Mitotic abnormalities precede microsatellite instability in lynch syndrome-associated colorectal tumourigenesis

Marjaana Pussila,[a,*] Aleksi Laiho,[b] Petri Törönen,[b] Pauliina Björkbacka,[c] Sonja Nykänen,[a] Kirsi Pylvänäinen,[e] Liisa Holm,[b] Jukka-Pekka Mecklin,[d,e] Laura Renkonen-Sinisalo,[f,g] Taru Lehtonen,[f] Anna Lepistö,[f,g] Jere Linden,[c] Satu Mäki-Nevala,[h] Päivi Peltomäki,[h,i] and Minna Nyström[a]

[a]Molecular and Integrative Biosciences Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland
[b]Organismal and Evolutionary Biology Research Program, Faculty of Biosciences, and Institute of Biotechnology, Helsinki Institute of Life Science (HiLIFE), University of Helsinki, Helsinki, Finland
[c]Department of Veterinary Biosciences, and Finnish Centre for Laboratory Animal Pathology (FCLAP), Helsinki Institute of Life Science (HiLIFE), University of Helsinki, Helsinki, Finland
[d]Well Being Services County of Central Finland, Department of Science, Jyväskylä, Finland
[e]Faculty of Sports and Health Sciences, University of Jyväskylä, Jyväskylä, Finland
[f]Department of Surgery, Helsinki University Hospital, Helsinki, Finland
[g]Applied Tumour Genomics, Research Programs Unit, University of Helsinki, Helsinki, Finland
[h]Department of Medical and Clinical Genetics, University of Helsinki, Finland
[i]HUSLAB Laboratory of Genetics, HUS Diagnostic Center, Helsinki University Hospital, Helsinki, Finland

eBioMedicine
2024;103: 105111
Published Online xxx
https://doi.org/10.1016/j.ebiom.2024.105111

Summary

Background Lynch syndrome (LS) is one of the most common hereditary cancer syndromes worldwide. Dominantly inherited mutation in one of four DNA mismatch repair genes combined with somatic events leads to mismatch repair deficiency and microsatellite instability (MSI) in tumours. Due to a high lifetime risk of cancer, regular surveillance plays a key role in cancer prevention; yet the observation of frequent interval cancers points to insufficient cancer prevention by colonoscopy-based methods alone. This study aimed to identify precancerous functional changes in colonic mucosa that could facilitate the monitoring and prevention of cancer development in LS.

Methods The study material comprised colon biopsy specimens (n = 71) collected during colonoscopy examinations from LS carriers (tumour-free, or diagnosed with adenoma, or diagnosed with carcinoma) and a control group, which included sporadic cases without LS or neoplasia. The majority (80%) of LS carriers had an inherited genetic *MLH1* mutation. The remaining 20% included *MSH2* mutation carriers (13%) and *MSH6* mutation carriers (7%). The transcriptomes were first analysed with RNA-sequencing and followed up with Gorilla Ontology analysis and Reactome Knowledgebase and Ingenuity Pathway Analyses to detect functional changes that might be associated with the initiation of the neoplastic process in LS individuals.

Findings With pathway and gene ontology analyses combined with measurement of mitotic perimeters from colonic mucosa and tumours, we found an increased tendency to chromosomal instability (CIN), already present in macroscopically normal LS mucosa. Our results suggest that CIN is an earlier aberration than MSI and may be the initial cancer driving aberration, whereas MSI accelerates tumour formation. Furthermore, our results suggest that MLH1 deficiency plays a significant role in the development of CIN.

Interpretation The results validate our previous findings from mice and highlight early mitotic abnormalities as an important contributor and precancerous marker of colorectal tumourigenesis in LS.

Funding This work was supported by grants from the Jane and Aatos Erkko Foundation, the Academy of Finland (330606 and 331284), Cancer Foundation Finland sr, and the Sigrid Jusélius Foundation. Open access is funded by Helsinki University Library.

Keywords: Lynch syndrome; Chromosomal instability; Colon cancer; Field defect; Mismatch repair

*Corresponding author. Molecular and Integrative Biosciences Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, PO Box 56, FI-00014 University of Helsinki, Finland.
E-mail address: marjaana.pussila@helsinki.fi (M. Pussila).

Research in context

Evidence before this study

Lynch syndrome mutation carriers have a high lifetime colorectal cancer risk, and regular surveillance is essential for clinical management. The high occurrence of interval cancers despite preventive colonoscopy and removal of visible precursors poses a challenge to current surveillance methods, particularly for individuals with *MLH1* mutations. This may be due to precursor lesions, such as flat lesions associated with *MLH1* mutation, that can go undetected during colonoscopy. Recent findings indicate that different Lynch syndrome susceptibility genes may lead to distinct differences in the pathways of carcinogenesis, both at molecular and morphological levels. This highlights the importance of studying precancerous expression profiles linked to inherited mutations in different MMR genes. In this way, possible differences in respective carcinogenesis pathways can be detected and help to enhance surveillance methods to achieve more effective cancer prevention and treatment in patients with Lynch syndrome.

Added value of this study

Our findings from *MLH1* mutation carriers confirm our previous results in mice, emphasizing the significance of early chromosomal instability in colorectal tumour development. Our findings indicate that *MLH1* haploinsufficiency may induce mitotic abnormalities by impacting the formation of a functional mitotic spindle, disrupting the proper regulation of spindle assembly checkpoint, and impairing the repair of DNA double-strand breaks. The evidence we provide of mitotic abnormalities as an early cancer driving aberration offers new insights to the prevailing models of Lynch syndrome tumourigenesis.

Implications of all the available evidence

The study provides insights into the molecular mechanisms underlying the development of chromosomally unstable colon cancer, particularly focusing on gene expression differences and cellular functions related to chromosomal instability and aneuploidy. Understanding these mechanisms may have implications for cancer prevention and risk assessment in genetically predisposed individuals, and potentially aiding in the development of targeted prevention strategies and treatments for chromosomally unstable colon cancer.

Introduction

Lynch syndrome (LS) is one of the most common hereditary cancer syndromes accounting for some 3% of all colorectal and endometrial cancer cases, and up to 15% of those with microsatellite instability (MSI) or absent MMR protein(s) as evidence of mismatch repair deficiency (dMMR) in tumours.[1] Dominantly inherited germline defects in DNA mismatch repair (MMR) genes, *MLH1*, *MSH2*, *MSH6* and *PMS2*, underlie LS.[2] Typically, the remaining functional allele of the involved MMR gene is inactivated by loss of heterozygosity (LOH) or a somatic mutation, leading to dMMR. MSI is a hallmark of LS-related cancers and present in nearly all LS colorectal cancers (CRCs).[1] However, it is noteworthy that, in contrast to cancers, around 25% of LS adenomas still retain MMR function,[3] suggesting a difference in the stage of tumour development.

Individuals who have inherited a mutation that causes LS (LS carriers) have a high lifetime risk of cancer, and regular surveillance is an integral part of clinical management. Despite undergoing preventive measures such as colonoscopy and removal of visible precursors, high incidence of CRC exists.[4–8] This may be attributed to missed precursor lesions, particularly flat lesions that may go undetected during colonoscopy.[8] Moreover, immunological,[9] genetic[10] or epigenetic[11] mechanisms in colonic mucosa may contribute to rapid cancer development within surveillance interval.

In this study, we employed RNA-sequencing (RNA-seq) to identify expression changes linked to the initiation of a neoplastic process in macroscopically normal colonic mucosa, which subsequently drives progressive tumour development. Such changes could potentially be utilized for monitoring of cancer risk among LS mutation carriers. Building upon our previous research, which examined cancer preceding field defects in colonic mucosa using a mouse model,[12] we compared the transcriptomes of tumour-free (unaffected) LS mutation carriers and sporadic cases to identify LS-specific alterations. We further examined the association of these alterations with tumour development by comparing the expression profiles of unaffected LS carriers with those of LS carriers diagnosed with colorectal adenoma or carcinoma. Finally, we investigated the significance of the LS-specific expression differences to various biological functions. By analysing the results with Gene Ontology and Pathway analyses, they clearly and consistently showed an increased propensity for mitotic abnormalities which was already present in LS colonic mucosa, thus validating our previous findings from mice.

Methods

Patient samples

The study material (Table 1, Supplementary Table S1) consisted of fresh frozen (liquid nitrogen, N2) or

Case category	Tissue type	Group ID	No. of samples	Mean age in years (SD)	Mutated MMR gene (n)
Unaffected LS carriers	Normal mucosa	LS mucUA	18	45 (12)	*MSH2* (2) *MLH1* (16)
LS patients with adenoma	Normal mucosa	LS mucADE	18	54 (14)	*MSH6* (3) *MSH2* (3) *MLH1* (12)
LS patients with CRC	Normal mucosa	LS mucCRC	8	52 (12)	*MSH2* (1) *MLH1* (7)
LS patients with adenoma or CRC	ADE	LS ADE	11	55 (12)	*MSH6* (1) *MSH2* (1) *MLH1* (9)
	CRC	LS CRC	5	56 (11)	*MSH2* (1) *MLH1* (4)
Unaffected sporadic cases	Normal mucosa	SP mucUA	11	57 (17)	–

ADE, adenoma; CRC, colorectal cancer; LS, Lynch syndrome; muc, normal mucosa; SP, sporadic; UA, unaffected.

Table 1: **Characteristics of the sample groups.**

RNAlater (Qiagen, Hilden, Germany) preserved colon biopsy specimens. The samples were collected from two separate groups of patients: from MMR gene mutation carriers from known LS families registered in the Finnish Hereditary Colon Cancer Registry, and from sporadic cases who underwent colonoscopy examinations at Helsinki University Central Hospital or Jyväskylä Central Hospital between 2011 and 2019. Here the term sporadic cases refers to individuals without LS or neoplasia, who underwent colonoscopy for various reasons but had no previous or concurrent diagnosis of colon adenomas, carcinomas, or inflammatory bowel disease, hence, referred to as unaffected. Samples from sporadic cases were used as a control group.

Mucosal biopsies were obtained from proximal colon (caecum to splenic flexure). Tumour biopsies were paired with colonic mucosa samples from the same colon location, avoiding immediate proximity to the tumour. Sampling criteria included intact whole colon without colectomy, absence of regular NSAID medication, potential tumour diagnosis located in the proximal colon, and successful RNA extraction with sufficient RIN values (average 8.3, range 6.3–9.9). In all study groups, the sex distribution was quite even, while mean age in patients with LS was clearly lower than in the sporadic control group (45 and 57 years, respectively). Only colonic mucosa samples were collected from the sporadic cases (n = 11).

For patients with LS, biopsies were taken during regular colonoscopy follow-ups and divided into three groups (Table 1, Supplementary Table S1): unaffected LS carriers with no history of colon adenomas, carcinomas, or inflammatory bowel disease at the time of sampling (n = 18); LS carriers diagnosed with adenoma (n = 18) at the time of sampling; and LS carriers diagnosed with carcinoma (n = 8) at the time of sampling. The majority (80%) of LS carriers had an inherited genetic *MLH1* mutation. The remaining 20% included *MSH2* mutation carriers (13%) and *MSH6* mutation carriers (7%) (Supplementary Table S1).

