Received: 3 January 2018 | Revised: 20 March 2018 | Accepted: 21 March 2018

DOI: 10.1002/gcc.22539

RESEARCH ARTICLE

WILEY

Mutation analysis of adenomas and carcinomas of the colon: Early and late drivers

Roger K. Wolff[1] | Michael D. Hoffman[1] | Erica C. Wolff[1] | Jennifer S. Herrick[1] | Lori C. Sakoda[2] | Wade S. Samowitz[3] | Martha L. Slattery[1]

[1]Department of Medicine, University of Utah, 383 Colorow, Salt Lake City, Utah, 84108

[2]Division of Research, Kaiser Permanente Northern California, Oakland, California

[3]Department of Pathology, University of Utah, Salt Lake City, Utah

Correspondence
Dr. Roger Wolff, Department of Internal Medicine, University of Utah School of Medicine, 383 Colorow Rd. RM 342, Salt Lake City, UT 84108.
Email: roger.wolff@hsc.utah.edu

Funding information
National Cancer Institute, Grant Number: P30CA042014; National Cancer Institute, Grant Number: CA163683

Abstract

Colorectal cancer (CRC) accounts for about 8% of all new cancer cases diagnosed in the US. We used whole exome sequence data from triplet samples (colon carcinoma, colon adenoma, and normal tissue) from 18 individuals to assess gene mutation rates. Of the 2 204 genes that were mutated, *APC*, *TTN*, *TP53*, *KRAS*, *OBSCN*, *SOX9*, *PCDH17*, *SIGLEC10*, *MYH6*, and *BRD9* were consistent with genes being an early driver of carcinogenesis, in that they were mutated in multiple adenomas and multiple carcinomas. Fifty-two genes were mutated in ≥12.5% of microsatellite stable (MSS) carcinomas but not in any of the adenomas, in line with the profile of a late driver event involved in tumor progression. Thirty-eight genes were sequenced in a larger independent set of 148 carcinoma/normal tissue pairs to obtain more precise mutation frequencies. Eight of the genes, *APC*, *TP53*, *ATM*, *CSMD3*, *LRP1B*, *RYR2*, *BIRC6*, and *MUC17*, contained mutations in >20% of the carcinomas. Interestingly, mutations in four genes in addition to *APC* that are associated with dysregulation of Wnt signaling, were all classified as early driver events. Most of the genes that are commonly associated with colon cancer, including *APC*, *TP53*, and *KRAS*, were all classified as being early driver genes being mutated in both adenomas and carcinomas. Classifying genes as potential early and late driver events points to candidate genes that may help dissect pathways involved in both tumor initiation and progression.

KEYWORDS
adenoma, colon cancer, mutation, whole exome sequencing

1 | INTRODUCTION

Colorectal cancer (CRC) is one of the most common cancers, accounting for about 8% of all new cancer cases diagnosed[1] in the US. The rate of new cases has been declining steadily over the past two decades[1] possibly due to increased screening and early detection and removal of precancerous lesions, adenomas or polyps.[2–5] Early mutations in colon adenomas can initiate tumor development.[6] Cancers usually develop from the accumulation of genetic alterations that allow for tumor progression. Mutations at both the initiation and progression stages of the carcinoma can be considered as driver events by providing a growth advantage to the cells that harbor them. However, other acquired mutations may be passengers, in that they occur coincidentally with the drivers that provide the growth advantage.

Some of the most common genomic changes in CRC include mutations in the *APC*, *TP53*, and *KRAS* genes, DNA hypermethylation leading to the CIMP (CpG Island Methylator Phenotype), and mismatch repair deficiency that leads to genetic hypermutability that can be detected by Microsatellite Instability (MSI). NextGen sequencing (NGS) approaches have allowed higher resolution analysis of tumor samples in the search for additional genes important in the carcinogenic process.

In this study, we focused on samples from 18 individuals with available germline cells (normal colonic mucosa or blood), a colonic adenoma, and a colon carcinoma available for whole exome sequencing

TABLE 1 Demographics of participants providing tissue samples

		Samples			
		Whole exome sequencing ($N = 18$)[a]		Targeted exome sequencing ($N = 148$)	
		Mean	SD	Mean	SD
Age		68.5	9.8	65.7	9.1
		N	%	N	%
Gender	Male	10	55.56	92	62.16
	Female	8	44.44	56	37.84
Race/ethnicity	Missing	3	16.67	9	6.08
	White, non Hispanic	13	72.22	119	80.41
	White, Hispanic	2	11.11	8	5.41
	Black	–	–	10	6.76
	Other	–	–	2	1.35
Tumor Location	Distal	8	44.44	75	50.68
	Proximal	10	55.56	72	48.65
	Unknown	–	–	1	0.68

[a]17 of the 18 whole exome sequencing samples also had adenoma tissue sequenced.

(WES). Sequencing from the normal samples allowed us to filter out germline genetic variation present in the individual enabling easier identification of somatically-derived mutations in the adenomas and carcinomas. Genetic mutations enriched in both the adenomas and carcinomas likely represent early driver events. Mutations present in only the carcinomas may indicate later driver mutations involved in tumor progression. Genes mutated in adenomas but not the carcinomas could represent either random mutation events not important to cancer initiation, or rare events not seen in the carcinomas due to the relatively small number of samples evaluated. We analyzed each gene for the number and type of mutations present in the adenomas and carcinomas and how they relate specifically to microsatellite stable (MSS) and MSI tumors. Thirty-eight genes were further evaluated in 148 individuals to more precisely estimate their frequencies in the population. This information allowed us to classify genes into likely early and late driver events and to assess differences among the different tumor phenotypes.

