CANCER GENOMICS

original reports abstract

# Clustered 8-Oxo-Guanine Mutations and Oncogenic Gene Fusions in Microsatellite-Unstable Colorectal Cancer

Russell W. Madison, MS[1]; Xiaoju Hu, PhD[2]; Vivek Ramanan, PhD[1]; Zhuxuan Xu, MS[2]; Richard S.P. Huang, PhD[1]; Ethan S. Sokol, PhD[1]; Garrett M. Frampton, PhD[1]; Alexa B. Schrock, PhD[1]; Siraj M. Ali, MD, PhD[1]; Shridar Ganesan, MD, PhD[2]; and Subhajyoti De, PhD[2]

**PURPOSE** Colorectal carcinomas (CRCs) with microsatellite-instability (MSI) are enriched for oncogenic kinase fusions (KFs), including *NTRK1*, *RET*, and *BRAF*, but the mechanism underlying this finding is unclear.

**METHODS** The genomic profiles of 32,218 advanced CRC tumor specimens were analyzed to assess the fusion breakpoints of oncogenic alterations including KFs in microsatellite-stable and microsatellite-unstable CRC. Genomic contexts of such alterations were analyzed to obtain mechanistic insights.

**RESULTS** Genomic analysis demonstrated that oncogenic fusion breakpoints in MSI tumors do not preferentially involve repetitive or low-complexity sequences. Instead, their junction regions showed pronounced guanine and cytosine bias and elevated mutation frequency at G:C contexts. Elevated mutation frequency at G:C bases in relevant introns predicted prevalence of associated oncogenic fusions in MSI CRCs. CRCs harboring mismatch repair signatures had enrichment of butyrate-producing microbial species, reported to be associated with induction of 8-oxoguanine lesions in the intestine.

**CONCLUSION** Detailed analysis of breakpoints in MSI-associated KFs support a model in which inefficient repair and/or processing of microbiome-induced clustered 8-oxoguanine damage in MSI CRC contributes to the increased incidence of specific oncogenic fusions.

**JCO Precis Oncol 6:e2100477. © 2022 by American Society of Clinical Oncology**



## INTRODUCTION

A subset of colorectal carcinoma (CRC) arises in the setting of an underlying defect in mismatch repair (dMMR) leading to microsatellite instability (MSI), a phenotype defined by variation in the length of microsatellite repeats. Such MSI cancers, of colonic origin or otherwise, may arise in patients with germline mutations in mismatch repair (MMR) genes (eg, Lynch syndrome) or by acquired somatic mutations or epigenetic silencing of MMR genes in sporadic cancers. Although dMMR tumors are often characterized by the MSI phenotype, these cancers are also characterized by a high tumor mutational burden (TMB), enrichment for specific mutational signatures when compared with microsatellite-stable (MSS) colon cancers.[1,2] Metastatic MSI colon cancers are very responsive to immune checkpoint blockade. Unfortunately, only about half of MSI CRC benefit from these agents, creating an unmet need for other treatment approaches for these cancers.[2]

MSI CRC is enriched for specific druggable oncogenic kinase fusions (KFs), particularly NTRK fusions,[3] presenting potential therapeutic opportunities. However, the mechanism of selective enrichment of these fusions in MSI CRCs but not other MSI cancers of noncolonic origin remains obscure—which motivated our investigation.

## METHODS

Approval for this study, including a waiver of informed consent and a Health Insurance Portability and Accountability Act waiver of authorization, was obtained from the Western Institutional Review Board (Protocol No. 20152817).

### Junction Sequence Determination

Comprehensive genomic profiling was performed using FoundationOne or FoundationOneCDx assays as previously described (see Appendix 1 for details). Sequences from breakpoint spanning reads were mapped to the partner gene DNA sequence to generate junction sequences. A minimum of five reads and consensus sequence length of at least 10 base pairs (bp) were used to identify the partner breakpoint. Junction sequences consisted of 50 bp from the kinase gene of interest and 50 from the partner gene and were aligned such to read in the positive direction of the kinase reference sequence. The target regions in the oncogenic introns were sufficiently large, and within these targeted regions, identified junction points

ASSOCIATED CONTENT

Appendix

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on April 4, 2022 and published at ascopubs.org/journal/po on May 18, 2022: DOI https://doi.org/10.1200/PO.21.00477

## CONTEXT

### Key Objective

A subset of colorectal carcinoma (CRC) shows a phenotype characterized by variation in the length of microsatellite repeats because of defects in mismatch repair pathways leading to microsatellite instability (MSI). Only about half of MSI CRCs benefit from immune checkpoint blockade, creating an unmet need for other treatment approaches for these cancers. Interestingly, MSI CRC is enriched for specific druggable oncogenic kinase fusions (KFs), particularly NTRK fusions, but underlying mechanisms are poorly understood.

### Knowledge Generated

A comprehensive analysis of breakpoints in MSI-associated KFs in CRCs suggests that inefficient repair and/or processing of clustered 8-oxoguanine damage contributes to the increased prevalence of specific oncogenic fusions in MSI CRCs.

### Relevance

Anecdotal evidence suggests that patients with MSI CRC harboring KFs may respond to targeted agents. It is possible that combination therapy targeting oncogenic fusions and immune checkpoint blockade may improve long-term outcome in KF-positive MSI CRC.

did not systematically segregate on the basis of their MSI status.

### Genomic and Microbial Profiling of the TCGA Cohorts

We obtained somatic mutation calls for 63 The Cancer Genome Atlas–colon adenocarcinoma (TCGA-COAD) samples with whole genome sequencing data from Yang et al[4] MMR gene (MSH2, MSH6, MLH1, and PMS2) mutation and hypermethylation status were reported in this study. We obtained the data on transcribed fusions from the major TCGA cancer cohorts and analyzed those attributed to MSI and MSS colon cancer samples.[5] We determined the extent of microbial abundance for the TCGA samples from Poore et al.[6] We analyzed microbial abundance at the level of order, except for cyanobacteria, which was reported at the level of order and family.

### Sequence Context Analysis

Annotation of known Gencode genes and repeat elements was obtained from the University of California Santa Cruz genome browser (hg19). Sequence complexity measures including entropy, Wootton & Federhen complexity (CWF) and Markov model sequence complexity (size N = 3, 4, 5, 6), guanine and cytosine (GC) content, GC skew, and CpG content were computed using SeqComplex.[7]

### Mutation Signature Analysis

We analyzed the weights of the COSMIC v3 mutation signatures from frequency of the somatic mutations in the trinucleotide contexts in the CRC samples using deconstructSigs.[8] We used the mutation signature attribution method[9,10] to probabilistically infer the most likely signature associated with individual mutations.

### Statistical Analysis

We used R version 3.5.2 and python to perform data integration and statistical analyses. For meaningful comparisons, statistical tests and *P* values are reported in the main text. *P* values are based on two-tailed tests, unless specified. False discovery rate correction for multiple testing was applied where necessary.