Colonic mucosa samples (n = 55) were analysed from all the sample donors, and tumours (n = 16) from patients with LS whenever available. Fresh tissue samples were supplemented with corresponding archival formalin-fixed and paraffin-embedded (FFPE) specimens for histopathological evaluation, immunohistochemistry analysis of MMR protein expression, and morphological examination of mitoses.

Ethics

All biopsies were taken after patients' informed consent. The study was approved by the Institutional Review Boards of the Central Finland Health Care District (Dnro 10U/2011, 3.5.2011) and the Helsinki and Uusimaa Health Care District (HUS/390/2021, 23.2.2022). The National Supervisory Authority for Welfare and Health (Valvira) approved the collection of archival specimens (Dnro 10741/06.01.03.01/2015, 14.1.2016).

Library construction and RNA-seq

The transcriptomes were analysed with RNA-sequencing. Total RNA was extracted with the Rneasy Plus Kit (Qiagen, Valencia, CA) or the AllPrep DNA/RNA Mini Kit (Qiagen, Valencia, CA), and concentrations were measured by the Qubit fluorometer (RRID:SCR_018095). RNA integrity was confirmed with Agilent 2100 Bioanalyzer (RRID:SCR_018043). Sequencing libraries were prepared from 500 ng of total RNA with Illumina TruSeq stranded mRNA-kit. Sequencing was performed on the Illumina Novaseq 6000 system (RRID:SCR_016387) on two separate runs.

Preprocessing of RNA-seq data

To generate read count data from acquired sequence reads, the initial step involved adapter sequence removal using the Trimmomatic program.[13] Trimmomatic was executed with Illumina paired-end adapter sequences, with pair end read type. Filter parameters were set using the string": 2:30:10:2:keepBothReads," and the minimum accepted read length was established at 20.[13] For

details on Trimmomatic parameters, please refer to its manual.[13] These filtering rules, though lenient, aim to effectively remove adapter sequences, trusting the subsequent STAR aligner[14] to discard any erroneous short sequences. The primary focus here is on the efficient removal of adapter sequences.

Following adapter removal, reads were mapped against the human genome using the STAR aligner. Genes were mapped to the ENSEMBL genome assembly corresponding to GenBank Assembly ID GCA_000 01405.27.

Count data from STAR alignments were generated using the HTSEQ python library with the ENSEMBL GTF file Homo_sapiens.GRCh38.96.gtf. Overlapping sequence features were processed with 'Intersection Strict' setting. The parallelization of HTSEQ was executed with GNU-parallel.[15]

RNA-seq count data processing

DESeq2 program[16] was employed for normalizing various read libraries. Initially, we tested DESeq2 to identify differentially expressed genes and mitigate batch effects. However, the results were unsatisfactory, revealing substantial correlations with batch groups. Consequently, we opted to use DESeq2 solely for performing library normalizations and generating the Variance Stabilizing Transformation (VST)[17] on the sequence read data. Here the VST is a mapping that aims to adjust variances to be independent of the mean.[17]

Batch correction with combat method

The Combat[18] method was used for batch correction. In ComBat, the data with batch effects are corrected with shift and scale corrections (γ and δ). These are estimated with Empirical Bayes model.

Analysis of differential gene expression

We used the Shrinkage-T test[19] for the analysis of differential expression. Shrinkage-T uses a James Stein estimate for variance estimation in T-test calculation.

Gene ontology and pathway analyses

To explore biological functions and pathways enriched among differentially expressed genes (DEGs) between different sample groups, we used the Gorilla gene ontology (GO) analysis tool[20] and two commonly used pathway enrichment analysis methods as suggested by Chicco et al.[21]: Reactome Knowledgebase (version 83)[22] and Ingenuity Pathway Analysis (IPA) (CARRID:SCR_008653).[23] Gorilla and Reactome represent threshold-free enrichment analysis methods whereas IPA was used as a threshold-based enrichment analysis.

Gorilla gene ontology analysis

Prior to GO enrichment analyses, DEGs were sorted with positive (upregulated genes) and negative (downregulated genes) shrink-T scores, and the sorted gene lists were used as separate inputs. In addition, for comparative analysis, the sorted gene lists were additionally organized based on fold change values. Gorilla[20] was run with default settings. The obsolete and recurring ontologies were reduced with Revigo using default settings, except that the resulting list was set to large (0.9).[24] The result list consisted of up- and down-regulated GO terms with FDR < 0.05.

Reactome Knowledgebase analysis

Reactome is a manually curated database that provides bioinformatics tools for visualization and interpretation of pathway data.[23] Quantitative pathway analysis was performed with the Reactome Gene Set Analysis (ReactomeGSA[25]) using batch corrected VST-transformed count data. ReactomeGSA performs threshold-free gene set analysis using Reactome pathway groups. We selected the CAMERA algorithm.[26] $FDR < 0.05$ was set as a threshold for significantly differing pathways in comparisons.

Qiagen Ingenuity Pathway Analysis

Qiagen IPA is a web-based, manually curated software for analysing and interpreting diverse data sources, including RNA-seq.[27] A total of 500 most up- and 500 most down-regulated genes between sample groups were used as input. The settings for the core analysis were ingenuity knowledge base (genes and endogenous chemicals) with direct and indirect relationships and default network interaction settings. Data sources were used with stringent confidence (experimentally observed, high predicted) and data from all species was selected with a relaxed filter. In IPA the p-value is calculated using the right-tailed Fisher's Exact Test.

Methods behind gene ontology and pathway analyses are discussed more in detail in the Statistics section.

Genome instability analyses

MSI analysis

MSI was assayed from tumours using mononucleotide repeat markers BAT25 and BAT26 and following the previously described protocol.[11] A tumour was assessed as MSI, if at least one of the two markers was unstable, and as microsatellite stable (MSS) if both markers were stable.

Morphological analysis of mitoses

Since chromosomal instability manifests as atypical mitoses,[28] we reasoned it could be demonstrated by measuring the size of mitoses as previously described by Kesarkar et al.[29] Feulgen staining with light green background stain was used as previously described[12] to visualize nuclear material and mitoses in 4 carcinoma (LS_CAR2, LS_CAR3, LS_CAR4, LS_CAR5), 3 adenoma (LS_ADE9, LS_ADE11, LS_ADE12), and 5 colonic mucosa samples from patients with LS diagnosed either with adenoma (LS_N9, LS_N23) or carcinoma (LS_N5,

LS_N6, LS_N7) (Supplementary Table S1). A sigmoid colon carcinoma previously confirmed as MMR proficient (pMMR) and CIN positive by comparative genomic hybridization study (case no. 56:2)[30] was used as a positive control for the comparison of mitotic sizes. Here, we analysed all available specimens with sufficient material.

The Feulgen-stained samples were scanned to whole slide images (WSIs) using 3D HISTECH Pannoramic 250 Flash III slide scanner with 40x objective (3DHISTEC, Budapest, Hungary). Mitoses were identified and diagnosed normal/abnormal by their morphology and size in the WSIs and mitoses indicating abnormality were confirmed under light microscope (Zeiss Axio Imager.A2, Carl Zeiss Microscopy GmbH, Oberkochen, Germany). Finally, the perimeters of metaphases (viewed from above) were measured and the mean perimeter differences between sample groups were compared with independent samples t-test after data was tested for normality (Q–Q plot) and homogeneity of variance (Levene's test) (IBM SPSS Statistics, version 28.0.0.0, Armonk, NY: IBM Corp).

Statistics

Statistical analysis of RNA-seq data was performed using R 4.3.1 (R Core Team (2021), Austria, Vienna). First, the count data was processed with DESeq2[16] to normalize the different count datasets obtained from each sample and to remove size deviations between samples. Next, DESeq2 was used to calculate variance estimates at different signal levels. RNA-seq data is known for higher variance among lower signal levels and could be balanced with these estimates using a Variance Stabilizing Transformation (VST).[17]

Finally, DESeq2 was used to test batch correction and DGE analysis. Since the obtained expression values showed strong correlations with batch signals, we next decided to use the Combat batch correction method[18] from the sva-package (R package version 3.47.0). We used (VST) processed data from DESeq2 as input for ComBat. Since our data contained two separate batch effects, we used ComBat to first correct the stronger batch from the separate sequencing runs and then the batch from the two sample storage techniques. After correction, the data showed good separation between colonic mucosa, adenoma, and carcinoma sample groups and an even distribution of batch sample groups in the visualizations (Supplementary Fig. S1).

For DGE analysis, we used the shrinkage-T test (Shrink-T)[19] from the st-package. Shrink-T provides a more reliable T-test for small sample sizes by using the Stein-estimate for variance. Because Shrink-T lacks p-value estimates, we generated a set of permuted scores for each gene and fitted a normal distribution to these data. The p-values were corrected for false discovery rate (FDR) using the Benjamini-Hochberg method,[31] with the R-function p.adjust.

To reduce the impact of errors and increase reliability when analysing pathways and biological groups from RNA-seq data, we used three different enrichment analysis methods in parallel. The simplest enrichment analysis method was the Qiagen IPA pathway analysis,[24,27] which tests a preselected set of genes for overrepresentation of different pathways. These pathways were predefined by Ingenuity and Fisher's Exact Test (FET) was used with a right tailed p-value. In IPA a pathway is tested for its overrepresentation in a null model of sampling without replacement. In the IPA analysis, we used 500 genes with strongest upregulation and 500 genes with strongest downregulation from Shrink-T in the IPA analysis. The group size was determined according to the recommendations provided by Qiagen Ingenuity Pathway Analysis.[32]

Our second enrichment method, GOrilla[20] was run from the website. It takes an ordered list of genes, tests it at all possible threshold positions for overrepresentation with FET, and selects the strongest score obtained as the result. This result is compared to all possible permutations of the input gene list to obtain a p-value estimate. This method does not require a predefined set of genes, as IPA pathway analysis does. It generates a p-value with all possible permutations, but it assumes that the genes in the list are independently distributed, like FET. We used the T-test scores from Shrink-T to order the genes. However, we also tested GOrilla with fold change ordered genes. Gene Ontology results obtained are often redundant and may contain outdated categories. These were removed with Revigo[24] using the default settings, except that the resulting list was set to large (0.9).

The third enrichment method CAMERA[26] was used from the REACTOME[25] website (https://reactome.org/) by providing it with batch-corrected data. CAMERA corrects the independence assumptions of previous methods by considering the correlations between genes in its analysis. Next, the p-values are obtained from an asymptotic distribution. However, these methods, and especially GOrilla, can outperform CAMERA if the pathway or process under investigation shows heterogeneous regulation. Since the three methods use different ways to define the tested gene groups, their parallel use gives us thorough picture of the different signals among different pathways and biological processes.

