somatic-hg38%2F?prefix=) was used as a normal reference to remove germline variants. Variant annotation and functional prediction were performed using SnpEff (version 5.2.c) (28). Only the variants predicted to have a higher functional impact on the protein function (nonsense, frameshift and missense variants) were considered as PPVs. In addition, only variants with a read depth >60 and variant allele frequency (VAF) >10% were considered.

Statistical analysis. To assess associations between PPVs and the CRA grade groups, the Fisher's exact test was used. $P<0.05$ was considered to indicate a statistically significant difference. The median read depth of the reported somatic PPVs was 239.5x (range: 63.0-560.0x) and the median VAF was 28.45% (range: 10.04-75.30%), as summarized in Table SII.

Results

Library preparation was successfully completed and sequenced for all samples. The present study focused on

Table I. Demographic and clinicopathological characteristics of patients with colorectal adenoma (n=67).

Characteristic	Value
Sex, n (%)	
Female	23 (34.0)
Male	44 (66.0)
Median age, years (IQR)	65 (55-72)
Histology, n (%)	
Hyperplastic	10 (14.9)
Tubular	30 (44.8)
Tubulovillous	24 (35.8)
Serrated	3 (4.5)
Grade, n (%)	
Hyperplastic polyp	10 (14.9)
Low-grade	44 (65.7)
High-grade	13 (19.4)
Location, n (%)	
Proximal colon[a]	25 (37.3)
Distal colon[b]	29 (43.3)
Rectum	13 (19.4)

[a]Cecum, colon ascendens, flexura hepatalis and colon transversum; [b]flexura lienalis, colon descendens, sigmoideum and rectosigmoideum.

mutations in 176 genes, primarily associated with CRC development in precancerous stages, using NGS methods. When comparing adenoma tissue with the corresponding adjacent mucosa focusing on somatic alterations, PPVs were identified in 44 out of 67 (66%) patients, the most frequent detection of these PPVs was identified in the *APC* (55.2%), *KRAS* (20.9%), *FBXW7* (6.0%), *BRAF* (4.5%) and *MAP2K4* (4.5%) genes (Fig. 2; Table II).

The *APC* gene exhibited the highest PPV frequency, with 47 distinct PPVs identified in CRA tissue (Table SII). The most common types were frameshift mutations (51.1%) and nonsense mutations (44.7%), found in a total of 37 patients. Notably, 19 patients carried >1 mutation in *APC*; one patient had four mutations, one patient had three mutations and 16 patients had two mutations. There was no significant difference in the *APC* mutation rate between low- and high-grade CRAs (P=0.52; Table III), supporting *APC* as the early driver mutation. The second most frequently mutated gene in CRA was *KRAS*, detected in 14 patients with eight mutation variants and predominantly in CRA samples with villous features. In addition, the *KRAS* mutation rate was significantly more frequent in high-grade CRAs compared with low-grade (P<0.01; Table III), suggesting an association between *KRAS* mutations and advanced dysplastic changes in CRA. No formal statistical analysis of mutation frequency according to histological subtype was performed due to the limited number of cases in individual histological categories.

In 10 CRA samples, mutations were detected in >3 different genes. The sample with the highest number of mutated genes, nine in total, was a tubulo-villous CRA obtained from the sigmoid colon of a 49-year-old man. The second highest number of mutations, six, was found in another tubulo-villous CRA sample collected from the transverse colon of a 53-year-old woman. Notably, both patients were relatively young. In addition, the patient with the highest number of mutations was monitored for 7 years. A recent follow-up total colonoscopy revealed no neoplastic lesions and only occasional diverticula.

Notably, mutations were also detected in three hyperplastic polyps in the entire sample set. In one case, a hyperplastic polyp located in the cecum of a 69-year-old man harbored a mutation in the *APC* gene, along with additional mutations in the *MAP2K4* and *FEN1* genes. In the second case, another hyperplastic polyp that was also retrieved from the cecum, was found in a 77-year-old male patient and carried mutations in *BRAF* and *POLQ* genes. In the third case, a *MAP2K4* mutation was identified in a hyperplastic polyp found in the rectum of a 43-year-old man.

