Mlh1 haploinsufficiency induces microsatellite instability specifically in intestine

Kul S. Shrestha[1,2], Minna M. Tuominen[1], Liisa Kauppi*[1,2]

[1]Systems Oncology (ONCOSYS) Research Program, Research Programs Unit, Faculty of Medicine, University of Helsinki, Helsinki, Finland

[2]Department of Biochemistry and Developmental Biology, Faculty of Medicine, University of Helsinki, Helsinki, Finland

*Corresponding author: Liisa.Kauppi@helsinki.fi

Conflicts of interest: The authors declare no potential conflicts of interest.

Abstract

Tumors of Lynch syndrome (LS) patients display high levels of microsatellite instability (MSI), which results from complete loss of DNA mismatch repair (MMR), in line with Knudson's two-hit hypothesis. Why some organs, in particular those of the gastrointestinal (GI) tract, are especially prone to tumorigenesis in LS remains unknown. We hypothesized that MMR is haploinsufficient in certain tissues, compromising microsatellite stability in a tissue-specific manner before tumorigenesis. Using mouse genetics, we tested how levels of MLH1, a central MMR protein, affect microsatellite stability *in vivo* and whether elevated MSI is detectable prior to loss of MMR function and to neoplastic growth. We assayed MSI by sensitive single-molecule PCR in normal jejunum and spleen of 4- and 12-month old *Mlh1*$^{+/+}$, *Mlh1*$^{+/-}$ and *Mlh1*$^{-/-}$ mice, accompanied by measurements of Mlh1 mRNA and MLH1 protein expression levels.

While spleen MLH1 levels of *Mlh1*$^{+/-}$ mice were, as expected, approximately 50% compared to wildtype mice, MLH1 levels in jejunum varied substantially between individual *Mlh1*$^{+/-}$ mice and decreased with age, due to sporadic and apparently progressive *Mlh1* promoter methylation. MLH1 levels (prior to complete loss of the protein) inversely correlated with MSI severity in *Mlh1*$^{+/-}$ jejunum, while in spleens of the same mice, MLH1 levels and microsatellites remained stable. Thus, *Mlh1* haploinsufficiency affects specifically the intestine where MLH1 levels are particularly labile, inducing MSI in normal cells long before neoplasia. A similar mechanism likely also operates in the human GI epithelium, and could explain the wide range in age of onset of LS tumorigenesis.

Introduction

Lynch syndrome (LS) is an autosomal-dominant cancer syndrome characterized by high risk of developing colorectal cancer (CRC), endometrial cancer and various other cancers. LS accounts for 1-5% of all CRC cases; individuals with LS have >80% lifetime risk of developing CRC, and are diagnosed with CRC at an average age of 44 years. LS patients carry germline heterozygous mutations in one of the DNA mismatch repair (MMR) genes, typically *MLH1* and *MSH2* [1-3]. MMR is required for the repair of single base pair mismatches and small insertion-deletion (indel) mutations arising from strand-slippage DNA replication errors [4, 5]. Microsatellites (short DNA tandem repeats), because of their repetitive nature, are highly prone to such replication errors [4, 6] leading to unrepaired indel mutations in MMR-defective cells, a phenotype known as microsatellite instability (MSI) [4, 7]. High-level MSI is a molecular hallmark of LS patients' tumors and reflects defective MMR [8, 9]. The conventional notion of MSI in LS-associated CRC is that it follows Knudson's two-hit hypothesis [10]: both MMR alleles must be defective in order to trigger MSI. In this model, the first hit is inherited as a germline defect in an MMR gene, and the remaining functional MMR allele is lost (i.e. second hit) by somatic mutations, epigenetic inactivation or loss of heterozygosity (LOH), leading to complete loss of MMR function (MMR deficiency) that initiates the MSI mutator phenotype, subsequently instigating tumorigenesis [11-13].

Certain observations in LS patients, however, suggest that MMR genes may be haploinsufficient, i.e. that loss of function of just one allele may lead to loss of genome integrity. LS patients show low-level of MSI in peripheral blood leukocytes, and in normal colon mucosa [14-16]. Additionally, cell-free extracts from *Mlh1*$^{+/-}$ mouse embryonic fibroblasts show decreased MMR activity [7], decreased MMR levels reduce MMR efficiency in non-neoplastic human cell lines [17, 18], and *Mlh1* hemizygosity increases indel mutations in cancer cell lines [19]. Given these hints of MMR haploinsufficiency, we hypothesized that reduced MMR protein levels, may provoke microsatellite stability in certain tissues, such as the highly proliferating (and tumor-prone) intestinal epithelium in mice, already prior to the "second hit".

The MSI mutator phenotype and MMR deficiency associated tumor spectrum has been extensively characterized in constitutional *Mlh1* knock-out mice (*Mlh1*$^{-/-}$ mice). These animals have a high incidence of MMR-deficient MSI-high GI tumors [7, 20-22]. Here, we used *Mlh1* heterozygous mice (*Mlh1*$^{+/-}$ mice) [7, 22] as model of LS to test tissue-specific MMR haploinsufficiency. As in LS, *Mlh1*$^{+/-}$ mice have early-onset of GI tract tumors, and increased mortality [21]. To assess putative MMR haploinsufficiency, we determined MSI and MLH1 levels in primary cells derived from

different organs, focusing on the comparison between the small intestine (highly proliferative [23] and tumor-prone) and spleen (less proliferating [24] and no MMR-dependent tumors reported). Additionally, we assayed *Mlh1* promoter methylation and tested for loss of heterozygosity (LOH) to obtain mechanistic insights to MLH1 protein expression regulation.

Methods

Mice and genotyping

Mlh1 mice (B6.129- *Mlh1*^{tm1Rak}, strain 01XA2, National Institutes of Health, Mouse Repository, NCI-Frederick) [7] were bred and maintained according to national and institutional guidelines (Animal Experiment Board in Finland and Laboratory Animal Centre of the University of Helsinki). *Mlh1* genotyping was performed using ear pieces. DNA lysate was prepared by incubating the ear pieces overnight in 100 µl of lysis buffer supplemented with 20 µg proteinase K. The next day, lysate was boiled for 10 minutes and spun down for 1 minute at 14,000 rpm at room temperature. 0.5 µl of DNA lysate was used for genotyping. Genotyping PCR was performed using Platinum green hot start PCR master mix (Invitrogen, Carlsbad, CA). Lysis buffer components, PCR program and primers [25] used for genotyping are listed in **Supplementary table 1.** In total 56 mice (39 males and 17 females) were used in this study, the number of mice for each experiment along with the genotype and the age is indicated in the respective figures in the results section.

Tissue collection for DNA, RNA and protein analysis

The small intestine was flushed with cold phosphate-buffered saline (PBS), cut open longitudinally and visually inspected for visible tumors. A 3-cm long piece of jejunum was cut approximately 15 cm from the pyloric sphincter. This tissue piece, henceforth referred to simply as “jejunum”, was approximately the center of the small intestine. Jejunum was inspected for any macroscopic tumor-like outgrowth under a stereoscope, and only normal-looking tissue pieces were used for the experiments. The jejunum was snap-frozen and stored at - 80°C until further use. Other tissues from *Mlh1*$^{+/-}$ mice (brain, kidney, liver and spleen) and *Mlh1*$^{-/-}$ intestinal tumors were also collected, snap-frozen and stored at - 80°C until further use. The frozen tissues were used for subsequent DNA, RNA and protein analysis.

DNA extraction and single-molecule MSI analysis by PCR

DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to manufacturer's instruction. Approximately 5 mg of tissue was used for DNA extraction. The extracted DNA was quantified using a Qubit fluorometer (Thermo Fisher Scientific, Waltham, MA) and diluted to 30 pg/µl concentration in 5 mM Tris-HCl (pH 7.5) supplemented with 5 ng/µl carrier herring sperm (Thermo Fisher Scientific). The MSI assay was performed using single-molecule PCR (SM-PCR) [26-28]. Three microsatellites, two mononucleotide repeats A27 and A33 [29], and a dinucleotide repeat D14Mit15 [7] were assayed for MSI. The number of amplifiable DNA molecules per reaction for each microsatellite was calculated as described previously [30]. SM-PCR programs and primers [7, 29] are listed in **Supplementary table 1**. D14Mit15 and A33 were assayed in the same PCR reaction, and a separate PCR was run for A27. Q5 High-Fidelity DNA Polymerase system (New England Biolabs, Ipswich, MA), supplemented with 1 ng/µl carrier herring sperm, was used for SM-PCR. PCR were performed in a 10 µl reaction volume. Fragment analysis was performed by combining 1µl of each PCR product, with an internal size standard (GeneScan™ 500 LIZ™ dye Size Standard, Applied Biosystems, Waltham, MA) on an ABI3730xl DNA Analyzer (Thermo Fisher Scientific). Between 116 and 205 amplifiable DNA molecules per tissue per mouse were assayed for MSI. Fragman R package was used for visualizing the electropherograms and for mutant scoring [31]. We used stringent mutant scoring criteria, adopted from [27, 32], and thus the mutation rates reported are likely a conservative estimate. Criteria for scoring microsatellite mutations were as follows:

I. A true microsatellite signal should have lower-intensity stutter peaks. Stutter peaks should display the expected size difference (that is, 1 base for mononucleotide repeats, and 2 bases for dinucleotide repeats) from the dominant peak. Reactions with peaks without stutter were considered artifacts.
II. In order for an allele to be considered as mutant, both the highest peak and the stutter peaks should shift as a single unit. Shift of the highest peak alone was not scored as a mutant.
III. If a wildtype and (apparently) mutant allele co-occurred in a single PCR, the reaction was scored as wildtype. Non-wildtype peaks were presumed to result from replication slippage during the early rounds of PCR, and thus considered artifacts.

