Comprehensive characterization of RAS mutations in colon and rectal cancers in old and young patients

Ilya G. Serebriiskii1,2, Caitlin Connelly3, Garrett Frampton3, Justin Newberg3, Matthew Cooke3, Vince Miller3, Siraj Ali3, Jeffrey S. Ross3,4, Elizabeth Handorf5, Sanjeevani Arora6, Christopher Lieu7, Erica A. Golemis1 & Joshua E. Meyer1,8

Colorectal cancer (CRC) is increasingly appreciated as a heterogeneous disease, with factors such as microsatellite instability (MSI), cancer subsite within the colon versus rectum, and age of diagnosis associated with specific disease course and therapeutic response. Activating oncogenic mutations in KRAS and NRAS are common in CRC, driving tumor progression and influencing efficacy of both cytotoxic and targeted therapies. The RAS mutational spectrum differs substantially between tumors arising from distinct tissues. Structure-function analysis of relatively common somatic RAS mutations in G12, Q61, and other codons is characterized by differing potency and modes of action. Here we show the mutational profile of KRAS, NRAS, and the less common HRAS in 13,336 CRC tumors, comparing the frequency of specific mutations based on age of diagnosis, MSI status, and colon versus rectum subsite. We identify mutation hotspots, and unexpected differences in mutation spectrum, based on these clinical parameters.
Colorectal cancer (CRC) is the fourth most common cancer in the United States, with an estimated 135,430 new cases diagnosed in 2017 and 50,260 deaths in the same year. While the overall survival (OS) at 5 years is 65%, in patients with metastatic disease this rate drops to 14%. Ongoing efforts to improve diagnosis and treatment for CRC depend on more accurate understanding of disease biology. For CRC, as for many cancers, a disease once considered as uniform is being increasingly divided into discrete subclases, based on the integration of molecular and clinical analyses. For example, microsatellite stable (MSS) CRC tumors have distinct etiology and treatment recommendations from CRC with a high level of microsatellite instability (MSI-H). Other potentially important distinctions are emerging, such as formation of tumors on the right or left side of the colon.

While typically grouped together, cancers arising in the colon differ from those arising in the rectum in several clinically important ways. First, early-stage patients with colon cancer demonstrate improved survival over rectal cancer in stage IIB, while in stages III and IV the opposite is true. Interestingly, although the incidence of CRC has been increasing over time, particularly among young patients, the increase has been more notable in rectal cancer. Second, rectal cancer is treated differently from colon cancer, with preoperative radiation with or without chemotherapy required to achieve acceptable loco-regional control in rectal cancer. Because rectal cancer is less common than colon cancer (28% of total CRC cases in the United States), data sets to analyze molecular differences have been limited.

Further complicating the analysis, the natural history of CRC appears different in old versus young patients, as evidenced by increased lymph node metastasis rates in younger patients, possibly suggesting a difference in biology. To date, examinations of age as a factor influencing molecular differences in CRC have also been limited, with the notable exception of MSI. MSI is more commonly found in younger patients, potentially due to the activity of inherited damaging variants in genes associated with DNA repair, which tends to lead to an early-onset disease. Additionally, one recent publication has shown increased alteration of TP53 and CTNNB1 in younger patients, with older patients demonstrating increases in APC, KRAS, BRAF and FAM123B alterations.

The three RAS GTPases (KRAS, NRAS, and HRAS), and particularly the family member KRAS, are the most commonly mutated oncogenes in cancer. KRAS and NRAS have been reported to have activating missense mutations in ~40 and 3–5% of CRCs, respectively. Both preclinical and clinical data show that the specific codon mutated, and the sequence to which it is mutated, differentially impact the activity of the protein, the prognosis, and the response to epidermal growth factor receptor (EGFR)-directed therapy in patients with CRC. In this study, we describe the frequency of missense mutations and copy number variation (CNV) of wild-type (wt) and mutated alleles of KRAS and NRAS in a very large dataset of 13,336 CRC tumors from patients profiled by Foundation Medicine (FM) Inc. We separately report results based on occurrence in MSS versus MSI-H tumors, in colon versus rectal cancer, and based on patient age. These data indicate non-equivalent profiles of specific CRC-associated mutations based on subsite, age, gender, and other criteria.

Results

Patient population: age, gender, tumor site, and MSI status.

Comprehensive genomic profiling (CGP) of tumor samples by next-generation sequencing was performed in the course of routine clinical care for CRC patients with advanced disease, resulting in the identification of genetic variants as well as biomarkers such as tumor mutation burden (TMB) and microsatellite stability. Characteristics of the patient population are summarized in Fig. 1a, Table 1, and Supplementary Fig. 1. The dataset analyzed here is nearly 2-fold greater than the combined previously analyzed patient cohorts reported in publications or developed through the GENIE consortium, or otherwise available at cBioPortal. Comparable to other publicly available sequencing data (PAD) for CRC, the group is 45.2% male, with a similar age distribution at the time of sequencing, and a similar proportion of 85.3% colon:14.7% rectum.

