Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Clinical and Molecular Medicine

Mariana Cecilia Magdalena Meza Vázquez

Evaluation of Variants of Uncertain Significance in *MLH1* and *MSH6* genes of the Mismatch Repair System

Master's thesis in Molecular Medicine
Supervisor: Professor Wenche Sjursen
Co-supervisor: Professor Frank Skorpen
June 2021

NTNU
Norwegian University of
Science and Technology

Mariana Cecilia Magdalena Meza Vázquez

Evaluation of Variants of Uncertain Significance in *MLH1* and *MSH6* genes of the Mismatch Repair System

Master's thesis in Molecular Medicine
Supervisor: Professor Wenche Sjursen
Co-supervisor: Professor Frank Skorpen
June 2021

Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Clinical and Molecular Medicine

NTNU
Norwegian University of
Science and Technology

Evaluation of Variants of Uncertain Significance in *MLH1* and *MSH6* genes of the Mismatch Repair System

Mariana Cecilia Magdalena Meza Vázquez

Thesis for the degree of Master of Science in Molecular Medicine
Supervisor Wenche Sjursen
Co-supervisor Frank Skorpen
Trondheim, June 2021

Norwegian University of Science and Technology
Faculty of Medicine and Health Sciences
Department of Clinical and Molecular Medicine

Table of contents

Abstract

Colorectal cancer (CRC) is one of the most common types of cancer both worldwide and in Norway. High-penetrance mutations lead to predisposition to colorectal cancer, comprising up to 5% of all colorectal cancer cases, e.g., defects in the DNA mismatch (MMR) genes causing Lynch syndrome. Variants in the MMR genes are classified into five classes, whereof the class three variants are named Variants of Uncertain Clinical Significance (VUS). VUSes cannot be used to diagnose a patient with Lynch syndrome, since it is uncertain whether they affect the proper function of the MMR system. . Databases and prediction tools as well as the classification systems that have been developed over time have been very useful to provide a better understanding of the possible effects that VUS variants could cause.

To verify and give a conclusive answer about the pathogenicity of MMR VUS variants, different functional assays have been developed. These assays include cell-free system assays, functional assays using human cell lines and yeast function assays.
The purpose for this master's thesis was to classify VUS Variants in *MLH1* and *MSH6* genes of the MMR system that were found at St. Olav's Hospital. Further aim was to make an overview of different types of functional assays from existing literature that has been used to evaluate the effect of MMR variants, and to eventually select one or two of these functional assays to establish them in the laboratory.

The results from the VUS Variant classification identified 12 variants with a high probability of pathogenicity. Within these pathogenic variants 6 were identified in the *MLH1* gene and 6 in *MSH6* gene. It was concluded that these possible pathogenic variants could affect the MMR repair capacity as a whole process complex. Therefore, functional assays that evaluate the MMR capacity were reviewed and prioritized. From these assays it was discussed which ones could be the best assays to evaluate the VUS variants. It was concluded that the best assay to evaluate the most probably pathogenic variants found was the CIMRA method.

Abbreviations

- **ACGS** Association for Clinical Genomic Science
- **ACMG** The American College of Medical Genetics and Genomics
- **CanViG-UK** Cancer Variant Interpretation Group UK
- **cDNAs** Complementary DNAs
- **CIMP** CpG Island methylator phenotype
- **CIMRA** Cell free in Vitro MMR activity
- **CIN** Chromosomal Instability
- **CRC** Colorectal Cancer
- **EPCAM** Epithelial Cell Adhesion Molecule
- **Exo1** Exonuclease 1
- **FAP** Familial Adenomatous Polyposis
- **gnomAD** The Genome Aggregation Database
- **HNPCC** Hereditary non-polyposis colorectal cancer
- **IARC** International Agency for Research on Cancer
- **InSiGHT** International Society for Gastrointestinal Hereditary Tumors
- **JPS** Juvenile polyposis syndrome
- **LOF** Loss of Function
- **LOH** Loss of Heterozygosity
- **LS** Lynch syndrome
- **MLH1** Mut L homologue 1
- **MMR** DNA Mismatch Repair
- **MNNG** Methylating agent N-Methyl-N-Nitro-N-Nitro-soguanidine
- **MSH2** MutS homologue 2
- **MSH6** MutS homologue 6
- **MSI** Microsatellite instability
- **MT** Methylation Tolerance-Based Functional Assays
- **NGS** Next Generation Sequencing
- **O^6MeG** O^6 – methylguanine
- **PCNA** Proliferating cell nucler antigen
- **PCR** Polymerase Chain Reaction
- **PJJ** Peutz-Jeghers syndrome
- **PMS2** Post meiotic segreagation increased 2
- **REF** RNA molecules that regulate gene expression
- **RFC** Replication Factor C
- **SDS** Sodium dodecylsulfate
- **SDS-PAGE** Sodium Dodecyl Sulfate polyacrylamide gel electrophoresis
- **Sf9** Spocloptera frugiperda
- **VUS** Variants of Uncertain Significance
- **WGS** Whole Genome Sequencing
- **WT** Wild -Type
- **6-TG** 6 - thioguanine

1 INTRODUCTION

1.1 Cancer development

Cancer can be defined as a disease in which cells in the human body start to grow uncontrollably and as well acquire properties to invade nearby tissues (1). Cancer is a genetic disease, because it is caused by mutations in genes that control cell growth, DNA repair, or otherwise are involved in the maintenance of cell integrity (1). Cancer development may also be associated with alterations in gene expression due to epigenetic changes, such as altered DNA methylation, histone modifications, as well as altered expression of various RNA molecules that regulate gene expression. The accumulation of genetic and epigenetic abnormalities leads to carcinogenesis. Thus, cancer cells develop the following characteristics: i) sustained proliferative signaling, ii) evasion of growth suppressors, iii) resistance to cell death, iv) replicative immortality, v) induction of angiogenesis, vi) invasion and metastasis (2).

Cancer can develop in any type of cell of the human body. Cancer cells can form masses of tissue or tumors. The majority of these tumors are considered malignant to humans because of the damage that they can produce to the human body. (1). The genetic changes in cancer often involve two type of genes: i) proto-oncogenes (oncogenes), and ii) tumor suppressor genes.

Proto oncogenes are genes which function is related to growth and division of the cell. Activating mutations in these genes will stimulate cell proliferation, and cells may start to grow and divide in an uncontrollable manner. Most of the activating mutations that occur in these genes are considered dominant gain of function mutations, meaning that mutation in only one gene copy is sufficient to stimulate cell growth. (1), (3). Tumor suppressor genes are genes which function is to prevent uncontrolled cell growth and to stimulate DNA repair if DNA is damaged. The loss of function of these genes (both alleles) caused by mutations or epigenetic changes may lead to increased mutational load and increased cell growth (1), (3).

1.2 Colorectal cancer

Colorectal cancer (CRC) accounts for approximately 10% of all annually diagnosed cancers and cancer-related deaths worldwide and it's the world's fourth most deadly cancer with around 900 000 deaths annually. CRC is the second most common cancer in women and the third most common in men (4). CRC commonly emerges from the glandular epithelial cells of the large

intestine. The main function of the colon is the reabsorption of water and minerals in the intestine.

Hereditary and environmental risk factors play an important role in the development of colorectal cancer. 10 – 20% of people with colorectal cancer presents family history for the development of this type of cancer (4). Environmental factors known to increase the risk of colorectal cancer include smoking, excessive alcohol, and increased bodyweight (4).

The cells of which CRC is derived are assumed to be stem cells. Stem cells are undifferentiated cells that are commonly present in the embryonic, fetal, and adult stages of life. Stem cells give rise to the different types of differentiated cells in our human body, and from these differentiated cells arise the development of the different tissues and organs (5). Stem cells are characterized by their ability to self-renewal or extensively proliferate, their clonality (usually arising from a single cell) and their potential to differentiate into different cell types (5). The accumulation of genetic and epigenetic alterations inactivates tumor-suppressor genes and activate oncogenes, leading to cancer stem cells (4). Most of CRC derive from a polyp (adenoma). The process starts when an aberrant crypt evolves into a precursor lesion (polyp) and eventually progressing to colorectal cancer (carcinoma) in 10-15 years (4) (Figure 1).

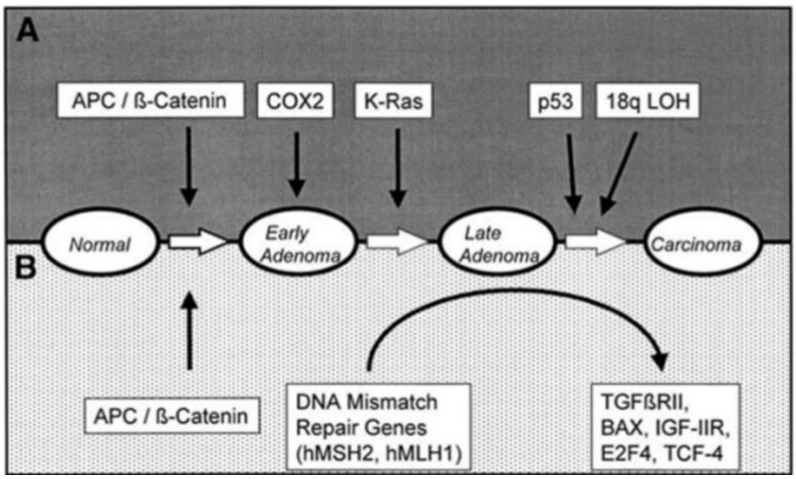


Figure 1. **Scheme representation of the key genetic events in colorectal tumorigenesis.** Representation of genes that are involved in the development of colorectal polyps and cancer. The genetic events are grouped into two categories A) The key genetic events associated with tumors exhibiting chromosomal instability and stability of microsatellite DNA B) The key Genetic events that represent tumors with Microsatellite instability (MSI). Gastroenterology 2000 (6) DOI: **doi.org/10.1053/gast.2000.16507**

Genomic instability is considered an important feature in the development of colorectal cancer (7) The principal pathogenic mechanisms associated with the development of CRC are chromosomal instability (CIN), microsatellite instability (MSI) and CpG island methylator phenotype (CIMP) (7).

CIN is considered the classical pathway of CRC. This pathway is characterized by the imbalances in the number of chromosomes, leading to tumor cells with aneuploidy and loss of heterozygosity (LOH) (7). The mechanism of the CIN pathway involves alterations in chromosome segregation, telomere dysfunction and DNA damage response which affect critical genes involved in the maintenance of the cell function such as *APC, KRAS, PI3K and TP53* (7) and figure 1).

The Microsatellite instability (MSI) phenotype is associated with loss of DNA repair mechanisms. During the synthesis of DNA, mutations occur in microsatellites due to the misalignment of repetitive sequences leading to the elongation or contraction of the microsatellite. This change in the length of microsatellite nucleotide repeats is referred to as microsatellite instability (MSI) (8). The ability to repair short DNA changes is decreased in tumors with microsatellite instability; therefore, mutations tend to accumulate in those regions.

The loss of expression as well as the malfunction of the mismatch repair genes (MMR) can be caused by spontaneous events (promoter hypermethylation and spontaneous mutations) or germinal mutations such as those found in Lynch syndrome, the most common hereditary syndrome causing CRC. (7). Epigenetic instability, which is responsible for the CpG island methylator phenotype (CIMP) is another feature in CRC. The main characteristic of CIMP tumors is the hypermethylation of tumor suppressor gene promoters, which leads to genetic silencing and the loss of protein expression (7).

There are two major precursor pathways in the development of sporadic CRC i) the adenoma-carcinoma pathway (the chromosomal instability pathway), accounting for 70-90% of colorectal cancer, and ii) the serrated neoplasia pathway accounting for 10-20% of colorectal cancers. Chromosomal instability often starts with an *APC* mutation, followed by the activations of the gene *RAS* and continue with the loss of function of *TP53* gene. In the serrated neoplasia pathway mutations in the *RAS* and *RAF* genes appear, as well as epigenetic instability that is characterized with CpG methylation, leading to microsatellite stable and instable cancers. (4) (Figure 2).

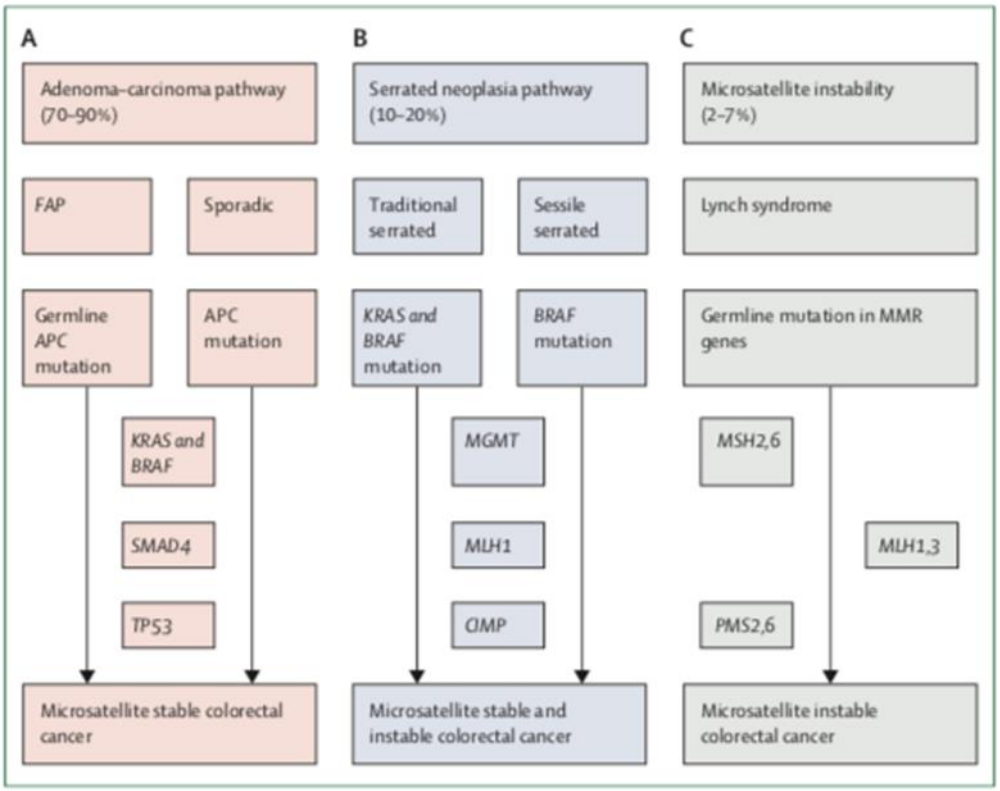


Figure 2. Colorectal cancer development pathways. A). Adenomas progress by the accumulation of genetic mutations and chromosomal instability causing microsatellite stable tumors. B) The serrated neoplasia pathway is initiated by genetic mutation of *BRAF* or *KRAS* genes but then progresses by methylation of tumor suppressing genes (CpG island methylator phenotype). C) Microsatellite instability is the result of defective DNA repair through inactivation of mismatch repair genes and is epitomized by the germline mutation of mismatch repair genes that is also seen in Lynch syndrome. Lancet Review 2019 (4) DOI: 10.1016/S0140-6736(19)32319-0.

1.2.1 Hereditary Colorectal Cancer

Hereditary cancers are derived from germline mutations. These mutations enter into the zygote via the sperm or the egg cell, and therefore, they will be present in every cell of the human body of the person that has it (9). Persons who already have a germline mutation are more susceptible to develop a cancer syndrome and have a higher risk of developing tumors. Cancer arises when further somatic mutations (including the wild type allele of the gene with a germline mutation) occur in genes of body cells through the mitosis cycle. The majority of hereditary cancer syndromes pursue an autosomal dominant inheritance in which first degree relatives have a 50% risk of carrying the mutation (9).

Hereditary colorectal cancer syndromes can be subdivided into non-polyposis (Lynch syndrome and familial colorectal cancer) and polyposis syndromes (9). Lynch syndrome have a feature of Microsatellite Instability (MSI) (8). The polyposis syndromes display chromosomal instability (CIN), and their tumors are characterized by being more aggressive. The two principal representatives of inherited CRC are Lynch syndrome and familial adenomatous polyposis (FAP). Other types of inherited CRC are MutY homolog (MUTYH)-associated polyposis (MAP), Peutz-Jeghers syndrome (PJS), juvenile polyposis syndrome (JPS), and Cowden/PTEN hamartoma syndrome (10).