To study CIN at the tissue level, the perimeters of metaphases (viewed from above) were measured, and mean perimeter differences were compared between sample groups. For continuous variables, an independent samples T-test was used after testing for normality and homogeneity of variance. Statistics were calculated using IBM SPSS Statistics, version 28.0.0.0, Armonk, NY: IBM Corp. Metaphase measurements were performed blinded for all available FFPE samples.

Because p-values have often been claimed to be misleading,[33] we included the Bayes Factor Bound (BFB) in our analysis.[34] It estimates the p-value as the Bayes Factor, the ratio between the signal model and the null model. We used R for the calculation.

In the visualization, we used Multi-Dimensional Scaling. It compresses multidimensional information into a low-dimensional representation while trying to preserve pairwise distances. This was done in R with the limma-package.[35]

The sample size was limited by the available patient material. Randomizations, when used, were done in the R environment by sampling without replacement. Patient genotype (Lynch syndrome mutation carriers, and sporadic non-carriers) was used to include each sample in the correct sample group.

Role of funders

The funding sources did not have any role in the study design, analyses, conclusions, and statements expressed herein.

Results

Differential gene expression analysis revealed distinct expression profiles for colonic mucosae

DGE analysis was performed to identify gene expression changes potentially associated with the initiation of tumourigenesis in LS. Gene ontology and cellular function analyses were then conducted for DEGs between different patient groups. To establish a baseline, we compared the gene expression profiles of colonic mucosa in unaffected LS carriers (LS mucUA) and unaffected sporadic cases (SP mucUA). A total of 1564 transcripts exhibited differential expression (214 upregulated, and 1350 downregulated) with $FDR < 0.05$. The top 100 DEGs with the highest absolute shrink-T scores were further analysed. Hierarchical clustering was applied, and a heat map (Fig. 1) was generated, showing the distinct difference between LS mucUA and SP mucUA.

Next, we compared LS mucUA with the expression profiles of patients with LS diagnosed with adenoma (LS mucADE) or CRC (LS mucCRC). For LS mucADE, 91 DEGs (70 upregulated, and 21 downregulated) with $FDR < 0.05$ were identified. Similarly, 302 DEGs (101 upregulated, and 201 downregulated) were found for LS mucCRC ($FDR < 0.05$). We focused on the top 100 genes with the highest absolute shrink-T scores by calculating their average ordinal number across all comparisons. Hierarchical clustering revealed distinct clusters representing unaffected LS carriers, patients with adenoma, and patients with CRC, although some overlap was observed (Supplementary Fig. S2).

Finally, we compared the expression profiles of LS mucUA to the expression profiles of LS adenomas and CRCs. In adenomas, 9958 DEGs were found with $FDR < 0.05$ (5394 upregulated and 4564 downregulated). Similarly, in carcinomas, 11 830 DEGs were expressed with FDR < 0.05 (4554 upregulated and 7276 downregulated). A heatmap of the top 100 genes (Supplementary Fig. S3) was generated and showed a clear distinction between the groups compared.

The Gene Ontology and Pathway Analyses Indicate Dysregulation of the Cell Cycle and Point to Chromosomal Instability in Colonic Mucosa from Unaffected LS carriers.

Here, we investigated the gene ontologies, Reactome pathways, and IPA pathways enriched in the colonic mucosa of unaffected LS carriers.

The entire set of 37 608 transcripts from the baseline comparison was used as input in the GOrilla gene ontology analysis, sorted separately by negative and positive shrink-T scores. As a result, 298 GO terms with FDR < 0.05 (Supplementary Table S2) were enriched. The GOrilla analysis of Fold Change sorted data yielded results that were quite similar to those obtained through the shrink-T analysis (Supplementary Table S3). The GO analysis indicated impaired cell cycle regulation and mitosis in LS mucUA (Table 2). The most significantly altered ontology was the Mitotic cell cycle process. Among the top enriched GO terms, other cell cycle-associated terms Regulation of mitotic cell cycle phase transition, Cell cycle process, Regulation of cell cycle phase transition, Regulation of cell cycle process, and Regulation of mitotic cell cycle, further suggested problems in cell cycle regulation. In addition, changes in cell division and chromosome organization were suggested by the enriched terms of Cell division, Chromosome organization, and Regulation of chromosome organization (Table 2). Other ontologies enriched among the top GO terms were linked to stress response, DNA metabolic process, including DNA repair and DNA replication, and immune system process.

Regarding the Reactome pathways, 286 pathways were altered in unaffected LS mucosa with an FDR < 0.05 and Bayes Factor Bound (BFB) values between 13 and 1.29×10^9. The most significantly altered hierarchical pathways DNA replication, Cell cycle, Metabolism of RNA, and DNA repair were all predicted to be downregulated. Cell cycle, and Cell Cycle Checkpoints, including all the three main cell cycle checkpoints G1/S, G2/M, and Spindle assembly checkpoint were downregulated in LS mucUA. The pathway analysis indicated downregulation of cell division, as the two major transitions G2/M transition and Mitotic Metaphase and Anaphase were downregulated when FDR < 0.05 was used. Also, DNA replication and DNA repair, including Mismatch repair, DNA Double-Strand Break Repair, Base Excision Repair, and Nucleotide Excision Repair were downregulated (Table 3, Supplementary Table S4).

In case of IPA pathways, we used 500 most up- and downregulated genes (n = 1000, Supplementary Table S5) between LS mucUA and SP mucUA as input in the Qiagen IPA. 500 annotations of diseases or

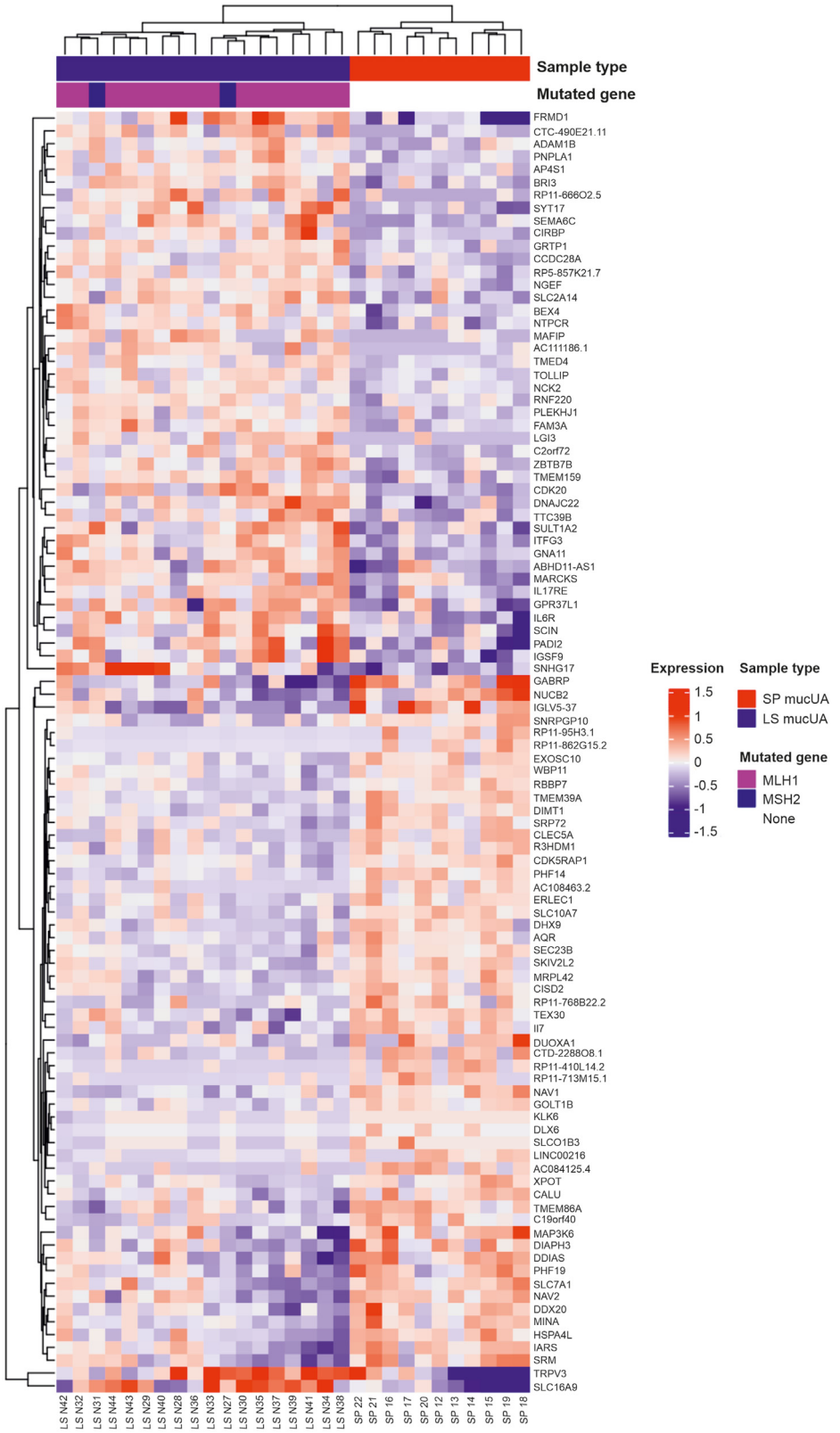


Fig. 1: Expression profiles of top 100 differentially expressed genes between colonic mucosa of unaffected LS carriers and unaffected sporadic cases.

GO Term	GO Term name	LS mucUA –SP mucUA	LS mucADE—LS mucUA	LS mucCRC—LS mucUA	LS ADE—LS mucUA	LS CRC—LS mucUA
GO:1903047	mitotic cell cycle process	−11.91	−6.52	−24.63	4.38	52.48
GO:0006950	response to stress	−11.09	3.14	4.56	16.31	16.98
GO:0006259	DNA metabolic process	−10.94	−1.35	−25.11	5.14	34.16
GO:0006281	DNA repair	−9.71	−2.54	−13.66	3.14	23.74
GO:0051276	chromosome organization	−9.73	−3.79	−12.29	2.75	27.74
GO:0002376	immune system process	−9.35	9.36	8.06	6.53	NA
GO:0080134	regulation of response to stress	−9.33	1.78	3.91	9.87	7.33
GO:0006974	cellular response to DNA damage stimulus	−9.33	−3.75	−12.29	5.51	28.88
GO:0006260	DNA replication	−9.14	−1.61	−18.35	NA	28.60
GO:0033554	cellular response to stress	−9.06	2.40	2.62	17.14	19.86
GO:1901990	regulation of mitotic cell cycle phase transition	−9.08	NA	−8.48	6.06	18.46
GO:0022402	cell cycle process	−9.05	−9.44	−22.65	NA	49.20
GO:0090304	nucleic acid metabolic process	−8.86	−2.22	−61.63	27.03	50.98
GO:0033044	regulation of chromosome organization	−8.53	−1.71	−8.24	5.42	12.83
GO:0051052	regulation of DNA metabolic process	−8.54	NA	−4.10	NA	18.43
GO:1901987	regulation of cell cycle phase transition	−8.49	NA	−8.28	6.21	18.32
GO:0010564	regulation of cell cycle process	−8.28	NA	−10.55	5.82	25.64
GO:0006139	nucleobase-containing compound metabolic process	−8.18	1.50	−58.02	30.94	49.40
GO:0043170	macromolecule metabolic process	−8.18	2.51	−23.14	31.48	33.20
GO:0046483	heterocycle metabolic process	−8.11	2.23	−57.11	30.71	45.95
GO:0051301	cell division	−7.61	−2.11	−20.31	4.52	29.97
GO:0071103	DNA conformation change	−7.60	−2.12	−7.14	NA	13.83
GO:0006725	cellular aromatic compound metabolic process	−7.30	2.75	−54.70	28.03	44.23
GO:0007346	regulation of mitotic cell cycle	−7.31	NA	−8.92	6.27	22.85
GO:0090329	regulation of DNA-dependent DNA replication	−7.08	NA	−2.67	NA	8.37

Values represent the absolute value of the logarithm of FDR multiplied by −1 or 1 (down- or upregulation of genes). The list is sorted by the biggest absolute values in the baseline comparison (LS mucUA—SP mucUA). mucADE, normal mucosa from LS patient with adenoma; mucCRC, normal mucosa from LS patient with colorectal cancer; SP, sporadic; ADE, adenoma; CRC, colorectal cancer; UA, unaffected.