2 | MATERIAL AND METHODS

2.1 | Study samples

Study samples come from a subset of participants in a larger epidemiologic study of colon cancer from Kaiser Permanente of Northern California (KPNC) and an eight-county area in Utah.[7] All participants analyzed in the study had previous data for *TP53* mutations,[8] *KRAS* mutations,[9] MSI,[10] and CIMP status.[11] Among the 18 individuals analyzed by WES, five of the carcinoma samples were uniquely categorized as *TP53* phenotype, five were uniquely *KRAS* phenotype, two had both the MSI and CIMP-high phenotypes, and two had only the CIMP-high phenotype. Four of the carcinomas were assigned a tumor molecular phenotype of 'None' previously detected as they did not contain any *TP53* or *KRAS* mutations in the genes analyzed, nor were they classified as MSI high or CIMP+. All non-MSI carcinomas were considered MSS and analyzed accordingly. The 18 participants were chosen to represent a spectrum of tumor phenotypes. The 148 additional individuals analyzed by targeted exome capture were not selected based on previous tumor phenotype data. The summary of demographic information on participants who provided samples is provided in Table 1. The University of Utah and KPNC Institutional Review Boards approved this study.

2.2 | DNA isolation

DNA was extracted from formalin-fixed paraffin embedded (FFPE) tissues for the carcinoma and adenoma samples. Slides were reviewed by the study pathologist to delineate carcinoma, normal mucosa, and adenoma tissue. Tissue for normal mucosa and adenoma were selected from the same subsite as the colon carcinoma; 17 of the 18 adenomas were considered independent of the carcinoma while one adenoma contained contiguous carcinoma material. This adenoma was dropped from analysis. Pathological characteristics of the adenomas is provided in Supporting Information Table 1.

Cells were dissected from three to five sequential sections on aniline blue stained slides using an H&E slide for reference. Genomic DNA was extracted, isolated, and purified using the RecoverAll Total Nucleic Acid isolation kit (Ambion). Normal genomic DNA was isolated from either FFPE normal colon mucosa (10 of the samples) or from blood (eight of the samples). For the 148 additional paired samples, the normal DNA was derived from blood for 119 individuals and from FFPE tissue of normal colon mucosa for 29 individuals.

2.3 | Library construction and exome capture

Library construction was performed on the 54 DNA samples (18 sets of carcinoma, adenoma, normal triplets) using Agilent Technologies SureSelect XT Reagent Kit, HSQ (cat # G9611A) as described below. Briefly, 200 ng genomic DNA was sheared in a 130 ul volume using a Covaris S2 Focused-ultrasonicator. Sheared DNA was converted to blunt-ended fragments and an A-base was added to prepare the fragments for adapter ligation. Adapters containing a T-base overhang were ligated to the A-tailed DNA fragments. Ligated fragments were PCR-amplified (12 cycles) and the amplified library purified using Agencourt AMPure XP beads. The concentration of the amplified library was measured on a Qubit fluorometer and an aliquot of the library was resolved on an Agilent 2200 Tape Station using a D1000 assay to define the size range.

Amplified libraries (750 ng) were hybridized with biotinylated RNA baits from the Agilent Technologies SureSelect (XT) Human All Exon v5 plus UTRs kit. Library molecules hybridized to the SureSelect baits were purified using Dynabeads MyOne Streptavidin T1 (Life Technologies). Index tags were added to the adapters of the captured library by PCR (12 cycles) using Agilent Herculase II DNA polymerase and primers from the SureSelect XT Reagent Kit. The amplified libraries were purified using Agencourt AMPure XP beads. Whole exome libraries were pooled for a multiplex of two in preparation for sequencing.

Targeted exome capture was performed on an additional 148 colon carcinoma-normal tissue pairs using Agilent Technologies SureSelect XT2 Reagent Kit, HSQ (cat # G9621A). We targeted 38 genes that were frequently mutated in carcinomas or adenomas based on our WES data. Genes were chosen as being frequently mutated in either MSS carcinomas (*KRAS* was omitted for space consideration on the platform as all the mutations that were identified were located at the hotspot codons 12 and 13), in both adenomas and carcinomas (potential early event genes), genes only mutated in the original MSS carcinomas (potential late event genes), genes identified as being mutated in both of the original two MSI-high carcinomas, or genes previously only mutated in the adenomas. Targeted exome libraries were pooled to sequence at a multiplex of 14–18 samples.

2.4 | Library sequencing

Sequencing libraries (25 pM) were chemically denatured and applied to an Illumina HiSeq v4 paired end flow cell using an Illumina cBot. Hybridized molecules were clonally amplified and annealed to sequencing primers with reagents from an Illumina HiSeq PE Cluster Kit v4-cBot. Following transfer of the flowcell to an Illumina HiSeq 2500 instrument (HCS v2.2.38 and RTA v1.18.61), a 125 cycle paired-end sequence run was performed using HiSeq SBS Kit v4 sequencing reagents.

2.5 | Variant calling

Sequence reads were aligned using BWA (Burrow-Wheeler Aligner) to GRCh37.[12] The alignments were postprocessed to calibrate base qualities and to realign INDELs using known published SNPs and INDELs as reference using the GATK software (Broad Institute). Duplicate reads were removed with Picard based on identical paired endpoints (Broad Institute). Variants were called using three programs: MuTect (SNPs),[13] SomaticSniper (SNPs),[14] and Strelka (INDELs).[15] Variants were annotated using AnnoVar for functional changes (nonsynonymous mutations) to *RefSeq* gene annotation.[16] For this analysis, mutation detection was restricted to variants identified within coding exons, excluding all noncoding exons and UTRs. All variants seen in $>5\%$ of sequence reads within the normal tissue were treated as genetic polymorphisms and not analyzed as potential mutations in that individuals adenoma and carcinoma. Additionally, we required that variants were present in the carcinoma or adenoma at a frequency of 10% or greater than in the normal tissue. We acknowledge that some mutations may have been excluded from our analysis as their Variant Allele Frequency (VAF) was below the 10% threshold either from intratumoral heterogeneity, contamination of normal cells in the tumor sample, or uneven allelic amplification during the capture procedure. However, we choose to concentrate on mutations as being identified at $>10\%$ frequency as a conservative approach to eliminate apparent mutations resulting from rare in vitro polymerase errors.

Variants present in the normal tissue for an individual were subtracted from those in their individual carcinoma and adenoma to further eliminate germline polymorphisms. The remaining variants in our data were defined as somatic mutations in the carcinoma and adenoma samples. Nonsynonymous mutations were defined as missense mutations, nonsense mutations, frameshifts due to insertions and deletions (INDELs), and insertions and deletions causing nonframeshift coding mutations.

