Reduced mismatch repair gene expression and functional deficiency as indicators of Lynch syndrome

Mariann Kasela

Molecular and Integrative Biosciences Research Programme

Faculty of Biological and Environmental Sciences

University of Helsinki, Helsinki, Finland

And

Doctoral Programme in Integrative Life Science

University of Helsinki, Helsinki, Finland

ACADEMIC DISSERTATION

To be presented for public examination with the permission of the Faculty of Biological and Environmental Sciences of the University of Helsinki in the auditorium **P673** of Porthania, Yliopistonkatu 3, Helsinki, on the **26th** of **June 2020** *(original examination scheduled 24th of April 2020, postponed due to global COVID-19 pandemic)* at 12 o'clock noon.

Dissertationes Scholae Doctoralis Ad Sanitatem Investigandam Universitatis Helsinkiensis 22/2020

The Faculty of Biological and Environmental Sciences, University of Helsinki, uses the Urkund system for plagiarism recognition to examine all doctoral dissertations.

Cover layout by: Anita Tienhaara
Cover image: DNA mismatch repair under the magnifying glass
Cover image by: Anni Dohlen (2020)

ISBN 978-951-51-5926-7 (paperback)
ISBN 978-951-51-5927-4 (PDF)
ISSN 2342-3161 (paperback)
ISSN 2342-317X (PDF)
http://ethesis.helsinki.fi
Printing house: Unigrafia Oy, Helsinki, Finland (2020)

Supervisors	Professor Minna Nyström, Ph.D. Molecular and Integrative Biosciences Research Programme Faculty of Biological and Environmental Sciences University of Helsinki, Helsinki, Finland Minttu Kansikas, Ph.D. Molecular and Integrative Biosciences Research Programme Faculty of Biological and Environmental Sciences University of Helsinki, Helsinki, Finland
Thesis Committee	Docent Katarina Pelin, Ph.D. Molecular and Integrative Biosciences Research Programme Faculty of Biological and Environmental Sciences University of Helsinki, Helsinki, Finland Research Director Juha Klefström, Ph.D. Translational Cancer Medicine Research Programme Research Programs Unit Faculty of Medicine University of Helsinki, Helsinki, Finland
Pre-Examiners	Professor Minna Pöyhönen, M.D., Ph.D. Department of Clinical Genetics HUSLAB HUS Diagnostic Center Helsinki University Hospital Helsinki, Finland and Department of Medical and Clinical Genetics Medicum University of Helsinki, Helsinki, Finland Professor Jukka Partanen, Ph.D. Research Director Finnish Red Cross Blood Service, Helsinki, Finland
Opponent	Professor Karl Heinimann, M.D., Ph.D. Institute for Medical Genetics and Pathology University Hospital Basel, Basel, Switzerland and Research Group Human Genomics, Department of Biomedicine, University of Basel, Basel, Switzerland
Custos	Professor Minna Nyström, Ph.D. Molecular and Integrative Biosciences Research Programme Faculty of Biological and Environmental Sciences University of Helsinki, Helsinki, Finland

"We're drowning in information and starving for knowledge."

Rutherford D. Rodgers

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

The thesis is based on the following original publications, referred to in the text by their roman numerals (**I-III**). The following articles are printed with the permission of the copyright holders.

I. Kansikas M, **Kasela M**, Kantelinen J, Nyström M. 2014. Assessing how reduced expression levels of the mismatch repair genes *MLH1*, *MSH2*, and *MSH6* affect repair efficiency. Hum Mutat 35(9):1123-7.

II. Tricarico R, **Kasela M**, Mareni C, Thompson BA, Drouet A, Staderini L, Gorelli G, Crucianelli F, Ingrosso V, Kantelinen J, Papi L, De Angioletti M, Berardi M, Gaildrat P, Soukarieh O, Turchetti D, Martins A, Spurdle AB, Nyström M, Genuardi M; InSiGHT Variant Interpretation Committee. 2017. Assessment of the InSiGHT Interpretation Criteria for the Clinical Classification of 24 *MLH1* and *MSH2* Gene Variants. Hum Mutat 38(1):64-77.

III. Kasela M, Nyström M, Kansikas M. 2019. *PMS2* expression decrease causes severe problems in mismatch repair. Hum Mutat 40(7):904-907.

ABBREVIATIONS

ACI	Amsterdam I criteria
ACII	Amsterdam II criteria
ACTB	Beta-actin
ADP	Adenosine diphosphate
APC	Adenomatous polyposis gene
AR%	Absolute repair percentage
ATP	Adenosine triphosphate
BG	Bethesda guidelines
CMMRD	Constitutional mismatch repair deficiency
CRC	Colorectal cancer
EC	Endometrial cancer
EPCAM	Epithelial cell adhesion molecule
EXO1	Exonuclease 1
FAP	Familial adenomatous polyposis
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GHKL	Gyrase, Hsp90, Histidine Kinase, MutL protein domain
GST	Glutathione S-transferase
GOI	Gene of interest
HNPCC	Hereditary nonpolyposis colorectal cancer
HPRT1	Hypoxanthine phosphoribosyltransferase 1
ICG-HNPCC	International Collaborative Group on HNPCC
IDLs	Insertion deletion loops
IHC	Immunohistochemistry
InSiGHT	International Society for Gastrointestinal Hereditary Tumours
KD	Knockdown
LOVD	Leiden Open Variation Database
LS	Lynch syndrome
MAP	MUTYH-associated polyposis
MIM	Mendelian Inheritance in Man
MLH	MutL homologue
MMR	Mismatch repair
MSH	MutS homologue
MSI	Microsatellite instability
MSI-L	Low microsatellite instability
MSI-H	High microsatellite instability
MSS	Microsatellite stable
NCI	National Cancer Institute
PCR	Polymerase chain reaction
PCNA	Proliferation cell nuclear antigen
PMS	Postmeiotic segregation increased
RFC	Replication factor C
RPA	Replication protein A
RR%	Relative repair percentage
Sf9	Spodoptera frugiperda 9 insect cells
STD	Standard deviation
VUS	Variant of uncertain significance
WT	Wild type

AMINO ACID ABBREVIATIONS

Amino acid	Three letter code	One letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

SUMMARY

Lynch syndrome (LS, previously known as hereditary non-polyposis colorectal cancer, HNPCC) is an inherited cancer predisposition syndrome caused by DNA mismatch repair (MMR) malfunction. MMR mechanism is a post-replicative repair pathway. The mismatch in the DNA is recognised and bound by MutSα, heterodimer of the proteins MSH2 and MSH6, or less often by MutSβ (MSH2+MSH3), after which the recruitment of MutLα (MLH1+PMS2) heterodimer initiates the repair process. Up to 90% of LS causing mutations are found in *MLH1*, *MSH2* and *MSH6* genes, whereas *PMS2* has been suggested to be only a low-risk LS susceptibility gene due to small number of families segregating a disease causing *PMS2* mutation. Cancer predisposition in Lynch syndrome is inherited dominantly through one defective MMR gene allele and tumorigenesis starts only after the loss of the second allele, giving rise most commonly to early-onset colorectal and endometrial cancers, and more rarely cancers of the uterine, stomach, urinary tract, ovary, small intestine or bile tract. The risk of developing colorectal cancer is higher in *MLH1* and *MSH2* mutation carriers than in *MSH6* and *PMS2* mutation carriers. However, the risk of developing endometrial cancer seems to be the highest in *MSH6* mutation carriers. Early diagnosis of LS families and MMR gene mutation carriers is extremely important, since risk-reducing clinical surveillance and prophylactic surgeries have been shown to reduce cancer-related mortality. LS diagnosis is generally based on cancer history of the family and on tumour studies, followed by genetic testing to determine a predisposing mutation. However, the atypical clinical phenotypes such as late age at onset, lower penetrance and different tumour spectrums associated with *MSH6* and *PMS2* families, as well as the increasing number of variants of uncertain significance (VUS) found in sequencing, complicate LS diagnosis and highlight the need for pathogenicity assessment.

The present work aimed to study the functional effect of lowered MMR gene expression as indication of Lynch syndrome and to assess the pathogenicity of MMR gene variants of uncertain clinical significance. Here, we studied how decreased *MLH1*, *MSH2*, *MSH6* and *PMS2* gene expression levels affect MMR efficiency. Using a stable shRNA knockdown approach we generated and studied altogether eleven cell lines retaining 23%, 50% and 74% of *MLH1*, 26%, 47% and 68% of *MSH2*, 50% and 79% of *MSH6*, and 19%, 33% and 53% of *PMS2* mRNA expression. The results of an *in vitro* MMR assay showed that the repair efficiency was not only associated with gene expression level but the expression decrease affected different genes differently. For *MSH2* and *MSH6* genes, an expression decrease to 75% already caused a significant decrease in MMR efficiency, while for *PMS2* the repair capability

decreased significantly near the mutation carrier level (~50%), and for *MLH1* gene only at the lowest expression level (23%). Unexpectedly, 19% and 33% of *PMS2* expression resulted in higher MMR efficiency than the carrier-like level, 53%, suggesting some kind of functional compensation for *PMS2* repair activity in the cell.

The functional significance of five *MLH1* and four *MSH2* VUSs found in suspected LS families was determined by the *in vitro* MMR assay. For that, nine recombinant MLH1 and MSH2 protein variants were used to complement MMR-deficient cancer cell lines lacking the normal respective protein. A MMR gene variant whose protein retained repair efficiency in the assay was determined MMR proficient, while variants that resulted in the lack of repair were determined MMR deficient. Here, the MLH1 variants p.Leu348Ser, p.Arg474Pro and p.Glu605Ala were shown to be MMR proficient and p.Gly101Ser and p.Leu260Arg deficient, while only one MSH2 variant p.Lys82Glu was MMR proficient and the variants p.Gly669Val, p.Phe694Ser and p.Pro696Leu deficient. Our results, together with the clinical and tumour data collected from the families, allowed the pathogenicity assessment of the MMR gene variants and Lynch syndrome diagnosis in the families.

Findings from these studies provide new insights into the severity of the malfunction that decreased levels of different MMR genes expression may cause. Furthermore, the results show that the functional assessment of variants of uncertain significance considerably helps their pathogenicity assessment. Both of these findings may have an important impact on Lynch syndrome diagnosis in future.

INTRODUCTION

Cancer is a major burden to public health worldwide, with colorectal cancer (CRC) being the third most commonly diagnosed cancer (International Agency for Research on Cancer (IARC), https://gco.iarc.fr/today/home). Cancers mostly evolve sporadically, yet in ~5-15% of cancers the cause is inherited [1,2]. Individuals with hereditary cancer syndromes often inherit one mutated and one healthy gene allele, and the carcinogenesis is a result of an acquired mutation in the remaining wild type (WT) allele [3]. The inheritance of gastrointestinal tumours was first recognized in the beginning of the 20th century, and the following research of Professor Henry Lynch and colleagues has since helped to understand the mechanisms of the "Cancer Family Syndrome" [4,5]. Lynch syndrome (LS, MIM#120435; previously known as hereditary non-polyposis colorectal cancer, HNPCC) is the most common cancer syndrome in the world, accounting for up to 3% of CRCs [6]. Mismatch repair (MMR) mechanism defects caused by mutations affecting the MMR genes *MLH1*, *MSH2*, *MSH6* or *PMS2* [7] lead to the development of a variety of cancers affecting the colon, endometrium, ovaries, gastrointestinal and urinary tracts [8,9]. In approximately 90% of diagnosed LS families, mutations affect *MLH1*, *MSH2*, and *MSH6* genes, whereas a relatively small number of LS families segregate a *PMS2* mutation [10], implying that *PMS2* is a moderate or low-risk LS susceptibility gene.

Early colorectal cancer diagnosis is important for providing cancer patients with accurate and preventative cancer surveillance [11]. LS diagnosis is based on international clinical criteria that account for the family history of colorectal and extra-colonic cancers, and the mismatch repair deficiency detected in the patient's tumour tissue [12]. However, a diagnosis is final only when a pathogenic MMR gene mutation is identified, usually guided by tumour immunohistochemistry results showing the expression loss of one or more MMR proteins [10]. As many as one third of the MMR gene mutations found have an unclear effect on the protein function, and their clinical significance has to be further determined with various *in silico* and biochemical assays [13,14]. Small insertions and deletions and splice site mutations constitute to these variants of uncertain significance (VUSs) [15], whose classification is facilitated by assessment models that combine data from multiple analyses [16]. The International Society for Gastrointestinal Hereditary Tumours (InSiGHT) facilitates gastrointestinal (GI) cancer diagnosis by incorporating MMR gene variants and their classification into a database [17].

This thesis aimed to study MMR deficiency, the characteristic of Lynch syndrome, by determining the functional effect of lowered *MLH1*, *MSH2*, *MSH6* and *PMS2* expression on the repair capability, and by assessing the pathogenicity of VUSs in *MLH1* and *MSH2* genes.

REVIEW OF THE LITERATURE

I LYNCH SYNDROME

Inherited colon cancer

Cancer is a heterogeneous disease characterised by aberrant cell growth. The third most common cancer is colorectal cancer (CRC), which is also the second most common cause of cancer-related deaths globally (International Agency for Research on Cancer (IARC), https://gco.iarc.fr/today/home). Over 1.8 million new cancer patients were diagnosed worldwide in 2018, and the CRC-related mortality was as high as 48% (https://gco.iarc.fr/today/home). In Finland, 42% of the 2142 CRCs diagnosed in 2017 caused death (https://cancerregistry.fi/statistics). These devastating statistics clearly demonstrate the need for better prevention strategies and improved understanding of the molecular mechanism of CRCs in order to reduce the cancer burden and lethality worldwide.

Carcinogenesis develops as a result of dysregulated balance between oncogenes (OCGs) and tumour suppressor genes (TSGs) [18]. OCGs promote tumour formation by inducing cell growth, while TSGs suppress tumour formation by regulating the cell proliferation via cell cycle, DNA repair mechanism and apoptosis. Thus, transformed cancer cells have acquired the ability to proliferate uncontrollably through the activation of OCGs and inactivation of TSGs [18]. According to Knudson's two hit hypothesis, both copies of a gene have to be mutated for carcinogenesis to occur [19]. Specifically in hereditary cancer predisposition syndromes, one mutated allele is inherited from either parent ("first hit"), while an acquired mutation in the remaining wild type allele ("second hit") during the carrier's lifetime induces the development of cancer [3]. More specifically, CRC progresses in a stepwise manner, also known as adenoma-to-carcinoma sequence [20]. Primarily, the changes in the normal colon mucosa lead to increased localized growth of colon epithelial cells (adenoma), and the subsequent malignant transformation results in invasive cancer.

Early diagnosis is essential for improving patient survival and quality of life. Recently, the CRC incidence and cancer-related death-rate have stabilized or slightly decreased in high-income countries, which is fostered by early detection and removal of premalignant lesions in the colon and thus preventing cancer progression [21,22]. Nevertheless, these rates continue to rise in developing countries. The identification of high risk individuals who have inherited a CRC predisposition is especially important, because early diagnosis helps to provide cancer surveillance and prevent cancer progression [1]. Moreover,

regular colonoscopy surveillance has been shown to reduce CRC incidence in patients with inherited predisposition [23] and increase the survival rate of individuals who have already been diagnosed with cancer [21].

The majority of CRCs are sporadic, while in 5%-15% the predisposition is inherited [1,2]. Individuals with a hereditary cancer predisposition have a much higher risk of developing cancer compared to the general population [2]. The main hereditary colon cancer syndromes are listed in **Table 1**. Briefly, they are classified as polyposis (up to thousands of polyps) or nonpolyposis (few or no polyps) syndromes [1]. The most common colon cancer syndrome belonging to nonpolyposis syndromes is Lynch syndrome (LS), previously known as hereditary nonpolyposis colorectal cancer (HNPCC), accounting for 1-3% of all colorectal cancers (**Table 1**) [6]. In LS, a dominant predisposition is inherited, and a germline mutation in a DNA mismatch repair (MMR) gene may increase the risk of CRC by up to 80%, and up to 50% for endometrial cancer (EC) and many other extra-colonic cancers, such as cancers of the stomach, ovary, small bowel, urinary and hepatobiliary tracts [10,24,25]. Recently, the frequency of pathogenic MMR gene mutations in the general population was estimated to be as high as 1 in 279, suggesting that Lynch syndrome is much more common than previously thought [26]. Mismatch repair gene mutations may also be biallelic (inherited from both parents), causing another predisposition syndrome called Constitutional Mismatch Repair Deficiency (CMMRD) (**Table 1**), which is much rarer than LS [27,28]. CMMRD is an early childhood or adolescence cancer syndrome associated with increased risk of brain, haematological and gastrointestinal tumours, as well as colorectal cancer with multiple adenomatous polyps.

Approximately 1% of CRCs are due to familial adenomatous polyposis (FAP), which is the most common polyposis syndrome – its incidence is 1 in 10 000 (**Table 1**) [29-31]. In FAP, the predisposition is dominantly inherited, and a mutation in the susceptibility gene (adenomatous polyposis coli, *APC*) causes over 80% chance of developing CRC over the person's lifetime [1]. A less severe disease phenotype (referred to as the attenuated FAP) is associated with mutations at the ends of the *APC* gene or with mutations specifically in exon 9. In addition to CRC, the risk of extra-colonic cancers affecting connective tissue (desmoid tumours), brain, thyroid and liver are common to patients suffering from this syndrome [1]. *MUTYH*-associated polyposis (MAP) is an example of a recessively inherited polyposis syndrome characterised by the development of hundreds of polyps in the colon, and up to 60-70% CRC risk [1]. Mutations in DNA polymerases ε, *POLE* (polymerase epsilon catalytic subunit), and δ, *POLD1* (polymerase delta 1), are rare causes of early-onset CRCs presenting multiple polyps [1]. It is advised that

these are searched for only when no mutations are found in other CRC predisposing genes, MMR, *APC* and *MUTYH* **(Table 1)** [32,33].

A distinct class of dominantly inherited rare polyposis syndromes is the Hamartomatous Polyposis Syndromes, including Peutz-Jeghers syndrome (PJS), Juvenile polyposis (JP) syndrome, *PTEN* hamartoma tumour syndrome, and hereditary mixed polyposis syndromes (**Table 1**) [34]. As the name implies, they are characterised by the development of benign (hamartomatous) polyps inside and outside of the gastrointestinal tract. Individuals suffering from PJS have an increased risk of gastrointestinal and extra-gastrointestinal cancers, including cumulative CRC risk by the age of 70 [35]. One unique feature of this syndrome is the pigmentation of oral mucosa [36]. Early-onset colorectal and extra-colorectal malignancies and the 40% lifetime risk of CRC is associated with JP [37]. *PTEN* hamartoma tumour syndrome is a collective name for several syndromes, including Cowden syndrome, Bannayan-Riley-Ruvalcaba syndrome, Proteus syndrome and Proteus-like syndrome [38]. This disease is characterised by the development of multiple hamartomas in different areas of the body and neurodevelopmental delays; the lifetime risk of CRC is 9% [34]. Hereditary mixed polyposis (*GREM1* mixed polyposis syndrome) is a rare syndrome, characterised by the development of various polyps such as juvenile, hyperplastic or adenomatous polyps, as well as an increased CRC risk [34]. The genetic cause is still unclear, however *GREM1* overexpression has been suggested to contribute to this syndrome. *BMPR1A* mutations have also been reported as a likely cause of a similar clinical phenotype as Juvenile polyposis.

Table 1. Main hereditary colon cancer syndromes

Syndrome	Inheritance	Mutated gene(s)	Incidence	MIM
Lynch syndrome	Dominant	*MLH1*, *MSH2*, *MSH6*, *PMS2*	1/279[26]	#120435
Constitutional Mismatch Repair Deficiency	Recessive	Biallelic mutations in *MLH1*, *MSH2*, *MSH6*, *PMS2*	1/1 000 000[28]	#276300
Familial Adenomatous Polyposis	Dominant	*APC*	1/10 000[31]	#175100
MUTYH-associated polyposis	Recessive	*MUTYH*	not available	#608456
POLE polyposis *POLD1* polyposis	Dominant	*POLE* *POLD1*	not available	#615083 #612591
Peutz-Jeghers syndrome	Dominant	*SKT11*	1/8300-29 000*	#175200
Juvenile polyposis syndrome	Dominant	*SMAD4/BMPR1A*	1/100 000-160 000*	#174900
PTEN hamartoma syndrome	Dominant	*PTEN*	1/200 000[34]	#158350
Hereditary mixed polyposis (*GREM1* mixed polyposis)	Dominant	*GREM1(BMPR1A)*	not available	#601228

*https://www.insight-group.org/syndromes

Lynch syndrome (LS)

Clinical and tumour pathological features of LS

The familiality of gastrointestinal and endometrial cancers was initially recognized in the beginning of the 20th century by the pathologist Aldred Scott Warthin. The subsequent research by Professor Henry T. Lynch in the 1960's determined that the "Cancer Family Syndrome" is inherited in an autosomal dominant manner [4,5]. Prior to the discovery of the molecular aetiology of this disease, Lynch named it "Hereditary Non-Polyposis Colorectal Cancer" (HNPCC) to differentiate it from the other familial colorectal cancer syndrome, "Familial Adenomatous Polyposis" [7,39] After the realization that the tumour spectrum is not limited to colorectal cancer, HNPCC was renamed as Lynch syndrome (LS), indicating the disease where the underlying MMR defect has been identified [7].

Lynch syndrome is the most common colon cancer syndrome in the world. Individuals suffering from LS usually have a familial history of colon and extra-colonic cancers [40]. Thus, the international clinical guidelines for diagnosing LS are based on strong family history of early-onset LS-associated tumours (**Table 3,** *chapter Lynch syndrome diagnosis*) [10]. The most common LS cancer is CRC, and the most common extra-colonic cancer is EC. The risk of developing CRC or EC for MMR gene mutation carriers is elevated (over 80% and up to 50%, respectively) [11,41] compared to the general population (2.2% and 1.3%, https://gco.iarc.fr/today/home). However, the LS cancer spectrum also includes ovarian, gastric, prostate, small bowel and urothelial tract cancers [8,9,42]. The association of breast cancer in LS has remained controversial [8,43-45].

LS is associated with a malfunction of DNA mismatch repair due to an inherited germline mutation in one of the MMR genes [46,47]. Recently, LS prevalence in the general population was estimated to be 1 in 279; a population-specific study from Iceland estimated the prevalence to be slightly higher – 1 in 226 [26,48]. The risk of cancer in mutation carriers has been shown to vary according to gender, as well as the mutated MMR gene. Thus far, no straightforward associations between the type or location of MMR gene mutation and the disease phenotype have been identified [6,9,49], yet the risk of developing LS-associated cancers has been determined to be considerably lower in *MSH6* (MIM# 600678) and *PMS2* (MIM# 600259) mutation carriers than in *MLH1* (MIM# 120436) and *MSH2* (MIM# 609309) mutation carriers [8]. Epimutations in *MSH2* and *MLH1* genes (changes to gene expression without affecting the DNA nucleotide sequence), which arise either *de novo* or as a consequence of genetic alteration, also cause LS [6].

