Original Article

Ann Coloproctol 2022;38(1):72-81
https://doi.org/10.3393/ac.2021.00290.0041

Annals of Coloproctology

pISSN 2287-9714 eISSN 2287-9722
www.coloproctol.org

Molecular characterization of dysplasia-initiated colorectal cancer with assessing matched tumor and dysplasia samples

Sungwon Jung[1,2,*], Jong Lyul Lee[3,*], Tae Won Kim[4], Jongmin Lee[1], Yong Sik Yoon[3], Kil Yeon Lee[5], Ki-hwan Song[6], Chang Sik Yu[3], Yong Beom Cho[4,7]

[1]Department of Genome Medicine and Science, Gachon University College of Medicine, Incheon; [2]Gachon Institute of Genome Medicine and Science, Gachon University Gil Medical Center, Incheon; [3]Department of Surgery, Asan Medical Center, University of Ulsan College of Medicine, Seoul; [4]Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul; [5]Department of Surgery, Kyung Hee University College of Medicine, Seoul; [6]Department of Surgery, Koo Hospital, Daegu; [7]Department of Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

Purpose: Ulcerative colitis (UC) is known to have an association with the increased risk of colorectal cancer (CRC), and UC-associated CRC does not follow the typical progress pattern of adenoma-carcinoma. The aim of this study is to investigate molecular characteristics of UC-associated CRC and further our understanding of the association between UC and CRC.
Methods: From 5 patients with UC-associated CRC, matched normal, dysplasia, and tumor specimens were obtained from formalin-fixed paraffin-embedded (FFPE) samples for analysis. Genomic DNA was extracted and whole exome sequencing was conducted to identify somatic variations in dysplasia and tumor samples. Statistical analysis was performed to identify somatic variations with significantly higher frequencies in dysplasia-initiated tumors, and their relevant functions were investigated.
Results: Total of 104 tumor mutation genes were identified with higher mutation frequencies in dysplasia-initiated tumors. Four of the 5 dysplasia-initiated tumors (80.0%) have *TP53* mutations with frequent stop-gain mutations that were originated from matched dysplasia. *APC* and *KRAS* are known to be frequently mutated in general CRC, while none of the 5 patients have *APC* or *KRAS* mutation in their dysplasia and tumor samples. Glycoproteins including mucins were also frequently mutated in dysplasia-initiated tumors.
Conclusion: UC-associated CRC tumors have distinct mutational characteristics compared to typical adenoma-carcinoma tumors and may have different cancer-driving molecular mechanisms that are initiated from earlier dysplasia status.

Keywords: *Ulcerative colitis; Colorectal neoplasms; Colitis-Associated Neoplasms; Genetic variation*

Received: Apr 21, 2021 • Revised: Jun 29, 2021 • Accepted: Jul 20, 2021

Correspondence to: Yong Beom Cho, M.D.
Department of Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, 81 Irwon-ro, Gangnam-gu, Seoul 06351, Korea
Tel: +82-2-3410-1669, Fax: +82-2-3410-6980
E-mail: gscyb@skku.edu
ORCID: https://orcid.org/0000-0002-9944-4706

Co-Correspondence to: Chang Sik Yu, M.D.
Department of Surgery, Asan Medical Center, University of Ulsan College of Medicine, 88 Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Korea
Tel: +82-2-3010-3494, Fax: +82-2-474-9027
E-mail: csyu@amc.seoul.kr
ORCID: https://orcid.org/0000-0001-9401-9981

*Sungwon Jung and Jong Lyul Lee contributed equally as co-first authors.

INTRODUCTION

Ulcerative colitis (UC), a type of inflammatory bowel disease, is a chronic inflammation of the colon with an unknown etiology [1]. UC is an idiopathic disorder of the colonic mucosa which generally expanded proximally in a continuous manner through part of, or the entire, colon [2]. The clinical process is unpredictable, marked by alternating periods of exacerbation and remission [3]. UC is associated with morbidity, mortality, and substantial cost to healthcare; therefore, several studies have tried to identify the disease. The incidence and prevalence rates of UC vary genetic, environmental, and geographical factors [4]. UC is the highest incidence rates from industrialized countries, such as North America (6–15.6 per 10^5 people) and Europe (1.5–20.3 per 10^5 people) [4]. Patients with UC are relatively rare in Asia; however, they are gradually increasing in Asia [4] such as Korea [5], Japan [6], Hong Kong [7], and India [8] because of the westernization of lifestyles. In Korea, the mean annual incidence of UC was 1.51 per 10^5 people from 1986 to 2005. From 2006 to 2012, the mean annual incidence of UC increased significantly to 3.2 per 10^5 people [5].

UC is associated with an increased risk of colorectal cancer (CRC) [9]. Previously studies indicated a standardized CRC incidence ratio when compared with expected CRC incidence in the general population [10, 11]. UC-associated CRC only accounted for 0.15% to 5% of all CRC diagnosed in the general population; however, the risk of CRC has been reported to markedly increase with increasing probabilities of 2% by 10 years, 8% by 20 years, and 18% by 30 years [10]. Additionally, UC-associated CRC accounts for 10% to 15% of all UC patients' mortality [9, 11]. In contrast to sporadic CRC, UC-associated CRC did not follow the typical "adenoma-carcinoma" sequences [2]. UC-associated CRC occurs as a result of repeated and prolonged inflammation in intestinal epithelial cells and has a sequence of "inflammation-dysplasia-carcinoma" from low-grade dysplasia and high-grade dysplasia to carcinoma [12, 13]. Dysplastic lesions occur in the form of 1 or 2 focal in the epithelial tissue where inflammation was induced, and pathological findings in the form of mucinous or signet ring cells are mainly observed [12, 13]. The main risk factors for UC-associated CRC include certain disease characteristics such as age at onset, extent and duration of disease, as well as non-UC characteristics such as family history of CRC [9]. However, excluding these factors, the biological factors and mechanisms of carcinogenesis have not been identified. Since limited understanding of the carcinogenesis of UC-associated CRC, the knowledge concerning the CRC risk in UC patients is still insufficient. Recent studies [14, 15] investigated UC-associated genomic alterations using targeted next-generation sequencing and reported certain discrepancy from UC-associated CRC in comparison with conventional CRC including higher frequency of *TP53* mutations and lower frequencies of APC mutations.

