Chapter 1

General introduction
Lynch syndrome introduction

Lynch syndrome (LS) is an autosomal dominant disorder caused by inherited defects in DNA mismatch repair (MMR) genes. It is characterized by early onset colorectal cancer (CRC) and endometrial cancer (EC), with affected individuals having a ten times elevated risk compared to the general population. It furthermore causes an increased risk of cancers of the bladder, ovary, small bowel, biliary tract, brain, sebaceous glands, ureter and renal pelvis. Although the syndrome is dominant at the phenotypic level, as patients have inherited only one defective copy of a MMR gene, at the genotypic level it is recessive, as MMR functions are not compromised until the somatic loss of the corresponding wild type/normal allele. The first publication on Lynch syndrome was from Aldred Warthin who, in 1913, described the pedigrees of three families with a very strong predisposition to cancer. Many years later, Henry Lynch described two large families with hereditary colorectal cancer. In the early 1990s it was discovered that LS is caused by mutations in the MMR genes. Not only is LS one of the earliest hereditary cancer syndromes that was discovered, it is also one of the most prevalent, accounting for approximately 1-3% of all colorectal cancer cases. It is estimated that roughly one million people in Europe are carriers of a MMR gene defect.

In the past, often the term Hereditary Non Polyposis Colorectal Cancer (HNPCC) was used to refer to this syndrome. This term however is not accurate since the cancer predisposition caused by mutations in MMR genes is not limited to colorectal cancer. Furthermore, these mutations can also result in the occurrence of multiple polyps. Therefore, there is now a consensus to refer to this cancer syndrome as Lynch syndrome, once it has been confirmed that the underlying genetic defect is a mutation that affects the MMR system.

Whereas LS patients have inherited one defective copy of a MMR gene, occasionally bi-allelic mutation carriers are detected. Constitutive MMR Deficiency (CMMRD) often leads to very early onset of leukemia or even embryonic death.

DNA mismatch repair

The main function of the MMR system is the recognition and repair of replication errors. For successful repair, it needs to recognize the mismatching base(s), distinguish between the parental and newly synthesized strand and orchestrate the removal of the mismatch and the subsequent resynthesis of the removed sequence. MMR also plays a role in the prevention of homeologous recombination (i.e., recombination between slightly diverged homologous DNA sequences) and recognition of certain
chemically modified bases caused by, among others, chemotherapeutic agents. These processes will be discussed in more detail below.

**Players**

The core of the MMR system is comprised of protein dimers that are conserved from bacteria to human. The first dimer, homodimer MutS in *Escherichia coli* and heterodimers MutSα or MutSβ in *Homo sapiens* and other eukaryotes, initiates the MMR reaction by recognition of the mismatch. The second dimer, homodimer MutL in *E. coli* and heterodimer MutLα in *H. sapiens*, is subsequently recruited to further direct repair by excision and gap filling proteins. From here on onwards, unless otherwise specified, only the eukaryotic system is described.

In eukaryotes, three MutS homologues or MSH proteins involved in MMR exist: MSH2, MSH3 and MSH6. The MutSα dimer consists of MSH2 and MSH6 and MutSβ of MSH2 and MSH3. The mismatch recognition capacities of these dimers only partially overlap. MutSα recognizes and binds mismatches and small extrahelical loops of 1-2 bases whereas MutSβ recognizes loops of 1-5 unpaired bases. The MSH dimers scan the DNA for incorrectly paired bases by nonspecific interactions with the DNA helix. It is thought that a mismatch is recognized because it facilitates bending of the helix, which stabilizes the MSH-DNA interaction. The actual binding of the mismatch differs between the two dimers. MSH6, in MutSα, is the only protein that actually makes contact with the mismatched bases through its conserved Phe-X-Glu motif. The phenylalanine stacks against the mismatch thereby causing bending of the DNA with a 60° angle. MSH3, in MutSβ, lacks this motif but instead has several basic and polar residues in the N-terminus that are inserted into the minor groove of the helix, which enables it to bind extrahelical loops of varying size. MutSβ binding to extrahelical bases bends the DNA much more severely than MutSα mismatch binding, ranging from 90 to 120 degrees depending on the size of the insertion-deletion loop that is bound.

**Structure**

The three-dimensional structure of the MSH heterodimers revealed striking similarity with the bacterial MutS homodimers of *E. coli* and *Thermus aquaticus*. Several domains can be recognized (see Figure 1A): the mismatch binding domain (blue/purple) which contacts the mismatch, the connector domain, the clamp, the lever, the ATPase domain and the C-terminal dimerization domain. The connector domain (green) harbors the interaction region between the MutS and the MutL dimers. The clamp (orange) contains the DNA binding domain, which makes nonspecific contacts with the backbone of the DNA helix. The levers (yellow/sand) connect the DNA binding domains to the ATPase domain and signal conformational changes between
Figure 1. Three dimensional structure of MutSα (A) and MutSβ (B). (A) MutSα with MSH2 on the left and MSH6 on the right. (B) MutSβ with MSH2 again on the left and MSH3 on the right. The different colors correspond to the different domains of the heterodimers. Blue/purple: the mismatch binding domain; orange: the clamp; yellow/sand: levers; green/mint: connector domain; red/pink: ATPase domain; (light)brown: dimerization domain (MutSβ only). Images were based on the structures published by Warren et al. and Gupta et al. The domains shown here in different colors are based on what these authors indicated in their papers, hence the small differences between both structures.