Table 2: **Top 25 gene ontologies enriched in the baseline comparison, and in the normal mucosa and the tumours from LS patients with adenoma or CRC.**

biological functions with $p < 0.05$ and BFB values between 16 and 6.6×10^{37} were made, of which 297/500 were related to cancer (Supplementary Table S6). Due to recurring gene lists among cancer category functions, we omitted the excess cancer-related functions except for the most significant one, Cancer ($p < 0.0001$ Fisher's Exact Test, BFB 6.60×10^{37}), which allowed us to identify other altered functions in LS mucUA. Like the Reactome pathway analysis, the G2/M phase of cell cycle, and RNA metabolism were downregulated. Furthermore, DNA repair was downregulated, and DNA damage response of cells was altered (no z-score). Overall, IPA predicted altered chromosomal functions in LS mucUA such as increased Chromosomal instability as well as decreased Cycling of centrosome, and decreased Duplication of centriole (Supplementary Table S6).

The results suggest significant dysregulation in cell cycle regulation, cell division, chromosome organization, DNA replication, and DNA repair processes in colonic mucosa of unaffected LS carriers highlighting the potential implications for cancer-related functions.

Altered Cellular Functions Associated with Lynch Syndrome are More Prominent in Colonic Mucosa of Patients with Colonic Tumours.

We reasoned that if the altered cellular functions in LS mucUA are linked to tumourigenesis, they should also be enriched in colonic mucosa and tumours of patients with LS. Therefore, we expanded our analyses to include colonic mucosa and tumours of patients with LS. Our focus was on functions related to tumourigenesis, thus representing a cancer preceding field defect. The analysis of colonic mucosa from patients with LS-adenoma or LS-CRC revealed significant alterations in gene ontologies and pathways associated with cellular functions related to tumourigenesis.

In comparison to SP mucUA, the gene ontologies enriched in LS mucUA were also found to be altered in LS mucosa with adenoma and/or CRC when FDR < 0.05 was used (Table 2). Notably, cell cycle-related functions such as Mitotic cell cycle process, and Cell cycle process as well as Cell division and Chromosome organization were further altered in LS mucADE and LS mucCRC. Moreover, DNA replication, DNA repair, and Cellular response to DNA damage stimulus, along with Immune effector process were significantly altered in both sample groups (Table 2, Supplementary Table S7) suggesting escalation of these phenomena within mucosa of patients with tumours.

Hierarchical reactome pathways	LS mucUA–SP mucUA	LS mucADE–LS mucUA	LS mucCRC–LS mucUA	LS ADE–LS mucUA	LS CRC–LS mucUA
DNA replication	−5.71	0.03	−0.87	5.53	4.23
DNA replication pre-initiation	−4.90	0.52	−0.44	5.64	2.98
Synthesis of DNA	−6.77	0.14	−1.46	8.10	6.18
Cell cycle	−4.52	−2.69	−1.91	2.04	5.02
Cell cycle, Mitotic	−4.54	−1.85	−1.51	2.32	4.82
Cell cycle checkpoints	−6.24	−1.68	−1.98	4.20	6.54
Chromosome maintenance	−3.21	−3.40	−2.72	1.74	3.90
Metabolism of RNA	−3.70	0.33	−4.42	13.64	9.45
Processing of Capped Intron-Containing Pre-mRNA	−3.88	−0.19	−2.09	7.14	7.78
Regulation of mRNA stability by proteins that bind AU-rich elements	−2.79	5.02	0.14	7.76	2.10
rRNA processing	−2.26	1.97	−13.37	22.42	12.07
tRNA processing	−4.00	−0.54	−3.85	6.26	7.78
Metabolism of non-coding RNA	−6.19	−0.34	−1.73	3.74	6.74
DNA repair	−2.67	−1.27	−1.79	1.61	3.42
Base excision repair	−1.94	−0.09	−0.79	1.34	1.39
DNA damage bypass	−2.34	−0.17	−0.39	1.13	1.72
DNA double-strand break repair	−1.99	−3.74	−2.19	0.49	2.88
Nucleotide excision repair	−1.70	−0.04	−0.92	2.98	2.85
Mismatch repair	−2.59	−0.76	−0.57	0.39	1.70
Drug ADME	1.10	0.32	−0.08	−0.62	−1.43
Muscle contraction	1.06	−0.30	0.22	−0.89	−0.35
Immune system	−0.68	1.31	1.36	0.24	0.03
Reproduction	−0.65	−1.02	−0.13	0.04	0.30
Programmed cell death	−0.63	3.29	1.52	1.01	−0.02
Organelle biogenesis and maintenance	−0.58	−0.41	−0.30	0.01	0.17
Cellular responses to stimuli	−0.57	2.05	0.01	2.55	0.52
Digestion and absorption	0.43	0.03	0.69	−0.48	−0.85
Transport of small molecules	0.42	0.87	1.12	−0.26	−1.27
Autophagy	0.38	2.63	1.62	−0.04	−1.43
Metabolism	0.38	2.26	0.02	−0.05	−0.94
Metabolism of proteins	−0.36	1.04	0.02	1.47	0.22
Sensory perception	0.34	−0.08	0.08	−0.48	−0.60
Cell–cell communication	0.28	0.18	1.43	−0.10	−0.73
Neuronal system	0.23	−0.19	0.08	−0.46	−0.34
Disease	−0.23	0.48	0.11	0.48	0.12
Protein localization	0.22	4.47	−0.18	0.06	−0.88
Extracellular matrix organization	−0.15	0.34	0.11	0.26	1.10
Circadian clock	0.14	−0.21	1.65	−1.03	−0.45
Hemostasis	−0.13	0.18	0.27	−0.02	−0.05
Developmental biology	0.12	0.12	0.00	0.08	−0.02
Chromatin organization	−0.06	−2.07	−0.15	−0.02	0.50
Vesicle-mediated transport	−0.06	0.93	1.19	−0.08	−0.95
Signal transduction	0.06	−0.08	0.62	−0.26	−0.11
Gene expression (Transcription)	−0.05	−1.73	−0.58	−0.10	0.39

Values represent the absolute value of the logarithm of FDR multiplied by −1 or 1 (the predicted down- or upregulation). Sub-pathways are shown for hierarchical pathways that were significantly altered in the baseline comparison. mucADE, normal mucosa from LS patient with adenoma; mucCRC, normal mucosa from LS patient with colorectal cancer; SP, sporadic; ADE, adenoma; CRC, colorectal cancer; UA, unaffected.

Table 3: **Hierarchical reactome pathways in normal mucosa of unaffected LS patients, LS patients with adenoma or CRC, and the tumours.**

The Reactome pathway analysis demonstrated that a considerable portion of the altered pathways observed in the baseline comparison were further altered in LS mucADE (48%, 138/286) and LS mucCRC (36%, 104/286) (Supplementary Table S8). The affected pathways were mainly related to the cell cycle and cell cycle checkpoints, with specific attention to the down-regulation of the Mitotic spindle checkpoint (SAC) in LS mucADE and LS mucCRC. Furthermore, DNA metabolism related functions DNA strand elongation

Diseases or functions annotation	LS mucUA—SP mucUA		LS mucADE—LS mucUA		LS mucCRC—LS mucUA		LS ADE—LS mucUA		LS CRC—LS mucUA	
	p value[a]	z-score	p value[a]	z-score	p value[a]	z-score	p value[a]	z-score	p value[a]	z-score
Cancer	6.02×10^{-41}	−0.152	1.40×10^{-54}	0.288	1.48×10^{-66}	1.128	5.78×10^{-80}	2.403	7.28×10^{-81}	1.98
Liver lesion	8.68×10^{-09}	−0.407	1.59×10^{-07}	−0.089	5.70×10^{-09}	−0.427	3.60×10^{-12}	2.095	3.51×10^{-24}	−0.059
Brain lesion	4.83×10^{-12}	−1	5.65×10^{-11}	0.58	7.05×10^{-24}	0.481	8.26×10^{-22}	1.732	2.98×10^{-21}	0.076
Organismal death	9.47×10^{-04}	1.939	2.36×10^{-03}	−1.754	4.95×10^{-19}	−1.269	1.47×10^{-11}	0.394	5.34×10^{-17}	−2.002
Organization of cytoplasm	2.04×10^{-03}	0.393	NA	NA	1.30×10^{-05}	0.587	9.99×10^{-08}	−0.363	1.03×10^{-14}	−0.364
Lymphoproliferative disorder	6.17×10^{-04}	0.947	NA	NA	5.27×10^{-10}	1.667	1.21×10^{-06}	0.663	3.89×10^{-14}	−0.165
Organization of cytoskeleton	2.57×10^{-03}	0.388	NA	NA	1.74×10^{-04}	0.587	4.24×10^{-06}	−0.256	4.38×10^{-14}	−0.364
Abnormality of endometrium	8.59×10^{-09}		4.44×10^{-03}		3.83×10^{-06}	0	9.36×10^{-10}		1.55×10^{-13}	
Viral Infection	1.91×10^{-04}	−0.242	7.86×10^{-05}	2.039	2.35×10^{-06}	−0.476	1.81×10^{-13}	1.414	3.23×10^{-12}	0.092
Chromosomal instability	2.62×10^{-03}	1.363	NA	NA	1.27×10^{-05}	0.257	NA	NA	3.39×10^{-11}	−0.326
Sensitivity of cells	1.52×10^{-06}	2.707	7.74×10^{-04}	0.752	1.33×10^{-04}	−0.794	NA	NA	1.84×10^{-08}	−2.185
Cell death of fibroblast cell lines	3.03×10^{-03}	−0.924	6.84×10^{-03}	0.505	3.75×10^{-08}	−1.062	NA	NA	7.84×10^{-08}	1.315
Cycling of centrosome	3.44×10^{-03}	−0.514	NA	NA	1.79×10^{-06}	−1.274	NA	NA	8.47×10^{-08}	1.536
Infection by RNA virus	3.37×10^{-03}	0.089	3.76×10^{-06}	2.269	2.85×10^{-05}	−0.034	2.73×10^{-12}	2.354	2.27×10^{-07}	−0.105
Transactivation	1.53×10^{-03}	−1.747	NA	NA	8.56×10^{-06}	−0.023	3.91×10^{-07}	−2.536	1.64×10^{-06}	−0.558
Repair of DNA	1.72×10^{-06}	−1.53	7.52×10^{-03}	−1.417	5.46×10^{-07}	0.91	NA	NA	2.01×10^{-06}	1.298
Replication of virus	1.94×10^{-03}	0.276	7.86×10^{-03}	1.04	1.44×10^{-04}	−1.082	NA	NA	2.28×10^{-06}	−0.11
Renal lesion	5.75×10^{-07}	0.692	5.71×10^{-03}	0.997	1.39×10^{-06}	2.042	3.00×10^{-06}	−0.212	4.21×10^{-06}	−0.624
Transcription of RNA	4.24×10^{-05}	−1.745	NA	NA	7.73×10^{-09}	2.774	7.38×10^{-08}	−0.693	6.98×10^{-06}	−0.934
Infection of cells	4.06×10^{-03}	−0.611	5.82×10^{-06}	2.416	1.54×10^{-05}	0.119	2.71×10^{-05}	2.711	7.91×10^{-06}	−0.245

