Discovery of genetic defects in unexplained colorectal cancer syndromes

Print: ProefschriftMaken || www.proefschriftmaken.nl
ISBN: 978-94-6295-913-2

The studies described in this thesis were supported by the Dutch Cancer Society (grant number: UL2012-5542).

Financial support for the printing of this thesis was provided by: Chipsoft BV and the department of Pathology, Leiden University Medical Center.

Discovery of genetic defects in unexplained colorectal cancer syndromes

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus Prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 29 mei 2018
klokke 16:15 uur

door

Anne Maria Lucia Jansen

geboren te Zoetermeer
in 1990

Promotores

Prof. dr. P. Devilee
Prof. dr. H. Morreau

Co-promotor

dr. F.J. Hes

Leden promotiecommissie

Prof. dr. C.J. van Asperen
Prof. dr. R.M.W. Hofstra (Erasmus Medical Center)
Prof. dr. R.H. Sijmons (University Medical Center Groningen)

Table of contents

Chapter 1	General introduction and thesis outline	7
Chapter 2	Whole gene capture analysis of 15 CRC susceptibility genes in suspected Lynch Syndrome patients	29
Chapter 3	Combined mismatch repair and *POLE/POLD1* defects explain unresolved suspected Lynch Syndrome cancers	53
Chapter 4	RNA analysis of cancer predisposing genes in formalin-fixed paraffin-embedded tissue determines aberrant splicing	73
Chapter 5	The complexity of screening *PMS2* in DNA isolated from formalin-fixed paraffin-embedded tissue	87
Chapter 6	Distinct patterns of somatic mosaicism in the *APC* gene in neoplasms from patients with unexplained adenomatous polyposis	97
Chapter 7	Concluding remarks and future perspectives	113
Chapter 8	Summary Nederlandse samenvatting Curriculum vitae List of publications Dankwoord	129

Chapter 1

General introduction and thesis outline

General introduction

Colorectal cancer (CRC) is the third most common cancer in men and second most common in women, with an incidence of 1.4 million cases worldwide in 2012.[1] An estimated 25-35% of all CRCs have heritable components, either pathogenic variants in high-risk CRC susceptibility genes (3-5%) or a positive family history (20-30%).[2-4] The underlying genetic cause in the patients without penetrant mutations but with family history is not well understood, but is expected to be a combination of environmental and inherited genetic factors with common, low-penetrant genetic alterations.[3, 4] With large genome-wide association studies (GWAS) more of these CRC susceptibility loci are being identified. These loci, often single-nucleotide polymorphisms (SNPs), slightly increase colorectal cancer risk. With information about the combined risk of multiple of these SNPs a personal CRC-risk profile can be created. An effort to calculate personal cancer risk scores (with a polygenic risk model) by combining risk scores of multiple (moderate) cancer susceptibility loci is already being done for breast cancer, and could be a possibility for colorectal cancer, if more CRC susceptibility loci are mapped.[5-7]

Colorectal carcinomas usually start as benign polyps that grow from normal colonic mucosa. The progression of normal colonic epithelial cells to adenocarcinomas usually follows the classical progression of precursor lesions with somatic, genetic and epigenetic changes. These changes often confer a growth advantage leading to clonal expansion of the altered cells. This process is better known as the adenoma-carcinoma sequence and typically spans over 15 years.[2, 8] One fundamental aspect of the tumorigenesis process is the acquisition of genomic instability, which can be present in one of these forms: microsatellite instability (MSI), chromosomal instability (CIN) or CpG island methylator phenotype (CIMP).[2, 9] MSI is caused by defects in the mismatch repair system, a characteristic of Lynch Syndrome tumors. CIN is described to be caused by a combination of oncogene activation (e.g. *KRAS* and *PICK3CA)* and tumor suppressor gene inactivation (e.g. *APC, TP53* and *SMAD4*).[10, 11] Over 80% of adenomas and CRCs are found to have inactivating *APC* variants.[10-13] These pathogenic variants result in Wnt signaling activation, a key early event in CRC tumorigenesis. CIMP is a subset of CRCs that result from epigenetic changes and that are characterized by the inactivation of multiple tumor suppressor genes and other tumor-related genes.[14, 15]

The three different types of genomic instability are a result of heritable factors, environmental factors and random mistakes during normal DNA replication.[16] Known heritable factors often increase the number of variants per replication of the cell by disabling correct proofreading, or by affecting one of the multiple DNA repair pathways present in the cell. Besides genetic variants directly affecting protein function, CRC susceptibility can occur through other forms of transcriptional silencing. The best known transcriptional silencer is epigenetic promoter methylation, described in *MLH1, MGMT, APC* and *P16/CDKN2A,* but transcriptional silencing can also occur through microRNAs (miRNAs).[17, 18] MiRNAs are small nucleotide sequences that participate in the regulation of cell differentiation, cell cycle progression, and apoptosis.[19, 20] Dysregulation of miRNAs has been shown to play a role in CRC tumorigenesis.[19, 21, 22] In an effort to create one consensus method to classify CRC subtypes, a large international consortium classified CRCs based on gene expression profiles.[23] Four consensus molecular subtypes (CMS) were defined: CMS1: MSI, hypermutated and strong immune activation, CMS2: epithelial with Wnt and Myc signaling activation, CMS3: epithelial with metabolic dysregulation and CMS4: TGF-b activation, stromal invasion and angiogenesis.[23] This classification was based on gene expression profiles and not on underlying genetic causes.

Syndrome	Abbr.	Genes	Pattern of inheritance	Prevalence	Phenotype	Screening
CRC syndromes						
Lynch Syndrome	LS	*MLH1, MSH2, MSH6, PMS2, EPCAM*	Dominant	2-5%	MMR-/MSI early onset CRC. Freq: EC, SBC, StC, OC, BlC	Colonoscopy every 1-2 years (start age 20-25).
Familial colorectal cancer type X	FCCTX	*SEMA4A, BMPR1A*	Unknown	Unk	MMR+ Amsterdam positive families.	Colonoscopy every 3-5 years (start 5-10 years before earliest age of onset in family).
Polyposis syndromes						
Familial adenomatous polyposis	FAP	*APC*	Dominant	~1%	100-1000 colorectal adenomas at early age. If untreated, CRC.	Colonoscopy every 1-2 years (start age 10-14).
MUTYH associated polyposis	MAP	*MUTYH*	Recessive	0.3-1%	10-100 colorectal adenomas, with possible CRC.	Colonoscopy every 2 years (start age 18-20)
Polymerase proofreading associated polyposis	PPAP	*POLE/POLD1*	Dominant	Unk	Early onset CRC with ultramutated phenotype, often also polyps	Colonoscopy every 1-2 years (start age 20-25). EC screening from age 40 for female *POLD1* carriers.
NTHL1-associated polyposis	NAP	*NTHL1*	Recessive	Unk	CRC with polyposis and frequent BrC, EC, SkC, DuoC and PrC.	Unknown
MSH3-associated polyposis	-	*MSH3*	Recessive	Unk	Unknown	Unknown
Juvenile polyposis	JPS	*SMAD4, BMPR1A*	Dominant	<1%	Spherical, lobulated polyps at young age (mean age 18.5 years)	Colonoscopy every 3 years, starting at age 15.
Peutz-Jeghers Syndrome	PJS	*STK11*	Dominant	<1%	Hamartomatous polyps and dark brown pigm. on lips, hands and feet. Freq: StC, BrC, OC, LuC	LB surveillance every 3 years (start age 18), UGI endoscopy every 3 years from age 25 and mammography from age 25.
PTEN hamartoma tumor syndrome	PTHS	*PTEN*	Dominant	<1%	Hamartomatous polyps in GI tract, macrocephaly, papillomas and fibromas	Colonoscopy every 5 years from age 35, yearly: thyroid (from age 18) and breast examination (from age 25-30).

Abbr = abbreviation, MMR = mismatch repair, CRC = colorectal cancer, EC = endometrium cancer, SBC = Small Bowel Cancer, StC = Stomach cancer, OC = ovarium cancer, BlC = Bladder cancer, SkC = skin cancer, DuoC = duodenum cancer, BrC = breast cancer, PrC = prostate cancer, LuC = lung cancer, LB = large bowel, UGI = upper gastrointestinal, Unk = unknown, MSI = microsatellite instability

Familial syndromes

A number of different hereditary colorectal cancer and polyposis syndromes have been defined on the basis of distinct clinical, pathological and molecular characteristics. Each of these has been linked to, or even named after, (a) specific gene(s) (Table 1), but the genotype-phenotype connection has become considerably more complicated in recent years. The most common familial colorectal cancer syndromes are discussed separately below.

Lynch Syndrome

Lynch Syndrome is the most common form of hereditary CRC, accounting for 2-5% of all CRCs in the general population, and is caused by heterozygous pathogenic germline variants in one of the mismatch repair (MMR) genes, *MLH1, MSH2, MSH6* and *PMS2*.[24-27] In addition, in approximately 20-25% of patients with immunohistochemical MSH2 loss but without pathogenic *MSH2* variants a germline deletion of the 3' end of the *EPCAM* gene is found, resulting in allele-specific hypermethylation and transcriptional inactivation of *MSH2* that is directly upstream from *EPCAM*.[28, 29] The function of the MMR pathway is to check DNA replication fidelity and repair DNA mismatches that occur due to replication errors.[30-32] This is necessary for maintaining genome stability. Although functioning in the same cellular pathway, the MMR proteins form distinct heterodimers. The MutS complex (MSH2/MSH6 or MSH2/MSH3) is responsible for recognition of mismatches and small insertions/deletions (indel). The MutL complex (MLH1/PMS2, MLH1/PMS1 or MLH1/MLH3) is responsible for forming a MutS/MutL/DNA complex, for endonuclease activity and for termination of mismatch-provoked excision.[30-32] Because MLH1 and MSH2 can heterodimerize with multiple proteins, these proteins are shown to be essential MMR components, while MSH6, PMS2, PMS1, MSH3 and MLH3 are important but partially redundant.[31, 32] When MMR ability is lost, cells develop a 'mutator' phenotype characterized by a 100-1000 times increase in mutation rate.[30-33] Microsatellites, repetitions of small DNA sequences, are more mutation-prone and become unstable if the MMR system is defective.[27, 30-33] This mutational signature known as microsatellite instability (MSI) is characteristic of MMR-deficient tumors.[27, 30-34] Other characteristics of MMR-deficient tumors are a high density of tumor infiltrating lymphocytes (TILs) and a proximal location in the colon.[33-38] TILs are related to a specific antigen-driven immune response, have been described to be activated and to have a cytotoxic nature and are associated with improved prognosis.[36, 38]

Approximately 15% of colorectal cancers display the MSI phenotype of whom the majority (>85%) are sporadic and result from somatic *MLH1* promoter hypermethylation.[33, 39-41] These sporadic *MLH1* methylated tumors commonly occur at relatively advanced age, and typically do not show a family history of CRC.[42-44] *MLH1* methylated cancers often carry the somatically acquired *BRAF* V600E variant. While *BRAF* testing has low specificity, it is still used in some centers as a pre-screening method to select cancers with methylated *MLH1* promoters.[45-47] Pathogenic germline MMR variants and somatic *MLH1* hypermethylation are usually described to be mutually exclusive, although rare cases of LS have been described with somatic promoter hypermethylation as a genetic 'second hit'.[47, 48] Furthermore, though rare, germline *MLH1* promoter hypermethylation has been described in young CRC patients with *MLH1* promoter hypermethylated tumors.[49-57] Inheritance of a constitutional epimutation has been previously described in at least three unrelated families.[58-61]

LS is inherited in an autosomal dominant fashion, with an average cumulative risk of developing CRC at the age of 70 years ranging from 34-67% for *MLH1* and *MSH2* mutation carriers[62, 63], 22-69% for *MSH6* mutation carriers[64, 65] and 11-20% for *PMS2* mutation carriers.[66, 67] Additionally, female mutation carriers have an average cumulative risk of developing endometrial cancer of 31-60% for *MLH1* and *MSH2* mutation carriers[68-71], 44-70% for *MSH6* mutation carriers[64, 65, 71] and 12-15% for *PMS2* carriers.[66, 67] Other cancer types, including small bowel, stomach, pancreas, ovary and bladder cancer occur, but less frequently.[63-69, 72] Recent studies also indicate an increased risk for prostate and breast cancer.[73-75] For patients carrying a pathogenic MMR variant, colonoscopy is recommend every 1 to 2 years starting at ages 20 to 25 years to reduce the incidence and mortality of this tumor.[76, 77] Previous studies have shown that MMR-deficient tumors have a better clinical prognosis and possibly a better response to novel regiments such as immunotherapy.[33, 34, 78, 79]

In the past, a set of criteria (Amsterdam Criteria) were used to clinically diagnose Lynch Syndrome families.[80] These criteria considered age of onset, number of CRC patients within a family and the relation between the affected family members (one should at least be a first-degree relative of the other).[80] However, even after inclusion of other LS-associated non-colorectal cancers (Amsterdam II Criteria), these diagnostic criteria were found to lack sensitivity and specificity in diagnosing Lynch Syndrome.[81, 82] In 1996 the Bethesda guidelines were introduced to identify individuals who should receive genetic testing for Lynch Syndrome.[83] These guidelines advised to screen patients fulfilling one of the following criteria: (1) individuals in families that meet Amsterdam Criteria, (2) patients with two Lynch-associated cancers, (3) patients with CRC and a first-degree relative (FDR) with an LS-associated cancer before age 45, (4) patients with right-sided undifferentiated or signet-ring cell type CRC before age 45 or (5) patients with adenomas before age 40.[83, 84] The Bethesda guidelines showed a high sensitivity (96%), an improvement on the previously used Amsterdam Criteria, but still a low specificity (27%).[85]

The introduction of the MSI analysis and immunohistochemical staining led to the revised Bethesda guidelines and a more sensitive method to determine MMR-deficiency.[86-89] Immunohistochemical staining of tissue slides from formalin-fixed paraffin-embedded tumors is used to determine the presence and location of a protein. The first step, tissue preparation, includes deparaffinization of the tissue slides, blocking of endogenous enzymatic activity and (heat-induced) antigen retrieval. After antigen retrieval specific primary and then secondary antibodies are added to the slides. Antibodies are detected with a chromogenic reaction, in which an enzyme label conjugated to the antibody reacts with a substrate to yield a colored precipitate. Horseradish peroxidase (HRP) is often used for this reaction, and the precipitating substrate DAB shows a typical brown-colored precipitate at the protein localization site. Finally, the slide is counterstained, commonly with hematoxylin, a compound creating a blue color. Brown-colored precipitate will indicate the presence of the target protein at a specific location, while only the blue counterstain will be seen when the protein is absent (see Figure 1). Importantly, non-tumor cells present in the sample should always show staining, since the target protein is still present in these cells. This allows these cells to be used as a positive control, i.e. to confirm that lack of staining is due to loss of protein and not due to technical artifacts.

For MSI analysis the National Cancer Institute microsatellite panel, consisting of at least 5 microsatellite markers, is recommended, although commercially available mononucleotide panels are routinely used.[87] Tumors are characterized to be MSI-low (MSI-L) if one marker shows instability, MSI-high (MSI-H) if two or more markers show instability and MSI-stable (MSS) when no markers show instability.[87, 90]

The revised Bethesda guidelines advise genetic testing of all CRCs with an age of onset younger than 50, as well as all MSI-H CRCs before 60 years. Furthermore, patients with one FDR with an LS-associated CRC before 50, or two or more first or second degree relatives regardless of age should also be tested for genetic variants. Combining revised Bethesda criteria with MSI analysis and IHC was found to result in a sensitivity of 82% and a specificity of 98%.[91] However, more recent studies advocate routine molecular screening of patients under 55 or even under 70, regardless of Bethesda criteria.[92, 93] This molecular screening is then often combined with *BRAF* V600E or *MLH1* promoter hypermethylation testing to exclude sporadic MLH1 loss.[45, 47]

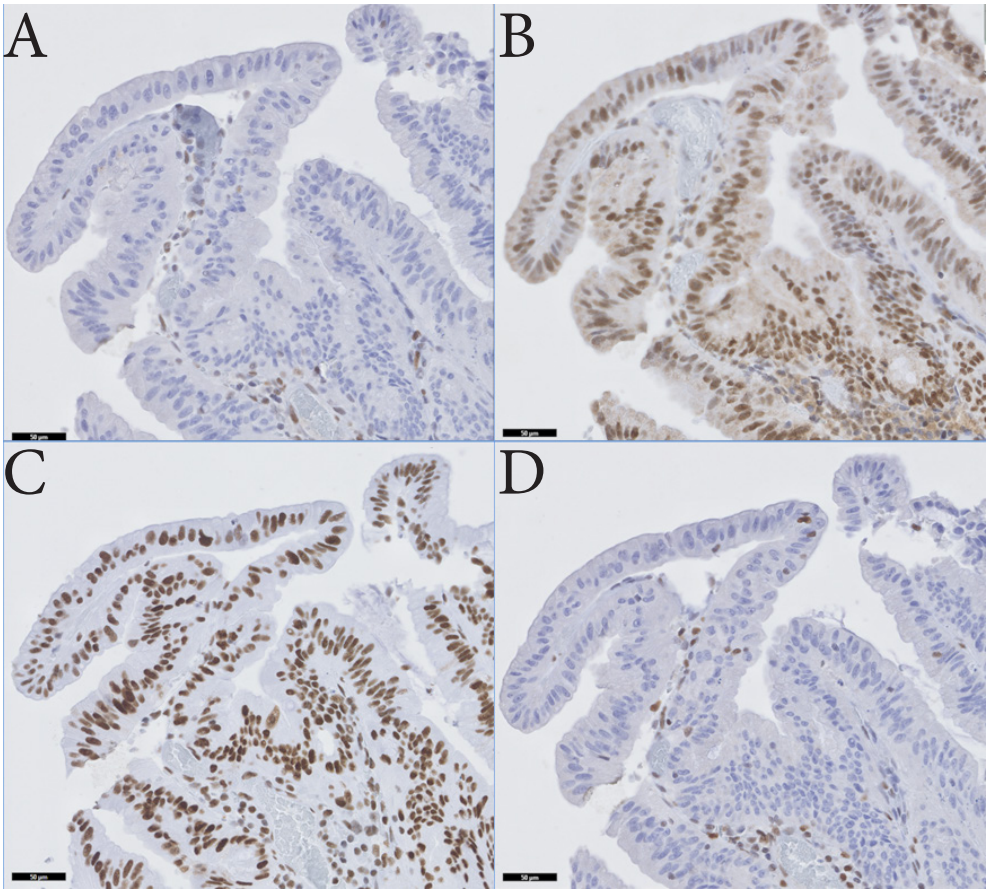


Figure 1: Colon high grade villous adenoma, IHC of the four MMR proteins
Immunohistochemical staining of [A] MLH1, [B] MSH2, [C] MSH6, [D] PMS2. Staining shows loss of MLH1 and PMS2 expression, and positive MSH2 and MSH6 expression.

Unexplained suspected Lynch Syndrome

A 2014 study estimated that up to 60% of MMR-deficient colorectal cancers do not carry germline MMR variants, nor can be explained by somatic *MLH1* promoter hypermethylation.[94] These patients are referred to as 'suspected Lynch Syndrome' (sLS) [94] or 'Lynch-like Syndrome' (LLS)[95, 96], and failure to determine the underlying (genetic) cause of disease has a major impact on the clinical management of these patients. Three potential reasons for MMR-deficient and/or MSI-H cancers of sLS patients discussed in literature are (1) missed variants in the MMR genes, (2) biallelic somatic inactivation of the MMR genes or (3) variants in other genes that can drive MSI.[96-99] These possible explanations are comprehensively discussed in Chapter 7 (concluding remarks). The cancer risk in families with sLS is found to be lower than that of families with LS but higher than that of families with sporadic CRCs and more research is needed into the potential (genetic) causes of these CRCs.[95-100]

Currently, many high-throughput screening efforts are being done to find the genetic cause in these sLS families, resulting in many (MMR) variants of uncertain clinical significance (VUS). These are the variants for with evidence is lacking to classify them as either (likely) benign or (likely) pathogenic. Characterization of MMR variants is done according to the standardized five-tiered scheme of the International Society for Gastrointestinal Hereditary Tumors (InSiGHT).[101] According to this scheme, variants can be classified to be not pathogenic (class 1), likely not pathogenic (class 2), uncertain (class 3), likely pathogenic (class 4) or pathogenic (class 5). While the scheme provides clear clinical guidelines for classes 1, 2, 4, and 5, many variants are assigned to class 3 for lack of good classification evidence, and clinical impact of these variants remains uncertain. In addition to clinical data (such as family history, co-segregation, immunohistochemistry, etc.), functional tests, such as minigene splicing assays or *in vitro* MMR assays, may help to interpret the clinical impact of these variants.[102-104]

Constitutional MMR-deficiency

Patients with homozygous or compound heterozygous variants in the MMR genes show a different phenotype than classical LS patients, known as constitutional MMR-deficiency syndrome (CMMRD).[105, 106] CMMRD patients develop a diverse spectrum of childhood cancers, including CRC but also hematological and brain malignancies.[105, 106] Another characteristic of CMMRD is café-au-lait maculae (CALM).[105, 106] Most CMMRD families have homozygous/compound heterozygous *PMS2* variants, but families with biallelic *MLH1* or *MSH6* variants have also been described.[105-109] The mean age at diagnosis in patients with CMMRD is 7-9 years for brain tumors, 16 years for CRC and 5-12 years for hematological cancers.[105, 106] Recommended surveillance consists of MRI scanning of the brain starting at the age of 2 years at an interval of 6-12 months and colonoscopies every 6 months from the age of 8 years.[105, 106] Siblings of CMMRD patients have a 25% risk of having the same genotype, and 50% chance of carrying a heterozygous MMR variant with an increased risk for LS-associated tumors in adulthood.[105] CMMRD is a severe disorder with a large spectrum of cancers. Depending on the type of pathogenic variant patients can have a severe phenotype with brain tumors in early childhood, or a milder phenotype with later age of onset.[105, 106] Surveillance will aid in early detection of tumors and guide proper treatment, but most patients will die from cancers in early childhood.[105, 106]

Muir-Torre

Muire-Torre syndrome (MTS) is an autosomal dominant skin condition characterized by sebaceous gland tumors or keratoacanthoma, with colorectal, endometrial, urological or upper gastrointestinal neoplasms.[110-113] Clinical evidence suggests there might be two types of MTS, one with MMR-proficient/MSS and one with MMR-deficient/MSI-H tumors.[110] The latter is possibly a clinical variant of Lynch Syndrome, and is often caused by pathogenic variants in the *MSH2* gene.[110, 112-114] Identical *MSH2* variants have been found in LS- and MTS-patients, and a possible explanation for the different phenotype could be that the *MSH2* variant in MTS-patients co-segregates with variants in other modulator genes involved in skin carcinogenesis or that inactivation of *MSH2* could result in molecular changes in other genes.[110, 113, 114] For the MMR-proficient type of MTS the genetic cause is still unknown, but biallelic *MUTYH* germline variants have been implicated as one of the possible genetic causes.[115, 116]

Familial colorectal cancer type X

Approximately half of the families positive for the Amsterdam I criteria carry MMR-proficient and MSS colorectal cancers.[117-119] These families are classified as familial colorectal cancer type X (FCCTX) and the molecular mechanism underlying these tumors is not well understood.[117-120] Specific clinical features of FCCTX (compared to LS) include absence of endometrial cancers and high prevalence of rectal cancers, lower incidence of CRC and diagnosis at a higher mean age (57.3 years, compared to 49.7 for LS).[119-124] FCCTX cancers have not (yet) been linked to one specific gene.[123] Recent studies indicate *SEMA4A* and *BMPR1A* as possible underlying genetic causes, explaining a small percentage of FCCTX cases.[125, 126] Other candidate genes are *CENPE, KIF24, GALNT12, ZNF367, GABBR2* and *BMP4,* but evidence for the involvement of these genes in FCCTX-tumorigenesis is lacking.[120, 124] One study suggests that FCCTX is not a single entity, but rather a name for a combination of multiple entities.[123] Surveillance recommendations currently include colonoscopies every 3-5 years, starting 5-10 years before the earliest age of onset in the family.[118, 124]

Familial adenomatous polyposis

Familial adenomatous polyposis (FAP) is a rare autosomal dominant disease accounting for ~1% of colorectal cancers.[24, 127] FAP is caused by germline heterozygous variants in the *APC* gene and is associated with the development of hundreds to thousands of colorectal adenomas at an early age.[24, 128] On average, cancers develop a decade after the first appearance of adenomas and the average age of CRC diagnosis if left untreated is 39 years.[4, 128, 129] *De novo* pathogenic *APC* variants are responsible for approximately 25% of FAP patients.[4, 130-132] *APC* is a tumor suppressor gene involved in the Wnt signaling pathway, which functions by negatively regulating the β-catenin oncoprotein.[128, 130] The 310 kDa protein has four β-catenin binding domains, and seven domains involved in binding and down-regulating β-catenin.[4, 130-133] In absence of APC β-catenin accumulates in the nucleus and interacts with factors that upregulate transcription of genes involved in cell cycle, proliferation, differentiation, migration, apoptosis and progression.[130] Notably, when *APC* is inactivated in the tumor, β-catenin overexpression is still kept in check by unknown mechanisms. Apparently, some residual downregulation is of great importance for tumor formation.[133] In addition, APC stabilizes microtubules and loss of APC leads to chromosomal instability, defective chromosome segregation and aberrant mitosis.[130]

Somatic inactivation of *APC* is a common molecular event in sporadic colorectal cancer and is present in about 80% of sporadic colorectal cancers.[128] The majority of pathogenic *APC* variants are truncating variants, either nonsense (26%), small insertions (10%) or small deletions (46%).[130]

FAP can be present in the classical, more severe form, or as attenuated FAP (AFAP) a milder form with fewer adenomas (<100) and a later age of onset. The severity of FAP is associated with the location of the *APC* variant within the *APC* gene. Variants resulting in an AFAP phenotype are located before codon 157, after codon 1595 and in the alternative spliced region of exon 9.[134, 135] Variants located in the DNA binding domain of the *APC* gene are described to lead to a severe type of FAP (>thousands of adenomas). The location of the variant however, is not the only predictive factor of the severity of polyposis. *APC* variants in a mosaic fashion are described to lead to an (attenuated) form of polyposis, irrespective of the variant's location.[136-138] The severity of polyposis in patients with mosaic *APC* variants depends on timing and origin of the mutation.[136, 139] Patients with FAP should be examined by colonoscopy every 1-2 years, beginning at age 10-14.[4, 130] Once adenomas are detected, annual follow-up is recommended.[4] Management of FAP includes endoscopic polypectomy and surgery.[130] Colectomy should be considered when more than 20 adenomas develop, when adenomas >1 cm are found, or when advanced histology (ulcerated, high grade dysplasias) appears.[4, 130]

MUTYH-associated polyposis

Besides FAP, another possible diagnosis for patients presenting with 10-100 adenomas is MUTYH-associated polyposis (MAP).[140, 141] MAP is an autosomal recessive disorder caused by biallelic variants in the base excision repair gene *MUTYH,* accounting for approximately 0.3 - 1% of all CRCs.[128, 134, 140-145] The base excision repair (BER) pathway has an important role in preventing variants associated with oxidative damage.[140, 141, 146] Reactive oxygen species (8-oxo-7,8-dihydro-2'-deoxyguanosine or 8-oxodG) can be incorporated in the DNA through direct oxidation of guanine or via incorporation of 8-oxodGTP from the nucleotide pool.[141] DNA polymerase incorporates adenosine opposite 8-oxodG, leading to G:C>T:A transversions.[146] The function of *MUTYH* within this pathway is to scan the daughter strand after replication and to remove adenosine residues mispaired with guanose or 8-oxoG.[141]

Pathogenic germline variants in the *MUTYH* gene were first detected in 2002 in one family in which 11 tumors from 3 affected siblings were screened for somatic *APC* variants.[140] In these 11 adenomas and carcinomas 18 inactivating somatic variants were found, of which the majority (n=15) were G:C>T:A variants. This was taken as a strong indication of a BER defect, previously described in yeast.[140, 147, 148] Sequencing of leukocyte DNA for variants in the BER-genes *MUTYH, OGG1* and *MTH*, led to the discovery of pathogenic compound heterozygous variants in *MUTYH.* Another hallmark of these tumors is the *KRAS* c.34G>T variant, found in 64% of MAP colorectal cancers.[142, 149, 150]

MAP is characterized by a greatly increased risk of lifetime colorectal cancer (43-100%), often in combination with colonic adenomas.[142] Monoallelic variants in *MUTYH* are present in 1-2% of the general population,[143, 151] and the cancer risk of these heterozygous carriers is still under debate.[152-157] Surveillance for MAP consists of colonoscopy every two years starting at age 18-20 years.[142]

Polymerase proofreading associated polyposis

Recently, variants in the exonuclease domains of *POLE* and *POLD1* genes have been described to be associated with colorectal carcinomas (CRC), endometrial cancer (EC) and colorectal polyposis.[158-165] *POLE* and *POLD1* are the genes that encode for the catalytic subunit of polymerases ε and δ, involved in DNA replication of the lagging and leading strand respectively. The exonuclease domain provides proofreading capabilities essential for maintenance of replication fidelity.[158-165] The mean age of onset of CRC is 40.7 years for *POLE* mutation carriers, and 35.9 years for *POLD1* carriers.[165] In a small percentage of *POLE/POLD1* mutation carriers brain tumors are diagnosed.[163-165] This variable phenotype has been coined polymerase proofreading associated polyposis (PPAP), a syndrome with high penetrance and dominant inheritance.[162-165] Interestingly, in contrast to the classical tumor development model, only a minority of tumors are found to have loss of the wildtype allele, or sustain other variants that could act as a 'second hit'.[158-160, 162] Somatic and germline variants in *POLE/POLD1* are believed to account for 3% of all CRCs and 7% of ECs.[160-164, 166] PPAP tumors are often MSS but MSI-H PPAP tumors have been described, where the MMR-deficiency supposedly resulted from somatic secondary MMR variants.[160, 167, 168] Guidelines recommend colonoscopies every 1-2 years starting at age 20-25, combined with endometrial cancer screening at age 40 for *POLD1* female carriers.[165]

Pathogenic *POLE* and *POLD1* variants have been described as inherited (PPAP) and somatically acquired, both leading to an 'ultramutated' phenotype with a variant incidence exceeding 100 variants/Mb.[158-164] While the majority of somatic variants are C:G>T:A variants, a particular increase in G:C>T:A transversions are characteristic of *POLE/POLD1* mutated tumors, with an elevated TCT>TAT and TCG>TTG mutational pattern.[158-163] Germline or somatic *POLE/POLD1* mutated tumors are significantly more immunogenic with increased lymphocyte infiltration and cytotoxic T-cell marker expression, and have a favorable prognosis.[169, 170]

NTHL1-associated polyposis

Due to many high-throughput sequencing efforts, new genes predisposing to familial colorectal syndromes continue to be found. In 2015, whole exome sequencing of 48 families with colorectal cancer and polyps led to the identification of NTHL1-associated polyposis (NAP).[171] NAP is a recessive disorder caused by biallelic inactivation of the *NTHL1* gene.[171-174] This gene is part of the base excision repair pathway and encodes for the NTHL1 glycosylase which is involved in removing oxidative pyrimidine lesions.[175] So far, only a few families with NAP have been described, and the prevalence and the exact phenotype remain unknown. Families with *NTHL1* variants appear to have a phenotype predominantly consisting of colorectal cancer with adenomatous polyposis, although breast, endometrial, duodenal, skin, prostate and pancreatic cancers have also been described in NAP patients.[128, 171] Furthermore, the mutational profile of these cancers resembles an MAP phenotype with G:C>T:A transversions.[171] Currently, only nonsense and splice site variants have been described, often the *NTHL1* p.Gln90* hotspot variant.[171-174]

MSH3-associated polyposis

In 2016 another new genetic underlying cause of unexplained polyposis was detected through whole exome sequencing (WES).[176] After WES on leukocyte DNA from 102 unrelated individuals with unexplained polyposis, two individuals with compound heterozygous *MSH3* loss of function variants were found.[176] Tumors of both patients showed high microsatellite instability in di- and tetranucleotides (EMAST) and immunohistochemical loss of MSH3.[176] Loss of MSH3 protein expression was already shown to be frequent in MSI-H tumors due to a microsatellite in *MSH3*, but no *MSH3* germline mutation carrier had been described until 2016.

Hamartomatous polyposis syndromes

Hamartomatous polyposis syndromes are a rare heterogeneous group of autosomal dominant disorders accounting for less than 1% of all hereditary colorectal cancer syndromes.[177, 178] Hamartomatous polyps are the main characteristic of these syndromes. While these polyps are benign they have the potential to become malignant and progress into carcinomas.[177, 178] This progression is through a hamartoma to carcinoma sequence in which stromal elements create a local environment that promotes epithelial dysplasia.[178] The different syndromes within this group all have different clinical phenotypes, each with different frequencies and location of the polyps, distinct organ-specific manifestations, and predispositions for the development of other malignancies. Proper distinction between these syndromes is of great importance for appropriate clinical management.

Juvenile polyposis syndrome

Juvenile polyposis syndrome (JPS) is characterized by the development of multiple gastrointestinal polyps in the colon.[177-182] These polyps generally vary in size from 5 mm to 50 mm and typically have a spherical, lobulated and pedunculated appearance.[179] JPS presents in the first or second decade of life, with an average age of diagnosis around 18.5 years.[178] Symptoms for JPS can include rectal bleeding, anemia, abdominal pain, constipation or change in stool size, shape or color, though some JPS patients remain asymptomatic.[177, 178] The cumulative lifetime risk for colorectal cancer is 40-70%.[177-179] If a pathogenic variant is present, surveillance with colonoscopy or endoscopy should start at the age of 15 years and should be repeated every 3 years.[177, 178] In about 20-60% of JPS patients a pathogenic germline variant in *SMAD4* or *BMPR1A* is found.[177-181] Both genes are involved in the TGF-beta signaling pathway. The majority of pathogenic germline *SMAD4* and *BMPR1A* variants are missense or small deletions and 15% is deletions of one or more exons.[179-181]

Peutz-Jeghers syndrome

Peutz-Jeghers syndrome (PJS) is characterized by the presence of hamartomatous polyps in the gastrointestinal tract and distinctive mucocutaneous pigmentation.[177, 183-186] A typical dark blue to dark brown pigmentation is present in 95% of PJS patients and is seen on the vermilion border of the lips, the buccal mucosa, hands and feet.[177, 185, 186] The small intestine is most commonly affected, although polyps can be found in colon, stomach, rectum, bladder, esophagus and gallbladder.[185-188] The polyps are generally between 5 to 50 mm in diameter.[177, 178, 186] PJS presents in the second or third decade of life.[177, 178, 186] Symptoms for PJS can include rectal bleeding, anemia, bowel obstruction and abdominal pain.[177, 185, 186] The risk of developing any cancer at age 65 is 47-93%, with an especially high risk of developing

stomach, small intestine, colon and breast cancer but also elevated risk of developing cancer in the esophagus, pancreas, lung, uterus and ovaries.[186, 189] Surveillance of the large bowel is recommended every three years, starting at age 18 and upper gastrointestinal endoscopies are recommended every three years starting at age 25.[190] Notably, since breast cancer risks are comparable to *BRCA1/BRCA2* mutation carriers (40-85% lifetime risk), PJS patients are recommended equal surveillance with monthly breast self-examination starting at age 18 and mammography starting at age 25.[186, 189]

Pathogenic germline variants in *STK11* gene (also known als *LKB1*) are found in 30-80% of PJS patients.[184-186, 188-192] STK11 is a serine-threonine kinase involved in regulation of cellular proliferation via G1 cell-cycle arrest, in WAS1 signaling and P53 mediated apoptosis.[185] A clinical diagnosis of PJS is made when a patient has at least two PJS polyps or one PJS polyp with a positive family history or mucocutaneous pigmentation.[185, 192] The mutation detection rate in patients who meet these criteria is 70-90% and the majority of variants detected are nonsense or frameshift deletions resulting in a truncated protein.[185, 186, 191]

PTEN hamartoma tumor syndrome
PTEN hamartoma tumor syndrome (PHTS) is an autosomal dominant disorder caused by germline variants in the tumor suppressor gene *PTEN*.[193] PHTS encompasses multiple overlapping syndromes as Cowden Syndrome (CS) and Bannayan-Riley-Ruvalcaba syndrome (BRRS).[177, 193] *PTEN* is a tumor suppressor gene with multiple, but incompletely understood, functions. As a lipid phosphatase it plays a role in the P13K/Akt signaling, involved in G1 cell-cycle arrest and apoptosis. As a protein phosphatase it can regulate cell-survival pathways.[194-196] It furthermore might play a role in cellular migration and focal adhesion.[194, 196]

CS is characterized by macrocephaly, mucocutaneous lesions, acral keratosis, papillomas and fibromas, and 30-85% of patients develop hamartomatous polyps in the gastrointestinal tract.[178, 194, 197, 198] CS patients have an increased lifetime risk of developing breast (25-50%), endometrial (5-28%), thyroid (3-17%), colon (9-16%), skin (6%) and renal cancers.[198, 199] With a range of possible tumors, management recommendations are broad, including yearly breast examinations from age 25-30, yearly thyroid examination or ultrasound starting at age 18 and colonoscopies every 5 years from age 35.[197, 199] BRRS patients develop lipomas, gastrointestinal hamartomatous polyps, macrocephaly, hemangiomas and developmental delay.[177] In 11 – 80% of patients meeting clinical criteria for PHTS, a pathogenic variant in *PTEN* is found.[178, 193, 194, 198, 199] In patients without a *PTEN* variant, pathogenic germline variants in *SDHB, SDHC, SDHD, AKT, PIK3CA* as well as hypermethylation of *KLLN* are found explaining the PHTS-like phenotype.[199, 200]

Thesis outline

Over the last few years, advances have been made in discovering the underlying genetic cause in unexplained CRC and polyposis patients. Three new CRC and polyposis syndromes—polymerase proofreading associated polyposis (PPAP), NTHL1-associated polyposis (NAP) and MSH3-associated polyposis—were discovered in the last three years, and new genes are still being described. The aim of this thesis was to find the underlying genetic cause in a large cohort of unexplained suspected Lynch Syndrome patients and a cohort of unexplained polyposis patients. We hypothesized that the suspected Lynch Syndrome (sLS) patients could be explained by missed variants in the mismatch repair (MMR) genes, by bi-allelic somatic inactivation of the MMR genes or by variants in other susceptibility genes.

Chapter 2 reports a whole gene capture effort in which we screened sLS patients for variants in the exonic- or intronic regions of 15 CRC susceptibility genes, including *MLH1, MSH2, MSH6* and *PMS2*. **Chapter 3** describes variants in the polymerase genes *POLE* and *POLD1* in sLS patients. Variants in the exonuclease domain of these genes result in hypermutated tumors. We hypothesize that the MMR-deficiency in these tumors is due to secondary MMR hits resulting from this hypermutated phenotype. **Chapter 4** shows a splicing assay to analyse the effect of variants (predicted) to result in splicing. For this assay RNA was isolated from formalin-fixed paraffin-embedded (FFPE) tissue, enabling analysis of somatic variants and variants in patients of which only FFPE is available. **Chapter 5** describes a practical guide on detecting and analysing variants in the *PMS2* gene in DNA isolated from FFPE. Analysis of this gene is complex due to the presence of many pseudogenes. **Chapter 6** focuses on unexplained polyposis patients. Half of unexplained patients with 20-100 adenomas could be explained by a mosaic *APC* variant, either present in leukocyte and colon, confined to the colon, or only detected in the adenomas but not in normal colonic mucosa of a patient. Finally, concluding remarks and future perspectives are presented in **Chapter 7**.

References

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015;136:E359-86.
2. Grady WM. Molecular Biology of Colon Cancer. In: Markman M, Saltz LB, eds. Colorectal Cancer: Evidence-Based Chemotherapy Strategies. Totowa, NJ: Humana Press, 2007:1-31.
3. Armelao F, de Pretis G. Familial colorectal cancer: a review. World J Gastroenterol 2014;20:9292-8.
4. Jasperson KW, Tuohy TM, Neklason DW, et al. Hereditary and familial colon cancer. Gastroenterology 2010;138:2044-58.
5. Michailidou K, Beesley J, Lindstrom S, et al. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nat Genet 2015;47:373-80.
6. Michailidou K, Hall P, Gonzalez-Neira A, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet 2013;45:353-61, 361e1-2.
7. Li H, Feng B, Miron A, et al. Breast cancer risk prediction using a polygenic risk score in the familial setting: a prospective study from the Breast Cancer Family Registry and kConFab. Genet Med 2017;19:30-35.
8. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell 1990;61:759-67.
9. Armaghany T, Wilson JD, Chu Q, et al. Genetic alterations in colorectal cancer. Gastrointest Cancer Res 2012;5:19-27.
10. Vogelstein B, Fearon ER, Hamilton SR, et al. Genetic alterations during colorectal-tumor development. N Engl J Med 1988;319:525-32.
11. Fodde R, Smits R, Clevers H. APC, signal transduction and genetic instability in colorectal cancer. Nat Rev Cancer 2001;1:55-67.
12. Powell SM, Petersen GM, Krush AJ, et al. Molecular diagnosis of familial adenomatous polyposis. N Engl J Med 1993;329:1982-7.
13. Miyoshi Y, Nagase H, Ando H, et al. Somatic mutations of the APC gene in colorectal tumors: mutation cluster region in the APC gene. Hum Mol Genet 1992;1:229-33.
14. Lao VV, Grady WM. Epigenetics and colorectal cancer. Nat Rev Gastroenterol Hepatol 2011;8:686-700.
15. Mojarad EN, Kuppen PJK, Aghdaei HA, et al. The CpG island methylator phenotype (CIMP) in colorectal cancer. Gastroenterology and Hepatology From Bed to Bench 2013;6:120-128.
16. Tomasetti C, Li L, Vogelstein B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. Science 2017;355:1330-1334.
17. de Vogel S, Weijenberg MP, Herman JG, et al. MGMT and MLH1 promoter methylation versus APC, KRAS and BRAF gene mutations in colorectal cancer: indications for distinct pathways and sequence of events. Ann Oncol 2009;20:1216-22.
18. Suehiro Y, Wong CW, Chirieac LR, et al. Epigenetic-genetic interactions in the APC/WNT, RAS/RAF, and P53 pathways in colorectal carcinoma. Clin Cancer Res 2008;14:2560-9.
19. Michael MZ, SM OC, van Holst Pellekaan NG, et al. Reduced accumulation of specific microRNAs in colorectal neoplasia. Mol Cancer Res 2003;1:882-91.
20. Yang L, Belaguli N, Berger DH. MicroRNA and colorectal cancer. World J Surg 2009;33:638-46.
21. Bandres E, Cubedo E, Agirre X, et al. Identification by Real-time PCR of 13 mature microRNAs differentially expressed in colorectal cancer and non-tumoral tissues. Mol Cancer 2006;5:29.
22. Schetter AJ, Okayama H, Harris CC. The role of microRNAs in colorectal cancer. Cancer J 2012;18:244-52.
23. Guinney J, Dienstmann R, Wang X, et al. The consensus molecular subtypes of colorectal cancer. Nat Med 2015;21:1350-6.
24. Gryfe R. Inherited colorectal cancer syndromes. Clin Colon Rectal Surg 2009;22:198-208.
25. de la Chapelle A. Genetic predisposition to colorectal cancer. Nat Rev Cancer 2004;4:769-80.
26. Underhill ML, Germansky KA, Yurgelun MB. Advances in Hereditary Colorectal and Pancreatic Cancers. Clin Ther 2016;38:1600-21.
27. Colas C, Coulet F, Svrcek M, et al. Lynch or not Lynch? Is that always a question? Adv Cancer Res 2012;113:121-66.
28. Ligtenberg MJ, Kuiper RP, Chan TL, et al. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. Nat Genet 2009;41:112-7.
29. Rumilla K, Schowalter KV, Lindor NM, et al. Frequency of deletions of EPCAM (TACSTD1) in MSH2-associated Lynch syndrome cases. J Mol Diagn 2011;13:93-9.
30. Martin-Lopez JV, Fishel R. The mechanism of mismatch repair and the functional analysis of mismatch repair defects in Lynch syndrome. Fam Cancer 2013;12:159-68.
31. Jiricny J. The multifaceted mismatch-repair system. Nat Rev Mol Cell Biol 2006;7:335-46.
32. Li GM. Mechanisms and functions of DNA mismatch repair. Cell Res 2008;18:85-98.