For each microsatellite, MSI was calculated separately for insertions and deletions (from here onwards termed deletion% and insertion%), as follows:

deletion% = (total no. of DNA molecules with deletion mutation observed / total DNA molecules analyzed) x 100

insertion% = (total no. of DNA molecules with insertion mutation observed / total DNA molecules analyzed) x 100

RNA extraction and RNA expression analysis

RNA was extracted using AllPrep DNA/RNA/Protein Mini Kit (Qiagen) according to manufacturer's instructions. Approximately 5 mg of tissue was used for RNA extraction. Extracted RNA was measured using NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA). 500 ng of extracted RNA was reverse-transcribed using SuperScript VILO cDNA Synthesis Kit (Invitrogen). Quantitative PCR (qPCR) was performed using SsoAdvance Universal SYBR Green Supermix system (Bio-Rad, Hercules, CA). PCR programs and primers used in qPCR can be found in **Supplementary table 1**. Gene expression data for Mlh1 was normalized to beta-actin. Data was analyzed using Bio-Rad CFX Maestro (version 1.1).

Western blot analysis

Frozen tissue pieces were thawed on ice, mechanically homogenized in RIPA buffer supplemented with protease inhibitor cocktail (Roche, Basel, Switzerland), incubated for 30 minutes on ice and centrifuged at 14000 rpm for 10 minutes at 4°C. The supernatant was collected and frozen until further use. Total protein concentration was measured using Pierce™ BCA™ protein assay kit (Thermo Fisher Scientific), 30 µg of total protein extract was used for western blotting. The denatured protein was run in 4–20% Mini-Protean TGX gels (Bio-Rad) and transferred to 0.2 µm nitrocellulose membrane using Trans-Blot Transfer Pack (Bio-Rad). To confirm complete protein transfer, membranes were stained with Ponceau solution for 5 minutes at room temperature. Membranes were blocked with 5% milk in PBS supplemented with 1µl/1ml Tween-20 (Thermo Fisher Scientific) for 1 hour at room temperature, and incubated overnight at 4°C with primary antibodies against MLH1 and Beta-actin, followed the next day by IRDye 800CW and IRDye 680RD secondary antibody incubation for 1 hour at room temperature. Details of antibodies and their dilutions are listed in **Supplementary table 1**. LI-cor Odyssey FC system (LI-COR, Nebraska, USA) was used to scan the membranes and LI-cor Image Studio lite (version 5.2) was used for image analysis. MLH1 protein signal intensity was normalized to Beta-actin signal intensity.

Immunohistochemistry (IHC) and histological image analysis

The small intestine was flushed with cold 1xPBS, fixed overnight in 4% paraformaldehyde, cut open longitudinally, embedded into paraffin blocks as a "Swiss roll" [33], and sectioned at 4µm thickness. Heat induced antigen retrieval was performed for 20 minutes using 10mM citrate buffer (pH 6).

MLH1 was detected using rabbit monoclonal antibody ab92312 (clone EPR3894, Abcam, Cambridge, UK, at 1:1500 dilution) and visualized using BrightVision Poly HRP goat anti-rabbit IgG (cat. no. DPVR55HRP, ImmunoLogic, Duiven, The Netherlands). Sections were counterstained with hematoxylin-eosin. Slides were scanned using 3DHistech Panoramic 250 FLASH II digital slide scanner (3DHistech, Budapest, Hungary) at 20X magnification, and images were visualized using CaseViewer (version 2.2). For each mouse, MLH1 protein levels were quantified for five randomly selected sites in the jejunum, each site consisting of approximately ten villus-crypt units, altogether approximately 50 villus-crypt units per jejunum. Quantification was performed using IHC Profiler plugin in ImageJ [34].

Mlh1 LOH and promoter methylation analysis by PCR

LOH and methylation analysis was performed from the same extracted genomic DNA that was used for MSI analysis. For LOH testing, 50 ng of DNA was used for PCR, and PCR was performed using *Mlh1* genotyping PCR conditions and primers. Genomic DNA from *Mlh1*$^{+/+}$ and *Mlh1*$^{-/-}$ jejuna were used as normal and mutant allele controls, respectively. PCR products were analyzed by gel electrophoresis on an ethidium bromide stained 1.5% agarose gel, and DNA samples were considered to have undergone LOH if the wildtype PCR product was absent [35].

Methylation analysis for *Mlh1* was performed using methylation-specific PCR (MSP) [36]. Universally methylated mouse DNA standard (cat. no. D5012, Zymo Research, Irvine, CA) and genomic DNA extracted from normal spleen was used as positive and negative controls, respectively. For MSP, 500 ng of DNA was bisulfite-converted using EZ DNA Methylation-Direct Kit (Zymo Research) according to manufacturer's instruction. 1µl of bisulfite-converted DNA was used for PCR, which was performed using 2xZymo Taq premix. Details of the PCR program and primers [37] are in **Supplementary table 1**. PCR products were analyzed by gel electrophoresis on an ethidium bromide stained 1.5% agarose gel. Methylation status of the *Mlh1* promoter was scored qualitatively, based on presence or absence of the methylated PCR product.

Statistical analysis

Differences between groups in deletion% and insertion% were tested using an unpaired t-test, and linear regression fitting was tested using F-test. Two-tailed P values < 0.05 were considered statistically significant. P values are either stated as numerical values, or indicated as follows: p-value ≥ 0.05 = not significant (ns), p-value < 0.05 = *, p-value <0.01 = **, and p-value < 0.005 = ***.

Results

Normal jejunum of Mlh1$^{+/-}$ mice displays MSI

To investigate tissue- and age-specific MSI in *Mlh1*$^{+/-}$ mice, we performed highly sensitive SM-PCR in jejunum and spleen of 4- and 12-month-old *Mlh1*$^{+/-}$ mice (**Fig. 1A**) on three microsatellite loci (**Fig. 1B**). Mononucleotide tract A27 is an intergenic microsatellite approximately 2 kb downstream to *Epas1* gene, and A33 is an intronic microsatellite within *Epas1* gene. D14Mit15 is an intergenic microsatellite at 40 kb distance from *Ptpn20* gene. Compared to age-matched *Mlh1*$^{+/+}$ jejuna, in *Mlh1*$^{+/-}$ jejuna deletion% at mononucleotide repeats was elevated, while the dinucleotide repeat D14Mt15 was stable (**Fig. 2A, Supplementary fig. 1A and Supplementary fig. 1B**). Deletions at the mononucleotide repeats in *Mlh1*$^{+/-}$ jejuna were predominantly single repeat unit, i.e. 1 bp, shifts (**Fig. 2A**), and the average deletion% was 6% and 7% at 4 and 12 months, respectively (**Fig. 2B**). Compared to *Mlh1*$^{+/+}$ jejuna, the deletion% in *Mlh1*$^{+/-}$ jejuna were 3.2- and 3.5-fold higher, and compared to *Mlh1*$^{-/-}$ jejuna 5.2- and 6.4- fold lower at 4 and 12 months, respectively. Unlike *Mlh1*$^{+/-}$ jejuna, *Mlh1*$^{+/-}$ spleens were microsatellite stable at both time points (**Fig. 2A and Fig. 2B**).

In contrast to *Mlh1*$^{+/-}$ tissues, in *Mlh1*$^{-/-}$ tissues, deletion mutant alleles had bigger size shifts and were more frequent at all three microsatellites at both time points (**Fig. 2A, Supplementary fig. 1A)**. *Mlh1*$^{-/-}$ jejuna showed bigger size shifts and a higher deletion% than *Mlh1*$^{-/-}$ spleens; both deletion size shifts and deletion% were further increased in *Mlh1*$^{-/-}$ intestinal tumor (**Fig. 2A, Supplementary fig. 1A**). In *Mlh1*$^{-/-}$ jejuna, the average deletion% at mononucleotide repeats were 32% and 46%, and in *Mlh1*$^{-/-}$ spleen, 21% and 32% at the 4- and 12-month time points, respectively (**Fig. 2B**). For comparison, the *Mlh1*$^{-/-}$ intestinal tumor showed 52% deletion (**Fig. 2B**). D14Mit15 displayed 9% and 22% deletions in *Mlh1*$^{-/-}$ jejuna, 3% and 6% in *Mlh1*$^{-/-}$ spleen at the 4- and 12- month time points, respectively, and 24% in the *Mlh1*$^{-/-}$ intestinal tumor. *Mlh1*$^{+/+}$ tissues showed 2% deletions at mononucleotide repeats, and no detectable MSI at D14Mit15 at either time point (**Fig. 2B**).

Unlike deletions, insertions at all three microsatellites were not influenced by *Mlh1* gene dosage: *Mlh1*$^{+/+}$, *Mlh1*$^{+/-}$, *Mlh1*$^{-/-}$ mice showed minimal or no difference in insertion% in jejunum and spleen, with the exception of A33 at the 12-month time point (**Fig. 2, Supplementary fig. 1A, Supplementary fig. 1B and Supplementary fig. 2**). At 12 months, A33 showed a decreasing trend in insertion% with decreased *Mlh1* gene dosage in both tissues (**Supplementary fig. 2**). Irrespective

of genotype, tissue-of-origin or age, insertions were almost exclusively single repeat unit in size (1 bp and 2 bp for mono- and dinucleotide repeats, respectively) (**Fig. 2A and Supplementary fig. 1A**).

One 4-month old *Mlh1*$^{+/-}$ mouse showed substantially elevated deletion% in jejunum (25% and 4% at mononucleotide markers and at D14Mit15, respectively; indicated by an arrow in **Fig. 2B**). With Grubbs' test, this *Mlh1*$^{+/-}$ mouse was classified as an outlier ($p < 0.05$). A slight increase in deletion% at mononucleotide repeats (5%) was also detected in the spleen of this mouse, while D14Mit15 was stable. As in other *Mlh1*$^{+/-}$ mice, deletions in this mouse consisted predominantly of single repeat shifts (**Supplementary fig. 3)**. Moreover, inter-individual variation in mononucleotide deletion% was evident in *Mlh1*$^{+/-}$ jejuna in both age groups (**Fig. 2B**).

