Mlh1 deficiency in normal mouse colon mucosa associates with chromosomally unstable colon cancer

Pussila, Marjaana

2018-06


http://hdl.handle.net/10138/236613
https://doi.org/10.1093/carcin/bgy056

Downloaded from Held, University of Helsinki institutional repository.
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.
ORIGINAL ARTICLE

Mlh1 deficiency in normal mouse colon mucosa associates with chromosomally unstable colon cancer

Marjaana Pussila1, Petri Törönen2, Elisabet Einarsdottir3,4, Shintaro Katayama3, Kaarel Krjutškov4,5, Liisa Holm1,2, Juha Kere3,4,6, Päivi Peltomäki7, Markus J.Mäkinen8,9, Jere Linden10 and Minna Nyström1,*

1Faculty of Biological and Environmental Sciences, Molecular and Integrative Biosciences Research Programme and 2Institute of Biotechnology, University of Helsinki, FI-00014 Helsinki, Finland, 3Department of Biosciences and Nutrition, Karolinska Institutet, SE-14183 Huddinge, Sweden, 4Folkhälsan Institute of Genetics, Molecular Neurology Research Program, University of Helsinki, FI-00014 Helsinki, Finland, 5Centre for Health Technologies, Tartu University, Tartu, Estonia, 6Department of Genetics and Molecular Medicine, King’s College London, London SE1 9RT, UK, 7Medicum, Department of Medical and Clinical Genetics, University of Helsinki, FI-00014 Helsinki, Finland, 8Cancer and Translational Medicine Research Unit, Department of Pathology, University of Oulu, FI-90014 Oulu, Finland, 9Medical Research Center Oulu, Oulu University Hospital, University of Oulu, FI-90014 Oulu, Finland and 10Department of Basic Veterinary Sciences, FCLAP, University of Helsinki, FI-00014 Helsinki, Finland

*To whom correspondence should be addressed. Tel: +358 2 941 59073; Fax: +358 9 191 59079; Email: minna.nystrom@helsinki.fi

Abstract

Colorectal cancer (CRC) genome is unstable and different types of instabilities, such as chromosomal instability (CIN) and microsatellite instability (MSI) are thought to reflect distinct cancer initiating mechanisms. Although 85% of sporadic CRC reveal CIN, 15% reveal mismatch repair (MMR) malfunction and MSI, the hallmarks of Lynch syndrome with inherited heterozygous germline mutations in MMR genes. Our study was designed to comprehensively follow genome-wide expression changes and their implications during colon tumorigenesis. We conducted a long-term feeding experiment in the mouse to address expression changes arising in histologically normal colonic mucosa as putative cancer preceding events, and the effect of inherited predisposition (Mlh1+/-) and Western-style diet (WD) on those. During the 21-month experiment, carcinomas developed mainly in WD-fed mice and were evenly distributed between genotypes. Unexpectedly, the heterozygote (B6.129-Mlh1tm1Rak) mice did not show MSI in their CRCs. Instead, both wildtype and heterozygote CRC mice showed a distinct mRNA expression profile and shortage of several chromosomal segregation gene-specific transcripts (Mlh1, Bub1, Mis18a, Tpx2, Rad9a, Pms2, Cenpe, Ncapd3, Odf2 and Dclre1b) in their colon mucosa, as well as an increased mitotic activity and abundant numbers of unbalanced/atypical mitoses in tumours. Our genome-wide expression profiling experiment demonstrates that cancer preceding changes are already seen in histologically normal colon mucosa and that decreased expressions of Mlh1 and other chromosomal segregation genes may form a field-defect in mucosa, which trigger MMR-proficient, chromosomally unstable CRC.

Introduction

Colorectal cancer (CRC) is the third most common cancer and the fourth most common cause of cancer-related deaths worldwide. The incidence rates increase significantly with age and interactions between genetic and environmental factors, including diet, are suggested to play a critical role in its etiology (1,2). Cancer development always includes lack of genomic integrity in cells and different types of genomic instability, such as chromosomal instability (CIN) and microsatellite instability...
(MIN, also called MSI) are thought to reflect distinct initiating mechanisms in cancer (3). Three different pathways leading to genomic instability in colon cancer have been described. Most CRCs represent CIN, where chromosomes fail to trigger the spindle assembly checkpoint (SAC) leading to aberrant chromosome segregation. In recent years, many new genes have been reported, whose mutations and expression changes disturb chromosomal stability causing aneuploidy and/or comprehensive loss of heterozygosity (LOH) and alterations in chromosome structure (4,5). About 15% of sporadic CRCs and over 95% of CRCs in Lynch syndrome (LS), the most common inherited colon cancer syndrome, represent MSI caused by a defective DNA mismatch repair (MMR) mechanism (6). MMR deficiency causes accumulation of point mutations in the genome and especially in short repeat sequences called microsatellites, and is thought to be the driver defect in MSI carcinomas (7). The third pathway, CpG Island Methylator Phenotype (CIMP), characterized by global genome hypermethylation and tumour suppressor gene silencing, is seen in 20–30% of CRCs (6).