In this study, we conducted a comprehensive analysis of 15 samples from 5 UC-associated CRC patients to understand the molecular characteristics of patients through a whole exome sequencing with tumor, dysplasia, and matched normal samples. To identify biological factors of UC-associated CRC, we analyzed mutation frequencies from tumor samples and biological function as well as generic alterations that were originated from earlier dysplasia status. Our study provides an in-depth understanding of UC-associated CRC and novel biological factors of carcinogenesis in UC-associated CRC.

METHODS

Patients and data collection

Five patients with surgically resected primary CRC were enrolled from Asan Medical Center (Seoul, Korea) between December 2014 and January 2017. The detailed clinical characteristics of the 5 patients are given in Table 1. Fifteen tissue sections were collected from recruited patients to identify somatic genetic variations in tumor and dysplasia samples, where tumor, dysplasia, and normal sections were collected from each patient's formalin-fixed paraffin-embedded (FFPE) tissues.

This study was conducted with the approval of the Institutional Review Board of Asan Medical Center (No. 2018-0436) with a waiver for informed consent.

Whole exome sequencing of tumor, dysplasia, and matched normal samples

Genomic DNA was extracted from the FFPE samples and at least 1.0 μg of genomic DNA was used for constructing sequencing library. Sequencing library was constructed using the SureSelect Human Exon V6 kit (Agilent Technologies, Santa Clara, CA, USA) by following the manufacturer's instructions. Sequencing was performed on the NextSeq platform (Illumina, San Diego, CA, USA) with a read length of 150 bp, where tumor and dysplasia samples were sequenced with the average target depth of 200× while matched normal samples were sequenced with the average target depth of 100×. From the sequencing, 90.0% of sequenced bases showed quality scores higher than Q30.

Processing of sequencing data for calling somatic variants

Filtering of sequenced reads based on the base quality was performed using Sickle [16]. Filtered reads were mapped to the human reference genome (hg19) using Burrow-Wheeler Aligner ver. 0.7.12 [17]. The aligned reads were further processed for deduplication (using Picard ver. 1.119 [18]), local realignment and recalibrating base quality scores (using Genome Analysis Toolkit ver. 3.5 [19]). Somatic variants were called by Mutect2 ver. 3.5 [20] with default parameters. Somatic variations on exonic regions and splicing sites were selected for downstream analysis, while synonymous variations were eliminated. Driver mutation analysis was performed for dysplasia samples, dysplasia-initiated tumor samples, and The Cancer Genome Atlas (TCGA) CRC samples using DriverPower ver.1.02 [21], where the false discovery rate-adjusted q-score of <0.1 was used as the threshold of driver mutations for the TCGA samples while low q-score of <0.1 was used for the threshold of driver mutations for our dysplasia and tumor samples due to small sample sizes. In order to detect somatic copy number variations from dysplasia and tumor samples, EXCAVATOR2 [22] was used with default parameters.

Statistical comparison of somatic variations with generic colorectal cancer

The somatic variation frequencies of individual genes were computed from the tumor samples of 5 patients in this study, and they were compared with the mutation frequencies from the Catalogue of Somatic Mutations in Cancer (COSMIC) [23] database for CRC. If a gene showed difference in mutation frequencies with a P-value of <0.01 by Fisher exact test, it was declared to have sta-

Table 1. Clinical characteristics of the 5 patients in this study

Characteristic	Data	Patient No.				
		S1	S2	S3	S4	S5
Clinical characteristic						
Sex		Female	Male	Male	Female	Female
Age at diagnosis (yr)	24.0	15	32	40	14	19
Age at first surgery (yr)	36.8	33	49	50	22	30
Period between diagnosis and first surgery (yr)	12.8	18	17	10	8	11
Family history	0 (0)	–	–	–	–	–
Extracolic manifestation	0 (0)	–	–	–	–	–
Primary sclerosing cholangitis	0 (0)	–	–	–	–	–
Preoperative steroid therapy	4 (80.0)	+	+	+	+	–
Characteristics on UC-associated CRC						
Location						
Right colon	3 (60.0)	+	+			+
Left colon	2 (40.0)			+	+	
Rectum	0 (0)					
CEA (ng/mL), at the time of surgery						
<6	5 (100)	+	+	+	+	+
≥6	0 (0)					
Tumor size (cm)	3.96	2.8	1.2	1.8	7.5	6.5
Histology						
Well-differentiated	4 (80.0)	+		+	+	+
Moderately differentiated	1 (20.0)		+			
Poorly differentiated	0 (0)					
Mucinous	0 (0)					
Stage						
I	2 (40.0)		+	+		
II	3 (60.0)	+			+	+
III	0 (0)					
IV	0 (0)					
Curability of the surgery						
R0	5 (100)	+	+	+	+	+
R1	0 (0)					
R2	0 (0)					

Values are presented as mean or number (%).
UC, ulcerative colitis; CRC, colorectal cancer; CEA, carcinoembryonic antigen.

tistically significant difference in mutation frequency in dysplasia-initiated tumor samples in this study.

