Genomic analysis for discovering genetic alterations in young Korean patients with double primary cancers of the stomach and colon

Yoon Young Choi

Department of Medicine

The Graduate School, Yonsei University

Genomic analysis for discovering genetic alterations in young Korean patients with double primary cancers of the stomach and colon

Yoon Young, Choi

Department of Medicine

The Graduate School, Yonsei University

Genomic analysis for discovering genetic alterations in young Korean patients with double primary cancers of the stomach and colon

Directed by Professor Jae-Ho Cheong

The Doctoral Dissertation
submitted to the Department of Medicine,
the Graduate School of Yonsei University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Yoon Young Choi

June 2019

This certifies that the Doctoral Dissertation of Yoon Young Choi is approved.

Thesis Supervisor: Jae-Ho Cheong

Thesis Committee Member#1: Hyunki Kim

Thesis Committee Member#2: Ho-Geun Yoon

Thesis Committee Member#3: Seung-Tae Lee

Thesis Committee Member#4: Ji Yeong An

The Graduate School
Yonsei University

June 2019

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. Jae-Ho Cheong who is my lifelong mentor and advisor for research and academia life. I also give my appreciation to my admirable senior mentor, Dr. Sung Hoon Noh, who has inspired me to be a surgeon–scientist. I would like to thank Dr. Ho-Geun Yoon, Seung-Tae Lee, Ji Yeong An, and Hyunki Kim for their advice and feedback of this work. Finally, I would like to dedicate this manuscript for my Doctoral dissertation to my family: my father and mother, my father- and mother-in-law, my lovely wife Dr. Min Hee Lee, and my loving sons, Jiu and Yeonu.

<TABLE OF CONTENTS>

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Genomic analysis for discovering genetic alterations in young Korean patients with double primary cancers of the stomach and colon

Yoon Young Choi

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Jae-Ho Cheong)

The incidence of multiple primary cancers has increased as the prognosis of patients with cancer has improved. The most common type of double primary cancer in Korea is the combination of stomach and colon cancers. It is the genetic risk of the individual that is associated with those who are affected by two primary cancers at an early age; however, the germline variant of patients with double primary cancer has not been evaluated in great detail. Two cancers in one individual share similar genomic characteristics, and if it is targeting a common variant then this would be a possible anti-cancer strategy. We evaluated the genomic characteristics, both germline and somatic variants, of patients with pathologically confirmed cancers in both the stomach and colon at Severance Hospital between January 2000 and December 2016. In the multi-gene germline target sequencing analysis, mismatch repair (MMR) related genes such as *MLH1*, *MSH2*, and *MSH6* pathogenic/likely pathogenic (P/LP) germline

variants were detected in nine patients (9/55, 16.4%). Young age (< 55 years old), Amsterdam II criteria, sum of the number of lesions, and microsatellite instability-high (MSI-H) were the significant risk factors of the P/LP germline variants in patients with double primary cancers of the stomach and colon. In the whole exome sequencing analysis of the normal, stomach, and colorectal cancers, a few shared common somatic variants were detected, mainly in the MSI-H tumors and not in the microsatellite stable (MSS) cancer type. In the mutational signature analysis of somatic variants in the stomach and colorectal cancers, MSI-related signatures such as dMMR, hypermutation, and MSI were shared when both the tumors were of the MSI-H type. Otherwise, no clear common mutational signatures were observed, except for age-related signatures despite the patients being diagnosed with two cancers at less than 55 years of age. Those with double primary cancers who were less than 55 years old, had a family history of gastric cancer, and the MSI-H tumor type would be recommended to undergo a germline genomic test. Common variants between stomach and colorectal cancers in one individual were rarely detected especially in the MSS type of cancer; consequently, the simultaneous targeting of both tumors in one patient would be a difficult strategy for clinical practice.

Key words: double primary cancer, stomach cancer, colorectal cancer, next generation sequencing, germline genomic alteration

Genomic analysis for discovering genetic alterations in young Korean patients with double primary cancer of the stomach and colon

Yoon Young Choi

Department of Surgery
The Graduate School, Yonsei University

(Directed by Professor Jae-Ho Cheong)

I. INTRODUCTION

The extended life span and improved survival of patients with cancers has led to a worldwide increase in the number of patients with double primary cancers.[1,2] The most common type of double primary cancers in Korea is that of the stomach and colon.[3] These cancers are the most common types of primary cancer in Korea;[4] and, aging would be the main cause of these double primary cancers. Lynch syndrome, a hereditary nonpolyposis colorectal cancer, mainly affects the colon, endometrium, and stomach. Microsatellite instability (MSI), which is related to Lynch syndrome, is one of the representative molecular subtypes of these three cancers.[5-7] Therefore, some cases of double primary

cancers of the stomach and colon are related to this hereditary cancer syndrome. According to a previous study regarding Lynch syndrome in Korea, the relative risk of obtaining gastric cancer in a family with a history of Lynch syndrome is 2–5 times higher than that of the normal population before the age of 60, whereas the relative risk of gastric cancer is similar between a Lynch syndrome family and the normal population when the age is over 60.[8] This could support the fact that cancer occurring at a young age is more related to genetic factors than it is to environmental factors. The above mentioned factors raise a clinical hypothesis that double primary cancers occurring at a young age are mainly related to genetic factors. Consequently, evaluating genomic characteristics of patients with double primary cancers of the stomach and colon might lead to a genetic risk being found. In addition, evaluating somatic variants of double primary cancers occurring in an individual would provide information for the biological mechanism of each cancer.

The primary purpose of the present study was to investigate the risk factors, including the genetic risk, of double primary cancers of the stomach and colon. The second purpose was to identify the genomic characteristics of double primary cancers of the stomach and colon.

II. MATERIALS AND METHODS

1. Population included in the study

The present study focused on the genomic characteristics of patients with double primary cancers of the stomach and colon in a young age group (< 55 years old). The patients were selected using the following criteria: 1) < 55 years old at diagnosis of second cancer, either gastric or colorectal, 2) undergone surgical resection including endoscopic resection for both tumors at Severance Hospital, Yonsei University College of Medicine, 3) formalin-fixed paraffin embedded (FFPE) normal and tumor tissues available, and 4) satisfactory DNA quality for genomic sequencing.

To evaluate the effect of age on germline variants, a control group was selected with the following selection criteria: 1) age > 55 years old at diagnosis of secondary cancer, either gastric or colorectal (1:4 randomly selected), 2) undergone surgical resection including endoscopic resection for both tumors at Severance Hospital, 3) FFPE normal tissues available, and 4) satisfactory DNA quality.

The clinical–pathological characteristics of the patients including age, sex, family history, location of tumors, number of tumors, histology and TNM stage,

and MSI status were evaluated.

2. Germline target sequencing analysis

Genomic DNA was extracted from each individual's normal confirmed FFPE sample using a QIAamp DNA Tissue Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Intact DNA was quantified and adjusted to a concentration of 5 ng/μL using a Qubit 2.0 fluorometer (Invitrogen, Waltham, MA) and a Qubit dsDNA HS Assay Kit (Invitrogen). Precapture libraries were constructed according to the manufacturer's sample preparation protocol. The genomic DNA of each patient was fragmented to a median size of 300 base pairs (bp). We used a customized targeted capture sequencing panel (OncoRisk, Celemics, Seoul, Korea) covering all coding sequences and intron–exon boundaries of the coding exons of 65 cancer susceptibility genes (Table 1). DNA fragments were end-repaired, phosphorylated, and adenylated at the 3′ ends. The index adaptors were ligated to the repaired ends, the DNA fragments were amplified, and fragments of 200 to 500 bp were isolated. Pooled libraries were sequenced on a MiSeq sequencer (Illumina, San Diego, CA) using a MiSeq Reagent Kit v2 (300 cycles).[9]

Variants were described based on the nomenclature recommendations of the Human Genome Variation Society (http://www.hgvs.org/mutnomen) and further categorized according to the American College of Medical Genetics and Genomics (ACMG) recommendations, with supporting linkage, biochemical, clinical, functional, and statistical data used for specific missense and intronic alterations. The variants were classified into pathogenic, likely pathogenic, variant of uncertain significance (VUS), likely benign, and benign/polymorphism, using the five-tier system following the guidelines of the ACMG. Initially, variants were filtered by their frequencies in the population control databases including ExAC (non-Cancer Genome Atlas dataset; frequencies were calculated based on ethnic subgroups), ESP6500, 1000 Genomes Project, and Korean Reference Genome Database. Variants with a minor allele frequency (MAF) greater than 5% in any of the population subgroups were classified as absolutely benign. Conventionally, variants with a MAF greater than 0.5% are considered as having a strong evidence of a benign variant, whereas the evidence supporting pathogenicity is considered moderate if these variants are shown to be absent from the general population. Furthermore, literature and database searches for previous reports and functional

studies were performed using Alamut Visual 2.6 software (Interactive Biosoftware, Rouen, France) and the Human Gene Mutation Database professional database. When all in silico analyses showed consistent predictions, the results were considered to demonstrate that a certain variant was benign or pathogenic.

We identified all small bp variations using Sanger sequencing on a 3730 DNA analyzer with a BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA). Sequencing data were aligned against appropriate reference sequences and analyzed using the Sequencher 5.3 software program (Gene Codes Corp., Ann Arbor, MI). Chromosomal copy number alterations were confirmed using the Infinium CytoSNP 850K array and BlueFuse Multi software (Illumina).