There are different ways to induce MMR malfunction and consequently lead to MSI. In sporadic tumours, the most common mechanism is the MLH1 (mutl. homolog 1) promoter hypermethylation often associated with BRAF V600E mutation (8). Also, in CIMP tumours the MLH1 is often inactivated by promoter hypermethylation but usually caused by aging and environmental factors (9). In LS, the first mutation is inherited, mainly in MLH1 (40%), MSH2 (34%) or MSH6 (18%) (10) and generally followed by LOH and somatic inactivation of the MMR gene (8). In LS, the second hit leading to MSI rarely happens through promoter hypermethylation (11). The fact that inactivation of MLH1 is associated with most MSI tumours irrespective of the inactivating mechanism emphasizes MLH1 deficiency as a key defect shared by sporadic and inherited CRC.

Colon cancer research focuses mainly on tumour characteristics, such as genomic instability, which can be utilized in treatment design. Recent findings have revealed that CIN and MSI pathways are not mutually exclusive (4,12), suggesting that also tumours with distinct features and instabilities may share initiative genomic aberrations while different tumour characteristics reflect subsequent alterations during cancer development. Since MLH1 deficiency seems to be a key defect shared by sporadic and inherited microsatellite unstable CRC, our aim was to follow how MLH1 would contribute to colon cancer development. Here, we used a mouse model to study cancer preceding expression changes in colon mucosa, Mlh1 phenotype in tumours and the effect of inherited predisposition (Mlh1+/−) and Western-style diet (WD) on those (13). We conducted a long-term feeding experiment with either a healthy rodent diet AIN-93G (AIN) or WD modified from AIN (Harlan Teklad, Madison, WI) (19) to resemble, on the nutritional level, the diet consumed in human Western population (high fat and energy content, low amounts of fiber, calcium and vitamin D) (Supplementary Table 1) (19) Twelve mice per each group (Mlh1+/+, AIN, Mlh1−/− AIN, Mlh1−/− WD, Mlh1+/+ WD) with equal representation of sexes, at time point (tp) 0 (5 weeks of age, Mlh1+/+, Mlh1−/−), tp1 (12 months of age), tp2 (18 months of age) and tp3 (21 months of age), 168 mice in total were sacrificed and sampled.

Collection of tumours and normal colon mucosa samples
All observed colon tumours were collected under dissecting microscope and preserved as FFPE blocks. If a tumour was large enough (3–5 mm in diameter), approximately half of it was embedded in O.C.T compound (VWR, Radnor, Pennsylvania) for cryo sampling. Histological studies, stainings and the grading of neoplasias were carried out at The Finnish Centre for Laboratory Animal Pathology (FCLAP), University of Helsinki, Finland. The neoplasias were graded as hyperplasias, adenomas and carcinomas according to criteria based on consensus rodent intestinal cancer nomenclature (21), (Supplementary Table 2). Longitudinal pieces (excluding the previously harvested tumorous sections), representing approximately one third of the proximal mouse colon were collected for normal mucosa. The mucosa was separated from the underlying submucosa and muscle layer under a dissecting microscope and samples for RNA extraction were stored in RNA later (Qiagen, Hilden, Germany) at −80°C (13).

Transcriptome analysis of normal mucosa
Transcriptome analysis was performed using RNA-sequencing method (RNA-seq). Total RNA was prepared from 0, 12 and 18 months old mice (14, 40 and 40 mice, respectively) using the RNeasy Plus Kit (Qiagen) with an extra DNase treatment (Qiagen). The RNA concentration was measured by Qubit 1.0 (Thermo Fisher Scientific, Waltham, MA) and RNA integrity with the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Only high quality RNA (RNA integrity number RIN ≥ 8) qualified for expression analysis (13).

RNA-seq method followed the single-cell tagged reverse transcription (STRT) protocol with modifications (22) (Supplementary Table 3). Briefly, 10 ng of total RNA was converted to cDNA and amplified to form an Illumina-compatible library. In total, 25 PCR cycles were used, but as four base-pair unique molecular identifiers were applied, only the absolute number of unique reads were included in the subsequent analysis. The samples were sequenced on a total of six lanes of Illumina HiSeq2000, further processed to fastq files by Casava 1.8.2 (both Illumina, San Diego, CA). Quality control was performed using the STRTprep pipeline (https://github.com/shka/STRTprep) (23).
processed reads were aligned by TopHat2 to the mouse RefSeq mm9 reference genome. STRT captures sequences at the 5′-end of poly(A)+ RNAs and the aligned reads therefore tend to be distributed close to the 5′-end (start site) of genes. STRTprep counts only the aligned reads at the 5′-untranslated region of protein-coding genes, or within the proximal (500 bp) upstream region (13).

