MARJAANA PUSSILA

CANCER-PRECEDING GENE EXPRESSION CHANGES IN MOUSE COLON MUCOSA

DIVISION OF GENETICS, DEPARTMENT OF BIOSCIENCES
FACULTY OF BIOLOGICAL AND ENVIRONMENTAL SCIENCES

INTEGRATIVE LIFE SCIENCE DOCTORAL PROGRAM
UNIVERSITY OF HELSINKI
HELSINKI, FINLAND

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

ACADEMIC DISSERTATION

To be presented for public examination with the permission of the Faculty of Biological and Environmental Sciences of the University of Helsinki
in Lecture Hall 2041, Viikinkaari 5, Helsinki
on the 16 of June 2017, at 12 noon.

Supervisors

Professor Minna Nyström, PhD
Department of Biosciences
Faculty of Biological and Environmental Sciences
University of Helsinki, Finland

Laura Sarantaus, PhD
Laboratory of Genetics
Helsinki University Hospital, Finland

and

Department of Biosciences
Faculty of Biological and Environmental Sciences
University of Helsinki, Finland

Reviewers

Adjunct Professor Minna Pöyhönen, MD, PhD
Department of Clinical Genetics
Helsinki University Hospital, Finland

and

Department of Medical and Clinical Genetics
Faculty of Medicine
University of Helsinki, Finland

Professor Suvi M. Virtanen, MD, PhD
Department of Public Health Solutions
National Institute for Health and Welfare, Finland

and

Faculty of Social Sciences
University of Tampere, Finland

Opponent

Professor Theodore Fotsis, MD, PhD
Foundation for Research & Technology-Hellas (FORTH)
Institute of Molecular Biology and Biotechnology (IMBB)
Department of Biomedical Research (Ioannina)

and

Laboratory of Biological Chemistry
Medical Faculty, School of Health Sciences
University of Ioannina, Greece

Custos Professor Minna Nyström

Thesis committee Professor Hannes Lohi, PhD
Department of Veterinary Biosciences
Faculty of Veterinary Medicine
University of Helsinki, Finland

and

Research Program for Molecular Neurology
Faculty of Medicine
University of Helsinki, Finland

Docent Minna Pöyhönen, MD, PhD

Cover layout by Anita Tienhaara
Cover image: Mouse and flute. Anni Moilanen 2016

Dissertationes Scholae Doctoralis Ad Sanitatem Investigandam Universitatis Helsinkiensis
ISBN 978-951-51-3501-8 (paperback)
ISBN 978-951-51-3502-5 (PDF)
ISSN 2342-3161 (paperback)
ISSN 2342-317X (PDF)
Unigrafia, Helsinki 2017

"Imagination is the beginning of creation"

George Bernard Shaw

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

I. **Pussila M***, Sarantaus L*, Dermadi Bebek D, Valo S, Reyhani N, Ollila S, Päivärinta E, Peltomäki P, Mutanen M, Nyström M. 2013, *PloS One*. Cancer-predicting gene expression changes in colonic mucosa of Western diet fed *Mlh1*$^{+/-}$ mice. *Equal contribution

II. **Pussila M**, Törönen P, Einarsdottir E, Katayama S, Krjutškov K, Holm L, Kere J, Peltomäki P, Mäkinen M, Linden J, Nyström M. 2017. *Mlh1* deficiency in normal mucosa associated with microsatellite stable colon cancer, *submitted*.

* Equal contribution

Study I: M.P. participated in designing the study, preparation of samples (e.g. mice operations, removal of the gut tissue and separation of colon mucosa), DNA methylation analysis (amplicon design, preparation of DNA samples for methylation analysis, and data analysis with Chipster software), and writing the manuscript. **M.P conducted** the extraction of RNA and DNA from colon mucosa, run the StellARray qPCR plates and analyzed the results with GPR-software, run the qPCR with TaqMan assays, and analyzed the results with Data-assist v2.0 software.

Study II: M.P. participated in designing the study, preparation of samples (e.g. mice operations, removal of the gut tissue and separation of colon mucosa), transcriptome data analysis, and writing the manuscript. **M.P conducted** the extraction of RNA and DNA (colon mucosa and tumors), pathway analysis (Ingenuity Pathway Analysis), data analysis, MSI, LOH, and immunohistochemical analyses.

The publications are referred to in the text by Roman numerals I-II

ABBREVIATIONS

ACF	Aberrant crypt foci
AFAP	Attenuated Familial Adenomatous Polyposis
AIN	AIN-93, American Institute of Nutrition, purified diet for laboratory rodents
APC	Adenomatous polyposis coli
BMI	Body mass index
bp	base pair
CGI	CpG island
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
CRC	Colorectal cancer
DEA	Differential expression analysis
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
FAP	Familial Adenomatous Polyposis
FFPE	Formalin fixed paraffin embedded
GPR	Global pattern recognition
HCL	Hydrogen chloride
HNPCC	Hereditary non-polyposis colorectal cancer
IDL	Insertion/deletion loop
IHC	Immunohistochemistry
IPA	Ingenuity pathway analysis
JPS	Juvenile Polyposis Syndrome
LOH	Loss of heterozygosity
LS	Lynch syndrome
MAP	MUTYH-associated polyposis
MDS	Multidimensional scaling
Min	Multiple intestinal neoplasia
MLH	MutL homologues
MMR	mismatch repair
mRNA	messenger RNA
MSH	MutS homologues
MSI	Microsatellite instability
MSS	Microsatellite stable
NMDS	Non-metric multi-dimensional scaling
OMIM	Online Mendelian Inheritance in Man
PJS	Peutz-Jeghers syndrome
PMS	Post meiotic segregation increased
qPCR	Quantitative polymerase chain reaction
RNA	Ribonucleic acid
RT	Reverse transcriptase

SAC	Spindle assembly checkpoint
seq	sequencing
STRT	Single-cell tagged reverse transcription
TSG	Tumor suppressor gene
WD	Western-style diet
WT	Wild-type

ABSTRACT

Colorectal cancer (CRC) is the second most common cause of cancer-related deaths in the Western world and interactions between genetic and environmental factors, including diet, are suggested to play a critical role in its etiology. Yet, the mechanisms by which diet impacts colorectal tumorigenesis remain largely unknown. Colorectal cancer evolves as a multistep process, which requires a series of genetic and epigenetic alterations in growth regulatory genes. The process is accelerated in individuals with inherited cancer predisposition such as Lynch syndrome (LS) which is one of the most common inherited cancer susceptibility syndromes and caused by inherited mutation in one of the DNA mismatch repair (MMR) genes. CRC is thought to develop via the so called adenoma-carcinoma sequence. However, the early events that occur in colon mucosa prior to polyp formation remain unknown.

The research presented here investigates the gene expression changes arising in histologically normal colonic mucosa as putative cancer-preceding events available for early detection. This was achieved by pursuing a long-term feeding experiment in the mouse. In the first study, the expression of 94 growth-regulatory genes previously linked to human CRC was studied at two time points (5 weeks and 12 months of age). The test animals were: heterozygote *Mlh1*$^{+/-}$ (B6.129-Mlh1tm1Rak) mice, an animal model for human Lynch syndrome, and the wild type *Mlh1*$^{+/+}$ littermates, fed with either Western-style (WD) diet containing high amounts of fat and reduced levels of fiber, calcium and vitamin D, or healthy AIN-93G control diet. Promoter CpG island methylation status was also studied for the genes which showed reduced expression. In mice fed for 12 months with WD, proximal colon mucosa, the predominant site of cancer formation in LS, exhibited a significant expression decrease in tumor suppressor genes, *Dkk1, Hoxd1, Slc5a8*, and *Socs1*, the latter two only in the *Mlh1*$^{+/-}$ mice. Furthermore, a reduced mRNA expression was accompanied by an increased CpG dinucleotide promoter methylation of the respective genes suggesting a cause for the mRNA down regulation. The strongest expression decrease together with a significant increase in its promoter methylation was seen in *Dkk1*, an antagonist of the canonical Wnt signaling pathway. Furthermore, the inactivation of *Dkk1* seemed to predispose to neoplasias in the proximal colon, since 4 out of the 6 neoplasms were found in mice

which showed Dkk1 inactivation, suggesting that the inactivation of *Dkk1* is a prominent early marker for colon oncogenesis.

Since no decrease in *Mlh1* expression was seen in the 12-month-old mice in study I, our aim was to comprehensively clarify the role of *Mlh1* expression during colon tumorigenesis, which is usually associated with Lynch syndrome and MSI. Here, the same mouse model and diets were used to study cancer-preceding expression changes in the colon mucosa of 12 and 18-month-old mice. Due to a longer diet experiment (up to 21 mo), more colon carcinomas develop in the mice as compared to study I. This enabled us to study the Mlh1 protein expression and MSI status in their colon carcinomas, and the effect of inherited predisposition ($Mlh1^{+/}$) and Western-style diet on those.

Carcinomas developed mainly in WD fed mice. Due to a low number of tumors in study I, the diet difference gave no statistically significant difference, yet suggesting that WD increases the number and probability of colonic tumors. In study II, at time points 12, 18 and 21 mo, 100%, 80% and 72% of CRCs developed in mice which were fed with WD, and 0%, 20% and 28% in AIN fed mice, respectively, indicating that Western-style diet also accelerates the progression of carcinogenesis.

CRC development always includes a lack of genomic integrity and the different types of genomic instabilities, such as chromosomal instability and MSI are thought to reflect distinct cancer initiating mechanisms. Interestingly, in the present study neither wildtype $Mlh1^{+/+}$ nor heterozygote $Mlh1^{+/-}$ mice lacked the Mlh1 protein or showed MSI in CRCs, while *Mlh1* RNA expression was already significantly decreased in their normal mucosa. Instead, CRC mice showed a distinct expression profile with shortage of *Mlh1* and several other chromosomal segregation gene-specific transcripts (*Bub1, Mis18a, Tpx2, Rad9a, Pms2, Cenpe, Ncapd3, Odf2,* and *Dclre1b*) in mucosa and aberrant mitosis in tumors. The genome wide expression profiling experiment demonstrated that cancer-preceding changes are already seen in histologically normal colon mucosa and a that decreased expression of *Mlh1* together with other chromosomal segregation genes may form a field-defect in mucosa and trigger MMR-proficient, chromosomally unstable CRC.

INTRODUCTION

Colorectal cancer is a significant cause of mortality worldwide being the third most common cancer and the fourth most common cause of cancer-related deaths globally.[1] The incidence rate is increasing in the industrialized world.[1] High CRC incidence is associated with aging, the so called Western lifestyle, and the consumption of Western-style diet which is recognized as a major risk factor for sporadic CRC.[2,3]

CRC evolves as a serial accumulation of genetic and epigenetic alterations during aging. Interactions between genetic and environmental factors, such as diet, seem to be in key position in its etiology[4,5] Yet, the earliest events in normal colon mucosa available for early detection and prevention of cancer development remain to be elucidated. Although inherited mutations in tumor suppressor genes (TSGs) such as *APC* (adenomatous polyposis coli), an important component of the Wnt/β-catenin signaling pathway, and *MLH1*, which controls the mutation rate in a cell[5], may confer a high lifetime risk of cancer with an early age at onset, colorectal cancer is clearly a disease of increasing age.[6,7] Accordingly, methylation changes of a small subset of TSGs have been detected in the aging colonic mucosa of normal healthy individuals, and this methylation involves genes which often become more substantially methylated in neoplastic cells[8-10], suggesting their role in cancer initiation and progression. In addition to aging, some exogenous compounds from dietary sources are important modifiers of methylation patterns in the colon.[11] This probably explains why Western populations consuming considerable amounts of red meat, saturated fat and sugar, and only moderate amounts of dietary fiber, vitamins and minerals (e.g. calcium, folate, and vitamin D), and plant derived nutrients show the highest CRC incidences in the world.[4] Epigenetic changes thus provide a potential link between nutrition and cancer[12] and emphasize the need to elucidate dietary effects on gene regulation in intestinal mucosa.

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, MSI), are thought to reflect distinct initiating mechanisms in cancer.[13] Colon cancer research focuses mainly on tumor characteristics, such as genomic instability, which can be utilized in treatment design. Recent findings however have revealed that CIN and MSI pathways are not mutually exclusive,[14-16] suggesting that also tumors with distinct features and

instabilities may share initiative genomic aberrations, while different tumor characteristics reflect subsequent alterations during cancer development.

Here, a mouse model is employed to study cancer-preceding expression changes in colon mucosa, *Mlh1* phenotype in tumors, and the effect of inherited predisposition (*Mlh1*$^{+/-}$) and Western-style diet on those. A long term feeding experiment was conducted with either a healthy rodent diet AIN-93G or Western-style diet modified from AIN. WD was used to ensure the development of colon carcinomas, since it has previously been shown to cause CRCs in mice even without any predisposing mutation or carcinogen treatment.[17-20] The mouse model provided a valuable tool to study the process of carcinogenesis from the earliest changes in colon mucosa until tumor development and characterization. Moreover, the use of an animal model enabled the detection of gene expression changes caused by different risk factors, such as age, inherited predisposition, and diet, as well as to distinguish those that signal carcinogenesis.

The aim of this thesis was to employ mouse models for LS and sporadic CRC to: 1) study the effects of Western-style diet, ageing, and genetic predisposition on gene expression in histologically normal mouse colon mucosa; and 2) and of those, to identify the expression changes and molecular mechanisms driving the CRC development. The obtained results provide new insights to the factors and mechanisms involved in CRC development by providing information on how diet, genetic predisposition, and aging affect genetic and epigenetic changes in the early stages of carcinogenesis. This opens new opportunities for determining the epidemiology of the disease, the level of risk, and the treatment of the disease, particularly in mutation carriers who already have the inherited predisposition to cancer.

REVIEW OF LITERATURE

COLORECTAL CANCER (CRC)

Epidemiology

Colorectal cancer (CRC) is a significant cause of morbidity and mortality worldwide being the third most common cancer and the fourth most common cause of cancer-related deaths globally and affecting more than 1.2 million new patients annually[1]. CRC is mainly a disease of developed countries and shows large geographic differences in the global distribution. The incidence rate varies up to 10 fold across the world and highest rates are found in North America, Western Europe, Australia, and New Zealand, whereas the lowest are in Africa and Asia[21] .

To date, tumor staging is the most important predictor of clinical outcome for patients with colorectal carcinoma. Tumors are graded according to their invasion to underlying tissues, lymph node involvement, and presence of distant metastases in other organs.[22] CRC survival is highly dependent on the stage of disease at diagnosis. The 5-year survival rate ranges from 90% in patients showing a localized stage tumor, to 70% in patients with regional metastatic cancer, and to 10% for patients with distant metastatic cancer[23]. In high-income countries, overall 5-year survival rate (65%) of CRC patients has improved during the past two decades due to improvements in cancer treatment and screening. In low-income countries, the survival rate is clearly lower, approximately 50%.[24,25] CRC mortality can be significantly reduced if cancers are diagnosed and cured early.[26] Early diagnosis is particularly challenging when there is no proper screening methods and practices, as is the case in many low-income countries. Screening usually aims to identify high-risk individuals with pre-cancerous formation.[27] Only two screening methods, fecal occult blood testing (FOBT) and flexible sigmoidoscopy, have been shown to reduce CRC mortality when used in randomized clinical trials.[28,29] FOBT has been shown to detect most early colorectal cancers and many advanced adenomas, as well as to substantially reduce colorectal cancer incidence and mortality[30], although it has low sensitivity for early polyps[28]. The method is non-invasive, feasible, widely available, and highly acceptable. The combination of annual FOBT with flexible sigmoidoscopy every five years[28] has been shown to be an especially effective screening method,

since the identification of colonic polyps can reduce CRC mortality through earlier diagnosis of cancers and the removal of polyps, the precursor lesions of CRC.[30]

Incidence rates increase significantly with age, although CRC has recently been shown to also be increasing among younger people[31,32]. CRC risk rises sharply after the age of 50 years, although the median age at diagnosis is 70 years.[3] The significant difference in incidence rates across geographic regions suggests the importance of environmental influences on risk to get colon cancer.[33] Epidemiological data shows that high CRC incidence is associated with the so called Western lifestyle, characterized by obesity, sedentary lifestyle, smoking, and high consumption of red meat and alcohol. Furthermore, inverse association is seen between CRC risk and diet rich in vegetables, fiber, dairy, and fish.[2,3]

The majority of CRCs are sporadic with no family history or inherited susceptibility to CRC. A considerable portion, approximately 30%, of all CRC cases are estimated to represent a familial form of the disease with an unusual aggregation of CRC among family members who lack a known inherited CRC syndrome.[5,34,35] Of all CRC cases, 5% are associated with dominantly inherited cancer syndromes caused by highly penetrant inherited mutations. The average lifetime risk for sporadic CRC is 6%. Population studies show a two-fold higher CRC risk in association with family history of one affected first-degree relative as compared with those without a family history.[36] Studies indicate that the risk consists of both hereditary and environmental risk factors in familial cancer.[37] In hereditary cancer syndromes, the cancer risk without any preventive actions may be up to 70-90%[38].

CRC development

The pathogenesis of CRC is a very complex and diverse process and influenced by multiple factors, some of which are related to genetic predisposition, while others are related to diet and lifestyle. To make the picture even more complex, some of these aberrations occur prior to the development of any visible histological growth.[39,40]

The colonic epithelia rapidly renews itself to ensure the proper absorption of nutrients from the lumen contents. Stem cells located at the very bottom of the

colonic crypts serve as a constant source of new daughter cells which differentiate to colonic epithelial cells and migrate towards the epithelial surface where they ultimately undergo apoptosis and shed to the lumen.[41] There is a fine balance between cellular proliferation and apoptosis in colonic epithelia. The disturbed homeostasis and increased proliferation rate or reduced apoptosis leads to a clone of cells that have escaped normal regulation of growth, proliferation, differentiation, and intercellular relationships.