33. Boland CR, Goel A. Microsatellite instability in colorectal cancer. Gastroenterology 2010;138:2073-2087 e3.
34. Boland CR. Recent discoveries in the molecular genetics of Lynch syndrome. Fam Cancer 2016;15:395-403.
35. Tougeron D, Fauquembergue E, Rouquette A, et al. Tumor-infiltrating lymphocytes in colorectal cancers with microsatellite instability are correlated with the number and spectrum of frameshift mutations. Mod Pathol 2009;22:1186-95.
36. Smyrk TC, Watson P, Kaul K, et al. Tumor-infiltrating lymphocytes are a marker for microsatellite instability in colorectal carcinoma. Cancer 2001;91:2417-22.
37. Thibodeau SN, Bren G, Schaid D. Microsatellite instability in cancer of the proximal colon. Science 1993;260:816-9.
38. Phillips SM, Banerjea A, Feakins R, et al. Tumour-infiltrating lymphocytes in colorectal cancer with microsatellite instability are activated and cytotoxic. British Journal of Surgery 2004;91:469-475.
39. Kuismanen SA, Holmberg MT, Salovaara R, et al. Genetic and Epigenetic Modification of MLH1 Accounts for a Major Share of Microsatellite-Unstable Colorectal Cancers. The American Journal of Pathology 2000;156:1773-1779.
40. Veigl ML, Kasturi L, Olechnowicz J, et al. Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. Proc Natl Acad Sci U S A 1998;95:8698-702.
41. Miyakura Y, Sugano K, Konishi F, et al. Extensive methylation of hMLH1 promoter region predominates in proximal colon cancer with microsatellite instability. Gastroenterology 2001;121:1300-1309.
42. Young J, Simms LA, Biden KG, et al. Features of colorectal cancers with high-level microsatellite instability occurring in familial and sporadic settings: parallel pathways of tumorigenesis. Am J Pathol 2001;159:2107-16.
43. Crucianelli F, Tricarico R, Turchetti D, et al. MLH1 constitutional and somatic methylation in patients with MLH1 negative tumors fulfilling the revised Bethesda criteria. Epigenetics 2014;9:1431-8.
44. Lynch HT, Lynch PM, Lanspa SJ, et al. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. Clin Genet 2009;76:1-18.
45. Newton K, Jorgensen NM, Wallace AJ, et al. Tumour MLH1 promoter region methylation testing is an effective prescreen for Lynch Syndrome (HNPCC). J Med Genet 2014;51:789-96.
46. Evaluation of Genomic Applications in P, Prevention Working G. Recommendations from the EGAPP Working Group: genetic testing strategies in newly diagnosed individuals with colorectal cancer aimed at reducing morbidity and mortality from Lynch syndrome in relatives. Genet Med 2009;11:35-41.
47. Parsons MT, Buchanan DD, Thompson B, et al. Correlation of tumour BRAF mutations and MLH1 methylation with germline mismatch repair (MMR) gene mutation status: a literature review assessing utility of tumour features for MMR variant classification. J Med Genet 2012;49:151-7.
48. Moreira L, Munoz J, Cuatrecasas M, et al. Prevalence of somatic mutl homolog 1 promoter hypermethylation in Lynch syndrome colorectal cancer. Cancer 2015;121:1395-404.
49. van Roon EH, van Puijenbroek M, Middeldorp A, et al. Early onset MSI-H colon cancer with MLH1 promoter methylation, is there a genetic predisposition? BMC Cancer 2010;10:180.
50. Goel A, Nguyen TP, Leung HC, et al. De novo constitutional MLH1 epimutations confer early-onset colorectal cancer in two new sporadic Lynch syndrome cases, with derivation of the epimutation on the paternal allele in one. Int J Cancer 2011;128:869-78.
51. Niessen RC, Hofstra RM, Westers H, et al. Germline hypermethylation of MLH1 and EPCAM deletions are a frequent cause of Lynch syndrome. Genes Chromosomes Cancer 2009;48:737-44.
52. Suter CM, Martin DI, Ward RL. Germline epimutation of MLH1 in individuals with multiple cancers. Nat Genet 2004;36:497-501.
53. Valle L, Carbonell P, Fernandez V, et al. MLH1 germline epimutations in selected patients with early-onset non-polyposis colorectal cancer. Clin Genet 2007;71:232-7.
54. Ward RL, Dobbins T, Lindor NM, et al. Identification of constitutional MLH1 epimutations and promoter variants in colorectal cancer patients from the Colon Cancer Family Registry. Genet Med 2013;15:25-35.
55. Miyakura Y, Sugano K, Akasu T, et al. Extensive but hemiallelic methylation of the hMLH1 promoter region in early-onset sporadic colon cancers with microsatellite instability. Clin Gastroenterol Hepatol 2004;2:147-56.
56. Hitchins MP. Finding the needle in a haystack: identification of cases of Lynch syndrome with MLH1 epimutation. Fam Cancer 2016;15:413-22.
57. Castillejo A, Hernandez-Illan E, Rodriguez-Soler M, et al. Prevalence of MLH1 constitutional epimutations as a cause of Lynch syndrome in unselected versus selected consecutive series of patients with colorectal cancer. J Med Genet 2015;52:498-502.
58. Haraldsdottir S, Hampel H, Wu C, et al. Patients with colorectal cancer associated with Lynch syndrome and MLH1 promoter hypermethylation have similar prognoses. Genet Med 2016;18:863-8.
59. Crepin M, Dieu MC, Lejeune S, et al. Evidence of constitutional MLH1 epimutation associated to transgenerational inheritance of cancer susceptibility. Hum Mutat 2012;33:180-8.

60. Hitchins MP, Wong JJ, Suthers G, et al. Inheritance of a cancer-associated MLH1 germ-line epimutation. N Engl J Med 2007;356:697-705.
61. Morak M, Schackert HK, Rahner N, et al. Further evidence for heritability of an epimutation in one of 12 cases with MLH1 promoter methylation in blood cells clinically displaying HNPCC. Eur J Hum Genet 2008;16:804-11.
62. Choi YH, Cotterchio M, McKeown-Eyssen G, et al. Penetrance of colorectal cancer among MLH1/MSH2 carriers participating in the colorectal cancer familial registry in Ontario. Hered Cancer Clin Pract 2009;7:14.
63. Dowty JG, Win AK, Buchanan DD, et al. Cancer risks for MLH1 and MSH2 mutation carriers. Hum Mutat 2013;34:490-7.
64. Baglietto L, Lindor NM, Dowty JG, et al. Risks of Lynch syndrome cancers for MSH6 mutation carriers. J Natl Cancer Inst 2010;102:193-201.
65. Hendriks YMC, Wagner A, Morreau H, et al. Cancer risk in hereditary nonpolyposis colorectal cancer due to MSH6 mutations: impact on counseling and surveillance. Gastroenterology 2004;127:17-25.
66. ten Broeke SW, Brohet RM, Tops CM, et al. Lynch syndrome caused by germline PMS2 mutations: delineating the cancer risk. J Clin Oncol 2015;33:319-25.
67. Senter L, Clendenning M, Sotamaa K, et al. The clinical phenotype of Lynch syndrome due to germ-line PMS2 mutations. Gastroenterology 2008;135:419-28.
68. Stoffel E, Mukherjee B, Raymond VM, et al. Calculation of risk of colorectal and endometrial cancer among patients with Lynch syndrome. Gastroenterology 2009;137:1621-7.
69. Quehenberger F, Vasen HF, van Houwelingen HC. Risk of colorectal and endometrial cancer for carriers of mutations of the hMLH1 and hMSH2 gene: correction for ascertainment. J Med Genet 2005;42:491-6.
70. Aarnio M, Sankila R, Pukkala E, et al. Cancer risk in mutation carriers of DNA-mismatch-repair genes. Int J Cancer 1999;81:214-8.
71. Ramsoekh D, Wagner A, van Leerdam ME, et al. Cancer risk in MLH1, MSH2 and MSH6 mutation carriers; different risk profiles may influence clinical management. Hereditary Cancer in Clinical Practice 2009;7:17-17.
72. Jenkins MA, Baglietto L, Dowty JG, et al. Cancer risks for mismatch repair gene mutation carriers: a population-based early onset case-family study. Clin Gastroenterol Hepatol 2006;4:489-98.
73. Win AK, Lindor NM, Jenkins MA. Risk of breast cancer in Lynch syndrome: a systematic review. Breast Cancer Res 2013;15:R27.
74. Bauer CM, Ray AM, Halstead-Nussloch BA, et al. Hereditary prostate cancer as a feature of Lynch syndrome. Fam Cancer 2011;10:37-42.
75. Haraldsdottir S, Hampel H, Wei L, et al. Prostate cancer incidence in males with Lynch syndrome. Genet Med 2014;16:553-7.
76. Lindor NM, Petersen GM, Hadley DW, et al. Recommendations for the care of individuals with an inherited predisposition to Lynch syndrome: a systematic review. JAMA 2006;296:1507-17.
77. Vasen HFA, Blanco I, Aktan-Collan K, et al. Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. Gut 2013.
78. Le DT, Uram JN, Wang H, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N Engl J Med 2015;372:2509-20.
79. Castro MP, Goldstein N. Mismatch repair deficiency associated with complete remission to combination programmed cell death ligand immune therapy in a patient with sporadic urothelial carcinoma: immunotheranostic considerations. J Immunother Cancer 2015;3:58.
80. Vasen HF, Mecklin JP, Khan PM, et al. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Dis Colon Rectum 1991;34.
81. Bellacosa A, Genuardi M, Anti M, et al. Hereditary nonpolyposis colorectal cancer: review of clinical, molecular genetics, and counseling aspects. Am J Med Genet 1996;62:353-64.
82. Syngal S, Fox EA, Eng C, et al. Sensitivity and specificity of clinical criteria for hereditary non-polyposis colorectal cancer associated mutations in MSH2 and MLH1. J Med Genet 2000;37:641-5.
83. Rodriguez-Bigas MA, Boland CR, Hamilton SR, et al. A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines. J Natl Cancer Inst 1997;89.
84. Hampel H, Frankel WL, Martin E, et al. Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). N Engl J Med 2005;352.
85. Terdiman JP, Gum JR, Jr., Conrad PG, et al. Efficient detection of hereditary nonpolyposis colorectal cancer gene carriers by screening for tumor microsatellite instability before germline genetic testing. Gastroenterology 2001;120:21-30.
86. Umar A, Boland CR, Terdiman JP, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal

cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst 2004;96:261-8.
87. Boland CR, Thibodeau SN, Hamilton SR, et al. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res 1998;58:5248-57.
88. Lindor NM, Burgart LJ, Leontovich O, et al. Immunohistochemistry versus microsatellite instability testing in phenotyping colorectal tumors. J Clin Oncol 2002;20.
89. Dietmaier W, Wallinger S, Bocker T, et al. Diagnostic microsatellite instability: definition and correlation with mismatch repair protein expression. Cancer Res 1997;57:4749-56.
90. Sood AK, Holmes R, Hendrix MJ, et al. Application of the National Cancer Institute international criteria for determination of microsatellite instability in ovarian cancer. Cancer Res 2001;61:4371-4.
91. Pinol V, Castells A, Andreu M, et al. Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer. JAMA 2005;293:1986-94.
92. Barnetson RA, Tenesa A, Farrington SM, et al. Identification and survival of carriers of mutations in DNA mismatch-repair genes in colon cancer. N Engl J Med 2006;354:2751-63.
93. van Lier MG, Leenen CH, Wagner A, et al. Yield of routine molecular analyses in colorectal cancer patients </=70 years to detect underlying Lynch syndrome. J Pathol 2012;226:764-74.
94. Buchanan DD, Rosty C, Clendenning M, et al. Clinical problems of colorectal cancer and endometrial cancer cases with unknown cause of tumor mismatch repair deficiency (suspected Lynch syndrome). Appl Clin Genet 2014;7:183-93.
95. Rodriguez-Soler M, Perez-Carbonell L, Guarinos C, et al. Risk of cancer in cases of suspected lynch syndrome without germline mutation. Gastroenterology 2013;144:926-932 e1; quiz e13-4.
96. Carethers JM. Differentiating Lynch-like from Lynch syndrome. Gastroenterology 2014;146:602-4.
97. Mensenkamp AR, Vogelaar IP, van Zelst-Stams WA, et al. Somatic mutations in MLH1 and MSH2 are a frequent cause of mismatch-repair deficiency in Lynch syndrome-like tumors. Gastroenterology 2014;146:643-646 e8.
98. Haraldsdottir S, Hampel H, Tomsic J, et al. Colon and endometrial cancers with mismatch repair deficiency can arise from somatic, rather than germline, mutations. Gastroenterology 2014;147:1308-1316 e1.
99. Geurts-Giele WR, Leenen CH, Dubbink HJ, et al. Somatic aberrations of mismatch repair genes as a cause of microsatellite-unstable cancers. J Pathol 2014;234:548-59.
100. Overbeek LI, Kets CM, Hebeda KM, et al. Patients with an unexplained microsatellite instable tumour have a low risk of familial cancer. Br J Cancer 2007;96:1605-12.
101. Thompson BA, Spurdle AB, Plazzer JP, et al. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nat Genet 2014;46:107-15.
102. van der Klift HM, Jansen AM, van der Steenstraten N, et al. Splicing analysis for exonic and intronic mismatch repair gene variants associated with Lynch syndrome confirms high concordance between minigene assays and patient RNA analyses. Mol Genet Genomic Med 2015;3:327-45.
103. Drost M, Zonneveld J, van Dijk L, et al. A cell-free assay for the functional analysis of variants of the mismatch repair protein MLH1. Hum Mutat 2010;31:247-53.
104. Drost M, Zonneveld JB, van Hees S, et al. A rapid and cell-free assay to test the activity of lynch syndrome-associated MSH2 and MSH6 missense variants. Hum Mutat 2012;33:488-94.
105. Vasen HF, Ghorbanoghli Z, Bourdeaut F, et al. Guidelines for surveillance of individuals with constitutional mismatch repair-deficiency proposed by the European Consortium "Care for CMMR-D" (C4CMMR-D). J Med Genet 2014;51:283-93.
106. Wimmer K, Kratz CP, Vasen HF, et al. Diagnostic criteria for constitutional mismatch repair deficiency syndrome: suggestions of the European consortium 'care for CMMRD' (C4CMMRD). J Med Genet 2014;51:355-65.
107. Ricciardone MD, Ozcelik T, Cevher B, et al. Human MLH1 deficiency predisposes to hematological malignancy and neurofibromatosis type 1. Cancer Res 1999;59:290-3.
108. Wang Q, Lasset C, Desseigne F, et al. Neurofibromatosis and early onset of cancers in hMLH1-deficient children. Cancer Res 1999;59:294-7.
109. Ripperger T, Beger C, Rahner N, et al. Constitutional mismatch repair deficiency and childhood leukemia/lymphoma--report on a novel biallelic MSH6 mutation. Haematologica 2010;95:841-4.
110. Ponti G, Ponz de Leon M. Muir-Torre syndrome. Lancet Oncol 2005;6:980-7.
111. Perera S, Ramyar L, Mitri A, et al. A novel complex mutation in MSH2 contributes to both Muir-Torre and Lynch Syndrome. J Hum Genet 2010;55:37-41.
112. Kruse R, Lamberti C, Wang Y, et al. Is the mismatch repair deficient type of Muir-Torre syndrome confined to

mutations in the hMSH2 gene? Hum Genet 1996;98:747-50.
113. Mangold E, Pagenstecher C, Leister M, et al. A genotype-phenotype correlation in HNPCC: strong predominance of msh2 mutations in 41 patients with Muir-Torre syndrome. J Med Genet 2004;41:567-72.
114. South CD, Hampel H, Comeras I, et al. The frequency of Muir-Torre syndrome among Lynch syndrome families. J Natl Cancer Inst 2008;100:277-81.
115. Guillen-Ponce C, Castillejo A, Barbera VM, et al. Biallelic MYH germline mutations as cause of Muir-Torre syndrome. Fam Cancer 2010;9:151-4.
116. Ponti G, Ponz de Leon M, Maffei S, et al. Attenuated familial adenomatous polyposis and Muir-Torre syndrome linked to compound biallelic constitutional MYH gene mutations. Clin Genet 2005;68:442-7.
117. Lindor NM, Rabe K, Petersen GM, et al. Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency - Familial colorectal cancer type X. Jama-Journal of the American Medical Association 2005;293:1979-1985.
118. Balmana J, Castells A, Cervantes A, et al. Familial colorectal cancer risk: ESMO Clinical Practice Guidelines. Ann Oncol 2010;21 Suppl 5:v78-81.
119. Shiovitz S, Copeland WK, Passarelli MN, et al. Characterisation of familial colorectal cancer Type X, Lynch syndrome, and non-familial colorectal cancer. Br J Cancer 2014;111:598-602.
120. Da Silva F, Wernhoff P, Dominguez-Barrera C, et al. Update on Hereditary Colorectal Cancer. Anticancer Res 2016;36:4399-405.
121. Lindberg LJ, Ladelund S, Frederiksen BL, et al. Outcome of 24 years national surveillance in different hereditary colorectal cancer subgroups leading to more individualised surveillance. J Med Genet 2017;54:297-304.
122. Spurdle AB, Couch FJ, Hogervorst FB, et al. Prediction and assessment of splicing alterations: implications for clinical testing. Hum Mutat 2008;29:1304-13.
123. Francisco I, Albuquerque C, Lage P, et al. Familial colorectal cancer type X syndrome: two distinct molecular entities? Fam Cancer 2011;10:623-31.
124. Dominguez-Valentin M, Therkildsen C, Da Silva S, et al. Familial colorectal cancer type X: genetic profiles and phenotypic features. Mod Pathol 2015;28:30-6.
125. Schulz E, Klampfl P, Holzapfel S, et al. Germline variants in the SEMA4A gene predispose to familial colorectal cancer type X. Nat Commun 2014;5:5191.
126. Nieminen TT, Abdel-Rahman WM, Ristimaki A, et al. BMPR1A Mutations in Hereditary Nonpolyposis Colorectal Cancer Without Mismatch Repair Deficiency. Gastroenterology 2011;141:E23-E26.
127. Groden J, Thliveris A, Samowitz W, et al. Identification and characterization of the familial adenomatous polyposis coli gene. Cell 1991;66:589-600.
128. Talseth-Palmer BA. The genetic basis of colonic adenomatous polyposis syndromes. Hered Cancer Clin Pract 2017;15:5.
129. Grady WM. Genetic testing for high-risk colon cancer patients. Gastroenterology 2003;124:1574-94.
130. Leoz ML, Carballal S, Moreira L, et al. The genetic basis of familial adenomatous polyposis and its implications for clinical practice and risk management. Appl Clin Genet 2015;8:95-107.
131. Aretz S, Uhlhaas S, Caspari R, et al. Frequency and parental origin of de novo APC mutations in familial adenomatous polyposis. Eur J Hum Genet 2004;12:52-8.
132. Ripa R, Bisgaard ML, Bulow S, et al. De novo mutations in familial adenomatous polyposis (FAP). Eur J Hum Genet 2002;10:631-7.
133. Albuquerque C, Bakker ER, van Veelen W, et al. Colorectal cancers choosing sides. Biochim Biophys Acta 2011;1816:219-31.
134. Nielsen M, Hes FJ, Nagengast FM, et al. Germline mutations in APC and MUTYH are responsible for the majority of families with attenuated familial adenomatous polyposis. Clin Genet 2007;71:427-33.
135. Nieuwenhuis MH, Vasen HF. Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): a review of the literature. Crit Rev Oncol Hematol 2007;61:153-61.
136. Hes FJ, Nielsen M, Bik EC, et al. Somatic APC mosaicism: an underestimated cause of polyposis coli. Gut 2008;57:71-6.
137. Aretz S, Stienen D, Friedrichs N, et al. Somatic APC mosaicism: a frequent cause of familial adenomatous polyposis (FAP). Hum Mutat 2007;28:985-92.
138. Farrington SM, Dunlop MG. Mosaicism and sporadic familial adenomatous polyposis. Am J Hum Genet 1999;64:653-8.
139. Tuohy TM, Burt RW. Somatic mosaicism: a cause for unexplained cases of FAP? Gut 2008;57:10-2.
140. Al-Tassan N, Chmiel NH, Maynard J, et al. Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors. Nat Genet 2002;30:227-32.
141. Chow E, Thirlwell C, Macrae F, et al. Colorectal cancer and inherited mutations in base-excision repair. The Lancet Oncology 2004;5:600-606.

142. Nielsen M, Morreau H, Vasen HF, et al. MUTYH-associated polyposis (MAP). Crit Rev Oncol Hematol 2011;79:1-16.
143. Kashfi SM, Golmohammadi M, Behboudi F, et al. MUTYH the base excision repair gene family member associated with colorectal cancer polyposis. Gastroenterol Hepatol Bed Bench 2013;6:S1-S10.
144. Lubbe SJ, Di Bernardo MC, Chandler IP, et al. Clinical implications of the colorectal cancer risk associated with MUTYH mutation. J Clin Oncol 2009;27:3975-80.
145. Croitoru ME, Cleary SP, Di Nicola N, et al. Association between biallelic and monoallelic germline MYH gene mutations and colorectal cancer risk. J Natl Cancer Inst 2004;96:1631-4.
146. Hayashi H, Tominaga Y, Hirano S, et al. Replication-associated repair of adenine:8-oxoguanine mispairs by MYH. Curr Biol 2002;12:335-9.
147. Thomas D, Scot AD, Barbey R, et al. Inactivation of OGG1 increases the incidence of G . C-->T . A transversions in Saccharomyces cerevisiae: evidence for endogenous oxidative damage to DNA in eukaryotic cells. Mol Gen Genet 1997;254:171-8.
148. Nghiem Y, Cabrera M, Cupples CG, et al. The mutY gene: a mutator locus in Escherichia coli that generates G.C----T.A transversions. Proc Natl Acad Sci U S A 1988;85:2709-13.
149. Nielsen M, de Miranda NF, van Puijenbroek M, et al. Colorectal carcinomas in MUTYH-associated polyposis display histopathological similarities to microsatellite unstable carcinomas. BMC Cancer 2009;9:184.
150. Lipton L, Halford SE, Johnson V, et al. Carcinogenesis in MYH-associated polyposis follows a distinct genetic pathway. Cancer Res 2003;63:7595-9.
151. Tenesa A, Campbell H, Barnetson R, et al. Association of MUTYH and colorectal cancer. Br J Cancer 2006;95:239-42.
152. Venesio T, Balsamo A, D'Agostino VG, et al. MUTYH-associated polyposis (MAP), the syndrome implicating base excision repair in inherited predisposition to colorectal tumors. Front Oncol 2012;2:83.
153. Win AK, Hopper JL, Jenkins MA. Association between monoallelic MUTYH mutation and colorectal cancer risk: a meta-regression analysis. Fam Cancer 2011;10:1-9.
154. Win AK, Cleary SP, Dowty JG, et al. Cancer risks for monoallelic MUTYH mutation carriers with a family history of colorectal cancer. Int J Cancer 2011;129:2256-62.
155. Khalaf R, Jones C, Strutt W, et al. Colorectal cancer in a monoallelic MYH mutation carrier. J Gastrointest Surg 2013;17:1500-2.
156. Farrington SM, Tenesa A, Barnetson R, et al. Germline susceptibility to colorectal cancer due to base-excision repair gene defects. Am J Hum Genet 2005;77:112-9.
157. Jones N, Vogt S, Nielsen M, et al. Increased colorectal cancer incidence in obligate carriers of heterozygous mutations in MUTYH. Gastroenterology 2009;137:489-94, 494 e1; quiz 725-6.
158. Palles C, Cazier JB, Howarth KM, et al. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nat Genet 2013;45:136-44.
159. Heitzer E, Tomlinson I. Replicative DNA polymerase mutations in cancer. Curr Opin Genet Dev 2014;24:107-13.
160. Church DN, Briggs SE, Palles C, et al. DNA polymerase epsilon and delta exonuclease domain mutations in endometrial cancer. Hum Mol Genet 2013;22:2820-8.
161. Shinbrot E, Henninger EE, Weinhold N, et al. Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. Genome Res 2014;24:1740-50.
162. Briggs S, Tomlinson I. Germline and somatic polymerase epsilon and delta mutations define a new class of hypermutated colorectal and endometrial cancers. J Pathol 2013;230:148-53.
163. Valle L, Hernandez-Illan E, Bellido F, et al. New insights into POLE and POLD1 germline mutations in familial colorectal cancer and polyposis. Hum Mol Genet 2014;23:3506-12.
164. Church JM. Polymerase proofreading-associated polyposis: a new, dominantly inherited syndrome of hereditary colorectal cancer predisposition. Dis Colon Rectum 2014;57:396-7.
165. Bellido F, Pineda M, Aiza G, et al. POLE and POLD1 mutations in 529 kindred with familial colorectal cancer and/or polyposis: review of reported cases and recommendations for genetic testing and surveillance. Genet Med 2015.
166. Kane DP, Shcherbakova PV. A common cancer-associated DNA polymerase epsilon mutation causes an exceptionally strong mutator phenotype, indicating fidelity defects distinct from loss of proofreading. Cancer Res 2014;74:1895-901.
167. Elsayed FA, Kets CM, Ruano D, et al. Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer. Eur J Hum Genet 2015;23:1080-4.
168. Billingsley CC, Cohn DE, Mutch DG, et al. Polymerase varepsilon (POLE) mutations in endometrial cancer: clinical outcomes and implications for Lynch syndrome testing. Cancer 2015;121:386-94.
169. Domingo E, Freeman-Mills L, Rayner E, et al. Somatic POLE proofreading domain mutation, immune re-

sponse, and prognosis in colorectal cancer: a retrospective, pooled biomarker study. The Lancet Gastroenterology & Hepatology 2016;1:207-216.
170. Bellone S, Centritto F, Black J, et al. Polymerase epsilon (POLE) ultra-mutated tumors induce robust tumor-specific CD4+ T cell responses in endometrial cancer patients. Gynecol Oncol 2015;138:11-7.
171. Weren RD, Ligtenberg MJ, Kets CM, et al. A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. Nat Genet 2015;47:668-71.
172. Rivera B, Castellsague E, Bah I, et al. Biallelic NTHL1 Mutations in a Woman with Multiple Primary Tumors. N Engl J Med 2015;373:1985-6.
173. Broderick P, Dobbins SE, Chubb D, et al. Validation of Recently Proposed Colorectal Cancer Susceptibility Gene Variants in an Analysis of Families and Patients-a Systematic Review. Gastroenterology 2017;152:75-77 e4.
174. Chubb D, Broderick P, Dobbins SE, et al. Rare disruptive mutations and their contribution to the heritable risk of colorectal cancer. Nat Commun 2016;7:11883.
175. Li J, Braganza A, Sobol RW. Base excision repair facilitates a functional relationship between Guanine oxidation and histone demethylation. Antioxid Redox Signal 2013;18:2429-43.
176. Adam R, Spier I, Zhao B, et al. Exome Sequencing Identifies Biallelic MSH3 Germline Mutations as a Recessive Subtype of Colorectal Adenomatous Polyposis. Am J Hum Genet 2016;99:337-51.
177. Zbuk KM, Eng C. Hamartomatous polyposis syndromes. Nat Clin Pract Gastroenterol Hepatol 2007;4:492-502.
178. Manfredi M. Hereditary Hamartomatous Polyposis Syndromes: Understanding the Disease Risks As Children Reach Adulthood. Gastroenterology & Hepatology 2010;6:185-196.
179. Brosens LA, Langeveld D, van Hattem WA, et al. Juvenile polyposis syndrome. World J Gastroenterol 2011;17:4839-44.
180. Calva-Cerqueira D, Chinnathambi S, Pechman B, et al. The rate of germline mutations and large deletions of SMAD4 and BMPR1A in juvenile polyposis. Clin Genet 2009;75:79-85.
181. Aretz S, Stienen D, Uhlhaas S, et al. High proportion of large genomic deletions and a genotype phenotype update in 80 unrelated families with juvenile polyposis syndrome. J Med Genet 2007;44:702-9.
182. van Hattem WA, Brosens LA, de Leng WW, et al. Large genomic deletions of SMAD4, BMPR1A and PTEN in juvenile polyposis. Gut 2008;57:623-7.
183. Giardiello FM, Welsh SB, Hamilton SR, et al. Increased risk of cancer in the Peutz-Jeghers syndrome. N Engl J Med 1987;316:1511-4.
184. Jenne DE, Reimann H, Nezu J, et al. Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. Nature Genetics 1998;18:38-44.
185. Beggs AD, Latchford AR, Vasen HF, et al. Peutz-Jeghers syndrome: a systematic review and recommendations for management. Gut 2010;59:975-86.
186. Giardiello FM, Trimbath JD. Peutz-Jeghers syndrome and management recommendations. Clin Gastroenterol Hepatol 2006;4:408-15.
187. Foster DR, Foster DB. Gall-bladder polyps in Peutz-Jeghers syndrome. Postgrad Med J 1980;56:373-6.
188. Hearle N, Schumacher V, Menko FH, et al. Frequency and spectrum of cancers in the Peutz-Jeghers syndrome. Clin Cancer Res 2006;12:3209-15.
189. Giardiello FM, Brensinger JD, Tersmette AC, et al. Very high risk of cancer in familial Peutz–Jeghers syndrome. Gastroenterology 2000;119:1447-1453.
190. Dunlop MG, British Society for G, Association of Coloproctology for Great B, et al. Guidance on gastrointestinal surveillance for hereditary non-polyposis colorectal cancer, familial adenomatous polypolis, juvenile polyposis, and Peutz-Jeghers syndrome. Gut 2002;51 Suppl 5:V21-7.
191. Lim W, Hearle N, Shah B, et al. Further observations on LKB1/STK11 status and cancer risk in Peutz-Jeghers syndrome. Br J Cancer 2003;89:308-13.
192. Volikos E, Robinson J, Aittomaki K, et al. LKB1 exonic and whole gene deletions are a common cause of Peutz-Jeghers syndrome. J Med Genet 2006;43:e18.
193. Smpokou P, Fox VL, Tan WH. PTEN hamartoma tumour syndrome: early tumour development in children. Arch Dis Child 2015;100:34-7.
194. Pilarski R. Cowden syndrome: a critical review of the clinical literature. J Genet Couns 2009;18:13-27.
195. Song MS, Salmena L, Pandolfi PP. The functions and regulation of the PTEN tumour suppressor. Nat Rev Mol Cell Biol 2012;13:283-96.
196. Tamguney T, Stokoe D. New insights into PTEN. J Cell Sci 2007;120:4071-9.
197. Mester J, Eng C. Cowden syndrome: recognizing and managing a not-so-rare hereditary cancer syndrome. J Surg Oncol 2015;111:125-30.
198. Pilarski R, Burt R, Kohlman W, et al. Cowden syndrome and the PTEN hamartoma tumor syndrome: system-

atic review and revised diagnostic criteria. J Natl Cancer Inst 2013;105:1607-16.
199. Gammon A, Jasperson K, Champine M. Genetic basis of Cowden syndrome and its implications for clinical practice and risk management. Appl Clin Genet 2016;9:83-92.
200. Lachlan KL. Cowden syndrome and the PTEN hamartoma tumor syndrome: how to define rare genetic syndromes. J Natl Cancer Inst 2013;105:1595-7.

Chapter 2

Whole gene capture analysis of 15 CRC susceptibility genes in suspected Lynch Syndrome patients

AML Jansen[1,2¶], MA Geilenkirchen[1¶], T van Wezel[2], SC Jagmohan-Changur[1], D Ruano[2], HM van der Klift[1], BEWM van den Akker[2], JFJ Laros[1], M van Galen[1], A Wagner[3], TGW Letteboer[4], EB Gómez-Garcia[5], CMJ Tops[6], HF Vasen[7], P Devilee[1], FJ Hes[6], H Morreau[2], JT Wijnen[1,6]

PLoS One, 2016; 11(6):e0157381

[¶]Authors contributed equally to this work

Departments of [1]Human Genetics, [2]Pathology, [6]Clinical Genetics and [7]Gastroenterology and Hepatology, Leiden University Medical Center, Leiden, The Netherlands

[3]Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands

[4]Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands

[5]Department of Clinical Genetics, University Hospital Maastricht, Maastricht, The Netherlands

Abstract

Background and Aims: Lynch Syndrome (LS) is caused by pathogenic germline variants in one of the mismatch repair (MMR) genes. However, up to 60% of MMR-deficient colorectal cancer cases are categorized as suspected Lynch Syndrome (sLS) because no pathogenic MMR germline variant can be identified, which leads to difficulties in clinical management. We therefore analyzed the genomic regions of 15 CRC susceptibility genes in leukocyte DNA of 34 unrelated sLS patients and 11 patients with *MLH1* hypermethylated tumors with a clear family history.

Methods: Using targeted next-generation sequencing, we analyzed the entire non-repetitive genomic sequence, including intronic and regulatory sequences, of 15 CRC susceptibility genes. In addition, tumor DNA from 28 sLS patients was analyzed for somatic MMR variants.

Results: Of 1979 germline variants found in the leukocyte DNA of 34 sLS patients, one was a pathogenic variant (*MLH1* c.1667+1delG). Leukocyte DNA of 11 patients with *MLH1* hypermethylated tumors was negative for pathogenic germline variants in the tested CRC susceptibility genes and for germline *MLH1* hypermethylation. Somatic DNA analysis of 28 sLS tumors identified eight (29%) cases with two pathogenic somatic variants, one with a VUS predicted to pathogenic and LOH, and nine cases (32%) with one pathogenic somatic variant (n=8) or one VUS predicted to be pathogenic (n=1).

Conclusions: This is the first study in sLS patients to include the entire genomic sequence of CRC susceptibility genes. An underlying somatic or germline MMR gene defect was identified in ten of 34 sLS patients (29%). In the remaining sLS patients, the underlying genetic defect explaining the MMR-deficiency in their tumors might be found outside the genomic regions harboring the MMR and other known CRC susceptibility genes.

Keywords: Genetics; Colorectal Cancer; Mismatch Repair Deficiency; Lynch Syndrome

Introduction

Lynch Syndrome (LS) is the most common form of hereditary colorectal cancer (CRC) and is caused by heterozygous pathogenic germline variants in one of the mismatch repair (MMR) genes (*MLH1*, *MSH2*, *MSH6*, *PMS2*).[1] In addition, a unique subgroup of LS patients carry deletions in the 3' end of *EPCAM*, a gene immediately upstream of *MSH2*.[2, 3] More than 95% of LS-associated CRCs display microsatellite instability (MSI), the molecular hallmark of LS.[4] Immunohistochemical analysis (IHC) of the tumor for loss of MMR protein expression points to a possible causative gene, with the diagnosis of LS confirmed once a pathogenic germline variant is identified. Patients suspect for LS are selected for genetic testing on the basis of clinical characteristics (Amsterdam or Bethesda criteria) and/or molecular diagnostic testing of the LS-associated tumors (LSAT).[5, 6] As opposed to familial colorectal cancer type X (FCCTX) families,[7] who also fulfill Amsterdam criteria, the patients suspect for LS do show MSI and loss of MMR gene expression in the tumor. LS patients have an increased risk of developing CRC, with estimates of lifetime risk ranging from 36% to 75%.[8-11] Women who carry pathogenic variants also face a high risk of endometrial cancer.[12] Several other cancer types, including small bowel, stomach, pancreas, ovary, renal, pelvis, ureter, bladder, brain, liver, bile duct, gall bladder and skin occur frequently.[11, 13-15] Recent studies also indicate an increased risk for prostate and breast cancer.[16-18] To achieve adequate cancer prevention, early identification of individuals with LS is essential. Intensive surveillance by colonoscopy every 1-2 years, starting at age 20 to 25, is now recommended and is known to reduce CRC morbidity and mortality.[19, 20] Accurate and timely identification of LS patients is therefore crucial to providing the correct treatment.[21]

A recent study estimated that, using current approaches, up to 60% of MMR-deficient colorectal cancer cases remain unexplained.[21] These patients are designated as 'suspected Lynch Syndrome' (sLS)[21], or also known as 'Lynch-Like Syndrome'[22], and failure to determine the underlying cause of disease has a major impact on the clinical management of these patients. In addition to germline variants, biallelic somatic variants may explain disease in up to 69% of MMR-deficient tumors that lack pathogenic germline variants or *MLH1* promoter hypermethylation.[23-25] MSI due to somatic hypermethylation of the promoter region of *MLH1* is also seen in up to 15% of sporadic CRC patients.[26] Sporadic *MLH1* methylated tumors commonly occur at a relatively advanced age and in absence of family history of CRC.[27-29] Patients with somatic *MLH1* promoter hypermethylated tumors rarely carry germline MMR variants, although exceptions have been published.[30-32] These studies indicate *MLH1* hypermethylation as a 'second-hit' mechanism already present in adenoma stage and demonstrate that *MLH1* hypermethylation does not exclude the presence of germline pathogenic CRC variants. *MLH1* hypermethylated tumors in young patients with a family history of CRC can also be due to germline *MLH1* hypermethylation. Though very rare, this phenomenon has been described before.[33-41] Inheritance of a constitutional epimutation has been shown in at least three unrelated families.[42-44]

The aim of our study was to identify an underlying genetic basis in a cohort of 34 sLS patients and 11 patients with *MLH1* hypermethylated tumors and a clear family history for LS. In an effort to discover previously undetected germline variants, the entire genomic sequences of four MMR genes and eleven CRC susceptibility genes were analyzed. In addition, tumor DNA from 28 sLS tumors was analyzed for somatic variants in the MMR genes.

Materials and Methods

Study subjects
Between 1998 and 2011, a total of 45 patients were recruited from five academic centers in The Netherlands, including Leiden University Medical Center (n=20), Maastricht University Medical Center (n=11), Erasmus Medical Center (n=7), University Medical Center Utrecht (n=6) and VU University Medical Center Amsterdam (n=1). Demographic and clinical data and informed consent were obtained during the consultation. Forty-three patients fulfilled the revised Bethesda criteria.[6] All patients had been previously screened for germline variants in the MMR gene that showed loss of expression (as indicated by immunohistochemical analysis) by Sanger sequencing or denaturing gradient gel electrophoresis (DGGE), without identification of a pathogenic germline variant. Large deletions/duplications in the MMR genes were excluded by analysis with multiplex ligation-dependent probe amplification (MLPA, MRC Holland, Amsterdam), or in some cases, with Southern blot analysis.

Immunohistochemical analysis (IHC) and microsatellite instability testing were routinely performed at the request of a board-certified Clinical Genetic medical specialist. Because routine testing of all four MMR proteins only became available around 2004, tumors recruited before 2004 were not fully tested by MMR immunohistochemistry. Leukocyte and tumor DNAs were retrieved from the archives for the current study. Immunohistochemistry data was complete for 18 sLS patients (53%), for 10 cases only PMS2 immunohistochemistry was missing and the remaining 6 tumors had incomplete IHC results (see Supplementary Table 1). Ten *MLH1* hypermethylated tumors (8 colorectal-, 2 endometrium-) showed IHC loss of MLH1 and PMS2 (PMS2 was not tested in sLS-68 - see Supplementary Table 1) and normal MSH2/MSH6 expression. The eleventh patient, sLS-81, showed loss of MLH1 expression (other MMR genes were not tested). All tumors except tumor sLS-48 (MSI not tested) displayed a microsatellite unstable phenotype (high or low instability, see Table 1 and Supplementary Table 1).

Family history data showed that 82% of the sLS patients and 100% of patients in the *MLH1* hypermethylated cohort had a first-degree relative with a Lynch Syndrome-associated tumor (LSAT). Unfortunately, no DNA could be obtained from these affected family members. Among the sLS cohort, 28 patients presented with colorectal cancer (CRC) as their first LSAT, while 5 patients had endometrial cancer (EC) and 1 patient had a sebaceous gland cancer. In the *MLH1* hypermethylated cohort, 9 patients presented with CRC and 2 with EC. The mean age of diagnosis of the first LSAT was 48,6 years for the sLS group and 63,2 years for the *MLH1* hypermethylated group (See Table 1). Leukocyte DNA isolated from peripheral blood was available for all patients. The study was approved by the local medical ethical committee of the LUMC (P01-019E).

Targeted genomic sequencing with next-generation sequencing
Targeted next-generation sequencing of leukocyte DNA was carried out using a custom designed set of SureSelect 120-mer target enrichment RNA oligonucleotides (baits) for in-solution hybrid selection (Agilent Technologies, Santa Clara, CA). Baits were designed against 15 CRC susceptibility genes, spanning the entire non-repetitive genomic region of the genes, including exons, introns, and UTRs, and 5 kb upstream and 3 kb downstream of the gene (see Table 2). The average coverage was > 95% for all coding regions, and 43% for overall

coverage. Libraries were prepared according to the manufacturers' protocols (NEBNext® and Illumina®, San Diego, California, USA). In brief, 2 µg of genomic DNA from each patient was fragmented to lengths of 300-500 bp using the Covaris S220 single tube sonicator (Life Technologies, Carlsbad, CA). Fragment ends were repaired and an A-tail added to the 3' end of the DNA fragments. Illumina® dual-barcoded adaptors (patient-specific) were ligated, and the adaptor-ligated DNA was then enriched by 10 cycles of PCR. PCR products derived from 4 to 5 patients were pooled in equimolar amounts and hybridized in solution to the custom baits. Captured targets were subsequently extracted and further enriched by 15 cycles of PCR. Paired-end sequencing of the PCR products was performed on the Illumina HiSeq® 2000.

Table 1: Clinicopathologic factors sLS- and *MLH1* hypermethylated cohort

Clinicopathologic factor	no of patients (%)	
	sLS	MLH1 hypermethylated
Number of patients	34	11
Patient characteristics		
Male	17 (50)	3 (27)
Female	17 (50)	8 (73)
Age, y	48,6	63,2
Clinical characteristics		
No Bethesda/Amsterdam II	1 (3)	1 (9)
Bethesda only	23 (68)	7 (64)
Amsterdam II	10 (29)	3 (27)
Tumor		
CRC	28 (82)	9 (82)
EC	5 (15)	2 (18)
Other	1 (3)	-
Family History		
FDR	28 (82)	11 (100)
NA	3 (9)	-
No	3 (9)	-
MSI		
MSI-High	25 (74)	9 (82)
MSI-Low	4 (12)	1 (9)
MSI-Stable	4 (12)	-
Unknown	1 (3)	1 (9)

Clinicopathologic factors of the 34 sLS- and 11 MLH1 hypermethylated patients. Patients presented with colorectal cancer (CRC), endometrial cancer (EC) or other LS-associated tumors (LSAT). Family history is defined as first degree relative with a LSAT (FDR), no family history of LS (No) or family history not available (NA).

Data analysis
Illumina HiSeq® 2000 sequences were exported as FASTQ files and separated using the barcodes. The sequence data was checked for quality using the quality control tool for high throughput sequence data, FastQC (*http://www.bioinformatics.babraham.ac.uk/projects/fastqc/*). Alignment of the Illumina sequences to the human reference genome (hg19, NCBI build GRCh37) was performed using the Burrows-Wheeler aligner (BWA) (*http://bio-bwa.sourceforge.net*) and sequence duplicates were marked with Picard (*http://picard.sourceforge.net/*). Variant calling on the resulting BAM files was performed by VarScan (*http://varscan.sourceforge.net/*) using the following settings: minimal coverage of 8, minimal reads of 2, minimal variant frequency of 0.2 and a minimal average quality of 20. Variants were functionally annotated using ANNOVAR.[45]

Variant filtering and classification
The full dataset was filtered by targeted region and variant frequency. Variants with a minor allele frequency (MAF) of >0.05, as reported in the NCBI dbSNP database version 142 (*http://www.ncbi.nlm.nih.gov/projects/SNP/*) were excluded from further analysis. Because analysis of *PMS2* variants is difficult due to interference by pseudogenes, variants located in one of the duplicated regions were excluded from further analysis.