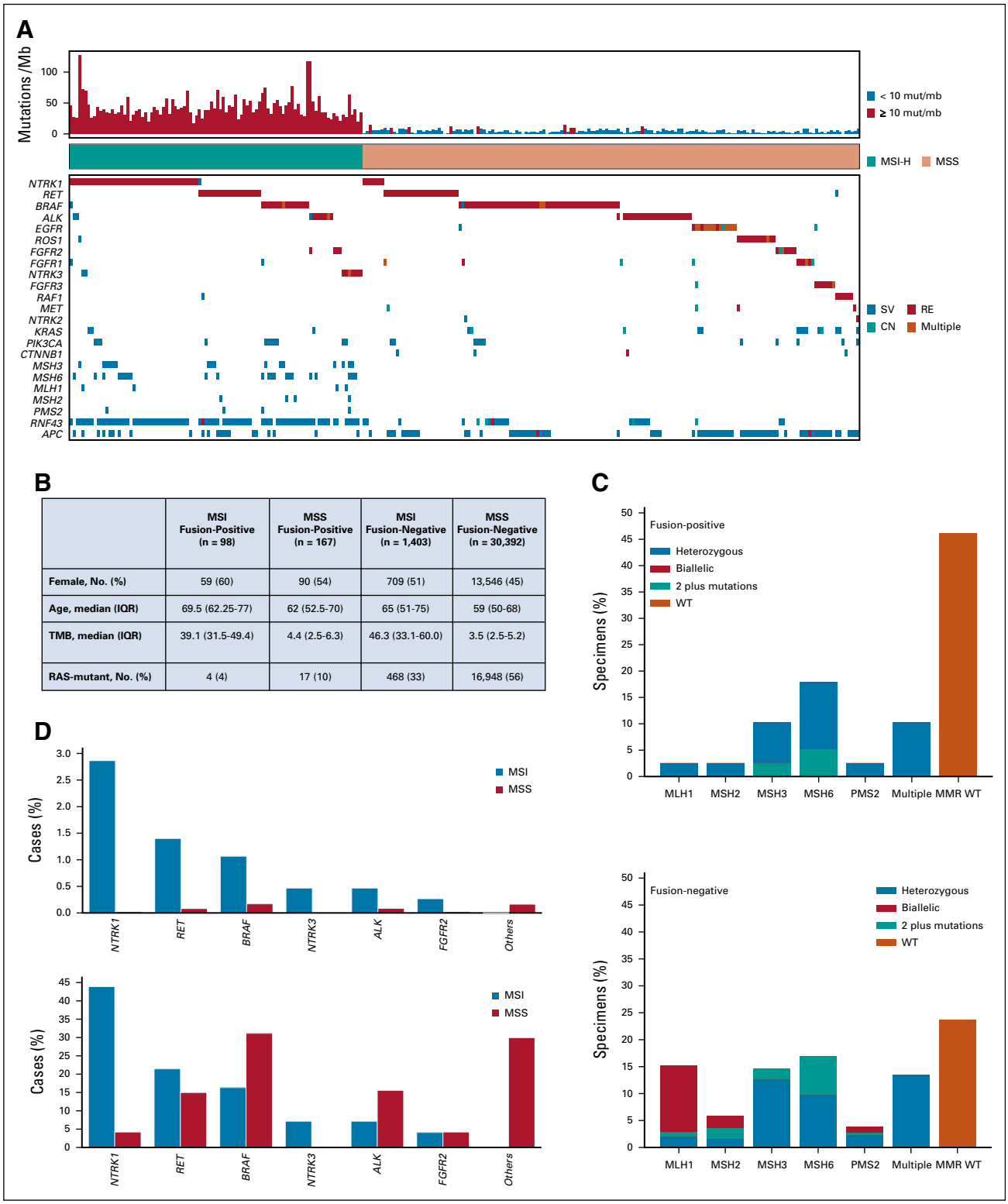
## RESULTS

### MSI Colon Cancers Are Enriched for Specific Oncogenic Fusions

The genomic profiles of 32,218 advanced CRC tumor specimens were reviewed to assess the genomic landscape of oncogenic alterations including KFs in MSI and MSS CRCs (Fig 1A). The available clinical and genomic features were consistent with previous studies (Fig 1B, Appendix Table A1, and Appendix Fig A1). Patients with KF-positive MSI CRC and KF-positive MSS CRC had higher median age compared with their KF-negative counterparts (MSI: 69.5 *v* 65 years, $P = 2.0E-04$; MSS: 62 *v* 59 years, $P = 1.0E-02$), and their tumors were less frequently *RAS*-mutated (MSI: 4% *v* 33%, $P = 9.1E-12$; MSS: 10% *v* 56%, $P = 4.5E-35$), supporting KFs as the primary oncogenic driver of these tumors. For KF-positive specimens, all MSI cases (100%, n = 98) had TMB > 10 mut/mb (range 18-128 mut/mb) while TMB > 10 mut/mb was rare in the MSS subset (9 of 167, 5%). Alterations in MMR genes were only seen in the KF-positive MSI cohort (KF-positive MSI: 43% *v* KF-positive MSS: 0%), and inactivating mutations in *RNF43* were enriched in MSI, whereas *APC* mutations were more frequent in MSS tumors (Fig 1A).

For a subset of evaluable MSI CRC (39 KF-positive and 477 KF-negative), allelic status of MMR alterations was assessed. Alterations in MMR genes were less frequent in KF-positive MSI tumors compared with the MSI KF-negative cohort (43% *v* 67%), and the frequency of biallelic MMR inactivation was also lower in KF-positive MSI cases (8% *v* 34%; Fig 1C). Among 311 KF-negative cases, germline mutations were detected in 15% (48 of 311) of cases while no germline mutations were seen in the KF-positive cases. These observations indirectly suggest that MMR defect in the KF-positive MSI cohort may be due to




| | MSI Fusion-Positive (n = 98) | MSS Fusion-Positive (n = 167) | MSI Fusion-Negative (n = 1,403) | MSS Fusion-Negative (n = 30,392) |
|---|---|---|---|---|
| Female, No. (%) | 59 (60) | 90 (54) | 709 (51) | 13,546 (45) |
| Age, median (IQR) | 69.5 (62.25-77) | 62 (52.5-70) | 65 (51-75) | 59 (50-68) |
| TMB, median (IQR) | 39.1 (31.5-49.4) | 4.4 (2.5-6.3) | 46.3 (33.1-60.0) | 3.5 (2.5-5.2) |
| RAS-mutant, No. (%) | 4 (4) | 17 (10) | 468 (33) | 16,948 (56) |

**FIG 1.** Select KFs are enriched in MSI CRC. (A) Patterns of oncogenic alterations including KFs and somatic mutation burden in MSI and MSS CRCs. Neither MMR alterations nor elevated tumor mutational burden was seen in KF-positive MSS tumors. Co-occurring oncogenic alterations were rare in KF-positive tumors. (B) Demographic and genomic features of the patients in this data set are consistent with previous studies. (C) Biallelic inactivation of the MMR pathway was less common in KF-positive MSI tumors (top) compared with KF-negative MSI tumors (bottom), consistent with the findings of Cocco et al that MLH1 methylation is enriched in this subset. Cases with two mutations in the same gene (2 plus mutations) were assumed to be biallelic. (D) Frequency of specific KFs in all MSI and MSS CRCs (top) and frequency of specific fusions with in KF-positive MSI and KF-positive MSS CRCs (bottom). CRC, colorectal carcinoma; IQR, interquartile range; KF, kinase fusion; MMR, mismatch repair; MSI-H, microsatellite instability high; MSS, microsatellite stable; TMB, tumor mutational burden; WT, wild type.

*MLH1* promoter methylation and rather than biallelic mutations in MMR genes, consistent with previous findings from Cocco et al.[3]

KFs were detected in 267 CRCs (0.83%) and involved 13 distinct kinase genes (Appendix Table A2). KFs were significantly more frequent in MSI tumors compared with MSS cases (6.5% *v* 0.5%, $P = 4.6E-61$); moreover, specific KFs were differentially enriched. *ALK* ($P = 7.1E-04$), *BRAF* ($P = 7.2E-08$), *FGFR2* ($P = 1.2E-3$), *NTRK1* ($P = 2.7E-50$), *NTRK3* ($P = 4.9E-10$), and *RET* ($P = 2.4E-16$) fusions were enriched in MSI tumors relative to MSS tumors (Fig 1D) while *ROS1*, *RAF1*, *MET*, *FGFR3*, *FGFR1*, *NTRK2*, and *EGFR* fusions occurred exclusively in MSS tumors. No notable differences between fusion partners were observed in MSI versus MSS cases (Apppendix Table A2).

Three thousand ninety-four (1.3%) non-CRC cases were MSI with 62 of these harboring fusions (2.0%) compared with 9,849 (4.5%) MSS non-CRC cases. The most common fusions in non-CRC MSI cases were in prostate acinar carcinoma and involved nonkinase *TMPRSS2*, and these fusions were enriched in MSS ($P = 3.5E-08$). Only four KFs (three *RAF1* and one *BRAF*) were detected in MSI prostate cases and were not enriched when compared with MSS cases. Oncogenic fusions in endometrial adenocarcinoma were identified in three unique kinases, none of which were enriched in MSI versus MSS tumors (Appendix Figure A2).

We surveyed KF-negative versus KF-positive MSI CRC cohorts for other oncogene mutations. *BRAF*-activating mutations, which are known to be associated with MSI CRC, were enriched in the KF-negative cohort (17% *v* 43%, $P = 2.1E-06$). Alterations in other oncogenes such as *PIK3CA* ($P = 1.1E-06$) and *CTNNB1* ($P = 2.3E-06$) were more frequent in KF-negative cases. Alterations in wingless-related integration site pathway regulator *RNF43* were more frequent in KF-positive cases while KF-negative cases had more frequent *APC* alterations (Appendix Fig A3).