Individuals carrying mutations in *MLH1* or *MSH2* are subject to the so-called classical LS phenotype. The overall lifetime risk of developing CRC is approximately 50%, and the mean age of onset is 45 years [8,50]. The CRC risk does not differ between *MLH1* and *MSH2* carriers [50]. Recently, the CRC risk in male *MLH1* carriers was shown to be higher compared to females, while there were no differences between genders for *MSH2* carriers [8]. Both *MLH1* and *MSH2* female carriers show an increased risk for endometrial cancer starting from middle age. Specifically, *MLH1* pathogenic mutation carriers have a higher risk of stomach and urinary tract cancers [8,51], while *MSH2* mutation carriers have a higher tendency of developing urinary and gastrointestinal tract cancers, as well as prostate and brain tumours [8,52].

Mutations in *MSH6* and *PMS2* genes are often associated with an atypical LS phenotype, which is characterised by lower penetrance of colorectal and endometrial cancers and a later age of cancer onset [53,54]. One explanation for the lower cancer penetrance in *MSH6* and *PMS2* families might be the protein homologues, MSH3 and MLH3, which are able to bind and partially function with their counterparts MSH2 and MLH1, respectively [55-57]. Although the CRC risk in *MSH6* mutation carriers is lower compared to *MLH1* and *MSH2* mutation carriers [8], endometrial cancer risk in *MSH6* mutation carriers might be the highest among all MMR gene mutation carriers [53]. Indeed, while the overall incidence of extra-colonic cancers in *MSH6* mutation carriers is lower compared to *MLH1* and *MSH2* [58], a recent large prospective study confirmed a high risk of gynaecological cancers in *MSH6* mutation carriers, with endometrial cancer being the most prevalent [8]. Compared to *MSH6* families, the risk has been suggested to be the same or lower for CRC and EC in *PMS2* carriers [53,54,59,60]. Currently, there is an ongoing debate about *PMS2* linked cancer risks. On one hand, it has been suggested that the increased risk is limited to CRC and EC [61]. However, others have suggested that there is no increased LS-associated cancer risk for heterozygous *PMS2* mutation carriers younger than 50 years old [8]. Overall, inconsistent segregation of cancers and unusually high prevalence of gastric, breast and prostate cancers are frequently observed in *PMS2* families [59,60,62].

The identification of LS families and individuals is complicated due to the highly heterogeneous phenotypes associated with LS [63]. The typical characteristics are early age of onset (45 years vs 65 years in sporadic CRC and 50 years vs 60 sporadic in EC), the predominant proximal (right-sided) location of tumours, and the tendency of multiple cancers, either synchronous or metachronous (https://www.cancer.net) [2,10,53,64]. Although LS cancers often have aggressive progression, survival seems to be better than in sporadic CRC [11,41].

MMR deficient tumours display features, which, although not necessarily unique to Lynch syndrome, help in diagnosis. Defective mismatch repair in a tumour leads to a microsatellite instability (MSI) phenotype, one of the main hallmarks of LS tumours [65], together with the loss of expression in one or more MMR proteins (*reviewed in Lynch syndrome diagnosis chapter*) [10].

Mismatch repair (MMR) malfunction as a hallmark of LS

Human cells are vulnerable to metabolic reactions, like hydrolysis or oxidation (internal lesions), and environmental, physical and chemical alterations (external lesions) [66,67]. Unrepaired DNA damage results in genomic instability, which is frequently seen in cancer cells [68]. Several major DNA repair mechanisms, including MMR, base-excision repair (BER), nucleotide-excision repair (NER), homologous recombination (HR) and non-homologous end joining (NHEJ) pathways, signal the presence of injuries to DNA strands, which initiates the repair process (**Table 2**) [66]. Single-stranded DNA damages are removed by either MMR, BER or NER. Non-helix-distorting lesions, such as alkylations, oxidations, deaminations or depurinations, are repaired via BER. Damages caused by ultraviolet radiation (UV) and/or chemicals are repaired by NER [69,70]. Double-stranded breaks (DSBs) are the result of various DNA damaging agents like ionizing radiation (IR), UV or chemicals, including those used in chemotherapy. DSBs are repaired through HR and NHEJ [71]. Mutations in genes encoding components of these repair pathways are associated with severe syndromes that are listed in **Table 2** [70,72,73].

Table 2. Major DNA repair pathways

Repair mechanism	Feature of lesion	Source of the lesion	Association with disease
Base excision repair	Single-stranded non-helix distorting lesions	Metabolic reactions	*MUTYH*-associated polyposis[70]
Nucleotide excision repair	Single-stranded helix-distorting lesions	UV light, chemicals	Cockayne syndrome, Trichothiodystrophy[72]
Mismatch repair	Single-stranded base-base mismatches and insertion/deletion loops	Replication	Lynch syndrome[10]
Homologous recombination	Double-strand breaks	Irradiation, UV light, chemicals	Breast cancer type 1 and 2 gene mutation-linked susceptibility in breast cancer[72]
Non-homologous end joining			Severe combined immune deficiency, ligase IV syndrome[73]

Modified from [72]

Every time a cell divides, its DNA is replicated and passed on to the descendant cells. The proofreading by DNA polymerases allows an average of only one mistake per 100 000 nucleotides. However, loss of proofreading may give rise to errors like single nucleotide mismatches or insertions/deletions in multiple nucleotides [74]. Fortunately, the majority of these mistakes are repaired by a post-replicative mismatch repair mechanism, thereby increasing the fidelity of replication [75].

The MMR malfunction is the cause of Lynch syndrome. Inactivating mutations that lead to mismatch repair deficiency were first found in bacterial genes known as mutL and mutS. In 1993, genetic defects in the human homologue of the mutS gene were linked to hereditary colon cancer syndrome previously known as HNPCC (presently LS) [10]. The first susceptibility gene identified in LS families was *MSH2* (mutS homologue 2) [76-78], and the second most commonly mutated gene, *MLH1* (mutL homologue 1), was identified shortly thereafter [79-81]. Human mutL homologues *PMS1* (post-meiotic segregation increased 1) and *PMS2* (post-meiotic segregation increased 2), as well as human mutS homologues *MSH6* (mutS homologue 6) and *MSH3* (mutS homologue 3) were subsequently discovered [82-84]. An additional human mutL homologue, *MLH3* (mutL homologue 3) was detected in 2000 [85]. Of these genes, *MSH6* was eventually determined to be the third most significant Lynch syndrome susceptibility gene [86]. Thus far, among mutL homologue genes, only *MLH1* and *PMS2* have been shown to be associated with LS [79,83], while the role of *MSH3* and *MLH3* mutations in LS predisposition is still somewhat unclear [87,88]. However, a biallelic *MLH3* gene variant was recently reported to be associated with predisposition to adenomatous polyposis syndrome [89]. In 2009, a novel cause of *MSH2* inactivation, the 3' deletion of a gene located upstream of *MSH2* (epithelial cell adhesion molecule; *EpCAM*), was detected in LS families with tumours exhibiting MSI and/or loss of MSH2 protein [90].

Lynch syndrome diagnosis

The ultimate aim of LS diagnosis is to reduce cancer morbidity and mortality. Time is the key to success – the earlier a MMR mutation carrier is identified and directed to surveillance, the better the chance of survival. Accurate diagnosis of LS is hampered by the variable clinical phenotypes associated with mutations MMR genes, especially the atypical characteristics and lower cancer penetrance associated with *MSH6* and *PMS2* [8]. Moreover, incomplete or lack of familial cancer history and/or tumour tissue of the proband contribute to under-recognition of LS. When a pathogenic LS mutation is detected, risk-reducing prophylactic surgeries such as the removal of polyps, the total removal of the colon, uterus or ovaries may be recommended to the patient [91]. Regular colonoscopies increase the

detection of precancerous polyps and thus reduce CRC incidence [23], yet there is no consensus on the appropriate surveillance intervals, not even in Europe. For example, screening intervals vary from 1 year in Germany to 1-2 years in the Netherlands, and 2-3 years in Finland [49,92,93]. Shorter intervals might be reasonable, as LS-associated CRCs progress through accelerated adenoma-to-carcinoma sequence [94], however shorter colonoscopy intervals have not been shown to decrease the number of detected CRCs [95,96].

Familial cancer history

A set of different criteria were developed to facilitate the identification of Lynch syndrome families, and these criteria are now being used worldwide [4]. The first clinical criteria for LS, Amsterdam I criteria (ACI), were established during the first International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer (ICG-HNPCC) meeting in Amsterdam in 1990 (**Table 3**) [2,97]. The criteria are based on early-onset colorectal cancers among successive generations in the family. The low sensitivity of mutation carrier detection demanded the need to update criteria to also include endometrial and other LS-associated extra-colonic cancers (Amsterdam II; (ACII)) (**Table 3**) [98]. Contrary to the Amsterdam criteria, the Bethesda Guidelines (BG), which were established in 1996 and revised in 2004, advised to test MMR-deficient tumours for microsatellite instability (**Table 3**) [99]. However, the traditional use of ACII and BG have been criticized for being inefficient and having less than optimal sensitivity [100].

Table 3. International criteria for Lynch syndrome diagnosis

Amsterdam I criteria[97]
at least three relatives with colorectal cancer
one should be the first-degree relative of the other two affected people
at least two successive generations affected
at least one of the cancers diagnosed before the age of 50
FAP should be excluded
Amsterdam II criteria[98]
at least three relatives with LS-associated cancer (CRC, EC, small bowel, ureter, renal pelvis)
one should be the first-degree relative of the other two affected people
at least two successive generations should be affected
at least one should be diagnosed before the age of 50
FAP should be excluded
tumour pathological verification
The revised Bethesda guidelines[99]
CRC diagnosed before the age of 50
presence of synchronous or metachronous CRCs or other LS-associated tumour, regardless of age
CRC with MSI-H histology diagnosed before the age of 60
CRC diagnosed in one or more first-degree relatives with LS-associated tumour, one of the cancers diagnosed before the age of 50
CRC diagnosed in two or more first- or second-degree relatives with LS-associated tumours, regardless of age

The identification of LS families may be improved by raising awareness in the general population and medical experts, as well as providing guidelines for genetic testing in cancer centres. In Western Europe, the identification of LS mutation carriers is primarily based on the assessment of familial cancer history. Bethesda guidelines are used to select high-risk individuals for further genetic testing [101]. The recommendation to test all new CRC cases by immunohistochemical staining and MSI analyses might be justified, as a small fraction of carriers have been missed using BG alone [12]. Once a carrier is subjected to genetic testing, the exact mutation is determined by sequencing approaches that are now more common in routine testing [102]. However, there is currently a lack of consensus on which screening approach is the most appropriate for LS diagnosis. The implementation of testing all newly diagnosed CRC and/or EC cases by immunohistochemistry and MSI analyses is challenged due to feasibility and cost effectiveness [100].

Tumour-based analyses

MMR malfunction leads to microsatellite instability (MSI) phenotype observed in LS tumours. MSI testing for characterising MMR deficient tumours involves the amplification of microsatellite markers using polymerase chain reactions (PCR) [99,103]. A panel consisting of two mononucleotide markers (BAT-25, BAT-26) and three dinucleotide markers (DS123, D5S346, D17S250) was proposed

by the National Cancer Institute (NCI). However, the Bethesda panel recommended the use of five quasi-monomorphic mononucleotide repeats (BAT-25, BAT-26, NR21, NR24, NR27), as the use of mononucleotides increases the efficiency of MSI testing [99]. The identical size between individuals makes mononucleotide repeats advantageous, since a control from a normal germline is unnecessary. Hence, it is advised to include at least three mononucleotide markers in the testing panel [104]. The MSI phenotypes are differentiated based on the number of unstable repetitive sequences found in the tumour tissue, and are classified as either high (MSI-H) (≥2/5 or 30% or more of unstable markers in a large panel), low (MSI-L) (1/5 or 10-30% in a large panel) or stable (MSS) (no unstable marker detected) [103].

The sensitivity and specificity of MSI testing are vulnerable to differences in marker panel, as well as testing protocols between laboratories [105]. In addition, a trained pathologist is required to differentiate between normal and tumour cells, which further impacts the sensitivity [103]. Moreover, MSI phenotype is not unique to LS, as sporadic cancers also show the presence of MSI. For example, in approximately 12% of sporadic CRCs MSI is found [106]. The presence of MSI among sporadic endometrial cancers is usually between 10-20% [107,108] but has been reported to be even 45% [109]. The MSI-H phenotype in sporadic gastric cancer is reported to be around 8.5% [110]. The cause for instability in sporadic cancers is gene silencing caused by *MLH1* promoter hypermethylation [111]. Nevertheless, majority of LS associated CRCs are microsatellite instable [103].

The loss of MMR protein expression due to genetic or epigenetic alterations is another characteristic of LS tumours [112]. Immunohistochemical staining detects the presence of MMR protein expression in a tumour tissue, and thus provides a supportive approach to MSI analysis. MMR proteins function as heterodimeric protein complexes. MLH1 and MSH2 are the major components of the MutLα and MutSα complexes, and are also more stable than their counterparts. Thus, the solitary loss of either PMS2 or MSH6 proteins indicates germline mutations in *PMS2* or *MSH6* (**Table 4**) [59,112], whereas the concomitant loss of either MLH1 and PMS2 or MSH2 and MSH6 indicates germline mutations in *MLH1* or *MSH2*, respectively (**Table 4**) [112]. However, the loss of MLH1 protein in a tumour tissue is not specific to LS, since many sporadic cancers also show inactivation of *MLH1* [111]. One of the most well-known epigenetic alterations in carcinogenesis is the methylation of CpG rich promoter areas, which leads to transcriptional silencing of gene expression [113]. This alternative mechanism for carcinogenesis is common to sporadic tumours, however *MLH1* epimutations have also been detected in Lynch syndrome-associated tumours – albeit rarely [114]. *MLH1* promoter hypermethylation and concomitant *BRAF* gene V600E mutation is associated with sporadic tumours [115]. As such, *BRAF* V600E mutation testing enables differentiation between sporadic and LS cases [116]. The immunohistochemical (IHC) analysis detects loss

of one or more MMR proteins, and gives indication which gene(s) should be preferentially tested [103]. Unfortunately, the intact protein expression determined with IHC does not necessarily mean that the expressed protein is functional. Consequently, a portion of cases could be missed [54]. In addition, the method is also influenced by tissue handling and staining procedure [117-119]. Nevertheless, compared to MSI, IHC has several advantages such as reduced cost and easier and quicker performance [103]. Although MSI and IHC testing provide the same results, using only one method often fails to identify LS mutation carriers [120].

Table 4. Interpretation of IHC and MSI testing results

MMR proteins				MSI	Next step	Interpretation
MLH1	**MSH2**	**MSH6**	**PMS2**			
+	+	+	+	MSS	No reason to proceed	LS excluded
-	+	+	-	MSI	*BRAF* mutation/methylation testing prior *MLH1* germline testing	LS suspicion
-	+	+	+	MSI		LS suspicion
+	-	-	+	MSI	*MSH2* germline testing	LS suspicion
+	+	-	+	MSI	*MSH6* germline testing	LS suspicion
+	+	+	-	MSI	*PMS2* germline testing	LS suspicion

Modified from [103]

Functional analyses to determine the pathogenicity of variants of uncertain significance

Once microsatellite instability and/or abnormal protein expression is detected, the suspicion of LS is confirmed and the high-risk individual can be guided to genetic testing. Currently, the next-generation sequencing (NGS) approach is becoming increasingly popular in clinical laboratories [10]. Major advantages of NGS include simultaneous testing of many genes, as well as cost effectiveness. However, NGS is not without limitations, which include the need for standardization for the universal testing in laboratories, and the abundance of uncertain findings. Also, epigenetic mutations and mutations outside coding regions cannot be found with NGS [102]. Various computational and laboratory analyses are available to assess the pathogenicity of variants that are found by sequencing, but whose clinical significance (i.e. effect on protein expression) remains undetermined (**Table 5**) [121]. For example, *in silico* methods are based on predicting the effect of variants of uncertain significance (VUSs) [122,123]. Biochemical and cell-based analyses enable the assessment of the main functions of mismatch repair in variant proteins, such as the binding between heterodimeric partners and binding of a variant protein onto mismatches or the MMR activity of a variant, thus providing definitive LS diagnosis for variants

that would be otherwise clinically uncertain [121]. A model has been proposed specifically for MMR gene variants, which facilitates the pathogenicity assessment of VUSs (*reviewed in chapter III*) [14].

Among MMR proteins, it is very common to find single amino acid substitutions that lead to missense mutations whose effect on the protein function is not obvious [16,123]. *In silico* methods can analyse the effect of missense variants in conserved sequences or algorithms by using sequence homology to predict the effect of amino acid substitution on protein functionality. Accordingly, such methods are widely used in MMR gene assessment (**Table 5**) [122,124-126]. Although computational methods are very efficient, potential discrepancies between different algorithms may require supporting methods like MMR activity assays to verify the significance of a certain variant [127].

Yeast- and cell-based functional analyses help to assess the MMR activity of a VUS (**Table 5**). The yeast-based methods are based on the evolutionary conservation between human and yeast mismatch repair. Thus, the human variants that are incapable of repair also in yeast MMR are interpreted as pathogenic [128,129]. Another approach introduces the identified MMR mutations to a yeast homologue, however since the homology between two hosts is restricted, the coverage is limited to conserved protein domains [130,131]. In cell-based MMR assays, the MMR deficient cell lines are complemented with the variant protein, and repair deficiency of a variant enables the pathogenicity to be determined [132,133]. This assay is based on using a substrate (plasmid containing a mismatch), which is then incubated with cell extracts including a MMR deficient cell line complemented with recombinant MMR protein. MMR proficiency of a variant is measured as restriction efficiency using gel electrophoresis [134]. Alternatively, fragment analysis may be used instead of gel electrophoresis [135].

Various biochemical analyses help to determine protein interactions as well as the stability of MMR proteins (**Table 5**) [121]. *MLH1* and *MSH2* mutations have been efficiently characterised by the yeast two-hybrid assay [136,137], which assesses the ability of MMR proteins to form protein complexes with their binding partners. This *in vivo* method is based on a "bait and catch" principal by fusing the protein of interest to a transcription factor's DNA binding domain, and its binding partner to a DNA-activation domain. Upon binding, the two domains will be in proximity to each other and hence capable of activating the transcription of a reporter gene [138,139]. The *in vitro* glutathione-s-transferase (GST) pull-down method uses a fusion-tagged protein as a "bait" to find its binding partner. Pathogenicity of a variant is determined by the inability to interact, and has been used to classify *MLH1* variants [140]. Finally, the co-immunoprecipitation method studies the binding of proteins by using an antibody that specifically binds to the protein of interest, and the interactions can be verified by Western blot (WB) analysis. Both mutL [88,134,141] and mutS [142,143] protein variants have been characterised with this analysis.

Insect and mammalian cell-based protein expression systems enable the study of variant protein stability (**Table 5**). In insect cell-based protein production, the recombinant proteins are introduced into plasmids, which are then amplified to collect baculoviruses expressing the variants. The viruses are then used to infect the *Spodoptera frugiperda* insect (*Sf9*) cells. Next, *Sf9* cells are grown and collected for variant protein extraction, from which protein expression can be measured and compared to the WT protein in WB analysis [88,136,141,143]. MMR deficient (MLH1-) human embryonic kidney 293 cells (HEK293T) are an attractive host for recombinant protein expression: their rapid division, high transfection efficiency and post-translational processing of proteins enables high protein yields [144]. Accordingly, they have been used to localize variant MLH1 proteins [145].

The nuclear localization of MMR proteins is crucial for their functional DNA repair, and can be analysed by expression of fluorescently labelled protein variants (**Table 5**). The variant MMR protein is first fused with the fluorescently labelled reporter protein, and then transiently expressed in the MMR deficient cell line to visualize their nuclear localization. MMR VUSs incapable of nuclear localization have been demonstrated [141,146,147]. However, MLH3 has shown conditional nuclear localization dependent on the cellular abundance of PMS2, thus the detection of MLH3 recombinant variants may be hampered [57].

The efficiency of mismatch binding by mutS homologues, which is important for the repair initiation process, can also be assessed with the electrophoretic mobility shift assay (EMSA) (**Table 5**). Insect cell [143] - or yeast strain-produced [129,130] recombinant proteins are incubated with mismatched DNA nucleotides to detect the binding efficiency of a protein variant. Protein variants that bind to DNA nucleotides have a slower migration rate through the polyacrylamide gel than the unbound nucleotides, which helps to determine the binding efficiency of a protein variant.

In addition to missense variants potentially altering protein function, VUSs in MMR genes may also affect splicing [148]. Splicing variants can be assessed with *in silico* programs that can predict the effect of a variant on donor and acceptor splice sites (**Table 5**) [123]. Pathogenicity of a variant determined via computational methods may be verified in a minigene assay [149,150]. This method is based on transient expression of a variant protein in a cell line, from which the RNA is isolated and splicing patterns of a variant can be compared to a control using RT-PCR.

Table 5. Overview of different analyses used in pathogenicity assessment of MMR gene variants

Assay	Purpose	Characteristics
in silico MAPP-MMR, PON-MMR, SIFT[122,124-126,151]	Determining the function of missense variants	Prediction of the effect of a gene variant using reference sequence
in vivo expression of mutant MMR genes in yeast strains[128,129]	Determining the MMR efficiency of a missense variant	Expression of the mutant human or corresponding yeast MMR in yeast strains
in vitro MMR assay[134]		Assessment of the MMR efficiency of an *in vitro* produced variant nuclear protein extracts by gel electrophoresis
in vitro cell-free MMR assay[135,152]		Assessing the MMR efficiency of *in vitro* produced variant extracts by fragment analysis
in vivo yeast two-hybrid assay[136,137]	Determining the interactions between MMR proteins	Physical interaction between proteins in yeast
in vitro GST pull-down assay[140]		Physical interaction between a GTS-tagged protein to a target protein
in vitro co-immunoprecipitation[88,143]		Physical interaction between proteins using antibodies
in vivo Sf9 baculovirus expression system[141]	Determining the MMR protein expression and stability	Recombinant proteins are amplified and *Sf9* cells are infected with variant producing baculoviruses
in vivo HEK293T expression system[145]		Transient expression of recombinant proteins into HEK293T cells
in vitro immunofluorescence[146,147]	Determining the nuclear localization of proteins	Expression of fluorescent MMR proteins in mammalian cells
in vitro electrophoretic mobility shift assay[153]	MMR protein and DNA interactions	Capacity of MMR heterodimers to recognize and bind mismatched DNA
in silico MaxEntScan, HumanSplicingFinder[123]	Detection of splicing aberrations	Predicting variant donor and acceptor sites using consensus sequences
in vivo minigene assay[150]		Assessing the effect of a variant on splicing using a minigene construct

II MISMATCH REPAIR MECHANISM AND GENES

MMR proteins and repair mechanism in human

Evolutionary conservation of the DNA repair pathway between prokaryotes and eukaryotes has facilitated human MMR research by *in vitro* reconstituted systems [75]. MMR mechanism functions as a post-replicative safeguard system by maintaining genomic integrity via correction of DNA errors [154]. High MMR activity is associated with replicative S-phase cells, but the errors initiated during repair or recombination are also eliminated by this mechanism [155]. Replication is a process controlled by DNA polymerases that incorporate new nucleotides in the daughter strand. The proofreading activity of these enzymes diminishes the number of errors made during replication, and additionally efficient MMR adds even more precision to the process [155].