Mlh1$^{+/-}$ *mice show sporadic decrease in Mlh1 expression levels in jejunum*

To assess whether the inter-individual variation in deletion% in *Mlh1*$^{+/-}$ jejuna can be explained by *Mlh1* expression levels, we quantified Mlh1 mRNA and MLH1 protein levels; *Mlh1*$^{+/-}$ spleens, which were microsatellite-stable (**Fig. 2B**), were used as control tissue. To further examine tissue-specific differences in *Mlh1* expression, we analyzed Mlh1 mRNA expression in brain, kidney and liver. We also assessed variation in MLH1 expression by immunohistochemistry (IHC) in jejunum of 4-month old mice.

In *Mlh1*$^{+/-}$ jejuna, average *Mlh1* expression (both at mRNA and protein levels) was less than the expected approximately 50% of *Mlh1*$^{+/+}$ jejuna. Further, average expression was lower in the older age group, implying age-dependent MLH1 depletion. Additionally, expression levels varied between individual *Mlh1*$^{+/-}$ mice; mice in the 4-month old group more inter-individual variation than those in the 12-month old group. At 4 months, *Mlh1*$^{+/-}$ jejuna showed on average 26% Mlh1 mRNA (range: 8%-46%) and 24% (range: 0%-54%) MLH1 protein expression compared to *Mlh1*$^{+/+}$ levels, and at 12 months the average Mlh1 mRNA and MLH1 protein expression decreased to 15% (range: 9%-23%) and 17% (range: 1%-29%), respectively **(Fig. 3A**). These age-specific decreases in Mlh1 mRNA and MLH1 protein were not statistically significant, however.

To investigate whether younger mice show similar aberration in *Mlh1* expression levels, we analyzed Mlh1 mRNA expression levels also at 1-month time point. As the older cohorts, 1- month *Mlh1*$^{+/-}$ jejuna also showed less-than-expected- and variable Mlh1 mRNA expression levels (average: 30%,

range: 21%-37%). 1-month old *Mlh1*$^{+/-}$ mice showed a significant decrease in Mlh1 mRNA levels in jejunum (p=0.0283) when compared to 12-month old *Mlh1*$^{+/-}$ mice (**Supplementary fig. 4**).

Irrespective of age, we observed approximately expected 50% Mlh1 mRNA and MLH1 protein levels in all the *Mlh1*$^{+/-}$ spleens (**Fig. 3A, Fig. 3C and Supplementary fig. 4)**. Jejunum of the outlier *Mlh1*$^{+/-}$ mouse showed only 8% Mlh1 mRNA expression (**Fig. 3A**) and no detectable MLH1 protein (**Fig. 3C**), while its spleen had the expected levels of Mlh1 mRNA and MLH1 protein (**Fig. 3A and Fig. 3C**).

We also assayed Mlh1 mRNA expression in brain, kidney and liver of two 4-month old *Mlh1*$^{+/-}$ mice that expressed below age-average Mlh1 mRNA levels in their jejunum; Mlh1 mRNA expression in these tissues were as expected (60%, 56% and 56%, respectively) (**Supplementary fig. 5**).

Substantial inter-individual variation in MLH1 protein expression in *Mlh1*$^{+/-}$ jejuna was also detectable by IHC. Based on staining intensity, jejuna of half the *Mlh1*$^{+/-}$ mice (n=6) were scored as MLH1-positive and the other half was scored as low MLH1-positive by IHC profiler (**Fig. 3D**). *Mlh1*$^{+/-}$ jejuna scored as low positive had low MLH1 staining intensity throughout the jejunum with minimal variation (see SD in stacked bar in **Fig. 3D**) indicating jejunum-wide lower expression of MLH1. Additionally, we observed an increase in the number of MLH1-negative cells in jejunum of these mice (stack-bars in **Fig. 3D**). Upon further analyzing intestinal crypts and villi separately, in MLH1 low-positive jejuna we observed decreased MLH1 staining intensity in both; moreover, all MLH1-negative cells were located in the villi **(Supplementary fig. 6)**.

Sporadic depletion of MLH1 in Mlh1$^{+/-}$ *jejuna is primarily due to Mlh1 promoter methylation*

The variable MLH1 protein levels in jejunum of *Mlh1*$^{+/-}$ mice ranging from approximately expected 50% of *Mlh1*$^{+/+}$ to no expression (**Fig. 3C**) prompted us to investigate the cause of this sporadic MLH1 depletion. To distinguish between a genetic (loss of heterozygosity) and epigenetic mechanism (methylation of the *Mlh1* promoter), we performed LOH and MSP assays. *Mlh1* promoter methylation in jejunum was common (detected in 7/12 and 9/11 *Mlh1*$^{+/-}$ mice at 4 and 12 months, respectively; see **Table 1)**. Irrespective of age, all *Mlh1*$^{+/-}$ mice with extremely low (less than 3%) MLH1 protein expression in jejunum (4/8 and 2/6 *Mlh1*$^{+/-}$ mice at 4 and 12 months, respectively) displayed *Mlh1* promoter methylation (**Fig. 4**). Additionally, three of the four 12-month old *Mlh1*$^{+/-}$ mice with higher

level of MLH1 expression (range: 18%-29%) showed *Mlh1* promoter methylation, while none (out of four) of the 4-month old *Mlh1*$^{+/-}$ mice with higher MLH1 levels (range: 18%-52%) had *Mlh1* promoter methylation (**Fig. 4**).

LOH was only observed in the outlier 4-month old *Mlh1*$^{+/-}$ mouse which also had *Mlh1* promoter methylation (**Fig. 4**). All other *Mlh1*$^{+/-}$ mice assayed (n=22) had retained their wildtype allele, i.e. no LOH had occurred (**Table 1**).

MSI correlates with MLH1 protein level in jejunum of Mlh1$^{+/-}$ *mice*

To explore the tissue- and age-specific relationship between mononucleotide deletion% and MLH1 protein expression, we performed linear regression analysis. In *Mlh1*$^{+/-}$ jejuna, deletion% was inversely correlated with MLH1 protein expression at both time points (F test: R^2 = 0.87 and 0.80, F test: p-value = 1.9×10^{-6} and 2.9×10^{-4} for 4 and 12 months, respectively) (**Fig. 5A**). Further, based on this correlation, at both the time points, *Mlh1*$^{+/-}$ mice could be divided into two sub-groups, henceforth termed as sub-grp.1 and sub-grp.2, and defined by comparatively higher MLH1 protein levels and lower deletion% at mononucleotide repeats, versus low MLH1 protein levels and higher deletion% in jejunum, respectively.

At the 4-month time point, sub-grp.1 showed close-to-expected (approximately 50% of wildtype) average MLH1 protein levels (average 41%, range: 18%-54%) and 5% deletions. Sub-grp.2 had very low MLH1 protein levels (average: 2%, range: 0%-3%) and showed 10% deletions (**Fig. 5B**). Compared to *Mlh1*$^{+/+}$ jejuna, sub-grp.1 and sub-grp.2 showed 2.5- and 5-fold increase in deletion%, respectively. At the 12-month time point, sub-grp.1 showed lower-than-expected average MLH1 expression (average 24%, range: 18%-27%) and 7% deletions. Sub-grp.2, similarly as observed at 4 months, showed very low MLH1 protein levels (average: 1%, and range: 0%-2%) and increase in 10% deletions (**Fig. 5B**). Compared to *Mlh1*$^{+/+}$ jejuna, sub-grp.1 and sub-grp.2 showed 3- and 5-fold increase in deletion%, respectively.

Irrespective of age, all *Mlh1*$^{+/-}$ spleens showed the expected approximately 50% MLH1 protein level, and no increase in deletion% compared to *Mlh1*$^{+/+}$ spleens (**Supplementary fig. 9**). The *Mlh1*$^{+/-}$ outlier mouse (indicated by arrow in **Fig. 5A** and **Fig. 5B**) had no detectable MLH1 protein in its jejunum, accompanied by frequent (25%) mononucleotide deletions.

Discussion

Mlh1 mice [7, 22] are powerful model to perform systematic and comprehensive MSI study, and to dissect molecular mechanisms contributing to the MSI phenotype. In line with previous studies [29, 38, 39], we observed mononucleotide repeats to be more unstable than dinucleotide repeats. Key to detecting subtle cellular phenotypes with pre-malignant potential, such as low-level MSI, is use of sufficiently sensitive molecular read-outs. The SM-PCR assay on unstable loci can detect MSI levels as low as 1%, whereas standard PCR-based has a detection limit of 20-25% MSI [40]. Use of less unstable microsatellite markers, combined with conventional MSI assays may result even in failure to detect MSI in *Mlh1*$^{+/-}$ tumors [35].

Compared to 4-month-old *Mlh1*$^{+/-}$ mice, no further substantial increase in deletions was observed in jejunum of 12-month-old *Mlh1*$^{+/-}$ mice; in *Mlh1*$^{-/-}$ mice this increase was seen prominently (**Supplementary fig. 10**). This observation i.e. no strong correlation between MSI and age in MMR heterozygous condition is in line with previous studies [14, 15]. This can be explained by age-independent sporadic MSI increase in MMR heterozygous condition. In our study, 42% of 4-month old *Mlh1*$^{+/-}$ mice already showed deletion frequency comparable to 12-month old *Mlh1*$^{+/-}$ mice. As with MSI levels, no substantial decrease in Mlh1 mRNA and MLH1 protein levels was observed in jejunum between the two age groups. Additionally, 1-month old *Mlh1*$^{+/-}$ mice also showed a jejunum-specific inter-individual variation in Mlh1 mRNA levels, and showed lower than expected Mlh1 mRNA levels (**Supplementary fig. 4**), suggesting that the cellular changes (including MSI) may occur even earlier.

