

Article

Enhancing the Landscape of Colorectal Cancer Using Targeted Deep Sequencing

Chul Seung Lee ¹, In Hye Song ², Ahwon Lee ², Jun Kang ², Yoon Suk Lee ¹, In Kyu Lee ¹, Young Soo Song ^{3,*} and Sung Hak Lee ^{2,*}

¹ Department of Surgery, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, 222 Banpo-daero, Seocho-gu, Seoul 06591, Republic of Korea; lcs0610@catholic.ac.kr (C.S.L); yslee@catholic.ac.kr (Y.S.L); cmcslee@catholic.ac.kr (I.K.L)
² Department of Hospital Pathology, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, 222 Banpo-daero, Seocho-gu, Seoul 06591, Republic of Korea; ihsongih@catholic.ac.kr (I.H.S); kleec@catholic.ac.kr (A.L); jkang@catholic.ac.kr (J.K)
³ Department of Pathology, College of Medicine, Konyang University, 158 Gwanjeodong-ro, Seo-gu, Daejeon 35365, Republic of Korea
* Correspondence: lifen@konyang.ac.kr (Y.S.S); hakjang@catholic.ac.kr (S.H.L)

Abstract: Targeted next-generation sequencing (NGS) technology detects specific mutations that can provide treatment opportunities for colorectal cancer (CRC) patients. We included 145 CRC patients who underwent surgery. We analyzed the mutation frequencies of common actionable genes and their association with clinicopathological characteristics and oncologic outcomes using targeted NGS. Approximately 97.9% (142) of patients showed somatic mutations. Frequent mutations were observed in TP53 (70%), KRAS (49%), and APC (47%). TP53 mutations were significantly linked to higher overall stage (p=0.038) and lower disease-free survival (DFS) (p=0.039). ATM mutation was significantly associated with higher tumor stage (p=0.012) and shorter overall survival (OS) (p=0.041). Stage 3 and 4 patients with ATM mutations (p=0.023) had shorter OS, and FBXW7 mutation was significantly associated with shorter DFS (p=0.002). In multivariate Cox regression analysis, ATM mutation was an independent biomarker for poor prognosis of OS (p=0.022). TP53 and FBXW7 mutations are independent biomarkers for poor prognosis of DFS (p=0.042 and 0.030, respectively). A comprehensive analysis of the molecular markers for CRC can provide insights into the mechanisms underlying disease progression and help optimize a personalized therapy.

Keywords: next-generation sequencing; colorectal cancer; ATM mutation; FBXW7 mutation

1. Introduction

Colorectal cancer (CRC) is the third most common malignancy worldwide [1]. Despite the advancements in CRC treatment and the decline in mortality rate over the past few decades, CRC remains the second most common cause of cancer death in women and third common cause of cancer death in men in Korea [2]. Although patients with localized stage CRC have a 5-year overall survival (OS) of 90%, cancer spread to distant organs carries a significantly worse prognosis with a 5-year OS of 14% [1]. Disease spread to distant organs is the major cause of morbidity and mortality in patients with CRC [3].

Several genomic alterations, including KRAS, NRAS, and BRAF mutations, are associated with resistance to targeted therapy with epidermal growth factor receptor (EGFR) monoclonal antibodies, providing a molecular basis for selecting appropriate agents in the treatment of metastatic CRC [4]. Subsequent studies identified other genetic mutations of the EGFR signaling pathways involving the HER2 and FGFR1 genes [5]. Genomic analysis showed that alterations in p53, WNT-β-catenin, TGF-β, EGFR, and downstream MAPK/ERK and PI3K/Akt signaling pathways are associated with CRC

tumorigenesis [6]. In the era of personalized medicine, an in-depth understanding of the molecular profiles and altered signaling pathways is important to identify the patients who may be able to benefit from such treatments.

Genetic or epigenetic alterations of the DNA mismatch repair (MMR) genes may have a predictive value in some cases with CRC. Although testing for MMR status in patients with CRC has been recommended as a workup test to evaluate the possible occurrence of Lynch syndrome, recent data revealed that microsatellite instability (MSI) is a predictive biomarker for immunotherapy [7].

The next-generation sequencing (NGS) approach allows the agnostic analysis of large portions of the genome and can identify multiple mutations with increased sensitivity [8]. This method is currently used in pathology laboratories as a routine molecular test modality [9]. Combined with various clinical information and advanced bioinformatic analysis, the NGS data could be used as a basis for establishing a personalized treatment for cancer patients.

We aimed to describe the mutational profile of patients with CRCs using the targeted NGS approach and analyze their potential correlations with clinicopathological factors. In addition, we aimed to assess the biological and clinical significance of low variant allele frequency (VAF) for small variants and compare them with those of The Cancer Genome Atlas (TCGA) dataset, a publicly available archive.

2. Results

2.1. Study population

A total of 145 patients (76 men and 69 women) with CRC were included in this study. The median age was 60.9 years (range, 25–88 years). Ninety-two (63.4%) of the tumors occurred in the colon, while the remaining 53 (36.6%) tumors developed in the rectum. The detailed clinicopathological characteristics of the study cohort are shown in Table 1.

Table 1. Clinicopathological data of 145 colorectal cancer patients.

Clinicopathological parameters		Number of Patients (N = 145, %)
Sex	Male	76 (52.4)
	Female	69 (47.6)
Age	≥60	78 (53.8)
	<60	67 (46.2)
Tumor site	Right colon	45 (31.0)
	Left colon	47 (32.4)
	Rectum	53 (36.6)
T stage	Tis	2 (1.4)
	T1	3 (2.1)
	T2	6 (4.1)
	T3	89 (61.4)
	T4	45 (31.0)
N stage	N0	35 (24.1)
	N1	47 (32.4)
	N2	63 (43.5)
M stage	M0	83 (57.2)
	M1	62 (42.8)

Stage	0	2 (1.4)
	1	7 (4.8)
	2	17 (11.7)
	3	57 (39.3)
	4	62 (42.8)
Lymphatic invasion	Negative	46 (31.7)
	Positive	81 (55.9)
	Not available	18 (12.4)
Vascular invasion	Negative	80 (55.2)
	Positive	47 (32.4)
	Not available	18 (12.4)
Perineural invasion	Negative	69 (47.6)
	Positive	58 (40.0)
	Not available	18 (12.4)
Differentiation	Well	12 (8.3)
	Moderately	104 (71.7)
	Poorly	15 (10.3)
	Not available	14 (9.7)
Follow-up	Recurrence	37 (25.5)
	Died	5 (3.4)
Initial CEA ^a	≥4	70 (48.3)
	<4	72 (49.6)
	Not available	3 (2.1)
EGFR immunohistochemistry	0	1 (0.7)
	1+	54 (37.2)
	2+	54 (37.2)
	3+	26 (17.9)
	Not available	10 (7.0)

^aCEA: carcinoembryonic antigen

2.2. Mutational profile analysis

Of 145 patients, 97.9% (142) showed somatic mutations. Frequent mutations were found in TP53 (70%), KRAS (49%), APC (47%), PIK3CA (23%), FBXW7 (13%), and SMAD4 (12%). The genes with a mutation frequency of >1% are presented in **Figure 1**. In comparison with the mutation frequencies reported by TCGA CRC dataset, the mutation frequencies of TP53, KRAS, and PIK3CA genes were higher in our cohort, and APC and FBXW7 mutations rarely occurred [10] (**Figure A1**). In the same manner, the KRAS and PIK3CA mutation rates were higher than the mutation rates reported in the Memorial Sloan Kettering Cancer Center (MSKCC) CRC dataset. On the contrary, our dataset indicated lower rates of APC and SMAD4 mutations [11] (**Figure A2**).

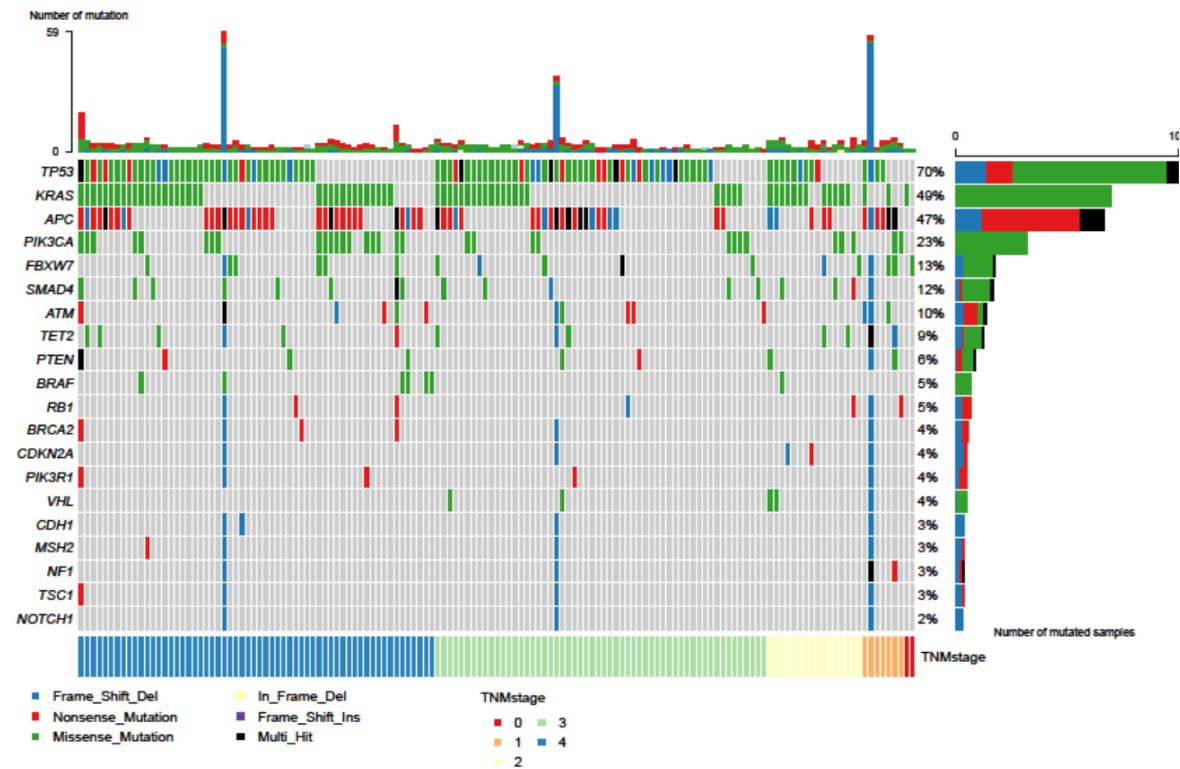


Figure 1. Oncoplot of the top 20 most frequently mutated genes in our cohort. This shows the list of top 20 genes arranged based on the total number of variants in each gene, and the percentage following each gene represents the ratio of tumor samples with its genetic alteration to the total samples. Colored squares indicate the mutated genes, while gray squares indicate the non-mutated genes.

Note: Variants annotated as Multi_Hit refer to those genes that mutated more than once in the same sample.

2.2.1. TP53 mutations

A total of 107 different TP53 variants were detected from 100 patients. Among pathogenic/likely pathogenic mutations, the most common alterations were those affecting the arginine residue 273 (13 cases; p.Arg273Cys and p.Arg273His) as well as the arginine residues 175 and 248 (8 and 7 cases; p.Arg175His and p.Arg248Trp, respectively) (Table 2). In both TCGA and MSKCC CRC datasets, p.Arg175His is the most common variant (15 and 67 cases, respectively) (Tables A1 and A2). The presence of missense-type mutant p53 at codon p.Arg175 and p.Arg273 played a key role in the submucosal invasion and metastasis of intestinal tumors through the gain-of-function mechanism [12].

Table 2. Mutation spectrum of driver genes detected in our cohort.

Gene	Cases (Variants)	Clinical significance	Variation
------	------------------	-----------------------	-----------

TP53	100 (107)	Pathogenic/ Likely pathogenic (n)	p.Arg175His (8), p.Arg196Ter (2), p.Arg209fs (1), p.Arg213Ter (1), p.Arg248Gln (4), p.Arg248Trp (7), p.Arg273Cys (9), p.Arg273His (4), p.Arg280fs (1), p.Arg282Gly (1), p.Arg282Trp (3), p.Arg306Ter (3), p.Arg337fs (1), p.Arg342Ter (2), p.Asn131fs (1), p.Cys124fs (1), p.Cys176Phe (1), p.Cys238Tyr (2), p.Cys275Phe (1), p.Gln167fs (1), p.Gln52fs (1), p.Glu285fs (1), p.Glu285Lys (4), p.Glu339Ter (1), p.Gly244Asp (1), p.Gly244Ser (2), p.Gly245Asp (2), p.Gly245Ser (3), p.Gly266fs (1), p.Ile195Thr (1), p.Leu257Gln (1), p.Met133fs (1), p.Ser241Tyr (1), p.Trp53Ter (1), p.Trp91Ter (1), p.Tyr220Asn (1), p.Tyr220Cys (2), p.Tyr234Cys (1), p.Val122fs (1), p.Val173Leu (1), p.Val73fs (1)
		Uncertain significance (n)	p.Arg249Ser (1), p.Gly244Cys (1), p.Gly262Val (1), p.Gly266Arg (1), p.Gly266Glu (1), p.Lys132Arg (1), p.Lys132Asn (1), p.Met237Ile (1), p.Pro250Leu (1), p.Tyr163His (1)
		Not provided (n)	p.Asn239_Ser240del (1), p.Cys135Ter (1), p.Glu224Ter (1), p.Glu258Asp (1), p.Glu287Ter (1), p.Glu326fs (1), p.Lys292Ter (1), p.Lys320fs (1), p.Ser240Cys (1), p.Ser240Ile (1), p.Thr256Lys (1), p.Tyr220fs (1), p.Tyr220Ter (1), p.Tyr236fs (1)
KRAS	75 (75) ^a	Pathogenic/ Likely pathogenic (n)	p.Ala146Val (1), p.Ala59Thr (2), p.Gln61His (1), p.Gln61Leu (1), p.Gln61Lys (2), p.Gly12Ala (3), p.Gly12Asp (29), p.Gly12Cys (3), p.Gly12Ser (3), p.Gly12Val (7), p.Gly13Asp (14)
		Uncertain significance (n)	p.Ala146Thr (3)
		Not provided (n)	p.Lys117Leu (1)
APC	67 (106)	Pathogenic/ Likely pathogenic (n)	p.Arg1114Ter (3), p.Arg1450Ter (5), p.Arg213Ter (6), p.Arg216Ter (1), p.Arg2204Ter (2), p.Arg2237Ter (1), p.Arg232Ter (3), p.Arg283Ter (2), p.Arg405Ter (1), p.Arg499Ter (2), p.Arg564Ter (4), p.Arg876Ter (2), p.Asn1455fs (1), p.Asn1830fs (1), p.Asn659fs (1), p.Gln1338Ter (2), p.Gln1367Ter (1), p.Gln1378Ter (2), p.Gln1429Ter (5), p.Gln473Ter (2), p.Glu1306Ter (1), p.Glu1309fs (2), p.Glu1309Ter (1), p.Glu1353Ter (1), p.Leu1488fs (1), p.Lys1030Ter (1), p.Lys1199fs (1), p.Ser1044Ter (1), p.Ser105fs (1), p.Ser1298fs (1), p.Ser1465fs (2), p.Ser1857fs (1), p.Ser457Ter (3), p.Tyr1075Ter (1), p.Tyr935Ter (2), p.Val452fs (1)
		Uncertain significance (n)	None
		Not provided (n)	p.Arg220fs (1), p.Asn1473fs (1), p.Asn1531fs (1), p.Asn1782fs (1), p.Asn741fs (1), p.Asp1486fs (1), p.Gln1035Ter (1), p.Gln1303Ter (1), p.Gln222fs (1), p.Gln264Ter (1), p.Glu1374Ter (2), p.Glu1577fs (1), p.Glu1577Ter (1), p.Glu211Ter (1), p.Glu225Ter (1), p.Glu460Ter (1), p.Glu918Ter (1), p.Glu941Ter (1), p.Gly1094fs (1), p.Gly1288Ter (1), p.Gly797fs (1), p.Leu1489fs (1), p.Leu1669Ter (1), p.Leu456fs (1), p.Lys1085Ter (1), p.Lys1308Ter (1), p.Lys1370Ter (2), p.Lys1518Ter (1), p.Lys2689fs (1), p.Lys993Ter (1),

			p.Ser1861Tyr (1), p.Ser671fs (1), p.Ser844fs (1), p.Tyr1147Ter (1), p.Tyr935fs (1), p.Val1472fs (1)
PIK3CA	32 (33)	Pathogenic/ Likely pathogenic (n)	p.Arg108His (1), p.Arg88Gln (3), p.Asn345Lys (1), p.Gln546His (1), p.Glu542Lys (3), p.Glu545Gly (1), p.Glu545Lys (11), p.His1047Arg (6), p.His1047Leu (2), p.Met1043Ile (1)
		Uncertain significance (n)	p.Asn1044Lys (1)
		Not provided (n)	p.Arg357Gln (1), p.Arg93Trp (1),
FBXW7	18 (21)	Pathogenic/ Likely pathogenic (n)	p.Arg465Cys (2), p.Arg465His (3), p.Arg479Gln (3), p.Arg505Cys (2), p.Arg505Leu (1)
		Uncertain significance (n)	None
		Not provided (n)	p.Arg224Ter (1), p. p.Arg689Trp (1), p.Asn598fs (2), p.Ile605fs (1), p.Pro201fs (1), Ser582Leu (3), p.Ser625fs (1)
SMAD4	17 (19)	Pathogenic/ Likely pathogenic (n)	p.Arg361Cys (2), p.Arg361His (6), p.Arg361Ser (1), p.Glu330Ala (1), p.His530fs (1)
		Uncertain significance (n)	p.Gln549Ter (1)
		Not provided (n)	p.Asp351Gly (1), p.Asp396fs (1), p.Glu520Ter (1), p.Gly395fs (1), p.Gly419Arg (1), p.Pro356Leu (2)
ATM	14 (38)	Pathogenic/ Likely pathogenic (n)	p.Arg1875Ter (1), p.Arg250Ter (1), p.Arg2993Ter (1), p.Arg3008His (1), p.Gly301fs (1), p.Lys2811fs (2), p.Pro1566fs (1), p.Ser1281fs (1), p.Ser1799fs (1)
		Uncertain significance (n)	p.Arg23Gln (1), p.Arg337Cys (1),
		Not provided (n)	p.Ala1942fs (3), p.Ala2225fs (1), p.Arg1619fs (1), p.Asn1062fs (1), p.Asn2586fs (1), p.Gln268fs (1), p.Gln446Ter (1), p.Glu376Ter (1), p.Gly2083Ter (1), p.Ile1659Ter (1), p.Ile1960fs (1), p.Ile598fs (1), p.Lys902fs (1), p.Phe2908fs (1), p.Phe2927fs (2), p.Phe492fs (1), p.Phe771fs (1), p.Pro453fs (1), p.Ser1360fs (1), p.Ser2581fs (1), p.Thr491fs (2), p.Trp308Ter (1)
TET2- AS1	13 (18)	Pathogenic/ Likely pathogenic (n)	None
		Uncertain significance (n)	None
		Not provided (n)	p.Ala893fs (1), p.Arg1404Ter (1), p.Arg544Ter (1), p.His222fs (1), p.Leu1276fs (1), p.Leu1457Ter (1), p.Lys1669fs (1), p.Phe329fs (1), p.Phe868Leu (9), p.Thr624fs (1)
PTEN	9 (12)	Pathogenic/ Likely pathogenic (n)	p.Arg130Gln (1), p.Arg173His (1), p.Gln261Ter (1), p.Gln298Ter (1), p.Glu7Ter (1), p.Ile101fs (1)
		Uncertain significance (n)	p.Pro38Thr (1)
		Not provided (n)	p.Ala151Thr (1), p.Asn340fs (1), p.Lys144Thr (1), p.Tyr177His (1), p.Tyr68Phe (1)

