







# The earliest events in *BRAF*-mutant colorectal cancer: exome sequencing of sessile serrated lesions with a tiny focus dysplasia or cancer reveals recurring mutations in two distinct progression pathways

Ame GC Bleijenberg<sup>1</sup> , Joep EG Ijspeert<sup>1</sup>, Jos BG Mulder<sup>2†</sup>, Paul Drillenburgh<sup>3</sup>, Herbert V Stel<sup>4</sup>, Elisabeth M Lodder<sup>5</sup> , Beatriz Carvalho<sup>6</sup> , Jade Jansen<sup>2</sup>, Gerrit Meijer<sup>6</sup> , Susanne van Eeden<sup>2</sup>, Evelien Dekker<sup>1</sup>  and Carel JM van Noesel<sup>2\*</sup> 

<sup>1</sup> Amsterdam University Medical Centers, Location AMC, Department of Gastroenterology and Hepatology, Cancer Center Amsterdam, University of Amsterdam, Amsterdam, The Netherlands

<sup>2</sup> Amsterdam University Medical Centers, Location AMC, Department of Pathology, University of Amsterdam, Amsterdam, The Netherlands

<sup>3</sup> Department of Pathology, Onze Lieve Vrouwen Gasthuis (OLVG), Amsterdam, The Netherlands

<sup>4</sup> Department of Pathology, Tergooi Ziekenhuizen, Hilversum, The Netherlands

<sup>5</sup> Amsterdam University Medical Centers, Core Facility Genomics, Department of Clinical Genetics, University of Amsterdam, Amsterdam, The Netherlands

<sup>6</sup> Department of Pathology, Netherlands Cancer Institute, Amsterdam, The Netherlands

\*Correspondence to: CJM van Noesel, Department of Pathology, Amsterdam University Medical Centers, Meibergdreef 9 | 1105 AZ, Amsterdam, The Netherlands. E-mail: [c.j.vannoesel@amsterdamumc.nl](mailto:c.j.vannoesel@amsterdamumc.nl)

†Deceased.

## Abstract

Around 15–30% of colorectal cancers (CRC) develop from sessile serrated lesions (SSLs). After many years of indolent growth, SSLs can develop dysplasia and rapidly progress to CRC through events that are only partially understood. We studied molecular events at the very early stages of progression of SSLs via the MLH1-proficient and deficient pathways to CRC. We collected a cohort of rare SSLs with a small focus (<10 mm) of dysplasia or cancer from the pathology archives of three hospitals. Whole-exome sequencing was performed on DNA from nonprogressed and progressed components of each SSL. Putative somatic driver mutations were identified in known cancer genes that were differentially mutated in the progressed component. All analyses were stratified by MLH1 proficiency. Forty-five lesions with a focus dysplasia or cancer were included, of which 22 (49%) were MLH1-deficient. Lesions had a median diameter of 10 mm (interquartile range [IQR] 8–15), while the progressed component had a median diameter of 3.5 mm (IQR 1.75–4.75). Tumor mutational burden (TMB) was high in MLH1-deficient lesions (23.9 mutations per MB) as compared to MLH1-proficient lesions (6.3 mutations per MB). We identified 34 recurrently mutated genes in MLH1-deficient lesions. Most prominently, *ACVR2A* and *RNF43* were affected in 18/22 lesions, with mutations clustered in three hotspots. Most lesions with *RNF43* mutations had concurrent mutations in *ZNRF3*. In MLH1-proficient lesions *APC* (10/23 lesions) and *TP53* (6/23 lesions) were recurrently mutated. Our results show that the mutational burden is exceptionally high even in the earliest MLH1-deficient lesions. We demonstrate that hotspot mutations in *ACVR2A* and in the *RNF43/ZNRF3* complex are extremely common in the early progression of SSLs along the MLH1-deficient serrated pathway, while *APC* and *TP53* mutations are early events in the the MLH1-proficient pathway.

© 2022 The Authors. *The Journal of Pathology* published by John Wiley & Sons Ltd on behalf of The Pathological Society of Great Britain and Ireland.

**Keywords:** colon; DNA sequencing; immunocytochemistry; colorectal cancer; serrated polyps; serrated neoplasia pathway; microsatellite instability; mismatch repair genes; *BRAF*<sup>V600E</sup>; *BRAF*

Received 2 August 2021; Revised 18 January 2022; Accepted 8 February 2022

*Conflict of interest statement:* ED has endoscopic equipment on loan from FujiFilm, receives a research grant from FujiFilm, has received a honorarium for consultancy from FujiFilm, Olympus, Tillots, GI Supply and CPP-FAP and a speakers' fee from Olympus, Roche and GI Supply. BC has several patents pending, which are not relevant to this study. GM has several patents pending, which are not relevant to this study. No other conflicts of interest were declared.

## Introduction

Traditionally, colorectal cancer (CRC) was considered to arise exclusively from tubular/tubulovillous adenomas. Far less is known about the ‘serrated neoplasia pathway,’ which is now held accountable for 15–30% of all CRCs [1]. CRCs that develop through this alternative pathway predominantly arise from sessile serrated lesions (SSL) [1–4]. Recent evidence suggests that these lesions remain dormant for over 15 years, with little change in size, before developing cytological atypia and dysplasia. After a brief stage of dysplasia, they then rapidly progress into full-blown CRC [4]. The molecular events driving this ‘sudden’ malignant progression are poorly understood.

Thus far, several hallmarks of the serrated neoplasia pathway have been identified [4]. An activating *BRAF*<sup>V600E</sup> mutation is considered the most common initiating mutation [1,2,4]. Aside from activating the mitogen-activated protein kinase (MAPK) pathway, this mutation also seems to mediate genomewide promoter methylation, known as CpG-island methylator phenotype (CIMP) [5,6]. Consequently, tumor suppressor genes (TSGs) are at risk of becoming epigenetically silenced [1,2]. In 33–75% of SSLs with a focus of dysplasia or cancer, the mismatch repair gene *MLH1* is silenced due to methylation of its promoter region, which results in microsatellite instability (MSI). This pathway will from hereon be referred to as the *MLH1*-deficient pathway and likely leads to consensus molecular subtype 1 (CMS 1) colorectal cancer [7]. Although the mechanism for genetic instability along this pathway is clear (i.e. MSI), the specific driver genes and the order by which they become mutated and drive early carcinogenesis are not known.

Alternatively, SSLs can progress into *BRAF*<sup>V600E</sup>/microsatellite stable (MSS) CRC through an *MLH1*-proficient pathway, in which *MLH1* expression is retained. Among others, somatic mutations in *TP53* and epigenetic silencing of *CDKN2A* (encoding p16) have been suggested to play a role in the progression of SSL into *BRAF*<sup>V600E</sup>/MSS cancer [2,8]. The tumors that evolve out of the *MLH1*-proficient serrated pathway have similarity with CMS type 4 colorectal cancer [7].