LS associated colorectal cancers are MMR-deficient and show high-level MSI, which led to the notion that two hits for MMR genes are required to instigate the MSI phenotype [9, 11]. However, recent evidence that LS-associated adenomas retain (some) MMR function suggests that the second hit may arise late in the multi-step tumorigenesis than previously appreciated [41], making the idea of pre-tumorigenic MMR haploinsufficiency more plausible. We demonstrate here that MSI is present in normal jejunum of *Mlh1*$^{+/-}$ mice with expected (i.e. 50% of wildtype) MLH1 levels, providing evidence for MMR haploinsufficiency. The second hit, when it eventually occurs, is primarily due to *Mlh1* promoter methylation, and occurs in a seemingly stochastic manner, likely giving rise to a fully MMR-deficient cell population. This stochastically acquired second hit can explain the inter-individual variation observed in Mlh1 mRNA, MLH1 protein, and MSI levels (**Fig. 5B**).

MSI reports on global MMR-dependent genome instability, but MSI *per se* does not necessarily lead to tumorigenesis [20]. Tumorigenesis initiates when MSI occurs in microsatellite-containing genes, giving the cells selective advantage and making them malignant. These genes include tumor suppressors and oncogenes involved in different regulatory pathways, including cell proliferation, cell cycle, apoptosis, and DNA repair [42, 43]. As genes (particularly exons) have more efficient MMR activity [44], it is unlikely that MSI-target genes would have early mutation accumulation, which could alter gene function. We did not observe any tumors in *Mlh1*$^{+/-}$ jejunum, although this observation was limited to mice that were sacrificed to harvest tissues. (We did not examine post-mortem the 30% *Mlh1*$^{+/-}$ mice which died in our care; **Supplementary fig. 11**) The MSI detected in intergenic/intronic microsatellites in normal *Mlh1*$^{+/-}$ jejuna reports on early-stage genomic instability prior to tumorigenesis. The 4-month-old outlier *Mlh1*$^{+/-}$ mouse with 25% mononucleotide deletions in its jejunum (comparable to age-matched *Mlh1*$^{-/-}$ mice) would presumably have had a higher chance of accumulating mutations in MSI-target genes that, over time, may have triggered tumorigenesis in the GI tract of this animal. Also, it seems likely that many, if not all, of those *Mlh1*$^{+/-}$ mice for which the cause of death is unknown (**Supplementary fig. 11**), succumbed to MMR-associated tumors.

Ours is the first systematic study that establishes the relationship between MLH1 levels and MSI in a variety of normal tissues. MSI in non-neoplastic mucosa [16] and crypt cells [45] has been reported in LS patient samples that likely were MMR-deficient. We demonstrate here that MSI is detectable – albeit at a relatively low level – in *Mlh1*$^{+/-}$ jejunum that still retains the expected 50% MLH1 protein level, implicating MMR haploinsufficiency (**Fig. 5B**) in normal, tumor-free *Mlh1*$^{+/-}$ intestine. High MSI only ensued upon substantial reduction of MMR protein levels. Thus, MMR function appears to rely on a critical threshold level of MMR proteins, which is regulated epigenetically, and is independent of age. Importantly, methylation-dependent lability of MLH1 expression affected the jejunum but not other tissues assayed.

Given that maintenance of genome stability is essential for all cells, it has been puzzling why germline mutations of key DNA repair pathways (mismatch repair, homologous recombination) give rise to cancer only in certain tissues. The tissue-specific haploinsufficiency of *Mlh1*, induced by promoter methylation, now provides an explanation for vulnerability of the GI tract to MMR-associated cancer. Heterozygosity of homologous recombination genes, namely *BRCA1* and *PALB2*, also confers subtle genome instability phenotypes that are detectable in pre-malignant cells of mutation carriers, provided that sufficiently sensitive assays are used [46, 47]. We propose that tissue-specific decrease in protein

levels, e.g. via promoter methylation, is an important factor in determining which organs are cancer-prone in heterozygous DNA repair mutation carriers.

Author contributions

K.S.S. designed and performed the experiments, analyzed the data, interpreted the results and wrote the manuscript. L.K. conceived and designed the study, interpreted the data, wrote the manuscript, acquired funding and supervised K.S.S. M.M.T. assisted in tissue collection, MSI assay, qPCR experiments, and genotyping.

Acknowledgments

We are grateful to Elli-Mari Aska, Erika Gucciardo, Manuela Tumiati, Päivi Peltomäki, Marjaana Pussila and Saara Ollila for critical reading of the manuscript, and to members of the Kauppi lab for constructive comments on the figures. We extend our sincere gratitude to the staff of the following core facilities at University of Helsinki: Laboratory Animal Center, Tissue preparation and histochemistry unit, Genome Biology Unit, and FIMM sequencing unit.

Disclosure of conflicts of interest

No conflicts of interest were disclosed.

Funding

This work was supported by the Academy of Finland grants: 263870, 292789, 256996, and 306026 (to L.K.), Sigrid Jusélius Foundation (to L.K.).

References

1. Fearon, E.R., *Molecular genetics of colorectal cancer.* Annu Rev Pathol, 2011. **6**: p. 479-507.
2. Burt, R.W., *Colon cancer screening.* Gastroenterology, 2000. **119**(3): p. 837-53.
3. Loukola, A., et al., *Microsatellite instability in adenomas as a marker for hereditary nonpolyposis colorectal cancer.* Am J Pathol, 1999. **155**(6): p. 1849-53.
4. Strand, M., et al., *Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair.* Nature, 1993. **365**(6443): p. 274-6.
5. Jiricny, J., *Postreplicative mismatch repair.* Cold Spring Harb Perspect Biol, 2013. **5**(4): p. a012633.
6. Pray, L.A., *DNA Replication and Causes of Mutation.* Nature Education, 2008. **1**(1).
7. Edelmann, W., et al., *Meiotic pachytene arrest in MLH1-deficient mice.* Cell, 1996. **85**(7): p. 1125-34.
8. Yamamoto, H. and K. Imai, *Microsatellite instability: an update.* Arch Toxicol, 2015. **89**(6): p. 899-921.
9. Aaltonen, L.A., et al., *Clues to the pathogenesis of familial colorectal cancer.* Science, 1993. **260**(5109): p. 812-6.
10. Knudson, A.G., Jr., *Mutation and cancer: statistical study of retinoblastoma.* Proc Natl Acad Sci U S A, 1971. **68**(4): p. 820-3.
11. Hemminki, A., et al., *Loss of the wild type MLH1 gene is a feature of hereditary nonpolyposis colorectal cancer.* Nat Genet, 1994. **8**(4): p. 405-10.
12. Mangold, E., et al., *Spectrum and frequencies of mutations in MSH2 and MLH1 identified in 1,721 German families suspected of hereditary nonpolyposis colorectal cancer.* Int J Cancer, 2005. **116**(5): p. 692-702.
13. Peltomaki, P., *Epigenetic mechanisms in the pathogenesis of Lynch syndrome.* Clin Genet, 2014. **85**(5): p. 403-12.
14. Alazzouzi, H., et al., *Low levels of microsatellite instability characterize MLH1 and MSH2 HNPCC carriers before tumor diagnosis.* Hum Mol Genet, 2005. **14**(2): p. 235-9.
15. Coolbaugh-Murphy, M.I., et al., *Microsatellite instability in the peripheral blood leukocytes of HNPCC patients.* Hum Mutat, 2010. **31**(3): p. 317-24.
16. Parsons, R., et al., *Mismatch repair deficiency in phenotypically normal human cells.* Science, 1995. **268**(5211): p. 738-40.
17. Kansikas, M., et al., *Assessing how reduced expression levels of the mismatch repair genes MLH1, MSH2, and MSH6 affect repair efficiency.* Hum Mutat, 2014. **35**(9): p. 1123-7.
18. Kasela, M., M. Nystrom, and M. Kansikas, *PMS2 expression decrease causes severe problems in mismatch repair.* Hum Mutat, 2019.
19. Wang, L., et al., *Whole-exome sequencing of human pancreatic cancers and characterization of genomic instability caused by MLH1 haploinsufficiency and complete deficiency.* Genome Res, 2012. **22**(2): p. 208-19.
20. Prolla, T.A., et al., *Tumour susceptibility and spontaneous mutation in mice deficient in Mlh1, Pms1 and Pms2 DNA mismatch repair.* Nat Genet, 1998. **18**(3): p. 276-9.
21. Edelmann, W., et al., *Tumorigenesis in Mlh1 and Mlh1/Apc1638N mutant mice.* Cancer Res, 1999. **59**(6): p. 1301-7.
22. Baker, S.M., et al., *Involvement of mouse Mlh1 in DNA mismatch repair and meiotic crossing over.* Nat Genet, 1996. **13**(3): p. 336-42.
23. Barker, N., M. van de Wetering, and H. Clevers, *The intestinal stem cell.* Genes Dev, 2008. **22**(14): p. 1856-64.
24. Crippen, T.L. and I.M. Jones, *Cell proliferation in the bone marrow, thymus and spleen of mice studied by continuous, in vivo bromodeoxycytidine labelling and flow cytometric analysis.* Cell Tissue Kinet, 1989. **22**(3): p. 203-12.
25. Pussila, M., et al., *Cancer-predicting gene expression changes in colonic mucosa of Western diet fed Mlh1+/- mice.* PLoS One, 2013. **8**(10): p. e76865.

