

UC San Diego

UC San Diego Previously Published Works

Title

Prevalence of cancer susceptibility variants in patients with multiple Lynch syndrome related cancers

Permalink

<https://escholarship.org/uc/item/6wj038fj>

Journal

Scientific Reports, 11(1)

ISSN

2045-2322

Authors

Choi, Yoon Young

Shin, Su-Jin

Lee, Jae Eun

et al.

Publication Date

2021

DOI

10.1038/s41598-021-94292-4

Peer reviewed



OPEN

Prevalence of cancer susceptibility variants in patients with multiple Lynch syndrome related cancers

Yoon Young Choi^{1,2,3}, Su-Jin Shin⁴, Jae Eun Lee³, Lisa Madlensky^{8,9}, Seung-Tae Lee^{5,6}, Ji Soo Park^{5,7}, Jeong-Hyeon Jo⁴, Hyunki Kim⁴, Daniela Nachmanson¹¹, Xiaojun Xu⁸, Sung Hoon Noh², Jae-Ho Cheong^{2,3}✉ & Olivier Harismendy^{8,10}✉

Along with early-onset cancers, multiple primary cancers (MPCs) are likely resulting from increased genetic susceptibility; however, the associated predisposition genes or prevalence of the pathogenic variants genes in MPC patients are often unknown. We screened 71 patients with MPC of the stomach, colorectal, and endometrium, sequencing 65 cancer predisposition genes. A subset of 19 patients with early-onset MPC of stomach and colorectum were further evaluated for variants in cancer related genes using both normal and tumor whole exome sequencing. Among 71 patients with MPCs, variants classified to be pathogenic were observed in 15 (21.1%) patients and affected Lynch Syndrome (LS) genes: *MLH1* (n = 10), *MSH6* (n = 2), *PMS2* (n = 2), and *MSH2* (n = 1). All carriers had tumors with high microsatellite instability and 13 of them (86.7%) were early-onset, consistent with LS. In 19 patients with early-onset MPCs, loss of function (LoF) variants in *RECQL5* were more prevalent in non-LS MPC than in matched sporadic cancer patients (OR = 31.6, 2.73–1700.6, $p = 0.001$). Additionally, there were high-confidence LoF variants at *FANCG* and *CASP8* in two patients accompanied by somatic loss of heterozygosity in tumor, respectively. The results suggest that genetic screening should be considered for synchronous cancers and metachronous MPCs of the LS tumor spectrum, particularly in early-onset. Susceptibility variants in non-LS genes for MPC patients may exist, but evidence for their role is more elusive than for LS patients.

The identification of inherited DNA variants in cancer predisposition genes is clinically important for both cancer prevention and treatment as it can reduce cancer morbidity and mortality in individuals carrying pathogenic variants¹. The prevalence of pathogenic variants in cancer predisposition genes is lower than 1% of the healthy population and 5–8% in patients diagnosed with cancer^{2–4}. Pathogenic variants in cancer predisposition genes are therefore rare. On the other hand, wide-spread surveillance screening may result in false positive findings or findings of uncertain significance, yielding to increased anxiety and unnecessary or ineffective clinical procedures. Thus, continued research supporting more effective and precise genetic screening guidelines is needed¹.

Multiple primary cancers (MPCs), referring to two or more histologically distinct cancers diagnosed in one individual, is gradually recognized as an important medical problem with a reported frequency of 2–17% in the cancer population^{5,6}. Along with early-onset cancer, MPC have been regarded as high-risk of heritability⁷ and those patients therefore represent excellent candidates for genetic screening. However, limited data is available for the prevalence and type of pathogenic variants in the cancer predisposition genes of these patients.

¹Department of Surgery, CHA University School of Medicine, Pocheon-si, Korea. ²Department of Surgery, Yonsei University Health System, Yonsei University College of Medicine, 50 Yonsei-ro, Seodaemun-gu, Seoul 120-752, Korea. ³Yonsei Biomedical Research Institute, Yonsei University Health System, Yonsei University College of Medicine, Seoul, Korea. ⁴Department of Pathology, Yonsei University Health System, Yonsei University College of Medicine, Seoul, Korea. ⁵Hereditary Cancer Clinic, Yonsei University Health System, Yonsei University College of Medicine, Seoul, Korea. ⁶Department of Laboratory Medicine, Yonsei University Health System, Yonsei University College of Medicine, Seoul, Korea. ⁷Department of Medicine, Yonsei University Health System, Yonsei University College of Medicine, Seoul, Korea. ⁸Moores Cancer Center and Division of Biomedical Informatics Department of Medicine, University of California San Diego School of Medicine, 3855 Health Sciences Dr, La Jolla, CA 92037, USA. ⁹Department of Family Medicine and Public Health, University of California San Diego School of Medicine, San Diego, CA, USA. ¹⁰Department of Medicine, University of California San Diego School of Medicine, San Diego, CA, USA. ¹¹Bioinformatics and Systems Biology Graduate Program, University of California San Diego School of Medicine, San Diego, USA. ✉email: JHCHEONG@yuhs.ac; oharismendy@health.ucsd.edu

Case ID	Cancer type	Age at second diagnosis [‡]	Gene	Nucleotide change	Amino acid change	Mutation type	ACMG criteria	Class [†]	MSI status
dou_002	CRC/GC	<55	<i>MLH1</i>	c.1721 T>C	p.Leu574Pro	Missense	PM2, PP1, PP2, PP3, PP5	LP	MSI-H/MSI-H
dou_003	GC/CRC	<55	<i>MLH1</i>	c.1758dupC	p.Met587HisfsTer6	Frameshift_insertion	PVS1, PM2	LP	MSI-H/MSI-H
dou_005	CRC/GC	<55	<i>MLH1</i>	c.208-1G>A	NA	Splice_acceptor_variant	PVS1, PM2, PP5	P	MSI-H/MSS
dou_006	GC/CRC	<55	<i>MLH1</i>	c.2041G>A	p.Ala681Thr	Missense	PM2, PM5, PP3, PP5	LP	MSI-H/MSI-H
			<i>BRCA1</i>	c.213-1G>A	NA	Splice_acceptor_variant	PVS1, PM2, PP5	P	
dou_011	GC/CRC	<55	<i>MLH1</i>	c.790 + 2 T>A	NA	Splice_donor_variant	PVS1, PM2, PP5	P	MSI-H/MSI-H
dou_016	GC/CRC	<55	<i>MLH1</i>	c.1758dupC	p.Met587HisfsTer6	Frameshift_insertion	PVS1, PM2	LP	MSI-H/MSI-H
dou_017	GC/CRC	<55	<i>MLH1</i>	c.1559-2A>C	NA	Splice	PVS1, PM2, PP5	P	MSI-H/MSI-H
dou_047	CRC/GC	>55	<i>MSH6</i>	c.829G>T	p.Glu277Ter	Splice_acceptor_variant	PVS1, PM2	LP	MSI-H/NA
dou_056	CRC/EC	>55	<i>PMS2</i>	c.1738A>T	p.Lys580Ter	Nonsense	PVS1, PM2, PP5	P	MSI-H/MSS
dou_061	CRC/EC	<55	<i>MSH6</i>	c.3477C>G	p.Tyr1159Ter	Nonsense	PVS1, PM2, PP5	P	NA/MSI-H
dou_062	CRC/EC	<55	<i>PMS2</i>	c.943C>T	p.Arg315Ter	Nonsense	PVS1, PM2, PP5	P	MSI-H/MSI-H
dou_065	CRC/EC	<55	<i>MLH1</i>	c.67G>T	p.Glu23Ter	Nonsense	PVS1, PM2, PP5	P	MSI-H/MSI-H
dou_069	CRC/EC/Klaskin	<55	<i>MLH1</i>	c.67G>T	p.Glu23Ter	Nonsense	PVS1, PM2, PP5	P	MSI-H/MSI-H/ MSI-H
dou_070	EC/CRC/GC	<55	<i>MLH1</i>	exon 7–9 deletion	NA	NA	PVS1, PM2	LP	MSI-H/NA/MSI-H
dou_071	CRC/GC/EC/Lung	<55	<i>MSH2</i>	c.942 + 3A>T	NA	Exon loss	PVS1, PM2	LP	MSI-H/MSI-H/NA/ MSI-H