Repair is initiated by mismatch recognition and binding by the MutS homologues (MSHs) (**Figure 1**). They function as dimers: MutSα (MSH2+MSH6) and MutSβ (MSH3+MSH6). They are functionally redundant, recognizing either base-base mismatches and small IDLs (≤3nt) or larger IDLs (≤13nt) (**Table 6**) [155]. Mismatch-bound MSHs undergo exchange from ADP to ATP, resulting in a conformational change that enables diffusion along the DNA strand and ultimately releases the mismatch [154]. This is necessary for the recruitment of additional repair factors like MutLα (MLH1+PMS2) – the most prevalent mutL heterodimer functioning in MMR (**Figure 1, Table 6**). Upon ATP binding, MutLα also undergoes conformational changes to encircle the DNA and form a ternary complex with DNA-bound MSHs [156,157]. The *in vitro* reconstitution of human MMR has suggested that the discrimination of the daughter strand is based on proliferating cell nuclear antigen (PCNA)-initiated activation of MutLα endonucleolytic activity, which is important for introducing nicks (**Table 6**) [154,158]. Replication factor C (RFC) catalyses the loading of PCNA to the DNA (**Table 6**) [154]. PCNA is a clamp protein that increases the ability of DNA polymerases to perform DNA synthesis without dissociation. PCNA is characterized by its rounded shape with a cavity in the middle, which enables it to surround the DNA (**Figure 1**). This physical interaction of PCNA with the DNA and other DNA-interacting proteins is crucial for accurate replication [159]. After efficient mismatch recognition and nicked DNA, the MSH- and MLH-activated single-strand 5' to 3' exonuclease EXO1 initiates the removal of the error-containing strand (**Figure 1, Table 6**) [160]. Replication protein A (RPA) is a single-stranded DNA-binding protein that binds to single-stranded (ss) DNA. This prevents the formation of a duplex and thereby facilitates the resynthesis [161]. Finally, DNA polymerase δ resynthesizes the missing section. After the gap is filled, the strand is ligated by a DNA ligase (**Figure 1**) [155].

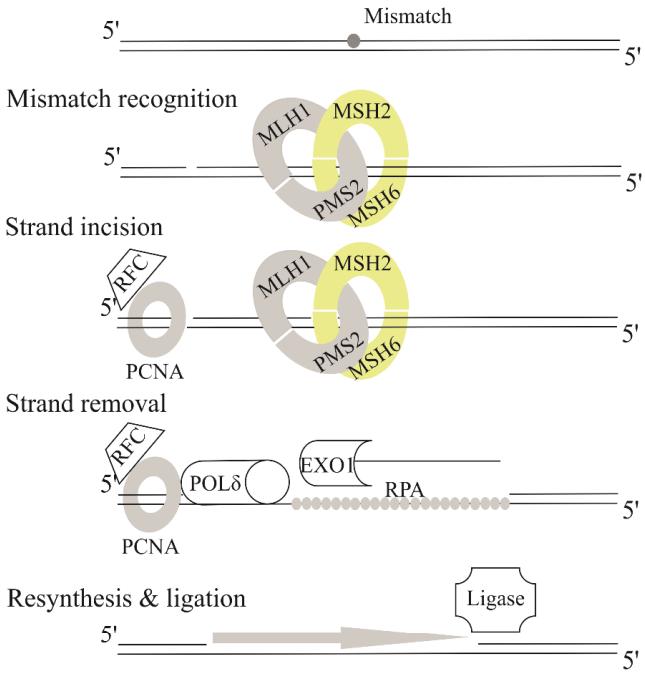

Figure 1. Schematic representation of MMR mechanism

RFC – replication factor C; PCNA – proliferating cell nuclear antigen; POL δ – polymerase delta; RPA – replication protein A; EXO1 – exonuclease 1; Ligase – DNA ligase

Table 6. The roles of proteins in MMR

Main protein complexes	**Functions in MMR**
MutSα (MSH2+MSH6)	Mismatch recognition
MutSβ (MSH3+MSH6)	Mismatch recognition
MutLα (MLH1+PMS2)	Complex formation with mutS homologues and DNA endonucleolytic activity
MutLβ (MLH1+PMS1)	No MMR activity
MutLγ (MLH1+MLH3)	Functioning as a back-up for MutLα
Other proteins	**Functions in MMR**
Replication factor C	Loading of PCNA and guiding excision polarity
Proliferating cell nuclear antigen	Recruitment of proteins to mismatch
Exonuclease 1	Excision of DNA strand
DNA polymerase delta δ	Resynthesis
Replication protein A	Binds to ssDNA during excision and DNA synthesis
DNA ligase	Ligation of the strand

Modified from [154]

Functions and characteristics of *MLH1*, *MSH2*, *MSH6* and *PMS2*

One of the main MMR genes, *MSH2*, is located on chromosome 2 short arm and consists of 22 exons (**Table 7**). Mutations in *MSH2* were first associated with Lynch syndrome susceptibility [76-78]. *MSH6* is another mutS homologue located on the same chromosome, and has 12 exons (**Table 7**) [82]. It is the third most significant Lynch syndrome susceptibility gene [86]. Proteins encoded by *MSH2* and *MSH6* participate in the DNA MMR mechanism as part of heterodimeric protein complex MutSα (MSH2+MSH6), recognizing base-base mismatches in the newly synthesized DNA strand [154].

MutS proteins belong to the ATP-binding cassette (ABC) protein family, and utilize ATP binding/hydrolysis energy for mismatch recognition and binding [162]. Specifically, MSH2 and MSH6 form an asymmetric protein complex, where the ATPase domains from both subunits are located opposite to each other. Mismatch recognition in MutSα is carried out specifically by MSH6 [163]. After mismatch recognition, the ADP exchange to ATP results in a conformational change in MutS proteins, allowing the mismatches to be liberated and the MutL proteins to be recruited [164,165].

MSH2 and MSH6 have a common protein structure, consisting of five different domains: N-terminal mismatch-binding domain, connector domain, lever domain, clamp domain and the C-terminal ATPase domain (**Figure 2**) [163]. The connector domain and ATPase domains facilitate the interactions with MutL proteins [166,167], while the DNA is bound by the lever and clamp domains. The most conserved protein domain – the ATPase domain – shares almost 50% identity with the bacterial counterpart and binds to nucleotides, ADP or ATP [163]. MSH2 interacts with MSH6 through two domains: the N-terminal part of MSH2 (aa 378 to 625) interacts with aa 326 to 575 of MSH6 and the C-terminal of MSH2 (aa 875 to 934) interacts to aa 953 to 1360 of MSH6 (**Figure 2**) [168].

Table 7. Four main mismatch repair genes[76-83,86]

Gene	Name	Chromosomal location	Exons	MIM
MSH2	mutS homologue 2	2p21-p16.3	22	#609309
MLH1	mutL homologue 1	3p22.2	21	#120436
PMS2	postmeiotic segregation increased 2	7p22.1	15	#600259
MSH6	mutS homologue 6	2p16.3	12	#600678

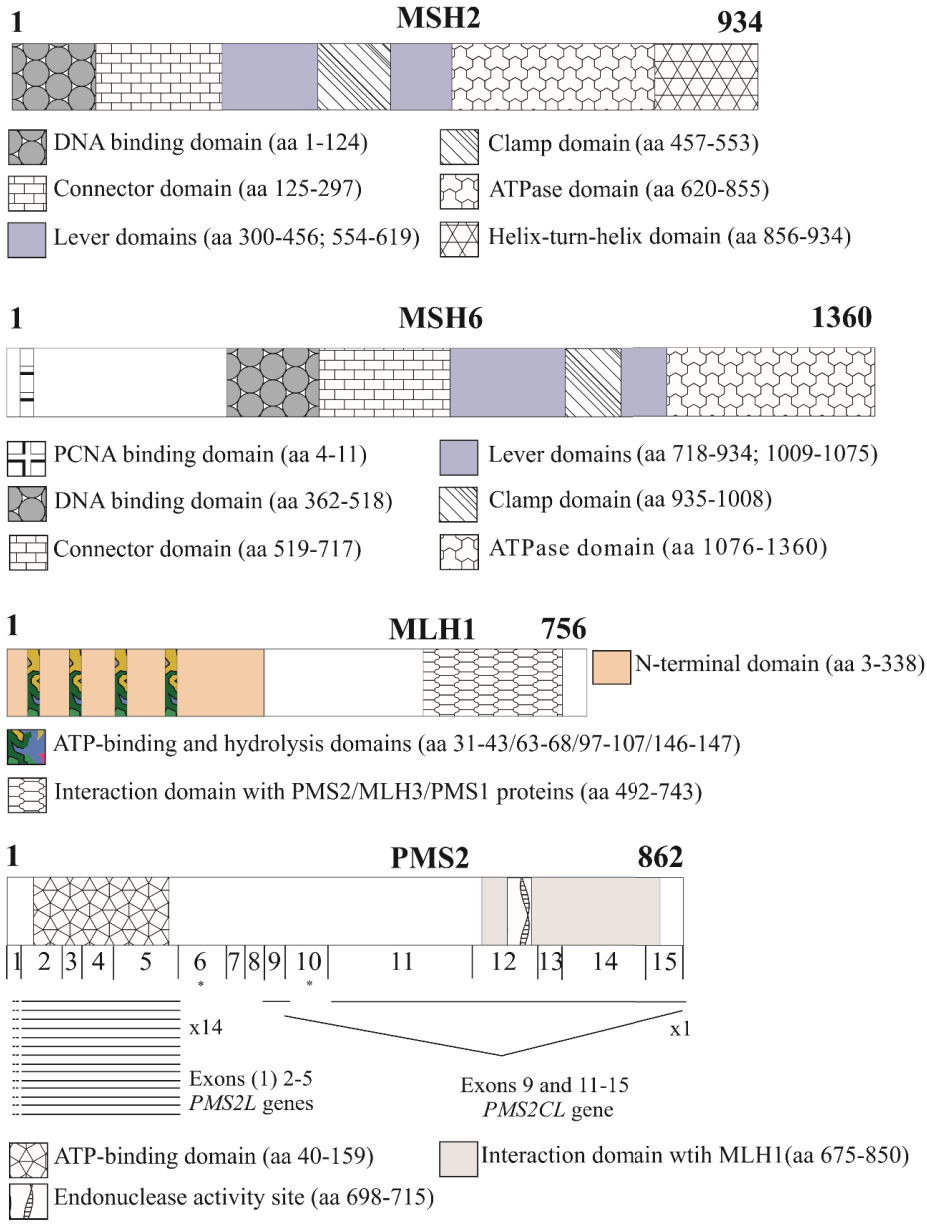

Figure 2. Protein domains of MSH2, MSH6, MLH1 and PMS2 and pseudogenes of *PMS2*

Mismatch repair protein domains and the corresponding amino acids shown for each protein. Fifteen *PMS2* pseudogenes are shown under PMS2 protein domains (modified from [15]).

Human *MLH1* gene was the second LS-associated locus and first human mutL homologue that was discovered and mapped to chromosome 3 short arm (3p21-23 region) (**Table 7**) [79-81]. The *MLH1* gene consists of 21 exons and encodes a protein product involved in MMR by coordinating protein-protein interactions to facilitate strand differentiation, as well as the process of removing the mismatch-containing strand. Another mutL homologue, *PMS2*, was discovered in 1994 (**Table 7**) [83]. MLH1 protein forms three different heterodimers with other mutL homologues through a conserved C-terminal dimerization domain [169]. Of these heterodimers, MutLα (MLH1+PMS2) contributes mainly to MMR activity. MutLβ (MLH1+PMS1) has no indication of MMR activity, and MutLγ (MLH1+MLH3) has been shown to partially compensate for the lack of MutLα [56,170]. Gyrase (type II DNA topoisomerase), chaperone Hsp90, mutL homologues and histidine kinases constitute an ATP-binding superfamily of proteins (GHKL family) [171]. Although each of these proteins regulate different processes in the cell, the ATP-binding domain (Bergerat fold) and the utilization of conformational energy through ATP hydrolysis or binding (or both) are common to all. With the exception of kinases, all of these proteins share a similar protein domain structure, where the N-terminal ATPase domain is connected to the C-terminal dimerization domain through a linker domain (**Figure 2**, only MLH1 and PMS2 proteins shown) [171].

The *PMS2* gene belongs to a gene family clustering on chromosome 7. Specifically, *PMS2* is located on chromosome 7p22.1 and consists of 15 exons [83]. Altogether, 15 highly homologous *PMS2* pseudogenes exist [172-174]. Fourteen of these *PMS2*-like (*PMS2L*) genes are homologous to the *PMS2* sequence at exons 2-5, whereas one (*PMS2CL*) is homologous to the C-terminal part (exons 9 and 11-15) (**Figure 2**) [172,174]. Although the *PMS2* gene was already identified in 1994, it took a decade before technology could differentiate between *PMS2* and its pseudogenes, and perform reliable sequencing and variant mutation detection [172,175,176].

MutLα (MLH1+PMS2) is central to MMR as a molecular matchmaker that links MMR error recognition to DNA repair [177]. MLH1 interacting domain with PMS2 is located in aa 492 to 742 and PMS2 interacting domain with MLH1 through amino acids 612 to 674 (**Figure 2**) [140,169]. The C-terminal activity sites of PMS2 contribute to the endonucleolytic activity of MutLα, which has been suggested to guide eukaryotic strand discrimination and downstream MMR events [178] while the ATP induced conformational changes at the N-terminal ATPase domain support protein interactions underlying the matchmaker role [157]. DQHA(X)2E(X)4E is the catalytic motif, along with three shorter motifs that constitute the active site responsible for endonuclease activity. This motif is found in both PMS2 and MLH3 proteins, suggesting functional redundancy [178]. Compared to MLH3 and PMS1, the abundance

of PMS2 is 60-fold and 10-fold higher, respectively. This strongly suggests that PMS2 is the main partner for MLH1 [56,170].

III MMR GENE MUTATIONS AND THEIR CLINICAL CLASSIFICATION

Mutation spectrum in MMR genes

Mutations in MMR genes are distributed throughout all domains, and no specific hotspots have been reported [6,179]. Furthermore, there is a wide range of mutation types that alter the exonic or intronic sequences of MMR genes (https://www.insight-database.org/). Nonsense and frameshift mutations leading to pathogenic alterations that disrupt normal gene function are common in MMR genes [6]. Single amino acid substitutions that lead to no or subtle changes at the protein level are also very abundant [6]. In addition, partial or whole gene deletion is a frequent second hit observed in Lynch syndrome [180]. *MLH1*, *MSH2*, *MSH6* and *PMS2* show a high proportion of *Alu* sequences prone to recombination events, which are detected as large genomic rearrangements. Predominantly, *MSH2* and *PMS2* are prone to these alterations, which is related to their higher *Alu* sequence density [181,182]. Rearrangements affecting *MLH1* and *MSH6* are detected less frequently [183,184].

Figure 3 illustrates the current distribution of unique variants in *MLH1* (1513 variants), *MSH2* (1377), *MSH6* (924) and *PMS2* (449) genes according to the InSiGHT database (https://databases.lovd.nl/shared/genes/; data accessed in August 2019). The most common type of mutations in *MLH1*, *MSH2* and *MSH6* genes are nonsense and/or frameshift mutations, followed by missense mutations, while silent mutations are frequently detected in *PMS2* (**Figure 3**). The abundance of variants with undetermined effect on the protein is quite high, ranging from 7% in *MSH6* to 19% in *MLH1* (**Figure 3**).

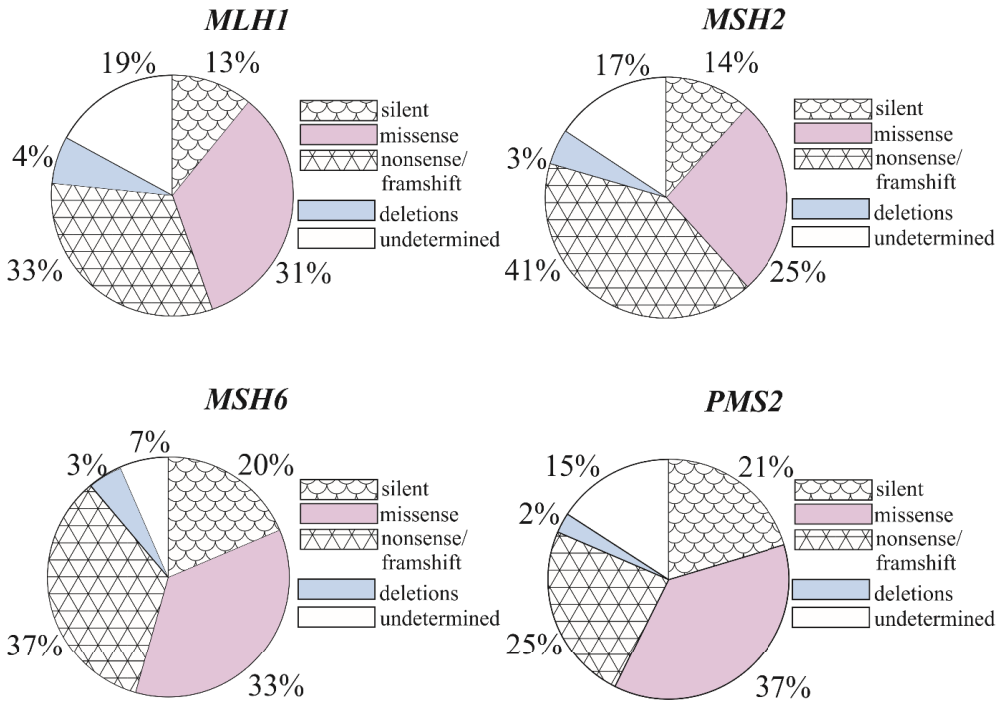

Figure 3. Mutation types in *MLH1*, *MSH2*, *MSH6* and *PMS2* genes

The distribution of different mutation types according to the effect at the protein level of unique variants listed for each gene in the LOVD database (https://databases.lovd.nl/shared/genes/; data accessed in August 2019).

The abundance of missense mutations among MMR genes represents a challenge, as a single amino acid substitution may not always result in a non-functional protein. Hence, the conclusive pathogenicity of a variant often needs to be assessed by several functional and biochemical assays [14,127]. A fraction of MMR gene variants remain unclear in relation to the disease phenotype, and are referred to as variants of uncertain significance (VUSs) [13]. These VUSs often complicate patient counselling and surveillance regimens.

Most of the abovementioned MMR gene sequence alterations are hereditary and unique to a given family [6]. However, some mutations arise recurrently in different families, populations and ethnic groups. They can be result of a *de novo* mutation or constitute founder mutations [6]. The latter have arisen in a common ancestor due to either natural selection or in very isolated populations [185]. Interestingly,

large deletions are also overrepresented in certain populations [186,187]. To date, over 50 founder mutations in LS have been described [185]. For example, the most prevalent *MLH1* founder mutations observed in more than 50% of Finnish LS individuals include c.320T>G, which leads to protein dysfunction [188]; c.454-1G>A, which causes aberrant splicing; and c.1732-2243_1896+404del, which is an exon 16 deletion [183,189]. Mutations in repetitive sequence motifs in *MSH2* have spontaneously arisen in Canadian (c.942+3A>T) [190,191] and Portuguese (c.388_389del) populations [192]. The existence of founder mutations in *MHS6* and *PMS2* could be more abundant than expected, due to their lower penetrance estimates and the technical difficulties in detecting *PMS2* mutations. In fact, a recent population-specific study demonstrated that *MSH6* and *PMS2* mutations prevail in Iceland, describing significantly increased cancer risks for one founder mutation in *MSH6* (p.Leu585Pro) and two in *PMS2* (p.Pro246Cysfs*3 and p.Met1?) [48]. Interestingly, the former shares an ancestral haplotype previously detected in Swedish, American and British study cohorts [193]. Furthermore, *MSH6* founder mutations are highly prevalent in Danish and Swedish populations [194,195].

Epigenetic dysregulation of either *MLH1* or *MSH2* genes can also result in LS-associated phenotypes [196,197]. Thus, heritable changes affecting gene expression without altering the DNA nucleotide sequence (epimutations) are an alternative means of hereditary cancer predisposition [198]. *MLH1* epimutations are referred to as primary because they arise mainly independently from genetic changes, whereas *MSH2* epimutations are secondary to genetic mutations. However, there are several rare cases that describe secondary *MLH1* epimutations that arise due to a genetic defect in close proximity to the *MLH1* gene [199,200]. In both *MLH1* epimutation carriers and conventional *MLH1* mutation carriers, it is common to find early-onset colorectal and endometrial cancers as well as the presence of MSI and/or loss of MLH1 protein expression. However, one major difference is the lack of family history in *MLH1* epimutation carriers [196,198]. Up to 10% of LS cases without identified MMR mutations have been attributed to *MLH1* epimutations that have arisen either *de novo* [114] or through non-Mendelian transmission between generations [201]. About 10% of LS-associated cancers result from the 3' deletion of *EPCAM*, a gene located upstream of *MSH2*, which causes *MSH2* promoter hypermethylation and gene inactivation [90,197,202]. The risk of colorectal cancer in *MSH2* epimutation carriers is similar to conventional *MSH2* families, but endometrial cancer risk is significantly lower. Interestingly, only larger *EPCAM* deletions that extend to the *MSH2* promoter have been associated with elevated EC risk [203,204].

Variants of uncertain significance (VUS)

Approximately one third of MMR gene variants have unidentified clinical significance. Most often, these VUSs are non-truncating small insertions or deletions, missense mutations or mutations in the proximity of the splice site. This is a major burden in LS diagnosis and a source of patient distress, because without knowledge of the effect on the disease development, no diagnosis can be made [15]. The primary clinical goal is to separate benign variants from harmful ones, however, the assessment of VUSs is labour intensive, as it needs multiple points of evidence prior to proving or disproving their possible pathogenicity [16]. The multigene panel testing options currently available have increased the number of incidental findings – including VUSs – further complicating the LS diagnosis [205].