### Characteristics of Sequence Context of Fusion Junction Sequences

Fusion junctions were clustered in certain intronic hotspots within the capture regions, as shown for *NTRK1* (Fig 2A), which are enriched in MSI CRCs but also detected in MSS CRCs (Figs 1A and 1D). For a given oncogene and its partner gene, the fusion breakpoints were not recurrent at the bp resolution within or between the MSI and MSS groups. For instance, fusion junction points in the *NTRK1* gene in MSI and MSS tumors from both CRC and non-CRC were mostly distributed within intronic regions between the 8th to the 12th exons and generally overlapped at kb-scale resolution. Fusion junction points or adjoining regions did not show over-representation of any specific polynucleotide motif (kmer size $> 5$). We observed similar results for other fusions in MSI and MSS CRCs (data not shown).

In both MSI and MSS CRCs, a majority of the fusions showed limited homology in the junction points between the oncogene and corresponding partner gene regions. Surprisingly, only 11.9% of the junction regions of fusions in MSI tumors overlapped with known repeat classes (simple repeats, Alu, L1, L2, satellite etc), which was lower than that observed in the MSS tumors (25.8%; $P < 9.0E-03$; Fig 2B). We performed a similar comparison for the junction regions of the partner genes, and no significant difference was observed (MSI: 53.2%, MSS: 54.9%; Fisher's exact test; Fig 2B).

Next, for each fusion, we extracted 101 bp genomic sequences centered at the junction point for the corresponding cancer gene and calculated entropy, CWF, and Markov model sequence complexity (size N = 3, 4, 5, 6) and also GC content, GC skew, and CpG content using SeqComplex.[7] We observed no significant difference between the MSI and MSS tumors in terms of sequence complexity for the junction region of the cancer gene involved (Wilcoxon test; Fig 2C). We repeated the analyses for the partner genes involved in fusions and found that CWF content was higher in MSS tumors ($P < 1.5E-02$). These results suggest that DNA breaks in low-complexity or repeat regions are unlikely to be a key driver underlying the enrichment of observed fusions in MSI CRCs (Fig 2C).

### Fusion Junction Regions Show GC Bias

GC bias of the junction regions of the fusion-associated cancer genes was significantly higher in MSI compared with MSS tumors (Fig 3A, $P < 5.0E-03$). We observed similar, although more modest, results for the partner gene (Wilcoxon test; $P < 5.0E-03$). There was no significant difference in CpG dinucleotide content between the groups. Rather, most cytosine-containing non-CpG dinucleotide contexts (AG+CT, CA+TG, CC+GG, and GpC) showed preferential enrichment in the MSI tumors. Next, we computed cumulative GC content of local sequence context at different distances from the fusion junction points (Fig 3B). We observed that the fusion junction point and its immediate proximal nucleotide positions had a pronounced GC bias in MSI tumors compared with the MSS tumors. The fusion partner genes also showed broadly similar differences between MSI and MSS groups, but with relatively modest effect sizes. Thus, fusions in MSI tumors are more likely to be associated with higher non–CpG-based GC content in their junction regions, and the effect is pronounced for the participating oncogene.

Analyzing mutational signatures associated with all somatic point mutations within the capture regions (Fig 3C), we found that the mutational landscapes of the MSS CRCs were dominated by CpG>TpG substitutions, which is attributed to deamination of 5-methylcytosine, a common mutational process in all somatic tissues; in contrast, the mutational landscapes of MSI CRCs were dominated by T>N substitutions, especially at TTT context, a classic MSI signature.[1] In addition, these samples also had an increase in C>T (also C>G and C>A) at non-CpG context, which are

