Received: 2 November 2020 | Revised: 15 January 2021 | Accepted: 20 January 2021

DOI: 10.1002/ijc.33492

CANCER GENETICS AND EPIGENETICS

IJC International Journal of Cancer

Somatic mutation profiles as molecular classifiers of ulcerative colitis-associated colorectal cancer

Satu Mäki-Nevala[1] | Sanjeevi Ukwattage[1] | Alisa Olkinuora[1] | Henrikki Almusa[2] | Maarit Ahtiainen[3] | Ari Ristimäki[4,5] | Toni Seppälä[6] | Anna Lepistö[6] | Jukka-Pekka Mecklin[7] | Päivi Peltomäki[1]

[1]Department of Medical and Clinical Genetics, University of Helsinki, Helsinki, Finland

[2]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland

[3]Department of Education and Research, Central Finland Central Hospital, Jyväskylä, Finland

[4]Department of Pathology, HUSLAB, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

[5]Applied Tumor Genomics Research Program, Research Programs Unit, University of Helsinki, Helsinki, Finland

[6]Department of Gastrointestinal Surgery, Helsinki University Hospital and Helsinki University, Helsinki, Finland

[7]Department of Sport and Health Sciences, University of Jyväskylä and Jyväskylä Central Hospital, Jyväskylä, Finland

Correspondence
Satu Mäki-Nevala, Department of Medical and Clinical Genetics, University of Helsinki, Helsinki, Finland.
Email: satu.maki-nevala@helsinki.fi

Funding information
Academy of Finland, Grant/Award Numbers: 294643, 330606, 331284; HiLIFE Fellows; Jane ja Aatos Erkon Säätiö; Sigrid Juséliuksen Säätiö; Syöpäsäätiö

Abstract

Ulcerative colitis increases colorectal cancer risk by mechanisms that remain incompletely understood. We approached this question by determining the genetic and epigenetic profiles of colitis-associated colorectal carcinomas (CA-CRC). The findings were compared to Lynch syndrome (LS), a different form of cancer predisposition that shares the importance of immunological factors in tumorigenesis. CA-CRCs (n = 27) were investigated for microsatellite instability, CpG island methylator phenotype and somatic mutations of 999 cancer-relevant genes ("Pan-cancer" panel). A subpanel of "Pan-cancer" design (578 genes) was used for LS colorectal tumors (n = 28). Mutational loads and signatures stratified CA-CRCs into three subgroups: hypermutated microsatellite-unstable (Group 1, n = 1), hypermutated microsatellite-stable (Group 2, n = 9) and nonhypermutated microsatellite-stable (Group 3, n = 17). The Group 1 tumor was the only one with *MLH1* promoter hypermethylation and exhibited the mismatch repair deficiency-associated Signatures 21 and 15. Signatures 30 and 32 characterized Group 2, whereas no prominent single signature existed in Group 3. *TP53*, the most common mutational target in CA-CRC (16/27, 59%), was similarly affected in Groups 2 and 3, but DNA repair genes and Wnt signaling genes were mutated significantly more often in Group 2. In LS tumors, the degree of hypermutability exceeded that of the hypermutated CA-CRC Groups 1 and 2, and somatic mutational profiles and signatures were different. In conclusion, Groups 1 (4%) and 3 (63%) comply with published studies, whereas Group 2 (33%) is novel. The existence of molecularly distinct subgroups within CA-CRC may guide clinical management, such as therapy options.

KEYWORDS
colorectal cancer, Lynch syndrome, microsatellite instability, somatic mutation, Ulcerative colitis

1 | BACKGROUND

Inflammatory bowel disease, comprising ulcerative colitis (UC) and Crohn's disease, is associated with an increased risk of colorectal

Abbreviations: CA-CRC, colitis-associated colorectal carcinoma; CCP, Comprehensive Cancer Panel; CIMP, CpG island methylator phenotype; CRC, colorectal carcinoma; LS, Lynch syndrome; MMR, DNA mismatch repair; MSI, microsatellite instability; MSI-H, high-degree microsatellite instability; MS-MLPA, methylation-specific multiplex ligation-dependent probe amplification; MSS, microsatellite-stable; UC, ulcerative colitis; VAF, variant allele frequency.

carcinoma (CRC).[1] UC-associated CRC (CA-CRC) develops in a multifactorial manner involving a complex imbalance of regulation and coordination of the human immune system, gut microbial composition and epithelial regeneration during the persistent inflammatory period.[1,2] Inflammation together with no mucosal healing predisposes to CA-CRC via inflammation-dysplasia-carcinoma sequence.[2] During inflammation-associated tumorigenesis, active inflammatory cells produce reactive oxygen species and reactive nitrogen intermediates, which induce mutations leading to genetic instability.[3] Cytokine production further enhances intracellular reactive oxygen species and reactive nitrogen intermediates in a malignant cell; moreover, it promotes epigenetic modifications that can accelerate tumor initiation by silencing DNA repair genes, for example.[3]

Up to 35% of CRC risk can be attributed to genetic factors, and some 5% of CRC cases represent hereditary single-gene disorders.[4] Germline defects in DNA mismatch repair (MMR) genes *MLH1*, *MSH2*, *MSH6* and *PMS2*,[5] or more rarely deletions in the 3′ end of *EPCAM* gene leading to hypermethylation of *MSH2* gene promoter,[6] cause Lynch syndrome (LS), the most prevalent form of hereditary CRC. Heterozygous germline defects lead to reduced levels of functional MMR proteins, which increases the risk for early-onset cancers, primarily CRC and endometrial cancer.[7] CRC in LS often, but not always, develops via the adenoma-carcinoma sequence.[8] The emergence of *de novo* somatic mutations contributes to high levels of neoantigens, which are thought to result in immune activation and later, immune evasion.[9] Thus, in analogy to CA-CRC, immunological alterations and inflammation accompany LS-associated colorectal tumorigenesis from the outset.

Recently, genomic analyses on cancers have revealed details of mutational processes and their timing in various cancers. Characterization of such events can be useful to understand molecular mechanisms of cancers and can help to determine plausible biomarkers for diagnostic and therapeutic use. In our study, we aim to determine the somatic mutational profiles and signatures for CA-CRC and compare the findings to LS-associated colorectal tumors, thereby covering two forms of early-onset colorectal cancer with different etiologies, but with a strong immunological component as a common denominator.

2 | MATERIALS AND METHODS

2.1 | Patients and samples

The study material consisted of formalin-fixed, paraffin-embedded tissue specimens from UC patients developing CRC (CA-CRC, n = 27) and LS patients (verified carriers of pathogenic or likely pathogenic germline variants of MMR genes) developing colorectal tumors (adenomas with high-grade dysplasia, n = 10, and CRCs, n = 18). Besides tumor material, we had the patients' normal colon tissue or blood specimens available. All the LS patients were represented in the nationwide Lynch Syndrome Registry of Finland. DNA was extracted using the nonenzymatic protocols, modified extraction protocol of the phenyl-chlorophorm method[10] and protocol described in Lahiri and Nurnberger[11] for formalin-fixed, paraffin-embedded and blood samples, respectively.

What's new?

Ulcerative colitis-associated colorectal carcinoma (CA-CRC) is a complex disease involving inflammation-associated tumorigenesis and genetic mutation. Despite extensive knowledge of germline defects linked to CA-CRC, however, molecular pathogenesis of the disease remains poorly defined. In this study, using tumor profiling, the authors describe three genetic and epigenetic CA-CRC subgroups, two of which are previously known and one that is novel. The novel subgroup consisted of hypermutated microsatellite-stable tumors, which displayed distinct mutational signatures compared to the remaining CA-CRC subgroups and Lynch syndrome tumors, suggesting pathophysiologic differences. The existence of molecular subgroups within CA-CRCs may inform treatment decisions.