Table 1. Summary of clinicopathologic characteristics of patients with multiple Lynch related primary cancers with pathogenic/likely-pathogenic germline variants by targeted panel. *The order of MSI/MMR status is matched to cancer type. [†]LP: likely pathogenic, P: Pathogenic. [‡]The age of the patients is described as a range to preserve the anonymity of the patients as recommended by institutional review board. ACMG, the American College of Medical Genetics and Genomics.

Gastric cancer (GC) and colorectal cancer (CRC) are the two most common cancers in Korea, with 5-year survival up to 70%⁸ and their combination is the most common instance of MPCs, representing ~2% of Korean GC or CRC cases^{9,10}. About 15% of each of the cancer types can be characterized by elevated microsatellite instability (MSI) caused by mismatch repair (MMR) deficiency^{11–15}. Together with Endometrial cancer (EC)—another cancer characterized by MSI¹⁶—GC and CRC are the main cancer types associated with Lynch Syndrome (LS), caused by pathogenic variants in MMR genes, resulting in up to an 80% lifetime risk of cancers^{17,18}. In particular, 2–5% of CRC and EC, as well as ~15% of early-onset CRC are thought to be associated with LS^{19,20}. MPCs with cancer types in the spectrum of LS therefore represent some of the strongest candidates for genetic screening, as the prevalence of LS among MPC is unknown.

Here we present the results of the genetic screening of 71 patients with MPC of the stomach, colorectal and endometrium. We evaluate the distribution and prevalence of pathogenic variants in 65 cancer predisposition genes, including LS susceptibility genes, as a function of age and clinicopathological features. We further characterize inherited coding variants focusing on DNA repair and cancer related genes in 19 early-onset MPC (eMPC) patients, proposing candidate MPC susceptibility genes based on the analysis of loss of function (LoF) variants in candidate tumor suppressor genes, their relative prevalence and association with the mutational landscape of the associated tumors.

Results

Identification of Lynch Syndrome patients by multigene targeting panel. We investigated the prevalence of LoF variants in cancer predisposition genes using a cohort of 71 patients with MPCs. Fifty-four (76.1%) patients were diagnosed with GC and CRC (23–42.6% synchronous) while 14 diagnosed with EC and CRC (N=13) or GC (N=1). Three patients had 3 or more types of cancers. Thirty (42.3%) patients were diagnosed with MPC before age 55 and classified as eMPCs. A family history of cancer was more prevalent in eMPCs: of the 56 patients for which family history was available, twenty-four (42.9%) patients had two or more first-degree relatives affected with any type of cancer, and 15 of them were eMPCs (Odds Ratio [OR]=8.58, 2.21–39.59, $p < 0.001$). Twenty-six (36.6%) patients had one or more MSI-H tumors, a higher proportion than the one observed in single-cancers of the same type (~15%)^{11–13,16} (Supplemental File 1, Table S1). This suggests an important contribution of defects in MMR to MPC development. The germline DNA of the patients was sequenced using a multigene targeting panel of 65 cancer predisposition genes. A total of 15/71 (21.1%) patients were found to carry variants classified to be pathogenic or likely pathogenic (P/LP) and all the variants affected LS related genes (Table 1, Table S2 in Supplemental File 1): *MLH1* (n=10), *MSH6* (n=2), *PMS2*

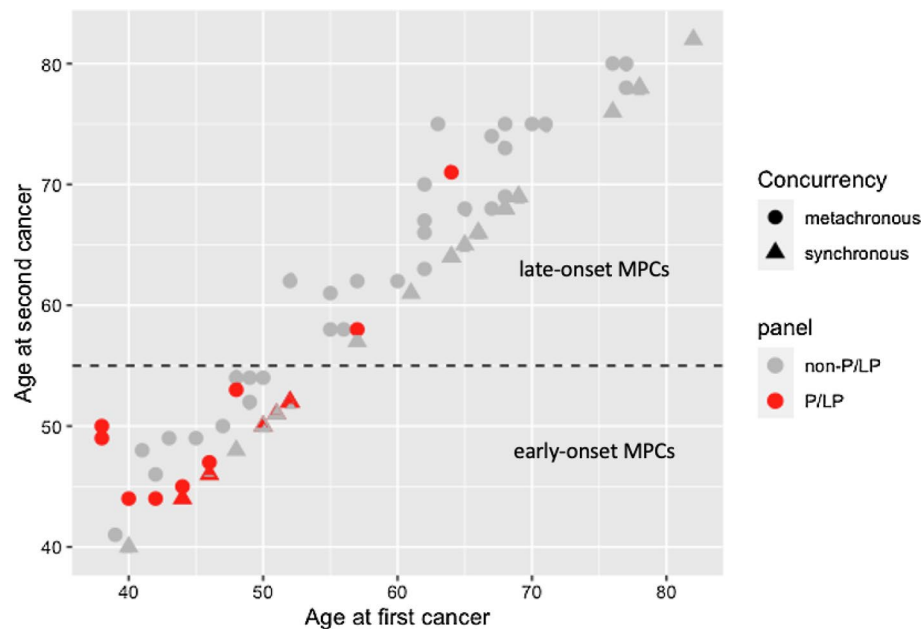


Figure 1. A scatter plot between age of first and second cancer by with/without pathogenic or likely pathogenic (P/LP) germline variants and concurrency of the tumors in each patient. There were 15 (21.1%) of patients with germline P/LP variants (red colors). Among early-onset multiple primary cancers (MPCs), 43.3% (13/30) patients had P/LP variants while only 4.9% (2/41) patients were related to P/LP variants in late-onset MPCs.