BRCA2	7 (30) ^b	Pathogenic/ Likely pathogenic (n)	p.Glu2183Ter (1), p.Glu510Ter (1), p.Ile2278fs (2), p.Lys1132fs (1), p.Val726fs (2)
		Uncertain significance (n)	p.Arg3384Ter (1)
		Benign (n)	p.Asn289His (1), p.Asn991Asp (1), p.Val2466Ala (1)
		Not provided (n)	p.Asn2463fs (1), p.Gln2354fs (1), p.Gln2655fs (3), p.Lys697fs (1), p.Pro1702fs (1), p.Pro2246fs (1), p.Val1532fs (1),

^aOf 75 patients, 5 showed amplification.

^bOf 30 cases, 6 had synonymous mutations, and 4 had intronic mutations.

Each SNV was classified according to the ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar/>).

2.2.2. KRAS mutations

Seventy-five mutations from 75 patients were identified in KRAS with the most common affecting glycine 12 residue (45 cases; p.Gly12Asp/Val/Ala/Cys/Ser), followed by changes in the glycine 13 residue (14 cases; p.Gly13Asp) (**Table 2**). Likewise, p.Gly12Asp mutation was the most common recurrent variant in both TCGA and MSKCC CRC datasets (**Tables A1 and A2**).

With regard to the overall RAS mutations, one mutation in codon 61 (p.Gln61Lys) of NRAS was also detected. However, no HRAS mutations were reported in our cohort.

2.2.3. APC mutations

APC mutations were detected in 67 patients (106 variants). APC mutations were not biased in a particular domain; however, the vast majority of them were truncating mutations. The mutations affecting the arginine 1450 residue, which is one of mutational hotspots for somatic APC mutations, were commonly reported in our study (6 cases; p.Arg213Ter) (**Table 2**) [13]. In TCGA and MSKCC CRC cohorts, the alterations affecting the arginine 1432 (19 cases; p.Arg1432Ter) and 876 residues (44 cases; p.Arg876Ter) were the most common mutations, respectively (**Tables A1 and A2**).

2.2.4. PIK3CA mutations:

Thirty-three PIK3CA mutations were detected in 32 patients. The most frequently reported mutations were changes in the glutamic residue 545 in exon 9 of the protein (12 cases; p.Glu545Lys and p.Glu545Gly) (**Table 2**). The mutations affecting the glutamic residue 545 (11 cases; p.Glu545Lys/Ala/Gly/Gln) commonly occurred in the TCGA CRC cohort. Likewise, p.Glu545Lys was the most frequent alteration (52 cases) in the MSKCC CRC dataset (**Tables A1 and A2**).

2.2.5. SMAD4 mutations

Nineteen mutations from seventeen patients were identified in the SMAD4 genes. Of these, pathogenic/likely pathogenic mutations were detected in 11 variants with the most common affecting arginine residue 361, which is the hotspot region for missense mutations (6 cases; p.Arg361His) (**Table 2**) [14]. In the same manner, the most frequent variant was also detected at arginine residue 361 in TCGA (8 cases; p.Arg361His) and MSKCC CRC datasets (32 cases; p.Arg361His and p.Arg361Cys) (**Tables A1 and A2**).

The remaining recurrent mutations identified are summarized in **Table 2**.

In the present study, the presence of TP53 mutations showed mutual exclusivity with that of PIK3CA mutations, which is also shown in both TCGA and MSKCC cohorts. These findings are in line with that evidence that PI3CA and TP53 alterations tend to be mutually exclusive in diverse

tumors [15]. On the contrary, the ATM mutations positively correlated with BRCA2 mutations, which was also confirmed with the MSKCC CRC data. Mutually exclusive or co-occurring set of genes was detected in our dataset, and TCGA and MSKCC cohorts are shown in **Figure 2**, and **Figures A3 and A4**, respectively.

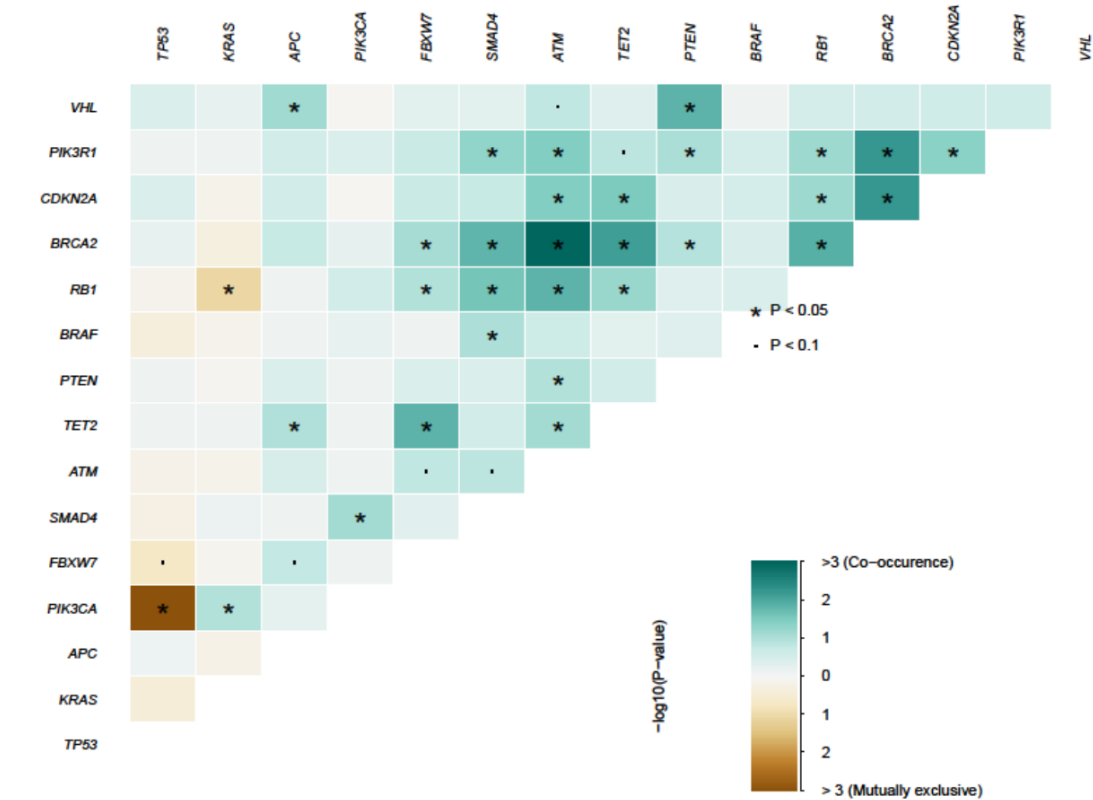


Figure 2. Mutually exclusive and co-occurring gene pairs in our dataset presented in a triangular matrix. Bluish green indicates a tendency toward co-occurrence, whereas brown indicates a tendency toward mutual exclusiveness.

2.3. Outcome analysis

We evaluated the prognostic impact of the mutational profile on clinical outcomes. TP53 mutations were significantly associated with lymph node metastasis ($p=0.01803$) and higher overall stage ($p=0.03813$). BRAF mutation was predominantly harbored by the presence of metastasis (0.04222). In addition, FBXW7 and ATM mutations were significantly associated with higher tumor stage ($p=0.03191$ and 0.01237 , respectively). The association of common gene mutations in CRC with TNM classification and overall stage are summarized in **Table 3**. In terms of the correlation between gene mutation and demographic and pathological parameters, APC mutations were significantly associated with vascular invasion ($p=0.01794$). In addition, the frequency of mutations affecting FBXW7 was higher in male patients ($p=0.01063$) (**Table 3**).

Table 3. Correlation of common gene mutations in CRC according to the clinicopathological data of our cohort

Mutations	T stage		N stage		M stage		Stage	
	1, 2	3, 4	0	1, 2	0	1	1, 2	3, 4
RAS	p = 0.8676		p = 0.4023		p = 0.5007		p = 0.7053	
Negative	6	63	14	55	42	27	11	58
Positive	5	71	21	55	41	35	15	61
BRAF	p = 1		p = 0.6752		p = 0.04222		p = 1	
Negative	11	127	33	105	82	56	25	113
Positive	0	7	2	5	1	6	1	6
PIK3CA	p = 0.7073		p = 0.7166		p = 0.05119		p = 0.9012	
Negative	8	105	26	87	70	43	21	92
Positive	3	29	9	23	13	19	5	27
TP53	p = 0.09612		p = 0.01803		p = 0.4125		p = 0.03813	
Negative	6	39	17	28	23	22	13	32
Positive	5	95	18	82	60	40	13	87
APC	p = 0.3726		p = 0.7935		p = 0.1024		p = 0.8235	
Negative	4	74	20	58	50	28	15	63
Positive	7	60	15	52	33	34	11	56
SMAD4	p = 1		p = 1		p = 0.904		p = 0.5087	
Negative	10	118	31	97	74	54	22	106
Positive	1	16	4	13	9	8	4	13
FBXW7	p = 0.03191		p = 0.1424		p = 0.9203		p = 0.09594	
Negative	7	120	28	99	72	55	20	107
Positive	4	14	7	11	11	7	6	12

ATM	p = 0.01237		p = 0.7439		p = 1		p = 0.717	
Negative	7	124	31	100	75	56	23	108
Positive	4	10	4	10	8	6	3	11

164

165 *Table 3 (continued)*

Mutations	Age		Sex		Differentiation ^a		Lymphatic invasion ^b		Vascular invasion ^b		Perineural invasion ^b	
	<60	≥60	M	F	W to M	Poorly	Absent	Present	Absent	Present	Absent	Present
RAS	p = 0.8369		p = 0.437		p = 0.9249		p = 0.1176		p = 0.9134		p = 0.972	
Negative	33	36	39	30	56	8	17	43	37	23	32	28
Positive	34	42	37	39	60	7	29	38	43	24	37	30
BRAF	p = 0.2491		p = 0.2577		p = 0.1002		p = 1		p = 1		p = 0.3746	
Negative	62	76	74	64	113	13	44	78	77	45	65	57
Positive	5	2	2	5	3	2	2	3	3	2	4	1
PIK3CA	p = 0.6054		p = 0.3622		p = 0.3053		p = 1		p = 0.8195		p = 0.7176	
Negative	54	59	62	51	90	14	36	64	64	36	53	47
Positive	13	19	14	18	26	1	10	17	16	11	16	11
TP53	p = 1		p = 0.6962		p = 0.1422		p = 0.08001		p = 0.2426		p = 0.201	
Negative	21	24	22	23	32	7	19	20	28	11	25	14
Positive	46	54	54	46	84	8	27	61	52	36	44	44
APC	p = 0.8782		p = 0.8369		p = 0.2135		p = 0.243		p = 0.01794		p = 1	
Negative	37	41	42	36	61	11	29	41	51	19	38	32
Positive	30	37	34	33	55	4	17	40	29	28	31	26

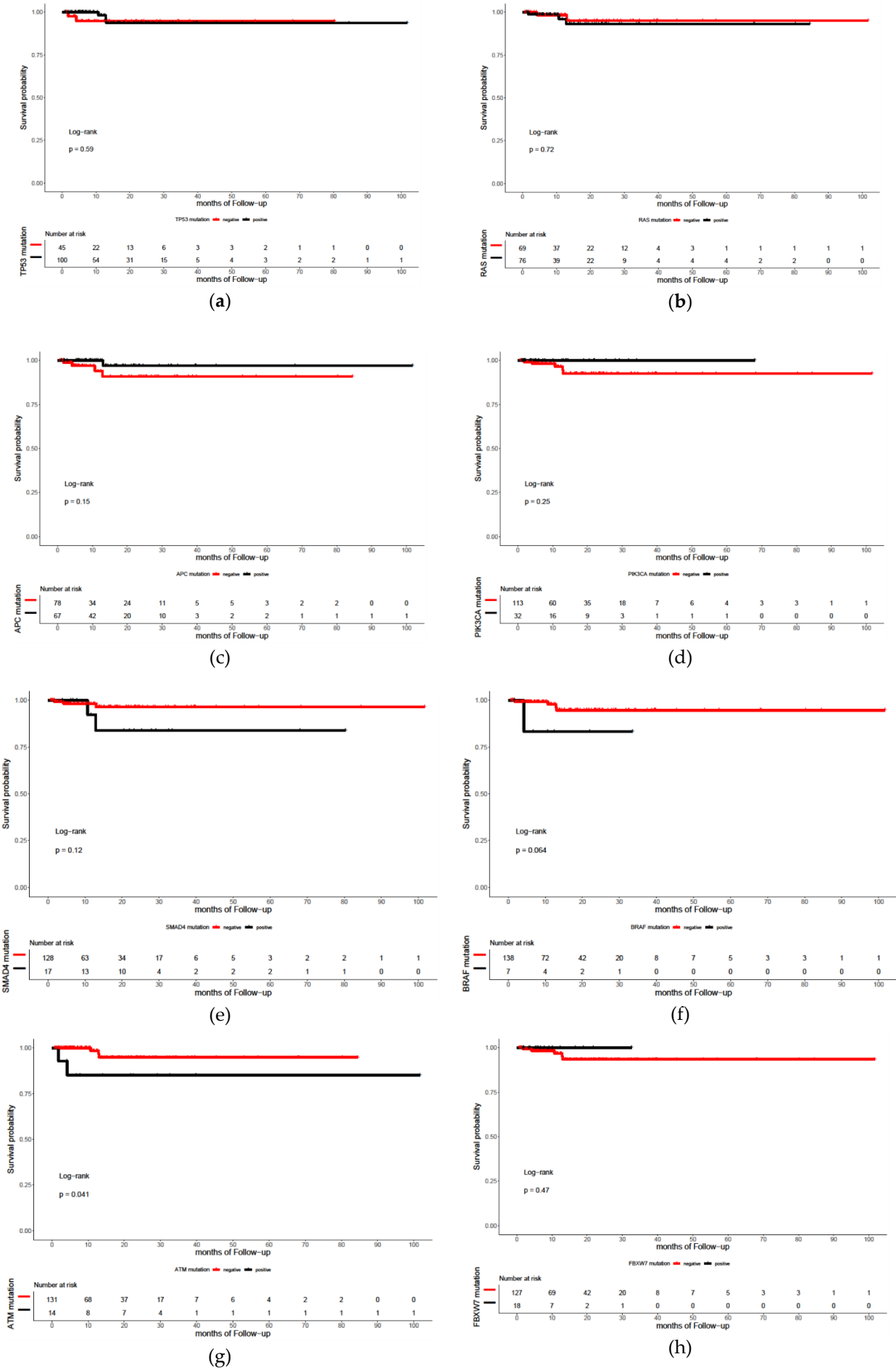
SMAD4	p = 0.8541		p = 0.832		p = 0.6913		p = 0.3429		p = 1		p = 0.9175	
Negative	60	68	68	60	100	14	38	73	70	41	61	50
Positive	7	10	8	9	16	1	8	8	10	6	8	8
FBXW7	p = 1		p = 0.01063		p = 1		p = 0.4667		p = 0.6693		p = 0.2363	
Negative	59	68	61	66	101	13	38	72	68	42	57	53
Positive	8	10	15	3	15	2	8	9	12	5	12	5
ATM	p = 0.986		p = 0.5129		p = 0.6284		p = 0.35		p = 0.7595		p = 0.5505	
Negative	60	71	67	64	106	13	40	75	73	42	61	54
Positive	7	7	9	5	10	2	6	6	7	5	8	4

^aA total of 14 cases have no available data.