them. The ATPase domain (red/pink) is the most conserved region and also the most extensively studied. All MutS dimers are ATPases containing two conserved composite ATP-binding domains consisting of a Walker A and B motif from one monomer and the ABC signature motif from the partner subunit. Although MSH2 and its binding partners, MSH3 and MSH6, have highly similar ATP binding domains, they bind ATP and ADP symmetrically with MSH6 and MSH3 contributing more to the total ATPase activity than MSH2. Based on single-molecule studies of fluorescently labelled MutS and yeast MutSα, it is thought that while MutSα scans the DNA, it travels along the helix in a corkscrew movement and ATP is bound and hydrolyzed by both subunits. When a mismatch is detected, MutSα pauses, the mismatch is bound by MSH6 and causes a conformational change with leads to an ADP for ATP exchange followed by inhibition of the ATP hydrolysis by both MSH2 and MSH6. The subsequent occupation of both ATPase domains by ATP causes another conformational change that disfavours stacking of the phenylalanine and the mismatch, resulting in the formation of a sliding clamp that can move freely along the helix. This latter
conformational change is also a requirement for binding of MutLα (reviewed in Jiricny [32]). Gupta et al. [19] suggest a similar mechanism for MutSβ. A recent study of the E. coli MutS/MutL crystal structure sheds more light on this process [33]. Their structure captures MutS in the sliding clamp conformation, which shows tilting of the MutS subunits across each other pushing the DNA into a new and larger channel. In this conformational state, MutS forms a loose ring around the DNA enabling it to diffuse freely on the DNA. The connector domain is now also in a new orientation which creates the binding interface for MutL.

Whereas the ATPase domain has been extensively studied, the structure of the C-terminus was only solved recently. In MutSα the extreme C-terminal regions of MSH2 and MSH6 did not adopt a conformation stable enough to be resolved by X-ray crystallography [18] and the C-terminal region even had to be removed from the E. coli MutS protein in order to obtain crystals needed to solve the structure [16]. In contrast, the human MutSβ structure solved by Gupta et al. [19] (Figure 1B) did contain the complete C-terminal domains of MSH2 and MSH3. Based on the earlier MutS and MutSα structures, the helical region C-terminal from the ATPase domain was called the helix-turn-helix (HTH) domain. Gupta et al. [19] noted that the newly solved region also consists of α-helices linked by short linkers, just as the HTH domain. In the case of MSH3, and based on secondary structure predictions also in MSH6, the last helix of the HTH domain and the first of the newly solved region form one long helix. Helices formed by C-terminal sequences from both subunits bind each other through hydrophobic interactions that are further stabilized by salt bridges. Therefore, the whole C-terminal region including the HTH domain is called the dimerization domain (DMD). The N-terminal part of the DMD also plays an important role in the communication between the two ATP-binding sites. It interacts directly with its “own” binding site, but also indirectly with the ATP-binding motif of the other subunit. As the movement of both DMD domains is coupled, binding of ATP by one subunit is channelled through the interacting DMD domains and also alters the binding site of the partner subunit [19].

Another region that is missing in the MutSα structure is the N-terminal region of MSH6. This domain harbors two important interaction domains. The first is the PWWP motif, which interacts with the histone mark H3K56me3 and as such recruits the dimer onto the chromatin [34]. This histone modification is most abundant during early S-phase and therefore enriches MutSα on replicating DNA [35,36]. The second motif, the PIP box located in the N-terminus of both MSH6 and MSH3, is also involved in the targeting of MutSα onto the DNA but this time through the binding of PCNA. This targets the MutS dimers to the replication factories [37,38] (Figure 2).
Figure 2. Schematic representation of eukaryotic mismatch repair. (1) MutSa is thought to be pre-loaded onto the chromatin through the interaction of its PWWP domain with the H3K56me3 histone mark. During replication, the interaction of the PIP box with PCNA targets the MutS dimer to the replication machinery. (2) MutSa binds to the mismatch, which causes and ADP for ATP exchange and subsequent recruitment of MutLa that introduces nicks into the nascent strand. (3) Exonuclease 1 removes the strand containing the mismatch and the resulting single-stranded DNA is temporarily coated by RPA. (4) The removed stretch of DNA is resynthesized by a replicative polymerase.
Mechanism