($n = 2$), and *MSH2* ($n = 1$), a distribution comparable to the previous reports¹⁷. There was one patient with multi-locus inherited neoplasia alleles syndrome (MINAS) harboring germline pathogenic variants in both *MLH1* and *BRCA1* gene²¹. One of recurrent P/LP variants of *MLH1* was c.1758dupC, a common variant in Korean LS patients²². There was a P/LP variant in *MSH2* (c.942 + 3A > T) that was reported as a frequent de novo mutation and founder variant in Newfoundland^{23,24}, in a patient diagnosed with four types of cancer (GC, CRC, EC, and lung cancer). Of the remaining 56 of patients, 42 had VUS altering 43 genes, leaving the contribution of these genes to MPC undetermined.

Characteristics of LS patients. All 15 LS patients had at least one MSI-H tumor in contrast to 11/56 non-LS patients. LS patients were also more likely to have two or more first-degree relatives (9/10, OR = 17.67, 2.12–836.73, $p = 0.001$, Supplemental file 1, Tables S1 and S2). The association of the variants with tumor characteristics were further investigated. The loss of expression of the protein corresponding to the gene with P/LP variants was confirmed in all 19 MSI-H tumors investigated from all LS patients (Supplemental file 1, Table S3)²⁵. The LS patients were more likely to be diagnosed with eoMPC (13/15, OR = 14.31, 2.79–144.22, $p < 0.001$, Fig. 1) but equally likely to present with synchronous tumors (6/15, OR = 0.89, 0.23–3.26, $p > 0.999$). This suggests that consistent with Lynch syndrome, LS MPC patients are diagnosed earlier. However, the variants are unlikely to influence the relative timing of the two cancer diagnosis. More importantly, these results showed that the majority of MPC patients do not have pathogenic variants known to cause LS. Even restricting to 13 patients with eoMPCs and strong family history of cancer (two or more first-degree relatives over two successive generations), cancer predisposition genes were not altered in nearly half of the patients (6/13, Table S2 in the Supplemental file 1). The likely genetic cause of MPC in these patients remains undetermined.

Identification of candidate cancer predisposition genes in early-onset MPC. To identify additional variants and genes underlying a possible cancer susceptibility in non-LS patients, we sequenced the whole exome of 19 eoMPC patients with gastric and colorectal cancer which is the most common combination of MPCs in Korea^{9,10}. Seven of the patients were LS patients, as determined above, and used as controls for the cancer predisposition genes discovery process. We identified 111,842 high-confidence variants across all patients, of which 3,211 were rare (MAF < 0.01) in East-Asian population and predicted to be damaging, affecting 2,675 genes. We investigated a set of 382 genes (referred to as cancer genes) involved in DNA repair, cancer susceptibility and progression (Supplemental file 1, Table S4)²⁶. Of those, 66 were altered by 82 variants, including 7 in the *MLH1* gene of the LS patients, therefore confirming the panel sequencing results and validating the initial WES variant calling and filtering approach.

For a given candidate cancer predisposition genes, we expect that the prevalence of LoF alterations observed in eoMPC cohort would be higher than in a matching sporadic cancer cohort. Despite some LS phenotypes being reported²⁶, TCGA gastric and colorectal cancer patients are mainly from sporadic cancers. We selected 70 of these patients with East-Asian ancestry to closely match the Korean ancestry of eoMPC cohort and reduce spurious findings resulting from population differences. In these patients, we identified a total of 87 LoF variants

Gene	N mutated cases			TCGA_EAS (n=70)	p-value* (TCGA)	p-value+ (TCGA)
	Non-Lynch (n=12)	Lynch (n=7)	Total (n=19)			
<i>MLH1</i>	0	7	7	2	> 0.999	< 0.0001
<i>RECQL5</i>	4	0	4	1	0.0013	> 0.999
<i>EME2</i>	2	0	2	1	0.0547	> 0.999
<i>EP300</i>	2	0	2	1	0.0547	> 0.999
<i>MUTYH</i>	2	0	2	1	0.0547	> 0.999
<i>MSH3</i>	2	0	2	2	0.1002	> 0.999
<i>COL7A1</i>	2	0	2	4	0.2108	> 0.999
<i>PTCH1</i>	2	0	2	7	0.6134	> 0.999

Table 2. Genes affected by Loss of function variants in two or more patients with eMPC at stomach and colon. $p < 0.01$ values are in bold. *Fisher's exact p -value for the frequency of each gene between non-lynch and TCGA. "+" Fisher's exact p -value for the frequency of each gene between lynch and TCGA.

affecting 47/382 cancer genes. We confirmed that LoF variants in *MLH1* were more prevalent in LS patients than in TCGA patients (7/7 vs 2/70 $p < 0.001$ Table 2, Table S5 in the Supplemental file 1). Importantly, 5/7 of the *MLH1* LoF variants in LS patients were predicted to be high-confidence LoF variants, indicating that most, but not all, known pathogenic variants can be identified through this approach. *RECQL5* was the only gene significantly more altered in non-LS patients than in TCGA patients (4/12 vs 1/70, OR:31.66, 2.73–1700.6, $p = 0.0013$, Table 2). In particular two of these patients had the same *RECQL5* variant (p.R441Q which is not seen in the TCGA patients ($p = 0.02$, Table S5 in the Supplemental file 1) The variant is rare, with a global minor allelic fraction of 3.6×10^{-4} . Interestingly 89 of the 101 observed minor alleles (out of 278,974 total alleles observed) in the gnomAD dataset²⁷ belong to the East-Asian population, suggesting a possible Korean-specific effect.

Bi-allelic alterations in corresponding tumors as a guide for cancer predisposition genes identification. According to Knudson's two-hit model²⁸, the wild-type allele of tumor suppressor genes is frequently lost or mutated in the tumor. Such combination of inherited deleterious variants with somatic loss—referred to as bi-allelic alteration—can help prioritize candidate cancer predisposition genes^{26,29}. We sequenced the exome of 37 tumors from all 19 patients with eMPC (Supplemental file 1, Table S7) and identified 35 somatic bi-allelic alterations by LoF mutations (N=5), loss of heterozygosity (N=27) or both (N=3) focusing on candidate germline LoF genes.