Sharing variant information has been facilitated by compiling all inherited colon cancer gene variants and their pathogenicity interpretations into the Leiden Open Variation database (LOVD; https://databases.lovd.nl/) [17]. Several hundreds of VUSs have been identified in MMR genes - it represents approximately 19% of *MLH1*, 23% of *MSH2*, 39% of *MSH6* and 40% of *PMS2* variants (**Figure 4**). A workflow has been introduced to decrease the number of VUSs and facilitate their classification by exploiting tumour pathological and RNA and protein based data [14].

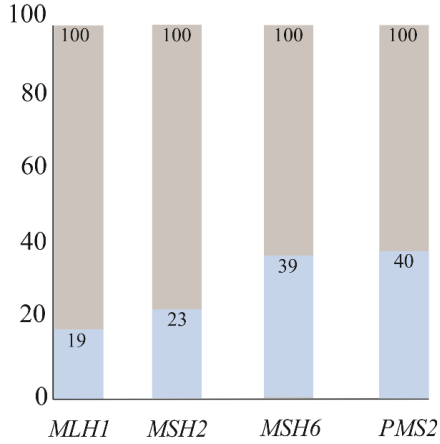

Figure 4. Proportion of variants of uncertain significance in four MMR genes

292/1513 of *MLH1*, 314/1377 of *MSH2*, 358/924 of *MSH6* and 178/449 *PMS2* unique variants listed in https://databases.lovd.nl/ are classified as VUS (data accessed in August 2019).

Combined with clinical data, results from *in vivo* yeast or *in vitro* mammalian assays (which assess the MMR efficiency of a variant) can provide valuable means for variant classification. A three step model combining data from multiple sources has been established to facilitate the classification of VUSs (**Figure 5**) [14]. Accordingly, variations are found via genetic testing preceded by results from MSI/IHC-testing supported by familial cancer history (known as STEP 1). If LS cannot be confirmed nor ruled out in the first step, the model continues with STEP 2 by combining data from functional (*in vitro* MMR assay) and computational assays (predictive *in silico* platforms). If MMR deficiency is confirmed by *in vitro* assay and *in silico* analysis in STEP 2, the diagnosis is final. However, the MMR proficient variant needs to be further characterised by additional biochemical assays in STEP 3 (**Figure 5**) [14]. The choice of assay and the interpretation of the result have to be considered critically, as a negative result is not necessarily an indication of non-pathogenicity when the variant is located in the domain that has no impact on the function studied with the analysis. In addition, the possible discrepancy between assays emphasizes the need to utilize various analyses [14].

This three step assessment model has been deemed sufficient for characterising the pathogenicity of MMR genes, as it was verified using 74 VUSs found in *MLH1*, *MSH2* and *MSH6* genes [127]. The first two steps in the workflow were sufficient for the majority of *MLH1* and *MSH2* variants. Obvious discrepancy between two *in silico* methods indicated the need to consider results from the functional assay. The model is an example of how to cope with unclassified variants in LS [127]. Nevertheless, it still presents several variables to be considered by clinicians, including the choice of patients who should undergo testing and which analyses to use [12,206,207].

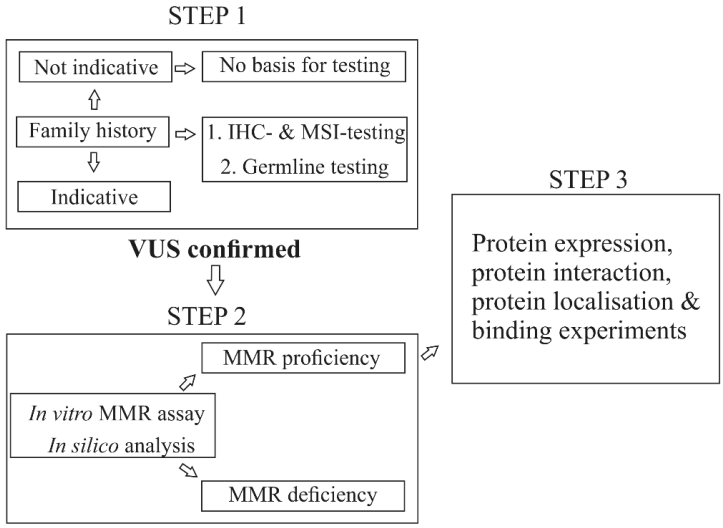

Figure 5. A proposed model to assess MMR gene variants [14]

Clinical classification of MMR gene variants

Scientists and healthcare professionals belonging to the International Society for Gastrointestinal Hereditary Tumours (InSiGHT) facilitate the screening of gastrointestinal (GI) tumours [17]. This society prioritizes global collaborations by collecting know-how from different clinics in order to integrate this knowledge into development of GI cancer diagnostics. The Leiden Open Variation Database (LOVD; https://databases.lovd.nl/), curated by the InSiGHT, is a repository of various mutations found in hereditary colorectal cancer genes, including *MLH1*, *MSH2*, *MSH6* and *PMS2*. Variant-associated molecular and clinical data contributes to variant interpretation, and thus the identification of LS families [17]. The InSiGHT data submission template, which lists all the data necessary for variant submission, illustrates the wide range of information that should be stored for a variant (https://databases.lovd.nl/). Accordingly, a variant could possibly be listed many times with evidence from multiple assays, thus complicating the interpretation process. In 2014, the InSiGHT Variant Interpretation Committee (VIC) undertook a comprehensive development, testing and standardization of variant classification in MMR genes using over two thousand variants listed in the LOVD database [16]. A five-step classification, based on the International Agency for Research on Cancer (IARC) recommendations, was used to classify the

MMR variants as pathogenic (Class 5), likely pathogenic (Class 4), VUS (Class 3), likely non-pathogenic (Class 2) or non-pathogenic (Class 1) [16,208]. Classes 2 and 4 aim to reduce the quantity of unclassified variants and to differentiate variants with sufficient evidence from variants of uncertain significance [208]. A definitive classification is based on patient-derived (age, gender, cancer history, segregation analysis), tumour-derived (MSI/IHC) and laboratory-derived (functional analyses) points of evidence. Solitary evidence from either the clinical or laboratory side is inefficient for VUS classification, as potential technical errors or misinterpreted results from one analysis can lead to misclassification [16]. Results from this large-scale study highlighted the importance of including unpublished data in pathogenicity assessment, as it altered the final classification of one third of the listed variants. Most importantly, the InSiGHT recommendations for variant classification are beneficial for at-risk families [16].

The current distribution of classified unique variants in MMR genes is shown in **Figure 6**. To date, the majority of the 4263 listed unique variants in the InSiGHT database are in *MLH1* (35.5%) and *MSH2* (32.3%) genes, while 21.7% are in *MSH6* and 10.5% in *PMS2* genes (data accessed in August 2019). The distribution of variant classes is variable among MMR genes, however, the prevalence of either VUSs or unclassified variants still remains high, emphasizing the continuous need for classification (**Figure 6**).

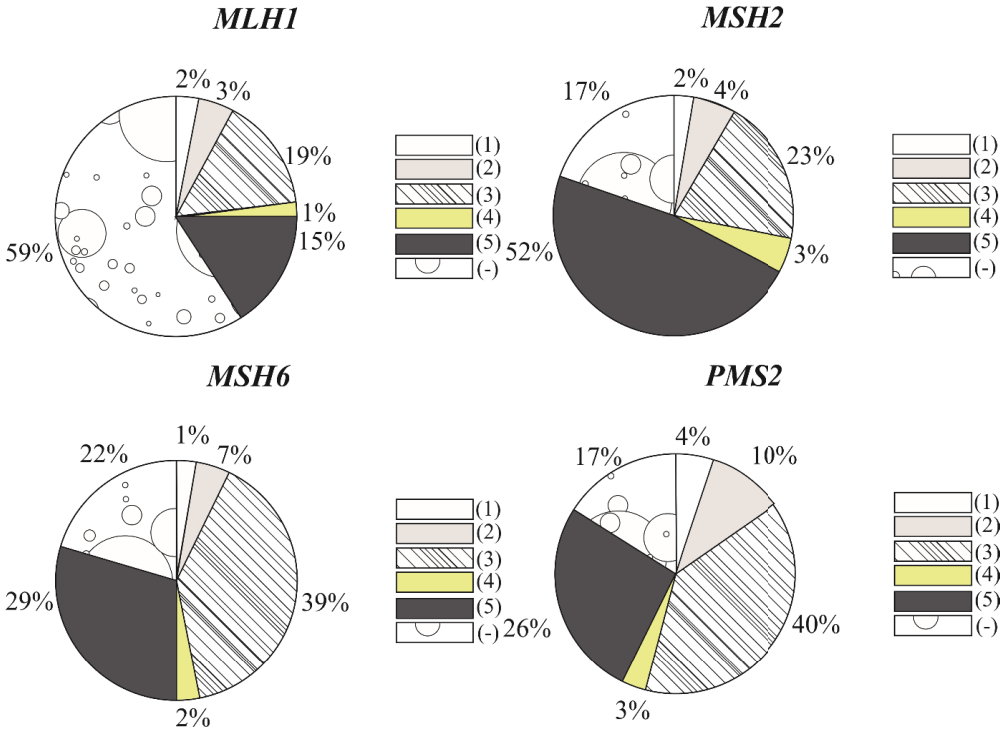

Figure 6. MMR gene variant classification according to InSiGHT

Current classification (%) of unique variants in MMR genes in LOVD database. Class 1 ((1), in the figure) – not pathogenic; class 2 (2) – likely not pathogenic; class 3 (3) – variants of uncertain significance; class 4 (4) – likely pathogenic; class 5 (5) – pathogenic; undetermined variants (-) (data accessed in August 2019).

AIMS OF THE STUDY

The present work aimed to study: 1) the functional effect of lowered MMR gene expression as an indication of Lynch syndrome, and 2) the pathogenicity of MMR gene variants of uncertain (clinical) significance.

The specific aims of the thesis were to assess the functional significance of:

- reduced *MLH1*, *MSH2*, *MSH6* and *PMS2* gene mRNA expression levels on MMR efficiency (**I, III**)
- nine VUSs detected in *MLH1* and *MSH2* genes (**II**)

MATERIALS AND METHODS

Functionally characterised variants in *MLH1* and *MSH2* (**II**)

Twenty four functionally assessed single-nucleotide substitutions in *MLH1* and *MSH2* genes were ascertained through cancer clinics from Italian families fulfilling the Bethesda guidelines. Clinical data and results from functional RNA- and protein-based analyses were used for final classification of variants. Nucleotide numbering of each variant reflects cDNA numbering with +1 corresponding to the A of the ATG translation initiation codon in the reference sequence, following the Human Genome Variation Society (HGVS) recommendations (https://varnomen.hgvs.org/) (**Table 8**). Clearly disease-causing variants, like splicing and truncating variants or large rearrangements, were excluded from the analysis. Six of the studied variants were novel and not yet reported in the InSiGHT database (https://www.insight-group.org/variants/databases). Based on the predicted effects on protein structure, alterations were divided into groups as amino acid substitutions, splice-site changes, synonymous or intronic variants (**Table 8**). The intronic and synonymous variants that cannot be assessed with the functional assay were excluded, and four missense changes were already verified previously. Finally, the stability and pathogenicity of five VUSs in *MLH1* (NM_000249.3) and four in *MSH2* (NM_000251.2) genes was determined by using Western blot (WB) analysis and a functional *in vitro* MMR assay (**Table 8**).

Table 8. *MLH1* and *MSH2* gene variants included in the functional assessment

Gene	Exonic location	DNA change	Predicted amino acid change	Mutation type	Protein stability and MMR activity
MLH1	3	c.301G>A	p.Gly101Ser	Amino acid substitution	**II**
	3i	c.307-19A>G		Intronic variant	na
	9	c.779T>G	p.Leu260Arg	Amino acid substitution	**II**
	11i	c.1039-8T>A		Intronic variant	na
	12	**c.1043T>C**	**p.Leu348Ser**	Amino acid substitution	**II**
	12	c.1217G>A	p.Ser406Asn	Amino acid substitution	[135,209]
	13	c.1421G>C	p.Arg474Pro	Amino acid substitution	**II**
	13i	c.1558+14G>A		Intronic variant	na
	15i	**c.1732-19T>C**		Intronic variant	na
	16	**c.1743G>A**	**p.Pro581=**	Synonymous variant	na
	16	**c.1814 A>C**	**p.Glu605Ala**	Amino acid substitution	**II**
	18	c.2041G>A	p.Ala681Thr	Amin acid substitution	[141,209-212]
	18	c.2059C>T	p.Arg687Trp	Amino acid substitution	[209,213]
Gene	**Exonic location**	**DNA change**	**Predicted amino acid change**	**Mutation type**	**Protein stability and MMR activity**
MSH2	2	**c.244A>G**	**p.Lys82Glu**	Amino acid substitution	**II**
	3	c.380A>G	p.Asn127Ser	Amino acid substitution	[153]
	8i	c.1387-8G>T		Intronic variant	na
	9i	c.1511-9AT		Intronic variant	na
	11	c.1666T>C	p.Pro556=	Synonymous variant	na
	11	c.1737A>G	p.Lys579=	Synonymous variant	na
	12i	c.2006-6T>C		Intronic variant	na
	13	c.2006G>T	p.Gly669Val	Potential splice site variant	**II**
	13	c.2081T>C	p.Phe694Ser	Amino acid substitution	**II**
	13	c.2087C>T	p.Pro696Leu	Amino acid substitution	**II**
	14	**c.2442T>G**	**p.Leu814=**	Synonymous variant	na

Unreported variants indicated in bold, variants assessed in this study are indicated with II, untested variants indicated with "na" (not available), and previously assessed variants indicated with references (as numbers).

Cell lines used for functional characterisation (**I-III**)

Spodoptera frugiperda (*Sf9*) insect cells were used to produce recombinant MLH1 and MSH2 proteins. MMR-deficient cell lines, HCT116 and LoVo, lacking MLH1 and MSH2 proteins respectively, were complemented with proteins to assess their ability to restore the MMR efficiency of MMR-deficient cell lines in the *in vitro* MMR assay (**II**). Additionally, cancer cell lines HCT116, LoVo and HEC-1-A were used as controls in the functional assay and WB analysis (**I-III**). Characteristic features of each cell line are summarised in **Table 9**.

Table 9. Cell lines

Cell line	Tissue of origin	Genetic phenotype	Origin	Publication
Sf 9	Insect cells	MMR proficient	ATCC	**II**
HCT116	Human colorectal adenocarcinoma cells	$MLH1^{-}/PMS2^{-}/MLH3^{-}$		**II, III**
LoVo		$MSH2^{-}/MSH3^{-}/MSH6^{-}$		**II**
HEC-1-A	Human endometrial adenocarcinoma cells	$PMS2^{-}/MLH3^{-}$		**III**
hTERT-1604	Human immortalized fibroblasts	MMR proficient	[214,215]	**I, III**

ATCC – American Type Culture Collection, Manassas, VA, USA
Roman numerals indicate in which publication the cell line was used

Stable shRNA knockdown of *MLH1*, *MSH2*, *MSH6* and *PMS2* (**I**, **III**)

Immortalized human fibroblasts (hTERT-1604) [214,215] were transfected with either *MLH1*, *MSH2*, *MSH6* or *PMS2* shRNA (SureSilencing™, QIAGEN) plasmids to generate KD cell lines with heterogeneous mRNA expression levels. Hygromycin selection at 150µg/µl (Invitrogen) and several rounds of monoclonalisation yielded cell lines specifically and stably expressing the shRNA plasmids of the gene of interest (GOI). Fibroblasts were also transfected with a non-targeting control vector (negative control, NC) throughout all experiments for each GOI.

After expanding and homogenizing about 3 x 10^6 cells (QIAshredder, Qiagen), the total RNA was extracted and reverse transcribed into cDNA (Superscript® VILO, Invitrogen) for quantitative PCR (Q-PCR) analysis. *MLH1*, *MSH2*, *MSH6* Taqman® assays (**Table 10**) using Taqman® Universal PCR Master Mix (Applied Biosystems) were run on a StepOnePlus™ device (Applied Biosystems), whereas *PMS2* and *MLH3* assays (**Table 10**) were run on a CFX96 Touch™ device (BioRad). Stable

housekeeping genes, *ACTB*, *GAPDH* and *HPRT1*, were chosen as reference genes after geNorm analysis. Expression levels of GOIs were quantified by normalization to the reference genes and calculated according to the $2^{-\Delta\Delta CT}$ method [216]. All Q-PCR assays for chosen cell lines were performed in at least two replicates with standard deviation (STD) of 0.2 or less. MMR gene-specific cell lines were chosen for functional analyses based on Q-PCR results and our inclusion criteria was that cell lines should retain at least 25% to 75% of mRNA expression for *MLH1*, *MSH2*, *MSH6* genes and at least ≤50% of mRNA expression for *PMS2.*

Table 10. TaqMan® assays used for quantitative analysis of MMR gene KD cell lines

Gene	RefSeq	TaqMan® assays
MLH1	NM_000249.3	Hs00179866_m1
MSH2	NM_000251.2	Hs00953523_m1
MSH6	NM_000179.2	Hs00264721_m1
PMS2	NM_000535.7	Hs00241053_m1
MLH3	NM_001040108.1	Hs00998139_m1
ACTB	NM_001101.5	Hs01060665_g1
GAPDH	NM_001256799.3	Hs02758991_g1
HPRT1	NM_000194.3	Hs02800695_m1

Production of wild type and recombinant MLH1 and MSH2 proteins (**II**)

The single-nucleotide changes in *MLH1* and *MSH2* genes were constructed into the wild type (WT) cDNA sequences via PCR-based site-directed mutagenesis (QuikChange Lightning®, Stratagene) and subsequently cloned into a pFastBac1 vector (Invitrogen) as previously described [134,142,217]. Prior to the production of recombinant proteins, the constructs were verified by sequencing (ABIPrism 3100 Genetic Analyzer; Applied Biosystems). Baculoviruses used to infect the *Sf9* insect cells were generated according to the manufacturer's instructions (Bac-to-Bac system; Invitrogen) and used for the production of recombinant proteins as previously described [134]. As MMR proteins function in complexes, the *Sf9* cells were co-infected with MLH1 and PMS2 viruses to yield MutLα, or with MSH2 and MSH6 viruses to yield MutSα heterodimers, respectively (**Table 11**). The total protein content of WT and variant proteins was extracted, and the protein expression levels were verified by WB analysis. Equal amounts of MutLα-WT and -variant or MutSα-WT and -variant total extracts (TEs) were loaded onto sodium dodecyl sulphate/polyacrylamide gel electrophoresis (SDS-PAGE) gels. After transferring

proteins to a nitrocellulose membrane (Hybond™ –P, PVDF, GE Healthcare), anti-MLH1, anti-PMS2, anti-MSH2 and anti-MSH6 antibodies were used for visualization. Equal loading of all proteins was verified with α-tubulin (details shown in **Table 11**).

Table 11. *Sf9* insect cell-based protein production of WT and recombinant proteins

	MLH1-WT	PMS2-WT	MLH1-variant		MSH2-WT	MSH6-WT	MSH2-variant
MutLα-WT	+	+		**MutSα-WT**	+	+	
MutLα^r		+	p.Gly101Ser	**MutSα^r**		+	p.Lys82Glu
MutLα^r		+	p.Leu260Arg	**MutSα^r**		+	p.Gly669Val
MutLα^r		+	p.Leu348Ser	**MutSα^r**		+	p.Phe694Ser
MutLα^r		+	p.Arg474Pro	**MutSα^r**		+	p.Pro696Leu
MutLα^r		+	p.Glu605Ala				

WT – wild type; MutLα^r – recombinant MutLα complex; MutSα^r – recombinant MutSα complex

An outline of the production of wild type and recombinant proteins by co-infection of *Sf9* cells, either with both MLH1-WT and PMS2-WT or MSH2-WT and MSH6-WT baculoviruses to yield a MutLα-WT or MutSα-WT protein complex, or MLH1 or MSH2 mutated baculovirus with PMS2-WT or MSH6-WT to yield recombinant protein complexes.

Nuclear protein extraction (**I-III**)

Recombinant protein total extracts were complemented with nuclear extracts (NEs) of MMR deficient cancer cell lines HCT116 (MLH1$^-$; used for *MLH1* variants), and LoVo (MSH2$^-$; used for *MSH2* variants), to assess their functional significance in the functional *in vitro* MMR assay (**II**). HEC-1-A NE was used as a positive control for MLH3 protein expression and a negative control for PMS2 protein expression in WB analysis (**III**). The *in vitro* MMR activity of *MLH1*, *MSH2*, *MSH6* and *PMS2* knockdown (KD) cell line NEs was verified by comparing the *in vitro* MMR activity of each GOI cell line to their respective control cell lines (**I, III**). Nuclear proteins of KD and NC cells, or HCT116, LoVo, HEC-1-A cell lines, were extracted as described previously, with minor modifications (**I-III**) [132,218]. KD and NC cell lines were extracted in parallel to allow comparison (**I, III**). Proteins were collected from approximately 1-2 x 10^8 cells, concentrated and cleared for use in the *in vitro* MMR assay. Nuclear extracts were quantified with a fluorimeter (Qubit 1.0 & 3.0; Thermo Fisher Scientific, Waltham, Massachusetts, USA) (**I-III**).

Western blot analysis (**I-III**)

The protein expression levels of MutLα and MutSα variants produced in *Sf9* cells, and the presence of MMR proteins and their heterodimeric partners in KD cell lines were assessed by sodium dodecyl sulfate/polyacrylamide gel electrophoresis (SDS-PAGE) and WB analysis (**I-III**). Equal amounts (adjusted according to loading control) of MutLα-WT and MutSα-WT and their variants (**II**) and gene-specific KD cell lines and their controls (**I, III**) were loaded on 7.5% SDS-PAGE gels. After electrophoresis, the proteins were transferred onto nitrocellulose membranes (Hybond, PVDF; Amersham Pharmacia Biotech) and visualized with anti-MLH1, anti-PMS2, anti-MSH2 and anti-MSH6 antibodies (listed in **Table 12**) (**I-III**). Commercial HCT116 cell line (ATCC® CCL-247™), lacking MLH1, PMS2 and MLH3 proteins, served as a negative control in the *PMS2* study (**III**). The presence of the mutL homologue MLH3 was specifically detected in *PMS2* KD cells to test the compensation hypothesis. Commercial HEC-1-A cell line (ATCC® HTB-112™) served as a positive control for MLH3 (lacking PMS2, but MLH1 is present) (**III**). α-tubulin was used as a loading control in all experiments (**I-III**).

