DRIVER GENE MUTATIONS IN STOOLS OF COLORECTAL CARCINOMA PATIENTS DETECTED BY TARGETED NEXT GENERATION SEQUENCING

Gemma Armengol* [a,b], Virinder Kaur Sarhadi* [a], Reza Ghanbari [c], Masoud Doghaei-Moghaddam [d], Reza Ansari [c,d], Masoud Sotoudeh [c,d], Pauli Puolakkainen [e], Arto Kokkola [e], Reza Malekzadeh [c,d], Sakari Knuutila [a]

[a] The University of Helsinki, Faculty of Medicine, Department of Pathology, Helsinki, Finland

[b] Unit of Biological Anthropology, Department of Animal Biology, Plant Biology and Ecology, Universitat Autònoma de Barcelona, Barcelona, Catalonia, Spain.

[c] Digestive Oncology Research Center, Digestive Diseases Research Institute, Tehran University of Medical Sciences, Tehran, Iran

[d] Sasan Alborz Biomedical Research Center, Masoud Clinic, Tehran, Iran

[e] The HUCH Gastrointestinal Clinic, The University Central Hospital of Helsinki, Helsinki, Finland

* both authors contributed equally to this work

Number of text pages: 19

Number of tables: 4

Number of figures: 2

Short running head: Mutations in stools of colorectal cancer patients

Sources of support: The study was supported by Novartis, Sigrid Jusélius Foundation, and Cancer Society of Finland.

Corresponding author:

Prof. Sakari Knuutila,
Haartmaninkatu 4, P.O.Box 105,
FI-00029 HUCH, Finland.
Phone: +358 504482797
Fax: +358 294126700
E-mail: sakari.knuutila@helsinki.fi

ABSTRACT

Detection of driver gene mutations in stool DNA represents a promising noninvasive approach for screening colorectal cancer (CRC). Amplicon-based next generation sequencing (NGS) is a good option to study mutations in many cancer genes simultaneously and from low amount of DNA. The aim of the present study was to assess the feasibility of identifying mutations in 22 cancer driver genes with Ion Torrent technology in stool DNA from a series of 65 CRC patients. The assay was successful in 80% of stool DNA samples. NGS results showed 83 mutations in cancer driver genes, 29 hotspot and 54 novel mutations. One to five genes were mutated in 75% of cases. *TP53*, *KRAS*, *FBXW7* and *SMAD4* were the top mutated genes, consistent with previous studies. Fifty four percent of samples with mutations presented concomitant mutations in different genes. PI3K/MAPK pathway genes were mutated in 70% of samples, with 58% having alterations in *KRAS*, *NRAS* or *BRAF*. Since mutations in these genes can compromise the efficacy of EGFR blockade in CRC patients, identifying mutations that confer resistance to some targeted treatments may be useful to guide therapeutic decisions. In conclusion, the data presented here show that NGS procedures on stool DNA represent a promising tool to detect genetic mutations that could be used in future for diagnosis, for monitoring or for treating CRC.

INTRODUCTION

Colorectal cancer (CRC) represents about 10% of all cancers diagnosed worldwide and is the fourth leading cause of death from cancer in both men and women, with 694.000 deaths in 2012. Despite its decreased mortality in the last decades, death rates still remain high, with more deaths in the less developed regions of the world [1]. Epidemiologic transition in the low- and middle-income countries that are witnessing socioeconomic progress with the adoption of a Western life style with higher intake of fat, meat, and total calories, along with increasing life expectancy, result in a significant increase in the burden of CRC. This cancer type can be curable if detected in stage I or II, usually asymptomatic stages; however, cancer is often detected few years later when it has reached an advanced stage. Therefore, detecting early CRC is vital to reduce mortality at a population level [2]. So far, colonoscopy is the most sensitive and specific test available. Nonetheless, it has some limitations, such as risk of complications, limited availability, need of visibly detectable lesions, manpower requirements, and probably the most important, low patient compliance for cathartic preparation and dietary restrictions, especially in healthy individuals [2]. Stool-based assays raise the possibility of avoiding these pitfalls. The currently used fecal immunochemical test detects occult blood in the stool, but it is also limited because bleeding from CRC is more common at late stages and can be intermittent, and moreover, other lesions can cause fecal blood, giving some false-negatives and false-positives and resulting in the need for repetitive testing [3,4]. On the other hand, stool testing for DNA abnormalities is potentially more attractive for population-based screening programs. This nonivasive procedure is based on tumor exfoliation: colonocytes are continuously shed into the large bowel lumen. What is more, the number of colonocytes exfoliated from malignant lesions is greater than from normal tissue [5]. The genetic alterations that are detected come from neoplastic lesions and represent the tumor cells themselves [6]. Up to now, single markers or combinations of different DNA mutations in *APC*, *KRAS* and/or *TP53* genes, aberrant methylation, microsatellite instability, and/or DNA integrity have been tested for early detection of CRC

with good sensitivities and specificities [6-8]. Stool DNA testing has been shown to be more sensitive than fecal occult blood test [7,8] and than testing on other circulating DNA such as plasma DNA [9,10]. Furthermore, a meta-analysis highlighted the diagnostic value of multiple markers over single markers [3].

The potential of stool DNA testing is not only for CRC screening, nevertheless. It can be a helpful diagnostic and prognostic tool, and in addition, it can be used as a method for the early detection of disease relapse after initial therapy, or even as a guide for preoperative treatment selection, for example through risk stratification based on DNA mutations [11,12]. However, for such purposes more than three genes should be tested. Recent genetic advances, such as massive parallel sequencing, could be applied to stool DNA to search for mutations in many cancer genes simultaneously. Moreover, in the last few years sequencing information has helped to develop targeted therapy drugs on the basis of specific molecular aberrations of neoplastic cells. Three such drugs (all monoclonal antibodies) are approved to treat CRC, mainly metastatic CRC: two of them inhibit EGFR, and the other inhibits the angiogenesis-promoting VEGF. However, some CRC patients become refractory to anti-EGFR therapies because of genetic mutations in tumor cells (e.g. *KRAS*, *NRAS* and *BRAF*) that activate MEK and ERK signaling, which persists after EGFR blockade [13]. Therefore, sequencing of stool DNA mutations could allow us to identify patients who may benefit or not from such alternative treatment strategies, even before surgery. Additionally, stool DNA mutation analysis could help in detecting early evidence of mutations that confer acquired resistance to targeted therapies and that may arise during treatment.

