doi:10.4149/neo_2021_210616N804

# Detection of therapeutically relevant and concomitant rare somatic variants in colorectal cancer

Peter MIKOLAJCIK[1], Zora LASABOVA[2,*], Dusan LODERER[3], Marian GRENDAR[3], Michal KALMAN[4], Ivana KASUBOVA[3], Vincent LUCANSKY[3], Alexander Johannes WIEDERHOLD[2], Juraj MARCINEK[4], Tatiana BURJANIVOVA[2], Eva KUDELOVA[1], Martin VOJTKO[1], Adam SVEC[1], Lukas PLANK[4], Jan JANIK[1], Ludovit LACA[1]

*[1]Clinic of Surgery and Transplant Center, Jessenius Faculty of Medicine and University Hospital in Martin, Comenius University in Bratislava, Martin, Slovakia; [2]Department of Molecular Biology and Genomics, Jessenius Faculty of Medicine in Martin, Comenius University in Bratislava, Martin, Slovakia; [3]Biomedical Center Martin, Jessenius Faculty of Medicine in Martin, Comenius University in Bratislava, Martin, Slovakia; [4]Institute of Pathological Anatomy, Jessenius Faculty of Medicine and University Hospital in Martin, Comenius University in Bratislava, Martin, Slovakia*

**Correspondence: zora.lasabova@uniba.sk*

**Received June 16, 2021 / Accepted August 23, 2021**

In colorectal cancer (CRC), clinically relevant biomarkers are known for genome-guided therapy that can be detected by both first and next generation methods. The aim of our work was to introduce a robust NGS assay that will be able to detect, in addition to standard predictive single nucleotide-based biomarkers, even rare and concomitant clinically relevant variants. Another aim was to identify truncating mutations in *APC* and pathogenic variants in *TP53* to divide patients into potentially prognostic groups. A multigene panel with hotspots in 50 cancer-critical genes was used. Finally, 86 patients diagnosed with primary or metastatic colorectal cancer were enrolled. In total, there were identified 163 pathogenic variants, among them in the genes most recurrent mutated in CRC such as *TP53* (49%), the RAS family genes *KRAS* and *NRAS* (47%), *APC* (43%), and *PIK3CA* (15%). In 30 samples, two driver mutations were present in one sample, 11 patients were without any mutations covered by this panel. In one patient, a novel variant in *BRAF* p.D594E was found, not previously seen in CRC, and was concomitant with *KRAS* p.G12A. In *KRAS*, a potentially sensitive mutation to anti-EGFR therapy p.A59T was found along with the *PIK3CA* missense variant p.E545K. It was possible to divide patients into groups based on the occurrence of truncating *APC* variant alone or concomitant with *TP53* or *KRAS*. Our results demonstrate the potential of small multigene panels that can be used in diagnostics for the detection of rare therapeutically relevant variants. Moreover, the division of patients into groups based on the presence of *APC* and *TP53* mutations enables this panel to be used in retrospective studies on the effectiveness of treatment with anti-EGFR inhibitors.

*Key words: colorectal cancer, next-generation sequencing, rare variants, concomitant variants, anti-EGFR therapy*

Colorectal cancer is the third most common type of cancer worldwide and the second leading cause of cancer death [1]. Multi-step carcinogenesis of colorectal cancer is generally accepted, and at least three driver mutations are required for cancer progression [2]. Sequencing of tumor genomes and exomes has revealed several highly prevalent driver variants that are recurrently mutated in tumors, and many less common driver variants [3]. Knowledge of molecular genetic pathogenesis and identification of driver variants have led to the development of personalized therapies. Biomarker-based targeted therapies are used to treat metastatic colorectal cancer (mCRC). Therapeutically relevant biomarkers include *KRAS* or *NRAS* mutations as negative predictors of anti-EGFR therapies, alterations in mismatch DNA repair proteins, and tumor mutation burden for treatment with checkpoint inhibitors, *NTRK*-fusion positive tumors, and tumors overexpressing or amplifying *HER2* [4–7]. Patients with the mCRC harboring *BRAF* V600E mutation are treated with combined targeted therapy with BRAF and MEK inhibitors, which in this case leads to a significantly longer overall survival [8]. These seven predictive markers provide key information for genomic-driven treatment decisions in metastatic colorectal cancer, and six of them can be detected using simple first-generation methods. The recently published recommendation for the use of NGS technology to detect these biomarkers does not yet see the need for daily NGS testing, however recommends multigene panels as an alternative to PCR if no extra costs

are incurred and if these panels can detect all mentioned markers [9]. It was also considered important that the panels should include all clinically relevant alterations inclusive of those for which drugs are under development, and clinical research centers should also reconsider the routine use of NGS [9]. The advantage of NGS is that patients can be tested for variants for which drugs are still under development, or detect variants that are outside the range of first-generation methods but patients could still benefit from therapy. A similar platform has been successfully established to identify the late-stage malignancies in cases of resistance to guideline-based treatment, resulting in the successful implementation of molecular profiling and enabling further therapy in 24% of cases [10].

A negative result in mutation analysis in the *KRAS* and *NRAS* genes is a prerequisite for the success of the anti-EGFR therapy however, about half of patients with wild-type (wt) genes, especially in the case of right-sided CRC, still fail to respond to this targeted therapy [11–13]. One of the causes may be mutations in the *BRAF* gene [14] but even these do not explain all the cases. A robust prognostic classification of CRC depending on the presence of truncating *APC* mutations and/or *KRAS* and *TP53* pathogenic variants has been published [15]. It was later shown that truncating *APC* variants could also play a predictive role in the refinement of drug-sensitive subpopulations to improve treatment with anti-EGFR monoclonal antibodies [16].

The aim of this work is to implement the NGS method to identify the most common mutations in colorectal cancer using a 50-gene panel so that validation of detected variants using the Sanger method is not necessary. The rationale for this approach is to test the assumption that this method detects common as well as rare therapeutically relevant variants, and concomitant variants with a potential effect on therapy, preferably anti-EGFR therapy. Another aim was to test whether this panel is suitable for the stratification of patients according to a prognostic algorithm based on mutations in the *APC* and *TP53* genes to divide patients into potentially prognostic groups for further studies.

**Table 1. Clinical characteristics of patients and samples included in the study.**

| | Samples from primary CRC n=56 | Samples from liver metastases n=32 |
|---|---|---|
| Age | 69 (SD ± 10) | 64 (SD ± 8) |
| Sex | | |
| female | 25 (45%) | 9 (28%) |
| male | 31 (55%) | 23 (72%) |
| BMI | 27.6 (SD ± 4.9) | 28.2 (SD ± 4.3) |
| Localization of the primary tumor | | |
| left-sided | 25 (45%) | 6 (19%) |
| right-sided | 31 (55%) | 26 (81%) |
| Grading | | |
| G1 | 13 (26%) | 1 (4%) |
| G2 | 27 (54%) | 23 (82%) |
| G3 | 10 (20%) | 4 (14%) |
| NA | 6 | 4 |
| T stage | | |
| T1 | 2 (4%) | 1 (3%) |
| T2 | 11 (20%) | 5 (16%) |
| T3 | 31 (55%) | 23 (72%) |
| T4 | 12 (21%) | 3 (9%) |
| N stage | | |
| N | 32 (57%) | 8 (24%) |
| N1 | 15 (27%) | 16 (52%) |
| N2 | 9 (16%) | 8 (24%) |
| M stage | | |
| M0 | 49 (88%) | NA |
| M1 | 7 (12%) | 32 (100%) |

## Patients and methods

**Patients.** Overall, 88 patients diagnosed with localized or metastatic colorectal cancer were recruited for the study. Finally enrolled were 55 patients diagnosed with colorectal cancer (localized or metastatic CRC), where the colon cancer samples were collected, and 31 patients with metastatic (m) CRC, where the liver metastases were obtained. The patients who underwent surgical intervention at the Surgical Clinic and Transplant Center of Jessenius Faculty of Medicine and University Hospital in Martin were informed about the study and signed informed consent. The study was approved by the Ethics Committee of Jessenius Faculty of Medicine. In the first sequence, the samples were processed as standard within the histopathological diagnostic protocols. Colorectal cancer as a diagnosis was confirmed by experienced pathologists (M.K., J.M.) by morphological examination of formalin-fixed paraffin-embedded (FFPE) sections using standard staining procedures [17]. The pathologist also decided to take a sample for research purposes with a high content of tumor cells (>70%). Demographic and clinical data of patients are summarized in Table 1.

**DNA extraction.** DNA was extracted from colon cancer or liver metastases surgical samples. First, tissue samples were disrupted with a TissueLyser LT II (Qiagen, Hilden, Germany) for 30 s at 24 Hz. Subsequently, DNA was extracted from the tissue lysate using the DNeasy Blood and Tissue DNA Isolation Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Samples were eluted in 50 µl of elution buffer and stored at –20 °C until use.