Normalizing the RNA-seq data

STRTprep pipeline generated a read count matrix, with genes as rows and samples as columns. Different sample library sizes were normalized using DESeq-style normalization (24). Next shifted log transformation ($x_n = \log(n+1)$) was done to generate more Gaussian like data and the ComBat program was used to filter batch effects. These preprocessing steps and alternative pipelines were evaluated by looking at the hierarchical clustering of samples and by plotting quantiles of expression values for each sample (Supplementary Information, available at Carcinogenesis Online) (13).

Tests for differential gene expression

Since the analyzed data was not any more integer count values after ComBat normalization, we tested three t-test based methods, Voom-Limma, Cyber-T and Shrinkage-T for analysis of differential gene expression. All these methods add a prior to variance estimate. Shrinkage-T is the only method here that allows also testing with unequal variance. This turned out to be important, as the genes with strongest separation between the sample groups had small variance in the analysed subset and medium variance among remaining samples.

The three methods were evaluated by viewing the separation of cancer samples from the remaining samples in the multidimensional Scaling (MDS) plots with top-k genes, which were selected using the evaluated statistic. Parameter $k$ was varied from 25 to few hundreds. Shrink-T showed the best separation in the generated plots across all values of $k$. Each methods’ ability to find correlations with Gene Ontology classes was also tested. We used t-test scores from each method separately as an input to enrichment analysis tool called GSZ (Gene Set Z-score) (25). Shrink-T again generated strongest results. We therefore applied only Shrink-T in the subsequent analyses (13).

Visualization of sample differences

To detect similarities and differences between the samples, we used MDS that generates a small-dimensional visualisation from the multidimensional data while trying to preserve the pair-wise distances of samples from the multi-dimensional data. PlotMDS distributed in the Limma package was used as a basis of the analysis (26), although modified so that we were able to use any selected score to pick the genes that were used to calculate pair-wise distances (13).

The activity of Mlh1 was visualized with ComBat normalized data. Samples were grouped based on the sample types (genotype, diet and time-point) to highlight the sample differences.

Pathway analysis

To study the biological functions and pathways enriched among the top separating genes, we used QIAGEN’s Ingenuity Pathway Analysis (IPA Software 7.0, Qigenex). Here, we analysed both the top 100 and top 300 genes, which were found to separate the normal mucosa expression patterns in carcinoma mice from the others. The settings for a core analysis were as follows: ingenuity knowledge base (genes and endogenous chemicals) with both direct and indirect relationships, default network interaction settings (include endogenous chemicals, 35 molecules per network and 25 networks per analysis). Data sources were used with stringent confidence (experimentally observed and high predicted) and data obtained in all species was selected with a relaxed filter (13).

MSI and LOH analyses

The MSI status was analysed from all the seven carcinomas (two Mlh1+/− WD, four Mlh1−/− WD and one Mlh1−/− AIN mice, Table 1) from which a cryo-preserved sample was available, using four dinucleotide (D18Mit15, D14Mit15, D10Mit12, D7Mit91) and two mononucleotide (BH104, U12235) markers (27). Tumour DNA samples were extracted from the cryo-preserved colon carcinomas using laser micro-dissection for cutting (Zeiss PALM MicroBeam, Carl Zeiss Microscopy GmbH, Jena, Germany) and normal DNA control samples from the tails of the same mice with QIAamp DNA micro Kit, and DNeasy Blood & Tissue Kit (Qiagen), respectively. The genomic DNA was amplified with 6-FAM labeled primers in 11.1+ PCR master mix (28) using the following PCR protocol: 1 min at 96°C, 30 cycles of 20 s at 96°C, 30 s at 62°C and 15 s at 70°C, and 7 min at 70°C. The fragments were analysed with ABI3730xl capillary electrophoresis (Thermo Fisher Scientific) and visualized with PeakScanner v1.0 (Thermo Fisher Scientific) (13).

The four colon carcinomas found in the heterozygote Mlh1+/− mice, of which the cryo sample was available, were also studied for loss of Mlh1.
Loss of heterozygosity was analysed using the genotyping primers M001 (TGT CAA TAG CCT GCC CTA GG), M002 (TGG AAG GAT TGG AGC TAC GG), and M003 (TTT TCA GTG CAG CCT ATG CTC), which produce two different length fragments, 350 and 500 bp, separating the normal (M001/M003) and the mutated allele (M001/M003), respectively (19) (Figure 1B, C). DNA was amplified with the 11.1× PCR master mix as described above and the fragments were visualized on 1% SB agarose gel.