Although hyperplastic polyps are generally considered lesions with a very low risk of progressing to CRC (29,30), the presence of mutations such as *APC* and *BRAF* genes warrants caution. These genetic alterations may indicate an increased risk for CRC development. In the patient with a *BRAF*-mutated hyperplastic lesion, no colorectal pathology was observed during a 4-year follow-up period, suggesting a stable course despite the molecular alteration. Conversely, the patient with an *APC*-mutated hyperplastic polyp underwent further endoscopic assessment after 4 years due to intestinal bleeding, though no advanced neoplasia was detected. The patient harboring a *MAP2K4* mutation, with a history of irritable bowel syndrome, histamine intolerance and suspected celiac disease, underwent repeated endoscopic assessments that revealed only mild chronic gastritis, focal intestinal metaplasia and transient terminal ileitis, without histological evidence of inflammatory bowel disease or advanced neoplasia.

Although *TP53* is often one of the earliest mutated genes in the development of CRC, in the present cohort, its mutation was detected only once, in a tubulovillous adenoma with high-grade dysplasia, obtained from the distal colon of a 76-year-old male patient.

In total, 23 out of 67 lesions showed no detectable mutations, comprising predominantly low-grade adenomas (n=13), followed by hyperplastic polyps (n=7) and a small subset of high-grade adenomas (n=3). While the absence of mutations in some high-grade adenomas is unexpected given their typical association with genetic alterations, this distribution still aligned with the general trend of an increasing mutational burden with lesion progression (6,31,32). Notably, no other common clinicopathological features were observed among these mutation-negative samples, suggesting heterogeneity in the underlying biology of early colorectal lesions.

Discussion

Adenoma-carcinoma sequences arising from healthy intestinal epithelia have previously been investigated at the level of somatic DNA mutations in our previous study (6). In that study, mutations in a panel of 11 key genes implicated in colorectal tumorigenesis (*APC*, *BRAF*, *EGFR*, *NRAS*, *KRAS*, *PIK3CA*, *POLE*, *POLD1*, *SMAD4*, *PTEN* and *TP53*) were analyzed using Sanger sequencing. Somatic alterations in these genes