2.2 | Microsatellite instability analysis

Microsatellite instability (MSI) was assayed using mononucleotide repeat markers *BAT25* and *BAT26* that are sensitive and specific markers of high-degree MSI (MSI-H).[12,13] Unstable *BAT25* and/or *BAT26* indicated MSI (specifically, MSI-H), whereas tumors in which both markers were stable were considered microsatellite-stable (MSS).

2.3 | CpG island methylator phenotype analysis

CpG island methylator phenotype (CIMP) status was assessed with commercial SALSA methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) ME042-B2 (LS samples) and ME042-C1 (UC samples) probemixes (MRC Holland, Amsterdam, the Netherlands) according to the manufacturer's protocol, and as described in Valo et al.[14] CIMP status was defined according to the Weisenberger panel[15]: a sample was considered CIMP positive when at least three out of five genes (*CACNA1G*, *IGF2*, *NEUROG1*, *RUNX3* and *SOCS1*) were methylated. A set of corresponding normal samples were examined to determine a threshold of hypermethylation for each probe with a stringency level II, details described in Valo et al[14] and values presented in Supplementary Table 1.

2.4 | Parallel sequencing

Sequencing was performed on 27 CA-CRC patients' paired tumor-normal samples (n = 54) and 28 LS tumors including adenomas with high-grade dysplasia and carcinomas, paired with the patients' normal tissue samples (n = 56). Prior to sequencing, the core facility (FIMM) conducted a LabChip gDNA analysis to evaluate the quality of each sample. CA-CRC samples were sequenced as follows: sequencing libraries were prepared using the Twist EF library kit and Twist custom capture kit (Twist

Bioscience, San Francisco, CA). The custom kit called Pan-cancer panel included probes (n = 17 314) in the regions of 999 cancer-associated genes totaling a 6.4 Mb design (Supplementary Table 2). The libraries were sequenced using a S4 cell and run on the NovaSeq platform (Illumina, San Diego, CA). Average data per sample was 6.4 Gb, and the average depth of targeted areas was 286. Detailed performance characteristics are given in Supplementary Table 3.

LS samples were sequenced as part of our earlier investigation.[16] Briefly, the Nimblegen Comprehensive Cancer Panel (CCP; Roche Diagnostics, Basel, Switzerland), a 4 Mb design with 578 cancer-related genes, was used together with ThruPLEX DNA-seq Kit (Rubicon Genomics, Ann Arbor, MI) for library preparation. Sequencing was done on the Illumina HiSeq 2500 platform. The mean target coverage of 41-fold was reached. Details of performance characteristics are described in Porkka et al.[16]

For both data sets, raw data were processed using an in-house pipeline called variant calling pipeline version 3.7,[17] and data were aligned to the human genome GRCh38. First, the adapters were trimmed from the reads as well as any bad quality nucleotides from the beginning or the end of the reads, removing any pair having read(s) smaller than 36 bp. Reads were then aligned to the GRCh38 reference genome with the BWA (version 0.6.2). Nonunique read pairs and nonunique single reads were removed and GATK (version 3.7) BaseRecalibrator was used to clean the alignment. Any potential PCR duplicates were removed using Picard (version 2.9.0). MarkDuplicates and GATK IndelRealigner were used for indel sites.

2.5 | Somatic mutation analysis

Paired tumor-normal data were compared and nonsynonymous somatic mutations (missense, nonsense, frameshift, in-frame coding deletion/insertion and splice site mutations) were extracted using the VarScan2 version 2.3.2. The following parameters and thresholds were applied: strand-filter 1, min-coverage-normal 8, min-coverage-tumor 6, somatic-*P*-value 1, normal-purity 1 and min-var-freq 0.05. Mutations were annotated using SnpEff version 4.0 with the Ensembl v86 annotation database. Misclassified germline variants were filtered out using the Database of Single Nucleotide Polymorphisms and common population variants were removed. Mutational signatures were extracted from somatic nonsynonymous mutation data using the MutationalPatterns package[18] in R. Mutation signatures were compared to sixty available single base substitution signatures of the Catalogue of Somatic Mutations in Cancer (https://cancer.sanger.ac.uk/cosmic/signatures/SBS/).[19]

2.6 | Statistical analysis

Statistical analyses were conducted using the SPSS software, version 25.0 (IBM SPSS Inc. Chicago, IL). Fisher's exact test was used to study pairwise comparisons of categorical variables. Normal distribution of continuous data was tested using the Shapiro-Wilk test. As data were largely not normally distributed and sample sizes were small, continuous variables were analyzed using the nonparametric Mann-Whitney *U* test. Correlation analyses were calculated with Spearman's or Pearson's correlation test. Exact two-sided *P* values were calculated. *P* values <.05 were considered statistically significant.

3 | RESULTS

3.1 | Study design

This investigation was undertaken to define the molecular pathogenesis of colorectal tumorigenesis in UC, an idiopathic chronic inflammatory bowel disease with accelerated tumor development by a "landscaper" mechanism.[20] The results were compared to LS, where impaired "caretaker" function due to inherited (and acquired) MMR gene defects is known to induce immunological/inflammatory alterations and result in rapid tumor development. Basic characteristics of the study series are described in Table 1.

3.2 | MSI and CIMP in CA-CRC

Only one CA-CRC tumor showed MSI-H (1/27, 4%). *MLH1* promoter methylation analysis was possible as part of the CIMP MS-MLPA protocol (see Materials and Methods). The tumor with MSI-H revealed hypermethylation of the proximal promoter region "C," which has been shown to be associated with a loss of MLH1 protein expression.[21] No other CA-CRC tumors had *MLH1* promoter methylation, consistent with stable microsatellites. The Weisenberger criteria[15] classified 11/27 (41%) of CA-CRCs as CIMP(+), which corresponds to the CIMP(+) frequency seen among LS tumors (Table 2). The CIMP MS-MLPA panel also includes probes for the *BRAF* Val600Glu mutation, and two CA-CRC tumors (7%) were mutation-positive, compared to none among LS tumors (Table 2); subsequent panel sequencing confirmed the results. The single MSI-H CA-CRC tumor did not harbor *BRAF* Val600Glu mutation, but it had *KRAS* Gly12Asp mutation (Supplementary Table 4).

3.3 | Somatic mutations in CA-CRC

By Pan-cancer panel sequencing, the average rate of nonsynonymous somatic mutations (VarScan2 somatic *P* value <.01) was 16.1 mutations/Mb (median 6.3, range 2.8-118.2/Mb) in CA-CRC tumors. Ten of 27 CA-CRC tumors (37%) were found to be hypermutated (>10 mutations/Mb). These included the single MSI tumor (84.8 mutations/Mb) and nine MSS tumors (average 29.5 mutations/Mb).