1.2.2 Lynch syndrome

Lynch syndrome (LS) was former named Hereditary non-polyposis colorectal cancer (HNPCC) it is diagnosed in 2%-3% of all Colorectal Cancer patients (4). Lynch syndrome is caused by germline mutations and loss of function in one DNA Mismatch Repair (MMR) gene as well as deletion mutations in the *EPCAM* gene. The MMR proteins associated with Lynch syndrome consists of mutS homologue 2 (MSH2), mutL homologue 1 (MLH1), mutS homologue 6 (MSH6), postmeiotic segregation increased 2 (PMS2).

Lynch syndrome is transmitted in an autosomal dominant manner (11),(2). As the mutation is usually inherited from one parent, every cell carries a defective copy of one of four genes that are involved in the MMR system (either in *MLH1, MSH2, MSH6 or PMS2)* and a functional copy gene that maintain the function of DNA repair in cells. A cell develops a DNA repair defect only when its second copy of the gene also becomes unfunctional due to a somatic mutation, a mechanism known as Knudson's two-hit hypothesis (12). 70 – 85% of Lynch syndrome cases are caused by mutation in *MLH1* or *MSH2* genes, and 10% - 20% of cases are caused by *MSH6* and *PMS2* mutations. (8).

Carriers with a mutation in *MLH1* or *MSH2* typically develop cancer at ages of 44-61 years, whereas carriers with a mutation in *MSH6* typically develop colorectal cancer at ages of 42-69 years (13).

Mutations in *MLH1* and *MSH2* genes have a larger effect on the DNA repair function than mutations in the genes *MSH6* and *PMS2.* Therefore, patients with mutations in *MLH1* or *MSH2* have a substantially higher risk of developing tumors than patients with *MSH6* mutations (12). So each gene of the MMR system presents different penetrance, leading to different risks of

developing CRC (outlined in more detail in Table 1) (14). Mutations in *PMS2* have lowest penetrance.

The mechanism of MSI is a common mark that is seen in patients who developed Lynch syndrome.

Table 1. Genes involved in Lynch Syndrome

Gene	Characteristics	Number of coding exons (ex) and Amino acids (aa) / Cytogenic location	Phenotypic heterogeneity	Penetrance
MLH1 **Disease MIM ID: 609310**	Is a highly conserved gene especially in the exons 1 – 7, the protein hetero dimerizes with PMS2 to form *MutLα*, a component of the post-replicative DNA mismatch repair system (MMR).	19 ex and 756 aa / 3p22.2	Heterozygous *MLH1* mutation; LS: CRC predominance; LS of a classical phenotype (fulfilling the Amsterdam I criteria) and cancer with high MSI.	Male carriers of a pathogenic *MLH1* gene have 50% risk to develop LS at the age of 60, and 70% risk to develop LS at the age of 75 years. Female carriers of a pathogenic *MLH1* gene have 50% risk to develop LS at the age of 55 years and 70% risk to develop LS at the age of 65 years.
MSH2 **Disease MIM ID: 120435**	The protein MSH2 forms two different heterodimers: MutSα (MSH2-MSH6 heterodimer) and MutSß (MSH2-MSH3 heterodimer) which binds to DNA mismatches thereby initiating DNA repair	16 ex and 934 aa / 2p21-p16.3	Heterozygous *MSH2* mutation; LS: greater frequency of extracolonic cancers. *MSH2* patients are linked to an augmented risk of Muir-Torre syndrome tumors-spectrum.	Male carriers of a pathogenic *MSH2* gene have 50% risk to develop LS at the age of 52, and 70% risk to develop LS at the age of 68 years. Female carriers of a pathogenic *MSH2* gene have 50% risk to develop LS at the age of 53 years and 70% risk to develop LS at the age of 62 years.
MSH6 **Disease MIM ID: 614350**	The protein MSH6 heterodimerizes with MSH2 to form *MutSα*, which binds to DNA mismatches thereby initiating DNA repair	10 ex and 1360 aa / 2p16.3	Heterozygous *MSH6;* LS: predominance of endometrial cancer; tumors sometimes exhibit low-level MSI.	Male carriers of a pathogenic *MSH6* gene have 10% risk to develop LS at the age of 50, and 30% risk to develop LS at the age of 72 years. Female carriers of a pathogenic *MSH6* gene have 20% risk to develop LS at

				the age of 55 years and 50% risk to develop LS at the age of 65 years.
PMS2 **Disease MIM ID:** **614337**	The protein PMS2 forms heterodimers with MLH1 to form *MutL*α heterodimer.	15 ex and 862 aa / 7p22.1	Heterozygous *PMS2;* LS: may contain excess colonic polyps; lower frequency of colorectal cancer.	Male carriers of a pathogenic PMS2 gene have 10% risk to develop LS at the age of 55 and 30% risk to develop LS at the age of 74 years. Female carriers of a pathogenic *MLH1* gene have 10% risk to develop LS at the age of 55 years and 30% risk to develop LS at the age of 75 years.
EPCAM **Disease MIM ID:** **613244**	*EPCAM* is located at 2p21 17 kb upstream of *MSH2*, encodes the EpCAM protein, expressed on the membrane of cells in epithelial tissues and plasma cells.		Heterozygous *EPCAM* deletion; LS: silences *MSH2* expression; often lower risk of extracolonic cancers albeit if the deletion is close to the *MSH2* gene, risk for endometrial cancer increases.	

Note. Adapted from references (15) and (16).

Other type of cancers (in addition to CRC) found frequently in patients with Lynch syndrome are endometrial, stomach, small intestine, pancreas, prostate, liver, kidney, urinary tract, brain and skin cancers (17). Endometrial cancer is the most common extracolonic cancer originating in the endometrium as well as is the most common gynecological tumor in the developed countries. Approximately 30 – 40% of endometrial cancers show a loss of the DNA mismatch repair proteins (18). Lynch syndrome is involved in the development of a single or few colorectal adenoma and carcinoma and clinically needs to be distinguished from sporadic tumors. Therefore clinical and familial criteria was earlier needed to identify patients with Lynch syndrome (12). Before the development of gene testing the patients, who met the Amsterdam criteria were suspected to be Lynch syndrome patients. Bethesda guidelines for testing colorectal tumors for MSI, Amsterdam I criteria that is focused on a strong family history of early-onset CRC, and Amsterdam II criteria conditioning having three relatives with a Lynch syndrome-associated cancer were the standard diagnostic criteria for Lynch syndrome

(19) (Table 2). Nowadays the gene testing method is the most useful way to detect patients with Lynch syndrome.

Table 2. Amsterdam I, Amsterdam II Criteria and Bethesda Guidelines

Amsterdam I	• For a diagnosis of LS, the Amsterdam I Criteria require at least three relatives with histologically identified CRC. • One is a first-degree relative of the other two. • At least two successive generations should be affected. • At least one of the relatives with CRC is diagnosed at <50 years of age. • Tumors should be confirmed by pathology. • FAP should be excluded.
Amsterdam II	• For a diagnosis of Lynch syndrome, the Amsterdam II Criteria require at least three relatives with a cancer associated with hereditary nonpolyposis colorectal cancer (colorectal, stomach, endometrial, ovary, ureter or renal-pelvis, small bowel, brain, hepatobiliary tract, or kin (sebaceous tumors)). • One needs to be a first-degree relative of the other two. • At least two successive generations need to be affected. • At least one of the relatives with CRC needs to have received the diagnosis before age 50. • Familial adenomatous polyposis should be excluded in any relative with colorectal cancer. • Tumors should be verified by pathology whenever possible.
Bethesda Guidelines for testing of colorectal tumors for MSI	• To verify MSI testing, the Bethesda Guidelines require the presence of synchronous or metachronous colorectal or other LS-associated tumors regardless of age. • CRC diagnosed in a patient who is younger than 50 years of age. • CRC with MSI-high histology diagnosed in a patient who is younger than 60 years of age. • CRC or LS-associated tumor diagnosed <50 years of age in at least one-first degree relative. • CRC or LS-associated tumor diagnosed at any age in two first- or second-degree relatives.

Note Adapted from [(20)]

1.3 Mismatch Repair System

The MMR system consists of several genes, including the CRC relevant genes *MLH1, MSH2, MSH6* and *PMS2* (Table 1). The mismatch repair system recognizes, removes, and re-synthesizes a mismatched site in the DNA during the DNA pre-replication and post-replication processes.

The base-base mismatches in the DNA double helix are recognized by MutSα (the heterodimer of MSH2-MSH6) (2). MutSα binds as a sliding clamp in the double-strand of the DNA. The ATP- activated state of MutSα interacts with MutLα (heterodimer of MLH1-PMS2) and together form a tetrameric complex. This tetrameric complex slides up and down of the double-strand DNA and searches for single-strand DNA gaps on the nascent or daughter strand that

recruits proliferating cell nuclear antigen (PCNA) and Replication factor C (RFC). MutLα can incise the nascent strand upon activation by PCNA. Then, the exonuclease 1 (Exo 1) is recruited and removes the daughter strand around the error region. The final step is the re-synthesis step and is carried out by DNA polymerase (Polδ or Polε) and the Ligase 1 (2) (Figure 3).

The *MSH2* and *MLH1* proteins have an ATPase domain which function in a biological reaction involved the hydrolysis of ATP. An ATP- hydrolysis reaction is necessary when MutSα recognizes a mismatch site or when MutLα forms a nick in the DNA strand. Therefore, the completion of the MMR system requires the utilization of energy.

Deletions including the polyadenylation site in exon 8 and 9 of *EPCAM* gene that is located upstream of *MSH2* is identified as a cause of Lynch syndrome. The *EPCAM* deletions lead to silencing of *MSH2* and is estimated to cause ~7% of Lynch syndrome cases without MMR mutation. (21)

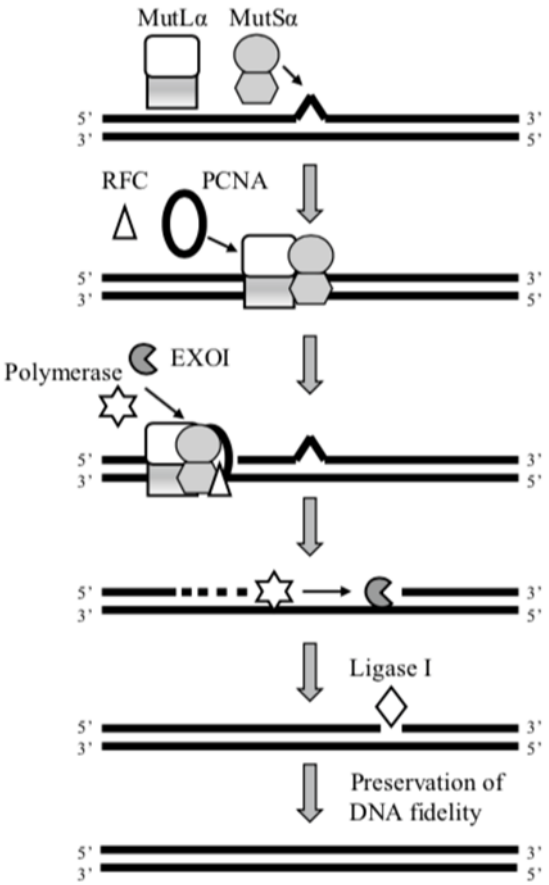


Figure 3. Molecular mechanism model of Mismatch Repair System

International Journal of Clinical Oncology **volume 24**, pages999–1011(2019) (2)
doi.org/10.1007/s10147-019-01494-y

Patients presenting a germline mutation in one of the four genes of the MMR system or EPCAM deletions are at high risk of developing Lynch syndrome associated cancers. For this reason, identification of these patients is of vital importance. Through this it would be easy to provide them a detail information about Lynch syndrome alongside with a detailed surveillance program.

1.4 Mutations in Lynch Syndrome

A large number of different mutations have been identified in Lynch syndrome patients including missense, frameshift, nonsense, and splicing mutations that result in truncated or altered protein structure (20). Functional studies have demonstrated that these variants are pathogenic because they alter the MMR capability of the encoded protein and they segregate in families (20). Genomic rearrangements (deletions and duplications) of *MSH2* and *MLH1* both resulting in the loss of the intact proteins are frequent causes of Lynch syndrome. (20).

1.5 Classification of mutations

In 2008, the International Agency for Research on Cancer (IARC) created the variant classification scheme to classify the cancer susceptibility genes (Table 3) (22).

Table 3. The IARC variant classification scheme

Category	**Synonym**
Pathogenic	
Likely Pathogenic	
VUS	Variant of Uncertain Significance
Likely not pathogenic	Likely benign
Not pathogenic	Benign

Note. Adapted from [(22)].

The scheme was adopted by the International Society for Gastrointestinal Hereditary Tumors (InSiGHT) and, by the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (23). If a variant is not associated with enough data to classify them as Pathogenic or Not Pathogenic, the variant will be classified as Variant of Uncertain Significance (VUS) (22).

To diagnose and verify that a cancer patient has Lynch syndrome, gene testing (sequencing and copy number variant analyses) of the four MMR genes and *EPCAM* are performed. The ACMG classified the sequencing variants into 5 classes of variants:

- Class 5: pathogenic
- Class 4: likely pathogenic
- Class 3: VUS
- Class 2: likely benign
- Class 1: benign

A loss of function variant (class 4 and 5) in one of the MMR genes is known to cause Lynch syndrome. However, a class 3 variant cannot be used to diagnose a patient with this syndrome.

1.6 VUS

VUS are defined as variants for which the clinical significance is unknown (24). These variants include missense variants, small in- frame deletions or insertions, synonymous nucleotide substitutions, as well as alterations in noncoding sequences (25). In contrast of pathogenic mutations, VUS can't firmly diagnose Lynch syndrome at the molecular level because many of them turn out to be neutral changes (26). The International database Insight lists more than 1,000 different VUS for MRR gene variants (26).

There are several ways to predict if a VUS variant could be pathogenic, for example by checking the frequency of the variant in the normal population. If it is a common variant in the general population then it is unlikely to cause the syndrome. On the other hand, if the variant is not reported before ("rare variant"), it could be either a new pathogenic mutation or a rare neutral variant. It's important to check the ethnicity and as well the geographical location of the reported populations because the frequency of the variants may vary between different ethnic groups and different populations or by geography location. Another way to indicate the probable pathogenicity of the variant is if the variant is observed in Lynch syndrome patients. A third way could be *in silico* or software tools to predict if the variant could change the protein function of the gene and these are just some of the ways that the pathogenicity of a VUS could be tested (26). Variants of Uncertain Significance represent 20 – 30 % of variants found in clinical testing of MMR genes (26).

1.7 Functional assays

As mentioned above in sections 1.4 and 1.5 a class 3 variant or a (VUS) cannot be used to diagnose a patient with Lynch syndrome. Therefore, it is important to verify whether a VUS is pathogenic or not, to know whether the variant has any effect on gene function which might cause an increased cancer risk, to avoid misinterpretation and not to obfuscate personalized health care (27). One possibility to characterize and reclassify a VUS, is to perform functional assays.

Most of the functional assays are designed and made to measure a specific effect the potential mutation may have on the biological function of the mutated MMR protein (28). The assays may determine the capacity of two MMR proteins to form for example protein – protein complexes and are commonly used with relevant MMR-deficient cell lines to determine the stabilization of the endogenous or exogenous binding partner (28). Other assays have been applied to reveal ADP-ATP cycling by MSH protein dimers.

As well as to determine whether the mutated protein can restore the repair capacity in MMR-deficient cells assays have been developed in which MMR-deficient human cells were complemented with an exogenous expressed MMR gene or genes and the repair capacity could be measured.

1.8 Aim of the study

Based on the premise that VUS are variants for which the clinical significance is unknown as well as they cannot be used to diagnose a patient with Lynch syndrome, this master project has two different aims.

The first aim is to characterize and classify *MLH1* and *MSH6* VUS variants identified at St Olavs Hospital, using multiple prediction tools, well-established classification guidelines, and publicly available databases to see whether some of these VUS could be re-classified as likely pathogenic or benign. The second aim is to perform a literature review to make an overview of functional assays that could validate the suspicious pathogenic variants suggested by the different tools in the first approach, with the objective to establish the best functional assay at St. Olav's Hospital.

2 MATERIALS AND METHODS

2.1 Study Samples

The *MLH1* and *MSH6* VUS variants (*n*=27 and 52, respectively) have been identified using Sanger sequencing or Next Generation Sequencing at the Department of Medical Genetics at St. Olavs Hospital.

