DNA methylation changes and somatic mutations as tumorigenic events in Lynch syndrome-associated adenomas retaining mismatch repair protein expression

Mäki-Nevala, Satu

2019-01


http://hdl.handle.net/10138/298910
https://doi.org/10.1016/j.ebiom.2018.12.018

Downloaded from Helda, University of Helsinki institutional repository.
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.
DNA methylation changes and somatic mutations as tumorigenic events in Lynch syndrome-associated adenomas retaining mismatch repair protein expression

Satu Mäki-Nevala a,*, Satu Valo a, Ari Ristimäki b,c, Virinder Sarhadi b, Sakari Knuutila b, Minna Nyström d, Laura Renkonen-Sinisalo e, Anna Lepistö e, Jukka-Pekka Mecklin f, Päivi Peltomäki a

a Department of Medical and Clinical Genetics, University of Helsinki, Helsinki, Finland
b Department of Pathology, HUSLAB, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
c Genome-Scale Biology Research Program, Research Programs Unit, University of Helsinki, Finland
d Department of Biosciences, University of Helsinki, Helsinki, Finland
e Department of Surgery, Helsinki University Hospital, Helsinki, Finland
f Sport and Health Sciences, University of Jyväskylä and Jyväskylä Central Hospital, Jyväskylä, Finland

Abstract

Background: DNA mismatch repair (MMR) defects are a major factor in colorectal tumorigenesis in Lynch syndrome (LS) and 15% of sporadic cases. Some adenomas from carriers of inherited MMR gene mutations have intact MMR protein expression implying other mechanisms accelerating tumorigenesis. We determined roles of DNA methylation changes and somatic mutations in cancer-associated genes as tumorigenic events in LS-associated colorectal adenomas with intact MMR.

Methods: We investigated 122 archival colorectal specimens of normal mucosa, adenomas and carcinomas from 57 LS patients. MMR-deficient (MMR-D, n = 49) and MMR-proficient (MMR-P, n = 18) adenomas were of particular interest and were interrogated by methylation-specific multiplex ligation-dependent probe amplification and Ion Torrent sequencing.

Findings: Promoter methylation of CpG island methylator phenotype (CIMP)-associated marker genes and selected colorectal cancer (CRC)-associated tumor suppressor genes (TSGs) increased and LINE-1 methylation decreased from normal mucosa to MMR-P adenomas to MMR-D adenomas. Methylation differences were statistically significant when either adenoma group was compared with normal mucosa, but not between MMR-P and MMR-D adenomas. Significantly increased methylation was found in multiple CIMP marker genes (IGF2, NEUROG1, CRABP1, and CDKN2A) and TSGs (SFRP1 and SFRP2) in MMR-P adenomas already. Furthermore, certain CRC-associated somatic mutations, such as KRAS, were prevalent in MMR-P adenomas.

Interpretation: We conclude that DNA methylation changes and somatic mutations of cancer-associated genes might serve as an alternative pathway accelerating LS-associated tumorigenesis in the presence of proficient MMR.

1. Introduction

Lynch syndrome (LS) is a hereditary cancer predisposing syndrome caused by germline defects in DNA mismatch repair (MMR) genes MLH1, MSH2, MSH6 and PMS2 [1], or rarely deletions in the 3' end of EPICAM gene that lead to hypermethylation of MSH2 gene promoter [2]. These germline defects cause a reduced level of MMR protein, i.e. haploinsufficiency, which accelerates the occurrence of de novo somatic mutations [3] or compromises other functions of MMR genes, such as apoptosis signaling [4], thereby increasing the risk for early onset malignancies. Those defects together with epigenetic events cause an increased risk of cancer [5], primarily colorectal cancer (CRC) and endometrial cancer [6].

The “second hit” leading to a loss of the remaining functional allele of a MMR gene is typically caused by loss of heterozygosity (LOH) or somatic mutations [7–9]. However, MMR deficiency is thought to appear as a secondary event in LS tumorigenesis, supported by multiple studies observing that 10–46% of adenomas show retained expression of MMR.

* Corresponding author at: Department of Medical and Clinical Genetics, P.O. Box 63 (Haartmaninkatu 8), 00014, University of Helsinki, Finland.
E-mail address: satu.maki-nevala@helsinki.fi (S. Mäki-Nevala).

© 2019 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Recent studies, including this study, have shed light on the neo-

implications of all the available evidence associated colorectal carcinomas. Ways, as supported by recent findings from Lynch syndrome—tumorigenesis in Lynch syndrome may be driven by different path-

retained protein expression. The emerging differences between mutations observed in mismatch repair proficient adenomas may indicate pattern of base substitutions. Utilization of such mutational preferences to determine if a given alteration has occurred before or after mismatch repair deficiency, together with other approaches, has resulted in varying conclusions of early or late appearance of mismatch repair defects in Lynch syndrome tumorigenesis. Based on the above cited and other (recent) literature retrieved from PubMed, together with our previous observation that a mismatch repair-proficient subset exists among adenomas from a nation-wide registry of Lynch syndrome individuals, this study was undertaken to explore the tumorigenic events in Lynch syndrome-associated colorectal adenomas retaining MMR protein expression.