The values indicate p values and z-scores for each comparison. Significantly altered pathways in the normal mucosa of LS CRC patients and their CRCs are presented. The list is sorted by the lowest p values in the baseline comparison (LS mucUA—SP mucUA). mucADE, normal mucosa from LS patient with adenoma; mucCRC, normal mucosa from LS patient with colorectal cancer; SP, sporadic; ADE, adenoma; CRC, colorectal cancer; UA, unaffected. [a]Fisher's exact test.

Table 4: **IPA pathways in normal mucosa of unaffected LS patients, LS patients with adenoma or CRC, and tumours.**

and Extension of telomeres were significantly downregulated in both sample groups. Interestingly, Mismatch Repair (MMR)-related pathways were not significantly altered in LS mucADE or LS mucCRC as compared to LS mucUA (Supplementary Table S8), indicating no substantial disruption of the MMR mechanism in the mucosa of LS carriers who had developed tumours. However, DNA double-strand break (DSB) repair was significantly downregulated in LS mucADE and LS mucCRC, and PCNA-dependent long patch base excision repair and Transcription-coupled nucleotide excision repair (TC-NER) were downregulated in LS mucCRC but not in LS mucADE.

In the IPA pathway analysis, a total of 231 out of 500 enriched functions were found to be associated with cancer in LS mucADE, while LS mucCRC exhibited 255 out 0f 500 enriched functions related to cancer (Supplementary Table S9). Again, to avoid redundancy, we excluded additional cancer-related biological functions, except for the most significant one, Cancer, which was also the most altered annotation in LS mucADE ($p <$ Moreover, with $p < 0.05$ metabolism 0.0001 Fisher's Exact Test, BFB 2.12×10^{51}) and LS mucCRC ($p < 0.0001$ Fisher's Exact Test, BFB 1.64×10^{63}) (Table 4, Supplementary Table S9). Moreover, with $p < 0.05$ (Fisher's Exact Test) metabolism of protein and DNA repair were significantly altered in LS mucADE and LS mucCRC. IPA also predicted increased Autophagy in both sample groups (Table 4, Supplementary Table S9) indicating cellular stress in the affected mucosa. Furthermore, Transcription and DNA damage response of cells were altered in LS mucCRC. Interestingly, IPA predicted increased Chromosomal instability (positive z-score) and decreased Cycling of centrosome (negative z-score) in LS mucCRC compared to LS mucUA suggesting enhanced functional aberration along with tumour diagnosis (Table 4, Supplementary Table S9).

The Gene Ontology and pathway analysis imply chromosomal instability in LS tumours

The enrichment analysis of gene ontologies revealed that a considerable portion of the ontologies enriched in LS mucUA were significantly altered in LS adenomas (52%) and carcinomas (70%) (Supplementary Table S7). These alterations were predominantly related to key cellular processes such as cell cycle and its regulation, cell division, metabolism, stress response, and immune system. Regarding the cell cycle, especially, mitotic cell cycle, chromosome segregation, kinetochore organization, spindle organization, and spindle assembly checkpoint were strongly affected in both adenomas and carcinomas. Additionally, DNA metabolic processes including DNA replication and DNA recombination, RNA metabolic processes, response to stress, cellular response to DNA damage stimulus, and immune effector processes showed significant changes in both tumour types (Supplementary Table S7).

Ensembl ID	Gene symbol	LS mucUA—SP mucUA		LS mucADE—LS mucUA		LS mucCRC—LS mucUA		LS ADE—LS mucUA		LS CRC—LS mucUA	
		shrink-T	FDR	shrink-T	FDR	shrink-T	FDR	shrink-T	FDR	shrink-T	FDR[a]
ENSG00000112742	TTK	−2.63	9.04×10^{-03}	−1.34	1.78×10^{-01}	−3.08	2.00×10^{-03}	2.93	4.10×10^{-03}	8.44	1.85×10^{-13}
ENSG00000184445	KNTC1	−2.48	1.06×10^{-02}	−1.34	1.66×10^{-01}	−2.25	1.80×10^{-02}	0.99	3.53×10^{-01}	4.59	3.14×10^{-05}
ENSG00000164109	MAD2L1	−2.21	2.52×10^{-02}	−0.46	6.60×10^{-01}	−3.4	7.00×10^{-04}	3.36	1.10×10^{-03}	10.21	0
ENSG00000159055	MIS18A	−2.12	2.66×10^{-02}	−0.49	6.27×10^{-01}	−2.49	1.09×10^{-02}	2.9	4.40×10^{-03}	10.45	9.29×10^{-20}
ENSG00000071539	TRIP13	−1.94	7.20×10^{-02}	−1.43	1.66×10^{-01}	−3.14	1.50×10^{-03}	4.59	1.24×10^{-05}	12.21	0
ENSG00000088325	TPX2	−1.68	1.09×10^{-01}	−1.48	1.46×10^{-01}	−3.63	4.00×10^{-04}	2.92	3.50×10^{-03}	9	9.07×10^{-15}
ENSG00000117724	CENPF	−1.67	1.14×10^{-01}	−2.79	3.00×10^{-03}	−3.01	2.80×10^{-03}	0.94	3.81×10^{-01}	7.38	1.70×10^{-10}
ENSG00000138778	CENPE	−1.65	1.19×10^{-01}	−2.66	5.40×10^{-03}	−2.74	6.10×10^{-03}	0.86	4.41×10^{-01}	7.76	1.70×10^{-11}
ENSG00000117399	CDC20	−1.41	1.86×10^{-01}	−0.77	4.69×10^{-01}	−3.49	5.00×10^{-04}	3.81	2.00×10^{-04}	8.9	0
ENSG00000178999	AURKB	−0.7	5.23×10^{-01}	−1.28	2.00×10^{-01}	−3.53	6.00×10^{-04}	3.28	1.10×10^{-03}	8.17	0
ENSG00000166851	PLK1	−0.64	5.39×10^{-01}	−1.31	2.00×10^{-01}	−3.51	6.00×10^{-04}	4.16	1.00×10^{-04}	9.78	0
ENSG00000136811	ODF2	0.06	9.64×10^{-01}	−1.35	1.92×10^{-01}	−2.57	9.50×10^{-03}	5.47	2.17×10^{-07}	13.32	0

The values indicate FDR values and shrink-T scores for each comparison. mucADE, normal mucosa from LS patient with adenoma; mucCRC, normal mucosa from LS patient with colorectal cancer; SP, sporadic; ADE, adenoma; CRC, colorectal cancer; UA, unaffected. [a]FDR-value < 2.3E-308 marked as 0.

Table 5: **Expression values of SAC genes in LS colonic samples.**

Of the Reactome pathways enriched in LS mucUA, 65% and 76% were significantly altered in LS adenomas and carcinomas, respectively (Supplementary Table S8) when FDR < 0.05 was used. The Reactome pathway analysis further highlighted the significance of cell cycle-related functions, such as mitotic cell cycle and cell cycle checkpoints, which were enriched in both adenomas and carcinomas (Supplementary Table S8). DNA strand elongation and DNA repair were upregulated in both tumour types, along with increased RNA metabolism, including rRNA processing. Chromosome maintenance and extension of telomeres among chromosomal functions were also significantly upregulated.

In the IPA pathway analysis, various cancer-related functions were again overrepresented as altogether 245/500 and 266/500 enriched functions in adenomas and carcinomas, respectively, were related to cancer (data not shown). As in previous comparisons, we omitted the redundant cancer related biological functions except for the most significant function, Cancer, which was also the most altered annotation in LS adenomas ($p < 0.0001$ Fisher's Exact Test, BFB 3.49×10^{76}) and carcinomas ($p < 0.0001$ Fisher's Exact Test, BFB 2.74×10^{77}). DNA repair was increased in carcinomas, and Transcription of RNA decreased both in adenomas and carcinomas. Importantly, IPA suggested alterations in chromosomal instability and cycling of centrosome (Table 4, Supplementary Table S9).

Altered spindle assembly checkpoint genes in colonic samples

The mitotic cell cycle and spindle assembly checkpoint were strongly altered in LS mucUA as compared to SP mucUA, and the effect was further enhanced in LS mucCRC as well as in the respective CRCs (Table 5). In our previous study[12] with *Mlh1*$^{+/-}$ mouse model we found a field defect of 10 chromosome segregation and SAC related genes (*Mlh1, Bub1, Mis18a, Tpx2, Rad9a, Pms2, Cenpe, Ncapd3, Odf2* and *Dclre1b*) in the colonic mucosa of mice with CRC. Here, we studied the expression changes of these candidate genes in the colonic samples of LS carriers. Of those *MIS18A* (MIS18 kinetochore protein A), *CENPE* (centromere protein E), *TPX2* (TPX2 microtubule nucleation factor) and *ODF2* (outer dense fiber of sperm tails 2) were downregulated in LS mucCRC with FDR < 0.05. Interestingly, *MIS18A* was downregulated already in the colonic mucosa of unaffected LS carriers suggesting an early role in the generation of chromosomal instability.