Functional annotation of genes with somatic variants

We selected the *tumor mutation* genes that have higher mutation frequencies from the 5 tumor samples than the frequencies from the COSMIC database [23] while showing statistical significance for the input of functional annotation. From the Molecular Signatures Database ver. 6.2 [24], 6,666 gene sets that represent canonical pathways, biological processes and molecular functions of Gene Ontology [25] were retrieved as reference of functional annotation. For each functional gene set, its overlap with the *tumor mutation* genes was computed, and the statistically significant P-value of overlap was evaluated based on the hypergeometric distribution. Functional gene sets that showed a hypergeometric P-value of < 0.01 were declared to have statistically significant associations with the *tumor mutation* genes.

We also evaluated the functional roles of the genes that showed somatic variations in both matched dysplasia and tumor samples. For the dysplasia and tumor samples from each patient, genes that showed the same somatic variations in both matched samples were declared as *dysplasia-tumor common mutation* genes for the patient. Each patient's *dysplasia-tumor common mutation* genes were functionally annotated using the same procedure that was used for the annotation of the *tumor mutation* genes. Among the functional gene sets that showed statistically significant associations, we identified *dysplasia-tumor common mutation*-specific functional gene sets by using the following random permutation approach. In order to evaluate the possibility of functional association by random chance, we randomly built 10,000 lists of genes of the same amount based on the mutation frequencies from the colon adenocarcinoma data set of TCGA. Each of 10,000 random lists of genes was functionally annotated with the same procedure, and its associated functional gene sets were identified. If a functional gene set showed statistically significant association with the patient's *dysplasia-tumor common mutation* genes and was also reported to have associations less than 100 of out 10,000 random trials (permutation P < 0.01), the functional gene set was declared to be specific to the patient's *dysplasia-tumor common mutation* genes.

RESULTS

Somatic mutational landscape of dysplasia-initiated colorectal tumors

By comparing the mutation frequencies from tumor samples with the mutation frequencies that are reported in the COSMIC database, a total of 104 *tumor mutation* genes were identified with showing significantly different mutation frequencies (Table 2). All 104 genes showed higher mutation frequencies than the mutation frequencies from the COSMIC database. Fig. 1 illustrates the mutational patterns of selected genes among the 104 genes for matched tumor and dysplasia samples, where the mutation patterns of *TP53*, *APC*, and *KRAS* are shown together as they are frequently mutated genes in general CRC patients [26]. Table 3 also shows the mutation frequencies of *TP53*, *APC*, and *KRAS*. Four out of the 5 (80.0%) dysplasia-initiated tumor samples have mutations of the *TP53* tumor suppressor gene, and this mutation frequency is higher than the mutation frequency of 31.0% that is reported in the COSMIC database even though the difference showed only marginal significance (P = 0.0347). Three of the 4 *TP53* mutations from the tumor samples were deleterious stop-gain mutations, and the matched dysplasia samples with the 3 tumor samples also have identical *TP53* stop-gain mutations. From the COSMIC database, *APC* mutations were reported with a frequency of 32.0% and *KRAS* mutations were reported with a frequency of 25.5%. However, none of the 10 tumor and dysplasia samples in this study showed mutation of *APC* or *KRAS* even though their frequency differences were not statistically significant. Regarding somatic copy number variations, no recurrent copy number variation was found from either dysplasia or tumor samples (data not shown).

Table 2. Tumor mutation genes that showed significantly different mutation frequencies in tumor samples compared to the mutation frequencies that are reported in the COSMIC database

Gene (somatic mutation frequency in tumor samples)[a]
AHNAK2 (60.0%), *ADAMTSL1* (40.0%), *ANKRD36C* (40.0%), *ASIC4* (40.0%), *C11orf80* (40.0%), *DCAF8L2* (40.0%), *DEAF1* (40.0%), *FTSJ3* (40.0%), *FUT9* (40.0%), (40.0%), *MIB2* (40.0%), *MUC19* (40.0%), *MUC4* (40.0%), *PAPD5* (40.0%), *SPATA31A6* (40.0%), *SPATA31D1* (40.0%), *STOX2* (40.0%), *TTN* (40.0%), *USP42* (40.0%), *WAS* (40.0%), *ZC3H12A* (40.0%), *ADGRA2, ADGRB2, ADRA2B, AJUBA, ANKEF1, AP5Z1, APOL4, ARHGAP45, ARHGEF40, GPRASP2, TTLL3, ASPSCR1, BAGE2, BCLAF1, BRINP2, BTG4, C16orf82, CAPN15, CEP162, CIT, CNIH1, DDX19A, ERN2, GATB, GPR179, GUCY2D, HMCN2, HOXC13, IFT20, INTS14, JADE3, KIAA1549, KIAA2012, KIR3DL1, KMT5A, LENG9, LILRA6, LILRB3, LMTK3, LOC728392, LRRK1, LY75, MELTF, MROH2A, MROH9, MS4A4E, MTCL1, MUC3A, MYO15B, NAXE, NBPF10, NECTIN1, NECTIN2, OR4C5, PERM1, PPP1R10, PRRC2C, R3HDM4, RABL6, RASGRP3, RBBP8NL, RFLNB, RNF212B, RPLP1, RPP25L, SELENOO, SPATA31C2, SPPL2C, SPTSSB, TENM3, TICRR, TNPO3, TOPORS, TPM4, TRAPPC12, TSACC, TUBB4A, TWNK, TYW1B, YBX3, ZBED8, ZNF112, ZNF177, ZNF714*

COSMIC database, Catalogue of Somatic Mutations in Cancer database.

[a] Genes without specified mutation frequencies have mutation frequencies of 20.0% (1 out of the 5 patients).

Biological functions associated with dysplasia-initiated tumor mutations

Fig. 2 shows the list of selected biological functions that are strongly associated with the 104 *tumor mutation* genes from the dysplasia-initiated tumors. Glycosylation-related functions (protein O-linked glycosylation and termination of O-glycan biosynthesis), functions related to maintaining cell structures (actin filament-based movement, actin myosin filament sliding, and extracellular matrix structural constituent), and immune-related functions (defense response and B-cell receptor signaling pathway) were strongly represented in the 104 dysplasia-initiated *tumor mutation* genes.