The *MLH1* gene was somatically altered in 13/15 tumors from LS patients and at least one tumor from each of the 7 LS patient. (Table 3). Importantly, 5/7 *MLH1* variants in LS patients were high-confidence LoF variants. This observation confirms the validity of the investigation of bi-allelic alteration to identify cancer predisposition genes, at least for high-confidence LoF variants in known predisposition genes. In contrast, the variants in the main candidate susceptibility gene in non-LS patients, *RECQL5*, were not high-confidence LoF variant and the gene was not affected in any of the 6 tumors from 4 variant carriers or any of the 22 tumors from all non-LS patients. Expanding the analysis to other candidate genes affected by LoF germline variants in non-LS patients, we identified 17 genes in 10 patients also affected by bi-allelic alterations in one or more tumors. Two were altered via mutations and 15 through LOH. Among these, 2 genes were affected by high-confidence LoF germline variants: *FANCG* (germline splice site variant, c.307 + 1G > C) in gastric tumor of dou_004 and *CASP8* (germline deletion variant, c.658_659del) in colorectal tumor of dou_010 (Table 3). Interestingly, this analysis also revealed somatic alterations of LS genes in absence of inherited LS mutation in a non-LS patient (dou_009) with both of MSI-H tumors in the stomach and colon. Indeed, somatic allelic imbalance was observed in *MLH1* (stomach and colon) and *PMS2* (stomach only) and the expression of both encoded proteins was lost from both tumors. Loss of one allele in the tumor would likely not be sufficient to impair MMR, raising the possibility of undetected germline LoF variants in these LS genes in this patient, perhaps through alterations not easily detected by WES³⁰. There was one tumor (colon cancer of dou_005) with double somatic mutations³¹ in MMR genes (c.582-591del at *MLH1* and c.493delT at *PMS2*, variant allele fraction was 0.186 and 0.188, respectively) and the patient harbors germline pathogenic variant at *MLH1*. The details of LoF germline variants and somatic alteration of the corresponding tumors were described in Table S6 (Supplementary file 1).

Discussion

As expected, the present result confirmed that individuals with MPC, particularly early onset, have a higher likelihood of LS: MMR related P/LP germline variants were observed in 21% of MPC (15/71) and in 43% of eMPC (13/30) patients, a proportion comparable to ~ 15% of early-onset CRC²⁰. The elevated extracolonic cancer risk following colorectal cancer in Lynch syndrome was reported (~ 5 and 40 times of gastric and endometrial cancer risk compared to general population, respectively)³², and this study supports those results. Therefore, due to the high prevalence of germline mutations individuals with MPC in the LS tumor spectrum should undergo germline testing, including when MSI status is unknown.

Because MSI-H was considered as a hallmarks of LS, universal screening of all CRC and EC has been recommended^{33,34}. In addition, MSI screening for gastric cancers in regions where gastric cancer is highly prevalent like Korea is a way to identify individuals and families who can benefit from germline testing of LS genes,

Patient group	Case ID	Genes with germline LoF variants (1)	N tumors (N sequenced)	Tumor MSI status (GC/CRC)	Bi-allelic alterations in at least one tumor (1)
Lynch group	dou_002	<i>MLH1*</i> , <i>FANCL</i> , <i>RAD54L</i> , <i>RFC3</i> , <i>TSC2</i> , <i>FANCA</i> , <i>ARID1B</i> , <i>POLN</i>	2 (1)	MSI-H/MSI-H	<i>MLH1</i>
	dou_003	<i>MLH1*</i> , <i>POLL</i> , <i>SMO</i>	3 (3)	MSI-H/MSI-H/MSI-H	<i>MLH1**</i> , <i>SMO**</i>
	dou_005	<i>MLH1*</i> , <i>FANCI</i> , <i>UNG</i> , <i>POLN</i> , <i>RFC3</i> , <i>ARID1A</i> , <i>FANCI</i>	2 (2)	MSS/MSI-H	<i>UNG</i> , <i>POLN</i> , <i>MLH1**</i> , <i>ARID1A**</i>
	dou_006	<i>MLH1</i> , <i>BRCA1</i> , <i>CBL</i> , <i>RAD50</i> , <i>UVSSA</i>	2 (2)	MSI-H/MSI-H	<i>MLH1</i> , <i>RAD50</i>
	dou_011	<i>MLH1*</i> , <i>ARID2</i> , <i>ERCC2</i> , <i>EGFR</i> , <i>TMEM127</i> , <i>RHBDF2</i>	2 (2)	MSI-H/MSI-H	<i>ARID2</i> , <i>RHBDF2</i> , <i>MLH1**</i>
	dou_016	<i>MLH1*</i> , <i>DCLRE1C</i> , <i>ERCC6</i> , <i>JAK3</i>	3 (3)	MSI-H/MSI-H/MSI-H	<i>MLH1***</i>
	dou_017	<i>MLH1*</i> , <i>GNAS</i>	2 (2)	MSI-H/MSI-H	<i>MLH1**</i>
Non-Lynch group	dou_001	<i>MUTYH</i> , <i>PBRM1</i> , <i>PTCH1</i> , <i>ABL1</i>	2 (2)	MSS/MSS	<i>MUTYH</i> , <i>PBRM1</i>
	dou_004	<i>CDH1</i> , <i>FANCI</i> , <i>FANCG</i> , <i>FANCM</i> , <i>RAD50</i> , <i>MSH3</i>	2 (2)	MSS/MSS	<i>CDH1</i> , <i>FANCI</i> , <i>FANCG</i> , <i>FANCM</i>
	dou_007	<i>BRCA2</i> , <i>MUTYH</i> , <i>EME2</i> , <i>MBD4</i> , <i>RECQL5*</i> , <i>WRN</i> , <i>COL7A1</i>	2 (1)	MSS/NA	
	dou_008	<i>AXIN1</i> , <i>COL7A1</i> , <i>RECQL5*</i> , <i>MSH4</i> , <i>PTCH1</i> , <i>NEIL3</i> , <i>KLF4</i>	2 (2)	MSS/MSS	<i>COL7A1</i>
	dou_009	<i>NOTCH2</i> , <i>REV3L</i> , <i>RECQL5*</i> , <i>SMARCA4</i> , <i>EP300</i> , <i>DOCK8</i>	2 (2)	MSI-H/MSI-H	<i>SMARCA4</i>
	dou_010	<i>CASP8</i> , <i>EME2</i> , <i>NOTCH1</i>	2 (2)	MSS/MSS	<i>CASP8</i>
	dou_012	<i>POLM</i> , <i>EME1</i>	2 (2)	MSS/MSS	<i>EME1</i>
	dou_013	<i>FLCN</i> , <i>CIC</i> , <i>NFATC2IP</i> , <i>XAB2</i>	2 (2)	MSS/MSS	<i>FLCN</i>
	dou_014	<i>EP300</i> , <i>XPC</i> , <i>RECQL5*</i>	2 (1)	MSS/NA	
	dou_015	<i>RAD52</i> , <i>CHEK2</i>	2 (2)	MSS/MSS	<i>RAD52</i>
	dou_018	<i>PARG</i> , <i>TET2</i> , <i>NF2</i> , <i>UVSSA</i> , <i>BCOR</i>	2 (2)	MSS/MSS	<i>PARG</i> , <i>TET2</i> , <i>UVSSA</i> , <i>NF2</i>
	dou_019	<i>MSH2</i> , <i>ERCC6</i> , <i>ALKBH2</i>	2 (2)	MSS/MSS	<i>ERCC6</i>

Table 3. Summary of candidate loss-of-function germline candidate genes and the status of loss of heterozygosity of corresponding tumors. (1) High confidence LoF prediction in bold. *Significantly recurrent in eoMPC. **Bi-allelic alteration was observed in 2 tumors. ***Bi-allelic alteration was observed in 3 tumors.

so that surveillance and risk reducing interventions can be undertaken along with cascade testing of family members. At the same time, MSI testing is valuable for other types of solid cancers for other purpose: MSI-H is a biomarker for response to immune checkpoint inhibitors that is a breakthrough for treating advanced solid cancers regardless of its origin³⁵. As MSI-H is predictive of LS across a broad tumor spectrum³⁶, screening of tumor MSI status in all patients with an initial cancer, especially for LS spectrum tumors and early-onset cancer, and consecutive germline testing for patients with MSI-H tumor will help refine the diagnosis, giving an opportunity to diagnose LS and avoid or detect the second cancer as early as possible.