26. Jeffreys, A.J., et al., *Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells.* Nucleic Acids Res, 1988. **16**(23): p. 10953-71.
27. Beal, M.A., et al., *Single-molecule PCR analysis of an unstable microsatellite for detecting mutations in sperm of mice exposed to chemical mutagens.* Mutat Res, 2015. **775**: p. 26-32.
28. Kraytsberg, Y. and K. Khrapko, *Single-molecule PCR: an artifact-free PCR approach for the analysis of somatic mutations.* Expert Rev Mol Diagn, 2005. **5**(5): p. 809-15.
29. Kabbarah, O., et al., *A panel of repeat markers for detection of microsatellite instability in murine tumors.* Mol Carcinog, 2003. **38**(4): p. 155-9.
30. Jeffreys, A.J., et al., *Complex gene conversion events in germline mutation at human minisatellites.* Nat Genet, 1994. **6**(2): p. 136-45.
31. Covarrubias-Pazaran, G., et al., *Fragman: an R package for fragment analysis.* BMC Genet, 2016. **17**: p. 62.
32. Berg, K.D., et al., *Detection of microsatellite instability by fluorescence multiplex polymerase chain reaction.* J Mol Diagn, 2000. **2**(1): p. 20-8.
33. Moolenbeek, C. and E.J. Ruitenberg, *The "Swiss roll": a simple technique for histological studies of the rodent intestine.* Lab Anim, 1981. **15**(1): p. 57-9.
34. Varghese, F., et al., *IHC Profiler: an open source plugin for the quantitative evaluation and automated scoring of immunohistochemistry images of human tissue samples.* PLoS One, 2014. **9**(5): p. e96801.
35. Pussila, M., et al., *Mlh1 deficiency in normal mouse colon mucosa associates with chromosomally unstable colon cancer.* Carcinogenesis, 2018. **39**(6): p. 788-797.
36. Herman, J.G., et al., *Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands.* Proc Natl Acad Sci U S A, 1996. **93**(18): p. 9821-6.
37. Fraga, M.F., et al., *A mouse skin multistage carcinogenesis model reflects the aberrant DNA methylation patterns of human tumors.* Cancer Res, 2004. **64**(16): p. 5527-34.
38. Webster, M.T., N.G. Smith, and H. Ellegren, *Microsatellite evolution inferred from human-chimpanzee genomic sequence alignments.* Proc Natl Acad Sci U S A, 2002. **99**(13): p. 8748-53.
39. Bacher, J.W., et al., *Use of mononucleotide repeat markers for detection of microsatellite instability in mouse tumors.* Mol Carcinog, 2005. **44**(4): p. 285-92.
40. Coolbaugh-Murphy, M., et al., *Estimating mutant microsatellite allele frequencies in somatic cells by small-pool PCR.* Genomics, 2004. **84**(2): p. 419-30.
41. Valo, S., et al., *DNA hypermethylation appears early and shows increased frequency with dysplasia in Lynch syndrome-associated colorectal adenomas and carcinomas.* Clin Epigenetics, 2015. **7**: p. 71.
42. Duval, A. and R. Hamelin, *Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability.* Cancer Res, 2002. **62**(9): p. 2447-54.
43. Vilkki, S., et al., *Screening for microsatellite instability target genes in colorectal cancers.* J Med Genet, 2002. **39**(11): p. 785-9.
44. Frigola, J., et al., *Reduced mutation rate in exons due to differential mismatch repair.* Nat Genet, 2017.
45. Kloor, M., et al., *Prevalence of mismatch repair-deficient crypt foci in Lynch syndrome: a pathological study.* Lancet Oncol, 2012. **13**(6): p. 598-606.
46. Pathania, S., et al., *BRCA1 haploinsufficiency for replication stress suppression in primary cells.* Nat Commun, 2014. **5**: p. 5496.
47. Nikkila, J., et al., *Heterozygous mutations in PALB2 cause DNA replication and damage response defects.* Nat Commun, 2013. **4**: p. 2578.

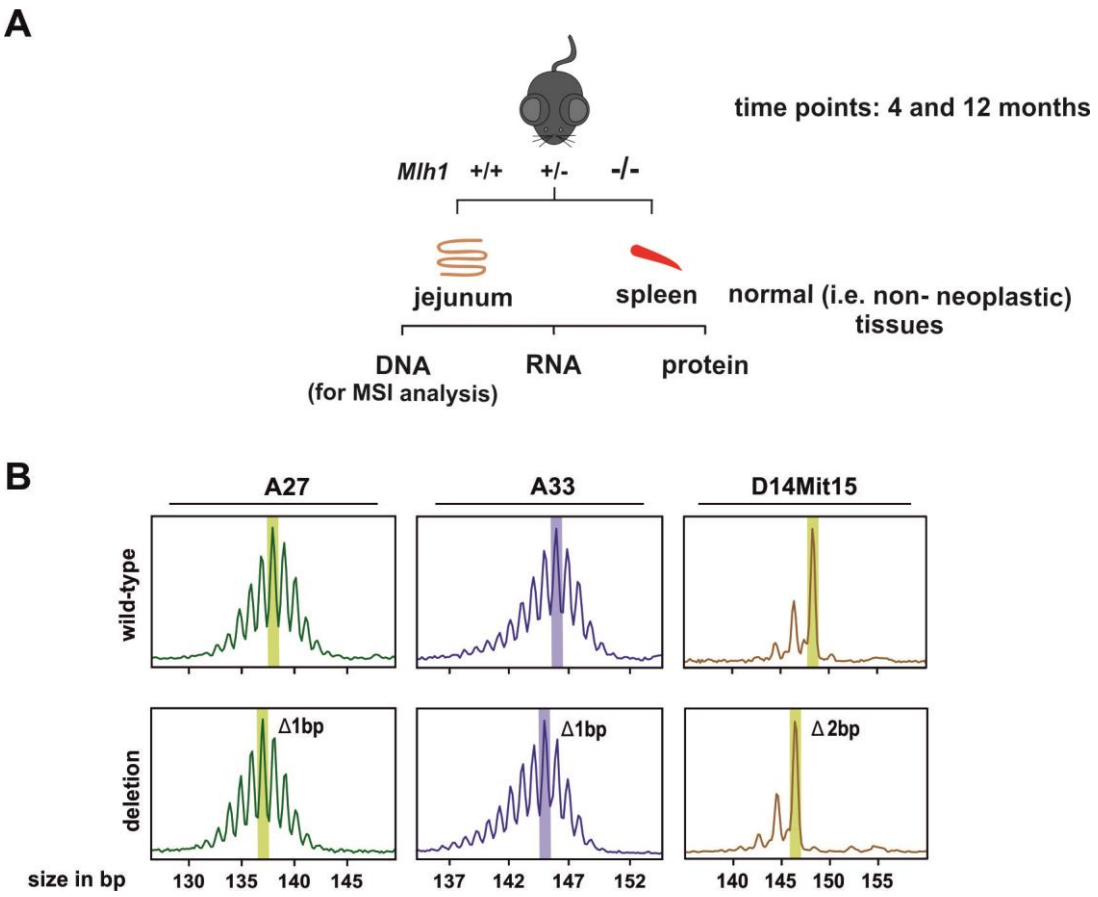


Fig. 1. Study approach. (**A**) Workflow of the study. We tested microsatellite instability (MSI) at single-DNA molecule resolution along with analyzing *Mlh1* expression both at mRNA and protein level. (**B**) Representative capillary electropherograms of single-molecule PCR (SM-PCR) based MSI assay. MSI was tested at three microsatellites: two mononucleotide repeats A27 and A33, and a dinucleotide repeat D14Mit15. Top and bottom panels show electropherograms scored as wildtype and single repeat unit deletion mutant alleles, respectively. Highlighted in yellow is highest peak that was scored; smaller peaks flanking the highest peak are stutter peaks, a typical feature of microsatellite markers.

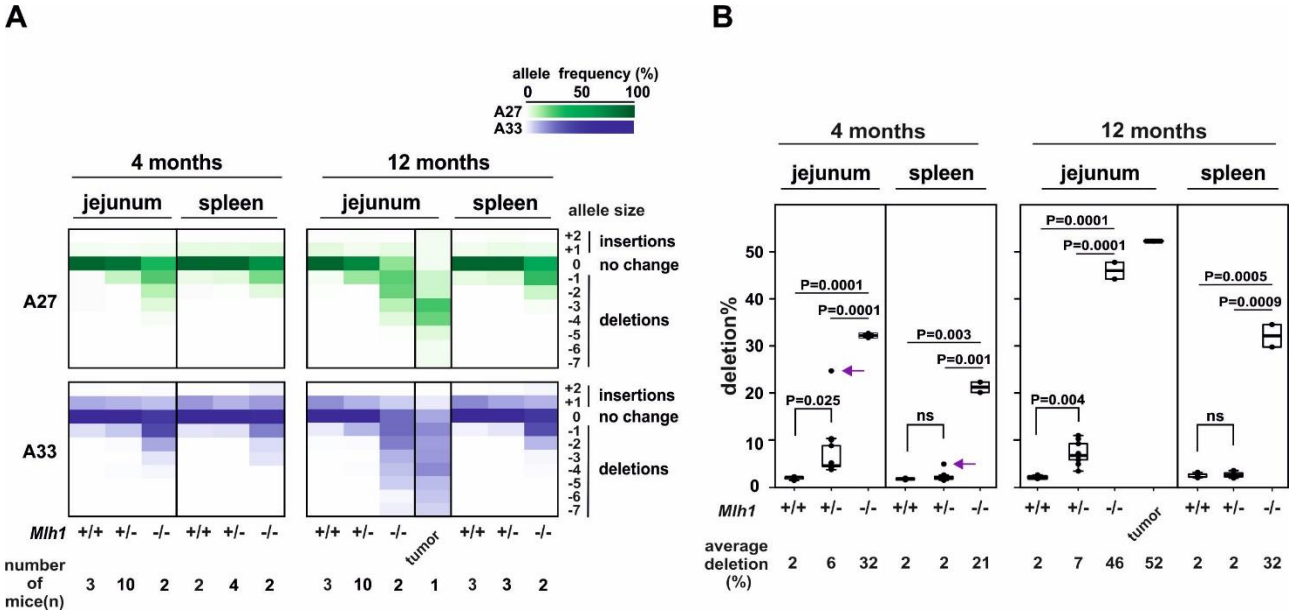


Fig. 2. Single-molecule MSI analysis at mononucleotide repeats in jejunum and spleen at 4- and 12-month time points. (**A**) Heat map of allele frequency (%) detected for microsatellites A27 and A33. The 4-month old *Mlh1* $^{+/-}$ outlier mouse (indicated by an arrow in **B**) was excluded from the heat map, and is shown separately in **Supplementary fig. 3.** (**B**) Deletion% in jejunum and spleen of *Mlh1*$^{+/-}$ mice. Arrow indicates the outlier *Mlh1*$^{+/-}$ mouse.