^bA total of 18 cases have no available data.

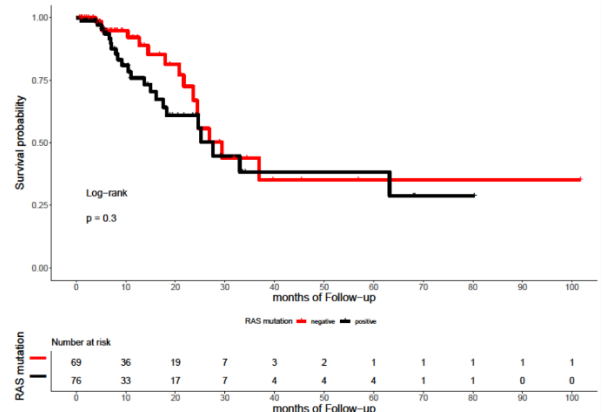
For TCGA dataset, PIK3CA mutations correlated with lower N stage and overall stage (p=0.0217 and 0.00253, respectively). In addition, ATM mutations were significantly associated with the absence of metastasis (p=0.00766) (**Table A3**).

The median time of follow-up was 16.5 months (range, 0.7–101.7 months). In the present study, the OS of patients with or without TP53, RAS, APC, PIK3CA, and SMAD4 mutations did not differ significantly (p=0.59, 0.72, 0.15, 0.25, and 0.12, respectively; **Figure 3**). Similarly, the DFS between patients with RAS, APC, PIK3CA, and SMAD4 mutations and those with wild-type mutations were not statistically different (p=0.3, 0.73, 0.13, and 0.59, respectively; **Figure 4**). However, the DFS of patients with TP53 mutation was significantly shorter than that of patients with TP53 wild type (p=0.039; **Figure 4A**). Furthermore, ATM mutation was significantly associated with shorter OS (p=0.041, **Figure 3G**).

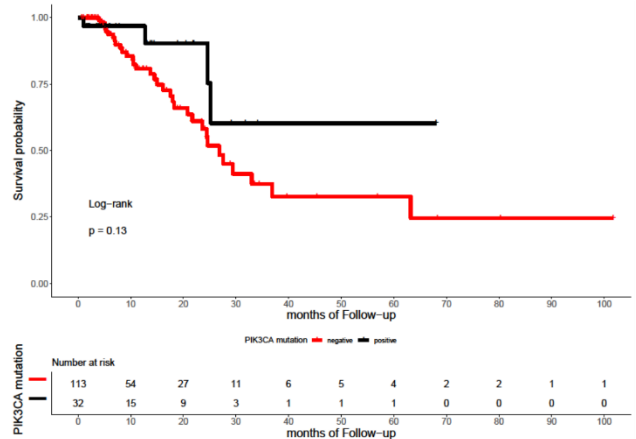


182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205

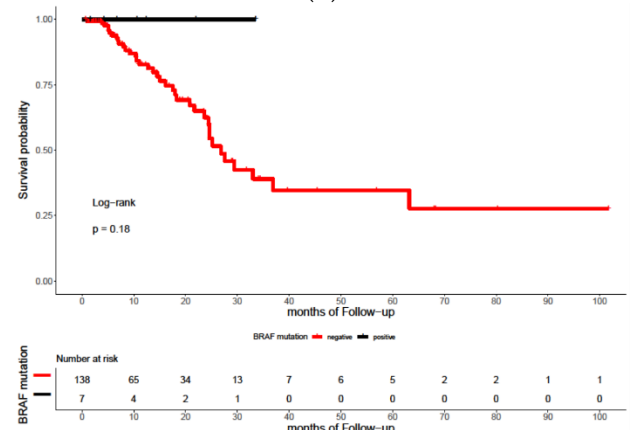
Figure 3. Kaplan–Meier curve for OS in 145 CRC patients by mutational status, including TP53 (A), RAS (B), APC (C), PIK3CA (D), SMAD4 (E), BRAF (F), ATM (G), and FBXW7 (H).



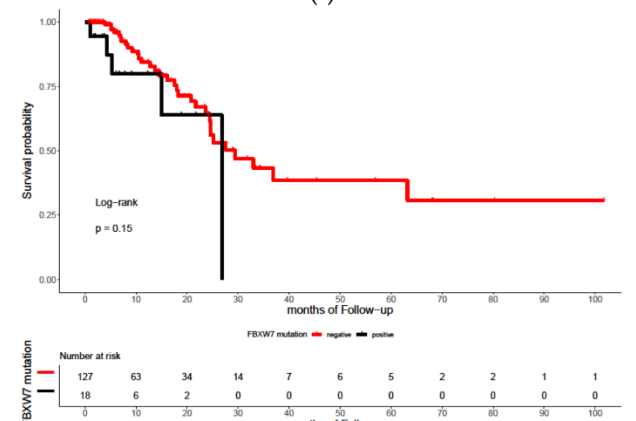
(b)



(d)



(f)



(h)

Figure 4. Kaplan–Meier curve for DFS in 145 CRC patients by mutational status, including TP53 (A), RAS (B), APC (C), PIK3CA (D), SMAD4 (E), BRAF (F), ATM (G), and FBXW7 (H).

With regard to the subset analysis of stage 3 and 4 patients, those with ATM mutation showed shorter OS ($p=0.023$, **Figure A5A**). BRAF mutation showed significant correlations with shorter OS ($p=0.042$, **Figure A5B**), which was not observed among the entire study population. The FBXW7 mutation was significantly associated with shorter DFS in patients with stage 3 and 4 CRC ($p=0.002$, **Figure A6A**). There was a trend toward a poor DFS for CRC in the group with TP53 mutation, although this trend was not statistically significant ($p=0.073$, **Figure A6B**).

For the TCGA CRC dataset, the OS was not significantly different between patients with TP53, KRAS, APC, PIK3CA, SMAD4, and NRAS mutations and their wild types ($p=0.58, 0.96, 0.89, 0.38, 0.19$, and 0.99 , respectively; **Figure A7**). For the MSKCC CRC cohort, there was no difference in OS between patients with TP53, KRAS, PIK3CA, SMAD4, and NRAS mutations and wild type ($p=0.42, 0.39, 0.4, 0.26$, and 0.16 , respectively; **Figure A8**). On the contrary, BRAF mutation was associated with shorter OS, compared with wild-type mutation ($p<0.0001$; **Figure A8F**). The absence of APC was correlated with poor clinical outcomes ($p=0.012$; **Figure A8C**). Contrary to our results, the mutational statuses of ATM and FBXW7 did not differ significantly in terms of OS ($p=0.34$ and 0.14 , respectively; **Figure A8H and A8I**).

In our cohort, the multivariate Cox’s regression analysis of OS showed that ATM mutation is an independent biomarker for poor prognosis with a hazard ratio of 19.637 ($p=0.022$, confidence interval [CI] = $1.547\text{--}249.26$). In the same manner, TP53 and FBXW7 mutations were significant poor prognostic factors for DFS with hazard ratios of 2.80 ($p=0.042$, CI = $1.037\text{--}7.6$) and 3.79 ($p=0.03$, CI = $1.142\text{--}12.6$), respectively. **Figure 5A and B** summarizes the role of clinicopathologic parameters and individual mutations on OS and DFS, respectively.

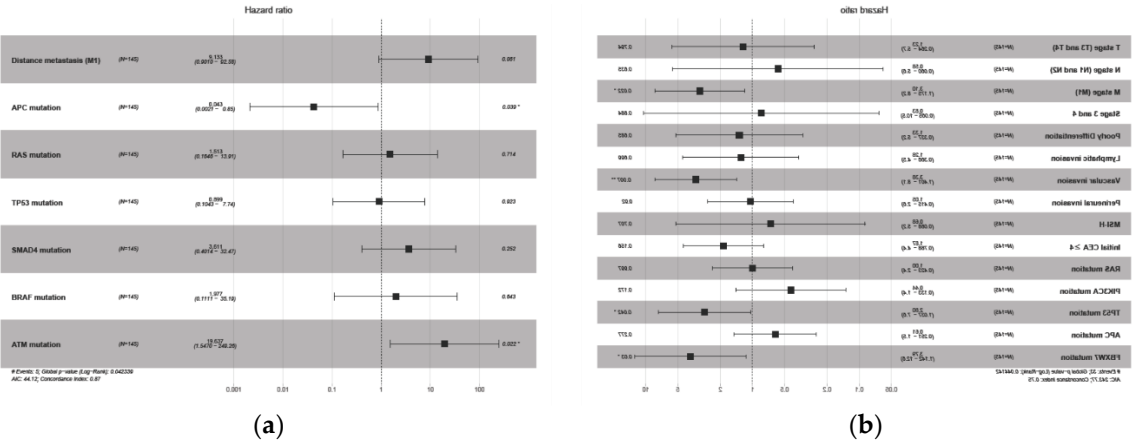


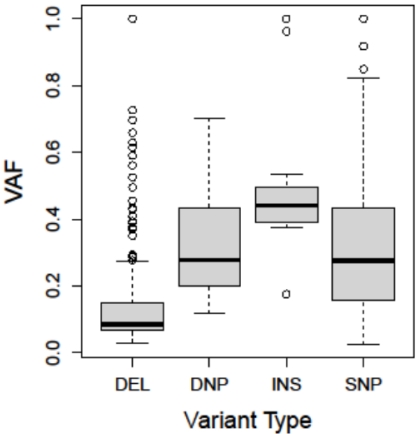
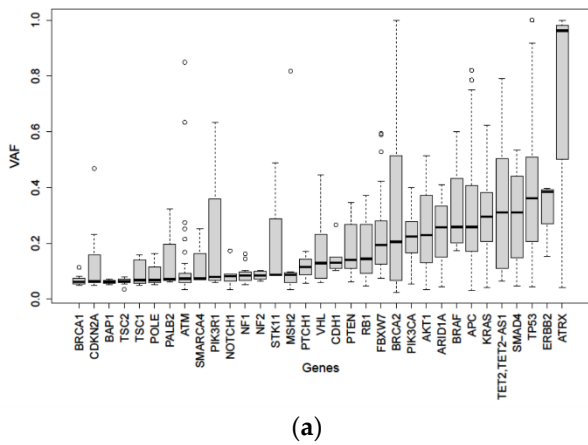
Figure 5. Forest plots for CRC prognostic markers. Multivariable Cox-proportional hazards analysis was performed to identify the prognostic markers that may predict OS (A) and DFS (B). Error bars represent hazard ratios and their 95% confidence intervals.

2.4. Variant allele frequency analysis

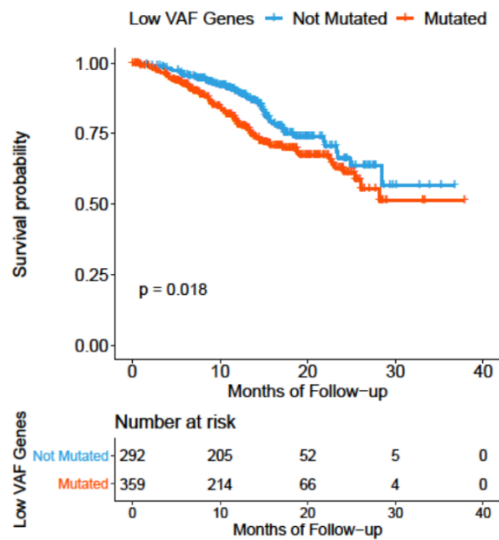
Based on the assumption that VAF is a potential biomarker [16], we investigated the biological and clinical significance of VAF for small variants (SNVs, DNVs, and indels) in our cohort as well as in the public datasets (TCGA and MSKCC cohorts). Among clinically actionable mutations, TP53 showed the highest VAF (**Figure A9**), and higher VAFs were significantly correlated with higher

stages in APC and SMAD4 mutations ($p=0.044$ and 0.047 , respectively). The distribution of median VAFs in our cohort were extremely left shifted (**Figure A10**). These trends were similar to those in the MSKCC cohort, but significantly different from those in TCGA CRC cohort ($p<0.001$), indicating the higher sensitivity of targeted deep sequencing over whole exome sequencing. Among clinically significant small variants in our cohort, the proportions of VAF less than 0.15 and 0.1 were 39.3% and 29.7%, respectively, suggesting that a low VAF variant should not be neglected.

We investigated whether certain genes are more likely to have either lower or higher VAF when mutated. When only 34 VAF-comparable genes (genes with more than two instances of VAFs) were considered, the gene-specific distribution of VAF were easily identified in spite of the large variability (**Figure 6A**). These trends were relatively preserved in the MSKCC cohort. Gene-wise median VAFs were weakly associated with gene-wise instances of VAFs, implying that more frequently mutated genes were more likely to have higher VAFs (**Figure A11A**). These findings straightforwardly suggest that the variations in VAF cannot be entirely explained by technical reasons, and some of the biological features of genes may be associated with VAF. To better understand VAF in relation to mutation type, we explored the distributions of VAF based on the type of mutation (**Figure 6B**). Although VAF seems significantly different in terms of mutation type, these trends were not observed in the MSKCC cohort (**Figure A11B**). These differences might be due to the variations in the NGS platform rather than in cancer biology.



(b)



(c)

Figure 6. Gene-wise distribution of variant allele frequencies (VAF) for genes with more than two instances of somatic mutations (A), distribution of VAFs depending on mutation types (B), and Kaplan–Meier curve for OS in the MSKCC cohort stratified by presence of genes with relatively lower VAF (C)

For the evaluation of clinical relevance of low VAF variants, we classified the 34 genes into low VAF genes and high VAF genes by sorting them according to their median VAFs (Table A4). Each of the clinical cases was classified into “low VAF gene predominant” and “high VAF gene predominant” depending on which type of genes is predominant among the list of mutated genes in each case. No significant difference was shown in OS and DFS between the subgroup with high VAF variants and that with low VAF variants (Figure not shown).

2.5. Comparison of clinicopathological parameters between MSI and MSS CRCs

Of 145 tumors, 133 were included in the MSI analysis. The correlations between MSI status and the clinicopathologic features as well as the mutations of major driver genes in CRC are summarized in Table 4. Microsatellite instability-high (MSI-H) CRCs revealed a marked predilection for the right colon (p=0.009651). MSI-H tumors showed lower T, N, and overall stages than MSI-L/MSS tumors (p=0.004249, 0.002777, and 0.003935, respectively). However, no significant differences were observed in the M stage between patients with MSI-H tumors and those with MSI-L/MSS tumors (p=0.08775). For genetic alterations, MSI-H tumors were characterized by infrequent KRAS mutation (p=0.04454). By contrast, no significant association was found between MSI status and PIK3CA or BRAF mutation (p=0.6935 and 0.3277, respectively). For the TCGA cohort, patients with PIK3CA and BRAF mutations were deviated in the MSI-H groups (p=0.03334 and <0.001, respectively, Table A5).

Table 4. Correlation of MSI status according to the clinicopathological data and RAS, PIK3CA, and BRAF mutations in our cohort.

Clinicopathological parameters	MSI status ^a	
	MSS/MSI-L	MSI-H
Age	p = 0.1853	
≥60	54	7
<60	69	3
Sex	p = 0.7448	
Male	62	6
Female	61	4
Tumor site	p = 0.009651	
Right colon	34	7
Left colon	89	3
T stage	p = 0.004249	
1, 2	7	4
3, 4	116	6
N stage	p = 0.002777	
0	27	7
1, 2	96	3
M stage	p = 0.08775	
0	72	9
1	51	1
Stage	p = 0.003935	
1, 2	20	6
3, 4	103	4

Differentiation^b	p = 1	
Well/ Moderately	102	9
Poorly	14	1
Lymphatic invasion^c	p = 0.7459	
Absent	41	4
Present	74	6
Vascular invasion^c	p = 0.09041	
Absent	70	9
Present	45	1
Perineural invasion^c	p = 0.02117	
Absent	59	9
Present	56	1
Initial CEA level (blood)^d	p = 0.01908	
≥4	60	9
<4	60	1
RAS mutation	p = 0.04454	
Absent	54	8
Present	69	2
PIK3CA mutation	p = 0.6935	
Absent	96	7
Present	27	3
BRAF mutation	p = 0.3277	
Absent	119	9

Present	4	1
---------	---	---

288 ^aA total of 12 patients have no MSI test result.

289 ^bA total of 14 patients have no available data .

290 ^cA total of 18 patients have no available data .

291 ^dThree patients have no available data.