**Library preparation, next-generation sequencing (NGS), and bioinformatic evaluation.** Prior to library preparation, DNA concentration was determined using the Qubit ds DNA BR Assay Kit (Thermo Fisher Scientific, Waltham, USA) according to the manufacturer's instructions. Samples were analyzed using the AmpliSeq for Illumina HotSpot Panel v2 targeted sequencing kit (Illumina, San Diego, USA), which contains 50 cancer-related genes and amplifies approximately

2,800 mutation hotspots of cancer critical genes generating 107 bp amplicons. When preparing the library, the manufacturer's instructions were followed. Briefly, libraries were prepared from 10 ng of DNA isolated from fresh surgical specimens, the DNA fragments were amplified after determining the DNA concentration, followed by partial digestion of the amplicons with FuPa reagent. In the next step, indices were ligated for dual-index sequencing, fragments were purified using AMP Pure XP Beads (Beckman Coulter, Brea, USA). Subsequently, a second amplification step of the amplified target fragments with adapters was performed, followed by a second clean-up with AMP Pure XP Beads and the library check. The quantity of the library was estimated with Agilent Bioanalyzer 2100 using Agilent DNA 1000 kit (Agilent Technologies, Santa Clara, USA), and the concentration was measured using the Qubit DNA HS Assay Kit (Thermo Fisher Scientific, Waltham, USA). After molarity calculation, the libraries were diluted to starting concentration of 2 nmol and 24 samples were pooled to a final loading concentration of 12 pmol each. Sequencing was performed on a MiSeq Reagent Kit v2 flow cell (300 cycles) (Illumina, San Diego, USA) on a MiSeq instrument. As an NGS reference, a control sample HD731 (Horizon Discovery, Cambridge, UK) with exact characterized variant frequencies from 30% down to 4% was used. The obtained FASTQ files were mapped to the reference sequence of the human genome hg19 using CLC Biomedical Workbench software (Qiagen, Hilden, Germany) – applying the Targeted Sequencing Algorithm (TSA), and somatic candidate variants were called. The frequency of variants in the HD731 was compared with real sequencing results. According to these results, filters were adapted to identify the variants. The variant calling used filtering for more than 400 reads and a variant frequency of at least 3.5% for single nucleotide variants (SNVs) and INDELs. Minimum coverage was defined as at least 5 on each side [18]. Homopolymers longer than 11 were excluded from the evaluation. The ClinVar and COSMIC databases were used to confirm the pathogenicity of the identified variants. Selected identified SNVs were chosen for the confirmation by the Sanger method. The patients were divided into prognostic classes according to the prognostic algorithm suggested by Schell et al. [15].

**Validation by Sanger sequencing.** Selected mutations identified by AmpliSeq for Illumina HotSpot Panel v2 that showed a frequency from 5% upwards, were validated by Sanger sequencing according to a standard protocol in our laboratory [19]. Briefly, PCR primers for regions of interest were designed and a classical endpoint PCR was performed in a reaction solution containing 1.5 mM $MgCl_2$, 0.2 mM primers, 0.1 mM dNTPs, 20 ng DNA, and 0.2 U of FasStart Taq polymerase (Roche Diagnostics, Indianapolis, USA). PCR had 30 cycles, denaturation at 95 °C for 30 s, annealing according to the calculated annealing temperature of the corresponding primers for 30 s, and polymerization at 72 °C for 30 s. Subsequently, the PCR products were purified with the NucleoSpin Extract II kit (Macherey-Nagel, Duren, Germany). For sequencing PCR, BigDye Termination kit v1.1 (Thermo Fisher Scientific, Waltham, USA) was used, and cycling at 95 °C for 1 min and 60 °C for 4 min was conducted. After denaturation in HiDi formamide (Thermo Fisher Scientific, Waltham, USA), samples were analyzed in a Genetic Analyzer AB 3500 and evaluated by Chromas software (Technelysium, South Brisbane, Australia).

**Cloning of partial *BRAF* exon 15 sequence for verification of *BRAF* codon 594 mutation by Sanger sequencing.** The fragment of exon 15 containing codon 594 was prepared by PCR using Phusion polymerase (Thermo Fisher Scientific, Waltham, USA) according to the manufacturer's protocol. The EcoRI restriction sites were included in overhanging parts of the primers (IDT) (F: cgggtaccgagctcgaattc TCATAATGCTTGCTCTGATAGGA and R: aaacgacggccagtgaattcGGCCAAAAATTTAATCAGTGGA) and cellular DNA (patient COCA4) served as a template. The fragment of the expected size was digested with EcoRI enzyme (New England Biolabs, Hitchin, UK); in parallel, plasmid pUC19 was linearized with EcoRI enzyme and treated with FastAP (Thermo Fisher Scientific, Waltham, USA). After electrophoresis in 2% low-melt agarose (Bio-Rad Laboratories, Hercules, USA) gel, both insert and vector was isolated by NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, Duren, Germany) and ligated in 3:1 ratio with Rapid DNA Ligation Kit (Thermo Fisher Scientific, Waltham, USA). Subsequently, the ligation mixture was transformed according to the manufacturer's protocol into Stbl3 competent bacteria (Thermo Fisher Scientific, Waltham, USA) and spread on agar plates supplemented with AMP (Thermo Fisher Scientific, Waltham, USA). After overnight incubation in an incubator at 37 °C, 5% $CO_2$, colonies were screened with the utilization of GoTaq G2 Green Master Mix (Promega, Madison, USA) and primers surrounding MCS of pUC19 (F: TGGAATTGTGAGCGGATAAC and R: ATTAAGTTGGGTAACGCCAG). Selected clones were grown in LB medium, shaking overnight at 37 °C, next day prepared as minipreps using PureLink™ Quick Plasmid Miniprep Kit (Invitrogen, Carlsbad, USA) and sequenced.

**Statistical evaluation.** Continuous variables were expressed as the means ± standard deviations, and nominal variables were expressed as frequencies and percentages. Fisher's exact test was used for the calculation of statistical significance. Results of the p-value $<0.05$ were considered statically significant.

## Results

**Quality evaluation of the targeted sequencing panel in three sequencing runs.** AmpliSeq for Illumina Cancer Hotspot Panel v2 contains 207 amplicons that overlap approximately 2,800 hotspot regions of 50 cancer-critical genes. The three test sequencing runs had a total coverage of 14,246,945 (run 1); 15,972,410 (run 2); and 19,888,474 (run

3). The average coverage for the sample was 547,959 in run 1; 638,896 for run 2; and 764,941 for run 3 (Supplementary Table S1). The quality of all runs was high, in the third run all amplicons had an average coverage greater than 1,000 in runs 1 and 2 there was only one amplicon with less coverage than 1,000 (Supplementary Table S1). The homogeneity of average coverage is summarized in Figure 1.

**Reliability of the targeted sequencing evaluation based on control sample with known frequencies and validation using Sanger sequencing.** The reliability of the runs was evaluated by the comparison of verified known frequencies and obtained really frequencies in commercially available control sample HD731 (Horizon Discovery, Cambridge, UK), which was sequenced in two different runs. According to the manufacturer, the minimum frequency of verified variants in the control sample HD731 was 4%, thus a 3% allelic frequency filter was used to determine the presence of verified mutations in this control sample. According to the manufacturer, there is a total of 14 verified mutations in 11 genes, with the H1047R variant in the *PIK3CA* gene having the highest frequency of 30%, and the V600R in the *BRAF* gene with the lowest frequency about 4% (Supplementary Table S2). The control sample also contains 4 variants of the *KRAS* gene, G13D with a frequency of 25%, another three G12C, G12D, and Q61H with a frequency of 5% each. Our sequence runs confirmed the presence of all variants. The *PIK3CA* gene variant had frequencies of 27.22% and 29.79% in the first and second run, respectively (Supplementary Table S2). The variants with the lowest frequency were the V600R of the *BRAF* gene with a frequency of 3.63% in the first run and 4.19% in the second run and the *EGFR* L861Q with a frequency of 3.79% in the first run and 3.7% in the second run. The *KRAS* G13D variant was expected to have a 25% frequency, in run 1 it was 27.19% and in the second run 25.9%. Other variants of the *KRAS* gene were all around 5% as expected. All variants in the control sample are also included in the gene panel. Our results further showed that the tested panel is reliable up to a frequency of about 3.5%, proving the robustness and reliability of the whole method.

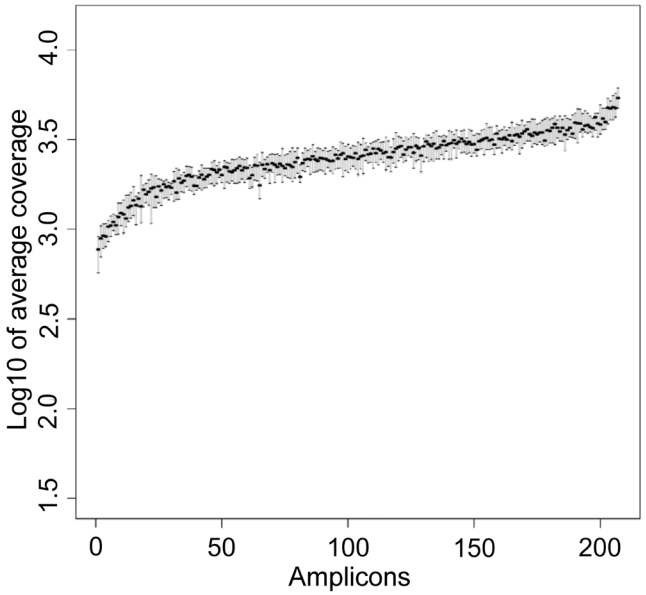


**Figure 1. Average coverage graph of the sequencing run. Average coverage of 207 amplicons of 26 first run samples that were sequenced using the AmpliSeq for Illumina HotSpot Panel gene panel. The average coverage for each amplicon and standard deviation are shown. The amplicons are sorted by ascending coverage.**

Using the Sanger method, we were able to identify variants that were present with a frequency of more than 8% based on the NGS results (data not shown), which represents a relatively high frequency of variants in somatic cells. For this reason, we decided to further rely on the results of HD731 and extrapolation of the results obtained by massive parallel sequencing of HD731 to other variants. Based on these results, we suggested filtering the NGS results in patients' tissue samples by setting the filter to 3.5% and more to find driver variants.