Table 12. List of antibodies used in protein expression analyses

Antigen	Type of antibody	Manufacturer	Publication
MLH1	Clone G168-15	BD Biosciences	**I-III**
MSH2	NA-26	Calbiochem	**I, II**
MSH6	Clone 44	BD Biosciences	**I, II**
PMS2	NA-30 & clone A16-4	Calbiochem & BD Biosciences	**II & III**
MLH3	(H2): sc-25313	Santa Cruz Biotechnology	**III**
m-IgG HRP	NXA931V	GE Healthcare	**I-III**
m-IgGκ BP-HRP	sc-516102	Santa Cruz Biotechnology	**III**
α-tubulin	DM1A	Sigma Aldrich	**I-III**

In vitro MMR assay (**I-III**)

The functional *in vitro* MMR assay is used to study the DNA mismatch repair activity in cells. The assay requires substrate (**Figure 7**) and cell extract preparations for the repair analysis. Single-stranded DNA and linearized plasmid DNA are re-annealed to create a circular heteroduplex molecule used as a substrate in the functional assay [219]. The same 5' G•T heteroduplex (5' G•T) construct was used in all studies to initiate the repair process (**I-III**). 5' G•T is a 3193bp molecule, in which the adenine has been replaced with guanine to create a mismatch, while the thymine in the complementary strand is left unchanged. A 5' nick is created with *Ban*II restriction enzyme located 369bp from the G•T mismatch (**Figure 7**). Successful repair reaction converts the 5' G•T heteroduplex substrate to an A•T homoduplex and makes the molecule susceptible to cleavage with the *Bgl*II restriction enzyme [132,219]. MMR efficiency can be quantified by the restriction efficiency of *Bgl*II after linearizing the substrate with the *Eco*3I. Hence, the repair efficiency is measured as *Bgl*II cleavage efficiency by gel electrophoresis.

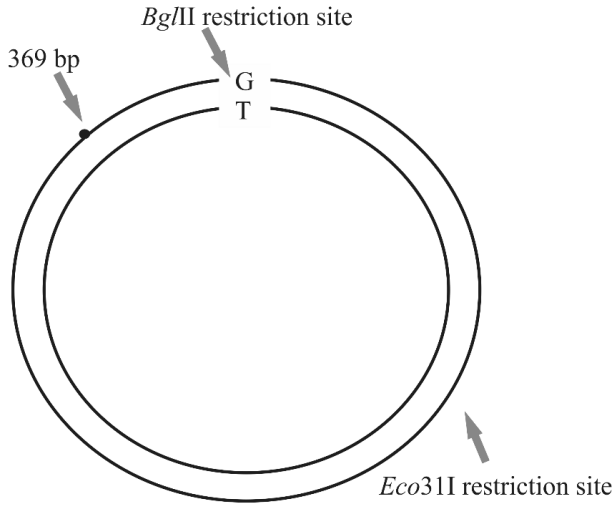

Figure 7. Heteroduplex substrate used in the functional MMR assay

The functional significance of MLH1 and MSH2 recombinant proteins was assessed as their ability to restore the repair efficiency of cell lines lacking either MLH1 (HCT116) or MSH2 (LoVo) proteins. For that, 100 ng of substrate was incubated with 75 µg of MMR deficient cell extract. This extract was complemented with *Sf9*-produced TE containing overexpressed MLH1 or MSH2 proteins (**II**). The MMR activity of *MLH1*, *MSH2*, *MSH6* and *PMS2* KD cell line nuclear extracts was assessed by comparing the MMR efficiencies to their respective negative controls. For that, 75 ng and 100 ng of substrate was incubated with either 50, 75 or 100 µg of the MMR gene KD cell line extracts and their controls (**I**, **III**). The digested DNA was run on a 1% agarose gel and visualized using Image-Pro 4.0 Software (Media Cybernetics) (**I-III**). The absolute repair percentage (AR%) of each nuclear extract was calculated as the amount of repaired DNA out of the total amount of DNA used in the reaction. This was done by using GeneTools (SynGene) version 3.08 (**I, II**) or version 4.03 (**III**). Average AR% was calculated from at least two independent experiments. The relative repair percentages (RR%) were calculated by comparing the AR% values of recombinant proteins to MutLα-WT or MutSα-WT protein AR% values. In KD cell lines, RR% values were calculated by comparing the AR% values of MMR gene KD cell lines to NC cell line AR% values. Statistical analyses were performed using Student's T-Test (**I-III**).

RESULTS AND DISCUSSSION

Diagnosis of Lynch syndrome relies heavily on familial history of cancer followed by pathological analyses of tumour and genetic tests. Screening based on the strict Amsterdam II criteria and the more sensitive revised Bethesda guidelines still misses a portion of LS carriers, especially *MSH6* and *PMS2* mutation carriers and families [59,60,64,220]. Indeed, up to 10% of *MSH6* and over 20% of *PMS2* carriers fail to meet these criteria and hence would have been missed if these were the only detection methods [59,60,195]. MSI-testing and the detection of MMR protein expression loss in tumour tissue are efficient ways to identify high risk individuals [221,222]. However, the sensitivity of MSI-testing depends on the choice of marker panel, which has been shown to be variable between cancer centres [221]. Furthermore, MSI-testing alone often fails to distinguish LS-associated tumours from sporadic ones, because MSI is not unique to LS [106]. Similarly, MSI-testing is not useful to identify tumours caused by *MSH6* deficiency, which often lack the MSI-H phenotype [64,223,224]. Immunohistochemical analyses of all the four MMR proteins have been shown to increase the sensitivity of LS identification, and is recommended to be used together with MSI when feasible [119,225]. However, expression of intact MMR proteins is not necessarily an indication of functionality and may therefore give false negative interpretations [62,141]. Moreover, *PMS2* studies are further challenged by the issue of reliable sequencing, which is difficult because of multiple highly homologous pseudogenes [172,175,176]. Due to these limitations in LS diagnosis, we aimed to determine whether reduced mRNA expression of *MLH1*, *MSH2*, *MSH6* **(I)** and *PMS2* **(III)** genes, comparable to MMR gene expression levels in LS mutation carriers, has an effect on repair efficiency, which could be detected with the functional test.

An additional challenge in the diagnosis and management of Lynch syndrome is the occurrence of variants whose effect on the protein function remains undetermined. In fact, up to one-third of MMR gene mutations are variants of uncertain significance (VUS), whose pathogenicity remains unclear [13]. However, the assessment of pathological nature of these variants is extremely important for Lynch syndrome diagnosis. It is needed to identify LS families, in order to provide counselling and accurate surveillance regimens to real risk individuals and to avoid unnecessary follow-up of non-LS individuals. Here, by functionally assessing the pathogenicity of *MLH1* and *MSH2* missense variants (**II**) we show the importance of the *in vitro* MMR assay in classifying VUSs.

Lowered MMR gene expression affects repair capability in a gene dependent manner (**I, III**)

Lynch syndrome mutation carriers inherit one healthy and one defective MMR gene allele. The assumption is that in a heterozygous cell in which the MMR gene mRNA expression level is decreased, the MMR capability is decreased accordingly. However, prior to the present studies, it had not been shown how the decrease in the expression levels of different MMR genes affects MMR efficiency. Here, the functional *in vitro* MMR assay was used to study the consequences of reduced mRNA expression on repair capability.

Transfection of four MMR gene shRNA plasmids into human fibroblasts and subsequent monoclonalisation yielded KD cell lines with various *MLH1*, *MSH2*, *MSH6* and *PMS2* mRNA expression levels. For *MLH1*, *MSH2* and *MSH6* genes, the aim was to functionally assess cell lines retaining approximately 25%, 50% and 75% of mRNA expression. Due to the lower penetrance estimates in *PMS2* families, the expression level for *PMS2* KD cell lines was assessed to be at mutation carrier level (50%) or less. Cell lines closest to these expected expression levels were chosen for the MMR assay. Thus, cells retaining 23%, 50%, and 74% of *MLH1* mRNA expression; 26%, 47%, and 68% of *MSH2* mRNA expression; 51% and 79% of *MSH6* mRNA expression; and 19%, 33%, and 53% of *PMS2* mRNA expression were chosen out of twenty four *MLH1*, sixteen *MSH2*, nineteen *MSH6* and sixteen *PMS2* KD clones (**Figure 8**). We were not able to establish *MSH6* KD clones with mRNA expression levels of approximately 25%, thus the repair efficiencies of only two different *MSH6* expression levels were analysed (**Figure 8**). The protein expression levels of each gene of interest (GOI) and their heterodimeric partners were verified by Western blot (WB) analysis from the nuclear protein extract of the KD cell lines and their respective controls. Each KD cell line had their own negative control (NC) throughout all analyses, which enabled us to compare the differences in repair capability and also to eliminate any possible experimental artefacts that might occur in such a multi-step experimental system.

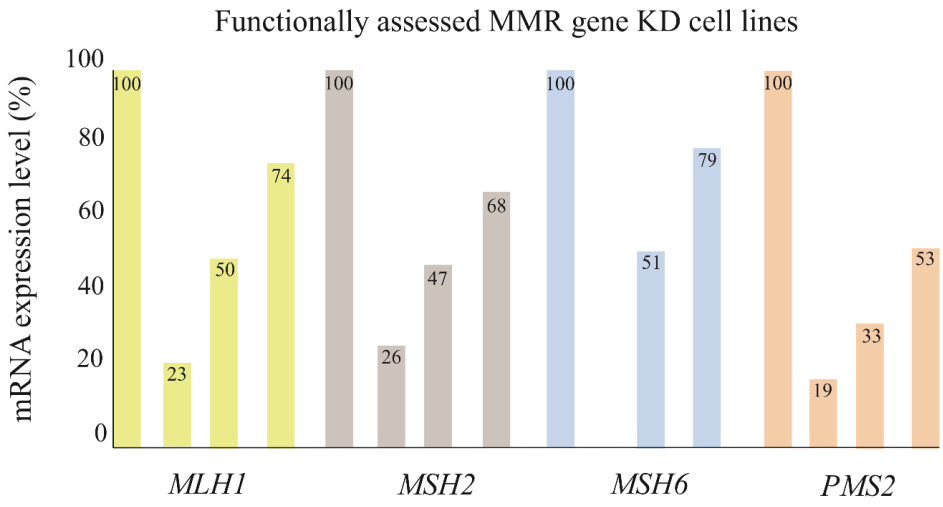

Figure 8. *MLH1*, *MSH2*, *MSH6* and *PMS2* KD cell lines chosen for functional analyses

The KD cell lines were chosen based on their mRNA expression levels verified by Q-PCR. The mRNA levels retained in KD cell lines are presented as percentages on the y-axis and in ascending order for each MMR gene. Only two *MSH6* KD cell lines met the inclusion criteria and were included in the analysis.

Nuclear extracts (NE) from each KD cell line were used to assess their repair efficiency in the *in vitro* MMR assay. Three different amounts of NE (50 µg, 75 µg or 100 µg) were used for *MLH1*, *MSH2* and *MSH6* assessments. For *PMS2*, only 75 µg of NE was used since it was already proved to be the optimal amount for usage in the functional assay (**I**). Here, the results of 75 µg of NE of each GOI analysed in the functional assay are presented in **Figure 9**. 23% and 26% of the mRNA expression level of *MLH1* and *MSH2*, respectively, led to complete MMR deficiency. This was confirmed by the finding that the increased quantity of NE from these cells did not rescue the MMR proficiency. The completely opposite result was detected in *PMS2* KD cells retaining 19% of *PMS2* mRNA expression level: these cells still maintained MMR proficiency, and surprisingly showed the highest repair capability among all *PMS2* KD lines analysed. Another *PMS2* cell line below the carrier-like expression level, which retained 33% of mRNA expression, presented a statistically significant decrease in MMR efficiency compared to its respective control in *PMS2* ($P = 0.001$). Carrier-like expression levels in *MSH2*, *MSH6* and *PMS2* led to a significant decrease in MMR efficiency ($P = 0.009$, 0.003, 0.0002 respectively), whereas the ~50% mRNA expression in *MLH1* led to only a slight decrease in repair. Overall, 75% of the GOI mRNA expression level was enough to maintain some repair in *MLH1*, *MSH2* and *MSH6*. However, *MSH2* and *MSH6* repaired significantly less efficiently than their controls ($P = 0.043$, 0.03 respectively),

while *MLH1* repaired at a level comparable to its control. No cell line in that range was established for the *PMS2* gene.

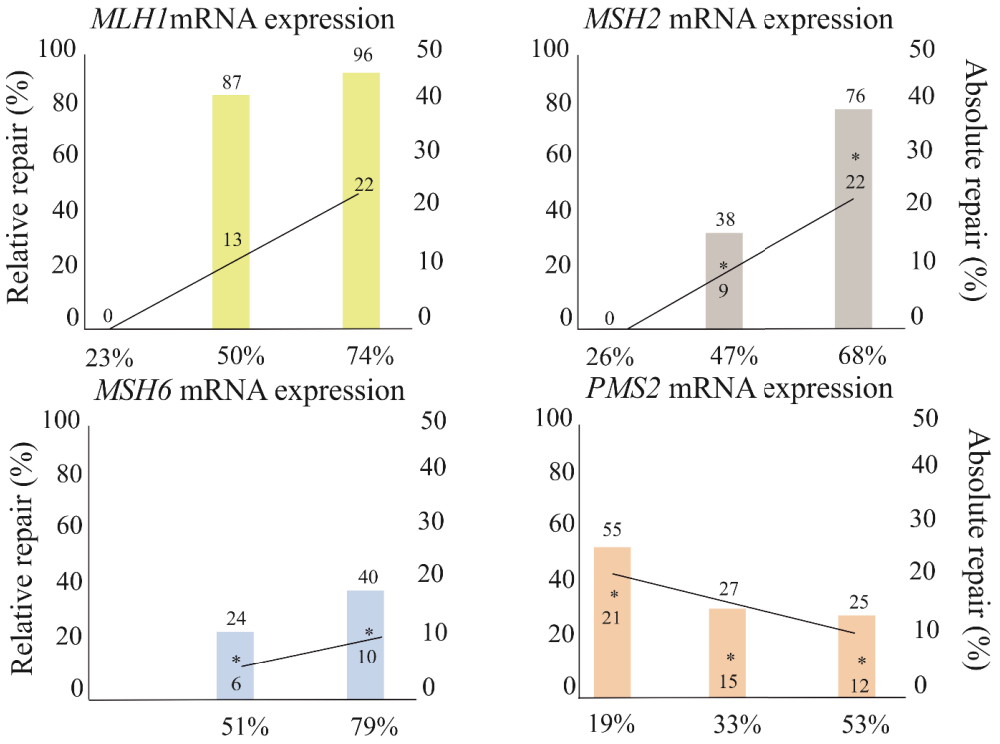

Figure 9. Relative (RR%) and absolute (AR%) repair efficiencies of *MLH1*, *MSH2*, *MSH6* and *PMS2* KD cell lines.

The percentage of repaired DNA of the total amount of DNA indicates absolute repair efficiencies (AR%). The average AR% of at least two independent assay results are depicted on the line corresponding to the y-axis on the right for each MMR gene assayed with 75µg of nuclear protein extract. Relative repair efficiencies (RR%) depicted on the y-axis on the left are a comparison of each KD cell line AR% value to their respective control AR% value. Cell lines where a significant decrease (based on p-value) was detected compared to the control are marked with *.

The full impairment of the MMR capability, seen in *MLH1* and *MSH2* KD cells retaining less than 50% of mRNA expression, could be explained by their central roles in MutLα and MutSα complexes, respectively [55,169]. Interestingly, such reductions in *PMS2* expression (19% and 33% respectively) resulted in higher MMR efficiency than the carrier-like level (53%). Cells that retained 53% of *PMS2* expression showed the lowest repair efficiency compared to two other *PMS2* cell lines. This result is unexpected and suggests that such a low level of *PMS2* in the cell may be compensated by

some homologous protein. Indeed, MLH3 (as part of the MutLγ (MLH1+MLH3) complex) functions as a backup for PMS2 [56]. Moreover, the nuclear localization of MLH3 has been demonstrated to be conditional and dependent on the abundance of PMS2 in the cell [57], further supporting this hypothesis. However, we could not confirm the hypothesized compensatory mechanism, since although slightly increased MLH3 protein expression was observed in the two *PMS2* KD cell lines retaining 19% and 33% of mRNA expression, the *MLH3* mRNA expression levels were decreased. Nevertheless, PMS2 has a substantial role in MMR as a component of MutLα. Our results suggesting that the *PMS2* heterozygous state has devastating effects on the repair efficiency supports the survival hypothesis, but needs further studies.

All MMR gene KD cell lines demonstrated reduced MMR efficiency with ~50% decrease in mRNA level, indicating that a carrier-level is efficiently recognized. Interestingly, *MLH1* KD cells indicated only a slight decrease in repair efficiency, while *MSH2*, *MSH6* and *PMS2* KD cell lines were highly sensitive to reduction. Immortalized lymphoblasts from *MLH1* mutation carriers have demonstrated similar repair capability as non-carrier lymphoblasts, whereas lymphoblast cells from *MSH2* heterozygous carriers revealed significantly reduced MMR capability as well as increased tolerance to methylating agent-treatment [226]. Moreover, stable expression of exogenous *MLH1* to 293T cell line ($MLH1^-$) demonstrated that low levels of *MLH1* are capable of restoring the MMR capability, further underscoring our findings [227]. A decrease of approximately 25% of *MLH1* expression was shown to have the weakest effect on MMR efficiency compared to *MSH2* and *MSH6* cells.

Here, by using a functional approach we have assessed the significance of lowered MMR gene expression on repair efficiency. The *in vitro* MMR assay proved to be sensitive for distinguishing MMR gene mRNA expression levels indicative of LS, hence we present a novel approach for assessing the role of *MLH1*, *MSH2*, *MSH6* and *PMS2* genes in LS.

Functional data helps to classify variants of uncertain significance (**II**)

As a part of an international collaboration project, the pathogenicity of 13 VUSs in *MLH1* and 11 in *MSH2* was assessed, and the results led to either the improvement of the existing classification or provided the classification for the first time. The functional assessment is based on previously developed standardized five-tiered system [16] using standards assigned by the International Agency for Research on Cancer (IARC) [208]. Each VUS was extensively assessed using population frequency, segregation, tumour pathological data in addition to RNA- and protein-based analyses where feasible. Here, we made recombinant proteins of five *MLH1* and four *MSH2* missense variants using site-directed mutagenesis and insect cell-based protein production. The protein expression levels of these missense variants were determined with WB analysis. Using the *in vitro* MMR assay, repair efficiencies were determined by complementing the MMR deficient cancer cell lines lacking either MLH1 or MSH2 protein with the produced recombinant proteins. Additional information on both IHC/MSI-testing and RNA/splicing assay results were used when possible to obtain a uniform classification for these nine *MLH1* or *MSH2* DNA variants.

Two out of the five MLH1 variants (p.Gly101Ser and p.Leu260Arg) demonstrated complete MMR deficiency in the functional assay. Of these, only the MLH1 p.Leu260Arg variant showed decreased MLH1 expression in WB, yet normal PMS2 protein expression. The MLH1 variants p.Gly101Ser and p.Glu605Ala indicated a slight reduction of MLH1 protein, yet the p.Glu605Ala variant retained MMR proficiency in the functional assay. Three out of the four MSH2 variants were incapable of repair (p.Gly669Val, p.Phe694Ser and p.Pro696Leu). The MHS2 p.Gly669Val variant had the same protein expression levels as the wild type (WT), while p.Phe694Ser and p.Pro696Leu variants indicated severe reduction in MutSα expression compared to the MutSα WT control. The MMR efficiencies of each recombinant MutLα (MLH1+PMS2) and MutSα (MSH2+MSH6) variants are demonstrated as relative repair efficiency (RR%) in **Figure 10**. RR% values are calculated from three independent experiments as the ratio of protein variant absolute repair (AR%) to WT control AR%, the latter set to 100. Two MMR proficient variants (MLH1 p.Leu348Ser and MSH2 p.Lys82Glu) demonstrated slightly reduced relative repair efficiencies (**Figure 10**), while the protein expression levels were comparable to the MutLα- and MutSα-WT levels, respectively.

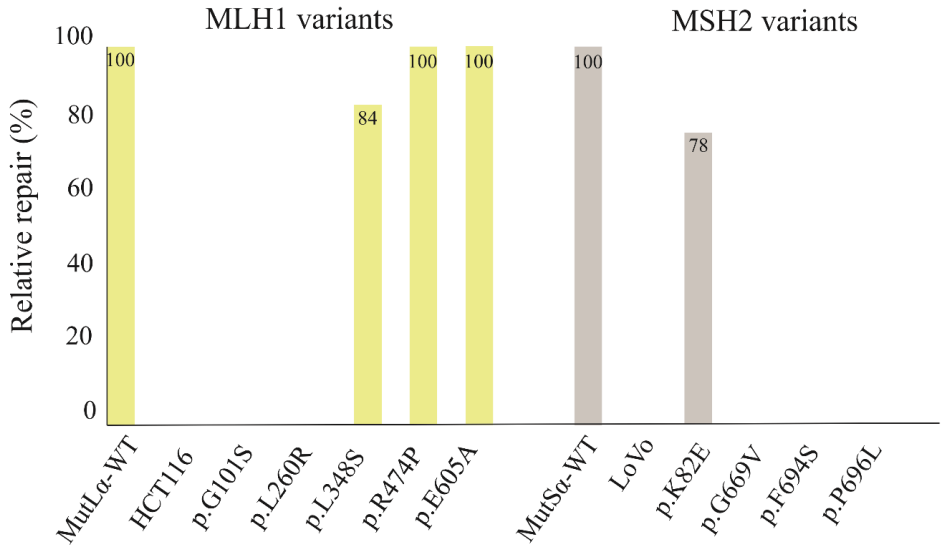

Figure 10. Relative repair efficiency of MLH1 and MSH2 protein variants.

The relative repair efficiencies of nine recombinant MutLα- and MutSα-variants. The amino acids are abbreviated (see beginning of the thesis). MutLα-WT (MLH1+PMS2) or MutSα-WT (MSH2+MSH6) was used as a positive control; HCT116 cell line (MLH1-) or LoVo (MSH2-) was used as a negative control for studying MLH1 or MSH2 variant proteins, respectively.

The data obtained from different analyses, including our protein expression and functional analyses, facilitated the final classification of the five *MLH1* and four *MSH2* variants. The results are presented in **Table 13** (excluding population frequency and multifactorial analysis results). The IHC/MSI-testing results available for four *MLH1* variants were contradictory to our findings, yet they supported our findings for all of the *MSH2* variants. More precisely, the two MMR proficient MLH1 variants (p.Arg474Pro and p.Glu605Ala) were indicative of LS, based on the tumour analyses. All MMR deficient MSH2 variants (p.Gly669Val, p.Phe694Ser and p.Pro696Leu) demonstrated the loss of MutSα expression and were associated with MSI-H tumours. The only MMR proficient MSH2 variant (p.Lys82Glu) demonstrated intact protein expression and MSS status in a tumour.