292

293

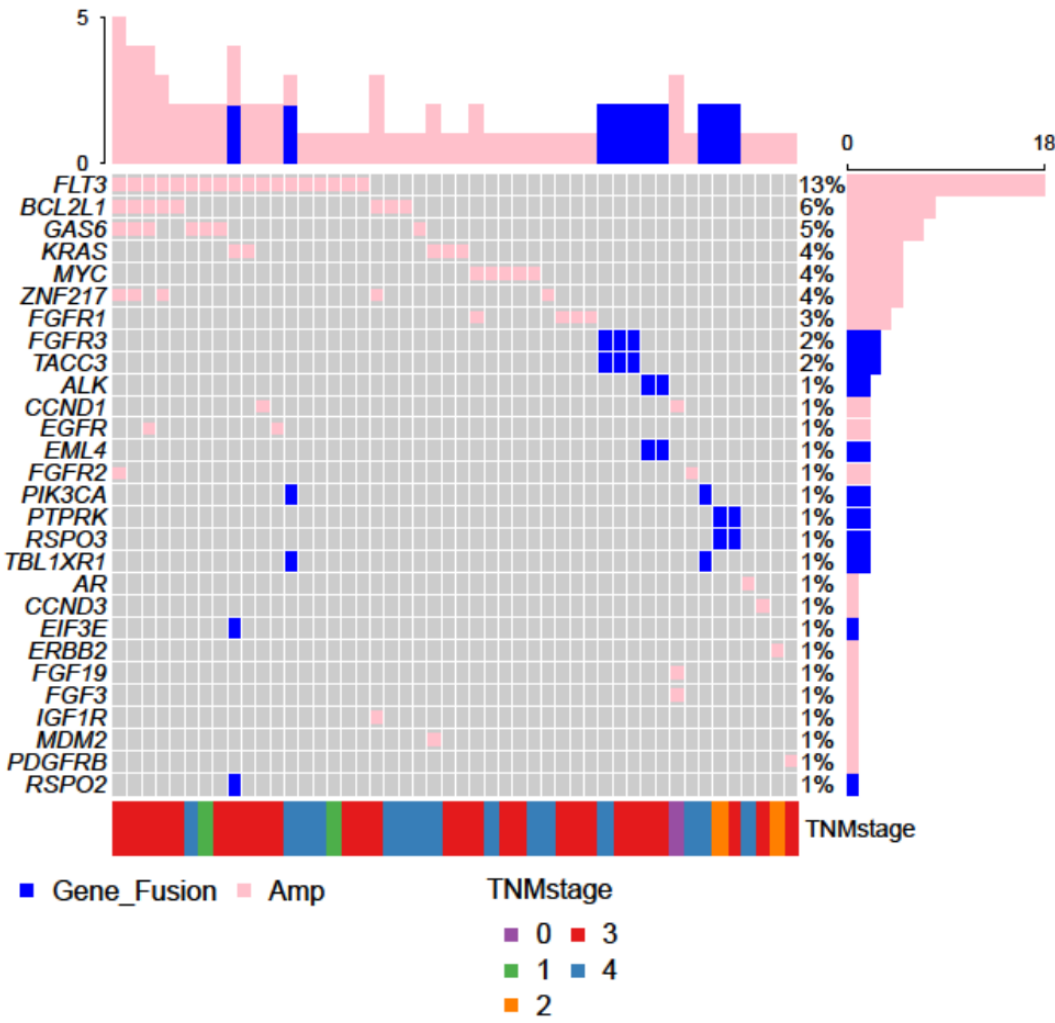
294 2.6. CNA and Fusion gene analysis

295 Before conducting the main analysis, as a sanity check, we investigated whether the degree of copy
296 number change in a case is correlated with a particular gene involved. The copy numbers of
297 sufficiently remote gene pairs would not be correlated without the occurrence of rare events such as
298 whole genome duplications. When gene pairs were randomly chosen in cases with more than two
299 copy number changes, the correlation of copy numbers was higher than expected when a
300 permutation test was performed by permuting the case labels and then the distribution of the
301 Spearman’s correlation of copy numbers was tested (p=0.022). These results indicate that the
302 suggested copy numbers of each gene should be carefully interpreted, and the possibility of some
303 technical artifacts should be considered.

304 Because the copy number itself may be less accurate, but amplification or deletion, whose property
305 is different from the copy number, is less dependent on the scaling factor, the main analysis was
306 performed only with the occurrence of either amplification or deletion. Copy number aberrations
307 were detected in 38 of 139 samples (27.3%), all of which being amplification (**Figure 7**). A total of 18
308 genes were amplified with FLT3 (being the most frequent) and BCL2L1, GAS6, KRAS, MYC,
309 ZNF217, and FGFR1 (being recurrent). All of these genes are well-known oncogenes, and most of
310 them are targetable with specific inhibitors.

311

312



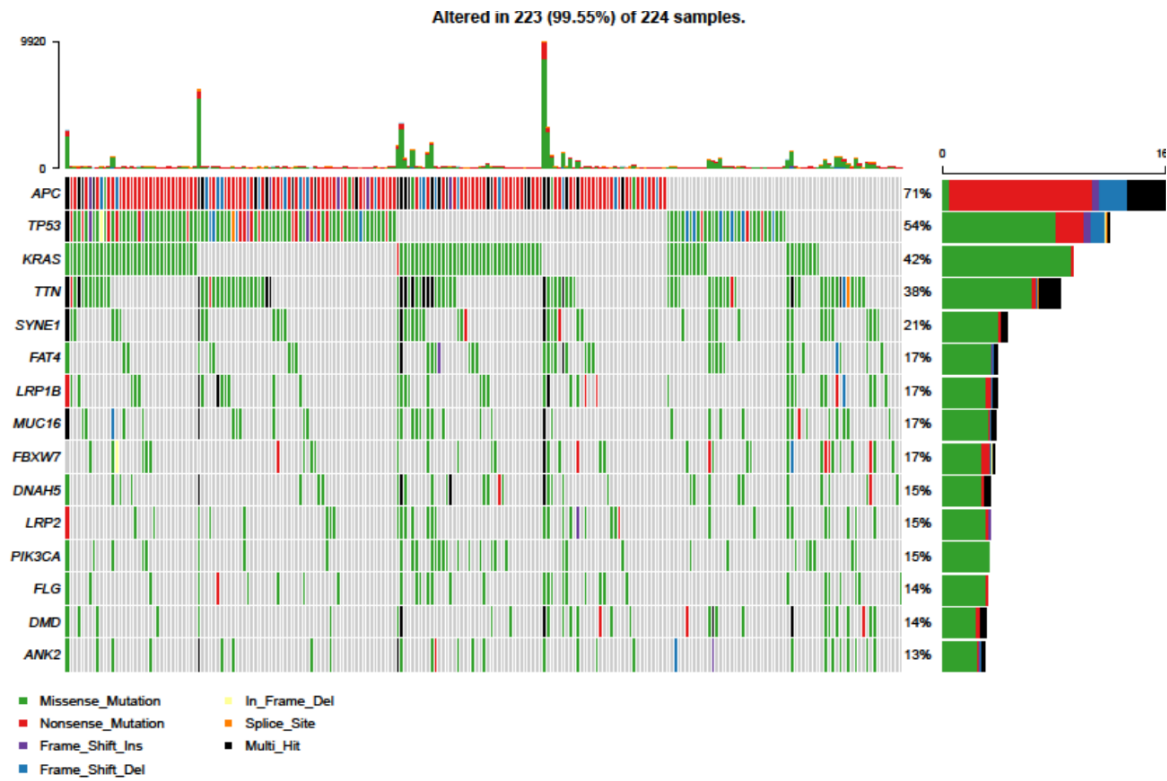


Figure 7. Oncoplot of all genes in which either gene fusion or copy number variation (CNV) was identified. This figure shows the list of genes arranged based on the total number of variants in each gene and the percentage of each gene, which is the ratio of tumor samples with this genetic alteration to total samples.

Fusion genes were identified in 10 cases, and a total of 10 genes were involved, with FGFR3 as the most frequent gene. At least one of the fusion partners were well-known oncogenes in all cases.

3. Discussion

In this study, we showed that TP53 is the most commonly mutated gene in CRC, followed by KRAS and APC. However, APC mutation, followed by TP53 and KRAS mutations, is mostly common in both TCGA and MSKCC CRC datasets. The rates of TP53 mutation are similar to those reported in our cohort and MSKCC dataset (70% vs 73%, respectively); however, the rate of TP53 mutation reported in the TCGA CRC dataset is lower (54%). The frequency of APC mutations reported in this study was 47%, which is significantly lower than that reported in the TCGA and MSKCC studies (71% and 76%, respectively) (**Figure 1 and Figures A1 and A2**). This difference in mutational frequency as well as distribution in codons among our study, TCGA and MSKCC CRC, may be attributable to the differences in the sample selection, ethnicity, and methodology for genomic studies.

KRAS mutated in 49% of CRC patients, which is in accordance with the frequencies reported in various studies on CRCs (35%–45%) including TCGA and MSKCC datasets (42% and 44%, respectively) [17, 18]. The presence of KRAS mutations is considered as predictive markers of negative pharmacological response to EGFR inhibitors, such as cetuximab or panitumumab [19, 20]. However, the prognostic role of KRAS mutations for DFS and OS in CRC patients remains controversial. Various retrospective studies did not show any prognostic significance, but some confirmed that KRAS mutations had negative prognostic role for DFS, OS, or liver metastases [21–23].

In our study, we did not indicate the relationship between RAS mutation and the clinicopathological parameters.

In our study, the BRAF mutation frequency was 4.8% (7/145), which was lower than that identified in the Western cohort (9.2%) but similar to the mutation rates identified in Asian countries (4.9%) [18]. Previous studies have widely demonstrated that patients with CRC who possess a BRAF mutation have significantly poorer clinical outcomes [4, 24]. This was observed also in the present study; BRAF mutations were not only associated with metastatic tumors, but also found to be a negative prognostic marker for OS in a subgroup analysis with advanced CRCs.

Besides the frequently mutated actionable genes, ATM and FBXW7 mutations tended to be a poor prognostic factor in the multivariate analyses of OS and DFS, respectively.

ATM is a member of the phosphatidylinositol-3 kinase-like family of serine/threonine protein kinases and plays a pivotal role in the cellular response to DNA damage by ionizing radiation, which results in DNA double-strand breaks [25]. With regard to the ATM gene alterations, 10% of our cohorts carried mutations. The TCGA and MSKCC dataset reported that ATM mutations were detected in 14% and 7% of CRC patients, respectively. Randon et al. found that ATM mutations are independently associated with longer OS in patients with metastatic CRC [26]. However, we revealed that ATM mutation is linked to poorer OS in patients with CRC. Furthermore, the subgroup analysis of stage 3 and 4 CRCs also identified that ATM mutation is significantly associated with shorter OS. In line with this, loss of ATM expression showed a tendency toward worse survival rate in patients with CIN CRC [27]. In addition, no differences were observed in the OS according to ATM mutation status in both TCGA and MSKCC cohorts. The discrepancy among studies might be due to the different prognostic role of ATM according to the disease stage and the confounding factors from the heterogeneity of available treatment strategies used for metastatic disease [26].

The frequency of FBXW7 mutation in the present study was 13%. This value is consistent with that in previous studies, which reported that 10% of patients with CRC have FBXW7 mutations [28]. FBXW7 is a potential tumor suppressor, and mutations in the gene are thought to impair cyclin E degradation resulting in uncontrolled cell division and growth, thus resulting in cancer progression [29]. A previous study suggested that a missense mutation was correlated with poor OS in CRC patients [30]. We revealed that FBXW7 mutations correlated with shorter DFS in the subgroup with advanced cancer stage.

MSI status is considered as an independent prognostic indicator. In our study, MSI-H CRCs showed significant association with lower T stage, N stage, and clinical stage. Likewise, previous studies generally revealed a better clinical outcome in patients with MSI-H CRCs compared with those with MSS tumors [31, 32]. In our cohort, KRAS-mutated tumors were more frequently found in the MSS group, which is consistent with the findings of previous studies [24, 33]. A previous study by Koyel et al. reported a strong association between high CEA (≥ 4) and MSI-H, which is also reflected in the present study [34].

In terms of cancer evolution, the driver mutations are likely to have a relatively high VAF due to their emergence in the earlier stage of carcinogenesis. On the contrary, actionable variants induced by a treatment may develop with a low VAF in patients with advanced cancer [16, 35]. Recently, the clinical relevance of low VAF variants on the survival of patients who received a targeted therapy was suggested. If the low VAF mutations are as important as high VAF mutations, only a slight difference in the survival curves is expected; on the contrary, if the low VAF variants are not so critical for cancer progression, there might be significant differences in the survival [16]. In our cohort, no significant differences were observed between high-VAF-gene-predominant cases and low-VAF-gene-predominant cases in terms of survival. When the same criteria were applied to the MSKCC cohort, low-VAF-gene predominant cases showed worse prognosis (**Figure 6C**). Even in our cohort, a significant association was found in some clinicopathologic parameters such as TNM stage ($p=0.037$) and lymph node metastasis ($p=0.026$), suggesting that high VAF genes might play a key role in these contexts.

Our study has some limitations. It has a relatively small sample size, a short follow-up time, and a retrospectively designed cohort. The frequency of NRAS mutation was too low to analyze the

association between mutational status and clinical impact. In addition, the retrospectively designed cohort might have influenced the results. Thus, prospective randomized trials are warranted to validate these conclusions. Moreover, the gene alterations tested were only confined to those included in the commercial NGS panel, which might be overcome by the application of cancer panels in which greater numbers of potential genes are covered or WES in future studies.

4. Materials and Methods

4.1. Tumor samples and DNA/RNA extraction

A total of 145 patients with CRC who previously underwent surgical resection in Seoul St. Mary's hospital between 2016 and 2019 were enrolled. All cases were sporadic, without any familial history of CRCs. The clinicopathological parameters were retrospectively reviewed from the medical records. The study protocol was reviewed and approved by the Institutional Review Board of the College of Medicine at the Catholic University of Korea (KC19RESI0669).

CRC areas with rich tumor cell content (at least 70%) from unstained formalin-fixed, paraffin-embedded (FFPE) tissue specimens were obtained for microdissection. DNA and RNA were then extracted using the Recover All Total Nucleic Acid Isolation kit (Ambion, Thermo Fisher Scientific) according to the manufacturer's instructions.

4.2. Library preparation and sequencing

Library preparation was accomplished using the Ion Chef System and Oncomine Comprehensive Assay v3 (Thermo Fisher Scientific, San Francisco, CA, USA) in accordance with the manufacturer's instructions. Genomic DNA was then amplified, and targeted gene sequencing was performed using the Ion S5 XL sequencer (Thermo Fisher Scientific, San Francisco, CA, USA).

The sequencing data were analyzed with the Ion Torrent Suite version 5.10.2 (Life Technologies, San Francisco, CA, USA) using the variantCaller plugin. To eliminate base calling errors, several filtering steps were applied to generate the final variant calling: minimum allele frequency of hotspot variant: $\geq 4\%$, minimum allele frequency of indel variant: $\geq 5\%$, minimum allele frequency of SNP variant: $\geq 5\%$, minimum read counts for fusions: ≥ 40 , CNV gain threshold: 4, and gain confidence level: 0.05. All identified variants were visually confirmed using the Integrative Genomics Viewer (IGV 2.8.6) [36].

4.3. MSI analysis

MSI assays were performed using paired normal and tumor samples as previously described [37]. The five microsatellite markers including two mononucleotide repeats (Bat-25 and Bat-26) and three dinucleotide repeats (D2S123, D5S346, and D17S250) recommended by the National Cancer Institute were amplified in a single multiplex PCR reaction. The PCR products were analyzed by capillary electrophoresis using an ABI 3500 automated sequencer (Applied Biosystems, Foster City, CA). Two of the five microsatellite markers demonstrated instability, and the tumor was considered to be high MSI (MSI-H). MSI at a single locus was defined as low MSI (MSI-L), while the absence of instability at any of the markers was defined as microsatellite stable (MSS). For statistical purposes, MSI-L tumors were considered together with MSS tumors, because of the similarity between MSI-L and MSS tumors [38].

4.4. Public datasets

In this study, we evaluated the clinicopathologic and genomic data from two public datasets and compared these data with those of our cohort.

The TCGA program offers great opportunity to identify the genotype-phenotype relationship, providing extensive archives of multi-omics data. We assessed the somatic mutation data of 459 patients with CRC (colon cancer: 341 and rectal cancer: 118) from the TCGA cohort. The TCGA data were downloaded from the Genomic Data Commons Data Portal (GDC portal, <https://portal.gdc.cancer.gov/>). In addition, data matrices and supporting data from two flagship articles (version 20120719 and version 20180409) were also downloaded [10, 39].

The MSKCC cohort contains a large-scale, prospective clinical sequencing data established by conducting a comprehensive assay, MSK-IMPACT, in 10,336 patients with 62 principal cancer types [11]. We downloaded the supplementary information and source data files available in the online version of the article, which include the data of 1,003 patients with CRCs.

4.5. Statistical analysis

All statistical analyses were performed using a combination of R 4.0.2 for Windows (R Core Team 2020, A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.r-project.org/>) and online chi-square/Fisher's exact test software (GraphPad, www.graphpad.com/quickcalcs/contingency1/). The chi-square test with Yates' correction or the chi-square test for association was used to determine the relationship between the clinicopathological characteristics and gene mutations. OS was defined as the period between initial diagnosis and death from any cause or last follow-up visit. Disease-free survival (DFS) was defined as the period between initial diagnosis and detection of recurrence or metastasis. Survival was analyzed by the Kaplan–Meier method using the log-rank test. Multivariate survival analyses and calculation of hazard ratios (HRs) were performed using the Cox proportional hazards model to determine the independent predictors of OS. A two-sided p-value of <0.05 was considered significant.

5. Conclusions

In conclusion, this study presented a mutational landscape of actionable genes in CRC, validating the prognostic role of genomic alterations in TP53, BRAF, ATM, NRAS, and FBXW7 as well as MSI status. In addition, we investigated the clinical relevance of low VAF variants. A comprehensive analysis of molecular markers for CRC can provide insights into the disease progression and help optimize a personalized therapy in the Korean population.

Supplementary Materials: Supplementary materials can be found at www.mdpi.com/xxx/s1.