**Somatic variants in colon cancer tissue and liver metastases.** NGS analysis was performed on 55 colon cancer tissue and 31 liver metastases surgical samples obtained from 86 patients with colorectal cancer. In total, we identified 163 somatic variants that were referred to as pathogenic/likely pathogenic (Supplementary Table S3), which represents an arithmetic mean of 1.943 and a median of 2 per sample. The number of mutations in all samples, as well as separately in colonic and metastatic samples is summarized in Table 2. As expected, we identified mutations in genes most often mutated in CRC such as *TP53* (49%), the *RAS* family (47%), of which *KRAS* was mutated in 40% and *NRAS* in 7%, *APC* (43%), and *PIK3CA* (15%) of samples. In colonic samples, the most frequent mutated genes were *TP53* genes (45%) and *RAS* family genes (45%), of which *KRAS* and *NRAS* represented 38% and 7%, respectively, followed by mutations in the *APC* gene (38%) and *PIK3CA* (13%). Pathogenic somatic variants were also identified in *FBXW7* (15%), *BRAF* (9%), *SMAD4* (5%), *CTNNB1* (4%). The most frequently mutated genes in liver metastases were *TP53* (54%), *APC* (52%), *RAS* family genes *KRAS* and *NRAS* (49%), *PIK3CA* (19%), *SMAD 4* (10%). We did not find a statistically significant difference between the occurrence of certain mutations when compared to the mutations in colonic samples and liver metastases (Table 2).

From the total number of 86 samples, there were 0 to 2 driver variants in 62 cases, of which the most common were 2 variants in 34 cases, and in 11 cases no driver variant was found. In 24 cases, we identified more than two driver variants in one sample (3 to 5); in three patients, we found five driver variants in one sample. The occurrence of driver variants between colon cancers samples and liver metastases with 0–2 and 3–5 variants, as well as the significance of gradual increase of driver variants in one sample were

analyzed, but no significant association was found (Supplementary Table S4).

The most common mutations found in the *TP53* gene were non-synonymous variants (36 of 44), among them with the highest prevalence were p.R282W, p.R273C (or H), and p.R175H. In the *APC* gene, there were in total 38 mutations, among them were 34 stop-gain or frameshift variants within the mutation cluster region and beta-catenin binding site between codons 1309 to 1556; in only 4 cases was found the stop-gain mutation in codon 876, in two cases in combination with L1450* and R1556fs, respectively. Missense mutations in the *CTNNB1* gene p.T41A and p.S45F were identified in only two cases in colon cancer tissue samples. The *KRAS* or *NRAS* mutations were detected in 40 cases with the most prevalent *KRAS* p.G12D in 7 cases, followed by *KRAS* p.G12V in 6 cases, *KRAS* p.G13D in 5 cases, and *KRAS* p.G12A in 4 cases. There were also identified two rare variants in the *KRAS* gene, p.A59T and *p.A146T*. Other often-occurring mutated genes in CRC include *PIK3CA*, with mutations most frequently present in exons 9 and 20. In our cohort, there were 8 out of 12 driver variants found in both exon 9 or 20, and four pathogenic variants in exon 1. The most frequent mutation in the *BRAF* gene p.V600E was found in three cases. In addition to these *BRAF* mutations, two rare variants p.D594E and p.G509A were detected as concomitant with *KRAS* and *NRAS* variants, respectively. The most common mutations in the *FBXW7* gene were found at codon 465 represented in four cases by the non-synonymous variant p.465C and in one case by p.465H. In the *SMAD4* gene were identified seven different, mainly non-synonymous variants. The detected variants are listed in detail based on the presence in the individual samples in Supplementary Table S3 and the number and percentage of detected variants are summarized in Table 2.

**Identification of concomitant variants.** When evaluating concomitant mutations, the most prevalent mutual presence of mutations was found in 19 cases for the genes *APC* and *TP53*. The *APC* and *TP53* genes were mutated in colon cancer tissues and metastatic samples in 10 and 9 cases, respectively. The difference was not statistically significant (p=0.2848). Other concomitant mutations events are summarized in Table 3. When focusing attention on the combinations referred to as mutually exclusive events, two such cases were observed, one case with the co-occurrence of *KRAS* p.G12A and *BRAF* p.D594E, and the other case with *NRAS* p.G13D and *BRAF* p.G509A (Table 3, patient ID COCA4 and COCA29). From 10 detected pathogenic variants in *PIK3CA* were 8 concomitants with mutations in *KRAS/NRAS/BRAF* and 4 with variants in *APC* (Table 3). *CTNNB1* mutations p.T41A and p.S45F co-occurred with *KRAS* p.G12S and *PIK3CA* H1047R, respectively, and were mutually exclusive with *APC* mutations.

**Identification of therapeutically relevant somatic high prevalent versus rare variants of potential therapeutic significance.** The AmpliSeq for Illumina HotSpot Panel v2 gene panel is designed mainly for cancer critical genes hotspots, therefore, as expected, it mainly detects recurrent driver variants. Mutations in codons 12/13/61 of the *KRAS* and *NRAS* genes were present in 32 and 6 patients

**Table 2. Calculated frequencies and total number of identified pathogenic/likely pathogenic variants in colon cancer tissue samples and liver metastases together, and separately in colon samples and liver metastases.**

| Most frequent mutated genes | % | Total (n=86) | % | total in pCRC (n=55) | % | total in mCRC (n=31) | p-value pCRC vs. mCRC |
|---|---|---|---|---|---|---|---|
| *TP53* | 49 | 42 | 45 | 25 | 54 | 17 | 0.5015 |
| *APC* | 43 | 37 | 38 | 21 | 52 | 16 | 0.2618 |
| *KRAS* | 40 | 34 | 38 | 21 | 42 | 13 | 0.8196 |
| *FBXW7* | 10 | 9 | 15 | 8 | 3 | 1 | 0.1474 |
| *PIK3CA* | 15 | 13 | 13 | 7 | 19 | 6 | 0.5324 |
| *BRAF* | 6 | 5 | 9 | 5 | 0 | 0 | NA |
| *NRAS* | 7 | 6 | 7 | 4 | 7 | 2 | 1.0000 |
| *PTEN* | 3 | 3 | 5 | 3 | 0 | 0 | NA |
| *SMAD4* | 7 | 6 | 5 | 3 | 10 | 3 | 0.6627 |
| *CTNNB1* | 2 | 2 | 4 | 2 | 0 | 0 | NA |
| *VHL* | 1 | 1 | 2 | 1 | 0 | 0 | NA |
| *ATM* | 1 | 1 | 2 | 1 | 0 | 0 | NA |
| *HNF1A* | 1 | 1 | 2 | 1 | 0 | 0 | NA |
| *STK11* | 1 | 1 | 2 | 1 | 0 | 0 | NA |
| *GNAS* | 2 | 2 | 2 | 1 | 3 | 1 | 1.0000 |

Note: number of mutations in individual genes and their proportions in % in the entire cohort, number of mutations in individual genes and their proportions in % in pCRC and mCRC

**Table 3. Summary of the occurrence of concomitant variants identified in samples of colon cancer tissues and liver metastases of individual patients.**

| Patient ID | *KRAS* | *NRAS* | *BRAF* | *PIK3CA* | *CTNNB1* | *APC* | *FBXW7* |
|---|---|---|---|---|---|---|---|
| COCA4 | p.G12A | | p.D594E | | | | |
| COCA5 | p.G12D | | | p.V344G | | | p.R465H |
| COCA16 | p.A59T | | | p.E545K | | L1488fs | |
| COCA21 | p.G12S | | | | | | p.R465C |
| COCA27 | p.G12S | | | p.E545K | | | |
| COCA29 | | p.G13D | p.G509A | p.N345K | | p.Q1429* | |
| COCA47 | p.G12S | | | | p.T41A | | |
| COCA54 | p.G12C | | | p.E545K | | p.E1309fs | |
| COCA72 | | | | p.R88Q | | p.T1556fs | |
| COCA76 | | | p.V600E | p.Q546H | | | |
| COCA77 | | p.G12D | | p.N345K | | | |
| COCA79 | p.A146T | | | p.E542K | | | |
| COCA83 | | | | p.H1047R | p.S45F | | |

of a total of 86, respectively (Supplementary Table S3). In addition, two rare missense variants were detected in the *KRAS* gene. Missense variant p.A146T is listed in ClinVar as NM_033360.4 (KRAS): c.436G>A (p.Ala146Thr), its frequency reported by the MyCancerGenome database is 0.36% in patients with CRC. The second rare variant found in the *KRAS* gene is p.A59T, listed in ClinVar as NM_033360.4 (KRAS): c.175G>A (p.Ala59Thr), and its frequency according to the MyCancerGenome database is 0.05% in patients with CRC. There was also identified concomitant occurrence of *NRAS* variant p.G13D with *BRAF* p.G509A (ClinVar record NM_001374258.1(BRAF):c.1526G>C (p.Gly509Ala), and a unique missense variant p.D594E in the *BRAF* gene (Figure 2A), listed in ClinVar as NM_004333.6 (BRAF): c.1782T>A (p.Asp594Glu), which has not yet been reported in patients with CRC according to the COSMIC database. To confirm the presence of *BRAF* c.1782T>A mutation, a partial sequence of exon 15 was cloned into the pUC19 vector and Sanger sequencing was performed. Out of 44 sequences, 4 clones contained mutated variant A in nucleotide 1782 thus Asp/Glu mutation in codon 594 (Figures 2B, 2C). The therapeutic significance of these variants will be discussed later.