Added value of this study

Our investigation is one of the very few existing studies focusing on mismatch repair-proficient colorectal adenomas from Lynch syndrome mutation carriers, with the aim to determine the roles of DNA methylation changes and somatic mutations in cancer-associated genes in such neoplastic lesions. We show that tumor initiation may precede the loss of mismatch repair protein and identify molecular alterations present in adenomas with retained mismatch repair protein expression. While many of these changes were more prevalent in mismatch repair deficient adenomas, some were less frequent or even absent in the latter. Specifically, we found that hypermethylation of IGF2, NEUROG1, CDKN2A, CRABP1, SFRP1, and SFRP2, and especially KRAS mutations may accelerate, or possibly initiate, Lynch syndrome-associated tumorigenesis when the wild-type allele of the predisposing mismatch repair gene is still present as judged from retained protein expression. The emerging differences between mismatch repair deficient and proficient adenomas suggest that tumorigenesis in Lynch syndrome may be driven by different pathways, as supported by recent findings from Lynch syndrome-associated colorectal carcinomas.

Implications of all the available evidence

Recent studies, including this study, have shed light on the neoplastic process associated with colon tumorigenesis in Lynch syndrome. The fact that adenoma development is possible in the absence of biallelic inactivation of the predisposing mismatch repair gene shows that the traditional Knudson’s two-hit mechanism is not the only one that applies to mismatch repair gene mutation-associated tumorigenesis. Genetic and epigenetic alterations observed in mismatch repair proficient adenomas may qualify for alternative tumor-initiating or promoting events. The findings have important clinical implications. As 27% of Lynch syndrome adenomas from our investigation and up to 46% of adenomas from published studies express the mismatch repair protein corresponding to the gene mutant in the germline, retained mismatch repair protein expression in a colorectal adenoma cannot be used to rule out Lynch syndrome in diagnostics. Genetic and epigenetic changes detected in mismatch repair proficient adenomas may provide potential biomarkers of increased tumor risk. Genetic heterogeneity in Lynch syndrome adenomas and carcinomas may be relevant for the design of targeted therapies or preventive measures. Larger sample sizes are required to confirm the recent findings described above and the specific contributions of the observed alterations to tumorigenesis. Further investigations are also needed to explore the extent to which the findings from Lynch syndrome may apply to sporadic colorectal tumorigenesis.

2. Materials and methods

2.1. Patient samples

Study material consisted of formalin-fixed, paraffin-embedded (FFPE) specimens of normal mucosae, adenomas and carcinomas from
DNA samples from adenoma specimens were quantified by the chlorophorm method [24]. DNA concentrations were measured by the Qubit™ system. A total of 63 DNA samples of FFPE adenoma specimens were eligible for the sequencing study. Libraries were prepared from 4 to 15 ng of input DNA according to the manufacturer’s protocol of the Ion AmpliSeq Colon and Lung Cancer Panel (Thermo Fisher Scientific). The panel targets mutational hotspots in 22 genes totaling 14.6 kb: AKTI, ALK, BRAF, CTNNB1, DDR2, EGRF, ERBB2, ERBB4, FBXW7, FGFR1, FGFR2, FGFR3, KRAS, MAP2K1, MET, NOTCH1, NRAS, PIK3CA, PTEN, SMAD4, STK11, and TP53. The libraries were purified using the AMPure XP magnetic beads (Beckman Coulter, Krefeld, Germany), and the quantification was performed with the Qubit fluorometer with the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). Library amplification and enrichment steps were performed according to the manufacturer’s protocol of the Ion PGM Hi-Q View OT2 Kit (Rev.C0) (Thermo Fisher Scientific) on the Ion OneTouch2 and Ion OneTouch ES instruments, respectively (Thermo Fisher Scientific). Sequencing was performed using the Ion PGM Hi-Q View Sequencing Kit (Rev.C0) on the Ion PGM 316v2 Chip (Thermo Fisher Scientific). The barcoded (Ion Xpress Barcode Adapters, Thermo Fisher Scientific) libraries were pooled in the sets of 8–16 libraries. The sequencing data were analyzed utilizing the Torrent Suite software (v5.2.2). Coverage analysis was performed by using the plug-in more detail in Valo et al. [9] Briefly, 150–250 ng of DNA was used for each reaction. MS-MLPA is based on the hybridizing probes including the recognition sequence of GC-rich CpG islands for the methylation-sensitive endonuclease Hhal. Methylation prevents digestion reaction to occur, and after the PCR amplification a signal peak appears if template DNA is methylated [28]. Methylation dosage ratios \( (D_m) \) were calculated by the following formula: \( D_m = (P_{\text{norm}} / P_{\text{inhom}}) \), where \( D_m \) is a methylation dosage ratio \( (\text{degree of methylation}) \), \( P_{\text{norm}} \) is a peak area of the given probe, \( P_{\text{inhom}} \) is a peak area of all control probes, \( D_m \) stands for normal mucosa samples determined a threshold value for the hypermethylation for each probe. In tumor samples, probes were defined methylated, if the degree of methylation exceeded the threshold value. The results from individual tumor samples were combined to define the relative frequency of hypermethylation.

CIMP status was investigated using the SALSA MLPA probemix ME042-B2 (MRG Holland) including 3–6 probes for the promoter of each marker gene: CACNA1G, IGF2, NEUROG1, RUNX3, SOCS1, CDNKN2A, MLH1 and CRABP1. The gene was considered methylated when one or more of the probes were methylated [29]. The tumor sample was classified CIMP(-) using the Weisenberger panel: three or more genes out of CACNA1G, IGF2, NEUROG1, RUNX3 and SOCS1 should be methylated [30].