A deeper understanding of the molecular events involved in the progression could facilitate development of more accurate (fecal) biomarker tests, improving early diagnosis of advanced SSLs. Existing studies have mainly focused on the full-blown serrated pathway cancers, making it difficult to discern which mutations are early and which are late events. In order to explore the earliest steps of the serrated neoplasia pathway we collected a unique cohort of SSLs that were caught in the act of malignant progression, as recognized by the presence of a very small focus of dysplasia or cancer. We compared somatic mutations in matched progressed and nonprogressed components of these SSLs on the verge of full-blown colorectal cancer.

## Materials and methods

### Study design and case selection

We retrieved formalin-fixed paraffin-embedded (FFPE) SSLs with a focus of dysplasia or cancer from the pathology units of three hospitals in the Netherlands (Amsterdam UMC, Amsterdam; Onze Lieve Vrouwe Gasthuis [OLVG], Amsterdam; Tergooi Hospital, Hilversum). Samples were reviewed by two expert pathologists (CvN and SvE). To ensure that only SSLs with an early focus of dysplasia or cancer were included, we only included lesions with (1) the presence of non-dysplastic SSL tissue with an uninterrupted transition to an area with dysplasia and/or cancer (Figure 1A), and a maximum diameter of the dysplastic/cancerous focus of ≤10 mm. We compared molecular alterations in the progressed versus the nonprogressed components within these lesions, as described previously [9,10].

Because of the observational nature of this study and because only somatic mutations were assessed, the Institutional Review Board of the Amsterdam University Medical Centers (Amsterdam UMC, previously Academic Medical Center [AMC]) decided that this study fell beyond the legislation regarding Medical Research Involving Human Subjects Act (WMO; Wet Medisch Wetenschappelijk Onderzoek), and that formal ethical approval was therefore not required. This study was performed in agreement with the Declaration of Helsinki [11].

### Clinicopathologic data collection

Patient and lesion characteristics were retrieved from pathology and patient reports. Data were anonymized and were stored online using Castor EDC [12]. During revision, a diagnosis of SSL and progression to dysplasia/carcinoma was based on the 2010 edition of the World Health Organization Classification of Tumors of the digestive system [13,14].

### Data analysis and outcome parameters

We stratified all included lesions by *MLH1*-proficiency, as determined by immunohistochemistry (Figure 1A).

### Immunohistochemistry

All selected tissue specimens were stained for *MLH1* (1:50; BD Pharmingen, San Diego, CA, USA), p16 (1:400; ImmunoLogic, Duiven, The Netherlands), *SMAD4* (1:200; Santa Cruz Biotechnology, Santa Cruz, CA, USA), and p53 (1:2000; Neomarkers, Fremont, CA, USA), as described previously (Figure 1A) [15]. *MLH1* deficiency was defined as a complete absence of nuclear *MLH1* staining. The complete absence of *SMAD4* or p16 staining was considered indicative of dysfunction of the TGF- $\beta$  pathway or p16, respectively. Dysfunction of p53 was defined as either strongly positive or absent p53 staining in ≥75% of the epithelial nuclei.



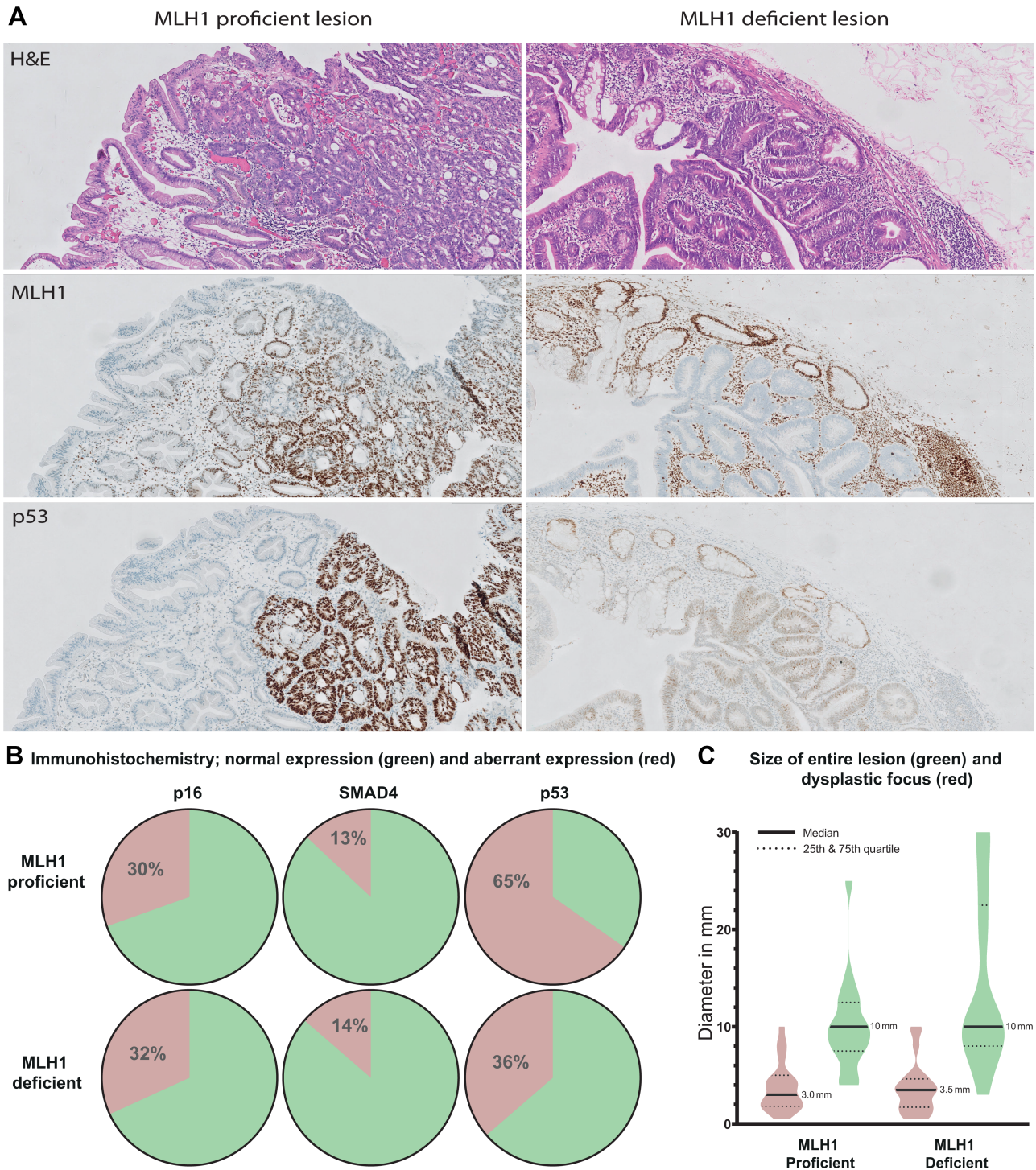


Figure 1. (A) Example of MLH1-proficient lesion (left) and MLH1-deficient lesion (right). Top to bottom: Hematoxylin/eosin-stained sections of transition zone of serrated and progressed components; MLH1 staining normal (left) and absent (right); p53 staining aberrant overexpression (left) and normal expression (right). (B) The proportion of lesions with aberrant expression of p16, SMAD4, and p53. (C) Median size in millimeters of entire lesion and dysplastic focus.