Figure 1 shows the most commonly mutated genes in CA-CRC by focusing either on genes affected by high-frequency mutations (variant allele frequency [VAF] at least 25%, which is characteristic of

TABLE 1 Characteristics of the patient samples

	CA-CRC	LS tumors combined[a]	LS HGD adenomas	LS carcinomas	*P* value CA-CRC vs LS tumors
No. of patients	27	24	10	15	NA
Male sex[b]	18 (66.7%)	11 (45.8%)	4 (40%)	8 (53.3%)	.164
Age at diagnosis, years (mean ± SD)[b]	51.1 (±10.6)	50.1 (±12.9)[c]	50.4 (±14)[c]	50.8 (±12.2)	.962
Years of colitis before CRC (mean ± SD)[b]	22.2 (±10.5)[d]	—	—	—	NA
Gene mutated in the germline[b]					
MLH1	—	19 (79.2%)	9 (90%)	11 (73.3%)	NA
MSH2	—	3 (12.5%)	1 (10%)	2 (13.3%)	NA
MSH6	—	2 (8.3%)	0	2 (13.3%)	NA
No. of tumors	27	28	10	18	NA
Tumor location					
Proximal[e]	12 (44.4%)	15 (53.6%)	2 (20.0%)	13 (72.2%)	.380
Distal	12 (44.4%)	12 (42.9%)	8 (100.0%)	4 (22.2%)	
NA	3 (11.1%)	1 (3.6%)	0	1 (5.6%)	
Stage of carcinomas					
I	12 (44.4%)	11 (61.1%)	—	11 (61.1%)	.354
II	5 (18.5%)	4 (22.2%)	—	4 (22.2%)	
III	7 (25.9%)	2 (11.1%)	—	2 (11.1%)	
IV	3 (11.1%)	0	—	0	
NA	0	1 (5.6%)	—	1 (5.6%)	

Note: Note regarding LS samples: Multiple samples were available from three LS patients (three carcinomas, two carcinomas, and a carcinoma plus adenoma from one patient each). If sampling took place at different time points (metachronous neoplasia), different ages were included in the calculation of age at diagnosis.

Abbreviations: HGD, high-grade dysplasia; NA, not available or applicable.

[a]LS-associated adenomas and carcinomas combined.

[b]Calculated per patients.

[c]Information for one case not available.

[d]Information for eight cases not available.

[e]From caecum to splenic flexure (included).

colon cancer driver genes),[22] or genes involved in at least 30% of tumors (a cutoff we have used in our previous studies on LS).[16] *TP53* was the most prominent mutational target in CA-CRC, being affected with high-frequency mutations in 14 samples out of 27 (52%) (Figure 1A) and in 18 samples (67%) if any VAF was considered (Figure 1B). Our VarScan2 annotations were based on the longest transcript and two *TP53* mutations included in Figure 1B were outside the primary transcript (ENST00000445888.6); 16/27 (59%) samples had *TP53* mutations involving the primary transcript. When VAF ≥25% and involvement in at least 30% of tumors were both required, only *TP53* fulfilled this criterion.

Figure 2 provides a detailed overview of exonic *TP53* mutations present in our CA-CRC tumors. The results are compared to mutations reported in recent investigations on UC or inflammatory bowel disease[23-27] (Figure 2A), and with *TP53* mutations detected in our LS tumors in this investigation (Figure 2B). Our CA-CRC tumors harbored 15 different mutations, mostly located in positions shown to be associated with impaired functionality when affected.[28] The most frequent mutation was p.Arg248Trp/Gln occurring in three samples with VAFs above 25% in all of them. Overall, the vast majority, 73% (11/15) of *TP53* mutations present in CA-CRC were high-frequency somatic mutations. One was a truncating mutation (p.Tyr205*) found in one sample (Figure 2A). For comparison, of *TP53* mutations present in LS-associated tumors, only 3/19 (16%) occurred with VAF of 25% or higher ($P = .0013$) (Figure 2B).

The total number of nonsynonymous somatic mutations (VarScan2 $P < .01$) per CA-CRC tumor did not significantly correlate with the number of methylated CIMP genes or probes. MMR deficiency explained hypermutability in a single tumor (the one with MSI described above). To identify possible molecular contributors for hypermutability in those CA-CRC tumors that were MSS, we investigated the prevalence of mutations in DNA repair pathway genes (n = 86) identified by Gene Ontology analysis (www.ebi.ac.uk/QuickGO). Mutations of DNA repair genes were significantly enriched in the nine hypermutated MSS cases compared to the 17 nonhypermutated MSS cases of CA-CRC: 142/189 (75%) of mutations were found in the hypermutant tumors ($P = .00001$) (Supplementary Table 5a). By a similar comparison, mutations in

TABLE 2 Comparison of central molecular features between CA-CRC and LS tumors

	CA-CRC[a]				LS tumors			*P* value CA-CRC (all) vs LS (all)
	Group 1	Group 2	Group 3	All	HGD-adenomas	CRCs	All	
No. of tumors	1	9	17	27	10	18	28	NA
MSI-H tumors	1 (100%)	0	0	1 (3.7%)	10 (100%)	16 (100%)[b]	26 (100%)[b]	<.00001
CIMP(+) tumors	0	3 (33.3%)	8 (47.1%)	11 (40.7%)	2 (20%)	9 (50%)	11 (39.3%)	.782
Average no. of somatic nonsynonymous mutations/Mb ($P < .01$)								
Pan-cancer panel	84.8	29.5	4.9	16.1	NA	NA	NA	NA
CCP	69.8	26.8	4.6	14.4	278.1	169.9	208.5	2.3E-12
No. of tumors with high-frequency[c] mutations								
BRAF Val600Glu (CCP)	0	0	2 (12%)	2 (7.4%)	0	0	0	.236
TP53 (CCP)	0	5 (55.6%)	9 (52.9%)	14 (51.9%)	1 (10%)	2 (11.1%)	3 (10.7%)	.0013
APC (CCP)	1 (100%)	2 (22.2%)	0	3 (11.1%)	6 (60%)	10 (35.7%)	16 (57.1%)	.0005
Predominant mutational signatures[d]	21, 46, 15	30, 32	(32)	30, 32	1, 46, 4, 10b, 6	1, 46, 6, 4, 10b	1, 46, 4, 10b, 6	NA

Abbreviations: HGD, high-grade dysplasia; NA, not available or applicable.

[a]Subdivision based on our molecular data as follows: Group 1, hypermutated MSI; Group 2, hypermutated MSS; and Group 3, nonhypermutated MSS tumors. Tumors with over 10 nonsynonymous somatic mutations/Mb were considered hypermutated.

[b]MSI information was unavailable for two cases, but immunohistochemical analysis suggested MMR defects in both samples.

[c]VAF 25% or higher.

[d]In decreasing order of contribution per tumor group (signatures were based on the entire Pan-Cancer panel in CA-CRC and the CCP subpanel in LS tumors).