LINE-1 hypomethylation was studied by the custom-made LINE-1-MS-MLPA assay described in Pavicic et al. [31] Briefly, the assay included in total of ten probes, of those three were methylation-sensitive probes targeting regions with the Hhal restriction site inside the LINE-1 promoter sequence. The remaining seven probes provided a set of control probes targeting other regions without the restriction site in the LINE-1 sequence. LINE-1 hypomethylation was calculated based on L1-M2 probe [31].

Similarly, the custom-made MS-MLPA panel was designed to study promoter methylation of the selected candidate TSGs involved in early colon tumorigenesis in a mouse model for LS: DKK1, SFRP2, CDH1, HOXD1, SFRP5, SLCSA8, and SFRP1 [23]. Detailed design is described in Valo et al. [9] The custom-made probes were used together with the reference probe mix of the SALSA MLPA P300-A2 Human DNA Reference-2 (MRG Holland).

2.6. Amplicon-based next generation sequencing (NGS)

In total of 63 DNA samples of FFPE adenoma specimens were eligible for the sequencing study. Libraries were prepared from 4 to 15 ng of input DNA according to the manufacturer’s protocol of the Ion AmpliSeq Library Kit 2.0 (Rev.C0) (Thermo Fisher Scientific) using the Ion AmpliSeq Colon and Lung Cancer Panel (Thermo Fisher Scientific). The library targets mutational hotspots in 22 genes totaling 14.6 kb: AKTI, ALK, BRAF, CTNNB1, DDR2, EGRF, ERBB2, ERBB4, FBXW7, FGFR1, FGFR2, FGFR3, KRAS, MAP2K1, MET, NOTCH1, NRAS, PIK3CA, PTEN, SMAD4, STK11, and TP53. The libraries were purified using the Ampure XP magnetic beads (Beckman Coulter, Krefeld, Germany), and the quantification was performed with the Qubit fluorometer with the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific). Library amplification and enrichment steps were performed according to the manufacturer’s protocol of the Ion PGM Hi-Q View OT2 Kit (Rev.C0) (Thermo Fisher Scientific) on the Ion OneTouch2 and Ion OneTouch ES instruments, respectively (Thermo Fisher Scientific). Sequencing was performed using the Ion PGM Hi-Q View Sequencing Kit (Rev.C0) on the Ion PGM 316v2 Chip (Thermo Fisher Scientific). The barcoded (Ion Xpress Barcode Adapters, Thermo Fisher Scientific) libraries were pooled in the sets of 8–16 libraries.

The sequencing data were analyzed utilizing the Torrent Suite software (v5.2.2). Coverage analysis was performed by using the plug-in
of coverageAnalysis (v5.2.1.2), and variant calling was conducted with the variantCaller (v5.2.1.39). Default analysis parameters were applied on signal processing, base calling, read alignment and variant calling. Reads were aligned at the human genome reference assembly of h19. Variant calling files were imported in the VarSeq software (GoldenHelix) and annotated with the databases of COSMIC Cancer Gene Census 71 [32], ExAC Variant Frequencies 0.3, dbSNP 149 (NCBI), and RefSeq Genes 105 Interim v1 (NCBI), and in silico functional prediction tools (SIFT, PolyPhen2 HVAR, MutationTaster, MutationAssessor, FATHMM, and FATHMM MKL). For the final step of variant analysis, only variants with quality score $\geq 30$, read depth $\geq 50$, and VAF $\geq 0.05$ were considered true positives. Moreover, all variants mapped outside the target regions, producing synonymous variant, or variants reported in the 1000 Genomes project (NCBI dbSNP 149) and/or in the ExaC database with minor allele frequency (MAF) of 0.01 or more, were filtered out.

2.7. Statistical analysis

Statistical analyses were performed using the SPSS software, version 24.0 (IBM SPSS Inc., Chicago, IL, USA). Group-wise comparisons for categorical variables were calculated by using the Fisher’s exact-test. For continuous variables, normal distribution was tested by Shapiro-Wilk test, and a parametric (Student’s t-test or one-way ANOVA with Tukey’s post hoc test, homogeneity of variances tested by Levene’s test) or non-parametric (Mann-Whitney U test/Kruskal-Wallis one-way ANOVA) test was used, as appropriate. Two-sided $p$ values $<0.05$ were considered significant. The Bonferroni correction method was applied on $p$ values obtained from multiple comparisons.

2.8. Data sharing statement

All relevant data supporting this study is published along the paper. Case-specific DNA methylation data is available upon request.

3. Results

3.1. MMR protein expression and MSI status

The colorectal specimen collection described in Table 1 was used for IHC investigations to evaluate if the remaining wild-type allele of the germline mutated MMR gene had undergone somatic inactivation. Protein expression of the relevant MMR genes was lost in all carcinomas (Table 2). Among adenomas, 35% (16/46) of low dysplasia cases and 10% (2/21) of high dysplasia cases retained the protein expression of the predisposing MMR gene (Table 2). Normal mucosae ($n=43$) were tested previously and all showed normal MMR protein expression [9].

Table 2
The proportions of retained MMR protein expression in adenomas and carcinomas by IHC (extended series of the previous study [9]).

<table>
<thead>
<tr>
<th></th>
<th>MLH1</th>
<th>MSH2</th>
<th>MSH6</th>
<th>Total*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenoma with low-grade dysplasia</td>
<td>11/34 (32%)</td>
<td>3/8 (38%)</td>
<td>2/4 (50%)</td>
<td>16/46 (35%)</td>
</tr>
<tr>
<td>Adenoma with high-grade dysplasia</td>
<td>1/17 (6%)</td>
<td>0/2 (0%)</td>
<td>1/2 (50%)</td>
<td>2/21 (10%)</td>
</tr>
<tr>
<td>Carcinoma</td>
<td>0/15 (0%)</td>
<td>0/3 (0%)</td>
<td>0/5 (0%)</td>
<td>0/23 (0%)</td>
</tr>
</tbody>
</table>

*a MMR IHC could not be performed on two carcinoma and one adenoma with low-grade dysplasia specimens, and thus, they are missing from this table compared to the numbers presented in Table 1.