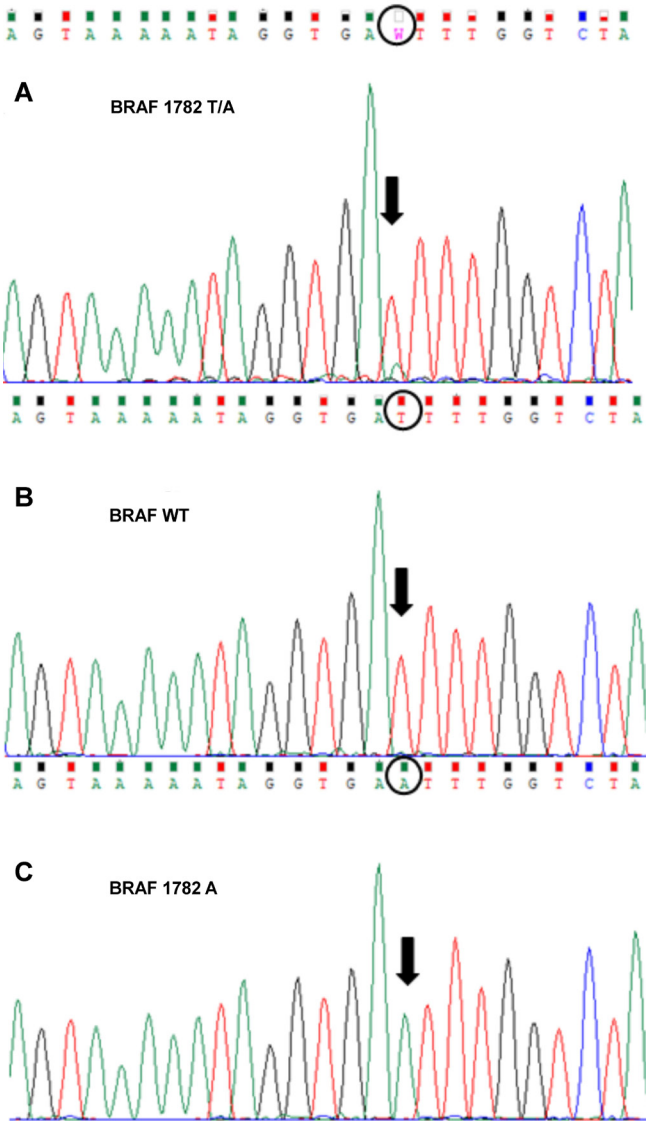


**Figure 2. Result of Sanger sequencing. Sanger sequencing electropherogram showing identification and confirmation of the novel variant NM_004333.6 (BRAF): c.1782T>A (p.Asp594Glu), first time confirmed in colon cancer tissue sample; A) sequence obtained from the patient sample, B) cloned wtDNA, C) cloned mutated DNA.**

**Truncating *APC* mutations as part of a prognostic algorithm and potentially predictive biomarker in colorectal cancer.** When evaluating truncating *APC* variants, it was shown that only the wtAPC allele was present in 51 cases out of 86, of which 35 were in colon cancer samples, and in 16 cases in liver metastases. One *APC* truncating variant alone or concomitant with either the *KRAS* or *TP53* variant was found in 27 cases, of which *APC* and *TP53* were co-occurring in 11 patients (Table 3). In six cases, truncating APC variants were concomitant with *KRAS* and *TP53*, and in two cases 2 truncating *APC* variants were concomitant with *KRAS* and *TP53*. The results are summarized in Table 4 and the patients are divided into potentially prognostic classes.

## Discussion

The accepted thesis of the genetic nature of tumors and subsequent research have shown that the most common driver events in tumors are SNVs and small-scale INDELs, which have also become one of the first targets in the development of advanced therapies. SNVs and small-scale INDELs can be easily detected by first-generation molecular methods including PCR, qPCR, and Sanger methods, but these are becoming impractical and slow, compared to the emerging new technologies such as NGS. NGS obtains a large amount of data and results and raises questions about how to deal with big data and its interpretation in clinical

**Table 4. Patient classification in the prognostic classes according to the algorithm by Schell et al. [15].**

| | Total | Colon cancer samples | Liver metastases samples |
|---|---|---|---|
| Class 0 | **51** | **35** | **16** |
| *APC*wt | 51 | 35 | 16 |
| Class 1 | **27** | **14** | **13** |
| *APC* (1) | 7 | 4 | 3 |
| *APC* (1) *KRAS* | 9 | 6 | 3 |
| *APC* (1) *TP53* | 11 | 4 | 7 |
| Class 2 *APC* (2) alone or + *KRAS* or *TP53* | **0** | **0** | **0** |
| Class 3 *APC* (1) *KRAS TP53* | **6** | **4** | **2** |
| Class 4 *APC* (2) *KRAS TP53* | **2** | **2** | **0** |

Note: the number of truncating mutations that were detected in the APC gene is given in brackets after APC

practice. The primary reason for using NGS should be the clinical utility and the implementation of results into existing and newly developed diagnostic and therapeutic algorithms [9, 20, 21]. Here, we report the implementation and validation of the NGS assay that detects relevant driver SNVs and INDELs using a 50-gene multigene panel capturing the most common hotspots in cancer-critical genes. We report the detection of known clinically relevant genetic biomarkers, but also the detection of rare variants and variants with a new clinical utility. Commercially available control samples and fresh surgical colon cancer or liver metastases were used for validation and implementation, respectively. The sequencing runs were of high quality and with the exception of one, all amplicons had average coverage of more than 1,000. These qualitative criteria correspond to the published reports [22]. The next step of the method validation was to determine the reliability of the detected variants at a frequency of 4%. We used a commercially available control sample with precisely defined frequencies of selected variants and were able to confirm these verified variants with minimal deviations from the variant frequencies specified by the manufacturer in our sequencing runs, thus confirming the robustness and reliability of our protocol. In contrast to the control sample with known frequencies of SNVs in NGS, we were unable to reliably validate the presence of variants present with allelic frequency less than 8% using the Sanger method. We consider this is a consequence of the method itself, for which the sensitivity is reported to be between 10–20% [23, 24].

According to recent research, at least three driver variants are required for the development of colorectal cancer [2]. In our samples, despite the limited number of multigene panel genes tested, we detected 163 variants in 86 samples, which is an average of 1,943 variants per sample, and the median represents two driver events in one sample. Since we used multigene panel with 50 cancer-critical genes, it cannot be expected that all driver events in each patient will be detected, it was not even the aim of this work. According to Vogelstein et al. [3], who summarized the results of whole-exome and whole-genome sequencing into a landscape image, tumors are driven by mutations in a relatively small number of genes that are altered with high prevalence in tumors. These variants are described as "mountains" or highly prevalent variants. In addition, driver mutations are found in many other less prevalent mutated genes representing "hills". Our results confirm that in colorectal tumors, the most common drivers of tumorigenesis are high-prevalent recurring mutations in "mountain" genes, especially SNVs and INDELs in genes such as *TP53, KRAS* or *NRAS, APC, PIK3CA,* and *BRAF*. There were no driver variants in 12.8% of cases, and it can be assumed that gene fusions, large deletions, CNVs, or rare variants of the SNV or INDELS type from the "hills" category are driver events [25], which, however, cannot be detected by our panel. Comprehensive genomic testing or exome sequencing, including RNA analyzes, would be a useful tool in the research of such cases.

For the interpretation of sequencing variants, the joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology is used, which introduced a 5-tier system for the classification of pathogenic variants in the diagnosis of monogenic diseases and in tumor tissues [26, 27]. The continuous increase in therapeutic targets and molecular stratification based on new biomarkers have shown that this classification will not be sufficient. For this reason, several variant classification systems have been proposed taking into account the clinical impact [27–29]. The most important clinically relevant variants are those where the therapy is approved for the tumor and the mutation is responsible for response or resistance, or clinically significant improvement is demonstrated in prospective randomized studies [27, 29]. In our cohort, we detected mainly negative predictors of the anti-EGFR therapy, mutations in codons 12/13/61 of the *KRAS* or *NRAS* genes. However, we found rare missense variants in the *KRAS* gene p.A59T and p.A146T, whose frequency in colorectal carcinomas are 0.05% and 0.36%, respectively [30]. In the literature, p.A59T variants have been reported as potentially pathogenic mutations, but the clinical relevance and utility are not entirely clear, as there are few such patients described in the literature. A single patient case report describes a remarkable clinical, radiographic, and CEA biomarker response after 8 months of treatment with the anti-EGFR inhibitor panitumumab and FOLFRI in the patient with mCRC harboring the A59T mutation. According to the authors, this variant could be a potential exception to the guidelines, but clinical studies focusing on this variant are needed to confirm this claim [31]. In this example, we see the advantage of using a multigene panel that allows the identification of rare but potentially therapeutically relevant SNVs.

*BRAF* is part of the MAPK pathway and the prevalence of mutations in CRC is 8.5% [30]. In our cohort, we detected *BRAF* mutations in 5 patients, which means a prevalence of 6%. The most common activating *BRAF* mutation in CRC is the missense variant p.V600E, which belongs to the class I *BRAF* mutations [31, 32], and in our cohort, it occurred in three out of five cases. Patients with the mCRC harboring V600E mutation are treated with a combination of targeted therapy with BRAF and MEK inhibitors, which in this case leads to a significantly longer overall survival [8].