Author Contributions: Conceptualization, Y.S.S., and S.H.L.; methodology, I.H.S., A.L., J.K., and S.H.L.; validation, S.H.L.; formal analysis, Y.S.S., and S.H.L.; investigation, C.S.L., and S.H.L.; resources, Y.S.L., and I.K.L.; data curation, C.S.L., and S.H.L.; writing—original draft preparation, C.S.L., and S.H.L.; writing—review and editing, Y.S.S. and S.H.L.; visualization, Y.S.S. and S.H.L.; project administration, S.H.L.; funding acquisition, S.H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science and ICT) (No. 2019R1F1A1062367).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

CRC: Colorectal cancer

OS: overall survival

DFS: disease-free survival

EGFR: epidermal growth factor receptor

MSI: microsatellite instability

NGS: next-generation sequencing

VAF: variant allele frequency

TCGA: The Cancer Genome Atlas

Appendix A

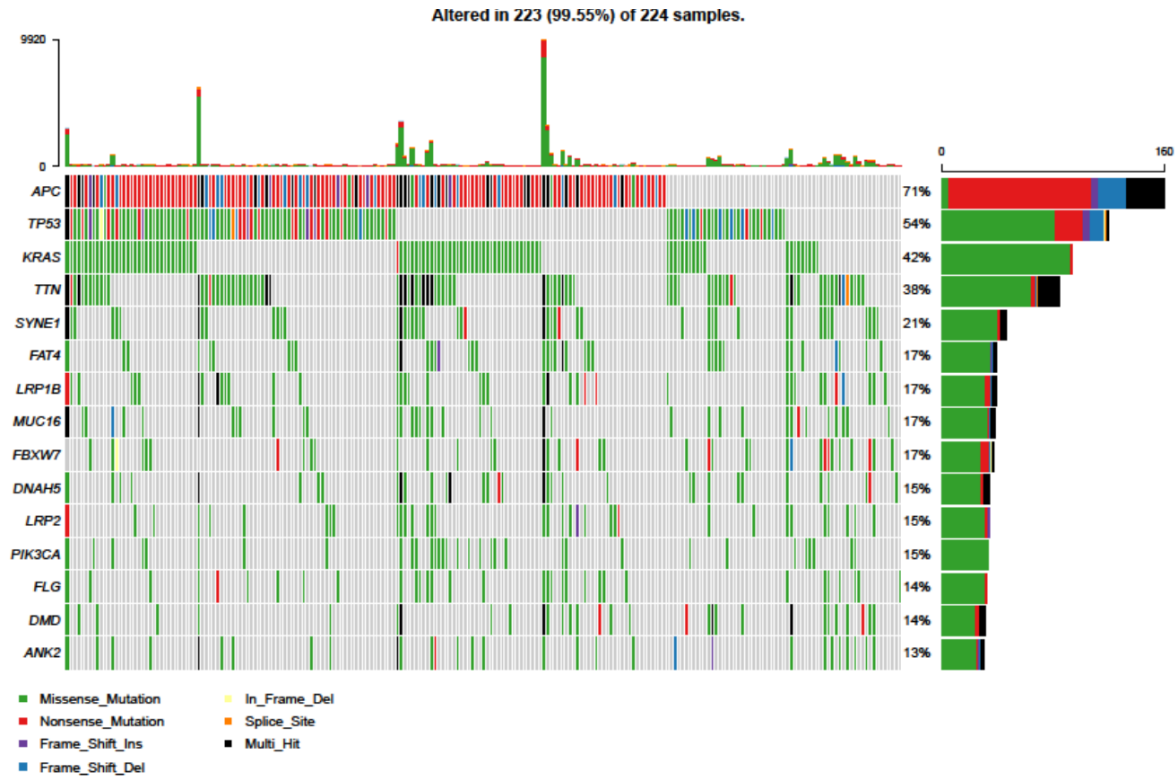


Figure A1. Oncoplot of the top 15 most frequently mutated genes in the TCGA cohort. The plot provided an overview of somatic mutations in particular genes (rows) affecting individual samples (columns). It shows 223 positive samples. TCGA, The Cancer Genome Atlas

489



490
491
492
493

Figure A2. Oncoplot of the top 20 most frequently mutated genes in the MSKCC cohort. There are 970 positive samples. MSKCC, Memorial Sloan Kettering Cancer Center

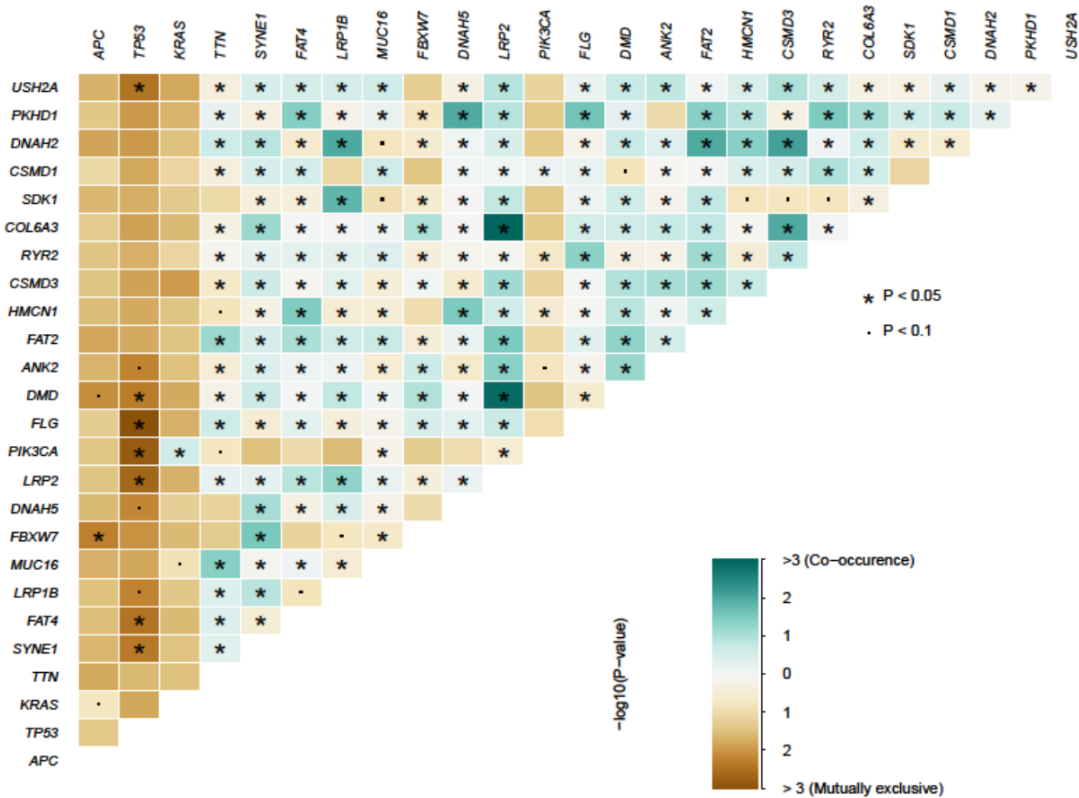


Figure A3. Mutually exclusive and co-occurring gene pairs in the TCGA dataset presented in a triangular matrix. Mutually exclusive/co-occurrence event on top 25 mutated genes are shown. Bluish green indicates tendency toward co-occurrence, whereas brown indicates tendency toward mutually exclusiveness.

494
495
496
497
498
499
500
501
502

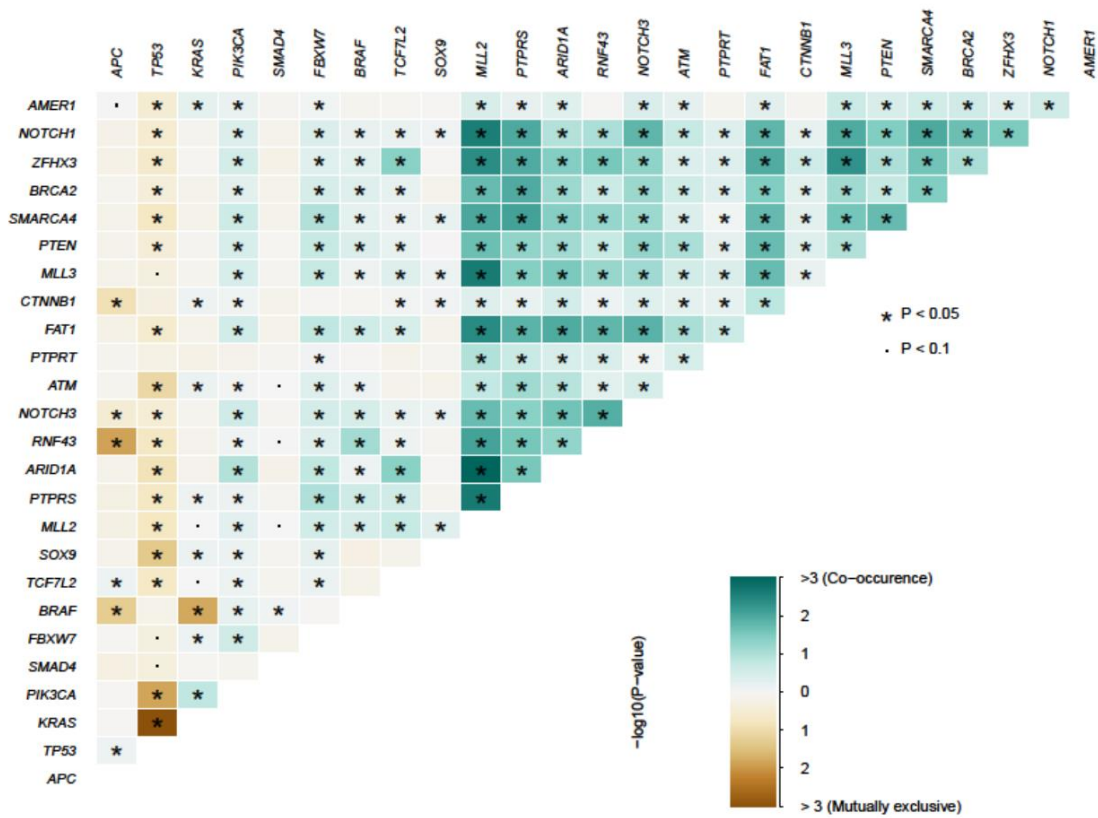


Figure A4. Mutually exclusive and co-occurring gene pairs in MSKCC dataset presented in a triangular matrix. The mutually exclusive/co-occurrence event in the 25 mutated genes are shown. Bluish green indicates a tendency toward co-occurrence, whereas brown indicates a tendency toward mutually exclusiveness.

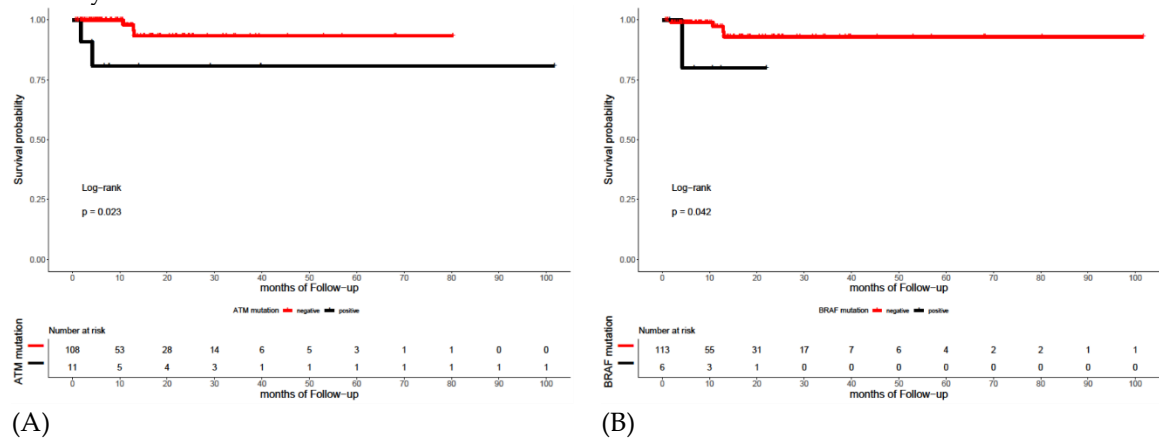
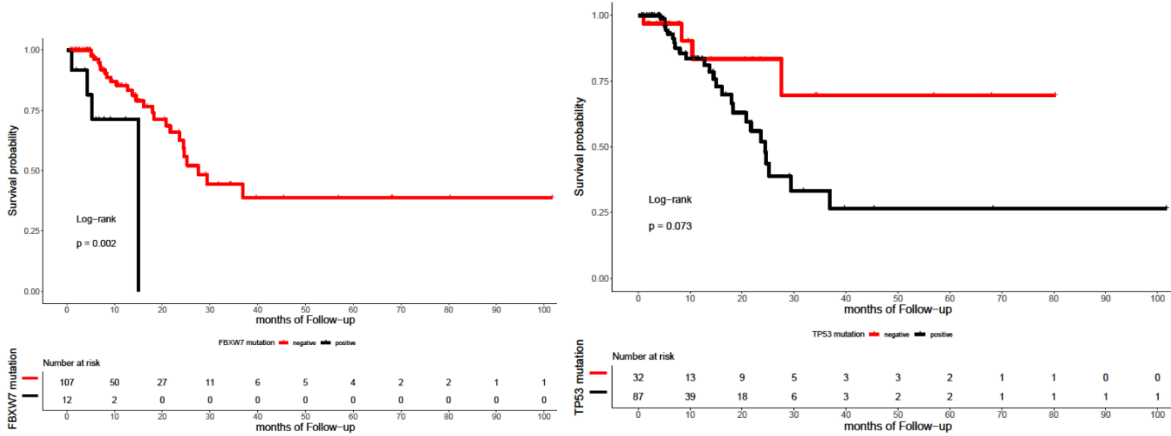
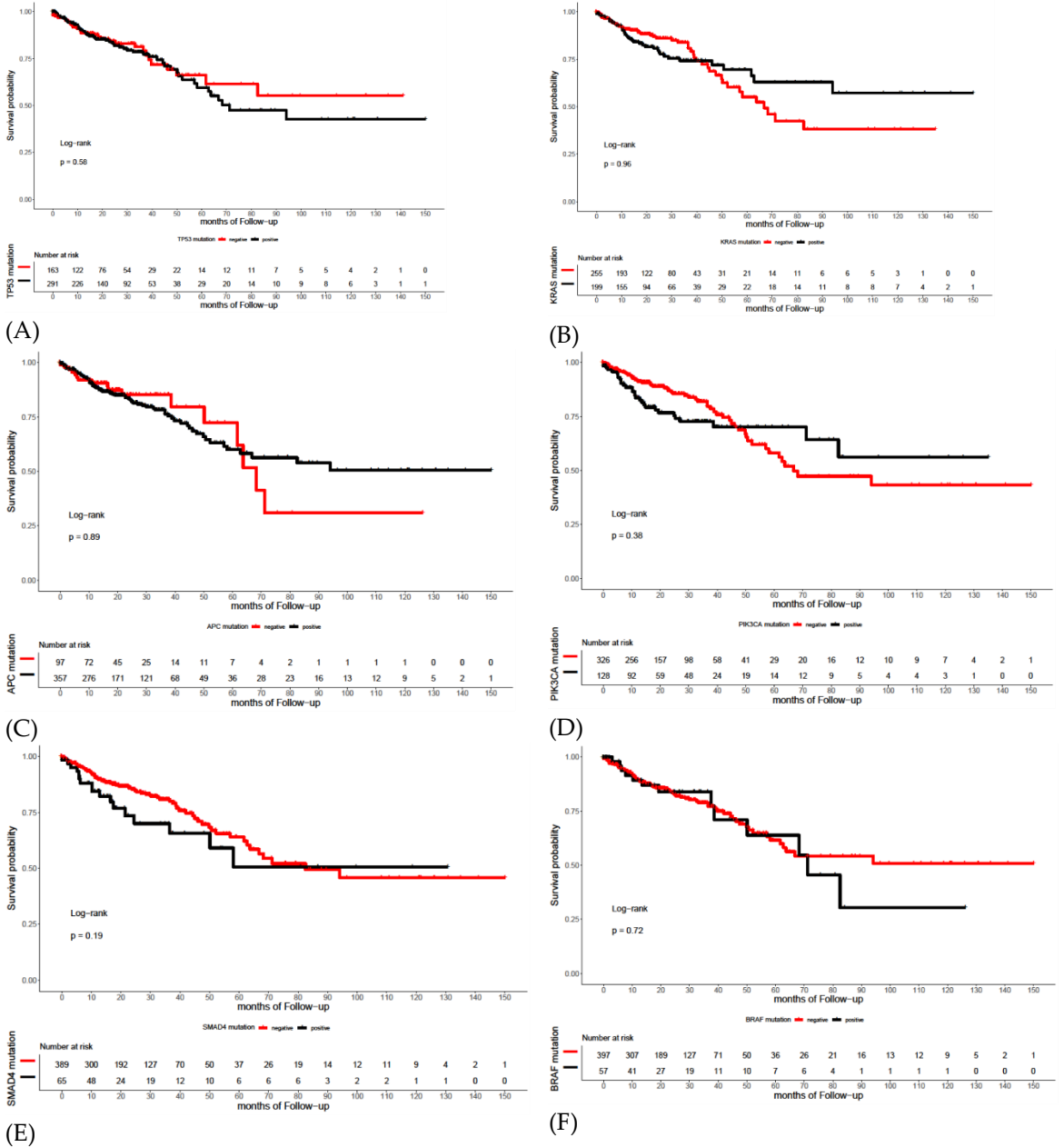


Figure A5. Kaplan-Meier curve for OS in stage 3 and 4 population by mutational status, including ATM (A) and BRAF (B).