Interestingly, two independent genome-scale studies aimed at identifying genetic changes in CRC by next generation sequencing (NGS) concluded that the pathways involved in CRC development are five: KRAS/NRAS/BRAF, p53, TGFB, PI3K, and WNT [14,15]. Moreover, after performing a hierarchical classification of more than 3,000 tumors from 12 tumor types on the basis of genomic alterations, Ciriello et al. [16] observed that oncogenic signature of CRC is dominated by point mutations over copy number changes. Therefore, using NGS technology

to search for point mutations in known colon cancer genes, such as targeted sequencing by Ion Torrent technology, would be suitable for the above-mentioned purposes, and much cheaper and faster than whole genome sequencing. Two previous studies have demonstrated that Ion AmpliSeq platform provides highly sensitive and quantitative mutation detection for cancer genes in CRC samples, even in formalin-fixed, paraffin-embedded (FFPE) specimens [17,18].

The aim of the present study was to assess the feasibility of identifying gene mutations in 22 cancer driver genes with the semiconductor-based Ion Torrent technology in stool DNA from a series of patients with CRC. We analyzed stool DNA from samples of both colon and rectum adenocarcinomas, as NGS analysis have revealed that colon and rectum cancers have very similar patterns of genomic alteration [15].

MATERIALS AND METHODS

Patient population

The study was carried out on a series of 65 stool samples from CRC patients, obtained at the time of diagnosis: 13 had the tumor located at the rectum and 52 at the colon. Sampling was done before any treatment. Patients were diagnosed and treated at Tehran University affiliated Shariati Hospital, in Tehran, Iran, from 2012 to 2014. The study protocol was approved by the ethics committee of Digestive Disease Research Institute at Shariati Hospital and an informed consent was obtained from all patients before being enrolled in the study. A list of tumor samples and their clinical characteristics is shown in Supplementary Table 1. None of the patients received EGFR-antibody treatment.

Gene mutation analysis by NGS

Stool DNA samples were sequenced with Ion Torrent technology (Life Technologies, Carlsbad, CA, USA), which has a very low input DNA requirement (as little as 10 ng of DNA). For a screening of CRC mutations, we used Ion AmpliSeq Colon and Lung Cancer Panel, designed and validated by eight cancer research groups from different European institutions; moreover, its accuracy has been proved recently in CRC patients [18].

DNA was isolated using the QIAamp DNA Stool Mini Kit (Qiagen GmbH, Hilden, Germany), following manufacturer's instructions. DNA concentration was assessed using Qubit® 2.0 Fluorometer (Life Technologies) and the Qubit® dsDNA HS Assay kit. Five samples did not show enough DNA quality to be sequenced. Therefore, deep sequencing was performed for the remaining 60 samples using the Ion Torrent platform, as follows. Ten ng of genomic DNA were used to prepare the library with the Ion AmpliSeq™ Library kit 2.0 (Life Technologies) and with a primer pool to analyze 504 mutational hotspots and targeted regions (totaling 14.6 kb) in 22 genes implicated in colon and lung cancer: *AKT1, ALK, BRAF, CTNNB1, DDR2, EGFR, ERBB2, ERBB4, FBXW7, FGFR1, FGFR2, FGFR3, KRAS, MAP2K1, MET, NOTCH1, NRAS, PIK3CA, PTEN, SMAD4, STK11,* and *TP53*. Amplified products were purified with Agencourt AMPure XP beads (Beckman Coulter Genomics, High Wycombe, UK). Concentrations of amplified and barcoded libraries were measured using the Qubit® 2.0 Fluorometer. After this, template preparation was performed with the Ion OneTouch™ 2 System (Life Technologies), an automated system for emulsion PCR, recovery of Ion Sphere™ Particles, and enrichment of template-positive particles. Finally, sequencing was carried out using Ion 316™ chips on the Ion Personal Genome Machine System (PGM™, Life Technologies) and with the Ion PGM™ Sequencing 200 kit v2. The Torrent Suite Software v.4.0.2 (Life Technologies) was used to assess run performance and data analysis. Fifty-two samples out of 60 were successfully sequenced, with a mean depth >20. Then, alignment to the hg19 human reference genome and variant calling was performed, according to the parameters set in the Torrent Suite Software. Only those variants with the following criteria were considered positive: variant frequency >5%, homopolymer length <6,

base phred quality > 15 (what means a minimum base call accuracy of 97%), coverage > 50, strand bias <0.9, common signal shift <0.3, and relative read quality >45. The variant frequency threshold was set at 5% to avoid false positive results. Recent reports have demonstrated that Ion Torrent platform is able to consistently detect mutations at the 5% level of mutant alleles, pinpointing true positive gene mutations [18,19]. Alignments were visually verified with the Integrative Genomics Viewer; IGV v.2.3.34, Broad Institute. The previously identified SNPs were determined using the NCBI dbSNP, HapMap or 1000 Human Genome Project databases and known polymorphic sites with a minor allele frequency >1% were removed from the analysis. Moreover, we filtered against synonymous and intron variants, as well as those variants that were considered to have both a neutral effect and a tolerated effect according to PROVEAN and SIFT prediction tools, respectively [20,21].

PCR and Sanger Sequencing

PCR was performed on DNA from the samples that showed *KRAS* codon 12, 13, or 20 or *EGFR* codon 706 or 757 mutations by NGS. Primers used and annealing temperature for PCR were: 5´-aggcctgctgaaaatgactgaatataa-3´ and 5´-caaagaatggtcctgcaccagtaatat-3´ (52ºC) for *KRAS* and 5´-tgtggagcctcttacaccca and 5´- gtgccagggaccttaccttatac-3´ (53ºC) for *EGFR.* The PCR products were cleaned and sequenced using Big Dye Terminator v3.1 sequencing kit (Applied Biosystems, Waltham, MA, USA) and run on 3130 XL Sequencer (Applied Biosystems).