Fig. 1. Frequencies of methylated genes in normal colon, and MMR-P and MMR-D adenoma specimens. a. Hypermethylation frequencies of CIMP marker genes and frequencies of CIMP positive phenotype. b. Hypermethylation frequencies of candidate tumor suppressor genes. Bonferroni corrected two-sided $p$ values are presented for pairwise comparisons. Abbreviations: MMR-D, MMR deficient; MMR-P, MMR proficient; NS, non-significant.

Abbreviations: MMR-D, MMR deficient; MMR-P, MMR proficient; NS, non-significant.
MSI status was determined by the mononucleotide markers BAT25 and BAT26. We found a clear association between IHC and MSI status ($p < 0.0001$). All tested carcinomas and all adenomas with absent MMR protein showed MSI, with one exception of adenoma with low dysplasia having stable microsatellites. All adenomas with intact MMR protein expression for which enough DNA was available for MSI analysis ($n = 11$) were MSS. Throughout this paper, adenomas with absent MMR protein or presence of MSI or both are considered MMR-deficient (MMR-D, $n = 49$) and those with neither abnormality MMR-proficient (MMR-P, $n = 18$).

The existence of significant MMR-proficient subsets of LS-adenomas indicates the necessity of other somatic events to initiate malignant transformation. The normal colon, MMR-P, and MMR-D adenoma specimens were compared in terms of the average degrees of methylation. The CIMP probes showing statistical significance ($p < 0.05$) with Kruskal-Wallis test are highlighted. The average degree of methylation of SFRP1 and SFRP2 is presented here with Bonferroni corrected two-sided $p$ values $< 0.10$ for the pairwise comparisons. The upper and lower edges of the boxes are the 75th and 25th percentiles, the horizontal line inside the box denotes the median, the whiskers indicate the highest and lowest values, and the asterisks and open circles stand for outliers. Abbreviations: Dm, methylation dosage; MMR-D, MMR deficient; MMR-P, MMR proficient.
transformation, leading us to investigate DNA methylation changes and somatic mutations in LS-associated adenomas (see below).

3.2. DNA methylation changes

3.2.1. CIMP markers

Methylation analysis of eight genes (CACA1G, IGF2, NEUROG1, RUNX3, SOCS1, CDKN2A, MLH1 and CRABP1) linked to CIMP was used to assess the frequency of CIMP, a high-degree coordinated methylation of CpG islands that are normally unmethylated. CIMP (+) was defined according to the Weisenberger panel [30]. The frequency of CIMP (+) increased along with the dysplasia grade, occurring in 0% (0/29), 11% (4/37), 17% (3/18), and 45% (10/22) of normal mucosae, adenomas with low-grade dysplasia, adenomas with high-grade dysplasia, and carcinomas, respectively.

Altered methylation was evaluated in two ways, by determining the frequencies of hypermethylated samples based on defined cut-off methylation dosage (Dm) values for hypermethylation for each marker gene with normal mucosa as a reference (see Materials and Methods) and alternatively, by treating the degree of methylation (by Dm values) as a continuous variable in the samples. Fig. 1a shows the frequencies of hypermethylation for eight CIMP markers in adenomas stratified by MMR status, with normal mucosa as a reference. Hypermethylation frequency of NEUROG1 was increased already in MMR-P adenomas compared to normal colon (p = 0.009), and methylation frequency was higher than in MMR-D adenomas, although statistical significance was not obtained between MMR-P and MMR-D cases. MMR-D adenomas showed significantly elevated frequencies of hypermethylation relative to normal mucosa at IGF2 (p < 0.0001), NEUROG1 (p = 0.009) and CRABP1 (p = 0.045) loci. Among adenomas with MLH1 methylation depicted in Figs. 1a and 2a, only one, a MMR-D adenoma with high-grade dysplasia, revealed hypermethylation with the probe targeting the promoter region “C” of MLH1 (MLH1-I probe); hypermethylation of this region is associated with loss of MLH1 protein [33]. In the remaining cases, other regions of the MLH1 promoter (MLH1-II, III, and IV probes), not associated with protein expression, were involved.

When examining the average Dm values for the individual probes contributing to each CIMP marker gene (Fig. 2a), at least one probe of the following marker genes revealed a significant increase already in MMR-P adenomas vs. normal colon: IGF2, NEUROG1, CDKN2A, and CRABP1. Of those, almost all probes showed significantly increased methylation also in MMR-D adenomas vs. normal colon. There were two probes, one for CRABP1 and one for NEUROG1, which showed significantly elevated methylation solely in MMR-P adenomas (vs. normal colon).

3.2.2. Candidate tumor suppressor genes

We studied seven candidate TSGs (DKK1, SFRP2, CDH1, HOXD1, SFRP5, SLC5A8, and SFRP1) associated with early colon oncogenesis and being downregulated by promoter methylation in an experimental mouse model for LS [23]. Similarly to CIMP marker analysis, we first compared the hypermethylation frequencies of each gene in normal colon, MMR-P and MMR-D adenomas. Here, both SFRP1 and SFRP2 showed significantly increased frequency of hypermethylation in both MMR-P and MMR-D tumors compared to normal colon (Fig. 1b). Likewise, in the analysis of average Dm values, SFRP1 and SFRP2 showed increased methylation levels in both MMR-P and MMR-D adenomas compared to normal colon (Fig. 2b, c).