**Immunohistochemical analysis of Mlh1 protein expression in carcinomas**

Formalin-fixed, paraffin embedded cancer tissue blocks were studied for Mlh1 expression. The 4 μm thick sections were deparaffinized and rehydrated and heat induced antigen retrieval was performed with 10 mM citrate buffer (pH 6). To detect Mlh1, the slides were incubated overnight at 4°C with the rabbit monoclonal antibody ab92312 (1:1500) (clone EPR3894, Abcam, Cambridge, UK). Stainings were visualized using UltraVision Detection System anti-rabbit HRP/DAB (ThermoFisher Scientific, Waltham, MA) by manufacturer’s instructions. Analysis of staining patterns was conducted at The Finnish Centre for Laboratory Animal Pathology (13).

**Analysis of mitoses in carcinomas**

A Feulgen with Midori green background stain was used to visualise nuclear material and mitoses in six carcinoma samples (E249, E314, E329, E333, E338 and E347). The samples were deparaffinized and rinsed in 1 M HCl. Mild acid hydrolysis was accomplished by using 60°C 1 M HCl and DNA was stained purple in Schiff’s reagent for 45 min. After several washes the samples were counterstained briefly with 1% Midori light green, dehydrated through alcohol series to xylene and mounted with xylene based mounting media. The stained samples were analysed under light microscope (Zeiss Axio Imager.A2, Carl Zeiss Microscopy GmbH) and the mitoses in the malignant areas of carcinomas were compared to mitoses in samples (E305, E311, E322, E323 and E346) from healthy mice (13).

**Statistical analysis**

Differential expression analysis (DEA) used modified t-tests (Limma, cyber-T and shrinkage-T). With Limma and cyber-T, we used their own P-value estimates. Shrinkage-T does not provide a P-value estimate, which were estimated by re-calculating Shrinkage-T with 1000 permutations for each gene separately. Normal distribution was fitted to the permutations and a one-tailed P-value was obtained from the cumulative distribution. Multiple testing correction was performed using false discovery rate. Importantly, we used DEA mainly to order the genes to most differentially regulated genes. All analysis was performed within the R-environment. Pathway enrichment analysis was done using IPA which uses Fisher’s exact test to analyse over-representation of genes from the analysed gene groups. Here, multiple testing correction was done using the Benjamin-Hochberg method (13).

**Results**

Carcinomas developed mainly and earlier in WD fed mice

The feeding study was done with offspring produced by crossing two isogenic strains, the heterozygote Mlh1+/− (B6.129-Mlh1tm1Rak)

### Table 1. Characteristics of mice and their tumours

<table>
<thead>
<tr>
<th>ID</th>
<th>Age</th>
<th>Genotype</th>
<th>Diet</th>
<th>Gender</th>
<th>Histopathology</th>
<th>MSI</th>
<th>LOH</th>
<th>Mlh1 expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>E249</td>
<td>12</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>M</td>
<td>Tubular ac</td>
<td>NA</td>
<td>NA</td>
<td>Positive</td>
</tr>
<tr>
<td>E314</td>
<td>18</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>M</td>
<td>Mucinous ac</td>
<td>NA</td>
<td>NA</td>
<td>Positive</td>
</tr>
<tr>
<td>E329</td>
<td>18</td>
<td>Mlh1+/−</td>
<td>AIN</td>
<td>F</td>
<td>Tubulovillous ac</td>
<td>NA</td>
<td>NA</td>
<td>Positive</td>
</tr>
<tr>
<td>E333</td>
<td>18</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>F</td>
<td>Tubular ac</td>
<td>NA</td>
<td>NA</td>
<td>Positive</td>
</tr>
<tr>
<td>E338</td>
<td>18</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>F</td>
<td>Tubular ac</td>
<td>MSS</td>
<td>—</td>
<td>Positive</td>
</tr>
<tr>
<td>E347</td>
<td>18</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>F</td>
<td>Serrated ac</td>
<td>MSS</td>
<td>—</td>
<td>Positive</td>
</tr>
<tr>
<td>E402</td>
<td>21</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>F</td>
<td>Tubular ac</td>
<td>MSS</td>
<td>NA</td>
<td>Positive</td>
</tr>
<tr>
<td>E409</td>
<td>21</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>M</td>
<td>Mucinous ac</td>
<td>NA</td>
<td>NA</td>
<td>Positive</td>
</tr>
<tr>
<td>E410</td>
<td>21</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>M</td>
<td>Tubular ac</td>
<td>NA</td>
<td>NA</td>
<td>Positive</td>
</tr>
<tr>
<td>E411</td>
<td>21</td>
<td>Mlh1+/−</td>
<td>AIN</td>
<td>F</td>
<td>Tubular ac</td>
<td>MSS</td>
<td>NA</td>
<td>Positive</td>
</tr>
<tr>
<td>E437</td>
<td>21</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>M</td>
<td>Mucinous ac</td>
<td>MSS</td>
<td>—</td>
<td>Positive</td>
</tr>
<tr>
<td>E444</td>
<td>21</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>F</td>
<td>Tubulovillous ac</td>
<td>NA</td>
<td>NA</td>
<td>Positive</td>
</tr>
<tr>
<td>E214</td>
<td>12</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>F</td>
<td>Adenoma</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>E244</td>
<td>12</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>F</td>
<td>Adenoma</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>E246</td>
<td>12</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>F</td>
<td>Adenoma</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>E321</td>
<td>18</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>F</td>
<td>Serrated adenoma</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>E337</td>
<td>18</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>F</td>
<td>Adenoma</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>E402</td>
<td>21</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>F</td>
<td>Adenoma</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>E411</td>
<td>21</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>F</td>
<td>Adenoma</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>E416</td>
<td>21</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>M</td>
<td>Adenoma</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>E446</td>
<td>21</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>F</td>
<td>Adenoma</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>E448</td>
<td>21</td>
<td>Mlh1+/−</td>
<td>AIN</td>
<td>F</td>
<td>Adenoma</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>E217</td>
<td>12</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>M</td>
<td>Hyperplasia</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>E307</td>
<td>18</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>M</td>
<td>Hyperplasia</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>E328</td>
<td>18</td>
<td>Mlh1+/−</td>
<td>AIN</td>
<td>F</td>
<td>Hyperplasia</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>E405</td>
<td>21</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>F</td>
<td>Serrated hyperplasia</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>E408</td>
<td>21</td>
<td>Mlh1+/−</td>
<td>AIN</td>
<td>F</td>
<td>Hyperplasia</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>E409</td>
<td>21</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>M</td>
<td>Hyperplasia</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>E410</td>
<td>21</td>
<td>Mlh1+/−</td>
<td>AIN</td>
<td>M</td>
<td>Hyperplasia</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>E433</td>
<td>21</td>
<td>Mlh1+/−</td>
<td>AIN</td>
<td>F</td>
<td>Hyperplasia</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>E454</td>
<td>21</td>
<td>Mlh1+/−</td>
<td>WD</td>
<td>F</td>
<td>Hyperplasia</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