**Microdissection and DNA isolation**

DNA was obtained separately from three components of each lesion: (1) normal colonic mucosa; (2) nondysplastic serrated tissue (hereon referred to as the ‘serrated component’); and (3) tissue from the focus of dysplasia or cancer (hereon referred to as the ‘progressed component’) (Figure 1A). Cells of each component were

isolated using manual microdissection, and DNA was extracted by proteinase K digestion (Hoffmann-La Roche, Basel, Switzerland; 3 µl proteinase K [1 mg/ml] in 27 µl proteinase K buffer with an overnight incubation at 50 °C). After incubation, proteinase K was inactivated at 95 °C for 3 min, and samples were then stored at -20 °C until further use. The fraction of lesional cells

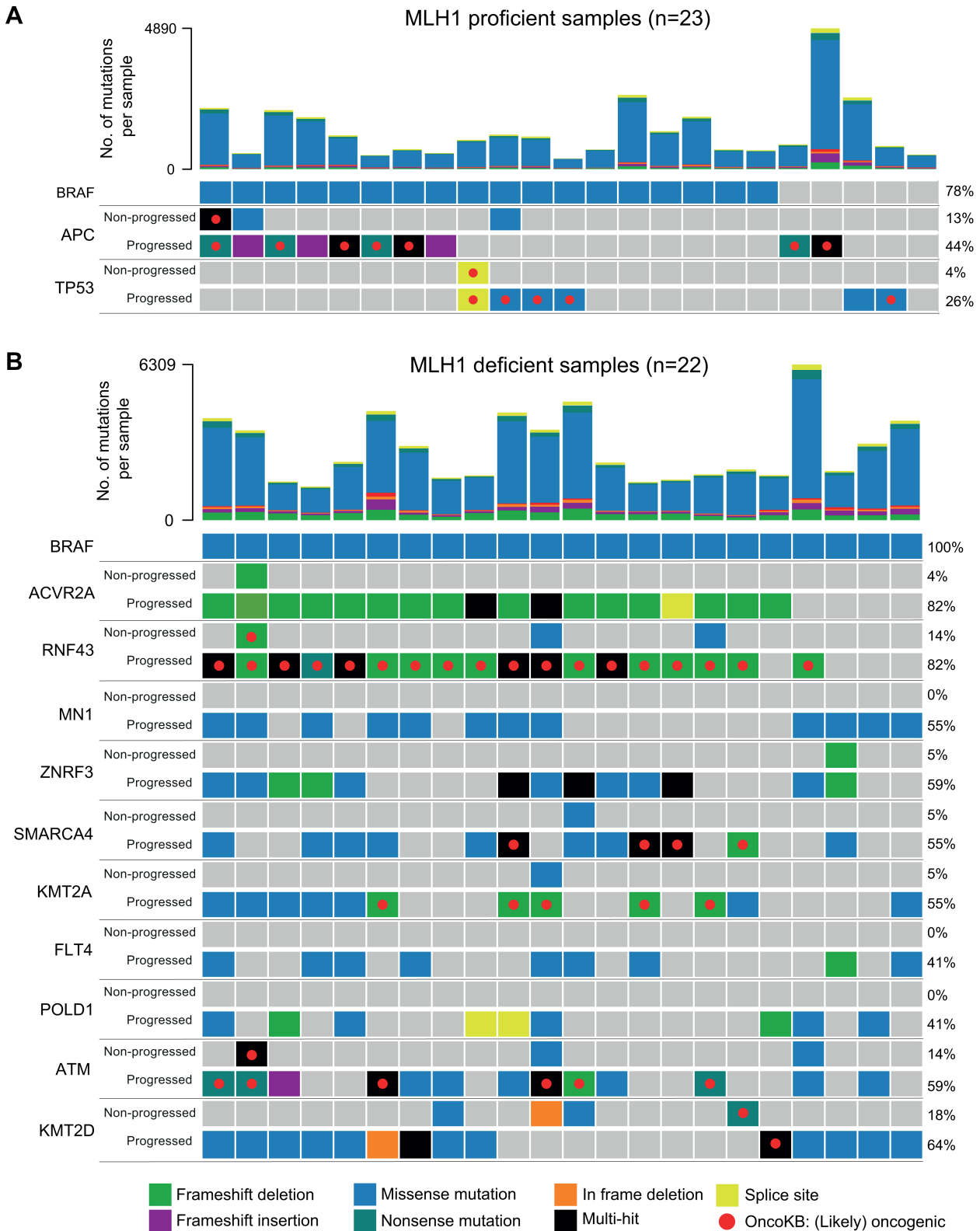


Figure 2. Oncoplots of differentially altered genes ( $p < 0.1$ ) in progressed versus the nonprogressed component. (A) MLH1-proficient lesions; (B) MLH1-deficient lesions. For each gene, the top row displays mutations in the serrated component ('S'), while the bottom row displays mutations in the progressed component ('P') of each lesion. Each column corresponds to one lesion.

was estimated during microdissection, and finally more exactly assessed based on the *BRAF* variant allele frequency (VAF), i.e. the percentage of sequence reads containing a *BRAF* mutation.

Microsatellite instability

Microsatellite (in)stability was assessed in the progressed components of each lesion by MSI analysis system v1.2 (Promega, Madison, WI, USA) according to



the manufacturer's instructions. This system uses five microsatellite markers (NR-21, NR-24, MONO-27, BAT25, and BAT26). Lesions were classified as microsatellite instable (MSI-high) if  $\geq 2$  markers were instable.

#### Whole-exome sequencing: sequencing protocol

The quality of the DNA samples was assessed using a Fragment Analyzer (Agilent, Santa Clara, CA, USA) and expressed as the Genomic Quality Number (GQN) score. Sample preparation and hybridization capture were performed on the SureSelectXT Target Enrichment System for Illumina Paired-End Sequencing Library, protocol v1.8 (G7530-90000), using the Bravo Liquid Handling system (Agilent Technologies, Santa Clara, CA, USA). The 67.3 megabase Agilent SureSelectXT Clinical Research Exome v2 capture library (5190-9493) was used (Agilent Technologies) [16]. Clustering and DNA sequencing using the NovaSeq6000 (Illumina, San Diego, CA, USA) was performed as a commercial service by GenomeScan (Leiden, The Netherlands) according to the manufacturer's protocols. NovaSeq control software NCS v1.6 was used. A median sequence depth of  $\geq 100$  and a horizontal coverage of  $\geq 98\%$  were considered adequate for the current study. Image analysis, base calling, and quality check were performed with the Illumina data analysis pipeline Real Time Analysis (RTA) v3.4.4 (Illumina) and Bcl2fastq v2.20 (Illumina).