Somatic mutations of glycoproteins including mucins in dysplasia-initiated tumor mutations

The most significantly associated biological function is protein O-linked glycosylation from Fig. 2, where 4 genes of *ADAMTSL1*, *MUC3A*, *MUC4*, and *MUC19* from the 104 dysplasia-initiated *tumor mutation* genes are included. *ADAMTSL1* encodes a secreted protein that resembles the members of a disintegrin and metalloproteinase with thrombospondin motif (ADAMTS) family, while containing different domains including the thrombos-

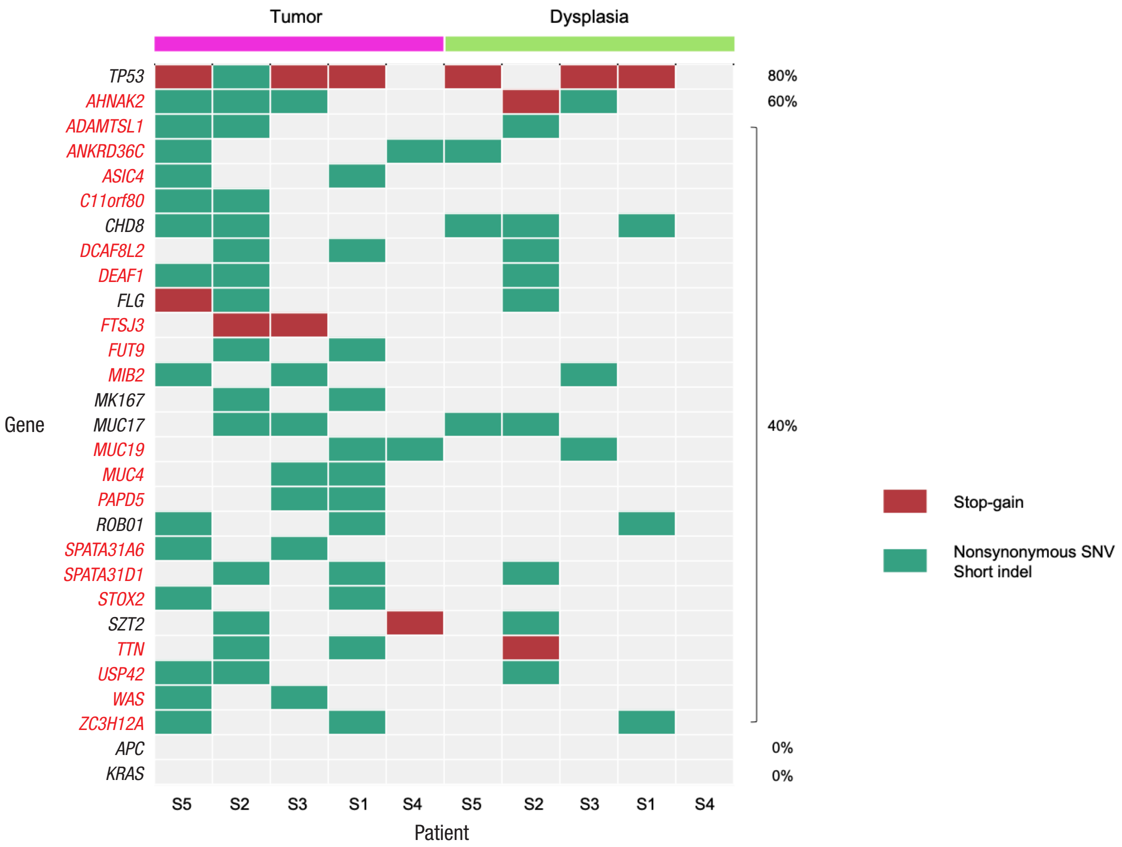

Fig. 1. Patterns of somatic mutations for selected genes from tumor and matched dysplasia samples of 5 patients. Genes that are mutated in more than one tumor sample are shown, together with *APC* and *KRAS*. If a gene has both stop-gain and other types of mutations for a sample, it is presented as a stop-gain mutation. Percentages show the frequencies of mutations in tumor samples. SNV, single nucleotide variant.

Table 3. Frequency of *TP53*, *APC*, and *KRAS* somatic mutation in dysplasia-initiated tumors and COSMIC database

Gene	Frequency of somatic mutation (%)	
	Dysplasia-initiated tumor	COSMIC
TP53	80	31
APC	0	32
KRAS	0	25

COSMIC database, Catalogue of Somatic Mutations in Cancer database.

pondin type 1 motif. *MUC3A*, *MUC4*, and *MUC19* are mucin genes that encode epithelial glycoproteins. Every dysplasia-initiated tumor sample has at least one mutation of these 4 genes (Fig. 3), and they are more frequently mutated in these dysplasia-initiated tumors than mutations in general CRCs from the COSMIC database (Table 4).

Common somatic mutations across matched dysplasia and tumor samples

Table 5 and Fig. 4 show the number of somatic mutations from matched tumor and dysplasia samples for each patient, and the number of common mutations between the 2 tissue types. On average, 22.5% of somatic mutations in dysplasia-initiated tumor samples were already present from the matched dysplasia samples. These common somatic variations across dysplasia and tumor samples did not show clear patterns of increase or decrease in their variant allele fractions (Fig. 5). Among the common somatic mutations from the 5 patients, 3 patients (60.0%) have *TP53* mutations and 2 patients (40.0%) have *AHNAK2* mutations commonly in their matched tumor and dysplasia samples (Fig. 1). The dysplasia sample of S2 also has the same mutation with the matched tumor sample other than the illustrated stop-gain mutation shown in Fig. 1.