(A) (B) Figure A6. Kaplan-Meier curve for DFS in stage 3 and 4 population by mutational status, including FBXW7 (A) and TP53 (B)



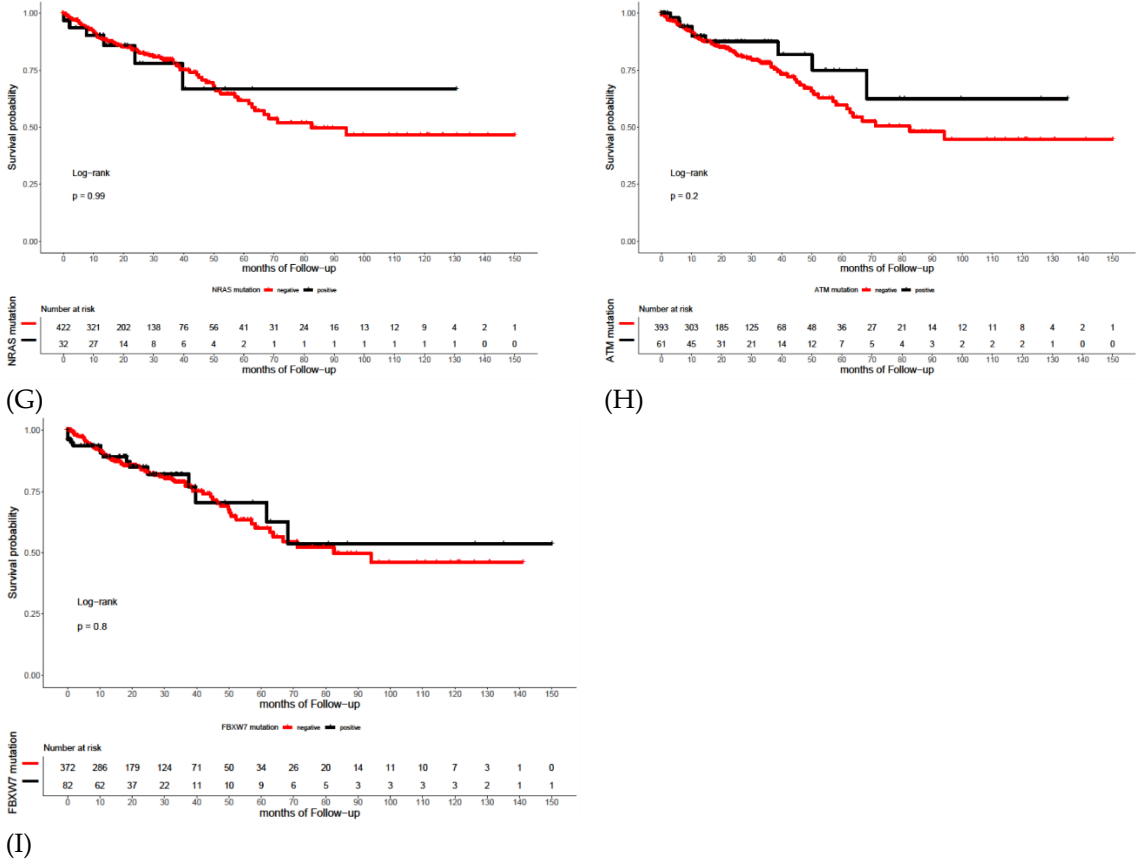
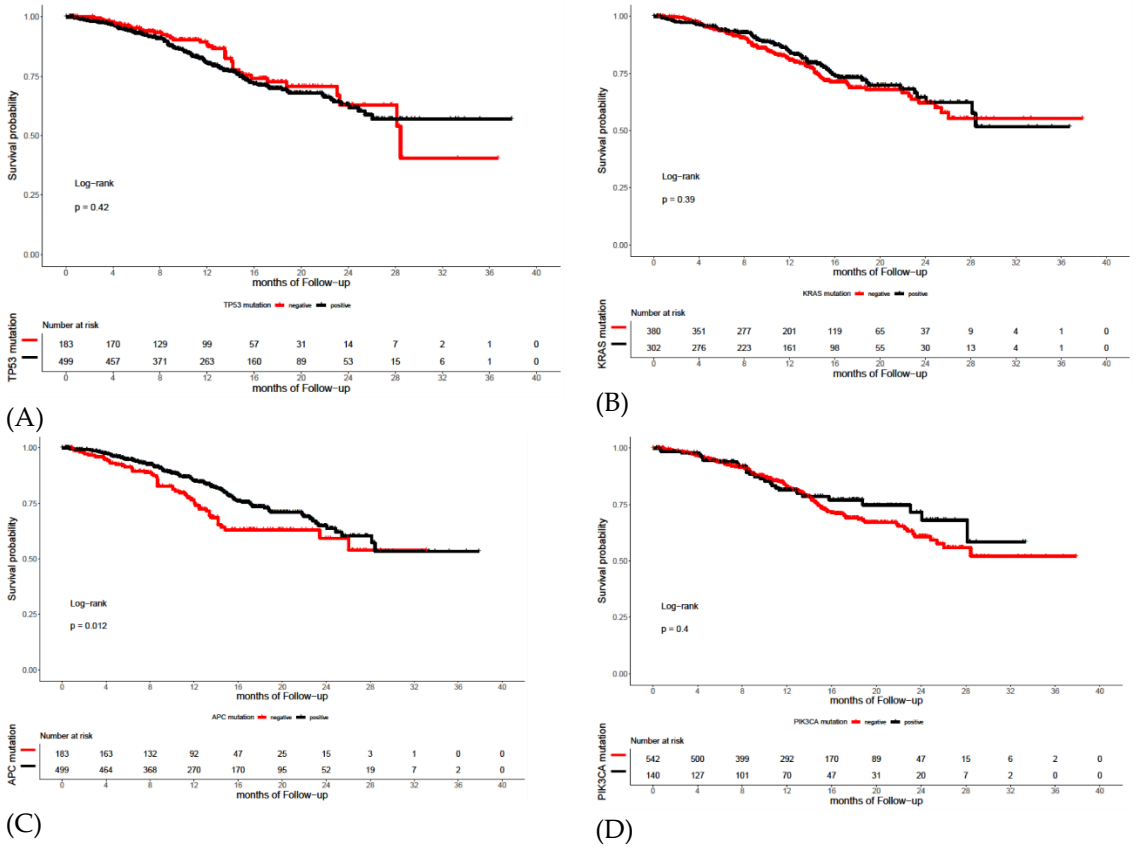


Figure A7. Kaplan-Meier curve for OS in the TCGA dataset by mutational status, including TP53 (A), KRAS (B), APC (C), PIK3CA (D), SMAD4 (E), BRAF (F), NRAS (G), ATM (H), and FBXW7 (I)



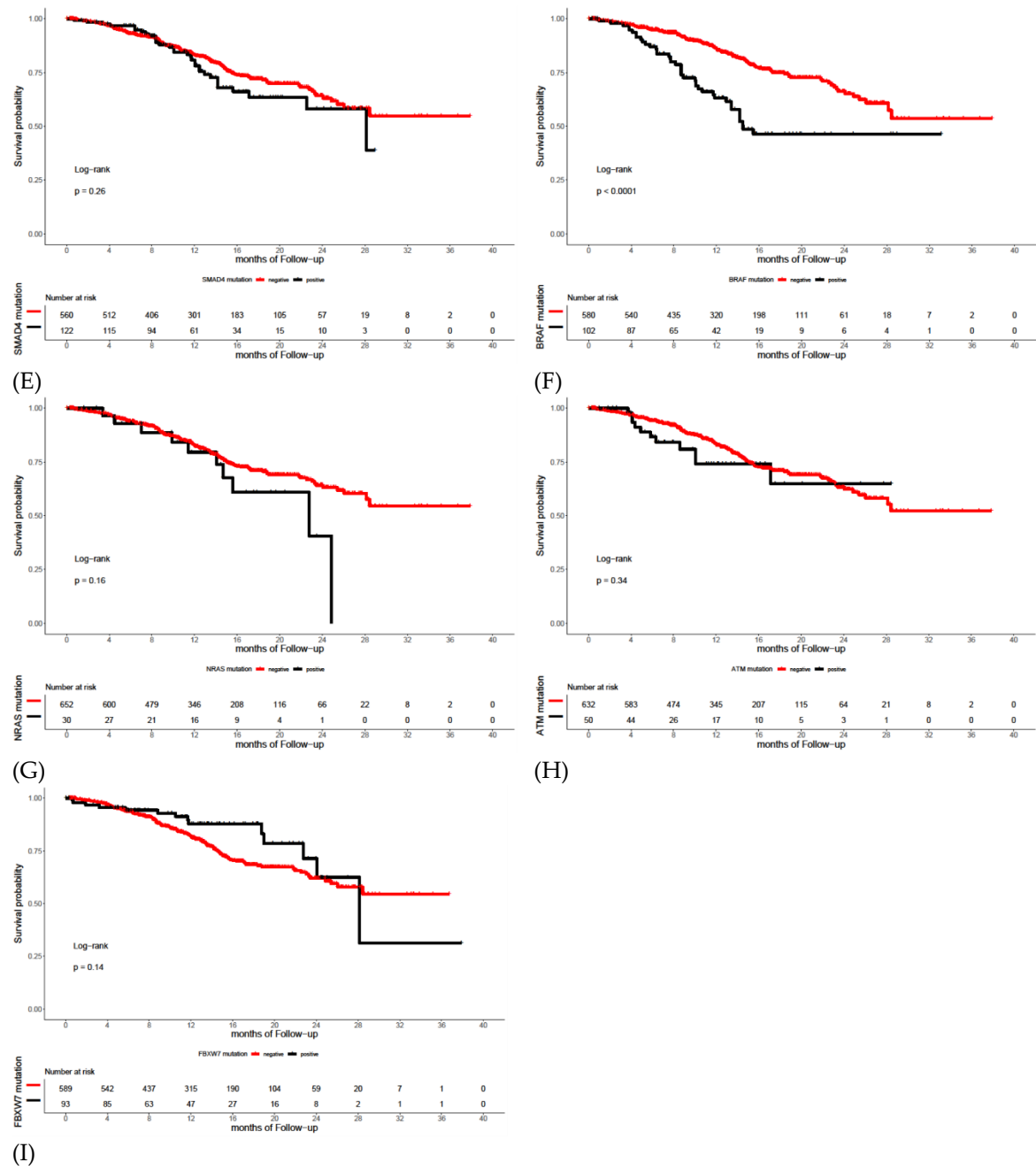


Figure A8. Kaplan-Meier curve for OS in the MSKCC dataset by mutational status, including TP53 (A), KRAS (B), APC (C), PIK3CA (D), SMAD4 (E), BRAF (F), NRAS (G), ATM (H), and FBXW7 (I)

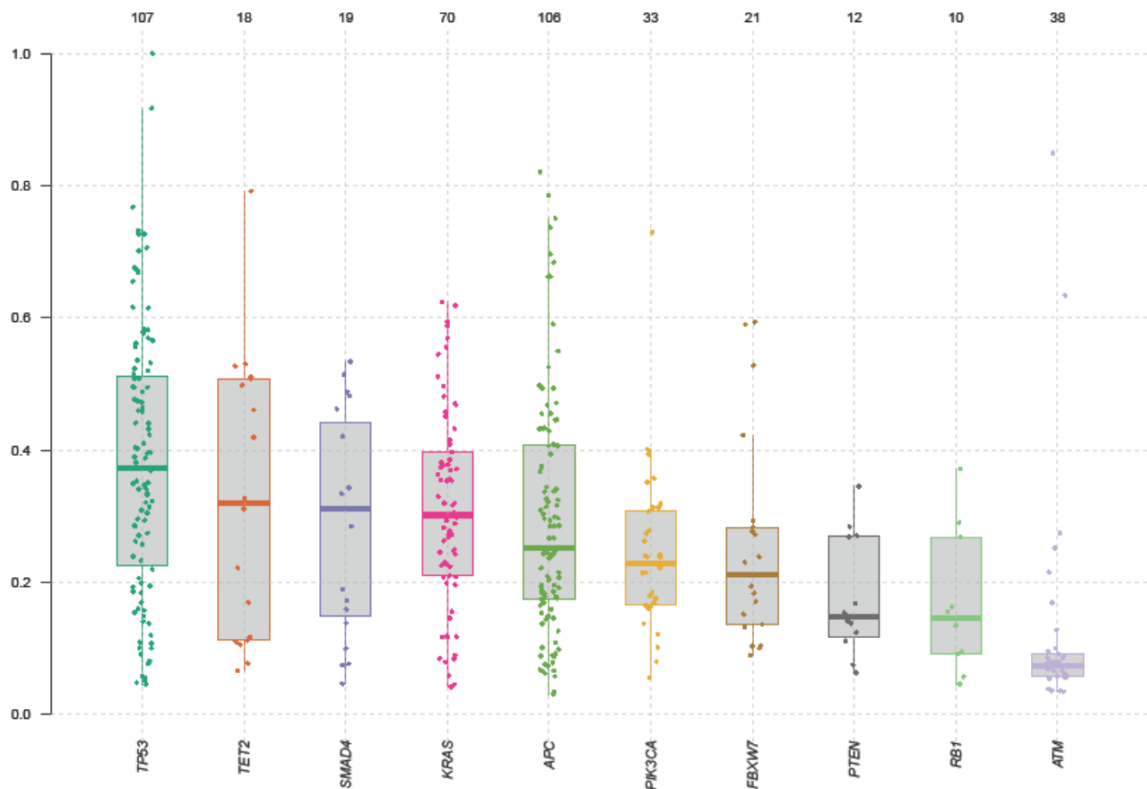


Figure A9. Distribution of VAFs (Y-axis) of SNPs (X-axis) among the top mutated genes in our study dataset. The numbers on upper part of the figure correspond to mutated patients.

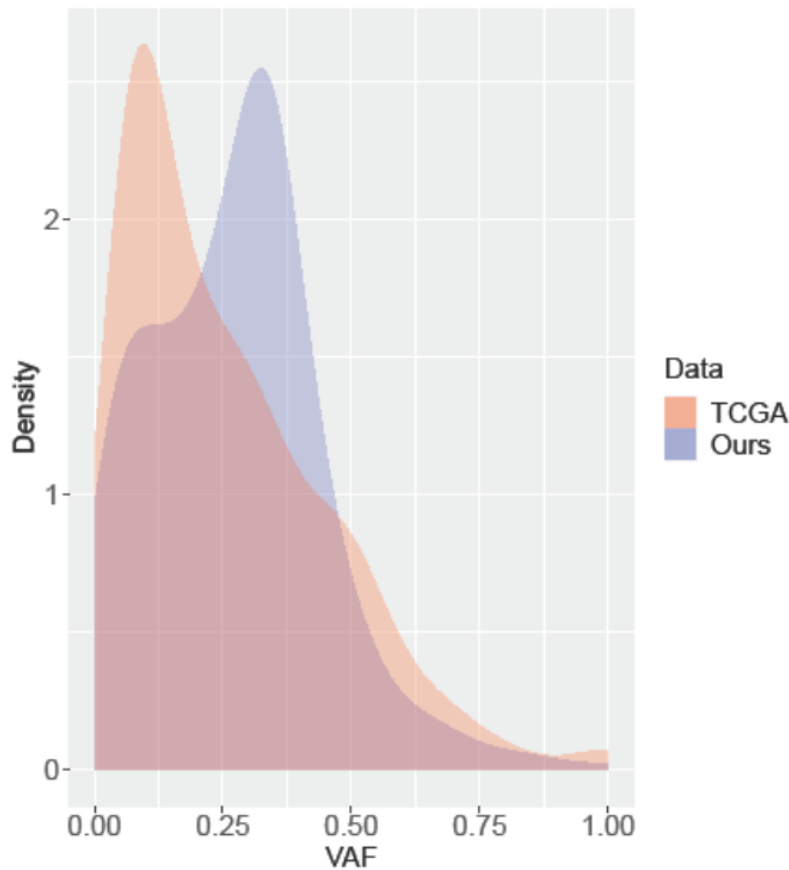
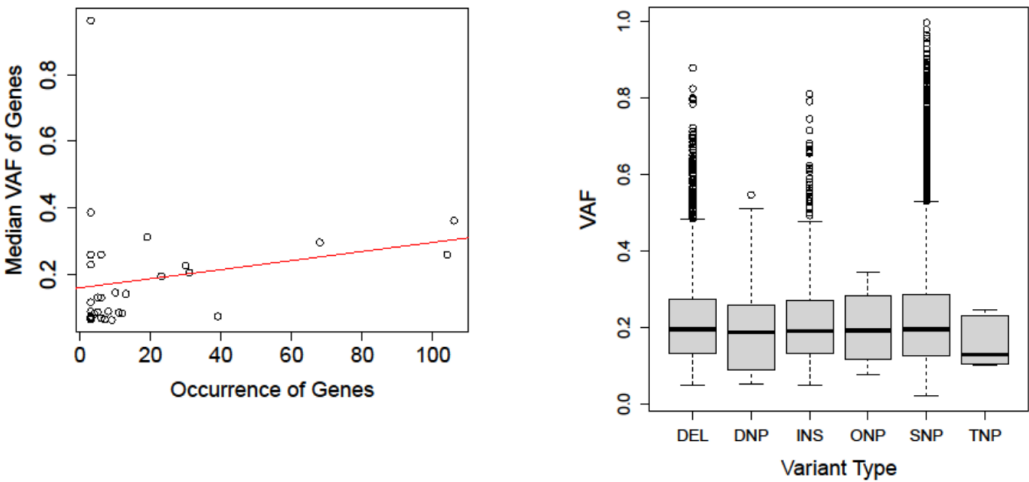


Figure A10. Comparison of the distribution of VAF for validated somatic mutations between the TCGA cohort and our cohort. The density is estimated by Gaussian kernel.