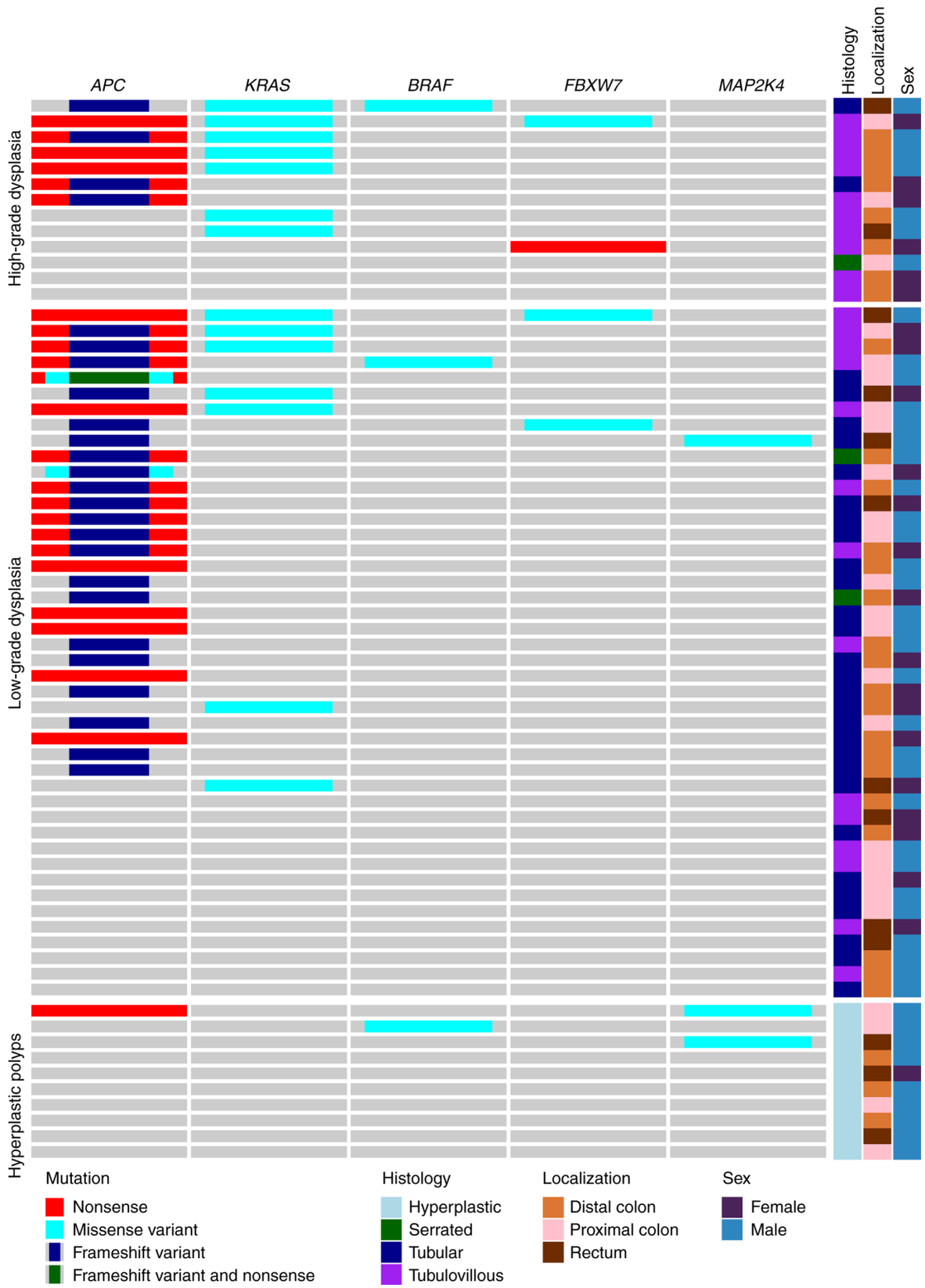

Figure 2. Distribution of genetic alterations detected in all colorectal adenoma samples. Each row represents a patient and each column represents a gene. Different mutation types, histology, localization and sex are indicated by different colors.

were identified not only in carcinoma *in situ* but also in precancerous lesions, with a notable association between mutation frequency and an increasing histological grade of dysplasia, particularly in *APC*, *KRAS* and *TP53* genes.

The present follow-up study was initiated based on the hypothesis that the previously observed low frequency of detectable mutations in CRA samples was largely attributable to the limited scope of the gene panel used. To overcome

Table II. Frequency of putatively pathogenic variants in the most commonly detected genes according to dysplasia status.

Gene name	HP (n=10), n (%)	LG adenoma (n=44), n (%)	HG adenoma (n=13), n (%)	All samples (n=67), %
APC	1 (10)	29 (65.9)	7 (53.8)	55.2
KRAS	0 (0)	7 (15.9)	7 (53.8)	20.9
FBXW7	0 (0)	2 (4.5)	2 (15.4)	6.0
BRAF	1 (10)	1 (2.3)	1 (7.7)	4.5
MAP2K4	2 (20)	1 (2.3)	0 (0)	4.5

HG, high-grade; HP, hyperplastic polyp; LG, low-grade.

Table III. Association between mutation frequency of *APC* and *KRAS* and dysplasia grade.

Gene name	P-value
APC	0.52
KRAS	0.01

P-value calculated by Fisher's exact test (low- vs. high-grade dysplasia).

this limitation, an NGS approach was employed, targeting a broader set of 176 genes, selected for their established or potential roles in colorectal carcinogenesis. These included genes associated with DNA damage response and repair mechanisms, regulation of the cell cycle, apoptosis and key oncogenic signaling pathways (6). Analysis was conducted on an independent cohort consisting exclusively of CRA samples, which allowed for more targeted insights into the molecular features of precancerous colorectal lesions. Library preparation was performed using a hybrid capture method with custom-designed probes, enabling deep coverage across all regions of interest.