Genomic instability assayed by MSI analysis and mitosis perimeter measurement

MSI results

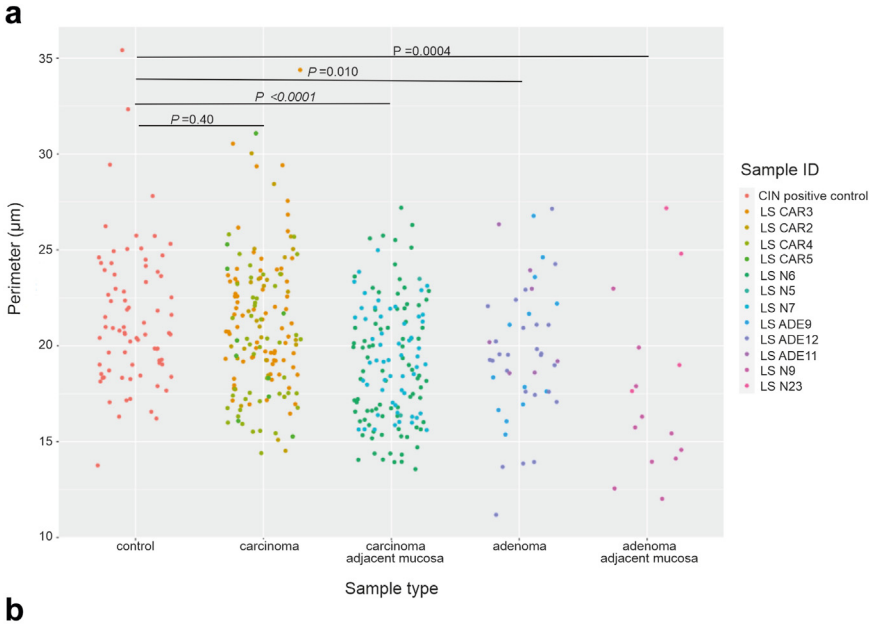
All carcinomas examined in our sample series (n = 5) showed MSI (Supplementary Table S1) and negative expression of the corresponding MMR protein by immunohistochemistry. However, five of 11 (45%) LS adenomas were MSS: two from *MSH6*, and three from *MLH1* mutation carriers (Supplementary Table S1).

Mean perimeters of metaphases increase along with malignancy

Archival tumour material was used for mitosis perimeter analyses as described in Materials and Methods. The perimeters were measured blindly for each identified metaphase (viewed from above) in LS mucADE (n = 2), LS mucCRC (n = 3), in LS adenoma (n = 3), and in LS CRC (n = 4) samples. A previously studied MMR proficient and CIN positive CRC sample was used as a positive control (case no. 56:2).[30]

Here, the average size of perimeters of metaphases increased significantly along with increasing malignancy being the highest in carcinomas and lowest in LS mucADE (Fig. 2). The mean perimeter of the CIN positive CRC sample (21.45 μm) and LS CRC samples (21.03 μm) was almost the same (Fig. 2). Representative

a



b

Sample type	Metaphases (n)	Mean (µm [SD])	Range (µm)	Mean difference (95% CI)	*P*
CIN positive control	76	21.46 (0.41)	(13.76-35.41)		
LS carcinoma	157	21.03 (0.29)	(14.52-34.38)	0.43 (-0.56-1.41)	0.40
Normal mucosa from LS carcinoma patients	143	19.11 (0.26)	(13.57-27.19)	2.35 (1.43-3.26)	<0.0001
LS adenoma	46	19.72 (0.51)	(15.43-27.17)	1.73 (0.42-3.05)	0.010
Normal mucosa from LS adenoma patients	15	17.6 (1.16)	(12.02-27.17)	3.85 (1.74-5.96)	0.0004

Fig. 2: The size distribution of metaphases in different sample groups. (a) A scatter plot showing the size distribution of metaphase perimeters in different sample groups. The CIN positive control and LS carcinomas exhibit nonsignificant difference in their size distributions of metaphases (p 0.397). (b)The number and average size of the metaphases increases along with malignancy being highest in LS carcinomas and lowest in the adenoma affected LS colonic mucosa. The p values represent the difference in mean perimeter of samples in comparison to CIN positive control. Statistical analyses used a t-test.

images of metaphases are shown in Supplementary Fig. S4.

Discussion

This study aimed to identify precancerous changes in macroscopically normal colonic mucosa that could facilitate the monitoring and prevention of cancer development in LS. To determine an LS specific baseline, we first compared the gene expression profiles of colonic mucosa in unaffected LS carriers (LS mucUA), and unaffected sporadic cases (SP mucUA) with no history of colon tumours or Lynch syndrome.

DGE analysis clearly revealed distinct clustering of the two unaffected sample groups (Fig. 1). Considering the total number of differentially expressed genes, the expression profile of unaffected LS mucosa differs more from that of unaffected sporadic mucosa than from tumour-affected LS mucosa. In the baseline comparison, 1564 genes were differentially expressed in the unaffected LS mucosa with $FDR < 0.05$. This represents approximately 17- and 5-fold difference in the number of DEGs when LS mucUA was compared to LS mucADE (1564/91 = 17.18) and LS mucCRC (1564/302 = 5.17), respectively. This suggests that LS mutation carriers have a robust, transcriptome-wide field defect in

their mucosa even before tumour development. The top genes with the largest absolute shrink-T scores (Supplementary Table S10) have roles in the immune system, metabolism, cell cycle, cellular stress and senescence, gene expression, mitochondrial function and DNA repair, reflecting the wide range of affected cellular functions in LS mucosa and tumours.[36–41]

Several cell functions that were already found to be abnormal in colonic mucosa of tumour-free LS carriers were enhanced in tumour-affected LS mucosa (Supplementary Tables S7, S9, and S11). In addition to extremely small p-values, the associated very large Bayes factors indicated robust evidence in favour of these pathway alterations being genuine rather than arising by chance (Supplementary Tables S4, S6, and S9). Mitotic cell cycle and cell cycle checkpoints were the most altered functions in mucosa from both tumour-free and tumour-affected patients with LS, and in LS tumours. According to GOrilla and Reactome, all the three major cell cycle checkpoints (G1/S, G2/M, and SAC) were downregulated in LS mucUA, and the effect was further enhanced in LS mucCRC (Supplementary Tables S7 and S8). The most altered checkpoint was spindle assembly checkpoint, SAC, which controls the proper attachment of metaphase chromosomes to the mitotic spindle apparatus. The malfunction of SAC genes allows cells to transit through mitosis prematurely, which inevitably leads to chromosomal instability and aneuploidy.[42,43] Accordingly, IPA pathway analysis predicted increased CIN in both LS mucUA and LS mucCRC (Supplementary Table S9).

Our results, suggesting increased occurrence of mitotic abnormalities in colonic mucosa, are consistent and confirm our previous study with *Mlh1*$^{+/-}$ mouse model, where we found a cancer preceding field defect of downregulated SAC genes (n = 10) that predisposed to CRC.[12] Of those, *MIS18A*, *CENPE*, *TPX2* and *ODF2* were significantly downregulated in LS mucCRC (Table 5) and *MIS18A* already in LS mucUA (Table 5). MIS18A has a crucial role in regulating CENP-A chromatin assembly, chromosome segregation, and epigenetic control of centromeric chromatin, and its shortage leads to severe defects in chromosome segregation[44] strongly suggesting chromosomal segregation defect in LS mucosa. The other two SAC related genes from the mouse study, *TPX2* and *ODF2,* are both involved in microtubule related processes in spindle formation.[45] *ODF2* is fundamental for the proper structure of centrioles[46] and its depletion causes the formation of tripolar spindles.[47] *TPX2* is also essential for the microtubule organization in mitotic spindle, and in centrosome maturation.[48,49] The expression level of *TPX2* is tightly regulated during mitosis, and abnormal expression, downregulation, and upregulation, are both linked to CIN and tumourigenesis.[47,50] When abnormally upregulated, *TPX2* promotes colon tumourigenesis, and metastasis.[51,52] Here, it was significantly upregulated in carcinomas as compared to LS mucUA (Table 5) reflecting abnormal chromosome segregation and increased oncogenic alterations in the carcinomas. CENPE, a kinesin-like motor protein, is an efficient stabilizer of microtubule capture at kinetochores and hence essential for chromosome alignment.[53,54] It is involved in the movement of chromosomes toward the metaphase plate during mitosis, and necessary for the mitotic checkpoint signal at the kinetochore to prevent chromosome loss.[54,55] The depletion of CENPE leads to mitotic catastrophe and cell death.[56] Yet also the upregulation of CENPE has been reported for several cancer types.[57–60] Here it was significantly downregulated in LS mucCRC and upregulated in the respective carcinomas (Table 5).

SAC-related genes *AURKB*, *CDC20*, *CENPF*, *KNTC1*, *MAD2L1*, *PLK1*, *TRIP13* and *TTK* were downregulated in LS mucCRC and upregulated in the respective carcinomas. Of those, *KNTC1*, *TTK* and *MAD2L1* were already downregulated in LS mucUA as compared to SP mucUA (Table 5), suggesting an early role in the development of CIN. Downregulation of SAC genes may result from tumour suppressor gene inactivation, epigenetic modifications, or dysregulation of signalling pathways in colonic mucosa. Many SAC genes being upregulated in the carcinomas reflect increased proliferation rates and oncogenic alterations in tumour cells, causing overexpression of mitotic regulators in carcinomas compared to the adjacent colonic mucosa.[50,61] Along with downregulation, also upregulation of mitotic genes impairs precise chromosome segregation in neoplastic cells[50,62] hence contributing to tumour progression.

The correct distribution and number of centrosomes is crucial for normal chromosome distribution, and an abnormal number of centrosomes can cause mitotic chromosome distribution defects and aneuploidy.[63] MMR proteins MLH3 and MSH2 are involved in the regulation of chromosome distribution by controlling the number of centrosomes.[64,65] Our findings, mostly derived from *MLH1* mutation carriers, suggest similar role for MLH1. According to GOrilla, the regulation of centriole replication, centrosome duplication, and centrosome cycle regulation were all altered (gene expression decreased) in LS mucUA (Supplementary Table S2). Further, IPA predicted decreased cycling of centrosomes in LS mucUA and during the development of cancer since the phenomenon was further increased in LS mucCRC (Supplementary Table S6). When centrosome amplification is accompanied with defective cell cycle checkpoints (Supplementary Table S9), the potential for malignant growth increases.[66] As known that the MLH1-mediated G2/M arrest and subsequent apoptosis require full complement of the protein67, downregulation of centrosome cycle in LS mucosa suggests a natural effort to maintain cellular homeostasis and counteract CIN. Additional mutations or

epigenetic aberrations may lead to a failure in controlling centrosome number, further increasing the tendency to aneuploidy and malignant process.[63]

Reactome indicated downregulation of MMR in LS mucUA compared to SP mucUA (Supplementary Table S4), which is explained by the inherited heterozygote dysfunction of the MMR mechanism in LS. Although, a second hit of the MMR gene is traditionally thought to be a prerequisite for cancer development in LS and although several studies link the existence of MMR-deficient colorectal crypts in histologically normal colonic mucosa to LS carrier status,[68,69] little information is available about the precise molecular alterations in these crypts or the associated clinicopathological correlates. In a recent study,[10] whole genome sequencing analysis of 132 non-neoplastic crypts and glands from LS individuals resulted in the detection of a single MMR-deficient crypt; this crypt lacked LOH or other plausible second hit, whereas copy number variants indicative of CIN were present in 15.2% of LS crypts.[10] In our study, MMR-related pathways were not significantly altered in LS mucADE or LS mucCRC when compared to LS mucUA. However, we did observe a significant overarching problem in DNA repair, specifically concerning double-strand break (DSB) repair. This issue is prevalent in both LS mucUA, LS mucADE, and LS mucCRC (Supplementary Tables S2, S4, S6, and S7–S9).