Splice variants
All remaining sequence variants with a MAF<0.05 were analyzed with Alamut software version 2.0 (Interactive Biosoftware, Roven, France), a software package that includes the splice site prediction algorithms SpliceSiteFinder, MaxEntScan (*http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html*), NNSPLICE (*http://www.fruitfly.org/seq_tools/splice.html*) and Human Splicing Finder (*http://www.umd.be/HSF/*). Variants can affect splicing by altering the canonical splice site sequence, by creation of new splice sites, activation of cryptic splice sites or by altering splice regulatory elements (SREs) [46]. In addition, branch point sequences and polypyrimidine tracts were investigated for possible variants. As a branch point is usually located 18 to 50 nt upstream of the splice acceptor site, all variants within 100 nt of the splice acceptor sites of *MLH1*, *MSH2*, *MSH6* or *PMS2* were visually inspected in Alamut.[47]

Missense prediction
All missense variants were filtered based on the predictions of *in silico* protein prediction software packages including Align GVGD, SIFT (*http://sift.jcvi.org/*), PolyPhen-2 (*http://genetics.bwh.harvard.edu/pph2/*), MutationTaster (*http://www.mutationtaster.org/*) and MutationAssessor.[48]

Promoter variants
Using the UCSC Genome browser, the putative promoter regions of *MLH1*, *MSH2*, *MSH6* and *PMS2* (as indicated by the histone mark H3K4me3 that is generally found near promoters) were analyzed for variants.

Variant classification
The Leiden Open Variation Database (LOVD, *http://www.lovd.nl/3.0/home*) and ClinVar[49] were consulted to identify previously described and classified variants. Variants that were not described in above-mentioned databases were classified according to the five-tiered InSiGHT

Table 2: Custom-designed baits used for Sureselect target enrichment of 15 CRC susceptibility genes

Chr	Genbank reference	Gene	Total target region (kb)	Chromosome		Total area covered (kb)	% of total target area	% repeated sequences
				Start	End			
1	NM_001128425.1	*MUTYH*	19,2	45791914	45811142	10,7	56%	33%
2	NM_002354.2	*EPCAM*	25,9	47591286	47617165	13,4	52%	55%
2	NM_000251.2	*MSH2*	88,1	47625262	47713360	26,1	30%	58%
2	NM_000179.2	*MSH6*	31,9	48005220	48037084	14,5	45%	42%
2	NM_000534.4	*PMS1*	101,5	190643810	190745354	48,4	48%	41%
3	NM_000249.3	*MLH1*	60,5	37034840	37095335	23,2	38%	52%
5	NM_002439.4	*MSH3*	230,3	79945293	80175633	80,6	35%	53%
5	NM_000038.5	*APC*	146,7	112038217	112184935	60,6	41%	47%
7	NM_000535.5	*PMS2*	36,7	6012370	6049037	9,8	27%	59%
10	NM_004329.2	*BMPR1a*	176,5	88511395	88687944	52,7	30%	58%
10	NM_000314.4	*PTEN*	103,3	89628194	89731531	41,8	40%	49%
14	NM_001040108.1	*MLH3*	43,8	75477466	75521235	21,3	49%	41%
16	NM_003502.3	*AXIN1*	73	334439	407464	39,1	54%	36%
17	NM_004655.3	*AXIN2*	118,5	63521684	63640183	65	55%	34%
18	NM_005359.5	*SMAD4*	62,8	48551582	48614409	31,7	50%	39%

Baits were designed against 15 CRC-susceptibility genes. Target region spans the entire genomic region, including 5 kb upstream and 3 kb downstream of the gene. Repeated sequences are not covered by custom-designed baits.

scheme: benign (class 1), likely benign (class 2), variant of unknown significance (class 3), likely pathogenic (class 4), and certainly pathogenic (class 5).[50]

Validation
All (likely) pathogenic or splice variants were visually inspected in the Integrative Genomics Viewer (IGV, *https://www.broadinstitute.org/igv/home*) and confirmed with Sanger Sequencing. Germline variants found in this study have been submitted to the appropriate LOVD database, available at *http://www.lovd.nl/3.0/home.*

MLH1 promoter hypermethylation
Methylation of the *MLH1* promoter region was analyzed using methylation specific PCR (MSP), with previously described primers.[51] Bisulfite conversion of tumor DNA was carried out using the EZ DNA methylation Gold kit (Zymo Research, Orange, US), following the manufacturer's standard protocol. Bisulfite-converted DNA was amplified using specific methylated and unmethylated primers in a PCR reaction, following a LUMC diagnostics protocol.[33]

Functional RNA analysis
To determine the effect on splicing of one splice site variant (*MLH1* c.1667+1delG), patient RNA was analyzed for aberrant splicing. RNA was isolated from short-term cultured peripheral blood lymphocytes, and analyzed with and without inhibition of nonsense-mediated decay.[46] In addition, a minigene splicing assay was performed to confirm the splicing patterns seen in the RNA of the patient as described by van der Klift et al.[46]

Somatic variant screening
DNA from 28 sLS tumors, isolated from formalin-fixed paraffin embedded tissue blocks, was screened for variants in the coding regions of *MLH1, MSH2, MSH6* and *PMS2* with the Ion PGM™ (Life Technologies, Carlsbad, CA). Next-generation sequencing was carried out according to the Ion PGM™ protocol, with supplier's materials. Primers were obtained from Life Technologies. The complete panel consisted of 162 amplicons, covering 98%, 90%, 98% and 75% of *MLH1, MSH2, MSH6* and *PMS2*, respectively.

Raw data analysis, alignments, and variant calling were carried out using the default parameters in Torrent Suite v4.0. The Variant Caller Parameter Setting was set on 'Somatic – PGM – Low Stringency'. Variants were functionally annotated using ANNOVAR.[45] Variants were visually inspected with IGV and (likely) pathogenic variants were confirmed with Sanger sequencing. The annotated dataset from the somatic screening was filtered in the same manner as the germline targeted next-generation sequencing dataset. For assessment of pathogenicity, the catalogue of somatic mutations in cancer (COSMIC, *http://cancer.sanger.ac.uk/cosmic*) was used additionally. Loss of heterozygosity (LOH) was determined as previously described.[52] Somatic data on 20 sLS tumors and 3 *MLH1* hypermethylated tumors (see Supplementary Table 1) has been described previously.[52] These twenty-three patients were also tested for variants in the *POLE/POLD1* exonuclease domain (EDM). Patient sLS-07 and sLS-09 were found to carry *POLE*-EDM variants, previously described to be pathogenic (respectively *POLE* c.2131 G>T and *POLE* c.857 C>G).[52, 53]

Results

Germline targeted next-generation sequencing

Coding variants

Targeted next-generation genomic sequencing of 15 CRC genes was performed in leukocyte DNA of 34 unrelated sLS patients and 11 patients with *MLH1* hypermethylated tumors. The average coverage was 101x. In total, 1,979 nucleotide variants were detected within the targeted region with a MAF < 0.05. All 15 genes were first analyzed for coding variants. After filtering, 52 coding variants remained, of which 16 were synonymous, 33 were missense and 3 were small (in-frame) insertions or duplications. All in-frame insertions/duplications occurred within a stretch of Ala-repeats in exon 1 of *MSH3* and were present in multiple patients and were classified as variants of unknown (clinical) significance (VUS). Eight of the 33 missense variants were found in the coding sequences of *MLH1, MSH2, MSH6* or *PMS2* and were described in the LOVD database as (likely) benign (class 1 or 2), except *MLH1* c.277A>G, which was classified as VUS (class 3). Of the remaining 25 missense variants, 20 were predicted to be benign by at least four out of five protein prediction programs. One of the remaining five variants, *EPCAM* c.50C>A was predicted to be pathogenic by two out of five prediction programs but was described to be benign.[49] The final 4 variants were found in *AXIN1*, *AXIN2, MSH3* and *MUTYH* and were classified as variants of uncertain significance (VUS), or as pathogenic (n=1; *MUTYH* c.1187G>A) (see Table 3).

Table 3: Patients with germline coding VUS or germline pathogenic variants

Patient	Tumor tested	IHC	MSI	Other tumors	Gene	Variant	Protein	Class
sLS-22	CRC54	MLH1[1]	H	-	*MUTYH* *MLH1*	c.1187G>A c.277A>G	p.(G396D) p.(S93G)	5 3
sLS-44	CRC41	MSH2/ MSH6	H	-	*AXIN2*	c.1168A>G	p.(S390G)	3
sLS-56	CRC64	MSH2/ MSH6	H	CRC64	*AXIN1*	c.2566G>A	p.(G856S)	3
sLS-72	CRC73	MSH2/ MSH6	H	Br60, EC68	*MUTYH*	c.1187G>A	p.(G396D)	5
sLS-88	CRC51	MLH1/ PMS2	H	Pr64	*MSH3*	c.982C>T	p.(R328W)	3
sLS-117	CRC20	PMS2	NP	-	*MLH1*	c.1667+1delG	p.(S556ins29)	5

[1]staining of MSH6 and PMS2 was not performed. Tumor tested represents tumor type, followed by the age of onset. Patients presented with colorectal- (CRC), endometrial cancer (EC), breast cancer (Br) and/or prostate cancer (Pr). MSI-status is defined as MSI-High (H) or not performed (NP). Classification of class 3 (VUS) and class 5 (pathogenic) is based on in silico protein predictions, as well as the LOVD Database. All variants were found in sLS patients.

Splice variants
For three variants the splice prediction algorithms predicted deviating splicing efficiencies compared to the wildtype sequences. An *MLH1* variant, in patient sLS-117 (see Table 3), was predicted to abolish the consensus splice site sequence (c.1667+1delG). Functional analysis of patient RNA revealed a mutant MLH1 transcript 87 nucleotides longer than the expected wild type transcript.[46] The 87 nt sequence corresponded to the intron sequence downstream of the splice variant, indicating activation of a cryptic donor splice site 88 nucleotides downstream of the canonical splice site. Translation of the aberrant mRNA leads to the in-frame incorporation of 29 amino acids in the protein-interacting domain of the MLH1 protein. The other variants predicted to affect splicing, a synonymous *APC* c.1959G>A change and the MUTYH c.1187G>A variant described above, only slightly lower the splicing efficiency according to prediction software. The *APC* variant is described in the LOVD database as having 'no known pathogenicity'.

In addition, branch point sequences and polypyrimidine tracts were investigated for possible variants with branch site prediction software SpliceSiteFinder. None of the variants found were predicted to change the existing consensus sequence or to create new branch points.

Promoter variants
Of the 22 promoter variants, 8 were known polymorphisms. The remaining 14 variants were present in single patients of which three were actually present in the specific MMR gene that showed loss of protein expression in the tumor: *MLH1* c.-1019A>C, *MLH1* c.116+730C>T and *MSH2* c.211+550G>C. These variants have not been described before, and functional significance of these variants is unknown according to the INSIGHT classification.[50]

Germline MLH1 methylation
Leukocyte DNA of patients with *MLH1* hypermethylated tumors were also investigated for possible germline methylation. No evidence of germline methylation was found in any of the patients

Somatic variant screening
Tumor DNA from 29 of the 34 sLS tumors was available for somatic DNA analysis. Patient sLS-117 was excluded from somatic variant screening due to the detection of a pathogenic germline *MLH1* variant (*MLH1* c.1667+1delG). Tumor and normal DNAs from the remaining 28 patients were sequenced for somatic MMR variants.

In total, two pathogenic somatic events were detected in eight tumors (29%), including either two variants (n= 3) or one variant together with LOH (n=5) (see Table 4 and Supplementary Table 1). One tumor was found to carry a VUS (predicted to be pathogenic) together with LOH. Nine tumors (32%) revealed one pathogenic somatic variant (n=8), or VUS predicted to be pathogenic (n=1), while no (likely) pathogenic somatic variants were found in seven of the tumors (25%) (see Table 4). Three tumors (11%) could not be analyzed due to poor tumor DNA quality. Seventeen out of the twenty-two somatic MMR variants were nonsense or frame shift variants and were classified as pathogenic (class 5). Of the remaining five somatic variants, two (*MLH1* c.790+1 G>A and *MLH1* c.2059C>T) were previously described to be pathogenic in the LOVD database; two (*MSH6* c.2876 G>A, and *MSH2* c.1166G>A) were not

previously described and were predicted to have a deleterious effect on function by at least four out of five protein prediction programs (See Table 4) and one was an in-frame deletion of three nucleotides (*MSH6* c.3974_3976delAGA), which was classified as having an uncertain effect on function (VUS, class 3).

Patient sLS-22 was previously found to carry a germline *MLH1* VUS (*MLH1* c.277 A>G), and analysis of the tumor DNA revealed an somatic *MLH1* frameshift variant located nearby the germline variant (*MLH1* c.281delT). NGS analysis showed that both variants were located on the same allele. Moreover, the tumor DNA displayed LOH with retention of both variants.

MMR mosaicism

To investigate the possibility of mosaic MMR variants, all cases in which a somatic MMR variant was identified were tested for mosaicism in the corresponding leukocyte DNA. The average coverage of the leukocyte DNA samples was more than a thousand reads per amplicon and no mosaic variant was detected.

Table 4: Patients screened for somatic variants

Patient	Tumor tested	Cohort	MSI	Family History	Gene	Variant	Amino acid alteration	%	Class	Functional annotation
sLS-06	CRC47	MSH2/MSH6	H	FDR	*MSH2*	c.1600_1601delCG	p.(R534*)	19%	5	Nonsense variant
					MSH2	c.2131 C>T	p.(R711*)	20%	5	Nonsense variant
sLS-07	CRC39	MSH6	S	FDR	*MSH6*	c.2876G>A	p.(R959H)	14%	3	VUS, 4/5 pred. path.
					MSH6	LOH				
sLS-22[a]	CRC54	MLH1/PMS2	H	FDR	*MLH1*	c.281delT	p.(S95Lfs*13)	93%	5	Frameshift variant
					MLH1	LOH				
sLS-38[a]	CRC30	MSH2/MSH6	H	FDR	*MSH2*	c.1140delA	p.(L380Ffs*32)	82%	5	Frameshift variant
					MSH2	LOH				
sLS-79[b]	EC57	MSH2/MSH6	H	FDR	*MSH2*	c.1600delC	p.(R534Vfs*9)	20%	5	Frameshift variant
					MSH2	c.2001delT	p.(T668Lfs*17)	20%	5	Frameshift variant
sLS-92	CRC45	MLH1/PMS2	H	FDR	*MLH1*	c.790+1 G>A	p.(E227_S295del)	78%	5	Pathogenic (LOVD)
					MLH1	LOH				
sLS-102[a]	CRC62	MLH1/PMS2	H	FDR	*MLH1*	c.869dupC	p.(F291Ifs*16)	53%	5	Frameshift variant
					MLH1	LOH				
sLS-104	SB47	MSH2/MSH6	H	No	*MSH2*	c.271delG	p.(D91Ifs*83)	92%	5	Frameshift variant
					MSH2	LOH				
sLS-111	EC58	MSH2/MSH6	H	FDR	*MSH2*	c.687delA	p.(I229Mfs*10)	44%	5	Frameshift variant
					MSH2	c.773 T>A	p.(L258*)	41%	5	Nonsense variant
sLS-09	CRC42	MSH6	L	FDR	*MSH2*	c.1166G>A	p.(R389Q)	38%	3	VUS, 4/5 pred. path.
					MSH6	c.2539G>T	p.(E847*)	36%	5	Nonsense variant

Table 4: continued

Patient	Tumor tested	Cohort	MSI	Family History	Gene	Variant	Amino acid alteration	%	Class	Functional annotation
sLS-55	EC47	MSH2/MSH6	H	FDR	*MSH6* *MSH2*	c.3971delAGA LOH	p.(L1325del)	24%	3	VUS, in-frame deletion
sLS-56	CRC64	MSH2/MSH6	H	NA	*MSH2*	c.1710T>A	p.(Y570*)	19%	5	Nonsense variant
sLS-58	CRC39	MLH1/PMS2	L	FDR	*MLH1*	c.790+1 G>A	p.(E227_S295del)	28%	5	Pathogenic (LOVD)
sLS-64	CRC48	MLH1/PMS2	H	FDR	*MLH1*	c.2059C>T	p.(R687W)	28%	5	Pathogenic (LOVD)
sLS-72	CRC73	MSH2/MSH6	H	No	*MSH2*	c.1576dupA	p.(T526Nfs*3)	29%	5	Frameshift variant
sLS-77	CRC45	MSH2/MSH6	H	No	*MSH2*	c.2470C>T	p.(Q824*)	38%	5	Nonsense variant
sLS-101	EC55	PMS2	H	FDR	*PMS2*	c.1687C>T	p.(R563*)	30%	5	Nonsense variant
sLS-127[a]	CRC45	MSH2/MSH6	H	NA	*MSH2*	c.2527delT	p.(C843Vfs*49)	37%	5	Frameshift variant
sLS-17	CRC39	MSH6	S	FDR	-	-	-	-		
sLS-20	CRC55	MSH2/MSH6	S	FDR	-	-	-	-		
sLS-43	CRC74	MSH2/MSH6	H	FDR	-	-	-	-		
sLS-44	CRC41	MSH2/MSH6	H	FDR	-	-	-	-		
sLS-62	CRC35	MSH2/MSH6	L	FDR	-	-	-	-		
sLS-82	CRC69	MLH1/PMS2	H	FDR	-	-	-	-		
sLS-120	CRC57	MLH1/PMS2	S	FDR	-	-	-	-		

[a]Somatic screening was performed by the Erasmus MC, Rotterdam. [b]Somatic screening was performed by the Radboud UMC, Nijmegen. Patients presented with colorectal cancer (CRC), endometrium cancer (EC) or sebaceous gland cancer (SB). Tumor tested shows tumor type followed by patients age of onset. Cohort gives an indication of IHC results. Detailed IHC results are shown in Supplementary Table 1. MSI-status is defined as MSI-High (H), MSI-Low (L) or MSS (S). Family history is defined as first-degree relatives with LSAT (FDR), no LSAT within the family (No), or family history not available (NA). Stopcodons are indicated with an asterisk (*). % shows the percentage of variant reads. All variants are validated with Sanger sequencing. Two variants are predicted to be pathogenic by at least 4 out of 5 of the following programs: Align GVGD, SIFT, MutationTaster, Polyphen and MutationAssessor.

Discussion

In this study we carried out an extensive sequencing analysis of the genomic regions of the four MMR and 11 other CRC susceptibility genes, including *MUTYH, EPCAM* and *MSH3*. We anticipated that this type of broad analysis, well beyond the boundaries of conventional mutation screening, would identify variants previously missed by standard techniques or would identify variants in genes other than the previously diagnostically tested MMR genes. As our patient cohort consisted mainly of cases with a first-degree relative with a LS-associated tumor, cancer susceptibility due to an underlying germline defect in these families seemed the most plausible explanation.

The approach used, Whole Gene Capture, yielded an average sequence depth up to 5-fold greater than whole exome sequencing, with sufficient depth to allow detection of mosaic and *de novo* variants. In total, 1,979 initial variants were detected. Many variants were classified as of uncertain significance and follow-up studies might reveal novel functional effects. After filtering by function and predicted pathogenicity, two likely pathogenic variants remained. An *MLH1* splice site variant, c.1667+1delG, was found in patient sLS-117, who was diagnosed with CRC at age 20. Patient sLS-117 presented with solitary PMS2 protein deficiency in the tumor and only *PMS2* had been previously screened with conventional mutation screening. IHC showed solitary PMS2 loss of expression, since the *MLH1* frame shift variant leads to a 29 amino acid insertion in the protein-protein interacting domain, resulting in an MLH1 transcript which is unable to bind PMS2. Analysis of family members demonstrated the variant in the patient's affected mother (CRC at age 44), whereas the patient's unaffected daughter tested negative for the variant.

The second pathogenic variant, *MUTYH* p.Gly396Asp, was present in a heterozygous state in two patients (patient sLS-22 and sLS-72, see Table 3). Monoallelic variants in *MUTYH* are present in 2% of the general population and are not found at increased frequencies in sLS patients.[54, 55] The role of monoallelic *MUTYH* variants is still under debate, and while some studies have indicated an increased cancer risk for carriers of a single *MUTYH* variant, the p.Gly396Asp variant alone is unlikely to be the explanation for the MSI-H and/or IHC status of the tumors in our patients[56, 57]. Moreover, both patients were found to have (likely) pathogenic somatic MMR variants (Supplementary Table 1) explaining the MMR-deficient phenotype.

In addition to the 34 sLS patients, eleven colorectal cancer patients with somatic *MLH1* hypermethylation and a family history suspected of LS were analyzed for possible underlying germline defects. *MLH1* promoter methylation in Lynch Syndrome patients has been described before, either co-occuring with a pathogenic germline *MSH6* variant in a patient with a urothelial carcinoma at age 70[30] , in a patient with a pathogenic germline *MLH1* variant in a CRC at the age of 59[31] or with pathogenic germline *MSH6* variant in a patient with multiple primary cancers, from the age of 56[32]. Another study describes *MLH1* hypermethylation in three LS-tumors, hypothesizing methylation is the second hit inactivating the wildtype allele.[58] These studies indicate that *MLH1* hypermethylation does not always exclude a diagnosis of LS. In our study we have not found support for above findings. Moreover, three families with germline *MLH1* hypermethylation in multiple affected family members have

been reported[42-44], indicating epigenetic inheritance of constitutional epimutations with a risk of transgenerational inheritance. All eleven patients with *MLH1* hypermethylated tumors in our cohort were tested for germline methylation, but no germline methylation was found.

Although this intensive study enabled the detection of variants within the intronic regions, UTRs and regions up- and downstream of the target genes, some limitations have to be considered. While the average coverage of the coding regions is over 95%, the overall average coverage is 43% (See Table 2). The lower overall coverage is due to the method used in which no baits were designed for the repetitive sequences such as the Alu- and Line-repeats within the introns. Therefore, missed intronic variants in these regions cannot be excluded. Moreover, we cannot exclude the possibility of large genomic rearrangements within the genes tested, which is a limitation of the method used in this study.

Screening of tumor DNA from 28 sLS patients for somatic variants revealed almost a third with two somatic variants (n=3) or a combination of a somatic variant and LOH (n=6). The frequency of biallelic inactivation in our cohort is lower than previously described[23-25], and might be due to differences in patient selection in the different study cohorts. While previous studies screened sLS patients irrespective of family history, the majority of patients in the present cohort had first-degree relatives with LS-associated tumors (see Supplementary Table 1) and eight families even fulfilled the Amsterdam II criteria. However, while biallelic somatic events may explain the MMR-deficiency of the tumor of the index patient, they cannot explain a family history of CRC. Ideally, a second affected family member in these families should be tested to see whether these patients can also be explained by somatic MMR inactivation. Unfortunately, no DNA could be obtained from affected family members. An underlying pathogenic germline gene variant outside these 15 CRC-susceptibility genes cannot be excluded in these families.

Besides somatic MMR variants, two sLS patients (sLS-07 and sLS-09) were recently found to carry somatic hotspot *POLE*-EDM variants (see Supplementary Table 1).[52] As *POLE/POLD1* EDM pathogenic variants give rise to ultramutated tumors, the somatic MMR variants apparently represent a second hit. Screening for germline or somatic *POLE/POLD1* –EDM variants, but also for variants in other genes recently described to be mutated in sLS CRCs such as *BRCA1/BRCA2*, *ATM* and *CHEK2*, may explain some of these sLS patients.[59, 60]

In conclusion, sequencing of the entire genomic region of 15 CRC susceptibility genes in 34 unrelated sLS patients and 11 patients with *MLH1* hypermethylated tumors, together with assessment of somatic variants, provides a broad impression of possible genetic causes of tumor formation in MSI-H and/or MMR-deficient tumors. No likely pathogenic MMR gene variants or germline *MLH1* hypermethylation were found that explained the familial aggregation of cancer susceptibility in any of the families with *MLH1* hypermethylated tumors. With the MMR-deficiency of around one-third of the 34 sLS tumors now explained, MMR-deficiency in two-thirds of sLS tumors remains genetically unaccounted for. A logical next step is whole exome sequencing (WES) or whole genome sequencing (WGS) to further elucidate the causative genetic defect(s) in the remaining patients.

References

1. Colas C, Coulet F, Svrcek M, et al. Lynch or not Lynch? Is that always a question? Adv Cancer Res 2012;113:121-66.
2. Ligtenberg MJ, Kuiper RP, Chan TL, et al. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. Nat Genet 2009;41:112-7.
3. Ligtenberg MJ, Kuiper RP, Geurts van Kessel A, et al. EPCAM deletion carriers constitute a unique subgroup of Lynch syndrome patients. Fam Cancer 2013;12:169-74.
4. Aaltonen LA, Peltomaki P, Mecklin JP, et al. Replication errors in benign and malignant tumors from hereditary nonpolyposis colorectal cancer patients. Cancer Res 1994;54:1645-8.
5. Vasen HF, Mecklin JP, Khan PM, et al. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Dis Colon Rectum 1991;34:424-5.
6. Umar A, Boland CR, Terdiman JP, et al. Revised Bethesda Guidelines for Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndrome) and Microsatellite Instability. JNCI Journal of the National Cancer Institute 2004;96:261-268.
7. Lindor NM, Rabe K, Petersen GM, et al. Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency - Familial colorectal cancer type X. Jama-Journal of the American Medical Association 2005;293:1979-1985.
8. Jenkins MA, Baglietto L, Dowty JG, et al. Cancer risks for mismatch repair gene mutation carriers: a population-based early onset case-family study. Clin Gastroenterol Hepatol 2006;4:489-98.
9. Quehenberger F, Vasen HF, van Houwelingen HC. Risk of colorectal and endometrial cancer for carriers of mutations of the hMLH1 and hMSH2 gene: correction for ascertainment. J Med Genet 2005;42:491-6.
10. Aarnio M, Sankila R, Pukkala E, et al. Cancer risk in mutation carriers of DNA-mismatch-repair genes. Int J Cancer 1999;81:214-8.
11. Stoffel E, Mukherjee B, Raymond VM, et al. Calculation of risk of colorectal and endometrial cancer among patients with Lynch syndrome. Gastroenterology 2009;137:1621-7.
12. de la Chapelle A. The incidence of Lynch syndrome. Fam Cancer 2005;4:233-7.
13. Dowty JG, Win AK, Buchanan DD, et al. Cancer risks for MLH1 and MSH2 mutation carriers. Hum Mutat 2013;34:490-7.
14. Baglietto L, Lindor NM, Dowty JG, et al. Risks of Lynch syndrome cancers for MSH6 mutation carriers. J Natl Cancer Inst 2010;102:193-201.
15. Senter L, Clendenning M, Sotamaa K, et al. The clinical phenotype of Lynch syndrome due to germ-line PMS2 mutations. Gastroenterology 2008;135:419-28.
16. Bauer CM, Ray AM, Halstead-Nussloch BA, et al. Hereditary prostate cancer as a feature of Lynch syndrome. Fam Cancer 2011;10:37-42.
17. Win AK, Lindor NM, Jenkins MA. Risk of breast cancer in Lynch syndrome: a systematic review. Breast Cancer Res 2013;15:R27.
18. Haraldsdottir S, Hampel H, Wei L, et al. Prostate cancer incidence in males with Lynch syndrome. Genet Med 2014;16:553-7.
19. Vasen HF, Abdirahman M, Brohet R, et al. One to 2-year surveillance intervals reduce risk of colorectal cancer in families with Lynch syndrome. Gastroenterology 2010;138:2300-6.
20. Lindor NM, Petersen GM, Hadley DW, et al. Recommendations for the care of individuals with an inherited predisposition to Lynch syndrome: a systematic review. JAMA 2006;296:1507-17.
21. Buchanan DD, Rosty C, Clendenning M, et al. Clinical problems of colorectal cancer and endometrial cancer cases with unknown cause of tumor mismatch repair deficiency (suspected Lynch syndrome). Appl Clin Genet 2014;7:183-93.
22. Rodriguez-Soler M, Perez-Carbonell L, Guarinos C, et al. Risk of cancer in cases of suspected lynch syndrome without germline mutation. Gastroenterology 2013;144:926-932 e1; quiz e13-4.
23. Mensenkamp AR, Vogelaar IP, van Zelst-Stams WA, et al. Somatic mutations in MLH1 and MSH2 are a frequent cause of mismatch-repair deficiency in Lynch syndrome-like tumors. Gastroenterology 2014;146:643-646 e8.
24. Haraldsdottir S, Hampel H, Tomsic J, et al. Colon and endometrial cancers with mismatch repair deficiency can arise from somatic, rather than germline, mutations. Gastroenterology 2014;147:1308-1316 e1.
25. Geurts-Giele WR, Leenen CH, Dubbink HJ, et al. Somatic aberrations of mismatch repair genes as a cause of microsatellite-unstable cancers. J Pathol 2014;234:548-59.
26. Pinol V, Castells A, Andreu M, et al. Accuracy of revised Bethesda guidelines, microsatellite instability, and immunohistochemistry for the identification of patients with hereditary nonpolyposis colorectal cancer. JAMA 2005;293:1986-94.

27. Young J, Simms LA, Biden KG, et al. Features of colorectal cancers with high-level microsatellite instability occurring in familial and sporadic settings: parallel pathways of tumorigenesis. Am J Pathol 2001;159:2107-16.
28. Crucianelli F, Tricarico R, Turchetti D, et al. MLH1 constitutional and somatic methylation in patients with MLH1 negative tumors fulfilling the revised Bethesda criteria. Epigenetics 2014;9:1431-8.
29. Lynch HT, Lynch PM, Lanspa SJ, et al. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. Clin Genet 2009;76:1-18.
30. Overbeek LI, Kets CM, Hebeda KM, et al. Patients with an unexplained microsatellite instable tumour have a low risk of familial cancer. Br J Cancer 2007;96:1605-12.
31. Rahner N, Friedrichs N, Steinke V, et al. Coexisting somatic promoter hypermethylation and pathogenic MLH1 germline mutation in Lynch syndrome. J Pathol 2008;214:10-6.
32. Raymond VM, Morris AM, Hafez KS, et al. MLH1 promotor hypermethylation does not rule out a diagnosis of Lynch syndrome: a case report. Fam Cancer 2015;14:77-80.
33. van Roon EH, van Puijenbroek M, Middeldorp A, et al. Early onset MSI-H colon cancer with MLH1 promoter methylation, is there a genetic predisposition? BMC Cancer 2010;10:180.
34. Goel A, Nguyen TP, Leung HC, et al. De novo constitutional MLH1 epimutations confer early-onset colorectal cancer in two new sporadic Lynch syndrome cases, with derivation of the epimutation on the paternal allele in one. Int J Cancer 2011;128:869-78.
35. Niessen RC, Hofstra RM, Westers H, et al. Germline hypermethylation of MLH1 and EPCAM deletions are a frequent cause of Lynch syndrome. Genes Chromosomes Cancer 2009;48:737-44.
36. Suter CM, Martin DI, Ward RL. Germline epimutation of MLH1 in individuals with multiple cancers. Nat Genet 2004;36:497-501.
37. Valle L, Carbonell P, Fernandez V, et al. MLH1 germline epimutations in selected patients with early-onset non-polyposis colorectal cancer. Clin Genet 2007;71:232-7.
38. Ward RL, Dobbins T, Lindor NM, et al. Identification of constitutional MLH1 epimutations and promoter variants in colorectal cancer patients from the Colon Cancer Family Registry. Genet Med 2013;15:25-35.
39. Miyakura Y, Sugano K, Akasu T, et al. Extensive but hemiallelic methylation of the hMLH1 promoter region in early-onset sporadic colon cancers with microsatellite instability. Clin Gastroenterol Hepatol 2004;2:147-56.
40. Castillejo A, Hernandez-Illan E, Rodriguez-Soler M, et al. Prevalence of MLH1 constitutional epimutations as a cause of Lynch syndrome in unselected versus selected consecutive series of patients with colorectal cancer. J Med Genet 2015;52:498-502.
41. Hitchins MP. Finding the needle in a haystack: identification of cases of Lynch syndrome with MLH1 epimutation. Fam Cancer 2016 Jul;15(3):413-22
42. Crepin M, Dieu MC, Lejeune S, et al. Evidence of constitutional MLH1 epimutation associated to transgenerational inheritance of cancer susceptibility. Hum Mutat 2012;33:180-8.
43. Hitchins MP, Wong JJ, Suthers G, et al. Inheritance of a cancer-associated MLH1 germ-line epimutation. N Engl J Med 2007;356:697-705.
44. Morak M, Schackert HK, Rahner N, et al. Further evidence for heritability of an epimutation in one of 12 cases with MLH1 promoter methylation in blood cells clinically displaying HNPCC. Eur J Hum Genet 2008;16:804-11.
45. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
46. van der Klift HM, Jansen AM, van der Steenstraten N, et al. Splicing analysis for exonic and intronic mismatch repair gene variants associated with Lynch syndrome confirms high concordance between minigene assays and patient RNA analyses. Mol Genet Genomic Med 2015;3:327-45.
47. Vreeswijk MP, Kraan JN, van der Klift HM, et al. Intronic variants in BRCA1 and BRCA2 that affect RNA splicing can be reliably selected by splice-site prediction programs. Hum Mutat 2009;30:107-14.
48. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res 2011;39:e118.
49. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res 2016;44:D862-8.
50. Thompson BA, Spurdle AB, Plazzer JP, et al. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nat Genet 2014;46:107-15.
51. Deng G, Peng E, Gum J, et al. Methylation of hMLH1 promoter correlates with the gene silencing with a region-specific manner in colorectal cancer. Br J Cancer 2002;86:574-9.
52. Jansen AML, van Wezel T, van den Akker BEWM, et al. Combined mismatch repair and POLE/POLD1 defects explain unresolved suspected Lynch syndrome cancers. European Journal of Human Genetics 2016 Jul;24(7):1089-92

53. Shinbrot E, Henninger EE, Weinhold N, et al. Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. Genome Res 2014;24:1740-50.
54. Tenesa A, Campbell H, Barnetson R, et al. Association of MUTYH and colorectal cancer. Br J Cancer 2006;95:239-42.
55. Castillejo A, Vargas G, Castillejo MI, et al. Prevalence of germline MUTYH mutations among Lynch-like syndrome patients. Eur J Cancer 2014;50:2241-50.
56. Khalaf R, Jones C, Strutt W, et al. Colorectal cancer in a monoallelic MYH mutation carrier. J Gastrointest Surg 2013;17:1500-2.
57. Win AK, Cleary SP, Dowty JG, et al. Cancer risks for monoallelic MUTYH mutation carriers with a family history of colorectal cancer. Int J Cancer 2011;129:2256-62.
58. Herman JG, Umar A, Polyak K, et al. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. Proc Natl Acad Sci U S A 1998;95:6870-5.
59. Elsayed FA, Kets CM, Ruano D, et al. Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer. Eur J Hum Genet 2014 Aug;23(8):1080-4.
60. Yurgelun MB, Allen B, Kaldate RR, et al. Identification of a Variety of Mutations in Cancer Predisposition Genes in Patients with Suspected Lynch Syndrome. Gastroenterology 2015 Sep; 149(3):604-1.e20.

Supplementary Table 1: Overview of germline and somatic variants

Patient	Gender	Tested tumor	IHC MLH1	IHC MSH2	IHC MSH6	IHC PMS2	MSI	Other tumors	Family History	MLH1 Methylation	BC/ACII
					sLS cohort						
sLS-04	Female	CRC42	-	+	+	NT	High	-	FDR	No	BC
sLS-06	Male	CRC47	INC	-	NT	NT	High	-	FDR	NA	ACII
sLS-07	Male	CRC39	INC	INC	-	NT	Stable	-	FDR	NA	BC
sLS-09	Male	CRC42	+	+	-	NT	Low	-	FDR	NA	BC
sLS-17	Female	CRC39	+	+	-	NT	Stable	Br47	FDR	NA	BC
sLS-20	Female	CRC55	+	INC	INC	NT	Stable	-	FDR	NA	ACII
sLS-22 (a)	Male	CRC54	-	+	NT	NT	High	Pol	FDR	No	BC
sLS-38 (a)	Male	CRC30	+	-	-	NT	High	Lymph15, Pol30	FDR	NA	ACII
sLS-43	Male	CRC74	INC	-	-	NT	High	CRC74	FDR	NA	BC
sLS-44	Female	CRC41	+	-	-	NT	High	-	FDR	NA	ACII
sLS-55	Female	EC47	+	-	-	+	High	OC48	FDR	NA	BC
sLS-56	Male	CRC64	+	-	-	+	High	CRC64	NA	NA	BC
sLS-58	Male	CRC39	-	+	+	-	Low	-	FDR	No	ACII
sLS-62	Male	CRC35	+	-	-	NT	Low	-	FDR	NA	BC
sLS-64	Male	CRC48	-	+	+	-	High	-	FDR	No	BC
sLS-72	Female	CRC73	+	-	-	+	High	Br60, EC68	No	NA	BC
sLS-74	Female	CRC30	+	-	-	+	Low	-	NA	No	BC
sLS-77	Female	CRC45	+	-	-	+	High	Pol	No	NA	BC
sLS-79 (b)	Female	EC57	+	-	-	NT	High	-	FDR	NA	BC
sLS-82	Female	CRC69	-	+	+	-	High	-	FDR	No	BC
sLS-88	Male	CRC51	-	+	+	-	High	Pr64	FDR	No	BC
sLS-92	Female	CRC45	-	+	INC	-	High	-	FDR	No	ACII
sLS-93	Male	CRC47	-	NT	NT	NT	High	-	FDR	NA	ACII

Supplementary Table 1: Overview of germline and somatic variants; continued

Patient	Gender	Tested tumor	IHC MLH1	IHC MSH2	IHC MSH6	IHC PMS2	MSI	Other tumors	Family History	MLH1 Methylation	BC/ACII
sLS-94	Male	CRC55	+	-	-	+	High	-	FDR	NA	ACII
sLS-95	Female	CRC47	+	+	+	-	High	EC68	FDR	NA	BC
sLS-99	Male	CRC49	+	-	-	+	High	-	FDR	NA	BC
sLS-101	Female	EC55	+	+	+	-	High	-	FDR	No	BC
sLS-102 (a)	Male	CRC62	-	NT	NT	-	High	-	FDR	No	BC
sLS-104	Male	SB47	+	-	-	+	High	-	No	NA	No
sLS-111	Female	EC58	+	-	-	+	High	-	FDR	NA	ACII
sLS-117	Female	CRC20	+	+	+	-	Unknown	-	FDR	NA	ACII
sLS-119	Female	EC50	NT	NT	-	NT	High	CRC64	FDR	NA	BC
sLS-120	Male	CRC57	-	+	+	-	Stable	-	FDR	No	BC
sLS-127 (a)	Female	CRC45	NT	-	-	NT	High	-	NA	NA	BC
MLH1 hypermethylated cohort											
sLS-33	Female	CRC73	INC	+	+	-	High	-	FDR	Yes	BC
sLS-47	Female	EC69	-	+	+	-	Low	Br70, Br76	FDR	Yes	No
sLS-48	Male	CRC52	-	+	+	-	Unknown	-	FDR	Yes	ACII
sLS-51	Female	CRC69	-	+	+	-	High	-	FDR	Yes	ACII
sLS-52	Female	EC67	-	+	+	-	High	-	FDR	Yes	BC
sLS-61	Female	CRC49	-	+	+	-	High	CRC51	FDR	Yes	BC
sLS-68	Female	CRC63	-	+	+	NT	High	EC68	FDR	Yes	BC
sLS-81	Male	CRC65	-	NT	NT	NT	High	Pol	FDR	Yes	ACII
sLS-103	Female	CRC66	-	+	+	-	High	-	FDR	Yes	BC
sLS-107	Female	CRC75	-	+	+	-	High	OC58, EC60, Br67, Br74	FDR	Yes	BC
sLS-110	Male	CRC62	-	+	+	-	High	Pol62	FDR	Yes	BC

Supplementary Table 1: continued

Patient	Germline variant	HET/ HOM	Class	Somatic variants	%	Conclusion	LOH MLH1	LOH MSH2	LOH MSH6	LOH PMS2	In Jansen et al, EJHG
sLS cohort											
sLS-04	-	-	-	-	-	Low tumor DNA quality	NP	NP	NP	NP	no
sLS-06	-	-	-	MSH2 c.1600_1601delCG MSH2 c.2131 C>T	19% 20%	2 somatic MMR variants	-	NP	-	-	yes
sLS-07	-	-	-	POLE c.1231 G>T MSH6 c.2876G>A MSH6 LOH	44% 14% -	somatic POLE variant with 2 somatic MMR variants	-	NP	3/3	-	yes
sLS-09	-	-	-	POLE c.857C>G MSH6 c.2539G>T MSH2 c.1166G>A	38% 36% 38%	somatic POLE variant with 1 somatic MMR variant	-	-	-	-	yes
sLS-17	-	-	-	-	-	No variants	-	-	-	-	yes
sLS-20	-	-	-	-	-	No variants	-	-	-	-	yes
sLS-22 (a)	MUTYH c.1187G>A MLH1 c.277A>G	HET HET	5 3	MLH1 c.281delT MLH1 LOH	93% -	2 somatic MMR variants					no
sLS-38 (a)	-	-	-	MSH2 c.1140delA MSH2 LOH	82% -	2 somatic MMR variants					no
sLS-43	-	-	-	-	-	No variants	-	-	-	-	yes
sLS-44	-	-	-	-	-	No variants	NP	NP	NP	NP	yes
sLS-55	-	-	-	MSH2 LOH MSH6 c.3974_3976del	- 24%	1 somatic MMR variant	-	1/1	NP	-	yes
sLS-56	AXIN c.2566G>A	HET	3	MSH2 c.1710T>A	19%	1 somatic MMR variant	-	-	-	-	yes
sLS-58	-	-	-	MLH1 c.790+1 G>A	28%	1 somatic MMR variant	NP	-	NP	NP	yes
sLS-62	-	-	-	-	-	No variants	-	-	-	-	yes
sLS-64	-	-	-	MLH1 c.2059C>T	28%	1 somatic MMR variant	-	NP	-	-	yes

Supplementary Table 1: Overview of germline and somatic variants; continued

Patient	Germline variant	HET/ HOM	Class	Somatic variants	%	Conclusion	LOH MLH1	LOH MSH2	LOH MSH6	LOH PMS2	In Jansen et al, EJHG
sLS-72	MUTYH c.1187G>A	HET	5	MSH2 c.1576dupA	29%	1 somatic MMR variant	NP	NP	NP	NP	yes
sLS-74	-	-	-	-	-	Low tumor DNA quality	NP	NP	NP	NP	no
sLS-77	-	-	-	MSH2 c.2470C>T	38%	1 somatic MMR variant	-	-	-	-	yes
sLS-79 (b)	-	-	-	MSH2 c.1600delC MSH2 c.2001delT	20% 20%	2 somatic MMR variants					no
sLS-82	-	-		-	-	No variants	-	NP	-	-	yes
sLS-88	EPCAM c.50C>A MSH3 c.982C>T	HET HET	3 3	-	-	No tumor DNA available	NP	NP	NP	NP	no
sLS-92	-	-	-	MLH1 c.790+1 G>A MLH1 LOH	78% -	2 somatic MMR variants	1/1	-	NP	-	yes
sLS-93	-	-	-	-	-	No tumor DNA available	NP	NP	NP	NP	no
sLS-94			-			Low tumor DNA quality	NP	NP	NP	NP	no
sLS-95	-	-	-	-	-	No tumor DNA available	NP	NP	NP	NP	no
sLS-99		-	3	-	-	No tumor DNA available	NP	NP	NP	NP	no
sLS-101	-	-	-	PMS2 c.1687C>T	30%	1 somatic MMR variant	NP	-	-	NP	yes
sLS-102 (a)	-	-	-	MLH1 c.869dupC MLH1 LOH	53% -	2 somatic MMR variants					no
sLS-104	-	-	-	MSH2 c.271delG MSH2 LOH	92%	2 somatic MMR variants	NP	1/1	NP	-	yes
sLS-111	-	-	-	MSH2 c.687delA MSH2 c.773 T>A	44% 41%	2 somatic MMR variants	NP	-	-	NP	yes

Supplementary Table 1: continued

Patient	Germline variant	HET/ HOM	Class	Somatic variants	%	Conclusion	LOH MLH1	LOH MSH2	LOH MSH6	LOH PMS2	In Jansen et al, EJHG
sLS-117	MLH1 c.1667+1delG	HET	5	-	-	LS - Germline MMR variant	NP	NP	NP	NP	no
sLS-119	-	-	-	-	-	No tumor DNA available	NP	NP	NP	NP	no
sLS-120	-	-	-	-	-	No variants	-	-	-	-	yes
sLS-127 (a)		-	-	MSH2 c.2527delT	37%	1 somatic MMR variant					no
MLH1 hypermethylated cohort											
sLS-33	-	-	-	NA	-	MLH1 methylation	NP	NP	NP	NP	no
sLS-47	-	-	-	NA	-	MLH1 methylation	NP	NP	NP	NP	no
sLS-48	-	-	-	NA	-	MLH1 methylation	NP	NP	NP	NP	no
sLS-51	-	-	-	NA	-	MLH1 methylation	NP	NP	NP	NP	yes
sLS-52	-	-	-	NA	-	MLH1 methylation	NP	NP	NP	NP	yes
sLS-61	-	-	-	NA	-	MLH1 methylation	NP	NP	NP	NP	yes
sLS-68	-	-	-	NA	-	MLH1 methylation	NP	NP	NP	NP	no
sLS-81	-	-	-	NA	-	MLH1 methylation	NP	NP	NP	NP	no
sLS-103	-	-	-	NA	-	MLH1 methylation	NP	NP	NP	NP	no
sLS-107	-	-	-	NA	-	MLH1 methylation	NP	NP	NP	NP	no
sLS-110	-	-	-	NA	-	MLH1 methylation	NP	NP	NP	NP	no

(a) somatic screening was performed by Erasmus MC, Rotterdam, (b) somatic screening was performed by Radboud, Nijmegen. Overview of germline and somatic variants found in 34 sLS patients and 11 patients with a MLH1 hypermethylated tumor. IHC is indicated as negative (-), positive (+), inconclusive (INC) or not tested (NT). Patients presented with colorectal cancer (CRC), endometrial cancer (EC), breast cancer (Br), Non-hodgkins lymphoma (lymph), prostate cancer (Pr), sebaceous gland cancer (SB) and/or polyps (pol). Tumor tested represents tumor type, followed by the age of onset. Family history is defined as first-degree relative with a LSAT (FDR), no LSAT in the family (No) or family history not available (NA).Germline variants are defined as heterozygous (HET) or homozygous (HOM). % shows the percentage of variant reads. MLH1 hypermethylation status is defined as yes, no or not applicable (NA). Patients fulfilled revised bethesda criteria (BC), Amsterdam II criteria (ACII) or neither (No). LOH was called when all heterozygous SNPs showed significant imbalance. LOH classified as no LOH (-), LOH, with the number of informative SNPs (#/#) or LOH analysis was not possible (NP).