Table 1. List of the 65 genes used for germline multi-gene next generation sequencing panel

APC	*CHEK2*	*POLE*	*VHL*	*ALK*	*FLCN*
ATM	*EPCAM*	*PRSS1*	*WT1*	*PHOX2B*	*GALNT12*
BARD1	*MEN1*	*PTEN*	*NF1*	*KIF1B*	*GPC3*
BLM	*MLH1*	*RAD50*	*NF2*	*LMO1*	*GREM1*
BMPR1A	*MRE11A*	*RAD51C*	*RB1*	*PAX6*	*MLH3*
BRCA1	*MSH2*	*RAD51D*	*RUNX1*	*CTNNB1*	*PMS1*
BRCA2	*MSH6*	*RET*	*KRAS*	*AXIN1*	*POLD1*
BRIP1	*MUTYH*	*SLX4*	*NRAS*	*NTRK1*	*PPM1D*
CDH1	*NBN*	*SMAD4*	*PTCH1*	*AXIN2*	*SDHAF2*
CDK4	*PALB2*	*STK11*	*SDHA*	*EXO1*	*RAD51*
CDKN2A	*PMS2*	*TP53*	*SDHB*	*FANCM*	

3. Whole exome sequencing (WES) analysis

Genomic DNA were extracted from the confirmed normal and tumor (gastric and colorectal) tissues of FFPE using a DNeasy Tissue Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's protocols. The DNA quality was checked by 1% agarose gel electrophoresis and by the PicoGreen® dsDNA Assay (Invitrogen, Carlsbad, CA, USA). For WES, SureSelect sequencing libraries were prepared according to the manufacturer's instructions (Agilent SureSelect All Exon V4 kit, Santa Clara, CA, USA) using the Bravo automated

liquid handler. The library qualities of both whole exomes were verified by capillary electrophoresis (Bioanalyzer, Agilent). After real-time polymerase chain reaction (qPCR) using SYBR Green PCR Master Mix (Applied Biosystems, Waltham, MA, USA), index-tagged libraries were combined in equimolar amounts in the pool. Cluster generation occurred in the flow cell on the cBot automated cluster generation system (Illumina™). The flow cell loaded on the HISEQ 2500 sequencing system (Illumina™) performed sequencing with read lengths of 2 × 100 bp.

4. MSI and mismatch repair (MMR) status

From all matched normal and tumor tissues, DNA was extracted for PCR amplification. The handling of all surgical specimens, sample preparation, PCR amplification, and fragment analysis were performed as previously described.[10,11] For DNA amplification, 20 × 1 L reaction solutions were used, containing 2 × 1 L of 103 buffer (Roche, Mannheim, Germany), 1.7–2.5 mmol/L of $MgCl_2$, 0.3 × 1 M of each primer pair, 250 × 1 M of deoxynucleotide triphosphates, and 2.5 units of DNA polymerase (Roche). Amplification involved an initial step at 94℃ for 5 min; 30 one-minute cycles at 94, 55, and 72℃ (step by step); and a final extension at 72℃ for 10 min. Subsequently,

0.7 × 1 L of the amplified samples were mixed with 0.3 × 1 L of GeneScan 500 size standard, and 9 × 1 L of HiDi formamide in an ABI Prism 3100 Genetic Analyzer for fragment separation. Electrophoresis was initiated when the temperature was 60℃, and 16 capillaries (36 cm in length and 50 cm in diameter) were used for the array. POP-4 was applied as the separation medium, and fluorescence was converted into digital information and sent to a workstation (ABI Prism 3100 Data Collection software). Two mononucleotide repeat markers (BAT25 and BAT26) and three dinucleotide repeat markers (D5S346, D2S123, and D17S250) were used for estimating the MSI status, as recommended by the National Cancer Institute consensus group.[12] When two or more mutated markers were identified, the tumor was classified as MSI-high (MSI-H). When MSI was demonstrated at only one marker, the tumor was classified as MSI-low (MSI-L). When there was no MSI, the tumors were classified as MSS.

Immuno-histochemistry was performed with a Ventana XT automated stained with antibodies for cytokeratin (1:300, AE1/AE3, DAKO, Carpinteria, CA, USA), *MLH1* (ready-to-use, clone M1, Roche, Indianapolis, IN, USA), *MSH2* (ready-to-use, clone G219-1129, Roche), *MSH6* (1:100, clone 44, Cell Marque,

Rocklin, CA, USA), and *PMS2* (1:40, clone MRQ28, Cell Marque). Sections were deparaffinized with EZ Prep solution (Ventana). CC1 standard [pH 8.4 buffer containing tris (hydroxymethyl) aminomethane–borate–EDTA] was used for antigen retrieval and blocked with 3% H_2O_2 for 4 min at 37℃. Slides were incubated with primary antibody for 40 min at 37℃ followed by a universal secondary antibody for 20 min at 37℃. Slides were incubated in streptavidin–horseradish peroxidase for 16 min at 37℃ and then the substrate, 3,30-diaminobenzidine tetrahydrochloride in H_2O_2, was added for 8 min followed by hematoxylin and bluing reagent counterstaining at 37℃. A loss of MMR protein expression (MMR deficiency) was defined as when none of the neoplastic epithelial cells showed nuclear staining, whereas normal expression was defined as the presence of nuclear staining of tumor cells, irrespective of the proportion or intensity. Infiltrating lymphocytes, stromal cells, and adjacent non-neoplastic epithelium served as internal positive controls. An MMR-deficient (dMMR) tumor was defined as a tumor showing loss of expression of any of the four MMR proteins. In the present study, either MSI-H or dMMR were considered as MSI-H in both stomach and colorectal cancers.

5. Data analyses

Continuous variables are presented as means and standard deviations and were analyzed by the Mann-Whitney U-test. Categorical variables are presented as numbers and percentages, and analyzed by the chi-square test or Fisher's exact test as appropriate. A *p*-value less than 0.05 was considered statistically significant. The analyses were performed using SPSS, version 23.0 software for Windows (SPSS, Chicago, IL, USA).

We aligned WES information using Burrows–Wheeler Aligner software. Then, we removed duplicated reads using Picard. Using Genome Analysis Toolkit software, indel realignment and base recalibration was conducted, then variant calling and filtering was undertaken. Variant annotation was performed by SnpEff.

III. RESULTS

A total of 19 patients with stomach and colorectal cancers who were < 55 years of age were included in the present study. For these patients, multi-gene germline targets next generation sequencing (NGS) analysis and WES analysis of gastric and colorectal cancers as well as for the paired normal tissues were conducted. To evaluate the influence of age for detecting pathogenic/likely

pathogenic (P/LP) germline variants, multi-gene target NGS analysis was conducted on 36 randomly selected (1:4) patients with stomach and colorectal cancers who were ≥ 55 years of age (Figure 1).

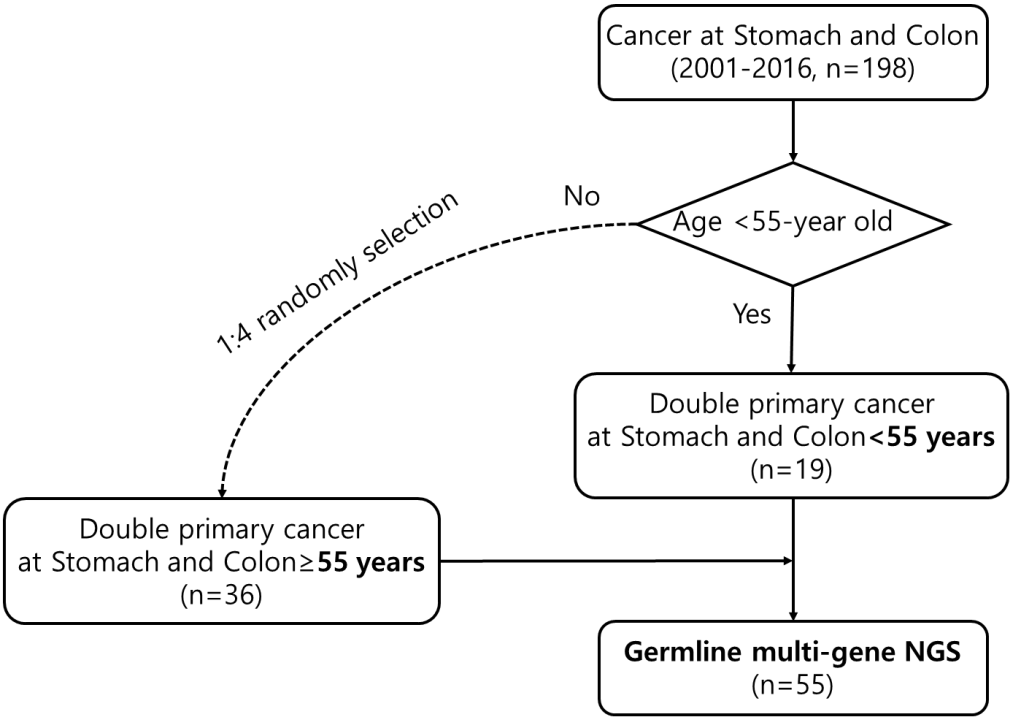

Figure 1. Flow diagram of the enrolled population in the present study

GC; gastric cancer, CRC: colorectal cancer, NGS; next generation sequencing, WES; whole exome sequencing

1. Germline variants of patients with double primary cancers of the stomach and colon

The baseline characteristics by age of the patients are described in Table 2. The mean age at diagnosis of gastric cancer and colorectal cancer of the young age group (< 55 years old) and control group was 46.0 and 48.3 years, and 68.0 and 69.0 years, respectively. Ten out of 13 (76.9%) patients in the young age group had a family history of gastric cancer, whereas 9 out of 34 (26.5%) patients in the control group had a family history of gastric cancer ($p = 0.003$). A total of 36% and 44% of patients in the young age group were of the MSI-H type for gastric cancer and colorectal cancer, respectively, whereas 11.1% and 20.6% of patients in the control group were of the MSI-H type for gastric cancer and colorectal cancer ($p = 0.067$ and 0.071, respectively). The P/LP germline variant was detected in 7 (36.8%) and 2 (5.6%) patients in the young age and control groups, respectively, which was statistically significant ($p < 0.001$).