Genetic analysis of families with high occurrence of cancer using cancer predisposition genes panels or exome sequencing has become standard to identify the underlying cancer susceptibility variants. But penetrance, interaction with environmental factors, and size of the pedigree may affect power of such linkage studies^{1,37–39}. Evaluating both germline and somatic alterations is another reasonable approach to find cancer predisposition genes^{26,29}. The present results showed that characterizing recurrent LoF genes, truncating variants and loss of heterozygosity are useful to prioritize candidate cancer predisposition genes; any combination of them unambiguously identified germline *MLH1*, a known cancer predisposition genes in LS patients, in multiple patients clinically diagnosed with LS. It suggests that MPC is a strong phenotype to find cancer predisposition genes, therefore this approach would be worthy to be expanded to other combinations of MPCs. However, the types of MPCs would be different by geographical regions or ancestry: region-specific environmental factors can contribute to the oncogenesis and lead to different cancer types within the LS spectrum. For instance, MPCs including GC could be a hallmark of LS in Korea but not in the United States where GC incidence is much lower. This could explain why GC risk in LS may have been underestimated^{17,40}. Therefore, geographical and/or ancestry specific cancer epidemiology needs to be accounted for in the genetic screening guidelines.

In non-LS patients with eoMPC, however, evidence of germline susceptibility was more elusive. *RECQL5* was a frequently recurrent gene and variant (p.R441Q) compared to East Asian sporadic cancer population (TCGA). Despite careful matching, some differences between the eoMPC and matched TCGA cohorts remained and could have confounded this observation. Notably, eoMPC patients were all younger than 55 and exclusively of Korean descent, while the TCGA patients included in the analysis were not selected for age and included non-Korean Asians. Importantly, the allele frequency of the *RECQL5* variant was 0.4% of EAS population in gnomAD, and there was no evidence of somatic LOH in the corresponding tumors, suggesting that additional analysis and experiment are required to establish its pathogenicity. *CASP8* and *FANCG* were high-confidence LoF germline gene with bi-allelic alteration in CRC and GC, respectively. *CASP8* encodes a member of caspase family and play a role in apoptosis, and some of its polymorphism have been reported as susceptibility to various cancers including CRC^{41–43}. *FANCG* is a gene encoding Fanconi anemia (FA) group G protein, and part of the FA DNA

damage repair pathway and in which germline pathogenic variants often predispose to cancers^{44–46}. Thus, the *CASP8* and *FANCG* variants identified in these two MPC cases could underlie their disease susceptibility though additional functional studies would be necessary to demonstrate their pathogenicity^{47–50}.

Despite evaluating all inherited coding variants in cancer genes and somatic changes in the corresponding tumors, we did not find a clear inherited predisposition that caused multiple cancers in over half of patients with eoMPC, even with strong family cancer history. Not only cancer predisposition genes, but also environmental factors such as smoking and alcohol also increase the risk of cancers^{51–53}.

Our study has inherent limitations. The present results did not cover epigenetic changes affecting candidate cancer predisposition genes in tumors, and methylation of the *MLH1* promoter in particular is known to be a mechanism for somatic loss of function in ~20% of CRC⁵⁴. Furthermore, we restricted the analysis to 382 well studied cancer genes, more likely to impact cancer susceptibility, therefore leaving the possibility to miss unexpected cancer predisposition genes. The complete analysis of whole exomes to identify rare disease susceptibility variants would indeed be intractable for the cohort under study. Careful analysis of the pedigree and genetic comparison of family members, a typical standard in such genetic susceptibility studies, was not possible due to the absence of family history information in many cases and to the retrospective nature of the study. However, this study is to our knowledge the largest study of patients with MPCs, evaluated using the same multigene targeting panel, and where both germline and multiple tumors DNA were investigated in the highest risk patients.

eoMPC of colorectal, endometrial, and gastric cancer are considerably enriched for LS patients, supporting the genetic screens of related family members as well as enhanced monitoring for younger patients after their first diagnosis. Routine tumor screening of MSI for patients with initial LS related cancers and consecutive germline testing for patients with MSI-H is worthy to diagnose LS and early detection or avoid second cancer. While susceptibility variants in non-LS genes for MPC patients may exist, evidence for their role is more elusive than for LS patients and would require deeper genetic investigation and complementary functional studies.

Methods

Population. We performed a retrospective population-based study targeting patients who were treated for two or more cancers in the stomach, colorectum, or endometrium at Severance Hospital, Yonsei University College of Medicine between January 2001 and December 2016. All methods were carried out in accordance with relevant guidelines and regulations. This study was approved by Institutional Review Board (IRB) of Severance hospital of the Yonsei University Health System and by the IRB of the University of California, San Diego (4-2017-0434, 191,543). The IRB of Severance hospital of the Yonsei University Health System waived the requirement for patient informed consent as the study is retrospective by design. The patients were selected using following criteria: (1) two or more cancers were pathologically confirmed, (2) multiple tumors were histologically different, not a metastatic or recurrent tumor from one cancer, (3) normal tissues were available and histologically confirmed for sequencing. DNA was obtained from formalin fixed paraffin embedded (FFPE) normal and tumor tissues. Early-onset MPC (eoMPC) was defined by the diagnosis of a second cancer at age 55 or younger. The clinico-pathologic characteristics of the patients and tumors including age, sex, family history of cancer, MSI or MMR status of tumors were evaluated. For evaluating MSI status, two mononucleotide repeat markers (BAT25 and BAT26) and three dinucleotide repeat markers (D5S346, D2S123, and D17S250) were used by polymerase chain reaction⁵⁵, and MSIsensor was used with cut-off of MSI score > 3.5 in Whole Exome Sequencing (WES) data⁵⁶. To evaluate MMR protein expression, immunohistochemistry (IHC) was conducted in four MMR genes; *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). An MMR-deficient (dMMR) tumor was defined as a tumor showing loss of expression of any of the four MMR proteins. If a tumor was classified as any one of MSI-high (MSI-H) or deficiency MMR (dMMR) it was considered as MSI-H. The concurrency of tumors was classified as synchronous tumors when the interval between tumors was less than 1 year, otherwise considered as metachronous tumors.