ac, adenocarcinoma; AIN, AIN-93G diet; F, female; M, male; NA, not analysed; WD, Western-style diet.
and the wild-type Mlh1+/− (C57BL/6) mice, and selecting an equal number of both genotypes to the study. Half of the mice fed WD and half the control diet, AIN-93G. In all 168 mice, 24 mice at time point 0 and 48 mice at time points 12, 18 and 21 months, were operated. Approximately 80% of all colon tumours, 10 out of 13 colon adenocarcinomas and 14 out of 20 adenomas and hyperplasias, developed in WD-fed mice (Table 1) (13).

At time points 12, 18 and 21 months, 80, 78 and 64% of all tumours and 100, 80 and 72% of CRCs were found in WD-fed mice, respectively. The overall number of colon tumours increased significantly with time (Supplementary Figure 1), being 5 at 12 months (one adenocarcinoma, two adenomas, one hyperplasia), 9 at 18 months (five adenocarcinomas, two adenomas, two hyperplasias) and 19 at the 21 months time point (seven adenocarcinomas, five adenomas, seven hyperplasias). Tumours were approximately evenly distributed between different genotypes since heterozygote Mlh1+/− mice showed 0, 40 and 43% of carcinomas and 50, 75 and 42% of adenomas and hyperplasias at different time points (Table 1). All the 13 carcinomas were found in the proximal part of colon and the majority of them were either tubular (54%) or mucinous (31%), two were tubulovillous and one carcinoma had serrated histology (Supplementary Table 2, Supplementary Figure 2, available at Carcinogenesis Online).

**Mlh1 mutation carriers did not show MSI, LOH and loss of MMR protein in tumours**

To check for the typical LS characteristics, seven carcinomas found in 18 and 21 months old mice, four in the Mlh1−/− mice (E338, E437, E437, E444) and three in the Mlh1+/− mice (E402, E410, E421), were analysed for MSI status and all the 13 carcinomas for Mlh1 expression. Surprisingly, all CRCs showed Mlh1 expression (Figure 1A), indicating that irrespective of the inherited mutation in one Mlh1 allele in the heterozygote mice, the normal allele was still present in the tumours. The presence of the normal allele was further confirmed with LOH study in all the four Mlh1−/− carcinomas (Figure 1B). To study whether the detected Mlh1 protein was functional and MMR proficient, we analysed the stability of six polymorphic microsatellite regions in the mouse genome. The markers and their amplified fragment sizes were as follows: D14Mit15 (148 bp, 150 bp), D18MIt15 (151 bp, 157 bp), D7Mit91 (139 bp, 147 bp), D10Mit2 (117 bp, 122 bp), JH104 (178 bp, 181 bp) and U12235 (79 bp, 83 bp). Altogether six out of seven CRCs were studied (E410 could not be amplified) and shown to be microsatellite stable, since no differences in the fragment lengths were observed between the tumour and corresponding normal DNA (Supplementary Figure 3, available at Carcinogenesis Online) (13).