Dysplasia-tumor common mutation-specific biological functions

Table 6 shows the selected list of biological functions that are strongly associated with the *dysplasia-tumor common mutation* genes, where only the biological functions that are specific to these dysplasia-tumor common mutations are shown. Patients' dysplasia-tumor common mutations show varying biological

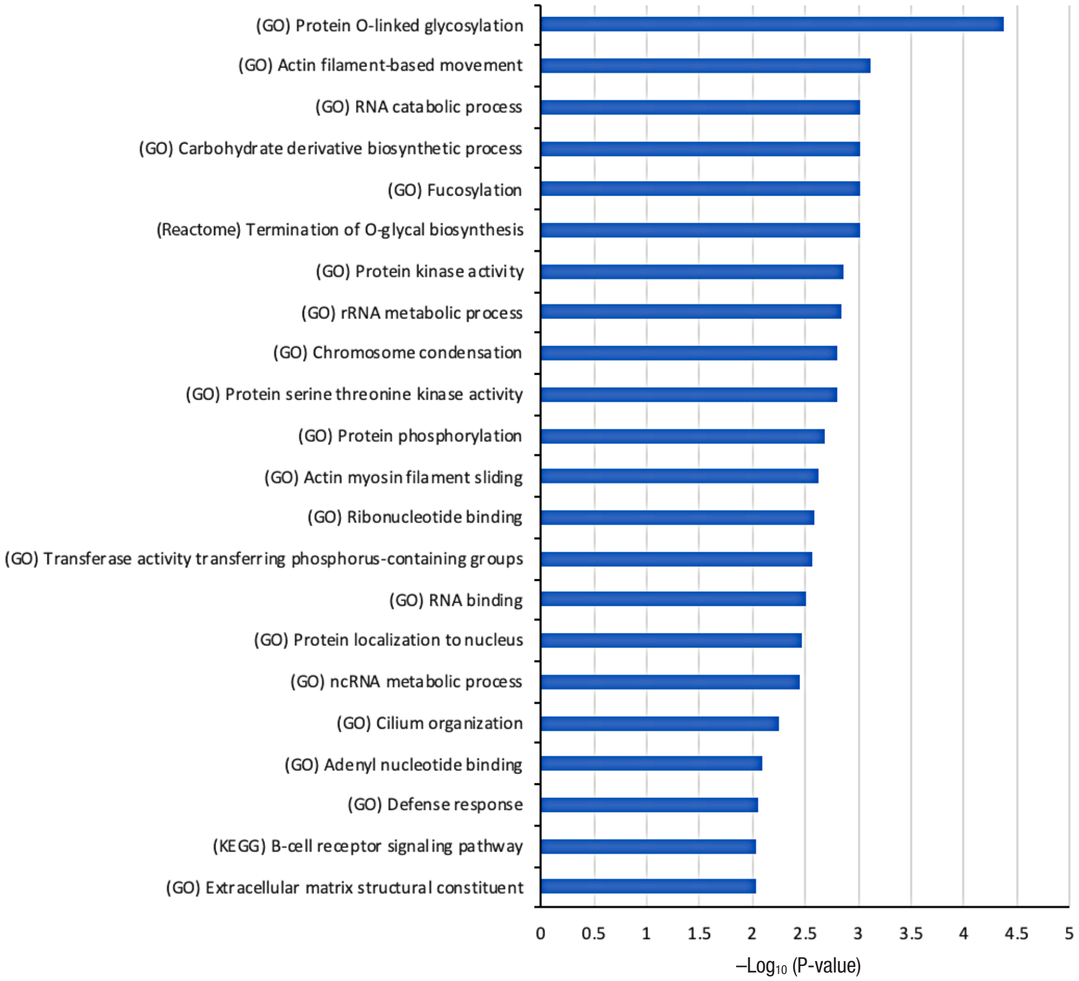

Fig. 2. Selected biological functions that are strongly associated with the 104 genes showing more frequent mutations in dysplasia-initiated tumors than general colorectal cancers reported in the Catalogue of Somatic Mutations in Cancer (COSMIC) database. The source of the biological function information is given in parentheses. Hypergeometric P-value in log-scale is shown as a measure of statistical significance. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.

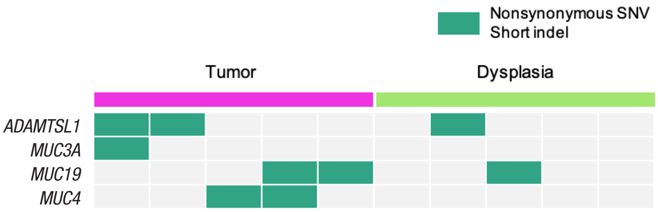

Fig. 3. Somatic mutation patterns of the 4 genes that are involved in the biological function of protein O-linked glycosylation, where it is the most significantly associated biological function with the 104 dysplasia-initiated tumor mutation genes. SNV, single nucleotide variant.

Table 4. Somatic mutation frequency of glycoproteins including mucins in dysplasia-initiated tumors and COSMIC database

Gene	Frequency of somatic mutation (%)	
	Dysplasia-initiated tumor	COSMIC
ADAMTSL1	40	3
MUC3A	20	0
MUC4	40	3
MUC19	40	1

COSMIC database, Catalogue of Somatic Mutations in Cancer database.

functions associated with them, while certain patients share similar functional profiles with their dysplasia-tumor common mutations. Biological functions related to stimulus response, apoptosis, *NOTCH* signaling, inflammation, and cell aging were associated with dysplasia-tumor common mutations from more than 2 patients.