Among the total set of 13,336 CRC tumors, information on microsatellite stability and quantifying TMB was available for 10,070 (Fig. 1b), of which 9615 (95.5%) were MSS and 455 displayed MSI-H (4.5%). While information for neither MSS/MSI-H nor TMB status was available for 267 patients, information on TMB status only was available for 3266 patients. Using the set of 10,070 tumors as a training set, we determined a threshold of <16 mutations separated 99% of the MSS from 99% of the MSI-H cases (Supplementary Fig. 2). Applying this threshold, we assigned 3096/3266 (95%) additional tumors as MSS microsatellite stable, MSI-H microsatellite instability high. Individuals with rectal cancer were slightly younger than colon cancer patients (Supplementary Fig. 1). Tumors with discordant sex were not included in the analysis.
KRAS, NRAS, and HRA S mutation hotspots in the entire cohort. Among the 13,336 tumors analyzed (Supplementary Data 1), there were 6926 KRAS alterations (defined as cumulative missense mutations, frameshifts, and copy number changes), including 6634 missense mutations, in the MT-L cohort of 12,711 patients, and 246 KRAS alterations (235 missense) among 625 MT-H TMB patients (Supplementary Data 1). Thirty-four alterations, including 32 missense mutations in NRAS, were detected in MT-H patients, and 560 alterations (537 missense mutations) in MT-L patients (Supplementary Data 1). For HRAS, 35 missense mutations (44 alterations in total) were identified among MT-H patients, and 32 missense mutations (53 alterations in total) among MT-L patients.

The complete set of missense mutations in KRAS, NRAS, and HRAS were then analyzed to identify mutational hotspots (Fig. 2a–c), and mapped to the protein structure for each RAS protein (Fig. 2d–k). This identified 16 mutational hotspots for KRAS (Fig. 2a); these included four sites previously reported from a dataset of >1267 CRC tumors29, which were a subset of 10 hotspots identified from an analysis of ~25,000 tumors of all types30, as well as 6 additional sites not previously apparent in analysis of smaller cohorts of CRC tumors (residues V9, V14, R68, R164, K176, and K180). For NRAS (Fig. 2b), this analysis identified one new hotspot for NRAS mutation (residue P185), in addition to three previously defined hotspots (residues G12, G13, and Q61). For HRAS, no new hotspots were unequivocally identified, although two residues (R164 and P167) were borderline significant (Fig. 2c). None of the identified mutations were associated with microsatellite regions.

RAS functions as a molecular switch between the GTP-bound active form and the GDP-bound inactive form. Missense mutations in RAS proteins alter the homeostatic balance between these two forms, either by reducing GTP hydrolysis or by increasing the rate of GTP loading31. The most common KRAS-activating mutations, in codons 12, 13, 61, 117, and 146, cluster around the nucleotide-binding pocket16. Display of the full set of mutational hotspots from KRAS on the three-dimensional (3D) structure of the KRAS protein (Fig. 2d–g) shows the majority of these clusters in functional domains. One new hotspot (V14) is located within 5 Å of K117; others (e.g., K176, K180) are localized within the C-terminal hypervariable region (HVR). Finally, our analysis did not identify two known hotspots in KRAS (residues K5 and A11), which were described in non-CRC tumors, reflecting known tissue bias in hotspots30. Interestingly, enrichment for mutations was observed within two short spans of six amino acids, although not targeting individual residues within the spans. This includes residues 47–52, which encompasses 10 mutations (p = 7.2E−03), and positions 62–67 (between the
known Q61 and new R68 hotspots), which encompasses 11 mutations ($p = 2.7 \times 10^{-4}$).

Display of the full set of mutational hotspots from the 3D structure of NRAS (Fig. 2h, i) shows that the majority of these cluster in functional domains. The new hotspot (P185) is localized within the C-terminal HVR. In addition, a marginally significant potential hotspot (E132; three mutations, $p = 0.018$) is located in the $\alpha$-helix between two amino acid motifs (G4, NKXD and G5, EXSAK) that are important for binding the guanine base of the nucleotide $^{32}$. KRAS does not have hotspots comparable to E132 and P185 of NRAS. The G60 hotspot, identified in other cancer types$^{30}$, was not found in our study;
however, mutations were enriched overall within the five amino acids immediately upstream of Q61 (amino acids 56–60, seven mutations, $p = 5.1 \times 10^{-04}$) (Fig. 2b).

For HRAS, of the four hotspots previously nominated based on a pan-cancer survey (G12, G13, Q61, K117), only one (G12) was also identified in our study (Fig. 2j, k). However, two previously overlooked residues (R161 and P167) were mutated at a higher frequency than expected (four mutations each; $p = 0.007$), and an additional four mutations were identified in residues 163 and 164. These above mutations are unusually enriched ($p = 2.6 \times 10^{-04}$) in a stretch of amino acids located near the N terminus of the HVR (Fig. 2c). The HVR is distinct between the RAS isoforms and this unusual enrichment of mutations in the N terminus of this region in the HRAS protein may play a role in disease pathogenesis.