CRC is thought to develop through an adenoma-carcinoma sequence, which is the classical example for stepwise progression of cancer and comprised of multiple well-defined histological stages.[42] The progressive accumulation of genetic and epigenetic alterations that affect genes controlling cell division, apoptosis, and DNA repair first gives a growth advantage to abnormal epithelial cells. This eventually leads to the formation of the first visible aberration in the mucosa, the aberrant crypt focus ACF[43] with dysplasia or hyperplasia, followed by benign and advanced adenomatous polyps or adenomas, a gland-like growths.[44] CRC typically starts from focal changes within precancerous polyps that with time grow in size and develop more severe dysplasia and develop into *in situ* carcinoma. Finally, the cells acquire the ability to invade the bowel wall and create a metastatic carcinoma.[45] While the most common genetic aberrations along the adenoma-carcinoma sequence are well characterized, the early molecular events which predispose to polyp formation and cancer remain unknown.[46]

MOLECULAR PATHWAYS IN COLON TUMORIGENESIS

Colorectal cancer is a heterogeneous disease its key feature being progressive accumulation of mutations and epigenetic alterations which activate oncogenes and inactivate tumor suppressor genes that regulate cell growth. According to the classical adenoma-carcinoma model[47] (Fig. 1) proposed by Fearon and Vogelstein, at least four to five different mutated genes are thought to be required for tumor development and the type or number of genetic alterations rather than their order determines the biological behavior of the tumor.[14]

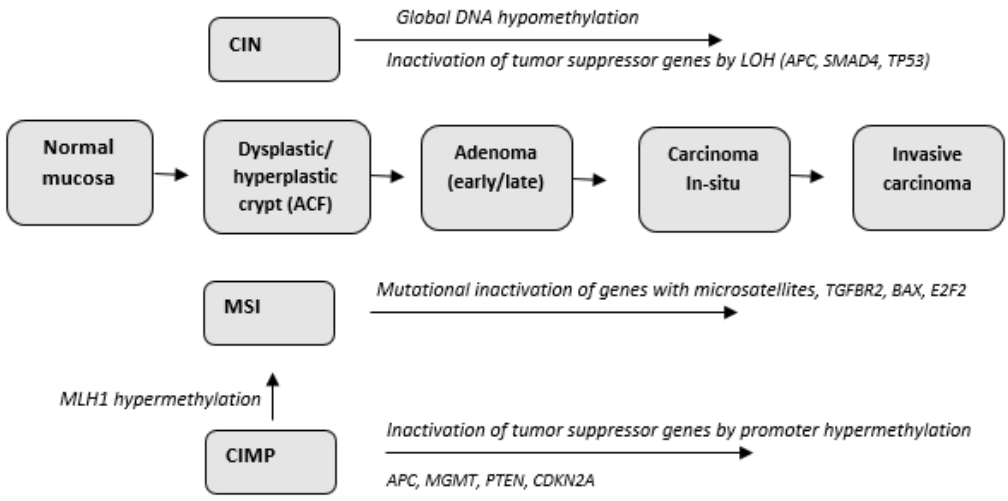

Figure 1. Molecular pathways in adenoma-carcinoma model of CRC development. Modified from Boland *et al* 2009.[48]

CRCs may be subdivided by the types of alterations which accumulate during carcinogenesis. Most CRCs show chromosomal instability (CIN) caused by a series of deletions, duplications, rearrangements and allelic losses. CIN is associated with aneuploidy, the degree of which correlates with the severity of the neoplasia.[49] Other pathways beyond CIN have been discovered: microsatellite instability (MSI, MIN) caused by defective DNA mismatch repair (MMR) mechanism, and the CpG island methylator phenotype (CIMP) associated with methylation of CpG rich regions in gene promoters, which causes silencing of genes.[14] Colon tumors show often both MSI and CIMP phenotype, since methylation of the MMR gene promoters also causes MSI.[50] The majority of microsatellite stable (MSS) tumors follow the CIN pathway of tumorigenesis.[51,52]

Chromosomal instability

Most CRCs (70-85%) follow the chromosomal instability pathway.[15,53] The precise cause of chromosomal instability is unknown[54], but it has been suggested to be a consequence of aberrations in the mitotic checkpoint, centrosome number and function, telomere function, DNA damage response, or loss of heterozygosity (LOH), which lead to large genomic aberrations.[51,55-57] The mitotic checkpoint, also known as the spindle assembly checkpoint (SAC), is the major cell cycle control mechanism that ensures fidelity of chromosome segregation by delaying the onset of anaphase until all pairs of duplicated chromatids are properly aligned on the metaphase plate. Defects in checkpoint signaling lead to chromosome missegregation and subsequent aneuploidy with abnormal numbers of chromosomes being distributed to daughter cells.[58]

CIN is associated with mutations in the *Adenomatous polyposis coli* (*APC*) and other genes that activate the Wnt signaling pathway[51]. *APC* mutation is thought to happen at a very early stage of tumor development and it is found in 80% of all adenomas and carcinomas[59,60]. A number of other key events associated with the development of CIN in CRC have been recognized. These include mutations in tumor suppressor genes and oncogenes such as *BRAF*, *TP53*, *KRAS*, *CTNNB1*, *PIK3CA*, and LOH in chromosome 18q which contains the tumor suppressor genes *SMAD2*, *SMAD4*, and *DCC*.[51,55-57,61,62] *KRAS* and *BRAF* mutations are particularly common events in sporadic CRC development. They belong to the intracellular RAS/RAF/MEK/mitogen-activated protein kinase (MAPK) cascade, which mediates cellular responses to growth signals. *KRAS* mutations are found in up to 50% of carcinomas and advanced adenomas with high grade dysplasia[63]. A single missense mutation of *BRAF* (BRAF V600E) is found in approximately 10% of CRCs.[64] BRAF mutations are common in sporadic colorectal cancers (CRCs) with DNA mismatch repair (MMR) deficiency caused by promoter methylation of the MMR gene *MLH1*, whereas *KRAS* mutations are common in MMR proficient CRCs.[65]

Microsatellite instability

Microsatellite instability represents another major pathway to CRC. Approximately 15% of all CRCs show MSI[53]. Of these, 20–25% represent hereditary cancer, Lynch syndrome (LS), and the rest are sporadic CRCs[66] . MSI is caused by MMR deficiency and is the hallmark of colorectal cancers in Lynch syndrome[67], although the phenomenon was initially detected in sporadic tumors[68]. In sporadic CRC, MSI is thought to accelerate rather than initiate tumorigenesis as opposed to Lynch syndrome where it is the driving cause.[69]

Microsatellites are repetitive DNA sequences including one to six nucleotides. They locate often in the non-coding regions of a genome, are repeated up to 100 times, and are unique and uniform in length in every tissue of a person. Due to their structure, microsatellites are more prone to mutations than the rest of the genome. Cells with MSI show a mutator phenotype with a 100 to 1000 fold increase in mutation rate when compared to normal cells.[70]

MSI is caused by defects in the DNA mismatch repair system. MMR recognizes and repairs single nucleotide mismatches and small insertion/deletion loops that arise at DNA replication and recombination and escape the proofreading of DNA polymerase[15]. Once both MMR gene alleles have been inactivated, the cell's propensity towards acquiring mutations increases, especially in genes carrying microsatellite repeats.[71] MMR malfunction is associated with mutational inactivation of DNA mismatch repair genes such as *mutL homologue 1* (*MLH1*), *mutS homologue 2* (*MSH2*), *mutS homologue 6* (*MSH6*), and *postmeiotic segregation increased 2* (*PMS2*) or hypermethylation induced silencing of the *MLH1* gene.[53,72]

MSI analysis is the primary method to identify tumors suspected to have an underlying MMR mutation.[66] According to the so-called "Bethesda Guidelines"[73], five microsatellite markers are used to identify MSI: three dinucleotide and two mononucleotide repeat regions. A tumor is classified as MSI-high (MSI-H) when two or more markers show new allele size, MSI-low (MSI-L) when one of the five markers has new allele size (fragment lengths different from normal tissue), and MSI-stable when none of the markers shows shift in the allele size.[66]

CpG island methylator phenotype

The third instability pathway in CRC, the CpG Island Methylator Phenotype (CIMP), is seen in approximately 20% of CRCs.[14] CIMP is caused by global hypermethylation in a subset of colon tumors[72]. DNA methylation, a covalent modification of genomic DNA with methyl groups, modifies gene expression and transmits epigenetic information through DNA replication and cell division. The majority of DNA methylation is associated with repetitive DNA sequences where it suppresses the activity and harmful effects of the sequences incorporated to the genome.[74-76] DNA methylation also has a role in several normal and important cellular and organismal functions such as transcriptional regulation, X-chromosome inactivation, imprinting, embryonic development and differentiation, and tissue-specific gene expression.[77-80] Two different types of aberrant methylation patterns are observed in cancer, global hypomethylation which aberrantly activates oncogenes, and hypermethylation which inactivates tumor suppressor genes[81]. CIMP is characterized by a genome-wide hypermethylation of CG-rich areas often located in gene promoters, the CpG islands (CGIs), resulting in the inactivation of several tumor suppressor genes. [7,14,51,82]

A large number of gene promoter CGIs are hypermethylated in the human genome during the normal aging process. This type A (aging) methylation is relatively common in normal colonic cells as well as in primary CRCs. The cancer specific type C (cancer) methylation, on the contrary, is only seen in colonic tumors, and is associated with the CIMP phenotype.[72] No standardized gene panel for CIMP exists, yet a robust panel of five CIMP marker genes, *CACNA1G, IGF2, NEUROG1, RUNX3,* and *SOCS1*, have been used to identify CRCs with high level of promoter methylation, MSI and $BRAF^{V600E}$ mutation.[83] As aberrant methylation happens already in the normal appearing mucosa, it serves as a potential marker for increased CRC risk.[84]

RISK FACTORS FOR CRC

Age

Colorectal cancer is clearly a disease of increasing age, although during the last years CRC has been shown to also be increasing among younger people.[31,32] CRC risk rises sharply after the age of 50 years, and people between 65 and 85 years are six times more likely to develop CRC than people younger than 50.[85] The median age at diagnosis is 70 years.[86,3] Interestingly, older patients present mostly with early-stage disease, whereas younger patients, usually in their 40s, carry a more aggressive form of the disease.[87]

The risk of developing CRC increases with advancing age. The risk is caused both by an age-related increase in somatic genetic aberrations, such as point mutations, single-strand breaks, DNA cross-links, insertions/deletions, oxidative damage, and epigenetic aberrations.[86,88] Furthermore, different DNA repair pathways such as MMR, nucleotide and base excision repair (NER and BER), and double strand brake repair (DSB), become less efficient with age leading to further accumulation of genetic mutations.[89] Along with genetic aberrations aging is also associated with accumulation of epigenetic aberrations such as genome-wide hypomethylation and gene-specific hypermethylation.[90] Growing evidence suggests that epigenetic changes might even play a bigger role than the genetic changes in aging and CRC development and be a major determinant in the origin of the tumor and tumor heterogeneity.[91] In fact, over half of the genetic defects that occur in cancer are epigenetic alterations as compared to genetic mutations.[92] Environmental factors, such as diet, are well known to influence gene expression and cancer development through epigenetic mechanisms.[93] Folate deficiency for example causes altered DNA methylation and histone modifications.[94] This emphasizes the importance of diet-induced accumulation of epigenetic aberrations during aging.[90]

Inherited predisposition

Lynch syndrome

Some individuals appear to be more prone to CRC than others. In fact, a considerable proportion, around 30%, of CRCs seem to have a heritable component.[38,95] Patients with CRC classified as familial have one or more relatives diagnosed with CRC, although germline mutations in known cancer susceptibility genes are found in only 5–6% of these cases.[38]

Lynch syndrome (OMIM 120435), previously known as hereditary non-polyposis colorectal cancer or HNPCC, is the most common inherited colorectal cancer syndrome.[96,97] It accounts for approximately 3-5% of all diagnosed CRC cases.[98] The dominantly inherited syndrome is caused by heterozygote germline mutation in one of the DNA mismatch repair genes, most often *MLH1* or *MSH2* (~90%), and less frequently in *MSH6* or *PMS2*[99]. Due to improper DNA mismatch repair, LS carriers have significantly increased lifetime risk of developing CRC (50-80%)[100] and endometrium cancer (40-60%)[101] but they may also develop cancers of other organs including ovarian, brain, urinary tract, stomach, and pancreas. In Lynch syndrome, the age at onset, cancer risk, and tumor spectrum vary depending on which MMR gene is mutated.[102, 103,104] *MLH1* mutation carriers have a higher incidence of CRC than *MSH2* mutation carriers, which are prone to have more extra-colonic manifestations such as the endometrium (lining of the uterus), ovaries, stomach, small intestine, liver, gallbladder duct, upper urinary tract, and brain. *MSH6* mutation carriers show a lower incidence of CRC but an excess of endometrial cancers as compared to *MLH1* and *MSH2* mutation carriers. The phenotypic consequences of PMS2 mutations are highly variable, often with childhood onset of atypical tumors.[105]

Microsatellite instability in tumors and early age at onset, ~45 years in LS as compared to ~65 years in sporadic CRC, are the common hallmarks of Lynch syndrome. LS tumors are histologically distinct from sporadic ones. They typically

locate at the proximal part of the colon, do not express the protein from the mutated MMR gene and exhibit MSI, lymphocytic infiltrate and mucinous histology with poor differentiation.[106,107] Synchronous and metachronous tumors are frequently observed.[108] Like sporadic ones, LS tumors also develop from colonic polyps, but due to the increased mutation rate, the adenoma-carcinoma sequence progresses much more rapidly (2-3 years) as compared with sporadic CRC (6-10 years). The microsatellite unstable LS tumors have a slightly better prognosis[109] and five-year survival than sporadic MSS CRC and they respond differently to chemotherapy.[53] One explanation for the better prognosis in LS compared to sporadic CRC is suggested to be the tumor infiltrating lymphocytes in MSI carcinomas.[110] The clinical phenotypes (age at onset and tumor spectrum) vary significantly among mutation carriers and even among the carriers of a similar germline mutation, [105,111] suggesting that environmental factors, such as diet, may modify the cancer risk even in LS.

In 1991, before the genetic basis for LS was discovered, the International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer established Amsterdam I criteria to set guidelines for clinical diagnosis of Lynch syndrome and to identify high risk families (Table 1).[112] In 1999, the guidelines were further revised to Amsterdam II criteria, when some extra-colonic tumors were added as qualifying criteria for LS.[113] In 1997, revised Bethesda guidelines were published (Table 1).[73] The evaluation of MSI status and/or immunohistochemistry (IHC) of MMR-proteins were included in order to better identify individuals who should be genetically tested for LS (Table 1).[73] The sensitivity and specificity for LS in those meeting any one of the guidelines is 82 and 77%, respectively.[114]

Table 1. *The international Amsterdam and revised Bethesda guidelines for diagnosis of Lynch syndrome*[99] and tumor MSI status classification

Amsterdam criteria I-II for the diagnosis of Lynch syndrome
Three or more relatives with histologically verified Lynch syndrome-associated cancer (colorectal, endometrial, small bowel, ureter, renal pelvis)
One is first degree relative
At least two successive affected generations
One or more cases diagnosed before age 50
FAP excluded
Tumors verified by pathological examination
The revised Bethesda guidelines for testing MSI in CRC
MSI testing of colorectal tumors if:
CRC diagnosed before 50 years of age
Synchronous/metachronous CRC or other LS-associated tumors detected
MSI-H CRC diagnosed in patient under 60 years of age
CRC diagnosed in at least one first degree relative before 50 years of age
CRC diagnosed in two or more first- or second degree relatives with LS-associated tumors
Tumor MSI status classification
MSI-H when two or more of five markers show new allele size
MSI-L when one of five markers show new allele size
MSS when none of five markers show new allele size

Other hereditary syndromes

The familial cancer-predisposing syndromes of the gastrointestinal tract are heterogeneous groups of diseases.[115] In addition to Lynch syndrome, the inherited CRC syndromes that express adenomatous polyps include familial adenomatous polyposis (FAP, OMIM 175100), attenuated FAP (AFAP, OMIM 175100), and MUTYH-associated polyposis (MAP, OMIM 608456). The primary lesions in Peutz-Jeghers syndrome (PJS, OMIM 175200) and juvenile polyposis syndrome (JPS, OMIM 174900) are hamartomatous polyps. A hyperplastic polyposis (HPP, OMIM 610069 and 601228) is a rare condition that has a substantial cancer risk but is rarely inherited.[101] FAP, attenuated FAP, Peutz-Jeghers syndrome, and juvenile polyposis syndrome are inherited in autosomal dominant manner and MAP in autosomal recessive manner.

FAP is the second-most common inherited CRC syndrome with a prevalence of 1 in 10,000 individuals accounting for less than 1% of all CRCs.[116] Characteristic features of FAP include the development of hundreds to thousands of colonic adenomas, beginning in late childhood / early adolescence, and 100% early onset CRC risk in untreated individuals.[117] Attenuated FAP is a less-severe form of the disease, characterized by an average 69% lifetime risk of CRC.[118,119] Both FAP and attenuated FAP are caused by germline mutations in tumor suppressor gene *APC* (Adenomatosis Polyposis Coli), which is an important regulator of the Wnt signaling pathway.

MUTYH-associated polyposis (MAP) is caused by mutations of the mutY homolog (MUTYH) gene that was firstly described in 2002 in three members of a British family.[120] It is clinically similar to the AFAP and characterized by early-onset of multiple adenomatous polyps of the colon and rectum which have a risk for malignant transformation, and infrequent extracolonic manifestations.[121]

Peutz-Jeghers syndrome and Juvenile polyposis are characterized by hamartomatous polyps that are rare compared to neoplastic and hyperplastic polyps and most often found in children.[122,123] Juvenile polyposis is a rare syndrome (1:100000-160000 live births) characterized by 1-100 hamartomatous polyps throughout the gastrointestinal tract, mostly in the colorectal segments, often diagnosed in young patients.[124] Patients with JPS have a significantly increased life

time risk for CRC.[125] Peutz-Jeghers syndrome is characterized by mucocutanous melanosis, polyposis of the GI-tract, luminal gastrointestinal cancer and extraintestinal cancer. The incidence has been estimated to be approximately 1:8300 and 1:200.000.[126] The patients have high risk of cancer in the GI-tract as well as extra-intestinal cancer.[127]

Hyperplastic polyposis is a rare condition characterized by multiple and/or large hyperplastic polyps of the colon whose etiology is unknown.[101] Familial cases of HPP have been reported, although these are rare.[128] The World Health Organization's criteria for HPP include 30 cumulative hyperplastic polyps of any size distributed throughout the colon, 5 or more hyperplastic polyps proximal to the sigmoid colon with at least 2 being greater than 10 mm in diameter, or at least 1 hyperplastic colonic polyp in an individual with a first-degree relative with HPP. Sessile serrated polyps/sessile serrated adenomas have also been added to the polyp histologic type.[129] Patients with HPP have increased CRC risk with a median age at onset of 50-60 years[128] and the carcinomas have a tendency to form in the proximal colon.[130] It is now widely accepted that the serrated neoplasia pathway exists in addition to the traditional adenoma–carcinoma sequence. Somatic changes within sessile serrated polyps, such as activating BRAF mutations and CIMP phenotype, with or without MSI, are important events in the serrated pathway.[131]

Lifestyle and dietary risks

Colon cancer is a disease of gene expression and only a minority of CRCs appear to be related to inherited single high-penetrance gene mutations. Consequently, a major determinant of cancer risk appears to be the interaction between genome and environment.[132] Indeed, the progression from normal colonic epithelium to cancer is a complex process involving genetics, epigenetics and environmental factors.[7] In addition to ageing and heredity, diet is noted to be one of the key players in CRC etiology. It is estimated that up to 80% of CRCs may be associated with diet.[133] Numerous epidemiological studies up to meta-analysis level have been conducted to provide evidence for dietary CRC prevention and to improve survival among CRC patients. However, the results are often indefinite or even contrary and

the molecular mechanisms through which diet affects colon mucosa and CRC development still needs to be elucidated.[21]

Food components may act directly as mutagens or alter the cellular milieu by modulating hormonal axes influencing the growth and proliferation of specific cell populations such as colonocytes.[132] Dietary factors such as folate, alcohol and methionine may be associated with colon cancer by directly influencing the expression of key genes such as *APC*, *CDKN2A*, *MGMT*, *MLH1*, and *RASSF1A* by causing abnormalities in DNA methylation,[132,134] synthesis, and repair.[80,135]

The wide geographic difference in the global distribution of CRC suggests the importance of environmental influences on colorectal carcinogenesis.[33] High CRC prevalence is associated with the consumption of a Western-style diet characterized by high intake of red meat and/or processed meat, high-fat dairy products, fast food, refined grains, and sweet foods and drinks.[2] Consequently, WD containing high amounts of fat, red and processed meat and ethanol from alcoholic drinks and low amounts of fiber, calcium, folate and vitamin D, is recognized as a major risk factor for sporadic CRC.[132,136] On the other hand inverse association is seen between CRC risk and diet rich in vegetables, fiber, dairy, and fish. According to the AICR/WCRF Colorectal Cancer Report 2011, physical activity and consumption of foods containing dietary fiber, milk, calcium, and vitamin D protects against colorectal cancer (Table 2).[2,3]

Table 2. *Lifestyle and dietary factors that affect CRC risk*

Decreases CRC risk	**Increases CRC risk**
Physical activity	Red meat
Foods containing dietary fiber	Processed meat
Milk	Alcoholic drinks
Calcium	Body fatness
Vitamin D_3	Abdominal fatness

Western-style diet with decreased levels of nutrients such as vitamin D and calcium has been shown to increase the incidence of pre-neoplastic intestinal lesions and malignant neoplasms in various mouse models of intestinal tumorigenesis.[137] In 1999, Lipkin et al. showed that wild-type (wt) mice developed whole colon-crypt hyperplasia when fed WD for short periods of time without carcinogen administration, suggesting that WD increases the cellular proliferation in colon epithelium.[137] In 2001 and 2009 they further showed that long term feeding with Western-style diet strongly induces both benign and malignant neoplasms in the colon of mice without any carcinogen treatment or genetic predisposition.[17,18] In a mouse feeding study by Lamprecht et al., supplementation of WD with calcium and vitamin significantly suppressed the diet-induced changes in wild-type mice and in genetically predisposed *Apc*-mutant mice.[138]

The reducing impact of fiber on CRC risk has been shown to be consistent in several cohort studies. Especially fiber from cereal, fruits, and grains have been shown to reduce the CRC risk.[139] Studies have demonstrated that in populations consuming high amounts of fat, the concomitant high intake of fibers significantly reduces the risk of CRC.[140,141] Actually, it has been estimated that CRC risk could be reduced up to 30-40% by increasing fiber intake.[142,143] Fiber has several effects in the gastrointestinal tract, but the precise mechanisms for its protective role remain poorly understood. Fiber dilutes fecal content, decreases its transit time, and increases stool weight by absorbing liquid, thus diminishing the direct contact time of epithelial colonocytes with colon contents. Furthermore, the intestinal microflora may be influenced by fiber or may modify the effects of fiber in colon which in turn may affect the CRC risk.[21,140] The gut flora produces fermentation products from fiber, especially short-chain fatty acids such as butyrate, which is the main source of energy for colonocytes. Short-chain fatty acids also induce apoptosis, cell cycle arrest, and differentiation.[144]

Calcium and vitamin D have been shown consistently in experimental studies on animal models to have anti-cancerous properties, including but not limited to stimulating cellular differentiation, reducing proliferation, and inducing apoptosis in colonic epithelial cells.[33,145] Yet, the results have been inconsistent in epidemiologic studies.[3] This is probably caused by the fact that the effects of vitamin D and calcium are strongly interrelated and, for example, calcium-mediated effects are strongly dependent on adequate levels of vitamin D which is needed for

its absorption.[146] To further clarify the impact of calcium and vitamin D in CRC prevention, the molecular mechanisms behind the joint beneficial effects still need to be elucidated.