2.2 Data analysis

The characterization and classification of the different VUS were performed using information from different databases and prediction tools. *MLH1* and *MSH6* VUS were categorized into a pathogenicity class according to the American College of Medical Genetics and Genomics (ACGM) classification standards and guidelines, based on information of each variant obtained from ClinVar, VarMap, Varsome, gnomAD and from the literature. The allele frequency of each variant was obtained from gnomAD. The different databases and prediction tools are described below.

2.2.1 ClinVar

ClinVar (https://www.cbi.nlm.nih.gov/clinvar/) is a large archive of information of clinically significant variants and their phenotypes, providing sequence variation and interpretation of the links between the variation and human diseases based on supporting evidences and classification guidelines (25). It is an open source where submitters can report variants and their phenotype freely, as well as some submitters could supply with the conclusion about the classifications of these variants and include the interpretations of them based on the literature.

2.2.2 VarMap

VarMap (https://bio.tools/VarMap) is a tool designed to map genomic coordinates to protein structures, providing information of protein 3D structures, annotation in structural information and potential molecular consequences of sequence variations (26). VarMap provides algorithms such as CADD which is a prediction tool estimating deleterious of variants, and PhastCons which is a program displaying conservation score of elements in variants. The data of VarMap was used to predict the impact of amino acid residue changes and disease propensity of the *MLH1* and *MSH6* missense variants in this study.

2.2.3 VarSome

VarSome (https://varsome.com/) is a search engine, and impact analysis tool for human genetic variation and a driven project aiming at sharing global expertise on human variants. VarSome enables the users to look up variants in their genomic context, collects data from multiple databases in a central location and most importantly, aims to enable to freely and easily share knowledge on human variation. Includes information from 30 external databases VarSome's databases consists of more than 33 billion data points describing 500 million variants (27).

2.2.4 gnomAD (The Genome Aggregation Database)

gnomAD (https://gnomad.broadinstitute.org/) consortium offers both exome and genome sequencing data from diverse large – scale sequencing projects and researches. The version 2.1.1 data set (GRCh37/hg19) from this website was used. This version spans 125,748 exome sequences and 15,708 whole-genome sequences from unrelated individuals sequenced as part of various disease-specific and population genetic studies, in the whole genome and in the protein coding regions, respectively with a totaling sequencing of 141,456 individuals (28). The population with the highest allele frequency from each *MLH1* and *MSH6* VUS variant obtained from version 2.1.1 of gnomAD was used in the classification criteria of the *MLH1* and *MSH6* VUS variants. A list with the information about the names, the number of Genomes and the number of Exomes of the populations used in the version 2.1.1 and in this thesis is shown in Appendix 1.

2.3 Interpretation of variants

2.3.1 ACMG (The American College of Medical Genetics and Genomics) standard classification

The classification of variants was identified according to the five-tier classification system of the ACMG. Following the variant classification of the ACMG, each variant was categorized as shown in Table 5 below:

Table 5. ACMG standard classification system

Classification	Description
Pathogenic (Class 5)	This variant directly contributes to the development of disease. Some pathogenic variants may not be fully penetrant. In this case of recessive or X-linked conditions, a pathogenic variant may not be solely sufficient enough to cause disease on its own.
Likely pathogenic (Class 4)	There is a high likelihood (greater than 90% certainty) that this variant is disease-causing. Additional evidence is expected to confirm this assertion of pathogenicity, however, there is a small possibility that new evidence may demonstrate that this variant does not have clinical significance.
Uncertain significance (Class 3, VUS)	There is not enough information at this time to support a more definitive classification of the variant.
Likely benign (Class 2)	This variant is not expected to have a major effect on disease; however, the scientific evidence is currently insufficient to prove this conclusively. Additional evidence is expected to confirm this assertion; however, we cannot fully rule out the possibility that new evidence may demonstrate that the variant can contribute to disease.
Benign (Class 1)	This variant does not cause disease.

Note. Adapted from (29).

The ACMG variant classification guidelines have provided standards for interpretation of sequence data from patients presenting common disease phenotypes to be identified whether they are predisposed to a disease with high penetrance. The interpretation of the ACMG guidelines was designed to recap conclusions generated by the ACMG combinatorial scoring method. The value of its criteria, however, has been constantly changed. ACGS (Association for Clinical Genomic Science, Ellard et al., 2020 (30) has suggested new standards for variant classification in rare disease based on the ACMG guidelines (Table 6, 7 and 8). The ACMG guidelines for classifying pathogenic or likely pathogenic variants and for classifying benign or likely benign variants have been revised by Tavtigian et al. (31).

Table 6. Criteria for classifying pathogenic variants from ACMG standards and guidelines

Evidence of pathogenicity	Category
Very strong	**PVS1: N**ull variant (nonsense, frameshift, canonical ± 1 or 2 splice sites, initiation codon, single or multi-exon deletion) in a gene where LOF is a known mechanism of disease.
Strong	**PS1:** Same amino acid change as a previously established pathogenic variant regardless of nucleotide change **PS2:** De novo (both maternity and paternity confirmed) in a patient with the disease and no family history **PS3:** Well – established in vitro or in vivo functional studies supportive of a damaging effect on the gene product **PS4:** The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls
Moderate	**PM1:** Located in a mutational hot spot and/or critical and well- established functional domain (e.g., active site of an enzyme) without benign variation **PM2:** Absent from controls (or at extremely low frequency if recessive) (table 5) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium **PM3:** For recessive disorders, detected in trans with a pathogenic variant **PM4:** Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants. **PM5:** Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before. **PM6:** Assumed de novo, but without confirmation of paternity and maternity.
Supporting	**PP1:** Co-segregation with disease in multiple affected family members in a gene definitively known to cause the disease

	PP2: Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism if disease **PP3:** Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.) **PP4:** Patient's phenotype or family history is highly specific for a disease with a single genetic etiology **PP5:** Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation

Table 7. Criteria for classifying benign variants from ACMG standards and guidelines

Evidence of benign impact	Category
Stand alone	**BA1:** Allele frequency is >5% in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium
Strong	**BS1:** Allele frequency is greater than expected for disorder **BS2:** Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder, with full penetrance expected at an early age **BS3:** Well – established in vitro or in vivo functional studies show no damaging effect on protein function or splicing **BS4:** Lack of segregation in affected members of a family
Supporting	**BP1:** Missense variant in a gene for which primarily truncating variants are known to cause disease **BP2:** Observed in trans with a pathogenic variant for a fully penetrant dominant gene/disorder or observed in cis with a pathogenic variant in any inheritance pattern **BP3:** In-frame deletions/insertions in a repetitive region without a known function

	BP4: Multiple line of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc.) **BP5:** Variant found in a case with an alternate molecular basis for disease **BP6:** Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation **BP7:** A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site and the nucleotide is not highly conserved

Note. Adapted from (30)

Table 8. Updated combining criteria for classifying pathogenic or likely pathogenic variants

Classification	Combining rules
Pathogenic (a)	1. 1 Very strong (PSV1) AND 1. ≥ 1 Strong (PS1-PS4) OR 2. ≥ 2 Moderate OR (c) ≥ 1 Moderate and ≥1 supporting OR (d) ≥ 1 Moderate OR (ii) ≥ 2 Supporting (PP1-PP5)
Pathogenic (b)	1. Strong (PS1 – PS4) AND 1. ≥ 3 Moderate (PM1 – PM6) OR 2. ≥ 2 Moderate (PM1 – PM6) AND ≥ 2 supporting (PP1-PP5) OR 3. ≥ 1 Moderate (PM1-PM6) AND ≥ 4 supporting (PP1 – PP5)
Likely Pathogenic (a)	≥ 2 Strong (PS1 – PS4)
Likely Pathogenic (b)	1. Strong (PS1-PS4) AND -2 moderate (PM1-PM6) OR ≥2 Supporting (PP1-PP5)
Likely Pathogenic (c)	1. ≥ 3 Moderate (PM1-PM6) OR (ii) 2 Moderate (PM1-PM6) AND ≥2 Supporting (PP1-PP5) OR (iii)1 Moderate (PM1-PM6) AND ≥4 Supporting (PP1-PP5)

Note. Adapted from (30)

The ACGS rules (Diagram from Ellard et al 2020) of Figure 4 was used with the purpose to categorize and select the VUS which are most likely to be pathogenic variants from the different *MLH1* and *MSH6* VUS variants. This diagram represents the different VUS categories according to the ACGS guidelines using a temperature gradient sub-classification going from Ice-cold (little evidence for pathogenicity) to hot (several evidences for pathogenicity, but not enough to be a class 4, likely pathogenic variant).

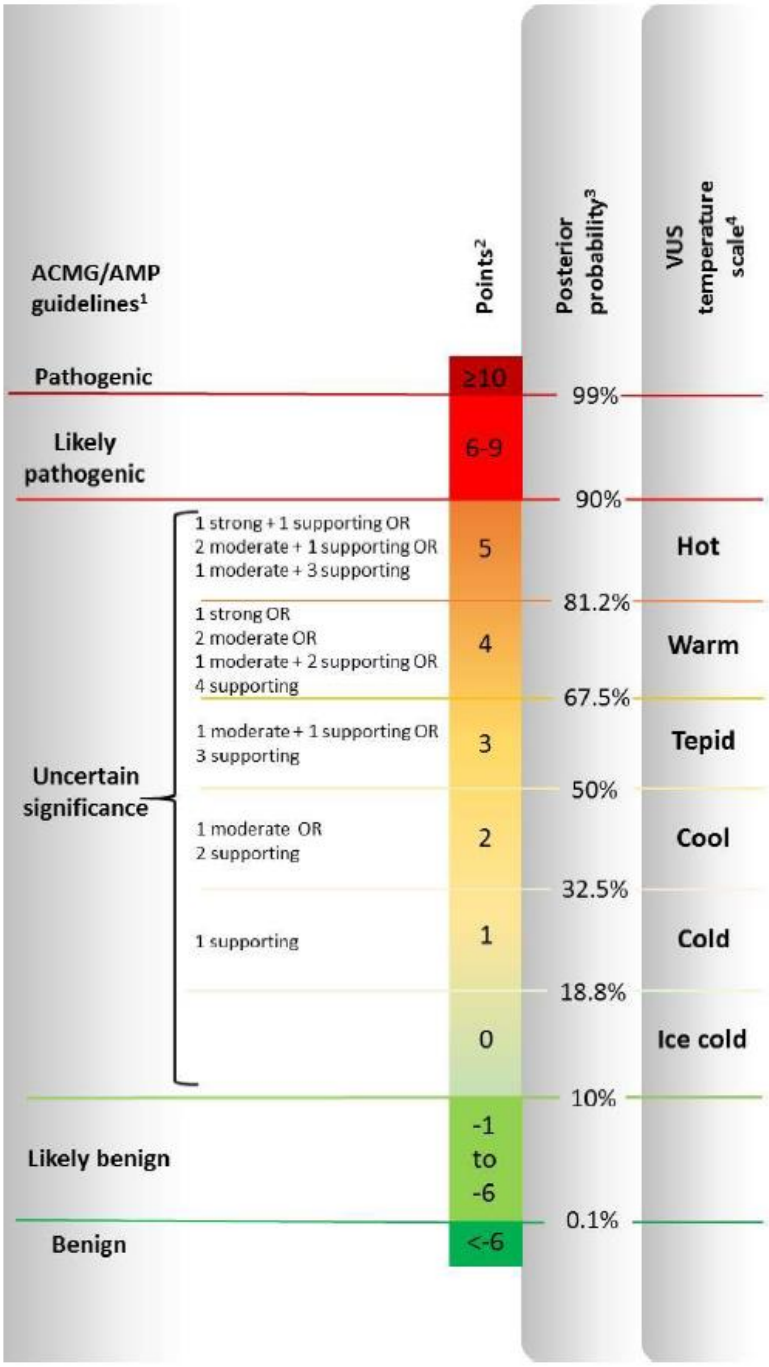


Figure 4. Diagram to illustrate the different ways to describe variants of uncertain significance with differing levels of evidence in support of pathogenicity and VUS temperature. Ellard et al 2020 (30)

The ACMG guidelines (Table 6 and 7) have been constantly changed. These rules were used for the interpretation of the variants. Now The Cancer Variant Interpretation Group UK (CanVIG-UK) has redefined the original ACMG-AMP guidelines with more precise and clear classification rules based on the original system of ACMG-AMP guidelines for classifying pathogenic or likely pathogenic variants and for classifying benign or likely benign variants. The CanVIG-UK classification was used to classify the MSH6 and MLH1 VUS variants.

Table 9. Criteria from CanVIG-UK for classifying pathogenic variants

Evidence of pathogenicity	Category
Very strong	**PVS1:** Null variant (nonsense, frameshift, canonical ± 1 or 2 splice sites, initiation codon, single or multi-exon deletion) in a gene where LOF is a known mechanism of disease.
Strong	**PS1:** Same amino acid change as a previously established pathogenic variant regardless of nucleotide change. Use at **Strong** for a missense variant under evaluation whereby there is a reference missense variant classified as (likely) pathogenic that results in the same amino acid change. Use at **Moderate** for an initiation codon variant under evaluation whereby there is a reference variant in the initiation codon classified as (likely) pathogenic. Use at **Supporting** for a donor/acceptor splice region variant under evaluation whereby there is a reference variant at the same base residue classified as (likely) pathogenic. The variant under evaluation must be predicted on in silico tools to be equally or more deleterious than the reference variant. **PS2:** De novo (both maternity and paternity confirmed) in a patient with the disease and no family history **PS3:** Well – established in vitro or in vivo functional studies supportive of a damaging effect on the gene product. **For assays of protein function:** Use at **Strong** when the relative protein activity assay or functional impact <25% compared to level for wildtype, controls ≥ 10 true positive ≥ 10 true negative, reproducibility ≥2 laboratories OR results demonstrably reproducible from a single laboratory. Use at **Moderate** when the relative protein activity assay or functional impact <25% compared to level for wildtype. Controls ≥ 5 true positive ≥ 5 true-negative, reproducibility ≥ 2 laboratories OR results demonstrably reproducible from a single laboratory.

	Use at **Supporting** when the relative protein activity assay or functional impact <25% compared to level for wildtype, controls ≥ 2 true positive ≥ 2 true negative, reproducibility single laboratory. **For assays of splicing function:** Use at **Very strong** when 2 orthogonal assays: exhibiting abnormal transcripts; no evidence of leakiness. Use at **Strong** when 1 assay: exhibiting abnormal transcripts; no evidence of leakiness. Use at **Moderate** when ≥ 1 assay: exhibiting abnormal/alternative transcripts; evidence of leakiness. Use at **Supporting** ≥ 1 assay: exhibiting abnormal/alternative transcripts which have been reported as present in normal controls (implying naturally occurring isoforms). Don't apply ≥ 1 assay: exhibiting abnormal/alternative transcripts with evidence of extreme leakiness. **PS4** The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls Used **Vstrong** $P_{exact} \leq 0.0025$ Used **Strong** $P_{exact} \leq 0.05$ Used **Moderate** $P_{exact} \leq 0.1$ Used **Supporting** $P_{exact} \leq 0.$
Moderate	**PM1:** Located in a mutational hot spot and/or critical and well- established functional domain (e.g., active site of an enzyme) without benign variation Use PM1 at **Moderate** for a variant in a mutational hotspot at which there is no benign variation. Use PM1 at **Supporting** for a variant in a mutational hotspot at which there is some benign variation. **PM2:** Absent from controls (or at extremely low frequency if recessive) (table 5) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium. Use at **Moderate** where 0 observations of the variant in control series >50,000 individuals. Use at **Supporting** where 1 observation on the variant in control series >50,000 individuals. **PM3:** For recessive disorders, detected in trans with a pathogenic variant. Use where variant found in trans with a pathogenic variant and the patient-level clinical features match those anticipated for the gene in question. Use at **Strong** where variant found in ≥ 2 unrelated cases, and the features are distinctive for that gene.