Mutational variation between MT-L and MT-H tumors. We compared the distribution of total alterations in KRAS, NRAS, and HRAS in CRC patients, as a factor of MT-L versus MT-H status (Fig. 3a). KRAS mutations were higher in MT-L than MT-H CRC. For NRAS, no significant differences in the total number of mutations were noted between MT-L and MT-H CRC tumors. Surprisingly, even though the total frequency of HRAS mutations was very low in CRC, the frequency of HRAS mutations in MT-H tumors was significantly higher than in MT-L tumors, comparable to that of NRAS mutations. This pattern differed significantly from the pattern observed for KRAS (where the mutation frequency was higher in MT-L CRC) or NRAS ($p < 2.2 \times 10^{-16}$).

The observed spectrum of KRAS missense mutations also significantly differed between MT-L and MT-H tumors ($p = 5.0 \times 10^{-04}$) (Fig. 3b). G12 mutations predominated in MT-L tumors (69.3%), whereas MT-H tumors had a greater distribution of sites targeted, including G12 (~34%), G13 (~35%), A146 (~9%), and other missense structural variants (SVs) (~22%). Among the G12 mutations, statistically significant differences in codon substitution preference were also observed, with over 77% of MT-H tumors having G12D mutations and no other specific codon substitution comprising an allele fraction >8%, versus MT-L tumors having ~44% G12D, ~31% G12V, ~11% G12C, ~6% G12A, and ~8% other G12 variants ($p = 2.4 \times 10^{-08}$) (Fig. 3c). This difference likely reflects the fact that G12D changes result from C>T nucleotide changes, which are known to be enriched in MT-H tumors. Variance in patterns of codon substitutions was also observed with other codons (e.g., A146), although with lesser significance due to smaller sample size.

In contrast to the lack of difference in total frequency of mutations, the NRAS mutational spectra differed significantly between MT-H and MT-L tumors ($p = 3.2 \times 10^{-13}$) (Fig. 3b). MT-H tumors were dominated by a large number of low-frequency mutations (~53%), followed by G12 (25%), Q61 (~16%), and G13 (~6%). The most common site of missense mutation in MT-L tumors was Q61 (~52%), with G12 (~33%), G13 (~10%), and other variants (~4%) making up the remainder of single codon substitutions. Due to the small sample size for NRAS alterations, further analysis of specific missense variants was not conclusive, with the exception of the most common variant, G12. In MT-L tumors, over 80% of 178 mutations affecting G12 are the substitutions G12A, G12V, and G12D, while in 8 MT-H tumors, 7 are substitutions to other residues, with G12C being predominant (5/8 samples, compared with only ~12% in MT-L (Fig. 3c).

The number of HRAS missense mutations observed in this dataset was too low for a significant analysis of mutation preference.

Missense mutational frequency and spectrum dependent on age. We analyzed RAS mutations in CRC patients as a continuous variable of age. In MT-L patients, the overall frequency of KRAS alterations increase with age ($p = 3.1 \times 10^{-03}$, Fig. 3d). In sharp contrast, the incidence of KRAS alternations is reduced with age in MT-H patients ($p = 2.1 \times 10^{-03}$, Fig. 3d). Among the MT-L patients, the frequency of missense variants in NRAS remained almost constant with age (Fig. 3d). For MT-H NRAS mutations, and for all HRAS mutations, there were insufficient numbers to identify significant age-related trends.

We next compared the profile of specific KRAS mutations based on the age of the patient at the time of sequencing. This was done both by considering age as continuous variable and by using age 40 as a stratification point in individuals from 8 to >85 years of age. Analysis of the MT-L cohort by both approaches revealed that the most significant difference in mutational spectra of younger and older patients was the greater incidence of G12 mutations prevalent in younger patients versus the more frequent incidence of A146, K117, and Q61 mutations in older patients, while the incidence of G13 remained the same ($p = 1.3 \times 10^{-03}$) (Fig. 4a, b). Among the G12 mutations (Supplementary Fig. 4a, b), the frequency of substitution to G12V fraction appears to slightly increase with age, while substitution to G12A and G12C decreases. Strikingly, the overall fraction of Q61 substitutions in the KRAS mutation pool more than doubled in the older group, with ~332/6153 (5.2%) variants observed among patients ≥40 years versus 12/482 (2.5%) in patients <40 years (Fig. 4a, c). Further, some variants, such as the highly transforming G61K, were only observed in patients ≥40 years, with a linear increase correlating with increasing age (Fig. 4e).
Similar numeric trends in mutational spectrum were observed in the MT-H population, including a greater proportion of G12 and a lesser proportion of A146 and Q61 mutations in the <40 years subgroup, and a greater diversity of mutations in older patients. However, these findings did not reach statistical significance due to lower overall numbers, with only 44/235 (18.5%) of MT-H patients <40 years (Supplementary Fig. 4c); further, trends based on analysis of age as continuous variable were not conclusive (Supplementary Fig. 4d).