#### Whole-exome sequencing (WES): analysis pipeline

Obtained sequences were deduplicated using Picard ([broadinstitute.github.io/picard/](https://broadinstitute.github.io/picard/)), adapter sequences were removed with cutadapt 2.9 ([cutadapt.readthedocs.io/en/stable/](https://cutadapt.readthedocs.io/en/stable/)) and were aligned to the human reference genome hg19 with GATK4 ([gatk.broadinstitute.org/hc/en-us](https://gatk.broadinstitute.org/hc/en-us)) with subsequent variants calling in the same pipeline. We applied left-normalization with bcftools ([github.com/samtools/bcftools](https://github.com/samtools/bcftools)) to the Insertions and Deletions before annotating all variants using Annovar ([annovar.openbioinformatics.org/](https://annovar.openbioinformatics.org/)). Only nonsynonymous exonic variants (NSVs) and variants at essential splice sites were considered. Hence, all variants that could theoretically result in altered proteins were included. Samples with more than 5,000 rare (Minor Allele Frequency [MAF]  $< 0.01$ ) variants in any of the three sections were deemed of too low quality and were excluded from downstream analysis. For downstream analysis, we filtered variants based on minor allele frequency (MAF  $< 4e-04$ ) in the following databases: Exome Aggregation Consortium (EXAC, [exac.broadinstitute.org/](https://exac.broadinstitute.org/)), Genome Aggregation Database (gnomAD, [gnomad.broadinstitute.org/](https://gnomad.broadinstitute.org/)), 1000 Genomes Project (1000G; [internationalgenome.org/](https://internationalgenome.org/)). Variants that were present in the germline were removed by excluding all variants that were found in the adjacent normal tissue. In other words, all variants found in normal tissue were subtracted from the variants found in both the nonprogressed and progressed parts of the lesion. Variant annotation was performed on the Lisa computer cluster ([userinfo.surfsara.nl/systems/lisa](https://userinfo.surfsara.nl/systems/lisa)). Variant filtering and

sorting was performed in R v.3.5.1 (R Foundation for Statistical Computing, Vienna, Austria) using packages R. utils ([github.com/HenrikBengtsson/R.utils](https://github.com/HenrikBengtsson/R.utils)), stringr ([github.com/tidyverse/stringr](https://github.com/tidyverse/stringr)), and dplyr ([github.com/tidyverse/dplyr](https://github.com/tidyverse/dplyr)).

#### Identification of potential driver genes

Potential drivers were identified within 724 known cancer genes listed in the Cancer Gene Census (CGC). The CGC contains an up-to-date list of genes that have been causally implicated in cancer development [17]. Potential driver mutations were defined as recurrently mutated CGC genes that were differentially mutated in the progressed versus the nonprogressed component in at least five lesions. Only mutations in genes with a Fisher's Exact  $P$  value  $< 0.1$  were considered potential drivers. To adjust for the fact that certain genes are, due to their length and genetic composition and location, more likely to acquire random passenger mutations than others and are thus more prone to be falsely identified as driver gene, we utilized the Genome Aggregation Database (gnomAD) database [18] to control for this background level of expected random mutations. Variants in the gnomAD database were filtered with the same criteria as the variants in our cohort. Differences from background levels of mutations were statistically assessed using Fisher's exact test for each of the CGC genes. Only genes with a Bonferroni-corrected  $p$  value  $< 0.05$  were considered as potential drivers. We utilized the OncoKB database to annotate variants [19]. Alterations listed as (likely) oncogenic have either been demonstrated or convincingly predicted to result in pathogenic alterations [20].

#### Tumor mutational burden

Tumor mutational burden (TMB) was defined as the number of mutations per sequenced megabase (Mb) and was compared between MLH1-proficient and -deficient lesions. We assessed whether the size of the progressed component or the grade of dysplasia were related with TMB. TMB was also compared with other publicly available cancer cohorts using data from the Cancer Genome Atlas (TCGA) [20].

#### Analysis software and R packages

Baseline polyp and patient characteristics were compared using the Statistical Package for Social Sciences (SPSS) v. 24 (IBM, Somers, NY, USA). All other analyses were performed using RStudio 1.2.1335 (RStudio, PBC, Boston, MA, USA), with R version 3.6.1.

## Results

### Baseline characteristics

A total of 65 SSLs with a small focus ( $\leq 10$  mm) of dysplasia or carcinoma were included for whole-exome analyses. After exome sequencing, 20 lesions had to be

excluded due to low sequencing quality, as evidenced by an exceptionally high number (i.e. >5,000) of rare variants in the normal, serrated, and/or progressed component of the sample, mostly due to fixation artifacts. After exclusion of these lesions, 45 SSLs from 40 individual patients were included. The SSLs had a focus of LGD in 27/45 (60%), HGD in 10/45 (22.2%), and carcinoma in 8/45 (17.8%), and were predominantly detected in females (77.8%) and approximately half (48.7%) occurred in a setting of serrated polyposis syndrome (Table 1). The majority of polyps (87.2%) were located proximal to the descending colon. Lesions had a median diameter of 10 mm (interquartile range [IQR] 8–15), while the progressed component had a median diameter of 3.5 mm (1.75–4.75 mm). Based on immunohistochemistry, 23 (51%) were classified as MLH1-proficient and 22 (49%) as MLH1-deficient. Based on methylation analysis, 20 of the 22 MLH1-deficient lesions showed methylation of the MLH1 promoter. The remaining two lesions did not show any somatic *MLH1* mutation.

MSI could be assessed for 41/45 samples (91.1%), after exclusion of four samples with limited DNA available. All 22 lesions from the group of MLH1-deficient lesions were MSI-high, compared to none of the MLH1-proficient lesions.

### Whole-exome sequencing (WES) quality metrics and variant calling

A total of 135 DNA samples (45 SSLs × 3 components) were included in our WES analyses. The median Genomic Quality Number (GQN) score of the FFPE DNA input was 1.0 (IQR 0.7–1.4). The median sequence

depth was 151 (IQR 108–182), with a median 100×, 50×, and 30× coverage of 68.7% (IQR 53.1–76.2%), 88.3% (IQR 80.0–91.9%), and 94.7% (IQR 90.6–96.5%), respectively. The median horizontal coverage was 99.8% (IQR 99.7–99.8%).

After applying our filtering pipeline, we found 54,716 variants within the progressed components, of which 3,506 (6.4%) were located within Cancer Gene Census (CGC) genes [17]. The median number of mutations in the CGC genes per sample was 72 (IQR 29.5–112.5). A total of 45,215 single-nucleotide polymorphisms (SNPs) were found; 11,639 C>A, 3,698 C>G, 19,282 C>T, 5,223 T>C, 2,722 T>A, and 2,651 T>G (supplementary material, Figure S1).