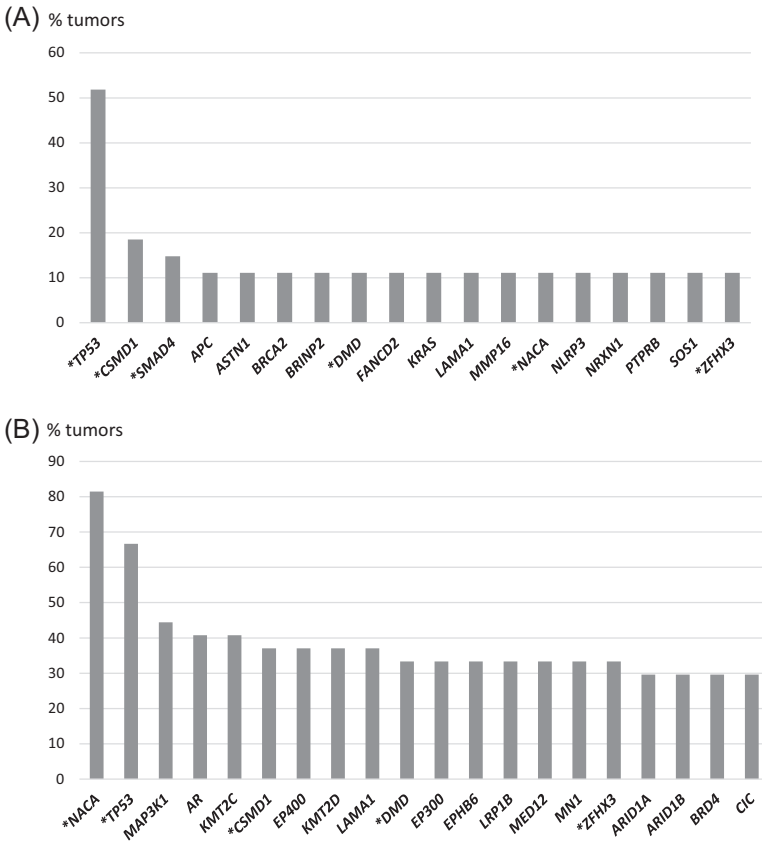

FIGURE 1 Top mutant genes in CA-CRC (n = 27 tumors), based on Pan-cancer panel (n = 999 genes) and VarScan2 annotations utilizing the longest RefSeq transcript. A, Selection criteria were variant allele frequency (VAF) 25% or higher, and at least 10% of tumors affected. B, Selection criterion was at least 30% of tumors affected (any VAF was considered). Asterisk indicates genes that fulfilled both (A) and (B)-specific criteria

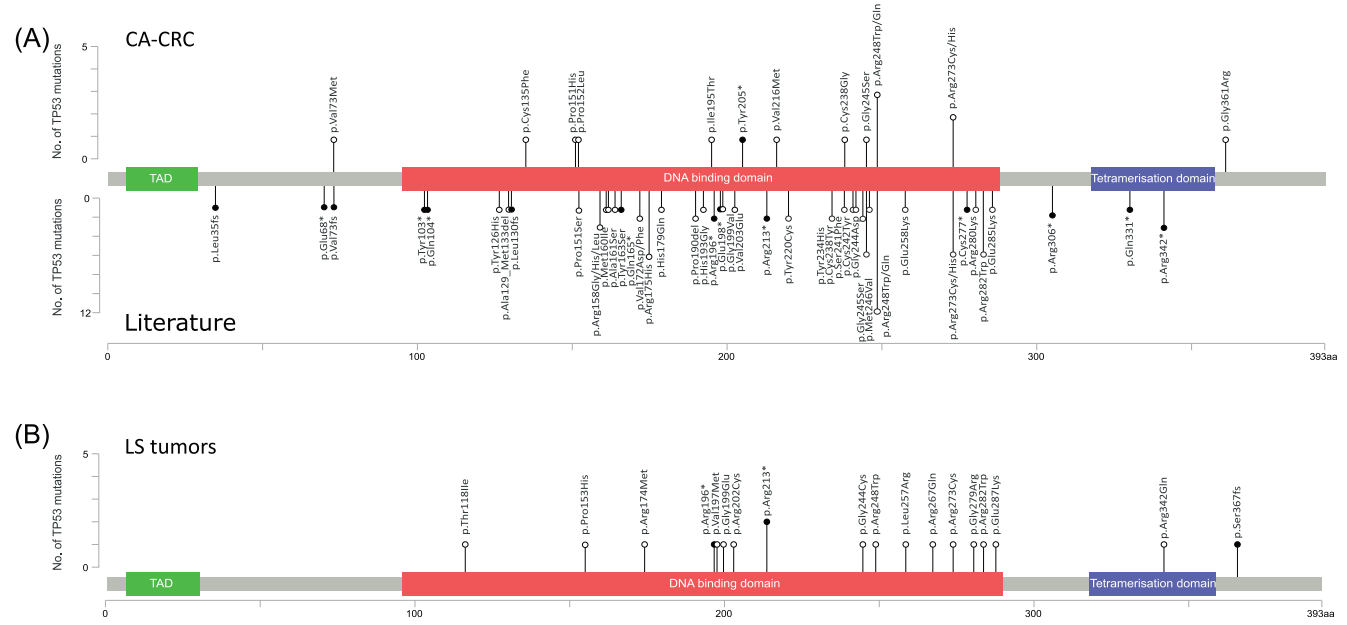

FIGURE 2 Number and distribution of exonic nonsynonymous *TP53* mutations along the protein domains (transcript ENST00000445888.6; any VAF was considered). A, Mutations found in our CA-CRC sample set are presented above the bar visualizing the protein domains. Mutations from recent studies on CA-CRC[23-27] are presented below the bar. B, Mutations found in our LS tumor material (n = 28) by CCP sequencing. White circles indicate nontruncating mutations and black circles refer to truncating mutations. TAD, transactivation domain [Color figure can be viewed at wileyonlinelibrary.com]

Wnt pathway genes (n = 44 genes) were mostly found in the hypermutated tumors: 115/134 (86%) (P = .00002) (Supplementary Table 5b). Detailed somatic mutation data for the CA-CRC samples can be found in Supplementary Table 6.

3.4 | Mutation signatures of CA-CRC tumors

Mutation signature analysis was performed on somatic mutation data (VarScan2 P < .01), and signatures were compared to known 60 single