2.6 | Calculations

The sequence coverage was calculated by multiplying the number of aligned reads to the capture target by 125 bp per read and dividing by $7.1x10^7$ bp (71MB capture target) for each sample. The mean was then calculated for all the samples.

For adenomas and carcinomas, nonsynonymous coding mutation rates were calculated by: (total number mutations in the samples subset x 10^6 bases/MB)/(33,018,000 bases (total capture length of coding sequence in bases). Mean mutation rates were calculated for adenomas, MSS carcinomas, and MSI carcinomas.

Calculations of fold differences for mutations types between MSI and MSS samples were calculated as a ratio of the mean number of a specific type of mutation per MSS carcinoma to that for the MSI carcinomas. This calculation was performed for both nonsynonymous and INDEL/frameshift mutations.

2.7 | *APC* mutation confirmation

All PCR amplifications were performed as previously described[17] for *APC* amplicons. Briefly, 12 μL reactions were amplified in 96-well tray format in GeneAmp 9700 PCR machines (Applied Biosystems, Foster City, CA) using AmpliTaq Gold DNA polymerase. In general, the thermocycler protocol used 10 min at 95°C initial denaturation, followed

TABLE 2 Mutation counts for adenomas and carcinomas by individual sample

Sample ID	MSS/MSI Status	Nonsynonymousmutation counts[a]		Mutation rates[b]	
		AD	CA	AD Mut/MB	CA Mut/MB
1	MSS	38	31	1.15	0.94
2	MSS	107	60	3.23	1.81
3	MSS	585	129	17.68	3.90
4	MSS	93	91	2.81	2.75
5	MSS	108	165	3.26	4.99
6	MSS	434	57	13.12	1.72
7	MSS	81	56	2.45	1.69
8	MSS	79	66	2.39	2.00
9	MSS	97	77	2.93	2.33
10	MSS	99	128	2.99	3.87
11	MSS	93	111	2.81	3.36
12	MSS	292	113	8.83	3.42
13	MSS	85	120	2.57	3.63
14	MSI	107	1464	3.23	44.26
15	MSI	131	863	3.96	26.09
16	MSS	60	71	1.81	2.15
17	MSS	91	128	2.75	3.87
18	MSS	–	73	–	2.21

[a]The total number of mutations are presented for adenoma (AD) and Carcinoma (CA).
[b]The mutation rates are calculate as the total number of mutations in a sample divided by the total number of MB coding sequence that was used in the exome capture.

by 30 cycles of 10 second denaturation at 95°C, 10 second annealing at primer specific temperatures, and a 20 second extension at 72°C. Nested PCR conditions utilized 2 μL of a 25-fold dilution of the primary PCR product with the addition of internal primers. All DNA used for subsequent *APC* confirmation by PCR/Sanger sequencing confirmation used the same DNA that was used in the Exome capture sequence that identified the mutations. PCR products were treated with ExoSAP and sequenced using an ABI 3730 capillary sequencer (Applied Biosystems, Foster City, CA). Sequence data were analyzed using Sequencher v.5.3 (Gene Codes Corporation, Ann Arbor, MI).

3 | RESULTS

3.1 | Whole exome sequencing (WES)

WES achieved a mean coverage of 244X with a range of 73–390X. The number of mutations identified in adenomas ranged from 38 to 585; for carcinoma samples the range of mutations was 31 to 1464 (Table 2). This corresponded to a mean adenoma mutation rate of 4.59 mutations/MB, a mean MSS carcinoma mutation rate of 2.79 mutations/MB, and a mean MSI carcinoma mutation rate of 35.17/MB. The large differences in mutation rates between MSI and MSS tumors led to the analysis of tumors in these classifications separately.

3.2 | Early driver events: Analysis of adenomas

In the 17 adenomas analyzed, 2204 genes contained non-synonymous mutations (Supporting Information Table 2). Two hundred-thirty-nine genes were mutated in two or more (≥11.76%) of the adenomas and thirty-two genes were mutated in three or more (≥17.65%) of the adenomas (Table 3). Of the 239 genes mutated in two or more adenomas, 69 were mutated in at least one of the sequenced MSS carcinoma samples. Ten of the 69 genes, *APC*, *TTN*, *TP53*, *KRAS*, *OBSCN*, *SOX9*, *PCDH17*, *SIGLEC10*, *MYH6*, and *BRD9*, were mutated in ≥11.76% of the adenomas and in two or more (≥12.5%) of the MSS carcinomas.

3.3 | Carcinoma analysis by MSS and MSI tumor molecular phenotype

A total of 1 273 genes were mutated among the 16 MSS carcinomas (Supporting Information Table 3). Ninety-four genes were mutated in two or more (≥12.5%) of the carcinomas. *APC*, *TTN*, *OBSCN*, *KRAS*, *TP53*, *PIK3CA*, *MYH6*, *PXDN*, *TMEM132C*, *SOX9*, and *PAPLN* were all mutated in three or more (≥18.75%) of the MSS carcinomas (Table 4). The most highly mutated gene, *APC*, was mutated in 14 of 16 MSS carcinomas (87.5%). All twenty-three *APC* mutations identified in the 14 carcinomas were classified as nonsense mutations or INDEL/frameshift

TABLE 3 Frequently mutated genes in adenomas (AD)