Based on simple dichotomization (Fig. 4a, right), or analysis with age as a continuous variable (Fig. 4b, right), the NRAS mutational spectra did not differ significantly by age in the MSS patients. However, analysis using age as continuous variable suggested a non-linear trend towards a greater proportion of Q61 mutations and a lower proportion of G13 mutants with increasing age, except for the oldest patients (Fig. 4c, right). More data are needed to confirm the validity of these trends.

**Missense mutation variation dependent on colorectal subsite.**
Among the 6634 defined missense mutations in KRAS in MT-L tumors, 983 (~15%) were from tumors originating in the rectum and 5651 (~85%) in the colon. For each CRC subsite, ~50% of the analyzed tumors had KRAS missense mutations. For the MT-H tumors, the corresponding numbers were 24 (10%) that arose in the rectum and 222 (90%) in the colon. The prevalence of KRAS missense mutations was much higher in patients with rectal cancer (21/35, ~60%) than in those with colon cancer (197/590,
More specific information about the anatomical subsite of tumors within the colon or rectum was not available, and sample size did not permit detailed analysis of mutation patterns in colon versus rectum for MT-H tumors. The overall KRAS and NRAS spectra for the MT-L subgroup did not differ between the colon and rectum subsites (Supplementary Fig. 4f, g). However, one specific codon substitution was observed more commonly in colon tumors than in rectal tumors: Q61K, found in 40 MT-L colon tumors and 4 MT-H colon tumors, but not in rectal tumors ($p = 2.9 \times 10^{-3}$, Supplementary Fig. 4e).

**CRC RAS missence mutation variation dependent on gender.**

Male patients with MT-L CRC had a slightly reduced frequency of KRAS mutations relative to female patients (49.4% and 53.5%, respectively; $p = 2.7 \times 10^{-5}$, Fig. 3a). Interestingly, age trends in the...
prevalence of KRAS missense mutations were distinct between males and females (Fig. 4d, left); an increase in mutational frequency with age was highly significant for males ($p = 3.8E^{-5}$), but not for females. Among MT-H cases, the total prevalence of KRAS mutations was higher in males (39.2% versus 30% in females, $p = 0.019$, Fig. 3a). The declining rate of such mutations with age did not reach statistical significance for males ($p = 0.036$), but did for females (Fig. 4d, center, $p = 7.3E^{-04}$). Among specific missense mutations, the overall spectrum was similar between males and females (Supplementary Fig. 4h), with the exception of variants in Q61, where Q61K mutations were more frequent in females ($p = 1.3E^{-03}$) (Fig. 4c).
In contrast to the pattern observed with KRAS mutations, the frequency of NRAS SV mutations in MT-H tumors did not differ significantly from that in MT-L tumors (ρ = 0.08), likely because of the lower number of mutations (Fig. 3a). Similarly, there was no conclusive evidence for gender-specific age trends for the prevalence of NRAS missense mutations (Fig. 4d, right), and specific missense amino acid substitutions within NRAS did not show gender-specific variance (Supplementary Fig. 4l). For HRAS, no gender-related difference in the prevalence of missense mutations was detected (Fig. 3a); there were not enough cases of patients with these mutations to explore gender-specific age trends or gender-specific mutation spectra.

Specific RAS mutation correlation with CNV. Although most studies of RAS activation in cancer focus on the activating role of point mutations, allelic imbalance arising from either the amplification of the mutated allele or loss of the wt allele can affect tumorigenicity and therapeutic outcome. Among all CRC tumors, 225 amplifications (defined as six or more copies) of wt KRAS, 21 of wt NRAS, and 11 of wt HRAS were observed among 12,711 MT-L tumors, while only one amplification of KRAS and none of NRAS or HRAS were detected in 625 MT-H tumors. Thus, amplifications of wt RAS genes are much more prevalent in MT-L tumors (ρ = 3.7E−04 for KRAS, and ρ = 8.3E−05 for all RAS genes combined) and comprise a much higher fraction of total RAS gene alterations in MT-L tumors (Fig. 5a), although the numbers did not reach statistical significance except for HRAS (ρ = 8.4E−04).

We analyzed RAS amplification for association with age or tumor subsite to the extent sample size allowed significant conclusions. There was no difference in the frequency of RAS amplifications between the colon and rectal subsets, overall. However, greater prevalence of KRAS amplification in CRC patients of younger age achieved borderline statistical significance (ρ = 5.9E−03; Fig. 5b). Dichotomization by age 40 years revealed that of 515 KRAS alterations in MT-L tumors from younger patients, 31 were amplifications (5.6%); from 6411 KRAS alterations in patients ≥40, 194 were amplifications in KRAS (2.8%), representing a 2-fold lower level in older patients (ρ = 6.8E−04) (Fig. 5c). Interestingly, KRAS amplifications also relatively frequently co-occurred with a RAS SV mutation in the same patient. Among 6322 MT-L CRCs with a single KRAS alteration (either amplification or SV), only 2.2% bear a KRAS amplification (140 occurrences). In contrast, in MT-L CRCs bearing two or more KRAS alterations, KRAS amplifications represent 83 out of 555 alterations (15%)—almost a 7-fold increase (ρ < 2.2E−16) (Fig. 5d).