RESULTS

Identification of mutations by NGS

The NGS assay was successfully carried out on 52/65 stool DNA samples (80% of success): five samples were removed from the analysis because of bad DNA quality and eight samples because of unsatisfactory sequencing quality. Results of gene mutation analysis showed both hotspot and novel mutations fulfilling the selection criteria. Overall, 83 mutations in

188 cancer driver genes were identified (Supplementary Table 2). Variant coverage ranged from
189 50 to 7324, with a mean value of 394 and a standard error of 101. About 65% (54 out of 83)
190 of mutations were new and had not been reported in COSMIC v.71, accessed 4 March, 2015
191 [22]. All mutations were protein-altering single-nucleotide substitutions: 87% transitions and
192 13% transversions. Only nine mutations corresponded to nonsense mutations, whereas all
193 the others were missense mutations.
194
195 The NGS analysis revealed that one to five genes were mutated in 39 of 52 (75%) cases,
196 whereas no mutations in the 22 tested genes were found in the 13 remaining cases
197 (Supplementary Table 3). Among the 22 genes, mutations were found in 18 genes (all except
198 *ALK*, *PIK3CA*, *NOTCH1* and *FGFR1*). There were some cases with two or even three
199 different co-existing mutations in the same gene (i.e. *TP53*, *KRAS*, *SMAD4*, *ERBB2*, *FGFR3*,
200 *MET*, and *MAP2K1*). No correlation could be found between the number and type of
201 mutations with tumor stage, site of tumor (colon/rectum) or patient overall survival.
202
203 Although quality control was performed to prevent false positive results, we performed some
204 validation experiments in order to check how the results were compared to Sanger
205 sequencing. We selected samples with *KRAS* mutations in exon 2 (codons 12, 13, and 20)
206 and samples with EGFR mutations (codons 706 and 757) for validation by PCR and Sanger
207 sequencing (Table 1). Only those samples with a good PCR amplification product were
208 sequenced. All samples that had mutant allele frequency higher than 10% were confirmed by
209 Sanger sequencing (it is known that Sanger sequencing is not sensitive enough to detect low
210 frequency mutations with a sensitivity threshold around 10% [23]). Figure 1 shows the
211 mutations detected by NGS and Sanger sequencing. Moreover, we could perform the
212 sequencing analysis of a tumor FFPE sample from case #44 and we were able to identify the
213 same mutations that were present in the corresponding stool sample (both in *KRAS* and
214 *TP53* genes) (Figure 2). Therefore, mutations identified in the present study by NGS are
215 most probably bona-fide mutations, and not false positive mutations.

Top mutated genes

The most frequently mutated genes were *TP53*, *KRAS*, *FBXW7*, *EGFR*, and *SMAD4*, all of them with at least 10% of samples with mutations. Unfortunately, the AmpliSeq Cancer Panel used in the present study does not include the *APC* gene or the *ATM* gene, and therefore mutations in these genes, which are frequent in CRC, could not be analyzed. *TP53* gene mutations occurred in 13 of 52 (25%) cases; most of them (nine) were hotspot mutations at the DNA binding domain (codons 94-292) and one was a nonsense mutation at the oligomerization domain (codon 342). The other three were novel mutations spread across all TP53 domains. *KRAS* gene was mutated in 12 of 52 (23%) cases. Surprisingly, *KRAS* hotspot mutations in exon 2 (codon 12 or 13) were identified in only seven of these 12 (58%) samples. Two patients had hotspot mutations in other codons (such as 117 or 146, both in exon 4) and the remaining three had novel mutations in exons 2 and 3 (codons 20, 43, and 63).Mutations in *FBXW7* were detected in nine of 52 (17%) cases. Three *FBXW7* mutations were hotspot mutations and the other novel mutations. Moreover, eight of the *FBXW7* mutations were located at the WD repeat domain. *EGFR* gene was mutated in eight samples (15% of cases), two of them with hotspot mutations. Interestingly, seven out of these eight mutations were located at the tyrosine kinase domain. *SMAD4* gene mutations were all novel and occurred in seven of 52 (13%) cases, one of them at the MH1 domain, five at the MH2 domain and one at the linker between these two domains. Less frequent mutations (4-8% of cases) were found in other seven genes (i.e. *PTEN*, *BRAF* (different than V600E), *ERBB2*, *FGFR3*, *MET*, *DDR2*, and *CTNNB1*). Mutations in the *ERBB4*, *FGFR2*, *STK11*, *AKT1*, *MAP2K1* and *NRAS* genes occurred very rarely (only one case each).

Concomitant mutations

Twenty one out of the 39 samples with mutations (54%) presented concomitant mutations in different genes. In some of the cases with concomitant mutations, no big differences were found in the variant frequency for the distinct genes, while in other cases variant frequencies

were quite different (Supplementary Table 2). For example, case #9 had 15% of alleles with a *TP53* p.C176Y mutation coexisting with a 12% of *KRAS* p.G13D mutation. On the other hand, for example, in case #22 the *KRAS* variant frequency was 6% and the *TP53* variant frequency was 23%. Moreover, the analysis revealed that 12 of the 18 cases with only one mutation (67%) harbored mutations in one of the most mutated genes in our series: *TP53*, *KRAS*, *FBXW7*, *EGFR* or *SMAD4* genes.

Pathways affected by mutated genes

Table 2 shows the concomitant mutations in every case ordered by pathways affected. Interestingly, the combined PI3K/MAPK pathway was altered in around 70% of samples. Among these, we found that more than half (58%) had alterations in *KRAS*, *NRAS* or *BRAF*, with a pattern of mutual exclusivity (only one sample harbored mutations in *KRAS* and *BRAF* simultaneously). Characteristics of mutations in these genes (with probable predictive value for anti-EGFR treatment) are summarized in Table 3.