Along with Western-style diet, alcohol consumption is a well-known and established risk factor for colorectal cancer.[147-149] The mechanisms by which alcohol consumption exerts its carcinogenic effect have not been defined fully, although plausible events include: a genotoxic effect of acetaldehyde, the main metabolite of ethanol; increased estrogen concentration; a role as solvent for tobacco carcinogens; production of reactive oxygen species and nitrogen species; changes in folate metabolism; and reduced absorption of other B vitamins (B1, B2, B12) which increases vulnerability to oxidative stress.[150,151] Alcohol is estimated to contribute to 17% and 4% of total CRC burden in men and in women, respectively. That is, when consumption exceeds the recommended upper limit of two drinks a day for men (about 24 g alcohol), and one for women (about 12 g alcohol).[152] Another major lifestyle factor related to increased CRC risk is tobacco smoking. Nicotine from tobacco smoke has been shown to increase cellular proliferation by changing the expression of receptors and their phosphorylation patterns in several different mitogenic pathways[153,154] and to stimulate angiogenesis and neovascularization in colon cancer.[153,155] In addition to nicotine, tobacco smoke contains carcinogenic compounds such as acetaldehyde, benz-pyrenes, aromatic amines, and N-nitrosamines that can bind DNA and disrupt normal cell functions.[156]

Sedentary lifestyle and consumption of energy dense foods high in animal fat and protein, refined grains, and added sugar in high-income countries has led to exceeding numbers of obese individuals.[157,158] Epidemiological evidence clearly demonstrates that excess intake of total energy[159,160] and consequent obesity, defined as a body mass index (BMI) greater than 30 kg/m2, are associated with a significantly increased risk (30-70%) of colon cancer, especially in men.[161] Excess body fat, especially abdominal fat, creates an environment that promotes carcinogenesis and discourages apoptosis by directly affecting the levels of circulating hormones, such as insulin, insulin like growth factors, and estrogens. It also stimulates the body's inflammatory response, which may contribute to the initiation and progression of cancers.[3]

Several studies suggest that physical activity reduces colorectal cancer risk.[162-166] Exercising just one hour per week has been shown to be associated with a lower prevalence of colonic polyps and adenomas when compared to people who exercised less or not at all[167], and at least 30 min of moderate to vigorous daily exercise has been shown to protect against CRC.[168] Regular physical activity elevates basal metabolism and improves tissue oxygenation. This leads to more efficient metabolism and finally to reduced body fat, insulin level, and insulin resistance which consequently reduces the risk of CRC. Indeed, individuals who take regular exercise have 24% smaller CRC risk, regardless of their BMI in comparison to people with more sedentary lifestyle.[169,170] In contrast, physical inactivity is associated with a status of low grade chronic inflammation or latent inflammation, and higher estrogen, androgen and insulin levels which are known to promote the proliferation of epithelial cells.[171] Physical inactivity also increases the gastrointestinal transit time of food components and the duration of direct contact between food derived carcinogens and the gut epithelium.[172]

MOUSE MODELS OF COLON CARCINOGENESIS

Today, many genetically engineered mouse strains exist and laboratory mouse (mus musculus) has become one of the most used animal models in biomedical research thanks to the abundant genetic information and advanced transgene and knock-out techniques. The mouse models enable the study of molecular mechanisms of colorectal carcinogenesis, to test potential preventative and therapeutic strategies, and to translate hypotheses derived from cell culture studies into the complex physiology of the colon.[173] Further advantage to use a mouse as a CRC model is its high reproducibility and that as in humans, CRC development seems to follow the adenoma-carcinoma sequence[174] The shared tumor phenotypes in human and genetically-engineered mouse suggest that the basic mechanisms of DNA repair and tumor suppression are conserved.[175] The short lifespan of mouse models allows the long-term study of the effects of various CRC risk factors, such as diet, *in vivo*.[173]

Mouse models of FAP

The first genetically engineered mouse model for human CRC studies is the *Apc*^{Min} mouse (multiple intestinal neoplasia Min), the model counterpart for human Familial adenomatous polyposis syndrome[176]. *Apc*^{Min} mice spontaneously form a number of benign adenomas in the small intestine due to a truncating point mutation at position 850 of the Apc gene[177]. After the *Apc*^{Min} mouse, many different Apc$^{+/-}$ strains have been constructed, e.g. Apc$^{\Delta 716}$ and Apc^{1638N} which develop 3-300 intestinal adenomas depending on the type and location of a mutation and function of other modifiers. [178,179,180,181] An important feature of these models is that, unlike in human FAP, the majority of intestinal adenomas develop in the small intestine instead of colon and the adenomas rarely progress into invasive adenocarcinomas.[182]

MMR-deficient mouse models

Today, there are several mouse strains with defective MMR capability resembling Lynch syndrome. Although heterozygous *Msh2, Msh6*, and *Mlh1* mice fail to develop early-onset tumors, homozygous knockout mice are cancer prone and develop tumors in multiple organs, including the gastrointestinal tract. These mice die prematurely due to aggressive lymphomas resembling patients carrying biallelic MMR gene mutations.[183]

Msh2$^{+/-}$ mice are unable to repair single base mismatches and small one to four base insertion/deletion loops (IDLs), which causes a severe reduction in survival and a strong cancer predisposition phenotype.[183,184] Homozygous *Msh2*$^{-/-}$ mice usually die in eight months from T-cell lymphomas. They develop adenomas and invasive adenocarcinomas, which show high MSI like LS tumors, but develop in small intestine.[183]

Homozygote *Mlh1*$^{-/-}$ mice show complete MMR deficiency and usually die after 9-12 months due to several lymphomas and intestinal carcinomas which show high MSI.[173,185,186] Also in *Mlh1*$^{-/-}$ mice the tumors are more often in small intestine than in colon.[173] Interestingly, when *Mlh1* and *Msh2* knockout mice are crossed to mice

heterozygous for a mutated *Apc* allele, intestinal tumorigenesis is markedly increased.[187,188]

Msh6 knock-out mice survive up to 18 months and develop tumors at an older age than *Msh2*- and *Mlh1*-deficient mice.[189] *Msh6*$^{-/-}$ mice develop lymphomas or epithelial tumors originating from the skin, but only rarely from the intestine.[190] Moreover, they often develop endometrial cancers and cancers with variable MSI phenotype like human *MSH6* mutation carriers[191,192]

Diet effect modelling in mouse intestine

Murine models provide an excellent starting point in studies on changes mediated through diet and for understanding the links between diet, genetics, and colorectal cancer.[193] Mouse models, especially *Apc*^{Min}, have been used to assess the effects of dietary components on cancer. Several studies have shown that so called Western-style diet, high in fat and total energy, induces gastrointestinal tumors in mouse models for familial intestinal cancer, and even in wild-type mice without any carcinogen treatment.[17,18,20] This is also seen in the *Apc*^{Min} mouse were small intestinal tumor numbers have been shown to increase 28% and 47% when a total fat in basal diet was increased from 3 to 10% and 15%, respectively.[194] Results were even more drastic in the colon of *Apc*^{Min} mice where a 207% increase in tumors occurred.[194] Both the total amount of fat and the type of fat can influence risk in genetically predisposed mice. Epidemiological evidence suggests that saturated fat increases CRC risk and, inversely, fats from vegetable sources may reduce the risk.[195]

Some essential nutrients and calorie restriction have been shown to have a protective effect against colon cancer. Significantly, feeding a 60% calorie-restricted diet resulted in a 60% reduction in small intestinal polyp numbers in *Apc*^{Min} mice.[196] The effect is probably based on the fact that calorie restriction reduces cell proliferation, enhances rate of apoptosis, and reduces inflammation,[197] which alone or in combination may reduce the CRC risk. Consumption of some fatty acids, however, namely steraridonic acid (SDA) and eicosapentaencic acid (EPA), and docosahexaenoic acid (DHA) appear to reduce colon cancer risk in *Apc*^{Min} mice by

reducing the number and size of colonic tumors.[198] In fact, SDA addition resulted in a significant 45% fewer small intestinal tumors and was as effective as the drug Sulindac (320 mg/kg) in reducing colonic tumors by 85%.[198] The ratio of omega-6 to omega-3 fatty acids may be key to determining a response to fatty acids.[199] When the ratio is low the amount of pro-inflammatory products produced from omega-6 fatty acids decreases. As a result cell proliferation is depressed and the number of visible tumors declines.[193]

Some natural compounds, such as sulforaphane a organosulfur compound from cruciferous vegetables, chafuroside a flavone derivative from oolong tea, and curcumin (diferuloylmethane) from turmeric, have been reported to significantly reduce the numbers of intestinal tumors in *Apc*^{Min} mice.[200,201] Sulforaphane appears to be an effective anti-cancer agent both in cell culture, and carcinogen-induced, and genetic cancer models[202] due to its ability to induce apoptosis and decrease cell proliferation[203], reduce inflammation[204] and as an antioxidant to protect against free radicals.[205,206] Curcumin, which acts as an antioxidant and anti-inflammatory factor through modulation of multiple signaling pathways[207] has been shown to reduce the incidence of intestinal adenomas *Apc*^{Min} mice.[201] Curcumin has several positive pharmacological effects such as anti-inflammatory, antioxidant, antimitogenic, anti-cancer, and antiviral properties. Curcumin's anti-inflammatory effects are thought to be caused by reducing trans-endothelial monocyte migration by reducing mRNA and protein expression of intercellular adhesion molecule-1, vascular cell adhesion molecule-1 and P-selectin and by modulating NFκB[208], JNK, p38, and STAT-3 in endothelial cells.[209] Chafuroside has also been reported to cause a significant inhibition of intestinal tumors in *Apc*^{Min} mice.[210] Its antitumor properties are thought to result from its ability to serve as a free-radical scavenger, reduce inflammation, and increase apoptosis.[211]

AIMS OF THE STUDY

Colorectal cancer evolves as a multistep process involving both inherited and environment-induced genome aberrations. The main focus of this study was to investigate the effects of Western-style diet, ageing, and genetic predisposition on gene expression in histologically normal mouse colon mucosa, and of those, to identify the expression changes and molecular mechanisms driving the CRC development.

Specific aims:

● To study whether and how WD, ageing, and genetic predisposition (*Mlh1* heterozygosity) induced tumorigenesis (tumor number and stages) in mouse colon (I,II)

● To define the genome-wide gene expression changes caused by WD and genetic predisposition in the histologically normal colon mucosa of aging mice (I,II)

● To study whether tumor suppressor gene silencing in histologically normal mouse mucosa was caused by their promoter hypermethylation (I)

● To define the characteristics in transcriptomes, which were associated with colorectal oncogenesis (II).

MATERIALS AND METHODS

MICE, FEEDING STUDY AND DIETS (I,II)

Heterozygote B6.129-*Mlh1*tm1Rak and wild-type C57BL/6 mice were obtained from NCI-MMHCC; National Institutes of Health, Mouse Repository, NCI-Frederick, MD. Altogether 12 animals, the *Mlh1*$^{+/-}$ mice and their wild-type C57BL/6 mates, formed six breeder pairs which produced the mouse colony used in the study. Mice were genotyped using genomic DNA extracted from earmarks according to the NCI – Mouse Repository protocol. Briefly, genotyping primers M001 (TGT CAA TAG GCT GCC CTA GG), M002 (TGG AAG GAT TGG AGC TAC GG), and M003 (TTT TCA GTG CAG CCT ATG CTC), produce two different length fragments, separating the normal 350 bp (M001/M003) and the mutated 500 bp alleles (M001/M002) (Fig. 2).

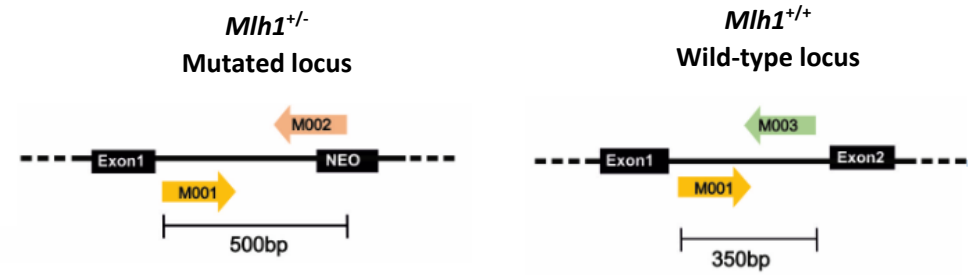

Figure 2. The structure of mutated and wild-type *Mlh1* alleles

The mice were bred and treated according to the study protocol approved by the National Animal Experiment Board in Finland (ESLH-2008-06502/Ym-23). At the beginning of the study, five-week old *Mlh1*$^{+/-}$ and *Mlh1*$^{+/+}$ mice were sampled as a baseline for the study. The remaining mice from both genotypes were divided into two dietary groups and fed *ad libitum* with either healthy rodent control diet AIN-93G (AIN) or Western-style diet (WD) modified from AIN (Harlan Teklad) (Table 3). WD was designed to resemble, on the nutritional level, the diet consumed in the human Western population (high fat and energy content, low amounts of fiber, calcium, and vitamin D_3).

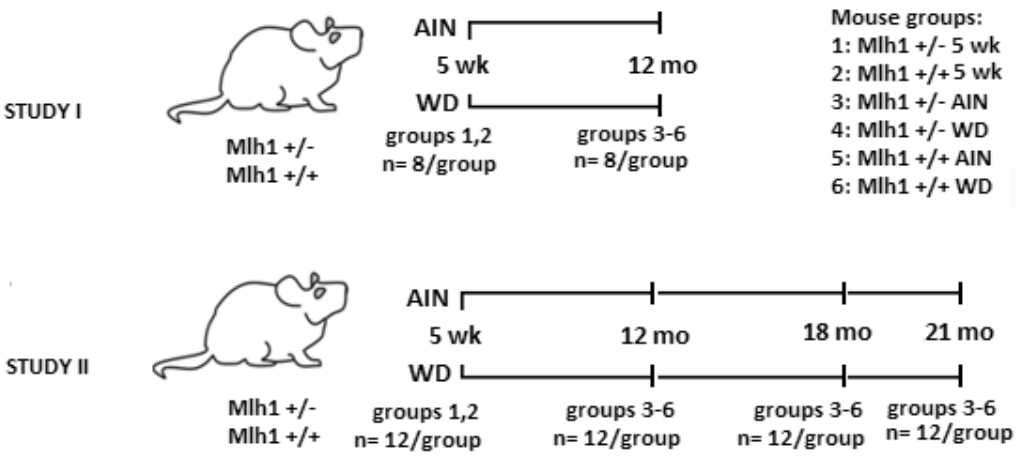

Figure 3. Experimental design of the feeding studies

Table 3. Diet nutritional information.

	AIN93-G	WD
Total energy content (kcal/g)	3.8	4.6
Kcal from carbohydrates (%)	63.9	42.3
Kcal from protein (%)	18.8	18.5
Kcal from fat (%)	17.2	39.2
Fat source (g/kg)		
Soybean oil	70	-
Anhydrous milkfat	-	133
Canola oil	-	55
Sunflower oil	-	12
Vitamin D (IU/kg)	1000	100
Folic acid (mg/kg)	2	0.2
Calcium (g/kg)	5	0.5
Fiber (cellulose) (g/kg)	50	20

RNA AND PROTEIN EXPRESSION ANALYSES (I,II)

Sample preparation (I,II)

Mice were sacrificed and sampled at the age of 12, 18, and 21 months (Fig. 3). Colons were removed, opened longitudinally, rinsed with 0.9% ice cold sodium chloride, and inspected for tumors under a dissecting microscope. Longitudinal pieces were collected for normal proximal mucosa. The mucosa was separated from the underlying submucosa and musculature under a dissecting microscope. Pieces representing approximately one third of the proximal colon were stored in RNAlater (Qiagen) at -80°C for RNA extraction and snap frozen in liquid nitrogen for DNA extraction. Total RNA was prepared using the RNeasy Plus Kit (Qiagen) with an extra DNase treatment (Qiagen) to remove traces of genomic DNA. RNA concentration was measured by Nanodrop 1000 (Thermo Fisher Scientific) and Qubit 1.0 (Thermo Fisher Scientific) and RNA integrity with the Agilent 2100 Bioanalyzer (Agilent technologies). Only high quality RNA (RNA integrity number RIN ≥ 8) qualified for expression analysis.

All observed colon tumors were collected under a dissecting microscope and preserved as FFPE blocks. If a tumor was large enough (three to five mm in diameter), approximately half of it was embedded in O.C.T compound (VWR) for cryo sampling. Histological investigation, staining and grading of neoplasias was 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.[212] The methods used in this study are explained in detail in the original publications I and II. A summary of the methods is presented in Table 4.

Table 4. Summary of methods used in the original publications

Method	Publication
Heterozygote B6.129-*Mlh1*^{tm1Rak} mice (NCI) and their wild-type (*Mlh1*$^{+/+}$) littermates	I, II
RNA and DNA extractions and quantification	I, II
StellARray gene expression array	I
TaqMan RT-qPCR	I
Global Pattern Recognition data analysis tool	I
Data-assist v2.0 software	
Transcriptome profiling with RNA-sequencing (single-cell tagged reverse transcription, STRT)	II
Ingenuity pathway analysis	II
Gene promoter methylation analysis (Sequenom MassARRAY EPITYPER^{TM} system)	I
FFPE sample cutting	I, II
Immunohistochemistry (IHC)	I, II
Laser capture microdissection (LCM)	II
Microsatellite instability (MSI) analysis	II
Loss of heterozygosity (LOH) analysis	II
Mitosis analysis with feulgen staining	II
Statistical analysis	I, II

RNA expression analysis of 94 growth regulatory genes (I)

Gene expression array

The RNA expression of histologically normal tissue samples from proximal colonic mucosa were analyzed by using a quantitative custom made StellARrayTM platform (Lonza Group Ltd, Bar Harbor BioTechnology). The StellARray included 94 genes (73 TSGs) previously associated with development of colorectal cancer and/or documented to exhibit CpG island (CGI) hypermethylation in CRC and other human cancers. APC was included in the array as an indicator of ongoing carcinogenesis.

Each StellARray plate well was loaded with 20 µl of SYBR Green master mix containing sample specific cDNA and DNA polymerase (Thermo Scientific). The RT-qPCR arrays were run on a Mx3000P cycler (Agilent technologies) using the following cycling parameters: 50°C for 2 min, 95°C for 7 min, 40 cycles of 95°C for 15 s, and 60°C for 1 min. The expression differences between different mouse groups were analyzed using the Global Pattern Recognition™ (GPR) software (Bar Harbor Biotechnology).

TaqMan assay

The results of statistically significant expression changes in StellARray were validated by using RT-qPCR and TaqMan gene expression assays. Each sample was assayed in triplicate for the target genes, as well as the endogenous reference genes using the following cycling parameters: 1 cycle of 95°C for 10 min, 40 cycles of 95°C for 15 s, and 60°C for 1 min. Thermal cycling and fluorescence data acquisition was performed with a StepOnePlus cycler (Life Technologies) and Cq values were called using the Data-assist v2.0 software (Life Technologies). The uniformly expressed reference genes were selected using the GeNorm algorithm from amongst the 94 genes included in the StellARray.

If the amount of template was too low to provide reliable results, the RT-qPCR validation was performed using pre-amplified cDNA. Samples were multiplex pre-amplified with TaqMan PreAmp Master Mix Kit (Life Technologies) following

manufacturer's instructions. The cycling parameters were as follows: 1 cycle of 95°C for 10 min and 10 pre-amplification cycles of 95°C for 15 s, and 60°C for 4 min.

Genome wide transcriptome analysis (II)

RNA-sequencing

Transcriptome analysis was performed using RNA-sequencing. RNA-seq followed the single-cell tagged reverse transcription (STRT) protocol with modifications. Briefly, total RNA from 5 week (n=14), 12 mo, and 18 mo (n=40 and 40) mice was converted to cDNA and amplified to form an Illumina-compatible library. Four base-pair unique molecular identifiers were applied and only the absolute number of unique reads were included in the subsequent analyses. The samples were sequenced on a Illumina HiSeq2000, further processed to fastq files by Casava 1.8.2 (Illumina), and quality control was performed using the STRTprep pipeline (https://github.com/shka/STRTprep). The processed reads were aligned by TopHat248 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. Therefore, STRTprep counts only the aligned reads at the 5'-untranslated region of protein-coding genes, or within the proximal (500 bp) upstream region.

RNA-seq data normalization and data analysis

A read count matrix, with genes as rows and samples as columns was obtained from STRTprep pipeline. Different library sizes were normalized using DESeq-style normalization[213]. The data were converted to a more Gaussian like with shifted log transformation ($x_{log} = \log(x + 1)$) and the ComBat program[214] was used to filter library batch effects. The preprocessing steps were evaluated by looking at the hierarchical clustering of samples and by plotting quantiles of expression values for each sample.