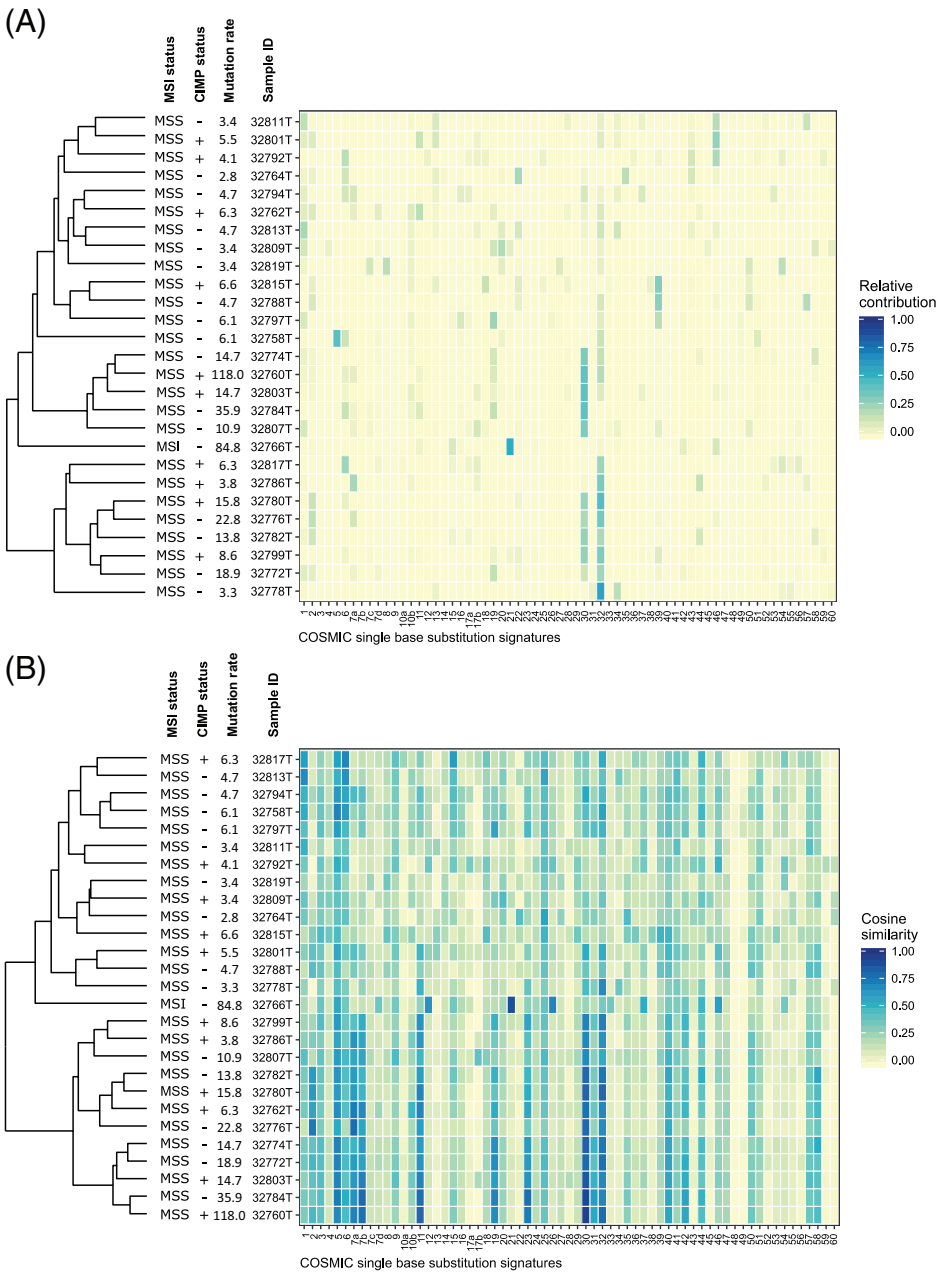

FIGURE 3 COSMIC signatures (v3) of all CA-CRC samples. A, Relative contribution heatmap. B, Cosine similarity heatmap. Mutation rate over 10/Mb identifies hypermutated samples [Color figure can be viewed at wileyonlinelibrary.com]

base substitution COSMIC signatures (v3). As evident from Figure 3, hypermutated and nonhypermutated CA-CRC tumors formed separate clusters. The single hypermutated MSI tumor clearly represented a different signature profile compared to other CA-CRC tumors. The MSI tumor showed a strong Signature 21 and a notable Signature 15 (Figure 3A), and cosine similarity additionally pinpointed Signature 26 (Figure 3B); all these three signatures are known to be associated with MSI.[29] Hypermutated MSS tumors revealed prominent Signatures 30 and 32 (Figure 3A,B). Additionally, cosine similarity to Signatures 7a and b, 11, 19 and 23 was high (Figure 3B). No single signature stood out among the nonhypermutated MSS tumors (Figure 3A,B). In group-wise analysis (Figure 4), comparison of the relative contributions of mutational signatures between hypermutated and nonhypermutated MSS CA-CRCs highlighted the predominance of Signatures 30 and 32 in the former group (relative contributions 36% and 20%, respectively) and the absence of a single dominant signature in the latter group (with the possible exception of Signature 32 that accounted for 16% of all signature contributions in this group) (Figure 4A).

3.5 | Molecular comparison of CA-CRC vs LS tumors

A study protocol analogous to CA-CRCs was applied to LS tumors, and Table 2 provides a comparative summary of the essential findings. All LS adenomas had high-grade dysplasia and did not significantly differ from LS-CRCs with respect to MSI, CIMP or somatic mutational load (Table 2); therefore, LS adenomas and carcinomas were combined to a single group ("LS tumors") throughout this article, unless

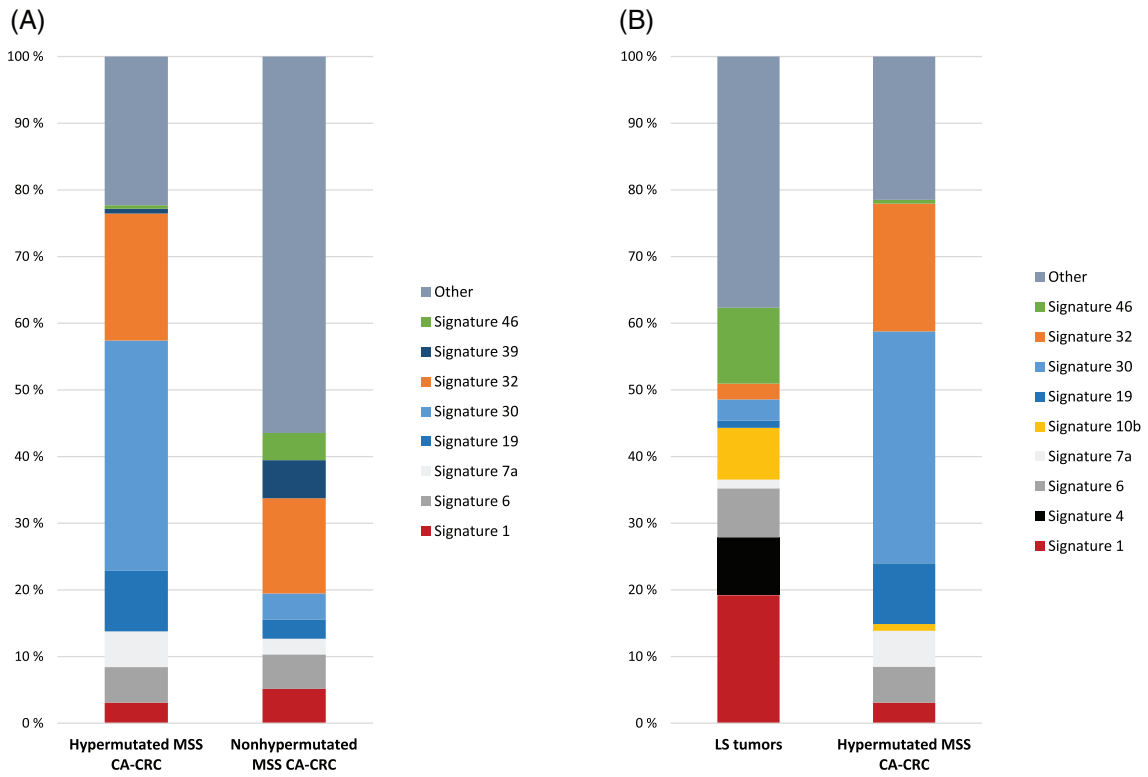

FIGURE 4 Relative contribution of mutational signatures, comparing the five largest contributors selected from each group. A, Top five signatures in hypermutated MSS CA-CRC (n = 9) and nonhypermutated MSS CA-CRC (n = 17) totaling eight signatures. B, Top five signatures in LS tumors (hypermutated MSI) (n = 28) and hypermutated MSS CA-CRC (n = 9) totaling nine signatures [Color figure can be viewed at wileyonlinelibrary.com]

specified otherwise. For LS tumors, raw sequence data based on CCP panel (n = 578 genes) was available from a previous investigation[16] and was reanalyzed by the same protocol used for the Pan-cancer data (see Materials and Methods) to enable comparison with CA-CRC results. For CA-CRC, CCP panel targets were extracted from the Pan-cancer sequencing data for the purposes of comparison.