Recent findings suggest that different LS susceptibility genes have significant differences in the associated carcinogenesis pathways, both molecularly and morphologically.[70] Furthermore, it is known that the most common somatic second hit in *MLH1* LS is LOH.[71,72] Our findings indicate that *MLH1* haploinsufficiency may induce mitotic abnormalities by impacting the formation of a functional mitotic spindle, disrupting the proper regulation of spindle assembly checkpoint, and impairing the repair of DNA double-strand breaks (Table 3, Supplementary Tables S2, S4, S6–S10). As MLH1 protein has been shown to play a role in DSB repair[73–76] the defective DSB repair and possibly also increased DSB propensity in *MLH1* heterozygous cells may therefore explain the marked increase in LOH[76,77] and the loss of the wild-type MMR gene allele which may eventually lead to dMMR and MSI. Here, the MSI status of LS adenomas and carcinomas was investigated using two mononucleotide repeat markers BAT25 and BAT26. Among adenomas, 45% were identified as MSS, and 55% were classified as MSI. In contrast, all LS CRCs were determined to be MSI. Markers were selected based on previous studies where they had been shown to function as sensitive and specific markers of high-degree MSI[78–80] and shown high concordance between MSI status and loss of MMR protein expression in Lynch syndrome-associated colorectal tumours.[11,81] Also in our analyses, all tumours identified as unstable according to MSI analysis showed a deficiency of corresponding MMR protein in IHC analysis.

Chromosomal instability manifests on tissue level as atypical mitoses which differ morphologically from normal, balanced mitotic figures.[28] An increase in mitotic perimeter directly indicates CIN, as it reflects irregular mitotic division, often resulting in chromosomal rearrangements and potential genetic abnormalities.[29] Here, the mean perimeter of metaphases was higher in LS mucCRC as compared to LS mucADE and increased along with increasing malignancy of the samples (Fig. 2). Also, morphologically abnormal mitoses with e.g., clumped chromatin but normal perimeter were observed. However, the decision was made to specifically compare the metaphase perimeter as an indication of CIN. Although chromosomal aberrations were not directly assessed at the DNA level and further genomic studies are required to verify our findings, the mean metaphase perimeter of LS CRCs (21.03 μm) was shown to be equal to that of a sigmoid colon carcinoma sample (21.45 μm), which was previously confirmed as MMR proficient and CIN positive by comparative genomic hybridization study.[30] Overall, the results showed that the perimeters of metaphases from adenoma- and carcinoma affected LS mucosa and LS tumours are highly consistent with the RNA-seq data suggesting an increased tendency to CIN with increasing malignancy in LS.

The main strengths in the present study are linked to sample collection and high-quality bioinformatics. Biopsy samples were collected from different stages of tumourigenesis, including colonic mucosa from unaffected patients, patients diagnosed with adenoma, and patients diagnosed with CRC at the time of sampling. Importantly, we were able to obtain mucosal samples from unaffected sporadic individuals without Lynch syndrome. This allowed us to compare events in colonic mucosa between LS carriers and non-carriers prior to any tumour formation. Our study is not without limitations. The mean age of unaffected sporadic cases[57] was 12 years higher than the mean age of unaffected LS mutation carriers[45] at inclusion in our investigation (Table 1) which might reflect tendency to earlier cancer occurrence in Lynch syndrome mutation carriers (~40–60 years) compared to the average population (~70 years). It is possible that the age difference may have some impact on the expression profiles of the respective sample groups. The other main difference between the two sample groups was the sample preservation method which may have had some impact on the expression profiles. In the LS mucUA, two different sample preservation methods (N2 and Rnalater) were used equally often, while in the SP mucUA group, all samples were preserved in RNAlater. However, we successfully corrected the technical variation (batch effect) arising from the preservation method as shown

in Supplementary Fig. S1 suggesting, that the observed differences were true biological signals.

Our results highlight two important findings. First, the primary role of MMR proteins has traditionally been attributed to DNA mismatch repair and thereby maintenance of genomic stability. Based on our results, at least MLH1 seems to play a significant role in the generation of chromosomal instability and aneuploidy. In this study, the majority of sample donors with Lynch syndrome had a mutation in *MLH1*, which is the main susceptibility gene in Finnish LS families. However, these results give reason to continue similar studies in patients with Lynch syndrome who have inherited mutations in other MMR genes. Second, previous studies on the timing of dMMR emergence in the multistage LS tumourigenesis have reached conflicting interpretations. Our results suggest that mitotic abnormalities are an earlier phenomenon than MSI and may be the initial cancer driving abnormality, while MSI accelerates tumour formation. Up to 25% of LS-adenomas and nearly 100% of LS-CRCs have MSI and mutator phenotype,[3] which is logical as wild-type allelic loss and dMMR increase the mutation load up to 100–1000 fold with each cell division.[82] Therefore, it is not unexpected that MSI becomes so common during tumourigenesis. From a clinical perspective, precancer expression profiles may help stratify genetically susceptible individuals according to their risk of developing CRC. The potential value of our findings in cancer prevention is still to be determined by further studies.

Contributors

Conceptualization: M.P. and M.N.
Patient data and samples: J-P.M., K.P., L.R-S., T.L., and A.L.
Resources: L.H., M.N., and P.P.
Experimental research work: M.P., P.B., S.N., J.L., and S.M-N.
Visualization: A.L., J.L., and P.T.
Formal analysis: A.L. and P.T.
Methodology: A.L., P.T., and J.L.
Funding acquisition: M.N., L.H., J-P.M., S.M-N., and P.P.
Supervision: M.P., S.M-N, P.P., and M.N.
Writing—original draft: M.P., M.N., and P.P.
Writing—review and editing: M.P, A.L, P.T., J.L, S.M-N., P.P., and M.N.
Accessed and verified the underlying data: M.P., A.L., P.T., S.M-N., M.N., and P.P.
All authors read and approved the final version of the manuscript.

Data sharing statement

All relevant data supporting this study are included in the manuscript and its supplements. Our IRB approvals do not allow sharing raw sequencing data from the patients. Requests to access additional material and datasets should be submitted to the corresponding author for consideration.

Declaration of interests

Minna Nyström is a member of the board of directors and a shareholder in LS CancerDiag Ltd. No disclosures were reported by the other authors.

Acknowledgements

This work was supported by grants from the Jane and Aatos Erkko Foundation (Päivi Peltomäki, and Minna Nyström), the Academy of Finland (330606 Päivi Peltomäki and 331284 Satu Mäki-Nevala); Cancer Foundation Finland sr (Päivi Peltomäki), and the Sigrid Jusélius Foundation (Päivi Peltomäki). Open access is funded by Helsinki University Library.

We thank Saila Saarinen for expert laboratory assistance and Maija Röntynen for collecting clinical data.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2024.105111.

References

1 Peltomäki P, Nyström M, Mecklin JP, Seppälä TT. Lynch syndrome genetics and clinical implications. *Gastroenterology*. 2023;164(5):783–799.
2 Thompson BA, Spurdle AB, Plazzer JP, et al. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nat Genet*. 2014;46(2):107–115.
3 Ahadova A, Gallon R, Gebert J, et al. Three molecular pathways model colorectal carcinogenesis in Lynch syndrome. *Int J Cancer*. 2018;143(1):139–150.
4 Dominguez-Valentin M, Sampson JR, Seppala TT, et al. Cancer risks by gene, age, and gender in 6350 carriers of pathogenic mismatch repair variants: findings from the prospective lynch syndrome database. *Genet Med*. 2020;22(1):15–25.
5 Moller P, Seppala T, Bernstein I, et al. Cancer incidence and survival in Lynch syndrome patients receiving colonoscopic and gynaecological surveillance: first report from the prospective lynch syndrome database. *Gut*. 2017;66(3):464–472.
6 Edwards KS. Research in Ohio. *Ohio Med*. 1989;85(3):182.
7 Möller P, Seppälä TT, Bernstein I, et al. Cancer risk and survival in path_MMR carriers by gene and gender up to 75 years of age: a report from the prospective lynch syndrome database. *Gut*. 2018;67(7):1306–1316.
8 Ahadova A, Seppala TT, Engel C, et al. The "unnatural" history of colorectal cancer in Lynch syndrome: lessons from colonoscopy surveillance. *Int J Cancer*. 2021;148(4):800–811.
9 Bohaumilitzky L, Kluck K, Huneburg R, et al. The different immune profiles of normal colonic mucosa in cancer-free Lynch syndrome carriers and Lynch syndrome colorectal cancer patients. *Gastroenterology*. 2022;162(3):907–919.e10.
10 Lee BCH, Robinson PS, Coorens THH, et al. Mutational landscape of normal epithelial cells in Lynch Syndrome patients. *Nat Commun*. 2022;13(1):2710.
11 Valo S, Kaur S, Ristimäki A, et al. DNA hypermethylation appears early and shows increased frequency with dysplasia in Lynch syndrome-associated colorectal adenomas and carcinomas. *Clin Epigenetics*. 2015;7(1):71.
12 Pussila M, Toronen P, Einarsdottir E, et al. Mlh1 deficiency in normal mouse colon mucosa associates with chromosomally unstable colon cancer. *Carcinogenesis*. 2018;39(6):788–797.
13 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–2120.
14 Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15–21.
15 Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166–169.
16 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
17 Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010;11(10):R106.
18 Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8(1):118–127.
19 Opgen-Rhein R, Strimmer K. Accurate ranking of differentially expressed genes by a distribution-free shrinkage approach. *Stat Appl Genet Mol Biol*. 2007;6:9.
20 Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*. 2009;10:48.
21 Chicco D, Agapito G. Nine quick tips for pathway enrichment analysis. *PLoS Comput Biol*. 2022;18(8):e1010348.
22 Gillespie M, Jassal B, Stephan R, et al. The reactome pathway knowledgebase 2022. *Nucleic Acids Res*. 2022;50(D1):D687–D692.