After ComBat normalization the data was no longer integer count values. Therefore, three T-test based methods, Voom-Limma, Cyber-T, and Shrinkage-T,[215-217] which all add a prior to a variance estimate, were considered for analysis of differential gene expression. Of these, shrinkage-T is the only method that also allows testing with unequal variance between data sets. The three methods were evaluated by viewing the separation of cancer samples from the remaining samples in the MDS plots with top-k genes, selected using the evaluated statistic. Parameter k was varied from 25 to a few hundred. Shrink-T showed the best separation in the generated plots across all values of k. Each method's ability to find correlations with Gene Ontology classes was also tested. T-test scores from each method were used separately as an input to enrichment analysis tool GSZ (Gene Set Z-score).[218] Shrink-T again generated strongest results. Therefore, only Shrink-T were used in the subsequent analyses.

To see the similarities and differences between the samples, visualizations were generated with Multi-dimensional Scaling (MDS). In short, MDS generates a small-dimensional visualization from the multidimensional data while trying to preserve the pairwise distances of samples from the multi-dimensional data. PlotMDS distributed in the Limma package was used with some modifications as a basis of the analysis.[219] The plotMDS was modified in a way that enabled us to use any selected score to pick the genes that were used to calculate pair-wise distances.

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.

Protein expression analysis in carcinomas (II)

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). Stainings were visualized using UltraVision Detection System anti-rabbit HRP/DAB (ThermoFisher Scientific).

GENE PROMOTER METHYLATION ANALYSIS (I)

For the DNA methylation analysis, only the genes which had shown a significant mRNA expression decrease (i.e. candidate hypermethylated TSGs) in association with inherited predisposition and/or WD were selected, and the methylation levels of their CpG islands (CGIs) were assessed individually for each mouse. Genomic DNA for the methylation analysis was extracted using DNeasy Blood & Tissue Kit (Qiagen) from colonic tissue samples from the immediate vicinity of those used for RNA expression studies.

The quantitative methylation analysis was performed with MassARRAY EPITYPER™ system (Sequenom) which is a bisulphate-based technology relying on base-specific cleavage of RNA and matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS) to determine the relative extent of methylation in DNA fragments containing either one or several subsequent CpG sites, which are referred to as CpG units. In the mass spectrum, a distinct signal pattern results from the methylated and non-methylated target sequence, and the EPITYPER software determines the individual methylation ratios (i.e. signal intensity ratios as percentage of methylated to non-methylated signals) for CpG sites/unit within a target sequence. The system is able to detect methylation levels as low as 5%.

The amplicons were primarily selected to cover CGIs annotated by the UCSC genome browser and to be overlapping with the transcription start site and the 5' UTR or be in their close proximity. Altogether, twelve different amplicons (lengths between 174 and 499 bp) were analyzed, of which eight belonged to the pre-validated Mouse Standard EpiPanel and four were custom DNA methylation assays designed using Sequenom's EpiDESIGNER software (Sequenom GmbH, www.epidesigner.com), for which primer sequences and target sites are provided in Table S4 Pussila et al 2013. In order to reduce methylation variability introduced during PCR[220], three replicate amplifications were performed and pooled for mass analysis. The initial methylation data was filtered by Sequenom to exclude poor quality measurements. CpG units that yielded data in greater than 75% of the samples passed the initial quality control. From these, samples that yielded data in greater than 80% for all CpG units within an amplicon were selected for that sample/amplicon pair. For further analysis, CpG units which had data available for

less than 50% of all samples and samples which had data available for less than 50% of all CpG units were excluded.

MICROSATELLITE INSTABILITY STUDY (II)

The microsatellite instability status was analyzed from carcinomas using four dinucleotide (D18Mit15, D14Mit15, D10Mit2, and D7Mit91) and two mononucleotide (JH104 and U12235) markers. Tumor DNA samples were extracted from the cryo-preserved colon carcinomas using laser micro-dissection for cutting (Zeiss PALM MicroBeam, Carl Zeiss Microscopy GmbH) 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 11.1X PCR master mix 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 analyzed with ABI3730xl capillary electrophoresis (Thermo Fisher Scientific) and visualized with PeakScanner v1.0 (Thermo Fisher Scientific).

LOSS OF *Mlh1* HETEROZYGOSITY STUDY (II)

Mlh1 Loss of heterozygosity was analyzed for colon carcinomas using the genotyping primers M001, M002, and M003, which produce two different length fragments separating the normal and the mutated allele (Fig. 2). DNA was amplified with the 11.1X PCR master mix 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, and the fragments were visualized on 1% SB agarose gel.

INGENUITY PATHWAY ANALYSIS (II)

Pathway analysis of the top genes which separated the carcinoma mice normal mucosa expression patterns from the others was performed using QIAGEN´s Ingenuity Pathway Analysis (IPA Software 7.0, Qiagen). 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.

ANALYSIS OF MITOTIC ABERRANCES IN CARCINOMAS (II)

Feulgen stain with Midori green background stain was used to visualize nuclear material and mitoses in carcinoma samples. 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 bisulfite washes, samples were counterstained briefly with 1% Midori light green, dehydrated through alcohols to xylene and mount with xylene-based mounting media. Stainings were analyzed under light microscope (Zeiss Axio Imager A2, Carl Zeiss Microscopy GmbH) and the mitoses in the malignant areas of the carcinoma samples were compared to samples from healthy control mice.

STATISTICAL ANALYSES (I,II)

In StellARray RT-qPCR, a common threshold was set across all the plates within the experiment. The expression differences between different mouse groups were analyzed using the Global Pattern Recognition™ (GPR) software which takes advantage of biological replicates to extract significant changes in gene expression. GPR executes efficiency correction and reference gene and control group normalizations (www.bhbio.com/BHB/dw/products.gpr.html). The StellARray/GPR

system does not work with user defined reference genes. Instead, the GPR software algorithm determines the best set of stably expressed reference genes within the experiment by comparing the expressions of all the genes included in the assay between test and control samples. It then generates a global pattern of expression changes, and creates a ranked list of significant changes between sample and control[221]. In this way, the selection of reference genes is unbiased since it enables the experimental data, and not the researcher, to define the stably expressed reference genes.

In TaqMan RT-qPCR, relative mRNA expression changes were analyzed using the comparative Ct (ΔΔCt) method, which presents the data as fold changes in gene expression normalized to endogenous reference genes and relative to the control group[222]. Data-assist v2.0 software (Life Technologies) was used for quality control and normalization of the quantification cycle (Cq) data, and a median permutation method[223] was used to determine the significance of the expression fold changes compared to the control group with significance level of $P < 0.05$.

The correlation between the StellARray expression patterns of 5 week old *Mlh1*$^{+/+}$ and *Mlh1*$^{+/-}$ mice was studied using Pearson correlation analysis (R value) of the PASW Statistics 18 system.

Chipster software[224] was used to visualize and analyze the methylation data. The Non-metric Multi-Dimensional Scaling (NMDS) tool was used to produce two-dimensional maps based on sample dissimilarity calculated using Euclidean distance, and the Dendrogram tool was used to create dendrograms of samples using normalized data with Pearson correlation and average linkage method. For Chipster analyses, missing methylation values of individual CpG units were defined as the median value of the CpG unit within the particular mouse group.

Comparison of the average methylation levels between different mouse groups was performed using Mann-Whitney test of the PASW Statistics 18 system.

In differential Expression Analysis (DEA) for RNA-seq data modified t-tests Limma, cyber-T, and shrinkage t were used. With Limma and cyber-T their own p-value estimates were used. 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. DEA was mainly used 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 analyze over-representation of genes from the analyzed gene groups. Here, multiple testing correction was done using the Benjamin-Hochberg method.

RESULTS AND DISCUSSION

WESTERN-STYLE DIET AND CRC RISK (I,II)

Colon tumors develop predominantly and earlier in WD fed mice (I, II)

A summary of colonic tumors found in mice is presented in Table 5. The age effect was clear in the mouse series since the overall number of tumors increased significantly with time (Fig. 4). Heterozygote *Mlh1*$^{+/-}$ mice showed 67% and 45% of all colonic tumors in feeding studies I and II, respectively. Importantly all the adenocarcinomas were found in the proximal part of colon and the majority were either tubular or mucinous in histology.

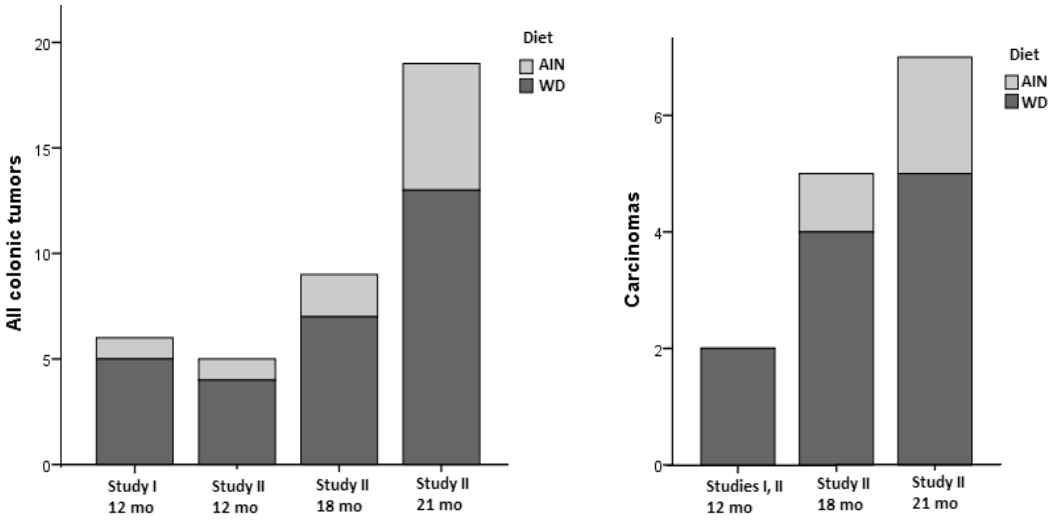

Figure 4. Number of colon tumors and carcinomas in different diet groups and ages in studies I and II

Table 5 Summary of colon tumors found in studies I and II

	Study I	**Study II**		
Time point (mo)	12	12	18	21
Total number of mice	48	48	48	48
Tumor incidence (%)*	12.5	10.4	18.8	33.3
Number of intestinal tumors				
Carcinomas	1	1	5	7
Adenomas	3	3	2	5
Hyperplasias	2	1	2	7
Total	6	5	9	19
Tumors in WD mice (%)	80	80	78	64
Carcinomas in WD mice (%)	100	100	80	72
Tumors in *Mlh1*$^{+/-}$ mice (%)	67	40	56	42

**Number (%) of mice with colon tumor*

The significance of WD on CRC risk in studies I and II is highlighted by the fact that 80% of mice with tumors were fed with WD, and that WD also caused tumor development in wild-type mice without inherited predisposition. In 32 mice included in study I, two hyperplasias, three adenomas, and one proximal adenocarcinoma were found after 12 mo of the feeding study, five out of the six tumors found in WD fed mice. The one control diet fed mouse had hyperplasia and was a mutation carrier. Due to low number of tumors the diet difference wasn´t statistically significant, yet suggesting that WD increases the number and probability of colonic tumors. In study II, altogether 33 colonic tumors developed during the 21 mo feeding experiment in 144 mice. Here, WD seemed to be a severe risk factor for CRC, since approximately 80% of all colon tumors, 10 out of 13 colon adenocarcinomas and 14 out of 20 adenomas and hyperplasias, developed in WD fed mice (Table 5). In study II, at time points 12, 18, and 21 mo, 100%, 80% and 72% of CRCs developed in WD fed mice, respectively, indicating that Western-style diet also accelerates the progression of carcinogenesis.

These findings follow the observations made by Newmark et al at 2001[18] that long term feeding with Western-style diet strongly induces both benign and malignant neoplasms in the colon of laboratory mice without any carcinogen treatment or genetic predisposition. The increased number of colon tumors developed in WD-fed mice is most probably caused by accumulation of cancer promoting genetic and epigenetic alterations in the colonic epithelium which precede tumor development. The diet used in this study mimics the new Western-style diet 1 (NWD1) used by Newmark et al[18] with few exceptions. In that study, the main fat source was corn oil, while in this study Western-style diet contained high amounts of saturated anhydrous milk fat. Although it was assumed that the hard animal fat used in the present study would have induced more tumorigenesis as compared to vegetable oil, after 18 mo feeding with NWD1 the tumor incidence reach 42% as compared to 18.8% in this study. The difference may partially be explained by strain differences, housing conditions, and the total number of mice in the experiment.

A specific set of tumor suppressor genes (Dkk1, Slc5a8, Hoxd1, and Socs1) show significantly altered methylation and mRNA expression in mouse colon mucosa (I)

In study I, a candidate gene approach was chosen instead of a genome wide expression analysis in order to focus on alterations already shown to be associated with human colon cancer. The normal mucosa expression of 94 growth-regulatory genes (73 tumor suppressor genes, TSG) previously linked to human CRC was studied at two time points (5 wk (week) and 12 mo of age) in the heterozygote *Mlh1*$^{+/-}$ mice and their wild-type littermates. The 5-week-old mice showed a strong correlation (Pearson's $R = 0.989$, $P < 0.01$) between the mRNA expression patterns at the outset (Fig. 5A). Distinct from the other 93 genes, but in accordance with the different genotypes, the expression of *Mlh1* was approximately 50% lower in the heterozygote mice than in the wt mice (Fig 5B). Overall, the isogenic background in the beginning made it justified to reason that expression differences that would appear between the different mice groups later were acquired and not the outcome of the inherited genetic constitution.

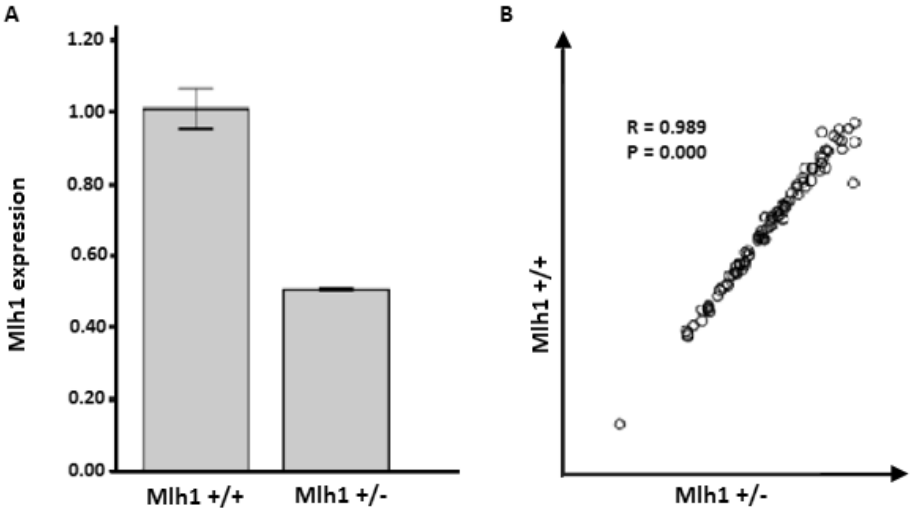

Figure 5. (A) The relative expression of *Mlh1* in the normal colon mucosa of 5 wk old *Mlh1*+/+ and *Mlh1*+/- mice and (B) the correlation between relative expression patterns of the two genotypes in study I.

The expression patterns of 5wk old mice were compared to different 12 mo mouse groups (*Mlh1*$^{+/-}$ AIN, *Mlh1*$^{+/-}$ WD, *Mlh1*$^{+/+}$ AIN, *Mlh1*$^{+/+}$ WD). As a result of aging, the expression of nine genes was found to be increased or decreased (Table 6) ($P < 0.05$). Decreased mRNA levels were seen in *Ccnd1* (Cyclin D1), *Cdh1* (Cadherin 1), and *Mal* (Myelin and lymphocyte protein, T cell differentiation protein), while increased expression was seen in *Axin2*, *Cdx1* (Caudal type homeobox 1), *Fzd10* [Frizzled homolog 10 (Drosophila)], *Mbd2* (Methyl-CpG binding domain protein 2), *Mbd4* (Methyl-CpG binding domain protein 4), and *Rasgrf2* (Ras protein-specific guanine nucleotide releasing factor 2). Since these changes were found in all mouse groups and did not depend on inherited predisposition or WD, they can be attributable to aging.

Table 6. The genes showing age-related down or up regulation in normal mouse colon mucosa (P < 0.05)

	Mlh1$^{+/+}$ AIN	**Mlh1$^{+/-}$ AIN**	**Mlh1$^{+/+}$ WD**	**Mlh1$^{+/-}$ WD**
Down regulated genes (fold change, *P*)				
Ccnd1	3.3 (0.003)	3.8 (0.000)	4.3 (0.001)	3.6 (0.000)
Cdh1	5.4 (0.002)	4.9 (0.000)	5.9 (0.002)	8.4 (0.000)
Mal	3.4 (0.025)	3.4 (0.014)	3.2 (0.034)	2.7 (0.028)
Up regulated genes (fold change, *P*)				
Axin2	2.6 (0.005)	3.0 (0.002)	1.8 (0.039)	1.6 (0.028)
Cdx1	4.1 (0.002)	5.6 (0.000)	3.4 (0.002)	5.4 (0.000)
Fzd10	4.8 (0.002)	7.6 (0.000)	3.4 (0.021)	5.0 (0.000)
Mbd2	3.3 (0.004)	4.7 (0.000)	3.2 (0.003)	5.7 (0.000)
Mbd4	3.0 (0.005)	6.0 (0.000)	3.0 (0.004)	3.3 (0.000)
Rasgrf2	2.3 (0.002)	1.8 (0.006)	2.3 (0.001)	1.5 (0.040)

Significant (P < 0.05) expression changes associated with inherited cancer predisposition and/or WD were observed in seven genes (Table 7). Decreased expression in *Dkk1* [Dickkopf homolog 1 (Xenopus laevis)], *Slc5a8* [(Solute carrier family 5 (iodide transporter), member 8], *Hoxd1* (Homeobox D1), and *Socs1* (Suppressor of cytokine signalling 1), and increased expression in *Dkk2* [Dickkopf homolog 2 (Xenopus laevis)], *Rprm* (Reprimo, TP53 dependent G2 arrest mediator candidate), and *Acaa1b* (acetyl-Coenzyme A acyltransferase 1B).

Table 7. The genes showing up or down regulation in association with inherited cancer predisposition and/or Western style diet in normal mouse colon mucosa

	Mlh1$^{+/+}$ AIN	Mlh1$^{+/+}$ WD	Mlh1$^{+/-}$ AIN	Mlh1$^{+/-}$ WD
Down regulated genes (fold change, *P*)				
Dkk1	2.0 (0.110)	6.2 (0.003)	5.1 (0.012)	7.3 (0.003)
Slc5a8	1.3 (0.078)	1.3 (0.224)	1.7 (0.041)	1.5 (0.032)
Hoxd1	1.1 (0.348)	2.1 (0.019)	2.0 (0.075)	1.5 (0.191)
Socs1	1.6 (0.460)	1.0 (0.485)	2.7 (0.067)	3.1 (0.026)
Up regulated genes (fold change, *P*)				
Dkk2	2.2 (0.108)	1.7 (0.329)	2.2 (0.037)	1.0 (0.637)
Rprm	1.4 (0.442)	1.1 (0.725)	2.0 (0.017)	3.3 (0.019)
Acaa1b	2.4 (0.120)	3.5 (0.124)	1.8 (0.140)	9.4 (0.000)

Significant expression differences (P < 0.05) marked with red.

Many of the genes showing age-related expression changes in study I associate with Wnt/β-catenin signaling. However, their changes did not necessarily imply Wnt signaling activation. The lack of colonic carcinomas (n=1) in study I hindered the investigation of whether the Wnt/β-catenin signaling pathway was aberrantly activated in mouse colon carcinomas as in most human CRCs.[225,226] Furthermore, no changes were observed in the expression of *Ctnnb1*, or *Apc*, a member of the β-catenin destruction complex in the normal mucosa.[227] Yet, the expression of *Cdh1* (cadherin 1, E-cadherin), which is a major component of adherent junctions and binds to free cytosolic β-catenin,[228] was significantly decreased in all mouse groups, downregulation being strongest in the *Mlh1*$^{+/-}$ WD group.

The silencing of Wnt antagonist *Dkk1* already in histologically normal colon mucosa was demonstrated for the first time. The mRNA expression of *Dkk1* was completely silenced in 16 mice and at least in 10 cases the reason could have been hypermethylation. The *Dkk1* silencing was associated with carcinogenesis since four out of the six colonic neoplasms found in study I developed in mice which showed *Dkk1* inactivation in normal colon mucosa. Furthermore, the fact that majority of the mice not expressing *Dkk1* in their colon mucosa were *Mlh1* heterozygotes (10/16) suggest a link to haploinsufficiency caused by a loss of function mutation in *Mlh1*. On the other hand, the decreased level of vitamin D, which has been shown to strongly regulate the human *DKK1* expression[226,229] may have contributed to the

decreased *Dkk1* mRNA levels and colonic neoplasias in mice fed with WD. The importance of vitamin D in the regulation of *Dkk1* by epigenetic modification was supported in a recent study where its intake was negatively associated with *DKK1* methylation in a large cohort of CRCs.[230]

Hoxd1, which is a target gene for Wnt[231] and Polycomb Group (PcG) proteins,[232] provides an illustrative example of dietary effects in study I. This is the first demonstration that the expression and methylation of *Hoxd1* is altered already in histologically normal colonic mucosa and especially in mice fed with WD, suggesting that dietary effects may create selective pressure for its silencing.