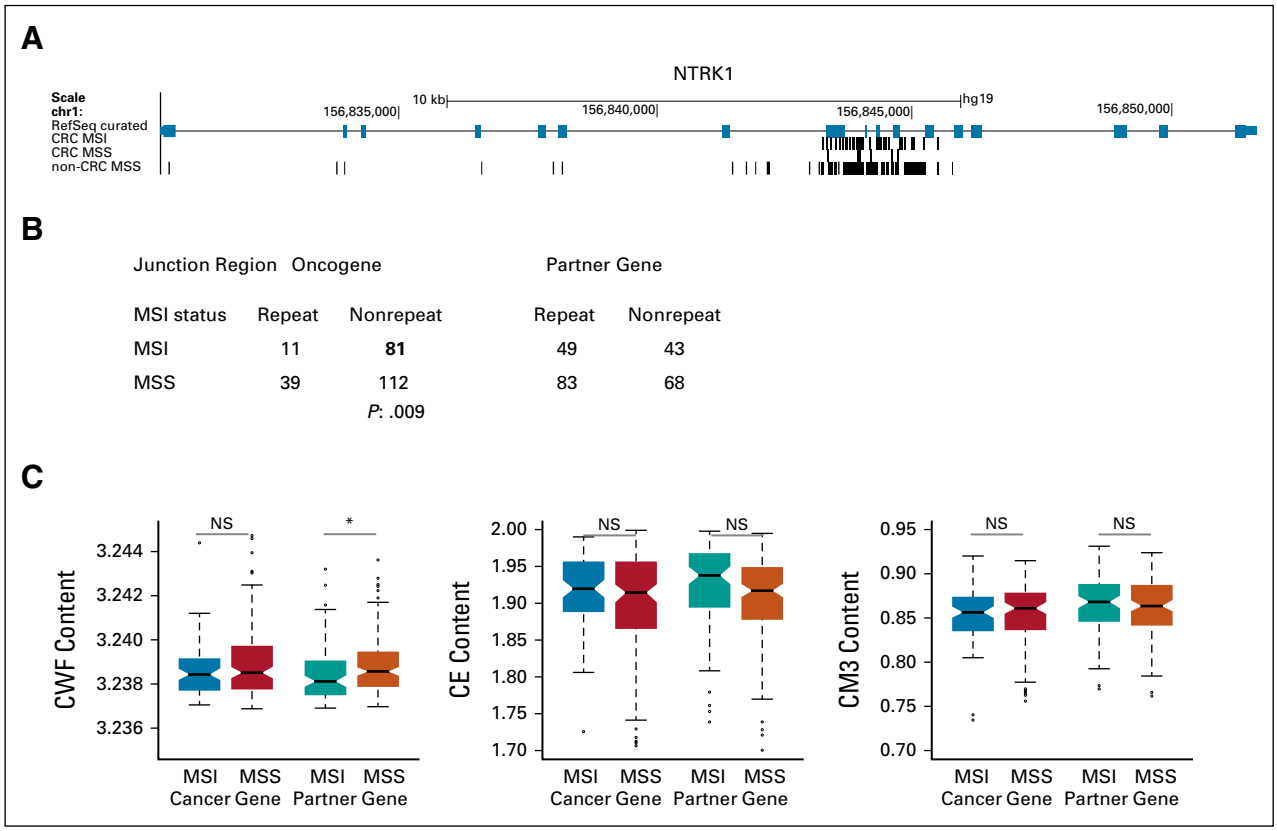


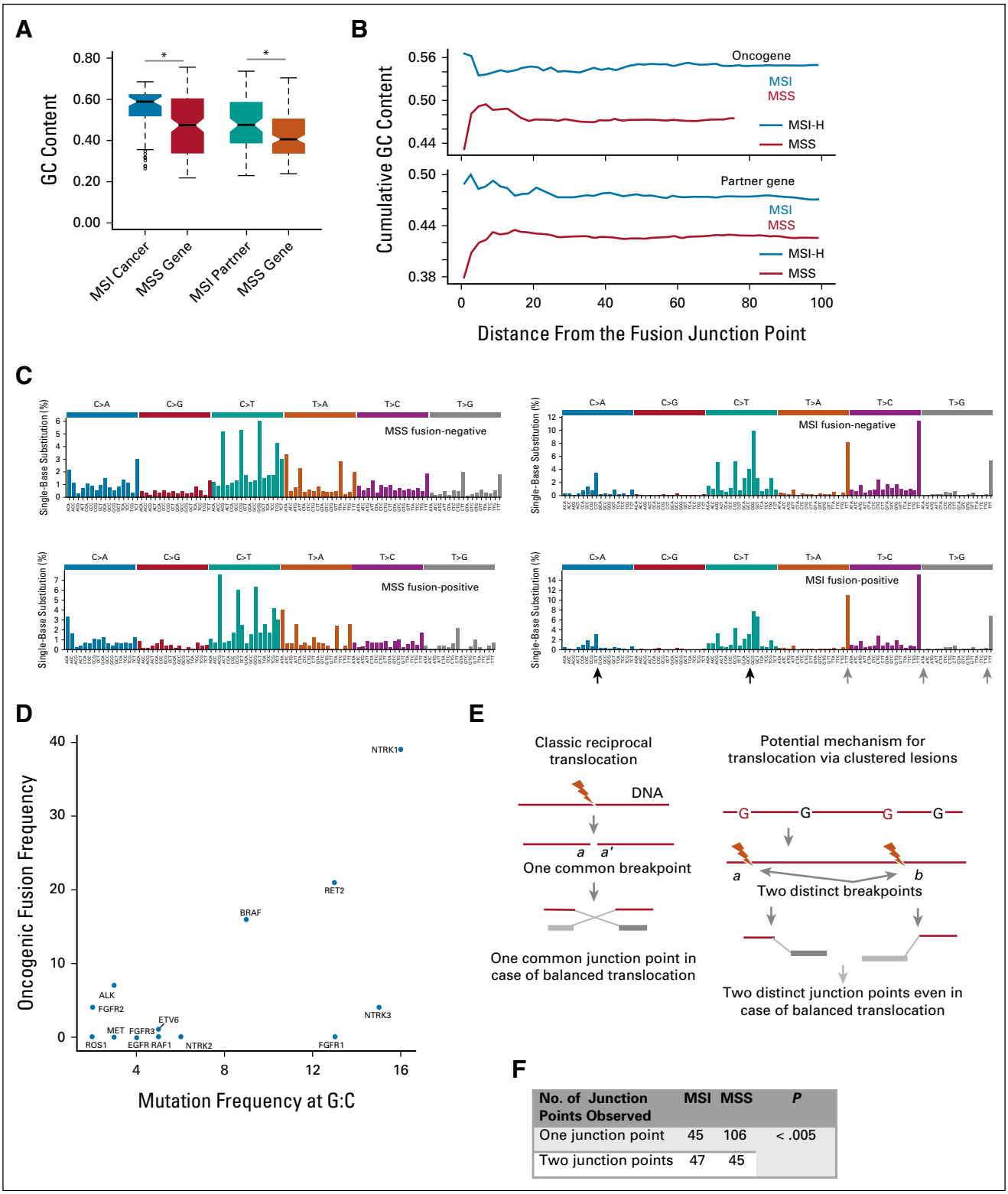

**FIG 2.** Junction regions of oncogenic fusions do not suggest a direct role of MSI. (A) Oncogenic fusion breakpoints in the NTRK1 locus shown in MSI and MSS CRCs and non-CRCs. (B) Table showing overlap of the fusion junctions with common repeat sequences. (C) Boxplot showing the extent of low-complexity sequences, which are common in repeat-rich regions in oncogenic junctions observed in MSI and MSS CRCs. * = $P < 0.05$ CE, calculated entropy; CM3, Markov model sequence complexity for N = 3; CRC, colorectal carcinoma; CWF, Wootton & Federhen complexity; MSI, microsatellite instability; MSS, microsatellite stable; NS, not significant.

similar to the MMR defect signatures SBS6 and SBS15 (cosine similarity: 0.70 and 0.62, respectively), which are particularly prevalent in CRC.[1] SBS12, a mutational signature of unknown etiology, was also present in both fusion-positive and -negative MSI CRCs, but not in the MSS tumors. We further observed somatic mutations with GC bias in the junction proximal regions of the candidate fusions in MSI CRCs.

## Context-Dependent Mutagenesis and DNA Breaks in MSI CRCs

Since the rearrangement events leading to oncogenic fusions do not necessarily preserve the original lesions, we analyzed the frequency of somatic mutations involving GC bases at junction proximal regions. The junction proximal capture region in our cohort did not cover the entire intron; thus, to avoid bias, we analyzed 63 TCGA-COAD samples with whole genome sequencing data.[4] In this cohort, there were 10 MSI and seven *POLE*-mutant tumors, and the remaining were MSS. A substantial proportion of the mutations in the 10 kb windows in the MSI tumors had signatures of dMMR. In general, mutations involving G:C bases were > 50% for both MSI and MSS tumors, and preference for mutation at G:C context was more prominent for genes that are more commonly involved in fusions in MSI tumors. For instance, in NTRK family genes, the preference was marginally stronger (MSI: 63%; MSS: 61%) than for other kinases (Fig 3D). Thus, higher GC content contributes to a higher frequency of somatic mutations involving G:C context in these regions, especially in the MSI tumors.

Interestingly, the frequency of somatic mutations was more pronounced in some genes that frequently undergo fusion in MSI tumors. For instance, 10 kb windows corresponding to junction regions of *NTRK1* had an excess of somatic mutations while corresponding regions of *ALK* had substantially fewer mutations. The frequency of somatic mutations at G:C content in the junction window in MSI tumors significantly correlated with the frequency of fusions (correlation: 0.63; $P < 1.3E-02$, Fig 3D); no equivalent correlation was observed for mutations involving non-GC bases or those in MSS tumors. This suggests that local mutation frequency at GC contexts in MSI tumors can predict the frequency of corresponding fusions. Although the genes enriched for fusions had high local somatic mutation frequency in the junction regions, the converse




| No. of Junction Points Observed | MSI | MSS | *P* |
|---|---|---|---|
| One junction point | 45 | 106 | < .005 |
| Two junction points | 47 | 45 | |

**FIG 3.** Junction regions of oncogenic fusions suggest preferential mutagenesis at G:C bases in MSI CRCs. (A) Boxplot showing significant differences in GC content in oncogenic junctions between the MSI and MSS CRCs. (B) Cumulative GC content of local sequence context at different distances from the fusion junction points. Fusion junction site in the cancer gene tends to have a significant excess of G:C in MSI tumors, compared with the MSS tumors; the partner gene showed modest difference between the two groups. (C) Substitution frequencies of all somatic point mutations within the capture regions of (left) MMR proficient, MSS and (right) MMR deficient, microsatellite-unstable CRCs. Although mutational landscape of the former is dominated by C>T mutations at CpG context, MSI CRCs not only show MMR defect signatures characterized by T>N substitutions (continued on following page).