Chapter 3

Combined mismatch repair and *POLE/POLD1* defects explain unresolved suspected Lynch Syndrome cancers

AML Jansen[1,2], T. van Wezel[1], BEWM van den Akker[1], M Ventayol Garcia[1], D Ruano[1], CMJ Tops[3], A Wagner[4], TGW Letteboer[5], EB Gómez-García[6], P Devilee[2], JT Wijnen[2,3], FJ Hes[3], H Morreau[1]

EJHG, 2016; 24 (7):1089-92

Departments of [1]Pathology, [2]Human Genetics and [3]Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands.

[4]Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands

[5]Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands

[6]Department of Clinical Genetics, University Hospital Maastricht, Maastricht, The Netherlands

Abstract

Many suspected Lynch Syndrome (sLS) patients who lack mismatch repair (MMR) germline gene variants and *MLH1* or *MSH2* hypermethylation are currently explained by somatic MMR gene variants or, occasionally, by germline *POLE* variants. To further investigate unexplained sLS patients, we analyzed leukocyte- and tumor DNA of a cohort of 62 sLS patients using gene panel sequencing including the *POLE, POLD1* and MMR genes. Forty tumors showed either one, two or more somatic MMR variants predicted to affect function. Nine sLS-tumors showed a likely ultramutated phenotype and were found to carry germline- (n=2) or somatic variants (n=7) in the *POLE/POLD1* exonuclease domain (EDM). Six of these *POLE/POLD1* mutated tumors also carried somatic MMR variants. Our findings suggest that faulty proofreading may result in loss of MMR and thereby in microsatellite instability.

Keywords: POLE; POLD1; suspected Lynch Syndrome; Colorectal Cancer; Mismatch Repair Deficiency

Introduction

Inactivation of the mismatch repair (MMR) genes *MLH1*, *MSH2*, *MSH6* and *PMS2* causes Lynch Syndrome (LS), an autosomal dominant predisposition for colorectal and endometrial cancer.[1] Inactivation of the mismatch repair pathway can also occur sporadically, through somatic *MLH1* methylation[1] or by acquired bi-allelic somatic inactivation (variant affecting function or loss of heterozygosity (LOH)) of the MMR genes.[2,3] Inaccurate DNA repair leads to a high frequency of somatic variants, with loss of MMR leading to 'hypermutated' tumors with 10-100 variants/Mb.[4] LS tumors are characterized by microsatellite instability (MSI) and immunohistochemical loss of expression of MMR proteins.[1] However, germline variants affecting function cannot be detected in up to 59% of patients displaying MSI and/or loss of MMR, referred to as 'suspected LS' (sLS).[5]

Recently, germline and somatic variants in the exonuclease domains (EDM) of DNA polymerase ε (*POLE*) and polymerase δ (*POLD1*) were described.[6-14] These *POLE*/*POLD1* variants affect proofreading function and lead to an ultramutated phenotype with a variant incidence exceeding 100 variants/Mb**.** Germline *POLE*-EDM variants can result in a LS phenotype and microsatellite unstable CRCs.[6,15] The exact role of somatic *POLE*/*POLD1* variants in tumors with high microsatellite instability (MSI-H) remains unclear.

The aim of our study was to identify the underlying genetic cause of disease in a cohort of 64 suspected LS cases - selected on the basis of MSI, loss of MMR, young onset and often a family history for LS - by screening the MMR, *POLE* and *POLD1* genes in both leukocyte and tumor DNA.

Subjects and Methods

This study included 64 patients with Lynch-associated tumors recruited in four academic centers in the Netherlands between 1990 and 2014: Leiden University Medical Center (n=36), Maastricht University Medical Center (n=11), Erasmus Medical Center (n=9) and University Medical Center Utrecht (n=7). Demographic and clinical data, as well as informed consent, were obtained at the time of diagnosis. Patients were selected based on loss of MMR (as indicated by immunohistochemical staining) and/or microsatellite instability (MSI). Unexplained tumors with low microsatellite instability, or tumors with inconclusive IHC results were also included in this study (See Supplementary Table 1 and Supplementary Methods).

Fifty-six (90%) patients fulfilled Bethesda criteria[16], and families of twenty-two (34%) patients also fulfilled Amsterdam II criteria.[17] Patients were previously screened in a diagnostic setting for germline MMR variants. While 58 patients showed no disease causing germline variants, six patients were found to have a germline variant of unknown significance (VUS). Of the total cohort, 75% of patients presented with colorectal cancer (CRC, n=48), 14% with endometrial cancer (EC, n=9), and 11% with another LS-associated tumor (See Supplementary Table 1).

The average age of onset was 52.1 years. Two patients were excluded from the analysis due to poor DNA quality. Of the remaining 62 tumors, tumor- and leukocyte DNA was sequenced for variants in the exonic regions of *MLH1*, *MSH2*, *MSH6*, *PMS2, POLE* and *POLD1* using the Ion PGM™ System (Life Technologies, Carlsbad, CA). Raw data analysis, alignments, and variant calling was carried out using the default parameters in Torrent Suite v4.0 (see Supplementary information for detailed description). Variants were functionally annotated using ANNOVAR.[18]

The full dataset was filtered and prioritized by variant frequency (>10%) and coverage (>50x). Interesting variants under the 10% were manually curated. *In silico* prediction programs were used to predict pathogenicity (see Supplementary Methods). All variants (likely) affecting function, including two variants with a 9% variant frequency, were validated with Sanger sequencing. For all *PMS2* variants, *PMS2* specific primers were created, to validate that the variant is present in *PMS2* and not in a *PMS2* pseudogene.

Loss of heterozygosity (LOH) was determined for every heterozygous SNP by comparing the ratio of allele A to allele B in leukocyte and tumor DNA samples. Furthermore, for every heterozygous SNP the allelic imbalance factor (AIF) [19] was calculated and a Fisher exact test was performed to determine whether the difference between normal and tumor is significant. If all heterozygous SNPs of one gene showed loss of heterozygosity with an AIF>2 and Fisher exact p-value <0.05 LOH was called. (Supplementary Table 1).

Results

Six patients with a germline MMR VUS (class 3) were included in this study (Supplementary Table 1). In all cases, the variant is detected with NGS, in leukocyte and tumor DNA. During the course of the study, three of these germline variants were reclassified as class 4 or 5 ((probably) affects function) by the International Society for Gastrointestinal Hereditary Tumors Incorporated (InSiGHT). Three of these patients displayed a second somatic MMR variant predicted to affect function. One tumor displayed LOH (Supplementary Table 1). Six cases with somatic *MLH1* hypermethylation fulfilling Revised Bethesda criteria (three from families fulfilling Amsterdam II criteria) were sequenced for underlying hereditary defects explaining the family history, but no germline variants were found. One of these *MLH1* methylated tumors carried a somatic *MLH1* variant likely to affect function and one displayed *MLH1* LOH (Supplementary Table 1).

One (n=27, 44%) or two (n=13, 21%) somatic aberrations (variant or LOH) in a MMR gene were found in a total of 62 tumors (See Supplementary Table 1). Bi-allelic inactivation was concordant with IHC. Twelve of the thirteen tumors with two somatic aberrations had variants in *MLH1* or *MSH2* and were MSI-H. The thirteenth tumor, sLS-07, showed expression loss of *MSH6* and was MSS. While the majority (81%) of tumors showed less than 10 somatic variants in the genomic region analyzed, ten cases displayed a larger number of somatic variants, ranging from 16 to 375 somatic variants within the sequenced area of 31 kb. Nine out of ten tumors showed a *POLE* or *POLD1* variant which (probably) affects function (Table 1). Of the highly mutated tumors, two carried novel germline heterozygous *POLE/POLD1* variants that are predicted to affect proofreading (Supplementary Table 2). Of these two germline cases, tumor sLS-67, was also found to carry two somatic *MLH1* variants, explaining the tumor phenotype (loss of MLH1 expression and MSI-H). The second tumor, sLS-16, was MSI-L, showed positive MMR expression and had no somatic MMR variants.

Seven of the highly mutated tumors showed somatic *POLE*/*POLD1* variants likely to affect function. Six tumors carried a somatic *POLE/POLD1*-EDM hotspot variant (*POLE*: c.857C>G, c.856C>T, c.1231G>T, c.1366G>C, c.1367C>T or c.1376C>T and *POLD1* c.1433G>A) that has previously been described to impair proofreading [4]. In the seventh tumor (sLS-105) a novel *POLE* c.846_847delinsTT variant was detected. This variant lies close to a known *POLE* hotspot site (*POLE* c.857) and is predicted to be affect function by 2 out of 3 prediction programs (Supplementary Table 2). All *POLE* variants were heterozygous, in agreement with previous research [7]. Four *POLE/POLD1-EDM* mutated tumors displayed additional somatic nonsense *POLE* variants outside the exonuclease domain (See Supplementary Table 3). Only one of these was upstream of the exonuclease domain (sLS-16).

Eight of the nine *POLE/POLD1*-EDM mutated tumors in our study showed microsatellite instability (3 MSI-H and 5 MSI-L). In six of these tumors IHC detected loss of at least one of the MMR proteins and all six tumors displayed somatic variants in the affected MMR gene likely to affect function. well as two somatic *PMS2* aberrations, while IHC showed solitary loss of PMS2 expression. Reanalysis of staining also showed ambiguous MLH1 staining (cytoplasmic enhancement and vague, focal nuclear staining).Tumor sLS-19 with two POLE variants, was found to have two somatic *MLH1* aberrations, as well as two somatic *PMS2* aberrations, while IHC showed solitary loss of PMS2 expression.

Table 1: Overview of patients with a *POLE/POLD1*-EDM mutated tumor

Patient	IHC	MSI	Onset age	Family history	# var	Gene	Variant	Amino acid alteration	%
Somatic *POLE/POLD1*-EDM variant									
sLS-05	None	L	62	FDR	330	*POLE*	c.1367C>T	p.(A456V)	28%
						MLH1	LOH	-	-
						PMS2	LOH	-	-
sLS-07	MSH6	S	39	TDR	37	*POLE*	c.1231G>T	p.(V411L)	44%
						MSH6	c.2735G>A	p.(W912*)	10%
						MSH6	c.2876 G>A	p.(R959H)	14%
sLS-09	MSH6	L	42	FDR	16	*POLE*	c.857C>G	p.(P286R)	38%
						MSH6	c.2539G>T	p.(E847*)	36%
sLS-19	PMS2	H	45	FDR	221	*POLE*	c.1376C>T	p.(S459F)	21%
						POLE	c.856C>T	p.(P286S)	9%
						MLH1	c.199G>A	p.(G67R)	19%
						MLH1	LOH	-	-
						PMS2	c.308C>T	p.(T103I)	11%
						PMS2	LOH	-	-
sLS-24	None	L	34	FDR	115	*POLE*	c.1366G>C	p.(A456P)	28%
sLS-66	MSH6	L	66	SDR	25	*POLD1*	c.1433G>A	p.(S478N)	32%
						MSH6	c.3600_3601del	p.(L1201Hfs*13)	28%
sLS-80	MSH2/ MSH6	H	52	FDR	5	*POLD1*	c.1429G>A	p.(V477M)	26%
						MSH6	c.3961A>G**	p.R1321G	52%
						MSH6	c.3186C>A	p.C1062*	25%
sLS-87	MSH2/ MSH6	H	49	FDR	9	*POLE*	c.1218C>G	p.(N406E)	16%
						MSH6	c.3473_3475del**	p.(C1158del)	53%
						MSH6	c.3311_3312del	p.(F1104Trpfs*3)	21%
sLS-101	PMS2	H	55	FDR	5	*POLD1*	c.1003A>G	p.(I335V)	22%
						PMS2	c.1687C>T	p.R563*	30%
sLS-105	MLH1/ PMS2	H	49	No	184	*POLE*	c.846_847delinsTT	p.(L283F)	13%
						MLH1	c.1614G>A	p.(W538*)	13%
Germline *POLE/POLD1*-EDM variant									
sLS-16	None	L	41	FDR	185	*POLD1*	c.961G>A	p.(G321S)	55%
sLS-67	MLH1/ PMS2	H	53	SDR	91	*POLE*	c.861T>A	p.(D287E)	50%
						MLH1	c.208-1G>A	p.?	14%
						MLH1	c.440_447del	p.(G147Dfs*22)	19%

EDM, exonuclease domain. MSI-status is defined as MSI-H (H), MSI-L (L) or MSS (S). Onset age is the age at which the first LS-associated tumor occurred. All patients presented with colorectal cancer, except patient sLS-87 and sLS-101 which presented with endometrial cancer. #var depicts the number of somatic variants with a frequency >10% identified in the sequenced region of 31 kb. Stop codons are indicated with an asterisk (*). Germline variants of unknown significance (VUS) are indicated with a double asterisk (**) % shows the percentage of variant reads. Family history is defined closest relative with LS-associated tumor, FDR: first-degree relative, SDR: second-degree relative, TDR: third-degree relative or no LS in the family (No).

Reanalysis of staining also showed ambiguous MLH1 staining (cytoplasmic enhancement and vague, focal nuclear staining). In three *POLE/POLD1* mutated tumors with positive MMR expression and MSI-L phenotype, no somatic MMR variants (likely) to affect function were found. However, in one of those three tumors (sLS-05) solitary *MLH1* LOH and *PMS2* LOH without variants was found (See Table 1).

In three non-ultramutated tumors a *POLE/POLD1*-EDM variant was found (sLS-80, sLS-87 and sLS-101, see Table 1). These variants have not been described before, but are predicted to affect function (See Supplementary Table 2). Two variants co-occur with a germline *MSH6* VUS and somatic *MSH6* variant (sLS-80 and sLS-87), while one (sLS-101) co-occurs with a somatic *PMS2* variant. Four additional non-ultramutated tumors showed *POLE* -(sLS-18, sLS-21) or *POLD1* LOH (sLS-12, sLS-49) in all heterozygous SNPs (Supplementary Table 1), without germline or somatic *POLE* variants. None of these tumors displayed an ultramutated phenotype.

Discussion

POLE-EDM variants are reported to be the mutagenic factor driving ultramutation in tumors.[20] The number of variants detected in the sequenced area in the present study implicates an ultramutated phenotype, with >100 variants/Mb in all *POLE/POLD1* mutated tumors in this cohort. Since only a limited region (31 Kb) was sequenced, we can only extrapolate the total number of variants per Mb. In our cohort the *POLE/POLD1* mutated MMR-deficient tumors display two deficient pathways increasing the mutational load. Comparing frequencies of the different variants found in these tumors, it might be concluded that faulty proofreading may be the initiating event in these tumors, possibly resulting in loss of MMR and thereby in microsatellite instability. Interestingly, four tumors show *POLE/POLD1* LOH without germline or somatic *POLE/POLD1* variants. These tumors however do not show the typical ultramutated phenotype, whereas single variants without LOH do show that. This phenomenon of LOH without variants affecting the exonuclease domain has not yet been described. Possible the remaining allele is enough to maintain proofreading. Furthermore, three tumors show somatic *POLE/POLD1* variants, without the ultramutated phenotype. All three variants are missense, but are predicted to affect function (Supplementary Table 2). Since these variants are not found in ultramutated tumors, evidence of pathogenicity is lacking.

In conclusion, targeted next-generation sequencing of 62 sLS cases led to the detection of nine higly mutated tumors with a germline- (n=2) or somatic- (n=7) *POLE/POLD1-EDM* variant. Even though *POLE* germline variants have previously been shown to co-occur with somatic MMR variants[11], in this study we found germline and somatic *POLE/POLD1* variants in a cohort selected for sLS characteristics. Importantly, while current literature mainly addressed *POLE/POLD1* variants in MSS tumors, somatic *POLE/POLD1* variants in sLS patients are likely to be overlooked. Screening of *POLE/POLD1* should be added to the current germline and somatic diagnostic screening for MSI-H and MMR-deficient cases and could resolve the causal defect in these presently unexplained cases.

Supplementary information accompanies this paper on European Journal of Human Genetics website (*http://www.nature.com/ejhg*)

References

1. Colas C, Coulet F, Svrcek M, et al. Lynch or not Lynch? Is that always a question? Adv Cancer Res 2012;113:121-66.
2. Mensenkamp AR, Vogelaar IP, van Zelst-Stams WA, et al. Somatic mutations in MLH1 and MSH2 are a frequent cause of mismatch-repair deficiency in Lynch syndrome-like tumors. Gastroenterology 2014;146:643-6 e8.
3. Geurts-Giele WR, Leenen CH, Dubbink HJ, et al. Somatic aberrations of mismatch repair genes as a cause of microsatellite-unstable cancers. J Pathol 2014;234:548-59.
4. Shinbrot E, Henninger EE, Weinhold N, et al. Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. Genome Res 2014;24:1740-50.
5. Buchanan DD, Rosty C, Clendenning M, Spurdle AB, Win AK. Clinical problems of colorectal cancer and endometrial cancer cases with unknown cause of tumor mismatch repair deficiency (suspected Lynch syndrome). Appl Clin Genet 2014;7:183-93.
6. Heitzer E, Tomlinson I. Replicative DNA polymerase mutations in cancer. Curr Opin Genet Dev 2014;24:107-13.
7. Palles C, Cazier JB, Howarth KM, et al. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nat Genet 2013;45:136-44.
8. Valle L, Hernandez-Illan E, Bellido F, et al. New insights into POLE and POLD1 germline mutations in familial colorectal cancer and polyposis. Hum Mol Genet 2014;23:3506-12.
9. Briggs S, Tomlinson I. Germline and somatic polymerase epsilon and delta mutations define a new class of hypermutated colorectal and endometrial cancers. J Pathol 2013;230:148-53.
10. Spier I, Holzapfel S, Altmuller J, et al. Frequency and phenotypic spectrum of germline mutations in POLE and seven other polymerase genes in 266 patients with colorectal adenomas and carcinomas. Int J Cancer 2015;137:320-31.
11. Haraldsdottir S, Hampel H, Tomsic J, et al. Colon and endometrial cancers with mismatch repair deficiency can arise from somatic, rather than germline, mutations. Gastroenterology 2014;147:1308-16 e1.
12. Yoshida R, Miyashita K, Inoue M, et al. Concurrent genetic alterations in DNA polymerase proofreading and mismatch repair in human colorectal cancer. Eur J Hum Genet 2011;19:320-5.
13. Church DN, Briggs SE, Palles C, et al. DNA polymerase epsilon and delta exonuclease domain mutations in endometrial cancer. Hum Mol Genet 2013;22:2820-8.
14. Billingsley CC, Cohn DE, Mutch DG, Stephens JA, Suarez AA, Goodfellow PJ. Polymerase varepsilon (POLE) mutations in endometrial cancer: clinical outcomes and implications for Lynch syndrome testing. Cancer 2015;121:386-94.
15. Elsayed FA, Kets CM, Ruano D, et al. Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer. European journal of human genetics : EJHG 2015 Aug; 23(8):1080-4
16. Umar A, Boland CR, Terdiman JP, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst 2004;96:261-8.
17. Park J-G, Vasen HF, Park Y, et al. Suspected HNPCC and Amsterdam criteria II: evaluation of mutation detection rate, an international collaborative study. International Journal of Colorectal Disease 2001;17:109-14.
18. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
19. Devilee P, Cleton-Jansen AM, Cornelisse CJ. Ever since Knudson. Trends Genet 2001;17:569-73.
20. Helleday T, Eshtad S, Nik-Zainal S. Mechanisms underlying mutational signatures in human cancers. Nat Rev Genet 2014;15:585-98.

Supplementary Methods

Study subjects
This retrospective cohort was collected between 1997 and 2014. Leukocyte DNA isolated from peripheral blood was available for all patients. Tumor DNA was isolated from formalin-fixed paraffin-embedded tumor tissue (FFPE) at time of diagnosis. Two patients were excluded from the analysis due to poor quality DNA. This study was approved by the local medical ethical committee of the LUMC (P01-019E).

Immunohistochemical analysis (IHC) and microsatellite instability testing was performed previously at request of board certified Clinical Genetics medical specialists. Routine testing of all four MMR proteins became available only around 2004. Therefore not of all tumors in our cohort complete MMR immunohistochemical results are available. At time of the current study leukocyte- and tumor DNA was retrieved from our archives. In a minority of cases FFPE-blocks were still available for retrospective testing. In 38 patients (61%) immunohistochemistry data was complete. In 11 tumors PMS2 immunohistochemistry was not performed. The remaining 13 tumors had one of more inconclusive immunohistochemical results. Patient sLS-05 and sLS-24 were initially included due to negative MSH6 staining, but reanalysis during the study showed no MSH6 expression loss. Patient sLS-11 was included due to a reported MSI-H status, but reanalysis showed no instability.

Targeted next-generation sequencing
Tumor and leukocyte DNA was sequenced for variants in *MLH1*, *MSH2*, *MSH6*, *PMS2, POLE* and *POLD1* using the Ion PGM™ System (Life Technologies). Ion AmpliSeq™ Custom Panels were designed with the Ion AmpliSeq™ Designer tool. The complete sequencing panel consisted of 307 amplicons (31094 bp), covering 98%, 90%, 98%, 75%, 95%, and 78% of the coding regions of *MLH1, MSH2, MSH6, PMS2, POLE, and POLD1*, respectively. Libraries were prepared with Ion AmpliSeq™ Library Kit 2.0 according to the manufacturer's protocol. For template preparation the Ion OneTouch™ 2 System and the Ion Chef™ System were used.

Data analysis
Raw data analysis, alignments, and variant calling was carried out using the default parameters in Torrent Suite v4.0. The Variant Caller Parameter Setting was set on 'Somatic – PGM – Low Stringency'. Variants were functionally annotated using ANNOVAR.[1] The following Genbank reference sequences were used: NM_000249.3 for *MLH1*, NM_000251.2 for *MSH2*, NM_000179.2 for *MSH6*, NM_000535.5 for *PMS2*, NM_006231.2 for *POLE* and NM_001256849.1 for *POLD1*. Recommendations of the Human Genome Variation Society (HGVS) to use the terms "variant" and "likely to affect function" instead of "mutation" and "pathogenic" were followed (*http://www.hgvs.org/mutnomen/recs.html*). Classification of the functional effects of the variants was done according to the five-tiered InSiGHT scheme.[2]

Variant filtering and validation
The full dataset was filtered by variant frequency (>10%) and coverage (>50x). Variants with a minor allele frequency (MAF) >0.05, as reported in the exome-database (*http://www.ncbi.nlm.nih.gov/projects/SNP/*), or a MAF >0.01, as reported in the Genome of the Netherlands (*http://nlgenome.nl*), were excluded from further analysis. In addition, coding variants

were filtered further based on predictions by *in silico* missense prediction software Align GVGD[3], SIFT (*http://sift.jcvi.org/*), PolyPhen-2 (*http://genetics.bwh.harvard.edu/pph2/*), MutationTaster (*http://www.mutationtaster.org/*) and MAPP (*http://mendel.stanford.edu/SidowLab/downloads/MAPP/index.html*). The Leiden Open Variation Database (LOVD) was consulted to find variants previously described and classified (*http://www.lovd.nl/3.0/home*). All predicted to affect function were visually inspected with the Integrative Genomics Viewer (*https://www.broadinstitute.org/igv/home*).Variants predicted to affect function were validated with Sanger sequencing. Germline variants of this study are submitted to the Leiden Open (source) Variation Database (LOVD) available at *http://www.lovd.nl/3.0/home* and somatic variants are submitted to the COSMIC database (*http://cancer.sanger.ac.uk/cosmic*) .

References

1. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research. 2010;38(16):e164.
2. Thompson BA, Spurdle AB, Plazzer JP, Greenblatt MS, Akagi K, Al-Mulla F, et al. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nature genetics. 2014;46(2):107-15.
3. Tavtigian SV, Deffenbaugh AM, Yin L, Judkins T, Scholl T, Samollow PB, et al. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. Journal of medical genetics. 2006;43(4):295-305.

Supplementary Table 1: Overview of all variants that are prediced to affect function

Patient	Gender	IHC MLH1	IHC MSH2	IHC MSH6	IHC PMS2	Tumor	Age of onset	MSI	Family History	# var	POLE/POLD1 variant	%	MMR variants	%
sLS-01	Male	-	+	+	-	CRC	54	Unknown	FDR	4	-	-	MLH1 c.1985_1988del	34%
sLS-03	Male	-	+	+	-	CRC	52	High	Unknown	5	-	-	MLH1 c.1252_1253delGA MLH1 LOH	64% -
sLS-05	Female	+	+	+	+	CRC	62	Low	FDR	330	POLE c.1367C>T	28%	MLH1 LOH PMS2 LOH	- -
sLS-06	Male	INC	-	NT	NT	CRC	47	High	FDR	8	-	-	MSH2 c.2131 C>T MSH2 c.1600_1601delCG	20% 19%
sLS-07	Male	INC	INC	-	NT	CRC	39	Stable	TDR	37	POLE c.1231 G>T	44%	MSH6 c.2876G>A MSH6 LOH	14% -
sLS-09	Male	+	+	-	NT	CRC	42	Low	FDR	16	POLE c.857C>G	38%	MSH6 c.2539G>T MSH2 c.1166G>A	36% 38%
sLS-11	Female	INC	INC	NT	NT	Ad	44	Stable	FDR	2	-	-	-	-
sLS-12	Male	+	+	NT	NT	CRC	30	High	FDR	0	POLD1 LOH	-	MLH1 LOH	-
sLS-14	Female	+	+	NT	NT	CRC	62	High	FDR	3	-	-	-	-
sLS-16	Male	+	+	NT	NT	CRC	41	Low	FDR	185	POLD1 c.961G>A*	55%	-	-
sLS-17	Female	+	+	-	NT	CRC	39	Stable	FDR	0	-	-	-	-
sLS-18	Female	+	+	NT	NT	CRC, Oes	62, 71	High	FDR	1	POLE LOH	-	MLH1 LOH	-
sLS-19	Male	AMB	+	+	-	CRC	45	High	FDR	221	POLE c.1376C>T POLE c.856C>T	21% 9%	MLH1 c.199G>A MLH1 LOH PMS2 c.308C>T PMS2 LOH	19% - 11% -
sLS-20	Female	+	INC	INC	NT	CRC	55	Low	FDR	1	-	-	-	-
sLS-21	Female	+	+	-	+	Ad	61	High	FDR	7	POLE LOH	-	MSH6 c.1134_1135delAA	31%
sLS-24	Female	+	+	+	+	CRC	34	Low	FDR	115	POLE c.1366G>C	28%	-	-
sLS-26	Female	+	-	-	NT	EC, OC	57, 57	Low	SDR	0	-	-	-	-

Supplementary Table 1: Overview of all variants that are prediced to affect function; continued

Patient	Gender	IHC MLH1	IHC MSH2	IHC MSH6	IHC PMS2	Tumor	Age of onset	MSI	Family History	# var	POLE/POLD1 variant	%	MMR variants	%
sLS-29	Female	+	-	-	NT	CRC	49	High	FDR	2	-	-	MSH2 c.1786_1788del* MSH2 LOH MSH6 LOH	46% - -
sLS-36	Female	+	+	+	NT	Ad	43	High	No	6	-	-	-	-
sLS-39	Female	+	+	+	NT	OC	68	High	FDR	1	-	-	-	-
sLS-40	Male	+	+	+	INC	CRC	52	High	TDR	3	-	-	-	-
sLS-42	Male	+	INC	INC	INC	CRC	69	High	FDR	47	-	-	PMS2 LOH	-
sLS-43	Male	INC	-	-	NT	CRC	74	High	FDR	1	-	-	-	-
sLS-44	Female	+	-	-	NT	CRC	41	High	FDR	13	-	-	-	-
sLS-45	Female	+	+	-	NT	CRC	47	High	No	2	-	-	MSH2 c.1638_1640del	30%
sLS-46	Female	+	+	+	NT	CRC, EC	76, 79	High	No	2	-	-	-	-
sLS-49	Male	-	+	+	-	CRC	44	High	SDR	4	POLD1 LOH	-	MLH1 c.885-2A>G MLH1 LOH	57% -
sLS-51	Female	-	+	+	-	CRC	69	High	FDR	4	-	-	-	-
sLS-52	Female	-	+	+	-	EC	67	High	FDR	2	-	-	-	-
sLS-53	Male	-	+	+	-	CRC	81	High	FDR	3	-	-	MLH1 c.769A>C	20%
sLS-54	Female	-	+	INC	-	EC	64	High	SDR	2	-	-	-	-
sLS-55	Female	+	-	-	+	OC, EC	48, 48	High	FDR	5	-	-	MSH2 LOH MSH6 c.3974_3976del	- 24%
sLS-56	Male	+	-	-	+	CRC 2x	64 2x	High	Unknown	5	-	-	MSH2 c.1710T>A	19%
sLS-58	Male	-	+	+	-	CRC	39	Low	FDR	6	-	-	MLH1 c.790+1 G>A	28%
sLS-60	Female	-	+	+	NT	OC	44	High	No	1	-	-	MLH1 LOH	-
sLS-61	Female	-	+	+	-	CRC 2x	49, 51	High	FDR	1	-	-	-	-
sLS-62	Male	+	-	-	NT	CRC	35	Low	FDR	2	-	-	-	-

Supplementary Table 1: continued

Patient	Gender	IHC MLH1	IHC MSH2	IHC MSH6	IHC PMS2	Tumor	Age of onset	MSI	Family History	# var	POLE/POLD1 variant	%	MMR variants	%
sLS-64	Male	-	+	+	-	CRC	48	High	FDR	1	-	-	MLH1 c.2059C>T	28%
sLS-66	Male	+	+	-	+	CRC	66	Low	TDR	25	POLD1 c.1433 G>A	32%	MSH6 c.3600_3601delAC	28%
sLS-67	Male	-	+	+	-	CRC 2x	57, 60	High	SDR	91	POLE c.861 T>A*	51%	MLH1 c.208-1 G>A MLH1 c.440_447del	14% 19%
sLS-71	Male	+	-	-	+	CRC, Ur	52, 52	High	FDR	1	-	-	MSH2 c.493T>G*	35%
sLS-72	Female	+	-	-	+	EC, CRC	68, 73	High	Unknown	1	-	-	MSH2 c.1576dupA	29%
sLS-77	Female	+	-	-	+	CRC, Pol	45, 45	High	No	5	-	-	MSH2 c.2470C>T	38%
sLS-80	Male	+	-	-	+	CRC	52	High	FDR	5	POLD1 c.1429G>A	-	MSH6 c.3961A>G** MSH6 c.3186C>A	52% 25%
sLS-82	Female	-	+	+	-	CRC	69	High	FDR	6	-	-	-	-
sLS-84	Female	-	+	+	-	EC	46	High	Unknown	1	-	-	MLH1 c.1989+1 G>A MLH1 c.2224 C>T	11% 9%
sLS-85	Male	+	+	-	+	Ki, CRC	41, 44	High	SDR	4	-	-	MSH6 c.3139T>G	30%
sLS-87	Female	+	-	-	+	EC	49	High	FDR	9	POLE c.1218C>G	-	MSH6 c.3473_3475del** MSH6 c.3311_3312delTT	53% 21%
sLS-89	Male	-	+	+	-	CRC 3x	48, 51, 61	High	TDR	7	-	-	MLH1 c.2041G>A* MLH1 c.677 G>A	50% 28%
sLS-90	Male	+	-	-	+	CRC	46	High	FDR	2	-	-	MSH2 c.490G>T** MSH2 c.1511-1G>T	40% 18%
sLS-91	Male	-	+	+	-	CRC	49	Unknown	No	3	-	-	MLH1 c.350 C>T MLH1 LOH	16% -
sLS-92	Female	-	+	INC	-	CRC	45	High	FDR	3	-	-	MLH1 c.790+1 G>A MLH1 LOH	78% -

Supplementary Table 1: Overview of all variants that are prediced to affect function;continued

Patient	Gender	IHC MLH1	IHC MSH2	IHC MSH6	IHC PMS2	Tumor	Age of onset	MSI	Family History	# vari	POLE/POLD1 variant	%	MMR variants	%
sLS-101	Female	+	+	+	-	EC	55	High	FDR	5	POLD1 c.1003A>G	-	PMS2 c.1687C>T	30%
sLS-104	Male	+	-	-	+	SB	47	High	No	1	-	-	MSH2 c.271delG MSH2 LOH	92% -
sLS-105	Male	-	+	+	-	CRC	49	High	No	184	POLE c.847_846delinsTT	13%	MLH1 c.1614G>A	13%
sLS-109	Female	-	+	+	-	CRC	49	High	No	2	-	-	MLH1 c.1192C>T MLH1 c.207+1 G>A	34% 19%
sLS-111	Female	+	-	-	+	EC	58	High	FDR	8	-	-	MSH2 c.687delA MSH2 c.773 T>A	44% 41%
sLS-113	Male	-	+	+	-	CRC	46	High	SDR	2	-	-	MLH1 c.982 C>T	16%
sLS-115	Female	-	+	+	-	CRC	38	High	Unknown	3	-	-	MLH1 c.1852_1854del MLH1 LOH	54% -
sLS-118	Female	-	+	+	-	EC	55	High	No	8	-	-	MLH1 c.1975C>T MLH1 c.2104-1G>C	33% 36%
sLS-120	Male	-	+	+	-	CRC	57	Stable	FDR	3	-	-	-	-
sLS-124	Male	+	+	-	+	CRC	50	High	FDR	4	-	-	MSH6 c.3484G>C**	48%

Supplementary Table 1: continued

Patient	MLH1 Meth.	Conclusion	ACII or BC	LOH MLH1	LOH MSH2	LOH MSH6	LOH PMS2	LOH POLE	LOH POLD1
sLS-01	No	1 somatic MMR variant	No	-	NP	NP	NP	NP	NP
sLS-03	No	2 somatic MMR variants	BC	1/1	NP	NP	-	NP	NP
sLS-05	NA	Somatic POLE variant	ACII	1/1	-	-	1/1	-	-
sLS-06	NA	2 somatic MMR variants	ACII	-	NP	-	-	NP	NP
sLS-07	NA	Somatic POLE variant	BC	-	NP	3/3	-	-	-
sLS-09	NA	Somatic POLE variant	BC	-	-	-	-	-	-
sLS-11	No	No variant	No	-		-	-	-	-
sLS-12	NA	1 somatic MMR variant	ACII	1/1	-	-	-	-	5/5
sLS-14	NA	No variant	ACII	NP	NP	-	NP	-	-
sLS-16	NA	Germline POLD1 variant	ACII	NP	NP	NP	NP	NP	NP
sLS-17	NA	No variant	BC	-	-	-	-	-	NP
sLS-18	NA	1 somatic MMR variant	BC	1/1	-	-	-	14/14	-
sLS-19	No	Somatic POLE variant	ACII	2/2	-	-	2/2	-	-
sLS-20	NA	No variant	ACII	-	-	-	-	NP	-
sLS-21	No	1 somatic MMR variant	ACII	NP	NP	NP	-	1/1	NP
sLS-24	NA	Somatic POLE variant	ACII	-	NP	-	NP	-	NP
sLS-26	NA	No variant	BC	NP	-	-	-	-	-
sLS-29	NA	Germline MMR variant with LOH	BC	NP	2/2	1/1	-	-	NP
sLS-36	NA	No variant	No	-	NP	-	-	-	-
sLS-39	NA	No variant	No	NP	NP	-	-	NP	-
sLS-40	NA	No variant	BC	-	NP	NP	-	-	NP
sLS-42	No	1 somatic MMR variant	ACII	-	-	-	4/4	NP	-
sLS-43	NA	No variant	BC	-	-	-	-	-	NP

Supplementary Table 1: Overview of all variants that are prediced to affect function; continued

Patient	MLH1 Meth.	Conclusion	ACII or BC	LOH MLH1	LOH MSH2	LOH MSH6	LOH PMS2	LOH POLE	LOH POLD1
sLS-44	NA	No variant	ACII	NP	NP	NP	NP	NP	NP
sLS-45	NA	1 somatic MMR variant	ACII	-	-	-	NP	-	-
sLS-46	NA	No variant	BC	NP	NP	-	NP	-	-
sLS-49	No	2 somatic MMR variants	BC	1/1	-	-	-	-	5/5
sLS-51	Yes	Sporadic - MLH1 Methylation	ACII	NP	-	NP	-	NP	NP
sLS-52	Yes	Sporadic - MLH1 Methylation	BC	-	-	-	-	-	NP
sLS-53	Yes	Sporadic - MLH1 Methylation	ACII	-	NP	NP	-	-	-
sLS-54	Yes	Sporadic - MLH1 Methylation	ACII	-	NP	-	-	NP	NP
sLS-55	NA	1 somatic MMR variant	BC	-	1/1	NP	-	NP	-
sLS-56	NA	1 somatic MMR variant	BC	-	-	-	-	-	NP
sLS-58	No	1 somatic MMR variant	ACII	NP	-	NP	NP	-	NP
sLS-60	Yes	Sporadic - MLH1 Methylation	BC	2/2	NP	-	-	NP	-
sLS-61	Yes	Sporadic - MLH1 Methylation	BC	NP	NP	-	-	-	NP
sLS-62	NA	No variant	BC	-	-	-	-	-	-
sLS-64	No	1 somatic MMR variant	BC	-	NP	-	-	-	-
sLS-66	NA	Somatic POLD1 variant	No	NP	-	-	-		-
sLS-67	No	Germline POLE variant	BC	-	-	-	NP	-	NP
sLS-71	NA	Germline MMR variant	ACII	-	NP	-	-	-	-
sLS-72	NA	1 somatic MMR variant	BC	NP	NP	NP	NP	NP	NP
sLS-77	NA	1 somatic MMR variant	BC	-	-	-	-	-	-
sLS-80	NA	Somatic POLD1 variant, germline VUS and 1 somatic MMR variant	ACII	-	-	-	-	-	-
sLS-82	No	No variant	BC	-	NP	-	-	-	NP
sLS-84	No	2 somatic MMR variants	BC	-	NP	-	-	-	-

Supplementary Table 1: continued

Patient	MLH1 Meth.	Conclusion	ACII or BC	LOH MLH1	LOH MSH2	LOH MSH6	LOH PMS2	LOH POLE	LOH POLD1
sLS-85	NA	1 somatic MMR variant	BC	NP	-	-	-	-	NP
sLS-87	NA	Somatic POLE variant, germline VUS and 1 somatic MMR variant	ACII	-	-	-	-	-	-
sLS-89	No	Germline MMR variant	BC	-	NP	-	-	-	-
sLS-90	NA	Germline VUS and 1 somatic MMR variant	ACII	-	-	-	-	-	-
sLS-91	No	1 somatic MMR variant	BC	1/1	NP	NP	-	-	-
sLS-92	No	2 somatic MMR variants	ACII	1/1	-	NP	-	-	NP
sLS-101	No	Somatic POLE variant	BC	NP	-	-	NP	-	-
sLS-104	NA	2 somatic MMR variants	No	NP	1/1	NP	-	-	-
sLS-105	No	Somatic POLE variant	BC	NP	NP	NP	NP	NP	NP
sLS-109	No	2 somatic MMR variants	BC	-	-	-	-	-	-
sLS-111	NA	2 somatic MMR variants	ACII	NP	-	-	NP	-	-
sLS-113	No	1 somatic MMR variant	ACII	-	NP	-	NP	-	NP
sLS-115	No	2 somatic MMR variants	BC	NP	-	-	-	-	NP
sLS-118	No	2 somatic MMR variants	BC	-	-	-	-	-	-
sLS-120	No	No variant	BC	-	-	-	-	-	-
sLS-124	NA	Germline VUS	BC	NP	NP	-	-	-	-

All variants described are nonsense, frameshift or missense variants that are likely to affect function in the MMR genes and POLE/POLD1 exonuclease domain. Patients presented with colorectal- (CRC), endometrial- (EC), ovarian- (OC), kidney- (Ki), sebaceous gland- (SB), urinary tract- (Ur), esophageal cancer (Oes) or adenoma with high grade dysplasia (Ad). Staining was defined as positive (+), negative (-), inconclusive (INC) or ambiguous (AMB). Family history is defined as first-degree relative with LS-associated tumors (FDR), second-degree relative with LS-associated tumors (SDR), third-degree relative with LS-associated tumor (TDR) or no Lynch-associated tumors in the family (No). Germline variants are indicated with an asterisk (*). Germline variants of unknown significance (VUS, class 3) are depicted with double asterisks (**). % shows the percentage of variant reads. MLH1 hypermethylation status is defined as yes, no or not applicable (NA). Patients fulfilled revised bethesda criteria (BC), Amsterdam II criteria (ACII) or neither (No). LOH was called when all heterozygous SNPs showed significant imbalance. LOH classified as no LOH (-), LOH, with the number of informative SNPs (#/#) or LOH analysis was not possible (NP).

Supplementary Table 2: Missense prediction of *POLE/POLD1* variants

Missense prediction									
***POLE/POLD1-EDM* variants found in hypermutated tumors**									
Patient	**Gene**	**Variant**	**Amino acid alteration**	**Observed**	**Align GVGD**	**SIFT**	**Mutation Taster**	**Previously described**	**Rs id**
sLS-05	*POLE*	c.1367C>T	p.(A456V)	Somatic	C0	Deleterious	Disease causing	Yes	-
sLS-07	*POLE*	c.1231G>T	p.(V411L)	Somatic	C0	Deleterious	Disease causing	Yes	-
sLS-09	*POLE*	c.857C>G	p.(P286R)	Somatic	C0	Deleterious	Disease causing	Yes	-
sLS-16	*POLD1*	c.961G>A	p.(G321S)	Germline	C55	Deleterious	Disease causing	No	rs41554817
sLS-19	*POLE*	c.856C>T	p.(P286S)	Somatic	C0	Deleterious	Disease causing	Yes	-
sLS-19	*POLE*	c.1376C>T	p.(S459F)	Somatic	C0	Deleterious	Disease causing	Yes	-
sLS-24	*POLE*	c.1366G>C	p.(A456P)	Somatic	C0	Deleterious	Disease causing	No	-
sLS-66	*POLD1*	c.1433G>A	p.(S478N)	Somatic	C45	Deleterious	Disease causing	Yes	rs397514632
sLS-67	*POLE*	c.861T>A	p.(D287E)	Germline	C0	Deleterious	Disease causing	No	rs139075637
sLS-110	*POLE*	c.846_847delinsTT	p.(L283F)	Somatic	C0	Deleterious	Disease causing	No	-
***POLE/POLD1-EDM* variants found in non-hypermutated tumors**									
sLS-80	*POLD1*	c.1429G>A	p.(V477M)	Somatic	C15	Deleterious	Disease causing	No	-
sLS-87	*POLE*	c.1218C>G	p.(N406E)	Somatic	C0	Deleterious	Disease causing	No	-
sLS-101	*POLD1*	c.1003A>G	p.(I335V)	Somatic	C25	Deleterious	Disease causing	No	-

Overview of missense prediction of POLE/POLD1 exonuclease (EDM) variants. Missense prediction is based on 3 prediction programs: Align GVGD [C0 - C65, with high values indicating likely deleterious variants], SIFT [Tolerated/Deleterious], MutationTaster [Polymorphism/Disease Causing].