For NRAS amplifications, no difference was found using dichotomization by age 40 years (Fig. 5c). However, comparative analysis of amplification rate in patients diagnosed at different age groups suggests that the rate of amplification may decline in tumors arising in older patients (Fig. 5b). The number of samples with HRAS amplifications was too low for detailed dissection of age–amplification relationships, but age 40 dichotomization also suggests an elevated prevalence in younger patients will emerge in analysis of a larger cohort (Fig. 5c).

Analysis of the profile of common KRAS and NRAS mutations in MT-L tumors in the context of allelic variation revealed some striking differences between distinct codon variants (Fig. 5e–g). Comparison of the pattern of missense mutations to the patterns of allelic gain and loss revealed specific correlations involving some of the more common missense mutations (Fig. 5e). Mutations affecting KRAS G12, present at 69.3% overall among missense mutations, were much less likely to be found when increasing copies of the mutant allele were present (representing 72.6% of all SV mutations found as one copy, but only 50.9% of the pool of KRAS SV mutations with five or more altered copies; association between copy number and G12 mutation fraction (ρ = 1.73E−11). The opposite pattern was observed with mutations in KRAS A146, representing 5.8% of overall mutations; 3.6% of samples with a single SV showed a mutation in A146, but 34.5% in samples with five or more mutated copies (ρ < 2E−16). In contrast, other missense variants, such as those affecting G13, do not have similar variance as a factor of copy number (Fig. 5e). A change in frequency of specific mutations based on amplification of the mutated allele was also observed with NRAS; surprisingly, here G12 mutations were more common in the context of higher copy number, whereas Q61 became less common (Fig. 5e).

Among the more common missense mutations, trends of association with CNV of the mutated allele were most strongly detected for missense variants of G12 and A146. In the setting of 0 copies of wt KRAS, G12 missense mutations were less likely to be amplified, and A146 more likely (Fig. 5f). As the number of wt copies of KRAS increased, these trends were lost, reflecting a significant relationship between wt copy number and specific mutations (ρ = 1.5E−10 for G12 and 3.8E−04 for A146). Among specific missense substitutions, G12D and G12V showed strong copy number dependence, but not G12A, C, or S; A146T showed a stronger copy number effect than the less common A146V (Fig. 5g). These patterns are also more prominent in tumors where the wt KRAS allele has been lost (Fig. 5g).

Discussion
Metastatic CRC serves as an example of the importance of molecular characterization in defining the optimal treatment paradigm for an individual patient. A limited number of targeted therapies have been approved for CRC, with anti-EGFR antibodies such as cetuximab and panitumumab the most successful to date. KRAS and NRAS are central downstream effectors of EGFR. Two landmark studies demonstrated improved OS and progression-free survival in patients with wt KRAS treated with anti-EGFR therapy, while showing no benefit to those with KRAS mutations. While initial studies investigated only mutations in codons 12 and 13 of the KRAS gene, additional studies have demonstrated other mutations that predict for resistance to EGFR-targeted treatment, those that do not predict for resistance, and the predictive value of expanded RAS testing, including NRAS and BRAF. Hence, there is considerable value in establishing the overall profile of RAS mutations relevant to CRC.

The large sample size in this study reinforces recent findings of other large studies, and also allows some statistically significant conclusions. As one point of interest, the rate of hypermutated patients observed in this study at 4.5% is significantly lower than that reported by the TCGA and others at ~13–15%. This difference reflects the fact that the patients analyzed by FM all presented with metastatic disease. MSI-H CRC has been reported to be less frequently metastatic than MSS CRC, with only 4–5% of stage 4 patients with MSI-H; the frequencies reported here are more likely to be relevant to this advanced population. Others have reported that MSI-H is more common in elderly females due to higher prevalence of CpG island methylator phenotype (CIMP) and MLH1 hypermethylation; this study did not identify such a correlation. This difference, also, may reflect differences in the study pool.

Others have found that KRAS mutant tumors were less likely than KRAS wt tumors to be deficient in mismatch repair pathways (7% versus 14% ρ < 0.001), agreeing with the findings of the present study. The data presented here both confirms previously identified mutational hotspots reported for CRC, or for other cancers, and also identifies some new hotspots, including
V9, V14, A68, R164, K176, and K180. Some of these (e.g., V1445) have previously been reported in sporadic tumors or in RASopathies, which include Noonan syndrome, Costello syndrome, cardiofacio-cutaneous syndrome, and neurofibromatosis type 116,45, and are typically associated with a milder RAS-activating function. Mutations that lead to these RASopathies impact intrinsic and GEF-induced nucleotide exchange, or affect intrinsic and/or GAP-induced hydrolysis. Modeling of these mutations on the 3D structure of KRAS places these residues adjacent to known critical functional areas for these processes (Fig. 2). Two previously unreported hotspot mutations, K176 and K180, target the positively charged lysine tract that characterizes the KRAS-4B isof orm; interestingly, one study has indicated that these residues interact with high affinity with the GDP-bound Switch1/Switch2 region, leading to the proposal they support auto-inhibition of KRAS46, which would provide one potential mechanism for activating function. Another interesting feature of the study is the recognition that specific common mutations are associated in characteristic ways with CNV of either the mutated or wild-type allele, expanding on earlier studies in this area using smaller cohorts.7,35