Once a mismatch is detected, discrimination between the nascent and template strand is essential to ensure genomic integrity, i.e., the erroneous strand and not the parental strand needs to be corrected. MutLα, which is recruited by MutSα or β following mismatch binding, is thought to play an important role in this process. The heterodimer consisting of MLH1 and PMS2 contains an N-terminal ATPase domain that is involved in the binding of MutSα/β. The C-terminal domain is important for dimerization and the PMS2 C-terminus furthermore harbors an endonuclease domain. This endonuclease domain plays a central part in strand discrimination. For the lagging strand it has been suggested that the gaps between Okazaki fragments can function as a discrimination signal. The leading strand however is replicated in a continuous manner. Small single-stranded gaps that occur after removal of erroneously incorporated ribonucleotides have been suggested to aid strand discrimination in the leading strand but they are unlikely to be a major factor. Instead, MutLα contributes to the discrimination process by introducing single stranded nicks in the nascent strand (Figure 2). These nicks are the entry point for Exonuclease 1 (EXO1), to date the only known exonuclease involved in MMR. The endonucleolytic activity of MutLα is activated by association with PCNA. PCNA is loaded at boundaries between double- and single-stranded DNA by RFC and it is always loaded with the same orientation respective to the direction of DNA synthesis. This way, the polarity of PCNA directs the endonuclease activity of PMS2 to the nascent strand. Although it has been shown that the efficiency of MMR is similar for the leading and lagging strand, the contribution of EXO1 to repair of the leading strand was found to be limited. This suggests that the repair process for the leading and lagging strand could be different. Kadyrov et al. found that a limited amount of MMR could still take place in the absence of EXO1, which is in accordance with the less severe phenotype of Exo1 knockout mice compared to Msh2 knockout mice. They found that the remaining MMR activity depended on the MutLα endonuclease activity and the presence of a DNA polymerase. They proposed a mechanism that involved strand displacement, and as such removal of the mismatch, by the polymerase that is loaded at the nicks created by MutLα.

Removal of the mismatch leads to a single-stranded stretch of DNA, which is protected against degradation by binding of RPA. The complementary strand is then resynthesized by a replicative polymerase that is loaded onto the DNA by PCNA and RFC (Figure 2).

Functions

Besides its primary function in postreplicative mismatch repair, the MMR system has two more functions that are important to maintain genomic...
integrity. The first is the suppression of homeologous recombination, which is recombination between homologous, but not identical (homeologous) DNA sequences. This suppression likely occurs through recognition and binding by MutSα of mismatches in heteroduplexes that arise upon strand exchange of the recombination partners. Although the extend at which the recombination is repressed increases with the amount of sequence divergence, it appears that less than 1% difference is already enough to elicit the MMR reaction that aborts the recombination 54. In both Msh2-/- and Msh6-/- cells, the recombination efficiencies of a 100% homologous and a 99.4% homologous targeting vector were highly similar 55,56. In MMR proficient or Msh3-/- cells, the targeting of the homeologous vector was strongly suppressed, which indicates that MutSα but not MutSβ is involved in this process 56. The mechanism by which MMR inhibits homeologous recombination is largely unknown. Recently, Honda et al. 57 found that MutSα is capable of binding mismatches in the context of a D-loop. The presence of RPA and Rad51 bound to displaced single stranded DNA did not abrogate mismatch recognition and furthermore ATP-bound sliding clamps can be formed on the heteroduplex region of the D-loop. However they also found that the residence time of the sliding clamp on the mismatch containing D-loop heteroduplex region was almost 3 times shorter than on mismatch containing double-stranded DNA. This suggests that MMR is able to function in the context of a D-loop as present during homologous recombination. But the shorter residence time points to heteroduplex rejection being less efficient than replication coupled MMR.

Another function of MMR is the mediation of toxicity of DNA damaging agents, in particular methylating agents such as N-methyl-N’-nitro-N-nitrosoguanidine (MNNG), N-methyl-N-nitrosourea (MNU) and chemotherapeutic agents as temozolomide, dacarbazine and procabazine. These agents methylate, amongst others, guanine residues at the O6 position, which, unless the methyl group is removed by O6-methylguanine DNA methyltransferase (MGMT), instructs DNA polymerases to incorporate thymidine instead of cytosine. A similar situation occurs when cells are exposed to 6-thioguanine (6-TG). The substitution of O6 for S6 in guanine residues in DNA leads to rapid methylation and preferential incorporation of thymidine instead of cytosine during replication. MMR proteins will bind the resulting O6-meG•T or S6-meG•T mismatches and remove the thymidines as these reside in the nascent strand. Since the lesion is still present, it will again give rise to a mismatch upon completion of the repair reaction, which MMR will again remove. Eventually, after futile cycles of MMR, the cell will go into the next S-phase with single strand gaps that lead to the formation of double strand breaks upon replication causing the cell to go into apoptosis 58,59. Thus, MMR
prevents the accumulation of mutations due to the mismatches that occur when the cell is exposed to methylating agents by promoting cell death. As a result, MMR-deficient cells are resistant to the toxicity but highly sensitive to the mutagenicity of methylating agents, whereas MMR-proficient cells are highly sensitive to cell killing by methylating agents\(^{55,60}\). MMR is also thought to recognize 5FdUTP:G mispairs that arise following exposure to 5-fluorouracil (5-FU), a core component of most chemotherapy regimens for CRC\(^{61-63}\). In vitro experiments have shown increased resistance to 5-FU in several MMR deficient (dMMR) cell lines\(^{62}\), which is consistent with a generally observed lack of benefit of 5-FU treatment for patients with dMMR CRC\(^{64,65}\). The precise mechanism behind this is yet unknown.