Table 2. Demographics of enrolled population

	< 55 years old (n = 19)	≥ 55 years old (n = 36)	P-value
Age			
GC	46.0 ± 4.6	68.0 ± 6.9	<0.001
CRC	48.3 ± 4.3	69.0 ± 6.6	<0.001
Family history			
GC (no/yes (%))	3/10 (23.1/76.9)	25/9 (73.5/26.5)	0.003
CRC (no/yes (%))	8/5 (61.5/38.5)	26/8 (76.5/23.5)	0.467
any cancer (no/yes (%))	0/13 (0/100)	11/23 (32.4/67.6)	0.021
Amsterdam I (no/yes (%))	12/1 (92.3/7.7)	33/1 (97.1/2.9)	0.481
Amsterdam II (no/yes (%))	8/5 (61.5/38.5)	27/7 (79.4/20.6)	0.209
Location of tumor			
GC (UB/M-LB (%)	1/18 (5.3/94.7)	9/27 (25.0/75.0)	0.139
CRC (Rt./Lt. (%))	7/12 (36.8/63.2)	15/21 (41.7/58.3)	0.779
Number of lesions			
GC (one/two or more)	18/1 (94.7/5.3)	33/3 (91.7/8.3)	>0.999
CRC (one/two or more)	17/2 (89.5/10.5)	31/5 (86.1/13.9)	>0.999
Histology			
GC (diff./undiff.)	6/13 (31.6/68.4)	16/19 (45.7/54.3)	0.391
CRC (W-MD/PD)	15/3 (83.3/16.7)	34/2 (94.4/5.6)	0.319
TNM stage *			
GC (I/II-III (%))	10/9 (52.6/47.4)	18/18 (50.0/50.0)	>0.999
CRC (I/II-III (%))	8/11 (42.1/57.9)	9/27 (25.0/75.0)	0.229
MSI status			

GC (MSS/MSI-H (%))	12/7 (63.2/36.8)	24/3 (88.9/11.1)	0.067
CRC (MSS/MSI-H (%))	10/8 (55.6/44.4)	27/7 (79.4/20.6)	0.071
Presence of P/LP variant			<0.001
no/yes (%)	12/7 (63.2/36.8)	34/2 (94.4/5.6)	
* AJCC8th			

GC; gastric cancer, CRC: colorectal cancer, MSI; microsatellite instability, MSS; microsatellite stable, MSI-H; MSI-high, P; pathogenic, LP; likely-pathogenic

The details of the detected germline variants are described in Table 3. In nine patients (seven in the young age group and two in the control group), the P/LP germline variants were detected, these being *MLH1* in seven patients, and *BML*, *BRCA1*, *MSH2*, and *MSH6* in one patient each. Other germline variants were classified as VUS. All nine patients who had the P/LP germline variant had one of the MMR-related germline variants such as *MLH1*, *MSH2*, or *MSH6*. There was at least one case of cancer history in their family, whereas the family history was not available for the other three patients. When comparing the frequency of the P/LP germline variants that were reported for the stomach cancer and colon cancer[13] to that of double primary cancers of the stomach and colon, *MLH1* germline variants were frequently observed in the double primary cancer patients (Table 4).

Table 3. Details of detected pathogenic/likely pathogenic germline variants in the enrolled population

Case_no.	Sex	Age		Family history			MSI status		ACMG classification	Gene	Accession	Nucleotide	Amino acid
		GC	CRC	GC	CRC	others	GC	CRC					
dou_002	M	44	40	NA	NA	NA	MSI-H	MSI-H	Likely pathogenic	*MLH1*	NM_000249.3	c.1721T>C	p.Leu574Pro
									Likely pathogenic	*BLM*	NM_000057.2	c.3651delA	p.Lys1217AsnfsTer62
dou_003	M	38	50	father	none	sister:brain, uterus, breast	MSI-H	MSI-H	Likely pathogenic	*MLH1*	NM_000249.3	c.1758dupC	p.Met587HisfsTer6
dou_005	F	44	42	father	none	none	MSS	MSI-H	Pathogenic	*MLH1*	NM_000249.3	c.208-1G>A	
dou_006	M	44	44	NA	NA	NA	MSI-H	MSI-H	Pathogenic	*BRCA1*	NM_007294.3	c.213-1G>A	
									Likely pathogenic	*MLH1*	NM_000249.3	c.2041G>A	p.Ala681Thr
dou_011	M	51	51	NA	NA	NA	MSI-H	MSI-H	Pathogenic	*MLH1*	NM_000249.3	c.790+2T>A	
dou_016	M	50	50	brother	brother, sister	mother: uterus	MSI-H	MSI-H	Likely pathogenic	*MLH1*	NM_000249.3	c.1758dupC	

dou_017	M	52	52	father	father	mother:uterus ca., brother: liver ca.	MSI-H	MSI-H	Pathogenic	*MLH1*	NM_000249.3	c.1559-2A>C	
dou_047	F	71	64	sister	none	father: liver ca., sister: uterus	NA	MSI-H	Likely pathogenic	*MSH6*	NM_000179.2	c.829G>T	p.Glu277Ter
dou_055	M	67	67	father, brother	none	none	MSI-H	MSI-H	Likely pathogenic	*MSH2*	NM_000251.2	c.965G>T	p.Gly322Val

Table 4. Comparison of the frequency of pathogenic/likely pathogenic germline variant between stomach and colon cancer in public data (TCGA) and double primary cancer of the stomach and colon

	Gastric Cancer (n = 443)	Colon Cancer (n = 419)	Double primary cancer (n = 55)	P-value	P-value
BRCA1	3 (0.68%)	1 (0.24%)	1 (1.8%)	0.375	0.219
BRCA2	4 (0.9%)	1 (0.24%)	0	>0.999	>0.999
ATM	7 (1.6%)	2 0.48%)	0	>0.999	>0.999
PALB2	5 (1.1%)	3 (0.72%)	0	>0.999	>0.999
MSH6	0	2 0.48%)	1 (1.8%)	0.11	0.39
SDHA	1 (0.23%)	1 (0.24%)	0	>0.999	>0.999
APC	1 (0.23%)	0	0	>0.999	>0.999
BLM	0	0	1 (1.8%)	0.11	0.116
MSH2	0	0	1 (1.8%)	0.11	0.116
MLH1	0	0	7 (12.7%)	<0.001	<0.001

TCGA; the cancer genome atlas

Table 5 shows the association between the clinic–pathologic characteristics of patients and the presence of the P/LP germline variants. Age under 55 years, family history of gastric cancer and Amsterdam II criteria, sum of number of lesions, histology of colorectal cancer, and MSI status of gastric cancer and colorectal cancer were significantly associated with the presence of the P/LP germline variants.

Table 5. Association between clinic–pathologic characteristics of patients and the presence of P/LP germline variants

	With P/LP (n = 9)	Without P/LP (n = 46)	P-value
Sex			0.473
Male	2 (22.2)	17 (37.0)	
Female	7 (77.8)	29 (63.0)	
Age			0.005
<55	7 (77.8)	12 (26.1)	
≥55	2 (22.2)	34 (73.9)	
Family history			
GC			0.003
No	0 (0)	28 (68.3)	
Yes	6 (100)	13 (31.7)	
CRC			>0.999
No	4 (66.7)	30 (73.2)	
Yes	2 (33.3)	11 (26.8)	
Any cancer			0.312
No	0 (0)	11 (26.8)	
Yes	6 (100)	30 (73.2)	
Amsterdam_I			>0.999

No	6 (100)	39 (95.1)	
Yes	0	2 (4.9)	
Amsterdam_II			0.003
No	1(16.7)	34 (82.9)	
Yes	5 (83.3)	7 (17.1)	
Sum of Number of lesions			0.003
2	4 (44.4)	42 (91.3)	
≥3	5 (55.6)	4 (8.7)	
Location of tumor			
GC			>0.999
MB/LB	8 (88.9)	37 (80.4)	
involving UB	1 (11.1)	9 (19.6)	
CRC			0.459
Right colon	5 (55.6)	17 (37.0)	
Left colon	4 (44.4)	29 (63.0)	
Histology			
GC			>0.999
differentiated	4 (44.4)	18 (40.0)	
undifferentiated	5 (55.6)	27 (60.0)	
CRC			0.028
WD/MD	6 (66.7)	43 (95.6)	
PD/mucinous	3 (33.3)	2 (4.4)	
TNM stage			
GC			>0.999
I	5 (55.6)	23 (50.0)	
II-III	4 (44.4)	23 (50.0)	
CRC			0.705
I	2 (22.2)	15 (32.6)	
II-III	7 (77.8)	31 (67.4)	
MSI status			
GC			<0.001

MSS	1 (12.5)	35 (92.1)	
MSI-H	7 (87.5)	3 (7.9)	
CRC			<0.001
MSS	0	37 (86.0)	
MSI-H	9 (100)	6 (14.0)	

P; pathogenic variant, LP; likely-pathogenic variant, GC; gastric cancer, CRC: colorectal cancer, MB; mid-body, LB; low-body, UB; upper-body, WD; well differentiated, MD; moderate differentiated, PD; poorly differentiated, MSI; microsatellite instability, MSS; microsatellite stable, MSI-H; MSI-high,

The accumulated number, incidence, and accumulated percentage of the P/LP germline variants at the age that the second cancer was diagnosed are depicted in Figure 2. The incidence of the P/LP germline variant steeply increased and reached a maximum at 44 years of age, and the accumulated percentage of P/LP reached 78% at 52 years of age. When comparing the association between the presence of the P/LP germline variant and the different cutoff points of ages, only 55 years old at the diagnosed last cancer was a statistically significant cutoff point ($p = 0.005$, Table 6).