DISCUSSION

Stool analysis is a non-invasive approach that promises improved simplicity, safety and patient compliance over other CRC screening methods. However, in the stool DNA analysis setting based on tumor exfoliation, challenges include the less than optimal DNA quality and the low DNA concentration. Ion Torrent NGS assay can overcome some of these problems, due to the low amount of DNA required, as demonstrated by the 80% success rate with the Ion AmpliSeq Colon and Lung Cancer Panel used in the present study. Nevertheless, technical improvements are needed in order to increase success rate and sequencing quality. Interestingly, none of the failed samples had rectum cancer, suggesting that rectum cancer cells could have a better preservation rate.

In our series, Ion Torrent technology with stool DNA detected tumor-specific mutations in 39 of 52 patients (75% of sensitivity) with CRC. After selecting mutations which probably have a functional consequence, based on SIFT and PROVEAN predictors, our analysis identified 83 mutations in cancer driver genes. Transitions were predominant over transversions, consistent with previous mutation reports on CRC [14,24]. Most of these mutations were new, as they had not been reported in COSMIC v.71 [22]. This agrees with previous reports, e.g. Seshagiri et al. [13] performed an exome sequencing analysis on 72 CRC samples and found that 98% of the detected mutations were new and not reported previously. It is expected that the adoption of NGS tools will make the occurrence of less common mutations to increase. The fact that several genes were mutated per sample and that most mutations were novel suggests that NGS studies applied to multiple driver genes are more accurate than current conventional analytical techniques, such as mutant-allele specific amplification, which are focused exclusively on the identification of known mutations in few markers.

Stool mutation pattern was quite similar to the reported in the literature for CRC tumor tissue, what reinforces the idea that fecal DNA mutations are representing mutations from tumor cells. *TP53*, *KRAS*, *FBXW7* and *SMAD4* were among the top mutated CRC genes, consistent with previous large-scale mutational analyses [14,15] and COSMIC database v.71 [22] (Table 4). *TP53* encodes a transcription factor which exerts antiproliferative functions. Most of the missense *TP53* mutations were found in the DNA-binding domain (83%). It has been reported that *TP53* mutations that cause cancer are mainly located at this domain, destroying the ability of the protein to bind to its target DNA sequences [25]. One mutation was located at the transactivation domain, and mutations in this domain are considered to alter the spectrum of target gene transactivation [26]. Another mutation was found in the tetramerization domain; various experimental studies have shown evidence that this domain is essential for DNA binding and protein-protein interactions [27]. In relation to *KRAS* mutations, activating mutations in the *KRAS* gene lead to cell transformation and increased resistance to chemotherapy and EGFR-targeted therapies [2,28]. We found *KRAS* mutations in codons 12,

13, 20, 63, 117, and 146 (all missense) and 43 (nonsense). It is well known that mutations in codons 12 and 13 (highly frequent in CRC), as well as in codon 63, cause perturbation of the intrinsic GTPase activity of the KRAS protein, while mutations in codons 117 and 146 cause a decrease in KRAS affinity for nucleotides, thereby affecting the rate of GDP/GTP exchange [28]. The p.T20M mutation detected in the present study has been reported in other two CRC samples [29,30], even though its effect is unknown. Finally, we detected a nonsense mutation in codon 43. Two other *KRAS* truncating mutations (in codons 22 and 150) have been reported in CRC in the literature [31,32]. Additionally, eight out of nine mutations in *FBXW7* gene were located at the WD repeat domain. This domain mediates interaction with substrates of the SCF E3 ubiquitin-protein ligase complex, which has important roles in the ubiquitination of proteins involved in cell division and growth. Loss of FBXW7 function causes chromosomal instability and tumorigenesis [33]. Finally, *SMAD4* gene had seven mutations: one at MH1 domain, required for DNA binding and five at MH2 domain, required for protein-protein interactions and for transcriptional regulation. SMAD4 is an essential factor of the TGF-β pathway and functions as a key tumor suppressor [34].

Stool-based assays applied to detect CRC-associated mutations may provide a safe, effective, noninvasive alternative for population-based CRC screening that is acceptable to patients. However, the clinical benefit of such tests could be further enhanced. NGS methodology applied to stool DNA might also have prognostic value (e.g. through risk stratification) or even treatment implications. In that sense, novel targeted therapeutic approaches could be applied to CRC patients before surgery or during patient follow-up, based on the genetic mutations detected by NGS. Included are WNT signaling inhibitors and/or ERB inhibitors, for example. On the other hand, it is known that mutations in *KRAS*, *BRAF*, and *NRAS* compromise the efficacy of EGFR blockade in patients with CRC and testing for these mutations is currently recommended before starting anti-EGFR therapy in patients with metastatic CRC [13,35,36]. Therefore, identifying mutations that confer resistance to some of these treatments may be useful to guide therapeutic decisions. In this series, nine

patients harbored mutations predictive of poor anti-EGFR response and thus these patients would likely not benefit from anti-EGFR targeted therapy (Table 3). Additionally, it remains to be demonstrated whether stool DNA analysis by NGS could be used to monitor disease progression or predict a relapse in pre-symptomatic individuals during follow-up. In conclusion, the data presented here show that NGS procedures on stool DNA represent a promising tool to identify CRC genetic mutations that could be used in future for diagnosis, for monitoring and for treating this deadly disease in a targeted way, if suitable.

The present study is focused on CRC samples, adenomas were not studied. It would be interesting to see whether advanced adenomas present also DNA mutations in stool samples that could be detected by NGS methods, and whether this technology allows to accurately identify adenomas with the potential for malignant transformation. In previous stool-based studies, two mutation marker panels testing for 19-21 CRC-associated DNA mutations in three genes and microsatellite instability showed significant sensitivity for advanced colonic polyps, especially those with high-grade dysplasia [37–39]. What is more, a meta-analysis study indicated that stool DNA testing for few markers had good ability for detecting pre-cancerous lesions in high-risk individuals [3]. Therefore, it is plausible to consider that NGS analysis on stool DNA might serve also as a noninvasive screening for pre-malignant adenomatous polyps in asymptomatic individuals who are at significant risk. This would reduce CRC incidence and mortality, as advanced adenomas could then be referred to colonoscopy and/or treated with targeted therapy.