	Use at **Moderate** where variant found in 1 case, and the features are distinctive for that gene. Use at **Supporting** where variant found in 1 case, and the features are distinctive for a set of genes. **PM4:** Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants. Use at **Moderate** for In-frame insertions/deletions for which PVS1 is not applicable. Use at **Supporting if** Reference variant is classified likely pathogenic and only reported in 1 individual. **PM5:** Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before. Use at **Moderate** if reference variant is classified as pathogenic [OR likely pathogenic and reported in >1 individual] Use at **Supporting** if reference variant is classified likely pathogenic and only reported in 1 individual. **PM6:** Assumed de novo, but without confirmation of paternity and maternity.
Supporting	**PP1:** Co-segregation with disease in multiple affected family members in a gene definitively known to cause the disease **PP2:** Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease Use PP2 at **Supporting** where there is overall constraint for missense variation at the level of the region/exon/gene (Z ≥3.09) **PP3:** Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.) Protein impact: using a **predefined** strategy of 3/3 tools (one tool may be marginally below threshold) SIFT (deleterious), Polyphen HumVar ≥ (probably damaging) plus: Align GVGD (C45, C65), (for BRCA1, BRCA2) OR MAPP (bad) (for MMR genes) OR CADD (>15) (for any other CSG) Or use Revel (> 0.7) as a single score Splicing impact: Intron-exon boundary: MaxEnt > 15% difference AND SSFL > 5% difference Deep intronic: predicted creation of a novel splice site of any strength, absent in the normal sequence **PP4:** Patient's phenotype or family history is highly specific for a disease with a single genetic etiology Level -, Points 0.5, Cellular/molecular phenotype: Moderately predictive for germline aberration of one of a small set of genes, Example: MSI (for mismatch repair deficiency).

	Level **Supporting,** Points 1, Cellular/molecular phenotype: Highly predictive for germline aberration of one of a small set of genes, Example: Aberration on mitomycin-induced chromosomal breakage (for genes related to Fanconi Anemia). Level **Supporting,** Points 1, Cellular/molecular phenotype: Moderately predictive for germline aberration of the specific gene, Example: LOH at chromosomal locus of tumor-suppressor gene and loss on immunohistochemistry of single protein e.g., MSH6, PMS2. Level **Moderate,** Points 2, Cellular/molecular phenotype: Highly predictive for germline aberration of the specific gene, Example: Depletion of BRCA2 in lymphocytes and aberration on mitomycin-induced chromosomal breakage (for BRCA2-related Fanconi Anemia), Loss on immunohistochemistry of paired mismatch repair proteins e.g. MSH2 and MSH6 and Loss of MLH1 + PMS2 on immunohistochemistry and normal MLH1 promoter methylation (for MLH1- related mismatch repair deficiency). **PP5:** Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation. Any classification of LP/P after 2016 from ≥ 2 accredited North American diagnostic laboratories OR A single North American diagnostic laboratory where the utilized evidence is clearly cited an approved ClinGEn Expert Group (3 star on ClinVar). When a single laboratory has classified as LP/P with provision of insufficient detail, it is advised that the individual laboratory is contacted to procure directly the evidence used for classification

Table 10. Criteria from CanVIG-UK for classifying benign variants

Evidence of benign impact	Category
Stand alone	Use BA1 as **Stand Alone** when allele frequency in a large dataset of heterogenous outbred population (>10,000 individuals) is: >1% or > 0.5% (BRCA1, BRCA2, MLH1, MSH2) Use BS1 as **Strong** when allele frequency in a heterogenous outbred population is > value specified for specific gene by respective expert group.
Strong	**BS1:** Allele frequency is greater than expected for disorder Use BS1 as **Strong** when allele frequency in a heterogenous outbred population is > value specified for specific gene by respective expert group **BS2:** Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder, with full penetrance expected at an early age. Use BS2 at **Supporting** where no further genotyping or clinical/cellular phenotyping is possible Use BS2 at **Strong** where Laboratory analysis has been repeated using an orthogonal approach (e.g. different primers) to confirm homozygosity for allele AND Patient has been actively examined to exclude relevant phenotype AND/OR had analysis of cellular phenotype OR the homozygote is observed in a specified control population in addition to a heterozygote frequency meeting BS1 **BS3** Well – established in vitro or in vivo functional studies show no damaging effect on protein function or splicing **For assays of protein function** Used at **Strong** when relative protein activity assay or functional impact >25% compared to level for wildtype, Controls ≥ 10 true positive ≥ 10 true negative, Reproducibility: ≥2 laboratories OR Results demonstrated as in single laboratory. Used at **Supporting** Relative protein activity assay or functional impact > 25% compared to level for wildtype, Controls: ≥ 10 true positive, ≥ 10 true negative, Reproducibility: ≥ 2 laboratories OR Results demonstrated as reproducible in single laboratory **For assays of splicing function** Used at **Strong when 1 assay:** with no evidence of abnormal transcripts (% normal transcript>90%), ISO accredited laboratory or recognized research laboratory with which direct consultation can be undertaken. Used at **Supporting when 1 assay:** with no evidence of abnormal transcripts (% normal transcripts>90%), alternative source of evidence (e.g. publication)

	BS4: Lack of segregation in affected members of a family For cancer susceptibility genes for which the phenotype is non-specific and/or feature age-related/reduced penetrance, phenocopies or hypomorphic variants, expert review is recommended for application of BS4 pertaining to non-segregation.
Supporting	**BP1:** Missense variant in a gene for which primarily truncating variants are known to cause disease Use at **Supporting** for genes/gene regions in which >95% of reported pathogenic variants are truncating **BP2:** Observed in trans with a pathogenic variant for a fully penetrant dominant gene/disorder or observed in cis with a pathogenic variant in any inheritance pattern Use BP2 at **Supporting** where no further genotyping or clinical/cellular phenotyping is possible Use BS2 at **Strong** where Alleles have been confirmed as in trans AND patient is of age at which biallelic pathogenic variants would be anticipated to be penetrant for a distinctive phenotype AND patient is of age at which biallelic pathogenic variants would be anticipated to be penetrant for a distinctive phenotype AND patient has been actively examined to exclude relevant phenotype AND/OR had analysis of cellular phenotype **BP3:** In-frame deletions/insertions in a repetitive region without a known function **BP4:** Multiple line of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc.) Splicing impact: Intron-exon boundary: Minimal difference in readings for each of MaxEnt AND SSFL AND no evidence of prediction of exonic/deep intronic novel splice site of any strength **AND** Protein impact: Using a predefined strategy of 3/3 tools (one tool may be marginally above threshold) SIFT (tolerated), Polyphen HumVar (benign) plus: Align GVGD (C0,C15), (for BRCA1, BRCA2) MAPP (good) (for MMR genes) CADD (<10) (for any other CSG) Or Revel (<0.4) as a single score **BP5:** Variant found in a case with an alternate molecular basis for disease This shouldn't be applied for autosomal dominant incompletely penetrant non-syndromic genes associated with common cancers e.g. HBOC (hereditary breast and ovarian cancer)

	BP6: Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation **BP7:** A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site and the nucleotide is not highly conserved Not to be used if any for suspicion of an impact on splicing

2.4 Functional assays

In order to validate the functional impact of the VUS variants in *MLH1* and *MSH6* genes different types of functional assays need to be evaluated. For this purpose, experimental articles about functional assays were searched and selected following two important criteria.

The first criterion was to make a query of Experimental Articles in PubMed using these key words:

1. Functional assays evaluating MMR in Lynch syndrome
2. Functional testing of MMR variants
3. Functional validation assays MMR genes
4. Functional validation assays for VUS in MMR genes
5. Functional testing of MMR variants in Colorectal Cancer
6. Functional assays evaluating the MMR activity of VUS variants in MMR genes

The second criterion was to select experimental functional assays articles according to this main aspect:

1.) Functional assays articles that determine whether the mutated protein can restore the Mismatch repair capacity.

2.) Functional assays articles that evaluate the specific effect of the potentially mutation on the biological or biochemical function of the mutated MMR protein.

To achieve the second criterion the articles that were displayed in PubMed from each of the five queries listed in the first criterion were selected according to these points:

1. Articles published from 1990 – until now
2. Articles using cell-free *in vitro* models, human cell line models, or mutational scanning methods.

2.5 Grantham distance

The Grantham Distance is a formula used for knowing the difference between amino acids. This formula combines properties that correlate with protein residue substitution such as composition, polarity and molecular volume. This formula was presented by Grantham R. (32). As well as this formula identify chemical factors that individually correlate best with evolutionary exchangeability of protein residues. The Grantham distance was used in this thesis to identify the impact of the changes of amino acids in the variants in order to obtain an assessment of pathogenicity.

3 RESULTS

3.1 *MLH1* and *MSH6* VUS variants classification

The classification for the VUS with highest potential to be likely pathogenic variants from *MLH1* and *MSH6* VUS variants are presented in the tables 9 and 10, respectively. Classifications were made using the information found for each variant in different databases such as: gnomAD, ClinVar, Varmap and VarSome as well as the ACGS rules. VUS temperature classification points for VUS variants from Ellard et al 2020 were useful tools to select the most promising likely pathogenic VUS variants.

The tepid, warm and hot VUS variants (Table 9 and 10) are described in more detail in the paragraphs below while the classification of the cold VUS are presented in Appendix 2 and 3.

3.1.1 Tepid and Hot MLH1 VUS Variants

Variant NM_000249.3(MLH1):c.43G>C, p.(Val15Leu)

This variant doesn't appear in the gnomAD database meaning that this variant has a Frequency of cero in the about 250.000 alleles, and for this reason was classified with a PM2 according to the ACMG-AMP classification. According to Varsome 11 in silico programs (BayesDel_addAF, DANN, DEOGEN2, EIGEN, FATHMM-MKL, LIST-S2, M-CAP, MVP, MutationTaster, PrimateAI and SIFT) classified the variants as pathogenic corresponding to a PP3 for the ACMG-AMP classification. According to the ACGS rules (Diagram from Ellard et al 2020) the sum of 1 Moderate (PM2) + 1 supportive (PP3) evidence from the criteria to classify pathogenic variants gives 3 points and a Tepid Temperature to this variant.

Variant NM_000249.3(MLH1):c.110A>C, p.(Glu37Ala)

The variant p.Glu37Ala doesn't appear in gnomAD database and this corresponds to a PM2 classification according to the ACMG-AMP classification. This codon has 1 pathogenic alternative variant p.(Glu37Lys) according to Drost et al 2010 (33) and Andersen Sofie et al 2012 (34) the variant p.(Glu37Lys) decrease the mismatch repair activity and loss of nuclear localization. Likewise, Shirts BH et al, 2018 (35) classified the variant p.(Glu37Lys) as pathogenic. For all these evidences presented a classification of PM5 was used. According to Varsome 11 in silico programs (BayesDel_addAF, DANN, DEOGEN2, EIGEN, FATHMM-MKL, LIST-S2, M-CAP, MVP, Mutation Assessor, Mutation Taster and SIFT) predict the

variant as pathogenic giving a PP3 classification. Summing up all this evidence from different databases and according to the Ellard et al 2020 diagram the sum of PM2 (1 Moderate) + PM5 (1 Moderate) and PP3 (1 supportive) results in 5 points and a VUS temperature of Hot.

Variant NM_000249.3(MLH1):c.794G>A p.(Arg265His)

The variant p.(Arg265His) has an a mean allele frequency of 0.00004772 from all populations that present this variant and the population with the highest allele frequency for this variant is (European non-Finnish) with an allele frequency of 0.00007033. Omar Soukarieh, et al 2016 (36) said that this variant didn't induce middle exon skipping hypothesizing that the effect on splicing of the variant depends on surrounding nucleotide context. This variant has 3 pathogenic alternative variants, p.(Arg265Ser), p.(Arg265Cys) and p.(Arg265Pro). According to Drost et al 2010 (33) the variant p.(Arg265Ser) decreased mismatch repair activity. According to Tournier Isabelle et al 2008 (37) the variant p.(Arg265Cys) has a partial of exon 10 skipping, as well as Drost et al 2010 (33) and Andersen Sofie et al 2012 (34) said that the variant p.(Arg265Cys) decrease the mismatch repair activity. Likewise, Nicole Köger, et al 2018 (38) consider the variant p.(Arg265Pro) as pathogenic with all this information the variant was classified as PM5. VarSome also showed that 11 in silico programs (BayesDel_addAF, DANN, DEOGEN2, EIGEN, FATHMM-MKL, LIST-S2, M-CAP, MVP, MutationAssessor, MutationTaster and SIFT) classified the variant as pathogenic giving a PP3 classification. Summing up all this evidence from different databases and according to the Ellard et al 2020 diagram the sum of PM5 (1 Moderate) and PP3 (1 supportive) results in 3 points and a VUS temperature of Tepid.

Variant NM_000249.3(MLH1):c.848 A>G p.(Tyr283Cys)

The variant p.(Tyr283Cys) doesn't appear in gnomAD database and this corresponds to a PM2 classification according to the ACMG-AMP classification. In the literature this variant has not been reported in individuals with MLH1 related disease. VarSome also showed that 11 in silico programs (BayesDel_addAF, DANN, DEOGEN2, EIGEN, FATHMM-MKL, LIST-S2, M-CAP, MVP, MutationAssessor, MutationTaster and SIFT) classified this variant as pathogenic giving a PP3 classification as well as this variant was found in 8 cases reports from patients that are tested for inherited cancer. The genetic testing was made by Ambry genetics, GeneDx, Invitae, Color Health, Inc., Illumina Clinical Laboratory, CeGaT Praxis fuer Humangenetik Tuebingen, Quest Diagnostics Nichols Institute and Integrated Genetics/Laboratory Corporation of America. According to the ACGS rules (Diagram from Ellard et al 2020) a PP3

(1 Supportive) plus PM2 (1 Moderate) give this variant 3 points and a VUS temperature of Tepid.

Variant NM_000249.3(MLH1):c.894 C>G p.(Ile298Met)

The variant p.(Ile298Met) doesn't appear in gnomAD database and this correspond to a PM2 classification according to the ACMG-AMP classification. This variant has not been reported in individuals with Lynch syndrome. VarSome also showed that 11 in silico programs (BayesDel_addAF, DANN, DEOGEN2, FATHMM-MKL, LIST-S2 and 6 more) classified this variant as pathogenic giving a PP3 classification. A PM2 (1 Moderate) and a PP3 (1 supportive) according to the ACGS classification rules (Diagram from Ellard et al 2020) give this variant 3 points and a VUS temperature of Tepid.

Variant NM_000249.3(MLH1):c.1493 G>A p.(Arg498Lys)

This variant p.(Arg498Lys) doesn't appear in gnomAD database and this correspond to a PM2 classification according to the ACMG-AMP classification. According to VarSome 9 in silico programs (bayesDel_addAF, DANN, FATHMM-MKL, LIST-S2, M-CAP, MVP, MutationAssessor, Mutation Taster and SIFT) classified this variant as pathogenic giving a PP3 classification. According with the ACGS classification rules a sum of PM2 + PP3 results in 3 points and a VUS temperature of Tepid.

Table 9. *MLH1* VUS variants classification

Variant (Reference Sequence: NM_000249.3)	gnomAD Mean Allele Frequency / Highest population Frequency	ClinVar	ACGS classification rules	VUS temperature
c.43 G>C p.(Val15Leu)	Frequency cero	In silico programs said that is disruptive	**PM2 + PP3**	**3 Points = Tepid**
c.110 A>C p.(Glu37Ala)	Frequency cero	**No information found**	**PM2 + PM5 + PP3**	**5 Points = Hot**

c.794G>A p.(Arg265His)	0.00004772 / (European non-Finnish) 0.00007033	**Information found with the version number NM_000249.4 2x Likely benign & 8x Uncertain significance**	**PM5 + PP3**	**3 points = Tepid**
c.848 A>G p.(Tyr283Cys)	Frequency cero	**Interpretation: Uncertain Significance**	**PM2 + PP3 +**	**3 points = Tepid**
c.894 C>G p.(Ile298Met)	Frequency cero	**Interpretation: Uncertain Significance**	**PM2 + PP3**	**3 points = Tepid**
c.1493 G>A p.(Arg498Lys)	Frequency cero	**N.I.**	**PM2 + PP3**	**3 points = Tepid**

The Hot and Tepid Variants in the MLH1 protein appear in 3 functional domains: The ATPase domain, MutSα interaction domain, and the EXO1 interaction domain (Figure 5). Four Tepid VUS variants are present in the MutSα interaction domain these variants are: p.Val15Leu, p.Arg265His, p.Ile298Met and p.Tyr283Cys. The Hot Variant p.Glu37Ala appears in the MutSα interaction and in the ATPase domain. Just one variant appears in the EXO1 interaction.