All LS tumors showed MSI-high and were hypermutated (>10 nonsynonymous somatic mutations/Mb). The overall mutation rate (VarScan2 somatic *P* value <.01) was significantly higher in LS-associated tumors than in CA-CRC, with averages of 208.5 mutations/Mb (median 175.6/Mb, range 15.8-528.8/Mb) and 14.4 mutations/Mb (median 5.5/Mb, range 2.8-108.8/Mb), respectively (*P* = 2.3E-12; Table 2). The same applied to high-frequency mutations (VAF 25% or higher), with the average and median numbers of 17.4 mutations/Mb and 12.4 mutations/Mb in LS tumors vs 2.1 mutations/Mb and 0.75 mutations/Mb in CA-CRC tumors (*P* = 5.6E-12). Among genes most often affected by high-frequency somatic mutations (Supplementary Figure 1), *TP53* showed significantly higher mutation rates in CA-CRC than LS tumors (14/27, 52% vs 3/28, 11%, *P* = .0013; Supplementary Figure 1A; Table 2). Conversely, mutations of *APC* were significantly more common in LS tumors compared to CA-CRC (16/28, 57% vs 3/27, 11%, *P* = .0005; Table 2). By mutation type (truncating vs nontruncating), truncating (ie, nonsense or frameshift) mutations were clearly more common in LS-associated tumors (Supplementary Figure 1C) compared to CA-CRC (Supplementary Figure 1B), consistent with MSI statuses (100% of LS and 4% of CA-CRC tumors had MSI; Table 2).

Somatic mutational signatures for the individual LS tumors are depicted in Supplementary Figure 2. LS tumors clustered into two major groups; signature profiles were evidently more uniform than between the subgroups of CA-CRC samples. In LS, the overall number of nonsynonymous mutations (below vs above 100 mutations/Mb) appeared to contribute to the pattern of sample clustering. The CIMP status seemed randomly distributed across the LS clusters (the same applied to CA-CRC tumors; Figure 3). Mutation signatures of LS-associated adenomas and carcinomas were very similar. Signatures 1 (aging-associated, "clock-like"), 4 (tobacco smoking-associated), 6 (MMR deficiency-associated), 10b (*POLE*-proofreading domain mutation-associated) and 46[19,29,30] (see also https://cancer.sanger.ac.uk/cosmic/signatures/SBS/) were well represented in the LS tumors (Figure 4B; Supplementary Figure 2). When compared to hypermutated MSS CA-CRCs (n = 9), LS tumors (n = 28, all hypermutated) showed clearly distinctive mutation signatures (Figure 4B; Supplementary Figure 3–A-C), compatible with different disease etiologies.

4 | DISCUSSION

Patients with UC like those with LS are predisposed to early onset CRC. Severity,[31] duration[32] and extent[33,34] of inflammation increase

the risk of CRC in UC patients. In LS, inherited MMR deficiency constitutes the main cancer susceptibility factor[5] and results in immunoactivation[9]; some observations suggest an even more primary tumorigenic role for immunological alterations in LS.[35] Contrary to 100% of LS tumors, only one CA-CRC (4%) showed MSI in our investigation. The low frequency of MSI high in CA-CRCs complies with published studies (9%-17%).[23,36,37] MSI coincided with elevated rates of somatic nonsynonymous mutations in the single MSI CA-CRC tumor (84.8 mutations/Mb; Figure 3) and even more so in the LS tumors (average 208.5 mutations/Mb); among the latter, association with truncating mutations was also evident (Supplementary Figure 1C).

MMR, CIMP, *BRAF* Val600Glu and *KRAS* mutation statuses are widely used molecular classifiers of CRC, where MSI is associated with favorable prognosis and *BRAF* mutation with poor prognosis, whereas the clinical significance of CIMP and *KRAS* mutations is less consistent.[38,39] Our CA-CRCs did not essentially differ from sporadic CRCs[38,39] or UC-associated CRCs[40] from published studies with respect to these characteristics. As far as CIMP status is concerned, reported frequencies of CIMP(+) vary a lot, even when the same classification system (such as the Weisenberger criteria) is used. This likely reflects the technique chosen (eg, MS-MLPA vs quantitative real-time PCR) and the cutoff values set for hypermethylation. We observed CIMP(+) prevalence of 41% in CA-CRC tumors (and a comparable frequency for LS tumors, 39%), whereas real-time PCR-based studies have arrived at considerably lower frequencies (5%-17%).[40,41] Our CIMP scoring method corresponds to that described in Berg et al[42] and the CIMP(+) frequency we observed for CA-CRCs falls within the range obtained for sporadic CRCs (24%-45% depending on the stringency of criteria).[42]

The broader somatic mutational profiles based on 999 cancer-relevant genes divided our CA-CRCs into three subgroups, consisting of hypermutated MSI tumors (1/27, 4%), hypermutated MSS tumors (9/27, 33%), and nonhypermutated MSS tumors (17/27, 63%). The strongest support comes from mutational signatures that reflect the underlying pathophysiologic processes.[29] Hypermutated MSI and nonhypermutated MSS subgroups are already evident from previous studies that addressed somatic mutation patterns in CA-CRC by panel or exome sequencing.[23,26,27] To our knowledge, a significant subgroup that is MSS, yet hypermutated, has not been pointed out among MSS CA-CRCs before. Identifying this subgroup in clinical practice might open a better targeted therapeutic selection with immune checkpoint inhibitors despite the lack of MSI. Tumors associated with polymerase proofreading defects[43] may be hypermutated in the absence of MSI. In our CA-CRC series, somatic low-frequency mutations in the *POLD1* and/or *POLE* genes were detected in four tumors, including the single hypermutated MSI tumor, two hypermutated MSS tumors and one nonhypermutated MSS tumor (Supplementary Table 4). None of the mutations affected the proofreading domain and none of the tumors was associated with any (notable) Signature 10b (the hallmark of *POLE* proofreading domain mutations),[29] arguing against functional relevance of these mutations. However, the possibility that some *POLD1* or *POLE* mutations may have contributed to hypermutability cannot be totally excluded. Two mutations had been deposited to COSMIC and/or ClinVar with the status of "uncertain significance" for both; these cooccurred, although with low allele frequencies (7% for *POLD1* p.Arg1079His and 6% for *POLE* c.5678+1G>A) in a MSS tumor with the highest somatic mutational load (118 nonsynonymous mutations/Mb) in our entire CA-CRC series.

Signatures 30 and 32 predominated among our hypermutated MSS CA-CRC tumors, whereas no truly dominant single signature was present in the nonhypermutated MSS subgroup (Figures 3 and 4). Signature 30 has been linked to inactivating mutations of *NTHL1* underlying hereditary base excision repair deficiency.[44] Our sequencing panel included the *NTHL1* gene, but no somatic mutations were found. Signature 32 is associated with transcriptional bias of C>T mutations with abundance of G mutations on the untranscribed strands, and its proposed etiology is linked to immunosuppression induced by azathioprine treatment. Many of our CA-CRC patients had a history of immunosuppression medication. Recently, Inman et al found Signature 32 in cutaneous squamous cell carcinoma, and its association with Signatures 11 and 30.[45] This pattern was evident in our hypermutated MSS CA-CRC samples as well (Figures 3 and 4). Signature 11 has been linked to alkylating agents; damage caused by these agents is corrected by the MGMT enzyme that is prone to frequent inactivation by promoter methylation in CA-CRCs.[46]

Analysis of selected candidate pathways as mutational targets offered some insights into possible mechanisms that might underlie hypermutated MSS CA-CRCs. Somatic mutations of DNA repair genes (Supplementary Table 5A) and those of Wnt signaling genes (Supplementary Table 5B) were significantly more prevalent in the hypermutated MSS group compared to the nonhypermutated CA-CRCs. While the cause and effect relationships are unknown, it is possible that disruption of DNA repair and/or Wnt signaling functions contributed to the elevated overall mutation rates seen in the hypermutated MSS CA-CRC group.