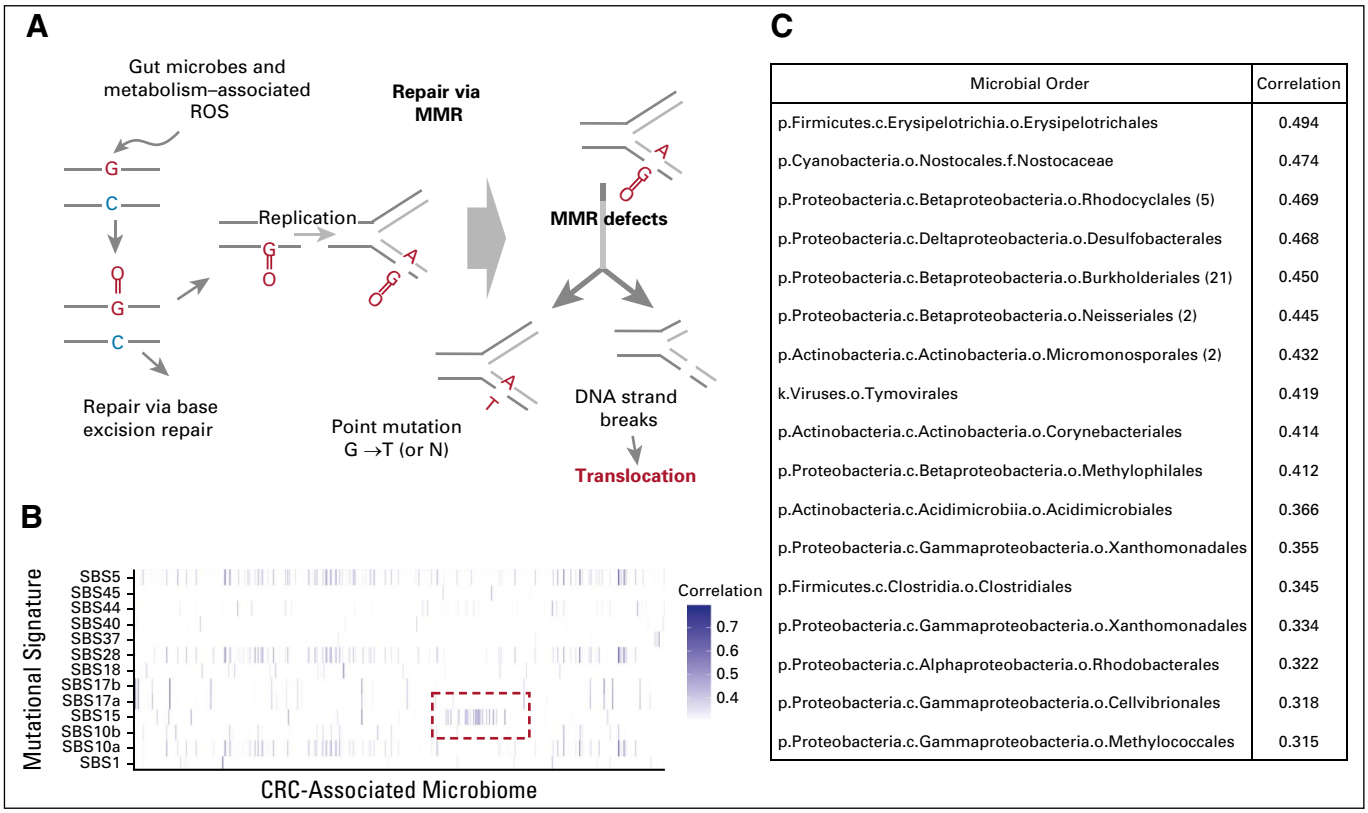
 

**FIG 3.** (continued). at TTT context, which is typically observed in microsatellite and other low-complexity sequence contexts, but also C>T and C>A substitutions in GC-rich contexts that are consistent with COSMIC mutational signatures of MMR defects (eg, SBS6 and SBS15). (D) Scatterplot showing association between the frequency of oncogenic fusion events in MSI CRC with mutation frequency at G:C sites in 10 kb junction-proximal regions. (E) Although in a classic case of balanced translocation with minimal end processing the junction points are equivalent, in case of DNA breaks arising from clusters of lesions (eg, 8-oxoguanine), there could be potentially two distinct junction points. (F) Oncogenic fusions in MSI are significantly more likely than MSS tumors to have two detectable translocation junction points between the oncogene partner gene pair (Fisher test; $P < .005$). *$P < .005$. A, adenosine; C, cytosine; CRC, colorectal carcinoma; G, guanine; GC, guanine and cytosine; MMR, mismatch repair; MSI-H, microsatellite instability high; MSS, microsatellite stable; T, thymine.

was not necessarily true; some cancer genes had high local mutation frequency but low frequency of fusions in the cohort, which might be due to selection on fusions during tumorigenesis.

### MMR-Mediated Repair of 8-Oxoguanine and DNA Breaks

A major source of mutations at the G:C context in CRC is oxidative DNA damage, particularly the formation of 8-oxoguanine (8-oxo-G). 8-oxo-G is repaired by the MMR system or OGG1-dependent base excision repair (BER) under physiologic conditions, but this repair is significantly impaired in the dMMR colon leading to elevation of 8-oxo-G level in dMMR CRC.[11,12] Error-prone repair of 8-oxo-G lesions can result in mutagenesis at GC-rich contexts, with a clustering of such mutations possibly leading to DNA breaks, which can be substrates for chromosomal rearrangements. If a DNA break is induced by clustered DNA damage and/or requires end processing, the two junctions may not be identical and can be separated by some distance (approximately $10^1$-$10^3$ bp). This contrasts with a balanced translocation event that results in two fusion products, such that the secondary one represents a reciprocal nononcogenic fusion event (Fig 3E). Consistent with an excess of 8-oxo-G leading to DNA breaks induced by clustered DNA damage, we found that MSI tumors had more separated junction points relative to MSS tumors




C

| Microbial Order | Correlation |
|---|---|
| p.Firmicutes.c.Erysipelotrichia.o.Erysipelotrichales | 0.494 |
| p.Cyanobacteria.o.Nostocales.f.Nostocaceae | 0.474 |
| p.Proteobacteria.c.Betaproteobacteria.o.Rhodocyclales (5) | 0.469 |
| p.Proteobacteria.c.Deltaproteobacteria.o.Desulfobacterales | 0.468 |
| p.Proteobacteria.c.Betaproteobacteria.o.Burkholderiales (21) | 0.450 |
| p.Proteobacteria.c.Betaproteobacteria.o.Neisseriales (2) | 0.445 |
| p.Actinobacteria.c.Actinobacteria.o.Micromonosporales (2) | 0.432 |
| k.Viruses.o.Tymovirales | 0.419 |
| p.Actinobacteria.c.Actinobacteria.o.Corynebacteriales | 0.414 |
| p.Proteobacteria.c.Betaproteobacteria.o.Methylophilales | 0.412 |
| p.Actinobacteria.c.Acidimicrobiia.o.Acidimicrobiales | 0.366 |
| p.Proteobacteria.c.Gammaproteobacteria.o.Xanthomonadales | 0.355 |
| p.Firmicutes.c.Clostridia.o.Clostridiales | 0.345 |
| p.Proteobacteria.c.Gammaproteobacteria.o.Xanthomonadales | 0.334 |
| p.Proteobacteria.c.Alphaproteobacteria.o.Rhodobacterales | 0.322 |
| p.Proteobacteria.c.Gammaproteobacteria.o.Cellvibrionales | 0.318 |
| p.Proteobacteria.c.Gammaproteobacteria.o.Methylococcales | 0.315 |

**FIG 4.** Characteristics of the junction regions of oncogenic fusions suggest association with impaired repair of 8-oxo-G in defect in MMR colon adenocarcinoma. (A) Schematic representation of a model showing the role of MMR in repair of 8-oxo-G bases on the template strand. Accordingly, MMR-mediated repair of multiple closely spaced oxoguanidine during replication can potentially result in DNA break. Microbial role in context-dependent 8-oxo-G formation also offers one possible explanation why the pattern of oncogenic fusions is unique to CRCs. (B) Heatmap showing correlation of abundance of CRC-associated microbes with the burden of mutational signatures detected in the tumors. MMR signature SBS15 is highlighted. (C) Microbial entities (phyla, clade, order) highly correlated with SBS15 burden are listed. If multiple families within an order show significant correlation, the number of such families is indicated in parenthesis. 8-oxo-G, 8-oxoguanine; CRC, colorectal carcinoma; G, guanine; MMR, mismatch repair; N, any nucleotide; T, thymine.

(MSI: 51.1%, MSS: 29.8%; Fisher's exact test; $P < 1.1E-03$, Fig 3F), corresponding to the fusion and its reciprocal nononcogenic fusion products.

### Colon-Specific DNA Damage and KF Preference

The fusion events associated with MSI CRC were much less frequent in other MSI non-CRCs, raising a possibility that mutagenic processes underlying fusions are tissue context specific.[13,14] Gut microbiota–related sustained inflammation, in part through production of butyrate and other metabolic products, triggers 8-oxo-G DNA lesions in colon epithelium, which is repaired via BER and MMR (Fig 4A).[11,12]

Our cohort did not have microbiome data, but we jointly analyzed microbiome data[6] and estimated proportional weights of the mutational signatures for TCGA CRC samples[15] and found that several microbial taxa, especially certain orders of firmicutes, beta-proteobacteria, cyanobacteria, are enriched in the tumors with dMMR signatures (Figs 4B and 4C; Spearman correlation $> 0.3$; $P < 5.0E-02$). Firmicutes produce butyrate, which has been shown to induce ROS and 8-oxo-G lesions in MMR-mutant mouse models.[12]

## DISCUSSION

We confirmed prior observations that MSI CRCs are enriched for fusions, specifically those involving *NTRK1* and *NTRK3*, as well as *RET*, *ALK*, and *BRAF*. Interestingly, this enrichment of fusions in MSI cancer seems to be limited to CRC, suggesting that the enrichment of specific fusions in MSI CRC may be both due to the MSI phenotype and a tissue of origin specific effect.

Our results suggest that lesions in GC-rich context in dMMR CRCs may contribute to mutagenesis and DNA double-strand breaks and subsequent fusion events. 8-oxo-G is a major source of mutagenesis in GC-rich context in CRCs, which require BER and MMR for repair.[11] The GC enrichment was predominantly in the oncogene while the diverse set of fusion partner genes suggests that staggered breaks in the GC-rich intron of the oncogene might be the initiating event of the rearrangement.