There is emerging awareness that the clinical behavior of colon cancer is impacted by the specific site of origin within the colon: for example, with tumors on the right side are associated with reduced survival relative to left-sided CRCs, data from multiple clinical trials demonstrating that left-sided colon cancer patients are more responsive to anti-EGFR therapy than right-sided cancers47, and an ongoing investigation into whether different mutational spectra exist in left versus right colore44,48,49 KRAS mutations, for example, have been reported to be more common in proximal and cecal tumors, associated with specific morphological features50,51. While the dataset in this study does not discriminate left from right colon, we did find notable differences in rectal cancers, as compared to colon cancers, including differing mutational spectra.

The age-related differences identified here are intriguing. Examination of young patients with CRC in small studies has shown a greater proportion of CIMP-low tumors in the sporadic (non-Lynch syndrome) cohort52. The Wnt/β-cat enin pathway has also been identified as more commonly activated in young patients53. To date, examination of the RAS pathway in these patients has been limited, and has been inconclusive, with a wide range of rates of KRAS mutations reported in young patients (4–54%)54,55, and no prior investigations into specific mutations are known. Finally, some of the data in this study hints at specific associations connecting age, gender, and tumor subsite with specific mutations. Of particular interest, the Q61K mutation is associated with patients who are older at testing, female, and diagnosed with colon cancer; this intriguing set of relationships suggests that investigation of cross-category linkages may identify unique disease subtypes that bear specific scrutiny. While data regarding specific indication for sequencing is unavailable for the samples in this study, the size of the dataset and varied provenance of the samples decreases the risk of the findings being due to selection bias. Finally, the clinical implications of these data would be illuminated with linked survival data, but these are not available for analysis. Nevertheless, this study and linked dataset should serve as a useful benchmark for future investigations.

Methods

Comprehensive genomic profiling. CGP was performed using the Foundatio-nOne assay (FM Inc., Cambridge, MA, USA), as previously described in detail56. Briefly, the pathologic diagnosis of each case was confirmed by review of hematoxylin- and eosin-stained slides and all samples that advanced to DNA extraction contained a minimum of 20% tumor cells. Hybridization capture of exonic regions from 315 cancer-related genes was applied to ≥50 ng of DNA extracted from formalin-fixed, paraffin-embedded clinical cancer specimens. These libraries were sequenced to high, uniform median coverage (>500×) and assessed for base substitutions, short insertions and deletions, copy number alterations, and gene fusions/rearrangements. MSI microsatellite track from the UCSC genome browser was aligned with short variants detected in KRAS and NRAS to determine whether specific mutations were associated with microsatellites. Comparison data sets for studies with information on KRAS and NRAS mutation status and, gender, age, and tumor subsite were collected from the cBioPortal database (http://www.cbioportal. org/index.do).

Statistical assessments of data. The relationships between mutations and patient characteristics were assessed using Fisher’s exact tests (including determination of significant difference between the mutational spectra of dichotomized age, gender, or subsite groups) and multivariable logistic regression models. For Fig. 4b, d and 5b, we determined whether age was associated with probability of mutations of interest using Pearson’s correlation, weighted to account for uneven sample sizes within age groups. To allow for multiple mutations within a patient, the predictors in these models were binary indicators for the presence/absence of particular mutations of interest. Separate models were conducted for the genes (KRAS/NRAS) and patient characteristics (gender, younger age, tumor site, MSI). Overall significance was assessed using a likelihood ratio test. Conditional inference trees57 were also used to determine which mutations were most strongly associated with patient characteristics of interest. CNV analyses were conducted using logistic regression analyses in the subgroup of patients with a KRAS mutation. Analyses were done overall and stratified by the number of KRAS wt CNVs. To determine if a frequently mutated site on the RAS proteins constitutes a mutational hotspot, we have used a binomial distribution model with a cumulative probability cutoff of 0.005.

Visualization of folded domains and mutation spectra. Figures were prepared using the program PyMOL58. The KRAS structure used is the GDP-bound form with KRASβ (PDB code: STAR)59. The NRAS structure used is the GppNHp-bound NRAS structure, which is 166 residues (PDB code: SUHV)60. The HVR is not represented in this resolved structure. Another available NRAS structure (PDB code: 3CON) was not used as it lacks information on residues 61–71. The HRAS structure61,62 (PDB code: 1CTQ) is used is the GppNHp-bound HRAS structure, which is 166 residues; this structure does not resolve the HVR.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the findings of this study appear in Supplementary Data 1. Clinical data for this cohort also appear in Supplementary Data 1. Because of the restrictions on disclosure based on the Health Insurance Portability and Accountability Act Privacy Rule and the Genetic Information and Nondiscrimination Act, we are unable to provide access to the raw sequencing reads because they include germline information. Furthermore, we provide qualified researchers access to additional detailed data through application to Foundation Medicine (https://www.foundationmedicine.com/insights-and-trials/foundation-insights).