A *BRAF* mutation was identified in 40 (88.9%) of the serrated components and in 40 (88.9%) of the progressed components. All but two detected *BRAF* variants were *BRAF*<sup>V600E</sup>. The two alternative variants were V600G and V600M, both detected in the progressed component of a single lesion in which a V600E was detected in the nonprogressed component. The median percentage of sequence reads with a *BRAF* mutation was 27.8% (IQR 22.4–37.4%), corresponding to a median estimated lesional DNA percentage of 55.6% (IQR 44.8–74.8%).

### MLH1-proficient versus MLH1-deficient lesions

#### Clinicopathologic differences

Comparing the 23 MLH1-proficient with the 22 MLH1-deficient lesions, we found no differences in lesional size (10 mm versus 10 mm,  $p = 0.35$ ), size of the dysplastic component (3 mm versus 3.5 mm,  $p = 0.78$ ), and grade of dysplasia ( $p = 0.30$ ), suggesting

Table 1. Baseline patient and polyp characteristics.

	All samples (n = 45)	MLH1-proficient (n = 23)	MLH1-deficient (n = 22)	P value
Age at diagnosis, median (IQR)	66 (63–70.8)	64 (55–68.3)	68.5 (63.8–72.3)	0.024*
Female gender, n (%)	35/45 (77.8%)	15/23 (65.2%)	20/22 (90.9%)	0.038†
Number of polyps resected (lifetime)				
Adenomas, median (IQR)	2 (1–3)	2 (1–3.5)	2 (1–3.5)	0.95‡
SPs, median (IQR)	6 (2–11)	3 (1.75–6.25)	6 (6–13.5)	0.009‡
Missing	6	5	1	
Fulfills diagnostic criteria SPS, n (%)	19/39 (48.7%)	4/18 (22.2%)	15/21 (68.2%)	0.002†
Missing	6	5	1	
Location				
Proximal to descending colon, n (%)	34/39 (87.2%)	11/18 (61.1%)	19/21 (90.5%)	
Distal to or in descending colon, n (%)	5/39 (12.8%)	7/18 (38.9%)	2/21 (9.5%)	0.030†
Missing	6	5	1	
Diameter				
Polyp, median (IQR)	10 mm (8–15)	10 mm (7.75–13.0)	10 mm (8–22.5)	0.35‡
Progressed component, median (IQR)	3.5 mm (1.75–4.75)	3.0 mm (1.75–3.0)	3.5 mm (1.7–4.6)	0.78‡
Grade of dysplasia/cancer				
LGD	27/45 (60%)	15/23 (65.2%)	12/22 (54.5%)	
HGD	10/45 (22.2%)	3/23 (13.0%)	7/22 (31.8%)	
Cancer	8/45 (17.8%)	5/23 (21.7%)	3/22 (13.6%)	0.30†
Microsatellite instability (MSI-high)				
Microsatellite instable	22/41 (53.7%)	0/19 (0%)	22/22 (100%)	<0.001†
Missing (insufficient DNA available)	4	4	0	

\*Independent samples t-test.

†Chi-squared test.

‡Mann-Whitney U test.

Significant p values ( $p < 0.05$ ) are indicated in bold font.

comparable groups with regard to stage of disease (Table 1, Figure 1C). Compared to MLH1-proficient lesions, MLH1-deficient lesions occurred more often in women (90.9% versus 65.2%,  $p = 0.038$ ), more often in patients that fulfilled the criteria of SPS (68.2% versus 22.2%,  $p = 0.002$ ), and more often in the proximal colon (90.5% versus 61.1%,  $p = 0.03$ ).

#### Immunohistochemistry

The absence of SMAD4 expression was found in 3/23 (13%) versus 3/22 (13.6%,  $p = 0.65$ ) of MLH1-proficient and MLH1-deficient lesions, respectively. Loss of p16 was found in 7/23 (30.4%) versus 7/22 (31.8%,  $p = 1.00$ ) of MLH1-proficient and MLH1-deficient lesions, respectively. Aberrant p53 staining was observed in 15/23 (65%) versus 8/22 (36.4%,  $p = 0.076$ ) of MLH1-proficient and MLH1-deficient lesions, respectively (Figure 1B).

#### Recurrently mutated genes

In MLH1-proficient lesions, we identified two genes that were differentially mutated in the progressed component: *APC* and *TP53* (Table 2, Figure 2, supplementary material, Table S1). Seven of 10 (70%) *APC* mutations and 5/6 (83%) *TP53* mutations were classified as (likely) oncogenic by OncoKB [19,20]. Four lesions harbored previously described hotspot *TP53* mutations (p.C275F, rs863224451; p.R248Q, rs11540652, and p.A138V, rs750600586) [19,21,22]. All mutations and their annotation are included in supplementary material, Table S2. Of interest, four of the five *BRAF*<sup>WT</sup> lesions harbored either a *TP53* or *APC* mutation. Manual

inspection revealed that the progressed component of two of these lesions had well-known pathogenic missense mutations in *KRAS* (p.G12D and p.G13D, respectively).

In MLH1-deficient lesions 34 recurrently mutated genes were identified (Table 2, Figure 2 and supplementary material, Table S1). Most prominently, mutations in *RNF43* and *ACVR2A* each occurred in 18/22 lesions. All the *RNF43* mutations were classified as (likely) oncogenic. Twelve of 18 lesions (66%) with *RNF43* mutations had concurrent *ZNRF3* mutations. The other top-10 recurrently mutated genes were *MN1* (12/22), *ZNRF3* (13/22), *SMARCA4* (12/22), *KMT2A* (12/22), *FLT4* (9/22), *POLD1* (9/22), *ATM* (13/22), and *KMT2D* (14/22). Mutations in *POLD1* were evenly spread throughout the gene, with 2/9 mutations occurring within the exonuclease domain (supplementary material, Table S2).

We identified several recurrently affected microsatellites in some of these genes: 15/18 lesions had a p.G659fs deletion in a 7xG homopolymer of *RNF43*, four lesions had a p.R117fs deletion in a 6xG homopolymer of *RNF43*, 17/18 lesions had a p.K437fs deletion in a 8xA homopolymer of *ACVR2A*, and three lesions had a p.P773fs deletion in a 7xG homopolymer of *KMT2A*. All variants in differentially altered genes are listed in supplementary material, Table S2.

#### Tumor mutational burden

The median TMB was substantially higher for MLH1-deficient lesions (23.9 per Mb) than for MLH1-proficient lesions (6.27 per Mb,  $p < 0.001$ ; Figure 3A). This corresponded with a higher number of mutations across all

Table 2. Top 10 differentially mutated cancer genes in progressed component compared to nonprogressed component of all samples stratified by MLH1 proficiency. Differentially altered genes were identified as follows: proportion of mutations within progressed versus nonprogressed component, within all CGC genes that are mutated in  $\geq 5$  genes.