ACKNOWLEDGEMENTS

The authors thank Tiina Wirtanen and Milja Tikkanen for her assistance in sequencing reactions. The study was supported by Novartis, Sigrid Jusélius Foundation, and Cancer Society of Finland.

354

355

REFERENCES

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman DD, Bray F: Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2014, 136:E359–E386

2. Kanthan R, Senger J-L, Kanthan SC: Fecal molecular markers for colorectal cancer screening. Gastroenterol Res Pract 2012, 2012:184343

3. Yang H, Xia BQ, Jiang B, Wang G, Yang YP, Chen H, Li BS, Xu AG, Huang YB, Wang XY: Diagnostic value of stool DNA testing for multiple markers of colorectal cancer and advanced adenoma: A meta-analysis. Can J Gastroenterol 2013, 27:467–475

4. Lee JK, Liles EG, Bent S, Levin TR, Corley DA: Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. Ann Intern Med 2014, 160:171

5. Davies RJ, Miller R, Coleman N: Colorectal cancer screening: prospects for molecular stool analysis. Nat Rev Cancer 2005, 5:199–209

6. Bosch LJW, Carvalho B, Fijneman RJA, Jimenez CR, Pinedo HM, van Engeland M, Meijer GA: Molecular tests for colorectal cancer screening. Clin Colorectal Cancer 2011, 10:8–23

7. Imperiale TF, Ransohoff DF, Itzkowitz SH, Turnbull BA, Ross ME: Fecal DNA versus fecal occult blood for colorectal-cancer screening in an average-risk population. N Engl J Med 2004, 351:2704–2714

8. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, Ahlquist DA, Berger BM: Multitarget stool DNA testing for colorectal-cancer screening. N Engl J Med 2014, 370:1287–1297

9. Ahlquist DA, Taylor WR, Mahoney DW, Zou H, Domanico M, Thibodeau SN, Boardman LA, Berger BM, Lidgard GP: The stool DNA test is more accurate than the plasma septin 9 test in detecting colorectal neoplasia. Clin Gastroenterol Hepatol 2012, 10:272–277.e1

10. Diehl F, Schmidt K, Durkee KH, Moore KJ, Goodman SN, Shuber AP, Kinzler KW, Vogelstein B: Analysis of mutations in DNA isolated from plasma and stool of colorectal cancer patients. Gastroenterology 2008, 135:489–498

11. Lansdorp-Vogelaar I, Knudsen AB, Brenner H: Cost-effectiveness of colorectal cancer screening - an overview. Best Pract Res Clin Gastroenterol 2010, 24:439–449

12. Avital I, Langan RC, Summers TA, Steele SR, Waldman SA, Backman V, Yee J, Nissan A, Young P, Womeldorph C, Mancusco P, Mueller R, Noto K, Grundfest W, Bilchik AJ, Protic M, Daumer M, Eberhardt J, Man YG, Brücher BL, Stojadinovic A: Evidence-based guidelines for precision risk stratification- based screening (PRSBS) for colorectal cancer: Lessons learned from the us armed forces: Consensus and future directions. J Cancer 2013, 4:172–192

13. Misale S, Arena S, Lamba S, Siravegna G, Lallo A, Hobor S, Russo M, Buscarino M, Lazzari L, Sartore-Bianchi A, Bencardino K, Amatu A, Lauricella C, Valtorta E, Siena S, Di Nicolantonio F, Bardelli A: Blockade of EGFR and MEK intercepts heterogeneous mechanisms of acquired resistance to anti-EGFR therapies in colorectal cancer. Sci Transl Med 2014, 6:224ra26

14. Seshagiri S, Stawiski EW, Durinck S, Modrusan Z, Storm EE, Conboy CB, Chaudhuri S, Guan Y, Janakiraman V, Jaiswal BS, Guillory J, Ha C, Dijkgraaf GJP, Stinson J, Gnad F, Huntley MA, Degenhardt JD, Haverty PM, Bourgon R, Wang W, Koeppen H, Gentleman R, Starr TK, Zhang Z, Largaespada DA, Wu TD, de Sauvage FJ: Recurrent R-spondin fusions in colon cancer. Nature 2012, 488:660–664

15. The Cancer Genome Atlas Network: Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012, 487:330–337

16. Ciriello G, Miller ML, Aksoy BA, Senbabaoglu Y, Schultz N, Sander C: Emerging landscape of oncogenic signatures across human cancers. Nat Genet 2013, 45:1127–1133

17. Zhang L, Chen L, Sah S, Latham GJ, Patel R, Song Q, Koeppen H, Tam R, Schleifman E, Mashhedi H, Chalasani S, Fu L, Sumiyoshi T, Raja R, Forrest W, Hampton GM, Lackner MR, Hegde P, Jia S: Profiling cancer gene mutations in clinical formalin-fixed, paraffin-embedded colorectal tumor specimens using targeted next-generation sequencing. Oncologist 2014, 19:336–343

18. Malapelle U, Vigliar E, Sgariglia R, Bellevicine C, Colarossi L, Vitale D, Pallante P, Troncone G: Ion Torrent next-generation sequencing for routine identification of clinically relevant mutations in colorectal cancer patients. J Clin Pathol 2015, 68:64–68

19. Tops B, Normanno N, Kurth H, Amato E, Mafficini A, Rieber N, Le Corre D, Rachiglio A, Reiman A, Sheils O, Noppen C, Lacroix L, Cree IA, Scarpa A, Ligtenberg M, Laurent-Puig P: Development of a semi-conductor sequencing-based panel for genotyping of colon and lung cancer by the Onconetwork consortium. BMC Cancer 2015, 15:26

20. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP: Predicting the functional effect of amino acid substitutions and indels. PLoS One 2012, 7:e46688