Although it is still true that mutations in the *KRAS/NRAS* and *BRAF* genes are mutually exclusive, in two cases we have identified just such concomitant events. In one patient, the *NRAS* p.G13D and *BRAF* p.G509A mutations co-occurred in the tumor sample, the other patient had concomitant *KRAS* p.G12A and *BRAF* p.D594E mutations. The *BRAF* mutation p.G509A is prevalent in CRC with 0.16% [30]. Mutations at codon 594 of the *BRAF* gene have been described as very rare in colorectal cancer, with an overall prevalence below 1% [30, 31]. Based on the analysis of available databases such as COSMIC, ClinVar, MyCancerGenome, we assume that

this variant has not yet been seen in CRC tissue. In addition to BRAF kinase activating class I mutations, mutations in *BRAF* are divided into class II mutations causing intermediate to high kinase activity, and class III mutations that have impaired or no kinase activity. The variant p.G509A and the variants in codons 594 belong to the class III *BRAF* mutations [31, 32], therefore we would also include a new variant p.D594E in this category. Because *BRAF* class III mutations have impaired or no kinase activity, activation of the signaling pathway is dependent on activation of the EGFR tyrosine kinase receptor. Due to this dependence, it has been suggested that these tumors could be sensitive to anti-EGFR therapy that prevents EGFR activation, despite the presence of an activating *KRAS* mutation [7, 10]. These findings also underline the importance of the use of a multigene panel in CRC, as they make it possible to identify patients with concomitant variants who would otherwise be excluded from the anti-EGFR therapy.

Wild-type (wt) *APC* in CRC is associated with poor overall survival regardless of the presence of mutations in the *RAS* or *BRAF* genes [15, 34]. In our cohort, we identified as many as 51 such patients (59%), which is more than published in the literature, where the proportion of patients with *wtAPC* is reported to be between 25% and 32% [15, 34]. Our gene panel, focused on SNVs and INDEL detection, cannot detect other typical mutations in the APC gene such as large deletions. Therefore, it will still be necessary to conduct analyzes in patients with *wtAPC* for the presence of large deletions in the *APC* gene, which represent a frequent loss-of-function mutation in this gene. Other factors affecting the WNT signaling pathway, such as methylation of the APC gene promoter and/or mutations in other genes involved in the WNT signaling pathway do not play a significant role [15]. Another important feature of concomitant variants is their relevance within the prognostic algorithm focused on the *APC, KRAS,* and *TP53* genes. Evaluation of these mutations has shown that patients with *wtAPC* and patients with a double truncating mutation in *APC* and a concomitant presence of mutations in the *KRAS* and *TP53* genes have the worst overall survival [15]. In our cohort, we identified truncating *APC* variants in 35 cases, and patients were divided according to the prognostic algorithm proposed in Schell et al. [15] into individual classes (Table 4), which represent different prognostic values, with the worst survival in patients of class 0 and 4, and the best survival of patients in class 1 and 2. Our study is prospective and we do not yet have data on overall survival, but in further studies, we would also focus on these outcomes. We also identified a subclass of patients with a single truncating APC mutation and a pathogenic variant in *TP53*, who should respond better to cetuximab therapy in wtKRAS cancer [16], and retrospective analysis of these data is currently underway, but results are not yet available. We believe that even a smaller multigene panel is useful in identifying such specific conditions, especially when we see that mutations in "mountain" genes are very common in patients with CRC and, based on the results of recent research, their clinical significance is clearly increasing.

For the CRC, the potential therapeutic significance and thus significantly higher clinical relevance can be expected from the presence of *PIK3CA v*ariants [34, 35], which were confidently detected by this panel. However, the use of this gene panel also has certain limitations. The gene panel for the detection of SNVs and INDEL in the hotspots of cancer-critical genes, which we used, does not allow detection of other clinically relevant variants in CRC such as *ERBB2* amplification, MSI or *NTRK* fusions, which proves to be a disadvantage of this panel. As mentioned in the introduction, these variants can be targeted by first-generation methods, but the effort to apply NGS should be a comprehensive detection of the most important clinically relevant variants. On the other hand, we have seen that the most common mutations in CRC are in "mountain" genes, and the advantage of this panel is that it allows the identification of the most common predictive biomarkers, therapeutically relevant rare, and concomitant variants in "mountain" genes, as well as new predictive biomarkers, which can be tested in retrospective studies.

In our work, we introduced a robust and reliable NGS method using a multigene panel across hotspots of cancer-critical genes and showed its clinical utility potential. In addition to detection of common genetic biomarkers predictive for treatment, this panel also allowed the identification of rare variants with potential therapeutic relevance, which can help in the therapeutic clinical decision in molecular tumor boards, and demonstrates the base for retrospective analyses to improve the prediction of treatment with some anti-EGFR inhibitors in case of identifying truncating variants in the *APC* and *TP53* genes.

**Supplementary information** is available in the online version of the paper.

Acknowledgments: This work was supported by grants APVV-16-0066, VEGA 1/0380/18, and project ITMS 313011V446, which was funded by EU resources. We thank Mrs. Vaňochová for her excellent technical assistance.

## References

[1] SUNG H, FERLAY J, SIEGEL RL, LAVERSANNE M, SOERJOMATARAM I et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021; 0: 1–41. https://doi.org/10.3322/caac.21660

[2] TOMASETTI C, MARCHIONNI L, NOWAK MA, PARMIGIANI G, VOGELSTEIN B. Only three driver gene mutations are required for the development of lung and colorectal cancers. Proc Natl Acad Sci USA 2015; 112: 118–123. https://doi.org/10.1073/pnas.1421839112

[3] VOGELSTEIN B, PAPADOPOULOS N, VELCULESCU VE, ZHOU S, DIAZ LA et al. Cancer genome landscape. Science 2013; 339: 1546–1558. https://doi.org/10.1126/science.1235122

[4] ROWLAND A, DIAS MM, WIESE MD, KICHENADASSE G, MCKINNON RA et al. Meta-analysis comparing the efficacy of anti-EGFR monoclonal antibody therapy between KRAS g13D and other mutant metastatic colorectal cancer tumours. Eur J Cancer 2016; 55: 122–130. https://doi.org/10.1016/j.ejca.2015.11.025

[5] OLIVEIRA AF, BRETES L, FURTADO I. Review of PD-1/PD-L1 inhibitors in metastatic dMMR/MSI-H colorectal cancer. Front Oncol 2019; 9: 396. https://doi.org/10.3389/fonc.2019.00396

[6] DOEBELE RC, DRILON A, PAZ-ARES L, SIENA S, SHAW AT et al. Entrectinib in patients with advanced or metastatic NTRK fusion-positive solid tumours: integrated analysis of three phase 1-2 trials. Lancet Oncol 2020; 21: 271–282. https://doi.org/10.1016/S1470-2045(19)30691-6

[7] MERIC-BERNSTAM F, HURWITZ H, RAGHAV KPS, MCAILLIAMS RR, FAKIH M et al. Pertuzumab and trastuzumab for HER2-amplified metastatic colorectal cancer: an updated report from MyPathway , a multicenter, open label, phase 2a multiple basket study. Lancet Oncol 2019; 20: 518–530. https://doi.org/10.1016/S1470-2045(18)30904-5

[8] KOPETZ S, GROTHEY A, YAEGER R, VAN CUTSEM E, DESAI T et al. Encorafenib, binimetinib, and cetuximab in BRAF V600E-mutated colorectal cancer. N Engl J Med 2019; 381: 1632–1643. https://doi.org/10.1056/NEJMos190875

[9] MOSELE F, REMON J, MATEO J, WESTPHALEN CB, BARLESI F et al. Recommendations for the use of next-generaton sequencing (NGS) for patients with metastatic cancers: a report from the ESMO Precision Medicine Working Group. Ann Oncol 2020; 31: 1491–1505. https://doi.org/10.1016/j.annonc.2020.07.014

[10] KIELER M, UNSELD M, BIANCONI D, WANECK F, MADER R et al. Interim analysis of real-world precision medicine platform for molecular profiling of metastatic or advanced cancers: MONDTI. ESMO Open 2019; 9: e000538. https://doi.org/10.1136/esmoopen-2019-000538

[11] DE ROOCK W, CLAES B, BERNASCONI D, DE SCHUTTER J, BIESMANS B et al. Effects of KRAS, BRAF, NRAS and PIK3CA mutations of the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. Lancet Oncol 2010; 11: 753–762. https://doi.org/10.1016/S1470-2045(10)70130-3

[12] VON EINEM JC, HEINEMANN V, VON WEIKERSTHAL LF, VEHLING-KAISER U, STAUCH M et al. Left-sided primary tumors are associated with favorable prognosis in patients with primary KRAS codon12/13 wild-type metastatic colorectal cancer treated with cetuximab plus chemotherapy: an analysis of the AIO KRK-0104 trial. J Cancer Res Clin Oncol 2014; 140: 1607–1614. https://doi.org/10.1007/s00432-014-1678-3

[13] TEJPAR S, STINZING S, CIARDIELLO F, TABERNERO J, VAN CUTSEM E et al. Prognostic and predictive relevance of primary tumor location in patients with RAS wild-type metastatic colorectal cancer. Retrospective analyses of the CRYSTAL and FIRE-3 trial. JAMA Oncol 2017; 3: 194–201. https://doi.org/10.1001/jamaoncol.2016.3797

[14] LAURENT-PUIG P, CAYRE A, MANCAEAU G, BUC E, BACHET JB et al. Analysis of PTEN, BRAF, and EGFR status in determining benefit from cetuximab in wild-type KRAS metastatic colorectal cancer. J Clin Oncol 2009; 27: 5924–5930. https://doi.org/10.1200/JCO.2008.21.6796

[15] SCHELL MJ, YANG M, TEER JK, LO FY, MADAN A et al. A multigene mutation classification of 468 colorectal cancers reveals a prognostic role for APC. Nat Commun 2016; 7: 11743. https://doi.org/10.1038/ncomms11743

[16] YANG M, SCHELL MJ, LOBODA A, NEBOZHYN M, LI J et al. Repurposing EGFR inhibitor utility in colorectal cancer in mutant APC and TP53 subpopulations. Cancer Epidemiol Biomarkers Prev 2019; 28: 1141–1152. https://doi.org/10.1158/1055-9965.EPI-18-1383