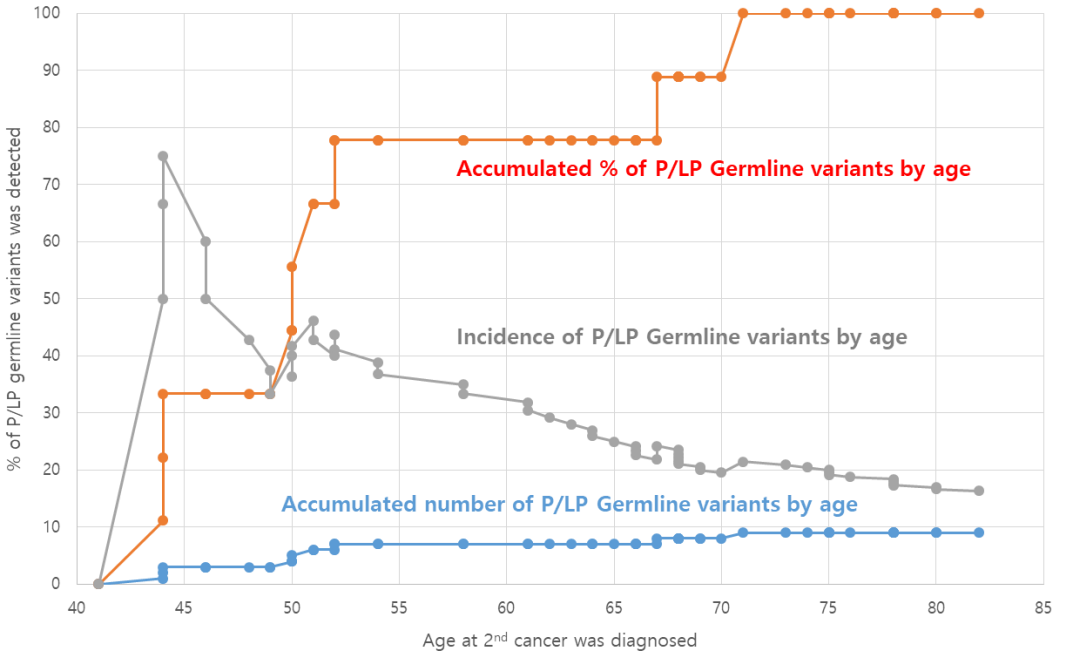

Figure 2. Accumulated number, incidence, and accumulated percentage of the P/LP germline variants at the age that the second cancer was diagnosed

P; pathogenic, LP; likely-pathogenic

Table 6. Association between age with various cutoff points and the presence of P/LP germline variants

Age (years)	With P/LP (n = 9)	Without P/LP (n = 46)	P-value
Age at dx of 2nd cancer			0.154
<50	3 (33.3)	6 (13.0)	
≥50	6 (66.7)	40 (87.0)	
Age at dx of 2nd cancer			0.005
<55	7 (77.8)	12 (26.1)	
≥55	2 (22.2)	34 (73.9)	
Age of any first cancer			0.092
<50	5 (55.6)	38 (82.6)	
≥50	4 (44.4)	8 (17.4)	

Dx; diagnosis, P; pathogenic variant, LP; likely-pathogenic variant

Figure 3 shows the representative pedigrees of the family of patients with the P/LP germline variant. In one family, we conducted a family germline test to identify the presence of the same P/LP germline variant in the family members. The patient with the *MLH1* LP germline variant underwent surgery for stomach and colorectal cancers at 40 years old, and died 7 years after treatment. His mother died from a type of uterine cancer, and her sister died of colorectal cancer in her 20's. His older brother had double primary cancers of the stomach and colon and had a confirmed *MLH1* germline variant. His niece was also confirmed as having the *MLH1* germline variant, although she has not had

cancer yet.

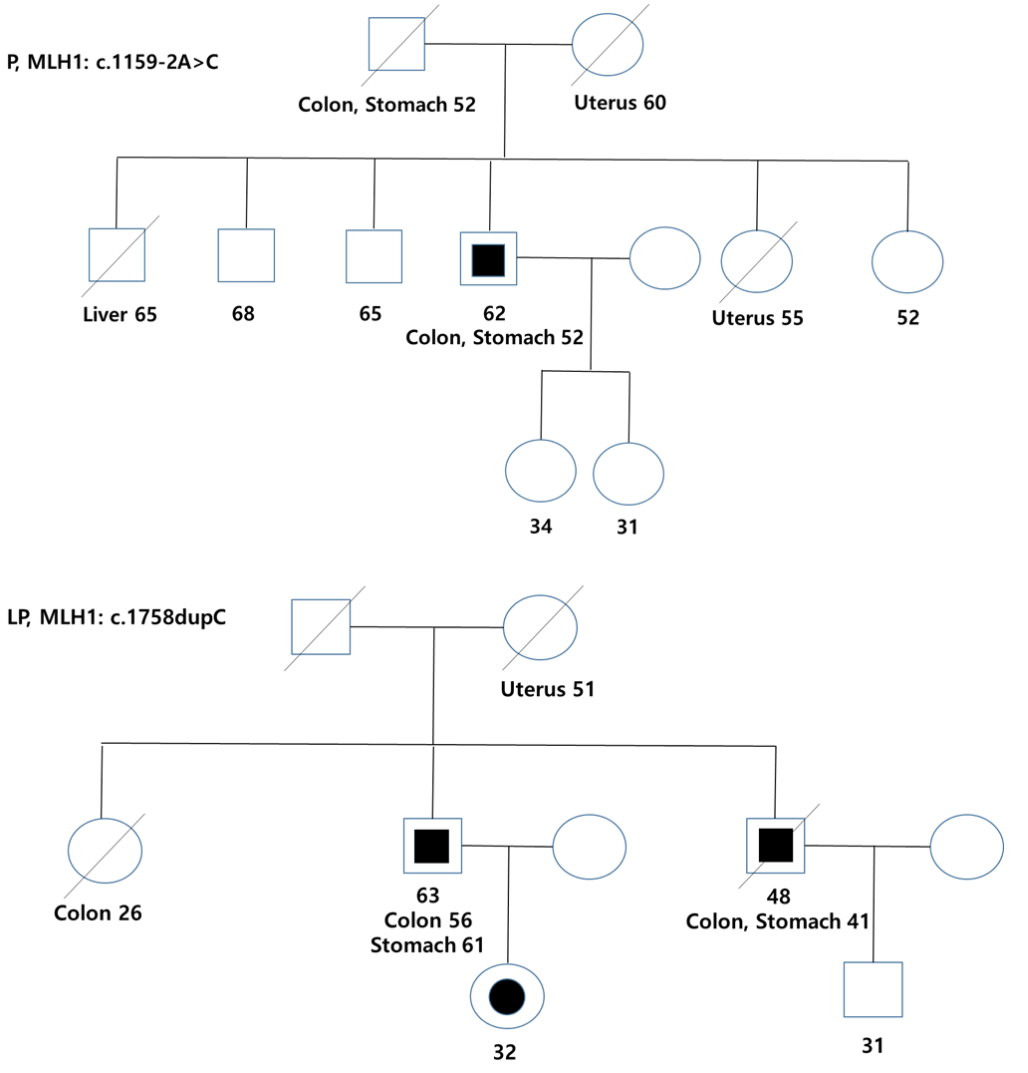

Figure 3. Pedigrees from a family with the *MLH1* germline LP variant. Filled square or circle represents confirmation of P/LP germline variant presence.

2. Somatic variants of young patients with double primary cancers of the stomach and colon

The WES analysis was conducted for patients with double primary cancers of the stomach and colon in the young age group (< 55 years) from their normal, gastric, and colorectal cancer tissue samples. Most of the patients with the P/LP germline variant were related to the MSI-H cancer type in both the stomach and colon; however, there was one patient with a MSS type of gastric cancer despite having the pathogenic *MLH1* germline variant (Table 7). In one patient, both gastric and colorectal cancers were of the MSI-H type; however, there was no P/LP germline variant or MMR-related VUS germline variant. Overall, the mutation burden (number of non-synonymous variants per megabase) was high in the MSI-H type of cancers. One patient with colorectal cancer was the MSS type but showed super-hyper mutations (dou_019), and this tumor was related to the *POLE* hotspot variant (p.Pro286Arg).

Table 7. Mutation burden of patients with double primary cancers of the stomach and colon in the young age group (< 55 years) and their MSI status and germline variants

	MSI/MMR status		Mutation burden (per Mb)		Multi-gene targeted NGS	
	GC	CRC	GC	CRC	Gene	ACMG
dou_001	MSS/pMMR	pMMR	4.12	5.01	APC	VUS
dou_002	**dMMR**	**dMMR**	**NA**	**74.33**	**MLH1, BLM1**	**Likely pathogenic**
dou_003	**MSI-H**	**MSI-H**	**35.41**	**NA**	**MLH1**	**Likely pathogenic**
dou_004	pMMR	NA	3.32	NA	CDH1	VUS
dou_005	**MSS**	**MSI-H**	**9.41**	**114.89**	**MLH1**	**Pathogenic**
dou_006	**dMMR**	**MSI-H**	**22.02**	**NA**	**BRCA1, MLH1**	**Pathogenic**
dou_007	pMMR	pMMR	7.16	NA	MSH6	VUS
dou_008	MSS/pMMR	pMMR	3.24	6.00	MUTYH	VUS
dou_009	MSI-	dMMR	29.94	NA	PAX6	VUS

	H/dMMR					
dou_010	pMMR	pMMR	5.74	5.66	CHEK2	VUS
dou_011	**dMMR**	**dMMR**	**29.94**	**43.25**	**MLH1**	**Pathogenic**
dou_012	pMMR	pMMR	3.85	5.47	None	None
dou_013	MSS/pMMR	pMMR	3.37	NA	FLCN	VUS
dou_014	pMMR	MSS	4.55	NA	CDKN2A	VUS
dou_015	MSS/pMMR	pMMR	3.94	5.06	MLH3	VUS
dou_016	**dMMR**	**MSI-H/dMMR**	**56.34**	**49.08**	**MLH1**	**Likely pathogenic**
dou_017	**dMMR**	**dMMR**	**26.28**	**43.24**	**MLH1**	**Pathogenic**
dou_018	NA	pMMR	NA	5.28	MLH3	VUS
dou_019	pMMR	pMMR	3.85	391.40	MSH2	VUS

GC; gastric cancer, CRC: colorectal cancer, MSI; microsatellite instability, MMR; mismatch repair, NGS; next generation sequencing, VUS; variant of unknown significance, NA; not available

When comparing somatic variants in the gene level between gastric and colorectal cancers in each patient, there was only a small number of intersection variant genes, with most of these detected in cancers with a high number of variants (Figure 4).