3.2.3. LINE-1

The average degree of methylation of LINE-1, a marker for global hypomethylation, decreased along with dysplasia (p = 0.005) (Fig. 3a). Statistically significant difference relative to normal colon was observed for carcinoma, but not for adenomas with high- or low-grade dysplasia, suggesting that LINE-1 methylation is a rather late event in LS-associated tumorigenesis. When comparing adenomas stratified by MMR status with normal colon, significantly decreased methylation level was observed for MMR-D adenomas (Fig. 3b), indicating that LINE-1 hypomethylation is likely to occur after the loss of MMR protein.

4. Effect of dysplasia grade on methylation results

To produce the results given in Figs. 1–3, adenomas with high- and low-grade dysplasia were combined. The same analyses were performed on adenomas with low-grade dysplasia only and are available in Supplementary Materials (Supplementary Fig. S1–S3). Adenomas with high-grade dysplasia included too few MMR-P cases (Table 2) and could not be studied as an independent group. As evident from Supplementary Figs. S1–S3, the results remained essentially similar to those derived from adenomas not stratified by dysplasia grade. Moreover, to exclude the effect of dysplasia grade on the results, we compared adenomas with high-grade dysplasia to those with low-grade dysplasia.
study produced good quality sequencing library and sequencing results, even in those samples with a very low DNA content. In this panel served as an excellent option to study somatic mutations (Methods). As the amount of sample material of adenoma specimens was limited, the result was weakly interpretable (MSI?).

### Table: 4.1. Somatic mutations

<table>
<thead>
<tr>
<th>SampleID</th>
<th>MMR status</th>
<th>CIMP status</th>
<th>CIMP-AMP</th>
<th>Dysplasia grade</th>
<th>KRAS</th>
<th>NRAS</th>
<th>BRAF</th>
<th>CTNNB1</th>
<th>FBXW7</th>
<th>FGFR1</th>
<th>FGFR2</th>
<th>FGFR3</th>
<th>MET</th>
<th>PIK3CA</th>
<th>ABHDG</th>
<th>NOTCH1</th>
<th>P53</th>
</tr>
</thead>
<tbody>
<tr>
<td>AdHGD</td>
<td>MMR-D</td>
<td>CIMP(+)</td>
<td>MSS</td>
<td>Low</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AdLGD</td>
<td>MMR-D</td>
<td>CIMP(+)</td>
<td>MSS</td>
<td>Low</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIMP(-)</td>
<td>Nonsense</td>
<td>Frameshift</td>
<td>Low</td>
<td>Low</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIMP(-)</td>
<td>Frameshift</td>
<td>Nonsense</td>
<td>Low</td>
<td>Low</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 4.** Mutation status of adenomas stratified by their MMR status and number of mutations. Numbers presented in this table are reported in the COSMIC database (v71). Numbers in the colored cells indicate the number of mutations affecting a given gene in the sample in question, and color distinguishes the type of mutation. Table below shows mutation frequencies in adenomas stratified by MMR status or dysplasia grade. Note: question mark after MSI or MSS status indicates that only one probe produced an interpretable result (for MSS?) or due to poor quality, the result was weakly interpretable (MSI?). Abbreviations: AdHGD, adenoma with high-grade dysplasia; AdLGD, adenoma with low-grade dysplasia; CIMP, CpG island methylator phenotype; D, deficient; MMR, DNA mismatch repair; MMR-D, MMR proficient; ND, not done; P, proficient.

regardless of MMR status, and no significant differences were observed (data not shown).

### 4.1. Somatic mutations

Adenomas were studied for somatic mutations targeting the mutational hotspots in 22 cancer-associated genes (see Materials and Methods). As the amount of sample material of adenoma specimens was limited, this panel served as an excellent option to study somatic mutations even in those samples with a very low DNA content. In total, 59 out of initial 63 samples that were eligible for the sequencing study produced good quality sequencing library and sequencing results, including 16 MMR-P adenomas, out of which a majority (n = 14) were adenomas with low-grade dysplasia. For the obtained sequencing data, mean values for the read depth, the mapping uniformity between the amplicons, and the percentage of reads mapped on targets were 1379 (± 176 CI 95%), 94% (± 2 CI 95%), and 92% (± 1 CI 95%) respectively.

Here, we present non-synonymous mutations that have previously been reported in the COSMIC database [32], emphasizing their cancer-associated nature. The most frequently mutated genes were TP53, followed by KRAS, SMAD4, and CTNNB1, found in 24% (n = 14), 22% (n = 13), 19% (n = 11), and 15% (n = 9) of the adenomas, respectively. Overall, 54% (32/59) of adenomas harbored at least one mutation. One MMR-P sample (249Ad) included clearly more mutations (n = 7) than other MMR-P adenomas (Fig. 4). This particular case was the only CIMP(+) MMR-P adenoma. In total, there were five CIMP(+)
adrenomas and three of those harbored several mutations in multiple genes (Fig. 4). A vast majority of the mutations were missense mutations (Fig. 4). The full list of our COSMIC mutations is available in Supplementary Table S3.