Western diet has been reported to induce oxidative stress responses and exert a pro-inflammatory stimulus in the colon long before tumors occur.[233] Metabolic errors may in essence initiate the epigenetic switch that contributes to carcinogenesis and cancer progression.[234] Changes that were observed in *Acaa1b* and *Socs1*, for example, may be related to inflammation and alterations in energy metabolism. Furthermore, a mild inflammation has been reported to accelerate colon carcinogenesis in *Mlh1*-deficient mice[235], a finding in line with the observation that most WD-related alterations are linked to *Mlh1* heterozygosity (Table 7).

To compare the methylation levels between 5 wk and 12 mo mice groups the mean methylation levels of each CGI for each mouse were calculated. Although the mean methylation levels were low in general they were significantly higher among 12 mo mice as compared to 5 wk mice, with the exception of *Mlh1*. This finding was in accordance with the observation that all other genes but *Mlh1* showed age-related decrease in mRNA expression. Overall, mean methylation levels varied between 3% and 8% at 5 wk, and between 6% and 15% at 12 mo, respectively. Mean methylation was lowest for *Mlh1* (5.6%), and highest in *Dkk1* (14.7%) and *Sfrp1* (14.9%) (Table 8).

The average methylation levels were visualized by NMDS plot (non-metric multi-dimensional scaling) in which the 5wk and 12mo mice segregated into different parts of the plots indicating differences in their methylation levels (Fig.6). When visualizing the methylation data by dendrogam and histogram, two distinct methylation clusters (higher and lower methylation cluster, i.e. Group 1 (G1) and Group 2 (G2), respectively) were observed in each CGI. Except for one mouse in *Sfrp1* and one in *Socs1* (B214 and B225, respectively) an identical set of 11 mice

clustered into the higher methylation group at CGIs of *Dkk1*, *SLc5a8, Hoxd1, Socs1*, and *Sfrp1*. Among those 11 mice, only two belonged to the control *Mlh1*$^{+/+}$ AIN group (B211, and B214), while the remaining nine mice had either the inherited predisposition to CRC (B201, B212, B225, and B232) were fed with WD* (B204, B219, B231) or had both risk factors (B215, B220). In addition to these 11 mice another 7 mice (B233, B236, B237, B243, B246, B248, and B252) belonged to the higher *Dkk1* methylation group which showed highest average methylation level at 12 mo time point.

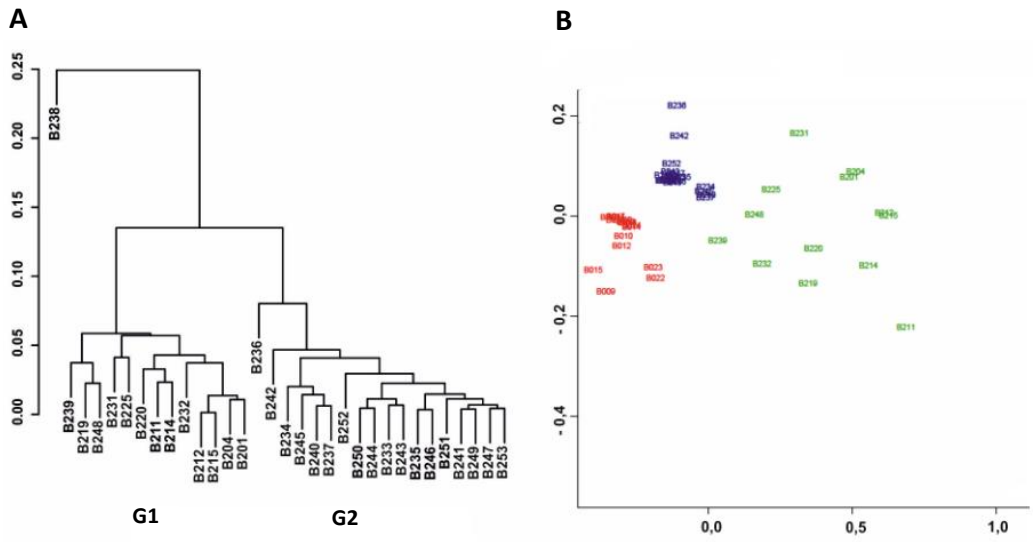

Figure 6. Representative example of higher (G1) and lower (G2) methylation clusters. (A) Histogram and (B) MDS plot of *Hoxd1* methylation groups G1 (green) and G2 (blue). The 5 week old mice marked with red.

Table 8. Descriptive and comparative statistics of the methylation data for genes, *Dkk1*, *Slc5a8*, *Hoxd1*, *Socs1*, *Sfrp1* and *Mlh1*, for 5wk and 12 mo old mice and for higher (G1) and lower (2) methylation cluster, respectively and comparison of the mean methylation values between the different mouse groups

Gene	Group	Mean methylation (%)	SD	5wk vs. 12mo (*P*)	G1 vs. G2
Dkk1	5wk (n =15)	3.4	0.024	0.000	0.000
	12mo (n =32)	14.7	0.075		
	G1 (n =18)	18.7	0.078		
	G2 (n =14)	9.4	0.013		
Slc5a8	5wk (n =15)	5.8	0.017	0.000	0.000
	12mo (n =32)	9	0.044		
	G1 (n =11)	14.2	0.037		
	G2 (n =21)	6.3	0.008		
Hoxd1	5wk (n =15)	3.7	0.008	0.000	0.000
	12mo (n =32)	9.3	0.032		
	G1 (n =13)	11.9	0.021		
	G2 (n =18)	6.9	0.006		
Socs1	5wk (n =15)	8.4	0.012	0.002	0.000
	12mo (n =32)	10.7	0.028		
	G1 (n =10)	14.4	0.021		
	G2 (n =22)	9.1	0.008		
Sfrp1	5wk (n =15)	7.1	0.014	0.000	0.000
	12mo (n =32)	14.9	0.040		
	G1 (n =11)	18.9	0.040		
	G2 (n =19)	12.6	0.016		
Mlh1	5wk (n =16)	4	0.016	0.089	0.000
	12mo (n =31)	5.6	0.029		
	G1(n =9)	8.6	0.023		
	G2 (n =22)	4.3	0.021		

P (Mann-Whitney test, Exact Sig., 2 tailed)

In study I, a significant age-related increase in CGI methylation levels of *Dkk1, Sfrp1, Slc5a8, Hoxd1*, and *Socs1* were observed. Interestingly, *Dkk1* as well as *Sfrp1* are secreted Wnt signaling antagonists[236,237] who act as epigenetic gatekeepers. Their

aberrant silencing may lock the cells into stem cell like states allowing time for genetic gatekeeper mutations in the downstream pathway genes to appear.[236-238] As opposed to other studied genes, *Mlh1*, which was still well-expressed at 12 months, did not show significantly increased CGI methylation suggesting that expression decreases of the other genes precede the involvement of *Mlh1*. Although *MLH1* methylation levels have been reported to increase with age in normal human colonic mucosa[8,72,84,239,240] the results rather comply with studies[240,241] classifying *Mlh1* as a type C gene, with methylation specificity for neoplasia.

Study I was based on the presumption that, together with aging, certain dietary components of the Western style diet are important modifiers of methylation patterns in the colon[11] and thus may affect the transcription of genes involved in tumorigenesis. We reasoned that in mutation carriers requiring just a second hit of the inherited susceptibility gene for malignant transformation, it might be possible to detect the earliest changes, which might even precede the second hit, and distinguish these from alterations occurring later in oncogenesis. These hypotheses proved valid when tested on a mouse model for Lynch syndrome. In summary, among 73 TSGs, a specific set of genes, and *Dkk1* in particular, were identified as promising candidates for altered methylation and expression, detectable already in histologically normal mucosa and potentiated by an inherited MMR gene mutation and Western style diet. Such changes, which are linked to human CRC and occur prior to the second hit in the predisposing gene, are associated with WD and may be among the very earliest alterations in multistep tumorigenesis.

Mlh1 seemed to have a surprisingly mild impact on colon tumorigenesis in 12mo old mice. Yet, two tumor suppressor genes, *Socs1* and *Slc5a8,* which are known to be hypermethylated in human CRC, were significantly down regulated only in the *Mlh1*$^{+/-}$ mice. This suggests a possible impact of *Mlh1* heterozygosity on the down-regulation of the respective genes. The results from study I highlight the interplay between genome, epigenome, and environment in colon tumorigenesis and encourage studies to explore the potential of the respective genes and alterations as biomarkers for diagnostic, prognostic, and therapeutic applications in humans.

THE *Mlh1* GENE AND CRC RISK (I,II)

The effect of *Mlh1* heterozygosity on mouse colon tumorigenesis (I,II)

Since in human Lynch syndrome *MLH1* heterozygosity is known predispose to colorectal cancer, the effect of *Mlh1* heterozygosity on mouse colon tumorigenesis was investigated. The tumors were approximately evenly distributed between the two genotypes. When the numbers of colonic findings from studies I and II were combined, the overall percentage of colonic tumors in *Mlh1*$^{+/-}$ mice was 49%. The distribution of carcinomas, adenomas, and hyperplasias in *Mlh1*$^{+/-}$ mice were 36% (n=5/14), 54% (n=7/13), and 58% (n=7/12), respectively. *Mlh1* heterozygosity did not increase the number of colonic tumors as such in our study, suggesting that the amount of Mlh1 protein in the heterozygote mice is enough to handle the DNA mismatch repair function of mouse *Mlh1*.[242]

Mouse colon tumors do not lack the Mlh1 protein or show MSI (II)

Due to an increased amount of colon tumors in study II we were able to continue to investigate the role of *Mlh1* in colon tumorigenesis. The typical Lynch syndrome characteristics, MSI status and Mlh1 protein expression[53], were analyzed from seven carcinomas found in 18 and 21 mo old mice, four in *Mlh1*$^{+/-}$ mice (E338, E347, E437, E444) and three in *Mlh1*$^{+/+}$ mice (E402, E410, E421). The inactivation of *MLH1* is a common event both in sporadic and inherited cancers leading to absence of the MLH1 protein and MSI.[53,243] Surprisingly, all the studied CRCs showed Mlh1 protein expression (Fig. 7A), indicating that irrespective of the inherited predisposition in the heterozygote mice, the functional allele was still present in the carcinomas. The presence of the normal allele was confirmed with LOH study in all the four *Mlh1*$^{+/-}$ carcinomas (Fig. 7B). To study whether the detected Mlh1 protein was functional and MMR proficient, the stability of six polymorphic microsatellite regions in the mouse genome were analyzed. 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 the seven CRCs were successfully studied and

shown to be microsatellite stable (one sample, E410, could not be amplified), since no differences in the fragment lengths were observed between the tumor and corresponding normal DNA (study II).

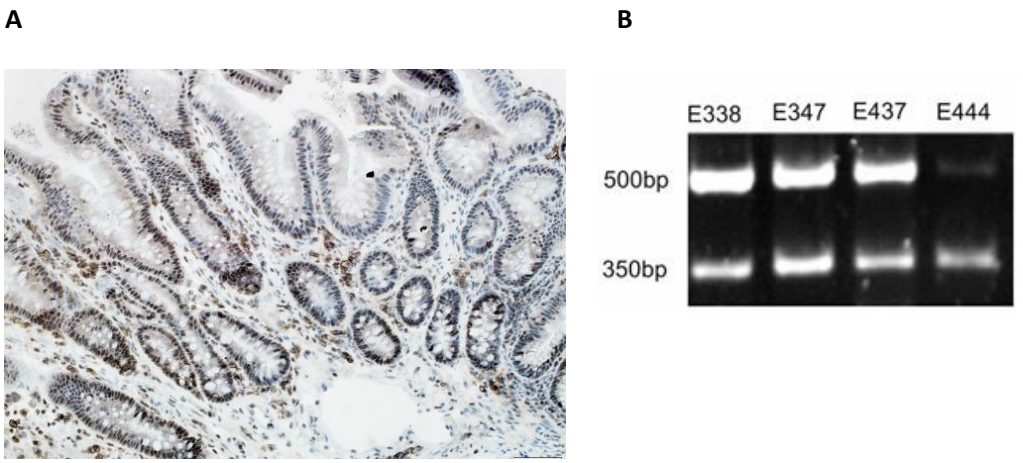

Figure 7. Mlh1 protein expression and loss of heterozygosity analyzes. (A) An example of a colon carcinoma showing positive Mlh1 expression analyzed by immunohistochemistry (mouse E347, serrated adenocarcinoma). (B) Four CRCs found in the heterozygote *Mlh1*$^{+/-}$ mice in study II showing that the normal Mlh1 allele (350bp) was still present in tumors.

Mlh1 mRNA expression is significantly decreased in normal colon mucosa of CRC mice (II)

Surprisingly, at the 12 mo time point no expression decrease was seen for *Mlh1* in normal mouse colon mucosa (Fig. 8, Fig. 9). Yet, at 18 mo time point several mice, including the mice with colonic carcinoma, showed extremely low *Mlh1* expression in their histologically normal mucosa (Fig. 9). In study I the expression was analyzed with TaqMan qPCR. In study II a genome wide transcriptome analysis was performed from 80 normal colon mucosa samples operated in 12 mo and 18 mo old mice. Analysis was performed with RNA sequencing (RNA-seq) using the single-

cell tagged reverse transcription method (STRT).[244,245] The 21 mo old mice were left out from the RNAseq study due to many health problems appearing because of their old age. Altogether 12,216 expressed transcripts were identified, including *Mlh1*. Contrary to varying *Mlh1* expression levels 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) (Fig. 9). The carcinoma mouse E347 had similar *Mlh1* expression level as mice with no CRC on average (Fig. 9).

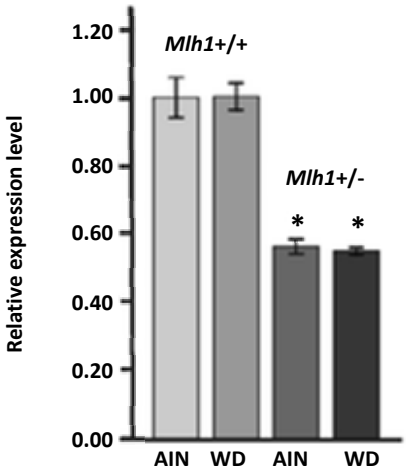

Figure 8. Relative *Mlh1* expression in different 12 mo old mouse groups ($Mlh1^{+/-}$ AIN, $Mlh1^{+/+}$ WD, and $Mlh1^{+/-}$ WD) in study I. Expression level is compared to the control group ($Mlh1^{+/+}$ AIN). Each sample is a mixture of eight RNA samples from eight different mice belonging to each mouse group. Data is presented as mean ± SEM (standard error of the mean) (n = 3), *significant difference compared to the control group. Median permutation method, $P < 0.05$. *Mlh1* shows the same 50% expression difference between the genotypes as at the starting point of the study. AIN, AIN-93G; WD, Western-style diet

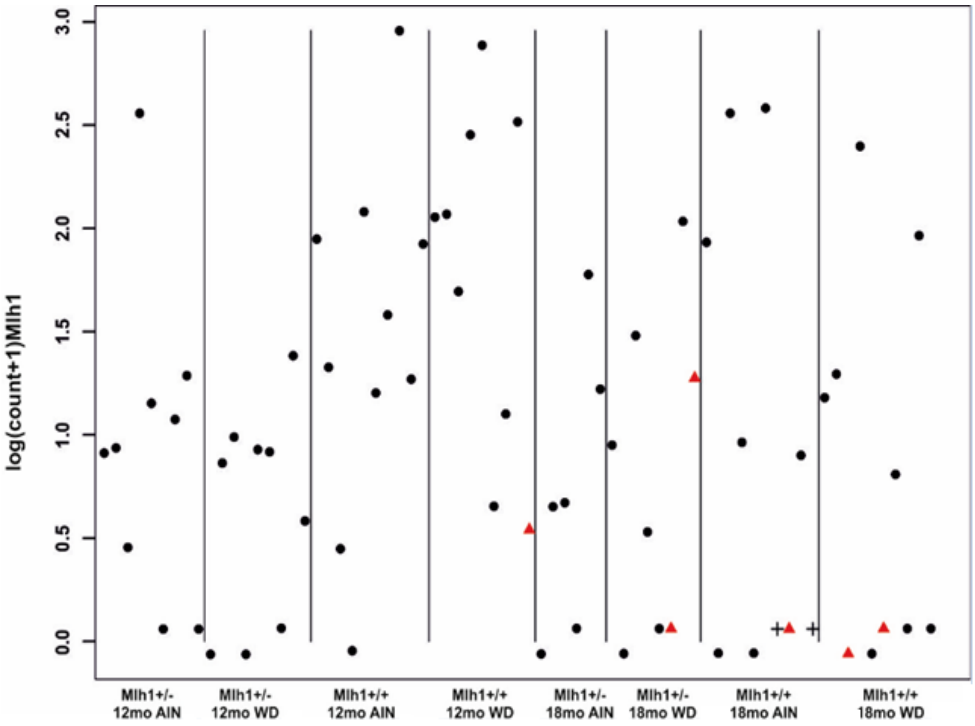

Figure 9. The *Mlh1* mRNA expression in normal mucosa in study II. The *Mlh1* gene expression levels detected in individual mice at different time points, genotypes and diet groups show that five out of six carcinoma mice (red triangle) and mice E325 and E332 (black cross) show very low Mlh1 expression. ComBat processed shifted-log values from STRT RNA-sequencing.

Expression profiles in colon mucosa form a distinct cluster for CRC mice (II)

Low *Mlh1* expression, although a prominent signal, seemed not to be 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 the genome-wide expression profiles in the six CRC mice were compared with the profiles of 74 12 and 18 mo old mice without CRC.

The normal mucosa expression profiles of the six CRC mice were strikingly different from the profiles of the other 12 and 18 mo old mice. Altogether 86% of the top 300 differentially regulated genes in CRC mice were down-regulated in normal mucosa and 14% were up-regulated as compared to other 12 and 18 mo old mice. Remarkably, the expression profiles of CRC mice formed a clearly distinct cluster (Fig. 10) as visualized by an MDS plot created with the 100 most altered/differentially regulated genes, indicating a field-defect of clonal abnormalities in the epithelial gene expression in normal colon mucosa that precede cancer development and predispose to it.[246-248]

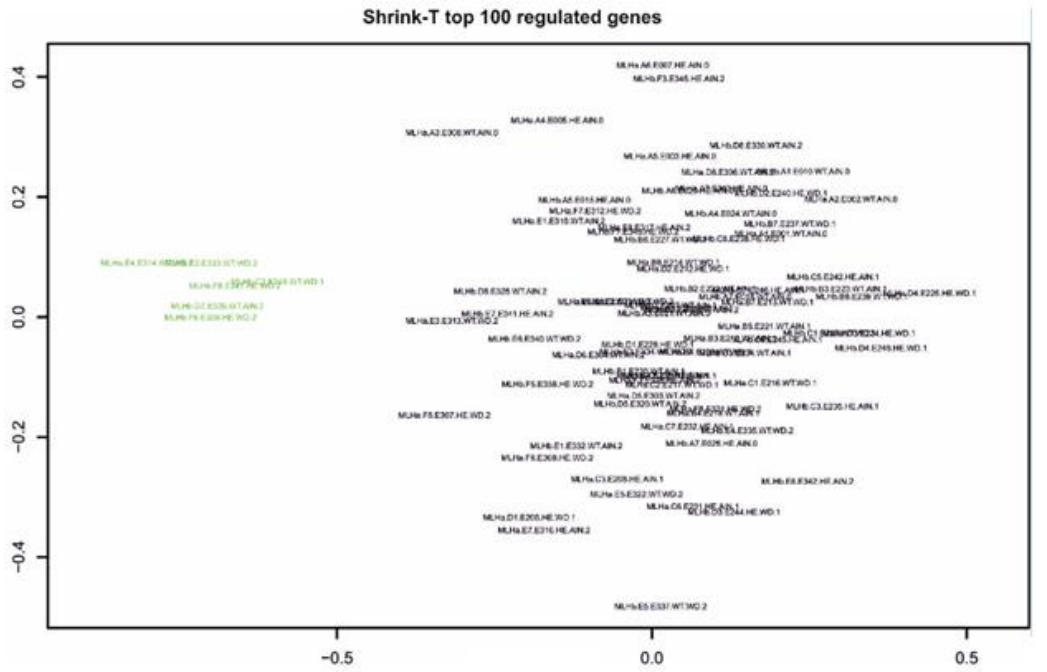

Figure 10. Genome wide expression profiles of normal colon mucosa in study II. An MDS plot created with the 100 most differentially expressed genes between CRC (green) mice and mice without cancer (black). The expression profiles of all six mice which developed carcinoma up to 18 mo of age form a distinct cluster in the plot.