COSMIC signatures[19,29] are based on single-base substitutions. To evaluate the role of copy number alterations in the pathogenesis of CA-CRC, Baker et al[23] used exome sequencing supplemented with SNP arrays and fluorescent in situ hybridization and found that copy number alterations started to increase at the transition from low-grade to high-grade dysplasia. Our approach did not allow a reliable detection of copy number alterations, which is a limitation of our investigation. *TP53* mutations, a surrogate for copy number status, were similarly distributed between hypermutated and nonhypermutated MSS CA-CRCs (Supplementary Table 4).

Contrary to sporadic CRCs, *TP53* mutations are proposed to occur as initiating events and *APC* mutations as late events in CA-CRC-related carcinogenesis.[26,47] The high VAFs and involvement in most CA-CRC tumors characteristic of *TP53* mutations we observed (Figure 1) is consistent with this notion. It has been proposed that *TP53* mutations in CA-CRC occur at different hotspots compared to sporadic CRC, although in both tumor types, DNA binding domain is the most common target of mutations.[23,26] As demonstrated in Figure 2, the locations of *TP53* mutations in our CA-CRCs were largely concordant with those reported in recent literature.[23-27] Yaeger et al[27] additionally found *MYC* to be more frequently mutated

in CA-CRC compared to sporadic cases, but we found only one case with a *MYC* mutation. Among CA-CRCs, *APC* mutations occurred exclusively in the hypermutated MSS subgroup (Supplementary Tables 4 and 5B). The relatively low overall prevalence of high-frequency mutations in *APC* (Figure 1A; Supplementary Table 4) and the fact that all samples exhibiting *APC* mutations harbored more than one *APC* mutation (Supplementary Table 4) are in agreement with the suggestion of *APC* mutations occurring late in CA-CRC-associated tumorigenesis.

Compared to hypermutated MSS CA-CRCs, mutational signatures of LS tumors were strikingly different (Figure 4B; Supplementary Figure 3A-C; Table 2). Signature 1 was prominent in LS tumors; this signature has been proposed to be associated with an endogenous mutational process of deamination of 5′ methylcytosine to thymine and is characterized by C>T transitions at methylated NpCpG sites. Formalin fixation and older age may increase this process.[29] However, the age range of our LS patients with a strong Signature 1 in tumor samples was 31 to 74 years, and there was no positive correlation between age and contribution of Signature 1 (r^2 = –0.17; P = .722). Among MMR deficiency-associated signatures, Signature 6 was prevalent among LS tumors (Supplementary Figure 2), whereas the single MSI CA-CRC sample exhibited Signatures 21, 15 and 26 (Figure 3). This agrees with the notion that Signature 6 is mainly associated with germline mutations of MMR genes, whereas Signatures 15, 20 and 26 are characteristic of somatic MMR deficiency.[48]

In conclusion, the patterns of somatic alterations stratified CA-CRCs into three subgroups: hypermutated MSI (4%), hypermutated MSS (33%) and nonhypermutated MSS (63%). Our observation of a significant hypermutated subgroup among MSS CA-CRCs, not explained by polymerase proofreading defects or defects of other currently known genes, is novel. Further studies with larger sample sets are necessary to confirm the proportional relationships of these subgroups. Molecular pathways that were differentially affected between the three CA-CRC subgroups (such as MMR and other DNA repair mechanisms; Supplementary Tables 4 and 5) can have prognostic or therapeutic significance in CA-CRC in analogy to their established relevance in sporadic CRCs or LS.[39,49]

ACKNOWLEDGMENTS

We thank Saila Saarinen for technical support and Satu Valo for coordinating UC sample collection. We also thank the Core facility Sequencing Unit at FIMM Technology Centre supported by University of Helsinki and Biocenter Finland. This study was supported by Jane and Aatos Erkko Foundation (to Päivi Peltomäki and Jukka-Pekka Mecklin); the Academy of Finland (grant numbers 294643 and 330606 to Päivi Peltomäki and 331284 to Satu Mäki-Nevala); the Finnish Cancer Foundation (to Päivi Peltomäki, Jukka-Pekka Mecklin and Ari Ristimäki); the Sigrid Juselius Foundation (to Päivi Peltomäki and Ari Ristimäki) and the HiLIFE Fellows 2017-2020 (to Päivi Peltomäki).

CONFLICT OF INTEREST

The authors declared no potential conflicts of interest.

DATA AVAILABILITY STATEMENT

All relevant data generated or analyzed during this study are included in this published article and its supplementary information files.

ETHICS STATEMENT

This study was performed in accordance with the Declaration of Helsinki. Samples were collected after informed consent. The Institutional Review Boards of the Helsinki University Central Hospital (466/E6/01, 2.10.2001 and amendment 17.12.2008), Jyväskylä Central Hospital (Dnro 10U/2011, 03.05.2011) and the Central Finland Health Care District (K-S shp Dnro4/2011) approved this research. The National Authority for Medicolegal Affairs (Dnro 1272/04/044/07) and the National Supervisory Authority for Welfare and Health (Valvira/Dnro 10741/06.01.03.01/2015, 14.1.2016) approved the collection of archival specimens.

ORCID

Alisa Olkinuora https://orcid.org/0000-0003-3987-9924

REFERENCES

1. Al Bakir I, Curtius K, Graham TA. From colitis to cancer: an evolutionary trajectory that merges maths and biology. *Front Immunol*. 2018;9:2368.
2. Rogler G. Chronic ulcerative colitis and colorectal cancer. *Cancer Lett*. 2014;345:235-241.
3. Hartnett L, Egan LJ. Inflammation, DNA methylation and colitis-associated cancer. *Carcinogenesis*. 2012;33:723-731.
4. Peters U, Bien S, Zubair N. Genetic architecture of colorectal cancer. *Gut*. 2015;64:1623-1636.
5. Thompson BA, Spurdle AB, Plazzer JP, et al. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nat Genet*. 2014;46:107-115.
6. Ligtenberg MJ, Kuiper RP, Chan TL, et al. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3′ exons of TACSTD1. *Nat Genet*. 2009;41:112-117.
7. Moller P, Seppala TT, Bernstein I, et al. Cancer risk and survival in path_MMR carriers by gene and gender up to 75 years of age: a report from the prospective Lynch syndrome database. *Gut*. 2018;67:1306-1316.
8. Ahadova A, Gallon R, Gebert J, et al. Three molecular pathways model colorectal carcinogenesis in Lynch syndrome. *Int J Cancer*. 2018;143:139-150.
9. Binder H, Hopp L, Schweiger MR, et al. Genomic and transcriptomic heterogeneity of colorectal tumours arising in Lynch syndrome. *J Pathol*. 2017;243:242-254.
10. Isola J, DeVries S, Chu L, Ghazvini S, Waldman F. Analysis of changes in DNA sequence copy number by comparative genomic hybridization in archival paraffin-embedded tumor samples. *Am J Pathol*. 1994;145:1301-1308.
11. Lahiri DK, Nurnberger JI Jr. A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res*. 1991;19:5444.
12. Esemuede I, Forslund A, Khan SA, et al. Improved testing for microsatellite instability in colorectal cancer using a simplified 3-marker assay. *Ann Surg Oncol*. 2010;17:3370-3378.
13. Loukola A, Eklin K, Laiho P, et al. Microsatellite marker analysis in screening for hereditary nonpolyposis colorectal cancer (HNPCC). *Cancer Res*. 2001;61:4545-4549.