We observed a somewhat higher average number of mutations in MMR-D adrenomas compared to MMR-P adrenomas (2-4 vs. 1-0), although no statistical difference was found (p = 0.3-23). As the most intriguing finding, KRAS mutations were clearly more prevalent in MMR-P adrenomas (Fig. 4) (5/16, 31% vs. 8/43, 19%), although the difference is statistically non-significant. Considering only KRAS mutations in the codons 12, 13, 61, or 146 regarded pathogenic [34-36], these were clearly more prevalent in MMR-P tumors (5/16, 31% vs. 7/43, 16%). The observation remained the same when only adrenomas with low-grade dysplasia were included: the frequency of pathogenic KRAS mutations was 29% (4/14) in MMR-P cases vs. 15% (4/27) in MMR-D cases. Detailed molecular characteristics of MMR-P adrenomas are presented case by case in Table 3.

Mutational frequencies between adrenomas with low-grade and high-grade dysplasia were mostly similar, with no statistically significant differences observed (Fig. 4). However, multiple genes, such as BRAF (4/41, 10% vs. 1/18, 6%) and SMAD4 (8/41, 20% vs. 3/18, 17%) were slightly more frequently mutated in adrenomas with low-grade dysplasia than in tumors with high-grade dysplasia. Instead, mutations in CTNNB1 (5/18, 28% vs. 4/41, 10%), KRAS (5/18, 28% vs. 8/41, 20%), and TP53 (6/18, 33% vs. 8/41, 20%) occurred more often in advanced stage adrenomas.

One adrena with low-grade dysplasia (256Ad) carried the BRAF V600E mutation (VAF 5%). The BRAF mutation was the only mutation detected in this specimen. Other BRAF missense mutations that were found in four cases occurred in the codons 455, 467, and 469. The BRAF V600E mutant tumor was characterized by MMR-D, CIMP(+), and germline point mutation in MLHI (c.543C > G). Moreover, this tumor harbored hypermethylation in multiple promoters, including all three probes of IGF2, two probes of NEUROG1, one CRABP1, all three probes of RUNX3, SP1, HOXD1, and SPRI2.

We also addressed the mutation signatures by Alexandrov et al. [37], taking the MMR deficiency related signature 6 as a reference. The adjacent bases for single base pair substitutions were determined and mutation patterns constructed for adrenomas based on the relative frequencies of mutation types among all mutations found (Fig. 5a, b). In this analysis, in addition to the COSMIC mutations, we also included other variants fulfilling the filtering criteria (see Materials and Methods).

Table 3: Molecular characteristics of Lynch syndrome adrenoma specimens with retained MMR protein expression, undergone somatic mutation analysis.

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Sample ID</th>
<th>Dysplasia grade</th>
<th>Germline mutation</th>
<th>Hypermethylated CIMP markers</th>
<th>Hypermethylated candidate TSGs</th>
<th>Somatic mutations (VAF %)</th>
<th>MMR status</th>
<th>CIMP status</th>
</tr>
</thead>
<tbody>
<tr>
<td>73</td>
<td>148Ad</td>
<td>LOW</td>
<td>MLHI c.1731 + 2247_1897-402del (Mutation 1)</td>
<td>IGF2</td>
<td>None</td>
<td>None</td>
<td>MSS</td>
<td>(−)</td>
</tr>
<tr>
<td>3</td>
<td>164Ad</td>
<td>LOW</td>
<td>MLHI c.1731 + 2247_1897-402del (Mutation 1)</td>
<td>ND</td>
<td>ND</td>
<td>None</td>
<td>MSS</td>
<td>ND</td>
</tr>
<tr>
<td>16</td>
<td>187Ad</td>
<td>LOW</td>
<td>MSH6 c.900dup</td>
<td>NEUROG1</td>
<td>SFRP1</td>
<td>CTNNB1 p.Gly455Phe (37)</td>
<td>MSS</td>
<td>(−)</td>
</tr>
<tr>
<td>95</td>
<td>197Ad</td>
<td>LOW</td>
<td>MLHI c.1731 + 2247_1897-402del (Mutation 1)</td>
<td>CDKN2A</td>
<td>SFRP2, SLC5A, SFRP1</td>
<td>KRAS p.Gly12Asp (7)</td>
<td>MSS</td>
<td>(−)</td>
</tr>
<tr>
<td>13</td>
<td>248Ad</td>
<td>LOW</td>
<td>MSH2 c.1667_1671del</td>
<td>CDKN2A</td>
<td>SFRP2, SLC5A, SFRP1</td>
<td>KRAS p.Gly12Asp (7)</td>
<td>MSS</td>
<td>(−)</td>
</tr>
<tr>
<td>45</td>
<td>259Ad</td>
<td>LOW</td>
<td>MLHI c.1731 + 2247_1897-402del (Mutation 1)</td>
<td>NEUROG1, CDKN2A, CRABP1</td>
<td>SFRP1</td>
<td>KRAS p.Gly12Val (29), SMAD4 p.Pro102Leu (18)</td>
<td>MSS</td>
<td>(−)</td>
</tr>
<tr>
<td>76</td>
<td>264Ad</td>
<td>LOW</td>
<td>MLHI c.1731 + 2247_1897-402del (Mutation 1)</td>
<td>NEUROG1, CDKN2A, CRABP1</td>
<td>SFRP1</td>
<td>KRAS p.Gly12Val (29), SMAD4 p.Pro102Leu (18)</td>
<td>MSS</td>
<td>(−)</td>
</tr>
<tr>
<td>95</td>
<td>268Ad</td>
<td>LOW</td>
<td>MLHI c.1731 + 2247_1897-402del (Mutation 1)</td>
<td>MLHI, IGF2, NEUROG1, CRABP1</td>
<td>None</td>
<td>CTNNB1 p.Thr41Ala (5)</td>
<td>MSS</td>
<td>(−)</td>
</tr>
<tr>
<td>101</td>
<td>271Ad</td>
<td>LOW</td>
<td>MLHI c.1731 + 2247_1897-402del (Mutation 1)</td>
<td>None</td>
<td>SFRP1</td>
<td>None</td>
<td>MSS</td>
<td>(−)</td>
</tr>
<tr>
<td>12</td>
<td>670Ad</td>
<td>LOW</td>
<td>MSH2 c.1552_1553_del</td>
<td>ND</td>
<td>SFRP2, CDH1, HOXD1, SLC5A, SFRP1</td>
<td>FEBXW7 p.Arg465His (7), KRAS p.Ala146Thr (10)</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>16</td>
<td>672Ad</td>
<td>LOW</td>
<td>MSH6 c.900dup</td>
<td>ND</td>
<td>SFRP2, CDH1, HOXD1, SLC5A, SFRP1</td>
<td>None</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>73</td>
<td>677Ad</td>
<td>HIGH</td>
<td>MLHI c.1731 + 2247_1897-402del (Mutation 1)</td>
<td>IGF2, NEUROG1, CDKN2A, CRABP1</td>
<td>SFRP2, SLC5A, SFRP1</td>
<td>KRAS p.Ala146Thr (45), TP53 p.Glu180Lys (5)</td>
<td>MSS</td>
<td>(−)</td>
</tr>
<tr>
<td>112</td>
<td>686Ad</td>
<td>HIGH</td>
<td>MSH6 c.3195_3199del</td>
<td>IGF2, NEUROG1, CDKN2A, CRABP1</td>
<td>SFRP2, SLC5A, SFRP1</td>
<td>KRAS p.Ala146Thr (45), TP53 p.Glu180Lys (5)</td>
<td>MSS</td>
<td>(−)</td>
</tr>
<tr>
<td>120</td>
<td>690Ad</td>
<td>LOW</td>
<td>MLHI c.1731 + 2247_1897-402del (Mutation 1)</td>
<td>None</td>
<td>SFRP2, CDH1, HOXD1, SLC5, SLC5A, SFRP1</td>
<td>None</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>123</td>
<td>693Ad</td>
<td>LOW</td>
<td>MLHI c.1731 + 2247_1897-402del (Mutation 1)</td>
<td>IGF2, NEUROG1, CDKN2A</td>
<td>SFRP2, CDH1, HOXD1, SLC5, SLC5A, SFRP1</td>
<td>None</td>
<td>MSS</td>
<td>(−)</td>
</tr>
</tbody>
</table>