CRC mice show shortage of *Mlh1* and chromosomal segregation gene transcripts in mucosa and aberrant mitoses in tumors (II)

To further understand the biological functions and pathways enriched among the top separating genes in CRC mice, the expression data were analyzed with Ingenuity Pathway Analysis (IPA). According to IPA, chromosome segregation (P = 2.92×10-5), aneuploidy of fibroblasts (P = 5.64×10-5), checkpoint control (P = 1.10×10-4), DNA replication checkpoint (P = 6.66×10-4) and morphology of mitotic spindle (P = 2.33×10-5) were among the most affected biological pathways. 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), cellular growth and proliferation (P = 3.07×10-3), and cellular compromise (P = 1.79×10-4). The analysis was also repeated with different RNA-seq data preprocessing (all mouse samples without ComBat normalization). These results confirmed the findings on chromosome segregation (P = 1.03×10-5), aneuploidy of fibroblasts (P = 4.57×10-4) and checkpoint control (P = 4.29×10-4).

The IPA results strongly indicated severe problems in cell cycle regulation and mitosis already in normal colon mucosa. In the six mice who developed carcinoma up to 18 mo in study II, along with *Mlh1*, which also triggers checkpoint activation,[71] the most differentially expressed genes that pointed to chromosome segregation and spindle assembly checkpoint (SAC) were *Bub1* (BUB1, mitotic checkpoint serine/threonine kinase), *Mis18a* (MIS18 kinetochore protein A), *Tpx2* (TPX2 microtubule associated), *Rad9a* (RAD9 checkpoint clamp component A), *Pms2* (postmeiotic segregation increased 2), *Cenpe* (centromere protein E), Ncapd3 (non-SMC condensing II complex subunit D3), *Odf2* (outer dens fiber of sperm tails 2), and *Dclre1b* (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 (Fig. 11). *Mlh1, Cenpe, Ncapd3, Odf2,* and *Dclre1b* showed variable levels of down regulation in two CRC mice (E347 and E249). In E249, *Mlh1* and *Dclr1b* showed, however, approximately 50% down regulation when compared to the average expression level in non-carcinoma mice (Fig. 11). In E347, whose carcinoma was histologically different from the others and

showed traditional serrated phenotype, *Mlh1, Cenpe, Ncapd3,* and *Odf2* expression levels were equal to non-carcinoma mice. Importantly, among all the 74 mice that did not develop colon carcinoma up to 12 or 18 mo, only two mice, E325 and E332, showed similar low expression of all the 10 genes (Fig. 11). Although, no colonic tumors were found in those mice, E325 had bloody feces and anemia, suggesting undefined mucosal pathology.

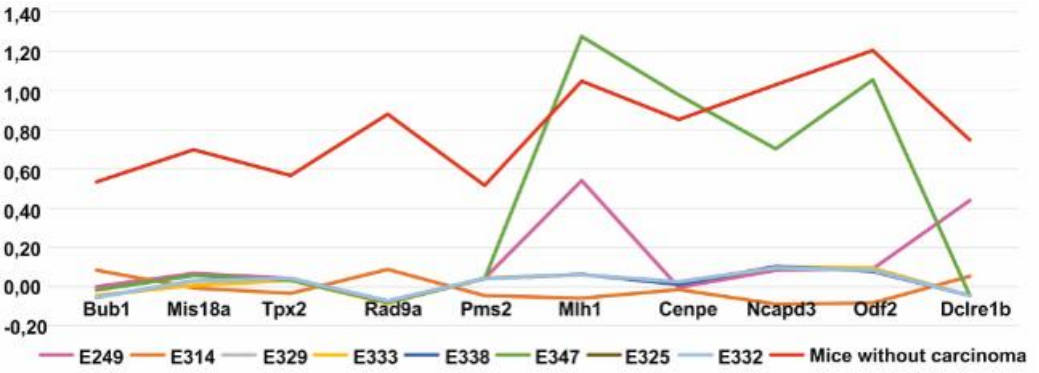

Figure 11. The expression levels of 10 chromosomal segregation-specific genes in colon mucosa in study II. The expression levels (gene expression values after ComBat processing) described as a line chart. In the carcinoma mice (E249, E314, E329, E333, E338, E347), two mice with similar expression profiles with CRC mice (E325 and E332) and the average levels of 74 mice without cancer.

Undisturbed mitosis is a central requirement of 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 chromosomal instability.[249] 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 (Fig. 12).

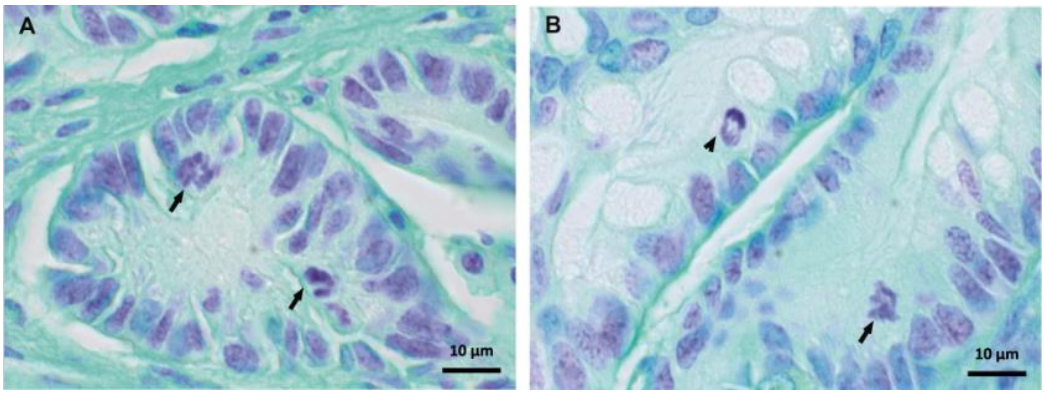

Figure 12. Abnormal mitoses in mouse colon carcinomas in study II. Representative pictures of abnormal mitoses (arrows) in (a) serrated adenocarcinoma (mouse E347) and (b) tubular adenocarcinoma (mouse E333), and a normal mitosis (arrow head).

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.[71] 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.[250] Indeed, cells with a diminished amount of MLH1 protein may still be MMR proficient, although they show defects in DNA damage signaling.[251] Consequently the damaged cells may not activate cell cycle checkpoints and enter apoptosis. The observation that low mRNA expression of *Mlh1* in carcinoma mice was associated with down regulation of several other genes related to chromosome segregation and checkpoint control, and consequent atypical mitoses in carcinomas supports the proposition that already decreased amount of Mlh1, when MMR is still functional, may be associated with aberrant chromosome segregation and play an important role in tumorigenesis.

CONCLUSIONS

These results on whether and how Western-style diet, ageing and genetic predisposition induce tumorigenesis in mouse colon provided the following conclusions:

- Western-style diet is a significant risk factor for colon carcinogenesis along with aging
- Cancer-preceding gene expression changes induced by aging, WD, and *Mlh1*$^{+/-}$ already occur in histologically normal mucosa.
- At least in some of the tumor suppressor genes the down regulation is associated with increase in the promoter hypermethylation of the respective genes.
- A decreased expression of *Mlh1* together with other chromosomal segregation genes may form a field-defect in mucosa and trigger MMR-proficient, chromosomally unstable CRC.
- *MLH1* mutation carriers develop also microsatellite stable colon cancers.
- The concomitant down regulation of *Mlh1* and several other genes related to chromosome segregation and checkpoint control supports the proposition that an already decreased amount of Mlh1, when MMR is still functional, may have an important role in tumorigenesis

FUTURE PROSPECTS

In light of the presented thesis, future prospects for the studies regarding the diet experiment may include:

In the presented thesis, several cancer-preceding gene expression changes were found already in the normal colon mucosa. However, since the observations were made in mice in order to utilize them as cancer predictive biomarkers, the findings and the significance of Mlh1 deficiency in chromosomal unstable tumors should be further validated in humans.

ACKNOWLEDGEMENTS

The work for this thesis was carried out under the supervision of Professor Minna Nyström and PhD Laura Sarantaus at the Division of Genetics, Department of Biosciences at the University of Helsinki. The work was funded by the European Research Council (2008-AdG-232635), Jenny and Antti Wihuri foundation, Jane and Aatos Erkko foundation, Cancer Society of Finland, and Sigrid Juselius Foundation.

I wish to express my sincere gratitude to both of my supervisors for a superior learning experience during these years. I am truly grateful to Professor Minna Nyström for the unique opportunity to be part of this fascinating project right from the beginning. I wish to thank her for continuous support and guidance throughout the project, and for the excellent opportunities to grow as a scientist in the field. I would like to thank PhD Laura Sarantaus for her supervision and unconditional support during my PhD project and for the valuable work done during the early years of the mouse project.

I am grateful to Professor Suvi Virtanen and adjunct Professor Minna Pöyhönen for their detailed revision of the thesis. I highly appreciate your time, and value the feedback you gave me. I also wish to warmly thank both members of my thesis committee, adjunct Professor Minna Pöyhönen and Professor Hannes Lohi for the time and valuable advice and energy during these years.

I would like to thank the Division of Genetics for providing a good working environment and functional facilities. I especially want to thank Docent Katarina Pelin for all the help and guidance and great collaboration while organizing the human genetics lab course, and Docent Pekka Heino for the help relating my PhD studies.

I wish to thank all co-authors and collaborators for their time and efforts related to our research. My special appreciation goes to Professor Päivi Peltomäki for the close collaboration and sharing her expertise in the field. I warmly thank PhD Petri Törönen for fruitful collaboration.

I wish to warmly thank Professor Marja Mutanen, PhD Essi Päivärinta, PhD Mikael Niku, and Mrs. Anu Heiman-Lindh for the excellent and pleasant collaboration in the mouse feeding experiment. Mikael Niku is also acknowledged for his help with immunohistochemistry.

Jukka, Minttu, Satu, Pauliina, Mariann, Anni, Emmi, and Laura, thank you for being such a great friends and colleagues during these years. It has always been a great pleasure to work with you guys. It is a blessing to have an open and joyful atmosphere at work.

Thank you mom, for being the best mother one could have ever hoped for and Susanna for being such a wonderful sister. The term girl power truly is personified in you two. I am grateful for your endless love, support and willingness to help no matter what, and proud of being part of our family. I hope I´m able to be at least half the mother to my child as you have been to me Liisa.

Thank you Aino and Arvo for taking me into your life and family. We have already created some unforgettable memories such as countless petrol scented weekends at car tracks and easygoing times spent in Nurmijärvi. I sincerely think you are the best possible "parents-in-law" one could have ever asked for, and I´m grateful for all the support you have given to us.

Last but not least, to my lion, Tarmo. Thank you for balancing my emotional character with a solid sense of reasoning. Thank you for your patience, for your open mind, and for always being interested to listen to my work related rambling and theories. Thank you for your love and support and for showing me a totally new way of thinking. And above all, thank you for giving me the most precious thing in my life, our darling "*Sintti*". I love you both more than words can express.

Helsinki, March 2017

Marjaana Pussila

REFERENCES

1 Brenner, H., Kloor, M. & Pox, C. P. Colorectal cancer. *Lancet* **383**, 1490-1502, doi:10.1016/S0140-6736(13)61649-9 (2014).

2 Yusof, A. S., Isa, Z. M. & Shah, S. A. Dietary patterns and risk of colorectal cancer: a systematic review of cohort studies (2000-2011). *Asian Pac J Cancer Prev* **13**, 4713-4717 (2012).

3 World Cancer Research Fund & American Institute for Cancer Research. Continuous Update Project Report. Food, Nutrition, Physical Activity, and the Prevention of Colorectal Cancer. (2011).

4 Bingham, S. & Riboli, E. Diet and cancer--the European Prospective Investigation into Cancer and Nutrition. *Nat Rev Cancer* **4**, 206-215, doi:10.1038/nrc1298 (2004).

5 Fearon, E. R. Molecular genetics of colorectal cancer. *Annu Rev Pathol* **6**, 479-507, doi:10.1146/annurev-pathol-011110-130235 (2011).

6 Feinberg, A. P., Ohlsson, R. & Henikoff, S. The epigenetic progenitor origin of human cancer. *Nat Rev Genet* **7**, 21-33, doi:10.1038/nrg1748 (2006).

7 Lao, V. V. & Grady, W. M. Epigenetics and colorectal cancer. *Nat Rev Gastroenterol Hepatol* **8**, 686-700, doi:10.1038/nrgastro.2011.173 (2011).

8 Ahuja, N. & Issa, J. P. Aging, methylation and cancer. *Histol Histopathol* **15**, 835-842 (2000).

9 Issa, J. P. *et al.* Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nat Genet* **7**, 536-540, doi:10.1038/ng0894-536 (1994).

10 Toyota, M. & Issa, J. P. CpG island methylator phenotypes in aging and cancer. *Semin Cancer Biol* **9**, 349-357, doi:10.1006/scbi.1999.0135 (1999).

11 Arasaradnam, R. P., Commane, D. M., Bradburn, D. & Mathers, J. C. A review of dietary factors and its influence on DNA methylation in colorectal carcinogenesis. *Epigenetics* **3**, 193-198 (2008).

12 Verma, M. Cancer control and prevention by nutrition and epigenetic approaches. *Antioxid Redox Signal* **17**, 355-364, doi:10.1089/ars.2011.4388 (2012).

13 Abbas, T., Keaton, M. A. & Dutta, A. Genomic instability in cancer. *Cold Spring Harb Perspect Biol* **5**, a012914, doi:10.1101/cshperspect.a012914 (2013).

14 Al-Sohaily, S., Biankin, A., Leong, R., Kohonen-Corish, M. & Warusavitarne, J. Molecular pathways in colorectal cancer. *J Gastroenterol Hepatol* **27**, 1423-1431, doi:10.1111/j.1440-1746.2012.07200.x (2012).

15 Colussi, D., Brandi, G., Bazzoli, F. & Ricciardiello, L. Molecular pathways involved in colorectal cancer: implications for disease behavior and prevention. *Int J Mol Sci* **14**, 16365-16385, doi:10.3390/ijms140816365 (2013).

16 Goel, A. *et al.* Characterization of sporadic colon cancer by patterns of genomic instability. *Cancer Res* **63**, 1608-1614 (2003).

17 Newmark, H. L. *et al.* Western-style diet-induced colonic tumors and their modulation by calcium and vitamin D in C57Bl/6 mice: a preclinical model for human sporadic colon cancer. *Carcinogenesis* **30**, 88-92, doi:10.1093/carcin/bgn229 (2009).

18 Newmark, H. L. *et al.* A Western-style diet induces benign and malignant neoplasms in the colon of normal C57Bl/6 mice. *Carcinogenesis* **22**, 1871-1875 (2001).

19 Yang, K. *et al.* Dietary induction of colonic tumors in a mouse model of sporadic colon cancer. *Cancer Res* **68**, 7803-7810, doi:10.1158/0008-5472.CAN-08-1209 (2008).

20 Yang, K. *et al.* Dietary components modify gene expression: implications for carcinogenesis. *J Nutr* **135**, 2710-2714 (2005).

21 Lee, D. H., Keum, N. & Giovannucci, E. L. Colorectal Cancer Epidemiology in the Nurses' Health Study. *Am J Public Health* **106**, 1599-1607, doi:10.2105/AJPH.2016.303320 (2016).

22 Fleming, M., Ravula, S., Tatishchev, S. F. & Wang, H. L. Colorectal carcinoma: Pathologic aspects. *J Gastrointest Oncol* **3**, 153-173, doi:10.3978/j.issn.2078-6891.2012.030 (2012).

23 Haggar, F. A. & Boushey, R. P. Colorectal cancer epidemiology: incidence, mortality, survival, and risk factors. *Clin Colon Rectal Surg* **22**, 191-197, doi:10.1055/s-0029-1242458 (2009).

24 Sankaranarayanan, R. *et al.* Cancer survival in Africa, Asia, and Central America: a population-based study. *Lancet Oncol* **11**, 165-173, doi:10.1016/S1470-2045(09)70335-3 (2010).

25 Siegel, R. *et al.* Cancer treatment and survivorship statistics, 2012. *CA Cancer J Clin* **62**, 220-241, doi:10.3322/caac.21149 (2012).

26 Levin, B. *et al.* Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *CA Cancer J Clin* **58**, 130-160, doi:10.3322/CA.2007.0018 (2008).

27 Wardle, J., Robb, K., Vernon, S. & Waller, J. Screening for prevention and early diagnosis of cancer. *Am Psychol* **70**, 119-133, doi:10.1037/a0037357 (2015).

28 Bond, J. H. Fecal occult blood test screening for colorectal cancer. *Gastrointest Endosc Clin N Am* **12**, 11-21 (2002).

29 Lieberman, D. *et al.* Screening for Colorectal Cancer and Evolving Issues for Physicians and Patients: A Review. *JAMA* **316**, 2135-2145, doi:10.1001/jama.2016.17418 (2016).

30 Bonnington, S. N. & Rutter, M. D. Surveillance of colonic polyps: Are we getting it right? *World J Gastroenterol* **22**, 1925-1934, doi:10.3748/wjg.v22.i6.1925 (2016).

31 O'Connell, J. B. *et al.* Rates of colon and rectal cancers are increasing in young adults. *Am Surg* **69**, 866-872 (2003).

32 O'Connell, J. B., Maggard, M. A., Livingston, E. H. & Yo, C. K. Colorectal cancer in the young. *Am J Surg* **187**, 343-348, doi:10.1016/j.amjsurg.2003.12.020 (2004).

33 Zhang, X. & Giovannucci, E. Calcium, vitamin D and colorectal cancer chemoprevention. *Best Pract Res Clin Gastroenterol* **25**, 485-494, doi:10.1016/j.bpg.2011.10.001 (2011).

34 Grady, W. M. Genetic testing for high-risk colon cancer patients. *Gastroenterology* **124**, 1574-1594 (2003).

35 Lichtenstein, P. *et al.* Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* **343**, 78-85, doi:10.1056/NEJM200007133430201 (2000).

36 Taylor, D. P., Burt, R. W., Williams, M. S., Haug, P. J. & Cannon-Albright, L. A. Population-based family history-specific risks for colorectal cancer: a constellation approach. *Gastroenterology* **138**, 877-885, doi:10.1053/j.gastro.2009.11.044 (2010).

37 Arvelo, F., Sojo, F. & Cotte, C. Biology of colorectal cancer. *Ecancermedicalscience* **9**, 520, doi:10.3332/ecancer.2015.520 (2015).

38 Stoffel, E. M. & Kastrinos, F. Familial colorectal cancer, beyond Lynch syndrome. *Clin Gastroenterol Hepatol* **12**, 1059-1068, doi:10.1016/j.cgh.2013.08.015 (2014).

39 Stoler, D. L. *et al.* The onset and extent of genomic instability in sporadic colorectal tumor progression. *Proc Natl Acad Sci U S A* **96**, 15121-15126 (1999).

40 Tsao, J. L. *et al.* Genetic reconstruction of individual colorectal tumor histories. *Proc Natl Acad Sci U S A* **97**, 1236-1241 (2000).

41 Simons, B. D. & Clevers, H. Stem cell self-renewal in intestinal crypt. *Exp Cell Res* **317**, 2719-2724, doi:10.1016/j.yexcr.2011.07.010 (2011).

42 Cho, K. R. & Vogelstein, B. Genetic alterations in the adenoma--carcinoma sequence. *Cancer* **70**, 1727-1731 (1992).

43 Bird, R. P. Observation and quantification of aberrant crypts in the murine colon treated with a colon carcinogen: preliminary findings. *Cancer Lett* **37**, 147-151 (1987).

44 Vogelstein, B. *et al.* Genetic alterations during colorectal-tumor development. *N Engl J Med* **319**, 525-532, doi:10.1056/NEJM198809013190901 (1988).

45 Muto, T., Bussey, H. J. & Morson, B. C. The evolution of cancer of the colon and rectum. *Cancer* **36**, 2251-2270 (1975).

46 Lynch, H. T., Snyder, C. L., Shaw, T. G., Heinen, C. D. & Hitchins, M. P. Milestones of Lynch syndrome: 1895-2015. *Nat Rev Cancer* **15**, 181-194, doi:10.1038/nrc3878 (2015).

47 Leslie, A., Carey, F. A., Pratt, N. R. & Steele, R. J. The colorectal adenoma-carcinoma sequence. *Br J Surg* **89**, 845-860, doi:10.1046/j.1365-2168.2002.02120.x (2002).

48 Boland, C. R., Shin, S. K. & Goel, A. Promoter methylation in the genesis of gastrointestinal cancer. *Yonsei Med J* **50**, 309-321, doi:10.3349/ymj.2009.50.3.309 (2009).