Next to these three functions, which all serve to protect genomic integrity, the MMR proteins are -seemingly contradictory- also involved in "mutagenic events" such as somatic hypermutation (SHM), class switch recombination (CSR) of immune-globulin genes and triplet repeat expansions. However, these processes fall beyond the scope of this thesis and will therefore not be discussed in detail.

**Consequences of MMR gene defects**

When MMR repair is deficient, the mutation rate increases 100-1000 fold, i.e., from \(10^{-10}\) till \(10^{-6}\)-\(10^{-7}\). Due to the partial redundancy of the system, the impact of the inactivation of each gene greatly differs.

**MMR deficient mice**

In mice, these differences are very pronounced. Both, \(Msh2\) and \(Msh6\) knockout mice have a clear predisposition to cancer but the cancer incidence and average lifespan differ. The majority of \(Msh2\) knockout mice develop T-cell lymphomas. A smaller subset, ±30%, develops intestinal adenoma and adenocarcinomas and all mice die within a year\(^{55,56}\) (Figure 3). \(Msh6\) knockout mice also have a strong cancer predisposition with a similar incidence and latency of lymphomagenesis as their \(Msh2\) counterparts. However some animals survive longer, up to 18 months\(^{56,66}\), which is due to a reduced incidence of intestinal tumors. This somewhat milder phenotype is due to the partial redundancy within the MMR system between MutS\(\alpha\) and MutS\(\beta\). This redundancy is especially clear in the phenotype of \(Msh3\) knockout mice whose survival curve overlaps with that of their wild-type littermates suggesting that MSH2/MSH6 activity alone is sufficient to suppress tumorigenesis\(^ {56}\). Only when combined with the loss of \(Msh6\), the effect of the \(Msh3\) defect becomes apparent as it accelerates intestinal tumorigenesis compared to \(Msh6\) only knockout mice\(^{56}\) (Figure 3).

The cancer predisposition and lifespan of \(Mlh1\) knockout mice is similar to that of \(Msh2^{-}\) mice. But in
contrast to the $\text{Msh}2^{-/-}$ mice these mice are infertile due to a defect in meiotic crossing over\textsuperscript{67,68}. $\text{Pms}2$ knockout mice are also infertile but only develop lymphomas later in life and no intestinal tumors \textsuperscript{69,70}.

On a functional and genetic level, the MMR systems of mice and humans are very similar. They involve the same players and for example murine $\text{MSH}2$ and $\text{MSH}6$ share 94\% and 96\% protein sequence similarity with their human counterparts, respectively. The close resemblance is further illustrated by the fact that the MMR defect in $\text{Mlh}1$ deficient mouse fibroblast can be complemented by human $\text{MLH}1$ cDNA\textsuperscript{71}.

**MMR defects in LS patients**

Just as the phenotype of the different MMR knockout mice differs, also in human LS patients the average age of onset and the risk of developing the different associated cancer types depend on the gene that is affected. The majority of LS cases is caused by inherited mutations in the MMR genes $\text{MLH}1$ and $\text{MSH}2$ (70-80\% of all LS-associated CRC cases). Mutations in the MMR genes $\text{MSH}6$ and $\text{PMS}2$ account for the remaining 20-30\% of LS-associated tumors (roughly equally distributed between both genes) \textsuperscript{72,73}.

Individuals, and males in particular, with $\text{MSH}2$ inactivating mutations have the highest risk of developing LS-associated cancers when the whole spectrum is taken into account \textsuperscript{74-77}. $\text{MLH}1$ mutation carriers have the highest risk of developing gastrointestinal cancers but develop few extracolonic cancers including endometrial carcinomas (EC) (Lindor (2014)\textsuperscript{4} and references therein).

In contrast to families carrying $\text{MLH}1$ and $\text{MSH}2$ mutations, families with $\text{MSH}6$ mutations often have a less severe and less obvious phenotype. Furthermore, the age of
onset is generally later, by approximately 10 years, and the risk for developing CRC is lower \(^{72,73}\), which is also seen in mouse models. There are reports of increased frequency of endometrial cancer in \( MSH6 \) mutation carriers versus \( MSH2 \) mutation carriers\(^{78}\), however, there are also several studies that contradict this \(^{72,73}\). The risk of developing CRC for \( PMS2 \) mutation carriers seems to be in the same, lower, range as that for \( MSH6 \) mutation carriers, while the risk of developing EC seems somewhat lower \(^{79,80}\).

Data on the prevalence of non-CRC and non-EC LS-associated cancer types in \( MSH6 \) and \( PMS2 \) mutation carriers is sparse \(^{81}\), but it seems that at least for \( MSH6 \), carriers mainly develop CRC and EC rather than other LS associated cancers \(^{76}\).