MLH1-proficient lesions Two genes with $p$ value $< 0.10$				
HUGO symbol	Serrated component	Progressed component	$P$ value	Adjusted $p$ value progressed component versus gnomAD*
APC	3/23 (13.0%)	10/23 (43.5%)	0.047	$4.87 \times 10^{-9}$
TP53	1/23 (4.45%)	6/23 (26.1%)	0.095	$2.34 \times 10^{-8}$
MLH1-deficient lesions 34 genes with $p$ value $< 0.10$ ; only top-10 displayed. See supplementary material, Table S1 for complete list.				
HUGO symbol	Serrated component	Progressed component	$P$ value	Adjusted $p$ value progressed component versus gnomAD*
ACVR2A	1/22 (4.5%)	18/22 (81.8%)	$< 0.001$	$7.52 \times 10^{-47}$
RNF43	3/22 (13.6%)	18/22 (81.8%)	$< 0.001$	$4.96 \times 10^{-35}$
MN1	0/22 (0%)	12/22 (54.5%)	$< 0.001$	$8.21 \times 10^{-19}$
ZNRF3	1/22 (4.5%)	13/22 (59.1%)	$< 0.001$	$5.45 \times 10^{-21}$
SMARCA4	1/22 (4.5%)	12/22 (54.5%)	$< 0.001$	$1.12 \times 10^{-18}$
KMT2A	1/22 (4.5%)	12/22 (54.5%)	$< 0.001$	$1.11 \times 10^{-15}$
FLT4	0/22 (0%)	9/22 (40.9%)	0.001	$1.02 \times 10^{-10}$
POLD1	0/22 (0%)	9/22 (40.9%)	0.001	$2.65 \times 10^{-10}$
ATM	3/22 (13.6%)	13/22 (59.1%)	0.004	$1.32 \times 10^{-14}$
KMT2D	4/22 (18.1%)	14/22 (63.6%)	0.005	$9.62 \times 10^{-13}$

\*To adjust for the fact that certain genes are, due to their length and genetic composition and location, more likely to acquire random passenger mutations than others and are thus more prone to be falsely identified as a driver gene. We utilized the Genome Aggregation Database (gnomAD) database to control for this background level of expected random mutations.



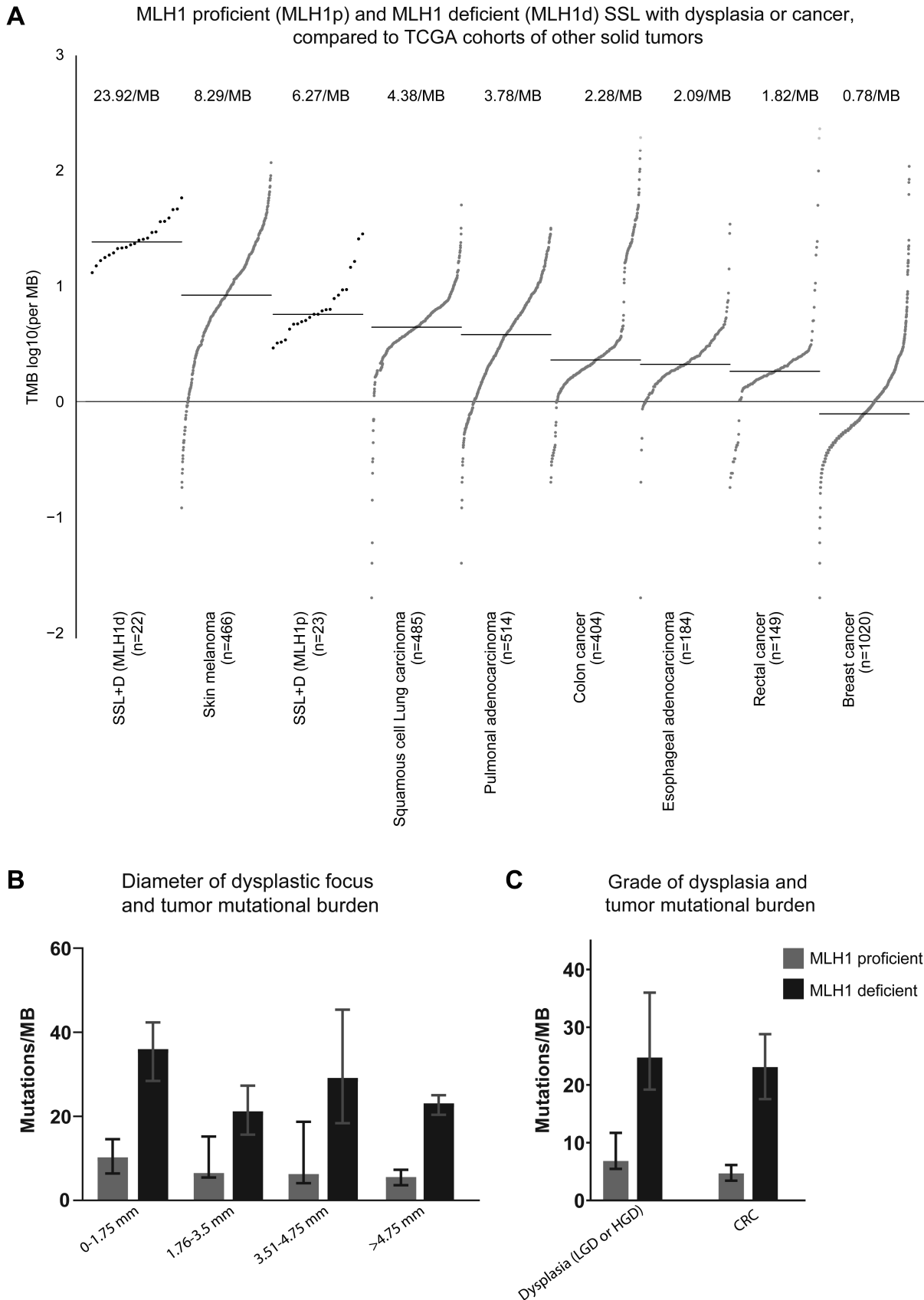


Figure 3. Tumor mutational burden. (A) MLH1-proficient (MLH1p) and MLH1-deficient (MLH1d) SSL with dysplasia or cancer, compared to TCGA cohorts of other solid tumors. (B) Stratified by diameter of progressed component. (C) Stratified by grade of dysplasia (LGD/HGD versus CRC). MLH1p, MLH1-proficient; MLH1d, MLH1-deficient; TCGA, The Cancer Genome Atlas; TMB, tumor mutational burden; SSL+D, sessile serrated lesion with dysplasia; LGD, low-grade dysplasia; HGD, high-grade dysplasia; CRC, colorectal cancer.

mutation types (i.e. insertions, deletions, missense, nonsense, and splice site mutations; supplementary material, Table S3). In our samples, the TMB value was not associated with grade of dysplasia (LGD, HGD, or CRC) or size of the progressed component (Figure 3B,C).