Supplementary Table 3: Additional nonsense variants in POLE outside the EDM-domain

	POLE/POLD1-EDM variants			Somatic MMR variants			Non-EDM *POLE/POLD1* variants			
Patient ID	**Gene**	**Variant**	**%**	**Gene**	**Variant**	**%**	**Gene**	**Variant**	**%**	**Location**
sLS-05	*POLE*	c.1367C>T	28%	*MLH1*	LOH	-	*POLE*	c.4933C>T	23%	A
				PMS2	LOH	-				
sLS-16	*POLD1*	c.961G>A*	55%	*MLH1*	c.208-1 G>A	14%	*POLE*	c.448C>T	24%	B
				MLH1	c.440_447del	19%				
sLS-19	*POLE*	c.1376C>T	21%	*MLH1*	c.199G>A	19%	*POLE*	c.6181C>T	23%	A
				MLH1	LOH	-				
	POLE	c.856C>T	9%	*PMS2*	c.308C>T	11%	*POLE*	c.5287C>T	14%	A
				PMS2	LOH	-				
sLS-105	*POLE*	c.847_846delinsTT	13%	*MLH1*	c.1614G>A	13%	*POLE*	c.2059C>T	17%	A
							POLE	c.2065C>T	17%	A

Patients with additional POLE nonsense variants outside the exonuclease (EDM) domain. % shows variant frequency. Asterisk (*) indicates germline variants. A/B shows whether nonsense variants were before (B) or after (A) the exonuclease domain.

Chapter 4

RNA analysis of cancer predisposing genes in formalin-fixed paraffin-embedded tissue determines aberrant splicing

AML Jansen[1,2], HM van der Klift[3], MAE Roos[1], JDH van Eendenburg[1], CMJ Tops[3], JT Wijnen[3], FJ Hes[3], H Morreau[2], T van Wezel[2]

EJHG, in press

Department of [1]Pathology, [2]Human Genetics and [3]Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands

Abstract

High-throughput sequencing efforts in molecular tumor diagnostics detect increasing numbers of novel variants, including variants predicted to affect splicing. *In silico* prediction tools can reliably predict the effect of variant disrupting canonical splice sites; however, experimental validation is required to confirm aberrant splicing. Here, we present RNA analysis performed for 13 canonical splice site variants predicted- or known to result in splicing in the cancer predisposition genes *MLH1*, *MSH2, MSH6, APC* and *BRCA1.* Total nucleic acid was successfully isolated for 10 variants from eight formalin-fixed paraffin-embedded (FFPE) tumor tissues and two B-cell lines. Aberrant splicing was confirmed in all six variants known to result in splicing. Of one known variant in the B-cell line, aberrant splicing could only be detected after formalin fixation, which indicated that formalin fixation could possibly inhibit RNA degradation. Aberrant splicing was concluded in three of four predicted splice variants of uncertain significance, supporting their pathogenic effect. With this assay, somatic splice variants can be easily and rapidly analysed, enabling retrospective analysis to support the pathogenicity of variants predicted to result in splicing when only FFPE material is available.

Keywords: Colorectal cancer; splice site; splice prediction; FFPE; RNA

Introduction

Most mammalian genes consist of multiple exons interspersed by long intronic sequences.[1] To create mature mRNA, the introns must be correctly identified and 'spliced out', and the exons joined together.[1] The spliceosome, the splicing machinery responsible for this process, recognizes conserved motifs at or near the intron ends and a branch site within the intron.[1] Splice regulatory elements (SRE) near exon-intron boundaries, such as SR (serine-arginine rich) or hnRNP (heterogeneous nuclear ribonucleoparticle) proteins, are indispensable for correct splice site identification.[2] These elements can enhance or repress splicing and play an important role in alternative splicing.[1, 2]

Of the variants that cause disease, 15-60% are proposed to disrupt splicing.[3] Included are variants in the canonical splice site, which directly alter the canonical splice site efficiency, but also intronic and exonic variants that alter the SREs or result in creation of a new splice site or activation of a cryptic splice site.[3, 4] The latter could result in inclusion of a pseudoexon, an intronic sequence wrongly interpreted as an exon. Exon skipping or inclusion of a pseudoexon often results in a shift of the open reading frame, resulting in a premature stop codon or leading to a non- or less-functional protein. The mechanism of nonsense-mediated decay (NMD) in which mRNAs with a premature termination codon are degraded can remove aberrant mRNAs encoding for truncated proteins, ensuring mRNA quality.[5]

Current assessment whether variants result in aberrant RNA transcripts often consists of *in silico* prediction with bioinformatics prediction tools, sometimes followed by reverse transcriptase PCR (RT-PCR) analysis of RNA extracted from blood[6-8] or functional splicing reporter minigene assays.[4, 9] Splicing microarrays can be used for large-scale identification of splicing differences but are not always implemented in current diagnostics.[3] Although high-quality patient RNA analysis is usually preferred, this RNA is not always available, or the analysis is hampered because of degradation of aberrant transcripts through NMD.[1]

With the current high-throughput sequencing methods applied in molecular tumor diagnostics, many variants are found, most of uncertain significance (VUS); hence, functional tests are required to classify these variants. Of these unclassified variants, a percentage is predicted to affect splicing. Specific kits are available to isolate RNA from formalin-fixed paraffin-embedded (FFPE) tissue blocks, and previous studies show that PCR, RT-PCR and even next-generation sequencing (NGS) are possible on these RNA samples.[10, 11] RNA analysis on RNA isolated from FFPE tissue is currently not standardly performed but would enable analysis of somatic splice site variants.

In the current study, the effect of splice site variants was examined in multiple cancer susceptibility genes, *MLH1, MSH2, MSH6, APC* and *BRCA1*. *MLH1, MSH2* and *MSH6* are part of the mismatch repair (MMR) pathway. Pathogenic heterozygous germ line variants in the MMR genes cause Lynch Syndrome, an autosomal dominant predisposition for colorectal, endometrial and other cancers.[12] Other known causes of MMR-deficiency are somatic *MLH1* promoter hypermethylation and the recently described biallelic somatic inactivation of the MMR genes caused by somatic variants.[13-15] Pathogenic germ line variants in the *APC* gene are known to result in familial adenomatous polyposis, a dominant disorder characterized

by the occurrence of hundreds to even thousands of adenomas throughout the colon.[16, 17] In a small percentage of patients, the tumor phenotype can be explained by mosaic *APC* variants.[18-20] These variants can be easily detected by screening multiple adenomas, because the *APC* variant is present with a higher variant allele frequency in the tumor.[18, 21] Pathogenic variants in the *BRCA1* gene, a key player in the nucleotide excision repair pathway, result in a high susceptibility to breast and ovarian cancers.[22, 23] Because BRCA-mutation status affects treatment strategies (PARP-inhibitors); the ability to detect and functionally assess both germ line and somatic mutations in *BRCA1* and *BRCA2* must increase.[24, 25] With the shift towards increased diagnostic screening of tumor tissue for all three syndromes, more somatic variants are found, which require functional tests to assess their pathogenicity.

The aim of our study was to investigate the possibility of analysing RNA isolated from FFPE tissue to assess the effect of germ line and somatic variants predicted to affect splicing. We hypothesized that formalin fixation could inhibit RNA degradation, enabling the detection of aberrant RNA in FFPE tissues.

Materials and methods

Selection of variants

In total, 13 variants of the cancer susceptibility genes *MLH1*, *MSH2, MSH6*, *APC* and *BRCA1* were tested for their effect on splicing (Supplementary Table 1). Of all variants, eight were somatic variants found between 2014 and 2017 with next-generation sequencing in a previous study[26] or through molecular tumor diagnostic NGS, with all having a variant allele frequency of at least 12%. Five were germ line splice site variants, all previously demonstrated to result in aberrant RNA (Supplementary Table 1). The *MLH1* c.454_545del, a germ line genomic exon six deletion, was added as a positive control.

RNA isolation and cDNA synthesis

For eleven variants, total nucleic acid was obtained from tissue cores punched from FFPE blocks embedded between 2009 and 2016. Tumor areas were marked on a hematoxylin and eosin stained slide by a pathologist. Tissue cores from the corresponding area on the FFPE block were punched with a 0.6 mm biopsy needle. Total nucleic acid was isolated from the obtained punches and microdissected areas with a Tissue Preparation System with VERSANT Tissue Preparation Reagents (Siemens Healthcare Diagnostics, Tarrytown, NY, USA).[27] For two variants (*MLH1* c.791-1G>C and *MSH2* c.1511-2A>G), no FFPE tissue was available, but EBV-transformed B-cells were cultured. Additionally, RNA from the *MSH2* c.1511-2A>G B-cell line was isolated after incubating the cells with 4% formalin for five hours. RNA from the B-cell lines and three colorectal cancer cell lines (SW480, SW837 and LS180) was isolated using a Nucleospin RNA isolation kit (Macherey-Nagel-06/2015, Rev.17, Düren, Germany) according to the manufacturer's protocol. Colorectal cancer cell lines were used as a positive control for RNA expression.

All cDNA was synthesized using OligoDT's and random primers with a SuperScript VILO cDNA synthesis Kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol.

Splice site prediction/variant nomenclature
For (canonical) splice site prediction, Alamut (Interactive Biosoftware, Rouen, France) was used. This software package includes the *in silico* splice site prediction algorithms SpliceSite Finder (SSF), MaxEntScan (MES) (*http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html*), NNSPLICE (*http://www.fruitfly.org/seq_tools/splice.html*) and Human Splicing Finder (HSF, *http://www.umd.be/HSF/*). Variants were annotated according to the Human Genetics Variation Society (HGVS) guidelines. Recommendations of the Human Genome Variation Society (HGVS) were followed to use the term "variant" instead of "mutation" (*http://www.hgvs.org/mutnomen/recs.html*). The following Genbank reference sequences were used: NM_000249.2 for *MLH1,* NM_000251.2 for *MSH2,* NM_000179.2 for *MSH6,* NM_000038.5 for *APC* and NM_007294.3 for *BRCA1.*

Primer design and PCR
For all variants with an (predicted) exon skip, two primer pairs were created to amplify two exon-exon boundaries. Primers were used in three combinations: Forward1/Reverse1, Foward2/Reverse2 and Forward1/Reverse2. For all variants with a partial exon skip or pseudoexon insertion as (predicted) RNA effect, primers were designed to amplify the exon-exon boundary. All primer sequences are listed in Supplementary Table 2. Real-time PCR was used to amplify the exon-exon boundaries and to assess the expression of the affected gene. All PCR reactions were performed on a CFX96 touch Realtime PCR machine (Bio-rad, Hercules, CA, USA) with the following PCR program: 95 °C for 5 min (1 cycle), 95 °C for 15 s, 60 °C for 30 s, and 72 °C for 30 s (38 cycles), followed by a melt curve from 65 °C to 95 °C with a 0.5 °C increment for 5 s with plate read. When no PCR product was detected, PCR was repeated with high cDNA input and 44 instead of 38 cycles. Because of limited cDNA, only F1/R2 was repeated for variants with two primer pairs, when the first PCR failed. All PCR products were analysed on a Qiaxcel capillary electrophoresis system (Qiagen, Hilden, Germany) and sequenced with Sanger Sequencing.

Results

MMR variants
Five MMR splice variants and one *MLH1* genomic exon deletion in RNA isolated from FFPE tissue were analysed for their effect on splicing (Table 1). Total nucleic acid was isolated from FFPE blocks and converted to cDNA using OligoDT's and random primers. The quality of cDNA was evaluated by detecting the expression of housekeeping genes (HKG) *HNRNPM* and/or *CPSF6.* From five FFPE tissue blocks, HKG expression was detected, and in three of the five, cDNA from the affected MMR gene could be amplified and analysed (*MLH1* c.454_545del, *MLH1* c.2104 G>C and *MSH6* c.3801+1_3801+5del). The amplified products of the three MMR cDNAs from FFPE tissues were measured with Qiaxcel (Figure 1A) and sequenced (Figure 1B). Size determination of the cDNAs carrying the *MLH1* c.454_545del and *MSH6* c.3801+1_3801+5del variants showed only a product size smaller than that of the WT control, whereas *MLH1* c.2104-1G>C only showed a product comparable in size with that of the WT product. Sequencing showed an aberrant product in two of the three FFPE samples, the *MLH1* genomic exon 6 deletion and a skip of exon 8 in the *MSH6* c.3801+1_3801+5del sample (Figure 1B), whereas for *MLH1* c.2104-1G>C, sequencing was normal, as was that for the WT control.

Additionally, RNA isolated from the two EBV-transformed B-cell lines carrying an *MSH2* c.1511-2A>G and an *MLH1* c.791-1G>C splice site variant was tested. Size determination of the cDNA from the two cell lines showed a product of approximately 350 bp (comparable with the WT control) and a product of 250 bp for the *MLH1* c.791-1G>C sample and a product comparable in size with that of the WT for the *MSH2* c.1511-2A>G sample. Sequencing detected an aberrant product only in the *MLH1* c.791-1G>C sample, which showed a skip of exon 10. To mimic FFPE conditions, 4% formalin was added to fixate EBV-transformed cells carrying the *MSH2* variant. After fixation, RNA was isolated, and cDNA was synthesized following the same protocol as that for the non-formalin fixed cells. cDNA from the formalin-fixed cells was tested and showed a size comparable with that of the WT (Figure 1A), and an aberrant product was detected with Sanger Sequencing (Figure 1B).

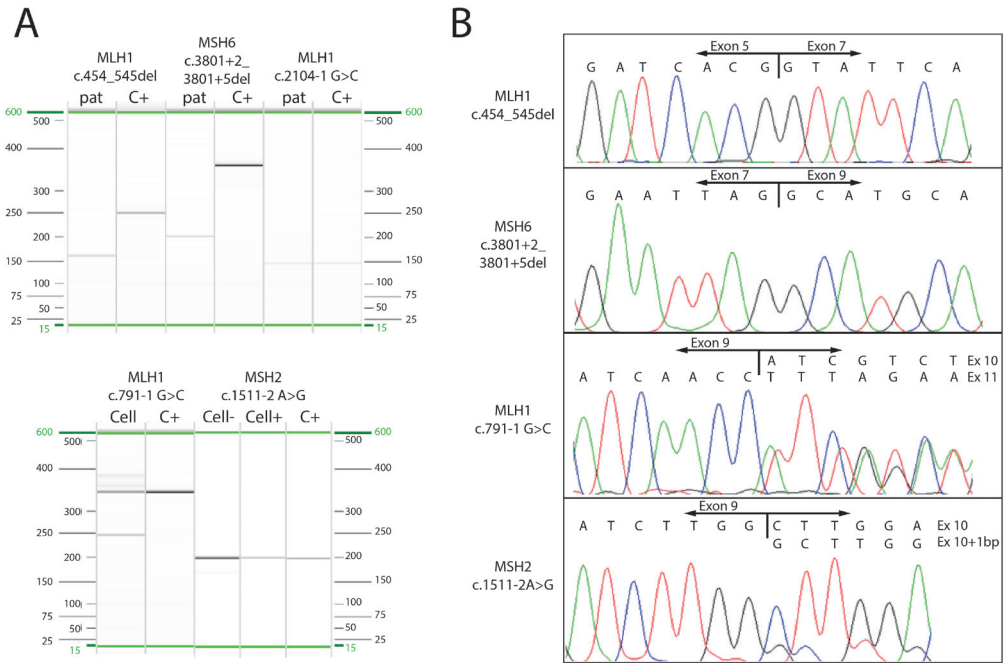


Figure 1A and B: Size determination and sequencing results MMR

Qiaxcel results showing the size of the MMR variants [A] of patient material (pat), cell lines (Cell) and control RNA isolated from colorectal cell lines (C+). The MSH2 cell line was analysed with (Cell+) and without (Cell-) formalin fixation of the cells. [B] Sanger sequencing results of variants showing aberrant products.

APC variants

Three *APC* variants were analysed for their effect on splicing (Table 1). RNA was successfully isolated from all three FFPE blocks, shown by detection of HKG expression. From all three samples, cDNA from *APC* could be amplified (Figure 1C) and sequenced (Figure 1D). Compared with the control without the variant, the difference in size of *APC* c.1548G>A was almost 125 bp. Sequencing showed an aberrant product for all three variants: a skip of the last 5 nucleotides of *APC* exon 7 in the *APC* c.834+2 T>A sample; a skip of the first nucleotide of exon 15 in the *APC* c.1959-1G>A sample; and an exon 11 skip in the *APC* c.1548G>A sample (Figure 1D).

Table 1: Results splicing assay

Gene	Variant	Source RNA	Variant frequency	(predicted) RNA effect	Result
			A. variants known to result in splicing		
MLH1	c.454_545del*	FFPE, CRC45	100%	Skip exon 6	Aberrant splicing as reported‡
MLH1	c.791-1G>C*	Cell line	50%	Skip exon 10	Aberrant splicing as reported
MLH1	c.793C>T*	FFPE, CRC63	NA	Skip exon 10 (major), WT (minor)	Unable to isolate nucleic acid
MLH1	c.1731G>A	FFPE, CRC51	35%	Skip exon 15	No MMR RNA expression detected
MSH2	c.1511-2A>G*	Cell line	80%	Last nt of intron 9 included	Aberrant splicing as reported§
MSH6	c.3801+1_ 3801+5del*	FFPE, Pol67	50%	Skip exon 8	Aberrant splicing as reported‡
APC	c.1548G>A	FFPE, Pol68	28%	Skip exon 11	Aberrant splicing as reported‡
BRCA1	c.213-12A>G	FFPE, EC74	93%	Insertion sequence corresponding to c.213-12_c.213-1	Aberrant splicing as reported‡
			B. variants predicted to result in splicing		
MLH1	c.885-2A>G	FFPE, CRC45	57%	Skip 5nt of exon 11†	No MMR RNA expression detected
MLH1	c.2104-1G>C	FFPE, EC55	36%	Skip 2nt of exon 19†	Only WT product
APC	c.834+2T>A	FFPE, Pol60	12%	Skip 5nt of exon 7†	Aberrant splicing as predicted
APC	c.1959-1G>A	FFPE, Pol67	40%	Skip 1nt of exon 15†	Aberrant splicing as predicted
BRCA1	c.212+3A>T	FFPE, OC46	54%	Skip 22nt of exon 5†	Aberrant splicing as predicted‡

*Source RNA shows cell line or formalin-fixed paraffin embedded (FFPE) tissue block, with type of tumor and age of onset. CRC = colorectal cancer, EC = endometrial cancer, Pol = polyps, WT = Wildtype, nt = nucleotides *Germline variants, variant allele frequency based on Sanger Sequencing of tumor material (estimated peak height. NA; not applicable, variant frequency in FFPE-tissue could not be assessed due to low quality †predicted RNA effect is based on nearest cryptic splice site, which could possibly act as a new 3' or 5' splice site. ‡no WT product detected §aberrant product only detected with formalin fixation of the cells.*

BRCA1 variants

Two *BRCA1* splice site variants were analysed for their effect on splicing (Table 1). RNA was successfully isolated from both FFPE blocks (shown by HKG detection), and *BRCA1* cDNA was amplified and analysed. PCR was performed with the same primers for both variants. Size determination indicated a smaller (*BRCA1* c.212+3A>T) and a slightly larger (*BRCA1* c.213-12A>G) product compared with that of the wild-type control (C+, Figure 1E). Sequencing showed a skip of the last 22 nucleotides of exon 5 (*BRCA1* c.212+3A>T) and an inclusion of the last 11 nucleotides of intron 5 corresponding to the *BRCA1* c.213-11_c.213-1 sequence (*BRCA1* c.213-12A>G, Figure 1F).

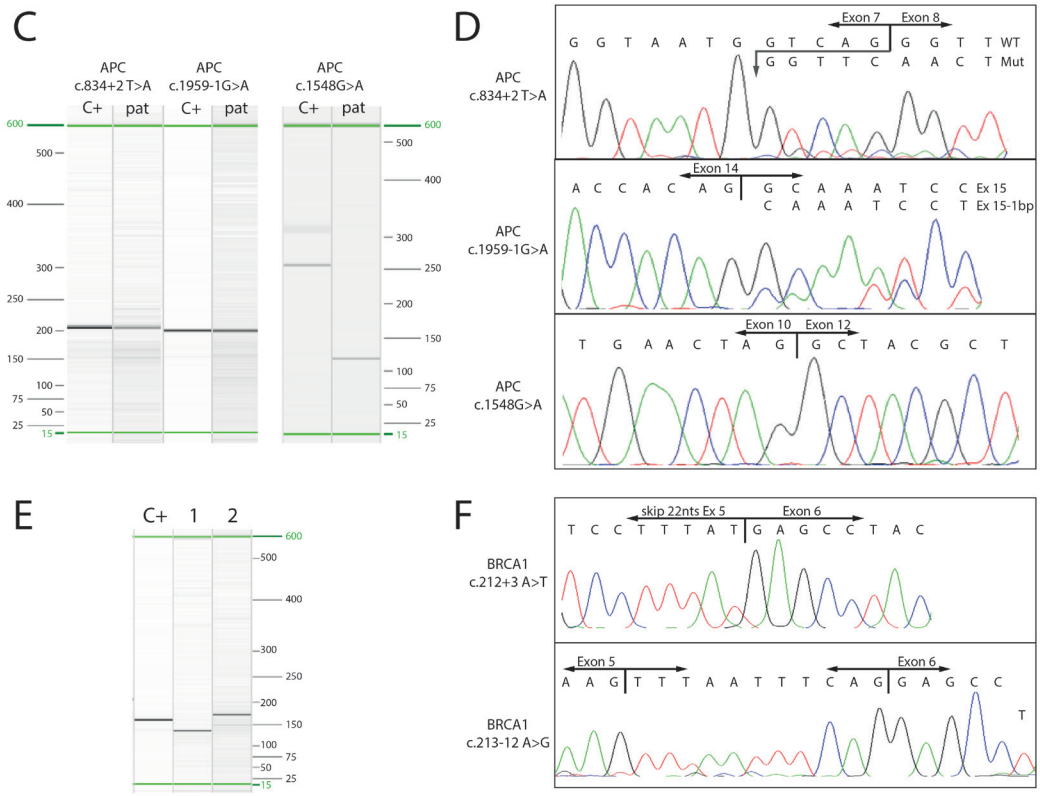


Figure 1C-F: Size determination and sequencing results *APC* and *BRCA1*.

Qiaxcel results showing the size of the APC variants [C] and BRCA1 variants [E] of patient material (pat) and control RNA isolated from colorectal cell lines (C+). The BRCA1 variants were analysed with the same primers, pat1 shows the BRCA1 c.212+3A>T and the pat2 shows the BRCA1 c.213-12A>G. [D] and [F] Sanger sequencing results of variants showing aberrant products.

Discussion

We performed RNA analysis for six splice site variants known to result in aberrant splicing and five variants predicted to result in aberrant splicing using RNA isolated from FFPE tissues. For the six variants shown previously to result in aberrant splicing, the reported splice effect was confirmed for four, and for the other two variants, RNA analysis was not possible because either RNA isolation from FFPE tissue failed (no expression of HKG) or the affected gene (i.e., *MLH1*) did not show expression in the presence of positive HKG expression. In all four confirmed splice effects, no WT product was identified. For two variants (*MLH1* c.454_545del and *BRCA1* c.213-12 A>G), this result was expected because of the high variant allele frequencies (100% and 93%, respectively). The *APC* c.1548G>A and *MSH6* c.3801+1_3801+5del had variant allele frequencies of 28% and 50%, respectively, but only aberrant product was detected, which could be due to preferential amplification of the smaller (aberrant) product or possible FFPE-induced RNA degradation. For the *APC* variant, the F2/R2 primers that amplified the boundary of exon 11-exon 12 (with a part of exon 11) produced a product, which showed that exon 11 was part of the cDNA.

For two variants, *MLH1* c.791-1G>C and *MSH2* c.1511-2A>G, RNA was isolated from EBV-transformed cell lines because FFPE tissue was not available. Both previously resulted in an (partial) exon skip,[4, 28-30] which changed the reading frame and led to a premature stop codon; however, the aberrant splicing was only confirmed for one of the two variants (*MLH1* c.791-1G>C), whereas only WT transcript was detected for the *MSH2* c.1511-2A>G. The EBV-transformed cells were cultured without NMD inhibitors, and we hypothesized that the aberrant MSH2 RNA was possibly degraded through nonsense-mediated decay (NMD). NMD of the *MSH2* c.1511-2A>G RNA and no NMD of the *MLH1* c.791-1G>C RNA is consistent with previous studies in which NMD-inhibitors were omitted.[31, 32] Notably, the aberrant RNA was detected after formalin fixation of the EBV cells carrying the *MSH2* c.1511-2A>G, which is consistent with our hypothesis that aberrant RNA in FFPE tissue can be detected because formalin fixation prevents the degradation of RNA.

For the five variants predicted to affect the canonical splice site, three resulted in aberrant splicing, showing the predicted splice effect. For the other two, in one (*MLH1* c.2104-1G>C), only the WT transcript was detected, and in one, no expression of the affected gene (*MLH1*) was detected in the presence of normal HKG expression. The *APC* c.834+2T>A and *APC* c.1959-1G>A showed WT and aberrant product, which can be explained by the variant allele frequencies (12% and 40%, respectively). The *BRCA1* c.212+3A>T had a variant allele frequency of 54%, although only aberrant product was detected, which was possibly due to preferential amplification of the smaller (aberrant) product or because no RNA expression of the WT allele occurred. However, FFPE-related RNA degradation could not be excluded. With the detection of aberrant RNA in three of four variants predicted to affect splicing, this assay confirmed that *in silico* splice prediction tools are reliable in their predictions, particularly for variants disrupting the canonical splice site, although experimental analysis is required.[4, 7, 33] The three predicted splice variants shown to result in aberrant splicing in this study are currently classified as variants of uncertain significance, but the RNA analysis in this study supports their pathogenic effect.

FFPE blocks were collected from the archives and were embedded between 2000 and 2016, with most (n=9) embedded after 2009. Notably, from the two samples embedded before 2009, no RNA was successfully isolated from the FFPE block, indicating that isolating RNA from FFPE tissue might not be possible in older blocks. From eight of nine blocks embedded after 2009, RNA was successfully isolated, showing WT and/or aberrant RNA, and combined with the formalin fixation results from the cell lines, showed that formalin fixation possibly inhibited RNA degradation. With analysis of RNA from FFPE tissues, the splicing effect of somatic variants that are only present in the tumor can be analysed. Furthermore, available material can be retrospectively analysed, without having to request RNA from leukocyte DNA, which is not always available.

However, a few limitations of this study deserve discussion. First, RNA from FFPE tissue blocks is often degraded to fragments less than ~300 bases in length.[34] To analyse this RNA, small amplicons must be designed instead of large amplicons containing multiple exons, as is preferred in leukocyte RNA testing. The design of the primers is very specific for the variant and is based on splicing prediction software. Although these algorithms are described as accurate for variants in the canonical splice site,[4, 7, 8] aberrant products that are not predicted and fall outside the amplification window of the assay will not be detected. Second, when an aberrant product is not detected, poor RNA quality, a wrongly predicted effect or no splice effect of the variant can be implicated, but no expression or RNA degradation of the mutant product by NMD can also occur. Therefore, the assay can confirm aberrant splicing, but from negative assay results, one cannot accurately conclude that no aberrant splicing occurs or that lack of aberrant product is due to other factors. Negative results should be confirmed with other methods, such as a minigene splicing assay.[4, 9] Further evidence of FFPE-based RNA analysis should be obtained from a larger study with more samples and variants. Targeted RNA-seq of RNA isolated from FFPE tissue might enable high-throughput analysis of somatic splice site variants.

RNA analysis on total nucleic acid isolated from FFPE tissue blocks is a valuable tool for the fast and easy detection of aberrant splicing, offering additional support for the pathogenicity of a (predicted) splice variant. With this assay, we correctly showed the splice effect of six known splice site variants and showed the splice effect of three variants predicted to affect splicing. This assay can be used to analyse somatic variants found in FFPE tumor tissue, with formalin fixation possibly inhibiting RNA degradation, and can be easily implemented in current molecular tumor diagnostics to help classify the high number of variants of uncertain significance currently found with high-throughput sequencing.

Supplementary information accompanies this paper on European Journal of Human Genetics' website (*http://www.nature.com/ejhg*).

Acknowledgements
The authors would like to acknowledge Marina Ventayol Garcia for technical assistance.

References

1. Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 2002;3:285-98.
2. Brillen AL, Schoneweis K, Walotka L, et al. Succession of splicing regulatory elements determines cryptic 5'ss functionality. Nucleic Acids Res 2016.
3. Wang GS, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. Nat Rev Genet 2007;8:749-61.
4. van der Klift HM, Jansen AM, van der Steenstraten N, et al. Splicing analysis for exonic and intronic mismatch repair gene variants associated with Lynch syndrome confirms high concordance between minigene assays and patient RNA analyses. Mol Genet Genomic Med 2015;3:327-45.
5. Hentze MW, Kulozik AE. A perfect message: RNA surveillance and nonsense-mediated decay. Cell 1999;96:307-10.
6. Whiley PJ, de la Hoya M, Thomassen M, et al. Comparison of mRNA splicing assay protocols across multiple laboratories: recommendations for best practice in standardized clinical testing. Clin Chem 2014;60:341-52.
7. Vreeswijk MP, Kraan JN, van der Klift HM, et al. Intronic variants in BRCA1 and BRCA2 that affect RNA splicing can be reliably selected by splice-site prediction programs. Hum Mutat 2009;30:107-14.
8. Spurdle AB, Couch FJ, Hogervorst FB, et al. Prediction and assessment of splicing alterations: implications for clinical testing. Hum Mutat 2008;29:1304-13.
9. Cooper TA. Use of minigene systems to dissect alternative splicing elements. Methods 2005;37:331-40.
10. Hedegaard J, Thorsen K, Lund MK, et al. Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. PLoS One 2014;9:e98187.
11. Gouveia GR, Ferreira SC, Ferreira JE, et al. Comparison of two methods of RNA extraction from formalin-fixed paraffin-embedded tissue specimens. Biomed Res Int 2014;2014:151724.
12. Lynch HT, de la Chapelle A. Hereditary colorectal cancer. N Engl J Med 2003;348:919-32.
13. Mensenkamp AR, Vogelaar IP, van Zelst-Stams WA, et al. Somatic mutations in MLH1 and MSH2 are a frequent cause of mismatch-repair deficiency in Lynch syndrome-like tumors. Gastroenterology 2014;146:643-646 e8.
14. Geurts-Giele WR, Leenen CH, Dubbink HJ, et al. Somatic aberrations of mismatch repair genes as a cause of microsatellite-unstable cancers. J Pathol 2014;234:548-59.
15. Kuismanen SA, Holmberg MT, Salovaara R, et al. Genetic and Epigenetic Modification of MLH1 Accounts for a Major Share of Microsatellite-Unstable Colorectal Cancers. The American Journal of Pathology 2000;156:1773-1779.
16. Powell SM, Petersen GM, Krush AJ, et al. Molecular diagnosis of familial adenomatous polyposis. N Engl J Med 1993;329:1982-7.
17. Groden J, Thliveris A, Samowitz W, et al. Identification and characterization of the familial adenomatous polyposis coli gene. Cell 1991;66:589-600.
18. Spier I, Drichel D, Kerick M, et al. Low-level APC mutational mosaicism is the underlying cause in a substantial fraction of unexplained colorectal adenomatous polyposis cases. J Med Genet 2016;53:172-9.
19. Yamaguchi K, Komura M, Yamaguchi R, et al. Detection of APC mosaicism by next-generation sequencing in an FAP patient. J Hum Genet 2015;60:227-31.
20. Out AA, van Minderhout IJ, van der Stoep N, et al. High-resolution melting (HRM) re-analysis of a polyposis patients cohort reveals previously undetected heterozygous and mosaic APC gene mutations. Fam Cancer 2015;14:247-57.
21. Jansen AM, Crobach S, Geurts-Giele WR, et al. Distinct Patterns of Somatic Mosaicism in the APC Gene in Neoplasms From Patients With Unexplained Adenomatous Polyposis. Gastroenterology 2017;152:546-549 e3.
22. Hall JM, Lee MK, Newman B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. Science 1990;250:1684-9.
23. Rahman N, Stratton MR. The genetics of breast cancer susceptibility. Annu Rev Genet 1998;32:95-121.
24. Mafficini A, Simbolo M, Parisi A, et al. BRCA somatic and germline mutation detection in paraffin embedded ovarian cancers by next-generation sequencing. Oncotarget 2016;7:1076-83.
25. Koczkowska M, Zuk M, Gorczynski A, et al. Detection of somatic BRCA1/2 mutations in ovarian cancer - next-generation sequencing analysis of 100 cases. Cancer Med 2016;5:1640-6.
26. Jansen AM, van Wezel T, van den Akker BE, et al. Combined mismatch repair and POLE/POLD1 defects explain unresolved suspected Lynch syndrome cancers. Eur J Hum Genet 2016;24:1089-92.
27. van Eijk R, Stevens L, Morreau H, et al. Assessment of a fully automated high-throughput DNA extraction method from formalin-fixed, paraffin-embedded tissue for KRAS, and BRAF somatic mutation analysis. Exp

Mol Pathol 2013;94:121-5.
28. Wijnen J, Khan PM, Vasen H, et al. Majority of hMLH1 mutations responsible for hereditary nonpolyposis colorectal cancer cluster at the exonic region 15-16. American Journal of Human Genetics 1996;58:300-307.
29. Wijnen J, Khan PM, Vasen H, et al. Hereditary nonpolyposis colorectal cancer families not complying with the Amsterdam criteria show extremely low frequency of mismatch-repair-gene mutations. Am J Hum Genet 1997;61:329-35.
30. Casey G, Lindor NM, Papadopoulos N, et al. Conversion analysis for mutation detection in MLH1 and MSH2 in patients with colorectal cancer. Jama 2005;293:799-809.
31. Bianchi F, Rosati S, Belvederesi L, et al. MSH2 splice site mutation and endometrial cancer. Int J Gynecol Cancer 2006;16:1419-23.
32. Thompson BA, Goldgar DE, Paterson C, et al. A multifactorial likelihood model for MMR gene variant classification incorporating probabilities based on sequence bioinformatics and tumor characteristics: a report from the Colon Cancer Family Registry. Hum Mutat 2013;34:200-9.
33. Colombo M, De Vecchi G, Caleca L, et al. Comparative in vitro and in silico analyses of variants in splicing regions of BRCA1 and BRCA2 genes and characterization of novel pathogenic mutations. PLoS One 2013;8:e57173.
34. Cronin M, Pho M, Dutta D, et al. Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. Am J Pathol 2004;164:35-42.

Supplementary Table 2: Primers

Variant	Forward primers		Reverse Primers	
MLH1 c.454_545del	F1 F2	GATGGAAAACTGAAAGCCCCT TTTTACAACATAGCCACGAG	R1 R2	GCCAACAACTTCCAAAATTTTCC GGTTGAGGCATTGGGTAGTG
MLH1 c.791-1G>C/ c.793C>T	F1 F2	TGGATGTGAGGATAAAACCC AGAAACAGTGTATGCAGCCT	R1 R2	AGGTACAGGAATGGGTGTGT CTGGAGGAATTGGAGCCCA
MLH1 c.885-2A>G	F1	AGAAACAGTGTATGCAGCCT	R1	CTGGAGGAATTGGAGCCCA
MLH1 c.1731G>A	F1	ATACCTTCTCAACACCACCA	R1	CATCTCAGCCTTCTTCTTCA
MLH1 c.2104-1G>C	F1	AAAGCCTCAGTAAAGAATGC	R1	TGTGTTCCACAGTCCACTTC
MSH2 c.1511-2A>G	F1	CCTTGTAAAACCTTCATTTGATCC	R1	CGAAGGACTTTTTCTTCCTTACA
MSH6 c.3801+1_ 3801+5del	F1 F2	GTGAAACTGCCAGCATACTCA CATTATTTTCAACTCACTACCAT	R1 R2	CATATGTCCTAGGCGCACAG TGGGAGATTAGCAAGCCTTG
APC c.834+2T>A	F1	GGTCATCTCAGAACAAGCAT	R1	CAGATGACTTGTCAGCCTTC
APC c.1548G>A	F1 F2	TCCTGCTGTGTGTGTTCTAA TTATTGCAAGTGGACTGTGA	R1 R2	CAAGTTTGTCAAAGCCATTC TAGTTGGGCCACAAGTGC
APC c.1959-1G>A	F1	AGCCAGACAAACACTTTAGC	R1	GATTCCACAAAGTTCCACAT
BRCA1 c.212+3A>T c.213-12A>G	F1	ATGCTGAAACTTCTCAACCA	R1	AACCTGTGTCAAGCTGAAAA

Supplementary Table 1: Splice variants tested

Gene	Variant	Protein	RNA effect	(Predicted) Effect on ss and/or ESE motifs	% decrease of can. ss strength	RNA effect previously described in:
			A. variants known to result in splicing			
MLH1	c.454_545del	p.(E153Ffs*8)	Skip exon 6	NA	NA	NA
MLH1	c.791-1G>C	p.(H264Lfs*2)	Skip exon 10	Loss of canonical 3' ss	(100, 100, 100, 100)	Wijnen et al, Am. J. Hum. Genet. 1996 & van der Klift et al, Mol Genet Genomic Med, 2015
MLH1	c.793C>T	p.(R265C)	Skip exon 10 (major), WT (minor)	Increase of 1 ESE	NA	van der Klift et al, Mol Genet Genomic Med, 2015
MLH1	c.1731G>A	p.(S556Rfs*14)	Skip exon 15	Decrease in canonical 5' ss, Loss of 1 ESE	(14, 23, 10, 11)	Tournier et al, Hum Mutat, 2008 & Auclair et al, Hum Mutat, 2006
MSH2	c.1511-2A>G	p.(D506Gfs*7)	Last nt of intron 9 included	Loss of canonical 3' ss	(100, 100, 100, 100)	Wijnen et al, Am. J. Hum. Genet. 1997 & Casey et al, JAMA, 2005
MSH6	c.3801+1_3801+5del	p.(R1217Mfs*6)	Skip exon 8	Loss of canonical 5' ss	(100, 100, 100, 24)	LOVD database, ID: MSH6_00908
APC	c.1548G>A	p.(G471Yfs*19)	Skip exon 11	Decrease in canonical 5' ss, loss SR940 site	(14,30, 20, 12)	Kaufmann et al, J Mol Diagn, 2009
BRCA1	c.213-12A>G	p.(?)	Insertion seq. corr. to c.213-12_c.213-1	Loss of canonical 3' ss	(100, 75, 100, 0)	Hoffman et al, Am J Med Genet, 1998
			B. variants predicted to result in splicing			
MLH1	c.885-2A>G	p.(?)	Skip 5nt of exon 11*	Loss of canonical 3' ss	(100, 100, 100, 100)	-
MLH1	c.2104-1G>C	p.(?)	Skip 2nt of exon 19*	Loss of canonical 3' ss	(100, 100, 100, 100)	-
APC	c.834+2T>A	p.(?)	Skip 5nt of exon 7*	Loss of canonical 5' ss	(100, 100, 100, 100)	-
APC	c.1959-1G>A	p.(?)	Skip 1nt of exon 15*	Loss of canonical 3' ss	(100, 100, 100, 100)	-
BRCA1	c.212+3A>T	p.(?)	Skip 22nt of exon 5*	Decrease in canonical 5' ss	(100, 41, 100, 18)	

*ss = splice site, Nt = nucleotides, NA = not applicable *predicted RNA effect is based on nearest cryptic splice site, which could possibly act as a new 3' or 5' splice site. Splice prediction programs SFF (Splice Site Finder), MES (MaxEntScan), NNS (NNSPLICE) and HSF (Human Splicing Finder) were used for splice effect prediction.*

Chapter 5

The complexity of screening *PMS2* in DNA isolated from formalin-fixed paraffin-embedded tissue

AML Jansen[1,2], CMJ Tops[3], D Ruano[1], R van Eijk[1],
JT Wijnen[3], FJ Hes[3], T van Wezel[1], H Morreau[1]

In submission

Department of [1]Pathology, [2]Human Genetics and [3]Clinical Genetics,
Leiden University Medical Center, Leiden, The Netherlands

Abstract

Background and Aims: Germline variants in the DNA mismatch repair (MMR) gene *PMS2* represent 1-14% of all MMR gene variants. Correct variant analysis of *PMS2* is complex due to the presence of multiple pseudogenes and the occurrence of gene conversion. The complexity of analysis increases in highly fragmented DNA from formalin-fixed paraffin-embedded (FFPE) tissue. We now describe and test a reliable approach to detect true *PMS2* variants in fragmented DNA.

Methods: A custom NGS panel meant for FFPE tissue was used targeting four MMR genes, *POLE* and *POLD1*. Amplicon design for *PMS2* was based on the position of paralogous sequence variants (PSV) that distinguish *PMS2* from its pseudogenes. We screened 125 MMR-deficient tumors for *PMS2* variants.

Results: We present an overview of PSVs that can be used for reliable distinction between *PMS2* and its pseudogenes. *PMS2* variants in exons 1-11 can be correctly curated on basis of this information. For exons 12-15 this is less reliable as these undergo gene conversion. Of the 125 tumors tested, six were unexplained MMR-deficient tumors with solitary PMS2 protein expression loss. In these six tumors three unclassified variants (class 3) and four (likely) pathogenic variants (class 4 and 5) were detected in *PMS2*. One microsatellite unstable tumor with positive staining for all MMR proteins was found to carry a frameshift *PMS2* variant. No pathogenic *PMS2* variants were detected in tumors with other patterns of MMR protein expression loss.

Conclusions: With a paired-end NGS approach with one or two PSVs in every amplicon, variants can reliably be detected in exons 1 to 11 of *PMS2*.

Keywords: *PMS2,* variant, next-generation sequencing, paralogous sequence variant.

Introduction

Pathogenic heterozygous germline variants in the MMR genes cause Lynch Syndrome (LS), an autosomal dominant predisposition for colorectal-, endometrial- and other cancers.[1] While the majority of the causal variants are found in *MLH1* and *MSH2*, variants in the less frequent mutated *PMS2* represent 1-14% of all MMR gene variants.[2, 3] The colorectal cancer (CRC) risk of *PMS2* variant carriers has shown to be much lower compared to *MLH1, MSH2* and *MSH6,* with risk of CRC around 11-19% by the age of 70 years.[4] Homozygous or compound heterozygous variants in the *PMS2* gene are seen more often in patients with constitutional mismatch repair deficiency (CMMRD), a recessive disorder characterized by CRC and childhood hematological- and brain malignancies.[5]

The analysis of *PMS2* is complex due to the presence of multiple pseudogenes.[3, 6, 7] Fourteen *PMS2*-pseudogenes share a high homology with the 5' end of *PMS2* (exon 1 to 5), while a fifteenth pseudogene (*PMS2CL*) shares high homology with *PMS2* exon 9 and exon 11 to 15.[2, 7-10] An additional complexity is added due to ongoing gene conversion events between *PMS2* and *PMS2CL*.[10] Germline variant screening strategies propose long-range PCR with a reverse primer in *PMS2* exon 6 or propose designing multiplex ligation-dependent amplification (MLPA) probes, and PCR primers based on paralogous sequence variants (PSVs) to distinguish *PMS2* exon 1 to 5 from the fourteen homologous pseudogenes.[2, 9, 11] These PSVs are specific nucleotides that differ between *PMS2* and the pseudogenes, and enable differentiation between two almost complete homologues sequences.[3, 8, 9] This strategy is not reliable in detecting variants in exon 12 to 15 due to gene conversion events between *PMS2* and *PMS2CL*.[10, 12] Through crossover the sequence corresponding to *PMS2* or *PMS2CL* could be present as the exon 12 to 15 sequence of *PMS2,* and subsequently expressed.[9-11] To determine which sequence is present, and expressed, long-range PCR on gDNA or cDNA is proposed using primers in the unique exon 10 and a nonspecific reverse primer in the 3' UTR.[9, 10, 12-14]

While this strategy is very suitable for reliable detection of *PMS2* variants in leukocyte DNA, it is not applicable when using DNA isolated from formalin-fixed paraffin-embedded (FFPE) tissue blocks, which is highly fragmented.[15] It has been recently shown that in a large proportion of MMR-deficient tumors without pathogenic germline MMR variant and without *MLH1* promoter hypermethylation, two somatic MMR variants can explain the MMR-deficiency.[16-18] This occurrence of somatic MMR inactivation also shows the need for reliable detection of somatic *PMS2* variants in DNA isolated from FFPE tissue. Testing DNA isolated from FFPE furthermore enables detection of variants in (deceased) index patients of which only FFPE is available. Furthermore, to implement reliable *PMS2* variant screening in molecular tumor diagnostics, a high-throughput strategy should be developed.