MMR is not only affected by mutations but can also be impaired by methylation of the \( MLH1 \) and \( MSH2 \) promoters. The methylation of the \( MSH2 \) promoter can be caused by deletion of the 3’ end of the neighbouring \( EPCAM \) gene \(^{82}\). \( MLH1 \) promoter methylation is not inherited in a mendelian fashion and should be excluded as the cause of the MMR defect prior to genetic testing. As \( MLH1 \) promoter methylation is often seen in tumors carrying the \( BRAF-V600E \) mutation, testing for this mutation is often added to the diagnostic tree for LS \(^{83-85}\).

### Diagnosing Lynch syndrome

Diagnosing LS is very important as it allows the identification of individuals at risk, to whom can be offered periodic screening, prophylactic surgery and life-style advice. Several clinical observations and guidelines are helpful in deciding whether MMR genes need to be sequenced in order to diagnose LS.

### Clinical characteristics of LS CRC

Colorectal tumors from LS patients have several clinical features that differ from sporadic CRC tumors. The first feature that was noted was the high percentage of colonic tumors located in the proximal colon \(^{86,87}\). Adenomas found in HNPCC/LS patients often show villous histology and high-grade dysplasia, which are features of increased risk of cancer\(^{10,88}\). Accelerated carcinogenesis is a hallmark of LS-associated colorectal cancer in which an adenoma develops into carcinoma within 2-3 years compared to 6-10 years in sporadic CRC\(^{88-91}\). This acceleration can be attributed to the loss of MMR, which causes rapid accumulation of additional mutations that promote tumor progression. Thus, the classical adenoma to carcinoma progression is less frequently seen in LS patients.

Another key feature that can be clearly linked to the loss of MMR is the high level of microsatellite instability (MSI) found in LS CRC tumors (reviewed in Boland \(^{92}\)). MSI is defined as frequent length alterations of repetitive sequences like \((A)_n\) or \((CA)_n\)
and is a hallmark of MMR deficiency. In the absence of MMR, slippage errors by DNA polymerases at repetitive sequences are no longer detected and repaired. In contrast, at the chromosomal level, the DNA content is nearly diploid \textsuperscript{93}. The MSI-high status is often accompanied by the presence of tumor-infiltrating lymphocytes and a Th1-associated cytokine-rich microenvironment \textsuperscript{94-96}. It has been suggested that the absence of MMR gives rise to aberrant proteins that activate the immune system \textsuperscript{97}. The presence of tumor-infiltrating lymphocytes may contribute to the survival advantage that has been found for patients with LS-associated CRC compared with sporadic CRC patients \textsuperscript{98-100}. A recent study found that this active immune micro-environment of MMR-deficient tumors may be counterbalanced by the strong expression of immune inhibitory signals such as mediated by the interaction between the immune checkpoint receptor PD-1 on activated T cells and its ligand PD-L1 on tumor cells \textsuperscript{101}. Most recently, it has been found that LS-associated tumors respond particularly well to the inhibition of this checkpoint by the anti-PD1 antibody pembrolizumab \textsuperscript{96}.

**Indications for MMR gene testing**

In 1991, the first set of clinical criteria was developed to identify families with what was then called HNPCC: the Amsterdam criteria \textsuperscript{102}. The set was created to standardize the inclusion criteria for patients in research settings. These criteria have been critical for the identification of HNPCC, the discovery of MSI in tumors and the identification of the underlying genetic defects in MMR genes \textsuperscript{103-105}. As the criteria were designed for specificity rather than sensitivity, they were much too strict for use in a clinical setting \textsuperscript{81}. Over the years the criteria have been adapted to implement new findings. Currently the revised Bethesda criteria \textsuperscript{106} are the most commonly used to select patients with CRC of whom the tumor should be further molecularly analyzed, with fulfillment of one criteria being sufficient for further analysis. These criteria are:

- CRC diagnosed in a patient before the age of 50.
- Presence of synchronous or metachronous CRC or other LS-associated tumors.
- CRC with MSI-high histology diagnosed in a patient younger than 60 years.
- CRC or LS-associated tumor diagnosed before the age of 50 in at least one first-degree relative.
- CRC or LS-associated tumor diagnosed at any age in at least two first- or second-degree relatives.

Unfortunately, even these revised criteria are said to be too complex and to lack specificity (77\%) and sensitivity (82\%)\textsuperscript{10,76}. Therefore, a group of European experts, called the Mallorca group came with the following recommendation: “The Mallorca group recommends investigation of all CRC
(or individuals with CRC<70 years) by immunohistochemistry (IHC) of the four MMR proteins or MSI (grade of recommendation C). These tests should be accompanied by methods that identify MLH1 promotor methylation. Investigation of all EC in individuals less than 70 years by IHC or MSI can be considered to improve identification\textsuperscript{76}. Also the Dutch guideline on hereditary colon cancer recommends that the possibility of LS is investigated by IHC in CRC and EC patients younger than 70 years of age (www.oncoline.nl). However, data in this thesis suggest that this is not sufficient as it may miss mutations that do not affect protein stability (normal IHC) but do show partial MMR defects (weak MSI).