Among the most frequently mutated genes were *APC*, *KRAS*, *BRAF*, *FBXW7* and *MAP2K4*, consistent with previous findings (6). *APC* mutations, a hallmark of early Wnt signaling dysregulation (33), were again the most prevalent. *KRAS* and *BRAF* mutations further underscore the activation of the MAPK pathway in early tumorigenesis (34). Notably, *FBXW7*, a tumor suppressor involved in the ubiquitin-proteasome pathway, functions as the substrate recognition component of the SCF (SKP1-CUL1-F-box) E3 ubiquitin ligase complex; it targets several key oncoproteins, including cyclin E, c-Myc, Notch and c-Jun, for phosphorylation-dependent ubiquitination and subsequent proteasomal degradation. Loss-of-function mutations in *FBXW7* impair this degradation process, leading to the accumulation of these oncogenic substrates and promoting cell cycle progression, genomic instability and tumorigenesis (35). Its identification in the present cohort highlights its potential role in early colorectal tumor progression.

An unexpected observation was the absence of detectable somatic PPVs in three high-grade adenomas. Given the advanced histological grade, this finding was biologically counterintuitive and likely reflected technical and biological limitations rather than a true lack of oncogenic alterations. First, the present analysis was restricted to a targeted panel of 176 genes and therefore did not capture the full spectrum of genomic alterations, including mutations outside the selected regions, structural variants or non-coding regulatory changes. In addition, stringent variant filtering criteria, while reducing false positives, may have limited sensitivity in samples with low lesion purity or a substantial admixture of non-neoplastic tissue, which can be particularly relevant in small or heterogeneous adenomas (36,37). Furthermore, driver events in these lesions could be mediated by copy number alterations, chromosomal instability or epigenetic mechanisms such as DNA methylation changes, none of which are reliably detected by targeted mutation profiling. Collectively, the absence of detectable mutations in a small number of high-grade adenomas should not be interpreted as evidence of molecular quiescence, but rather highlights the complexity of early colorectal tumorigenesis and the methodological constraints of targeted sequencing approaches.

This broader panel also enabled the detection of mutations in genes not commonly covered in standard panels, such as *FEN1*, *MAP2K4* (a kinase in the JNK signaling cascade) and *POLQ* (a DNA polymerase involved in alternative end-joining repair). These findings not only supported the role of genomic instability in early lesions but also highlighted a more diverse molecular landscape than previously acknowledged (13,38).

Overall, the present results suggested that broader sequencing approaches can uncover additional alterations of potential relevance for early CRC risk stratification and biomarker development. The observed mutational diversity emphasized the importance of comprehensive molecular profiling in precancerous lesions and may inform future efforts to refine non-invasive diagnostic strategies.

Although relatively few studies (13,14,39) have specifically investigated the mutational landscape of CRC precursor lesions (40), these findings are consistent with the present observations. For example, Wolff *et al* (41) conducted whole-exome sequencing (WES) on DNA isolated from 18 individuals bearing both CRA and matched tumor tissues. Mutations in CRA were identified at frequencies exceeding 11.76% in genes such as *APC*, *TTN*, *TP53*, *KRAS*, *OBSCN*, *SOX9*, *PCDH17*, *SIGLEC10*, *MYH6* and *BRD9*. Notably, mutations in *APC*, *TP53*, *KRAS* and *SOX9* were also observed in tumor samples, whereas *FBXW7* mutations were restricted to CRC tissues, with a frequency of 11%. By contrast, *FBXW7* mutations were already detectable at the CRA stage in the present cohort.

Consistently, Sievers *et al* (42) employed targeted sequencing of the 50 most commonly mutated cancer-associated genes from 36 asymptomatic patients identified at normal CRC screening and identified *APC*, *KRAS*/*NRAS*, *BRAF*, *FBXW7* and *TP53* as the most frequently mutated genes in small intestinal polyps. Furthermore, Smit *et al* (43) demonstrated, in an *in vitro* model using intestinal organoids, that mutations in *APC*, *KRAS*, *SMAD4* and *TP53* were key drivers of global mRNA deregulation during the adenoma-to-carcinoma transition.

In another study, Lin *et al* (13) performed WES on DNA derived from 149 CRA samples and their matched blood samples, and subsequently proposed a 20 gene panel for the detection of precancerous CRA lesions. This panel included numerous genes also identified in the present study, including *TP53*, *KRAS*, *APC*, *PIK3CA* and *FBXW7*. The application of CRC- and CRA-specific targeted gene panels may therefore represent a valuable strategy for the early detection of colorectal neoplasia (23).