Germline multigene targeting panel. Germline DNA of patients with MPCs was evaluated using a customized targeted capture sequencing panel (OncoRisk, Celeomics, Seoul, Korea) covering all coding sequences and intron–exon boundaries of the coding exons of known 65 cancer predisposition genes (Supplemental file 1, Table S8)⁵⁷. Both structural variants and nucleotide variants were evaluated. The germline variants were classified into pathogenic, likely pathogenic, variant of uncertain significance (VUS) and reported following the guidelines of the American College of Medical Genetics and Genomics 2017⁵⁸. Only pathogenic or likely pathogenic (P/LP) germline variants in cancer predisposition genes were considered in the analysis. Read-depth based detection of structural variants was conducted using the ExomeDepth software. Chromosomal copy number variations (CNVs) detected by ExomeDepth were further crosschecked using our custom pipeline; this retrieved base-level depth-of-coverage for each bam file using SAMtools software and normalized the depths against those of other samples in the same batch. There were no significant pathogenic CNV detected in the studied samples.

Whole exome sequencing. *Data generation.* WES of normal and tumors was conducted for patients with eoMPCs in the stomach and colon. Genomic DNA was extracted from the confirmed normal and tumor tissues of FFPE. SureSelect sequencing libraries were prepared according to the manufacturer's instructions (Agilent SureSelect All Exon V6 kit, Santa Clara, CA, USA), and HISEQ2500 sequencing system (Illumina™, San Diego, CA, USA) performed sequencing with read lengths of 2 × 100 bp. The statistics and quality metrics of normal and tumors were described in Supplemental file 1, Tables S7 and S9.

Data analysis. The reads were aligned to hg19 reference genome by Burrows-Wheeler Aligner software (BWA 0.7.17) and duplicated reads were removed by Picard (MarkDuplicates) through bcio-nextgen (v.1.2.3)⁵⁹ For germline and somatic analysis, variants were called by Genome Analysis Toolkit Haplotype joint caller (GATK v3.9) and mutect2^{60,61}, respectively. Variants were annotated by refGene using ANNOVAR⁶², CADD scores (CADD13 and CADDindel)⁶³, population allelic frequency in ExAC v3.0⁶⁴, and membership to ClinVar⁶⁵.

To discover novel cancer predisposition genes, we focused on high-confidence (TLOD > 12, FS < 10, SEQQ > 60, MQ ≥ 60, STRANDQ > 40, DP > 10), rare (minor allele fraction was < 0.01 in ExAC of Eastern Asian population), and damaging (exonic/splicing, non-synonymous variants with > 20 of CADD score) variants in 382 genes of 11 cancer-relevant pathways²⁶, and they were defined as LoF germline variants. High-confidence of LoF variants were predicted by Loss-Of-Function Transcript Effect Estimator (LOFTEE) plugin for Ensembl VEP (v.99) that targeting stop-gained, splice site disrupting, frameshift variants^{66,67}. Copy number alterations of tumors compared to normal was estimated using CNVkit (V0.8)⁶⁸. The gene level copy number ratio was calculated as the weighted mean of all bins covered by the whole segment overlapping the gene. Genes had ≥ 3 segments of copy number changes were included, and log₂ copy number < -0.3 was defined as deletion. Somatic allelic imbalance (AI) in tumors of a given heterozygous germline variants were estimated using hapLOHseq⁶⁹. Germline variants of tumors were obtained by GATK haplotype caller and filtering out variants were not observed in germline variants in normal sample. Phase of genotypes were estimated with a companion phasing utility with hapLOHseq. Event prevalence was set 0.1, and AI was defined when posterior probability of being in AI was over 90%. To assess loss of heterozygosity (LOH) of a given heterozygous germline variants in tumors as a secondary hit mechanism of inactivation of cancer predisposition genes^{26,28}, we evaluated germline and somatic bi-allelic alteration of corresponding tumors. When there was any one of somatic damaging mutation or AI event in candidate germline cancer predisposition genes, it was considered as bi-allelic alteration.

To compare the frequency of recurrent cancer predisposition genes in patients of eMPCs to that of the cancer genome atlas (TCGA) cohort, germline data of East-Asian patients (≥ 80% of admixture)⁷⁰ with stomach cancer or colorectal cancer in TCGA cohort was used^{26,71}. Variants were filtered for rare and damaged variants similar to LoF germline variants in MPCs cohort. For recurrent variants and genes in eMPCs cohort, the frequency was compared between eMPCs cohort and selected TCGA cohort.

Statistical analysis. Continuous variables were compared by two-sided Mann-Whitney test and categorical variables were compared by two-sided Fisher's exact test. $P < 0.05$ was considered as statistically significant and R version 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria) was used for generating figures and statistical analysis.

Received: 28 February 2021; Accepted: 5 July 2021

Published online: 20 July 2021

References

- Rahman, N. Realizing the promise of cancer predisposition genes. *Nature* **505**, 302–308. <https://doi.org/10.1038/nature12981> (2014).
- Huang, K. L. *et al.* Pathogenic germline variants in 10,389 adult cancers. *Cell* **173**, 355–370. <https://doi.org/10.1016/j.cell.2018.03.039> (2018).
- Evans, J. P., Powell, B. C. & Berg, J. S. Finding the rare pathogenic variants in a human genome. *JAMA* **317**, 1904–1905. <https://doi.org/10.1001/jama.2017.0432> (2017).
- Mersch, J. *et al.* Prevalence of variant reclassification following hereditary cancer genetic testing. *JAMA* **320**, 1266–1274. <https://doi.org/10.1001/jama.2018.13152> (2018).
- Cybulski, C., Nazarali, S. & Narod, S. A. Multiple primary cancers as a guide to heritability. *Int J Cancer* **135**, 1756–1763. <https://doi.org/10.1002/ijc.28988> (2014).
- Vogt, A. *et al.* Multiple primary tumours: challenges and approaches, a review. *ESMO Open* **2**, e000172. <https://doi.org/10.1136/esmoopen-2017-000172> (2017).
- Whitworth, J. *et al.* A clinical and genetic analysis of multiple primary cancer referrals to genetics services. *Eur. J. Hum. Genet.* **23**, 581–587. <https://doi.org/10.1038/ejhg.2014.157> (2015).
- Hong, S. *et al.* Cancer statistics in Korea: incidence, mortality, survival, and prevalence in 2017. *Cancer Res. Treat.* **52**, 335–350. <https://doi.org/10.4143/crt.2020.206> (2020).
- Cho, I. *et al.* Risk factors for double primary malignancies and their clinical implications in patients with sporadic gastric cancer. *Eur. J. Surg. Oncol.* **40**, 338–344. <https://doi.org/10.1016/j.ejso.2013.10.027> (2014).
- Yun, H. R. *et al.* Double primary malignancy in colorectal cancer patients—MSI is the useful marker for predicting double primary tumors. *Int. J. Colorectal. Dis.* **24**, 369–375. <https://doi.org/10.1007/s00384-008-0541-x> (2009).
- Cancer Genome Atlas Research. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209. <https://doi.org/10.1038/nature13480> (2014).
- Cancer Genome Atlas. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337. <https://doi.org/10.1038/nature11252> (2012).
- Choi, Y. Y. *et al.* Microsatellite instability and programmed cell death-ligand 1 expression in stage II/III gastric cancer: post hoc analysis of the CLASSIC randomized controlled study. *Ann. Surg.* **270**, 309–316. <https://doi.org/10.1097/SLA.0000000000002803> (2019).
- Kim, C. G. *et al.* Effects of microsatellite instability on recurrence patterns and outcomes in colorectal cancers. *Br. J. Cancer* **115**, 25–33. <https://doi.org/10.1038/bjc.2016.161> (2016).
- Guastadisegni, C., Colafranceschi, M., Ottini, L. & Dogliotti, E. Microsatellite instability as a marker of prognosis and response to therapy: a meta-analysis of colorectal cancer survival data. *Eur. J. Cancer* **46**, 2788–2798. <https://doi.org/10.1016/j.ejca.2010.05.009> (2010).
- Cancer Genome Atlas Research. Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73. <https://doi.org/10.1038/nature12113> (2013).