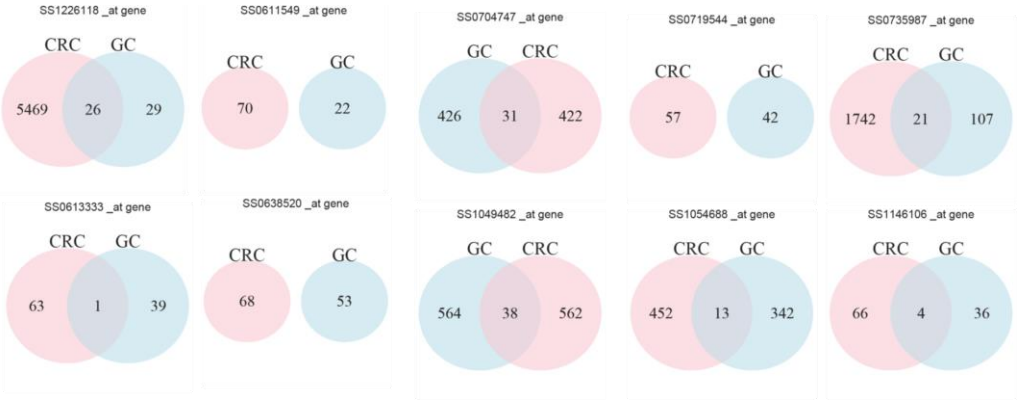

Figure 4. Number of intersections and union variants between gastric (GC) and colorectal cancers (CRC) in each patient

The list of recurrently detected overall, gastric cancer, and colorectal cancer specific somatic non-synonymous variants in patients with double primary cancers of the stomach and colon in the young age group (< 55 years) are depicted in Figure 5. *APC*, *MYCBP2*, *PCDH15*, and *RYR2* were the most frequently observed somatic non-synonymous variants in patients (n = 8). Figure 6 shows the top list of gastric cancer- and colorectal cancer-specific somatic variants (non-synonymous) in patients with double primary cancers of

the stomach and colon in the young age group (< 55 years old). Some somatic variants were detected only in stomach cancer or in colorectal cancer.

Figures 7–10 show the results from the public data (the cancer genome atlas) of somatic variants by age (< 55 years vs. ≥ 55 years) for stomach and colorectal cancers. Some somatic non-synonymous variants were observed with similar frequency by the age of patients in both stomach and colorectal cancers. However, *PCDH10* was significantly frequently observed in the older age group (≥ 55 years old) of patients with colorectal cancer (Figure 8) and *GPR98 and ZBTB20* in stomach cancer (Figure 10). *CDH1* somatic variants in stomach cancer were more frequently observed in the young age group than in the old age group (Figure 10).

Figures 11 and 12 show the comparison of the mutational signature analysis[14] between gastric cancer and colorectal cancer in the young age group in each patient by with/without P/LP germline variants. In patients with the P/LP germline variants, because most of the tumors were related to their MSI-H status, gastric cancer and colorectal cancer shared similar mutational signatures related to MMR, hypermutation, and MSI. In the patient who had the pathogenic MLH1 germline variant but gastric cancer was of the MSS type, there was no similar

mutational signature between gastric cancer and colorectal cancer. In the patients who did not have the P/LP germline variants, only age was the common mutational signature between gastric and colorectal cancers, despite the age of the patients being < 55 years.

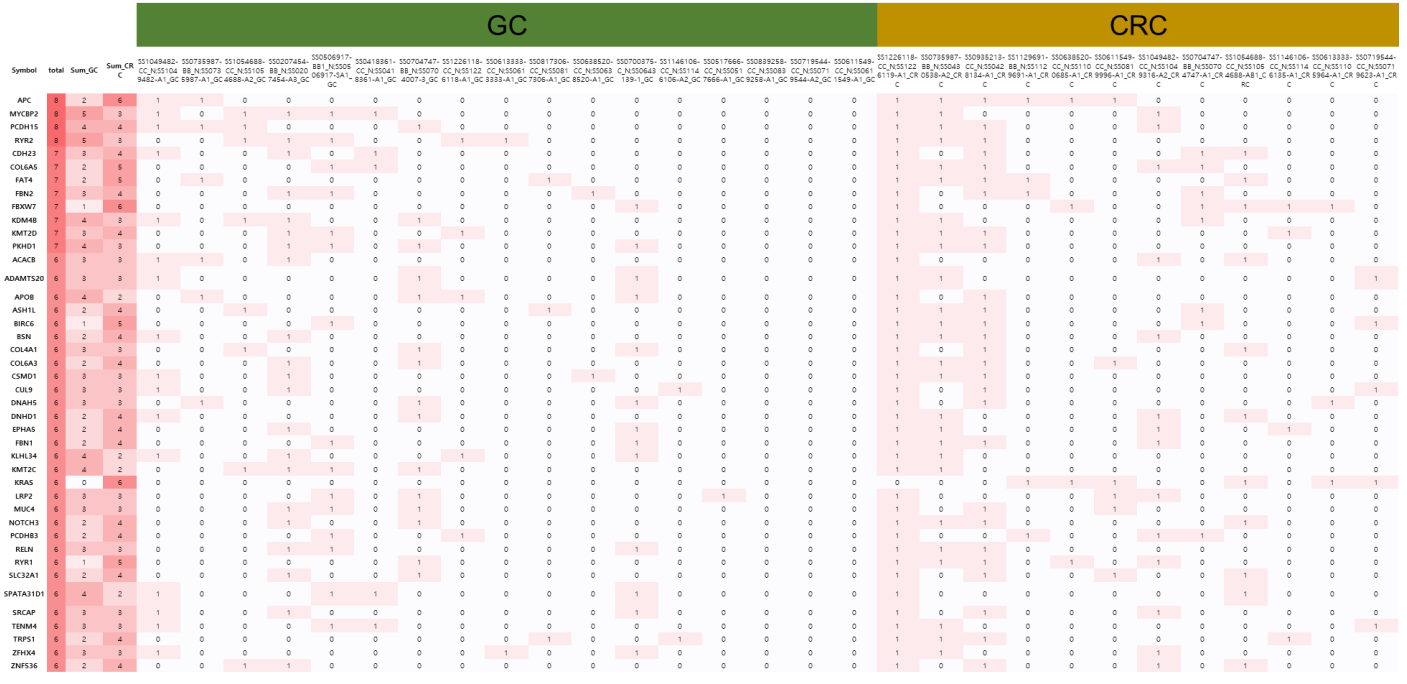

Figure 5. List of recurrently detected overall, gastric cancer, and colorectal cancer specific somatic non-synonymous variants in patients with double primary cancer of the stomach (GC) and colon (CRC) in the young age group (< 55 years old)

GC-specific genes

Symbol	total	Sum_GC	Sum_CRC
ABCG8	3	3	0
ADAMTSL3	3	3	0
AUTS2	3	3	0
CACNA1C	3	3	0
DND1	3	3	0
DPP6	3	3	0
MAP1S	3	3	0
MCF2L	3	3	0
NRROS	3	3	0
TMC8	3	3	0
UBXN6	3	3	0
ABCA2	2	2	0
ACACA	2	2	0
AGPAT3	2	2	0
ANKRD30A	2	2	0

CRC-specific genes

Symbol	total	Sum_GC	Sum_CRC
KRAS	6	0	6
CTNNB1	5	0	5
PCDH10	5	0	5
C10orf90	4	0	4
DYNC1H1	4	0	4
ECM2	4	0	4
IGFN1	4	0	4
MUC12	4	0	4
MYH13	4	0	4
PCDH7	4	0	4
PDZD4	4	0	4
PLXNA4	4	0	4
SCAF8	4	0	4
SKOR2	4	0	4
TBCEL	4	0	4

Figure 6. Top list of gastric cancer (GC)- and colorectal cancer (CRC)-specific somatic variants (non-synonymous) in patients with double primary cancers of the stomach and colon in the young age group (< 55 years old)

TCGA COAD
Top 40 genes per age group

coad_<55	coad_≥55
APC	APC
TP53	TP53
KRAS	KRAS
PIK3CA	PIK3CA
CSMD3	FAT4
RYR2	ZFHX4
FAT4	RYR2
DNAH5	PCLO
GPR98	DNAH5
KMT2D	LRP1B
ATM	CSMD1
DMD	ABCA13
DNAH2	CSMD3
TRPS1	RYR1
ZFHX3	FBXW7
SPTA1	RYR3
VPS13B	FAT3
DOCK2	BRAF
KMT2C	LRP2
RELN	SDK1
HYDIN	DCHS2
LRP1B	UNC80
ZFHX4	GPR98
FAT3	HYDIN
LRP2	PCDH15
CUBN	SOX9
ABCA13	ANK2
UNC13C	CACNA1E
PTEN	XIRP2
RNF213	MDN1
RYR3	KIAA1109
BSN	AMER1
FNDC1	DMD
PCDH18	KMT2D
ZNF469	ATM
PKHD1	COL6A3
FBXW7	DNAH8
DYNC2H1	ANK3
SACS	SPTA1
DNAH1	FAT2