We propose a model (Fig 4A) in which environmental-induced 8-oxo-G lesions are an important source of genomic mutations and preference for specific oncogenic fusions in the intestinal cell of origin of CRC. In setting of dMMR, clustered 8-oxo-G lesions may accumulate in GC-rich regions, such as key introns of *NTRK1*, and lead to staggered DNA double-strand breaks that increase the chance of generating genomic rearrangements and oncogenic fusions. Microbial-derived butyrate can induce 8-oxo-G lesions in intestinal epithelium, suggesting that the composition of the gut microbiome may play an important role in the pathogenesis of dMMR CRCs.[12] Consistent with this hypothesis, there is evidence that 8-oxo-G mutations occur more frequently in intestinal epithelium due to microbiome-derived reactive oxygen species and that MMR may play an important role in proper repair of these lesions.[11,12] It is possible that dMMR may impede end processing of the DNA breaks engendered by clustered 8-oxo-G lesions, further contributing to formation of genomic rearrangements. Fusion-generating rearrangements remain rare but are strongly selected for during tumorigenesis in MSI CRC.

MSI CRCs associated with epigenetic silencing of *MLH1* are more likely to harbor oncogenic fusions, which suggests a role of *MLH1* in repair of 8-oxo-G lesions and/or preventing rearrangements associated with clustered 8-oxo-G lesions. However, since *MLH1* methylation data were not available for this study, further studies are needed.

The presence of oncogenic fusions in a subset of MSI cancers has important therapeutic implications. Although immune checkpoint blockade has great activity in MSI cancers, only approximately 50% have initial response highlighting a need for additional therapies in MSI cancers.[16] TRK inhibitors have shown efficacy in CRC, but the MSI status for the tumors was not reported.[17,18] Anecdotal evidence suggests that patients with MSI CRC harboring KFs may respond to targeted agents, and combinations of targeted therapy and immune checkpoint therapy are showing impressive activity in several cancer types.[19,20] It is possible that combined therapy targeting oncogenic fusion genes and immune checkpoint blockade may lead to improved long-term outcome in KF-positive MSI CRC.


## AFFILIATIONS

[1]Foundation Medicine, Cambridge, MA
[2]Rutgers Cancer Institute of New Jersey and Rutgers Robert Wood Johnson Medical School, Rutgers University, New Brunswick, NJ

## CORRESPONDING AUTHOR

Subhajyoti De, PhD, Rutgers Cancer Institute of New Jersey, Rutgers, the State University of New Jersey, 195 Little Albany Street, New Brunswick, NJ 08901; e-mail: sd948@cinj.rutgers.edu.

## EQUAL CONTRIBUTION

R.W.M. and X.H. contributed equally to this work. S.G. and S.D. are cosenior authors to this work.



## PRIOR PRESENTATION

Presented as abstract #457PD, ESMO 2018 Congress, Munich, Germany, October 19-23, 2018.


## AUTHOR CONTRIBUTIONS

**Conception and design:** Garrett M. Frampton, Siraj M. Ali, Shridar Ganesan, Subhajyoti De
**Provision of study materials or patients:** Richard S.P. Huang
**Collection and assembly of data:** Russell W. Madison, Vivek Ramanan, Shridar Ganesan
**Data analysis and interpretation:** Russell W. Madison, Xiaoju Hu, Zhuxuan Xu, Richard S.P. Huang, Ethan S. Sokol, Garrett M. Frampton, Alexa B. Schrock, Siraj M. Ali, Shridar Ganesan, Subhajyoti De
**Manuscript writing:** All authors

**Final approval of manuscript:** All authors
**Accountable for all aspects of the work:** All authors

## AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/po/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

**Russell W. Madison**
**Employment:** Foundation Medicine
**Stock and Other Ownership Interests:** Roche

**Vivek Ramanan**
**Employment:** Foundation Medicine
**Stock and Other Ownership Interests:** Foundation Medicine

**Richard S.P. Huang**
**Employment:** Roche/Foundation Medicine
**Stock and Other Ownership Interests:** Roche
**Patents, Royalties, Other Intellectual Property:** Patent on IHC, provisional patent on biomarkers and biomarker methodology

**Ethan S. Sokol**
**Employment:** Foundation Medicine
**Stock and Other Ownership Interests:** Roche
**Patents, Royalties, Other Intellectual Property:** Submitted patent for HRD calling methodology (Inst)

**Garrett M. Frampton**
**Employment:** Foundation Medicine
**Stock and Other Ownership Interests:** Roche

**Alexa B. Schrock**
**Employment:** Foundation Medicine
**Stock and Other Ownership Interests:** Foundation Medicine, Roche

**Siraj M. Ali**
This author is a member of the *JCO Precision Oncology* Editorial Board. Journal policy recused the author from having any role in the peer review of this manuscript.
**Employment:** EQRX
**Leadership:** Incysus, Elevation Oncology, Pillar Biosciences, Droplet Biosciences
**Stock and Other Ownership Interests:** Exelixis, Merus NV, Pfizer
**Consulting or Advisory Role:** Azitra, Princeptx, Archer
**Patents, Royalties, Other Intellectual Property:** patents via FOundation Medicine, Patents via Seres Health on microbiome stuff in non neoplastic disease

**Shridar Ganesan**
This author is an Associate Editor for *JCO Precision Oncology*. Journal policy recused the author from having any role in the peer review of this manuscript.
**Employment:** Merck
**Stock and Other Ownership Interests:** Ibris, Inspirata, Merck, Silagene
**Consulting or Advisory Role:** Inspirata, Novartis, Roche, Foghorn Therapeutics, Foundation Medicine, Merck Sharp & Dohme, Silagene, EQRx, EMD Serono, KayoThera
**Research Funding:** M2Gen (Inst)
**Patents, Royalties, Other Intellectual Property:** I hold two patents for digital imaging that may be lisenced to Ibris Inc and Inspirata Inc
**Other Relationship:** NIH/NCI

**Subhajyoti De**
**Patents, Royalties, Other Intellectual Property:** S.D. has provisional US patent (US Patent Application No. 63/177,696 and 63/177,808) that are unrelated to this publication

No other potential conflicts of interest were reported.

## REFERENCES


1. Alexandrov LB, Kim J, Haradhvala NJ, et al: The repertoire of mutational signatures in human cancer. Nature 578:94-101, 2020
2. Schrock AB, Ouyang C, Sandhu J, et al: Tumor mutational burden is predictive of response to immune checkpoint inhibitors in MSI-high metastatic colorectal cancer. Ann Oncol 30:1096-1103, 2019
3. Cocco E, Benhamida J, Middha S, et al: Colorectal carcinomas containing hypermethylated MLH1 promoter and wild-type BRAF/KRAS are enriched for targetable kinase fusions. Cancer Res 79:1047-1053, 2019
4. Yang L, Wang S, Lee JJK, et al: An enhanced genetic model of colorectal cancer progression history. Genome Biol 20:168, 2019
5. Gao Q, Liang WW, Foltz SM, et al: Driver fusions and their implications in the development and treatment of human cancers. Cell Rep 23:227-238.e3, 2018
6. Poore GD, Kopylova E, Zhu Q, et al: Microbiome analyses of blood and tissues suggest cancer diagnostic approach. Nature 579:567-574, 2020
7. Caballero J, Smit AFA, Hood L, et al: Realistic artificial DNA sequences as negative controls for computational genomics. Nucleic Acids Res 42:e99, 2014
8. Rosenthal R, McGranahan N, Herrero J, et al: DeconstructSigs: Delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol 17:31, 2016
9. Singh VK, Rastogi A, Hu X, et al: Mutational signature SBS8 predominantly arises due to late replication errors in cancer. Commun Biol 3:421, 2020
10. Morganella S, Alexandrov LB, Glodzik D, et al: The topography of mutational processes in breast cancer genomes. Nat Commun 7:11383, 2016
11. Colussi C, Parlanti E, Degan P, et al: The mammalian mismatch repair pathway removes DNA 8-oxodGMP incorporated from the oxidized dNTP pool. Curr Biol 12:912-918, 2002
12. Irrazabal T, Thakur BK, Kang M, et al: Limiting oxidative DNA damage reduces microbe-induced colitis-associated colorectal cancer. Nat Commun 11:1802, 2020
13. Hu X, Xu Z, De S: Characteristics of mutational signatures of unknown etiology. NAR Cancer 2:zcaa026, 2020
14. Coleman N, De S: Mutation signatures depend on epigenomic contexts. Trends Cancer 4:659-661, 2018
15. Cancer Genome Atlas Network: Comprehensive molecular characterization of human colon and rectal cancer. Nature 487:330-337, 2012
16. André T, Shiu K-K, Kim TW, et al: Pembrolizumab in microsatellite-instability-high advanced colorectal cancer. N Engl J Med 383:2207-2218, 2020
17. Drilon A, Laetsch TW, Kummar S, et al: Efficacy of larotrectinib in TRK fusion-positive cancers in adults and children. N Engl J Med 378:731-739, 2018