17. Moller, P. *et al.* Cancer incidence and survival in Lynch syndrome patients receiving colonoscopic and gynaecological surveillance: first report from the prospective Lynch syndrome database. *Gut* **66**, 464–472. <https://doi.org/10.1136/gutjnl-2015-309675> (2017).
18. Koornstra, J. J. *et al.* Management of extracolonic tumours in patients with Lynch syndrome. *Lancet Oncol.* **10**, 400–408. [https://doi.org/10.1016/S1470-2045\(09\)70041-5](https://doi.org/10.1016/S1470-2045(09)70041-5) (2009).
19. Cohen, S. A., Pritchard, C. C. & Jarvik, G. P. Lynch syndrome: from screening to diagnosis to treatment in the era of modern molecular oncology. *Ann. Rev. Genom. Hum. Genet.* **20**, 293–307. <https://doi.org/10.1146/annurev-genom-083118-015406> (2019).
20. Pearlman, R. *et al.* Prevalence and spectrum of germline cancer susceptibility gene mutations among patients with early-onset colorectal cancer. *JAMA Oncol.* **3**, 464–471. <https://doi.org/10.1001/jamaoncol.2016.5194> (2017).
21. Whitworth, J. *et al.* Multilocus inherited neoplasia alleles syndrome: a case series and review. *JAMA Oncol.* **2**, 373–379. <https://doi.org/10.1001/jamaoncol.2015.4771> (2016).
22. Shin, Y. K. *et al.* Germline mutations in MLH1, MSH2 and MSH6 in Korean hereditary non-polyposis colorectal cancer families. *Hum. Mutat* **24**, 351. <https://doi.org/10.1002/humu.9277> (2004).
23. Desai, D. C. *et al.* Recurrent germline mutation in MSH2 arises frequently de novo. *J. Med. Genet.* **37**, 646–652. <https://doi.org/10.1136/jmg.37.9.646> (2000).
24. Woods, M. O. *et al.* High frequency of hereditary colorectal cancer in Newfoundland likely involves novel susceptibility genes. *Clin. Cancer Res.* **11**, 6853–6861. <https://doi.org/10.1158/1078-0432.CCR-05-0726> (2005).
25. Sarode, V. R. & Robinson, L. Screening for lynch syndrome by immunohistochemistry of mismatch repair proteins: significance of indeterminate result and correlation with mutational studies. *Arch. Pathol. Lab Med.* **143**, 1225–1233. <https://doi.org/10.5858/arpa.2018-0201-OA> (2019).
26. Buckley, A. R., Ideker, T., Carter, H., Harismendy, O. & Schork, N. J. Exome-wide analysis of bi-allelic alterations identifies a Lynch phenotype in The Cancer Genome Atlas. *Genome Med.* **10**, 69. <https://doi.org/10.1186/s13073-018-0579-5> (2018).
27. Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443. <https://doi.org/10.1038/s41586-020-2308-7> (2020).
28. Knudson, A. G. Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc. Nat. Acad. Sci. USA* **68**, 820–823. <https://doi.org/10.1073/pnas.68.4.820> (1971).
29. Park, S., Supek, F. & Lehner, B. Systematic discovery of germline cancer predisposition genes through the identification of somatic second hits. *Nat. Commun.* **9**, 2601. <https://doi.org/10.1038/s41467-018-04900-7> (2018).
30. Rhee, J., Arnold, M. & Boland, C. R. Inversion of exons 1–7 of the MSH2 gene is a frequent cause of unexplained Lynch syndrome in one local population. *Fam Cancer* **13**, 219–225. <https://doi.org/10.1007/s10689-013-9688-x> (2014).
31. Pearlman, R. *et al.* Clinical characteristics of patients with colorectal cancer with double somatic mismatch repair mutations compared with Lynch syndrome. *J. Med. Genet.* **56**, 462–470. <https://doi.org/10.1136/jmedgenet-2018-105698> (2019).
32. Win, A. K. *et al.* Risks of primary extracolonic cancers following colorectal cancer in lynch syndrome. *J. Nat. Cancer Inst.* **104**, 1363–1372. <https://doi.org/10.1093/jnci/djs351> (2012).
33. Gupta, S. *et al.* NCCN Guidelines insights: genetic/familial high-risk assessment: colorectal, Version 3.2017. *J Natl Compr Canc Netw* **15**, 1465–1475. <https://doi.org/10.6004/jnccn.2017.0176> (2017).
34. Hampel, H. *et al.* Screening for Lynch syndrome (hereditary nonpolyposis colorectal cancer) among endometrial cancer patients. *Cancer Res.* **66**, 7810–7817. <https://doi.org/10.1158/0008-5472.CAN-06-1114> (2006).
35. Le, D. T. *et al.* PD-1 blockade in tumors with mismatch-repair deficiency. *N Engl. J. Med.* **372**, 2509–2520. <https://doi.org/10.1056/NEJMoa1500596> (2015).
36. Latham, A. *et al.* Microsatellite instability is associated with the presence of lynch syndrome pan-cancer. *J. Clin. Oncol.* **37**, 286–295. <https://doi.org/10.1200/JCO.18.00283> (2019).
37. Galvan, A., Ioannidis, J. P. & Dragani, T. A. Beyond genome-wide association studies: genetic heterogeneity and individual predisposition to cancer. *Trends Genet.* **26**, 132–141. <https://doi.org/10.1016/j.tig.2009.12.008> (2010).
38. Pirinen, M., Donnelly, P. & Spencer, C. C. Including known covariates can reduce power to detect genetic effects in case-control studies. *Nat. Genet.* **44**, 848–851. <https://doi.org/10.1038/ng.2346> (2012).
39. Hindorf, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **106**, 9362–9367. <https://doi.org/10.1073/pnas.0903103106> (2009).
40. Chen, S. *et al.* Prediction of germline mutations and cancer risk in the Lynch syndrome. *JAMA* **296**, 1479–1487. <https://doi.org/10.1001/jama.296.12.1479> (2006).
41. Sun, T. *et al.* A six-nucleotide insertion-deletion polymorphism in the CASP8 promoter is associated with susceptibility to multiple cancers. *Nat. Genet.* **39**, 605–613. <https://doi.org/10.1038/ng2030> (2007).
42. Zhang, Y. *et al.* A systematic analysis of the association studies between CASP8 D302H polymorphisms and breast cancer risk. *J. Genet.* **96**, 283–289. <https://doi.org/10.1007/s12041-017-0774-y> (2017).
43. Ying, Y., Xu, J., Qi, Y., Zhang, M. & Yang, Y. CASP8 rs3834129 (-652 6N insertion/deletion) polymorphism and colorectal cancer susceptibility: an updated meta-analysis. *J. Cancer* **9**, 4166–4171. <https://doi.org/10.7150/jca.27110> (2018).
44. Chen, H., Zhang, S. & Wu, Z. Fanconi anemia pathway defects in inherited and sporadic cancers. *Transl. Pediatr.* **3**, 300–304. <https://doi.org/10.3978/j.issn.2224-4336.2014.07.05> (2014).
45. Esteban-Jurado, C. *et al.* The Fanconi anemia DNA damage repair pathway in the spotlight for germline predisposition to colorectal cancer. *Eur. J. Hum. Genet.* **24**, 1501–1505. <https://doi.org/10.1038/ejhg.2016.44> (2016).
46. Huang, J. P. *et al.* FANCA D1359Y mutation in a patient with gastric polyposis and cancer susceptibility: A case report and review of literature. *World J. Gastroenterol.* **24**, 4412–4418. <https://doi.org/10.3748/wjg.v24.i38.4412> (2018).
47. Soung, Y. H. *et al.* CASPASE-8 gene is inactivated by somatic mutations in gastric carcinomas. *Cancer Res.* **65**, 815–821 (2005).
48. Soung, Y. H. *et al.* Caspase-8 gene is frequently inactivated by the frameshift somatic mutation 1225_1226delTG in hepatocellular carcinomas. *Oncogene* **24**, 141–147. <https://doi.org/10.1038/sj.onc.1208244> (2005).
49. Ramanagoudr-Bhojappa, R. *et al.* Multiplexed CRISPR/Cas9-mediated knockout of 19 Fanconi anemia pathway genes in zebrafish revealed their roles in growth, sexual development and fertility. *PLoS Genet.* **14**, e1007821. <https://doi.org/10.1371/journal.pgen.1007821> (2018).
50. Pulliam-Leath, A. C. *et al.* Genetic disruption of both Fancg and Fancj in mice recapitulates the hematopoietic manifestations of Fanconi anemia. *Blood* **116**, 2915–2920. <https://doi.org/10.1182/blood-2009-08-240747> (2010).
51. Hashibe, M. *et al.* Contribution of tobacco and alcohol to the high rates of squamous cell carcinoma of the supraglottis and glottis in Central Europe. *Am. J. Epidemiol.* **165**, 814–820. <https://doi.org/10.1093/aje/kwk066> (2007).
52. Gandini, S. *et al.* Tobacco smoking and cancer: a meta-analysis. *Int. J. Cancer* **122**, 155–164. <https://doi.org/10.1002/ijc.23033> (2008).
53. Services, U. S. D. o. H. a. H. 14th report on carcinogens. <https://ntp.niehs.nih.gov/whatwestudy/assessments/cancer/roc/index.html> (2016).
54. Li, X. *et al.* MLH1 promoter methylation frequency in colorectal cancer patients and related clinicopathological and molecular features. *PLoS ONE* **8**, e59064. <https://doi.org/10.1371/journal.pone.0059064> (2013).
55. Umar, A. *et al.* Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J. Natl. Cancer Inst.* **96**, 261–268. <https://doi.org/10.1093/jnci/djh034> (2004).
56. Niu, B. *et al.* MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. *Bioinformatics* **30**, 1015–1016. <https://doi.org/10.1093/bioinformatics/btt755> (2014).