**Mlh1 RNA expression was significantly decreased in normal mucosa of CRC mice**

After finding that irrespective of the mouse genotype, functional Mlh1 was still expressed in carcinomas, our interest was to look for potential early drivers of tumorigenesis on a genome-wide scale. Altogether 71 out of 80 normal colon mucosa samples operated from 12- and 18-month old mice were qualified for genome-wide transcriptome analysis. Analysis was done with RNAseq using the single-cell tagged reverse transcription method (STRT) (23,29). The 21-month old mice were left out from the RNAseq study due to many health problems most probably because of their old age. Altogether 12 216 expressed transcripts were identified in the samples. First, we analyzed the Mlh1 gene expression levels from the STRT data (Figure 2). In our previous study (19), we showed that in the beginning of the feeding experiment (at time point 0), the Mlh1 heterozygote mice showed exactly 50% lower Mlh1 expression than the Mlh1−/− mice. Contrary to varying Mlh1 expression levels in mice in general, 5/6 mice (E249, E314, E329, E333, E338) who developed carcinoma showed remarkably low Mlh1 expression in their normal colon mucosa (P = 0.03) (Figure 2) (13). The mouse E347 whose carcinoma had serrated histology had, however, higher Mlh1 expression level as non-carcinoma mice on average (Table 1; Figure 2).

**Expression profiles in normal mucosa formed a distinct cluster for CRC mice**

After finding that carcinoma mice had extremely low levels of Mlh1 transcripts in their mucosa, we next compared their genome-wide expression profiles with profiles of all other 12 and 18 months old mice. The normal mucosa expression profiles of the six CRC mice were strikingly different from the profiles of the other mice and formed a distinct cluster as visualized by an MDS plot created with the 100 most altered/differentially regulated genes (Figure 3, Supplementary Table 3, available at Carcinogenesis Online). Altogether 86% of the top 300 differentially regulated genes in CRC mice were down-regulated and 14% were up-regulated (Supplementary Table 3, available at Carcinogenesis Online) (13).

**Pathway analysis and shortage of chromosomal segregation gene-specific transcripts suggest problems in cell cycle regulation and mitosis**

To further understand the biological functions and pathways enriched among the top separating genes in CRC mice, the expression data were analysed with Ingenuity Pathway Analysis. According to Ingenuity Pathway Analysis, chromosome segregation (P = 2.92 × 10−6), aneuploidy of fibroblasts (P = 5.31 × 10−6), checkpoint control (P = 1.10 × 10−5), DNA replication checkpoint (P = 1.88 × 10−5) and morphology of mitotic spindle (P = 6.45 × 10−5) were among the most affected biological functions. In network analysis, the most affected molecular and cellular functions included cell cycle (P = 9.24 × 10−5), cellular assembly and organization (P = 9.24 × 10−5), DNA replication, recombination and repair (P = 9.24 × 10−5), cell death and survival (P = 7.30 × 10−5) and cellular growth and proliferation (P = 3.07 × 10−5) (Supplementary Table 4, available at Carcinogenesis Online). The analysis was also repeated with different RNA-seq data preprocessing (all mouse samples without ComBat normalization, Supplementary Table 4, available at Carcinogenesis Online). These results confirmed our findings on chromosome segregation (P = 1.03 × 10−3), aneuploidy of fibroblasts (P = 4.57 × 10−4) and checkpoint control (P = 4.29 × 10−5) (13).

The Ingenuity Pathway Analysis results strongly indicated that there are severe problems in cell cycle regulation and mitosis already in colon mucosa. In the six mice who developed carcinoma up to 18 months, the most altered/differentially expressed genes that pointed to chromosome segregation and SAC were Bub1 (BUB1, mitotic checkpoint serine/threonine kinase), Mis18a (MIS18A kinetochore protein A), Tpx2 (TPX2 microtubule associated), Rad9a (RAD9 checkpoint clamp component A), Pms2 (post-meiotic segregation increased 2), Mlh1 (MUTL homolog 1, along with MMR function also triggers checkpoint activation), Cenpe (centromere protein E), Ncapd3 (non-SMC condensing II complex subunit D3), Dof2 (outer dens fiber of sperm tails 2) and Dcrl1b (DNA cross-link repair 1B). Five of these 10 genes, Bub1, Mis18a, Tpx2, Rad9a and Pms2, were strongly down regulated in all of the six carcinoma mice (Figure 4). Mlh1, Cenpe, Ncapd3, Dof2 and Dcrl1b showed variable level of expression in two CRC mice
M.Pussila et al.

In E249, \( \textit{Mlh1} \) and \( \textit{Dclr1b} \) showed approximately 50% down regulation when compared to the average expression level in non-carcinoma mice (Figure 4). In E347, whose carcinoma was histologically different from the others and showed typical serrated phenotype, \( \textit{Cenpe}, \textit{Ncapd3} \) and \( \textit{Odf2} \) expression levels were equal and \( \textit{Mlh1} \) expression level higher than in the non-carcinoma mice. Importantly, among all the 74 mice that did not develop colon carcinoma up to 18 months, only two mice, E325 and E332, showed similar low expression of all the 10 genes (Figure 4). Although, no colonic tumours were found in those mice, E325 had bloody feces and anemia, suggesting undefined mucosal pathology (13).