Received: 18 July 2018 Accepted: 14 June 2019
Published online: 19 August 2019

References

1. SEER. Surveillance, Epidemiology, and End Results (SEER) Program (www. seer.cancer.gov) SEER*Stat Database: Incidence—SEER 9 Regs Research Data, November 2016 Sub (1973–2014) <Katrina/Rita Population Adjustment—> Linked to County Attributes—Total U.S., 1969–2015 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2017, based on the November 2016 submission. Accessed 7 February 2018.
2. Stigliano, V., Sanchez-Mete, L., Martayan, A. & Anti, M. Early-onset colorectal cancer: a sporadic or inherited disease? World J. Gastroenterol. 20, 12420–12430 (2014).
3. Arnold, D. et al. Prognostic and predictive value of primary tumour site in patients with RAS wild-type metastatic colorectal cancer treated with chemotherapy and EGFR directed antibodies in six randomized trials. Ann. Oncol. 28, 1713–1729 (2017).
4. Lopukhina, F. et al. Primary tumor location as a prognostic factor in metastatic colorectal cancer. J. Natl Cancer Inst. 107 https://doi.org/10.1093/jnci/dju427 (2015).
5. Lee, Y. C., Lee, Y. L., Chuang, J. P. & Lee, J. C. Differences in survival between colon and rectal cancer from SEER data. PLoS ONE 8, e78709 (2013).
6. Benedix, F. et al. Comparison of 17,641 patients with right- and left-sided colon cancer: differences in epidemiology, perioperative course, histology, and survival. Dis. Colon Rectum. 53, 57–64 (2010).
7. Meyer, J. E., Cohen, S. J., Ruth, K. J., Sigurdson, E. R. & Hall, M. J. Young age increases risk of lymph node positivity in early-stage rectal cancer. J. Natl. Cancer Inst. 100, https://doi.org/10.1093/jnci/djy284 (2016).
8. Bailey, C. E. et al. Increasing disparities in the age-related incidences of colon and rectal cancers in the United States, 1975–2010. JAMA Surg. 150, 17–22 (2015).
9. NCCNRCV. https://www.nccn.org/professionals/physician_gls/pdf/rectal.pdf. Accessed 20 April 2018.
10. Siegel, R. L. et al. Colorectal cancer statistics, 2017. CA Cancer J. Clin. 67, 177–193 (2017).
11. Grady, W. M. & Pritchard, C. C. Molecular alterations and biomarkers in colorectal cancer. Toxicol. Pathol. 42, 124–139 (2014).
12. Lieu, C. H. et al. Comprehensive genomic landscapes in early and later onset colorectal cancer. Clin Cancer Res. clancarenses.0899.2019 (2019).
13. Hobbs, G. A., Der, C. J. & Rossman, K. L. RAS isoforms and mutations in cancer at a glance. J. Cell Sci. 129, 1287–1292 (2016).
14. Arrington, A. K. et al. Prognostic and predictive roles of KRAS mutation in colorectal cancer. Int. J. Mol. Sci. 13, 12153–12168 (2012).
15. Domingo, E. et al. Use of multivariate analysis to suggest a new molecular classification of colorectal cancer. J. Pathol. 229, 441–448 (2013).
16. Haigis, K. M. KRAS alleles: the devil is in the detail. Trends Cancer 3, 686–697 (2017).
17. Janakiraman, M. et al. Genomic and biological characterization of exon 4 KRAS mutations in human cancer. Cancer Res. 70, 5901–5911 (2010).
18. Guerrero, S. et al. K-Ras codon 12 mutation induces higher level of resistance to apoptosis and predisposition to anchorage-independent growth than codon 13 mutation or proto-oncogene overexpression. Cancer Res. 60, 6750–6756 (2000).
19. Csepedes, M. V. et al. K-ras Asp12 mutant neither interacts with Raf, nor signals through Erk and is less tumorigenic than K-ras Val12. Carcinogenesis 27, 2190–2200 (2006).
20. Benesova, L., Minarik, M., Janancarikova, D., Belsanova, B. & Pesek, M. Multiplicity of EGFR and KRAS mutations in non-small cell lung cancer (NSCLC) patients treated with tyrosine kinase inhibitors. Anticancer Res. 30, 1667–1671 (2010).
21. Fiola, O. et al. GI2V and GI2A KRAS mutations are associated with poor outcome in patients with metastatic colorectal cancer treated with bevacizumab. Tumour Biol. 37, 6823–6830 (2016).
22. Imanura, Y. et al. Specific mutations in KRAS codons 12 and 13, and patient prognosis in 1075 BRAF wild-type colorectal cancers. Clin Cancer Res. 18, 4753–4763 (2012).
23. Giannakis, M. et al. Genomic correlates of immune-cell infiltrates in colorectal carcinoma. Cell Rep. 15, 857–865 (2016).
24. Yaeger, R. et al. Clinical sequencing defines the genomic landscape of metastatic colorectal cancer. Cancer Cell 33, 125–136 e123 (2018).