542



(A) (B)

543

544

545

546

547

548

549

Figure A11. Scatterplot representing the relationships between median VAF of genes and occurrence of corresponding genes (A) and boxplot of VAF according to mutation types in the MSKCC cohort (B). Genes with less than three instances of occurrences are excluded. The red line indicates the results of linear regression. SNP: Single nucleotide polymorphism; DNP: Double nucleotide polymorphism; TNP: Triple nucleotide polymorphism; ONP: Oligo-nucleotide polymorphism; INS: Insertion; DEL: Deletion

550

Appendix B

551

Table A1. Mutation spectrum of driver genes in the TCGA cohort

Gene	Case (Variants)	Variation
APC	160 (251 ^a)	p.1219_1219del (1), p.1275_1275del (1), p.1289_1291del (1), p.1291_1291del (1), p.1293_1293del (1), p.A1089V (1), p.A2028T (1), p.A209V (1), p.A2D (1), p.A513V (1), p.C1252X (1), p.C1392S (1), p.C1392X (1), p.C507X (1), p.D2501G (1), p.D756G (2), p.E1191X (2), p.E1239X (1), p.E1268X (1), p.E1277fs (1), p.E1277X (1), p.E1288X (3), p.E1291X (2), p.E1304X (1), p.E1335X (1), p.E1356X (3), p.E1361X (1), p.E1379X (1), p.E1390X (2), p.E146K (2), p.E1518X (2), p.E1555fs (1), p.E1560fs (1), p.E190fs (1), p.E200X (1), p.E829fs (2), p.E923X (1), p.E966X (1), p.F1336fs (2), p.F2766C (1), p.G1270X (1), p.G1312fs (1), p.G549X (1), p.H1472fs (1), p.H289fs (1), p.I1236N (1), p.I1286fs (1), p.I1644I (1), p.I2497V (1), p.K1067X (1), p.K1352X (1), p.K2174T (1), p.K652X (2), p.K921X (1), p.K975X (1), p.L1471X (1), p.L235X (1), p.L478fs (1), p.M1395fs (1), p.N1143K (1), p.P1301fs (1), p.P1351P (1), p.P1453fs (1), p.P1742R (1), p.P2338H (1), p.P2743L (2), p.Q1017X (1), p.Q1023E (1), p.Q1134X (1), p.Q1173X (1), p.Q1210X (1), p.Q1226X (1), p.Q1242X (1), p.Q1244fs (1), p.Q1276X (3), p.Q1285X (1), p.Q1310X (1), p.Q1320X (2), p.Q1349X (4), p.Q1360fs (1), p.Q1360X (3), p.Q1378X (1), p.Q1388X (4), p.Q1411X (1), p.Q1451X (1), p.Q246X (1), p.Q394X (1), p.Q462X (1), p.Q524X (1), p.Q636X (1), p.Q649X (1), p.Q739X (1), p.Q883X (1), p.Q960X (2), p.R1096X (6), p.R1381H (1), p.R1417fs (1), p.R1432X (19), p.R1450X (1), p.R216X (7), p.R223X (5), p.R2311I (1), p.R232X (4), p.R2413K (1), p.R265X (2), p.R284X (3), p.R330X (1), p.R387X (1), p.R481X (2), p.R499X (1), p.R536X (2), p.R546X (6), p.R554X (1), p.R640W (1), p.R787X (6), p.R827H (1), p.R858X (11), p.R906X (1), p.S1145fs (1), p.S1180X (1), p.S1263X (1),

		p.S1264X (1), p.S1297X (2), p.S1328X (4), p.S1380fs (1), p.S1382L (2), p.S1382X (1), p.S1393fs (1), p.S1401N (1), p.S1403fs (2), p.S2011Y (1), p.S2411F (1), p.S2464Y (1), p.S569S (1), p.S578X (1), p.S695X (1), p.S752X (1), p.S836fs (1), p.S945fs (1), p.S960fs (1), p.T1283fs (1), p.T1380fs (1), p.T1420fs (1), p.T1469fs (1), p.T646T (1), p.T916fs (1), p.V1334fs (1), p.V291fs (1), p.V539A (1), p.V579fs (1), p.V812fs (1), p.Y1165X (1), p.Y1358X (1), p.Y917X (2), p.Y935_N936delinsX (1), p.Y938C (1)
TP53	122 (126)	p.122_122del (1), p.173_173del (1), p.177_183del (1), p.212_212del (1), p.289_291del (1), p.98_99del (1), p.A138V (1), p.A159V (1), p.C124G (1), p.C135S (1), p.C141R (1), p.C141S (1), p.C141Y (1), p.C176F (1), p.C176X (1), p.C238Y (1), p.C275Y (1), p.D207N (1), p.E154K (1), p.E258X (1), p.E285X (1), p.E286G (1), p.F109C (1), p.F113C (1), p.G154C (1), p.G244D (2), p.G245D (1), p.G245S (4), p.G266X (1), p.H178fs (1), p.I195T (1), p.I251F (1), p.I63fs (1), p.K132N (1), p.K132T (1), p.K187fs (1), p.L130F (1), p.L194H (1), p.L35fs (1), p.M237I (1), p.P151H (1), p.P152R (1), p.P278A (1), p.P278R (1), p.P27fs (1), p.P322fs (1), p.P60fs (1), p.Q199fs (1), p.Q331H (1), p.Q331X (1), p.R175C (2), p.R175H (15), p.R196X (3), p.R213X (7), p.R248Q (1), p.R248W (9), p.R267W (1), p.R273C (4), p.R273H (7), p.R282W (3), p.R306X (4), p.R335fs (2), p.R337C (1), p.R342X (1), p.S127P (1), p.S33fs (1), p.T125M (2), p.T125T (1), p.T231T (1), p.V143A (1), p.V173M (1), p.V274L (2), p.Y205C (1), p.Y236X (1)
KRAS	94 (95 ^b)	p.A146T (8), p.A146V (1), p.E98X (1), p.G12A (1), p.G12C (6), p.G12D (31), p.G12R (1), p.G12S (4), p.G12V (23), p.G13D (10), p.K117N (2), p.Q22K (1), p.Q61L (2), p.R68S (1)
FBXW7	39 (50)	p.321_322del (1), p.A235A (1), p.D440N (1), p.D440Y (1), p.D520Y (2), p.E112A (1), p.E113D (1), p.E207V (1), p.G357R (1), p.H415P (1), p.K159Q (2), p.K167T (2), p.P4P (1), p.Q501X (1), p.R13X (1), p.R14Q (2), p.R259X (1), p.R287X (2), p.R385C (4), p.R385H (8), p.R393fs (1), p.R399X (1), p.R425C (3), p.R425G (1), p.R425H (1), p.R578Q (1), p.R578X (2), p.R609W (1), p.R99H (1), p.S396R (1), p.S502L (2), p.T144T (1)
PIK3CA	34 (42)	p.D258N (1), p.D350G (1), p.D925D (1), p.E542K (2), p.E545A (1), p.E545G (1), p.E545K (8), p.E545Q (1), p.E81K (1), p.G106R (1), p.G364R (1), p.H1047R (5), p.H1047R (4), p.M1043I (1), p.N345K (2), p.P471L (1), p.Q546K (1), p.R357Q (1), p.R777M (1), p.R88Q (4), p.R951C (1), p.T1052K (1), p.V344A (1), p.V344G (1), p.V955V (1), p.Y1021C (1)
SMAD4	26 (31)	p.30_30del (1), p.36_45del (1), p.A327V (1), p.C324R (1), p.D355G (1), p.D537G (1), p.D537V (1), p.D537Y (1), p.E134K (1), p.E33X (1), p.E520X (1), p.G510R (1), p.G65E (1), p.L104F (1), p.L533R (1), p.L57V (1), p.P356L (1), p.R361H (8), p.R496H (1), p.R97H (1), p.S357P (1), p.V136A (1), p.V437D (1), p.Y513X (1)
ATM	25 (50)	p.2693_2694del (1), p.A1024T (1), p.A920V (1), p.C1168C (1), p.D2721N (1), p.E1666X (1), p.E1822X (1), p.E2039K (1), p.E2676X (1), p.F1837fs (1), p.F2140V (1), p.F2839L (1), p.G2425fs (1), p.G2891R (1), p.H2038H (1), p.H2872R (1), p.K477N (1), p.L1217M (2), p.L1408I (1), p.L1488L (2), p.L2077I (1), p.L2147I (2), p.L2251I (2), p.L2557V (1), p.L822V (1), p.L942I (2), p.N2435K (1), p.P2353T (1), p.Q1117P (1), p.Q499E (1), p.R1150I (2), p.R1466X (2), p.R1489H (1), p.R1730X (3), p.R2060H (1), p.R250X (2), p.R2598Q (1), p.S614N (1), p.T2771I (1), p.W2769X (1)
BRAF	21 (22)	p.F247L (1), p.K205Q (1), p.V600E (20)
NRAS	20 (21)	p.E132K (1), p.G12A (1), p.G12C (3), p.G12D (2), p.G13R (2), p.Q61H (1), p.Q61K (7), p.Q61L (2), p.Q61R (1), p.R164C (1)

^bThree patients have no data on amino acid changes.

Each SNV was classified according to the ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar/>).

Table A2. Mutation spectrum of driver genes in the MSKCC cohort

Gene	Case (variants)	Variation (n ≥ 5)
APC	737 (1203)	p.R876* (44), p.R1450* (41), p.R216* (33), p.T1556Nfs*3 (31), p.R213* (30), p.E1309Dfs*4 (29), p.R232* (24), p.R283* (21), p.S1465Wfs*3 (21), p.Q1367* (16), p.Q1429* (16), p.R1114* (15), p.R554* (15), p.R499* (14), p.R805* (14), p.Y935* (14), p.E1322* (12), p.E1353* (11), p.Q1406* (11), p.R564* (11), p.E1309* (10), p.E1397* (10), p.Q1338* (10), p.R302* (10), p.E1306* (9), p.E1379* (9), p.P1319Lfs*2 (8), p.Q1378* (8), p.E1345* (7), p.L1488Yfs*19 (7), p.Q1303* (7), p.I1307Nfs*8 (6), p.K534* (6), p.S1315* (6), p.S1465Rfs*9 (6), p.E1408* (5), p.E1464Vfs*8 (5), p.E941* (5), p.S1346* (5)
TP53	710 (780)	p.R175H (67), p.R273H (40), p.R248Q (35), p.R282W (34), p.R273C (32), p.G245S (28), p.R213* (28), p.R248W (22), p.R342* (18), p.R196* (16), p.R306* (9), p.P152L (8), p.G266E (7), p.M237I (7), p.V173L (7), p.X307_splice (7), p.P151S (6), p.R158H (6), p.R209Kfs*6 (6), p.X331_splice (6), p.C176F (5), p.R337C (5), p.S127F (5), p.S241F (5), p.T125M (5), p.X126_splice (5)
KRAS	432 (447)	p.G12D (122), p.G12V (87), p.G13D (77), p.G12C (31), p.A146T (29), p.G12A (22), p.G12S (19), p.Q61H (9), p.G12R (7), p.K117N (7), p.A146V (6), p.Q61K (5)
PIK3CA	190 (219)	p.E545K (52), p.E542K (29), p.H1047R (28), p.Q546K (9)
SMAD4	153 (170)	p.R361H (19), p.R361C (13), p.D537G (7), p.A118V (5), p.E520* (5)
FBXW7	114 (130)	p.R465C (13), p.R465H (13), p.R505C (13), p.S668Vfs*39 (6), p.R278* (5)
BRAF	107 (128)	p.V600E (76), p.D594G (6)
ATM	73 (101)	p.R337C (7), p.R3008C (5)
NRAS	44 (44)	p.Q61K (9), p.G12D (7)

Each SNV was classified according to the ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar/>).

Table A3. Correlation of common gene mutations with tumor stages in TCGA cohort

Mutations	T stage ^a		N stage ^b		M stage ^c		Stage ^d	
	1, 2	3, 4	0	1, 2	0	1	1, 2	3, 4
KRAS	p = 0.7224		p = 0.8708		p = 0.6389		p = 0.5531	
Negative	55	202	143	113	184	31	134	119
Positive	39	160	113	85	145	29	108	84
NRAS	p = 0.3883		p = 0.8406		p = 0.7946		p = 1	
Negative	85	339	239	183	304	55	225	188
Positive	9	23	17	15	25	5	17	15
BRAF	p = 0.6445		p = 0.00398		p = 0.05844		p = 0.00195	
Negative	80	317	212	183	281	57	199	188
Positive	14	45	44	15	48	3	43	15
PIK3CA	p = 0.6271		p = 0.0217		p = 0.1388		p = 0.00253	
Negative	70	258	173	154	229	48	161	162
Positive	24	104	83	44	100	12	81	41
TP53	p = 0.06349		p = 0.00025		p = 0.0747		p = 0.00016	
Negative	42	122	111	52	125	15	106	53
Positive	52	240	145	146	204	45	136	150
APC	p = 0.6315		p = 0.3294		p = 0.4598		p = 0.2612	
Negative	18	80	60	38	72	10	57	38

Positive	76	282	196	160	257	50	185	165
SMAD4	p = 0.3371		p = 0.09645		p = 1		p = 0.243	
Negative	84	307	226	163	282	51	212	169
Positive	10	55	30	35	47	9	30	34
FBXW7	p = 0.2788		p = 0.25		p = 0.07132		p = 0.2284	
Negative	81	292	204	167	260	54	192	171
Positive	13	70	52	31	69	6	50	32
ATM	p = 0.8081		p = 0.2109		p = 0.00766		p = 0.2312	
Negative	80	314	216	176	276	58	204	180
Positive	14	48	40	22	53	2	38	23

^aThree patients have no available data.

^bFive patients have no available data.

^cA total of 70 patients have no available data.

^dA total of 14 patients have no available data.

Table A4. Classification of genes according to median VAF

Low VAF genes	High VAF genes
BRCA1, CDKN2A, BAP1, TSC2, TSC1, POLE, PALB2, ATM, SMARCA4, PIK3R1, NOTCH1, NF1, NF2, STK11, MSH2, LOC100507346,PTCH1	PTCH1, VHL, CDH1, PTEN, RB1, FBXW7, BRCA2, PIK3CA, AKT1, ARID1A, BRAF, APC, KRAS, TET2, TET2-AS1, SMAD4, TP53, ERBB2, ATRX

Table A5. Correlation of MSI status according to the clinicopathological data and KRAS, NRAS, PIK3CA, and BRAF mutations in the TCGA cohort

Clinicopathological parameters	MSI status ^a	
	MSS/MSI-L	MSI-H
Age^b	p = 0.1627	
≥60 years	262	50
<60 years	127	15
Sex	p = 0.3295	
Male	204	29
Female	188	36
T stage^c	p = 1	
1, 2	81	13
3, 4	308	52
N stage^d	p = 5.222e-05	
0	202	52
1, 2	185	13
M stage^e	p = 0.05575	
0	279	50
1	57	3
Stage^f	p = 7.857e-06	
1, 2	189	51
3, 4	191	12
Tumor sites	p = 4.554e-14	
Right colon	125	51
Left colon	239	7
KRAS mutation	p = 0.03194	
Absent	212	45
Present	180	20
NRAS mutation	p = 0.6002	

Absent	363	62
Present	29	3
PIK3CA mutation	p = 0.03334	
Absent	289	39
Present	103	26
BRAF mutation	p < 2.2e-16	
Absent	371	27
Present	21	38

^aTwo cases have no available data.
^bThree cases have no available data.
^cThree cases have no available data.
^dFive cases have no available data.
^eA total of 70 cases have no available data.
^fA total of 14 cases have no available data.
^gThree cases have no available data.

References

1. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2019. *CA Cancer J Clin* **2019**, *69*, 7–34.

2. Jung, K.W.; Won, Y.J.; Kong, H.J.; Lee, E.S. Prediction of cancer incidence and mortality in Korea, 2019. *Cancer Res Treat* **2019**, *51*, 431–437.

3. Mitry, E.; Guiu, B.; Coscinea, S.; Jooste, V.; Faivre, J.; Bouvier, A. M. Epidemiology, management and prognosis of colorectal cancer with lung metastases: a 30-year population-based study. *Gut* **2010**, *59*, 1383–1388.

4. Sepulveda, A.R.; Hamilton, S.R.; Allegra, C.J.; Grody, W.; Cushman-Vokoun, A.M.; Funkhouser, W.K.; Kopetz, S.E.; Lieu, C.; Lindor, N.M.; Minsky, B.D.; Monzon, F.A.; Sargent, D.J.; Singh, V.M.; Willis, J.; Clark, J.; Colasacco, C.; Bryan Rumble, R.; Temple-Smolkin, R.; Ventura, C.B.; Nowak, J.A. Molecular biomarkers for the evaluation of colorectal cancer: Guideline From the American Society for Clinical Pathology, College of American Pathologists, Association for Molecular Pathology, and American Society of Clinical Oncology. *Arch Pathol Lab Med* **2017**, *141*, 625–657.

5. Bertotti, A.; Papp, E.; Jones, S.; Adleff, V.; Anagnostou, V.; Lupo, B.; Sausen, M.; Phallen, J.; Hruban, C.A.; Tokheim, C.; Niknafs, N.; Nesselbush, M.; Lytle, K.; Sassi, F.; Cottino, F.; Migliardi, G.; Zanella, E.R.; Ribero, D.; Russolillo, N.; Mellano, A.; Muratore, A.; Paraluppi, G.; Salizzoni, M.; Marsoni, S.; Kragh, M.; Lantto, J.; Cassingena, A.; Li, Q.K.; Karchin, R.; Scharpf, R.; Sartore-Bianchi, A.; Siena, S.; Diaz, L.A., Jr.; Trusolino, L.; Velculescu, V.E. The genomic landscape of response to EGFR blockade in colorectal cancer. *Nature* **2015**, *526*, 263–267.