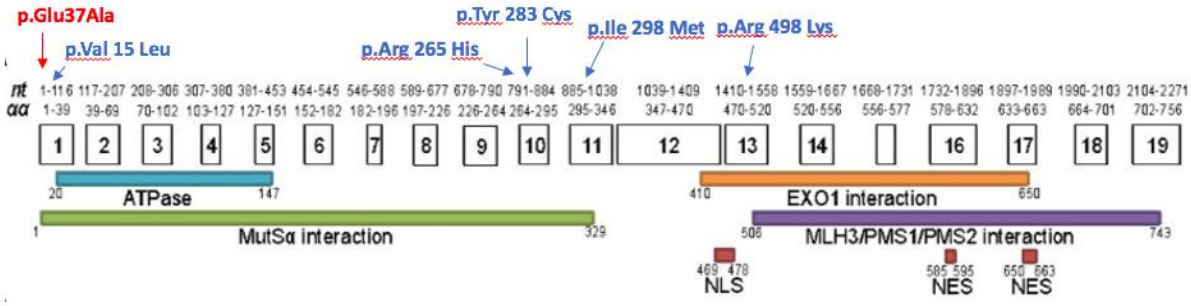


Figure 5. Diagram representing the Linear schematic of MLH1 mismatch repair gene structure and functional domains. The different MLH1 VUS variants classified as the most probably damaging ones are indicated above. Variants with a Tepid temperature appear in color blue, Hot variants are presented in red. Adapted from (39)

3.1.2 Tepid and Hot MSH6 VUS variants

Variant NM_000179.2(MSH6):c.234A>G p.(Arg78=)

This variant doesn't appear in gnomAD database meaning that this variant has a frequency of cero in about 245,000 alleles and for this reason was classified with a PM2 classification according to the ACMG-AMP classification. Algorithms to predict the effect of sequence changes on RNA splicing suggest that this variant create or strengthen a splice site receiving a PP3 classification according to the ACMG-AMP classification. A PM2 (1 Moderate) and a PP3 (1 supportive) according to the ACGS classification rules (Diagram from Ellard et al 2020) give this variant 3 points and a VUS temperature of Tepid.

Variant NM_000179.2(MSH6):c. 1137 A>C p.(Arg379Ser)

This variant doesn't appear in gnomAD database and this correspond to a PM2 classification according to the ACMG-AMP classification. This variant hasn't been reported in individuals with MSH6-related conditions in the literature. According to VarSome nine in silico programs (BayesDel_addAF, DANN, DEOGEN2, FATHMM-MKL, LIST-S2, M-CAP, MVP, MutationTaster and SIFT) predict that this variant is pathogenic giving a PP3 classification. A PM2 (1 Moderate) and a PP3 (1 supportive) according to the ACGS classification rules (Diagram from Ellard et al 2020) give this variant 3 points and a VUS temperature of Tepid.

Variant NM_000179.2(MSH6):c. 1885 G>T p.(Asp629Tyr)

This variant doesn't appear in gnomAD database and this correspond to a PM2 classification. According to Varosme ten pathogenic predictions from BayesDel_addAF, DANN, DEOGEN2, EIGEN, FATHMM-MKL, LIST-S2, M-CAP, MutationAssessor, MutationTaster and SIFT predict that this variant is pathogenic classifying it as PP3. According to the ACGS classification 1 Moderate (PM2) and 1 Supportive (PP3) evidence of pathogenicity classify this variant with 3 points and with a Tepid temperature.

Variant NM_000179.2(MSH6):c. 2054 G>A p.(Gly685Asp)

This variant doesn't appear in gnomAD database therefore its frequency is cero classifying it with PM2. No information was found in ClinVar. According to Varsome 12 in silico programs (BayesDel_addAF, DANN, DEOGEN2, EIGEN, FATHMM-MKL, LIST-S2, M-CAP, MVP, MutationAssessor, MutationTaster, PrimateAI and SIFT) predict that this variant is pathogenic

and classifying it as PP3. According to the ACGS classification 1 Moderate (PM2) and 1 Supportive (PP3) evidence of pathogenicity classify this variant with 3 points and with a Tepid temperature.

Variant NM_000179.2(MSH6):c. 2295 C>G p.(Cys765Trp)

This variant doesn't appear in gnomAD database and this correspond to a PM2 classification. In silico models agree that this variant is deleterious as well algorithms developed for the *MSH6* gene suggests that the change is likely to be deleterious giving a PP3 classification. According to the ACGS classification 1 Moderate (PM2) and 1 Supportive (PP3) evidence of pathogenicity classify this variant with 3 points and with a Tepid temperature.

Variant NM_000179.2(MSH6):c.3656C>T p.(Thr1219Ile)

This variant doesn't appear in gnomAD database and this correspond to a PM2 classification. According with Houlleberghs Hellen et al, 2017 (40) this variant abrogates MMR activity comparing with 5 pathogenic controls. Likewise this alteration has demonstrated to result in less than 25% mismatch repair efficiency in comparison with the wild type protein by two independent functional studies (Drost M et al, 2012 (41); Geng H et al, 2012, (42)) with all this information it was classified as PS3_Moderate. This alteration is predicted to be deleterious by in silico analysis. In addition, the CoDP in silico tool predicts this alteration is likely to impair molecular function giving a PP3 classification. According to the ACGS classification 2 Moderate (PM2, PS3_Moderate) and 1 Supportive (PP3) evidence of pathogenicity classify this variant with 5 points and with a Hot temperature.

Table 10. *MSH6* VUS variants classification

Variant (Reference Sequence)	gnomAD Mean Allele Frequency/Highest population Frequency	ClinVar	ACGS classification rules	VUS temperature
NM_000179.2(MSH6):c.234A>G p.(Arg78=)	Frequency cero	Algorithms developed to predict the effect of sequence changes on RNA splicing suggest that this variant may create or strengthen a splice site, but this prediction has not been	**PM2 + PP3**	**3 points = Tepid**

		confirmed by published transcriptional studies.		
NM_000179.2(MSH6):c. 1137 A>C p.(Arg379Ser)	Frequency cero	Information found with the version number NM_000179.3 Interpretation: Uncertain significance	**PM2 + PP3**	**3 points = Tepid**
NM_000179.2(MSH6):c. 1885 G>T p.(Asp629Tyr)	Frequency cero	Information found with the version number NM_000179.3 Interpretation: Uncertain significance	**PM2 + PP3**	**3 POINTS = TEPID**
NM_000179.2(MSH6):c. 2054 G>A p.(Gly685Asp)	Frequency cero	No Information Found	**PM2 + PP3**	**3 Points = Tepid**
NM_000179.2(MSH6):c. 2295 C>G p.(Cys765Trp)	Frequency cero	Interpretation: Conflicting interpretation of pathogenicity Likely pathogenic (1); Uncertain significance (1)	**PP3 + PM2**	**3 POINTS = TEPID**
NM_000179.2(MSH6):c.3656C>T p.(Thr1219Ile)	Frequency cero	1x Likely pathogenic: 1x Uncertain significance Conflicting	**PM2 + PP3 + PS3_Moderate**	**5 points = Hot**

The Tepid variants in the MSH6 protein appear in four functional domains: DNA binding, MSH2 interaction, Connector and Lever domains (Figure 6). The variant p.Arg379Ser appears in the DNA binding and in the MSH2 interaction domain, the variants p.Asp629Ser and the p.Cys765Trp variant appears in the Lever domain. The hot variant p.Gly685Asp appears in the Lever domain and p.Thr1219Ile appears in the ATPase domain.

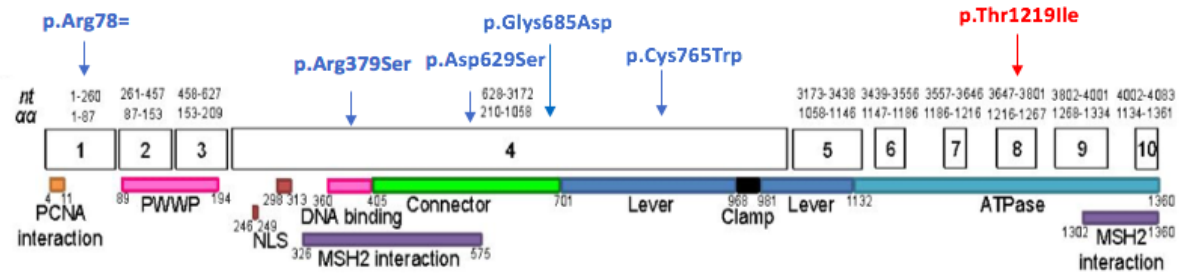


Figure 6. Diagram representing the Linear schematic of MSH6 mismatch repair gene structure and functional domains. The different MSH6 VUS variants classified as the most probably damaging ones are indicated above. Variants with a Tepid temperature appear in color blue, Hot variants are presented in red. Adapted from (39)

3.2 Functional assays

A total of 26 articles were found in PubMed using the criteria outlined in section 2.5 in the Materials and Methods section. From these articles the followingstudies: Olilla Saara, et al 2006 (48); Nyström-Lahti Minna, et al 2002 (49), Kariola Reetta, et al, 2002 (50) Kantelinen Jukka et al, 2011 (51), AR Ellison et al 2001 (45), Shimodaira H et al (46), Cervelli Tiziana et al 2020 (43), Drost M et al 2020 (47), Drost M et al 2019 (26), Cliaj N et al, 2002 (52), Bouvet D. et al. 2019 (53), Jia X et al 2021 (54) Ellison AR et al 2001 (45) were used to make an overview of functional assays that might be suitable to evaluate the MSH6 and MLH1 VUS variants.

Different functional assays exist to assess the effect of MMR variants, and they may be divided into two groups. The first group corresponds to functional assays that evaluate a biochemical or biological function (MMR capacity) of just one or two MMR proteins independently. These kinds of functional assays measure different functions of an MMR protein such as:

i) The binding capacity of two MMR proteins to each other.
ii) The heterodimer binding to a mismatch DNA.
iii) The ATP ↔ ADP cycling
iv) The MMR cellular localization

The specific functional assays exist to measure the different biological or biochemical functions of the MMR genes, are described in more details in table 11.

The second group corresponds to functional assays that evaluate the MMR capacity as a whole. These assays are classified as follows:

i) Functional assays using yeast
ii) Functional assays using cell-free system
iii) Functional assays using human cell lines

Also, as in the second group specific functional assays exist to measure the repair capacity of the MMR system, a detailed description of the specific characteristics of these assays are shown in table 11.

Table 11. Classification of Mismatch Repair (MMR) Functional Assays

Classification of functional assays	Subclassification	Specific functional assay and Features
Group I: Tests for a specific biological or biochemical function of an MMR protein.	Measures the capacity of two MMR proteins binding to each other (heterodimer formation)	Yeast two-hybrid assay (in vivo assay) Glutathione S-transferase (GST) pull-down assay (in vitro assay) Transient expression of MMR genes in relevant MMR-deficient cell line (in vivo assay)
	Measures the capacity of two MMR proteins (Heterodimer binding) to recognized and bind to a mismatched DNA	DNA mobility shift assays (gel shift assay) (in vitro assay) Binding to immobilized DNA (1D) (in vitro assay on the surface of a biosensor)
	ATP ◄⟷► ADP cycling ATP assays	This assay test: 1.- ATP binding, 2.- ATP⟶►ADP exchange 3.- ATPase catalytic efficiency 4.- ATP induced dissociation from a mismatch 5.- Conformational change in the presence of ATP

		6.- Efficiency of loading multiple sliding clamps on to a circular DNA strand
	MMR protein subcellular localization	Localization experiments. These assays test the expression of fluorescent MMR proteins in mammalian cells to localize the distribution of these proteins in the cell.
Group II: Tests for MMR repair capacity as a complete process	Functional assays using yeast	Expression of mutant yeast MMR genes in haploid yeast strains. These assays test the expression of mutant human or corresponding yeast MMR in yeast strains to monitor the repair capacity as a whole (based on the homology of human and yeast MMR proteins).
	Functional assays using cell-free systems	Three types: 1.- Expression of mutant human MMR genes in haploid yeast strains 2.- Expression of mutant yeast MMR genes in diploid yeast strains 3.- Cell- free in vitro MMR assays. These three types of assays tests in vitro the repair of mismatched DNA by protein extracts. Mostly baculovirus infected insect cell extracts are used to complement MMR-deficient cell extract.
	Functional assays using human cell lines	Cell based in vitro MMR functional assay using a human expression system These kinds of assays test the expression of mutant human MMR genes in homologous human cell lines to monitor the repair capacity as a whole.

Adapted from (28)

3.2.1 Functional assays to evaluate MLH1 and MSH6 VUS variants

To evaluate the pathogenicity of the VUS variants in this study, functional assays that evaluate the MMR repair capacity as a complete process were prioritized since they measure the repair activity of the MMR system as a whole and indicate whether the MMR system can perform its main function, which is to repair mis-paired bases. The majority of the variants are located in protein domains that are critical for the overall repair capacity of the complex. For this reason

assays that measures the MMR capacity as a complete process (group 2 in table 11) were selected and described in further detail. Eventually one of these assays will be selected for implementation as an assay to evaluate the VUS variants described in the section 3.1

To evaluate the MMR capacity as complete process, the assays were selected according to this classification:

1) Functional assays using yeast
2) Functional assays using cell-free system
3) Functional assays using human cell lines

3.2.2 Evaluation of the MMR capacity as a complete process

3.2.2.1 Functional assays using yeast

To evaluate the MMR capacity as a complete process functional assays using yeast were made. In yeast assays human genes are introduced and expressed on a plasmid within the yeast. Another way of introducing the human genes involve replacing the yeast genomic copy with the human genomic copy. These types of approaches are used when there is homology of the human genes with the yeast as is the case of the genes of the MMR system.

In yeast-type assays, the mutation rate of the variant is measured by comparing it with the mutation rate and the mutant phenotype of the Wild Type (WT) strain. One way to measure the effect of the mutator phenotype in for MMR genes is to use forward and reverse mutation assays (Figure 7). In forward mutation assays the mutation frequency is evaluated by counting the number of colonies resistant to Canavine (CAN1 to can1) (Figure 7 A) or 5-fluoroorotic acid (5-FOA, URA3 to ura3) (Figure 7 B). In the URA3 gene using the URA3- forward mutation assay (44), G-T mispairs are in-frame inserted in the 5'end of the gene. If the insertion is not repaired by the MMR system the frequency of mutation increases and therefore there is a change from URA3 to ura3. The mutation rate can be observed by the mutant phenotype of the yeast colonies. Colonies with a pathogenic variant will show a high mutation rate and will become into 5-FOA resistant colonies (Figure 7 B). The mutation rate in the reverse mutation assays is measured by counting the number of yeast colonies capable of growing on different selective media such as medium lacking threonine (hom3-10 to HOM3) colonies becoming blue (lacZ to LACZ), colonies becoming green (gfp to GFP), colonies becoming white in medium lacking adenine (ade2 to ADE2) colonies becoming able to grow in medium lacking

lysine (lys2 to LYS2) (Figure 7 G) and colonies becoming able to grow in medium lacking histidine (hys7.2 to HIS7) (Figure 7 H). One way to assess the pathogenicity of variants using reverse mutation assays is by comparing the mutator phenotype of each variant with the human MMR WT (hMMRWT). The hMMRWT confers a mutant phenotype in yeast (45), depending on the assays different phenotypes will be present in the yeast colonies. Depending of the assay used the colonies of hMMRWT will have a blue color in case of LACZ assays or green color in GFP assays or white color using ADE2 assays. The conferred mutant phenotype of the hMMRWT is suppressed when a pathogenic variant is expressed (45) giving a different phenotype to the colonies white color in the case of lacz and gfp or red color in the case of ade2 indicating that the wild type mutant phenotype was suppressed. The reverse mutations are represented in (Figure 7 D-H).