[17] LASABOVA Z, KALMAN M, HOLUBEKOVA V, GRENDAR M, KASUBOVA I et al. Mutation analysis of POLE gene in patients with early onset colorectal cancer revealed a rare silent variant within the endonuclease domain with potential effect on splicing. Clin Exp Med 2019; 19: 393–400. https://doi.org/10.1007/s10238-019-0058-7

[18] SEKINE S, YAMASHITA S, TANABE T, KASHIMOTO T, YOSHIDA H et al. Frequent PTPRK-RSPO3 fusion and RNF43 mutations in colorectal traditional serrated adenoma. J Pathol 2016; 239: 133–138. https://10.1002/path.4709

[19] LASABOVA Z, STANCLOVA A, GRENDAR M, MIKOLAJCIKOVA S, CALKOVSKA A et al. Genetic association of single nucleotide polymorphisms of FZD4 and BDNF genes with retinopathy of prematurity. Ophthalmic Genet 2018; 39: 332–337. https://doi.orf/10.1080/13816810.2018.1432064

[20] REMON J, DIENSTMANN R. Precision oncology: separating the wheat from the chaff. ESMO Open 2018; 3: e000446. https://org.doi/10.1136/esmoopen-20198-00046

[21] VAN CUTSEM E, CERVANTES A, ADAM R, SOBRERO A, VANKRIEKEN JH et al. ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. Ann Oncol 2016; 27: 1386–13422. https://doi.org/10.1093/annonc/mdw235

[22] SIMEN BB, YIN L, GOSWAMI CP, DAVIS KO, BAJAJ R et al. Validation of a next-generation-sequencing cancer panel for use in the clinical laboratory. Arch Pathol Lab Med 2015; 139: 508–517. https://odi.org/10.5858/apra.2013-0710-OA

[23] TSIATIS AC, NORRIS-KIRBY A, RICH RG, HAFEZ MJ, GOCKE CD et al. Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations. J Mol Diagn 2010; 12: 425–432. https://doi.org/jmoldx:2010.090188

[24] VANOVA B, KALMAN M, JASEK K, KASUBOVA I, BURJANIVOVA T et al. Droplet digital PCR revealed high concordance between primary tumors and lymph node metastases in multiplex screening of KRAS mutations in colorectal cancer. Clin Exp Med 2019; 19: 219–224. https://doi.org/10.1007/s10238-019-00545-y

[25] CANCER GENOME ATLAS NETWORK. Comprehensive molecular characterization of human colon and rectal cancer. Nature 487; 2012: 330–337. https://doi.org/10.1038/nature11252

[26] RICHARDS S, NAZNEEN A, BALE S, NICK D, DAS S 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 2015; 17: 405–424. https://doi.org/10.1038/gim.2015.30

[27] LI MM, DATTO M, DUNCAVAGE EJ, KULKARNI S, LINDEMAN NI et al. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: A joint consensus recommendation of the Associating for Molecular Pathology, America Society of Clinical Oncology, and College of America Pathologist. J Mol Diagn 2017; 1: 4–23. https://doi.org/10.1016/j.jmoldx.2016.10.002

[28] HORAK P, KLINK B, HEINING C, GROSCHEL S, HUTTER B et al. Presicion oncology based on omics data: The NCT Heidelberg experience. Int J Cancer 2017; 141: 877–886. https://doi.org/10.1002/ijc.30828

[29] MATEO J, CHARKRAVATRY D, DIENSTMANN R, JEZDIC S, GONZALEZ-PEREZ A et al. A framework to rank genomic alteration as targets for cancer precision medicine: the ESMO Scale for Clinical Actionability of molecular Targets (ESCAT). Annals Oncol 2018; 29: 1896–1902. https://doi.org/10.1093/annonc/mdy263

[30] THE AACR PROJECT GENIE CONSORTIUM. AACR Project GENIE: Powering precision medicine through an International Consortium. Cancer Discov 2017; 7: 818–831. https://doi.org/10.1158/2159-8290.CD-17-0151

[31] LOU E, D´SOUZA D, NELSON AC. Therapeutic response of metastatic colorectal cancer harboring a KRAS missense mutation after combination chemotherapy with the EGFR inhibitor panitumumab. J Natl Compr Canc Netw 2017; 15: 427–432. https://doi.org/10.6004/jnccn.2017.0043

[32] ZHENG G, TSENG LH, CHEN G, HALEY L, ILLEI P et al. Clinical detection and categorization of uncommon and concomitant mutations involving BRAF. BMC Cancer 2015; 15: 779. https://doi.org/10.1186/s12885-015-1811-y

[33] BRACHT JWP, KARACHALIOU N, BIVONA T, LANMAN RB, FAULL I et al. BRAF mutations classes I, II, and III in NSCLC patients included in the SLLIP trial: The need for a new pre-clinical treatment rationale. Cancers 2019; 11: 1381. https://doir.org/10.3390/cancers11091381

[34] JURIC D, RODON J, TABERNERO J, JANKU F, BURRIS HA et al. Phospatidylinosytol 3-kinase α-selective inhibition with alpelisib (BYL719) in PIK3CA-altered solid tumors: Results from the first-in-human study. J Clin Oncol 2018; 36: 1291–1299. https://doi.org/10.1200/JCO.2017.72.7107

[35] MISHRA R, PATEL H, ALANAZI S, KILROY MK, GARRETT JT. PIK3CA inhibitors in cancer: Clinical implicaions and adverse effects. Int J Mol Sci 2021; 22: 3464. https://doi.org/10.3390/ijms22073464

https://doi.org/10.4149/neo_2021_210616N804

# Detection of therapeutically relevant and concomitant rare somatic variants in colorectal cancer

Peter MIKOLAJCIK[1], Zora LASABOVA[2,*], Dusan LODERER[3], Marian GRENDAR[3], Michal KALMAN[4], Ivana KASUBOVA[3], Vincent LUCANSKY[3], Alexander Johannes WIEDERHOLD[2], Juraj MARCINEK[4], Tatiana BURJANIVOVA[2], Eva KUDELOVA[1], Martin VOJTKO[1], Adam SVEC[1], Lukas PLANK[4], Jan JANIK[1], Ludovit LACA[1]

## Supplementary Information

**Supplementary Table S1. The total number of reads per sequencing run and the sample average coverage with the minimum and maximum coverage, and the number of amplicons for each coverage.**

| | total coverage | sample average coverage (min;max) | total number of amplicons | number of amplicons with average coverage >1000 | total number of amplicons with minimal coverage | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | >1000 | 500-999 | 200-499 | <200 |
| run1 | 14,246,945 | 547,959 (382,162; 714,703) | 207 | 206 | 123 | 58 | 23 | 3 |
| run2 | 15,972,410 | 638,896 (455,682; 751,167) | 207 | 207 | 191 | 15 | 2 | 0 |
| run3 | 19,888,474 | 764,941 (542,296; 1,005,114) | 207 | 207 | 203 | 4 | 0 | 0 |

**Supplementary Table S2. Comparison of expected and obtained allelic frequencies of the control sample HD731 with the Ampliseq for Illumina Hot Spot Panel in two independent sequencing runs. All variants were present in the gene panel.**

| Gene | Mutation | Expected Allelic Frequency according Horizon in % | Run 1 Coverage | Run 1 Frequency | Run 2 Coverage | Run 2 Frequency |
|---|---|---|---|---|---|---|
| *PIK3CA* | H1047R | 30 | 1543 | 27.22 | 3058 | 29.79 |
| *KRAS* | G13D | 25 | 1953 | 27.19 | 4073 | 25.90 |
| *EGFR* | G719S | 16-Jul | 3801 | 17.76 | 2766 | 17.50 |
| *BRAF* | V600E | 8 | 845 | 07-Oct | 4190 | Jul-64 |
| *ABL1* | T315I | 5 | 1731 | Apr-85 | 3128 | Apr-60 |
| *IDH2* | R172K | 5 | 956 | Apr-39 | 3664 | May-35 |
| *KIT* | D816V | 5 | 7840 | Apr-91 | 3772 | 05-Apr |
| *KRAS* | G12C | 5 | 1951 | Apr-25 | 4105 | Apr-46 |
| *KRAS* | G12D | 5 | 1957 | May-16 | 4098 | 05-Mar |
| *KRAS* | Q61H | 5 | 2304 | Apr-77 | 4821 | 05-Aug |
| *NRAS* | Q61R | 5 | 11493 | May-60 | 3065 | Apr-80 |
| *PDGFRA* | D842V | 5 | 2198 | May-41 | 4522 | May-31 |
| *EGFR* | L861Q | 04-Feb | 1345 | Mar-79 | 4405 | Mar-70 |
| BRAF | V600R | 4 | 853 | Mar-63 | 4229 | Apr-19 |