Gene	N^a (%) (N = 17)		AD mutated samples; total mutations	Missense	Nonsense	INDEL/frameshift	CA mutated sample; N^2 (%) (N = 16)	
APC	12	70.59	22	1	13	8	14	87.50
TTN	7	41.18	12	7	0	5	8	50.00
TP53	5	29.41	6	6	0	0	7	43.75
KRAS	5	29.41	5	4	0	1	7	43.75
AMER1	4	23.53	5	0	5	0	1	6.25
HYDIN	4	23.53	5	5	0	0	0	0.00
LRP1B	4	23.53	4	3	1	0	1	6.25
RYR2	4	23.53	4	2	1	1	1	6.25
MGAT4B	4	23.53	4	3	0	1	0	0.00
SLIT3	4	23.53	4	3	0	1	0	0.00
DMXL2	3	17.65	4	3	0	1	0	0.00
ZNF536	3	17.65	4	3	0	1	0	0.00
OBSCN	3	17.65	3	2	0	1	7	43.75
SOX9	3	17.65	3	0	1	2	3	18.75
PCDH17	3	17.65	3	2	0	1	2	12.50
SIGLEC10	3	17.65	3	3	0	0	2	12.50
ARMCX4	3	17.65	3	2	0	1	1	6.25
GOLGB1	3	17.65	3	2	0	1	1	6.25
NAV3	3	17.65	3	3	0	0	1	6.25
PHACTR1	3	17.65	3	3	0	0	1	6.25
ANKLE1	3	17.65	3	2	0	1	0	0.00
CACNA1A	3	17.65	3	1	1	1	0	0.00
CSMD2	3	17.65	3	2	0	1	0	0.00
DOCK4	3	17.65	3	2	0	1	0	0.00
FAM9A	3	17.65	3	2	0	1	0	0.00
HERC2	3	17.65	3	2	0	1	0	0.00
KCNT1	3	17.65	3	3	0	0	0	0.00
KIAA1109	3	17.65	3	2	0	1	0	0.00
MUC3A	3	17.65	3	1	0	2	0	0.00
NACAD	3	17.65	3	0	0	3	0	0.00
RASGRF1	3	17.65	3	3	0	0	0	0.00
ZNF835	3	17.65	3	3	0	0	0	0.00

[a]The number and frequency of mutations in these genes that were seen in adenomas (AD) is listed. The mutation counts are subdivided into mutation types.

The mutation count and mutation frequency seen in MSS carcinomas (CA) for these genes are also listed. The specific mutations and their codon effects are listed in Supporting Information Table 2.

mutations. Other genes exhibiting only nonsense and INDEL mutations were *SOX9* with three frameshift mutations present in three carcinomas, *CTNNB1* with two nonsense mutations, *TGIF1* with two frameshift mutations, and *BCL9L* with one frameshift and one nonsense mutation. Most mutations in other genes were skewed towards missense mutations. Eighty-four of the genes were mutated in two or more of the MSS carcinomas (>12.50%) but mutated in one or less (≤5.88%) of the adenomas; 52 of the mutated genes in MSS samples were not mutated

TABLE 4 Frequently mutated genes in MSS carcinomas

Gene	N[a] (N = 16)	(%)	Total mutation	Missense	Nonsense	INDEL/frameshift
APC	14	87.50	23	0	16	7
TTN	8	50.00	8	7	0	1
OBSCN	7	43.75	8	8	0	0
KRAS	7	43.75	7	7	0	0
TP53	7	43.75	7	4	1	2
PIK3CA	4	25.00	4	4	0	0
MYH6	3	18.75	4	4	0	0
PXDN	3	18.75	4	2	0	2
TMEM132C	3	18.75	4	3	1	0
SOX9	3	18.75	3	0	0	3
PAPLN	3	18.75	3	3	0	0

[a]The number and frequency of mutations in these genes that were seen in MSS carcinomas (CA) is listed.
The specific mutations and their codon effects are listed in Supporting Information Table 3.

in any of the adenomas, suggesting a mutation pattern consistent with their being late drivers in the carcinogenic process.

Two carcinomas, classified as MSI-high,[18] were found to be hypermutated with 1 464 and 863 mutations each (Table 2). The mean number of base substitution mutations was 785.5, 9.7-fold higher than the mean in the MSS carcinomas, and the mean number of INDEL/frameshifts mutations of 378 was 33.6 fold higher than the MSS carcinomas. Of the 2 031 genes that were mutated in the two MSI-high carcinomas (Supporting Information Table 4), 121 of the genes were mutated in both samples. Of these, 22 genes contained the same mutation in the two MSI tumors, and 18 of the 22 genes *(ERRFI1*, *PHF2*, *RGL2*, *RNF43*, *ACVR2A*, *ARFGAP3*, *ARSJ*, *BRD3*, *C7orf49*, *CBX5*, *DDX27*, *EFCAB5*, *GRIN2A*, *MSH3*, *PROM1*, *RBM23*, *RPL22*, and *ZNRF3)* shared the same location of the frameshift mutation at short mononucleotide repeats (Table 5). The two MSI cancers also shared *MAP4K1* (P808A), *NEK4* (S385C), *BRAF* (V600E), and *KRT3* (G579S) missense mutations.

3.4 | Confirmation of NGS data

We attempted to confirm 43 of the mutations that we identified within the *APC* gene in adenomas and carcinomas by performing Sanger sequencing of PCR products. We confirmed 37 (86%) of the mutations found in the WES analysis. Thus, the majority of the mutations identified were validated by the use of an alternative sequencing method.

Five *KRAS* mutations had been previously identified in our carcinoma samples[9] and we validated all five of these mutations using the NGS of our exome captured DNA. We also identified two additional samples with *KRAS* mutations that were previously undetected. Both of these mutations were within the codon 12 and 13 hotspot previously scanned by Sanger methods. Seven samples had *TP53* mutations, with three of the seven mutated samples previously identified from our hotspot exon 5–8 Sanger sequencing screen.[8] Of the four newly identified *TP53* mutations, two were within the hotspot exons previously assessed by Sanger methods and the other two resided in exons 3 and 4 which were not previously scanned. Furthermore, two *TP53* mutations previously identified in exon 5 were not identified in our WES sequence.

3.5 | Validation of mutation frequency

We chose to replicate mutation analysis to better ascertain mutation frequencies by performing exome capture sequencing on a targeted set of genes. Thirty-eight genes were sequenced in an additional 148 carcinoma/normal DNA pairs and evaluated for codon effecting mutations. The 148 carcinomas had previously been characterized for tumor molecular phenotype (*TP53* mutation containing, *KRAS* mutation containing, MSI status, CIMP status, or "None") allowing analysis within specific phenotypes as well as for MSS and MSI.[8–11] Targeted exome capture sequencing achieved a mean depth of coverage of 1214X with a range of 489–2038X. The distribution of frequency of mutant genes in MSS and MSI carcinomas are shown in Table 6 (the frequencies of mutations for tumor phenotypes *TP53*, *KRAS*, CIMP, and for none previously identified are included in Supporting Information Table 5). Specific mutations for each gene are presented in Supporting Information Table 6. All of the genes were found to be mutated in greater than 10% of at least one tumor phenotype in the carcinomas with the exception of *BRD9* and *ANKLE1*. When looking overall, eight of the genes evaluated were mutated in greater than 20% of the 148 carcinomas (*APC*, 67.6%; *TP53*, 46.6%; *ATM*, 20.9%; *CSMD3*, 20.3%; *LRP1B*, 25.0%; *RYR2*, 33.8%; *BIRC6*, 20.3%; *MUC17*, 21.6%).