14. Valo S, Kaur S, Ristimaki A, et al. DNA hypermethylation appears early and shows increased frequency with dysplasia in Lynch syndrome-associated colorectal adenomas and carcinomas. *Clin Epigenetics*. 2015;7:71.
15. Weisenberger DJ, Siegmund KD, Campan M, et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet*. 2006;38:787-793.
16. Porkka N, Valo S, Nieminen TT, et al. Sequencing of Lynch syndrome tumors reveals the importance of epigenetic alterations. *Oncotarget*. 2017;8:108020-108030.
17. Sulonen AM, Ellonen P, Almusa H, et al. Comparison of solution-based exome capture methods for next generation sequencing. *Genome Biol*. 2011;12:R94.
18. Blokzijl F, Janssen R, van Boxtel R, Cuppen E. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med*. 2018;10:33.
19. Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. *Nature*. 2020;578:94-101.
20. Kinzler KW, Vogelstein B. Landscaping the cancer terrain. *Science*. 1998;280:1036-1037.
21. Deng G, Peng E, Gum J, Terdiman J, Sleisenger M, Kim YS. Methylation of hMLH1 promoter correlates with the gene silencing with a region-specific manner in colorectal cancer. *Br J Cancer*. 2002;86:574-579.
22. Williams MJ, Werner B, Barnes CP, Graham TA, Sottoriva A. Identification of neutral tumor evolution across cancer types. *Nat Genet*. 2016;48:238-244.
23. Baker AM, Cross W, Curtius K, et al. Evolutionary history of human colitis-associated colorectal cancer. *Gut*. 2019;68:985-995.
24. Chakrabarty S, Varghese VK, Sahu P, et al. Targeted sequencing-based analyses of candidate gene variants in ulcerative colitis-associated colorectal neoplasia. *Br J Cancer*. 2017;117:136-143.
25. Fujita M, Matsubara N, Matsuda I, et al. Genomic landscape of colitis-associated cancer indicates the impact of chronic inflammation and its stratification by mutations in the Wnt signaling. *Oncotarget*. 2018;9:969-981.
26. Robles AI, Traverso G, Zhang M, et al. Whole-exome sequencing analyses of inflammatory bowel disease-associated colorectal cancers. *Gastroenterology*. 2016;150:931-943.
27. Yaeger R, Shah MA, Miller VA, et al. Genomic alterations observed in colitis-associated cancers are distinct from those found in sporadic colorectal cancers and vary by type of inflammatory bowel disease. *Gastroenterology*. 2016;151:278-87.e6.
28. Kotler E, Shani O, Goldfeld G, et al. Systematic p53 mutation library links differential functional impact to cancer mutation pattern and evolutionary conservation. *Mol Cell*. 2018;71:178-90.e8.
29. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature*. 2013;500:415-421.
30. Alexandrov LB, Jones PH, Wedge DC, et al. Clock-like mutational processes in human somatic cells. *Nat Genet*. 2015;47:1402-1407.
31. Rutter MD, Saunders BP, Wilkinson KH, et al. Thirty-year analysis of a colonoscopic surveillance program for neoplasia in ulcerative colitis. *Gastroenterology*. 2006;130:1030-1038.
32. Eaden JA, Abrams KR, Mayberry JF. The risk of colorectal cancer in ulcerative colitis: a meta-analysis. *Gut*. 2001;48:526-535.
33. Ekbom A, Helmick C, Zack M, Adami HO. Ulcerative colitis and colorectal cancer. A population-based study. *N Engl J Med*. 1990;323:1228-1233.
34. Gyde SN, Prior P, Allan RN, et al. Colorectal cancer in ulcerative colitis: a cohort study of primary referrals from three centres. *Gut*. 1988;29:206-217.
35. Chang K, Taggart MW, Reyes-Uribe L, et al. Immune profiling of premalignant lesions in patients with Lynch syndrome. *JAMA Oncol*. 2018;4:1085-1092.
36. Fleisher AS, Esteller M, Harpaz N, et al. Microsatellite instability in inflammatory bowel disease-associated neoplastic lesions is associated with hypermethylation and diminished expression of the DNA mismatch repair gene, hMLH1. *Cancer Res*. 2000;60:4864-4868.
37. van Dieren JM, Wink JC, Vissers KJ, et al. Chromosomal and microsatellite instability of adenocarcinomas and dysplastic lesions (DALM) in ulcerative colitis. *Diagn Mol Pathol*. 2006;15:216-222.
38. Murcia O, Juarez M, Rodriguez-Soler M, et al. Colorectal cancer molecular classification using BRAF, KRAS, microsatellite instability and CIMP status: prognostic implications and response to chemotherapy. *PLoS One*. 2018;13:e0203051.
39. Phipps AI, Limburg PJ, Baron JA, et al. Association between molecular subtypes of colorectal cancer and patient survival. *Gastroenterology*. 2015;148:77-87.e2.
40. Sanchez JA, Dejulius KL, Bronner M, Church JM, Kalady MF. Relative role of methylator and tumor suppressor pathways in ulcerative colitis-associated colon cancer. *Inflamm Bowel Dis*. 2011;17:1966-1970.
41. Olaru AV, Cheng Y, Agarwal R, et al. Unique patterns of CpG island methylation in inflammatory bowel disease-associated colorectal cancers. *Inflamm Bowel Dis*. 2012;18:641-648.
42. Berg M, Hagland HR, Soreide K. Comparison of CpG island methylator phenotype (CIMP) frequency in colon cancer using different probe- and gene-specific scoring alternatives on recommended multigene panels. *PLoS One*. 2014;9:e86657.
43. Comprehensive molecular characterization of human colon and rectal cancer. *Nature*. 2012;487:330-337.
44. Grolleman JE, de Voer RM, Elsayed FA, et al. Mutational signature analysis reveals NTHL1 deficiency to cause a multi-tumor phenotype. *Cancer Cell*. 2019;35:256-66.e5.
45. Inman GJ, Wang J, Nagano A, et al. The genomic landscape of cutaneous SCC reveals drivers and a novel azathioprine associated mutational signature. *Nat Commun*. 2018;9:3667.
46. Scarpa M, Scarpa M, Castagliuolo I, et al. Aberrant gene methylation in non-neoplastic mucosa as a predictive marker of ulcerative colitis-associated CRC. *Oncotarget*. 2016;7:10322-10331.
47. Kameyama H, Nagahashi M, Shimada Y, et al. Genomic characterization of colitis-associated colorectal cancer. *World J Surg Oncol*. 2018;16:121.
48. Grolleman JE, Diaz-Gay M, Franch-Exposito S, Castellvi-Bel S, de Voer RM. Somatic mutational signatures in polyposis and colorectal cancer. *Mol Aspects Med*. 2019;69:62-72.
49. Das S, Ciombor KK, Haraldsdottir S, Goldberg RM. Promising new agents for colorectal cancer. *Curr Treat Options Oncol*. 2018;19:29.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Mäki-Nevala S, Ukwattage S, Olkinuora A, et al. Somatic mutation profiles as molecular classifiers of ulcerative colitis-associated colorectal cancer. *Int. J. Cancer*. 2021;148:2997–3007. https://doi.org/10.1002/ijc.33492