The *BRAF* V600E mutation testing and/or promoter methylation analyses were undertaken specifically for *MLH1* variants, and the results supported our findings. Specifically, in one MMR deficient variant (p.Leu260Arg), neither the *BRAF* mutation nor the promoter methylation was detected. However, the *BRAF* mutation was confirmed in the MMR proficient variant (p.Arg474Pro). This variant also presented the loss of the other allele (**Table 13**).

Two MMR deficient variants also indicated problems in RNA analyses. A potential splice site variant, *MSH2* c.2006G>T (predicted missense change p.Gly669Val), which was assessed as MMR deficient, indicated complete exon 13 exclusion in a minigene assay (**Table 13**). Only partial skipping was observed in established lymphoblastoid cell lines from a carrier (data not shown). The MMR deficient MLH1 p.Gly101Ser variant also indicated partial exon skipping.

Table 13. MLH1 and MSH2 protein variants, assay results and proposed classification

Gene	Exonic location	DNA change	Predicted amino acid change	Protein expression	MMR efficiency	IHC	MSI	*BRAF* p.V600E	*MLH1* promoter methylation	LOH	Minigene splicing assay	Proposed classification
MLH1	3	c.301G>A	p.Gly101Ser	$MLH1^{+/-}$	deficient	nt	MSI-H (1)	nt	WT	nt	Total inclusion of exon 3 with concomitant loss of alternative 5' ss	Class 5*
	9	c.779T>G	p.Leu260Arg	$MLH1^{+/-}$	deficient	$MLH1^{-}$	MSI-H (6)/ MSS (1)	WT	WT	no	No effect	Class 5+
	12	**c.1043T>C**	**p.Leu348Ser**	WT	proficient	normal (1)	MSS (1)	nt	nt	no	No effect	Class 3n
	13	c.1421G>C	p.Arg747Pro	WT	proficient	MLH1-	MSI-H	mut	nt	variant allele	No effect	Class 3+
	16	**c.1814A>C**	**p.Glu605Ala**	$MLH1^{+/-}$	proficient	MLH1-	MSI-H (1)	nt	nt	no	No effect	Class 4n
Gene	**Exonic location**	**DNA change**	**Predicted amino acid change**	**Protein expression**	**MMR efficiency**	**IHC**	**MSI**	***BRAF* p.V600E**	***MLH1* promoter methylation**	**LOH**	**Minigene splicing assay**	**Proposed classification**
MSH2	2	**c.244A>G**	**p.Lys82Glu**	WT	proficient	normal	MSS (1)	nt	nt	no	No effect	Class 2n
	13	c.2006G>T	p.Gly669Val	WT	deficient	$MSH2^{-}$ /$MSH6^{-}$ (1)	MSI-H (1)	nt	nt	no	Complete exclusion of exon 13	Class 5+
	13	c.2081T>C	p.Phe694Ser	$MSH2^{-}$ $MSH6^{+/-}$	deficient	$MSH2^{-}$ /$MSH6^{-}$ (3)	MSI-H (3)	nt	nt	nt	No effect	Class 5+
	13	c.2087C>T	p.Pro696Leu	$MSH2^{+/-}$ $MSH6^{+/-}$	deficient	$MSH2^{-}$ /$MSH6^{-}$ (1)	MSI-H (3)	nt	nt	no	No effect	Class 5+

Previously unclassified variants are indicated in **bold**; the results obtained in this study are highlighted in dark grey.
IHC - immunohistochemistry; MSI - microsatellite instability; LOH - loss of heterozygosity
nt - not tested; WT – wild type; mut – mutated; ss – splice site
*-Previous InSiGHT classification changed and new proposal based on results from this study
+-Previous InSiGHT classification confirmed and proposal based on results from this study
n-novel variant classified in this study and a proposed classification
Class 2 – likely not pathogenic; Class 3 – VUS; Class 4 – likely pathogenic; Class 5 – pathogenic

The former InSiGHT Class 5 (pathogenic) was confirmed for one *MLH1* and three *MSH2* variants. The tumour pathological data, methylation/*BRAF* testing as well as our analyses unanimously indicated that the MLH1 p.Leu260Arg variant is LS causative. However, the tumour data together with our *in vitro* MMR assay also suggested pathogenicity for MSH2 variants (p.Gly669Val, p.Phe694Ser and p.Pro694Leu), while the protein expression levels detected by WB were variable. Thus, the detected MSH2 protein expression in a tumour may not be an indication of protein functionality, as was previously shown with *MSH2* missense mutation [143]. Interestingly, all of the three *MSH2* variants classified were located in the ATPase domain, which is suggestive that this protein domain of MSH2 is somehow sensitive to alterations. Indeed, as the binding and hydrolysis is essential for MutSα's MMR activity, it could be anticipated that structural changes in the ATPase domain cause harmful effects on the protein function [47,228]. Four previously studied *MSH2* variants located in the ATPase domain were assessed as pathogenic [217]. Although no clear associations between certain protein domains and mutation frequencies in MMR genes have been determined [6], currently slightly more than 50% of the listed unique *MHS2* exon 13 missense alterations are classified as pathogenic in the InSiGHT database (https://databases.lovd.nl/shared/variants/MSH2/), thus indicating that mutations in the ATPase domain of MSH2 have adverse effects on protein function. One MLH1 variant (p.Arg474Pro) remained as a variant of uncertain significance based on observed discrepancies between the tumour pathological data (indicating pathogenicity) and the functional analysis (indicating proficiency).

The inclusion of new data on tumour characteristics and/or segregation data in addition to functional protein analysis led to the successful reclassification of one Class 4 (likely pathogenic) MLH1 variant (p.Gly101Ser) to a Class 5 (pathogenic variant). This alteration does not affect the screening/surveillance recommendations for a putative mutation carrier, and it is highly unlikely that a pathogenic variant would change into a non-pathogenic one with the inclusion of additional data. Thus, current guidelines for variant interpretation recommend the collection of as much information as possible for variants assigned to either Class 2 (likely not pathogenic) or to Class 4 (likely pathogenic), in order to provide a definitive classification [16,208].

For two novel MLH1 variants (p.Leu348Ser, p.Glu605Ala) and for one novel MSH2 variant (p.Lys82Glu), the MSI status was available from only one tumour. The MLH1 p.Leu348Ser variant had the least available data and was thus moved to Class 3 (variant of uncertain clinical significance). Although, the MMR proficiency and MSS tumour are an indication of non-pathogenicity, as was the lack of major splicing alterations, we followed the interpretation guidelines for breast cancer genes that

advise not to classify variants with limited clinical/laboratory evidence and instead to consider them as VUS until more evidence is available for assignment into a clinically actionable class (https://enigmaconsortium.org/). Although clear discrepancies were observed between tumour pathological/multifactorial analysis data compared to our functional MMR assay for the MLH1 variant p.Glu605Ala, it was classified as a likely pathogenic variant (Class 4). Unfortunately, no *BRAF* mutation/methylation testing was carried out for this variant – a result from either test could have aided the final classification. Finally, MSH2 p.Lys82Glu variant's tumour features and MMR activity indicated its proficiency and led to being assigned as a likely not pathogenic variant (Class 2). Nevertheless, neither of these variants are assigned permanently to these classes due to insufficient evidence, and are hence regarded for the time being as VUSs (https://enigmaconsortium.org/).

The LOVD database (https://databases.lovd.nl/shared/genes/), curated by the InSiGHT, facilitates the classification of MMR gene variants, thus promoting the identification of LS families. Importantly, VUS pathogenicity has been demonstrated on several occasions using functional analyses that indicate the effect of the protein variant on MMR efficiency [141,153]. Here, we have also demonstrated the importance of the functional assay in determining the significance of a variant. In comparison to silent mutations or intronic variants, the effect of a non-truncating MMR gene variant on the functionality is more complex to interpret. As such, obtaining multiple points of evidence is extremely important to facilitate the final classification. By assessing one of the hallmarks of LS associated tumours – MMR deficiency – we have efficiently determined the significance of the majority of the tested variants. Alterations that remained as VUS after MMR activity assessment require additional clinical evidence for determining their possible non-pathogenicity or pathogenicity.

CONCLUSIONS AND FUTURE PROSPECTS

The main aim of the present thesis was to elucidate the impact of lowered gene expression in mismatch repair genes *MLH1*, *MSH2*, *MSH6* and *PMS2* on mismatch repair capability, and to determine the functional significance of five *MLH1* and four *MSH2* missense variants found in suspected Lynch syndrome families.

The *in vitro* MMR assay enables the assessment of the repair capability of cells, and is therefore suitable for analysing the effect of MMR gene expression decrease on the repair efficiency. For the first time, we applied the stable shRNA knockdown approach to achieve cell lines with variable MMR gene expression, which allowed us to assess the effect of MMR gene expression change on the repair capability. We demonstrated that the *in vitro* MMR assay is sensitive enough to detect repair differences due to lowered *MLH1*, *MSH2*, *MSH6* and *PMS2* mRNA expression. Interestingly, the effect of decreased MMR gene expression on repair efficiency varied between genes. Mismatch repair capability of the cells was most sensitive to reductions in *MSH2* and *MSH6* expression, with carrier-like (~50%) expression levels significantly decreasing repair efficiency. Cells were less sensitive to ~50% *MLH1* expression reduction, while the repair capability was lost when only approximately 25% of *MLH1* or *MSH2* expression was left. Remarkably, carrier-like *PMS2* expression level in the cell caused severe effects on the repair compared to cell lines retaining lower than 50% of *PMS2* expression. These findings confirm that with the functional assessment we were able to distinguish MMR gene expression reductions, similar to MMR gene expression decrease in Lynch syndrome carriers, and highlight the importance of gene-specific characteristics that contribute to MMR deficiency.

Variants whose pathogenicity remains undetermined after sequencing require further analyses for classification as non-pathogenic and pathogenic. The International Society for Gastrointestinal Hereditary Tumours continues to significantly contribute to the classification of variants of uncertain significance (VUSs), by gathering MMR gene variant-related information to a database available to medical experts and scientists. Our results demonstrate that the functional *in vitro* MMR assessment of VUSs contributes significantly for their pathogenicity classification. Indeed, of the functionally analysed five missense variants in *MLH1* and four in *MSH2*, three variants were novel, and thus previously unclassified by InSiGHT. In addition, we were able to give a definitive classification to one *MLH1* variant and confirm previous InSiGHT classifications to five VUSs. Thus, the functional analyses helped to determine the clinical significance of nine missense variants.

Findings from these studies provide new insights into the severity of the malfunction that decreased levels of MMR gene expression can cause. Furthermore, the results show that the functional assessment of variants of uncertain significance considerably helps their pathogenicity assessment. Both of these findings may have an important impact in Lynch syndrome diagnosis in the future. Since the carrier-like expression level of the *PMS2* gene was shown to cause surprisingly severe mismatch repair malfunction, one of the future prospects in our studies includes the pathogenicity assessment of *PMS2* mutations/VUSs with the functional test. The functional test itself has already been further developed to be of high quality and sensitive, and taken into use for variant pathogenicity assessment, while our future prospects include its use for variant classification.

ACKNOWLEDGEMENTS

This work has been carried out under the supervision of Professor Minna Nyström and postdoctoral researcher Minttu Kansikas, at the Faculty of Biological and Environmental Sciences, University of Helsinki. The present study was financially supported by the European Research Council and Jane and Aatos Erkko Foundation. Doctoral Programme in Integrative Life Science and the University of Helsinki are acknowledged for supporting participation in scientific conferences.

My sincere gratitude to Professor Minna Nyström for the opportunity to participate in these fascinating projects. Thank you for introducing me to the world of Lynch syndrome. I am truly honoured to be entrusted with all the tasks that are included in this thesis and I will always remember the encouragement and support during these years. I also want to acknowledge my second supervisor, postdoctoral researcher Minttu Kansikas. I thank her for teaching me all the methods in the laboratory. Her diligence and the ability to always be professional are qualities that I admire. I wish to thank her for being a role model to me. Her passion for science is motivating.

Professor Minna Pöyhönen and Professor Jukka Partanen are acknowledged for pre-reviewing this thesis and for providing valuable comments. I am also thankful to thesis committee members, docent Katarina Pelin and Professor Juha Klefström, for all the constructive criticism and guidance.

I am grateful to all the people that I have worked with. Professor Maurizio Genuardi and his colleagues are thanked for the opportunity to study the *MLH1* and *MSH2* VUSs in Italian Lynch syndrome families. Professor Päivi Peltomäki and her research group is thanked for long-term collaboration and inspiring science in the field of Lynch syndrome. The friendly staff at the Faculty of Biological and Environmental Sciences is acknowledged for their support. Specifically I would like to mention laboratory coordinator Outi Kostia, research technician Arja Ikävalko and university lecturer Pekka Heino. I would like to thank all my former and present colleagues for their endless support and friendly atmosphere at work. My sincere gratitude to postdoctoral researcher Jukka Kantelinen for teaching me everything in the laboratory. Jukka, you were my non-official supervisor. It has been very interesting to get to know you. Thanks to you I started speaking Finnish quite soon after moving to Finland, which has made my life here so much easier. Thank you for everything! I would especially like to thank postdoctoral researcher Laura Vähätalo. I will always remember our small PCR meetings and your selfless guidance with any scientific problems that I encountered. Thank you for all the non-

scientific talks we had, that kept me sane. You are a true Friend and I hope that we always keep close contact. Postdoctoral researcher Jaana Putula, thank you for always being so friendly and cheerful. Your spirit has a positive influence on the working days. Anni Dohlen, thank you so much for the thesis cover image and your friendship. Marjaana, Sonja, Valtteri and Laura K. – your support and companionship were valuable.

I wish to express my gratitude to my friends outside the laboratory who have stayed by my side even when I have not been so active myself. Friends, who have always found time for me and who have made me smile. My dear Kerttu, Milen, Ester and Annika. Ester, your support during the year 2019 has been remarkable – I am so grateful to you!

Above all, my special appreciation to my relatives for their endless love, support and interest into my work. Special acknowledgments to the extended KASELA family. I express my warmest thanks to my loving grandparents Tiiu and Tõnis, great-aunt Reet, uncle Jüri's and aunt Krista's families. Especially I would like to thank my dear grandfather Tõnis for believing in me and for celebrating even the smallest accomplishments. My parents Leelo and Marti, my brothers Markus, Mattias and Magnus and our lovely dog Sheba deserve my special appreciation. Thank you for being patent with me and for teaching me not to give up. Your unconditional love means the world to me!

Finally, I dedicate this work to my own little family. Kristjan, I would not be writing these words here if you had not been by my side the whole time. I do not have enough words to thank you for all your support and understanding during these years. You have seen the good, the bad - all! Thank you for being a father and a mother to our sons while I was concentrating on my thesis. This thesis is literally our first child that has fallen to the third place. Last but not least, my boys - Karl Kristjan and Karel Kaspar. Thank you for all the laughter, playfulness and questions – please, never stop questioning. You have already taught me more than I can ever teach you. Through the moments we have shared I know that together we can accomplish everything! Love you to the moon and back!

Helsinki, January 2020

Mariann Kasela

REFERENCES

1. Vasen HF, Tomlinson I, Castells A. Clinical management of hereditary colorectal cancer syndromes. *Nat Rev Gastroenterol Hepatol*. 2015;12(2):88-97. doi: 10.1038/nrgastro.2014.229 [doi].

2. Lynch HT, de la Chapelle A. Hereditary colorectal cancer. *N Engl J Med*. 2003;348(10):919-932. doi: 10.1056/NEJMra012242 [doi].

3. Knudson AG,Jr. Hereditary cancer, oncogenes, and antioncogenes. *Cancer Res*. 1985;45(4):1437-1443.

4. Boland CR, Lynch HT. The history of lynch syndrome. *Fam Cancer*. 2013;12(2):145-157. doi: 10.1007/s10689-013-9637-8 [doi].

5. Lynch HT, Krush AJ. Cancer family "G" revisited: 1895-1970. *Cancer*. 1971;27(6):1505-1511. doi: 10.1002/1097-0142(197106)27:63.0.co;2-l [doi].

6. Peltomaki P. Update on lynch syndrome genomics. *Fam Cancer*. 2016;15(3):385-393. doi: 10.1007/s10689-016-9882-8 [doi].

7. Boland CR. Evolution of the nomenclature for the hereditary colorectal cancer syndromes. *Fam Cancer*. 2005;4(3):211-218. doi: 10.1007/s10689-004-4489-x [doi].

8. Dominguez-Valentin M, Sampson JR, Seppala TT, et al. Cancer risks by gene, age, and gender in 6350 carriers of pathogenic mismatch repair variants: Findings from the prospective lynch syndrome database. *Genet Med*. 2019. doi: 10.1038/s41436-019-0596-9 [doi].

9. Barrow E, Robinson L, Alduaij W, et al. Cumulative lifetime incidence of extracolonic cancers in lynch syndrome: A report of 121 families with proven mutations. *Clin Genet*. 2009;75(2):141-149. doi: 10.1111/j.1399-0004.2008.01125.x [doi].

10. Lynch HT, Snyder CL, Shaw TG, Heinen CD, Hitchins MP. Milestones of lynch syndrome: 1895-2015. *Nat Rev Cancer*. 2015;15(3):181-194. doi: 10.1038/nrc3878 [doi].

11. Watson P, Lin KM, Rodriguez-Bigas MA, et al. Colorectal carcinoma survival among hereditary nonpolyposis colorectal carcinoma family members. *Cancer*. 1998;83(2):259-266. doi: 10.1002/(SICI)1097-0142(19980715)83:23.0.CO;2-L [pii].

12. Vasen HF, 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;62(6):812-823. doi: 10.1136/gutjnl-2012-304356 [doi].

13. Sijmons RH, Greenblatt MS, Genuardi M. Gene variants of unknown clinical significance in lynch syndrome. an introduction for clinicians. *Fam Cancer*. 2013;12(2):181-187. doi: 10.1007/s10689-013-9629-8 [doi].

14. Couch FJ, Rasmussen LJ, Hofstra R, et al. Assessment of functional effects of unclassified genetic variants. *Hum Mutat*. 2008;29(11):1314-1326. doi: 10.1002/humu.20899 [doi].

15. Borras E, Chang K, Pande M, et al. In silico systems biology analysis of variants of uncertain significance in lynch syndrome supports the prioritization of functional molecular validation. *Cancer Prev Res (Phila)*. 2017;10(10):580-587. doi: 10.1158/1940-6207.CAPR-17-0058 [doi].

16. Thompson BA, Spurdle AB, Plazzer JP, et al. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nat Genet*. 2014;46(2):107-115. doi: 10.1038/ng.2854 [doi].

17. Plazzer JP, Sijmons RH, Woods MO, et al. The InSiGHT database: Utilizing 100 years of insights into lynch syndrome. *Fam Cancer*. 2013;12(2):175-180. doi: 10.1007/s10689-013-9616-0 [doi].

18. Bashyam MD, Animireddy S, Bala P, Naz A, George SA. The yin and yang of cancer genes. *Gene*. 2019;704:121-133. doi: S0378-1119(19)30376-2 [pii].

19. Knudson AG,Jr. Mutation and cancer: Statistical study of retinoblastoma. *Proc Natl Acad Sci U S A*. 1971;68(4):820-823. doi: 10.1073/pnas.68.4.820 [doi].

20. Leslie A, Carey FA, Pratt NR, Steele RJ. The colorectal adenoma-carcinoma sequence. *Br J Surg*. 2002;89(7):845-860. doi: 2120 [pii].

21. Garborg K. Colorectal cancer screening. *Surg Clin North Am*. 2015;95(5):979-989. doi: 10.1016/j.suc.2015.05.007 [doi].

22. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut*. 2017;66(4):683-691. doi: 10.1136/gutjnl-2015-310912 [doi].

23. Jarvinen HJ, Renkonen-Sinisalo L, Aktan-Collan K, Peltomaki P, Aaltonen LA, Mecklin JP. Ten years after mutation testing for lynch syndrome: Cancer incidence and outcome in mutation-positive and mutation-negative family members. *J Clin Oncol*. 2009;27(28):4793-4797. doi: 10.1200/JCO.2009.23.7784 [doi].

24. Watson P, Lynch HT. The tumor spectrum in HNPCC. *Anticancer Res*. 1994;14(4B):1635-1639.

25. Watson P, Vasen HFA, Mecklin JP, et al. The risk of extra-colonic, extra-endometrial cancer in the lynch syndrome. *Int J Cancer*. 2008;123(2):444-449. doi: 10.1002/ijc.23508 [doi].

26. Win AK, Jenkins MA, Dowty JG, et al. Prevalence and penetrance of major genes and polygenes for colorectal cancer. *Cancer Epidemiol Biomarkers Prev*. 2017;26(3):404-412. doi: 10.1158/1055-9965.EPI-16-0693 [doi].

27. Wimmer K, Etzler J. Constitutional mismatch repair-deficiency syndrome: Have we so far seen only the tip of an iceberg? *Hum Genet*. 2008;124(2):105-122. doi: 10.1007/s00439-008-0542-4 [doi].

28. 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(6):355-365. doi: 10.1136/jmedgenet-2014-102284 [doi].

29. Jarvinen HJ. Epidemiology of familial adenomatous polyposis in finland: Impact of family screening on the colorectal cancer rate and survival. *Gut*. 1992;33(3):357-360. doi: 10.1136/gut.33.3.357 [doi].

30. Bisgaard ML, Fenger K, Bulow S, Niebuhr E, Mohr J. Familial adenomatous polyposis (FAP): Frequency, penetrance, and mutation rate. *Hum Mutat*. 1994;3(2):121-125. doi: 10.1002/humu.1380030206 [doi].

31. Carr S, Kasi A. Familial adenomatous polyposis. In: *StatPearls.* Treasure Island (FL): StatPearls Publishing LLC; 2019. NBK538233 [bookaccession].

32. Elsayed FA, Tops CMJ, Nielsen M, et al. Low frequency of POLD1 and POLE exonuclease domain variants in patients with multiple colorectal polyps. *Mol Genet Genomic Med*. 2019;7(4):e00603. doi: 10.1002/mgg3.603 [doi].

33. 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(13):3506-3512. doi: 10.1093/hmg/ddu058 [doi].

34. Jelsig AM, Qvist N, Brusgaard K, Nielsen CB, Hansen TP, Ousager LB. Hamartomatous polyposis syndromes: A review. *Orphanet J Rare Dis*. 2014;9:101-1172-9-101. doi: 10.1186/1750-1172-9-101 [doi].

35. van Lier MG, Wagner A, Mathus-Vliegen EM, Kuipers EJ, Steyerberg EW, van Leerdam ME. High cancer risk in peutz-jeghers syndrome: A systematic review and surveillance recommendations. *Am J Gastroenterol*. 2010;105(6):1258-64; author reply 1265. doi: 10.1038/ajg.2009.725 [doi].

36. Beggs AD, Latchford AR, Vasen HF, et al. Peutz-jeghers syndrome: A systematic review and recommendations for management. *Gut*. 2010;59(7):975-986. doi: 10.1136/gut.2009.198499 [doi].