Driver mutations analyzed from dysplasia and tumor samples

We investigated driver mutation genes from dysplasia samples and tumor samples. Four genes of *C4BPB*, *SIGLEC7*, *TET3*, and *TP53* were identified as driver mutation genes from the dysplasia samples. From the tumor samples, 19 genes of *AC016757.3*, *AHNAK2*, *ASIC4*, *ATP13A4*, *COL2A1*, *C11orf80*, *DCAF8L2*, *FLG*, *FUT9*, *HIF1AN*, *MUC17*, *NOTCH2*, *PLIN4*, *RAX2*, *SPATA31A6*, *SPATA31D1*, *SPNS3*, *TP53*, and *WSCD2* were identified as driver mutations. Only *TP53* was a common driver mutation gene between the dysplasia and tumor samples. We also analyzed driver mutations for the 399 CRC samples with reported somatic mutations from TCGA, where 87 genes were identified as driver mutation

Table 5. The number of somatic mutations in each tissue and dysplasia sample, and the number of common somatic mutations

Patient No.	No. of tissues		No. of common somatic mutations (% from tumor mutations)
	Tumor	Dysplasia	
S1	120	64	18 (15.0)
S2	409	491	232 (56.7)
S3	103	97	19 (18.4)
S4	45	32	6 (13.3)
S5	226	195	21 (9.3)

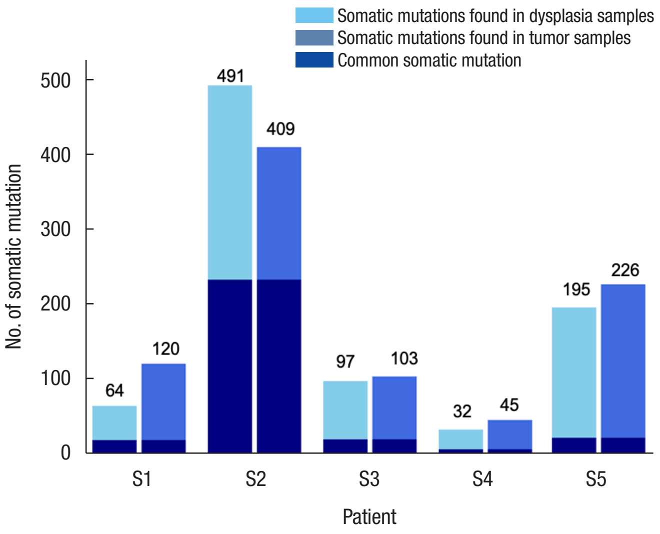

Fig. 4. The amount of somatic mutations from each subject, separately shown for dysplasia and tumor samples.

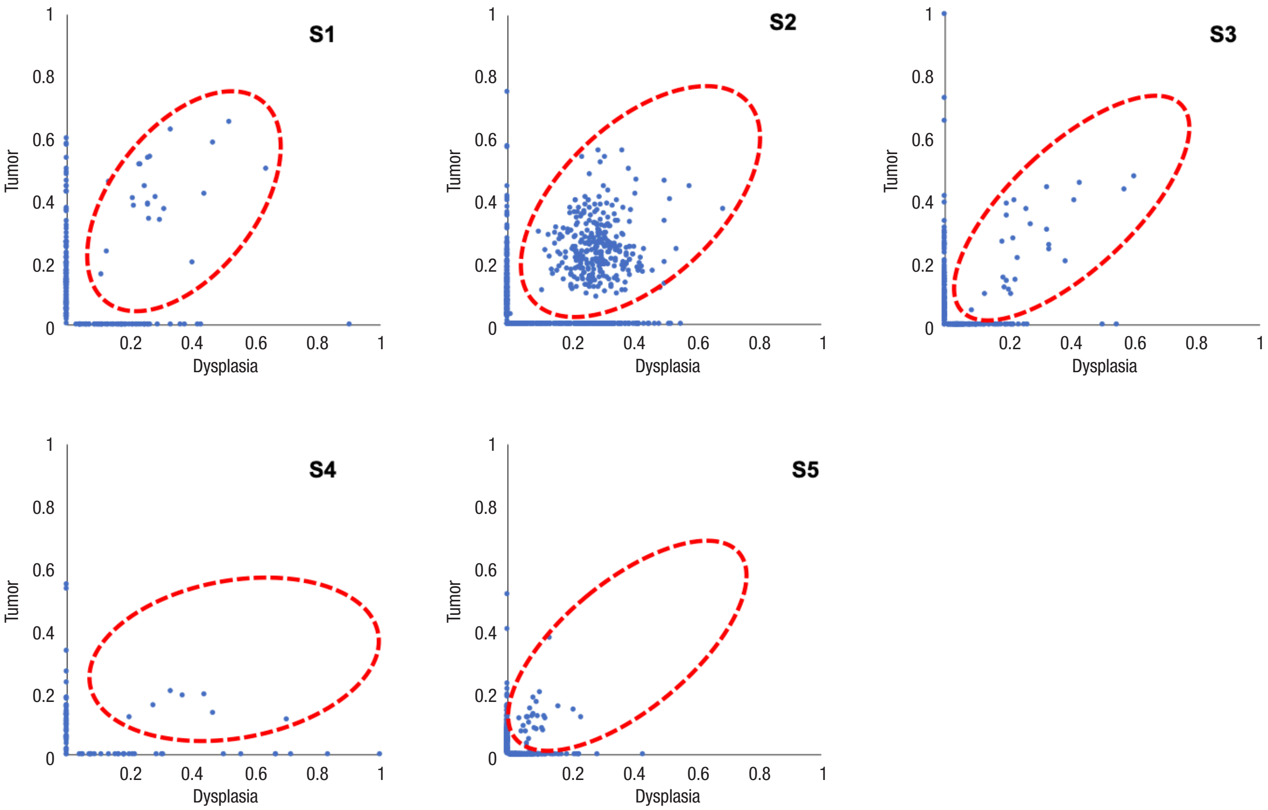

Fig. 5. Variant allele fractions of somatic variations are shown for matched dysplasia and tumor samples for each patient. Red-dotted circles mean the dysplasia-tumor common mutation genes.