18. Doebele RC, Drilon A, Paz-Ares L, et al: Entrectinib in patients with advanced or metastatic NTRK fusion-positive solid tumours: Integrated analysis of three phase 1-2 trials. Lancet Oncol 21:271-282, 2020
19. Yakirevich E, Resnick MB, Mangray S, et al: Oncogenic ALK fusion in rare and aggressive subtype of colorectal adenocarcinoma as a potential therapeutic target. Clin Cancer Res 22:3831-3840, 2016
20. Rini BI, Plimack ER, Stus V, et al: Pembrolizumab plus axitinib versus sunitinib for advanced renal-cell carcinoma. N Engl J Med 380:1116-1127, 2019
21. Frampton GM, Fichtenholtz A, Otto GA, et al: Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Nat Biotechnol 31:1023-1031, 2013
22. Single Nucleotide Polymorphism Database (dbSNP). http://www.ncbi.nlm.nih.gov/SNP
23. ExAC Database. http://exac.broadinstitute.org
24. Chalmers ZR, Connelly CF, Fabrizio D, et al: Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome Med 9:34, 2017
25. Sun JX, He Y, Sanford E, et al: A computational approach to distinguish somatic vs. germline origin of genomic alterations from deep sequencing of cancer specimens without a matched normal. PLoS Comput Biol 14:e1005965, 2018
26. Trabucco SE, Gowen K, Maund SL, et al: A novel next-generation sequencing approach to detecting microsatellite instability and pan-tumor characterization of 1000 microsatellite instability–high cases in 67,000 patient samples. J Mol Diagn 21:1053-1066, 2019

# APPENDIX

## APPENDIX 1. SUPPLEMENTARY METHODS

### Sequencing and Fusion Calling

Clinical genomic profiling was performed in a Clinical Laboratory Improvement Amendments–certified, New York State and College of American Pathologists–accredited laboratory (Foundation Medicine Inc, Cambridge, MA). Approval for this study, including a waiver of informed consent and a Health Insurance Portability and Accountability Act waiver of authorization, was obtained from the Western Institutional Review Board (Protocol No. 20152817). In brief, ≥ 50 ng of DNA was extracted from 40 microns of formalin-fixed, paraffin-embedded tissue blocks from 33,360 colorectal carcinoma specimens and 238,891 specimens from other solid tumors. The samples were assayed by clinical genomic profiling using adapter ligation, and hybrid capture was performed for all coding exons of 310 or 385 cancer-related genes plus select introns from 31 or 34 genes frequently rearranged in cancer. Sequencing of captured libraries was performed to a mean exon coverage depth of > 500×, and resultant sequences were analyzed for genomic alterations including short variants (base substitutions, insertions, and deletions), copy number alterations (focal amplifications and homozygous deletions), and select gene fusions or rearrangements, as previously described.[21]

Tumor mutational burden was determined on 1.14 or 0.83 Mb of sequenced DNA using a mutation burden estimation algorithm that, on the basis of the genomic alterations detected, extrapolates to the exome or the genome as a whole. For purposes of mutation burden estimation, all base substitutions and indels, including synonymous alterations, are counted. Subtracted from this number are functionally oncogenic or germline alterations, as defined below. Germline alterations are those listed in the dbSNP database,[22] those with two or more counts in the ExAC database,[23] or those predicted by a somatic germline zygosity algorithm to be germline in the specimen being assessed. Tumor mutational burden is reported as mutations per megabase (mut/Mb).[24,25]

Microsatellite instability (MSI) was measured by evaluating the changes to 114 loci selected from a total set of 1,897 that have adequate coverage. In a large training set of data from clinical specimens, we then used principal components analysis to project the 228-dimension data onto a single-dimension (the first principal component) that maximizes the data separation, producing an next generation sequencing-based MSI score.[26]

Allelic status for mutations was computationally predicted without matched normal tissue; in validation testing of 480 tumor-only predictions against matched normal specimens, accuracy was 95% for somatic and 99% for germline predictions. For specimens with sufficient tumor purity, no single nucleotide polymorphism contamination, and copy number modeling not skewed by GC bias, mismatch repair (MMR) alteration allelic status was categorized as biallelic, monoallelic, or unknown. Biallelic/monoallelic predictions were attempted for all MMR short variants; for large (full gene or multi-exon) deletions in MMR genes, only homozygous events were included; and for genomic rearrangements, neither allelic status was not predicted.[25]

**TABLE A1.** Genomic and Demographic Characteristics of Full-CRC Cohort

| Characteristic | All CRC (N = 32,218) | Colon Adeno (n = 26,945) | Rectum Adeno (n = 5,273) |
|---|---|---|---|
| Female, No. (%) | 14,485 (45) | 12,521 (46) | 1,964 (37) |
| Age, median (IQR) | 59 (50-68) | 60 (50-68) | 57 (49-66) |
| TMB, median (IQR) | 3.75 (2.50-6.09) | 3.75 (2.50-6.09) | 3.47 (1.74-5.00) |
| MSI, No. (%) | 1,501 (4.7) | 1,432 (5.3) | 69 (1.3) |
| RAS-mutant, No. (%) | 17,522 (54) | 14,630 (54) | 2,892 (55) |

Abbreviations: CRC, colorectal carcinoma; IQR, interquartile range; MSI, microsatellite instability; TMB, tumor mutational burden.

**TABLE A2.** Specific Oncogenes Involved in Kinases Fusions Within Colorectal Carcinoma

| Kinase | MSI Status | Count, No. (%) | Partner Genes (No.) |
|---|---|---|---|
| ALK | MSI | 7 (0.47) | STRN (3), EML4 (2), DIAPH2 (1), SLMAP (1) |
| | MSS | 26 (0.09) | EML4 (8), STRN (8), CAD (3), ATIC (1), DIAPH2 (1), GPHN (1), MAPRE3 (1) |
| BRAF | MSI | 16 (1.07) | TRIM24 (6), AGAP3 (3), CUL1 (3), AKAP9 (2), GKAP1 (1), TARDBP (1) |
| | MSS | 52 (0.17) | TRIM24 (19), AGAP3 (7), SND1 (3), CUX1 (2), MKRN1 (2), AGAP1 (1), AKAP9 (1), ANO10 (1), ARGLU1 (1), ARMC10 (1), CSNK1A1 (1), CUL1 (1), DAAM1 (1), DENND2A (1), DLG1 (1), GTF2IRD1 (1), JHDM1D (1), LMTK2 (1), MAP3K9 (1), NFIA (1), RHEB (1), TAX1BP1 (1), TMEM44 (1), TRAK1 (1) |
| | MSI ambiguous | 1 (0.63) | AGAP3 (1) |
| EGFR | MSS | 15 (0.05) | SEPT14 (13), COBL (1), FLJ45974 (1) |
| FGFR1 | MSS | 7 (0.02) | TACC1 (3), BAG4 (1), DDHD2 (1), HOOK3 (1), PLXDC1 (1) |
| FGFR2 | MSI | 4 (0.27) | POF1B (2), CIT (1), ZNF608 (1) |
| | MSS | 7 (0.02) | BICC1 (2), DDX21 (1), ERC1 (1), NRBF2 (1), PAWR (1), TACC2 (1) |
| FGFR3 | MSS | 7 (0.02) | TACC3 (7) |
| MET | MSS | 1 (0.003) | DIAPH2 (1) |
| NTRK1 | MSI | 43 (2.86) | TPM3 (22), LMNA (12), TPR (4), PLEKHA6 (2), CCDC88C (1), IRF2BP2 (1), TPM1 (1) |
| | MSS | 7 (0.02) | TPM3 (3), LMNA (2), PLEKHA6 (1), TPR (1) |
| NTRK2 | MSS | 1 (0.003) | ACO1 (1) |
| NTRK3 | MSI | 7 (0.47) | ETV6 (7) |
| RAF1 | MSS | 6 (0.02) | ATG7 (2), ANO10 (1), CAPN7 (1), DENND5A (1), PDZRN3 (1) |
| RET | MSI | 21 (1.40) | NCOA4 (12), CCDC6 (4), TRIM24 (3), SNRNP70 (1), TNIP2 (1) |
| | MSS | 25 (0.08) | NCOA4 (19), CCDC6 (5), TNIP1 (1) |
| | MSI ambiguous | 1 (0.63) | NCOA4 (1) |
| ROS1 | MSS | 13 (100) | GOPC (9), TTC28 (2), MCM9 (1), SRPK1 (1) |

Abbreviations: MSI, microsatellite instability; MSS, microsatellite stable.