Abbreviations: CIMP, Cpg island methylator phenotype; MSS, microsatellite stable; MSI, microsatellite instable; ND, not done; TSGs, tumor suppressor genes; VAF, variant allele frequency.

* Only one of the two microsatellite markers produced an interpretable result.
Methods), although not found in the COSMIC database. The full list of the mutations used in this analysis is available in Supplementary Table S3. Based on the total of 310 missense or nonsense mutations, the following observations were made. C → T transitions were clearly the most prevalent substitutions in both MMR-P and MMR-D adenomas. However, there were some differences in base triplets harboring the transitions. Signature 6 is characterized by C → T transitions at NpCpG sites [37], thus, we investigated the frequencies of these mutations among all mutations found in MMR-P and MMR-D adenomas. Indeed, they were more prevalent in MMR-D adenomas (53/277, 19%) vs. 3/33, 9%), although statistical significance was not reached. Intriguingly, the mutation with the highest probability to occur in MMR-D tumors according to the mutation signature, C → T at GpCpG, was missing in MMR-P adenomas (0/16), but was present (12/43, 28%) in MMR-D adenomas (p = 0.025). Moreover, we observed ApCpC, CpCpC, and GpTpa sites to be enriched with mutations in MMR-P adenomas compared to MMR-D adenomas with >5% difference in the relative mutation frequency. Signature 6 is characterized also by small indels, but in our study, only three deletions (1–3 bp) were found, all in MMR-D adenomas.

In addition, we examined pathogenic KRAS mutations (codons 12, 13, 61, and 146) that can cause constitutive activation of KRAS protein [34,35], and all possible mutations based on the flanking nucleotides (Fig. S5C). KRAS G13D mutations (n = 4) occurred exclusively in MMR-D adenomas, in agreement with published findings of this mutation being associated with MSI tumors [38,39]. Codon 12 and 146 mutations were more prevalent in MMR-P adenomas (see Discussion).

5. Discussion

In the majority of colorectal cancers, the adenoma-carcinoma sequence is a well-established path, where adenomatous polyps have a high risk of developing into carcinomas. However, in LS, recent findings such as frequent development of colorectal cancer despite regular colonoscopic surveillance to remove adenomas has challenged the classical view [40]. Furthermore, the tumorigenic stage at which MMR deficiency appears is controversial. Observations of MMR protein loss associated with high-grade dysplastic and larger polyps supports deficient MMR being a late event (this study and [10,13–15,41]). However, the discovery of histologically normal colonic crypts with a loss of MMR protein expression, i.e. MMR-deficient crypt foci (MMR-DCF) [42,43], as well as a recent finding of MMR-deficiency preceding adenoma formation in LS suggest early appearance of MMR deficiency [44]. In our investigation, all LS carcinomas showed a complete loss of the MMR protein expression, whereas 10% of adenomas with high-grade dysplasia and 35% adenomas with low-grade dysplasia maintained the expression, implying that MMR protein loss occurs relatively late. Ahadova et al. [38] proposed a model of three different pathways driving LS-associated tumorigenesis, which could reconcile the contradictory findings of the adenoma-carcinoma sequence and the timing of MMR-deficiency. The model covers late appearance of deficient MMR (MMR-P adenoma with secondary MMR inactivation) as well as early MMR deficiency (MMR-DCF evolving into carcinoma either via MMR-D adenoma or directly without any polyp precursor).