Table 6. Selected list of biological functions that are specific to dysplasia-tumor common mutations for each patient

Patient	Biological function	Hypergeometric P-value
S1	(GO) Positive regulation of execution phase of apoptosis	< 0.001
	(GO) Cellular response to glucose starvation	< 0.001
	(GO) Response to external stimulus	< 0.001
	(GO) Inflammatory response	0.0016
	(GO) Replicative senescence	0.0063
S2	(GO) Central nervous system development	< 0.001
	(GO) Response to external stimulus	< 0.001
	(GO) Negative regulation of peptidase activity	< 0.001
	(GO) Regulation of cilium assembly	< 0.001
	(Reactome) *ZINC* transporters	0.0046
S3	(GO) Positive regulation of histone deacetylation	2.94E-05
	(GO) Positive regulation of chromatin modification	0.0011
	(GO) Regulation of apoptotic signaling pathway	0.0012
	(Reactome) Signaling by *NOTCH*	0.0016
	(GO) Inflammatory response	0.0022
S4	(GO) Central nervous system development	< 0.001
	(GO) Cell fate commitment	< 0.001
	(Reactome) Signaling by *NOTCH3*	0.0022
	(Reactome) *NOTCH - HLH* transcription pathway	0.0024
	(GO) Positive T-cell selection	0.0039
S5	(GO) Replicative senescence	< 0.001
	(BioCarta) Telomerase pathway	< 0.001
	(GO) Cell cycle arrest	< 0.001
	(GO) Regulation of DNA-templated transcription in response to stress	< 0.001
	(PID) *C-Myc* activation pathway	0.0011

GO, Gene Ontology; PID, Pathway Interaction Database.

genes; but they showed no common driver mutation gene with the results from our dysplasia and tumor samples (data not shown).

DISCUSSION

In this study, we investigated the somatic mutational characterization of 5 dysplasia-initiated CRC patients where mutational profiles of their tumors and matched dysplasia samples were analyzed. From the comparison with the mutational frequencies that are reported for general CRC in the COSMIC database, 104 genes were identified to show different mutation frequencies from the dysplasia-initiated tumor samples. All the 104 genes showed higher mutation frequencies compared to general CRCs, but a notable observation is the frequent mutation of the tumor suppressor gene *TP53* (mutated from 4 of the 5 tumor samples) even though its mutation frequency was only marginally different from that of general CRC (80% vs. 31%, $P = 0.0347$). More importantly, 3 of the 4 *TP53* mutations were deleterious stop-gain mutations, and they were originally initiated from earlier dysplasia status (Fig. 1). This high mutation frequency and early establishment of somatic mutation from dysplasia status for *TP53* strongly suggest that the aberration of *TP53* plays an important role in driving dysplasia-initiated CRC. It is generally recognized that somatic mutations of other genes such as *APC* work as the main driver of initiating CRC, and the mutation of *APC* was reported with a frequency of 32% from the COSMIC database. However, none of the tumors or matched dysplasia samples in this study has *APC* mutation (Fig. 1). This is another evidence that dysplasia-initiated tumors have different tumor driving mechanisms with general CRC. Our result of driver mutation analysis also supports this point, where *TP53* was the common driver mutation gene across matched dysplasia and tumor samples while no common driver mutation gene was identified from conventional CRC samples. Additionally, we could not observe any *KRAS* mutation from the tumor and matched dysplasia samples in this study (Fig. 1), while *KRAS* is often mutated in general CRC (25% from the COSMIC database) and causes resistance to several therapeutic agents. These findings of the important role of *TP53* in UC-associated CRC and lower mutation frequencies of *APC* and *KRAS* are in concordance with previous studies [14, 15].

When investigating the biological functions associated with the highly mutated 104 genes from the dysplasia-initiated tumors, we observed multiple biological functions that can be related to dysplasia and inflammation, such as cell structure maintenance, and immune-related functions. We also observed strong aberration in glycosylation-related functions where it can be related to aberrated mucus formation on epithelial surfaces. Secreted extracellular proteins are often glycosylated, and it is known that epithelial tissue-produced mucins are heavily glycosylated. Mucins are major constituents of mucus, the viscous secretion that covers epithelial surfaces as those in colon. As glycoproteins including mucins play important roles in the protection of the epithelial cells, the aberration of glycosylation function suggests its strong correlation with the physiology of dysplasia-initiated CRC. We observed that 4 genes (*ADAMTSL1*, *MUC3A*, *MUC4*, and *MUC19*) are involved in this strong aberration in glycosylation, and we found that every dysplasia-initiated tumor in this study has mutations of at least one of these genes while general CRCs rarely have these mutations (Fig. 3). This suggests that dysplasia-initiated CRC can be associated with the disturbed ability to protect epithelial surfaces of colon due to mutated mucins and aberrated glycosylation function.

When we compared the somatic mutations of tumors and their matched dysplasia samples, we found that many somatic mutations already arose from dysplasia status, and a certain portion of them (22.5% on average) are carried to later established tumors. Biological functions that are specific to dysplasia-tumor common mutations include many functions related to cancer establishment and progression. The most frequently observed dysplasia-tumor common mutation was the mutations of *TP53* (3 of 5 patients) and *AHNAK* (2 of 5 patients). *TP53* is a well-known tumor suppressor gene, and it is noteworthy that its stop-gain mutations are established earlier from the dysplasia status as mentioned earlier in this section. *AHNAK2* is an *AHNAK* nucleoprotein 2, and it may play a role in calcium signaling by associating with calcium channel proteins. The relevance between the *AHNAK2* mutation and dysplasia-initiated CRC is not clear, but we will investigate its role in future studies.