25. Consortium, A. P. G. AACR Project GENIE: powering precision medicine. J. Clin. Oncol. 36, 2200 (2018).
26. Kamat, K. P. et al. KRAS and BRAF mutations in predicting recurrence and benefits from chemotherapy in colorectal cancer. J. Clin. Oncol. 29, 1261–1270 (2011).
27. Schram, T. et al. High-affinity interaction of the K-RasB hypervariable region with the Ras active site. Biophys. J. 109, 2602–2613 (2015).
28. Loree, J. M. & Kopeta, S. Recent developments in the treatment of metastatic colorectal cancer. Ther. Adv. Med. Oncol. 9, 551–564 (2017).
29. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. Nature 487, 330–337 (2012).
30. Luo, J. et al. Differentially expressed protein expression profiles, KRAS and BRAF mutation, and prognosis in right-sided colon, left-sided colon and rectal cancer. Sci. Rep. 7, 7882 (2017).
31. Yamauchi, M. et al. Assessment of colorectal cancer molecular features along bowel subwells challenges the conception of distinct dichotomy of proximal versus distal colorectal. Gut 61, 847–854 (2012).
32. Rosty, C. et al. Colorectal carcinomas with KRAS mutation are associated with distinctive morphological and molecular features. Mod. Pathol. 26, 825–834 (2013).
33. Ballester, V., Rashtak, S. & Boardman, L. Clinical and molecular features of young-onset colorectal cancer. World J. Gastroenterol. 22, 1736–1744 (2016).
34. Kirzin, S. et al. Sporadic early-onset colorectal cancer is a specific sub-type of cancer: a morphological, molecular and genetics study. PLoS ONE 9, e103159 (2014).
35. Chang, D. T. et al. Clinicopathologic and molecular features of sporadic early-onset colorectal adenocarcinoma: an adenocarcinoma with frequent signet ring cell differentiation, rectal and sigmoid involvement, and adverse morphologic features. Mod. Pathol. 25, 1128–1139 (2012).
36. Watson, R., Liu, T. C. & Ruizinova, M. B. High frequency of KRAS mutation in early onset colorectal adenocarcinoma: implications for pathogenesis. Hum. Pathol. 56, 163–170 (2016).
37. Freampton, G. M. et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Nat. Biotechnol. 31, 1023–1031 (2013).
38. Hothorn, T., Hornik, K. & Zeileis, A. Ctree: Conditional Inference Trees. The Comprehensive R Archive Network (2015). https://rdrr.io/rforge/parttykit/f/inst/doc/ctree.pdf.
39. PyMOL. The PyMOL Molecular Graphics System, Version 2.0 (Schrodinger, LLC), https://pymol.org/2/.
40. Dharmia, S. et al. Structural basis of recognition of farnesylated and methylated KRAS by PDEdelta. Proc. Natl Acad. Sci. USA 113, E6766–E6775 (2016).
41. Scheidig, A. J., Burmester, C. & Goody, R. S. The pre-hydrolysis state of p21 Ras GTPase in complex with GTP: new insights into the role of water molecules in MAP kinase dependence in cancer. Cell 168, 817–829 e815 (2017).
42. Karapetis, C. S. et al. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. N. Engl. J. Med. 359, 1757–1765 (2008).
43. Amado, R. G. et al. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. J. Clin. Oncol. 26, 1626–1634 (2008).
44. Douillard, J. Y. et al. Panitumumab-FOLFIRX treatment and RAS mutations in colorectal cancer. N. Engl. J. Med. 369, 1023–1034 (2013).
45. Kras and KRAS. https://www.nature.com/naturecommunications. ARTICLE

Acknowledgements
We thank I. Karanicolas and R. Dunbrack for consultation on analysis of protein structure. The authors were supported by NCI Core Grant P30 CA06927 (to Fox Chase Cancer Center), NIH R01 DK108195 (to E.A.G.), NIH R01 CA229259-01 (to C.L.), by a subsidy of the Russian Government to support the Program of Competitive Growth of Kazan Federal University (to I.G.S.), and Department of Defense Cancer Development Award W81XWH-18-1-0148 (to S.A.), and by the Colorectal Cancer Alliance (to E.A.G., J.E.M.).
Author contributions
C.C., G.F., M.C., V.M., S.Ali, and J.S.R. collected data. I.G.S., E.H. S.Arora., and J.S.R. performed data analysis. E.A.G., I.G.S. S.Arora., C.L. and J.E.M. designed the study and wrote the manuscript.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-11530-0.

Competing interests: C.C., G.F., M.C., V.M., S.Ali, and J.S.R. are employed by FM Inc., and own stock in FM Inc. The remaining authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Peer review information: Nature Communications thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019