**Supplementary Table S3. Spreadsheet of all identified driver variants obtained by targeted sequencing with gene panel AmpliSeq for Illumina HotSpot Panel v2 in 86 individual fresh surgical specimens collected from CRC patients; legend of row 3: 1 = primary colonic tissue; 2 = liver metastases**

| Genes present in AmpliSeq for Illumina Hot Spot Gene Panel | COCA1 | COCA2 | COCA3 | COCA4 | COCA5 | COCA6 | COCA7 | COCA8 | COCA9 | COCA10 | COCA11 | COCA12 | COCA13 | COCA14 | COCA15 | COCA16 | COCA17 | COCA19 | COCA20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| primary1/mts2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 |
| MPL | | | | | | | | | | | | | | | | | | | |
| NRAS | | | | | | | | | | | | | G12S | | | | | | |
| ALK | | | | | | | | | | | | | | | | | | | |
| IDH1 | | | | | | | | | | | | | | | | | | | |
| ERBB4 | | | | | | | | | | | | | | | | | | | |
| VHL | | | | | | | | I147V | | | | | | | | | | | |
| MLH1 | | | | | | | | | | | | | | | | | | | |
| CTNNB1 | | | | | | | | | | | | | | | | | | | |
| PIK3CA | | | | | V344G | | | | | | | | | H1047R | | E545K | | | |
| FGFR3 | | | | | | | | | | | | | | | | | | | |
| PDGFRA | | | | | | | | | | | | | | | | | | | |
| KIT | | | | | | | | | | | | | | | | | | | |
| KDR | | | | | | | | | | | | | | | | | | | |
| FBXW7 | | | | | | | | R465C | | | | | | | | R465H | | | |
| APC | | | G1309fs | R1432* | T1556fs | | A1492Pfs*15 | R876* | | | | R876*+ R1450* | | | Q1429* | L1488fs | | | |
| CSF1R | | | | | | | | | | | | | | | | | | | |
| NPM1 | | | | | | | | | | | | | | | | | | | |
| EGFR | | | | | | | | | | | | | | | | | | | |
| MET | | | | | | | | | | | | | | | | | | | |
| SMO | | | | | | | | | | | | | | | | | | | |
| BRAF | | | | D594E | | | | | | | | | | | | | | | |
| EZH2 | | | | | | | | | | | | | | | | | | | |
| FGFR1 | | | | | | | | | | | | | | | | | | | |
| JAK2 | | | | | | | | | | | | | | | | | | | |
| CDKN2A | | | | | | | | | | | | | | | | | | | |
| GNAQ | | | | | | | | | | | | | | | | | | | |
| ABL1 | | | | | | | | | | | | | | | | | | | |
| NOTCH1 | | | | | | | | | | | | | | | | | | | |
| RET | | | | | | | | | | | | | | | | | | | |
| PTEN | | | | | | | | | | | | | | | | | | | |
| FGFR2 | | | | | | | | | | | | | | | | | | | |
| HRAS | | | | | | | | | | | | | | | | | | | |
| ATM | | | | | | | | | | | | | | | | | | | |
| KRAS | | | | G12A | G12D | G12V | | | | Q61L | | G12D | | | G13D | A59T | | | |
| PTPN11 | | | | | | | | | | | | | | | | | | | |
| HNF1A | | | | | | | | | | | | | | | | | | | |
| FLT3 | | | | | | | | | | | | | | | | | | | |
| RB1 | | | | | | | | | | | | | | | | | | | |
| AKT1 | | | | | | | | | | | | | | | | | | | |
| IDH2 | | | | | | | | | | | | | | | | | | | |
| CDH1 | | | | | | | | | | | | | | | | | | | |
| TP53 | C176F | | c.673-1G splice | | | | R273C | | | R175H | | R282W | T273C | | | | | P82fs* | R282W |
| ERBB2 | | | | | | | | | | | | | | | | | | | |
| SMAD4 | | | | | | | | | | | | | | | | | | | |
| STK11 | | | | | | | | | | | | | | | | | | | |
| GNA11 | | | | | | | | | | | | | | | | | | | |
| JAK3 | | | | | | | | | | | | | | | | | | | |
| SRC | | | | | | | | | | | | | | | | | | | |
| GNAS | | | | | | | | | | | | | | | | | | | |
| SMARCB1 | | | | | | | | | | | | | | | | | | | |

**Supplementary Table S3.** ***Continued ...***

| Genes present in AmpliSeq for Illumina Hot Spot Gene Panel | COCA21 | COCA22 | COCA23 | COCA24 | COCA25 | COCA27 | COCA28 | COCA29 | COCA30 | COCA31 | COCA33 | COCA36AM | COCA35 | COCA36 | COCA37 | COCA38 | COCA39 | COCA40 | COCA41 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| primary1/mts2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 2 | | 2 | 2 | 1 | 1 | 2 | 2 | 2 |
| MPL | | | | | | | | | | | | | | | | | | | |
| NRAS | | | | | | | | G13D | | | | | | | | | | | |
| ALK | | | | | | | | | | | | | | | | | | | |
| IDH1 | | | | | | | | | | | | | | | | | | | |
| ERBB4 | | | | | | | | | | | | | | | | | | | |
| VHL | | | | | | | | | | | | | | | | | | | |
| MLH1 | | | | | | | | | | | | | | | | | | | |
| CTNNB1 | | | | | | | | | | | | | | | | | | | |
| PIK3CA | | | | | | E545K | | N345K | | | | | | | | | | | |
| FGFR3 | | | | | | | | | | | | | | | | | | | |
| PDGFRA | | | | | | | | | | | | | | | | | | | |
| KIT | | | | | | | | | | | | | | | | | | | |
| KDR | | | | | | | | | | | | | | | | | | | |
| FBXW7 | R465C | | | | | | R505C | | R505C | | | | | | | | | | |
| APC | L1488fs | E1353* | E1344* | | | | Q1378* | Q1429* | Q1411* | T1556fs | S1436fs | E1288* | | E1288* | | | | | |
| CSF1R | | | | | | | | | | | | | | | | | | | |
| NPM1 | | | | | | | | | | | | | | | | | | | |
| EGFR | | | | | | | | | | | | | | | | | | | |
| MET | | | | | | | | | | | | | | | | | | | |
| SMO | | | | | | | | | | | | | | | | | | | |
| BRAF | | | | | | | | G509A | | | | | | | | | | | |
| EZH2 | | | | | | | | | | | | | | | | | | | |
| FGFR1 | | | | | | | | | | | | | | | | | | | |
| JAK2 | | | | | | | | | | | | | | | | | | | |
| CDKN2A | | | | | | | | | | | | | | | | | | | |
| GNAQ | | | | | | | | | | | | | | | | | | | |
| ABL1 | | | | | | | | | | | | | | | | | | | |
| NOTCH1 | | | | | | | | | | | | | | | | | | | |
| RET | | | | | | | | | | | | | | | | | | | |
| PTEN | Y336* | | | | | | | | | | | | | | | | | | |
| FGFR2 | | | | | | | | | | | | | | | | | | | |
| HRAS | | | | | | | | | | | | | | | | | | | |
| ATM | | | | | | | | | | | | | | | | | | | |
| KRAS | G12S | | | | G12A | G12S | G12D | | G12D | Q61H | | | | | | G12D | | G13C | |
| PTPN11 | | | | | | | | | | | | | | | | | | | |
| HNF1A | | | | | | | | | | | | | | | | | | | |
| FLT3 | | | | | | | | | | | | | | | | | | | |
| RB1 | | | | | | | | | | | | | | | | | | | |
| AKT1 | | | | | | | | | | | | | | | | | | | |
| IDH2 | | | | | | | | | | | | | | | | | | | |
| CDH1 | | | | | | | | | | | | | | | | | | | |
| TP53 | C135S | R282W | R282W | | | | | | | P113L | R273C | R170fs | R282W | R170fs | | | R273C | | R273H |
| ERBB2 | | | | | | | | | | | | | | | | | | | |
| SMAD4 | | | | | R135* | | | | | R361H | | | L495P | | | | | | |
| STK11 | | | | | | | | | | | | | | | | | | | |
| GNA11 | | | | | | | | | | | | | | | | | | | |
| JAK3 | | | | | | | | | | | | | | | | | | | |
| SRC | | | | | | | | | | | | | | | | | | | |
| GNAS | | | | | | | | | | | | | | | | | | | |
| SMARCB1 | | | | | | | | | | | | | | | | | | | |

**Supplementary Table S3.** ***Continued ...***

| Genes present in AmpliSeq for Illumina Hot Spot Gene Panel | COCA43 | COCA44 | COCA45 | COCA46 | COCA47 | COCA48A | COCA49A | COCA50A | COCA52A | COCA53 | COCA54A1 | COCA55 | COCA56 | COCA57 | COCA58 | COCA59 | COCA60A | COCA61A | COCA63A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| primary1/mts2 | 1 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 2 |
| MPL | | | | | | | | | | | | | | | | | | | |
| NRAS | | | | | | | | | Q61K | G13V | | | | | | | | | |
| ALK | | | | | | | | | | | | | | | | | | | |
| IDH1 | | | | | | | | | | | | | | | | | | | |
| ERBB4 | | | | | | | | | | | | | | | | | | | |
| VHL | | | | | | | | | | | | | | | | | | | |
| MLH1 | | | | | | | | | | | | | | | | | | | |
| CTNNB1 | | | | | T41A | | | | | | | | | | | | | | |
| PIK3CA | | | | | | | | | | | E545K | | | | | | | | |
| FGFR3 | | | | | | | | | | | | | | | | | | | |
| PDGFRA | | | | | | | | | | | | | | | | | | | |
| KIT | | | | | | | | | | | | | | | | | | | |
| KDR | | | | | | | | | | | | | | | | | | | |
| FBXW7 | | | | | | | | | | | | | | | | | | | |
| APC | | P1319fs | E1322* | | | | | | | | E1309fs | E1309fs | | R876*, T1556fs | Q1367* | | | I1580fs | T1493fs |
| CSF1R | | | | | | | | | | | | | | | | | | | |
| NPM1 | | | | | | | | | | | | | | | | | | | |
| EGFR | | | | | | | | | | | | | | | | | | | |
| MET | | | | | | | | | | | | | | | | | | | |
| SMO | | | | | | | | | | | | | | | | | | | |
| BRAF | | | | | | | | | | | | | | | | V600E | | | |
| EZH2 | | | | | | | | | | | | | | | | | | | |
| FGFR1 | | | | | | | | | | | | | | | | | | | |
| JAK2 | | | | | | | | | | | | | | | | | | | |
| CDKN2A | | | | | | | | | | | | | | | | | | | |
| GNAQ | | | | | | | | | | | | | | | | | | | |
| ABL1 | | | | | | | | | | | | | | | | | | | |
| NOTCH1 | | | | | | | | | | | | | | | | | | | |
| RET | | | | | | | | | | | | | | | | | | | |
| PTEN | | | | | | | | | | | | | | | | | | | |
| FGFR2 | | | | | | | | | | | | | | | | | | | |
| HRAS | | | | | | | | | | | | | | | | | | | |
| ATM | | | | | | | | | | | | | | | | | | | |
| KRAS | | G12A | | Q61H | G12S | | G13D | G12V | | | G12C | | G13C | G13D | | | G12A | | G12D |
| PTPN11 | | | | | | | | | | | | | | | | | | | |
| HNF1A | | | E275de-lAAG | | | | | | | | | | | | | | | | |
| FLT3 | | | | | | | | | | | | | | | | | | | |
| RB1 | | | | | | | | | | | | | | | | | | | |
| AKT1 | | | | | | | | | | | | | | | | | | | |
| IDH2 | | | | | | | | | | | | | | | | | | | |
| CDH1 | | | | | | | | | | | | | | | | | | | |
| TP53 | Y103N | | P177T | P151T | | | | F134V | R248Q | R141C | | R248W | | R248W | | | N288fs | | |
| ERBB2 | | | | | | | | | | | | | | | | | | | |
| SMAD4 | | | | | | | | | | | | | | | A118V | | | | |
| STK11 | | | | | | | | | | | | | | | | | | | |
| GNA11 | | | | | | | | | | | | | | | | | | | |
| JAK3 | | | | | | | | | | | | | | | | | | | |
| SRC | | | | | | | | | | | | | | | | | | | |
| GNAS | | | | | | | | | R201K | | | | | | | | | | |
| SMARCB1 | | | | | | | | | | | | | | | | | | | |