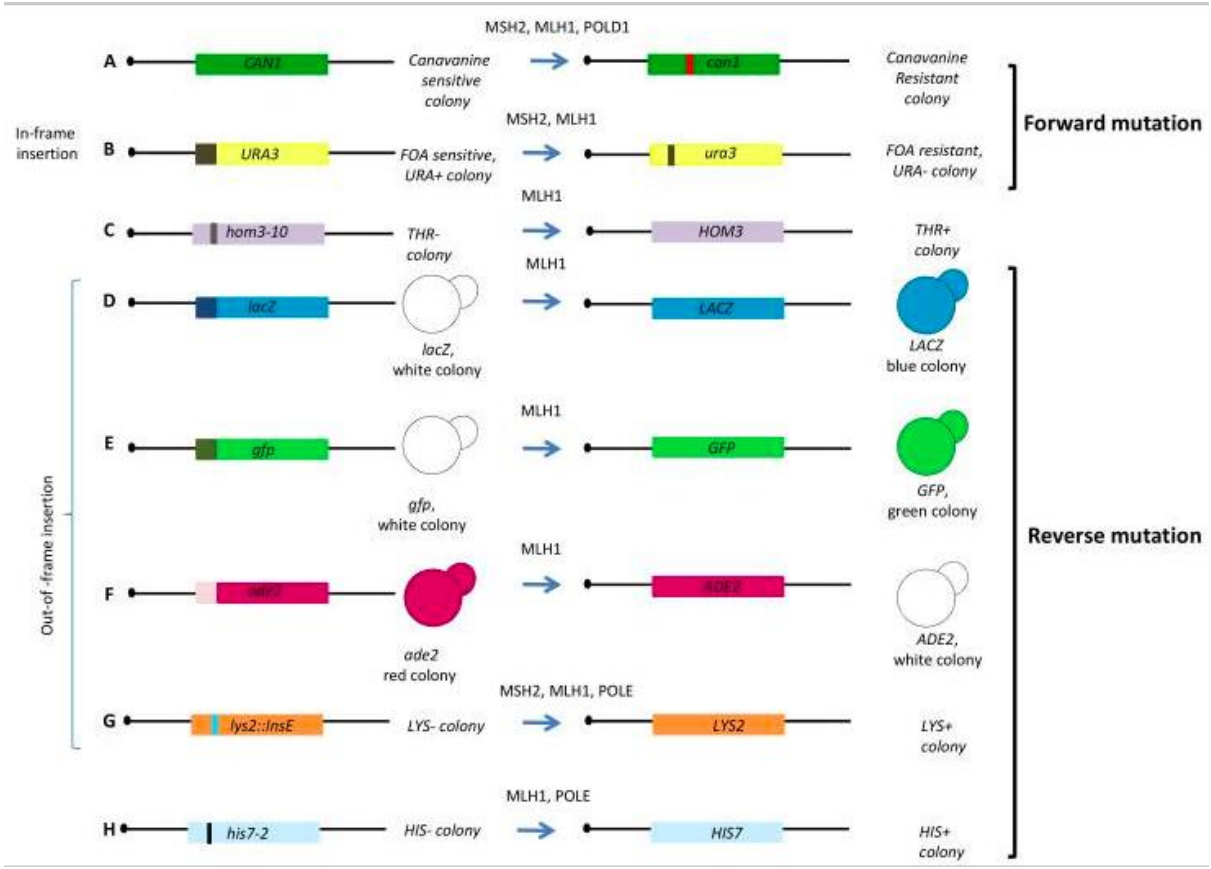


Figure 7. Schematic representation of the yeast functional assays for MMR genes. Forward mutation frequency/rate is evaluated by scoring the number of colonies becoming resistant to Canavanine (CAN1 to can1) (A) or 5-fluoroorotic acid (5-FOA, URA3 to ura3) (B). Reverse mutation is assessed by counting the number of colonies becoming able to grow in selective medium lacking threonine (THR+; hom3-10 to HOM3) (C). Colonies becoming blue (lacz to LACZ) (D), colonies becoming green (gfp to GFP) (E), colonies becoming white in medium lacking adenine (ADE+; ade2 to ADE2) **(**F), colonies becoming able to grow in medium lacking lysine (LYS+; ly2 to LYS2) (G) and colonies becoming able to grow in medium lacking histidine (HIS+ hys7-2 to HIS7) (H). The URA3 gene used for forward reversion (B) contains an in-frame insertion of several nucleotides, therefore the gene is WT. Constructs of D,E,F and G contain out-of-frame insertions of several nucleotides, therefore the gene is mutated. Constructs of C and H contain a point mutation. Above the arrow are given the

names of the human proteins studied with the assay. Cervelli Tiziana et al., 2020 (43). doi: 10.15698/mic2020.07.721

3.2.2.2 Functional assays using cell-free system

CIMRA Assay

Another way to evaluate the MMR capacity as a whole process and as well the pathogenicity of the VUS variants is the biochemical test named cell-free in vitro MMR activity (CIMRA) that was carried out by Drost M. et al, 2019 (46), Drost M. et al, 2020 (47).

The overall principle of the CIMRA functional assay works as follows:

1. cDNA for the MMR protein that should be analyzed is used as template to recreate the allelic variant(s) to be studied (e.g. by PCR using mutagenic oligonucleotides).
2. The mutant MMR protein is expressed from the cDNA *in vitro*.
3. The mutant protein is added to a nuclear cell extract (cell line deficient of the MMR protein to be analyzed) together with a fluorescent mis-match substrate (e.g. containing a G-T mispair)
4. The mixture is incubated for the reparative reaction of the protein to take place.
5. The MMR repair capacity of the MMR mutated protein is measured using fluorescent analysis.

The principle of the CIMRA assay to evaluate the MMR capacity is shown in Figure 8.

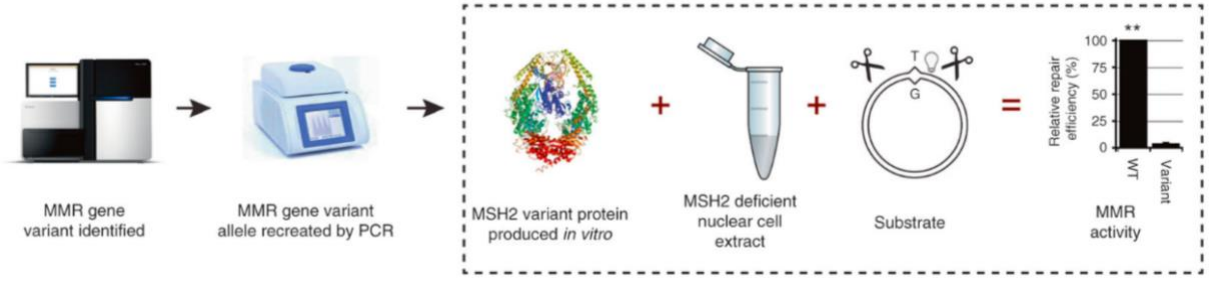


Figure 8. Cell-free in vitro MMR activity (CIMRA) assay. The figure describes testing of an *MSH2* variant. However, in the same way *MLH1* variants could be tested. Genetics in Medicine, 2019 (46). DOI: 10.1038/s41436-018-0372-2

Many functional assays are based on the principle of cell-free in vitro MMR activity (CIMRA) Some studies (Olilla Saara, et al 2006 (48); Nyström-Lahti Minna, et al 2002 (49), Kariola Reetta, et al, 2002 (50) and Kantelinen Jukka et al, 2011 (51) have used this methodology to evaluate MMR VUS variants.

3.2.2.3 Functional assays using cell lines

Methylation Tolerance-Based Functional Assay.

The Methylation Tolerance-Based Functional Assay can be used to evaluate the fidelity of the DNA repair that MMR system carries out. This assay is based on the function of the MMR system to induce apoptosis caused by DNA damage induced for example by methylating agents. The methylation Tolerance-Based Functional Assay methylate DNA at different positions creating a critical O^6 – methylguanine (O^6-MeG) lesion that causes miscoding when the site is replicated by DNA polymerase, creating the mismatch O^6-MeG-T. Since the methylated base remains in the DNA the repair synthesis will fail because of the inability to find a correctly matching nucleotide leading the cell to undergo apoptosis (52).

The Methylation Tolerance-Based Functional Assay (MT Assay) was carried out by Bouvet Delphine et al., 2019 (53). In this assay they assessed the response to the cytotoxic effect of the methylating agent N-Methyl-N'-Nitro-N-Nitrosoguanidine (MNNG). Before carrying out the Methylation Tolerance-Based Functional Assay (MT Assay) the *MLH1* variants were generated by site-directed mutagenesis.

Principle of the MT assay:

The variants were introduced into *MLH1*-null human colorectal cancer cells by transient transfection. The transfected cells that express the reporter gene mCherry present in the expression vector were selected, reseeded and exposed to the methylating agent MNNG. The MMR activity was evaluated by a clonogenic (colony formation) assay. The cells that resist the cytotoxic effect of MNNG grew into resistant colonies and were considered as pathogenic variants. Conversely, the cells that don't resist the MNNG effect didn't grow and didn't form a colony or very few colonies were able to grow in the presence of MNNG for this reason they were considered as non-pathogenic variants.

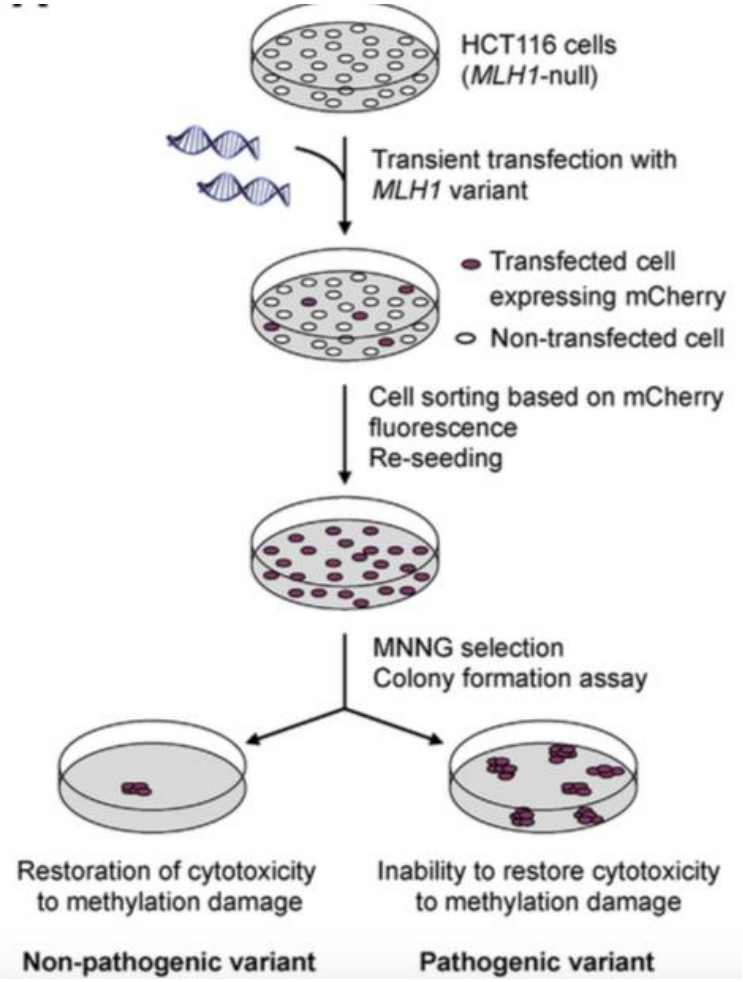


Figure 9. Schematic diagram representing the principle of the MT assay.

Gastroenterology 2019 (53). DOI: 10.1053/j.gastro.2019.03.071

Chemical selection for MMR dysfunction and deep sequencing.

These assays are based on chemical selection for mismatch repair dysfunction using 6-thioguanine (6-TG) and deep sequencing to identify the variants that are resistant and survive the exposure of 6-TG and with this being able to classify them into neutral, pathogenic or benign variants.

Jia Xiaoyan et al., 2021 (54) utilized this assay to evaluate MSH2 missense variants, but the assay can be extended to evaluate variants in other key Lynch syndrome genes such as *MLH1* and *MSH6.* The principle of this assay is described below and represented in figure 10.

Overall principle of this assay:

1. MMR knockout cell lines and MMR missense variants libraries were made and packaged into lentivirus (lentiviral packaging).

2. The corresponding Wildtype (WT) of the MMR gene is used as a functional control. And known pathogenic variants for the MMR gene were used as positive controls.

3. Both controls (negative and functional) and the missense library were transduced into the knockout cell lines and then will be exposed to the cytotoxic agent (6-TG).

4. After the exposure with the cytotoxic agent the variants resistant to the agent will be select for deep sequencing.

5. Deep sequencing will be performed before and after the treatment with 6-TG.

6. A Loss of Function [LOF] score will be made for each missense variant. Positive LOF scores will indicate a deleterious variant and negative LOF scores will be indicate a benign variant.

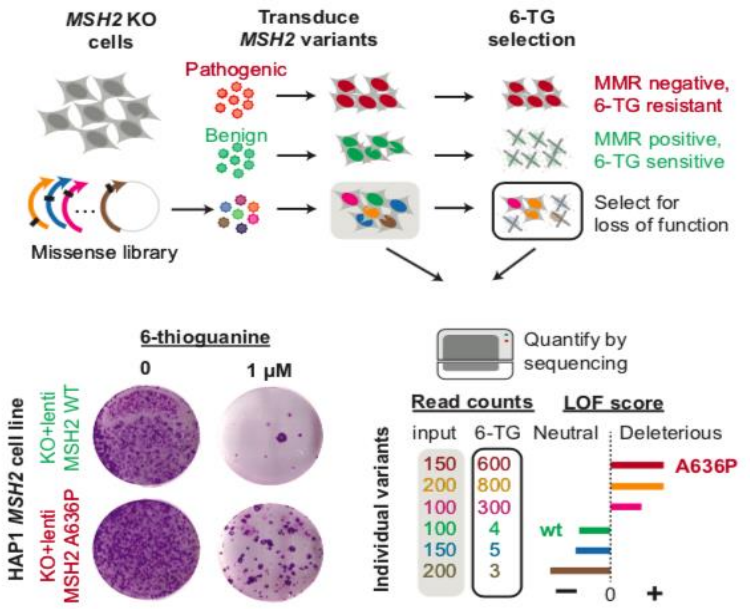


Figure 10. Overview of MSH2 functional screen. The figure describes the steps to find deleterious and neutral LOF scores by deep sequencing. The American Journal of Human Genetics. 2021 (54)

DOI: 10.1016/j.ajhg.2020.12.003

4 DISCUSSION

4.1 Overview

In this project, 27 MLH1 and 52 MSH6 VUS variants of the Mismatch Repair System (MMR) from St. Olav's Hospital were evaluated. Each VUS variant of each gene was investigated in different databases such as gnomAD, ClinVar, VarSome, VarMap to enlighten the potential consequences of pathogenicity of each variant and therefore could generate Lynch syndrome. Six possible pathogenic VUS variants have been identified in each gene.

As part of this thesis, an overview of the best functional assays found in literature which could be used to evaluate the possible pathogenic variants was made. One of the most important aspects to select these functional assays were mainly based on the evaluation of how the most likely pathogenic variants found in this work could disrupt the repair activity of MMR and the pathways that are essential for maintaining DNA fidelity.

In the following discussion it will be argued why the variants present in the paragraphs below in this Discussion part of the thesis are the most probably pathogenic variants found in this project. These variants were predicted as pathogenic by in silico tools, prediction programs and also mentioned and shown as pathogenic in different scientific articles in the genes *MLH1* and *MSH6* that could affect the Mismatch Repair function and activity. Another aspect to be argued is which functional assay would be the best to evaluate the VUS variants that are predicted to be the most pathogenic ones. A comparison of each functional assay will also be made and as well as the benefits and limitations of each one of them.

4.2 Most probable pathogenic variants found in MLH1 protein.

MLH1 p.Glu37Ala

This variant is predicted to changes Glutamate to an Alanine at coding position 37 within an alfa strand motif in two functional domains the ATPase domain and the MutSα interaction. Glutamate is negatively charged and hydrophilic. Alanine is an aliphatic residue, non-polar, small fairly non-reactive and hydrophobic. The different properties of the two amino acids make them poor substitutes for each other. For this reason, this change is considered as a big change. The position is highly conserved meaning that it's very likely to be important for the protein's function. The variant has not been published, but another variant p.Glu37Lys located

in the same codon has been classified as class 5 by two study reports (33, (35). In the (33) study this variant was classified as pathogenic because decrease the MMR activity. The Grantham's distance between Glutamate and Lysine is 56, and the Grantham's distance between Glutamate and Alanine is 107, meaning that the difference is two-fold higher between the substituted amino acids in the present study compared to the pathogenic variant p.Glu37Ala. This is a medium score in the distance of change of amino acids according to Grantham's distance for the variant p.Glu37Ala indicating a possibly deleterious change. As well as this variant is not been reported in gnomAD database indicating that the variant is not commonly present in any population meaning that this variant is probably a rare variant. With all these evidences the probability that this variant would be a pathogenic variant is quite high.