Molecular diagnosis of LS

In the Netherlands, MSI analysis is usually the starting point of molecular diagnosis\textsuperscript{12}. It is indicative of MMR function since slippage errors at mono- and dinucleotide repeats remain unrepaired if MMR is inactive. MSI was generally tested using a set of five different mono and dinucleotide markers called the Bethesda panel\textsuperscript{106}. Tumors are classified as MSI-high when two or more markers show length alterations. No length alterations leads to an MSS classification and one altered marker is labelled MSI-low. Nowadays, the significance of the MSI-low status is unclear and therefore testing of additional markers is warranted.

The use of dinucleotide repeats however poses some problems, as they are highly polymorphic and also subject to a PCR artefact called stutter. This may cause misclassification of MSI-low tumors as MSI-high. On the other hand, the MSH6 deficient tumors mainly show instability at mononucleotide repeats instead of dinucleotide repeats and may therefore be misclassified as MSH-low or even MSS. Therefore these days often a set of five mononucleotide repeats is used\textsuperscript{107}.

MSI analysis can indicate a MMR defect but it gives no clear information on which gene is affected. IHC of the four main MMR proteins can immediately give insight into which gene is affected since the absence of nuclear staining of a specific MMR protein can indicate a genetic defect in the corresponding gene. Absence of MSH2 or MLH1 though coincides with absent staining of their binding partners MSH6 and PMS2, respectively, since the stability of these MMR proteins depends on the presence of their binding partner\textsuperscript{56}. Absence of MSH6 or PMS2 staining without absence of their binding partner directly pinpoints the affected gene. However it has to be noted that IHC may fail to reveal mutations that do not impact on protein stability, interaction or nuclear localization but do abrogate MMR activity. It is also possible that the absence of MSH6 or PMS2 staining is not due to mutations in these genes but to mutations in their binding partners MSH2 and MLH1 that do not directly impact the stability of MSH2 or MLH1 itself, but do affect the binding with MSH6 or PMS2.
After detection of a MMR defect

Once LS has been diagnosed, mutation carriers are offered a screening regimen. For colorectal cancer it is recommended to undergo regular colonoscopies with a 1-2 year interval starting between 20-25 years of age 10,12,76. In contrast to screening for CRC, which is proven to be effective 108, the value of screening for EC is still unknown. Nonetheless, surveillance of the endometrium starting from the age of 35-40 may lead to early detection and should be offered to mutation carriers 10,76. The lack of benefit from screening for EC is worrisome as the risk of developing EC for female patients equals or may even exceed the risk of developing CRC 109. Therefore both European and US experts recommend the option of prophylactic hysterectomy and bilateral oophorectomy is discussed with female carriers who feel they have completed their families 10,76.

When a mutation in one of the MMR genes is found, the type of action that is taken depends on the effect the mutation has or is thought to have on MMR function. Large deletions, frameshifts and truncating mutations abrogate protein function and are classified as pathogenic whereas synonymous mutations, unless located in possible splice donor or acceptor sites can generally be considered polymorphic. Besides overtly deleterious mutations, also missense mutations are frequently found affecting a single amino acid but leaving the remainder of the protein intact. Such variants are often detected in patients with only a weak suspicion for familial cancer (e.g., young age but no family history) and their number is rising by increased diagnostic sequencing. The latter is also fueled by the possible implications a MMR defect may have on the choice of therapeutic intervention. As long as the phenotypic consequences are unclear, the diagnosis LS cannot be made and carriers of such variants of uncertain clinical significance (VUS) need to be considered LS patients who are advised to undergo regular burdensome screening.

Variants of uncertain clinical significance (VUS)

A system of five classes was developed by the International Agency for Research on Cancer (IARC) working group for unclassified variants 110 that provides an estimation of the chance that a MMR gene variant is pathogenic or not. The classes in this system are linked to clinical recommendations with the advice for class 5 (pathogenic) and 4 (likely pathogenic) being clinical testing and full high-risk surveillance. Although not specifically mentioned, this should also apply to class 3 (uncertain) mutation carriers. The advice for class 1 (non pathogenic) and class 2 (likely non pathogenic) is to treat as if “no mutation associated with disease has been detected”. It also advocates acquiring additional data on mutations within class 2, 3 and 4 to improve classification.
Recently, the Variant Interpretation Committee from the International Society for Gastrointestinal Hereditary Tumours (InSiGHT) undertook a collaborative effort to develop a standardized classification scheme for constitutional variants found in MLH1, MSH2, MSH6 and PMS2. Using clinical data and data from functional assays, both published and unpublished, they were able to classify many variants listed in the InSiGHT database in classes 1, 2, 4 or 5. In spite of this massive effort, about a third of variants largely consisting of non-synonymous variants still had to be assigned to class 3. As class 3 mutation carriers cannot safely be discharged from regular testing, robust tests to determine the significance of class 3 variants are urgently needed.

**Clinical data**

Clinical variables are classically used to calculate likelihood ratios. These are based on tumor characteristics such as MSI, the IHC status, in trans co-occurrence with known deleterious mutations, loss of heterozygosity, and population/family based information such as co-segregation of the variant with disease and the allele frequency in unaffected controls. Unfortunately, as we also see in chapter 4 of this thesis, the MSI and IHC status can vary between tumors from mutation carriers from the same family. Furthermore, small family size and insufficient data preclude determination of co-segregation. Therefore, classification solely based on clinical data is often an unsecure and difficult undertaking.