**Abnormal mitoses and CIN in carcinomas**

Undisturbed mitosis is a central requirement of the normal cell cycle and division. In cancer cells, mitoses are often aberrant,
showing aneuploidy caused by unequal segregation of chromosomes and/or structural changes in chromosomes, both of which lead to CIN. To validate the RNA sequencing results, which suggested impaired cell cycle regulation and mitosis in CRC mice, all the 13 carcinomas were stained with feulgen and analyzed for mitotic aberrations. Although all the carcinomas were well-differentiated early cancers with limited submucosal invasion and relatively lenient cytological changes, they exhibited increased mitotic activity and abundant numbers of unbalanced/atypical mitoses in contrast to normal tissue samples (Figure 5) (13).

Discussion

The instability status and other characteristics of colon tumours have been used to define the mechanisms behind colon cancer formation (3). Our present genome-wide expression profiling experiment demonstrates that cancer preceding changes occur and can be detected already in normal colon mucosa. We show that these changes may form a field-defect in histologically normal mucosa and trigger MMR-proficient, chromosomally unstable CRC both in Mlh1+/− and Mlh1+/− mice.
One common event linked to both sporadic and inherited MSI cancers is the inactivation of MLH1, although the inactivation mechanism varies from genetic mutations in LS to epigenetic silencing in sporadic CRC (7,30). Our mouse study was designed to comprehensively clarify the role of Mlh1 expression during colon tumorigenesis and the effects of genetic predisposition and other risk factors such as WD and aging on it. Mlh1 protein expression was studied in colon tumours and Mlh1 gene expression in histologically normal mucosa. Approximately 70% of all tumours and 80% of colon carcinomas developed in WD-fed mice indicating a strong diet effect on cancer predisposition. 33% of CRCs and 75% of adenomas and hyperplasias found in mice up to 18 months of age was found in Mlh1+/− mice. Surprisingly, Mlh1 protein was present and there was no MSI in their cancers. Genome-wide expression profiling of histologically normal mucosa however showed that 5/6 mice who developed CRC up to 18 months had significantly decreased mucosal Mlh1 RNA expression. Only in the carcinoma mouse E347 the Mlh1 expression level was higher than the average level of 65 mice without cancer.

Low Mlh1 expression, although a prominent signal, seemed not to be an absolute requirement or sufficient alone to cause colon cancer since several mice without CRC had low Mlh1 expression as well. In order to identify other genes and pathways involved in CRC development, we compared the genome wide expression profiles in the six CRC mice with the profiles of 65 mice without CRC (13). Remarkably, the expression profiles of CRC mice formed a clearly distinct cluster (Figure 3), indicating a field-defect in normal colon mucosa (31,32). By network analysis of top 100 CRC mice separating genes, Mlh1 expression in normal mucosa from CRC mice was found to be low together with significant down regulation of several cancer related genes and pathways (Supplementary Table 3 and Tables 4, available at Carcinogenesis Online) and especially of chromosomal segregation genes, Bub1, Mis18a, Tpx2, Rad9a, Pms2, Cenpe, Ncapd3, Odf2 and Dclre1b. Only two (E325 and E332) of the 65 mice without cancer shared the expression profile of the CRC mice related to chromosomal segregation (Figure 4). Although no colonic tumours were found in those mice, carcinogenesis might have been happening in their mucosa. For example, E325 had bloody feces and anemia suggesting pathological problems in mucosa. Differing from the other CRC mice, E347, which did not show decrease in Mlh1 expression, showed decrease only in the expressions of Bub1, Mis18a, Tpx2, Rad9a, Pms2 and Dclre1b, suggesting their remarkable importance in serrated carcinogenesis. Furthermore, in the mouse E249 the Mhl1 and Dclre1b genes showed approximately 50% lower expression than was detected in the non-carcinoma mice on average. Here, the milder decrease may reflect young age of onset.