We have analyzed the mutational characteristics and their importance in dysplasia-initiated CRC in this study by investigating somatic mutations of matched tumor and dysplasia samples from 5 patients. There have been previous studies that investigated genomic characteristics of UC-associated CRC using next-generation sequencing, and our result is in consistency with such studies. Nevertheless, our study can be distinguished with previous studies in 2-folds, where whole exome sequencing was used in our study to investigate entire genes while previous studies used only limited targeted-sequencing, and matched normal, dysplasia, and tumor samples were used in our study while only tumor tissues were analyzed in previous studies. One limitation of our study is the limited number of analyzed samples, as it is challenging to suggest stronger statistical evidence with 5 patients when comparing to general CRC populations. We aim to provide more sufficient analytical evidence in future studies with additional tumor and dysplasia samples.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGMENTS

The results shown here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. This study was funded by the Korean Society of Coloproctology. This work was also supported in part by the National Research Foundation of Korea, funded by the Ministry of Science and ICT, Republic of Korea (grant number: 2020M3C9A5085206).

REFERENCES

1. Kunovszki P, Milassin Á, Gimesi-Országh J, Takács P, Szántó K, Bálint A, et al. Epidemiology, mortality and prevalence of colorectal cancer in ulcerative colitis patients between 2010-2016 in Hungary: a population-based study. PLoS One 2020;15:e0233238.
2. Baker KT, Salk JJ, Brentnall TA, Risques RA. Precancer in ulcerative colitis: the role of the field effect and its clinical implications. Carcinogenesis 2018;39:11-20.
3. Ordás I, Eckmann L, Talamini M, Baumgart DC, Sandborn WJ. Ulcerative colitis. Lancet 2012;380:1606-19.
4. Loftus EV Jr. Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. Gastroenterology 2004;126:1504-17.
5. Baek SJ, Lee KY, Song KH, Yu CS. Current status and trends in inflammatory bowel disease surgery in Korea: analysis of data in a nationwide registry. Ann Coloproctol 2018;34:299-305.
6. Morita N, Toki S, Hirohashi T, Minoda T, Ogawa K, Kono S, et al. Incidence and prevalence of inflammatory bowel disease in Japan: nationwide epidemiological survey during the year 1991. J Gastroenterol 1995;30 Suppl 8:1-4.
7. Chow DK, Leong RW, Tsoi KK, Ng SS, Leung WK, Wu JC, et al. Long-term follow-up of ulcerative colitis in the Chinese population. Am J Gastroenterol 2009;104:647-54.
8. Sood A, Midha V, Sood N, Bhatia AS, Avasthi G. Incidence and prevalence of ulcerative colitis in Punjab, North India. Gut 2003;52:1587-90.
9. de Campos Silva EF, Baima JP, de Barros JR, Tanni SE, Schreck T, Saad-Hossne R, et al. Risk factors for ulcerative colitis-associated colorectal cancer: a retrospective cohort study. Medicine (Baltimore) 2020;99:e21686.
10. Jess T, Rungoe C, Peyrin-Biroulet L. Risk of colorectal cancer in patients with ulcerative colitis: a meta-analysis of population-based cohort studies. Clin Gastroenterol Hepatol 2012;10:639-45.
11. Zhou Q, Shen ZF, Wu BS, Xu CB, He ZQ, Chen T, et al. Risk of colorectal cancer in ulcerative colitis patients: a systematic review and meta-analysis. Gastroenterol Res Pract 2019;2019:5363261.
12. Chung DC. The genetic basis of colorectal cancer: insights into critical pathways of tumorigenesis. Gastroenterology 2000;119:854-65.
13. Sharan R, Schoen RE. Cancer in inflammatory bowel disease. An evidence-based analysis and guide for physicians and patients. Gastroenterol Clin North Am 2002;31:237-54.
14. Kameyama H, Nagahashi M, Shimada Y, Tajima Y, Ichikawa H, Nakano M, et al. Genomic characterization of colitis-associated colorectal cancer. World J Surg Oncol 2018;16:121.
15. Yaeger R, Shah MA, Miller VA, Kelsen JR, Wang K, Heins ZJ, et al. Genomic alterations observed in colitis-associated cancers are distinct from those found in sporadic colorectal cancers and vary by type of inflammatory bowel disease. Gastroenterology 2016;151:278-87.
16. Joshi NA, Fass JN. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (ver. 1.33) [Software]. 2011.
17. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-60.
18. Broad Institute. Picard toolkit [Internet]. Cambridge, MA: Broad Institute; c2019 [cited 2021 Apr 21]. Available from: http://broadinstitute.github.io/picard/.
19. Van der Auwera GA, O'Connor BD, editors. Genomics in the cloud: using docker, GATK, and WDL in Terra. Sebastopol, CA: O'Reilly Media; 2020.

20. Benjamin D, Sato T, Cibulskis K, Getz G, Stewart C, Lichtenstein L. Calling somatic SNVs and indels with Mutect2. bioRxiv 2019: 861054.
21. Shuai S; PCAWG Drivers and Functional Interpretation Working Group, Gallinger S, Stein L; PCAWG Consortium. Combined burden and functional impact tests for cancer driver discovery using DriverPower. Nat Commun 2020;11:734.
22. D'Aurizio R, Pippucci T, Tattini L, Giusti B, Pellegrini M, Magi A. Enhanced copy number variants detection from whole-exome sequencing data using EXCAVATOR2. Nucleic Acids Res 2016;44: e154.
23. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res 2019;47:D941-7.
24. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545-50.
25. Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res 2021;49:D325-34.
26. Armaghany T, Wilson JD, Chu Q, Mills G. Genetic alterations in colorectal cancer. Gastrointest Cancer Res 2012;5:19-27.