37. Brosens LA, van Hattem A, Hylind LM, et al. Risk of colorectal cancer in juvenile polyposis. *Gut*. 2007;56(7):965-967. doi: gut.2006.116913 [pii].

38. Tan MH, Mester JL, Ngeow J, Rybicki LA, Orloff MS, Eng C. Lifetime cancer risks in individuals with germline PTEN mutations. *Clin Cancer Res*. 2012;18(2):400-407. doi: 10.1158/1078-0432.CCR-11-2283 [doi].

39. Lynch HT, Drouhard TJ, Schuelke GS, Biscone KA, Lynch JF, Danes BS. Hereditary nonpolyposis colorectal cancer in a navajo indian family. *Cancer Genet Cytogenet*. 1985;15(3-4):209-213. doi: 0165-4608(85)90164-5 [pii].

40. Boland PM, Yurgelun MB, Boland CR. Recent progress in lynch syndrome and other familial colorectal cancer syndromes. *CA Cancer J Clin*. 2018;68(3):217-231. doi: 10.3322/caac.21448 [doi].

41. Jass JR, Smyrk TC, Stewart SM, Lane MR, Lanspa SJ, Lynch HT. Pathology of hereditary non-polyposis colorectal cancer. *Anticancer Res*. 1994;14(4B):1631-1634.

42. Seppala T, Pylvanainen K, Renkonen-Sinisalo L, et al. Diagnosis and treatment of lynch syndrome. *Duodecim*. 2016;132(3):233-240.

43. Espenschied CR, LaDuca H, Li S, McFarland R, Gau CL, Hampel H. Multigene panel testing provides a new perspective on lynch syndrome. *J Clin Oncol*. 2017;35(22):2568-2575. doi: 10.1200/JCO.2016.71.9260 [doi].

44. Roberts ME, Jackson SA, Susswein LR, et al. MSH6 and PMS2 germ-line pathogenic variants implicated in lynch syndrome are associated with breast cancer. *Genet Med*. 2018;20(10):1167-1174. doi: 10.1038/gim.2017.254 [doi].

45. Ten Broeke SW, Suerink M, Nielsen M. Response to roberts et al. 2018: Is breast cancer truly caused by MSH6 and PMS2 variants or is it simply due to a high prevalence of these variants in the population? *Genet Med*. 2019;21(1):256-257. doi: 10.1038/s41436-018-0029-1 [doi].

46. de la Chapelle A. Genetic predisposition to colorectal cancer. *Nat Rev Cancer*. 2004;4(10):769-780. doi: nrc1453 [pii].

47. Kunkel TA, Erie DA. DNA mismatch repair. *Annu Rev Biochem*. 2005;74:681-710. doi: 10.1146/annurev.biochem.74.082803.133243 [doi].

48. Haraldsdottir S, Rafnar T, Frankel WL, et al. Comprehensive population-wide analysis of lynch syndrome in iceland reveals founder mutations in MSH6 and PMS2. *Nat Commun*. 2017;8:14755. doi: 10.1038/ncomms14755 [doi].

49. Engel C, Loeffler M, Steinke V, et al. Risks of less common cancers in proven mutation carriers with lynch syndrome. *J Clin Oncol*. 2012;30(35):4409-4415. doi: 10.1200/JCO.2012.43.2278 [doi].

50. Barrow E, Hill J, Evans DG. Cancer risk in lynch syndrome. *Fam Cancer*. 2013;12(2):229-240. doi: 10.1007/s10689-013-9615-1 [doi].

51. Vasen HF. Review article: The lynch syndrome (hereditary nonpolyposis colorectal cancer). *Aliment Pharmacol Ther*. 2007;26 Suppl 2:113-126. doi: 10.1111/j.1365-2036.2007.03479.x [doi].

52. Grindedal EM, Moller P, Eeles R, et al. Germ-line mutations in mismatch repair genes associated with prostate cancer. *Cancer Epidemiol Biomarkers Prev*. 2009;18(9):2460-2467. doi: 10.1158/1055-9965.EPI-09-0058 [doi].

53. Hendriks YM, 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(1):17-25. doi: S0016508504005785 [pii].

54. 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(2):312-322. doi: S0016-5085(05)02232-8 [pii].

55. Acharya S, Wilson T, Gradia S, et al. hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. *Proc Natl Acad Sci U S A*. 1996;93(24):13629-13634. doi: 10.1073/pnas.93.24.13629 [doi].

56. Cannavo E, Marra G, Sabates-Bellver J, et al. Expression of the MutL homologue hMLH3 in human cells and its role in DNA mismatch repair. *Cancer Res*. 2005;65(23):10759-10766. doi: 65/23/10759 [pii].

57. Korhonen MK, Raevaara TE, Lohi H, Nystrom M. Conditional nuclear localization of hMLH3 suggests a minor activity in mismatch repair and supports its role as a low-risk gene in HNPCC. *Oncol Rep*. 2007;17(2):351-354.

58. Baglietto L, Lindor NM, Dowty JG, et al. Risks of lynch syndrome cancers for MSH6 mutation carriers. *J Natl Cancer Inst*. 2010;102(3):193-201. doi: 10.1093/jnci/djp473 [doi].

59. Senter L, Clendenning M, Sotamaa K, et al. The clinical phenotype of lynch syndrome due to germline PMS2 mutations. *Gastroenterology*. 2008;135(2):419-428. doi: 10.1053/j.gastro.2008.04.026 [doi].

60. 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(4):319-325. doi: 10.1200/JCO.2014.57.8088 [doi].

61. Ten Broeke SW, van der Klift HM, Tops CMJ, et al. Cancer risks for PMS2-associated lynch syndrome. *J Clin Oncol*. 2018;36(29):2961-2968. doi: 10.1200/JCO.2018.78.4777 [doi].

62. Truninger K, Menigatti M, Luz J, et al. Immunohistochemical analysis reveals high frequency of PMS2 defects in colorectal cancer. *Gastroenterology*. 2005;128(5):1160-1171. doi: S001650850500168X [pii].

63. Watson P, Lynch HT. Cancer risk in mismatch repair gene mutation carriers. *Fam Cancer*. 2001;1(1):57-60. doi: 322810 [pii].

64. Wagner A, Hendriks Y, Meijers-Heijboer EJ, et al. Atypical HNPCC owing to MSH6 germline mutations: Analysis of a large dutch pedigree. *J Med Genet*. 2001;38(5):318-322. doi: 10.1136/jmg.38.5.318 [doi].

65. Aaltonen LA, Peltomaki P, Leach FS, et al. Clues to the pathogenesis of familial colorectal cancer. *Science*. 1993;260(5109):812-816. doi: 10.1126/science.8484121 [doi].

66. Chatterjee N, Walker GC. Mechanisms of DNA damage, repair, and mutagenesis. *Environ Mol Mutagen*. 2017;58(5):235-263. doi: 10.1002/em.22087 [doi].

67. Tubbs A, Nussenzweig A. Endogenous DNA damage as a source of genomic instability in cancer. *Cell*. 2017;168(4):644-656. doi: S0092-8674(17)30005-3 [pii].

68. Jeggo PA, Pearl LH, Carr AM. DNA repair, genome stability and cancer: A historical perspective. *Nat Rev Cancer*. 2016;16(1):35-42. doi: 10.1038/nrc.2015.4 [doi].

69. Spivak G. Nucleotide excision repair in humans. *DNA Repair (Amst)*. 2015;36:13-18. doi: S1568-7864(15)00178-0 [pii].

70. Wallace SS. Base excision repair: A critical player in many games. *DNA Repair (Amst)*. 2014;19:14-26. doi: 10.1016/j.dnarep.2014.03.030 [doi].

71. Shrivastav M, De Haro LP, Nickoloff JA. Regulation of DNA double-strand break repair pathway choice. *Cell Res*. 2008;18(1):134-147. doi: cr2007111 [pii].

72. Torgovnick A, Schumacher B. DNA repair mechanisms in cancer development and therapy. *Front Genet*. 2015;6:157. doi: 10.3389/fgene.2015.00157 [doi].

73. Sishc BJ, Davis AJ. The role of the core non-homologous end joining factors in carcinogenesis and cancer. *Cancers (Basel)*. 2017;9(7):10.3390/cancers9070081. doi: E81 [pii].

74. McCulloch SD, Kunkel TA. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res*. 2008;18(1):148-161. doi: 10.1038/cr.2008.4 [doi].

75. Jiricny J. The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol*. 2006;7(5):335-346. doi: nrm1907 [pii].

76. Fishel R, Lescoe MK, Rao MR, et al. The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. *Cell*. 1993;75(5):1027-1038. doi: 0092-8674(93)90546-3 [pii].

77. Leach FS, Nicolaides NC, Papadopoulos N, et al. Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. *Cell*. 1993;75(6):1215-1225. doi: 0092-8674(93)90330-S [pii].

78. Peltomaki P, Aaltonen LA, Sistonen P, et al. Genetic mapping of a locus predisposing to human colorectal cancer. *Science*. 1993;260(5109):810-812. doi: 10.1126/science.8484120 [doi].

79. Bronner CE, Baker SM, Morrison PT, et al. Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. *Nature*. 1994;368(6468):258-261. doi: 10.1038/368258a0 [doi].

80. Lindblom A, Tannergard P, Werelius B, Nordenskjold M. Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. *Nat Genet*. 1993;5(3):279-282. doi: 10.1038/ng1193-279 [doi].

81. Papadopoulos N, Nicolaides NC, Wei YF, et al. Mutation of a mutL homolog in hereditary colon cancer. *Science*. 1994;263(5153):1625-1629. doi: 10.1126/science.8128251 [doi].

82. Drummond JT, Li GM, Longley MJ, Modrich P. Isolation of an hMSH2-p160 heterodimer that restores DNA mismatch repair to tumor cells. *Science*. 1995;268(5219):1909-1912. doi: 10.1126/science.7604264 [doi].

83. Nicolaides NC, Papadopoulos N, Liu B, et al. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. *Nature*. 1994;371(6492):75-80. doi: 10.1038/371075a0 [doi].

84. Watanabe A, Ikejima M, Suzuki N, Shimada T. Genomic organization and expression of the human MSH3 gene. *Genomics*. 1996;31(3):311-318. doi: S0888-7543(96)90053-1 [pii].

85. Lipkin SM, Wang V, Jacoby R, et al. MLH3: A DNA mismatch repair gene associated with mammalian microsatellite instability. *Nat Genet*. 2000;24(1):27-35. doi: 10.1038/71643 [doi].

86. Miyaki M, Konishi M, Tanaka K, et al. Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nat Genet*. 1997;17(3):271-272. doi: 10.1038/ng1197-271 [doi].

87. Duraturo F, Liccardo R, Cavallo A, De Rosa M, Grosso M, Izzo P. Association of low-risk MSH3 and MSH2 variant alleles with lynch syndrome: Probability of synergistic effects. *Int J Cancer*. 2011;129(7):1643-1650. doi: 10.1002/ijc.25824 [doi].

88. Korhonen MK, Vuorenmaa E, Nystrom M. The first functional study of MLH3 mutations found in cancer patients. *Genes Chromosomes Cancer*. 2008;47(9):803-809. doi: 10.1002/gcc.20581 [doi].

89. Olkinuora A, Nieminen TT, Martensson E, et al. Biallelic germline nonsense variant of MLH3 underlies polyposis predisposition. *Genet Med*. 2019;21(8):1868-1873. doi: 10.1038/s41436-018-0405-x [doi].

90. Ligtenberg MJ, Kuiper RP, Geurts van Kessel A, Hoogerbrugge N. EPCAM deletion carriers constitute a unique subgroup of lynch syndrome patients. *Fam Cancer*. 2013;12(2):169-174. doi: 10.1007/s10689-012-9591-x [doi].

91. Celentano V, Luglio G, Antonelli G, Tarquini R, Bucci L. Prophylactic surgery in lynch syndrome. *Tech Coloproctol*. 2011;15(2):129-134. doi: 10.1007/s10151-010-0666-0 [doi].

92. Jarvinen HJ, Aarnio M, Mustonen H, et al. Controlled 15-year trial on screening for colorectal cancer in families with hereditary nonpolyposis colorectal cancer. *Gastroenterology*. 2000;118(5):829-834. doi: S0016508500404555 [pii].

93. 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(7):2300-2306. doi: 10.1053/j.gastro.2010.02.053 [doi].

94. Jass JR, Young J, Leggett BA. Evolution of colorectal cancer: Change of pace and change of direction. *J Gastroenterol Hepatol*. 2002;17(1):17-26. doi: 10.1046/j.1440-1746.2002.02635.x [doi].

95. Engel C, Vasen HF, Seppala T, et al. No difference in colorectal cancer incidence or stage at detection by colonoscopy among 3 countries with different lynch syndrome surveillance policies. *Gastroenterology*. 2018;155(5):1400-1409.e2. doi: S0016-5085(18)34818-2 [pii].

96. Seppala T, Pylvanainen K, Evans DG, et al. Colorectal cancer incidence in path_MLH1 carriers subjected to different follow-up protocols: A prospective lynch syndrome database report. *Hered Cancer Clin Pract*. 2017;15:18-017-0078-5. eCollection 2017. doi: 10.1186/s13053-017-0078-5 [doi].

97. Vasen HF, Mecklin JP, Khan PM, Lynch HT. The international collaborative group on hereditary non-polyposis colorectal cancer (ICG-HNPCC). *Dis Colon Rectum*. 1991;34(5):424-425. doi: 10.1007/bf02053699 [doi].

98. Vasen HF, Watson P, Mecklin JP, Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, lynch syndrome) proposed by the international collaborative group on HNPCC. *Gastroenterology*. 1999;116(6):1453-1456. doi: S0016508599005715 [pii].

99. 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(4):261-268. doi: 10.1093/jnci/djh034 [doi].

100. Pasche B, Pennison MJ, DeYoung B. Lynch syndrome testing: A missed opportunity in the era of precision medicine. *JAMA*. 2016;316(1):38-39. doi: 10.1001/jama.2016.8291 [doi].

101. Vasen HF, Moslein G, Alonso A, et al. Recommendations to improve identification of hereditary and familial colorectal cancer in europe. *Fam Cancer*. 2010;9(2):109-115. doi: 10.1007/s10689-009-9291-3 [doi].

102. Tafe LJ. Targeted next-generation sequencing for hereditary cancer syndromes: A focus on lynch syndrome and associated endometrial cancer. *J Mol Diagn*. 2015;17(5):472-482. doi: 10.1016/j.jmoldx.2015.06.001 [doi].

103. Buecher B, Cacheux W, Rouleau E, Dieumegard B, Mitry E, Lievre A. Role of microsatellite instability in the management of colorectal cancers. *Dig Liver Dis*. 2013;45(6):441-449. doi: 10.1016/j.dld.2012.10.006 [doi].

104. Palomaki GE, McClain MR, Melillo S, Hampel HL, Thibodeau SN. EGAPP supplementary evidence review: DNA testing strategies aimed at reducing morbidity and mortality from lynch syndrome. *Genet Med*. 2009;11(1):42-65. doi: 10.1097/GIM.0b013e31818fa2db [doi].

105. Coelho H, Jones-Hughes T, Snowsill T, et al. A systematic review of test accuracy studies evaluating molecular micro-satellite instability testing for the detection of individuals with lynch syndrome. *BMC Cancer*. 2017;17(1):836-017-3820-5. doi: 10.1186/s12885-017-3820-5 [doi].

106. Boland CR, Goel A. Microsatellite instability in colorectal cancer. *Gastroenterology*. 2010;138(6):2073-2087.e3. doi: 10.1053/j.gastro.2009.12.064 [doi].

107. Duggan BD, Felix JC, Muderspach LI, Tourgeman D, Zheng J, Shibata D. Microsatellite instability in sporadic endometrial carcinoma. *J Natl Cancer Inst*. 1994;86(16):1216-1221. doi: 10.1093/jnci/86.16.1216 [doi].

108. Saeki H, Hlaing MT, Horimoto Y, et al. Usefulness of immunohistochemistry for mismatch repair protein and microsatellite instability examination in adenocarcinoma and background endometrium of sporadic endometrial cancer cases. *J Obstet Gynaecol Res*. 2019. doi: 10.1111/jog.14061 [doi].

109. MacDonald ND, Salvesen HB, Ryan A, Iversen OE, Akslen LA, Jacobs IJ. Frequency and prognostic impact of microsatellite instability in a large population-based study of endometrial carcinomas. *Cancer Res*. 2000;60(6):1750-1752.

110. An JY, Kim H, Cheong JH, Hyung WJ, Kim H, Noh SH. Microsatellite instability in sporadic gastric cancer: Its prognostic role and guidance for 5-FU based chemotherapy after R0 resection. *Int J Cancer*. 2012;131(2):505-511. doi: 10.1002/ijc.26399 [doi].

111. Esteller M, Levine R, Baylin SB, Ellenson LH, Herman JG. MLH1 promoter hypermethylation is associated with the microsatellite instability phenotype in sporadic endometrial carcinomas. *Oncogene*. 1998;17(18):2413-2417. doi: 10.1038/sj.onc.1202178 [doi].

112. Hampel H. Genetic testing for hereditary colorectal cancer. *Surg Oncol Clin N Am*. 2009;18(4):687-703. doi: 10.1016/j.soc.2009.08.001 [doi].

113. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev*. 2002;16(1):6-21. doi: 10.1101/gad.947102 [doi].

114. Hitchins M, Williams R, Cheong K, et al. MLH1 germline epimutations as a factor in hereditary nonpolyposis colorectal cancer. *Gastroenterology*. 2005;129(5):1392-1399. doi: S0016-5085(05)01787-7 [pii].

115. Parsons MT, Buchanan DD, Thompson B, Young JP, Spurdle AB. 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(3):151-157. doi: 10.1136/jmedgenet-2011-100714 [doi].

116. Thiel A, Heinonen M, Kantonen J, et al. BRAF mutation in sporadic colorectal cancer and lynch syndrome. *Virchows Arch*. 2013;463(5):613-621. doi: 10.1007/s00428-013-1470-9 [doi].

117. Kheirelseid EA, Miller N, Chang KH, et al. Mismatch repair protein expression in colorectal cancer. *J Gastrointest Oncol*. 2013;4(4):397-408. doi: 10.3978/j.issn.2078-6891.2013.021 [doi].

118. Pearlman R, Markow M, Knight D, et al. Two-stain immunohistochemical screening for lynch syndrome in colorectal cancer may fail to detect mismatch repair deficiency. *Mod Pathol*. 2018;31(12):1891-1900. doi: 10.1038/s41379-018-0058-y [doi].

119. Shia J. Immunohistochemistry versus microsatellite instability testing for screening colorectal cancer patients at risk for hereditary nonpolyposis colorectal cancer syndrome. part I. the utility of immunohistochemistry. *J Mol Diagn*. 2008;10(4):293-300. doi: 10.2353/jmoldx.2008.080031 [doi].

120. Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. 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(1):35-41. doi: 10.1097/GIM.0b013e31818fa2ff [doi].

121. Heinen CD, Juel Rasmussen L. Determining the functional significance of mismatch repair gene missense variants using biochemical and cellular assays. *Hered Cancer Clin Pract*. 2012;10(1):9-4287-10-9. doi: 10.1186/1897-4287-10-9 [doi].

122. Niroula A, Vihinen M. Classification of amino acid substitutions in mismatch repair proteins using PON-MMR2. *Hum Mutat*. 2015;36(12):1128-1134. doi: 10.1002/humu.22900 [doi].

123. Spurdle AB, Couch FJ, Hogervorst FB, Radice P, Sinilnikova OM, IARC Unclassified Genetic Variants Working Group. Prediction and assessment of splicing alterations: Implications for clinical testing. *Hum Mutat*. 2008;29(11):1304-1313. doi: 10.1002/humu.20901 [doi].

124. Ali H, Olatubosun A, Vihinen M. Classification of mismatch repair gene missense variants with PON-MMR. *Hum Mutat*. 2012;33(4):642-650. doi: 10.1002/humu.22038 [doi].

125. Chao EC, Velasquez JL, Witherspoon MS, et al. Accurate classification of MLH1/MSH2 missense variants with multivariate analysis of protein polymorphisms-mismatch repair (MAPP-MMR). *Hum Mutat*. 2008;29(6):852-860. doi: 10.1002/humu.20735 [doi].

126. Stone EA, Sidow A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res*. 2005;15(7):978-986. doi: gr.3804205 [pii].

127. Kansikas M, Kariola R, Nystrom M. Verification of the three-step model in assessing the pathogenicity of mismatch repair gene variants. *Hum Mutat*. 2011;32(1):107-115. doi: 10.1002/humu.21409 [doi].

128. Shimodaira H, Filosi N, Shibata H, et al. Functional analysis of human MLH1 mutations in saccharomyces cerevisiae. *Nat Genet*. 1998;19(4):384-389. doi: 10.1038/1277 [doi].

129. Clark AB, Cook ME, Tran HT, Gordenin DA, Resnick MA, Kunkel TA. Functional analysis of human MutSalpha and MutSbeta complexes in yeast. *Nucleic Acids Res*. 1999;27(3):736-742. doi: gkc184 [pii].

130. Drotschmann K, Clark AB, Tran HT, Resnick MA, Gordenin DA, Kunkel TA. Mutator phenotypes of yeast strains heterozygous for mutations in the MSH2 gene. *Proc Natl Acad Sci U S A*. 1999;96(6):2970-2975. doi: 10.1073/pnas.96.6.2970 [doi].

131. Shcherbakova PV, Kunkel TA. Mutator phenotypes conferred by MLH1 overexpression and by heterozygosity for mlh1 mutations. *Mol Cell Biol*. 1999;19(4):3177-3183. doi: 10.1128/mcb.19.4.3177 [doi].

132. Holmes J,Jr, Clark S, Modrich P. Strand-specific mismatch correction in nuclear extracts of human and drosophila melanogaster cell lines. *Proc Natl Acad Sci U S A*. 1990;87(15):5837-5841. doi: 10.1073/pnas.87.15.5837 [doi].

133. Li GM, Modrich P. Restoration of mismatch repair to nuclear extracts of H6 colorectal tumor cells by a heterodimer of human MutL homologs. *Proc Natl Acad Sci U S A*. 1995;92(6):1950-1954. doi: 10.1073/pnas.92.6.1950 [doi].

134. Nystrom-Lahti M, Perrera C, Raschle M, et al. Functional analysis of MLH1 mutations linked to hereditary nonpolyposis colon cancer. *Genes Chromosomes Cancer*. 2002;33(2):160-167. doi: 10.1002/gcc.1225 [pii].

135. 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(3):247-253. doi: 10.1002/humu.21180 [doi].

136. Gammie AE, Erdeniz N, Beaver J, Devlin B, Nanji A, Rose MD. Functional characterization of pathogenic human MSH2 missense mutations in saccharomyces cerevisiae. *Genetics*. 2007;177(2):707-721. doi: genetics.107.071084 [pii].