6. Koveitypour, Z.; Panahi, F.; Vakilian, M.; Peymani, M.; Seyed Forootan, F.; Nasr Esfahani, M.H.; Ghaedi, K. Signaling pathways involved in colorectal cancer progression. *Cell Biosci* **2019**, *9*, 97.

7. Luchini, C.; Bibeau, F.; Ligtenberg, M.J.L.; Singh, N.; Nottegar, A.; Bosse, T.; Miller, R.; Riaz, N.; Douillard, J.Y.; Andre, F.; Scarpa, A. ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: a systematic review-based approach. *Ann Oncol* **2019**, *30*, 1232–1243.

8. Deans, Z.C.; Costa, J.L.; Cree, I.; Dequeker, E.; Edsjo, A.; Henderson, S.; Hummel, M.; Ligtenberg, M.J.; Loddo, M.; Machado, J.C.; Marchetti, A.; Marquis, K.; Mason, J.; Normanno, N.; Rouleau, E.; Schuuring,

- E.; Snelson, K.M.; Thunnissen, E.; Tops, B.; Williams, G.; van Krieken, H.; Hall, J.A.; ASBL, I.Q.N.P. Integration of next-generation sequencing in clinical diagnostic molecular pathology laboratories for analysis of solid tumours; an expert opinion on behalf of IQN Path ASBL. *Virchows Arch* **2017**, *470*, 5–20.
9. Lee, A.; Lee, S.H.; Jung, C.K.; Park, G.; Lee, K.Y.; Choi, H.J.; Min, K.O.; Kim, T.J.; Lee, E.J.; Lee, Y.S. Use of the Ion AmpliSeq Cancer Hotspot Panel in clinical molecular pathology laboratories for analysis of solid tumours: With emphasis on validation with relevant single molecular pathology tests and the Oncomine Focus Assay. *Pathol Res Pract* **2018**, *214*, 713–719.
10. Cancer Genome Atlas, N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **2012**, *487*, 330–337.
11. Zehir, A.; Benayed, R.; Shah, R.H.; Syed, A.; Middha, S.; Kim, H.R.; Srinivasan, P.; Gao, J.; Chakravarty, D.; Devlin, S.M.; Hellmann, M.D.; Barron, D.A.; Schram, A.M.; Hameed, M.; Dogan, S.; Ross, D.S.; Hechtman, J.F.; DeLair, D.F.; Yao, J.; Mandelker, D.L.; Cheng, D.T.; Chandramohan, R.; Mohanty, A.S.; Ptashkin, R.N.; Jayakumaran, G.; Prasad, M.; Syed, M.H.; Rema, A.B.; Liu, Z.Y.; Nafa, K.; Borsu, L.; Sadowska, J.; Casanova, J.; Bacares, R.; Kiecka, I.J.; Razumova, A.; Son, J.B.; Stewart, L.; Baldi, T.; Mullaney, K.A.; Al-Ahmadie, H.; Vakiani, E.; Abeshouse, A.A.; Penson, A.V.; Jonsson, P.; Camacho, N.; Chang, M.T.; Won, H.H.; Gross, B.E.; Kundra, R.; Heins, Z.J.; Chen, H.W.; Phillips, S.; Zhang, H.; Wang, J.; Ochoa, A.; Wills, J.; Eubank, M.; Thomas, S.B.; Gardos, S.M.; Reales, D.N.; Galle, J.; Durany, R.; Cambria, R.; Abida, W.; Cercek, A.; Feldman, D.R.; Gounder, M.M.; Hakimi, A.A.; Harding, J.J.; Iyer, G.; Janjigian, Y.Y.; Jordan, E.J.; Kelly, C.M.; Lowery, M.A.; Morris, L.G.T.; Omuro, A.M.; Raj, N.; Razavi, P.; Shoushtari, A.N.; Shukla, N.; Soumerai, T.E.; Varghese, A.M.; Yaeger, R.; Coleman, J.; Bochner, B.; Riely, G.J.; Saltz, L.B.; Scher, H.I.; Sabbatini, P.J.; Robson, M.E.; Klimstra, D.S.; Taylor, B.S.; Baselga, J.; Schultz, N.; Hyman, D.M.; Arcila, M.E.; Solit, D.B.; Ladanyi, M.; Berger, M.F. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* **2017**, *23*, 703–713.
12. Nakayama, M.; Oshima, M. Mutant p53 in colon cancer. *J Mol Cell Biol* **2019**, *11*, 267–276.
13. Aghabozorgi, A.S.; Bahreyni, A.; Soleimani, A.; Bahrami, A.; Khazaei, M.; Ferns, G.A.; Avan, A.; Hassanian, S.M. Role of adenomatous polyposis coli (APC) gene mutations in the pathogenesis of colorectal cancer; current status and perspectives. *Biochimie* **2019**, *157*, 64–71.
14. Fleming, N.I.; Jorissen, R.N.; Mouradov, D.; Christie, M.; Sakthianandeswaren, A.; Palmieri, M.; Day, F.; Li, S.; Tsui, C.; Lipton, L.; Desai, J.; Jones, I.T.; McLaughlin, S.; Ward, R.L.; Hawkins, N.J.; Ruskiewicz, A.R.; Moore, J.; Zhu, H.J.; Mariadason, J.M.; Burgess, A.W.; Busam, D.; Zhao, Q.; Strausberg, R.L.; Gibbs, P.; Sieber, O.M. SMAD2, SMAD3 and SMAD4 mutations in colorectal cancer. *Cancer Res* **2013**, *73*, 725–735.
15. Valentino, E.; Bellazzo, A.; Di Minin, G.; Sicari, D.; Apollonio, M.; Scognamiglio, G.; Di Bonito, M.; Botti, G.; Del Sal, G.; Collavin, L. Mutant p53 potentiates the oncogenic effects of insulin by inhibiting the tumor suppressor DAB2IP. *Proc Natl Acad Sci U S A* **2017**, *114*, 7623–7628.
16. Shin, H.T.; Choi, Y.L.; Yun, J.W.; Kim, N.K.D.; Kim, S.Y.; Jeon, H.J.; Nam, J.Y.; Lee, C.; Ryu, D.; Kim, S.C.; Park, K.; Lee, E.; Bae, J.S.; Son, D.S.; Joung, J.G.; Lee, J.; Kim, S.T.; Ahn, M.J.; Lee, S.H.; Ahn, J.S.; Lee, W.Y.; Oh, B.Y.; Park, Y.H.; Lee, J.E.; Lee, K.H.; Kim, H.C.; Kim, K.M.; Im, Y.H.; Park, K.; Park, P.J.; Park, W.Y. Prevalence and detection of low-allele-fraction variants in clinical cancer samples. *Nat Commun* **2017**, *8*, 1377.

651 17. Dinu, D.; Dobre, M.; Panaitescu, E.; Birla, R.; Iosif, C.; Hoara, P.; Caragui, A.; Boeriu, M.; Constantinoiu,
652 S.; Ardeleanu, C. Prognostic significance of KRAS gene mutations in colorectal cancer--preliminary
653 study. *J Med Life* **2014**, *7*, 581–587.

654 18. Al-Shamsi, H.O.; Jones, J.; Fahmawi, Y.; Dahbour, I.; Tabash, A.; Abdel-Wahab, R.; Abousamra, A.O.;
655 Shaw, K.R.; Xiao, L.; Hassan, M.M.; Kipp, B.R.; Kopetz, S.; Soliman, A.S.; McWilliams, R.R.; Wolff, R.A.
656 Molecular spectrum of KRAS, NRAS, BRAF, PIK3CA, TP53, and APC somatic gene mutations in Arab
657 patients with colorectal cancer: determination of frequency and distribution pattern. *J Gastrointest Oncol*
658 **2016**, *7*, 882–902.

659 19. Heinemann, V.; Stintzing, S.; Kirchner, T.; Boeck, S.; Jung, A. Clinical relevance of EGFR- and KRAS-
660 status in colorectal cancer patients treated with monoclonal antibodies directed against the EGFR.
661 *Cancer Treat Rev* **2009**, *35*, 262–271.

662 20. Zhao, B.; Wang, L.; Qiu, H.; Zhang, M.; Sun, L.; Peng, P.; Yu, Q.; Yuan, X. Mechanisms of resistance to
663 anti-EGFR therapy in colorectal cancer. *Oncotarget* **2017**, *8*, 3980–4000.

664 21. Andreyev, H.J.; Norman, A.R.; Cunningham, D.; Oates, J.; Dix, B.R.; Iacopetta, B.J.; Young, J.; Walsh, T.;
665 Ward, R.; Hawkins, N.; Beranek, M.; Jandik, P.; Benamouzig, R.; Jullian, E.; Laurent-Puig, P.; Olschwang,
666 S.; Muller, O.; Hoffmann, I.; Rabes, H.M.; Zietz, C.; Troungos, C.; Valavanis, C.; Yuen, S.T.; Ho, J.W.;
667 Croke, C.T.; O'Donoghue, D.P.; Giaretti, W.; Rapallo, A.; Russo, A.; Bazan, V.; Tanaka, M.; Omura, K.;
668 Azuma, T.; Ohkusa, T.; Fujimori, T.; Ono, Y.; Pauly, M.; Faber, C.; Glaesener, R.; de Goeij, A.F.; Arends,
669 J.W.; Andersen, S.N.; Lovig, T.; Breivik, J.; Gaudernack, G.; Clausen, O.P.; De Angelis, P.D.; Meling, G.I.;
670 Rognum, T.O.; Smith, R.; Goh, H.S.; Font, A.; Rosell, R.; Sun, X.F.; Zhang, H.; Benhattar, J.; Losi, L.; Lee,
671 J.Q.; Wang, S.T.; Clarke, P.A.; Bell, S.; Quirke, P.; Bubb, V.J.; Piris, J.; Cruickshank, N.R.; Morton, D.; Fox,
672 J.C.; Al-Mulla, F.; Lees, N.; Hall, C.N.; Snary, D.; Wilkinson, K.; Dillon, D.; Costa, J.; Pricolo, V.E.;
673 Finkelstein, S.D.; Thebo, J.S.; Senagore, A.J.; Halter, S.A.; Wadler, S.; Malik, S.; Krtolica, K.; Urosevic, N.
674 Kirsten ras mutations in patients with colorectal cancer: the "RASCAL II" study. *Br J Cancer* **2001**, *85*,
675 692–696.

676 22. Umeda, Y.; Nagasaka, T.; Mori, Y.; Sadamori, H.; Sun, D.S.; Shinoura, S.; Yoshida, R.; Satoh, D.; Nobuoka,
677 D.; Utsumi, M.; Yoshida, K.; Yagi, T.; Fujiwara, T. Poor prognosis of KRAS or BRAF mutant colorectal
678 liver metastasis without microsatellite instability. *J Hepatobiliary Pancreat Sci* **2013**, *20*, 223–233.

679 23. Chang, Y.Y.; Lin, P.C.; Lin, H.H.; Lin, J.K.; Chen, W.S.; Jiang, J.K.; Yang, S.H.; Liang, W.Y.; Chang, S.C.
680 Mutation spectra of RAS gene family in colorectal cancer. *Am J Surg* **2016**, *212*, 537–544 e3.

681 24. Taieb, J.; Le Malicot, K.; Shi, Q.; Penault-Llorca, F.; Bouche, O.; Tabernero, J.; Mini, E.; Goldberg, R.M.;
682 Folprecht, G.; Luc Van Laethem, J.; Sargent, D.J.; Alberts, S.R.; Emile, J.F.; Laurent Puig, P.; Sinicrope,
683 F.A. Prognostic value of BRAF and KRAS Mutations in MSI and MSS stage III colon cancer. *J Natl Cancer*
684 *Inst* **2017**, *109*, 272.

685 25. Shiloh, Y. ATM: expanding roles as a chief guardian of genome stability. *Exp Cell Res* **2014**, *329*, 154–161.

686 26. Randon, G.; Fuca, G.; Rossini, D.; Raimondi, A.; Pagani, F.; Perrone, F.; Tamborini, E.; Busico, A.;
687 Peverelli, G.; Morano, F.; Niger, M.; Antista, M.; Corallo, S.; Saggio, S.; Borelli, B.; Zucchelli, G.; Milione,
688 M.; Pruneri, G.; Di Bartolomeo, M.; Falcone, A.; de Braud, F.; Cremolini, C.; Pietrantonio, F. Prognostic
689 impact of ATM mutations in patients with metastatic colorectal cancer. *Sci Rep* **2019**, *9*, 2858.

690 27. Beggs, A.D.; Domingo, E.; McGregor, M.; Presz, M.; Johnstone, E.; Midgley, R.; Kerr, D.; Oukrif, D.;
691 Novelli, M.; Abulafi, M.; Hodgson, S.V.; Fadhil, W.; Ilyas, M.; Tomlinson, I.P. Loss of expression of the
692 double strand break repair protein ATM is associated with worse prognosis in colorectal cancer and
693 loss of Ku70 expression is associated with CIN. *Oncotarget* **2012**, *3*, 1348–1355.

28. Grim, J.E. Fbxw7 hotspot mutations and human colon cancer: mechanistic insights from new mouse models. *Gut* **2014**, *63*, 707–709.

29. Sailo, B.L.; Banik, K.; Girisa, S.; Bordoloi, D.; Fan, L.; Halim, C.E.; Wang, H.; Kumar, A.P.; Zheng, D.; Mao, X.; Sethi, G.; Kunnumakkara, A.B. FBXW7 in cancer: what has been unraveled thus far? *Cancers (Basel)* **2019**, *11*, 246.

30. Korphaisarn, K.; Morris, V.K.; Overman, M.J.; Fogelman, D.R.; Kee, B.K.; Raghav, K.P.S.; Manuel, S.; Shureiqi, I.; Wolff, R.A.; Eng, C.; Menter, D.; Hamilton, S.R.; Kopetz, S.; Dasari, A. FBXW7 missense mutation: a novel negative prognostic factor in metastatic colorectal adenocarcinoma. *Oncotarget* **2017**, *8*, 39268–39279.

31. Gafa, R.; Maestri, I.; Matteuzzi, M.; Santini, A.; Ferretti, S.; Cavazzini, L.; Lanza, G. Sporadic colorectal adenocarcinomas with high-frequency microsatellite instability. *Cancer* **2000**, *89*, 2025–2037.

32. Elsaleh, H.; Shannon, B.; Iacopetta, B. Microsatellite instability as a molecular marker for very good survival in colorectal cancer patients receiving adjuvant chemotherapy. *Gastroenterology* **2001**, *120*, 1309–1310.

33. Hu, J.; Yan, W.Y.; Xie, L.; Cheng, L.; Yang, M.; Li, L.; Shi, J.; Liu, B.R.; Qian, X.P. Coexistence of MSI with KRAS mutation is associated with worse prognosis in colorectal cancer. *Medicine (Baltimore)* **2016**, *95*, e5649.

34. Koyel, B.; Priyabrata, D.; Rittwika, B.; Swati, D.; Soma, M.; Jayasri, B.; Ashis, M. Deterministic role of CEA and MSI status in predicting outcome of CRC patients: a perspective study amongst hospital attending Eastern Indian populations. *Indian J Surg Oncol* **2017**, *8*, 462–468.

35. Hata, A.N.; Niederst, M.J.; Archibald, H.L.; Gomez-Caraballo, M.; Siddiqui, F.M.; Mulvey, H.E.; Maruvka, Y.E.; Ji, F.; Bhang, H.E.; Krishnamurthy Radhakrishna, V.; Siravegna, G.; Hu, H.; Raoof, S.; Lockerman, E.; Kalsy, A.; Lee, D.; Keating, C.L.; Ruddy, D.A.; Damon, L.J.; Crystal, A.S.; Costa, C.; Piotrowska, Z.; Bardelli, A.; Iafrate, A.J.; Sadreyev, R.I.; Stegmeier, F.; Getz, G.; Sequist, L.V.; Faber, A.C.; Engelman, J.A. Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition. *Nat Med* **2016**, *22*, 262–269.

36. Robinson, J.T.; Thorvaldsdottir, H.; Wenger, A.M.; Zehir, A.; Mesirov, J.P. Variant review with the integrative genomics viewer. *Cancer Res* **2017**, *77*, e31–e34.

37. Berg, K.D.; Glaser, C.L.; Thompson, R.E.; Hamilton, S.R.; Griffin, C.A.; Eshleman, J.R. Detection of microsatellite instability by fluorescence multiplex polymerase chain reaction. *J Mol Diagn* **2000**, *2*, 20–28.

38. Alexander, J.; Watanabe, T.; Wu, T.T.; Rashid, A.; Li, S.; Hamilton, S.R. Histopathological identification of colon cancer with microsatellite instability. *Am J Pathol* **2001**, *158*, 527–535.

39. Liu, Y.; Sethi, N.S.; Hinoue, T.; Schneider, B.G.; Cherniack, A.D.; Sanchez-Vega, F.; Seoane, J.A.; Farshidfar, F.; Bowlby, R.; Islam, M.; Kim, J.; Chatila, W.; Akbani, R.; Kanchi, R.S.; Rabkin, C.S.; Willis, J.E.; Wang, K.K.; McCall, S.J.; Mishra, L.; Ojesina, A.I.; Bullman, S.; Pedamallu, C.S.; Lazar, A.J.; Sakai, R.; Cancer Genome Atlas Research, N.; Thorsson, V.; Bass, A.J.; Laird, P.W. Comparative molecular analysis of gastrointestinal adenocarcinomas. *Cancer Cell* **2018**, *33*, 721–735 e8.