4 | DISCUSSION

Sequence analysis of all coding genes in colon carcinomas, adenomatous polyps, and paired normal DNAs is helpful in understanding their potential roles in carcinogenesis. Analysis of both adenomas and

TABLE 5 Mutation hotspots in MSI carcinomas

Gene[a]	Base change[b]	Amino acid change[c]
ERRFI1	c.404delA	p.N135fs
PHF2	c.1468_1469delAA, c.1468delA	p.K490fs
RGL2	c.362delG, c.362insG	p.G121fs
RNF43	c.1976delG	p.G659fs
ACVR2A	c.1303insA, c.1303delA	p.K435fs
ARFGAP3	c.896delA	p.N299fs
ARSJ	c.1736delA	p.K579fs
BRAF	c.1799T>A	p.V600E
BRD3	c.71delC, c.71insC	p.P24fs
C7orf49	c.389delG	p.G130fs
CBX5	c.317delA	p.K106fs
DDX27	c.2065delA	p.K689fs
EFCAB5	c.2427dupC	p.R809fs
GRIN2A	c.4021delA	p.S1341fs
KRT3	c.1735G>A	p.G579S
MAP4K1	c.2422C>G	p.P808A
MSH3	c.1141delA	p.K381fs
NEK4	c.1153A>T	p.S385C
PROM1	c.1697delA	p.N566fs
RBM23	c.44delA	p.K15fs
RPL22	c.2106delG	p.K702fs
ZNRF3	c.404delA	p.N135fs

[a]Genes containing mutations at the same location in both MSI Carcinomas.
[b]The mutation base change.
[c]Amino acid consequence are shown.

carcinomas helps discern whether a gene has a potential early or late driver role in carcinogenesis, or whether it is a likely passenger gene.

Among the MSS carcinomas, 93 genes contained mutations in two or more of the samples (≥12%); the most commonly mutated genes were *APC* (87.5%), *TTN* (50.0%), *OBSCN* (43.8%), *KRAS* (43.8%), *TP53* (43.8%), and *PIK3CA* (25.0%). With the exception of *OBSCN*, this finding is consistent with the colon carcinoma sequencing results from The Cancer Genome Atlas Network (TCGA).[19] In addition, they saw less frequently mutated *FBXW7* (11%), *SMAD4* (10%), *TCF7L2* (9%) and *NRAS* (9%). In our sample set we also confirm that these genes were mutated in one of the carcinomas analyzed each (6.25%) (Supporting Information Table 3), with the exception of *NRAS* which we did not see mutated, perhaps due to our smaller sample size. Our data also revealed that *MYH6*, *PXDN*, *SOX9*, *TMEM132C*, *PAPLN* were frequently mutated (18.8% of samples each). All of the genes that we identified as mutated frequently in MSS carcinomas were also mutated in at least one adenoma with the exception of *PAPLN*, consistent with them not only being commonly mutated but also occurring early in carcinogenesis. Interestingly, *APC*, *SOX9*, and *CTNNB1* (β-catenin), all part of the Wnt-signaling pathway, and *BCL9L* and *TGIF1* contained mutations that were exclusively loss of function mutations (nonsense and frameshift mutations) in our 16 MSS carcinoma samples. This appeared to be unique to these five genes with all other genes biased towards missense mutations.

Inactivation of the tumor suppressor gene *APC* is the most common mutation found in colon cancer and leads to an accumulation of the β-catenin transcription factor in the nucleus. Wild type (wt) *SOX9* competes with TCF/LEF for binding to β-catenin and subsequent degradation of β-catenin.[20] Thus, it is logical that a truncated *SOX9* could lead to an accumulation of β-catenin. *CTNNB1* mutations have been described in colon cancer that activate β-catenin and make it insensitive to wt *APC*.[21] Inhibition by wt *APC* is mediated through phosphorylation of serine/threonine residues. Mutation of these residues leads to activation of β-catenin independent of mutant *APC*. We observed three mutations in *CTNNB1*. One was identified in an adenoma where codon 45, a serine, was deleted (S45Del). This deletion of a phosphorylation site was reported by Morin et al.[21] as an activating mutation. The other two mutations were nonsense mutations identified in carcinomas (R95X and Q773X). One of these carcinomas also contained a single *APC* mutation while the other harbored two *APC* mutations. As these two carcinomas also likely contained a wt copy of *CTNNB1* it remains possible that the *APC* mutations could still be driving the canonical Wnt-signaling pathway. While over 40 truncating mutations in *CTNNB1* have been reported in the COSMIC database,[22] it is unlikely that they all work through the transcription factor activity associated with canonical Wnt signaling. Perhaps the truncating mutations contribute to tumor formation through β-catenin's alternate role in adheren junctions by disrupting cellular contact inhibition.[23]

To evaluate whether genes are likely early drivers in the carcinogenesis process of colon carcinomas, we determined which genes were mutated in both adenomas and carcinomas. Many of the known commonly mutated genes in colon cancer were among these early events, including *APC*, *KRAS*, and *TP53*. It is worth noting that, in addition to *APC*, three other members of the Wnt-signaling pathway *SOX9*, *AMER1*, and *PRICKLE2* are among these potential early driver genes.