57. Eoh, K. J. *et al.* Detection of germline mutations in patients with epithelial ovarian cancer using multi-gene panels: beyond BRCA1/2. *Cancer Res. Treat.* **50**, 917–925. <https://doi.org/10.4143/crt.2017.220> (2018).
58. Li, M. M. *et al.* Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a joint consensus recommendation of the association for molecular pathology, American society of clinical oncology, and college of American pathologists. *J. Mol. Diagn.* **19**, 4–23. <https://doi.org/10.1016/j.jmoldx.2016.10.002> (2017).
59. BCBio-nextgen v1.2.3. <https://doi.org/10.5281/zenodo.3743344>.
60. VanderAuwera, G. A. *et al.* From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinf.* **43**, 111011–111033. <https://doi.org/10.1002/0471250953.bi1110s43> (2013).
61. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498. <https://doi.org/10.1038/ng.806> (2011).
62. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164. <https://doi.org/10.1093/nar/gkq603> (2010).
63. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894. <https://doi.org/10.1093/nar/gky1016> (2019).
64. Karczewski, K. J. *et al.* The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res.* **45**, D840–D845. <https://doi.org/10.1093/nar/gkw971> (2017).
65. Landrum, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42**, D980–985. <https://doi.org/10.1093/nar/gkt1113> (2014).
66. McLaren, W. *et al.* The ensembl variant effect predictor. *Genome Biol.* **17**, 122. <https://doi.org/10.1186/s13059-016-0974-4> (2016).
67. MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828. <https://doi.org/10.1126/science.1215040> (2012).
68. Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput. Biol.* **12**, e1004873. <https://doi.org/10.1371/journal.pcbi.1004873> (2016).
69. Vattathil, S. & Scheet, P. Haplotype-based profiling of subtle allelic imbalance with SNP arrays. *Genome Res.* **23**, 152–158. <https://doi.org/10.1101/gr.141374.112> (2013).
70. Harismendy, O., Kim, J., Xu, X. & Ohno-Machado, L. Evaluating and sharing global genetic ancestry in biomedical datasets. *J. Am. Med. Inf. Assoc.* **26**, 457–461. <https://doi.org/10.1093/jamia/ocy194> (2019).
71. Buckley, A. R. *et al.* Pan-cancer analysis reveals technical artifacts in TCGA germline variant calls. *BMC Genom.* **18**, 458. <https://doi.org/10.1186/s12864-017-3770-y> (2017).

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government. (2017R1D1A1B03032553) We would like to thank to Severance Research Initiative project from Yonsei University College of Medicine to support this study.

Author contributions

Y.Y.C., J.H.C., and O.H. designed and supervised the research. X.X., D.N., Y.Y.C., J.E.L. generated sequencing data and Y.Y.C. performed the statistical analysis, prepared the figures and tables, and drafted the manuscript. Y.Y.C., J.H.C., and O.H. designed the experiments. S.S., J.E.L. conducted experiment. S.L., J.S.P., J.J., H.K., S.H.N. screened patients and collected specimen. Y.Y.C., L.M. and O.H. wrote the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-94292-4>.

Correspondence and requests for materials should be addressed to J.-H.C. or O.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021