49 Heim, S. & Mitelman, F. Primary chromosome abnormalities in human neoplasia. *Adv Cancer Res* **52**, 1-43 (1989).

50 Kane, M. F. *et al.* Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines. *Cancer Res* **57**, 808-811 (1997).

51 Nazemalhosseini Mojarad, E., Kuppen, P. J., Aghdaei, H. A. & Zali, M. R. The CpG island methylator phenotype (CIMP) in colorectal cancer. *Gastroenterol Hepatol Bed Bench* **6**, 120-128 (2013).

52 Worthley, D. L. & Leggett, B. A. Colorectal cancer: molecular features and clinical opportunities. *Clin Biochem Rev* **31**, 31-38 (2010).

53 Boland, C. R. & Goel, A. Microsatellite instability in colorectal cancer. *Gastroenterology* **138**, 2073-2087 e2073, doi:10.1053/j.gastro.2009.12.064 (2010).

54 Negrini, S., Gorgoulis, V. G. & Halazonetis, T. D. Genomic instability--an evolving hallmark of cancer. *Nat Rev Mol Cell Biol* **11**, 220-228, doi:10.1038/nrm2858 (2010).

55 Pino, M. S. & Chung, D. C. The chromosomal instability pathway in colon cancer. *Gastroenterology* **138**, 2059-2072, doi:10.1053/j.gastro.2009.12.065 (2010).

56 Grady, W. M. Genomic instability and colon cancer. *Cancer Metastasis Rev* **23**, 11-27 (2004).

57 Rajagopalan, H. & Lengauer, C. Aneuploidy and cancer. *Nature* **432**, 338-341, doi:10.1038/nature03099 (2004).

58 Ganem, N. J., Godinho, S. A. & Pellman, D. A mechanism linking extra centrosomes to chromosomal instability. *Nature* **460**, 278-282, doi:10.1038/nature08136 (2009).

59 Jass, J. R. Pathogenesis of colorectal cancer. *Surg Clin North Am* **82**, 891-904 (2002).

60 Worthley, D. L., Whitehall, V. L., Spring, K. J. & Leggett, B. A. Colorectal carcinogenesis: road maps to cancer. *World J Gastroenterol* **13**, 3784-3791 (2007).

61 Smith, G. *et al.* Mutations in APC, Kirsten-ras, and p53--alternative genetic pathways to colorectal cancer. *Proc Natl Acad Sci U S A* **99**, 9433-9438, doi:10.1073/pnas.122612899 (2002).

62 Fodde, R. *et al.* Mutations in the APC tumour suppressor gene cause chromosomal instability. *Nat Cell Biol* **3**, 433-438, doi:10.1038/35070129 (2001).

63 Raskov, H., Pommergaard, H. C., Burcharth, J. & Rosenberg, J. Colorectal carcinogenesis--update and perspectives. *World J Gastroenterol* **20**, 18151-18164, doi:10.3748/wjg.v20.i48.18151 (2014).

64 Barras, D. BRAF Mutation in Colorectal Cancer: An Update. *Biomark Cancer* **7**, 9-12, doi:10.4137/BIC.S25248 (2015).

65 Nagasaka, T. *et al.* Colorectal cancer with mutation in BRAF, KRAS, and wild-type with respect to both oncogenes showing different patterns of DNA methylation. *J Clin Oncol* **22**, 4584-4594, doi:10.1200/JCO.2004.02.154 (2004).

66 Lynch, H. T. *et al.* Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin Genet* **76**, 1-18, doi:10.1111/j.1399-0004.2009.01230.x (2009).
67 Aaltonen, L. A. *et al.* Clues to the pathogenesis of familial colorectal cancer. *Science* **260**, 812-816 (1993).
68 Thibodeau, S. N., Bren, G. & Schaid, D. Microsatellite instability in cancer of the proximal colon. *Science* **260**, 816-819 (1993).
69 Homfray, T. F. *et al.* Defects in mismatch repair occur after APC mutations in the pathogenesis of sporadic colorectal tumours. *Human mutation* **11**, 114-120, doi:10.1002/(SICI)1098-1004(1998)11:2<114::AID-HUMU3>3.0.CO;2-J (1998).
70 Parsons, R. *et al.* Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell* **75**, 1227-1236 (1993).
71 Cejka, P. *et al.* Methylation-induced G(2)/M arrest requires a full complement of the mismatch repair protein hMLH1. *The EMBO journal* **22**, 2245-2254, doi:10.1093/emboj/cdg216 (2003).
72 Toyota, M. *et al.* CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A* **96**, 8681-8686 (1999).
73 Umar, A. *et al.* Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst* **96**, 261-268 (2004).
74 Feinberg, A. P. & Vogelstein, B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301**, 89-92 (1983).
75 Ostertag, E. M. & Kazazian, H. H., Jr. Biology of mammalian L1 retrotransposons. *Annu Rev Genet* **35**, 501-538, doi:10.1146/annurev.genet.35.102401.091032 (2001).
76 Chalitchagorn, K. *et al.* Distinctive pattern of LINE-1 methylation level in normal tissues and the association with carcinogenesis. *Oncogene* **23**, 8841-8846, doi:10.1038/sj.onc.1208137 (2004).
77 Gonzalgo, M. L. & Jones, P. A. Mutagenic and epigenetic effects of DNA methylation. *Mutat Res* **386**, 107-118 (1997).
78 Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* **13**, 335-340 (1997).
79 Razin, A. & Kafri, T. DNA methylation from embryo to adult. *Prog Nucleic Acid Res Mol Biol* **48**, 53-81 (1994).
80 Crider, K. S., Yang, T. P., Berry, R. J. & Bailey, L. B. Folate and DNA methylation: a review of molecular mechanisms and the evidence for folate's role. *Adv Nutr* **3**, 21-38, doi:10.3945/an.111.000992 (2012).
81 Ehrlich, M. DNA methylation in cancer: too much, but also too little. *Oncogene* **21**, 5400-5413, doi:10.1038/sj.onc.1205651 (2002).
82 Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev* **25**, 1010-1022, doi:10.1101/gad.2037511 (2011).

83 Weisenberger, D. J. *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet* **38**, 787-793, doi:10.1038/ng1834 (2006).

84 Ahuja, N., Li, Q., Mohan, A. L., Baylin, S. B. & Issa, J. P. Aging and DNA methylation in colorectal mucosa and cancer. *Cancer Res* **58**, 5489-5494 (1998).

85 Kemppainen, M., Raiha, I., Rajala, T. & Sourander, L. Characteristics of colorectal cancer in elderly patients. *Gerontology* **39**, 222-227 (1993).

86 SEER Program (National Cancer Institute (U.S.)), National Institutes of Health (U.S.) & National Cancer Institute (U.S.). Surveillance Program. in *NIH publication* v. (U.S. Dept. of Health and Human Services, Public Health Institute, Bethesda, Md.).

87 MacGillivray, D. C., Swartz, S. E., Robinson, A. M., Cruess, D. F. & Smith, L. E. Adenocarcinoma of the colon and rectum in patients less than 40 years of age. *Surg Gynecol Obstet* **172**, 1-7 (1991).

88 Milholland, B., Auton, A., Suh, Y. & Vijg, J. Age-related somatic mutations in the cancer genome. *Oncotarget* **6**, 24627-24635, doi:10.18632/oncotarget.5685 (2015).

89 Gorbunova, V., Seluanov, A., Mao, Z. & Hine, C. Changes in DNA repair during aging. *Nucleic Acids Res* **35**, 7466-7474, doi:10.1093/nar/gkm756 (2007).

90 Tollefsbol, T. O. Dietary epigenetics in cancer and aging. *Cancer Treat Res* **159**, 257-267, doi:10.1007/978-3-642-38007-5_15 (2014).

91 Chan, A. O. *et al.* CpG island methylation in aberrant crypt foci of the colorectum. *Am J Pathol* **160**, 1823-1830, doi:10.1016/S0002-9440(10)61128-5 (2002).

92 Issa, J. P. Cancer prevention: epigenetics steps up to the plate. *Cancer Prev Res (Phila)* **1**, 219-222, doi:10.1158/1940-6207.CAPR-08-0029 (2008).

93 Herceg, Z. Epigenetics and cancer: towards an evaluation of the impact of environmental and dietary factors. *Mutagenesis* **22**, 91-103, doi:10.1093/mutage/gel068 (2007).

94 Aslam, M. N. & Varani, J. The Western-Style Diet, Calcium Deficiency and Chronic Disease. *Journal of Nutrition & Food Sciences*, doi:10.4172/2155-9600.1000496 (2016).

95 Armelao, F. & de Pretis, G. Familial colorectal cancer: a review. *World J Gastroenterol* **20**, 9292-9298, doi:10.3748/wjg.v20.i28.9292 (2014).

96 Southey, M. C. *et al.* Use of molecular tumor characteristics to prioritize mismatch repair gene testing in early-onset colorectal cancer. *J Clin Oncol* **23**, 6524-6532, doi:10.1200/JCO.2005.04.671 (2005).

97 Wagner, A. *et al.* Molecular analysis of hereditary nonpolyposis colorectal cancer in the United States: high mutation detection rate among clinically selected families and characterization of an American founder genomic deletion of the MSH2 gene. *Am J Hum Genet* **72**, 1088-1100, doi:10.1086/373963 (2003).

98 Tiwari, A. K., Roy, H. K. & Lynch, H. T. Lynch syndrome in the 21st century: clinical perspectives. *QJM* **109**, 151-158, doi:10.1093/qjmed/hcv137 (2016).

99 Lynch, H. T. & de la Chapelle, A. Hereditary colorectal cancer. *N Engl J Med* **348**, 919-932, doi:10.1056/NEJMra012242 (2003).
100 Stoffel, E. *et al.* Calculation of risk of colorectal and endometrial cancer among patients with Lynch syndrome. *Gastroenterology* **137**, 1621-1627, doi:10.1053/j.gastro.2009.07.039 (2009).
101 Jasperson, K. W., Tuohy, T. M., Neklason, D. W. & Burt, R. W. Hereditary and familial colon cancer. *Gastroenterology* **138**, 2044-2058, doi:10.1053/j.gastro.2010.01.054 (2010).
102 Barrow, E., Hill, J. & Evans, D. G. Cancer risk in Lynch Syndrome. *Fam Cancer* **12**, 229-240, doi:10.1007/s10689-013-9615-1 (2013).
103 Watson, P. & Lynch, H. T. The tumor spectrum in HNPCC. *Anticancer Res* **14**, 1635-1639 (1994).
104 Watson, P. & Riley, B. The tumor spectrum in the Lynch syndrome. *Fam Cancer* **4**, 245-248, doi:10.1007/s10689-004-7994-z (2005).
105 Lynch, H. T. *et al.* Phenotypic and genotypic heterogeneity in the Lynch syndrome: diagnostic, surveillance and management implications. *Eur J Hum Genet* **14**, 390-402, doi:10.1038/sj.ejhg.5201584 (2006).
106 Jass, J. R. *et al.* Pathology of hereditary non-polyposis colorectal cancer. *Anticancer Res* **14**, 1631-1634 (1994).
107 Jass, J. R. Pathology of hereditary nonpolyposis colorectal cancer. *Ann N Y Acad Sci* **910**, 62-73; discussion 73-64 (2000).
108 Jang, E. & Chung, D. C. Hereditary colon cancer: lynch syndrome. *Gut Liver* **4**, 151-160, doi:10.5009/gnl.2010.4.2.151 (2010).
109 Ribic, C. M. *et al.* Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. *N Engl J Med* **349**, 247-257, doi:10.1056/NEJMoa022289 (2003).
110 Dolcetti, R. *et al.* High prevalence of activated intraepithelial cytotoxic T lymphocytes and increased neoplastic cell apoptosis in colorectal carcinomas with microsatellite instability. *Am J Pathol* **154**, 1805-1813, doi:10.1016/S0002-9440(10)65436-3 (1999).
111 Peltomaki, P., Gao, X. & Mecklin, J. P. Genotype and phenotype in hereditary nonpolyposis colon cancer: a study of families with different vs. shared predisposing mutations. *Fam Cancer* **1**, 9-15 (2001).
112 Vasen, H. F., Mecklin, J. P., Khan, P. M. & Lynch, H. T. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). *Dis Colon Rectum* **34**, 424-425 (1991).
113 Vasen, H. F., Watson, P., Mecklin, J. P. & Lynch, H. T. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology* **116**, 1453-1456 (1999).
114 Pinol, V. *et al.* Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary

nonpolyposis colorectal cancer. *JAMA* **293**, 1986-1994, doi:10.1001/jama.293.16.1986 (2005).

115 Jung, I., Gurzu, S. & Turdean, G. S. Current status of familial gastrointestinal polyposis syndromes. *World J Gastrointest Oncol* **7**, 347-355, doi:10.4251/wjgo.v7.i11.347 (2015).

116 Nieuwenhuis, M. H. *et al.* Genotype predicting phenotype in familial adenomatous polyposis: a practical application to the choice of surgery. *Dis Colon Rectum* **52**, 1259-1263, doi:10.1007/DCR.0b013e3181a0d33b (2009).

117 Anaya, D. A., Chang, G. J. & Rodriguez-Bigas, M. A. Extracolonic manifestations of hereditary colorectal cancer syndromes. *Clin Colon Rectal Surg* **21**, 263-272, doi:10.1055/s-0028-1089941 (2008).

118 Burt, R. W. *et al.* Genetic testing and phenotype in a large kindred with attenuated familial adenomatous polyposis. *Gastroenterology* **127**, 444-451 (2004).

119 Knudsen, A. L., Bisgaard, M. L. & Bulow, S. Attenuated familial adenomatous polyposis (AFAP). A review of the literature. *Fam Cancer* **2**, 43-55 (2003).

120 Al-Tassan, N. *et al.* Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. *Nat Genet* **30**, 227-232, doi:10.1038/ng828 (2002).

121 Leoz, M. L., Carballal, S., Moreira, L., Ocana, T. & Balaguer, F. The genetic basis of familial adenomatous polyposis and its implications for clinical practice and risk management. *Appl Clin Genet* **8**, 95-107, doi:10.2147/TACG.S51484 (2015).

122 Brosens, L. A., Langeveld, D., van Hattem, W. A., Giardiello, F. M. & Offerhaus, G. J. Juvenile polyposis syndrome. *World J Gastroenterol* **17**, 4839-4844, doi:10.3748/wjg.v17.i44.4839 (2011).

123 Hoffenberg, E. J., Sauaia, A., Maltzman, T., Knoll, K. & Ahnen, D. J. Symptomatic colonic polyps in childhood: not so benign. *J Pediatr Gastroenterol Nutr* **28**, 175-181 (1999).

124 Jelsig, A. M. *et al.* Hamartomatous polyposis syndromes: a review. *Orphanet J Rare Dis* **9**, 101, doi:10.1186/1750-1172-9-101 (2014).

125 Howe, J. R., Mitros, F. A. & Summers, R. W. The risk of gastrointestinal carcinoma in familial juvenile polyposis. *Ann Surg Oncol* **5**, 751-756 (1998).

126 van Lier, M. G. *et al.* High cancer risk in Peutz-Jeghers syndrome: a systematic review and surveillance recommendations. *Am J Gastroenterol* **105**, 1258-1264; author reply 1265, doi:10.1038/ajg.2009.725 (2010).

127 Resta, N. *et al.* Cancer risk associated with STK11/LKB1 germline mutations in Peutz-Jeghers syndrome patients: results of an Italian multicenter study. *Dig Liver Dis* **45**, 606-611, doi:10.1016/j.dld.2012.12.018 (2013).

128 Yeoman, A., Young, J., Arnold, J., Jass, J. & Parry, S. Hyperplastic polyposis in the New Zealand population: a condition associated with increased colorectal cancer risk and European ancestry. *N Z Med J* **120**, U2827 (2007).

129 Boparai, K. S. *et al.* Increased colorectal cancer risk during follow-up in patients with hyperplastic polyposis syndrome: a multicentre cohort study. *Gut* **59**, 1094-1100, doi:10.1136/gut.2009.185884 (2010).
130 Chow, E. *et al.* Hyperplastic polyposis syndrome: phenotypic presentations and the role of MBD4 and MYH. *Gastroenterology* **131**, 30-39, doi:10.1053/j.gastro.2006.03.046 (2006).
131 Minoo, P. *et al.* Extensive DNA methylation in normal colorectal mucosa in hyperplastic polyposis. *Gut* **55**, 1467-1474, doi:10.1136/gut.2005.082859 (2006).
132 Wiseman, M. The second World Cancer Research Fund/American Institute for Cancer Research expert report. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. *Proc Nutr Soc* **67**, 253-256, doi:10.1017/S002966510800712X (2008).
133 Willett, W. C. Diet, nutrition, and avoidable cancer. *Environ Health Perspect* **103 Suppl 8**, 165-170 (1995).
134 Slattery, M. L. *et al.* Diet and lifestyle factor associations with CpG island methylator phenotype and BRAF mutations in colon cancer. *Int J Cancer* **120**, 656-663, doi:10.1002/ijc.22342 (2007).
135 Weinstein, S. J. *et al.* Null association between prostate cancer and serum folate, vitamin B(6), vitamin B(12), and homocysteine. *Cancer Epidemiol Biomarkers Prev* **12**, 1271-1272 (2003).
136 Doll, R. & Peto, R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J Natl Cancer Inst* **66**, 1191-1308 (1981).
137 Lipkin, M. *et al.* Preclinical mouse models for cancer chemoprevention studies. *Ann N Y Acad Sci* **889**, 14-19 (1999).
138 Lamprecht, S. A. & Lipkin, M. Chemoprevention of colon cancer by calcium, vitamin D and folate: molecular mechanisms. *Nat Rev Cancer* **3**, 601-614, doi:10.1038/nrc1144 (2003).
139 Bingham, S. A. *et al.* Dietary fibre in food and protection against colorectal cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC): an observational study. *Lancet* **361**, 1496-1501 (2003).
140 Jensen, O. M., MacLennan, R. & Wahrendorf, J. Diet, bowel function, fecal characteristics, and large bowel cancer in Denmark and Finland. *Nutr Cancer* **4**, 5-19, doi:10.1080/01635588209513733 (1982).
141 Reddy, B. S. *et al.* Metabolic epidemiology of colon cancer: effect of dietary fiber on fecal mutagens and bile acids in healthy subjects. *Cancer Res* **47**, 644-648 (1987).
142 Howe, G. R. *et al.* Dietary intake of fiber and decreased risk of cancers of the colon and rectum: evidence from the combined analysis of 13 case-control studies. *J Natl Cancer Inst* **84**, 1887-1896 (1992).
143 Trock, B., Lanza, E. & Greenwald, P. Dietary fiber, vegetables, and colon cancer: critical review and meta-analyses of the epidemiologic evidence. *J Natl Cancer Inst* **82**, 650-661 (1990).

144 Heerdt, B. G., Houston, M. A. & Augenlicht, L. H. Short-chain fatty acid-initiated cell cycle arrest and apoptosis of colonic epithelial cells is linked to mitochondrial function. *Cell Growth Differ* **8**, 523-532 (1997).
145 Lamprecht, S. A. & Lipkin, M. Cellular mechanisms of calcium and vitamin D in the inhibition of colorectal carcinogenesis. *Ann N Y Acad Sci* **952**, 73-87 (2001).
146 Heaney, R. P. Vitamin D and calcium interactions: functional outcomes. *Am J Clin Nutr* **88**, 541S-544S (2008).
147 Chen, J. *et al.* A methylenetetrahydrofolate reductase polymorphism and the risk of colorectal cancer. *Cancer Res* **56**, 4862-4864 (1996).
148 Ferrari, P. *et al.* Lifetime and baseline alcohol intake and risk of colon and rectal cancers in the European prospective investigation into cancer and nutrition (EPIC). *Int J Cancer* **121**, 2065-2072, doi:10.1002/ijc.22966 (2007).
149 Fedirko, V. *et al.* Alcohol drinking and colorectal cancer risk: an overall and dose-response meta-analysis of published studies. *Ann Oncol* **22**, 1958-1972, doi:10.1093/annonc/mdq653 (2011).
150 Boffetta, P. & Hashibe, M. Alcohol and cancer. *Lancet Oncol* **7**, 149-156, doi:10.1016/S1470-2045(06)70577-0 (2006).
151 Haas, S. L., Ye, W. & Lohr, J. M. Alcohol consumption and digestive tract cancer. *Curr Opin Clin Nutr Metab Care* **15**, 457-467, doi:10.1097/MCO.0b013e3283566699 (2012).
152 Schutze, M. *et al.* Alcohol attributable burden of incidence of cancer in eight European countries based on results from prospective cohort study. *BMJ* **342**, d1584, doi:10.1136/bmj.d1584 (2011).
153 Wong, H. P. *et al.* Nicotine promotes cell proliferation via alpha7-nicotinic acetylcholine receptor and catecholamine-synthesizing enzymes-mediated pathway in human colon adenocarcinoma HT-29 cells. *Toxicol Appl Pharmacol* **221**, 261-267, doi:10.1016/j.taap.2007.04.002 (2007).
154 Ye, Y. N. *et al.* Nicotine promoted colon cancer growth via epidermal growth factor receptor, c-Src, and 5-lipoxygenase-mediated signal pathway. *J Pharmacol Exp Ther* **308**, 66-72, doi:10.1124/jpet.103.058321 (2004).
155 Ye, Y. N., Wu, W. K., Shin, V. Y. & Cho, C. H. A mechanistic study of colon cancer growth promoted by cigarette smoke extract. *Eur J Pharmacol* **519**, 52-57, doi:10.1016/j.ejphar.2005.07.009 (2005).
156 Jensen, K., Afroze, S., Munshi, M. K., Guerrier, M. & Glaser, S. S. Mechanisms for nicotine in the development and progression of gastrointestinal cancers. *Transl Gastrointest Cancer* **1**, 81-87, doi:10.3978/j.issn.2224-4778.2011.12.01 (2012).
157 Hurt, R. T., Kulisek, C., Buchanan, L. A. & McClave, S. A. The obesity epidemic: challenges, health initiatives, and implications for gastroenterologists. *Gastroenterol Hepatol (N Y)* **6**, 780-792 (2010).
158 Malik, V. S., Willett, W. C. & Hu, F. B. Global obesity: trends, risk factors and policy implications. *Nat Rev Endocrinol* **9**, 13-27, doi:10.1038/nrendo.2012.199 (2013).