In addition, we identified 52 genes that were only seen mutated in two or more MSS carcinoma but not in any of the adenomas. These could represent genes that drive events later in the carcinogenic process. It is worth noting that these 52 genes do not include any of the commonly known mutated colon cancer genes. Three of the genes from the set of late candidates were chosen for additional sequencing in the 148 additional samples: *BIRC6*, *CCDC105*, and *RBMXL3*. *BIRC6* and *RBMXL3* both confirmed their frequent mutation rates in MSS carcinomas (18.75% vs. 12.5% in the first cohort and 14.06% vs. 12.5%, respectively). No adenomas were included in the replication set, precluding confirmation of the mutations as strictly late events. Notably, all three of these genes were mutated frequently in the additional replication set of MSI carcinomas [*BIRC6* (30%), *CCDC105* (15%), and *RBMXL3* (25%)]. All of these mutations were unique and none represented apparent slippage events at coding mononucleotide repeats.

TABLE 6 Validation of mutation frequencies by targeted exome capture

	MSI ($N = 128$)		MSI ($N = 20$)		Choice[a]	% in whole exome sequencing samples[b]		
	Mutated samples		Mutated samples					
	N	%	*N*	%		MSS ($N = 16$)	AD ($n = 17$)	MSI ($n = 2$)
APC	92	71.88	8	40.00	MSS	87.50	70.59	0
PIK3CA	19	14.84	6	30.00	MSS	25.00	5.88	0
PXDN	9	7.03	4	20.00	MSS	18.75	5.88	50
SOX9	7	5.47	4	20.00	MSS	18.75	17.65	0
TMEM132C	9	7.03	2	10.00	MSS	18.75	5.88	50
TP53	63	49.22	6	30.00	MSS	43.75	29.41	0
AMER1	13	10.16	3	15.00	Early	6.25	23.53	0
ARID1A	18	14.06	10	50.00	Early	12.50	5.88	100
ARMCX4	19	14.84	3	15.00	Early	6.25	17.56	0
ATM	27	21.09	4	20.00	Early	12.50	5.88	50
BRD9	2	1.56	1	5.00	Early	12.50	11.76	0
CSMD3	24	18.75	6	30.00	Early	6.25	11.76	100
LRP1B	29	22.66	8	40.00	Early	6.25	23.53	0
MAGEC1	27	21.09	1	5.00	Early	12.50	5.88	0
NOTCH2	10	7.81	4	20.00	Early	6.25	11.76	0
PCDH17	7	5.47	4	20.00	Early	12.50	17.65	50
PLEC	14	10.94	10	50.00	Early	12.50	5.88	50
RPS3A	11	8.59	4	20.00	Early	6.25	5.88	0
RYR2	37	28.91	13	65.00	Early	6.25	23.53	100
BIRC6	24	18.75	6	30.00	Late	12.50	0	50
CCDC105	3	2.34	3	15.00	Late	12.50	0	50
RBMXL3	18	14.06	5	25.00	Late	12.50	0	50
CACNA1C	8	6.25	5	25.00	MSI	6.25	5.88	100
CDHR3	9	7.03	3	15.00	MSI	6.25	0	100
EFCAB5	12	9.38	1	5.00	MSI	0	11.76	100
FAM47A	10	7.81	6	30.00	MSI	0	5.88	100
MAP4K1	17	13.28	4	20.00	MSI	0	5.88	100
MSH3	11	8.59	8	40.00	MSI	0	5.88	100
PIEZO2	18	14.06	5	25.00	MSI	6.25	5.88	100
RNF43	11	8.59	9	45.00	MSI	6.25	5.88	100
SCRIB	8	6.25	7	35.00	MSI	0	0	100
ANKLE1	6	4.69	1	5.00	AD only	0	17.65	50
CSPG4	12	9.38	1	5.00	AD only	0	11.76	0
DIAPH1	19	14.84	2	10.00	AD only	0	11.76	50
KCNT1	6	4.69	2	10.00	AD Only	0	17.65	0
MUC17	27	21.09	5	25.00	AD only	0	11.76	0
OR2L3	5	3.91	0	0.00	AD Only	0	11.76	0
ZNF536	8	6.25	2	10.00	AD Only	0	17.65	0

[a]Choice of Gene—MSS: genes chosen for frequently mutated in MSS; Early: mutant in adenomas and carcinomas; Late: only mutant in carcinoma; MSI: mutant in both the MSI carcinomas. Some of these genes can fit into more than one category.
[b]Percent mutated in original WES cohort.

The two previously identified MSI carcinomas were hypermutated with an excess of both missense/nonsense mutations and INDEL/frameshift mutations as compared to MSS tumors studied (9.7 fold and 33.6 fold, respectively). The high mutation frequency is consistent with observations of other groups.[19] One hundred and twenty-one genes were found to be mutated in both of the MSI tumors in this study. Twenty-one of these genes likely harbor a hotspot for mutation as the same location was mutated in both samples. Only three of these genes were highlighted by the TCGA as frequent targets for mutation, *ACVR2A*, *MSH3*, and *BRAF* (V600E). Of the 21 genes with the same location mutated, all but four were frameshift mutations occurring at mononucleotide repeats as others have seen in mismatch repair deficient MSI tumors.[24] The four genes with recurrent missense mutations were *NEK4*, *KRT3*, *BRAF*, and *MAP4K1*. Only the *BRAF* (V600E), *MSH3* (K381fs), and *RNF43* (G659fs) have been previously reported to be a hotspot mutation in many cancers.[25–29]