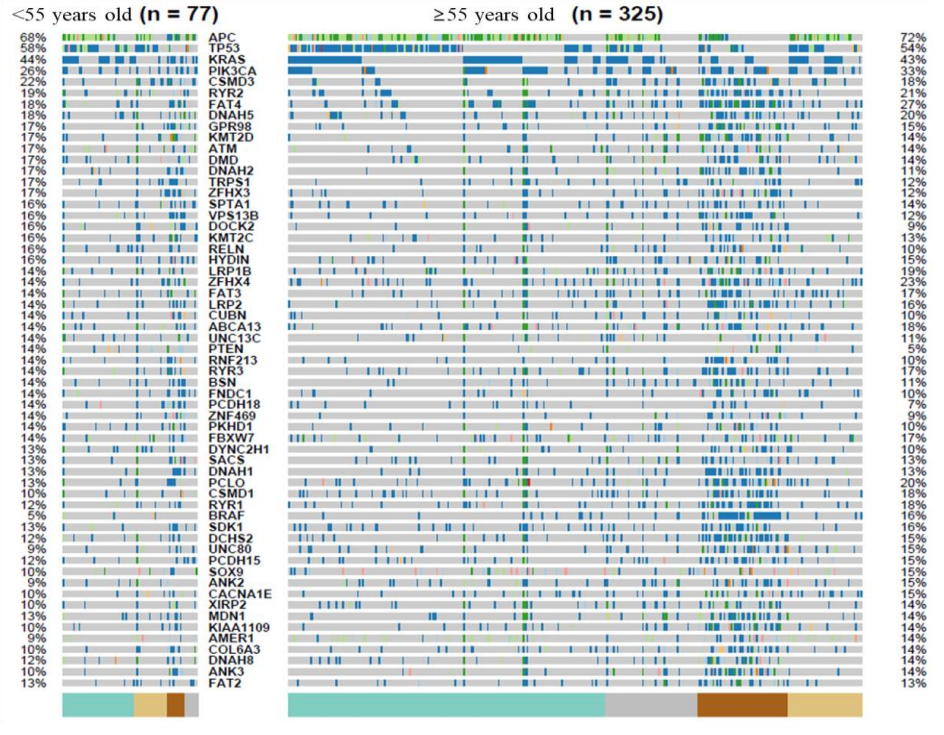

Figure 7. List of most frequent somatic variants in colorectal cancer by age (< 55 and ≥ 55 years old) from the public database (TCGA)

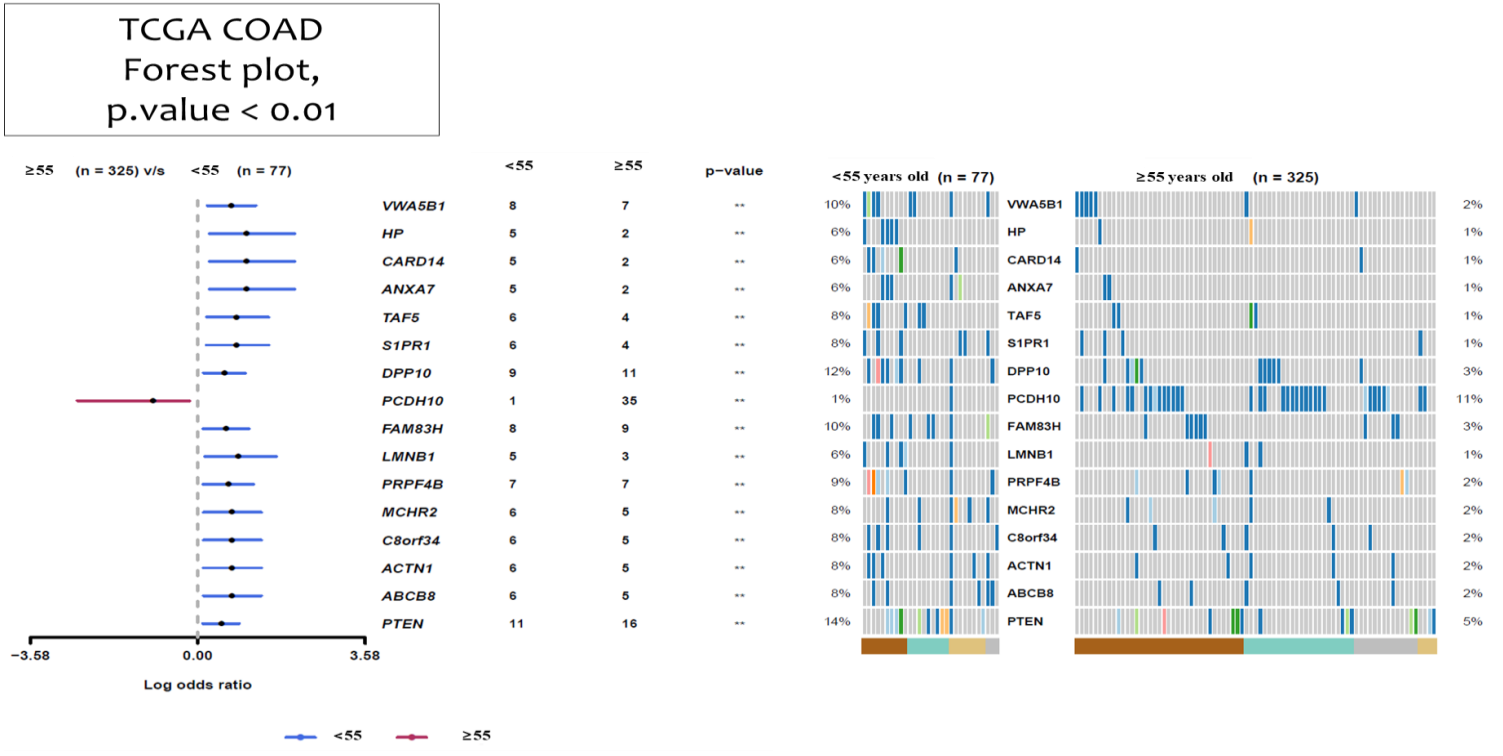

Figure 8. Somatic variants of colorectal cancer where the incidence was significantly different between young age group (< 55 years old) and others in the public database (TCGA)

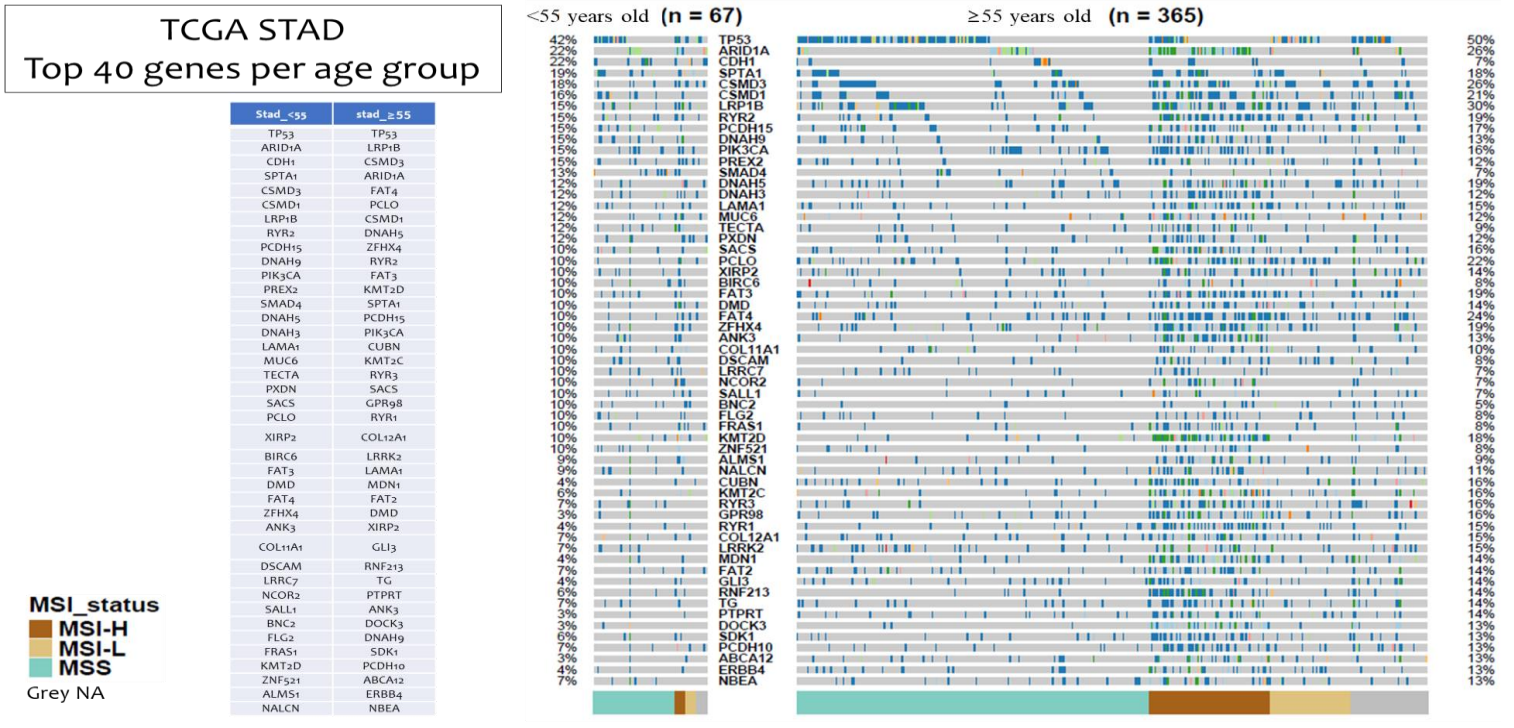

Figure 9. List of most frequent somatic variants in stomach cancer by age (< 55 and ≥ 55 years old) in the public database (TCGA)