Some studies only focus on screening for variants in *MLH1, MSH2* and *MSH6,* possibly because of the complexity of screening for true *PMS2* variants.[16] We now describe possible pitfalls in *PMS2* variant detection and a next-generation sequencing (NGS)-based approach for reliable somatic and germline *PMS2* testing in FFPE DNA.

Materials and Methods

Study Cohort

Colorectal and endometrial cancers and when available matching normal tissue of 40 patients with unexplained MMR-deficiency was screened for DNA variants with NGS to detect variants in the MMR genes in a diagnostic setting. All tumors were pre-screened with immunohistochemical (IHC) staining of the four MMR proteins and the majority (83%) showed expression loss of one or two of the MMR proteins. Many of these tumors were previously screened for microsatellite instability and showed to have high microsatellite instability (MSI-H). All MLH1/PMS2 negative tumors were tested for *MLH1* promoter hypermethylation, and somatic NGS was performed if no methylation was detected. Four tumors had solitary immunohistochemical expression loss of PMS2. Additionally, DNA isolated from FFPE tissue blocks of 85 unexplained suspected Lynch Syndrome patients (without germline MMR variants and without *MLH1* promoter hypermethylation) were screened with NGS for variants in the MMR genes in a research setting. Two of the latter tumors showed isolated PMS2 expression loss with IHC.

NGS panel

A custom paired end NGS library was designed covering *MLH1, MSH2, MSH6, PMS2, POLE* and *POLD1*. Ion AmpliSeq™ Custom Panels were designed with the Ion AmpliSeq™ Designer tool. Libraries were prepared with Ion AmpliSeq™ Library Kit 2.0 according to the manufacturer's protocol. The panel used in a diagnostic setting slightly differs from the research panel. The diagnostic panel covers the exonic regions with 99.2% coverage of *MLH1*, 99.3% coverage of *MSH2*, 100% coverage of *MSH6* and 76.5% of *PMS2* (exon 1-12) and the exonuclease domain of *POLE* (exon 7-14) and *POLD1* (exon 8-13). The research panel is comparable but covers 100% of *MLH1,* 94.9% of *MSH2,* 97.7% of *MSH6,* 79.1% of *PMS2* (exons 1-11 and exon 14) and *POLE* and *POLD1* completely. Next-generation sequencing was performed with the Ion Proton™ System (Life Technologies, Carlsbad, CA, USA).

NGS annotation

Raw data analysis, alignments, and variant calling was carried out using the default parameters in Torrent Suite. The unaligned BAM files generated by the Proton sequencer were mapped against the human reference genome (GRCh37/hg19) using the TMAP 5.0.7 software with default parameters (*https://github.com/iontorrent/TS*). A read is assigned to the genomic location with the highest mapping score. In case that a particular read gets the same alignment score at multiple locations, it will be randomly assigned to one of the loci. All (likely) pathogenic *PMS2* variants were visually inspected with the Integrative Genomics Viewer (IGV).[19, 20] The following Genbank reference sequences were used: NM_000249.3 for *MLH1*, NM_000251.2 for *MSH2*, NM_000179.2 for *MSH6*, NM_000535.5 for *PMS2*, NM_006231.2 for *POLE* and NM_001256849.1 for *POLD1*. Classification of the functional effects of the variants was done according to the five-tiered InSiGHT scheme.[21]

Results

A custom paired-end MMR panel was designed for detecting variants in DNA isolated from formalin-fixed paraffin-embedded (FFPE) tissue. On average, amplicons are 100-175 bp in size in order to be able to amplify fragmented DNA. For *PMS2*, a reliable screening panel could be made covering exons 1-11 only, with one or two paralogous sequence variants (PSV) in every amplicon. With exons 1-11 as target, a 72.9% coverage can be achieved. An overview of PSVs present in these eleven exons is shown in Table 1.

Table 1: paralogous sequence variants (PSVs) in *PMS2*

Target	Size (bp)	Amplicons needed	# pseudogenes	PSV
Exon 1	23	1	5	c.-13G>C (4/5) and c.-9G>A (1/5) c.1A>T (4/5) and c.-4_-5delinsAG (1/5)
Exon 2	140	1 or 2	13	c.24-4C>T (13/13) c.89A>C (11/13) c.117A>G (13/13) c.121G>A (9/13) c.125T>A (9/13)
Exon 3	87	1 or 2	14	c.164-12delT (3/14) and c.209A>G (11/14) c.187G>A (10/14) c.195T>C (14/14) c.240C>T (3/14) and c.250+8G>A (11/14)
Exon 4	103	2	14	c.251-11C>G (13/14) and c.251-13C>T (1/14) c.299A>G (10/14) and c.298C>G (3/14) c.353+22C>T (14/14)
Exon 5	184	2	14	c.396A>G, c.406A>G, c.418G>T, c.429T>C, c.452G>C, c.478C>A, c.492C>T (all 14/14)
Exon 6	168	2	0	-
Exon 7	98	1 or 2	0	-
Exon 8	100	1 or 2	0	-
Exon 9	85	1 or 2	1 (PMS2CL)	c.924G>C, c.932A>G and c.934A>G
Exon 10	156	2	0	-
Exon 11	862	8 or 9	1 (PMS2CL)	c.1238_1239delAAinsGG, c.1360_1361delCTinsTC, c.1379G>A, c.1556A>G, c.1559C>T, c.1567T>A, c.1688_1689delGAinsAG, c.1714G>A, c.1717A>T, c.1730dupA, c.1732C>T, c.1740A>G, c.1760G>A, c.1771T>C, c.1795G>A, c.1798A>G, c.1855G>A, c.1863_1864delTA, c.1952A>G, c.2006+26C>A and c.2006+28delC

Overview of paralogous sequence variants (PSVs) in PMS2, #/#; number of pseudogenes with the variant/total number of pseudogenes for this exon.

Using the strategy described above and in the Material and Methods, 125 MMR-deficient tumors and, when available matching normal colonic mucosa, were screened for *PMS2* variants. Six tumors showed solitary immunohistochemical PMS2 expression loss. Four (likely) pathogenic *PMS2* variants (class 4 or 5), and three variants of uncertain significance (VUS class 3) were detected (Figure 1A and Table 2). The class 3 *PMS2* c.308C>T and c.1687C>T variants were both found in tumors with a variant in the exonuclease domain of *POLE*, where the *PMS2* variant is expected to be secondary to the *POLE* variant.[22] Additionally, one tumor with positive staining for all MMR proteins and an MSI-H phenotype was found to carry a frameshift *PMS2* c.325dupG variant. In remaining cases with combined MLH1/PMS2, combined MSH2/MSH6 or solitary MSH6 expression loss no pathogenic *PMS2* variant was detected.

For all variants the IGV was used to determine the presence of PSVs. An example is shown in Figure 1B for an exon 9 *PMS2* c.955C>A variant. The NGS amplicon containing the *PMS2* c.955C>A also contains three PSVs, c.934A>G, c.932A>G and c.924G>C. None of the reads showed any of these three PSVs, indicating that this variant is truly present in *PMS2* and not in the pseudogene *PMS2CL*. This was done for all eight *PMS2* variants shown in Table 2, and all variants were found to be present in *PMS2* and not one of the pseudogenes.

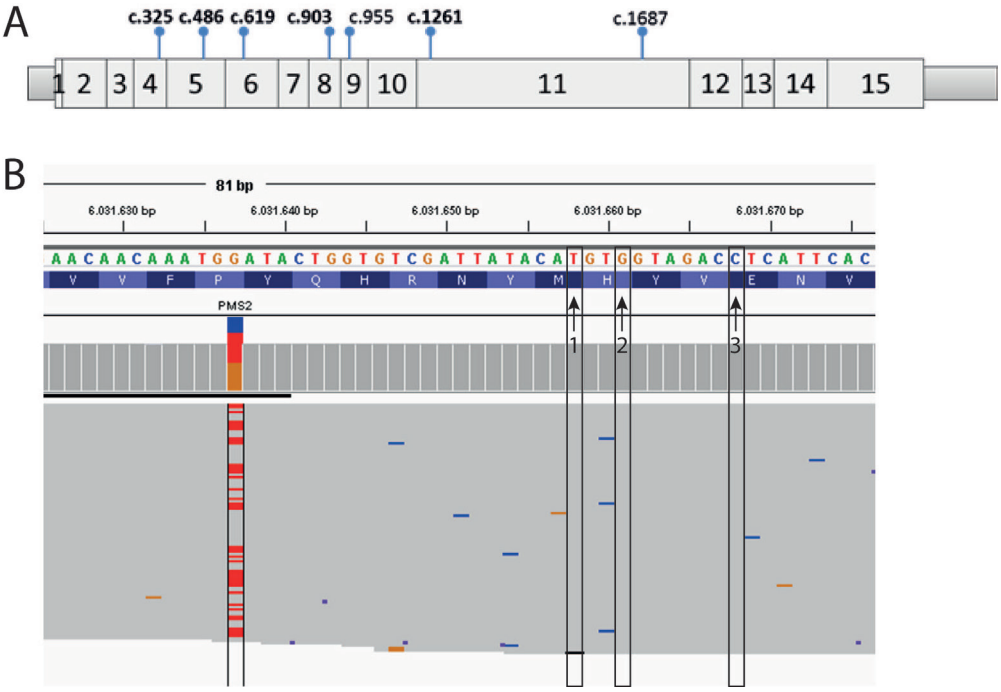


Figure 1: *PMS2* variants detected with NGS

[A] PMS2 variants found with NGS, Class 4 and class 5 variants shown in bold. [B] IGV printout of the PMS2 c.955C>A shown (in red). Arrows show the location of three PSVs present in the amplicon (1. c.934A>G, 2. c.932A>G and 3. c.924G>C). All three are absent in the reads, indicating that this variant is present in PMS2 and not one of the pseudogenes. PMS2 is shown in reverse complement, because PMS2 is present on the reverse strand.

Table 2: Overview PMS2 variants

Tumor characteristics	Variant	Exon	%	Class	PSVs in amplicon
CRC45, PMS2-, MSI-H	*PMS2* c.308C>T†	4	11%	3	c.299, c.298
CRC38, MMR+, MSI-H	*PMS2* c.325dupG	4	77%	4	c.299, c.298
CRC31, PMS2-, MSI-H	*PMS2* c.486delA	5	34%	4	c.406, c.418, c.429, c.452, c.478, c.492
CRC48, PMS2-, MSI unknown	*PMS2* c.619G>T	6	48%	4	NA
CRC67, PMS2-, MSI-H	*PMS2* c.903G>T	8	52%	4	NA
	PMS2 c.1261C>T	11	26%	5	c.1238_1239, c.1360_1361*
EC58, PMS2-, MSI unknown	*PMS2* c.955C>A	9	41%	3	c.924, c.932, c.934
EC55, PMS2-, MSI-H	*PMS2* c.1687C>T†	11	30%	3	c.1556, c.1559, c.1567, c.1688_1689

*Tumor is shown as type of tumor, followed by age of onset; CRC, colorectal cancer; EC, endometrial cancer; MSI, microsatellite instability, high (H) or unknown; PMS2-, PMS2 negative staining; MMR+, positive IHC MMR staining; % is variant allele frequency; NA, not applicable, this exon is unique; *, PSV is present in primer sequence; †variant present in POLE mutated tumor.*

Discussion

We now describe how to interpret *PMS2* variants present in DNA isolated from formalin-fixed paraffin-embedded (FFPE) tissue using paired-end targeted NGS with paralogous sequence variants (PSVs) in every amplicon. Six of the eight *PMS2* variants detected were located in exons that have high homology with one or more of the *PMS2* pseudogenes. Of all variants it could be concluded that they were truly present in *PMS2* and not in the pseudogenes by analysing the presence of PSVs in the amplicon (Table 2). Through automatically assigning the read to the genomic location with the highest mapping score, a reliable distinction could be made between *PMS2* and the pseudogenes, for *PMS2* exon 1 to 11. For exons 12 to 15 of *PMS2* it is not possible to reliably detect variants in FFPE derived DNA due to the existence of continuous gene conversion targeting these exons. The only solution to this challenge is long range PCR of fragments covering *PMS2* exons 12-15, but this is not applicable on the fragmented FFPE derived tissue DNA.[9, 10]

Studies that aim to detect *PMS2* variants in DNA from FFPE tissues are very limited. Only five studies describe somatic analysis of *PMS2*.[17, 18, 22-24] We and others achieve a coverage of 75-80% and do not sequence *PMS2* exon 12 to 15 completely, because variants cannot be curated in this region due to sequence exchange events. One previous study suggested full sequencing (100%) coverage of *PMS2* in tumor tissue, but did not fully explain how was coped with gene conversion of exons 12 to 15 (*http://tests.labmed.washington.edu/COLOSEQ#Introducing_ColoSeq.2BISI_Tumor*).[24] One *PMS2* splice site variant in intron 12 was shown without confirmation of its presence in *PMS2* and not in *PMS2CL* through gene conversion (previously shown to occur in 69% of tested individuals).[10] This example typically highlights the existing problem with sequencing of *PMS2* exons 12 to 15. Consensus should be reached whether it is preferable to test patients for variants in this region, without being able to confirm that the variant is expressed, or to not sequence this region and possibly missing somatic *PMS2* variants. In the currently described research panel exons 12 and exon 14 are included, but caution is needed when analysing the variants, since it cannot be established whether this variant is expressed. It could be considered that when a (likely) pathogenic *PMS2* exon 12 to 15 variant is detected in a tumor with solitary PMS2 loss of expression with no other *PMS2* variants, this variant is likely expressed and the cause of the immunohistochemical loss of PMS2 expression. Additionally, since expressed genes have elevated mutation rates, if a somatic variant is detected in *PMS2* exon 12-15 it is likely that *PMS2* is expressed.[25] However, it cannot be confirmed whether *PMS2* is truly expressed.

In conclusion, with a custom NGS panel with one or two PSVs we were able to reliably detect eight variants in *PMS2* exon 1 to 11 in six tumors with solitary PMS2 loss, and one tumor with positive MMR staining and microsatellite instability. Previous studies describe comprehensive strategies for accurate mutation detecting in *PMS2,* but mainly focus on testing genomic DNA extracted from blood.[9, 26] However, since recent studies have shown biallelic somatic inactivation of the MMR genes, there is a growing need for reliable detection of somatic variants in *PMS2*.[16-18, 22] With this guide we show a reliable method to detect *PMS2* variants in DNA from FFPE tissue.

References

1. Lynch HT, de la Chapelle A. Hereditary colorectal cancer. N Engl J Med 2003;348:919-32.
2. Borras E, Pineda M, Cadinanos J, et al. Refining the role of PMS2 in Lynch syndrome: germline mutational analysis improved by comprehensive assessment of variants. J Med Genet 2013;50:552-63.
3. Hendriks YM, Jagmohan-Changur S, van der Klift HM, et al. Heterozygous mutations in PMS2 cause hereditary nonpolyposis colorectal carcinoma (Lynch syndrome). Gastroenterology 2006;130:312-22.
4. ten Broeke SW, Brohet RM, Tops CM, et al. Lynch syndrome caused by germline PMS2 mutations: delineating the cancer risk. J Clin Oncol 2015;33:319-25.
5. Vasen HF, Ghorbanoghli Z, Bourdeaut F, et al. Guidelines for surveillance of individuals with constitutional mismatch repair-deficiency proposed by the European Consortium "Care for CMMR-D" (C4CMMR-D). J Med Genet 2014;51:283-93.
6. Kondo E, Horii A, Fukushige S. The human PMS2L proteins do not interact with hMLH1, a major DNA mismatch repair protein. J Biochem 1999;125:818-25.
7. Nicolaides NC, Carter KC, Shell BK, et al. Genomic organization of the human PMS2 gene family. Genomics 1995;30:195-206.
8. De Vos M, Hayward BE, Picton S, et al. Novel PMS2 pseudogenes can conceal recessive mutations causing a distinctive childhood cancer syndrome. Am J Hum Genet 2004;74:954-64.
9. van der Klift HM, Mensenkamp AR, Drost M, et al. Comprehensive Mutation Analysis of PMS2 in a Large Cohort of Probands Suspected of Lynch Syndrome or Constitutional Mismatch Repair Deficiency Syndrome. Hum Mutat 2016;37:1162-1179.
10. van der Klift HM, Tops CM, Bik EC, et al. Quantification of sequence exchange events between PMS2 and PMS2CL provides a basis for improved mutation scanning of Lynch syndrome patients. Hum Mutat 2010;31:578-87.
11. Wernstedt A, Valtorta E, Armelao F, et al. Improved multiplex ligation-dependent probe amplification analysis identifies a deleterious PMS2 allele generated by recombination with crossover between PMS2 and PMS2CL. Genes Chromosomes Cancer 2012;51:819-31.
12. Hayward BE, De Vos M, Valleley EM, et al. Extensive gene conversion at the PMS2 DNA mismatch repair locus. Hum Mutat 2007;28:424-30.
13. Clendenning M, Walsh MD, Gelpi JB, et al. Detection of large scale 3' deletions in the PMS2 gene amongst Colon-CFR participants: have we been missing anything? Fam Cancer 2013;12:563-6.
14. Etzler J, Peyrl A, Zatkova A, et al. RNA-based mutation analysis identifies an unusual MSH6 splicing defect and circumvents PMS2 pseudogene interference. Hum Mutat 2008;29:299-305.
15. Daugaard I, Kjeldsen TE, Hager H, et al. The influence of DNA degradation in formalin-fixed, paraffin-embedded (FFPE) tissue on locus-specific methylation assessment by MS-HRM. Experimental and Molecular Pathology 2015;99:632-640.
16. Mensenkamp AR, Vogelaar IP, van Zelst-Stams WA, et al. Somatic mutations in MLH1 and MSH2 are a frequent cause of mismatch-repair deficiency in Lynch syndrome-like tumors. Gastroenterology 2014;146:643-646 e8.
17. Geurts-Giele WR, Leenen CH, Dubbink HJ, et al. Somatic aberrations of mismatch repair genes as a cause of microsatellite-unstable cancers. J Pathol 2014;234:548-59.
18. Jansen AM, Geilenkirchen MA, van Wezel T, et al. Whole Gene Capture Analysis of 15 CRC Susceptibility Genes in Suspected Lynch Syndrome Patients. PLoS One 2016;11:e0157381.
19. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. Nat Biotech 2011;29:24-26.
20. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics 2013;14:178-192.
21. Thompson BA, Spurdle AB, Plazzer JP, et al. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nat Genet 2014;46:107-15.
22. Jansen AM, van Wezel T, van den Akker BE, et al. Combined mismatch repair and POLE/POLD1 defects explain unresolved suspected Lynch syndrome cancers. Eur J Hum Genet 2015.
23. Stelloo E, Jansen AM, Osse EM, et al. Practical guidance for mismatch repair-deficiency testing in endometrial cancer. Ann Oncol 2016.
24. Haraldsdottir S, Hampel H, Tomsic J, et al. Colon and endometrial cancers with mismatch repair deficiency can arise from somatic, rather than germline, mutations. Gastroenterology 2014;147:1308-1316 e1.
25. Zhang J, Yang J-R. Determinants of the rate of protein sequence evolution. Nat Rev Genet 2015;16:409-420.
26. Li J, Dai H, Feng Y, et al. A Comprehensive Strategy for Accurate Mutation Detection of the Highly Homologous PMS2. J Mol Diagn 2015;17:545-53.

Chapter 6

Distinct patterns of somatic mosaicism in the APC gene in neoplasms from patients with unexplained adenomatous polyposis

AML Jansen[1,2], S Crobach[1], WRR Geurts-Giele[3], BEWM van den Akker[1], M Ventayol Garcia[1], D Ruano[1], M Nielsen[4], CMJ Tops[4], JT Wijnen[2,4], FJ Hes[4], T van Wezel[1], WNM Dinjens[3], H Morreau[1]

Gastroenterology, 2017; 152:546-549

Department of [1]Pathology, [2]Human Genetics, [4]Clinical Genetics,
Leiden University Medical Center, Leiden, The Netherlands
[3]Department of Pathology, Erasmus MC Cancer Institute, Rotterdam, The Netherlands

Abstract

To further unravel mechanisms of *APC* mosaicism in 27 genetically unexplained patients with colonic neoplasms, we used deep sequencing to analyze at least two adenomas or carcinomas per patient. Identical mosaic *APC* variants were identified in adenomas of 9/18 patients with 21-~100 adenomas. Mosaic variants were then variably detected in leukocyte DNA and normal mucosa or were confined to normal mucosa. In one patient comprehensive analysis found no evidence for the mosaic *APC* variant in normal mucosa. One patient was found to carry the mosaic *APC* variant in 10/16 adenomas, underlying the importance of screening ≥ two adenomas.

Keywords: unexplained polyposis; APC mosaicism

Main body

Somatic *APC* mosaicism is estimated to occur in 20% of *de novo* adenomatous polyposis cases.[1-3] Current methods to detect *APC* mosaicism are designed to test for germline variants in leukocyte DNA and the effectiveness of these methods depends on both method-specific sensitivity and the relative contribution of the mosaic variant to the tissue analyzed. Recently, a next-generation sequencing (NGS) study of at least two colonic adenomas in twenty patients found somatic pathogenic *APC* variants in five patients (25%).[4] Analysis of leukocyte DNA confirmed the mosaic *APC* variant in 4/5 cases, but no other tissues were analyzed.

With the aim of identifying *APC* mosaicism we analyzed at least two adenomas or carcinomas with NGS in six patients with 5-20 adenomas, 18 patients with 21-~100 adenomas, one patient with >1000 adenomas, two patients with multiple primary colorectal carcinomas and two positive controls.[5] Nine patients with 21-~100 adenomas (50%) and the two positive controls (APC-08 and APC-09), showed somatic mosaicism, with identical *APC* variants in adenomas tested (Table 1). No difference was observed in location (left/right) or distribution (segmental/whole colon) of the adenomas in patients with or without somatic *APC* mosaicism.

Leukocyte DNA was available for 10/11 of the mosaic patients, including one of the controls. Five of these previously tested negative for a germline mosaic *APC* variant (APC-03, APC-08, APC-09, APC-17 and APC-18, Supplementary Table 1). With NGS deep sequencing a pathogenic variant was identified in three of 10 leukocyte DNAs (APC-03, APC-17 and APC-18) at a 1-4% variant frequency, indicating the detection limit of previously used techniques. Initial testing of adenomas allows detection of very low frequency germline variants and mosaic variants confined to the colon, while initial leukocyte deep sequencing is shown to result in a high percentage false positives.[4] Human gastrulation, the process by which the three germ layers are established, is thought to occur at approximately day 16.[3, 6] The presence of a variant in both leukocyte DNA (mesoderm) and colon mucosa (endoderm) would imply the appearance of this variant before day 16. Primordial germ cells are thought to arise from the primary ectoderm during the second week of development. Therefore, the presence of a somatic variant in both leukocyte DNA and colon mucosa indicates that the variant arose early enough to potentially also be present in germ cells and therefore transmissible to the next generation.

In leukocyte DNA of six mosaic *APC* patients (APC-07, APC-09, APC-15, APC-21, APC-24 and APC-25), the *APC* variant could not be detected by deep sequencing. In four patients (APC-07, APC-09, APC-21 and APC-24), with available normal colonic mucosa, the mosaic *APC* variant was either not detected in normal mucosa (APC-07) or was present at a frequency of 2-29% (APC-09, APC-21 and APC-24, Table 1). In patient APC-24 the variant was present in one of the normal mucosa samples with a 2% frequency, while absent from the other sample, probably due to sampling bias. Variants only present in adenoma- and normal colonic DNA, but absent from leukocyte DNA , indicate a somatic event after the third week of embryogenesis and are unlikely present in the germ cells. For the other two mosaic patients without positivity in leukocyte DNA (APC-15 and APC-25), normal mucosa was not available. DNA from buccal mucosa, fibroblasts and urine from patient APC-09 and APC-21 tested negative for the mosaic *APC* variant.

Patient APC-04 and APC-08 were both found to carry an *APC* c.4666dupA variant in the adenomas initially sequenced with NGS. This variant is located in a homopolymeric nucleotide sequence and could not be reliably detected with the NGS technique alone.[7] However, the variant could be confirmed using the less sensitive HRMA and Sanger sequencing, with a 30-70% variant frequency in both adenomas of patient APC-08 and 10/16 tested adenomas of patient APC-04. In the six adenomas of patient APC-04 not carrying the c.4666dupA variant, other *APC* variants were detected with NGS (Supplementary Table 2, APC-04/P4-P9). The pathology report did not give the exact location of each adenoma and investigating possible clustering of adenomas with and without the *APC* c.4666dupA variant was therefore not possible. Nevertheless, this unusual pattern of adenomas with and without the *APC* mosaic variant illustrates the possible co-occurrence of sporadic adenomas within a mosaic environment, and highlights the importance of screening multiple adenomas. Additionally, leukocyte DNA of the children of patient APC-04, APC-07 and APC-21 was tested for the mosaic *APC* variant detected in the parent (Table 1), but none showed the mosaic variants. These findings are consistent with the *APC* variants occurring after primordial germ cell specification, as would be predicted by the absence of the *APC* variants in leukocyte DNA in APC-07 and APC-21.

In three patients (APC-03, APC-07 and APC-17) the mosaic pattern was subjected to a comprehensive analysis (Figure 1). Deep sequencing of normal mucosa samples identified an *APC* variant at frequencies between 0.4-29% (APC-03) and 0.8-15% (APC-17). In patient APC-07, the *APC* c.3340C>T variant was not found in DNA from the muscularis propria of the ileum, one draining lymph node of the colon, rectal mucosa, small bowel mucosa or seven normal colon mucosa samples. The lymph node sample was analyzed to a 700,000 sequence depth, but still the variant was not detected in this sample. A possible explanation for the anatomically widespread occurrence of the *APC* variant could be field cancerization, a process in which the normal cell population is replaced with tumorigenic clones.[8] This mechanism has been previously described in patients with inflammatory bowel disease, in whom "pre-tumor" clones with identical *TP53* founder mutations, but different driver mutations spread throughout the colon. In agreement with this idea, patient APC-07 showed other additional driver mutations in the different adenomas tested (Supplementary Table 2). A possible mechanism of this clonal expansion of tumorigenic cells is crypt fission, and chronic inflammation of the intestine as seen in inflammatory bowel disease patients is a likely growth stimulus.[8] It is not known whether chronic inflammation was present in patient APC-07. It is also conceivable that previous endoscopies may have aided the spread of tumorigenic cells throughout the intestine, although this does not explain the presence of 20-30 adenomas at first colonoscopy.[8] The size gradient from distal to proximal further accords with the notion that the underlying mechanism of polyposis in this patient was field cancerization developing from one distal tumorigenic cell in the colon and then spreading more proximally.

In detecting APC mosaic variants we underline the importance of testing at least two but preferentially more adenomas. We detected remarkable patterns of mosaicism, with either sporadic adenomas in a mosaic environment or spread of tumorigenic clones throughout the colon suggesting a mechanism of field cancerization.

Table 1: Patients with one APC variant in multiple adenomas

Patient ID	Age	#ad	#T	Mutation	Normal mucosa			Leukocyte		Other tissues tested	Offspring
					n	Freq.	Cov.	Freq.	Cov.		
APC-03	54	~100	6/6	c.3904delC	5	0.4-29%	>2000	1%	2564	LN[2]	0/0
APC-04	54	38	10/16	c.4666dupA	11	NP	NP	NP	NP		0/2
APC-07	48	25	23/23	c.3340C>T	7	0%	NA	0%	NA	I, LN	0/1
APC-08*	35	~100	2/2	c.4666dupA	0	-	-	-	-		0/0
APC-09*	25	62	2/2	c.4057G>T	1	14%	5589	0%	4442	BM, F, U	0/0
APC-15	33	30	3/3	c.2566_2572delCGCGGAA	0	-	-	0%	11446		0/0
APC-17	42	~100	8/8	c.4110_4111delAA	12	0.8 – 15%	>2000	4%	1478		0/0
APC-18	47	80	2/2	c.2493dupA	2	NP	NP	3%[1]	162		0/0
APC-21	67	40	6/6	c.1959-1G>A	2	15-29%	>10000	0%	97256	BM, U	0/1
APC-24	51	58	3/3	c.1974_1975delGA	2	0-2%	>10000	0%	100312		0/0
APC-25	67	30	3/3	c.3211C>T	0	-	-	0%	120062		0/0

*Cumulative adenomas at age of last visit. #ad, number of adenomas; #T, samples with variant/ samples tested; n, number of normal mucosa samples; Freq, frequency; cov, coverage; "-", DNA not available; NP, sequencing not possible due to variant; NA, not applicable - tested with mutation-specific PCR; LN, lymph nodes; I, ileum muscularis propria; F, fibroblasts; BM, buccal mucosa; U, urine. Offspring, children positive for the variant/#children tested *positive control [1]percentage estimated from KASPar assay (Supplementary Figure 1) [2]variant present in 2-3%*

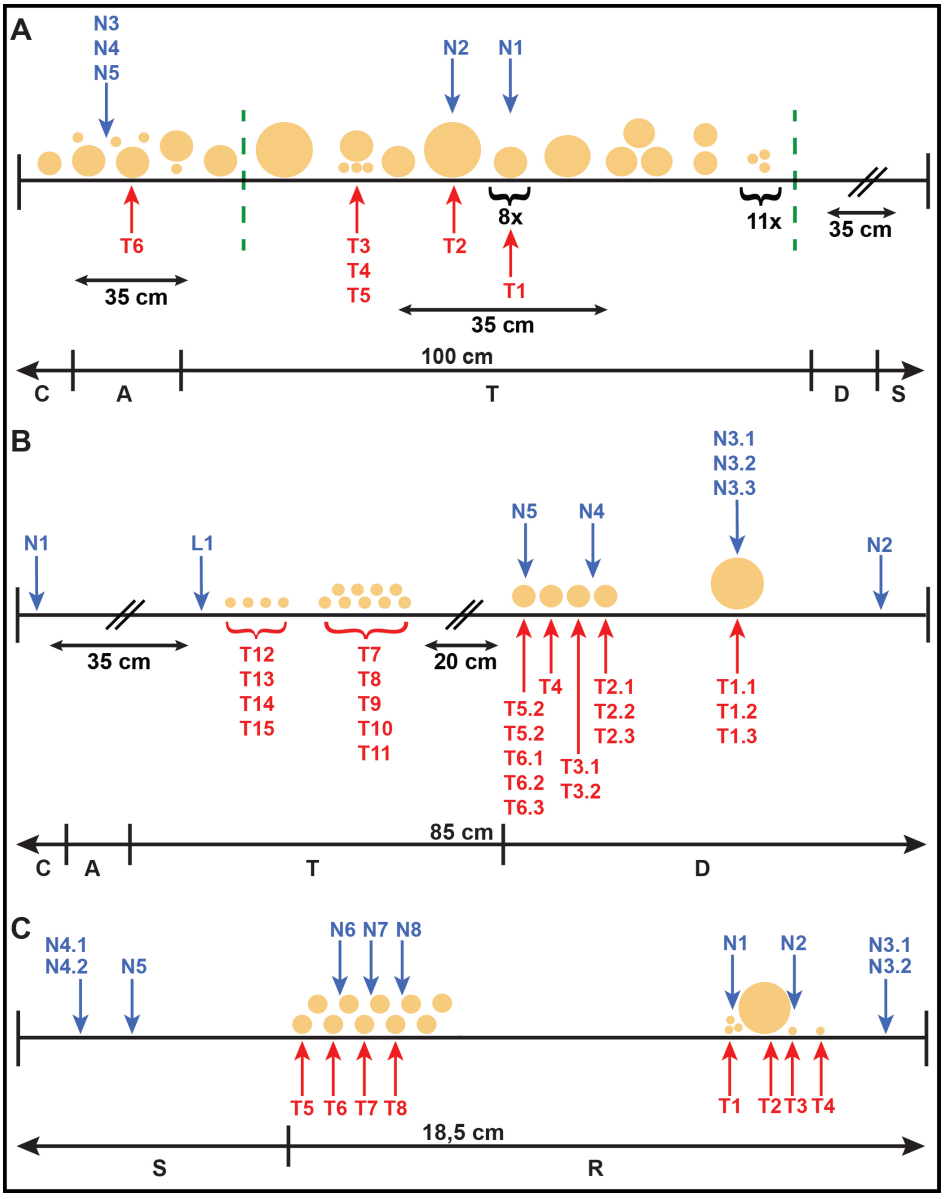


Figure 1: Mosaic patterns of patient APC-03, APC-07 and APC-17

Representation of hemicolectomy patient APC-03(A) and APC-07(B) and rectosigmoid-resection of patient APC-17(C) with location adenoma-derived (T), normal mucosa (N) and lymph node (LN) samples. C, cecum; A, colon ascendens; T, colon transversum; D, colon descendens; S, sigmoid colon; R, rectum. Size of the adenomas in the figure is proportional to real adenoma size, ranging between 4 mm and 3 cm diameter.

References

1. Hes FJ, Nielsen M, Bik EC, et al. Somatic APC mosaicism: an underestimated cause of polyposis coli. Gut 2008;57:71-6.
2. Aretz S, Stienen D, Friedrichs N, et al. Somatic APC mosaicism: a frequent cause of familial adenomatous polyposis (FAP). Hum Mutat 2007;28:985-92.
3. Tuohy TM, Burt RW. Somatic mosaicism: a cause for unexplained cases of FAP? Gut 2008;57:10-2.
4. Spier I, Drichel D, Kerick M, et al. Low-level APC mutational mosaicism is the underlying cause in a substantial fraction of unexplained colorectal adenomatous polyposis cases. J Med Genet 2016;53:172-9.
5. Out AA, van Minderhout IJ, van der Stoep N, et al. High-resolution melting (HRM) re-analysis of a polyposis patients cohort reveals previously undetected heterozygous and mosaic APC gene mutations. Fam Cancer 2015;14:247-57.
6. Gilbert SF. Early Mammalian Development. Developmental Biology 6th edition ed. Sunderland (MA): Sinauer Associates, 2000.
7. Bragg LM, Stone G, Butler MK, et al. Shining a light on dark sequencing: characterising errors in Ion Torrent PGM data. PLoS Comput Biol 2013;9:e1003031.
8. Galandiuk S, Rodriguez-Justo M, Jeffery R, et al. Field cancerization in the intestinal epithelium of patients with Crohn's ileocolitis. Gastroenterology 2012;142:855-864 e8.
9. Hes FJ, Ruano D, Nieuwenhuis M, et al. Colorectal cancer risk variants on 11q23 and 15q13 are associated with unexplained adenomatous polyposis. J Med Genet 2014;51:55-60.
10. Elsayed FA, Kets CM, Ruano D, et al. Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer. Eur J Hum Genet 2014 Aug; 23(8):1080-4
11. Lurkin I, Stoehr R, Hurst CD, et al. Two multiplex assays that simultaneously identify 22 possible mutation sites in the KRAS, BRAF, NRAS and PIK3CA genes. PLoS One 2010;5:e8802.
12. Wang K, Mingyao L, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research 2010;38:e164.
13. Tavtigian SV, Deffenbaugh AM, Yin L, et al. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. J Med Genet 2006;43:295-305.

Supplementary Methods

Study subjects
In total, 27 patients and two positive controls were included. Patients presented with 5-20 adenomas (n=6), 21-100 (n=18), 1000 adenomas (n=1) or multiple colorectal carcinomas (n=2), while the two positive controls both had 21-100 adenomas. All patients were negative for germline *APC/MUTYH* variants by Sanger sequencing and Multiplex Ligation-dependent Probe Amplification (MLPA). In addition, the cohort included patients previously screened for germline (mosaic) *APC* variants by denaturing gradient gel electrophoresis (DGGE)[1], the protein truncation test (PTT)[1] and high-resolution melting analysis (HRMA)[5]. Patients were also screened for CRC risk-associated SNPs[9] and *POLE* and *POLD1* hotspot mutations[10] (see Supplementary Table 1). No pathogenic germline (including mosaic) *APC*, *MUTYH, POLE* or *POLD1* variants were detected in these patients.

Demographic and clinical data and informed consent were obtained during the consultation. The study was approved by the LUMC medical ethical committee (P01-019E). Patients presented with polyposis affecting the entire colon (n= 12) or segmental polyposis, either right-sided (n=2) or left-sided (n= 9) (see Table 1). The location of the adenomas was unknown for four patients, and two patients had multiple colorectal cancers but no adenomas.

Of the 29 patients included, 11 patients represented a retrospective cohort collected between 2000 and 2013. Nine of these patients presented with multiple colorectal adenomas (5-100), while one patient showed 1000 adenomas distributed throughout the colon. The final patient in this group (APC-12) presented with two primary colorectal carcinomas before the age of 50. The average age at diagnosis was 46.8 years (range 26 – 63 years). In addition, sixteen patients were prospectively recruited.

All but one presented with 20-100 adenomas, with an average age at diagnosis of 60.1 years (range 33-75 years). One patient (APC-11) was diagnosed with four primary colorectal carcinomas at the age of 53. Adenomas from two positive controls were also included. These patients (APC-08 and APC-09) were previously tested for mosaic *APC* variants in a pilot study using high resolution melting analysis (HRMA).[5] Patient APC-08 carried an *APC* frameshift variant (c.4666dupA) in both adenomas tested, while patient APC-09 displayed an *APC* nonsense variant (c.4057G>T) in both adenomas tested (see Supplementary Table 1).

Twenty-three tumor or adenoma-derived DNA samples from patient APC-07 were analyzed with Sanger sequencing for six *APC* hotspot variants (c.3340C>T, c.3927_3931delAAAG, c.4348C>T, c.4348C>G, c.4391_4394delAGAG and c.4666delA) at the Erasmus MC, Rotterdam. DNA isolated from the muscularis propria of the ileum, one draining lymph node from the colon and several normal mucosa samples throughout the colon and near the adenomas was tested with SNAPshot[11] and mutation-specific PCR (*APC* mutation-specific primers available on request). Additionally, DNA from rectal mucosa, small bowel mucosa and normal colon mucosa was analyzed by deep sequencing (see Figure 1).

Tissue micro-dissection and DNA extraction
Formalin-fixed paraffin embedded (FFPE) tissue blocks were collected for all patients. Hematoxylin and eosin-stained tumor tissue slides were examined for tissue sections with tumor percentages >20%. After examination, 5 to 10 μm sections were prepared and stained with hematoxylin (eosin staining was omitted to preserve the integrity of the DNA). After staining, slides were visualized with an inverted microscope and manually microdissected with a sharp, pointed knife. When frozen tissue was used, tumor enrichment was achieved by removing non-tumorous tissue as much as possible after frozen section analysis. DNA was isolated with the Nucleospin® Tissue kit (Bioké, Leiden, the Netherlands). If possible, DNA from normal colon mucosa was also isolated from FFPE tissue blocks.

Target enrichment, DNA sequencing and data analysis
A custom *APC* Panel was designed with the Ion AmpliSeq™ Designer tool. The complete sequencing panel consisted of 115 amplicons (11216 bp), covering 99.3% of the coding regions of *APC*. Next-generation sequencing of *APC* was performed with the Ion PGM™ System (Life Technologies, Carlsbad, CA, USA). Raw data analysis, alignments, and variant calling was done using the default parameters in Torrent Suite v4.0. The Variant Caller Parameter Setting was set on 'Somatic – PGM – Low Stringency'. Variants were functionally annotated using ANNOVAR.[12] Variants were annotated to the Genbank reference sequence NM_000038.5. Coding variants were analyzed for their effect on function with *in silico* protein prediction software Align GVGD[13], SIFT (*http://sift.jcvi.org/*), PolyPhen-2 (*http://genetics.bwh.harvard.edu/pph2/*) and MutationTaster (*http://www.mutationtaster.org/*). The Leiden Open Variation Database (LOVD, http://www.lovd.nl/APC) and the catalogue of somatic mutations in cancer (COSMIC, *http://cancer.sanger.ac.uk/cosmic*) were consulted to find variants previously described and classified. All variants predicted to affect function were visually inspected with the Integrative Genomics Viewer (IGV, *https://www.broadinstitute.org/igv/home*) and confirmed by Sanger sequencing. Loss of heterozygosity (LOH) was called if all *APC* SNPs showed a variant frequency between 10 – 30% or between 60-90% (Supplementary Table 2).

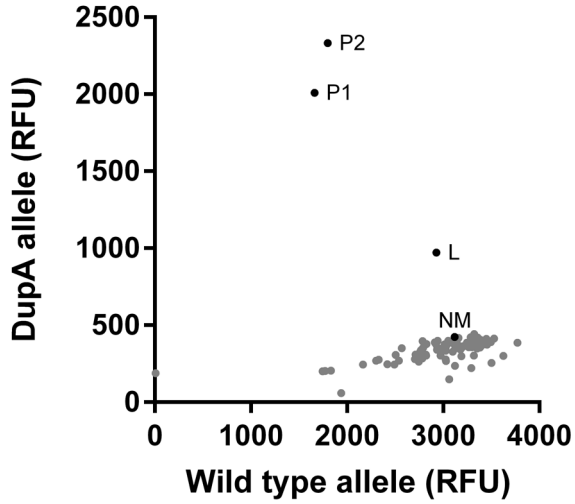
Leukocyte and/or DNA from normal colon mucosa of all patients with identical *APC* variants in multiple adenomas was screened for a specific mosaic *APC* variant. Primers amplifying specific amplicons containing the variant were used for deep sequencing (coverage minimal 1000x). All other available tumor or adenoma-derived DNA was also tested for the mosaic *APC* variant in order to study the pattern of mosaicism (Supplementary Table 1).

Confirmation of variants using KASPar or HRMA
Confirmation of the *APC* c.2493dupA variant (patient APC-18) was done using the competitive allele-specific PCR (KASPar) assay, following the manufacturer's protocol (LGC Genomic, Berlin, Germany). The primers were designed using Primerpicker (KBioscience, Hoddesdon, UK) and the following primers were used: APC_c.2493dupA_C1; **5'** GGC AAC ATG ACT GTC CTT TCA CCA T-3', APC_c.2493dupA_A1; 5'- GAA GGT GAC CAA GTT CAT GCT CCT CTT GAT GAA GAG GAG CTG GGT A-3' and APC_c.2493dupA_A2; 5' GAA GGT CGG AGT CAA CGG ATT CTC TTG ATG AAG AGG AGC TGG GTT- 3'. Variants were identified using CFX manager software v3.0 (Bio-rad, Veenendaal, the Netherlands). Confirmation of the variant is shown in Supplementary Figure 1.

The *APC* c.4666dupA variant (patients APC-04 and APC-08) is a duplication in a homopolymeric nucleotide stretch and could in many cases be a false positive as the NGS workflow used (PGM) is known to produce errors in homopolymeric sequences.[7] This variant was confirmed by high resolution melting analysis (HRMA). Sixteen adenoma and 11 matched normal mucosa DNA samples (normal colon mucosa isolated from the same FFPE block as matched tumor) were available for patient APC-04 and were tested for the variant. HRMA and HRMA data analysis were performed according to the LUMC clinical diagnostics protocol, as previously described.[5]

Testing offspring

The offspring above the age of 18 years of all patients within the cohort was offered genetic screening if a mosaic variant was found in the index patient, irrespective whether the mutation was present in adenomas, normal mucosa and/or leukocyte DNA. (Supplementary Table 1)



Supplementary Figure 1: KASPar assay APC c.2493dupA

KASPar assay of the APC c.2493dupA variant. In black two positive controls (P1 and P2), the leukocyte DNA sample (L) with 7% mutant allele and the normal mucosa (NM) sample. Cluster analysis of the endpoint fluorescence of the KASPar assay shows three distinct clusters. The two adenoma samples cluster together and display a heterozygous phenotype, containing the wildtype and mutant allele. All negative controles (in grey) cluster together and are shown to be homozygous wildtype. The normal mucosa sample falls within this cluster. The leukocyte DNA sample shows a low, but certainly visible percentage of mutant allele.