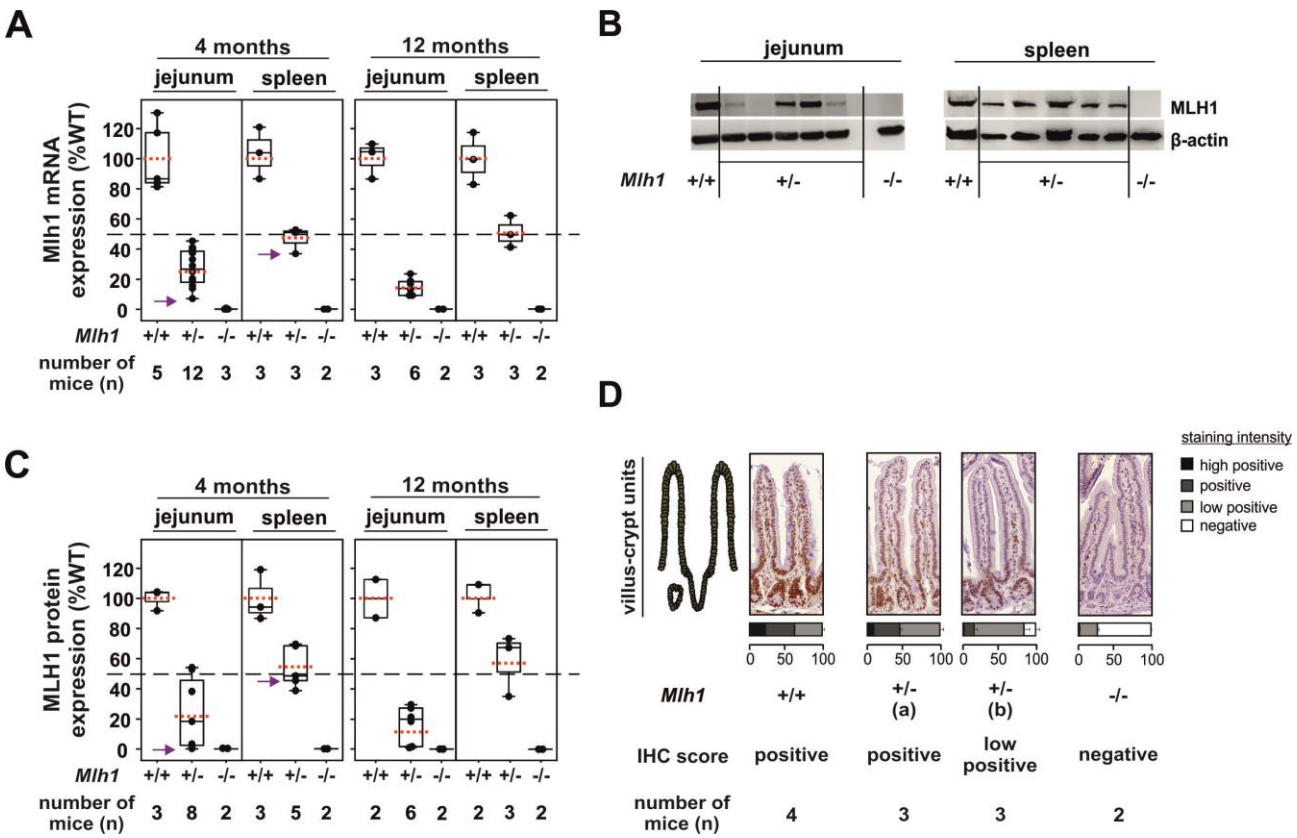


Fig. 3. *Mlh1* gene expression analysis in jejunum and spleen of *Mlh1*$^{+/-}$ mice age 4 and 12 months. (**A**) Mlh1 mRNA expression analysis. Arrow indicates data from the outlier *Mlh1*$^{+/-}$ mouse with high deletion% at mononucleotide repeats. Red dotted line across each box-plot indicates the average. Dashed black line across the chart marks 50% expression level. (**B**), (**C**) and (**D**) MLH1 protein levels analysis. (**B**) Representative image of western blot. The corresponding lanes in the two adjacent western blots are tissues from the same mouse. Empty lanes in western blot do not necessarily mean absence of MLH1 but possibly MLH1 protein levels below detectable range of our western blot assay (**Supplementary fig. 7**). (**C**) MLH1 protein expression analysis. The red-dotted line across each box-plot represents the average value. The dashed-horizontal line across the chart represents the 50% MLH1 protein expression. (**D**) Representative IHC images of jejunum. Middle two IHC images shows a side-by-side comparison of villus-crypt units scored as (a) positive and (b) low positive by IHC profiler. Bars below the representative IHC images show distribution of MLH1 staining classes (mean + SD) per genotype.

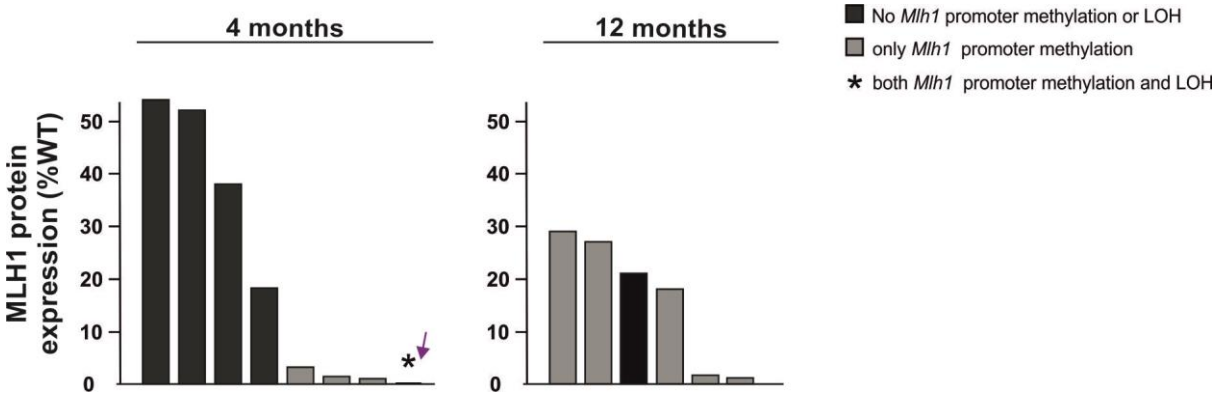


Fig. 4. Promoter methylation commonly underlies MLH1 depletion in *Mlh1*$^{+/-}$ jejuna. Each bar represents an individual *Mlh1*$^{+/-}$ mouse. Arrow indicates the outlier *Mlh1*$^{+/-}$ mouse.