MLH1 p.Arg265His

The variant is predicted to change an Arginine to a Histidine at coding position 265 within a Beta strand motif in the MutSα interaction functional domain. Arginine has a positive charge and is hydrophilic. Histidine as well present positive charge and is hydrophilic. The change from an Arginine to a Histidine is not a big change. The Grantham's distance between Arginine and Histidine is 29. However, the residue Arginine at this position is highly conserved. So, it's very likely to be important for the protein's function. The variant presents another 3 variants in the same codon position p.Arg265Ser, p.Arg265Cys and p.Arg265Pro has been classified as class 5 by six studies. The variant p.(Arg265Ser) and p.(Arg265Cys) decreased mismatch repair activity (33, 35). According to (36) (37) both variants are considered pathogenic because increase exon 10 skipping leading to a premature stop codon in exon 11 and degradation of the aberrant transcript also the skipping of exon 10 cause a drastic loss of MLH1 protein. As well as, the variant p.(Arg265Ser) presents a loss of activity of the mismatch system (35). For the variant p.Arg265Pro the study (38) reports this variant as pathogenic. Eleven in silico tools predict the change to be pathogenic. However, just one study (55) contradicts the possible pathogenicity of the variant p.Arg265His, they state that this variant is MMR proficient after doing an in vitro MMR activity. With all these evidences the probability that this variant would be a pathogenic variant is high.

MLH1 p.Tyr283Cys

The variant is predicted to changes a Tyrosine to a Cysteine at the codon position 283. This variant is located in the MutSα interaction domain. Tyrosine is aromatic partially hydrophobic,

large and prefers to reside in hydrophobic cores. It contains a reactive hydroxyl group that interacts with non-carbon atoms. Cysteine is small, polar and highly dependent on cellular localization. The residue at the position 283 is well conserved. So, it's very likely to be important for the protein's function. Eleven in silico programs predict the change to be disruptive. The Grantham's distance between the residues is 192. This is a significantly high score indicating a deleterious change. This variant doesn't appear in gnomAD database indicating that this variant is a rare variant. However, this variant has been found in genetic testing with a total of 8 submission to ClinVar. Therefore, this can indicate that this variant it may be pathogenic. However, non-publications in the literature nor functional impact by in vivo/vitro studies has been found to support the pathogenicity of this variant, meaning that their clinical significance is still uncertain. With all these evidences the probability that this variant would be a pathogenic variant it could be medium high.

4.3 Most probable pathogenic variants found in *MSH6* protein

MSH6 p.Arg379Ser

The variant is predicted to change an Arginine to a Serine in the conserved 379 coding position. The variant is located within two important functional domains in the DNA binding domain and the MSH2 interaction domain of the MSH6 protein. For the stabilization of the MSH6 protein is necessarily to be able to bind to the MSH2 protein. Otherwise MSH6 will be destabilized, as well as for the MMR system to work properly MSH2 must bind correctly to MSH6 (56). Likewise, the DNA binding domain in MSH6 is important for efficient MMR activity (56). Serine is small and neutral. Arginine is a large polar and often positively charged. For this reason, the change from Arginine to Serine is very big and might result in a change to the protein's function. The Arginine residue is highly conserved so it's very likely to be important for the protein's function. This variant doesn't appear in gnomAD database indicating that is a rare variant. Nine in silico tools predict that this change is pathogenic. The Grantham's distance between these two amino acids is 110. This is a high score indicating that this variant could have a deleterious change and be pathogenic. However, this variant has not been reported in individuals with MSH6-related conditions. With all these evidences the probability that this variant would be a pathogenic variant is quite high.

MSH6 p.Cys765Trp

This variant is predicted to change a Cysteine in codon 765 to a Tryptophan. Cysteine is tiny, polar, neutral and has a side chain capable of forming a disulphide bond with another cysteine providing a strong structural support for the protein. Tryptophan is large and has an aromatic side chain. For this reason, the change from a Cysteine to a Tryptophan is a large one and might result in a change to the protein's function. Also, the Cysteine at this position is conserved. So, it's very likely to be important for the protein's function. This variant is located in the lever domain of the MSH6 protein. One study investigating the functional effects of MSH6 missense mutations found that mutations within the lever domain affect the proper coordination of DNA binding and nucleotide processing indicating that signaling between domains is disrupted (57). This variant doesn't appear in gnomAD database indicating that this variant is a very rare variant. In silico tools and algorithms developed for the gene MSH6 suggest that the change is deleterious. With all these evidences the probability that this variant would be a pathogenic variant is high.

MSH6 p.Thr1219Ile

The variant is predicted to change Threonine in codon 1219 to an Isoleucine within the ATPase domain of MSH6. The position of this variant in the ATPase domain is a conserved position in the protein. Threonine is small, neutral and polar. It forms hydrogen bonds with polar substrates and is quite common in function centres. Isoleucine has an aliphatic side chain and is hydrophobic. This position is conserved and the difference in structure and properties of the residues could distort the protein folding or function. Three different studies said that this variant reduced the mismatch repair activity and efficiency (37, 38, 39). This variant is not been reported in the gnomAD database indicating that the variant is not commonly present in the allele frequency of any population meaning that this variant is probably a rare variant among populations and probably pathogenic. In silico tools predict that this variant is deleterious. With all these evidences the probability that this variant would be a pathogenic variant is high.

4.4 Functional assays

The genes of the MMR system underlying Lynch syndrome are among the most screened genes in the clinical setting. Therefore, there is a need to interpret the variants in these genes that

could be pathogenic. For this reason, functional assays are very useful to evaluate and to give a classification of them.

Different functional studies exist to evaluate the MMR variants. In this thesis functional assays that evaluate the MMR capacity as a complete process were chosen, since the variants that were probably pathogenic and were found in this project most likely would be critical for the overall repair capacity as a whole complex. The functional studies that evaluate the MMR capacity as a whole process involves three different methodological approaches: 1) Functional assays using yeast, 2) Functional assays using cell-free system and 3) Functional assays using human cell lines.

Functional assays using yeast have been widely used since the MMR system is evolutionarily conserved in all eukaryotic organisms. The easy genetic manipulation and the conservation of cellular functions in yeast and mammals are a great advantage when using these types of organisms to evaluate the pathogenicity of cancer-related genes. Likewise, another advantage of these assays is that they make it possible to analyze a large number of variants could be possible. However, a big downside and limitation of using these assays is that not every human MMR factor complements it yeast ortholog, and many human variants occur outside of the conserved domains and they cannot be analyzed. Just the mutations that appeared in conserved domain regions between humans and yeasts could be analyzed.

The second type of functional assays that evaluate the MMR repair capacity as a complete process are cell-free systems. These assays allow the identification of pathogenic variants. An important upside of these kind of studies is that the assays are fast. A second advantage of these assays is that they can evaluate many mutants at the same time and in parallel. One big disadvantage of this type of assays is that they don't evaluate specific MMR defects such as splicing, intracellular location, etc. As a complement to these assays, other assays assessing residual activity should be useful for a final diagnosis of VUS. For example, the CIMRA assay can efficiently evaluate and predict exactly which variants are pathogenic and which ones are benign within the MMR system.

The third and last type of assays evaluated in this thesis were assays involving human cell lines. These assays involved the use of chemical agents such as MNNG and 6-TG, both of which are considered methylating agents. The MMR system is involved in two pathways that are essential for maintaining DNA fidelity: *i*) correction of replication errors that escape proofreading of

DNA polymerase and *ii*) MMR-dependent cell death following specific DNA damage induced by methylating agents. A defective MMR system will lead to an increased range of cellular tolerance to DNA lesions that are induced by methylating agents. This phenomenon is known as methylation tolerance. This type of assay can evaluate the pathogenicity of the VUS variants and could discriminate which VUS are harmful and which are neutral or benign depending on which colonies can resist or not to the exposure of these chemical agents.

The functional assay using MNNG as a cytotoxic agent has 95% sensitivity and 100% specificity when discriminating and classifying MMR variants as pathogenic or neutral. This type of assay is simple, reliable and fast to develop. Two other advantages offered by this type of assay is that the material derived from the patient is not necessarily required because the variants analyzed could be obtained from cell lines.

The second assay using human cell lines utilized 6-TG as a cytotoxic chemical agent and a method of deep mutational scanning named massively parallel sequencing. In this study a combination of this deep mutational scanning with an MMR assay was used to establish a big effect map of missense variants in this case using the gene MSH2. This approach could evaluate also the genes *MLH1, PMS2* and *MSH6.* Two big upsides of this assay are that the study can correct the misclassification done before from different clinical databases. The second upside is that this study has the potential to accurately classify MMR variants.

One of the limitations of this type of assays is in the measurement of LOF (Loss of Function) scores, since this type of measurement doesn't offer a direct measure of the rate of mutations associated with each variant. Another downside appears when you want to scale the approach to genes bigger than MSH2 this could require including different strategies such as landing pads that circumvent the size limitations of the lentiviral packaging.

Among the limitations of these two types of assays using methylating agents are that they can only evaluates one function of the MMR system, which is DNA damage inducing cell death. The first pathway involved in the correction of replication pathway is missed in this assay.

With all this information mentioned in the previous paragraphs and considering the advantages and disadvantages of the different types of functional assays, the CIMRA assay is considered to be the best functional assay to evaluate the most probable pathogenic VUS variants found in

this project. All the functional assays can evaluate and give a good classification on the pathogenicity of the different variants. However, the CIMRA assay can effectively evaluate the reparative activity of each variant in a very effective way just as all the different variants to be analyzed can be efficiently generated. With this method many variants can be analyzed at the same time and in parallel.

The functional assays using yeast could be a fast method, and the yeast can be genetically manipulated in a simple way. However, the experimental model to evaluate the variants needs to mimics as closely as possible the human organism. Not every human MMR factor complements its yeast orthologs and this could be a big disadvantage when using this type of assays. For this reason, this assay couldn't be the best option to evaluate all VUS variants. Now, considering the functional assays using cytotoxic agents (MNNG and 6-TG) their advantages are that they could evaluate the pathogenicity of the VUS variants efficiently in a general way, they are also relatively fast and simple methods to perform in the laboratory and they use human cell lines. However, it only analyzes the second pathway of the DNA fidelity of the MMR system and it does not indicate if a mismatch base pair could repair such as the case of evaluation the repairment of the mismatch bases G.T to T.A that can be observed, evaluated, analyzed and quantified in the CIMRA test.
For all these reasons the best method to evaluate the most pathogenic VUS variants in this project is the CIMRA method.

To be able to define which specific part of the MMR system is altered, more specific assays have to be used in combination and to complement the CIMRA assay. For instance, ATP assays can be performed since one hot variant were found within the ATPase functional domain in both genes.

5. CONCLUSION

The purpose of this study was to evaluate and analyze the VUS variants using different databases and identify the most probably pathogenic VUS variants within the cohort as well as describe the best functional assays to evaluate these variants. A total of 12 variants were identified as the most probably pathogenic variants. The variants MLH1, p.Glu37Ala, MSH6, p.Gly685Asp and MSH6, p.Thr1219Ile were identified as the three Hot variants with highest probabilities of pathogenicity of the 12 potential pathogenic VUS variants. To investigate the

pathogenicity of these variants a research in literature of functional assays were made. From the literature study it was concluded that due to its efficiency, and for its usefulness to evaluate, analyze and classify MMR missense VUS variants the CIMRA method would be the best functional assay to evaluate the VUS variants.

6 REFERENCES:

1. Institute NC. Cancer Dictionary: National Cancer Institute; 1937 [updated April 6, 2018. The National Cancer Institute (NCI) provides access to all individuals seeking information on www.cancer.gov. The NCI is the federal government's principal agncy for cancer research and training.]. Available from: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/cancer.
2. Tamura K, Kaneda M, Futagawa M, Takeshita M, Kim S, Nakama M, et al. Genetic and genomic basis of the mismatch repair system involved in Lynch syndrome. Int J Clin Oncol. 2019;24(9):999-1011.
3. Lee EY, Muller WJ. Oncogenes and tumor suppressor genes. Cold Spring Harb Perspect Biol. 2010;2(10):a003236.
4. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. Lancet. 2019;394(10207):1467-80.
5. Kolios G, Moodley Y. Introduction to stem cells and regenerative medicine. Respiration. 2013;85(1):3-10.
6. Chung DC. The genetic basis of colorectal cancer: insights into critical pathways of tumorigenesis. Gastroenterology. 2000;119(3):854-65.
7. Mármol I, Sánchez-de-Diego C, Pradilla Dieste A, Cerrada E, Rodriguez Yoldi MJ. Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer. Int J Mol Sci. 2017;18(1).
8. Cox VL, Saeed Bamashmos AA, Foo WC, Gupta S, Yedururi S, Garg N, et al. Lynch Syndrome: Genomics Update and Imaging Review. Radiographics. 2018;38(2):483-99.
9. Rahner N, Steinke V. Hereditary cancer syndromes. Dtsch Arztebl Int. 2008;105(41):706-14.
10. Byrne RM, Tsikitis VL. Colorectal polyposis and inherited colorectal cancer syndromes. Ann Gastroenterol. 2018;31(1):24-34.
11. Carethers JM, Stoffel EM. Lynch syndrome and Lynch syndrome mimics: The growing complex landscape of hereditary colon cancer. World J Gastroenterol. 2015;21(31):9253-61.
12. Steinke V, Engel C, Büttner R, Schackert HK, Schmiegel WH, Propping P. Hereditary nonpolyposis colorectal cancer (HNPCC)/Lynch syndrome. Dtsch Arztebl Int. 2013;110(3):32-8.
13. Idos G, Valle L. Lynch Syndrome. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Mirzaa G, et al., editors. GeneReviews(®). Seattle (WA): University of Washington, Seattle

14. Aggarwal N, Donald ND, Malik S, Selvendran SS, McPhail MJ, Monahan KJ. The Association of Low-Penetrance Variants in DNA Repair Genes with Colorectal Cancer: A Systematic Review and Meta-Analysis. Clin Transl Gastroenterol. 2017;8(7):e109.
15. Seifert BA, McGlaughon JL, Jackson SA, Ritter DI, Roberts ME, Schmidt RJ, et al. Determining the clinical validity of hereditary colorectal cancer and polyposis susceptibility genes using the Clinical Genome Resource Clinical Validity Framework. Genet Med. 2019;21(7):1507-16.
16. Dominguez-Valentin M, Sampson JR, Seppälä TT, Ten Broeke SW, Plazzer JP, Nakken S, et al. Cancer risks by gene, age, and gender in 6350 carriers of pathogenic mismatch repair variants: findings from the Prospective Lynch Syndrome Database. Genet Med. 2020;22(1):15-25.
17. Strafford JC. Genetic testing for lynch syndrome, an inherited cancer of the bowel, endometrium, and ovary. Rev Obstet Gynecol. 2012;5(1):42-9.
18. Morice P, Leary A, Creutzberg C, Abu-Rustum N, Darai E. Endometrial cancer. Lancet. 2016;387(10023):1094-108.
19. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Rüschoff J, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. J Natl Cancer Inst. 2004;96(4):261-8.
20. Lynch HT, Snyder CL, Shaw TG, Heinen CD, Hitchins MP. Milestones of Lynch syndrome: 1895-2015. Nat Rev Cancer. 2015;15(3):181-94.
21. Tutlewska K, Lubinski J, Kurzawski G. Germline deletions in the EPCAM gene as a cause of Lynch syndrome - literature review. Hered Cancer Clin Pract. 2013;11(1):9.
22. Valle L, Vilar E, Tavtigian SV, Stoffel EM. Genetic predisposition to colorectal cancer: syndromes, genes, classification of genetic variants and implications for precision medicine. J Pathol. 2019;247(5):574-88.
23. Dominguez-Valentin M, Sampson JR, Seppälä TT, Ten Broeke SW, Plazzer JP, Nakken S, et al. Correction: Cancer risks by gene, age, and gender in 6350 carriers of pathogenic mismatch repair variants: findings from the Prospective Lynch Syndrome Database. Genet Med. 2020;22(9):1569.
24. Hernâni-Eusébio J, Barbosa E. Phenotypic Heterogeneity by Germline Mismatch Repair Gene Defect in Lynch Syndrome Patients. Acta Med Port. 2016;29(10):587-96.
25. Moghadasi S, Eccles DM, Devilee P, Vreeswijk MP, van Asperen CJ. Classification and Clinical Management of Variants of Uncertain Significance in High Penetrance Cancer Predisposition Genes. Hum Mutat. 2016;37(4):331-6.
26. Sijmons RH, Greenblatt MS, Genuardi M. Gene variants of unknown clinical significance in Lynch syndrome. An introduction for clinicians. Fam Cancer. 2013;12(2):181-7.
27. Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat. 2008;29(11):1282-91.
28. Ou J, Niessen RC, Lützen A, Sijmons RH, Kleibeuker JH, de Wind N, et al. Functional analysis helps to clarify the clinical importance of unclassified variants in DNA mismatch repair genes. Hum Mutat. 2007;28(11):1047-54.
29. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-24.