## Discussion

In this study we analyzed the clinical, pathological, and molecular characteristics of a cohort 45 SSLs with a small focus of dysplasia or early cancer. Clinicopathologic characteristics suggested that SSLs occurring in the distal colon, in male patients and in the absence of comorbid serrated polyposis syndrome, had the highest chance of developing via the MLH1-proficient pathway to dysplasia or early cancer. Our whole-exome analyses revealed several recurrently mutated genes, which differed substantially between the MLH1-proficient and -deficient pathways. P53 was found to be recurrently affected in MLH1-proficient lesions, based on both immunohistochemical as well as mutation analyses. In addition, we found *APC* to be mutated in 44% of MLH1-proficient SSLs, which suggests that WNT-pathway activation occurs via *APC* mutations in the MLH1-proficient pathway. Interestingly, neither *TP53* nor *APC* were found to be recurrently mutated in MLH1-deficient lesions.

MLH1-deficient lesions were without exception microsatellite instable and had a much higher number of mutations (median TMB 23.9 versus 6.3, Figure 3A), including mutations in cancer-related genes ( $n = 34$ , Table 2 and Figure 2). This indicates that MLH1 loss causes a surprisingly rapid accumulation of mutations. The degree of genetic instability in these tiny lesions is impressive and very similar to that of fully fledged MSI colorectal cancer (Figure 3A). The genetic instability most notably manifested as recurring mutations in *RNF43* and *ACVR2A* in almost all MLH1-deficient lesions. These mutations were clustered in two hotspots in *RNF43* (p.G659fs and p.R117fs) and one hotspot in *ACVR2A* (p.K437fs), all of which were located in homopolymeric microsatellites. *RNF43* and *ACVR2A* hotspot mutations have been described in microsatellite instable tumors previously [23,24]. More recent studies demonstrated that frameshift *ACVR2A* mutations indeed invoke loss of *ACVR2A* expression [25], which is linked to poor prognosis and metastatic potential of CRC [26]. Interestingly, it was claimed that *ZNRF3* mutations in CRC most frequently co-occur with *RNF43* mutations. This co-occurrence was also clearly seen in our results (Figure 2). *RNF43* and *ZNRF3* act together in a complex as critical negative feedback regulators of WNT-signaling [27–29]. Tumors with mutations in the *ZNRF3/RNF43* complex are, due to decreased WNT-receptor degradation, hypersensitive to WNT-signaling [29]. Although mutations in either gene can cause WNT-signaling activation, combined

knockout of both enzymes led to rapid formation of intestinal adenomas in a murine model [28]. Mutations in these genes might have clinical implications: an ongoing clinical trial is studying the antitumor activity of WNT inhibitor WNT974 in patients with *BRAF<sup>mut</sup>/RNF43<sup>mut</sup>* CRCs ([clinicaltrials.gov/ct2/show/NCT02278133](https://clinicaltrials.gov/ct2/show/NCT02278133), NCT02278133, date last accessed 2 February 2022), which most likely progressed through the MLH1-deficient serrated pathway.

It is unclear whether the mutations we describe in the MSI lesions represent actual driver mutations. They could also represent random passenger mutations. The frequency of *RNF43* and *ACVR2A* mutations, as well as the co-occurrence of *RNF43* and *ZNRF3* mutations, do suggest they result in a survival benefit for derailed SSLs. This is supported by the OncoKB database, which includes the *RNF43* hotspot mutations as being likely oncogenic. Also, both the *RNF43* G659 and R117 frameshift mutations occur mutually exclusive with *APC* mutations in CRC, suggesting an *APC<sup>mut</sup>*-like effect [30]. There is also evidence that *ACVR2A* mutations in MSI-high CRC are oncogenic [25,26]. However, MSI tumors acquire large numbers of mutations as they progress, including many passenger mutations. Indeed, Tu and colleagues suggest the *RNF43* G659fs mutation to be a passenger mutation instead of a driver, resulting in functional RNF43 protein [31]. Although our study provides *candidate* driver genes, the recurring mutations we found could certainly represent passenger mutations as well. Their actual biologic effect should therefore be the topic of future experimental studies, for example, using immunohistochemistry, gene expression analysis, or the *CRISPR-cas9* organoid model described by Lannagan *et al* [32,33].

Our results have to be interpreted with caution. First, our cohort was relatively small, and not powered to identify subtle differences between MLH1-proficient and -deficient lesions. To reduce the risk of false-discovery and increase confidence, we confined our analyses to a limited number of known cancer genes. In addition, massively parallel sequencing of FFPE material might lead to an increased number of sequence artifacts due to, e.g., formaldehyde-induced crosslinks and DNA fragmentation [34]. Although we cannot rule out sequence artifacts, the quality (i.e. GQN score) of our FFPE-derived DNA was relatively high. Moreover, WES analyses of matched fresh-frozen and FFPE-derived tumor DNA showed high concordance in multiple previous studies [35–37]. Third, two MLH1-deficient lesions did not show MLH1 promoter methylation or MLH1 mutation. Although no known Lynch syndrome patients were included, these two patients could theoretically harbor germline MLH1 mutations. Finally, although many of our findings showed agreement with previously published literature, several novel findings warrant further validation. Especially the effect of the recurring mutations on cancer signaling pathways should be assessed in experimental settings.

In conclusion, our study provides novel insights in the very first events that are involved in the malignant

progression of SSLs, the precursor of *BRAF*-mutant CRC. We show that even in the earliest stages of dysplasia, these lesions rapidly acquire extensive microsatellite instability, and an exceptionally high mutational burden following MLH1 loss. These tiny lesions, measuring no more than 3.5 mm, were already fully fledged MSI lesions. We found recurring hotspot mutations in *ACVR2A* and the *RNF43/ZNRF3*-complex in MLH1-deficient lesions, while *APC* and *TP53* mutations were frequently observed as early events in MLH1-proficient lesions.

## Acknowledgements

We thank Roy Reinten, Lisette Hoogendijk, Mireille de Wit, Jitske Grundell-Weegenaar, and Alex Musler for their technical support in tissue processing and molecular analyses. We thank Stephanie van den Oever for her support in performing the whole-exome sequencing. This work was funded by grants from the Dutch Cancer Society (KWF, grant number 19221) and the Sascha Swarttouw-Hijmans Foundation

## Author contributions statement

AGCB, JEGIJ, ED and CJMN designed this study. CJMN supervised this study. AGCB, JEGIJ and JJ carried out the experiments. SVE and CJMN performed expert pathologic revision. AGCB, JEGIJ, PD, HVS, JJ and CJMN collected all required samples and clinical data. AGCB and EML carried out all statistical analyses. AGCB, JEGIJ, EML, BC, GM, ED and CJMN were responsible for data interpretation. AGCB performed all literature searches. AGCB designed all figures and tables. Writing of the manuscript was done by AGCB and CJMN; critical revision of the manuscript was done by all other authors. Funding acquisition was done by JEGIJ, BC, GM, ED and CJMN.