Only limited information is available of molecular alterations present in LS adenomas lacking MMR deficiency as an obvious driver of tumorigenesis (MMR-P adenomas). We addressed DNA methylation changes and somatic mutations of cancer-associated genes as possible cancer-initiating or promoting events. Methylation was studied for eight CIMP markers, seven selected candidate TSGs [23], and LINE-1. The frequency of CIMP(+) and the average degree of LINE-1 hypomethylation increased with the degree of dysplasia. In general, they were rather rare events in adenomas with low-grade dysplasia or in MMR-P adenomas, indicating that wide-spread hypo- and hypermethylation
occurs at later stages of tumorigenesis. Nevertheless, MMR-P adenomas revealed significantly elevated methylation levels (expressed as average Dm values) relative to normal mucosa in four CIMP marker genes: IGF2, CRABP1, NEUROG1, and CDKN2A (Fig. 2a) and the candidate TSGs SFRP1 (Fig. 2b) and SFRP2 (Fig. 2c). Our findings from LS adenomas are supported in the sporadic context by Ibrahim et al. [21]. They recognized hypermethylation of NEUROG1 and SFRP2 as early events in sporadic colorectal carcinogenesis, with NEUROG1 being hypermethylated in hyperplastic polyps and SFRP2 in adenomatous polyps. Moreover, in our previous study comparing LS individuals without vs. with colorectal cancer, SFRP1 hypermethylation occurred as a “field defect” in high-risk normal mucosa from the latter group [9]. The baseline level of LINE-1 methylation may vary according to the clinical category of colorectal cancer [31,45]. Progressive LINE-1 hypomethylation from normal mucosa to adenoma to carcinoma accompanies both LS-associated (Fig. 3a) and sporadic colon tumorigenesis [22].

We used a sensitive technique (amplicon-based next-generation sequencing) to screen the hotspot regions of 22 established cancer-associated genes for somatic mutations in LS adenomas. Variant alleles occurring with 5% or higher frequencies were considered. Mutations associated genes for somatic mutations in LS-adenomas. Variant alleles sequencing) to screen the hotspot regions of 22 established cancer-according to the clinical category of colorectal cancer [31,45]. Progressive LINE-1 hypomethylation from normal mucosa to adenoma to carcinoma accompanies both LS-associated (Fig. 3a) and sporadic colon tumorigenesis [22].

We identified a strong predisposing defect (inherited MMR gene mutation) in every cell but the absence of tumor development until additional hits (to the predisposing MMR gene and other genes) occur in a target tissue, makes LS a valuable model to dissect the multistep tumorigenesis of colon and other organs. Our findings of MMR-P adenomas shows, first, that tumor formation in LS is possible even in the absence of a second hit to inactivate the wild-type allele of the predisposing MMR gene. Second, we identify possible alternative events prevalent in MMR-P adenomas that may serve as early alterations accelerating, or possibly initiating, LS-associated tumorigenesis when MMR protein expression is still intact. Hypermethylation of IGF2, NEUROG1, CDKN2A, CRABP1, SFRP1, and SFRP2, and especially KRAS mutations may represent such events. Third, the emerging epigenetic and genetic differences between MMR-P and MMR-D adenomas imply the existence of different alternative routes to cancer development in LS, consistent with observations of genetic heterogeneity in LS carcinomas as discussed above. Further studies with large sample sizes are necessary to deepen the understanding of tumorigenic
sequences leading to LS cancers, thereby improving early recognition of increased tumor risk and cancer prevention in LS.

Acknowledgements

We thank the patients contributing to this study. We thank Beatriz Alcala-Repo and Kirsi Pylvänäinen for collecting the samples and patient information. Salla Saarinen and Omar Youssef are thanked for laboratory support. The work was funded by the Jane and Aatos Erkko Foundation (MN, J-PM, and PP), Academy of Finland (grant no. 294643) (PP), the Cancer Foundation Finland (J-PM and PP), the Sigrid Juselius Foundation (JL), and HIIFLE (PP). Funding sources had no involvement in the study design, in the collection, analysis, or interpretation of data, in the writing, nor in the decision to submit the paper for publication.

Declaration of interests

During the conduct of the study, Dr. Nyström received a grant from the Jane and Aatos Erkko Foundation, Dr. Melckin received grants from the Jane and Eirkko Foundation and the Cancer Foundation Finland, and Dr. Peltonäki received grants from the Jane and Aatos Erkko Foundation, the Academy of Finland, the Cancer Foundation Finland, the Sigrid Juselius Foundation, and the HIIFLE. Other authors declare no conflicts of interests.

Author contributions

SM-N, SV, and PP designed the study. LR-S, AL, and J-PM provided the patient data and samples. AR performed the histopathological analyses. VS and SK provided sequencing expertise and access to the facilities. MN shared epigenetic expertise and information of epigenetic markers derived from a mouse model of LS. SM-N and SV performed the laboratory experiments, and data analyses. SM-N and PP wrote the manuscript. All authors read and approved the final manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2018.12.018.

References