Overall, the present findings were in line with published mutational profiling studies of CRA, which consistently identify *APC* alterations and MAPK pathway activation (*KRAS*/*BRAF*) as common early events, with additional recurrent involvement of *TP53* and *FBXW7* in subsets of lesions. In the present cohort, *APC*, *KRAS* and *BRAF* represented the most frequently altered genes, supporting their central role in early colorectal neoplasia. Notably, *FBXW7* alterations were already detected at the CRA stage, whereas a number of studies have reported *FBXW7* enrichment predominantly in CRC tissues, which may reflect differences in cohort composition, lesion size/grade or sequencing design. Differences across studies are also expected given the variation in gene panel breadth, sequencing depth, variant calling thresholds, tissue purity and sample processing (for example, formalin-fixed, paraffin-embedded versus fresh, and microdissected versus bulk tissue).

Notably, PPVs (in genes such as *BRAF* or *MAP2K4*) were detected in a small subset of hyperplastic polyps. Given the limited number of PPVs in hyperplastic lesions in the present cohort and the absence of progression to advanced neoplasia in the available follow-up data, these findings should be interpreted cautiously. In addition, the role of certain variants (such as *MAP2K4* missense changes) in the earliest phases of colorectal tumor initiation has not been well established. Thus, rather than implying a definite neoplastic trajectory, the present results suggested that a minority of lesions classified as hyperplastic/serrated may harbor early molecular alterations the clinical relevance of which remains uncertain, and should be evaluated in larger longitudinal cohorts and functional studies (44-46).

It should also be noted that the classification of variants as putatively pathogenic in the present study was based primarily on predicted functional impact rather than established clinical pathogenicity frameworks. While nonsense and frameshift variants are generally expected to disrupt protein function, the biological relevance of missense variants is more variable and context-dependent (47). In the absence of functional validation or clinical outcome data, some of the reported variants may represent passenger events or alterations with limited biological importance. Therefore, the identified variants should be interpreted as candidates for further functional and longitudinal evaluation rather than as definitive clinically pathogenic mutations.

Future studies should therefore aim to refine variant classification by incorporating additional annotation layers and validation strategies. This includes the use of numerous *in silico* prediction tools, such as SIFT, PolyPhen-2, MutationTaster and CADD, population frequency filtering in large reference cohorts, and standardized classification frameworks such as American College of Medical Genetics and Genomics/Association for Molecular Pathology guidelines (47). Notably, functional validation of selected candidate variants, particularly missense changes in genes such as *MAP2K4*, *FEN1* and *POLQ*, will be necessary to distinguish true driver events from passenger or benign alterations. Integration of molecular data with longitudinal clinical follow-up will further enable assessment of the prognostic and biological relevance of these variants in the context of colorectal lesion progression.

The present study is limited by the lack of longitudinal follow-up and functional validation of the identified variants, which restricted the ability to draw causal conclusions. Nevertheless, a key strength lies in the use of the extended targeted sequencing panel of 176 cancer-associated genes applied to real-world screening samples, which enabled the detection of a broad mutational spectrum and rare molecular alterations detected in hyperplastic polyps traditionally considered low-risk, the clinical relevance of which remains to be determined. In conclusion, these findings provide novel insights into the molecular landscape of precancerous colorectal lesions, and highlight the potential of comprehensive mutational profiling to refine risk stratification and guide early detection strategies.

Acknowledgements

The authors would like to thank Ms. Frances Zatrepalkova (Institute of Experimental Medicine of the Czech Academy of Sciences, Prague, Czech Republic) for English language proofreading of the manuscript.

Funding

The present work was financially supported by the Czech Science Foundation (grant no. 22-05942S) and the Project National Institute for Cancer Research (EXCELES program; grant no. LX22NPO5102) funded by the European Union (NextGenerationEU) and the Cooperatio program (research area: 'oncology and hematology').

Availability of data and materials

The data generated in the present study may be found in the National Center for Biotechnology Information under accession number PRJNA1441829 or at the following URL: https://www.ncbi.nlm.nih.gov/sra/PRJNA1441829.