## References

- Bettington M, Walker N, Clouston A, *et al*. The serrated pathway to colorectal carcinoma: current concepts and challenges. *Histopathology* 2013; **62**: 367–386.
- Bettington M, Walker N, Rosty C, *et al*. Clinicopathological and molecular features of sessile serrated adenomas with dysplasia or carcinoma. *Gut* 2017; **66**: 97–106.
- Ijspeert JEG, Vermeulen L, Meijer GA, *et al*. Serrated neoplasia-role in colorectal carcinogenesis and clinical implications. *Nat Rev Gastroenterol Hepatol* 2015; **12**: 401–409.
- Crockett SD, Nagtegaal ID. Terminology, molecular features, epidemiology, and management of serrated colorectal neoplasia. *Gastroenterology* 2019; **157**: 949–966.e4.
- Fang M, Ou J, Hutchinson L, *et al*. The BRAF oncoprotein functions through the transcriptional repressor MAFG to mediate the CpG Island Methylator phenotype. *Mol Cell* 2014; **55**: 904–915.
- Bond CE, Liu C, Kawamata F, *et al*. Oncogenic BRAF mutation induces DNA methylation changes in a murine model for human serrated colorectal neoplasia. *Epigenetics* 2018; **13**: 40–48.
- Guinney J, Dienstmann R, Wang X, *et al*. The consensus molecular subtypes of colorectal cancer. *Nat Med* 2015; **21**: 1350–1356.
- Bond CE, Umapathy A, Ramsnes I, *et al*. p53 mutation is common in microsatellite stable, BRAF mutant colorectal cancers. *Int J Cancer* 2012; **130**: 1567–1576.
- Hermesen M, Postma C, Baak J, *et al*. Colorectal adenoma to carcinoma progression follows multiple pathways of chromosomal instability. *Gastroenterology* 2002; **123**: 1109–1119.
- Carvalho B, Postma C, Mongera S, *et al*. Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression. *Gut* 2009; **58**: 79–89.
- World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 2013; **310**: 2191–2194.
- Castor EDC Castor Electronic Data Capture; 2019. [Accessed 1 October 2021]. Available from: <https://castoredc.com>
- Snover DC, Ahnen DJ, Burt RW. Serrated polyps of the colon and rectum and serrated polyposis syndrome. In *WHO Classification of Tumours of the Digestive System*, Bosman FT, Carneiro F, Hruban RH, *et al*. (eds). IARC: Lyon, 2010; 160–165.
- Rex DK, Ahnen DJ, Baron JA, *et al*. Serrated lesions of the colorectum: review and recommendations from an expert panel. *Am J Gastroenterol* 2012; **107**: 1315–1329; quiz 1314, 1330.
- Boparai KS, Dekker E, Polak MM, *et al*. A serrated colorectal cancer pathway predominates over the classic WNT pathway in patients with hyperplastic polyposis syndrome. *Am J Pathol* 2011; **178**: 2700–2707.
- Hsu F, Kent WJ, Clawson H, *et al*. The UCSC known genes. *Bioinformatics* 2006; **22**: 1036–1046.
- Sondka Z, Bamford S, Cole CG, *et al*. The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer* 2018; **18**: 696–705.
- Karczewski KJ, Francioli LC, Tiao G, *et al*. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020; **581**: 434–443.
- Chakravarty D, Gao J, Phillips SM, *et al*. OncoKB: a precision oncology knowledge base. *JCO Precis Oncol* 2017; **2017**: PO.17.00011.
- Mayakonda A, Lin DC, Assenov Y, *et al*. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res* 2018; **28**: 1747–1756.
- Chang MT, Asthana S, Gao SP, *et al*. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat Biotechnol* 2016; **34**: 155–163.
- Chang MT, Bhattarai TS, Schram AM, *et al*. Accelerating discovery of functional mutant alleles in cancer. *Cancer Discov* 2018; **8**: 174–183.
- Maruvka YE, Mouw KW, Karlic R, *et al*. Analysis of somatic microsatellite indels identifies driver events in human tumors. *Nat Biotechnol* 2017; **35**: 951–959.
- Bond CE, McKeone DM, Kalimutho M, *et al*. RNF43 and ZNRF3 are commonly altered in serrated pathway colorectal tumorigenesis. *Oncotarget* 2016; **7**: 70589–70600.
- Jung B, Doctolero RT, Tajima A, *et al*. Loss of activin receptor type 2 protein expression in microsatellite unstable colon cancers. *Gastroenterology* 2004; **126**: 654–659.
- Zhuo C, Hu D, Li J, *et al*. Downregulation of activin A receptor type 2A is associated with metastatic potential and poor prognosis of colon cancer. *J Cancer* 2018; **9**: 3626–3633.
- Hao HX, Jiang X, Cong F. Control of Wnt receptor turnover by R-spondin-ZNRF3/RNF43 signaling module and its dysregulation in cancer. *Cancers (Basel)* 2016; **8**: 54.
- Koo BK, Spit M, Jordens I, *et al*. Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors. *Nature* 2012; **488**: 665–669.



29. Koo BK, van Es JH, van den Born M, *et al.* Porcupine inhibitor suppresses paracrine Wnt-driven growth of Rnf43;Znrf3-mutant neoplasia. *Proc Natl Acad Sci U S A* 2015; **112**: 7548–7550.
30. Giannakis M, Hodis E, Jasmine Mu X, *et al.* RNF43 is frequently mutated in colorectal and endometrial cancers. *Nat Genet* 2014; **46**: 1264–1266.
31. Tu J, Park S, Yu W, *et al.* The most common RNF43 mutant G659Vfs\*41 is fully functional in inhibiting Wnt signaling and unlikely to play a role in tumorigenesis. *Sci Rep* 2019; **9**: 18557.
32. Lannagan TRM, Lee YK, Wang T, *et al.* Genetic editing of colonic organoids provides a molecularly distinct and orthotopic preclinical model of serrated carcinogenesis. *Gut* 2019; **68**: 684–692.
33. Bleijenberg A, Dekker E. Reverse-engineering the serrated neoplasia pathway using CRISPR-Cas9. *Nat Rev Gastroenterol Hepatol* 2018; **15**: 522–524.
34. Do H, Dobrovic A. Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. *Clin Chem* 2015; **61**: 64–71.
35. Bonfiglio S, Vanni I, Rossella V, *et al.* Performance comparison of two commercial human whole-exome capture systems on formalin-fixed paraffin-embedded lung adenocarcinoma samples. *BMC Cancer* 2016; **16**: 692.
36. Bonnet E, Moutet ML, Baulard C, *et al.* Performance comparison of three DNA extraction kits on human whole-exome data from formalin-fixed paraffin-embedded normal and tumor samples. *PLoS One* 2018; **13**: e0195471.
37. Hedegaard J, Thorsen K, Lund MK, *et al.* Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. *PLoS One* 2014; **9**: e98187.

### SUPPLEMENTARY MATERIAL ONLINE

**Figure S1.** Single basepair substitutions in the progressed component, in all samples and stratified by MLH1 status

**Table S1.** Differentially altered CGC genes in MLH1-proficient and MLH1-deficient lesions

**Table S2.** Variants in differentially altered genes (both MLH1-proficient and MLH1-deficient lesions)

**Table S3.** Tumor mutational burden and number of mutations in Cancer Gene Census genes, overall and stratified by MLH1 proficiency