**Supplementary Table S3.** ***Continued ...***

| Genes present in AmpliSeq for Illumina Hot Spot Gene Panel | COCA64A | COCA65A | COCA66A | COCA67A | COCA68A | COCA69A | COCA70A | COCA71A | COCA72A | CO-CA73A | CO-CA74A | COCA75 | COCA76 | COCA77 | COCA78 | COCA79 | COCA80 | COCA81 | CO-CA82A |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| primary1/mts2 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1,2? | 1 | 2 |
| MPL | | | | | | | | | | | | | | | | | | | |
| NRAS | | | | | | | | | | | | | | G12D | | | | | |
| ALK | | | | | | | | | | | | | | | | | | | |
| IDH1 | | | | | | | | | | | | | | | | | | | |
| ERBB4 | | | | | | | | | | | | | | | | | | | |
| VHL | | | | | | | | | | | | | | | | | | | |
| MLH1 | | | | | | | | | | | | | | | | | | | |
| CTNNB1 | | | | | | | | | | | | | | | | | | | |
| PIK3CA | | | | | | | | | R88Q | | | | Q546H | N345K | | E542K | | | |
| FGFR3 | | | | | | | | | | | | | | | | | | | |
| PDGFRA | | | | | | | | | | | | | | | | | | | |
| KIT | | | | | | | | | | | | | | | | | | | |
| KDR | | | | | | | | | | | | | | | | | | | |
| FBXW7 | | | | | | | | | | | | | R465C | | | | | | |
| APC | | | | I1580fs | | T1493fs | G1312* | | T1556fs | E1286* | E1306* | | | | E1309fs | | R876* | | |
| CSF1R | | | | | | | | | | | | | | | | | | | |
| NPM1 | | | | | | | | | | | | | | | | | | | |
| EGFR | | | | | | | | | | | | | | | | | | | |
| MET | | | | | | | | | | | | | | | | | | | |
| SMO | | | | | | | | | | | | | | | | | | | |
| BRAF | | | | | | | | | | | | | V600E | | | | | | |
| EZH2 | | | | | | | | | | | | | | | | | | | |
| FGFR1 | | | | | | | | | | | | | | | | | | | |
| JAK2 | | | | | | | | | | | | | | | | | | | |
| CDKN2A | | | | | | | | | | | | | | | | | | | |
| GNAQ | | | | | | | | | | | | | | | | | | | |
| ABL1 | | | | | | | | | | | | | | | | | | | |
| NOTCH1 | | | | | | | | | | | | | | | | | | | |
| RET | | | | | | | | | | | | | | | | | | | |
| PTEN | | | | | | | | | K267fs | | | | | | | | | | |
| FGFR2 | | | | | | | | | | | | | | | | | | | |
| HRAS | | | | | | | | | | | | | | | | | | | |
| ATM | | | | | | | | | | | | | | | | | | | |
| KRAS | | | | | | G13D, G13C | G12C | | | G12V | | | | | | A146T | | G12V | G12D |
| PTPN11 | | | | | | | | | | | | | | | | | | | |
| HNF1A | | | | | | | | | | | | | | | | | | | |
| FLT3 | | | | | | | | | | | | | | | | | | | |
| RB1 | | | | | | | | | | | | | | | | | | | |
| AKT1 | | | | | | | | | | | | | | | | | | | |
| IDH2 | | | | | | | | | | | | | | | | | | | |
| CDH1 | | | | | | | | | | | | | | | | | | | |
| TP53 | Q100* | R174*, R361C | R248Q | | | R273C V218del | R175H | | | C135Y | | | | | R248W | | | | R273C |
| ERBB2 | | | | | | | | | | | | | | | | | | | |
| SMAD4 | | | | | | | | | | P356L | P356R | | | | | | | | |
| STK11 | | | | | | | | | | | | | | | | | | | |
| GNA11 | | | | | | | | | | | | | | | | | | | |
| JAK3 | | | | | | | | | | | | | | | | | | | |
| SRC | | | | | | | | | | | | | | | | | | | |
| GNAS | | | | | | | | | | | | | | | | R201H | | | |
| SMARCB1 | | | | | | | | | | | | | | | | | | | |

**Supplementary Table S3.** ***Continued ...***

| Genes present in AmpliSeq for Illumina Hot Spot Gene Panel | COCA83A | COCA84A | COCA86A | COCA87A | COCA88A | COCA89A | COCA90A | COCA91A | COCA92A | COCA93A |
|---|---|---|---|---|---|---|---|---|---|---|
| primary1/mts2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 |
| MPL | | | | | | | | | | |
| NRAS | | | | | | | | | | |
| ALK | | | | | | | | | | |
| IDH1 | | | | | | | | | | |
| ERBB4 | | | | | | | | | | |
| VHL | | | | | | | | | | |
| MLH1 | | | | | | | | | | |
| CTNNB1 | S45F | | | | | | | | | |
| PIK3CA | H1047R | | | E545G | | | | | | |
| FGFR3 | | | | | | | | | | |
| PDGFRA | | | | | | | | | | |
| KIT | | | | | | | | | | |
| KDR | | | | | | | | | | |
| FBXW7 | H500R | S282* | | | | | | | | |
| APC | | | | | | E1353* | Q1367* | S1346* | | |
| CSF1R | | | | | | | | | | |
| NPM1 | | | | | | | | | | |
| EGFR | | | | | | | | | | |
| MET | | | | | | T1010I | | | | |
| SMO | | | | | | | | | | |
| BRAF | | | | | | | | | | V600E |
| EZH2 | | | | | | | | | | |
| FGFR1 | | | | | | | | | | |
| JAK2 | | | | | | | | | | |
| CDKN2A | | | | | | | | | | |
| GNAQ | | | | | | | | | | |
| ABL1 | | | | | | | | | | |
| NOTCH1 | | | | | | | | | | |
| RET | | | | | | | | | | |
| PTEN | L325P | | | | | | | | | |
| FGFR2 | | | | | | | | | | |
| HRAS | | | | | | | | | | |
| ATM | | | | | S333F | | | | | |
| KRAS | | G12V | | | | | | G13D | G12V | |
| PTPN11 | | | | | | | | | | |
| HNF1A | | | | | | | | | | |
| FLT3 | | | | | | | | | | |
| RB1 | | | | | | | | | | |
| AKT1 | | | | | | | | | | |
| IDH2 | | | | | | | | | | |
| CDH1 | | | | | | | | | | |
| TP53 | | R175H | | Y126H | R282W | P301* | R248Q H179R | S127F | R248Q | P151H |
| ERBB2 | | | | | | | | | | |
| SMAD4 | | | | | | | | | | |
| STK11 | V197M | | | | | | | | | |
| GNA11 | | | | | | | | | | |
| JAK3 | | | | | | | | | | |
| SRC | | | | | | | | | | |
| GNAS | | | | | | | | | | |
| SMARCB1 | | | | | | | | | | |

**Supplementary Table S4. Result of the statistical testing. The occurrence of driver variants between colon cancers samples and liver metastases with 0-2 and 3-5 variants, as well as the significance of gradual increase of driver variants in one sample is without statistical significance**

| | Number of driver variants (%) | | | | | Total | p-value (Fisher´s exact test) | with Yates correction |
|---|---|---|---|---|---|---|---|---|
| | **0** | **1** | **2** | **3** | **4 to 5** | | | |
| colon cancer tissue | 8 (7.03) | 10 (10.87) | 20 (21.74) | 9 (9.59) | 8 (5.76) | 55 | | |
| | 38 | | | 17 | | | | |
| liver metastases | 3 (3.97) | 7 (6.13) | 14 (12.26) | 6 (5.41) | 1 (3.24) | 31 | 0.4619 | |
| | 24 | | | 7 | | | 0.4084 | 0.5643 |