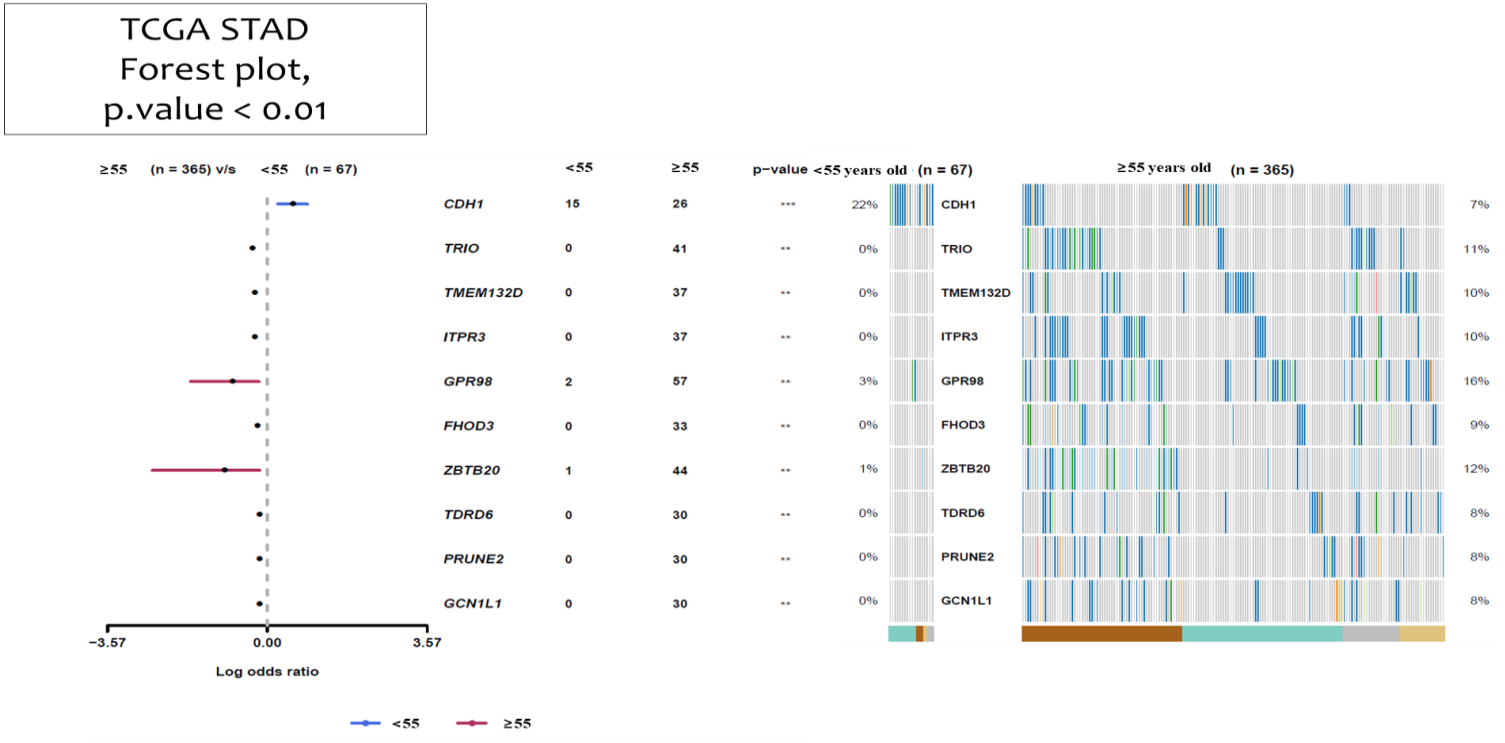

Figure 10. Somatic variants of stomach cancer where the incidence was significantly different between the young age group (< 55 years old) and others in the public database (TCGA)

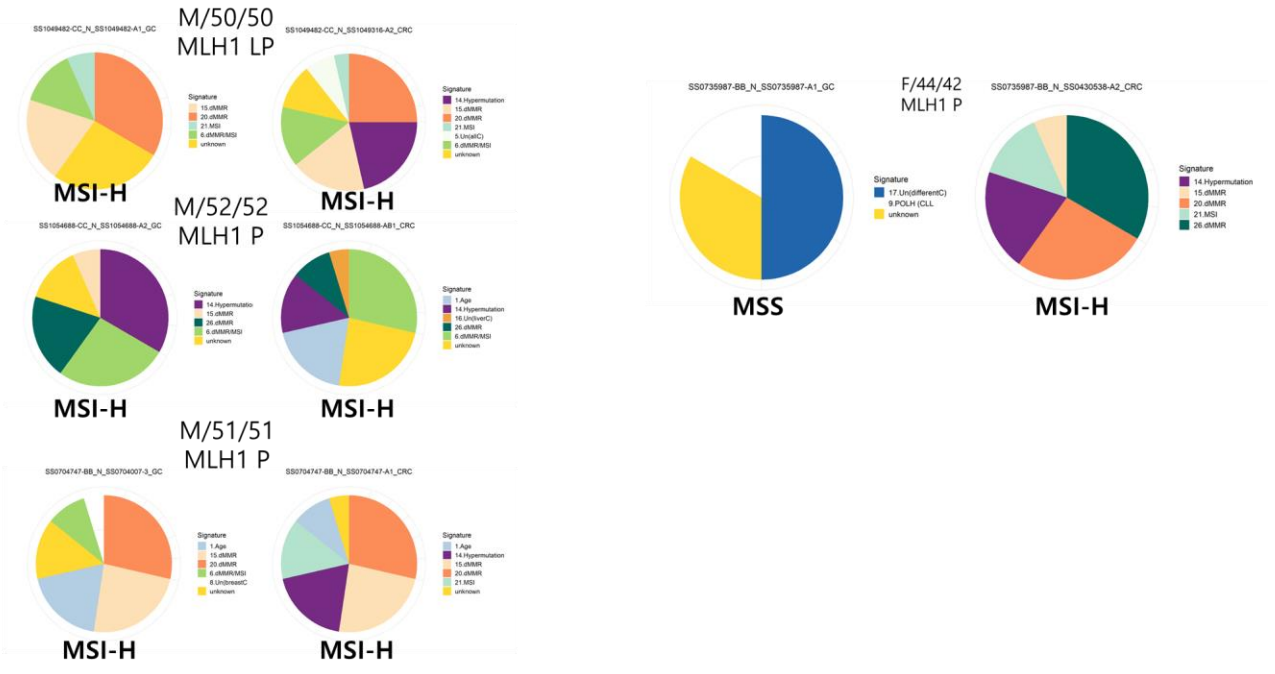

Figure 11. Comparison of mutational patterns and etiology analysis between gastric and colorectal cancer in each patient with double primary cancers of the stomach and colon in the young age group with the P/LP germline variants

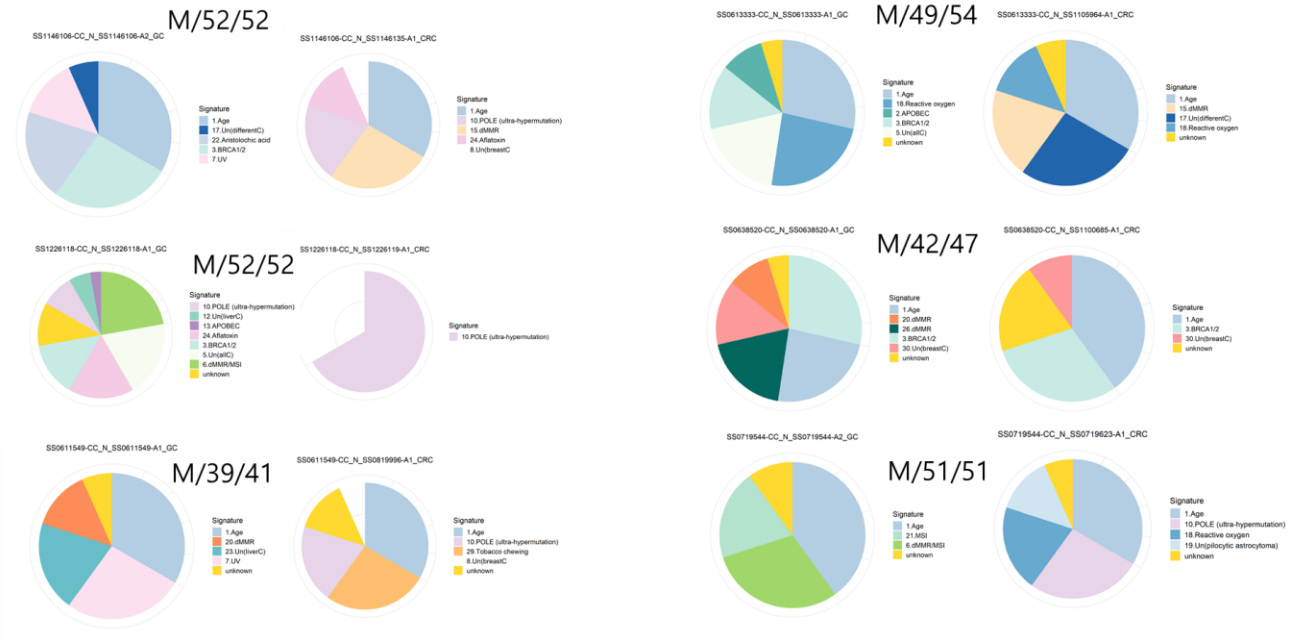

Figure 12. Comparison of mutational patterns and etiology analysis between gastric and colorectal cancer in each patient with double primary cancers of the stomach and colon in the young age group without the P/LP germline variants

IV. DISCUSSION

In the present study, Lynch syndrome related P/LP germline variants, such as *MLH1*, *MSH2*, and *MSH6*, were mainly observed in young patients with double primary cancers of the stomach and colon. Lynch syndrome is well known to be related to colorectal cancer, endometrial cancer, and stomach cancer; however, stomach cancer-related Lynch syndrome has rarely been evaluated because stomach cancer does not frequently occur in western countries.[15,16] The detection rate for the P/LP germline variants by single/multi-gene NGS panel has been reported as 1% in the overall general population and 5%–8% in patients with cancer.[13,17] In the present study, the P/LP germline variants were detected in 16% of patients with double primary cancers of the stomach and colon; therefore, targeting double primary cancer patients would be an effective way to screen the germline test to identify the super-high risk group of the cancer population. In addition, the present results showed that there are some factors that should be considered for indicating the germline test, e.g., family history, age, MSI status. The current Amsterdam criteria recommends that having a diagnosed cancer at less than 50 years old is one of the criteria for undertaking the germline test;[18] however, the present results showed the possibility that extending the age criteria to 55 years for the second primary cancer would be

effective for detecting more patients and families that were affected by the family cancer syndrome, as 37% of patients (7 out of 19) with double primary cancers of the stomach and colon were in the young age group (< 55 years old). MSI was one of the risk factors for the P/LP germline variant in the present study and most of the detected P/LP germline variants were MMR related. However, not all the patients with MMR-related P/LP germline variants had MSI-H tumors, and not all MSI-H tumors were related to the P/LP germline variant because there were sporadic MSI-H tumors. Similar to the well-known mechanism of cancer, the two-hit mechanism of inactivation of tumor suppressor genes, germline and somatic bi-allelic alteration, is required to induce carcinogenesis and a single germline "hit" is not enough.[19-22] In this sense, this finding is understandable. Therefore, integrative consideration of the risk factors and incidence of double primary cancer at a young age, as related to MSI-H tumor and family history, is required for clinical practice.