21. Kumar P, Henikoff S, Ng PC: Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 2009, 4:1073–1081

22. Forbes SA, Tang G, Bindal N, Bamford S, Dawson E, Cole C, Kok CY, Jia M, Ewing R, Menzies A, Teague JW, Stratton MR, Futreal PA: COSMIC (the Catalogue of Somatic Mutations In Cancer): A resource to investigate acquired mutations in human cancer. Nucleic Acids Res 2009, 38:D652–D657

23. Tsiatis AC, Norris-Kirby A, Rich RG, Hafez MJ, Gocke CD, Eshleman JR, Murphy KM: Comparison of Sanger Sequencing, Pyrosequencing, and Melting Curve Analysis for the Detection of KRAS Mutations. J Mol Diagnostics 2010, 12:425–432

24. Bass AJ, Lawrence MS, Brace LE, Ramos AH, Drier Y, Cibulskis K, Sougnez C, Voet D, Saksena G, Sivachenko A, Jing R, Parkin M, Pugh T, Verhaak RG, Stransky N, Boutin AT, Barretina J, Solit DB, Vakiani E, Shao W, Mishina Y, Warmuth M, Jimenez J, Chiang DY, Signoretti S, Kaelin WG, Spardy N, Hahn WC, Hoshida Y, Ogino S,

DePinho RA, Chin L, Garraway LA, Fuchs CS, Baselga J, Tabernero J, Gabriel S, Lander ES, Getz G, Meyerson M: Genomic sequencing of colorectal adenocarcinomas identifies a recurrent VTI1A-TCF7L2 fusion. Nat Genet 2011, 43:964–968

25. Joerger AC, Fersht AR: Structure-function-rescue: the diverse nature of common p53 cancer mutants. Oncogene 2007, 26:2226–2242

26. Ganguly D, Chen J: Modulation of the Disordered Conformational Ensembles of the p53 Transactivation Domain by Cancer-Associated Mutations. PLOS Comput Biol 2015, 11:e1004247

27. Chène P: The role of tetramerization in p53 function. Oncogene 2001, 20:2611–2617

28. Jančík S, Drábek J, Radzioch D, Hajdúch M: Clinical relevance of KRAS in human cancers. J Biomed Biotechnol 2010, 2010:150960

29. Brink M, de Goeij AFPM, Weijenberg MP, Roemen GMJM, Lentjes MHFM, Pachen MMM, Smits KM, de Bruïne AP, Goldbohm RA, van den Brandt PA: K-ras oncogene mutations in sporadic colorectal cancer in The Netherlands Cohort Study. Carcinogenesis 2003, 24:703–710

30. Giannakis M, Hodis E, Jasmine Mu X, Yamauchi M, Rosenbluh J, Cibulskis K, Saksena G, Lawrence MS, Qian ZR, Nishihara R, Van Allen EM, Hahn WC, Gabriel SB, Lander ES, Getz G, Ogino S, Fuchs CS, Garraway LA: RNF43 is frequently mutated in colorectal and endometrial cancers. Nat Genet 2014, 46:1264–1266

31. Naser WM, Shawarby MA, Al-Tamimi DM, Seth A, Al-Quorain A, Al Nemer AM, Albagha OME: Novel KRAS Gene Mutations in Sporadic Colorectal Cancer. PLoS One 2014, 9:e113350

32. Palmirotta R, Savonarola A, Formica V, Ludovici G, Del Monte G, Roselli M, Guadagni F: A novel K-ras mutation in colorectal cancer. A case report and literature review. Anticancer Res 2009, 29:3369–3374

33. Welcker M, Clurman BE: FBW7 ubiquitin ligase: a tumour suppressor at the crossroads of cell division, growth and differentiation. Nat Rev Cancer 2008, 8:83–93

34. Yang G, Yang X: Smad4-mediated TGF-beta signaling in tumorigenesis. Int J Biol Sci 2010, 6:1–8

35. De Roock W, Claes B, Bernasconi D, De Schutter J, Biesmans B, Fountzilas G, Kalogeras KT, Kotoula V, Papamichael D, Laurent-Puig P, Penault-Llorca F, Rougier P, Vincenzi B, Santini D, Tonini G, Cappuzzo F, Frattini M, Molinari F, Saletti P, De Dosso S, Martini M, Bardelli A, Siena S, Sartore-Bianchi A, Tabernero J, Macarulla T, Di Fiore F, Gangloff AO, Ciardiello F, Pfeiffer P, Qvortrup C, Hansen TP, Van Cutsem E, Piessevaux H, Lambrechts D, Delorenzi M, Tejpar S: Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. Lancet Oncol 2010, 11:753–762

36. Douillard J-Y, Oliner KS, Siena S, Tabernero J, Burkes R, Barugel M, Humblet Y, Bodoky G, Cunningham D, Jassem J, Rivera F, Kocákova I, Ruff P, Błasińska-Morawiec M, Šmakal M, Canon JL, Rother M, Williams R, Rong A, Wiezorek J, Sidhu

R, Patterson SD: Panitumumab-FOLFOX4 treatment and RAS mutations in colorectal cancer. N Engl J Med 2013, 369:1023–1034

37. Berger BM, Vucson BM, Ditelberg JS: Gene mutations in advanced colonic polyps: potential marker selection for stool-based mutated human DNA assays for colon cancer screening. Clin Colorectal Cancer 2003, 3:180–185

38. Berger BM, Schroy 3rd PC, Rosenberg JL, Lai-Goldman M, Eisenberg M, Brown T, Rochelle RB, Billings PR: Colorectal cancer screening using stool DNA analysis in clinical practice: early clinical experience with respect to patient acceptance and colonoscopic follow-up of abnormal tests. Clin Color Cancer 2006, 5:338–343

39. Tagore KS, Lawson MJ, Yucaitis JA, Gage R, Orr T, Shuber AP, Ross ME: Sensitivity and specificity of a stool DNA multitarget assay panel for the detection of advanced colorectal neoplasia. Clin Colorectal Cancer 2003, 3:47–53

493 Table 1. Validation of NGS-identified mutations (at specific codons of *KRAS* and *EGFR*

494 genes) by PCR and Sanger sequencing.