Authors' contributions

AV contributed to methodology, performed DNA extraction and library preparation for sequencing, wrote the original

draft and contributed to figure preparation. MU contributed to methodology, management and preprocessing of sequencing data, and the reviewing and editing of the manuscript. JH was responsible for validation, formal analysis, preprocessing and quality control of sequencing data, and graphical representation of sequencing data. JJ, JK, TH, VM, SS, PK and RM contributed to the provision of patient samples and clinicopathological data, and to reviewing and editing the manuscript. PV contributed to the conceptualization and design of the study, interpretation of the data, and critically revised the manuscript for important intellectual content. VV contributed to funding acquisition, conceptualization, writing the original draft, reviewing and editing, and supervision. MU and JH confirm the authenticity of all the raw data. All authors have read and approved the final version of the manuscript.

Ethics approval and consent to participate

The present study involving humans was conducted in accordance with The Declaration of Helsinki, and was reviewed and approved by the Institute for Clinical and Experimental Medicine and Thomayer University Hospital, Institute of Experimental Medicine Ethics Committee (Prague, Czech Republic; approval no. G-17-03-02). Written informed consent was provided by all patients.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Siegel RL, Giaquinto AN and Jemal A: Cancer statistics, 2024. CA Cancer J Clin 74: 12-49, 2024.
2. Siskova A, Cervena K, Kral J, Hucl T, Vodicka P and Vymetalkova V: Colorectal adenomas-genetics and searching for new molecular screening biomarkers. Int J Mol Sci 21: 3260, 2020.
3. Fearon ER and Vogelstein B: A genetic model for colorectal tumorigenesis. Cell 61: 759-767, 1990.
4. Nguyen LH, Goel A and Chung DC: Pathways of colorectal carcinogenesis. Gastroenterology 158: 291-302, 2020.
5. Li Q, Geng S, Luo H, Wang W, Mo YQ, Luo Q, Wang L, Song GB, Sheng JP and Xu B: Signaling pathways involved in colorectal cancer: Pathogenesis and targeted therapy. Signal Transduct Target Ther 9: 266, 2024.
6. Jungwirth J, Urbanova M, Boot A, Hosek P, Bendova P, Siskova A, Svec J, Kment M, Tumova D, Summerova S, *et al*: Mutational analysis of driver genes defines the colorectal adenoma: In situ carcinoma transition. Sci Rep 12: 2570, 2022.
7. Hetzel JT, Huang CS, Coukos JA, Omstead K, Cerda SR, Yang S, O'Brien MJ and Farraye FA: Variation in the detection of serrated polyps in an average risk colorectal cancer screening cohort. Am J Gastroenterol 105: 2656-2664, 2010.
8. O'Brien MJ, Yang S, Mack C, Xu H, Huang CS, Mulcahy E, Amorosino M and Farraye FA: Comparison of microsatellite instability, CpG island methylation phenotype, BRAF and KRAS status in serrated polyps and traditional adenomas indicates separate pathways to distinct colorectal carcinoma end points. Am J Surg Pathol 30: 1491-1501, 2006.
9. Murcia O, Juarez M, Hernandez-Illan E, Egoavil C, Giner-Calabuig M, Rodríguez-Soler M and Jover R: Serrated colorectal cancer: Molecular classification, prognosis, and response to chemotherapy. World J Gastroenterol 22: 3516-3530, 2016.
10. Rex DK, Ahnen DJ, Baron JA, Batts KP, Burke CA, Burt RW, Goldblum JR, Guillem JG, Kahi CJ, Kalady MF, *et al*: Serrated lesions of the colorectum: review and recommendations from an expert panel. Am J Gastroenterol 107: 1315-1329, 2012.
11. Snover DC, Jass JR, Fenoglio-Preiser C and Batts KP: Serrated polyps of the large intestine: A morphologic and molecular review of an evolving concept. Am J Clin Pathol 124: 380-391, 2005.