We employed a targeted exome capture and sequencing strategy to further evaluate 38 genes on an additional 148 colon carcinoma/normal paired samples. Six genes were chosen for their mutation frequencies in MSS carcinomas. All were seen to be mutated in both MSS and MSI carcinomas, consistent with the TCGA data. Both *APC* and *TP53* exhibited lower mutation frequencies in MSI tumors than in MSS carcinomas. In this larger sample set, *PXDN*, *SOX9*, and *TMEM132C* all showed a MSS mutation rate below 10%; however, *PXDN* had an 11.76% mutation rate in *KRAS* mutation containing carcinomas and *SOX9* had a 21.05% mutation rate in carcinomas exhibiting the "none previously identified" phenotype (Supporting Information Table 5). Perhaps *SOX9* plays a more important role in carcinomas lacking the commonly defined tumor molecular phenotypes. Thirteen genes were selected from our results classifying early driver mutations. Only *BRD9* failed to show an appreciable mutation frequency compared to our whole exome results (1.56% vs. 12.5%, respectively). All the other potential early driver genes appear to have appreciable mutation frequencies across both MSS and MSI tumor molecular phenotypes ($\geq$5%). Three of the genes defined as late drivers also were sequenced. Only *CCDC105* failed to show an appreciable mutation rate in the MSS tumors but did have a 15% mutation rate in MSI tumors. Nine genes were chosen as being frequently mutated in MSI tumors. Only *EFCAB5* demonstrated a low mutation rate (5%) in the MSI tumors. Four of the genes analyzed as frequently mutated in MSI carcinomas, contained potential hotspots for mutation. The same shared mutations were found among the 20 MSI carcinomas in the validation set. Eight of nine *RNF43* mutations found in the MSI carcinomas resulted in p.G659fs, seven of eight *MSH3* mutations were p.K381fs, three of five *MAP4K1* mutations were p.P808A, and one of one *EFCAB5* mutation was p.R809fs. We evaluated seven genes that were only mutated in adenomas and not in carcinomas (Table 3 and Supporting Information Table 2). In the validation data set, they were all found to be mutated in both MSS and MSI samples at low ($\leq$10%) frequency with the exception of *DIAPH1* and *MUC17* that were more frequently mutated (14.8% in MSS and 10% in MSI tumors for *DIAPH1* and 21.09% in MSS and 25% in MSI for *MUC17*). These genes may represent additional early event genes in carcinogenesis.

One of the most frequently mutated genes overall in the replication set was *RYR2* which was part of both the early MSS mutated genes and MSI mutated genes. In the replication set of carcinomas it was mutated in 28.91% of MSS carcinomas and 65% of MSI carcinomas. None of the MSI mutations were the result of a mononucleotide repeat INDEL. Overall 50 of the carcinomas were found to contain a total of 76 *RYR2* mutations. Twenty-two of the mutations were nonsense or INDEL/frameshifts, thus likely inactivating. The remaining 54 mutations were categorized as missense, thus it is unknown if they are inactivating. The *RYR2* gene codes for a large calcium channel and inhibition of its function protects cells against apoptosis,[30] thereby promoting growth. Thus, these mutations in *RYR2* may affect the cells from going into an apoptotic state. In a study looking at tolerance of functional mutation across all genes, Petrovski et al. found that the *RYR2* gene was very intolerant to functional variation in healthy individuals (top 0.05% of genes), further supporting the potential significance of somatic mutation within this gene.[31]

The mean number of codon effecting mutations in adenomas was 151.8, exceeding the mean identified in MSS carcinomas at 92.3 (Table 2). This difference might be due to the few adenomas exhibiting a high number of mutations, perhaps they are beginning to exhibit a hypermutable phenotype. While testing for MSI was not performed on the adenomas, the adenoma from sample ID 3 which contained 585 non-synonymous mutations contained both a *POLE* (L1561fs) and a *POLD3* mutation (Q104fs) consistent with a hypermutable phenotype.[32] No known cause for hypermutability was identified in the adenoma from sample ID 6 which contained 434 non-synonymous mutations. Removing the two adenomas with a high mutation count (ID 3 and ID 6) from the means calculation would lower the resulting value to 104.1. This would suggest that adenomas and carcinomas contain a very similar number of mutations. Among the commonly mutated colon cancer genes, *APC*, *KRAS*, and *TP53*, all revealed high mutation rates in adenomas and carcinomas, suggesting they may be necessary for initiation of adenomas, but are unlikely to be the gene that drives transition to become a carcinoma. Further examination of potential late driver genes may shed light on pathways involved in tumor progression.

It is interesting that six genes from the Wnt-signaling pathway were identified as being mutated in our original carcinoma screen (*APC*, *CTNNB1*, *SOX9*, *AMER1*, *PRICKLE2*, and *RNF43*), all except *CTNNB1* can be defined as early driver events based on being mutated in both carcinomas and adenomas. *CTNNB1* could not be categorized as an early event gene as the adenoma mutation was considered an activating mutation and the MSS carcinomas contained inactivating mutations. Disruption within the Wnt-signaling pathway, would appear to be a necessary event early in tumorigenesis but not sufficient to lead to carcinoma formation. Other specific mutations would be required for that transition to occur.

RNF43 a negative regulator of Wnt signaling was mutated in one MSS carcinoma, one adenoma, and two MSI carcinomas in our WES sample set. The adenoma and MSS mutations (D595V and L12fs, respectively) were not mutated at a coding mononucleotide repeat, while the two mutations identified in the MSI samples involved frameshifts in coding mononucleotide repeats (G659fs and R117fs). Further

analysis within our validation set revealed that 8.6% of MSS carcinoma samples had acquired *RNF43* mutations, none of which were coding mononucleotide frameshifts, where 45% of MSI carcinomas were mutated with 10 (8 G659fs and 2 R117fs) out of 13 mutations occurring at the two previously identified coding mononucleotide repeats.

While whole exome capture sequencing performed quite well compared to Sanger sequencing for the *APC* gene (Sanger confirmed 37 of 43 mutations; 86%), and did better at detecting mutations at *KRAS* (detecting all previously detected mutations detected by Sanger methods plus two that were missed by previous Sanger analysis). However, *TP53* did not fare as well as *APC* or *KRAS*. Four new mutations were found by WES, two within the exon 5–8 hotspot and two that were outside of the hotspot region previously assessed by Sanger methods. In addition, two samples with *TP53* mutations previously detected by Sanger were not confirmed by capture and NGS sequencing. Further sequencing of the 148 carcinomas at *TP53* revealed that capture and NGS missed an additional 21 out of 76 previously identified mutations. Upon further review of the data, these mutations were originally identified in the NGS but rejected as being mutations through secondary filtering/processing based on quality scores. This phenomenon seems to have affected *TP53* more than other genes that were reviewed.

Other studies of colon cancer mutation such as the TCGA have evaluated large numbers of samples by WES. A strength of this study is that we were also able to compare mutation rates for genes seen in adenomas and carcinomas from the same individual to better ascertain whether genes were likely early drivers of carcinogenesis versus those that occur later in the carcinogenic process. Although our whole exome study included only 18 carcinomas, most of the genes were found to be mutated frequently in the larger validation set of 148 carcinomas. Our study both confirms the findings of other WES studies on colon carcinomas, but also puts forth additional candidate genes that are frequently mutated in colon carcinogenesis and at various stages of the process. We encourage other to further investigate the role that these genes play in colon cancer.