Sample	Mutation by NGS	Mutant allele frequency	PCR band quality	Mutation by Sanger sequencing
9	*KRAS* p.G13D	11 %	Good	*KRAS* p.G13D
27	*KRAS* p.T20M	8%	Bad	Not sequenced
32	*KRAS* p.G12D	9 %	Bad	Not sequenced
44	*KRAS* p.G12A	24 %	Good	*KRAS* p.G12A
50	*KRAS* p.G12D	26%	Good	*KRAS* p.G12D
52	*KRAS* p.G12S	5 %	Bad	Not sequenced
56	*KRAS* p.G12S & p.E31K	7% & 7%	Good	No mutation
65	*KRAS* p.G13D & p.Q25*	7% & 8%	Good	No mutation
11	*EGFR* p.I706-576-653-661T	7%	Bad	Not sequenced
34	*EGFR* p.K757-627-704-712E	11%	Bad	Not sequenced
59	*EGFR* p.I706-576-653-661T	8%	Good	No mutation

495

496

497

498

499

Table 2. Mutation pattern in the pathways typically altered in CRC for the 39 samples with mutations. For each sample, it is shown with an X if at least one mutation was identified for that particular gene. For the sake of clarity, number of genes mutated per sample were scored (counts) in the two pathways with more than one gene analyzed.

A. Combined PI3K and MAPK pathway

	1	6	7	9	10	11	15	18	19	21	22	24	27	28	30	32	33	34	35	37	39	41	44	46	47	49	50	51	52	53	54	55	56	57	58	59	60	63	65
KRAS				X							X		X			X		X	X				X			X	X		X				X					X
NRAS														X																									
BRAF							X													X														X					
EGFR						X			X							X		X														X			X	X			X
ERBB2					X									X																				X					
ERBB4								X																															
FGFR2									X																														
FGFR3																																			X	X			
MET																									X						X								
PTEN		X																		X																X			X
ALK																																							
PIK3CA																																							
MAP2K1																																							X
AKT1																																							X
Counts	0	1	0	1	1	1	1	1	2	0	1	0	1	2	0	2	0	2	1	2	0	0	1	0	1	1	1	0	1	0	1	1	2	1	2	3	0	0	5

B. WNT pathway

	1	6	7	9	10	11	15	18	19	21	22	24	27	28	30	32	33	34	35	37	39	41	44	46	47	49	50	51	52	53	54	55	56	57	58	59	60	63	65
FBXW7	X									X	X			X							X			X			X		X					X					
CTNNB1			X								X																												
Counts	1	0	1	0	0	0	0	0	0	1	2	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0

C. TGFβ pathway

	1	6	7	9	10	11	15	18	19	21	22	24	27	28	30	32	33	34	35	37	39	41	44	46	47	49	50	51	52	53	54	55	56	57	58	59	60	63	65
SMAD4									X					X	X		X					X										X				X			

D. TP53 pathway

	1	6	7	9	10	11	15	18	19	21	22	24	27	28	30	32	33	34	35	37	39	41	44	46	47	49	50	51	52	53	54	55	56	57	58	59	60	63	65
TP53			X	X	X	X					X	X										X	X					X	X	X		X						X	

516 Table 3. Characteristics of mutations in genes predictive of EGFR antibody treatment
517 identified from stool DNA of colorectal cancer patients.

Gene	Exon	Mutation	COSMIC ID / Novel	Resistant/sensitive mutation	No. of samples
KRAS	2	p.G12A	COSM522	resistant	1
	2	p.G12D	COSM521	resistant	2
	2	p.G12S	COSM517	resistant	2
	2	p.G13D	COSM532	resistant	2
	2	p.T20M	novel	unknown	1
	3	p.Q43*	novel	unknown	1
	3	p.E63G	novel	unknown	1
	4	p.K117N	COSM19940	resistant	1
	4	p.A146T	COSM19404	resistant	1
NRAS	2	p.L6M	novel	unknown	1
BRAF	11	p.R444Q	novel	unknown	1
	11	p.R462G	novel	unknown	1
	11	p.G466E	COSM453	unknown	1

518

519

520 Table 4. Percentage of colorectal cancer samples with mutations in the genes from the Ion
521 AmpliSeq Colon and Lung Cancer Panel according to COSMIC database v.71 in comparison
522 with the present study.

		Colon adenocarcinoma	Rectum adenocarcinoma	Present study
Census cancer genes	*TP53*	48%	62%	25%
	KRAS	35%	35%	23%
	FBXW7	20%	21%	17%
	SMAD4	23%	19%	13%
	PTEN	25%	21%	8%
	BRAF	14%	4%	6%
	PIK3CA	22%	12%	0%
Other genes	*EGFR*	8%	3%	15%
	ERBB2	3%	3%	6%
	CTNNB1	10%	5%	4%
	MET	8%	5%	4%
	FGFR3	4%	1%	4%
	DDR2	3%	2%	4%
	AKT1	1%	0%	2%
	FGFR2	6%	5%	2%
	NRAS	5%	9%	2%
	STK11	7%	4%	2%
	ERBB4	13%	5%	2%
	MAP2K1	2%	2%	2%
	FGFR1	4%	3%	0%
	ALK	12%	6%	0%