12. Dos Santos W, Dos Reis MB, Porto J, de Carvalho AC, Matsushita M, Oliveira G, Syrjänen K, Reis RM and Guimarães DP: Somatic targeted mutation profiling of colorectal cancer precursor lesions. BMC Med Genomics 15: 143, 2022.
13. Lin SH, Raju GS, Huff C, Ye Y, Gu J, Chen JS, Hildebrandt MAT, Liang H, Menter DG, Morris J, *et al*: The somatic mutation landscape of premalignant colorectal adenoma. Gut 67: 1299-1305, 2018.
14. Alquati C, Prossomariti A, Piazzi G, Buttitta F, Bazzoli F, Laghi L and Ricciardiello L: discovering the mutational profile of early colorectal lesions: A translational impact. Cancers (Basel) 13: 2081, 2021.
15. Voorham QJ, Carvalho B, Spiertz AJ, Claes B, Mongera S, van Grieken NC, Grabsch H, Kliment M, Rembacken B, van de Wiel MA, *et al*: Comprehensive mutation analysis in colorectal flat adenomas. PLoS One 7: e41963, 2012.
16. Horackova K, Frankova S, Zemankova P, Nehasil P, Cerna M, Neroldova M, Otahalova B, Kral J, Hovhannisyan M, Stranecky V, *et al*: Low frequency of cancer-predisposition gene mutations in liver transplant candidates with hepatocellular carcinoma. Cancers (Basel) 15: 201, 2022.
17. Soukupova J, Zemankova P, Lhotova K, Janatova M, Borecka M, Stolarova L, Lhota F, Foretova L, Machackova E, Stranecky V, *et al*: Validation of CZECANCA (CZEch CAncer paNel for Clinical Application) for targeted NGS-based analysis of hereditary cancer syndromes. PLoS One 13: e0195761, 2018.
18. FDA: 510(k) premarket notification: PGDx™ elio tissue complete. US Food and Drug Administration, 2020. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K192063. Accessed January 31, 2024.
19. NantHealth: NantHealth Announces FDA Marketing Authorization of Omics CoreSM: The Nation's First Tumor-Normal Mutation Profiling of Overall Tumor Mutational Burden from Whole Exome Sequencing in Solid Tumors. 2019. https://nanthealth.com/resources/press-releases/nanthealth-announces-fda-marketing-authorization-of-omics-core/. Accessed January 31, 2024.
20. Yoshii Y, Okazaki S and Takeda M: Current status of next-generation sequencing-based cancer genome profiling tests in japan and prospects for liquid biopsy. Life (Basel) 11: 796, 2021.
21. Memorial Sloan Kettering Cancer Center: Next-Generation Tumor Genetic Testing. https://www.mskcc.org/cancer-care/diagnosis-treatment/diagnosing/next-generation-tumor-genetic-testing. Accessed January 31, 2024.
22. Milbury CA, Creeden J, Yip WK, Smith DL, Pattani V, Maxwell K, Sawchyn B, Gjoerup O, Meng W, Skoletsky J, *et al*: Clinical and analytical validation of FoundationOne(R)CDx, a comprehensive genomic profiling assay for solid tumors. PLoS One 17: e0264138, 2022.
23. Gong B, Li D, Kusko R, Novoradovskaya N, Zhang Y, Wang S, Pabón-Peña C, Zhang Z, Lai K, Cai W, *et al*: Cross-oncopanel study reveals high sensitivity and accuracy with overall analytical performance depending on genomic regions. Genome Biol 22: 109, 2021.
24. Ewels P, Magnusson M, Lundin S and Kaller M: MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32: 3047-3048, 2016.
25. Bolger AM, Lohse M and Usadel B: Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120, 2014.
26. Li H and Durbin R: Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26: 589-595, 2010.
27. van derAuwera G and O'Connor BD: Genomics in the cloud: using docker, GATK, and WDL in Terra. 1st edition. O'Reilly Media, Sebastopol, CA, USA, pp 467, 2020.
28. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X and Ruden DM: A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6: 80-92, 2012.
29. Sullivan BA, Noujaim M and Roper J: Cause, epidemiology, and histology of polyps and pathways to colorectal cancer. Gastrointest Endosc Clin N Am 32: 177-194, 2022.