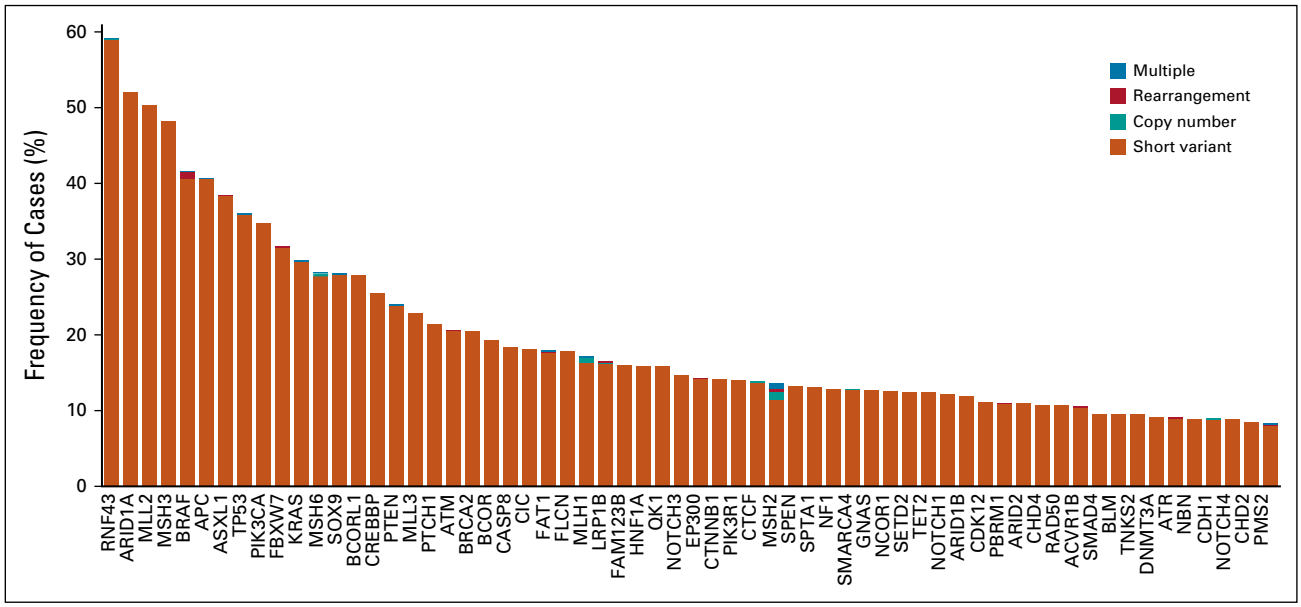



**FIG A1.** Genomic landscape of microsatellite-instable colorectal carcinoma..

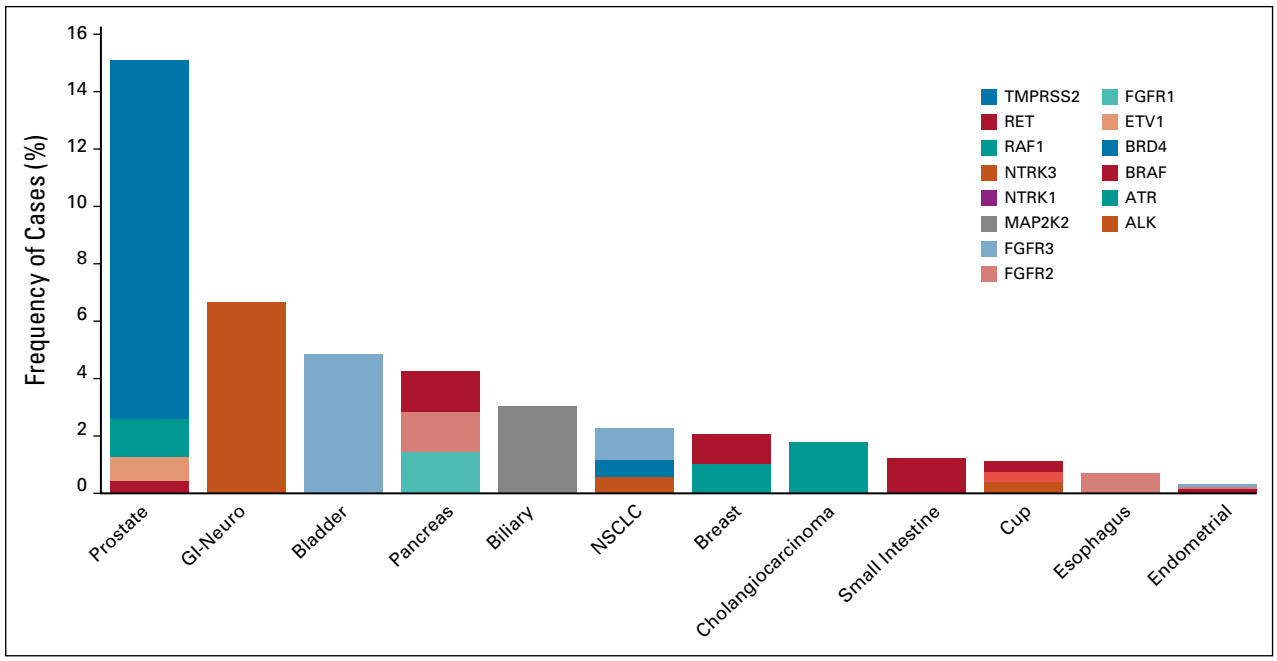



**FIG A2.** Fusions in MSI tumors other than CRC are rare, suggesting that co-occurrence of kinase fusions and MSI is distinct to CRC. CRC, colorectal carcinoma; MSI, microsatellite instability; NSCLC, non–small-cell lung cancer.

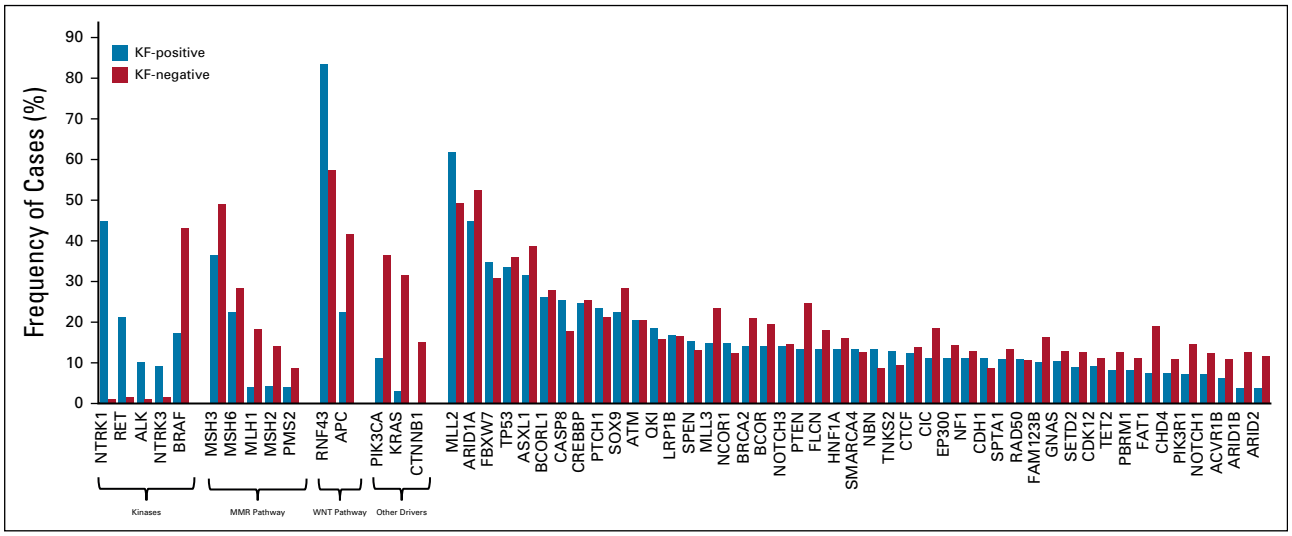



**FIG A3.** Comparing genomic landscapes of fusion-positive and fusion-negative microsatellite instability colorectal carcinoma supports oncogenic nature of these fusions. KF, kinase fusion; MMR, mismatch repair; WNT, wingless-related integration site.