Supplementary Table 1: Cohort information

ID	Age	#ad	Age*	#ad*	Location	Type	Tumor	#T	#NM	Previously tested
Patients with mosaic *APC* variant										
APC-03	54	100	-	-	**CATDS**R	Whole colon	-	6	5	DGGE, HRM, POLE/POLD1, PTT, Risk-SNPs
APC-04	54	38	-	-	CATD**SR**	Segmental (Left)	CRC54	16	11	-
APC-07	48	25	-	-	**CATD**SR	Whole colon	-	23	7	-
APC-08*	27	80	35	100	Unknown	Unknown	-	2	-	HRM, POLE/POLD1
APC-09*	17	35	25	62	CATD**SR**	Segmental (Left)	-	6	1	DGGE, HRM, POLE/POLD1, PTT
APC-15	33	30	-	-	**CADSR**	Whole colon	-	3	-	POLE/POLD1, Risk-SNPs
APC-17	37	10	42	100	CATD**SR**	Segmental (Left)	Sig37, Rect37	8	12	DGGE, POLE/POLD1, PTT, Risk-SNPs
APC-18	35	20	47	80	**S**	Segmental (Left)	-	4	2	DGGE, HRM, POLE/POLD1, PTT, Risk-SNPs
APC-21	67	40	-	-	**CATDSR**	Whole colon	-	6	-	-
APC-24	51	58	-	-	**T**	Whole colon	CRC51 3x	3	-	POLE/POLD1
APC-25	67	30	-	-	**SR**	Segmental (Left)	-	3	-	-
Patients without mosaic APC variant										
APC-01	45	5	-	-	**CATS**	Whole colon	CRC41, CRC45, Des52	2	-	-
APC-02	54	5	-	-	CATD**S**R	Segmental (Left)	-	4	-	-
APC-05	33	1000	-	-	**ATSR**	Whole colon	-	13	-	POLE/POLD1
APC-06	63	10	-	-	CATD**SR**	Segmental (Left)	CRC63	4	-	-
APC-10	65	>20	-	-	**ATDR**	Whole colon	Rect64, 3x Ad duo	2	-	-
APC-11	53	0	-	-	C**ATD**SR	No polyposis	CRC53 4x	3	-	-
APC-12	40	0	-	-	**T**	No polyposis	CRC40, CRC49	2	-	DGGE, POLE/POLD1, PTT
APC-13	62	5	-	-	Unknown	Unknown	Sig62, Abd.desm.	2	-	-
APC-14	26	40	30	65	Unknown	Unknown	-	12	-	POLE/POLD1, Risk-SNPs

Supplementary Table 1: continued

ID	Age	#ad	Age*	#ad*	Location	Type	Tumor	#T	#NM	Previously tested
APC-16	76	50	-	-	**AS**	Whole colon	CRC74, Sig74	2	-	-
APC-19	61	20	-	-	**CT**	Segmental (right)	-	2	-	-
APC-20	66	20	-	-	**CATDSR**	Whole colon	Pr65	4	-	POLE/POLD1
APC-22	51	34	-	-	Unknown	Unknown	-	2	-	-
APC-23	63	>27	-	-	**TD**	Segmental (left)	CRC63, BHD	4	-	POLE/POLD1
APC-26	67	>50	71	100	**CATDS**	Whole colon	CRC67	3	-	-
APC-27	75	21	-	-	**CAT**	Segmental (right)	-	3	-	-
APC-28	64	>45	-	-	**S**	Segmental (left)	-	3	-	-
APC-29	67	17	68	23	**CATD**	Whole colon	CRC67	4	-	POLE/POLD1

Age and #ad indicates the age and number of adenomas at first diagnosis. Age and #ad* shows cumulative age and cumulative number of adenomas ('-' if no additional data was available). The column 'Location' shows the location of the adenomas and/or carcinomas in bold/underlined if adenomas were present, not in bold if no adenomas were present. If a letter is absent, the location is not described in pathology report, and no information is available. c = cecum, a = colon ascendens, t = colon transversum, d= colon descendens, s= sigmoid, r = rectum. Tumor shows tumortype followed by age of diagnosis. CRC = colorectal carcinoma, Des = desmoid carcinoma, Sig = sigmoid carcinoma, Ad duo = adenoma in duodenum, Rect = rectum carcinoma, Abd.desm = abdominal desmoids, Pr = prostate cancer, BHD = Birt-Hogg-Dubé Syndrome. Some patients were previously screened for mosaic APC variants with denaturing gel gradient electrophoresis (DGGE), high-resolution melting (HRM), the protein truncation test (PTT) or were tested for variants in POLE/POLD1 (POLE/POLD1) or for the presence of CRC-risk associated SNPs in APC (Risk-SNPs). *positive control*

Supplementary Table 2: Overview of *APC* variants found in tested adenomas

Patient-ID		P1-Variant	%	Cov.	P2-Variant	%	Cov.	P3-Variant	%	Cov.	P4-Variant	%	Cov.
APC-01	**2**	-	-	-	c.3337_3338del	17%	1257						
APC-02	**2**	c.4069G>T LOH?	24% -	1635	c.3964G>T LOH?	38% -	40						
APC-03	**2**	c.3904delC c.694C>T	31% 34%	2445 7954	c.3904delC LOH	83% -	2533	c.3904delC	74%	27			
APC-04	**4**	c.4666dupA LOH	58% -	1818	c.4666dupA LOH	45% -	967	c.4666dupA LOH	41% -	443	c.4655_4656del	51%	542
APC-05	**4**	c.3880C>T c.4221delT	34% 24%	1490 2541	-	-	-	c.1279_1289del	32%	1762	c.4580_4589del	37%	1205
APC-06	**4**	847C>T	31%	2241	c.2496delC c.4099C>T	39% 43%	2731 2784	c.4611_4612delAG c.4501delT	37% 39%	894 512			
APC-07	**23**												
APC-08	**2**	c.4666dupA	42%	487	c.4666dupA	40%	2061						
APC-09	**2**	c.4057G>T	27%	699	c.4057G>T	28%	953						
APC-10	**2**	c.3030delT	64%	4251	c.2991T>A LOH	68% -	2789 -						
APC-11	**3**	c.4468delC c.1660C>T	23% 25%	1658 1667	c.637C>T c.4228_4229insT	37% 37%	1052 2648	c.3639_3650delinsT	42%	2202			
APC-12	**2**	-	-	-	-	-	-						
APC-13	**2**	c.4314delA LOH	66% -	1296 -	-	-	-						
APC-14	**2**	-	-	-	-	-	-						
APC-15	**3**	c.2566_2572del c.4429C>T	28% 26%	1555 >2000	c.2566_2572del c.4737delT	28% 27%	2000 1979	c.2566_2572del c.4660G>T	44% 40%	>2000 1934			
APC-16	**2**	-	-	-	-	-	-						
APC-17	**2**	c.4110_4111del LOH	66% -	2165	c.4110_4111del LOH	59% -	865						

Supplementary Table 2 : continued

Patient-ID		P1-Variant	%	Cov.	P2-Variant	%	Cov.	P3-Variant	%	Cov.	P4-Variant	%	Cov.
APC-18	**2**	c.2493dupA	38%	846	c.2493dupA	33%	803						
APC-19	**2**	c.4348C>T c.3467_3470del	17% 26%	451 1984	-	-	-						
APC-20	**4**	c.4233delT	30%	2222	-			-			c.1349delT	31%	3322
APC-21	**6**	c.1959-1G>A	30%	>2000	c.1959-1G>A	14%	>2000	c.1959-1G>A c.3982C>T	30% 30%	>2000 2000	c.1959-1G>A c.4062_4063del c.1972G>T	30% 40% 30%	>2000 1982 1998
APC-22	**2**	c.2626C>T c.2805C>A	23% 22%	1999 1990	-	-	-	c.591_592del c.4157delG	17% 18%	1985 22			
APC-23		c.3991A>T c.847C>T	16% 16%	>2000 >2000	c.2626C>T c.4240delG	47% 31%	2000 62	c.2413C>T c.4233delT	19% 52%	>2000 58	c.1495C>T	41	>2000
APC-24	**3**	c.1974_1975del LOH	70% -	>2000 -	c.1974_1975del	27%	>2000	c.1974_1975del c.994C>T	40% 39%	>2000 >2000			
APC-25	**3**	c.3211C>T	26%	>2000	c.3211C>T c.4057G>T	37% 24%	>2000 >2000	c.3211C>T	34%	1987			
APC-26	**3**	-	-	-	-	-	-	-	-	-			
APC-27	**4**	c.5771_5778del c.3927_3931del	32% 26%	2000 >2000	-	-	-	-	-	-	c.694C>T c.1213C>T	12 11	319 >2000
APC-28	**3**	-	-	-	c.694C>T	28%	94	c.2626C>T c.4271delC	42% 45%	1998 1998			
APC-29	**4**	c.2626C>T c.4348C>T	33% 34%	1998 2000	c.3682C>T c.4733_4734del	37% 36%	2000 1979	c.2932C>T c.4358delC	12% 12%	1998 1952	c.1269G>A c.4459dupA	46 41	1996 1993

n indicates the number of adenomas initially tested. In patients with identical mutations in multiple adenomas, all available tumor material was tested. % shows the variant frequency, cov. indicates the coverage in reads per amplcion. '–' adenoma is sequenced, but no APC variant was detected. LOH indicates loss of heterozygosity. Patient APC-07 was analyzed for APC hotspot mutations and the c.3340C>T was found in all 23 adenoma-derived samples (not shown in this table

Chapter 7

Concluding remarks and future perspectives

Concluding remarks and future perspectives

In this thesis, underlying genetic causes of unexplained suspected Lynch Syndrome and unexplained colonic adenomatous polyposis are presented. Furthermore, practical guidelines are presented to curate variants detected in *PMS2* in DNA isolated from formalin-fixed paraffin-embedded (FFPE) tissue. Additionally, an assay determining the functional effect of splice site variants using RNA isolated from FFPE is presented, all with the ultimate goal to determine the underlying genetic cause of colorectal cancer syndromes. New CRC susceptibility genes and new underlying mechanisms currently explain a portion of the previously unexplained cases, but still a large part of the seemingly familial colorectal cancers remains unexplained. The newly discovered causes, current achievements of determining other susceptibility factors and future perspectives are discussed in this chapter.

Unexplained suspected Lynch Syndrome

In 2014, a study described that in up to 60% of mismatch repair (MMR)-deficient and/or microsatellite unstable tumors no *MLH1* methylation or germline MMR variants are detected.[1] These patients are referred to as 'suspected Lynch Syndrome' (sLS)[1] or Lynch-Like Syndrome (LLS)[2] and clinical management of these patients remains difficult. Little is known about the cancer risk for these patients, although one study showed they have a lower risk of cancer than patients from LS families, but higher than those from families with sporadic CRC.[2] Still, without an exact genetic diagnosis, determining surveillance strategies for patients and their relatives is difficult, and can lead to over- as well as undertreatment of family members, e.g. intensive cancer screening in those without an increased CRC risk.[3] Several theories have been suggested as to what could be the cause of the MMR-deficiency in these tumors. Three potential reasons for MMR-deficient and/or MSI-H tumors in sLS patients are discussed here.

Missed MMR variants

The most obvious explanation of MMR-deficiency in sLS tumors is missed variants in one of the MMR genes. Since most diagnostic and research testing only screen the coding regions of the MMR genes, (intronic) variants can be missed. One study shows a previously missed deep-intronic *MSH2* variant resulting in inclusion of a pseudoexon.[4] However, as we have shown in Chapter 2, sequencing of intronic regions results in a large number of intronic variants of uncertain significance (VUS), and classification of these variants is challenging.[5] Furthermore, while some of these intronic variants might affect splicing, this is difficult to predict using most splice-site prediction software, because these depend strongly on the presence of a canonical splice site nearby the variant. More experimental data are needed to optimize existing prediction tools for deep intronic variants.[6-8] According to the variant classification guidelines created by the InSiGHT (International Society for Gastrointestinal Hereditary Tumors) variant interpretation committee, intronic variants, but also missense variants, silent variants and promoter variants, are automatically classified as Class 3 (uncertain significance), until proof of pathogenicity is delivered.

Another category of missed variants are large genomic rearrangements. Detecting large insertions/deletions (indel) with next-generation sequencing (NGS) has been shown to be challenging, and these indels might be missed in current molecular tumor diagnostics.[9] Some studies describe previously missed inversions[10, 11] and rearrangements[12, 13] in sLS patients.

Other explanations include the presence of variants for which the impact on MMR protein function is presently not clear. For example, many studies have investigated the effect that promoter variants can have on gene expression. One well studied promoter variant, *MLH1* c.-27C>A, has been described to confer a CRC-risk by dominant inheritance of a constitutional MLH1 epimutation, to co-segregate with CRC in multiple families affected with CRC and to lead to reduced MLH1 expression.[14-17] While in the past described as pathogenic, it is currently in the LOVD database classified as a VUS because of 'insufficient evidence'. Other known promoter variants are the *MLH1* c.-28A>G/-7C>T and the *MLH1* c.-93G>A variants described to show partial loss of expression (c.-28)[18] and increased CRC risk due to possible epigenetic silencing (-93).[19-21] The latter, however, was classified as benign by the InSiGHT group in 2013, because of the high minor allele frequency (MAF) in the general population.

Variants of uncertain significance remain a concern and functional assays are needed to assess pathogenicity, especially for rare variants found in only a few families or moderately penetrant variants which do not show complete co-segregation in a family. Laboratory efforts capable of assessing the effect of a VUS on various aspects of MMR protein function are cell-free assays determining mismatch repair activity[22-24], cell-based assays showing (lack of) expression in CRC cell lines[25-27], nuclear localization assays and RNA splicing assays using patient RNA[28, 29] or minigene splicing assays.[7, 30, 31] Consensus about pathogenicity of a variant is important for proper clinical management.

Somatic inactivation

A quite prevalent cause of MMR-deficiency in sLS patients appears to be biallelic somatic inactivation of the MMR-genes. Multiple studies (Table 1) have shown somatic inactivation in 11% to 100% of the tested sLS cohort.[32-37] Of the 68 patients described with biallelic somatic inactivation of one of the MMR genes in these six studies, 37 tumors (54%) had one somatic variant and additional loss of heterozygosity. Patients presented with colorectal, endometrial or small bowel cancer with a mean age of 54.6 years (range 27 - 81 years). Two studies also tested for variants in the exonuclease domain of *POLE* and *POLD1*[34, 35], in the other four an underlying variant in one of these genes cannot be excluded. We, and others, have detected

Table 1: Studies on biallelic somatic inactivation in sLS patients

	No. of sLS patients†	No. of biallelic som. MMR	2 MMR	1MMR/ LOH	Average age of onset (years)
Chika et al[32]	2	2	2	0	75.0
Geurts-Giele et al[33]	40	21	5	16	58.9
Haraldsdottir et al[34]	27*	19	11	8	53.8
Jansen et al[35]	53*	10	5	5	48.1
Mensenkamp et al[36]	25	13	5	8	47.7
Sourrouille et al[37]	27	3	3	0	55.0
Total	**174**	**68**	**31**	**37**	**54.6**

†Only patients without germline MMR variants and without MLH1 promoter hypermethylation are shown
**sLS patients were tested for POLE and POLD1 variants, patients with a variant were excluded. 2MMR; two somatic variants, 1MMR/LOH; one variant and loss of heterozygosity (LOH) of the WT-allele.*

two somatic variants in patients with a family history of CRC, showing that biallelic somatic inactivation does indeed explain the occurrence of a tumor. However, it cannot explain other occurrences of colon cancer in the family and it could be that another underlying genetic cause was missed.[35, 36] In some cases, the somatic MMR variants are secondary to another (germline) defect that results in a higher mutational load. This has previously been described in patients with MMR-deficient/MSI-H tumors, found to carry biallelic *MUTYH* variants impairing base excision repair (BER).[38-42] In the studies where further testing was performed on the *MUTYH* mutated/MMR-deficient tumors, somatic MMR variants[38] or *MLH1* promoter hypermethylation[40] explained the MMR-deficiency. The somatic MMR variants were MAP-specific G>T variants, indicating that the impaired BER was the primary defect followed by MMR-deficiency.[38] This mechanism of secondary MMR-deficiency is also seen in *POLE/POLD1* tumors. These tumors have a high mutational load, leading to many variants that could possibly inactivate genes. These findings underline the importance of screening for variants in other CRC susceptibility genes in patients with biallellic somatic inactivation of the MMR genes but with a positive family history of CRC, especially if the other family members do not show MMR-deficiency/MSI in the tumor. Another possibility is that these somatic variants are present as mosaic variants in the leukocyte DNA, something that has been (albeit very rarely) described before. Sourrouille et al describes one patient where a somatic *MSH2* variant found in the tumor of the mother (but not in blood) was also detected in leukocyte- and tumor DNA of the affected son.[37] No other germinal mosaic MMR variants have been described.

Variants in other CRC susceptibility genes

Recent advances in next-generation sequencing (NGS) and whole exome sequencing (WES) or even whole genome sequencing (WGS) resulted in the detection of additional genes possibly involved in tumorigenesis of sLS tumors. As mentioned before, we (Chapter 3), and others, have shown germline and somatic variants in *POLE* or *POLD1* in MMR-deficient tumors.[34, 35, 43] Variants in the exonuclease domain (EDM) of these genes encoding for polymerase ε and δ respectively, have been shown to result in a high mutational burden, often with a high number of C>A transversions.[44] In sLS tumors it has been hypothesized that a somatic or germline *POLE/POLD1* variant can result in somatic MMR variants, subsequently resulting in microsatellite instability.[35, 43] Notably, a recent study shows synergistic increase in mutation rate when a pathogenic *POLD1* variant (*POLD1* R689W) was combined with MMR-deficiency, indicating that the *POLD1* mutator effect results from a high rate of replication errors.[45] This variant is not located in the *POLD1*-EDM but does show impaired nucleotide selectivity, showing that even variants outside the *POLE/POLD1*-EDM might confer an increased CRC susceptibility.[45, 46]

Other genes that have been implicated as the underlying cause of suspected Lynch Syndrome tumors are *BRCA1, BRCA2, MUTYH, APC, STK11, MLH3* and *EXO1*.[38, 47-49] As mentioned before, homozygous and compound-heterozygous *MUTYH* variants have been described in multiple sLS patients, some of which were shown to have secondary MMR-deficiency through somatic MMR variants or *MLH1* promoter hypermethylation.[38, 40] *BRCA1, BRCA2, APC* and *STK11* are known dominant cancer predisposition genes previously described to be involved in hereditary breast cancer, familial adenomatous polyposis and Peutz-Jeghers Syndrome respectively.[50-52] While the increased risk of CRC for *BRCA1* and *BRCA2* mutation

carriers is still under debate, some studies indicate an up to five fold increased risk of CRC.[53-58] However, the 'suspected Lynch Syndrome' patients described in these studies found to carry these *BRCA1, BRCA2, APC* and *STK11* variants were selected based on personal and family history of CRC but no IHC of the four MMR genes or MSI was performed.

MLH3 and *EXO1* are both mismatch repair genes. MLH3 interacts with MLH1, is believed to participate in insertion deletion loop (IDL) repair and has been shown to exhibit MSI in cell culture.[59, 60] The role of the mismatch repair gene *MLH3* in colorectal tumorigenesis is under debate.[61-64] Whereas one study describes germline *MLH3* variants in sLS patients[61] (with positive MLH1, MSH2 and MSH6 staining, but with MSI), more evidence is needed before it can be regarded as a causal (possible reduced penetrant) Lynch Syndrome gene. The exonuclease 1 (*EXO1*) gene encodes a 5' -> 3' exonuclease that is involved in multiple DNA repair pathways.[65, 66] In MMR, EXO1 interacts with MSH2 and in yeast it shows a mutator phenotype when lost.[66] *EXO1* has been extensively studied in CRC[48, 64, 67-71], and is often found not to be associated with (suspected) Lynch Syndrome.[67, 70, 71] Single nucleotide polymorphisms (SNPs) in this gene have been described to confer a slightly increased risk of CRC in the general population, but have not been associated with LS.[68, 69] Wu et al describes germline *EXO1* variants in 14 patients fulfilling Amsterdam criteria, with six of the tumors showing microsatellite instability.[48] Twelve of thirteen tested tumors showed loss of heterozygosity with retention of the wildtype allele. The authors suggest that complete loss of EXO1 is lethal to the cell, and that a haploinsufficiency effect can give rise to tumors. Other studies screened for *EXO1* variants in sLS patients, but no other carriers were found.[72, 73] This, together with the fact that patients were not screened for variants in established LS/CRC susceptibility genes (*PMS2, POLE, POLD1),* shows little evidence to include *EXO1* as a LS gene, although a role as a low penetrant cancer susceptibility gene cannot be excluded at this point in time.

For Lynch Syndrome tumorigenesis, it remains unclear when the MMR-deficient phenotype (e.g. loss of the second allele in germline MMR carriers) is acquired during tumorigenesis. Multiple studies tested adenomas from MMR-mutation carriers for immunohistochemical loss of the MMR proteins or MSI and found a Lynch phenotype in 40-80% of tested adenomas, often associated with adenoma size.[74-80] Ahadova et al proposed two pathways in Lynch Syndrome (see Figure 1), one where adenomatous polyps precede MMR-deficiency with an initiating event as *APC* and/or *KRAS* gene variants and one where MMR gene inactivation is the initial event leading to non-polypous precursor lesions and secondary *CTNNB1* hits are needed for tumorigenesis.[81] While it can be debated whether polypous growth precedes MMR gene inactivation in LS tumors, this is probably the preferred pathway in suspected Lynch Syndrome patients with somatic MMR inactivation. In these patients the underlying genetic defects (*MUTYH, POLE, POLD1)* often result in a mutator phenotype, with the MMR-deficiency acting as a secondary defect resulting in tumorigenesis. It is conceivable that for these patients adenomas can arise before mismatch repair is completely inactivated. Immunohistochemical staining and MSI analysis of these tumors gives a bias to expect an MMR defect, while the true underlying germline defect could be missed. Especially when patients present with early onset colorectal cancer and a positive family history of CRC, tumors should be screened for variants in other CRC susceptibility genes. Whole exome sequencing of unexplained suspected Lynch Syndrome patients might detect CRC susceptibility genes

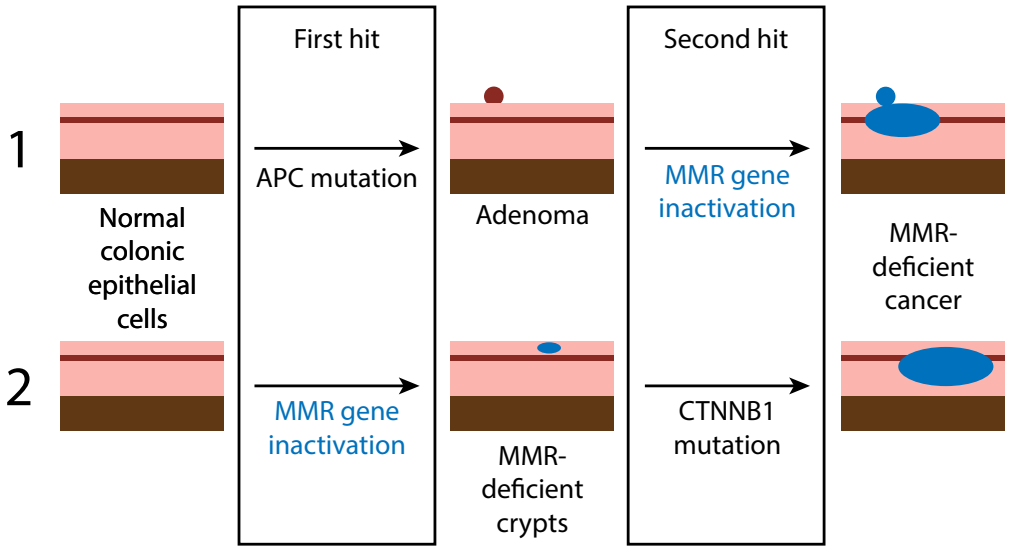


Figure 1: proposed pathways of colorectal tumorigenesis in Lynch Syndrome as described by Ahadova et al, Familial Cancer, 2016.[81]

or loci which are currently not known, although this may require very large study sizes. The expectation is that through these approaches, rare genetic variants or high-risk combinations of CRC susceptibility SNPs can be detected, an effort that is already being done in MMR-proficient, microsatellite stable tumors.[82, 83]

Unexplained polyposis

In approximately 76-82% of the severe and typical FAP patients a pathogenic germline variant in *APC* is found, while the majority of patients with atypical or attenuated polyposis can be explained by *APC* or homozygous/compound-heterozygous germline *MUTYH* variants.[84-86] Still, in a small fraction of patients no genetic predisposition can be found.

In recent years three new polyposis syndromes have been described which may account for a part of these unexplained polyposis cases. These syndromes are polymerase proofreading associated polyposis (PPAP)[87], caused by variants in the proofreading domain of *POLE* and *POLD1,* NTHL1-associated polyposis (NAP)[88], caused by homozygous or compound-heterozygous variants in the base excision repair gene *NTHL1* and the recently described MSH3-associated polyposis[89] caused by biallelic germline inactivation of the mismatch repair gene *MSH3.* Both PPAP and NAP have a variable phenotype with polyps but also with the occurrence of colorectal cancers and other extracolonic tumors. Both are also described to confer a specific mutation spectrum, with a *POLE/POLD1* defect resulting in an ultramutated phenotype with an increase of C:T>A:G transversions and *NTHL1* deficient tumors showing an increase in C:G>T:A transitions.[44, 88] Compound-heterozygous *MSH3* variants have only been described in two unrelated individuals, one female with over 40 polyps and a history of thyroid adenoma (age 35), uterine leiomyoma (age 44) and polyps in corpus uteri and

duodenum, and one female with multiple adenomas at age 32, astrocytoma (age 26), ovarian and dermoid cysts and follicular thyroid adenomas at age 42.[89] Both also had small intraductal papillomas in the mammary gland. Whether these neoplasms are specific for an *MSH3* defect is unknown, and can only be concluded after more *MSH3* variant carriers are detected.

Of all patients carrying a pathogenic *de novo APC* variant, explaining the FAP phenotype, approximately 20% are estimated to have somatic mosaicism.[90-92] This phenomenon, in which (*APC)* variants are only present in a fraction of the cells, was already described in 1999 in a study where parents of five *de novo* FAP patients were tested for the *APC* variant found in the index patient.[93] Multiple studies have been conducted since then, always focusing on testing leukocyte DNA for variants with a low variant allele frequency, or on analyzing parents of *de novo* patients.[90-92, 94-99] We (Chapter 5) and others have recently shown a high number of missed mosaic *APC* patients, by sequencing multiple adenomas of the same patients.[100, 101] If two or more adenomas carry an identical *APC* variant, this might indicate an underlying *APC* mosaicism. This strategy has been proven to be more sensitive and specific than sequencing leukocyte DNA for variants with a low variant frequency, and can detect mosaicism confined to the colon.[100, 101] Early detection of patients with somatic *APC* mosaicism is important and will help guide clinical management. Patients often show an attenuated FAP phenotype but if the variant is inherited by the patients offspring, they will show a full blown FAP phenotype with possibly 100 to more than 1000 adenomas.[90, 91, 100]

Besides the recently described colorectal cancer syndromes with *POLE, POLD1, NTHL1* and *MSH3* germline variants, other genes have been implicated in unexplained familial polyposis syndromes. Three recent studies performing whole exome sequencing or genome-wide SNP genotyping in unexplained adenomatous polyposis patients describe variants or copy number variations in the *CNTN6, FOCAD, HSPH1, KIF26B, MCM3AP, YBEY, ARHGAP, CTNNB1, DSC2, PIEZO1, ZSWIM7* and *MCM9* genes.[102-104] Variants in *DSC2*, *PIEZO1* and *ZSWIM7* were first detected in a cohort of 7 unrelated polyposis patients with 20-40 adenomas. Sequencing these three genes in a validation cohort of 191 unexplained polyposis patients resulted in the detection of 4 additional *DSC2* variants and 4 additional *PIEZO1* variants.[104] Copy number variant (CNV) analysis in 221 unexplained polyposis patients showed rare, non-recurrent germline CNVs in 77 proteins.[103] Targeted NGS found point-mutations in 10 of the 77 investigated genes (*CNTN6, FOCAD, HSPH1, KIF26B, MCM3AP, YBEY, CTNNB1* and three genes from the *ARHGAP* family).[103] Of these 10 genes, only *FOCAD* and *CTNNB1,* involved in Wnt signaling, have previously been related to CRC or colorectal adenoma predisposition.[105, 106] The third study shows homozygous variants in the *MCM9* gene in two sisters with multiple polyps and metastatic CRC at young age. The *MCM9* gene encodes a DNA helicase involved in homologous recombination (HR) and mismatch repair (MMR).[107] However, an independent study sequencing *MCM9* in suspected Lynch Syndrome patients only detected variants of uncertain significance.[108] For all newly discovered genes more evidence is needed to establish them as CRC or polyposis susceptibility genes.

Other factors in colorectal cancer

While genetic factors are expected to play an important role in families with multiple affected patients, other factors are known to increase CRC susceptibility. Diet is an important factor, and it has been shown that a Western diet characterized by high intake of meat, refined

sugar and saturated fat, but lacking in fiber, contributes to development of CRC.[109-112] Other lifestyle factors contributing to CRC risk are excess body weight, low physical activity, alcohol consumption and smoking.[111-113] Calcium, fiber, milk and whole grains on the other hand, have been associated with a lower risk of CRC.[112-114]

Another factor involved in increased CRC is the gut microbiota. The human intestinal tract is inhabited by trillions of microorganisms and the presence of specific bacteria or the dysbiosis of bacteria can aid in development and progression of colorectal cancer.[115, 116] Bacteria can affect CRC tumorigenesis by secretion of toxins that can induce DNA damage and secretion of metabolites that affect translation, gene regulation and cell proliferation. [115, 116] The gut microbiome is also shown to affect innate immunity through activation of toll-like receptors and through influencing T-cell differentiation.[115, 116] Interestingly, diet and alcohol consumption have been shown to be able to dysregulate microbiota, possibly explaining the link between diet and colorectal cancer.[115-117]

Abovementioned non-genetic factors can result in believing multiple cancers in a family are due to an underlying genetic cause, while in fact the affected patients are phenocopies, i.e. displaying characteristics of a certain genotype but produced by environmental factors. This underlines a challenge in determining the underlying genetic cause in unexplained families with CRC. While WES and WGS enable the detection of rare pathogenic variants, it is difficult to determine whether a family with a few affected family members is indeed the result of one dominant DNA variation. Colorectal cancer is a common disease, and many environmental, as well as polygenic, factors can modulate CRC risk. A recent study shows that families with exactly two first-degree relatives only have a moderate probability of being due to segregating familial disease and advises to first focus on families with three or more members to increase the probability of finding genetic factors.[118]

Final conclusions

The estimated worldwide incidence of colorectal cancer is 746.000 new cases annually.[119] Approximately 15% of these colorectal cancers will display MSI, while 3% of these CRCs will carry a germline MMR variant. Routine molecular screening of all early-onset CRCs with immunohistochemical staining of the four MMR proteins and/or MSI analysis yields a high number of suspected Lynch Syndrome cases of which the majority will likely remain suspected after conventional germline mutation screening and MSI analysis. In this thesis we describe possible genetic causes of unexplained suspected Lynch Syndrome (sLS) and unexplained adenomatous polyposis. We hypothesize that unexplained sLS patients could be explained by (1) missed MMR variants (Chapter 2), (2) by biallelic somatic inactivation (Chapter 2 and 3) or (3) by variants in other CRC susceptibility genes (Chapter 3).

We (Chapter 2 and 3), and others, show that biallelic somatic inactivation could possibly explain up to half of all unexplained sLS patients. Screening for somatic MMR variants should be broadly introduced in molecular tumor diagnostics, giving more insight in these sporadic MMR-deficient cases. Families should be critically assessed, so no underlying genetic variants are missed in these seemingly sporadic cases. Recent advances in (whole) exome and targeted next-generation sequencing enable detection of rare variants in previously unknown CRC susceptibility genes, and show that MMR-deficiency could be due to (secondary) somatic events, sometimes with underlying germline gene variants in genes such as *POLE, POLD1* or *MUTYH*. In future research it would be interesting to test affected family members of the patients of whom the tumor occurrence is explained by somatic events, to see whether these patients also have MMR-deficient tumors and to find a common underlying genetic cause in these family members.

While missed MMR variants seem a rare event, explaining only a fraction of patients (Chapter 2), it cannot be excluded that the remaining unexplained sLS tumors still carry undetected MMR variants. These variants could be currently misunderstood and classified as uncertain significance (VUS) or could be missed by conventional screening, because they are in intronic regions or in close proximity to the MMR genes. Functional assessment of VUS is critical. In Chapter 4 we show an easy method to detect aberrant splicing in formalin-fixed paraffin-embedded tissue, showing opportunities for the future. However, while this method enables functional assessment of variants predicted to result in splicing, it cannot detect aberrant splicing when the mutation status is unknown. A high-throughput NGS-based assay sequencing all exon-exon boundaries in RNA to detect any possible RNA aberration could possibly detect the cause of the MMR-deficiency. Also, it could indicate missed deep-intronic variants resulting in an aberrant RNA product. Furthermore, the current study relies on *in silico* prediction tools and effects that fall outside the amplification window will not be detected, while a high throughput assay would analyse the RNA in an unbiased manner. Other currently undiscovered mechanisms leading to aberrant RNA could be detected with this unbiased NGS approach.

The introduction of the population-based screening leads to an increasing number of patients with a low number of polyps (5-40) at older age (60+) which would previously go undiscovered. We (Chapter 6), and others, recently showed that many of these patients with

an attenuated FAP (AFAP) phenotype carry somatic mosaic *APC* variants. Current screening consisting of screening leukocyte DNA for variants with a low variant allele frequency is shown to miss many of these patients, because the variant is present with a too low variant allele frequency or because the mosaicism is confined to the colon. In-depth analyses of adenomas of AFAP patients could lead to the detection of more mosaic *APC* carriers, affecting clinical management of their children, who, if the variant is inherited, will show a full blown FAP phenotype.

While the exact incidence of biallelic somatic MMR variants, *POLE/POLD1* variants with secondary MMR-deficiency and somatic mosaicism still needs to be assessed, a large fraction of the previously unexplained sLS patients can now be explained and receive proper clinical management. However, in the remaining, approximately, 20-40% of sLS patients the underlying (familial) cause is still unknown. Whole exome or genome sequencing in the future will possibly lead to the detection of more rare gene variants, variants with moderate increase in CRC susceptibility or variants with moderate penetrance. Joint efforts screening for variants in larger cohorts and data sharing are essential in finding these (low penetrant) CRC susceptibility loci and might enable establishment of a CRC polygenic risk model with a personal cancer risk score.

References

1. Buchanan DD, Rosty C, Clendenning M, et al. Clinical problems of colorectal cancer and endometrial cancer cases with unknown cause of tumor mismatch repair deficiency (suspected Lynch syndrome). Appl Clin Genet 2014;7:183-93.
2. Rodriguez-Soler M, Perez-Carbonell L, Guarinos C, et al. Risk of cancer in cases of suspected lynch syndrome without germline mutation. Gastroenterology 2013;144:926-932 e1; quiz e13-4.
3. Castells A, Castellví–Bel S, Balaguer F. Concepts in Familial Colorectal Cancer: Where Do We Stand and What Is the Future? Gastroenterology;137:404-409.
4. Clendenning M, Buchanan DD, Walsh MD, et al. Mutation deep within an intron of MSH2 causes Lynch syndrome. Fam Cancer 2011;10:297-301.
5. Jansen AM, Geilenkirchen MA, van Wezel T, et al. Whole Gene Capture Analysis of 15 CRC Susceptibility Genes in Suspected Lynch Syndrome Patients. PLoS One 2016;11:e0157381.
6. Colombo M, De Vecchi G, Caleca L, et al. Comparative in vitro and in silico analyses of variants in splicing regions of BRCA1 and BRCA2 genes and characterization of novel pathogenic mutations. PLoS One 2013;8:e57173.
7. van der Klift HM, Jansen AM, van der Steenstraten N, et al. Splicing analysis for exonic and intronic mismatch repair gene variants associated with Lynch syndrome confirms high concordance between minigene assays and patient RNA analyses. Mol Genet Genomic Med 2015;3:327-45.
8. Vreeswijk MP, Kraan JN, van der Klift HM, et al. Intronic variants in BRCA1 and BRCA2 that affect RNA splicing can be reliably selected by splice-site prediction programs. Hum Mutat 2009;30:107-14.
9. Ye K, Wang J, Jayasinghe R, et al. Systematic discovery of complex insertions and deletions in human cancers. Nat Med 2016;22:97-104.
10. Rhees J, Arnold M, Boland CR. Inversion of exons 1-7 of the MSH2 gene is a frequent cause of unexplained Lynch syndrome in one local population. Fam Cancer 2014;13:219-25.
11. Mork ME, Rodriguez A, Taggart MW, et al. Identification of MSH2 inversion of exons 1-7 in clinical evaluation of families with suspected Lynch syndrome. Fam Cancer 2017;16:357-361.
12. Perez-Cabornero L, Borras Flores E, Infante Sanz M, et al. Characterization of new founder Alu-mediated rearrangements in MSH2 gene associated with a Lynch syndrome phenotype. Cancer Prev Res (Phila) 2011;4:1546-55.
13. Plaschke J, Ruschoff J, Schackert HK. Genomic rearrangements of hMSH6 contribute to the genetic predisposition in suspected hereditary non-polyposis colorectal cancer syndrome. J Med Genet 2003;40:597-600.
14. Hitchins MP, Rapkins RW, Kwok CT, et al. Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5'UTR. Cancer Cell 2011;20:200-13.
15. Kwok CT, Vogelaar IP, van Zelst-Stams WA, et al. The MLH1 c.-27C>A and c.85G>T variants are linked to dominantly inherited MLH1 epimutation and are borne on a European ancestral haplotype. Eur J Hum Genet 2014;22:617-24.
16. Ward RL, Dobbins T, Lindor NM, et al. Identification of constitutional MLH1 epimutations and promoter variants in colorectal cancer patients from the Colon Cancer Family Registry. Genet Med 2013;15:25-35.
17. Raevaara TE, Korhonen MK, Lohi H, et al. Functional significance and clinical phenotype of nontruncating mismatch repair variants of MLH1. Gastroenterology 2005;129:537-49.
18. Hesson LB, Packham D, Kwok CT, et al. Lynch syndrome associated with two MLH1 promoter variants and allelic imbalance of MLH1 expression. Hum Mutat 2015;36:622-30.
19. Raptis S, Mrkonjic M, Green RC, et al. MLH1 -93G>A promoter polymorphism and the risk of microsatellite-unstable colorectal cancer. J Natl Cancer Inst 2007;99:463-74.
20. Wang T, Liu Y, Sima L, et al. Association between MLH1 -93G>a polymorphism and risk of colorectal cancer. PLoS One 2012;7:e50449.
21. Chen H, Taylor NP, Sotamaa KM, et al. Evidence for heritable predisposition to epigenetic silencing of MLH1. Int J Cancer 2007;120:1684-8.
22. Drost M, Zonneveld J, van Dijk L, et al. A cell-free assay for the functional analysis of variants of the mismatch repair protein MLH1. Hum Mutat 2010;31:247-53.
23. Drost M, Zonneveld JB, van Hees S, et al. A rapid and cell-free assay to test the activity of lynch syndrome-associated MSH2 and MSH6 missense variants. Hum Mutat 2012;33:488-94.
24. Drost M, Koppejan H, de Wind N. Inactivation of DNA mismatch repair by variants of uncertain significance in the PMS2 gene. Hum Mutat 2013;34:1477-80.
25. Trojan J, Zeuzem S, Randolph A, et al. Functional analysis of hMLH1 variants and HNPCC-related mutations using a human expression system. Gastroenterology 2002;122:211-9.