**In silico and splicing analysis**

Another classification tool that has become very important over the past decade is *in silico* analysis. This makes use of algorithms that aim to predict the impact of a missense mutation mainly based on evolutionary conservation and physicochemical characteristics. They assess the severity of the amino acid substitution based on the physicochemical differences between the two amino acids, such as polarity, size etc. Using protein multiple sequence alignments they analyze the sequence conservation but also the range of variation that is evolutionarily tolerated at this position. Some algorithms also take structural features into account. Examples of such *in silico* programs used for the analysis of MMR variants are polyphen-2, SIFT, Align-GVGD, MAPP-MMR and PON-MMR(2), of which the latter two are specifically developed or modified for MMR. These programs all rely on the quality of the alignment used to make the prediction and on their calibration. However, in spite of all the effort put into the development and optimization of these programs, they do not have sufficient predictive value by themselves to base classification upon. Furthermore, they do not take possible splicing alterations caused by the variation into account. This is addressed by splice prediction programs and splicing assays. The first are also *in silico* analyses that predict...
disruption of splice sites and the creation of new splice sites due to the sequence variation. They make use of the current knowledge of splice donor and acceptor sites (reviewed by Spurdle [124]). Nucleotide variants that alter the consensus splice sites are currently often considered pathogenic without further analysis [111]. If a splice prediction program predicts an effect for an intronic or exonic variant outside known splice sites, splicing assays (often relying on RT-PCR) can be performed. Nowadays, many clinical laboratories routinely perform such assays [125,126].

**Functional assays**

As the mismatch repair proteins are evolutionarily conserved, yeast has been used as a model system to analyze the effects of MMR missense mutations [127,128], in part because gene targeting in yeast is relatively easy. The mismatch repair capacity of the mutated protein is generally studied using reporter assays with resistance markers that will only convey (or regain) resistance once they are mutated [129-131]. Although this system is fairly fast and straightforward, there are several caveats. First, it can only be used for amino acids that are conserved between the yeast and human proteins and even more preferably, are located in highly conserved regions instead of a single conserved amino acid in an (relatively) unconserved domain. This rules out a substantial amount of VUS. Secondly, inaccurately high expression levels of the variant protein may mask (partial) MMR defects. Also, post-translational modifications such as phosphorylation may differ between the human and yeast proteins. Mutations affecting these modifications, as well as mutations affecting other aspects of MMR regulation such as protein-protein interactions, may not be conserved across distantly-related species and therefore cannot be studied.

A large disadvantage of the yeast-based systems, the lack of conservation in significant regions of the proteins, is circumvented by the use of human cell lines. In the cell-based systems, MMR deficient cell lines are complemented with ectopically expressed variant proteins and subsequently studied for MMR capacity. The cell lines that are used are mainly cancer cell lines that lack expression of one MMR protein (e.g. HCT116 [131] or HEK293T [132] for MLH1 and LoVo or Hec59 for MSH2 [133]). While the ability to recreate every VUS found in (suspected) Lynch syndrome patients is a clear advantage over the yeast system, the disadvantage of overexpression is still present in this setup as the mutated protein is expressed ectopically. Another advantage over the yeast system is the wider array of MMR assays that can be conducted in this system, covering not only the ability to repair mismatches but for example also the response to methylating agents which is modulated by the MMR system. However, the use of transformed cancer cell lines creates a variable genetic background that may affect the MMR system.
Another system that uses human proteins is the so-called cell-free system. In these assays, nuclear extracts from cell lines missing a specific MMR protein are complemented with the purified variant of that protein \(^{134-137}\). Often these variant proteins are expressed in and purified from \(Sf9\) insect cells. Using this system, you don’t have the disadvantage of the unpredictable genetic background of the cancer cell lines. On the other hand, the number of assays that can be done to test MMR function is fairly limited. Generally, an \textit{in vitro} repair assay is done in which a plasmid with an inactive restriction site is added to the extract. The restriction site can only be restored when the mismatch that renders it inactive is repaired. Hence the amount of digested plasmid is a measure for MMR activity. These assays also pose the risk of overexpression masking a partial MMR defect. Furthermore, the production of the variant proteins requires laborious cloning procedures. To circumvent cloning procedures, Drost \textit{et al.} \(^{138,139}\) have developed an \textit{in vitro} cell-free system in which the variant genes are produced \textit{in vitro} using site-directed mutagenesis. Subsequently proteins are produced from these PCR products using a commercial \textit{in vitro} expression kit. This greatly reduces the time involved, however, as it is still a cell-free system, only the repair capacity of the variant protein can be tested, which leaves the other functions of MMR unaddressed.
**Thesis outline**

This thesis presents a novel method to test *Msh2* and *Msh6* VUS found in families with suspected Lynch syndrome. To obtain clear and reliable results we chose to develop a system that allowed us to introduce the variant allele at the endogenous locus in a convenient cell type, mouse embryonic stem cells (ESCs), which allowed subsequent analysis of critical MMR functions at the cellular level.