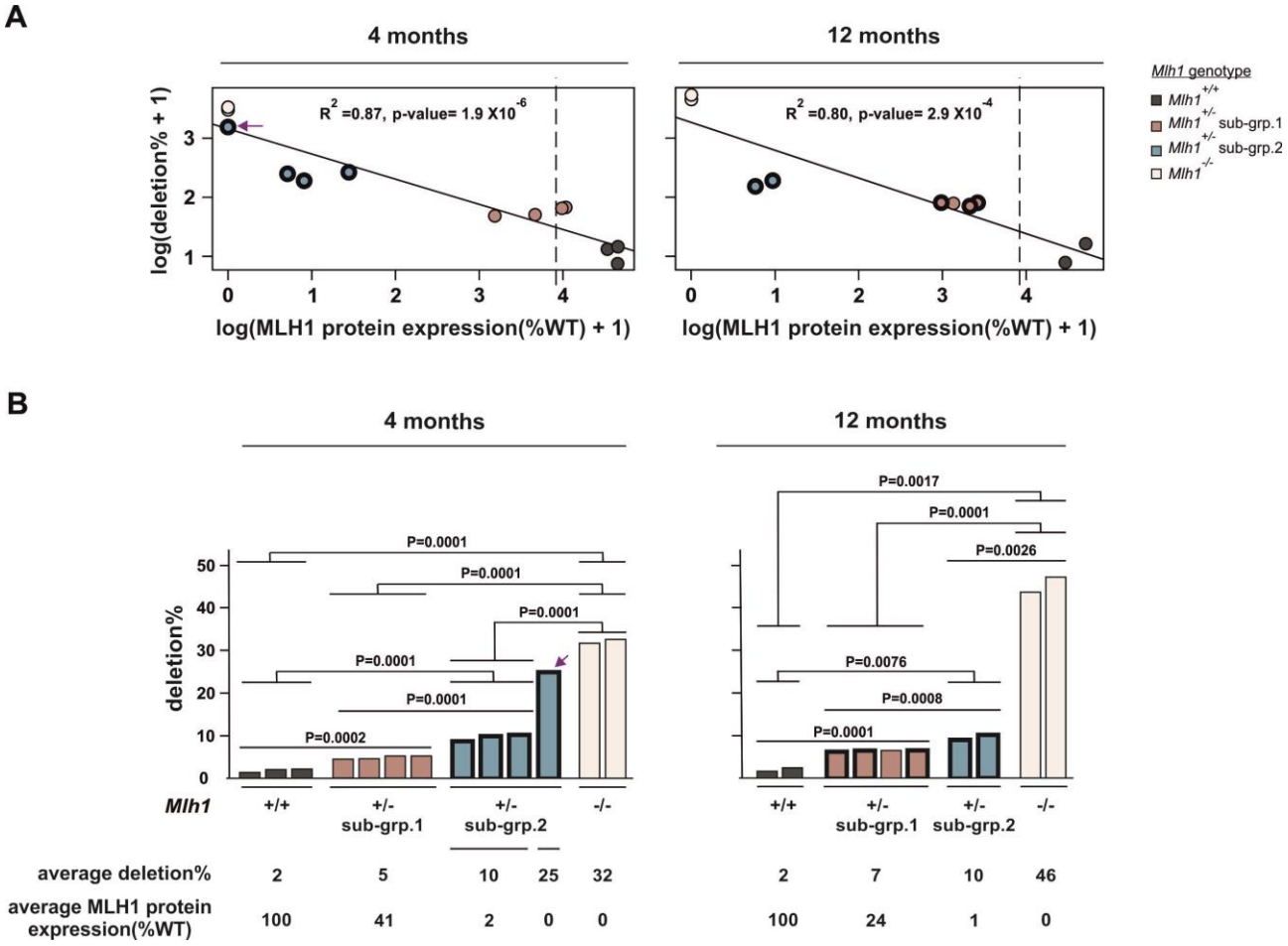


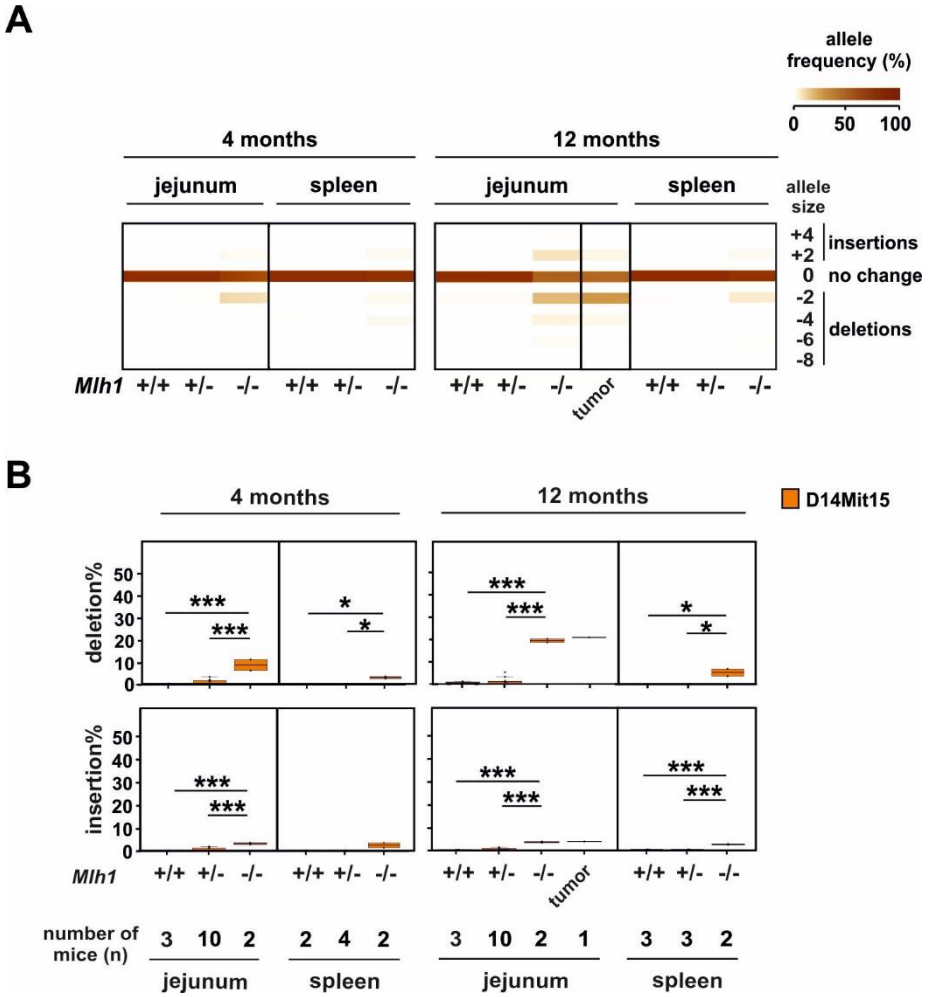
Fig. 5. Deletion% at mononucleotide repeats and MLH1 protein levels inversely correlate in jejunum of *Mlh1*$^{+/-}$ mice. (**A**) Linear regression fitting curve of log(deletions%+1) ~ log(MLH1 protein expression(%WT)+1)) at 4- and 12-month time point. Dashed vertical line marks 50% MLH1 expression level. (**B**) Comparison of deletion% between the two *Mlh1*$^{+/-}$ sub-groups. The arrow indicates the outlier *Mlh1*$^{+/-}$ mice. In both **(A)** and **(B)** thick black borders indicates those *Mlh1*$^{+/-}$ jejunum samples where *Mlh1* promoter methylation was detected.

Table 1. Percentage of *Mlh1*$^{+/-}$ jejuna with *Mlh1* promoter methylation and/or LOH. For representative images of MSP assay and LOH assay, see **Supplementary fig. 8.**

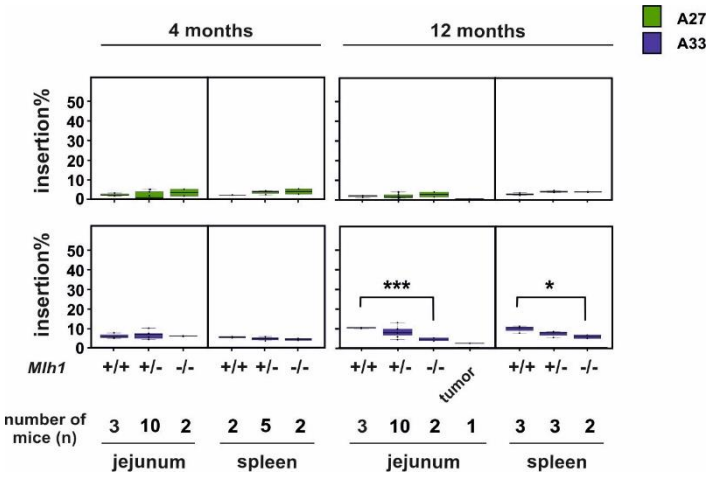
	4 months	**12 months**
***Mlh1* promoter methylation**	7 out of 12 (58.3%)	9 out of 11 (81.8%)
***Mlh1* LOH**	1* out of 12 (8.3%)	0 out of 11 (0%)

* Outlier 4-month *Mlh1*$^{+/-}$ mouse

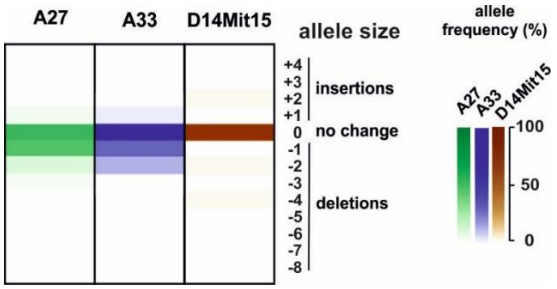
Supplementary figures:



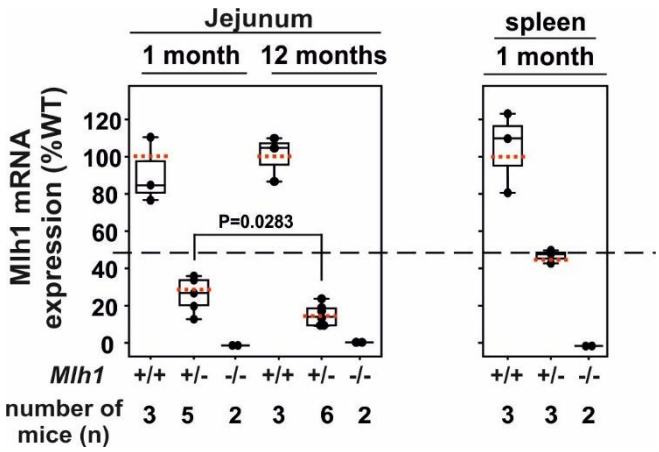
Supplementary fig. 1. Single-molecule MSI analysis at dinucleotide D14Mit15. (**A**) Heat map of the % of different-sized alleles at microsatellite D14Mit15. (**B**) MSI in normal jejunum and spleen. Indicated are significant p-values for the deletion% (top panel) and insertion% (bottom panel) comparisons of *Mlh1*$^{+/+}$ vs. *Mlh1*$^{+/-}$, *Mlh1*$^{+/+}$ vs. *Mlh1*$^{-/-}$, and *Mlh1*$^{+/-}$ vs. *Mlh1*$^{-/-}$ mice.



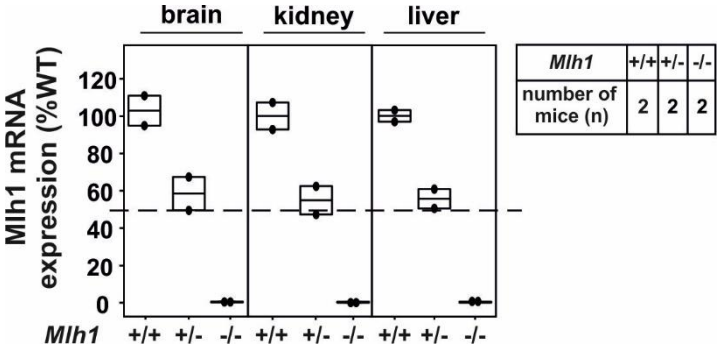
Supplementary fig. 2. Insertion% at mononucleotide repeats in jejunum and spleen. Top and bottom panel shows insertion% at A27 and A33, respectively. Indicated are statistically significant p-values for insertion% comparisons of *Mlh1*$^{+/+}$ and *Mlh1*$^{+/-}$, *Mlh1*$^{+/+}$ and *Mlh1*$^{-/-}$, and *Mlh1*$^{+/-}$ and *Mlh1*$^{-/-}$ mice at 4 and 12 month age.