26. Takahashi M, Shimodaira H, Andreutti-Zaugg C, et al. Functional analysis of human MLH1 variants using yeast and in vitro mismatch repair assays. Cancer Res 2007;67:4595-604.
27. Perera S, Bapat B. The MLH1 variants p.Arg265Cys and p.Lys618Ala affect protein stability while p.Leu749Gln affects heterodimer formation. Hum Mutat 2008;29:332.
28. Sharp A, Pichert G, Lucassen A, et al. RNA analysis reveals splicing mutations and loss of expression defects in MLH1 and BRCA1. Hum Mutat 2004;24:272.
29. Pagenstecher C, Wehner M, Friedl W, et al. Aberrant splicing in MLH1 and MSH2 due to exonic and intronic variants. Hum Genet 2006;119:9-22.
30. Auclair J, Busine MP, Navarro C, et al. Systematic mRNA analysis for the effect of MLH1 and MSH2 missense and silent mutations on aberrant splicing. Hum Mutat 2006;27:145-54.
31. Tournier I, Vezain M, Martins A, et al. A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. Hum Mutat 2008;29:1412-24.
32. Chika N, Eguchi H, Kumamoto K, et al. Prevalence of Lynch syndrome and Lynch-like syndrome among patients with colorectal cancer in a Japanese hospital-based population. Jpn J Clin Oncol 2017;47:108-117.
33. Geurts-Giele WR, Leenen CH, Dubbink HJ, et al. Somatic aberrations of mismatch repair genes as a cause of microsatellite-unstable cancers. J Pathol 2014;234:548-59.
34. Haraldsdottir S, Hampel H, Tomsic J, et al. Colon and endometrial cancers with mismatch repair deficiency can arise from somatic, rather than germline, mutations. Gastroenterology 2014;147:1308-1316 e1.
35. Jansen AM, van Wezel T, van den Akker BE, et al. Combined mismatch repair and POLE/POLD1 defects explain unresolved suspected Lynch syndrome cancers. Eur J Hum Genet 2016;24:1089-92.
36. Mensenkamp AR, Vogelaar IP, van Zelst-Stams WA, et al. Somatic mutations in MLH1 and MSH2 are a frequent cause of mismatch-repair deficiency in Lynch syndrome-like tumors. Gastroenterology 2014;146:643-646 e8.
37. Sourrouille I, Coulet F, Lefevre JH, et al. Somatic mosaicism and double somatic hits can lead to MSI colorectal tumors. Fam Cancer 2013;12:27-33.
38. Morak M, Heidenreich B, Keller G, et al. Biallelic MUTYH mutations can mimic Lynch syndrome. Eur J Hum Genet 2014;22:1334-7.
39. Lefevre JH, Rodrigue CM, Mourra N, et al. Implication of MYH in colorectal polyposis. Ann Surg 2006;244:874-9; discussion 879-80.
40. Colebatch A, Hitchins M, Williams R, et al. The role of MYH and microsatellite instability in the development of sporadic colorectal cancer. Br J Cancer 2006;95:1239-43.
41. O'Shea AM, Cleary SP, Croitoru MA, et al. Pathological features of colorectal carcinomas in MYH-associated polyposis. Histopathology 2008;53:184-94.
42. Cleary SP, Cotterchio M, Jenkins MA, et al. Germline MutY human homologue mutations and colorectal cancer: a multisite case-control study. Gastroenterology 2009;136:1251-60.
43. Elsayed FA, Kets CM, Ruano D, et al. Germline variants in POLE are associated with early onset mismatch repair deficient colorectal cancer. Eur J Hum Genet 2015;23:1080-4.
44. Shinbrot E, Henninger EE, Weinhold N, et al. Exonuclease mutations in DNA polymerase epsilon reveal replication strand specific mutation patterns and human origins of replication. Genome Res 2014;24:1740-50.
45. Mertz TM, Baranovskiy AG, Wang J, et al. Nucleotide selectivity defect and mutator phenotype conferred by a colon cancer-associated DNA polymerase delta mutation in human cells. Oncogene 2017;36:4427-4433.
46. Mertz TM, Sharma S, Chabes A, et al. Colon cancer-associated mutator DNA polymerase delta variant causes expansion of dNTP pools increasing its own infidelity. Proc Natl Acad Sci U S A 2015;112:E2467-76.
47. Yurgelun MB, Allen B, Kaldate RR, et al. Identification of a Variety of Mutations in Cancer Predisposition Genes in Patients With Suspected Lynch Syndrome. Gastroenterology 2015;149:604-13 e20.
48. Wu Y, Berends MJ, Post JG, et al. Germline mutations of EXO1 gene in patients with hereditary nonpolyposis colorectal cancer (HNPCC) and atypical HNPCC forms. Gastroenterology 2001;120:1580-7.
49. Castillejo A, Vargas G, Castillejo MI, et al. Prevalence of germline MUTYH mutations among Lynch-like syndrome patients. Eur J Cancer 2014;50:2241-50.
50. Groden J, Thliveris A, Samowitz W, et al. Identification and characterization of the familial adenomatous polyposis coli gene. Cell 1991;66:589-600.
51. Hall JM, Lee MK, Newman B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. Science 1990;250:1684-9.
52. Jenne DE, Reimann H, Nezu J, et al. Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. Nature Genetics 1998;18:38-44.
53. Thompson D, Easton DF. Cancer Incidence in BRCA1 Mutation Carriers. JNCI: Journal of the National Cancer Institute 2002;94:1358-1365.
54. Sopik V, Phelan C, Cybulski C, et al. BRCA1 and BRCA2 mutations and the risk for colorectal cancer. Clin

Genet 2015;87:411-8.
55. Brose MS, Rebbeck TR, Calzone KA, et al. Cancer Risk Estimates for BRCA1 Mutation Carriers Identified in a Risk Evaluation Program. JNCI: Journal of the National Cancer Institute 2002;94:1365-1372.
56. Ford D, Easton DF, Bishop DT, et al. Risks of cancer in BRCA1-mutation carriers. The Lancet 1994;343:692-695.
57. Phelan CM, Iqbal J, Lynch HT, et al. Incidence of colorectal cancer in BRCA1 and BRCA2 mutation carriers: results from a follow-up study. Br J Cancer 2014;110:530-534.
58. Kadouri L, Hubert A, Rotenberg Y, et al. Cancer risks in carriers of the BRCA1/2 Ashkenazi founder mutations. J Med Genet 2007;44:467-71.
59. Lipkin SM, Wang V, Jacoby R, et al. MLH3: a DNA mismatch repair gene associated with mammalian microsatellite instability. Nat Genet 2000;24:27-35.
60. Chen PC, Kuraguchi M, Velasquez J, et al. Novel roles for MLH3 deficiency and TLE6-like amplification in DNA mismatch repair-deficient gastrointestinal tumorigenesis and progression. PLoS Genet 2008;4:e1000092.
61. Wu Y, Berends MJ, Sijmons RH, et al. A role for MLH3 in hereditary nonpolyposis colorectal cancer. Nat Genet 2001;29:137-8.
62. Hienonen T, Laiho P, Salovaara R, et al. Little evidence for involvement of MLH3 in colorectal cancer predisposition. Int J Cancer 2003;106:292-6.
63. Liu HX, Zhou XL, Liu T, et al. The role of hMLH3 in familial colorectal cancer. Cancer Res 2003;63:1894-9.
64. Kim JC, Roh SA, Yoon YS, et al. MLH3 and EXO1 alterations in familial colorectal cancer patients not fulfilling Amsterdam criteria. Cancer Genet Cytogenet 2007;176:172-4.
65. Goellner EM, Putnam CD, Kolodner RD. Exonuclease 1-dependent and independent mismatch repair. DNA Repair (Amst) 2015;32:24-32.
66. Tishkoff DX, Boerger AL, Bertrand P, et al. Identification and characterization of Saccharomyces cerevisiae EXO1, a gene encoding an exonuclease that interacts with MSH2. Proc Natl Acad Sci U S A 1997;94:7487-92.
67. Thompson E, Meldrum CJ, Crooks R, et al. Hereditary non-polyposis colorectal cancer and the role of hPMS2 and hEXO1 mutations. Clin Genet 2004;65:215-25.
68. Yamamoto H, Hanafusa H, Ouchida M, et al. Single nucleotide polymorphisms in the EXO1 gene and risk of colorectal cancer in a Japanese population. Carcinogenesis 2005;26:411-416.
69. Zhang M, Zhao D, Yan C, et al. Associations between Nine Polymorphisms in EXO1 and Cancer Susceptibility: A Systematic Review and Meta-Analysis of 39 Case-control Studies. Scientific Reports 2016;6:29270.
70. Jagmohan-Changur S, Poikonen T, Vilkki S, et al. EXO1 variants occur commonly in normal population: evidence against a role in hereditary nonpolyposis colorectal cancer. Cancer Res 2003;63:154-8.
71. Alam NA, Gorman P, Jaeger EE, et al. Germline deletions of EXO1 do not cause colorectal tumors and lesions which are null for EXO1 do not have microsatellite instability. Cancer Genet Cytogenet 2003;147:121-7.
72. Talseth-Palmer BA, Bauer DC, Sjursen W, et al. Targeted next-generation sequencing of 22 mismatch repair genes identifies Lynch syndrome families. Cancer Med 2016;5:929-41.
73. Chen W, Swanson BJ, Frankel WL. Molecular genetics of microsatellite-unstable colorectal cancer for pathologists. Diagnostic Pathology 2017;12:24.
74. Sekine S, Mori T, Ogawa R, et al. Mismatch repair deficiency commonly precedes adenoma formation in Lynch Syndrome-Associated colorectal tumorigenesis. Mod Pathol 2017;30:1144-1151.
75. De Jong AE, Morreau H, Van Puijenbroek M, et al. The role of mismatch repair gene defects in the development of adenomas in patients with HNPCC. Gastroenterology 2004;126:42-48.
76. Pino MS, Mino-Kenudson M, Wildemore BM, et al. Deficient DNA Mismatch Repair Is Common in Lynch Syndrome-Associated Colorectal Adenomas. The Journal of Molecular Diagnostics 2009;11:238-247.
77. Walsh MD, Buchanan DD, Pearson SA, et al. Immunohistochemical testing of conventional adenomas for loss of expression of mismatch repair proteins in Lynch syndrome mutation carriers: a case series from the Australasian site of the colon cancer family registry. Mod Pathol 2012;25:722-30.
78. Yurgelun MB, Goel A, Hornick JL, et al. Microsatellite instability and DNA mismatch repair protein deficiency in Lynch syndrome colorectal polyps. Cancer Prev Res (Phila) 2012;5:574-82.
79. Halvarsson B, Lindblom A, Johansson L, et al. Loss of mismatch repair protein immunostaining in colorectal adenomas from patients with hereditary nonpolyposis colorectal cancer. Mod Pathol 2005;18:1095-1101.
80. Iino H, Simms L, Young J, et al. DNA microsatellite instability and mismatch repair protein loss in adenomas presenting in hereditary non-polyposis colorectal cancer. Gut 2000;47:37-42.
81. Ahadova A, von Knebel Doeberitz M, Blaker H, et al. CTNNB1-mutant colorectal carcinomas with immediate invasive growth: a model of interval cancers in Lynch syndrome. Fam Cancer 2016;15:579-86.
82. Ku C-S, Cooper DN, Wu M, et al. Gene discovery in familial cancer syndromes by exome sequencing: prospects for the elucidation of familial colorectal cancer type X. Mod Pathol 2012;25:1055-1068.
83. de Voer RM, Hahn MM, Weren RD, et al. Identification of Novel Candidate Genes for Early-Onset Colorectal

Cancer Susceptibility. PLoS Genet 2016;12:e1005880.
84. Nielsen M, Hes FJ, Nagengast FM, et al. Germline mutations in APC and MUTYH are responsible for the majority of families with attenuated familial adenomatous polyposis. Clin Genet 2007;71:427-33.
85. Aretz S, Uhlhaas S, Goergens H, et al. MUTYH-associated polyposis: 70 of 71 patients with biallelic mutations present with an attenuated or atypical phenotype. Int J Cancer 2006;119:807-14.
86. Friedl W, Aretz S. Familial adenomatous polyposis: experience from a study of 1164 unrelated german polyposis patients. Hered Cancer Clin Pract 2005;3:95-114.
87. Palles C, Cazier JB, Howarth KM, et al. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nat Genet 2013;45:136-44.
88. Weren RD, Ligtenberg MJ, Kets CM, et al. A germline homozygous mutation in the base-excision repair gene NTHL1 causes adenomatous polyposis and colorectal cancer. Nat Genet 2015;47:668-71.
89. Adam R, Spier I, Zhao B, et al. Exome Sequencing Identifies Biallelic MSH3 Germline Mutations as a Recessive Subtype of Colorectal Adenomatous Polyposis. Am J Hum Genet 2016;99:337-51.
90. Hes FJ, Nielsen M, Bik EC, et al. Somatic APC mosaicism: an underestimated cause of polyposis coli. Gut 2008;57:71-6.
91. Aretz S, Stienen D, Friedrichs N, et al. Somatic APC mosaicism: a frequent cause of familial adenomatous polyposis (FAP). Hum Mutat 2007;28:985-92.
92. Tuohy TM, Burt RW. Somatic mosaicism: a cause for unexplained cases of FAP? Gut 2008;57:10-2.
93. Farrington SM, Dunlop MG. Mosaicism and sporadic familial adenomatous polyposis. Am J Hum Genet 1999;64:653-8.
94. Davidson S, Leshanski L, Rennert G, et al. Maternal mosaicism for a second mutational event--a novel deletion--in a familial adenomatous polyposis family harboring a new germ-line mutation in the alternatively spliced-exon 9 region of APC. Hum Mutat 2002;19:83-4.
95. Baert-Desurmont S, Piton N, Bou J, et al. A remarkable APC mosaicism with two mutant alleles in a family with familial adenomatous polyposis. Am J Med Genet A 2011;155A:1500-2.
96. Jones AC, Sampson JR, Cheadle JP. Low level mosaicism detectable by DHPLC but not by direct sequencing. Hum Mutat 2001;17:233-4.
97. Necker J, Kovac M, Attenhofer M, et al. Detection of APC germ line mosaicism in patients with de novo familial adenomatous polyposis: a plea for the protein truncation test. J Med Genet 2011;48:526-9.
98. Out AA, van Minderhout IJ, van der Stoep N, et al. High-resolution melting (HRM) re-analysis of a polyposis patients cohort reveals previously undetected heterozygous and mosaic APC gene mutations. Fam Cancer 2015;14:247-57.
99. Yamaguchi K, Komura M, Yamaguchi R, et al. Detection of APC mosaicism by next-generation sequencing in an FAP patient. J Hum Genet 2015;60:227-31.
100. Spier I, Drichel D, Kerick M, et al. Low-level APC mutational mosaicism is the underlying cause in a substantial fraction of unexplained colorectal adenomatous polyposis cases. J Med Genet 2016;53:172-9.
101. Jansen AM, Crobach S, Geurts-Giele WR, et al. Distinct Patterns of Somatic Mosaicism in the APC Gene in Neoplasms From Patients With Unexplained Adenomatous Polyposis. Gastroenterology 2017;152:546-549 e3.
102. Goldberg Y, Halpern N, Hubert A, et al. Mutated MCM9 is associated with predisposition to hereditary mixed polyposis and colorectal cancer in addition to primary ovarian failure. Cancer Genetics 2015;208:621-624.
103. Horpaopan S, Spier I, Zink AM, et al. Genome-wide CNV analysis in 221 unrelated patients and targeted high-throughput sequencing reveal novel causative candidate genes for colorectal adenomatous polyposis. Int J Cancer 2015;136:E578-89.
104. Spier I, Kerick M, Drichel D, et al. Exome sequencing identifies potential novel candidate genes in patients with unexplained colorectal adenomatous polyposis. Familial Cancer 2016;15:281-288.
105. Krepischi AC, Achatz MI, Santos EM, et al. Germline DNA copy number variation in familial and early-onset breast cancer. Breast Cancer Res 2012;14:R24.
106. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nat Med 2004;10:789-99.
107. Traver S, Coulombe P, Peiffer I, et al. MCM9 Is Required for Mammalian DNA Mismatch Repair. Mol Cell 2015;59:831-9.
108. Liu Q, Hesson LB, Nunez AC, et al. Pathogenic germline MCM9 variants are rare in Australian Lynch-like syndrome patients. Cancer Genet 2016;209:497-500.
109. Magalhaes B, Peleteiro B, Lunet N. Dietary patterns and colorectal cancer: systematic review and meta-analysis. Eur J Cancer Prev 2012;21:15-23.
110. Aleksandrova K, Pischon T, Jenab M, et al. Combined impact of healthy lifestyle factors on colorectal cancer: a large European cohort study. BMC Med 2014;12:168.
111. Mehra K, Berkowitz A, Sanft T. Diet, Physical Activity, and Body Weight in Cancer Survivorship. Med Clin North Am 2017;101:1151-1165.

112. Kerr J, Anderson C, Lippman SM. Physical activity, sedentary behaviour, diet, and cancer: an update and emerging new evidence. Lancet Oncol 2017;18:e457-e471.
113. Song M, Garrett WS, Chan AT. Nutrients, Foods, and Colorectal Cancer Prevention. Gastroenterology 2015;148:1244-1260.e16.
114. Bradbury KE, Appleby PN, Key TJ. Fruit, vegetable, and fiber intake in relation to cancer risk: findings from the European Prospective Investigation into Cancer and Nutrition (EPIC). Am J Clin Nutr 2014;100 Suppl 1:394s-8s.
115. Garrett WS. Cancer and the microbiota. Science 2015;348:80-6.
116. Mima K, Ogino S, Nakagawa S, et al. The role of intestinal bacteria in the development and progression of gastrointestinal tract neoplasms. Surgical Oncology 2017;26:368-376.
117. Dubinkina VB, Tyakht AV, Odintsova VY, et al. Links of gut microbiota composition with alcohol dependence syndrome and alcoholic liver disease. Microbiome 2017;5:141.
118. Dudbridge F, Brown SJ, Ward L, et al. How many cases of disease in a pedigree imply familial disease? Ann Hum Genet 2017.
119. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015;136:E359-86.

Chapter 8

Summary/Nederlandse samenvatting
Curriculum vitae
List of publications
Dankwoord

Summary

Colorectal cancer (CRC) is the third most common cancer in men and second most common cancer in women worldwide. An estimated 20-30% of all CRC patients have a positive family history of colorectal cancer, which in 3-5% of all CRCs can be explained by inherited germline variants in highly penetrant CRC genes. The most common form of hereditary CRC is Lynch Syndrome, caused by pathogenic germline variants in the mismatch repair (MMR) genes, *MLH1, MSH2, MSH6* and *PMS2.* MMR-deficient tumors characteristically show instability in the microsatellites, small repeated sequences in the DNA, and immunohistochemical loss of the MMR gene that is mutated. Additionally, in approximately 20-25% of patients with immunohistochemical loss of MSH2 but without a germline *MSH2* variant a germline deletion in the *EPCAM* gene is found. This deletion results in transcriptional inactivation of the *MSH2* gene, that is directly upstream from *EPCAM.* MMR-deficiency can also occur sporadically, through *MLH1* promoter hypermethylation, shown in >85% of MMR-deficient tumors. Still, in up to 60% of patients with MMR-deficient tumors without *MLH1* promoter hypermethylation no germline pathogenic MMR gene variant is detected. These patients are referred to as 'suspected Lynch Syndrome' (sLS) patients and clinical management of these patients and their families remains difficult. This thesis focusses on explaining the MMR deficiencies and finding underlying genetic causes in these patients and their relatives. We hypothesize that these sLS patients can be explained by (1) missed variants in the MMR genes, (2) biallelic somatic inactivation of the MMR genes or (3) variants in other genes that subsequently lead to secondary MMR-deficiency.

Accordingly, in **Chapter 2**, we describe an effort to detect missed germline intronic or promoter variants in the *MLH1, MSH2, MSH6* and *PMS2* genes, but also germline variants in other CRC susceptibility genes, such as *MUTYH, BMRP1A, PTEN* and *APC.* Whole Gene Capture on leukocyte DNA of 45 sLS patients showed 1979 germline variants, of which the majority (97%) was intronic. One patient was found to carry a missed variant in *MLH1,* resulting in a 29 amino acid incorporation in the protein-interacting domain. This patient showed solitary PMS2 protein loss in the tumor, and *MLH1* had not been screened before. Additionally, germline variants of uncertain significance (VUS) were found in *EPCAM, MSH3, MUTYH* and *AXIN1,* but no further testing was done to assess the functional relevance of these variants.

In **Chapter 3** we describe a more prevalent explanation for these suspected Lynch Syndrome patients. Of 62 sLS patients leukocyte and tumor DNA was analysed with next-generation sequencing (NGS) to detect somatic and germline variants in the MMR genes, but also the *POLE* and *POLD1* genes. Variants in the exonuclease domain (EDM) of these genes encoding for polymerase ε and δ respectively, had been described in microsatellite stable tumors before but the prevalence of these type of variants in MMR-deficient/MSI-H tumors was unknown. With NGS 10 tumors were found to carry two somatic aberrations (two pathogenic variants or one pathogenic variant with concomittant loss of heterozygosity) in the MMR gene that showed immunohistochemical loss. Additionally, in nine patients a germline- (n=2) or somatic (n=7) variant was detected in the EDM of *POLE* or *POLD1.* All *POLE/POLD1* mutated tumors seemed to show a hypermutated phenotype, concordant with previous studies on *POLE/POLD1* EDM variants.

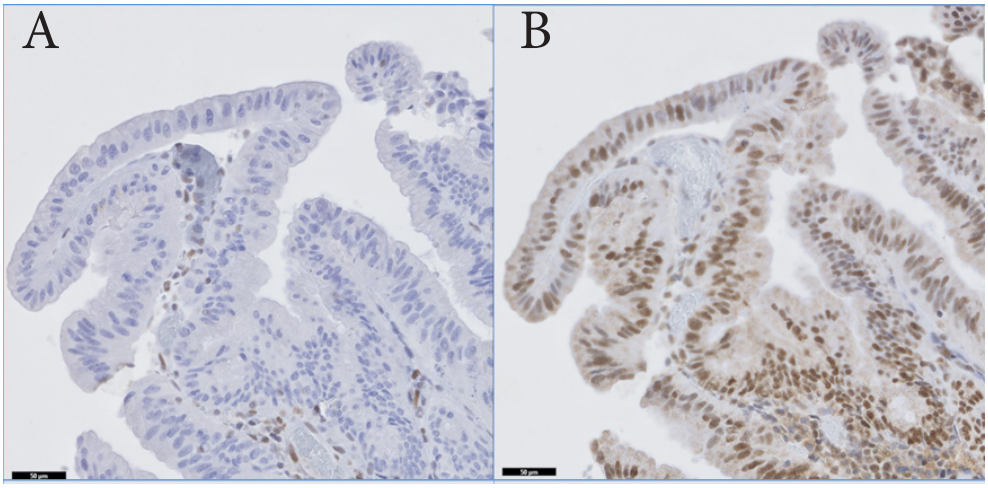
In **Chapter 4** we present a method to verify the predicted splicing effect of splice site variants found in *MLH1, MSH2, MSH6*, *APC* and *BRCA1*. For 11 variants RNA was isolated from formalin-fixed paraffin-embedded (FFPE) tissue blocks, and for two additional variants RNA was isolated from EBV-transformed B-cells. Of the total 13 splice site variants, eight variants were previously described to result in splicing, while five were only predicted to result in splicing, but functional evidence was lacking. For all variants specific primers were designed, and cDNA was synthesized and analysed with PCR. For six out of eight samples carrying known splice site variants RNA could be successfully isolated, and all six showed the previously reported aberrant splicing. Notable, one variant for which only RNA from EBV-transformed cell lines was available only showed an aberrantly spliced product after formalin fixation of the cells, indicating that formalin fixation inhibits RNA degradation. For four of the five variants predicted to result in aberrant splicing of RNA cDNA products could successfully be amplified and aberrant splicing was seen in three. It could be concluded that these variants should be regarded as (likely) pathogenic and not of uncertain significance. A simple addition to the analysis toolkit can thus answers the potential pathogenicity of variants involved in aberrant splicing.

In **Chapter 5** we describe a practical guide to detect and analyse variants in *PMS2* in DNA isolated from FFPE tissue. DNA isolated from FFPE is often very fragmented, and the standard diagnostic routine for sequencing leukocyte DNA is not feasible in these samples. This brings extra complexity for the analysis of *PMS2*, which in itself is already complex due to the presence of multiple pseudogenes. Fourteen *PMS2*-pseudogenes share a high homology with the 5' end of *PMS2* (exon 1 to 5), while a fifteenth pseudogene (*PMS2CL*) shares high homology with *PMS2* exon 9 and exon 11 to 15. *PMS2* is distinguishable through paralogous sequence variants (PSVs), a small number of nucleotides that are specific for *PMS2* and not present in the pseudogenes. By designing a custom NGS FFPE-suitable library with small amplicons containing one or two PSVs each, exon 1 to exon 11 can be reliably sequenced.

In **Chapter 6** we aimed to detect the underlying genetic cause in unexplained adenomatous polyposis patients negative for germline defects using conventional testing for CRC susceptibility genes. DNA of two or more colonic adenomas of each patient was tested for *APC* variants with NGS, with the hypothesis that if multiple adenomas carry an identical *APC* variant, this might indicate underlying mosaicism. In nine of 18 patients, and two positive controls, all with 21 to 100 adenomas an identical *APC* variant was detected in multiple adenomas. Testing of DNA from different germ layers (mesoderm, ectoderm and endoderm) could help identify the approximate time point at which the mosaicism arose. Different patterns of mosaicism were identified. In three patients the *APC* variant present in the adenomas could be detected in leukocyte DNA with a very low variant allele frequency, in one patient the *APC* variant was confined to the colon and in six patients the variant was not present in leukocyte, and no normal colonic mucosa was available or testing was not possible. In one patient the variant was only detected in adenomas but not in leukocyte DNA or normal colonic mucosa. In this patient we propose an underlying mechanism of field cancerization where one tumorigenic clone migrates through the colon. Lastly, there was one patient that had mixed mosaic adenomas with sporadic adenomas with unique APC variants.

Nederlandse samenvatting

Darmkanker is de op twee na meest voorkomende kanker bij mannen en de op één na meest voorkomende kanker bij vrouwen. Ongeveer 20-30% van alle darmkankerpatiënten heeft een familiegeschiedenis van darmkanker en 3-5% van alle patiënten kan verklaard worden door een verandering in het DNA (mutatie) in één van de bekende darmkankergenen. De meest voorkomende vorm van erfelijke darmkanker is Lynch Syndroom en wordt veroorzaakt door mutaties in de DNA-herstelgenen: *MLH1, MSH2, MSH6* en *PMS2.* Van deze genen worden de DNA-hersteleiwitten afgelezen, die er voor zorgen dat foutjes in het DNA worden gevonden en worden hersteld. Een mutatie in het gen kan ervoor zorgen dat er geen eiwit wordt gemaakt, wat er voor zorgt dat foutjes in het DNA niet meer hersteld worden. Om te kijken of de DNA -hersteleiwitten aanwezig zijn in een tumor, wordt er een kleuring gedaan. Hierbij wordt een stofje toegevoegd dat zich bindt aan één specifiek eiwit. Als het eiwit aanwezig is, en het stofje zich kan binden, is dat te zien aan een bruine kleur (Figuur 1B). Als het eiwit niet aanwezig is, is alleen de blauwe achtergrondkleuring te zien (Figuur 1A).



Figuur 1- Kleuring van (A) MLH1 en (B) MSH2.
Bruin laat zien dat het eiwit aanwezig is, terwijl blauw de achtergrondkleuring is om de cellen zichtbaar te maken. A en B laten dezelfde tumor zien, maar met negatieve kleuring voor MLH1 (A) en positieve kleuring voor MSH2 (B).

Negatieve kleuring – het ontbreken van een kleur – duidt op het ontbreken van het eiwit. Dit betekent meestal een onderliggende erfelijke mutatie. Tumoren met een DNA-hersteldefect hebben daarnaast instabiliteit van hun DNA-microsatellieten. Microsatellieten zijn kleine stukjes DNA met een herhalend patroon, zoals CACACACA. Als er een defect is in de DNA-herstelfunctie, kan de lengte van deze microsatellieten veranderen, iets wat we microsatelliet-instabiliteit of MSI noemen. Somatische, niet erfeljke, mutaties – mutaties die tijdens het leven ontstaan en niet van de ouders komen – kunnen ook zorgen voor verlies van de DNA-herstelfunctie. Een veelvoorkomende somatische verandering is hypermethylatie van de promoter van *MLH1.* Dit houdt in dat het eerste stukje van het *MLH1*-gen (de promoter) wordt afgedekt met methylgroepen (CH_3-groepen). Daardoor wordt het hele gen geïnactiveerd, en wordt er geen eiwit gemaakt. Meer dan 85% van de tumoren waarin we geen MLH1-eiwit

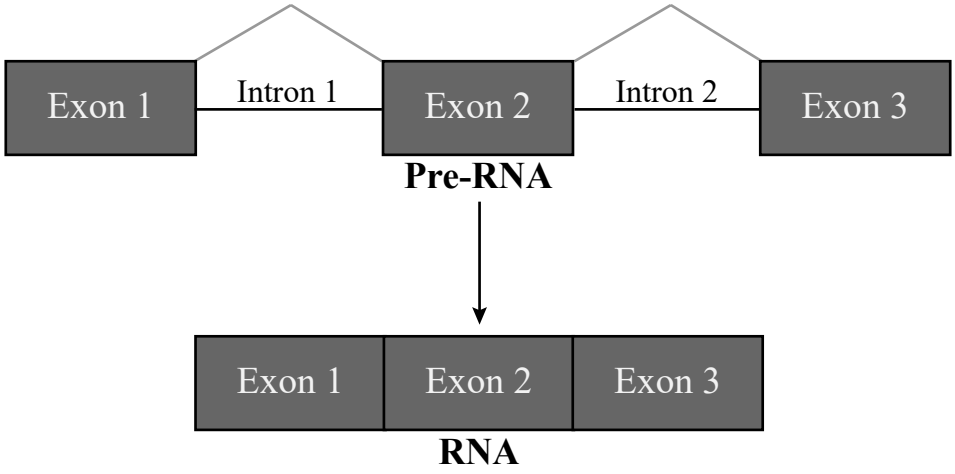
kunnen detecteren, kunnen verklaard worden door *MLH1* promoter hypermethylatie. In 60% van de tumoren met negatieve kleuring voor de DNA-hersteleiwitten en/of MSI wordt geen erfelijke mutatie en geen *MLH1* promoter hypermethylatie gevonden. Deze patiënten noemen we 'verdacht Lynch Syndroom'-patiënten. Omdat de oorzaak van de tumoren in deze patiënten onbekend is, is het moeilijk om screeningsrichtlijnen op te stellen. Ook kunnen familieleden niet getest worden op dragerschap van een erfelijke mutatie. Dit proefschrift richt zich op het vinden van de onderliggende erfelijke oorzaak van de darmkanker van deze 'verdacht Lynch Syndroom'-patiënten en hun familie. We stellen dat dit kan komen door: (1) mutaties in de DNA-herstelgenen die voorheen gemist zijn, (2) andere (somatische) manieren van inactivatie van de DNA-herstelgenen, (3) mutaties in andere genen die microsatelliet instabiliteit kunnen veroorzaken.

Zodoende beschrijft **hoofdstuk 2** een poging tot detectie van gemiste erfelijke mutaties in de DNA-herstelgenen. DNA bestaat uit coderende gedeeltes (exonen) en niet-coderende gedeeltes (intronen). Van dit DNA (een dubbele helix) wordt eerst RNA gemaakt (een directe kopie van één van de twee strengen), waarna de intronen eruit 'geknipt' worden. Het RNA dat overblijft wordt 'gelezen' en vertaald naar aminozuren, die in een lange keten een eiwit vormen. Vaak wordt gedacht dat de niet-coderende gedeeltes niet belangrijk zijn voor het uiteindelijke eiwit, en wordt alleen gekeken naar mutaties in het coderende gedeelte: de exonen. Om te kijken of we daardoor mutaties missen die de darmkanker wel kunnen verklaren, hebben we gekeken naar erfelijke mutaties in de exonen én intronen van de DNA-herstelgenen: *MLH1, MSH2, MSH6* en *PMS2* maar ook in elf andere bekende darmkankergenen. Door deze 15 darmkankergenen te analyseren bij 45 verdacht Lynch Syndroom patiënten, vonden we 1979 mutaties, waarvan het grootste gedeelte (97%) in het niet-coderende gedeelte van de genen lag. In één patiënt werd een eerder gemiste erfelijke mutatie in het *MLH1*-gen gevonden. Deze mutatie zorgde voor 29 extra aminozuren in het MLH1 eiwit waardoor het niet meer met PMS2 kon binden, en niet meer als DNA-hersteleiwit kon functioneren. Omdat kleuring van de tumor alleen liet zien dat er geen PMS2 eiwit was, was er voorheen niet getest op mutaties in het *MLH1*-gen. Naast deze mutatie werden erfelijke mutaties met onbekend effect gevonden in vier andere darmkankergenen, maar er werden geen verdere testen gedaan naar de het effect van deze mutaties.

In **hoofdstuk 3** beschrijven we een veelvoorkomende verklaring voor deze 'verdacht Lynch Syndroom'-tumoren, waarbij we de tumoren testen voor somatische mutaties. Recentelijk is ontdekt dat als er twee somatische mutaties ontstaan in hetzelfde DNA-herstelgen, dit leidt tot dezelfde soort tumoren als verdacht Lynch Syndroom patiënten hebben. Of verdacht Lynch Syndroom patienten met een familiegeschiedenis ook verklaard konden worden door twee somatische mutaties, was voorheen niet bekend. In de studie beschreven in **hoofdstuk 3** werden 62 verdacht Lynch Syndroompatiënten getest op mutaties in het DNA geïsoleerd uit bloed (op zoek naar erfelijke mutaties) én de tumor (op zoek naar somatische mutaties) in de DNA-herstelgenen, maar ook in de *POLE* en *POLD1* genen. Mutaties in deze genen zijn eerder beschreven in een ander type darmkanker, maar zijn niet eerder onderzocht in tumoren met Lynch Syndroomkarakteristieken. Door analyse vonden we in 10 patiënten twee somatische mutaties in het DNA-herstelgen, dat verlies van kleuring liet zien. Daarnaast hadden negen patiënten een mutatie in het *POLE* of *POLD1*-gen, waarvan twee erfelijke- en zeven somatische mutaties waren. Alle tumoren met een mutatie in het *POLE*- of *POLD1*-

gen lieten heel veel somatische DNA-mutaties zien, iets wat eerder beschreven is in dit type tumoren en we 'gehypermuteerd' noemen.

In **hoofdstuk 4** is onderzoek gedaan naar mutaties die een effect hebben op RNA splicing. Als DNA wordt afgelezen wordt eerst pre-RNA gemaakt met alle intronen er nog in. Deze intronen moeten eruit geknipt worden om RNA te maken. Het eruit knippen van de intronen noemen we splicing (Figuur 1). Mutaties aan de rand van het exon of het intron kunnen een effect hebben op de splicing. Ze kunnen ervoor zorgen dat de exon-intron rand niet goed herkend wordt en er niet correct geknipt wordt. Zo kan het gebeuren dat er per ongeluk een exon uitgeknipt wordt, of dat een intron blijft zitten. In de studie in **hoofdstuk 4** hebben we gekeken naar varianten die op de rand van het intron of exon liggen. Hierbij zijn niet alleen de DNA-herstelgenen geanalyseerd, maar ook *BRCA1* (een bekend borstkanker gen) en *APC* (een bekend darmkankergen). Voor 11 mutaties werd RNA geïsoleerd uit tumormateriaal. Dit tumormateriaal was gefixeerd in formaline en daarna omhuld in paraffine, om het te kunnen bewaren. Sommige tumoren waren al meer dan 20 jaar oud. Daarnaast werd voor twee mutaties RNA geïsoleerd uit gekweekte bloedcellen. Acht van de 13 mutaties waren bekende mutaties, en eerdere studies lieten zien dat deze invloed hebben op splicing. Van de overige vijf mutaties was door predictieprogramma's voorspeld dat ze splicing beïnvloedden, maar dit was nog niet bewezen. Van de acht bekende mutaties kon voor zes mutaties RNA geïsoleerd worden. Alle zes lieten het splice-effect zien zoals eerder beschreven. Eén van deze mutaties, waarvan het RNA geïsoleerd was uit de gekweekte bloedcellen, liet alleen het specifieke splice-effect zien nadat de gekweekte cellen gefixeerd waren met formaline. Dit laat zien dat formaline de afbraak van RNA tegengaat. Van de vijf mutaties waarvan voorspeld was



Figuur 2: Splicing
Pre-RNA, nog bestaande uit exonen en intronen, wordt omgezet in RNA door de intronen ertussenuit te knippen.

dat ze invloed hadden op splicing, kon voor vier mutaties RNA geïsoleerd worden, en werd een afwijkende splicing aangetoond voor drie mutaties.

In **hoofdstuk 5** beschrijven we hoe mutaties in het *PMS2*-gen gedetecteerd en geanalyseerd kunnen worden. Het gaat hierbij om mutaties in DNA geïsoleerd uit formaline gefixeerd tumorweefsel dat ingebed is in paraffine. DNA uit dit weefsel is vaak erg gefragmenteerd, waardoor de standaardmethodes die gebruikt worden voor DNA uit bloed niet te gebruiken zijn op dit materiaal. Dit maakt de analyse van *PMS2* complexer, terwijl *PMS2* op zichzelf al een complex gen is. *PMS2* heeft namelijk veel pseudogenen. Een pseudogen is een gen dat heel erg lijkt op een ander gen, maar vaak niet functioneel is, en waar vaak geen eiwit van wordt gemaakt. *PMS2* heeft 14 pseudogenen met bijna dezelfde sequentie als de sequentie van *PMS2* exon 1 tot exon 5, en één pseudogen dat bijna hetzelfde is als *PMS2* exon 9 en exon 11 tot 15. *PMS2* is te onderscheiden van zijn pseudogenen door 'paraloge sequentie variaties' (PSVs): een klein aantal specifieke nucleotides die wel in *PMS2* zitten, maar niet in de pseudogenen. We analyseren het DNA stukje voor stukje, en door ervoor te zorgen dat elk stukje zo'n PSV bevat, kunnen we zeggen of het bij *PMS2* of bij één van de pseudogenen hoort.

In **hoofdstuk 6** wilden we de erfelijke oorzaak in een groep patiënten met onverklaarde polyposis vinden. Polyposispatiënten hebben veel poliepen in de darm, soms wel meer dan 1000. Vaak hebben deze patiënten een mutatie in het *APC*-gen. In het verleden zijn ook patiënten beschreven met een mozaïeke mutatie in *APC.* Een mozaïeke mutatie is een mutatie die ergens tijdens de ontwikkeling is ontstaan, vaak in de embryonale fase. Hierdoor zit de mutatie niet in elke cel in het lichaam, maar slechts in een deel van de cellen. Om te zoeken naar een mozaïeke mutatie werd er voorheen altijd gekeken naar bloed, en ging men op zoek naar een mutatie die aanwezig was in een deel van de cellen. In deze studie zijn 27 polyposis patiënten zonder *APC*-mutatie onderzocht of zij drager waren van een mozaïeke *APC*-mutatie. Hierbij werd niet gekeken naar DNA uit bloed, maar juist naar DNA geïsoleerd uit de poliepen. De hypothese was dat als twee of meerdere poliepen dezelfde mutatie hadden, er een onderliggende mozaïeke mutatie kon zijn. In negen van de 18 patiënten met 21 tot 100 poliepen in de darm, en in twee patiënten waarvan we wisten dat ze een mozaïeke *APC*-mutatie droegen, konden we dezelfde *APC*-mutatie vinden in meerdere poliepen. Om te kijken of deze mozaïeke mutatie in het hele lichaam zat, of alleen in de darmen, hebben we DNA uit verschillende weefsels (bloed, speeksel, normaal darmweefsel) getest op de *APC*-mutatie aanwezig in de poliepen. In drie patiënten was de mutatie die in de poliepen zat ook terug te vinden in DNA uit bloed, in 1 patiënt zat de mutatie niet in bloed maar wel in normaal darmweefsel en in zes patiënten zat de mutatie niet in DNA uit bloed en was geen DNA uit normaal darmweefsel aanwezig, of was testen van dit weefsel niet mogelijk. In één patiënt werd een mutatie gevonden in de poliepen die niet aanwezig was in het normale darmweefsel. Voor deze patiënt verwachten we dat 1 gemuteerde cel zich verspreidde door de darm, zich deelde en uiteindelijk uitgroeide tot een poliep. Uit deze poliep liet dan weer een cel los die zich verder verspreidde zodat er meerdere poliepen ontstonden.

Curriculum Vitae

Anne Maria Lucia Jansen werd geboren op 24 april 1990 te Zoetermeer. In 2008 behaalde zij haar vwo-diploma aan het Alfrink College te Zoetermeer, waarna zij met haar bachelor 'Life Science and Technology' begon aan de Universiteit Leiden en Technische Universiteit Delft. Tijdens haar bachelor volgde ze de minor Educatie, waarbij ze haar tweedegraads onderwijsbevoegdheid behaalde voor het vak scheikunde. Na afronding van haar bachelor in 2011, begon ze met haar master 'Life Science and Technology' aan de Universiteit Leiden. Hier volgde ze, naast een onderzoeksprogramma, het Science Communication and Society programma, waarbij ze tijdens haar stage onderzoek deed naar de kwaliteit van patiënteninformatie gericht aan kinderen. Haar onderzoeksstage liep ze in de groep van Prof. Dr. Peter Devilee in het Leids Universitair Medisch Centrum (LUMC). Tijdens deze stage van negen maanden deed ze onderzoek naar de pathogeniteit van genvarianten met behulp van het minigene splicing assay, onder supervisie van Heleen van der Klift. Na het behalen van haar masterdiploma, was ze practicumassistent voor het practicum Biochemie 1 van de studie Biofarmaceutische wetenschappen tot ze in februari 2014 haar PhD startte. Haar PhD-project voerde zij uit in de groep van Prof. Dr. Hans Morreau en Dr. Tom van Wezel bij de afdeling Pathologie en in de groep van Prof. Dr. Peter Devilee bij de afdeling Humane Genetica, beide verbonden aan het LUMC. Eveneens kreeg zij begeleiding van Dr. Frederik J. Hes en Dr. Juul T. Wijnen. Het resultaat van dit onderzoek is beschreven in dit proefschrift. Verder bezocht Anne, in het kader van haar promotietraject, verscheidene congressen in het binnen- en buitenland, waar ze de kans kreeg haar werk te presenteren aan een internationaal publiek. Anne zal haar carrière voortzetten als postdoc in de onderzoeksgroep van Prof. Dr. Ajay Goel in het Baylor Scott & White Health Institute in Dallas, Texas in de Verenigde Staten.

List of publications

Jansen AML, van der Klift HM, Roos MAE, van Eendenburg JDH, Tops CMJ, Wijnen JT, Hes FJ, Morreau H, van Wezel T. RNA analysis of cancer predisposing genes in formalin-fixed paraffin-embedded tissue determines aberrant splicing. *Eur J Hum Genet* in press.

Crobach S, **Jansen AML**, Ligtenberg MJL, Koopmans M, Nielsen M, Hes FJ, Wijnen JT, Dinjens WNM, van Wezel T, Morreau H. Excluding Lynch syndrome in a female patient with metachronous DNA mismatch repair deficient colon- and ovarian cancer. *Fam cancer* 2017.

Jansen AML, Crobach S, Geurts-Giele WR, van den Akker BE, Garcia MV, Ruano D, Nielsen M, Tops CM, Wijnen JT, Hes FJ, van Wezel T, Dinjens WN, Morreau H. Distinct Patterns of Somatic Mosaicism in the APC Gene in Neoplasms From Patients With Unexplained Adenomatous Polyposis. *Gastroenterology* 2017;152:546-549 e3.

Stelloo E, **Jansen AML**, Osse EM, Nout RA, Creutzberg CL, Ruano D, Church DN, Morreau H, Smit VT, van Wezel T, Bosse T. Practical guidance for mismatch repair-deficiency testing in endometrial cancer. *Ann Oncol* 2017;28(1):96-102.

Jansen AML, van Wezel T, van den Akker BE, Ventayol Garcia M, Ruano D, Tops CM, Wagner A, Letteboer TG, Gómez-Garcia EB, Devilee P, Wijnen JT, Hes FJ, Morreau H. Combined mismatch repair and POLE/POLD1 defects explain unresolved suspected Lynch syndrome cancers. *Eur J Hum Genet* 2016;24(7):1089-92

Jansen AML, Geilenkirchen MA, van Wezel T, Jagmohan-Changur SC, Ruano D, van der Klift HM, van den Akker BE, Laros JF, van Galen M, Wagner A, Letteboer TG, Gómez-Garcia EB, Tops CM, Vasen HF, Devilee P, Hes FJ, Morreau H, Wijnen JT. Whole Gene Capture Analysis of 15 CRC Susceptibility Genes in Suspected Lynch Syndrome Patients. *PLoS One* 2016;11:e0157381.

van der Klift HM, **Jansen AML**, van der Steenstraten N, Bik EC, Tops CM, Devilee P, Wijnen JT. Splicing analysis for exonic and intronic mismatch repair gene variants associated with Lynch syndrome confirms high concordance between minigene assays and patient RNA analyses. *Mol Genet Genomic Med* 2015;3:327-45.

Dankwoord

In dit dankwoord wil ik iedereen van harte bedanken die een bijdrage heeft geleverd aan de totstandkoming van dit proefschrift. Enkele mensen wil ik in het bijzonder bedanken voor hun bijdragen.

Allereerst mijn wil ik mijn uitgebreide begeleidingsteam bedanken: promotoren prof. dr. Hans Morreau en prof. dr. Peter Devilee, co-promotor dr. Frederik Hes, dr. Juul Wijnen en dr Tom van Wezel. Bedankt voor jullie vertrouwen in mij, en voor alles wat jullie mij de afgelopen jaren hebben geleerd. Bedankt dat jullie mij de kans gaven vanuit een klinisch en pathologisch oogpunt naar darmkanker en polyposis te kunnen kijken.

Alle co-auteurs wil ik bedanken voor hun input en de kritische beoordeling van de manuscripten. Marina, bedankt voor al je hulp met verschillende technische aspecten van mijn projecten. Dina en Jan, bedankt voor alle onmisbare bio-informatische hulp. Carli, Maartje en Sanne, bedankt voor jullie klinische blik en onze afdelingoverschrijdende meetings.

Zonder al het beschikbare darmkankerweefsel kon dit onderzoek niet worden uitgevoerd. Hartelijk dank aan alle patiënten voor het deelnemen aan wetenschappelijk onderzoek, en alle specialisten die betrokken waren in het verzamelen en het inbedden van het weefsel.

Tijdens mijn PhD had ik het geluk met veel verschillende mensen te mogen werken, collega's die het dagelijkse werk een stuk beter maakten. Ik wil iedereen van de afdeling Pathologie bedanken voor de tijd die ik op deze afdeling heb morgen doorbrengen. Bedankt aan iedereen uit de tumorgenetica groep, waar ik me altijd welkom voelde. Ook hartelijk dank voor iedereen van de moleculaire diagnostiek voor het helpen met zoeken van samples, hoe vaak ik het ook vroeg. Bedankt voor alle leden van de kantinecommissie, voor de leuke feestdagen en labuitjes die ik met jullie heb morgen organiseren. En in het bijzonder wil ik de collega's van kamer P1-39 bedanken: Noel, Arnoud, Johnny, Yvonne, Elleke, Fanny, Karin en Marie, bedankt voor alle hulp, alle discussies, dineetjes en reizen. Marie, thank you for bringing your gigantic enthusiasm and heart to our room. Bedankt mijn lieve paranimfen, Marieke en Lisa, voor al jullie hulp en steun.

Ten slotte wil ik mijn familie bedanken. Dank jullie voor jullie steun en liefde, zelfs toen mijn werk mij meenam naar de andere kant van de wereld. Malou en Britta, bedankt voor jullie commentaar op de Nederlandse samenvatting. Als laatste Tom, bedankt dat je er altijd voor me bent. Waar de toekomst ons ook brengt, samen kunnen we alles aan.