159 Cust, A. E. *et al.* Total dietary carbohydrate, sugar, starch and fibre intakes in the European Prospective Investigation into Cancer and Nutrition. *Eur J Clin Nutr* **63 Suppl 4**, S37-60, doi:10.1038/ejcn.2009.74 (2009).
160 Satia-Abouta, J. *et al.* Associations of total energy and macronutrients with colon cancer risk in African Americans and Whites: results from the North Carolina colon cancer study. *Am J Epidemiol* **158**, 951-962 (2003).
161 Bardou, M., Barkun, A. N. & Martel, M. Obesity and colorectal cancer. *Gut* **62**, 933-947, doi:10.1136/gutjnl-2013-304701 (2013).
162 Boyle, P. & Leon, M. E. Epidemiology of colorectal cancer. *Br Med Bull* **64**, 1-25 (2002).
163 Shephard, R. J. Exercise in the prevention and treatment of cancer. An update. *Sports Med* **15**, 258-280 (1993).
164 Martinez, M. E. *et al.* Leisure-time physical activity, body size, and colon cancer in women. Nurses' Health Study Research Group. *J Natl Cancer Inst* **89**, 948-955 (1997).
165 Thune, I. & Lund, E. Physical activity and risk of colorectal cancer in men and women. *Br J Cancer* **73**, 1134-1140 (1996).
166 Vena, J. E., Graham, S., Zielezny, M., Brasure, J. & Swanson, M. K. Occupational exercise and risk of cancer. *Am J Clin Nutr* **45**, 318-327 (1987).
167 Sanchez, N. F. *et al.* Physical activity reduces risk for colon polyps in a multiethnic colorectal cancer screening population. *BMC Res Notes* **5**, 312, doi:10.1186/1756-0500-5-312 (2012).
168 McTiernan A Cancer Prevention and Management Through Exercise and Weight Control. Taylor & Francis Group. (2006).
169 Tarraga Lopez, P. J., Albero, J. S. & Rodriguez-Montes, J. A. Primary and secondary prevention of colorectal cancer. *Clin Med Insights Gastroenterol* **7**, 33-46, doi:10.4137/CGast.S14039 (2014).
170 Wolin, K. Y., Yan, Y., Colditz, G. A. & Lee, I. M. Physical activity and colon cancer prevention: a meta-analysis. *Br J Cancer* **100**, 611-616, doi:10.1038/sj.bjc.6604917 (2009).
171 Baena, R. & Salinas, P. Diet and colorectal cancer. *Maturitas* **80**, 258-264, doi:10.1016/j.maturitas.2014.12.017 (2015).
172 Anand, P. *et al.* Cancer is a preventable disease that requires major lifestyle changes. *Pharm Res* **25**, 2097-2116, doi:10.1007/s11095-008-9661-9 (2008).
173 Johnson, R. L. & Fleet, J. C. Animal models of colorectal cancer. *Cancer Metastasis Rev* **32**, 39-61, doi:10.1007/s10555-012-9404-6 (2013).
174 Rosenberg, D. W., Giardina, C. & Tanaka, T. Mouse models for the study of colon carcinogenesis. *Carcinogenesis* **30**, 183-196, doi:10.1093/carcin/bgn267 (2009).
175 Felton, K. E., Gilchrist, D. M. & Andrew, S. E. Constitutive deficiency in DNA mismatch repair. *Clin Genet* **71**, 483-498, doi:10.1111/j.1399-0004.2007.00803.x (2007).
176 Moser, A. R., Pitot, H. C. & Dove, W. F. A dominant mutation that predisposes to multiple intestinal neoplasia in the mouse. *Science* **247**, 322-324 (1990).

177 Su, L. K. *et al.* Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. *Science* **256**, 668-670 (1992).
178 Fodde, R. *et al.* A targeted chain-termination mutation in the mouse Apc gene results in multiple intestinal tumors. *Proc Natl Acad Sci U S A* **91**, 8969-8973 (1994).
179 Oshima, M. *et al.* Loss of Apc heterozygosity and abnormal tissue building in nascent intestinal polyps in mice carrying a truncated Apc gene. *Proc Natl Acad Sci U S A* **92**, 4482-4486 (1995).
180 Moser, A. R. *et al.* ApcMin: a mouse model for intestinal and mammary tumorigenesis. *Eur J Cancer* **31A**, 1061-1064 (1995).
181 Karim, B. O. & Huso, D. L. Mouse models for colorectal cancer. *Am J Cancer Res* **3**, 240-250 (2013).
182 Caldwell, C. M., Green, R. A. & Kaplan, K. B. APC mutations lead to cytokinetic failures in vitro and tetraploid genotypes in Min mice. *J Cell Biol* **178**, 1109-1120, doi:10.1083/jcb.200703186 (2007).
183 Reitmair, A. H. *et al.* Spontaneous intestinal carcinomas and skin neoplasms in Msh2-deficient mice. *Cancer Res* **56**, 3842-3849 (1996).
184 de Wind, N., Dekker, M., Berns, A., Radman, M. & te Riele, H. Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. *Cell* **82**, 321-330 (1995).
185 Baker, S. M. *et al.* Involvement of mouse Mlh1 in DNA mismatch repair and meiotic crossing over. *Nat Genet* **13**, 336-342, doi:10.1038/ng0796-336 (1996).
186 Edelmann, W. *et al.* Meiotic pachytene arrest in MLH1-deficient mice. *Cell* **85**, 1125-1134 (1996).
187 Edelmann, W. *et al.* Tumorigenesis in Mlh1 and Mlh1/Apc1638N mutant mice. *Cancer Res* **59**, 1301-1307 (1999).
188 Reitmair, A. H. *et al.* MSH2 deficiency contributes to accelerated APC-mediated intestinal tumorigenesis. *Cancer Res* **56**, 2922-2926 (1996).
189 Edelmann, W. *et al.* Mutation in the mismatch repair gene Msh6 causes cancer susceptibility. *Cell* **91**, 467-477 (1997).
190 de Wind, N. *et al.* HNPCC-like cancer predisposition in mice through simultaneous loss of Msh3 and Msh6 mismatch-repair protein functions. *Nat Genet* **23**, 359-362, doi:10.1038/15544 (1999).
191 Wijnen, J. *et al.* Familial endometrial cancer in female carriers of MSH6 germline mutations. *Nat Genet* **23**, 142-144, doi:10.1038/13773 (1999).
192 Kolodner, R. D. *et al.* Germ-line msh6 mutations in colorectal cancer families. *Cancer Res* **59**, 5068-5074 (1999).
193 Tammariello, A. E. & Milner, J. A. Mouse models for unraveling the importance of diet in colon cancer prevention. *The Journal of nutritional biochemistry* **21**, 77-88, doi:10.1016/j.jnutbio.2009.09.014 (2010).
194 Wasan, H. S., Novelli, M., Bee, J. & Bodmer, W. F. Dietary fat influences on polyp phenotype in multiple intestinal neoplasia mice. *Proc Natl Acad Sci U S A* **94**, 3308-3313 (1997).

195 Johnson, I. T. & Lund, E. K. Review article: nutrition, obesity and colorectal cancer. *Aliment Pharmacol Ther* **26**, 161-181, doi:10.1111/j.1365-2036.2007.03371.x (2007).

196 Mai, V. *et al.* Calorie restriction and diet composition modulate spontaneous intestinal tumorigenesis in Apc(Min) mice through different mechanisms. *Cancer Res* **63**, 1752-1755 (2003).

197 Hursting, S. D., Lavigne, J. A., Berrigan, D., Perkins, S. N. & Barrett, J. C. Calorie restriction, aging, and cancer prevention: mechanisms of action and applicability to humans. *Annu Rev Med* **54**, 131-152, doi:10.1146/annurev.med.54.101601.152156 (2003).

198 Petrik, M. B., McEntee, M. F., Johnson, B. T., Obukowicz, M. G. & Whelan, J. Highly unsaturated (n-3) fatty acids, but not alpha-linolenic, conjugated linoleic or gamma-linolenic acids, reduce tumorigenesis in Apc(Min/+) mice. *J Nutr* **130**, 2434-2443 (2000).

199 Bartram, H. P. *et al.* Missing anti-proliferative effect of fish oil on rectal epithelium in healthy volunteers consuming a high-fat diet: potential role of the n-3:n-6 fatty acid ratio. *Eur J Cancer Prev* **4**, 231-237 (1995).

200 Hu, R. *et al.* Cancer chemoprevention of intestinal polyposis in ApcMin/+ mice by sulforaphane, a natural product derived from cruciferous vegetable. *Carcinogenesis* **27**, 2038-2046, doi:10.1093/carcin/bgl049 (2006).

201 Mahmoud, N. N. *et al.* Plant phenolics decrease intestinal tumors in an animal model of familial adenomatous polyposis. *Carcinogenesis* **21**, 921-927 (2000).

202 Clarke, J. D., Dashwood, R. H. & Ho, E. Multi-targeted prevention of cancer by sulforaphane. *Cancer Lett* **269**, 291-304, doi:10.1016/j.canlet.2008.04.018 (2008).

203 Liu, K. C. *et al.* Sulforaphane Induces Cell Death Through G2/M Phase Arrest and Triggers Apoptosis in HCT 116 Human Colon Cancer Cells. *Am J Chin Med* **44**, 1289-1310, doi:10.1142/S0192415X16500725 (2016).

204 Greaney, A. J., Maier, N. K., Leppla, S. H. & Moayeri, M. Sulforaphane inhibits multiple inflammasomes through an Nrf2-independent mechanism. *J Leukoc Biol* **99**, 189-199, doi:10.1189/jlb.3A0415-155RR (2016).

205 Frydoonfar, H. R., McGrath, D. R. & Spigelman, A. D. Sulforaphane inhibits growth of a colon cancer cell line. *Colorectal Dis* **6**, 28-31 (2004).

206 Yanaka, A. Role of Sulforaphane in Protection of Gastrointestinal Tract against H.pylori- and NSAID-Induced Oxidative Stress. *Curr Pharm Des*, doi:10.2174/1381612823666170207103943 (2017).

207 Hatcher, H., Planalp, R., Cho, J., Torti, F. M. & Torti, S. V. Curcumin: from ancient medicine to current clinical trials. *Cell Mol Life Sci* **65**, 1631-1652, doi:10.1007/s00018-008-7452-4 (2008).

208 Rajitha, B. *et al.* Inhibition of NF-kappaB translocation by curcumin analogs induces G0/G1 arrest and downregulates thymidylate synthase in colorectal cancer. *Cancer Lett* **373**, 227-233, doi:10.1016/j.canlet.2016.01.052 (2016).

209 Karimian, M. S., Pirro, M., Johnston, T. P., Majeed, M. & Sahebkar, A. Curcumin and Endothelial Function: Evidence and Mechanisms of Protective Effects. *Curr Pharm Des*, doi:10.2174/1381612823666170222122822 (2017).

210 Niho, N. *et al.* Inhibition of intestinal carcinogenesis by a new flavone derivative, chafuroside, in oolong tea. *Cancer Sci* **97**, 248-251, doi:10.1111/j.1349-7006.2006.00167.x (2006).

211 Landis-Piwowar, K. R., Milacic, V. & Dou, Q. P. Relationship between the methylation status of dietary flavonoids and their growth-inhibitory and apoptosis-inducing activities in human cancer cells. *J Cell Biochem* **105**, 514-523, doi:10.1002/jcb.21853 (2008).

212 Washington, M. K. *et al.* Pathology of rodent models of intestinal cancer: progress report and recommendations. *Gastroenterology* **144**, 705-717, doi:10.1053/j.gastro.2013.01.067 (2013).

213 Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol* **11**, R106, doi:10.1186/gb-2010-11-10-r106 (2010).

214 Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127, doi:10.1093/biostatistics/kxj037 (2007).

215 Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* **15**, R29, doi:10.1186/gb-2014-15-2-r29 (2014).

216 Kayala, M. A. & Baldi, P. Cyber-T web server: differential analysis of high-throughput data. *Nucleic Acids Res* **40**, W553-559, doi:10.1093/nar/gks420 (2012).

217 Opgen-Rhein, R. & Strimmer, K. Accurate ranking of differentially expressed genes by a distribution-free shrinkage approach. *Stat Appl Genet Mol Biol* **6**, Article9, doi:10.2202/1544-6115.1252 (2007).

218 Mishra, P., Toronen, P., Leino, Y. & Holm, L. Gene set analysis: limitations in popular existing methods and proposed improvements. *Bioinformatics* **30**, 2747-2756, doi:10.1093/bioinformatics/btu374 (2014).

219 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).

220 Coolen, M. W., Statham, A. L., Gardiner-Garden, M. & Clark, S. J. Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. *Nucleic Acids Res* **35**, e119, doi:10.1093/nar/gkm662 (2007).

221 Akilesh, S., Shaffer, D. J. & Roopenian, D. Customized molecular phenotyping by quantitative gene expression and pattern recognition analysis. *Genome Res* **13**, 1719-1727, doi:10.1101/gr.533003 (2003).

222 Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408, doi:10.1006/meth.2001.1262 (2001).

223 Efron, B., Tibshirani, R., Storey, J. D. & Tusher, V. Empirical Bayes analysis of a microarray experiment. *J Am Stat Assoc* **96**, 1151-1160, doi:Doi 10.1198/016214501753382129 (2001).

224 Kallio, M. A. *et al.* Chipster: user-friendly analysis software for microarray and other high-throughput data. *BMC Genomics* **12**, 507, doi:10.1186/1471-2164-12-507 (2011).

225 Gregorieff, A. & Clevers, H. Wnt signaling in the intestinal epithelium: from endoderm to cancer. *Genes Dev* **19**, 877-890, doi:10.1101/gad.1295405 (2005).

226 Pendas-Franco, N., Aguilera, O., Pereira, F., Gonzalez-Sancho, J. M. & Munoz, A. Vitamin D and Wnt/beta-catenin pathway in colon cancer: role and regulation of DICKKOPF genes. *Anticancer Res* **28**, 2613-2623 (2008).

227 Burgess, A. W., Faux, M. C., Layton, M. J. & Ramsay, R. G. Wnt signaling and colon tumorigenesis--a view from the periphery. *Exp Cell Res* **317**, 2748-2758, doi:10.1016/j.yexcr.2011.08.010 (2011).

228 Schneider, M. R. *et al.* A key role for E-cadherin in intestinal homeostasis and Paneth cell maturation. *PloS one* **5**, e14325, doi:10.1371/journal.pone.0014325 (2010).

229 Aguilera, O. *et al.* The Wnt antagonist DICKKOPF-1 gene is induced by 1alpha,25-dihydroxyvitamin D3 associated to the differentiation of human colon cancer cells. *Carcinogenesis* **28**, 1877-1884, doi:10.1093/carcin/bgm094 (2007).

230 Rawson, J. B. *et al.* Vitamin D intake is negatively associated with promoter methylation of the Wnt antagonist gene DKK1 in a large group of colorectal cancer patients. *Nutr Cancer* **64**, 919-928, doi:10.1080/01635581.2012.711418 (2012).

231 Janssens, S., Denayer, T., Deroo, T., Van Roy, F. & Vleminckx, K. Direct control of Hoxd1 and Irx3 expression by Wnt/beta-catenin signaling during anteroposterior patterning of the neural axis in Xenopus. *Int J Dev Biol* **54**, 1435-1442, doi:10.1387/ijdb.092985sj (2010).

232 Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441**, 349-353, doi:10.1038/nature04733 (2006).

233 Erdelyi, I. *et al.* Western-style diets induce oxidative stress and dysregulate immune responses in the colon in a mouse model of sporadic colon cancer. *J Nutr* **139**, 2072-2078, doi:10.3945/jn.108.104125 (2009).

234 Hitchler, M. J. & Domann, F. E. Metabolic defects provide a spark for the epigenetic switch in cancer. *Free Radic Biol Med* **47**, 115-127, doi:10.1016/j.freeradbiomed.2009.04.010 (2009).

235 Taniguchi, K. *et al.* Mild inflammation accelerates colon carcinogenesis in Mlh1-deficient mice. *Oncology* **71**, 124-130, doi:10.1159/000100522 (2006).

236 Baylin, S. B. & Ohm, J. E. Epigenetic gene silencing in cancer - a mechanism for early oncogenic pathway addiction? *Nat Rev Cancer* **6**, 107-116, doi:10.1038/nrc1799 (2006).

237 Gonzalez-Sancho, J. M. *et al.* The Wnt antagonist DICKKOPF-1 gene is a downstream target of beta-catenin/TCF and is downregulated in human colon cancer. *Oncogene* **24**, 1098-1103, doi:10.1038/sj.onc.1208303 (2005).
238 Suzuki, H. *et al.* Epigenetic inactivation of SFRP genes allows constitutive WNT signaling in colorectal cancer. *Nat Genet* **36**, 417-422, doi:10.1038/ng1330 (2004).
239 Kawakami, K. *et al.* DNA hypermethylation in the normal colonic mucosa of patients with colorectal cancer. *Br J Cancer* **94**, 593-598, doi:10.1038/sj.bjc.6602940 (2006).
240 Worthley, D. L. *et al.* DNA methylation within the normal colorectal mucosa is associated with pathway-specific predisposition to cancer. *Oncogene* **29**, 1653-1662, doi:10.1038/onc.2009.449 (2010).
241 Issa, J. P., Ahuja, N., Toyota, M., Bronner, M. P. & Brentnall, T. A. Accelerated age-related CpG island methylation in ulcerative colitis. *Cancer Res* **61**, 3573-3577 (2001).
242 Kawate, H. *et al.* A defect in a single allele of the Mlh1 gene causes dissociation of the killing and tumorigenic actions of an alkylating carcinogen in methyltransferase-deficient mice. *Carcinogenesis* **21**, 301-305 (2000).
243 Ollikainen, M. *et al.* Mechanisms of inactivation of MLH1 in hereditary nonpolyposis colorectal carcinoma: a novel approach. *Oncogene* **26**, 4541-4549, doi:10.1038/sj.onc.1210236 (2007).
244 Islam, S. *et al.* Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. *Nat Protoc* **7**, 813-828, doi:10.1038/nprot.2012.022 (2012).
245 Krjutskov, K. *et al.* Single-cell transcriptome analysis of endometrial tissue. *Hum Reprod* **31**, 844-853, doi:10.1093/humrep/dew008 (2016).
246 Park, S. K. *et al.* Field Cancerization in Sporadic Colon Cancer. *Gut Liver* **10**, 773-780, doi:10.5009/gnl15334 (2016).
247 Hawthorn, L., Lan, L. & Mojica, W. Evidence for field effect cancerization in colorectal cancer. *Genomics* **103**, 211-221, doi:10.1016/j.ygeno.2013.11.003 (2014).
248 Braakhuis, B. J., Tabor, M. P., Kummer, J. A., Leemans, C. R. & Brakenhoff, R. H. A genetic explanation of Slaughter's concept of field cancerization: evidence and clinical implications. *Cancer Res* **63**, 1727-1730 (2003).
249 Orr, B., Godek, K. M. & Compton, D. Aneuploidy. *Curr Biol* **25**, R538-542, doi:10.1016/j.cub.2015.05.010 (2015).
250 Wang, L. *et al.* Whole-exome sequencing of human pancreatic cancers and characterization of genomic instability caused by MLH1 haploinsufficiency and complete deficiency. *Genome Res* **22**, 208-219, doi:10.1101/gr.123109.111 (2012).
251 He, W. *et al.* Rad9 plays an important role in DNA mismatch repair through physical interaction with MLH1. *Nucleic Acids Res* **36**, 6406-6417, doi:10.1093/nar/gkn686 (2008).