To substitute single codons at the endogenous MMR gene, we made use of a recently developed gene modification technique in mouse ESCs that allows single codon or even single nucleotide substitution at any desired location in the genome. The method uses synthetic oligodeoxynucleotides (ssODN) of ±35 residues that are complementary to an endogenous target sequence except for the centrally located nucleotide(s) that comprise the desired modification. Gene modification is believed to occur by annealing of the ssODN to its complementary chromosomal sequence when this is transiently exposed as single-stranded DNA during replication. Subsequently, the annealed oligonucleotide serves as a primer for DNA synthesis leading to its stable integration into the genome. However, the mismatches at the position of the mutating nucleotides elicit a MMR reaction that aborts the gene modification reaction and restricts gene modification to an efficiency of only $10^{-7}$. To circumvent this problem, ESCs can be made permissive for ‘oligo targeting’ by transient suppression of *MSH2* or *MLH1* through RNA interference. Upon transient knockdown of MSH2, mostly used in this thesis, residual MSH2 level (±10% of wild-type) is still sufficient to recognize single base-pair mismatches. However 3-4 adjacent mismatches are no longer recognized and therefore an entire codon can effectively be substituted. In case of MLH1 knockdown, single base-pair mismatches are no longer recognized, allowing effective single-nucleotide substitution. Both protocols allow subtle gene modification at a frequency of $10^{-5}$. Cells carrying the planned modification can be purified from the excess of non-modified cells by seeding ssODN-exposed ESCs in 96-wells plates at a density of 5000 cells/well. Wells containing cells carrying the planned modification can be identified by mutation-specific real-time PCR. Subsequent reseeding of positive wells at lower density leads to enrichment for modified cells and in 3-4 steps to a purified cell clone.

Using the oligo targeting technique we closely mimic the human situation by generating the missense mutation found in patients in the endogenous MMR gene, which ensures physiological expression levels. Subsequent inactivation of the wild-type allele yielded cells that exclusively expressed the mutant allele. To assess the MMR capacity of mutant cells, we performed a set of functional assays interrogating the main MMR functions. This way we could directly study the
effect of the mutation on the MMR system in an in vivo setting.

In chapter 2 we set up the system by creating four Msh2 missense mutations found in suspected Lynch syndrome patients. These mutations were selected based on the available literature and included two mutations that were strongly suspected to be pathogenic or polymorphic, respectively. Our results classified three mutations as polymorphic and one as pathogenic and the data from the functional assays corresponded with the clinical data from the carriers that was reported in the literature.

In chapter 3 we used a similar experimental setup to see if our method would also obtain reliable results for Msh6 missense mutations. Again, our results were in accordance with clinical data, which led to the conclusion that our setup also works for Msh6.

With the reliability of our approach now confirmed, we moved on and applied our method to a clinical setting. Through a collaboration with three Dutch centers for clinical genetics, we obtained requests for the analysis of three Msh2 missense mutations found in Dutch families with suspected Lynch syndrome. In all three cases, conventional classification had been insufficient to classify these VUS, which hampered counseling. In chapter 4 we present the clinical data along with the results from our analysis and again show large similarities between the two sets. Surprisingly, two VUS showed a clear but partial MMR defect while the third behaved as wild-type and was classified as a polymorphism. Even though the defects of the other two VUS were partial, they were significant which leads us to believe that these VUS are the underlying cause of the disease although they might be low penetrance alleles.

In chapter 5 we exploit one of the advantages of using mouse ESCs combined with oligo targeting by making a mutant mouse. We selected a VUS that caused a deletion of the last 60 amino acids of MSH2. For this VUS, we did not limit this study to a single cell line. Because of the underlying fundamental question regarding the role of the C-terminus of MSH2, we made additional cell lines in which we inactivated either Msh3 or Msh6. Our analysis showed that the interaction between the truncated MSH2 protein and its binding partners was affected but not completely abolished as there was some remaining MMR activity seen in the functional assays. The mutant mice showed a clear tumor predisposition and a shortened lifespan. In agreement with the functional data, however, the effect was less severe than a complete loss of Msh2. This led to the conclusion that the MSH2 C-terminal 60 amino acids are essential for interaction with MSH6 and MSH3 and effective DNA mismatch repair.
References


Fischer, F., Baerenfaller, K. & Jiricny, J. 5-Fluorouracil is efficiently removed from DNA by the base
Umar, A.
Hewish, M., Lord, C. J., Martin, S. A., Cunningham, D. & Ashworth, A. Mismatch repair deficient
Baker, S. M.
Honda, M.
Meyers, M.
Mojas, N., Lopes, M. & Jiricny, J. Mismatch repair-dependent processing of methylation damage


116 Tavtigian, S. V., Greenblatt, M. S., Lesueur, F., Byrnes, G. B. & Group, I. U. G. V. W. In silico analysis of missense substitutions using sequence-alignment based methods. *Hum Mutat* 29, 1327-1336,