Determining the individuals with the P/LP germline variant would be a very effective strategy against cancer because of the following reasons:[1,9,23] 1) one individual with the P/LP germline variant represents that 2–4 generations of his/her relatives will be affected by the same P/LP germline variant, 2) over 80% of the affected individuals with the P/LP germline variant will be diagnosed with

cancer during their lifetime, and 3) prevention and early detection of cancer is the most effective way to treat cancer. Consequently, the germline test must be expanded to individuals who are suspected to be related by the family cancer syndrome, and we must establish a system of integrative care for this syndrome.

One hypothesis of the present study was that the double primary cancer that occurred in one individual would have similar genomic characteristics. However, we failed to find the intersection between stomach and colorectal cancers; therefore, it is difficult to provide a strategy that simultaneously targets both gastric and colorectal cancers. In the mutational signature analysis, similarity was observed in cases in which both cancers were of the MSI-H type and most of them were related to the Lynch syndrome. Despite a patient having the P/LP germline variant, if one of the tumors in the stomach and colon was not MSI-H, there was no similar mutational signature between the tumors in one individual. Therefore, the origin and etiology of the double primary cancers are different even if they occurred in one individual at an early age.

One of the most commonly detected mutational signatures in the double primary cancers of the stomach and colon was the age-related signatures, despite these cancers occurring at a relatively young age (< 55 years old). When we consider that the mean age of obtaining stomach and colorectal cancers in Korea

is approximately 60 years of age, we need to think about the meaning of age-related signatures. It might be possible that there is a cancer-related biological age rather than a chronological age, with further studies on this required in the future.

There were some limitation to the present study. Despite this being the first study, to the best of our knowledge, to target genomic characteristics of patients with double primary cancer, the population number was too small to provide strong evidence. In addition, we failed to find a novel germline variant that caused either gastric or colon cancers.

V. CONCLUSION

The MMR-related germline P/LP variants were mainly detected in patients with double primary cancers of the stomach and colon in the young age group. Those who have had double primary cancer and who were less than 55 years, have a family history of gastric cancer, and MSI-H type of tumors are recommended to undergo a germline genomic test. The common variants between stomach and colorectal cancers in one individual were rarely detected, especially in the MSS type of cancer; consequently, simultaneously targeting both tumors in one patient would be a difficult strategy for clinical practice.

References

1. Wood ME, Vogel V, Ng A, Foxhall L, Goodwin P, Travis LB. Second malignant neoplasms: assessment and strategies for risk reduction. *J Clin Oncol.* 2012;30(30):3734-3745.
2. Vogt A, Schmid S, Heinimann K, et al. Multiple primary tumours: challenges and approaches, a review. *ESMO Open.* 2017;2(2):e000172.
3. Cho I, An JY, Kwon IG, et al. Risk factors for double primary malignancies and their clinical implications in patients with sporadic gastric cancer. *Eur J Surg Oncol.* 2014;40(3):338-344.
4. Jung KW, Won YJ, Kong HJ, Lee ES. Cancer statistics in Korea: Incidence, mortality, survival, and prevalence in 2016. *Cancer Res Treat.* 2019;51(2):417-430.
5. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature.* 2012;487(7407):330-337.
6. Cancer Genome Atlas Research N. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature.* 2014;513(7517):202-209.
7. Cancer Genome Atlas Research N, Kandoth C, Schultz N, et al. Integrated genomic characterization of endometrial carcinoma. *Nature.* 2013;497(7447):67-73.
8. Park YJ, Shin KH, Park JG. Risk of gastric cancer in hereditary nonpolyposis colorectal cancer in Korea. *Clin Cancer Res.* 2000;6(8):2994-2998.
9. Eoh KJ, Kim JE, Park HS, et al. Detection of germline mutations in patients with epithelial ovarian cancer using multi-gene panels: Beyond BRCA1/2. *Cancer Res Treat.* 2018;50(3):917-925.
10. An JY, Kim H, Cheong JH, Hyung WJ, Kim H, Noh SH. Microsatellite instability in sporadic gastric cancer: its prognostic role and guidance for 5-FU based chemotherapy after R0 resection. *Int J Cancer.* 2012;131(2):505-511.
11. Kim SY, Choi YY, An JY, et al. The benefit of microsatellite instability is attenuated by chemotherapy in stage II and stage III gastric cancer: Results from a large cohort with subgroup analyses. *Int J Cancer.* 2015;137(4):819-825.
12. Umar A, Boland CR, Terdiman JP, et al. Revised Bethesda guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst.* 2004;96(4):261-268.

13. Huang KL, Mashl RJ, Wu Y, et al. Pathogenic germline variants in 10,389 adult cancers. *Cell.* 2018;173(2):355-370 e314.
14. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature.* 2013;500(7463):415-421.
15. Charames GS, Millar AL, Pal T, Narod S, Bapat B. Do MSH6 mutations contribute to double primary cancers of the colorectum and endometrium? *Hum Genet.* 2000;107(6):623-629.
16. Cybulski C, Nazarali S, Narod SA. Multiple primary cancers as a guide to heritability. *Int J Cancer.* 2014;135(8):1756-1763.
17. Mersch J, Brown N, Pirzadeh-Miller S, et al. Prevalence of variant reclassification following hereditary cancer genetic testing. *JAMA.* 2018;320(12):1266-1274.
18. Samadder NJ, Smith KR, Wong J, et al. Cancer risk in families fulfilling the Amsterdam criteria for Lynch syndrome. *JAMA Oncol.* 2017;3(12):1697-1701.
19. Polak P, Kim J, Braunstein LZ, et al. A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. *Nat Genet.* 2017;49(10):1476-1486.
20. Riaz N, Blecua P, Lim RS, et al. Pan-cancer analysis of bi-allelic alterations in homologous recombination DNA repair genes. *Nat Commun.* 2017;8(1):857.
21. Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A.* 1971;68(4):820-823.
22. Buckley AR, Ideker T, Carter H, Harismendy O, Schork NJ. Exome-wide analysis of bi-allelic alterations identifies a Lynch phenotype in the Cancer Genome Atlas. *Genome Med.* 2018;10(1):69.
23. Beitsch PD, Whitworth PW, Hughes K, et al. Underdiagnosis of hereditary breast cancer: Are genetic testing guidelines a tool or an obstacle? *J Clin Oncol.* 2019;37(6):453-460.

ABSTRACT(IN KOREAN)

한국인 다빈도 젊은 위.대장 중복암 위험인자발굴과 유전체 특징 규명

<지도교수 정재호>

연세대학교 대학원 의학과

최윤영

암치료 전략의 발전에 따라 암환자의 예후가 좋아지면서 두 가지 암 이상이 발생하는 중복암 환자가 늘어나고 있으며, 한국에 가장 흔한 중복암은 위암과 대장암의 조합이다. 비교적 젊은 연령에 두 가지 암이 발생한 경우 환경적 요인보다는 유전적 요인에 영향을 받았을 가능성이 크지만, 아직까지 중복암의 유전적 변이에 대한 것은 잘 알려지지 않았다. 또한 두 암이 한 환자에서 발생했을 경우 유전체적 유사성이 있을 수 있으며, 만일 두 암이 치료가 가능한 공통된 유전체 변이를 공유할 경우 새로운 암치료 전략이 될 수 있을 것이다. 이에 우리는 2000년부터 2016년 까지 세브란스병원에서 위암 및 대장암으로 치료를 받은 환자들의 생식세포 및 체세포의 유전체적 특성에 대한 연구를 하였다. 생식세포에 대한 다유전자 염기서열분석에서 mismatch repair(MMR) 관련된 유전자의 생식세포 변이가 55명 중 9명에서(16.4%) 확인되었다. 위.대장 중복암 환자 중에서도 55세 이전에 두 암이 모두 발생한 경우, Amsterdam_II 적응증, 위 혹은 대장암이 두 개 이상인 경우, 그리고 현미부수체불안전성(microsatellite instability) 암인 경우 생식세포 변이가 있는 위험성이 높았다. 55세 미만에 발생한 중복암 환자들의 전엑솜염기서열 분석을 이용한 체세포

유전체 분석상, 한 환자에서 발생한 위암 및 대장암 간에 공통된 체세포 변이가 매우 드물게 발견 되었고, 이는 주로 현미부수체 불안전성 암인 경우였고, 현미부수체 안전성 암일 경우에는 공통변이가 없었다. 체세포 변이 시그니처 분석 상에서는 현미부수체불안전성암에서 주로 MMR 결핍, 다빈도변이(hypermutation), 및 현미부수체불안전성 관련 시그니처가 주로 발견되었다. 이외에는 연령을 제외하고는 두 암 간의 공통되는 시그니처가 없었는데, 환자들이 모두 55세 미만의 젊은 연령에 암이 발생한 점을 고려하면 특이한 점이라 할 수 있다. 55세 미만에 발생한 중복암, 위암의 가족력, 현미부수체불안전성암을 가진 경우 유전자 검사를 시행하는 적응증이 될 수 있을 것이다. 한 환자에서 발생했다고 하더라도 위암 및 대장암 간의 공통되는 유전체 변이는 현미부수체불안전성이 아닌 경우 드문 것으로 보여, 하나의 약물로 두 암을 모두 치료하는 전략은 임상적으로 어려울 것으로 생각된다.

핵심되는 말: 중복암, 위암, 대장암, 차세대염기서열분석, 유전체