ACKNOWLEDGMENTS

The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official view of the National Cancer Institute. We acknowledge Sandra Edwards and Lila Mullany for data oversight and study management and Dr. Bette Caan and the staff at Kaiser Permanente Northern California for sample and data collection. Sequencing was performed at the High Throughput Genomics and Bioinformatics Core Facility, University of Utah and partially supported by Award Number P30CA042014 from the National Cancer Institute. This study was supported by National Cancer Institute grant CA163683 to M.S.

ORCID

Roger K. Wolff http://orcid.org/0000-0003-3218-1953
Martha L. Slattery http://orcid.org/0000-0002-1655-6543

REFERENCES

[1] Howlader NNA, Krapcho M, Miller D, eds. *SEER Cancer Statistics Review, 1975–2013*. Available at: http://seer.cancer.gov/csr/1975_2013. Bethesda, MD: National Cancer Institute; 2016.

[2] Slattery ML, Edwards SL, Ma KN, Friedman GD. Colon cancer screening, lifestyle, and risk of colon cancer. *Cancer Causes Control*. 2000;11:555–563.

[3] Kavanagh AM, Giovannucci EL, Fuchs CS, Colditz GA. Screening endoscopy and risk of colorectal cancer in United States men. *Cancer Causes Control*. 1998;9:455–462.

[4] Winawer SJ, Zauber AG, Ho MN, et al. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. *N Engl J Med*. 1993;329:1977–1981.

[5] Muller AD, Sonnenberg A. Prevention of colorectal cancer by flexible endoscopy and polypectomy. A case-control study of 32,702 veterans. *Ann Intern Med*. 1995;123:904–910.

[6] Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell*. 1990;61:759–767.

[7] Slattery ML, Caan BJ, Potter JD, et al. Dietary energy sources and colon cancer risk. *Am J Epidemiol*. 1997;145:199–210.

[8] Samowitz WS, Curtin K, Ma K-N, et al. Prognostic significance of p53 mutations in colon cancer at the population level. *Int J Cancer*. 2002;99:597–602.

[9] Samowitz WS, Curtin K, Schaffer D, Robertson M, Leppert M, Slattery ML. Relationship of Ki-ras mutations in colon cancers to tumor location, stage, and survival: a population-based study. *Cancer Epidemiol Biomarkers Prev*. 2000;9:1193–1197.

[10] Samowitz WS, Curtin K, Ma KN, et al. Microsatellite instability in sporadic colon cancer is associated with an improved prognosis at the population level. *Cancer Epidemiol Biomarkers Prev*. 2001;10:917–923.

[11] Samowitz WS, Albertsen H, Herrick J, et al. Evaluation of a large, population-based sample supports a CpG island methylator phenotype in colon cancer. *Gastroenterology*. 2005;129:837–845.

[12] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–1760.

[13] Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31:213–219.

[14] Larson DE, Harris CC, Chen K, et al. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics*. 2012;28:311–317.

[15] Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*. 2012;28:1811–1817.

[16] Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38:e164.

[17] Samowitz WS, Slattery ML, Sweeney C, Herrick J, Wolff RK, Albertsen H. APC mutations and other genetic and epigenetic changes in colon cancer. *Mol Cancer Res*. 2007;5:165–170.

[18] Samowitz WS, Holden JA, Curtin K, et al. Inverse relationship between microsatellite instability and K-ras and p53 gene alterations in colon cancer. *Am J Pathol*. 2001;158:1517–1524.

[19] Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012;487:330–337.

[20] Akiyama H, Lyons JP, Mori-Akiyama Y, et al. Interactions between Sox9 and beta-catenin control chondrocyte differentiation. *Genes Dev*. 2004;18:1072–1087.

[21] Morin PJ, Sparks AB, Korinek V, et al. Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. *Science*. 1997;275:1787–1790.

[22] Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res*. 2015;43:D805–D811.

[23] Brembeck FH, Rosario M, Birchmeier W. Balancing cell adhesion and Wnt signaling, the key role of beta-catenin. *Curr Opin Genet Dev*. 2006;16:51–59.

[24] Woerner SM, Gebert J, Yuan YP, et al. Systematic identification of genes with coding microsatellites mutated in DNA mismatch repair-deficient cancer cells. *Int J Cancer*. 2001;93:12–19.

[25] Davies H, Bignell GR, Cox C, et al. Mutations of the BRAF gene in human cancer. *Nature*. 2002;417:949–954.

[26] Giannakis M, Hodis E, Jasmine Mu X, et al. RNF43 is frequently mutated in colorectal and endometrial cancers. *Nat Genet*. 2014;46:1264–1266.

[27] Swisher EM, Mutch DG, Herzog TJ, et al. Analysis of MSH3 in endometrial cancers with defective DNA mismatch repair. *J Soc Gynecol Investig*. 1998;5:210–216.

[28] Akiyama Y, Tsubouchi N, Yuasa Y. Frequent somatic mutations of hMSH3 with reference to microsatellite instability in hereditary nonpolyposis colorectal cancers. *Biochem Biophys Res Commun*. 1997;236:248–252.

[29] Malkhosyan S, Rampino N, Yamamoto H, Perucho M. Frameshift mutator mutations. *Nature*. 1996;382:499–500.

[30] Mariot P, Prevarskaya N, Roudbaraki MM, et al. Evidence of functional ryanodine receptor involved in apoptosis of prostate cancer (LNCaP) cells. *Prostate*. 2000;43:205–214.

[31] Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet*. 2013;9:e1003709.

[32] Robles AI, Traverso G, Zhang M, et al. Whole-exome sequencing analyses of inflammatory bowel disease-associated colorectal cancers. *Gastroenterology*. 2016;150:931–943.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Wolff RK, Hoffman MD, Wolff EC, et al. Mutation analysis of adenomas and carcinomas of the colon: Early and late drivers. *Genes Chromosomes Cancer*. 2018;57:366–376. https://doi.org/10.1002/gcc.22539