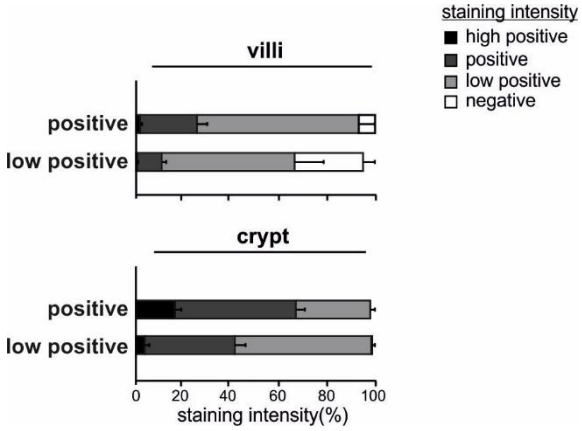
Supplementary fig. 3. MSI in the outlier *Mlh1*$^{+/-}$ mouse. Heat map of % of allele sizes for microsatellites A27, A33 and D14Mit15 observed in jejunum of outlier *Mlh1*$^{+/-}$ mouse with high deletion% at mononucleotide repeats.



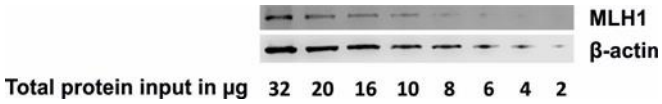
Supplementary fig. 4. Comparison of Mlh1 mRNA expression in 1- and 12-month old mice. The dashed horizontal line indicates 50% Mlh1 mRNA expression. The red dotted line across each box-plot shows the average value for each genotype group. Indicated is the only statistically significant p-value for within-genotype comparisons of Mlh1 mRNA levels in the jejunum in 1- vs. 12-month old mice.



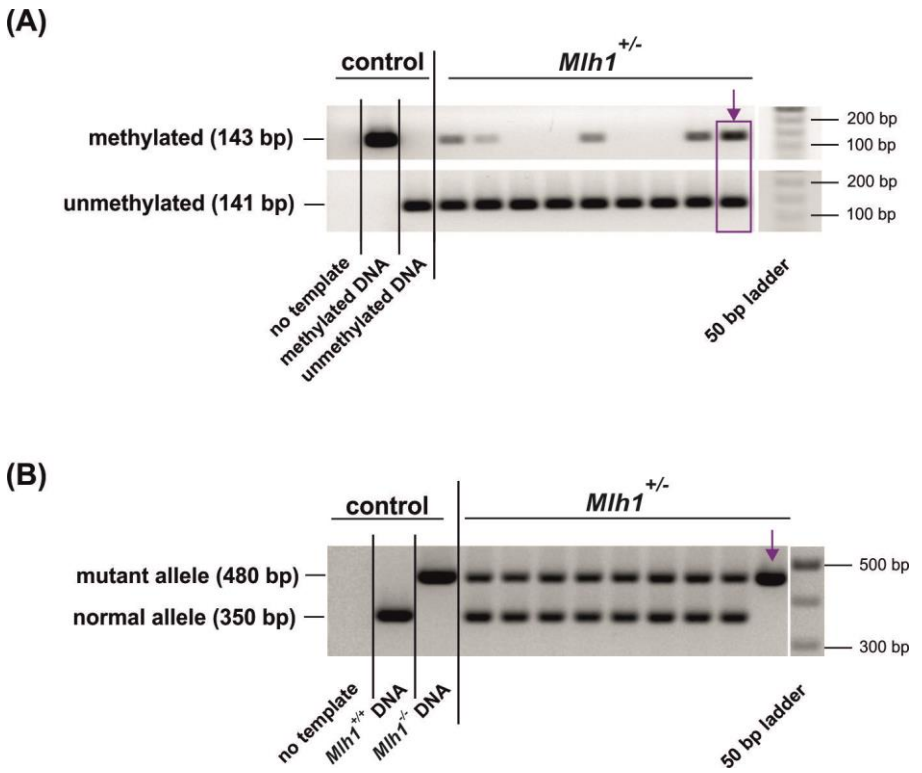
Supplementary fig. 5. Mlh1 mRNA expression in brain, kidney and liver of those *Mlh1*$^{+/-}$ mice with low Mlh1 mRNA levels in their jejunum. The dashed horizontal line indicates 50% of wild-type Mlh1 mRNA expression.



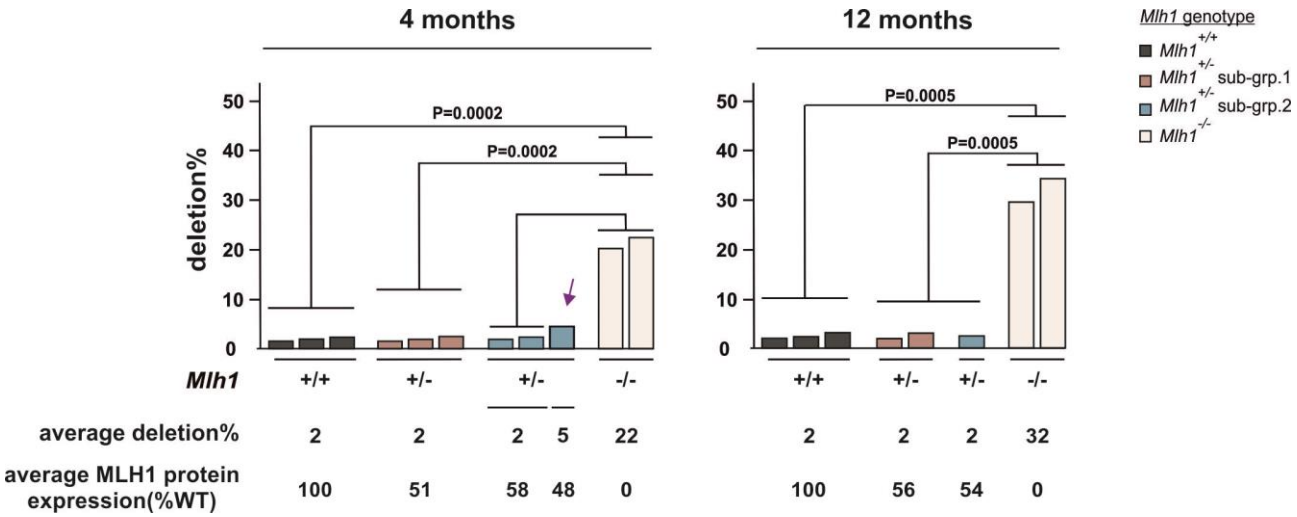
Supplementary fig. 6. Comparison of MLH1 staining intensity in crypts and villi of MLH1-positive and MLH1-low positive *Mlh1*$^{+/-}$ jejunum samples.



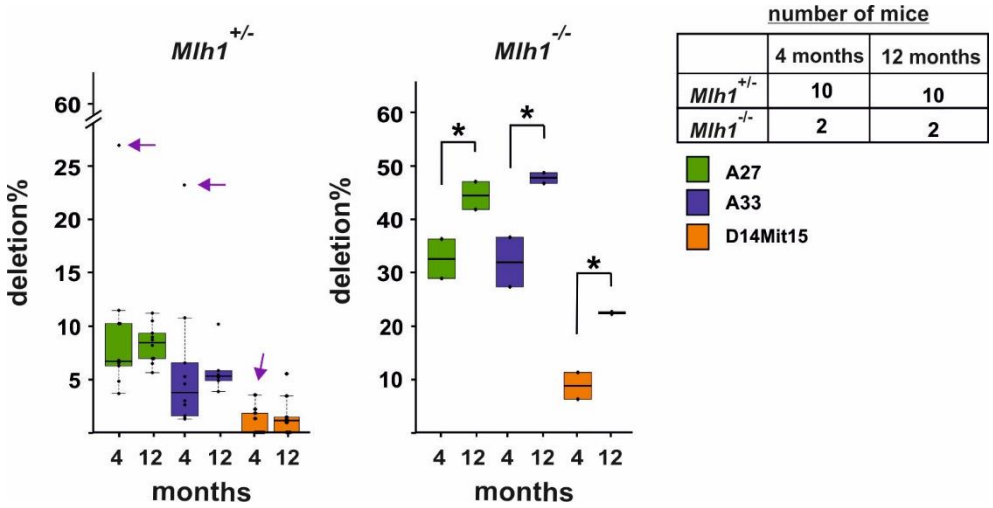
Supplementary fig. 7. Sensitivity of western blot assay. A dilution series of protein was run to determine sensitivity of the western blot assay. Infra-red signal was not detectable below 4 µg of MLH1 protein.



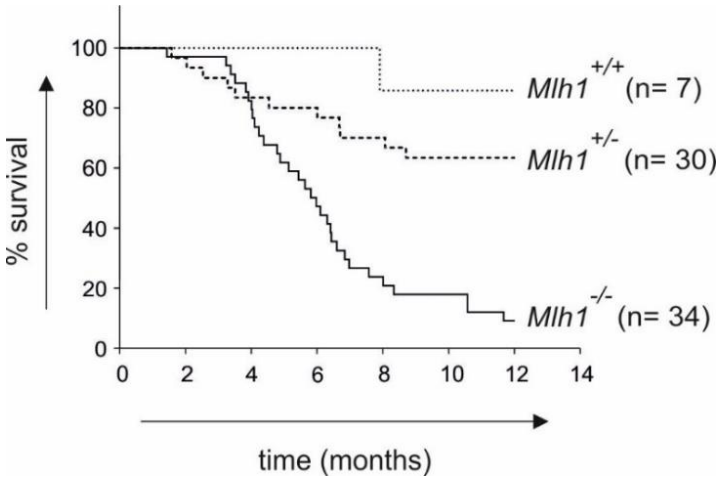
Supplementary fig. 8. Representative images of (A) MSP assay (B) LOH assay. The arrow indicates the outlier *Mlh1*$^{+/-}$ mice.



Supplementary fig. 9. Relationship between MSI and MLH1 protein levels in spleen of *Mlh1*$^{+/-}$ mice. Arrow indicates the outlier *Mlh1*$^{+/-}$ mouse.



Supplementary fig. 10. Deletion% and age. The arrow indicates the outlier *Mlh1*$^{+/-}$ mouse with high deletion%.



Supplementary fig. 11. Kaplan-Meier survival curves for *Mlh1* mutant mice.