	NOTCH1	7%	0%	0%

523

524

525

526 FIGURE LEGENDS

527

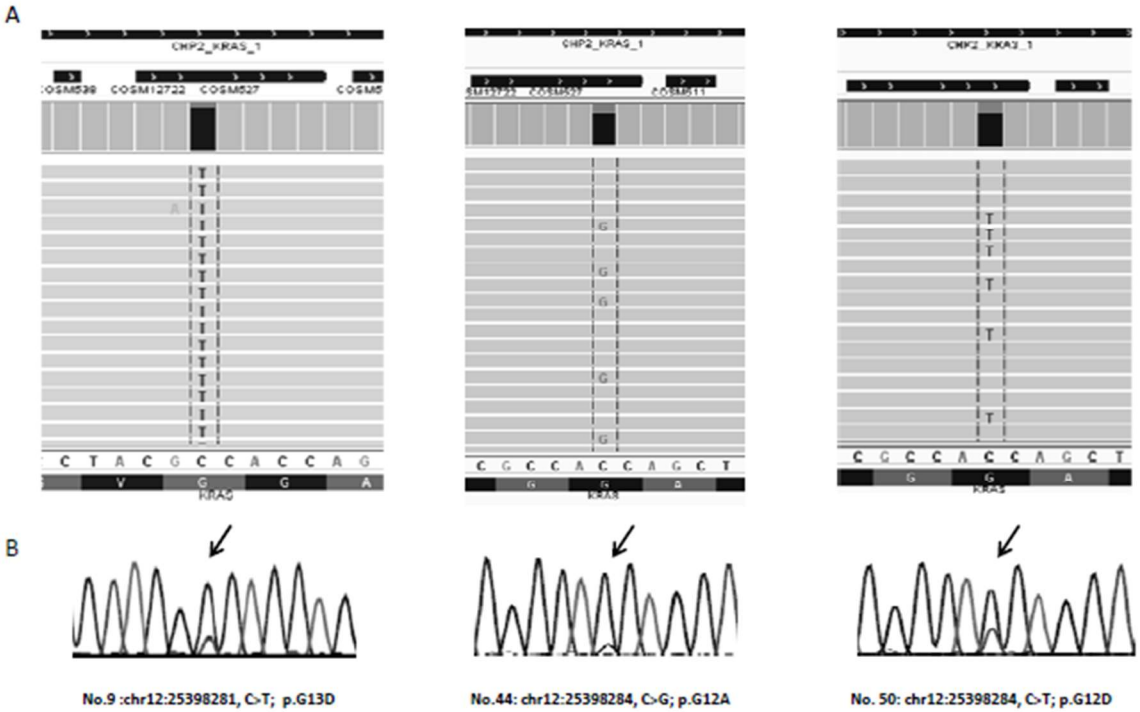
528 Figure 1. *KRAS* exon 12 mutations observed by both NGS (A) and Sanger sequencing (B) in
529 stool DNA from cases #9, #44, and #50.

530

531 Figure 2. *KRAS* (A, C) and *TP53* (B, D) mutations observed by NGS in a tumorous sample
532 (A, B) and the corresponding stool sample (C, D) from case #44.

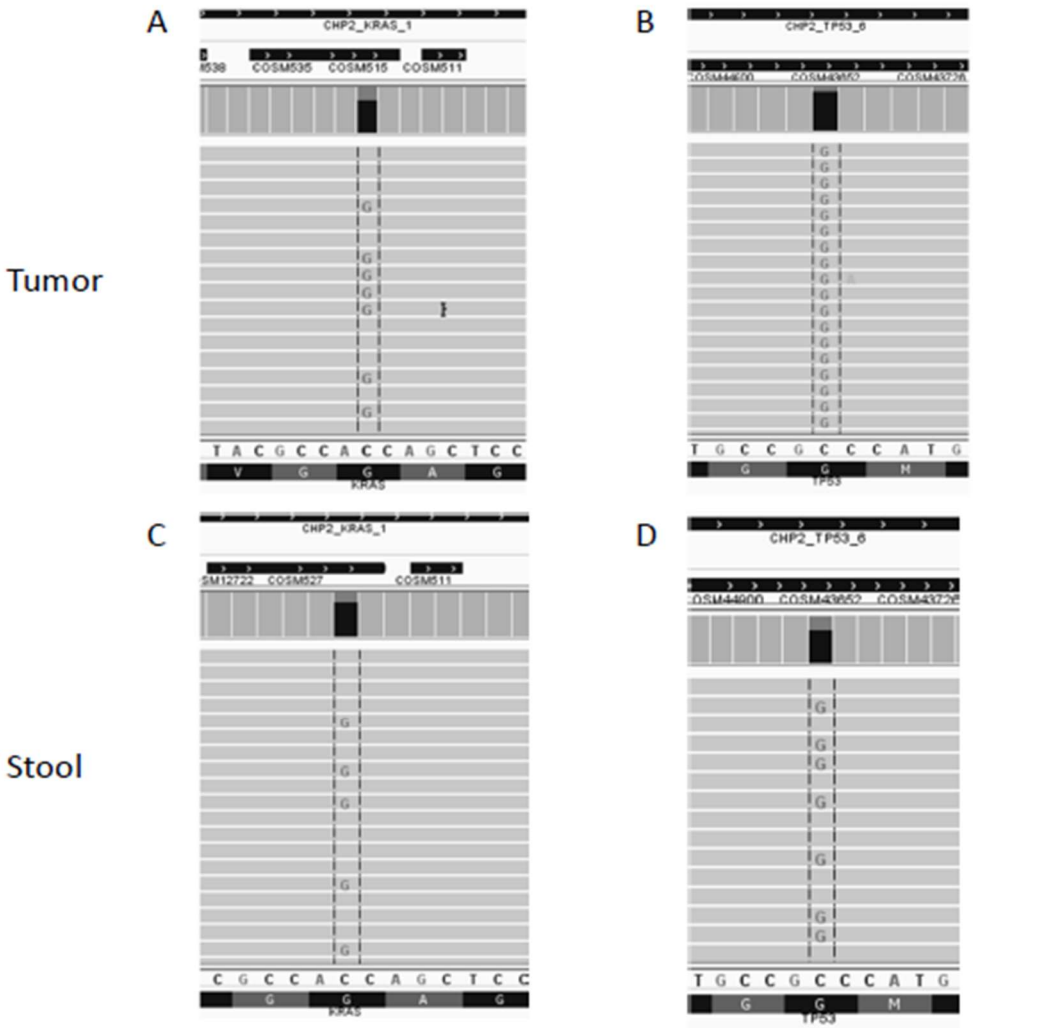
533

534 Figure 1



535

536 Figure 2



No.44: chr12:25398284, C>G; p.G12A **No.44: chr17:7674232, C>G; p.G244A**

537