30. Mallick TR and Hasan M: Serrated polyps in the colorectum: Clinical significance and future directions. BMJ Open Gastroenterol 12: e001927, 2025.
31. Cho KR and Vogelstein B: Genetic alterations in the adenoma-carcinoma sequence. Cancer 70: 1727-1731, 1992.
32. Karczmarski J, Goryca K, Pachlewski J, Dabrowska M, Pysniak K, Paziewska A, Kulecka M, Lenarcik M, Mroz A, Mikula M and Ostrowski J: Mutation profiling of premalignant colorectal neoplasia. Gastroenterol Res Pract 2019: 2542640, 2019.
33. Song P, Gao Z, Bao Y, Chen L, Huang Y, Liu Y, Dong Q and Wei X: Wnt/β-catenin signaling pathway in carcinogenesis and cancer therapy. J Hematol Oncol 17: 46, 2024.
34. Singh SR, Bhaskar R, Ghosh S, Yarlagadda B, Singh KK, Verma P, Sengupta S, Mladenov M, Hadzi-Petrushev N, Stojchevski R, *et al*: Exploring the genetic orchestra of cancer: The interplay between oncogenes and tumor-suppressor genes. Cancers (Basel) 17: 1082, 2025.
35. Al-Shamsi HO, Jones J, Fahmawi Y, Dahbour I, Tabash A, Abdel-Wahab R, Abousamra AO, Shaw KR, Xiao L, Hassan MM, *et al*: Molecular spectrum of KRAS, NRAS, BRAF, PIK3CA, TP53, and APC somatic gene mutations in Arab patients with colorectal cancer: Determination of frequency and distribution pattern. J Gastrointest Oncol 7: 882-902, 2016.
36. Aran D, Sirota M and Butte AJ: Systematic pan-cancer analysis of tumour purity. Nat Commun 6: 8971, 2015.
37. Sottoriva A, Kang H, Ma Z, Graham TA, Salomon MP, Zhao J, Marjoram P, Siegmund K, Press MF, Shibata D and Curtis C: A big bang model of human colorectal tumor growth. Nat Genet 47: 209-216, 2015.
38. Borras E, San Lucas FA, Chang K, Zhou R, Masand G, Fowler J, Mork ME, You YN, Taggart MW, McAllister F, *et al*: Genomic landscape of colorectal mucosa and adenomas. Cancer Prev Res (Phila) 9: 417-427, 2016.
39. Dhali A, Maity R and Biswas J: Expanding the genetic landscape of colorectal polyposis: Progress and challenges. World J Gastroenterol 31: 112220, 2025.
40. Mamlouk S, Simon T, Tomas L, Wedge DC, Arnold A, Menne A, Horst D, Capper D, Morkel M, Posada D, *et al*: Malignant transformation and genetic alterations are uncoupled in early colorectal cancer progression. BMC Biol 18: 116, 2020.
41. Wolff RK, Hoffman MD, Wolff EC, Herrick JS, Sakoda LC, Samowitz WS and Slattery ML: Mutation analysis of adenomas and carcinomas of the colon: Early and late drivers. Genes Chromosomes Cancer 57: 366-376, 2018.
42. Sievers CK, Zou LS, Pickhardt PJ, Matkowskyj KA, Albrecht DM, Clipson L, Bacher JW, Pooler BD, Moawad FJ, Cash BD, *et al*: Subclonal diversity arises early even in small colorectal tumours and contributes to differential growth fates. Gut 66: 2132-2140, 2017.
43. Smit WL, Spaan CN, de Boer RJ, Ramesh P, Garcia TM, Meijer BJ, Vermeulen JLM, Lezzerini M, MacInnes AW, Koster J, *et al*: Driver mutations of the adenoma-carcinoma sequence govern the intestinal epithelial global translational capacity. Proc Natl Acad Sci USA 117: 25560-25570, 2020.
44. van Herwaarden YJ, Koggel LM, Simmer F, Vink-Börger EM, Dura P, Meijer GA, Nagengast FM, Hoogerbrugge N, Bisseling TM and Nagtegaal ID: RNF43 mutation analysis in serrated polyposis, sporadic serrated polyps and Lynch syndrome polyps. Histopathology 78: 749-758, 2021.
45. Kambara T, Simms LA, Whitehall VL, Spring KJ, Wynter CV, Walsh MD, Barker MA, Arnold S, McGivern A, Matsubara N, *et al*: BRAF mutation is associated with DNA methylation in serrated polyps and cancers of the colorectum. Gut 53: 1137-1144, 2004.
46. Aiderus A, Barker N and Tergaonkar V: Serrated colorectal cancer: Preclinical models and molecular pathways. Trends Cancer 10: 76-91, 2024.
47. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, *et al*: 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 17: 405-424, 2015.