During cell division the SAC, which is the major target of mitotic alterations, maintains the genome stability by delaying cell division until all chromosomes are accurately aligned in the spindle (33). Aberrant expression of mitotic genes leads to mitotic abnormalities including centrosome defects and improper spindle checkpoint leading to CIN (34) and tumour formation in multiple tissues including colon (35). In this study, the mRNA expression was significantly decreased in five SAC associated genes, Mlh1, Bub1, Rad9a, Dclre1b and Cenpe. Of those, Bub1 is a major player and activator in SAC and its haploinsufficiency (heterozygosity) is known to be responsible for chromosome segregation defects and aneuploidy (35). During mitosis, Bub1 is required for the recruitment of other checkpoint and motor proteins, such as Cenpe, to a kinetochore (36). There is evidence suggesting that inaccurate chromosome segregation with causal implication of Bub1 deficiency drives tumorigenesis through tumour-suppressor gene LOH (37), perfectly in line with our findings that the majority of the carcinoma mice distinguishing genes were tumour suppressor genes, which were down regulated (Supplementary Table 3, available at Carcinogenesis Online). Cenpe, a kinesin-like motor protein which is an efficient stabilizer of microtubule capture at kinetochores and hence essential for metaphase chromosome alignment (38), was strongly down regulated in the mice with carcinoma. While it plays an important role in the movement of chromosomes toward the metaphase plate during mitosis, it is also necessary for the mitotic checkpoint signal at the kinetochore to prevent chromosome loss (39). Dclre1b has a central role in telomere maintenance and protection during S-phase through its 5′-3′ exonuclease activity. Moreover, in case of spindle stress, Dclre1b like Bub1 is involved in prophase checkpoint (40,41).

RAD9A, a component of the 9-1-1 cell cycle checkpoint response complex, plays a major role in DNA repair and participates in multiple cell cycle checkpoints and apoptosis and its aberrant expression has been linked to tumorigenesis of multiple tissues (42). Interestingly, Rad9 also physically interacts with the MMR protein MLH1 (43). The MMR mechanism is so essential for normal cell function that it may explain why even a small amount of MLH1 appears to be sufficient for MMR function, whereas its checkpoint activation role seems to require a full complement of the protein (44). It has been argued that the MLH1 heterozygosity/haploinsufficiency may drive the development of cancer by accumulation of insertion/deletion mutations in other gatekeeper genes prior to MSI (45). Indeed, cells with diminished amount of MLH1 protein may still be MMR proficient, although they show defects in DNA damage signalling (44). Consequently, the damaged cells may not activate cell cycle checkpoints and enter apoptosis. Our observation that low mRNA expression of Mlh1 in carcinoma mice together with down regulation of several other genes related to chromosome segregation and checkpoint control supports the proposition that already decreased amount of Mlh1, when MSI is still functional, may have an important role in tumorigenesis (13).

Low expression of Tpx2, Mis18a, Ncapd3 and Odf2 reflects problems in formation of the nuclear spindle and chromosome segregation. Tpx2, Ncpd3 and Odf2, a general scaffolding protein (46), are all involved in microtubules related processes in spindle formation. Tpx2 plays a role in microtubule organization and is involved in centrosome maturation (47). In fact, Tpx2-depleted cells fail to form proper mitotic spindles (48). Recent findings suggest that Tpx2 also plays an important role in promoting colon tumorigenesis (49). In this study results support a driver role for Tpx2, since it was strongly down regulated in colon mucosa in all carcinoma mice. Ncapd3 functions in the condensin II complex and is needed to establish the chromosomal architecture necessary for proper spindle assembly and chromosome segregation. Chromosome condensation and resolution are compromised when condensin is depleted (50). The MIS18 complex accumulates at the centromere during anaphase to early G1 phase, slightly ahead of the histone H3 variant CENPA and is an absolute requirement for the localization of CENPA at centromeres. Importantly, Mis18a knockout causes severe chromosomal missegregation, lack of CENPA and ultimately cell death (51). Here, along with Mis18a, Cenpa was significantly down regulated in the normal colon mucosa of the CRC mice (Supplementary Table 3, available at Carcinogenesis Online) supporting the finding of improper chromosome segregation.
Our RNA sequencing results suggest that cancer preceding changes occur already in histologically normal mucosa and trigger MMR-proficient chromosomally unstable CRC. Problems in cell cycle regulation and mitosis were confirmed by tumour analysis, which showed increased mitotic activity and abundant numbers of atypical mitoses in carcinomas. Altogether 71 mice were included in genome-wide expression profiling and only two of the 65 mice without cancer shared the expression profile of the CRC mice related to chromosomal segregation, indicating that the aberrant expression of this gene set could be used as signalling carcinogenesis in colon mucosa. In addition, the finding of unexpected tumour phenotypes in Mlh1+/- mice implies that stable microsatellites and normal MLH1 expression is not necessarily excluder inherited MLH1 deficiency.

Supplementary material
Supplementary data can be found at Carcinogenesis online.

Funding
This study was supported by grants from Jane and Aatos Erkko Foundation (M.N.), and Helsinki Institute of Life Science (P.T.).

Acknowledgements
The authors wish to thank Laura Sarantaus for help in planning the feeding study and sample collection, Satu Valo and Denis Dermadi-Bebek for help in sample collection, Marja Mutanen for the feeding study and sample collection, Satu Valo and Denis Dermadi-Bebek for help with the diet planning, Mikael Niku for help with immunohistochemistry, Liisa Kauppi for help in planning the MSI mark preparation. Liisa Kauppi for help in planning the MSI mark preparation.

Conflict of Interest Statement: None declared.

References