137. Kondo E, Suzuki H, Horii A, Fukushige S. A yeast two-hybrid assay provides a simple way to evaluate the vast majority of hMLH1 germ-line mutations. *Cancer Res*. 2003;63(12):3302-3308.

138. Bruckner A, Polge C, Lentze N, Auerbach D, Schlattner U. Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci*. 2009;10(6):2763-2788. doi: 10.3390/ijms10062763 [doi].

139. Ou J, Niessen RC, Lutzen A, et al. Functional analysis helps to clarify the clinical importance of unclassified variants in DNA mismatch repair genes. *Hum Mutat*. 2007;28(11):1047-1054. doi: 10.1002/humu.20580 [doi].

140. Guerrette S, Acharya S, Fishel R. The interaction of the human MutL homologues in hereditary nonpolyposis colon cancer. *J Biol Chem*. 1999;274(10):6336-6341. doi: 10.1074/jbc.274.10.6336 [doi].

141. Raevaara TE, Korhonen MK, Lohi H, et al. Functional significance and clinical phenotype of nontruncating mismatch repair variants of MLH1. *Gastroenterology*. 2005;129(2):537-549. doi: S0016-5085(05)01103-0 [pii].

142. Kariola R, Raevaara TE, Lonnqvist KE, Nystrom-Lahti M. Functional analysis of MSH6 mutations linked to kindreds with putative hereditary non-polyposis colorectal cancer syndrome. *Hum Mol Genet*. 2002;11(11):1303-1310. doi: 10.1093/hmg/11.11.1303 [doi].

143. Ollila S, Dermadi Bebek D, Jiricny J, Nystrom M. Mechanisms of pathogenicity in human MSH2 missense mutants. *Hum Mutat*. 2008;29(11):1355-1363. doi: 10.1002/humu.20893 [doi].

144. Ooi A, Wong A, Esau L, Lemtiri-Chlieh F, Gehring C. A guide to transient expression of membrane proteins in HEK-293 cells for functional characterization. *Front Physiol*. 2016;7:300. doi: 10.3389/fphys.2016.00300 [doi].

145. 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(1):211-219. doi: S0016508502075790 [pii].

146. Andersen SD, Liberti SE, Lutzen A, et al. Functional characterization of MLH1 missense variants identified in lynch syndrome patients. *Hum Mutat*. 2012;33(12):1647-1655. doi: 10.1002/humu.22153 [doi].

147. Lutzen A, de Wind N, Georgijevic D, Nielsen FC, Rasmussen LJ. Functional analysis of HNPCC-related missense mutations in MSH2. *Mutat Res*. 2008;645(1-2):44-55. doi: 10.1016/j.mrfmmm.2008.08.015 [doi].

148. Pagenstecher C, Wehner M, Friedl W, et al. Aberrant splicing in MLH1 and MSH2 due to exonic and intronic variants. *Hum Genet*. 2006;119(1-2):9-22. doi: 10.1007/s00439-005-0107-8 [doi].

149. Thompson BA, Martins A, Spurdle AB. A review of mismatch repair gene transcripts: Issues for interpretation of mRNA splicing assays. *Clin Genet*. 2015;87(2):100-108. doi: 10.1111/cge.12450 [doi].

150. 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(4):327-345. doi: 10.1002/mgg3.145 [doi].

151. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31(13):3812-3814. doi: 10.1093/nar/gkg509 [doi].

152. Drost M, Zonneveld JB, van Hees S, Rasmussen LJ, Hofstra RM, de Wind N. A rapid and cell-free assay to test the activity of lynch syndrome-associated MSH2 and MSH6 missense variants. *Hum Mutat*. 2012;33(3):488-494. doi: 10.1002/humu.22000 [doi].

153. Ollila S, Dermadi Bebek D, Greenblatt M, Nystrom M. Uncertain pathogenicity of MSH2 variants N127S and G322D challenges their classification. *Int J Cancer*. 2008;123(3):720-724. doi: 10.1002/ijc.23573 [doi].

154. Jiricny J. Postreplicative mismatch repair. *Cold Spring Harb Perspect Biol*. 2013;5(4):a012633. doi: 10.1101/cshperspect.a012633 [doi].

155. Liu D, Keijzers G, Rasmussen LJ. DNA mismatch repair and its many roles in eukaryotic cells. *Mutat Res*. 2017;773:174-187. doi: S1383-5742(17)30033-9 [pii].

156. Guarne A. The functions of MutL in mismatch repair: The power of multitasking. *Prog Mol Biol Transl Sci*. 2012;110:41-70. doi: 10.1016/B978-0-12-387665-2.00003-1 [doi].

157. Tomer G, Buermeyer AB, Nguyen MM, Liskay RM. Contribution of human mlh1 and pms2 ATPase activities to DNA mismatch repair. *J Biol Chem*. 2002;277(24):21801-21809. doi: 10.1074/jbc.M111342200 [doi].

158. Pluciennik A, Dzantiev L, Iyer RR, Constantin N, Kadyrov FA, Modrich P. PCNA function in the activation and strand direction of MutLalpha endonuclease in mismatch repair. *Proc Natl Acad Sci U S A*. 2010;107(37):16066-16071. doi: 10.1073/pnas.1010662107 [doi].

159. Zhuang Z, Ai Y. Processivity factor of DNA polymerase and its expanding role in normal and translesion DNA synthesis. *Biochim Biophys Acta*. 2010;1804(5):1081-1093. doi: 10.1016/j.bbapap.2009.06.018 [doi].

160. Genschel J, Bazemore LR, Modrich P. Human exonuclease I is required for 5' and 3' mismatch repair. *J Biol Chem*. 2002;277(15):13302-13311. doi: 10.1074/jbc.M111854200 [doi].

161. Krasikova YS, Rechkunova NI, Lavrik OI. Replication protein A as a major eukaryotic single-stranded DNA-binding protein and its role in DNA repair. *Mol Biol (Mosk)*. 2016;50(5):735-750. doi: 10.7868/S0026898416030083 [doi].

162. Hopfner KP, Tainer JA. Rad50/SMC proteins and ABC transporters: Unifying concepts from high-resolution structures. *Curr Opin Struct Biol*. 2003;13(2):249-255. doi: S0959440X0300037X [pii].

163. Warren JJ, Pohlhaus TJ, Changela A, Iyer RR, Modrich PL, Beese LS. Structure of the human MutSalpha DNA lesion recognition complex. *Mol Cell*. 2007;26(4):579-592. doi: S1097-2765(07)00254-7 [pii].

164. Heinen CD, Cyr JL, Cook C, et al. Human MSH2 (hMSH2) protein controls ATP processing by hMSH2-hMSH6. *J Biol Chem*. 2011;286(46):40287-40295. doi: 10.1074/jbc.M111.297523 [doi].

165. Mazur DJ, Mendillo ML, Kolodner RD. Inhibition of Msh6 ATPase activity by mispaired DNA induces a Msh2(ATP)-Msh6(ATP) state capable of hydrolysis-independent movement along DNA. *Mol Cell*. 2006;22(1):39-49. doi: S1097-2765(06)00112-2 [pii].

166. Lenhart JS, Pillon MC, Guarne A, Simmons LA. Trapping and visualizing intermediate steps in the mismatch repair pathway in vivo. *Mol Microbiol*. 2013;90(4):680-698. doi: 10.1111/mmi.12389 [doi].

167. Mendillo ML, Hargreaves VV, Jamison JW, et al. A conserved MutS homolog connector domain interface interacts with MutL homologs. *Proc Natl Acad Sci U S A*. 2009;106(52):22223-22228. doi: 10.1073/pnas.0912250106 [doi].

168. Guerrette S, Wilson T, Gradia S, Fishel R. Interactions of human hMSH2 with hMSH3 and hMSH2 with hMSH6: Examination of mutations found in hereditary nonpolyposis colorectal cancer. *Mol Cell Biol*. 1998;18(11):6616-6623. doi: 10.1128/mcb.18.11.6616 [doi].

169. Kondo E, Horii A, Fukushige S. The interacting domains of three MutL heterodimers in man: hMLH1 interacts with 36 homologous amino acid residues within hMLH3, hPMS1 and hPMS2. *Nucleic Acids Res*. 2001;29(8):1695-1702. doi: 10.1093/nar/29.8.1695 [doi].

170. Raschle M, Marra G, Nystrom-Lahti M, Schar P, Jiricny J. Identification of hMutLbeta, a heterodimer of hMLH1 and hPMS1. *J Biol Chem*. 1999;274(45):32368-32375. doi: 10.1074/jbc.274.45.32368 [doi].

171. Dutta R, Inouye M. GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem Sci*. 2000;25(1):24-28. doi: S0968-0004(99)01503-0 [pii].

172. De Vos M, Hayward BE, Picton S, Sheridan E, Bonthron DT. Novel PMS2 pseudogenes can conceal recessive mutations causing a distinctive childhood cancer syndrome. *Am J Hum Genet*. 2004;74(5):954-964. doi: 10.1086/420796 [doi].

173. Horii A, Han HJ, Sasaki S, Shimada M, Nakamura Y. Cloning, characterization and chromosomal assignment of the human genes homologous to yeast PMS1, a member of mismatch repair genes. *Biochem Biophys Res Commun*. 1994;204(3):1257-1264. doi: S0006-291X(84)72598-8 [pii].

174. Nicolaides NC, Carter KC, Shell BK, Papadopoulos N, Vogelstein B, Kinzler KW. Genomic organization of the human PMS2 gene family. *Genomics*. 1995;30(2):195-206. doi: S0888-7543(85)79885-0 [pii].

175. 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(5):578-587. doi: 10.1002/humu.21229 [doi].

176. Vaughn CP, Robles J, Swensen JJ, et al. Clinical analysis of PMS2: Mutation detection and avoidance of pseudogenes. *Hum Mutat*. 2010;31(5):588-593. doi: 10.1002/humu.21230 [doi].

177. Jiricny J, Nystrom-Lahti M. Mismatch repair defects in cancer. *Curr Opin Genet Dev*. 2000;10(2):157-161. doi: S0959-437X(00)00066-6 [pii].

178. Kadyrov FA, Dzantiev L, Constantin N, Modrich P. Endonucleolytic function of MutLalpha in human mismatch repair. *Cell*. 2006;126(2):297-308. doi: S0092-8674(06)00812-9 [pii].

179. Thodi G, Fostira F, Sandaltzopoulos R, et al. Screening of the DNA mismatch repair genes MLH1, MSH2 and MSH6 in a greek cohort of lynch syndrome suspected families. *BMC Cancer*. 2010;10:544-2407-10-544. doi: 10.1186/1471-2407-10-544 [doi].

180. Smith MJ, Urquhart JE, Harkness EF, et al. The contribution of whole gene deletions and large rearrangements to the mutation spectrum in inherited tumor predisposing syndromes. *Hum Mutat*. 2016;37(3):250-256. doi: 10.1002/humu.22938 [doi].

181. Charbonnier F, Olschwang S, Wang Q, et al. MSH2 in contrast to MLH1 and MSH6 is frequently inactivated by exonic and promoter rearrangements in hereditary nonpolyposis colorectal cancer. *Cancer Res*. 2002;62(3):848-853.

182. Vaughn CP, Baker CL, Samowitz WS, Swensen JJ. The frequency of previously undetectable deletions involving 3' exons of the PMS2 gene. *Genes Chromosomes Cancer*. 2013;52(1):107-112. doi: 10.1002/gcc.22011 [doi].

183. Nystrom-Lahti M, Kristo P, Nicolaides NC, et al. Founding mutations and alu-mediated recombination in hereditary colon cancer. *Nat Med*. 1995;1(11):1203-1206.

184. Wijnen J, van der Klift H, Vasen H, et al. MSH2 genomic deletions are a frequent cause of HNPCC. *Nat Genet*. 1998;20(4):326-328. doi: 10.1038/3795 [doi].

185. Ponti G, Castellsague E, Ruini C, Percesepe A, Tomasi A. Mismatch repair genes founder mutations and cancer susceptibility in lynch syndrome. *Clin Genet*. 2015;87(6):507-516. doi: 10.1111/cge.12529 [doi].

186. Castellvi-Bel S, Castells A, Strunk M, et al. Genomic rearrangements in MSH2 and MLH1 are rare mutational events in spanish patients with hereditary nonpolyposis colorectal cancer. *Cancer Lett*. 2005;225(1):93-98. doi: S0304-3835(05)00105-9 [pii].

187. Zavodna K, Krivulcik T, Bujalkova MG, et al. Partial loss of heterozygosity events at the mutated gene in tumors from MLH1/MSH2 large genomic rearrangement carriers. *BMC Cancer*. 2009;9:405-2407-9-405. doi: 10.1186/1471-2407-9-405 [doi].

188. Nystrom-Lahti M, Wu Y, Moisio AL, et al. DNA mismatch repair gene mutations in 55 kindreds with verified or putative hereditary non-polyposis colorectal cancer. *Hum Mol Genet*. 1996;5(6):763-769. doi: 6d0002 [pii].

189. Moisio AL, Sistonen P, Weissenbach J, de la Chapelle A, Peltomaki P. Age and origin of two common MLH1 mutations predisposing to hereditary colon cancer. *Am J Hum Genet*. 1996;59(6):1243-1251.

190. Froggatt NJ, Joyce JA, Davies R, et al. A frequent hMSH2 mutation in hereditary non-polyposis colon cancer syndrome. *Lancet*. 1995;345(8951):727-6736(95)90900-1. doi: S0140-6736(95)90900-1 [pii].

191. Froggatt NJ, Green J, Brassett C, et al. A common MSH2 mutation in english and north american HNPCC families: Origin, phenotypic expression, and sex specific differences in colorectal cancer. *J Med Genet*. 1999;36(2):97-102.

192. Pinheiro M, Pinto C, Peixoto A, et al. The MSH2 c.388_389del mutation shows a founder effect in portuguese lynch syndrome families. *Clin Genet*. 2013;84(3):244-250. doi: 10.1111/cge.12062 [doi].

193. Clendenning M, Senter L, Hampel H, et al. A frame-shift mutation of PMS2 is a widespread cause of lynch syndrome. *J Med Genet*. 2008;45(6):340-345. doi: 10.1136/jmg.2007.056150 [doi].

194. Cederquist K, Emanuelsson M, Goransson I, et al. Mutation analysis of the MLH1, MSH2 and MSH6 genes in patients with double primary cancers of the colorectum and the endometrium: A population-based study in northern sweden. *Int J Cancer*. 2004;109(3):370-376. doi: 10.1002/ijc.11718 [doi].

195. Ramsoekh D, Wagner A, van Leerdam ME, et al. A high incidence of MSH6 mutations in amsterdam criteria II-negative families tested in a diagnostic setting. *Gut*. 2008;57(11):1539-1544. doi: 10.1136/gut.2008.156695 [doi].

196. Hitchins MP, Ward RL. Constitutional (germline) MLH1 epimutation as an aetiological mechanism for hereditary non-polyposis colorectal cancer. *J Med Genet*. 2009;46(12):793-802. doi: 10.1136/jmg.2009.068122 [doi].

197. 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(1):112-117. doi: 10.1038/ng.283 [doi].

198. Hitchins MP. Constitutional epimutation as a mechanism for cancer causality and heritability? *Nat Rev Cancer*. 2015;15(10):625-634. doi: 10.1038/nrc4001 [doi].

199. 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(2):200-213. doi: 10.1016/j.ccr.2011.07.003 [doi].

200. Morak M, Koehler U, Schackert HK, et al. Biallelic MLH1 SNP cDNA expression or constitutional promoter methylation can hide genomic rearrangements causing lynch syndrome. *J Med Genet*. 2011;48(8):513-519. doi: 10.1136/jmedgenet-2011-100050 [doi].

201. Hitchins MP, Wong JJ, Suthers G, et al. Inheritance of a cancer-associated MLH1 germ-line epimutation. *N Engl J Med*. 2007;356(7):697-705. doi: 356/7/697 [pii].

202. Kovacs ME, Papp J, Szentirmay Z, Otto S, Olah E. Deletions removing the last exon of TACSTD1 constitute a distinct class of mutations predisposing to lynch syndrome. *Hum Mutat*. 2009;30(2):197-203. doi: 10.1002/humu.20942 [doi].

203. Kempers MJ, Kuiper RP, Ockeloen CW, et al. Risk of colorectal and endometrial cancers in EPCAM deletion-positive lynch syndrome: A cohort study. *Lancet Oncol*. 2011;12(1):49-55. doi: 10.1016/S1470-2045(10)70265-5 [doi].

204. Lynch HT, Riegert-Johnson DL, Snyder C, et al. Lynch syndrome-associated extracolonic tumors are rare in two extended families with the same EPCAM deletion. *Am J Gastroenterol*. 2011;106(10):1829-1836. doi: 10.1038/ajg.2011.203 [doi].

205. 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(3):604-13.e20. doi: 10.1053/j.gastro.2015.05.006 [doi].

206. Hampel H, Frankel WL, Martin E, et al. Screening for the lynch syndrome (hereditary nonpolyposis colorectal cancer). *N Engl J Med*. 2005;352(18):1851-1860. doi: 352/18/1851 [pii].

207. Legolvan MP, Taliano RJ, Resnick MB. Application of molecular techniques in the diagnosis, prognosis and management of patients with colorectal cancer: A practical approach. *Hum Pathol*. 2012;43(8):1157-1168. doi: 10.1016/j.humpath.2012.03.003 [doi].

208. Plon SE, Eccles DM, Easton D, et al. Sequence variant classification and reporting: Recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat*. 2008;29(11):1282-1291. doi: 10.1002/humu.20880 [doi].

209. Takahashi M, Shimodaira H, Andreutti-Zaugg C, Iggo R, Kolodner RD, Ishioka C. Functional analysis of human MLH1 variants using yeast and in vitro mismatch repair assays. *Cancer Res*. 2007;67(10):4595-4604. doi: 67/10/4595 [pii].

210. Xie J, Guillemette S, Peng M, Gilbert C, Buermeyer A, Cantor SB. An MLH1 mutation links BACH1/FANCJ to colon cancer, signaling, and insight toward directed therapy. *Cancer Prev Res (Phila)*. 2010;3(11):1409-1416. doi: 10.1158/1940-6207.CAPR-10-0118 [doi].

211. Hardt K, Heick SB, Betz B, et al. Missense variants in hMLH1 identified in patients from the german HNPCC consortium and functional studies. *Fam Cancer*. 2011;10(2):273-284. doi: 10.1007/s10689-011-9431-4 [doi].

212. Hinrichsen I, Brieger A, Trojan J, Zeuzem S, Nilbert M, Plotz G. Expression defect size among unclassified MLH1 variants determines pathogenicity in lynch syndrome diagnosis. *Clin Cancer Res*. 2013;19(9):2432-2441. doi: 10.1158/1078-0432.CCR-12-3299 [doi].

213. Christensen LL, Kariola R, Korhonen MK, et al. Functional characterization of rare missense mutations in MLH1 and MSH2 identified in danish colorectal cancer patients. *Fam Cancer*. 2009;8(4):489-500. doi: 10.1007/s10689-009-9274-4 [doi].

214. Ouellette MM, McDaniel LD, Wright WE, Shay JW, Schultz RA. The establishment of telomerase-immortalized cell lines representing human chromosome instability syndromes. *Hum Mol Genet*. 2000;9(3):403-411. doi: ddd049 [pii].

215. McDaid JR, Loughery J, Dunne P, et al. MLH1 mediates PARP-dependent cell death in response to the methylating agent N-methyl-N-nitrosourea. *Br J Cancer*. 2009;101(3):441-451. doi: 10.1038/sj.bjc.6605186 [doi].

216. Bustin SA. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol*. 2000;25(2):169-193. doi: JME00927 [pii].

217. Ollila S, Sarantaus L, Kariola R, et al. Pathogenicity of MSH2 missense mutations is typically associated with impaired repair capability of the mutated protein. *Gastroenterology*. 2006;131(5):1408-1417. doi: S0016-5085(06)01804-X [pii].

218. Kantelinen J, Kansikas M, Korhonen MK, et al. MutSbeta exceeds MutSalpha in dinucleotide loop repair. *Br J Cancer*. 2010;102(6):1068-1073. doi: 10.1038/sj.bjc.6605531 [doi].

219. Lahue RS, Au KG, Modrich P. DNA mismatch correction in a defined system. *Science*. 1989;245(4914):160-164. doi: 10.1126/science.2665076 [doi].

220. Hampel H, Frankel WL, Martin E, et al. Feasibility of screening for lynch syndrome among patients with colorectal cancer. *J Clin Oncol*. 2008;26(35):5783-5788. doi: 10.1200/JCO.2008.17.5950 [doi].

221. Burt RW. Diagnosing lynch syndrome: More light at the end of the tunnel. *Cancer Prev Res (Phila)*. 2012;5(4):507-510. doi: 10.1158/1940-6207.CAPR-12-0087 [doi].

222. Shia J, Holck S, Depetris G, Greenson JK, Klimstra DS. Lynch syndrome-associated neoplasms: A discussion on histopathology and immunohistochemistry. *Fam Cancer*. 2013;12(2):241-260. doi: 10.1007/s10689-013-9612-4 [doi].

223. Berends MJ, Wu Y, Sijmons RH, et al. Molecular and clinical characteristics of MSH6 variants: An analysis of 25 index carriers of a germline variant. *Am J Hum Genet*. 2002;70(1):26-37. doi: S0002-9297(07)61280-4 [pii].

224. Wu Y, Berends MJ, Mensink RG, et al. Association of hereditary nonpolyposis colorectal cancer-related tumors displaying low microsatellite instability with MSH6 germline mutations. *Am J Hum Genet*. 1999;65(5):1291-1298. doi: S0002-9297(07)62135-1 [pii].

225. Ward R, Meldrum C, Williams R, et al. Impact of microsatellite testing and mismatch repair protein expression on the clinical interpretation of genetic testing in hereditary non-polyposis colorectal cancer. *J Cancer Res Clin Oncol*. 2002;128(8):403-411. doi: 10.1007/s00432-002-0361-2 [doi].

226. Marra G, D'Atri S, Corti C, et al. Tolerance of human MSH2+/- lymphoblastoid cells to the methylating agent temozolomide. *Proc Natl Acad Sci U S A*. 2001;98(13):7164-7169. doi: 10.1073/pnas.121136498 [doi].

227. Cejka P, Stojic L, Mojas N, et al. Methylation-induced G(2)/M arrest requires a full complement of the mismatch repair protein hMLH1. *EMBO J*. 2003;22(9):2245-2254. doi: 10.1093/emboj/cdg216 [doi].

228. Hsieh P, Yamane K. DNA mismatch repair: Molecular mechanism, cancer, and ageing. *Mech Ageing Dev*. 2008;129(7-8):391-407. doi: 10.1016/j.mad.2008.02.012 [doi].