23 Yang Z, Li S, Liu H, et al. Identification of key genes and pathways associated with diabetes of the exocrine pancreas. *Medicine (Baltimore)*. 2022;101(34):e29781.
24 Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*. 2011;6(7):e21800.
25 Griss J, Viteri G, Sidiropoulos K, Nguyen V, Fabregat A, Hermjakob H. ReactomeGSA - efficient multi-omics comparative pathway analysis. *Mol Cell Proteom*. 2020;19(12):2115–2125.
26 Wu D, Smyth GK. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res*. 2012;40(17):e133.
27 Kramer A, Green J, Pollard J Jr, Tugendreich S. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics*. 2014;30(4):523–530.
28 Matsuda Y, Aida J, Ishikawa N, Takubo K, Ishiwata T, Arai T. Morphological markers of chromosomal instability. In: *Chromosomal abnormalities - a hallmark manifestation of genomic instability*. IntechOpen; 2017:204.
29 Kesarkar K, Tamgadge A, Peirera T, Tamgadge S, Gotmare S, Kamat P. Evaluation of mitotic figures and cellular and nuclear morphometry of various histopathological grades of oral squamous cell carcinoma: comparative study using crystal violet and feulgen stains. *Sultan Qaboos Univ Med J*. 2018;18(2):e149–e154.
30 Abdel-Rahman WM, Ollikainen M, Kariola R, et al. Comprehensive characterization of HNPCC-related colorectal cancers reveals striking molecular features in families with no germline mismatch repair gene mutations. *Oncogene*. 2005;24(9):1542–1551.
31 Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J R Stat Soc B*. 1995;57(1):289–300.
32 Core analysis setup. Available from: https://qiagen.my.salesforce-sites.com/KnowledgeBase/KnowledgeNavigatorPage?id=kA41i000000L6rMCAS&categoryName=IPA.
33 Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature*. 2019;567(7748):305–307.
34 Sellke T, Bayarri MJ, Berger JO. Calibration of values for testing precise null hypotheses. *Am Stat*. 2001;55(1):62–71.
35 Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
36 Jaberi SA, Cohen A, D'Souza C, et al. Lipocalin-2: structure, function, distribution and role in metabolic disorders. *Biomed Pharmacother*. 2021;142:112002.
37 Kotani T, Akabane S, Takeyasu K, Ueda T, Takeuchi N. Human G-proteins, ObgH1 and Mtg1, associate with the large mitochondrial ribosome subunit and are involved in translation and assembly of respiratory complexes. *Nucleic Acids Res*. 2013;41(6):3713–3722.
38 Pan G, Zhang K, Geng S, et al. PHF14 knockdown causes apoptosis by inducing DNA damage and impairing the activity of the damage response complex in colorectal cancer. *Cancer Lett*. 2022;531:109–123.
39 Tran Q, Park J, Lee H, et al. TMEM39A and human diseases: a brief review. *Toxicol Res*. 2017;33(3):205–209.
40 Xia Y, Liu Y, Yang C, et al. Dominant role of CDKN2B/p15INK4B of 9p21.3 tumor suppressor hub in inhibition of cell-cycle and glycolysis. *Nat Commun*. 2021;12(1):2047.
41 Xie G, Peng Z, Liang J, et al. Zinc finger protein 277 is an intestinal transit-amplifying cell marker and colon cancer oncogene. *JCI Insight*. 2022;7(4):e150894.
42 Klaasen SJ, Kops G. Chromosome inequality: causes and consequences of non-random segregation errors in mitosis and meiosis. *Cells*. 2022;11(22):3564.
43 Cohen J. Sorting out chromosome errors. *Science*. 2002;296(5576):2164–2166.
44 Kim IS, Lee M, Park KC, et al. Roles of Mis18alpha in epigenetic regulation of centromeric chromatin and CENP-A loading. *Mol Cell*. 2012;46(3):260–273.
45 Ishikawa H, Kubo A, Tsukita S, Tsukita S. Odf2-deficient mother centrioles lack distal/subdistal appendages and the ability to generate primary cilia. *Nat Cell Biol*. 2005;7(5):517–524.
46 Kashihara H, Chiba S, Kanno SI, Suzuki K, Yano T, Tsukita S. Cep128 associates with Odf2 to form the subdistal appendage of the centriole. *Gene Cell*. 2019;24(3):231–243.
47 Yang K, Tylkowski MA, Huber D, Contreras CT, Hoyer-Fender S. ODF2 maintains centrosome cohesion by restricting beta-catenin accumulation. *J Cell Sci*. 2018;131(20):jcs220954.
48 De Luca M, Lavia P, Guarguaglini G. A functional interplay between Aurora-A, Plk1 and TPX2 at spindle poles: plk1 controls centrosomal localization of Aurora-A and TPX2 spindle association. *Cell Cycle*. 2006;5(3):296–303.
49 Karsenti E, Vernos I. The mitotic spindle: a self-made machine. *Science*. 2001;294(5542):543–547.
50 Perez de Castro I, Malumbres M. Mitotic stress and chromosomal instability in cancer: the case for TPX2. *Genes Cancer*. 2012;3(11–12):721–730.
51 Wei P, Zhang N, Xu Y, et al. TPX2 is a novel prognostic marker for the growth and metastasis of colon cancer. *J Transl Med*. 2013;11:313.
52 Neumayer G, Belzil C, Gruss OJ, Nguyen MD. TPX2: of spindle assembly, DNA damage response, and cancer. *Cell Mol Life Sci*. 2014;71(16):3027–3047.
53 Putkey FR, Cramer T, Morphew MK, et al. Unstable kinetochore-microtubule capture and chromosomal instability following deletion of CENP-E. *Dev Cell*. 2002;3(3):351–365.
54 She ZY, Yu KW, Zhong N, et al. Kinesin-7 CENP-E regulates chromosome alignment and genome stability of spermatogenic cells. *Cell Death Discov*. 2020;6:25.
55 Weaver BA, Bonday ZQ, Putkey FR, Kops GJ, Silk AD, Cleveland DW. Centromere-associated protein-E is essential for the mammalian mitotic checkpoint to prevent aneuploidy due to single chromosome loss. *J Cell Biol*. 2003;162(4):551–563.
56 Iegiani G, Gai M, Di Cunto F, Pallavicini G. CENPE inhibition leads to mitotic catastrophe and DNA damage in medulloblastoma cells. *Cancers (Basel)*. 2021;13(5):1028.
57 El-Arabey AA, Salama SA, Abd-Allah AR. CENP-E as a target for cancer therapy: where are we now? *Life Sci*. 2018;208:192–200.
58 Kung PP, Martinez R, Zhu Z, et al. Chemogenetic evaluation of the mitotic kinesin CENP-E reveals a critical role in triple-negative breast cancer. *Mol Cancer Ther*. 2014;13(8):2104–2115.
59 Balamuth NJ, Wood A, Wang Q, et al. Serial transcriptome analysis and cross-species integration identifies centromere-associated protein E as a novel neuroblastoma target. *Cancer Res*. 2010;70(7):2749–2758.
60 Liang Y, Ahmed M, Guo H, et al. LSD1-Mediated epigenetic reprogramming drives CENPE expression and prostate cancer progression. *Cancer Res*. 2017;77(20):5479–5490.
61 Cunningham CE, Li S, Vizeacoumar FS, et al. Therapeutic relevance of the protein phosphatase 2A in cancer. *Oncotarget*. 2016;7(38):61544–61561.
62 Nath S, Ghatak D, Das P, Roychoudhury S. Transcriptional control of mitosis: deregulation and cancer. *Front Endocrinol (Lausanne)*. 2015;6:60.
63 Lingle WL, Lukasiewicz K, Salisbury JL. Deregulation of the centrosome cycle and the origin of chromosomal instability in cancer. *Adv Exp Med Biol*. 2005;570:393–421.
64 Campbell MR, Wang Y, Andrew SE, Liu Y. Msh2 deficiency leads to chromosomal abnormalities, centrosome amplification, and telomere capping defect. *Oncogene*. 2006;25(17):2531–2536.
65 Roesner LM, Mielke C, Faehnrich S, et al. Localization of MLH3 at the centrosomes. *Int J Mol Sci*. 2014;15(8):13932–13937.
66 Loffler H, Lukas J, Bartek J, Kramer A. Structure meets function–centrosomes, genome maintenance and the DNA damage response. *Exp Cell Res*. 2006;312(14):2633–2640.
67 Cejka P, Stojic L, Mojas N, et al. Methylation-induced G(2)/M arrest requires a full complement of the mismatch repair protein hMLH1. *EMBO J*. 2003;22(9):2245–2254.
68 Brand RE, Dudley B, Karloski E, et al. Detection of DNA mismatch repair deficient crypts in random colonoscopic biopsies identifies Lynch syndrome patients. *Fam Cancer*. 2020;19(2):169–175.
69 Staffa L, Echterdiek F, Nelius N, et al. Mismatch repair-deficient crypt foci in Lynch syndrome–molecular alterations and association with clinical parameters. *PLoS One*. 2015;10(3):e0121980.
70 Engel C, Ahadova A, Seppala TT, et al. Associations of pathogenic variants in MLH1, MSH2, and MSH6 with risk of colorectal adenomas and tumors and with somatic mutations in patients with Lynch syndrome. *Gastroenterology*. 2020;158(5):1326–1333.
71 Zhang J, Lindroos A, Ollila S, et al. Gene conversion is a frequent mechanism of inactivation of the wild-type allele in cancers from MLH1/MSH2 deletion carriers. *Cancer Res*. 2006;66(2):659–664.
72 Ahadova A, Stenzinger A, Seppala T, et al. A "Two-in-One hit" model of shortcut carcinogenesis in MLH1 Lynch syndrome carriers. *Gastroenterology*. 2023;165(1):267–270.e4.
73 Chahwan R, van Oers JM, Avdievich E, et al. The ATPase activity of MLH1 is required to orchestrate DNA double-strand breaks and end processing during class switch recombination. *J Exp Med*. 2012;209(4):671–678.

74 Hoffmann ER, Eriksson E, Herbert BJ, Borts RH. MLH1 and MSH2 promote the symmetry of double-strand break repair events at the HIS4 hotspot in Saccharomyces cerevisiae. *Genetics*. 2005;169(3):1291–1303.
75 Bannister LA, Waldman BC, Waldman AS. Modulation of error-prone double-strand break repair in mammalian chromosomes by DNA mismatch repair protein Mlh1. *DNA Repair*. 2004;3(5):465–474.
76 Guan J, Lu C, Jin Q, et al. MLH1 deficiency-triggered DNA hyperexcision by exonuclease 1 activates the cGAS-STING pathway. *Cancer Cell*. 2021;39(1):109–121.e5.
77 Moynahan ME, Jasin M. Loss of heterozygosity induced by a chromosomal double-strand break. *Proc Natl Acad Sci USA*. 1997;94(17):8988–8993.
78 Brennetot C, Buhard O, Jourdan F, Flejou JF, Duval A, Hamelin R. Mononucleotide repeats BAT-26 and BAT-25 accurately detect MSI-H tumors and predict tumor content: implications for population screening. *Int J Cancer*. 2005;113(3):446–450.
79 Esemuede I, Forslund A, Khan SA, et al. Improved testing for microsatellite instability in colorectal cancer using a simplified 3-marker assay. *Ann Surg Oncol*. 2010;17(12):3370–3378.
80 Loukola A, Eklin K, Laiho P, et al. Microsatellite marker analysis in screening for hereditary nonpolyposis colorectal cancer (HNPCC). *Cancer Res*. 2001;61(11):4545–4549.
81 Maki-Nevala S, Valo S, Ristimaki A, et al. DNA methylation changes and somatic mutations as tumorigenic events in Lynch syndrome-associated adenomas retaining mismatch repair protein expression. *EBioMedicine*. 2019;39:280–291.
82 Modrich P. Mechanisms and biological effects of mismatch repair. *Annu Rev Genet*. 1991;25:229–253.