30. Sian Ellard ELB, Alison Callaway, Ian Berry, Natalie Forrester, Clare Turnbull, Martina Owens, Diana M Eccles, Stephen Abbs, Richard Scott, Zandra C Deans, Tracy Lester, Jo Campbell, William G Newman, Simon Ramsden and Dominic J McMullan. ACGS Best Practice Guidelines for Variant Classification in Rare Disease 2020 Assocation for Clinical Genomic Science. 2020:32.
31. Tavtigian SV, Greenblatt MS, Harrison SM, Nussbaum RL, Prabhu SA, Boucher KM, et al. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. Genet Med. 2018;20(9):1054-60.
32. Grantham R. Amino acid difference formula to help explain protein evolution. Science. 1974;185(4154):862-4.
33. Drost M, Zonneveld J, van Dijk L, Morreau H, Tops CM, Vasen HF, et al. A cell-free assay for the functional analysis of variants of the mismatch repair protein MLH1. Hum Mutat. 2010;31(3):247-53.
34. Andersen SD, Liberti SE, Lützen A, Drost M, Bernstein I, Nilbert M, et al. Functional characterization of MLH1 missense variants identified in Lynch syndrome patients. Hum Mutat. 2012;33(12):1647-55.
35. Shirts BH, Konnick EQ, Upham S, Walsh T, Ranola JMO, Jacobson AL, et al. Using Somatic Mutations from Tumors to Classify Variants in Mismatch Repair Genes. Am J Hum Genet. 2018;103(1):19-29.
36. Soukarieh O, Gaildrat P, Hamieh M, Drouet A, Baert-Desurmont S, Frébourg T, et al. Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. PLoS Genet. 2016;12(1):e1005756.
37. Tournier I, Vezain M, Martins A, Charbonnier F, Baert-Desurmont S, Olschwang S, et al. A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. Hum Mutat. 2008;29(12):1412-24.
38. Köger N, Paulsen L, López-Kostner F, Della Valle A, Vaccaro CA, Palmero EI, et al. Evaluation of MLH1 variants of unclear significance. Genes Chromosomes Cancer. 2018;57(7):350-8.
39. Tricarico R, Kasela M, Mareni C, Thompson BA, Drouet A, Staderini L, et al. Assessment of the InSiGHT Interpretation Criteria for the Clinical Classification of 24 MLH1 and MSH2 Gene Variants. Hum Mutat. 2017;38(1):64-77.
40. Houlleberghs H, Goverde A, Lusseveld J, Dekker M, Bruno MJ, Menko FH, et al. Suspected Lynch syndrome associated MSH6 variants: A functional assay to determine their pathogenicity. PLoS Genet. 2017;13(5):e1006765.
41. 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):488-94.
42. Geng H, Sakato M, DeRocco V, Yamane K, Du C, Erie DA, et al. Biochemical analysis of the human mismatch repair proteins hMutSα MSH2(G674A)-MSH6 and MSH2-MSH6(T1219D). J Biol Chem. 2012;287(13):9777-91.
43. Cervelli T, Lodovichi S, Bellè F, Galli A. Yeast-based assays for the functional characterization of cancer-associated variants of human DNA repair genes. Microb Cell. 2020;7(7):162-74.
44. Ellison AR, Lofing J, Bitter GA. Functional analysis of human MLH1 and MSH2 missense variants and hybrid human-yeast MLH1 proteins in Saccharomyces cerevisiae. Hum Mol Genet. 2001;10(18):1889-900.

45. Shimodaira H, Filosi N, Shibata H, Suzuki T, Radice P, Kanamaru R, et al. Functional analysis of human MLH1 mutations in Saccharomyces cerevisiae. Nat Genet. 1998;19(4):384-9.
46. Drost M, Tiersma Y, Thompson BA, Frederiksen JH, Keijzers G, Glubb D, et al. A functional assay-based procedure to classify mismatch repair gene variants in Lynch syndrome. Genet Med. 2019;21(7):1486-96.
47. Drost M, Tiersma Y, Glubb D, Kathe S, van Hees S, Calléja F, et al. Two integrated and highly predictive functional analysis-based procedures for the classification of MSH6 variants in Lynch syndrome. Genet Med. 2020;22(5):847-56.
48. Ollila S, Sarantaus L, Kariola R, Chan P, Hampel H, Holinski-Feder E, et al. Pathogenicity of MSH2 missense mutations is typically associated with impaired repair capability of the mutated protein. Gastroenterology. 2006;131(5):1408-17.
49. Nyström-Lahti M, Perrera C, Räschle M, Panyushkina-Seiler E, Marra G, Curci A, et al. Functional analysis of MLH1 mutations linked to hereditary nonpolyposis colon cancer. Genes Chromosomes Cancer. 2002;33(2):160-7.
50. Kariola R, Raevaara TE, Lönnqvist KE, Nyström-Lahti M. Functional analysis of MSH6 mutations linked to kindreds with putative hereditary non-polyposis colorectal cancer syndrome. Hum Mol Genet. 2002;11(11):1303-10.
51. Kantelinen J, Hansen TV, Kansikas M, Krogh LN, Korhonen MK, Ollila S, et al. A putative Lynch syndrome family carrying MSH2 and MSH6 variants of uncertain significance-functional analysis reveals the pathogenic one. Fam Cancer. 2011;10(3):515-20.
52. Claij N, Te Riele H. Methylation tolerance in mismatch repair proficient cells with low MSH2 protein level. Oncogene. 2002;21(18):2873-9.
53. Bouvet D, Bodo S, Munier A, Guillerm E, Bertrand R, Colas C, et al. Methylation Tolerance-Based Functional Assay to Assess Variants of Unknown Significance in the MLH1 and MSH2 Genes and Identify Patients With Lynch Syndrome. Gastroenterology. 2019;157(2):421-31.
54. Jia X, Burugula BB, Chen V, Lemons RM, Jayakody S, Maksutova M, et al. Massively parallel functional testing of MSH2 missense variants conferring Lynch syndrome risk. Am J Hum Genet. 2021;108(1):163-75.
55. Trojan J, Zeuzem S, Randolph A, Hemmerle C, Brieger A, Raedle J, et al. Functional analysis of hMLH1 variants and HNPCC-related mutations using a human expression system. Gastroenterology. 2002;122(1):211-9.
56. Edelbrock MA, Kaliyaperumal S, Williams KJ. Structural, molecular and cellular functions of MSH2 and MSH6 during DNA mismatch repair, damage signaling and other noncanonical activities. Mutat Res. 2013;743-744:53-66.
57. Cyr JL, Heinen CD. Hereditary cancer-associated missense mutations in hMSH6 uncouple ATP hydrolysis from DNA mismatch binding. J Biol Chem. 2008;283(46):31641-8.

7 ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my supervisors. Wenche Sjursen and Frank Skorpen.

It has been an honour to learn from you. I am grateful for your support.

I would like to extend my special thanks to my colleagues and the faculty of medicine and health sciences at NTNU.

And I send my deepest thanks to my family and friends for their endless and unconditional support and love.

Appendices

Appendix 1

Number of Genomes and the number of Exomes of the populations used in the version 2.1.1 of gnomAD

Population Description	Genomes	Exomes	Total
African/African-American	4,359	8,128	12, 487
Latino/Admixed American	424	17,296	17,720
Ashkenazi Jewish	145	5,040	5,185
East Asian	780	9,197	9,977
Koreans	0	1,909	1,909
Japanese	0	76	76
Other East Asian	780	7,212	7,992
Finnish	1,738	10,824	12,562
Non-Finnish European	7,718	56,885	64,603
Bulgarian	0	1,335	1,335
Estonian	2,297	121	2,418
North-Western European	4,299	21,111	25,410
Southern European	53	5,752	5,805
Swedish	0	13,067	13,067
Other non-Finnish European	1,069	15,499	16,568
South Asian	0	15,308	15,308

Other (population not assigned)	544	3,070	3,614
Total	15,708	125,748	141,456

Appendix 2
MLH1 VUS Variant Classification

Variant (Reference Sequence)	**gnomAD Mean allele frequency/ Highest pop. Frequency**	**ClinVar**	**VarMap**	**VarSome**	**ACGS classification rules**	**INSIGHT MMR classifcation**	**VUS temperature**
NM_000249.3 (MLH1):c.-42C>G	Frequency cero (no information found)	Interpretation: Uncertain Significance	-		PM2	-	**Cool**
NM_000249.3(MLH1):c.-42C>T	0.00004951/ European non- Finnish 0.0001007 Allele Frequency is higher than 0.0001, numbers above or round this number are considered BS1 for CanVIG	1xLikely Pathogenic & 6xVUS Conflicting Morak Monika et al 2018 classify the variant as class 2 (Likely benign) bi allelic expression was found it.	-	-	BS3 + BS1	Class 1 Bi-allelic expression, shown by cDNA experiment	**Class 1 Benign**

NM_000249.3 (MLH1): c.-28 A>G	0.001665/ European Finnish 0.008160 European Finnish population Allele Frequency higher than 0.001, are considered BA1 for CanVIG	2 x Benign; 3x Likely Benign & 5 x Uncertain Significance Conflicting	-	Benign computational verdict based on 1 benign prediction from DANN vs no pathogenic predictions	BA1 Benign	-	Class 1

NM_000249 .3(MLH1): c.-7C>T	0.001248/ European Finnish 0.008042 considered BA1 for CanVIG	3x Benign & 1x Uncertain Significance Three other clinical diagnostic laboratories have submitted clinical-significance assessments for this variant without evidence for independent evaluation. They cited the variant as benign	-		BA1 + BP6 = Benign	-	**Class 1**
NM_000249 .3(MLH1):c. 51C>T p.(Asn17=)	0.00001061/ European non-Finnish 0.00002323	Interpretation: Likely Benign benign by multiple in silico algorithms lack of segregation with disease. lack of disease association in case-control studies.	-	-	BS4 + BP4 = Likely benign	-	**Class 2 Likely benign**

NM_000249.3(MLH1):c.65G>C p.(Gly22Ala)	0.0001308/ (Other) 0.0002769 considered BS1 for CanVIG	Interpretation: Benign In silico algorithms predicted that the variant doesn't alter the protein function	Wouldn't be expected to change the protein function no disease associated on no interaction with any ligans, metal DNA/RNAor prot-prot	-	BP4 + BS1	Multifactorial likelihood analysis probability <0.001 = Class 1 no pathogenic no clinical significance	**Class 2 Likely Benign**
NM_000249.3(MLH1):c.69A>G p.(Glu23=)	0.000007959 / (European non-Finnish) 0.00001761	5xLikely benign ; 1x Uncertain significance Conflicting The p.Glu23= variant is not expected to have clinical significance because it does not result in a change of amino acid and is not located in a known consensus splice site. In addition, in silico or computational prediction software programs (SpliceSiteFinder, MaxEntScan, NNSPLICE, GeneSplicer,) do not predict a difference in splicing.	-		BP7	-	**Ice cold**

NM_000249.3(MLH1):c.290 A>G p.(Tyr97Cys)	0.0002427/ (South Asian) 0.001927 considered BS1 for CanVIG	1x Likely benign; 4x Uncertain significance In silico analyses predict that this variant is probably damaging to protein structure and function	Located in MUTSα interaction and ATPase domain this residue interacts with a protein and is not located in any catalytic residue	No Information	BS1 + PP3	-	**Cold**
NM_000249.3(MLH1):c.318C>T p.(Ser106=)	0.000007955 / (South Asian) 0.00006533	Interpretation: Likely benign	N.I.	Synonymous variant (in transcript NM_000249.4), not predicted splicing (not found in scSNV), and the position is not strongly conserved (GERP++ rejected substitutions = 5.15 is less than 5.5).	BP7	-	**Ice cold**
NM_000249.3(MLH1):c.409G>A p.(Ala137Thr)	0.00002784/ (European non-Finnish) 0.00005276	Interpretation: Uncertain Significance	No known disease associated variant known natural variant Low disease propensity not located in any functional domain	Pathogenic computational verdict based on 7 pathogenic predictions	PP3	-	**Cold**

NM_000249.3(MLH1):c.595G>C p.(Glu199Gln)	0.00009197/ European non-Finnish 0.0001783 Allele Frequency BS1 for CanVIG	An experimental study has shown that this variant doesn't affect MLH1 mismatch repair activity or protein expression in vitro. (Masanobu Takahashi et al., 2007)	No known disease associated variant known natural variant not located in any functional domain	-	BS1 + BS3	-	**Class 1 Benign**
NM_000249.3(MLH1):c.739 T>G p.(Ser247Ala)	0.00003582/ (Other) 0.0001632 Allele Frequency considered BS1 for CanVIG	Information found with the version number: NM_000249.4 Interpretation: Uncertain Significance	Disease associated variant: Pro1 disease-associated: linked with disease: Hereditary nonpolyposis colorectal cancer located in DNA mismatch repair protein (MutSα interaction) The Ser residue at position 247 is very highly conserved, so it's very likely to be important for the protein's function	Alternative variant chr3: 37055984 T⇒C (Ser247Pro) is classified Pathogenic by UniProt Variants (and confirmed using ACMG).	PM5 + BS1	-	**Cool**

NM_000249.3(MLH1):c.945C>G p.(His315Gln)	0.00001768/ European (non-Finnish) 0.00003872	Interpretation: Uncertain Significance	Located in the DNA mismatch repair protein in the C-terminal domain The His residue at position 315 is very highly conserved. no interaction with a ligand, a metal, DNA/RNA or protein	Pathogenic computational verdict based on 9 pathogenic predictions	PP3	-	**Cold**
NM_000249.3(MLH1):c.1007G>A p.(Gly336Asp)	0.00001194/ (European non-Finnish) 0.00002641	Interpretation: Uncertain significance	Likely deleterious high propensity of disease not located in a functional domain	Pathogenic computational verdict based on 12 pathogenic predictions	PP3	-	**Cold**
NM_000249.3(MLH1):c.1040C>A p.(Thr347Asn)	0.00007275/ (European non-Finnish) 0.0001384 considered BS1 for CanVIG	Information found with the version number NM_000249.4 Interpretation: Benign Multifactorial likelihood analysis posterior pathogenicity probability <0.001	Low disease propensity not located in any functional domain	10 pathogenic predictions	PP3 + BS1	No pathogenic/ no clinical significance	**Cold**
NM_000249.3(MLH1):c.1136A>G p.(Tyr379Cys)	0.00005307/ Latino 0.0001411 Latino Allele Frequency considered	Interpretation: Likely benign Multifactorial likelihood analysis	Not located in any functional domain	Pathogenic computational verdict based on 10 pathogenic predictions	PP3 + BS1	Likely not pathogenic/little clinical significance	**Cold**

	BS1 for CanVIG	posterior probability <0.05 (0.005) In silico analyses predict a deleterious effect					
NM_000249.3(MLH1):c.1409+23C>T	0.000004004 / (Other) 0.0001650 Allele Frequency considered BS1 for CanVIG	No information found		Benign computational verdict based on 1 benign prediction	BS1	- -	**Ice Cold**
NM_000249.3(MLH1):c.1490G>A p.(Arg497Gln)	0.00002784/ South Asian 0.0001307 South Asian Allele Frequency considered BS1 for CanVIG	Interpretation: Uncertain significance In silico tools supports a deleterious effect	Variant no known disease-associated variants high disease propensity possibly deleterious doesn't interact with any protein, ligand, metal or DNA/RNA		PP3 + BS1	- -	**Cold**
NM_000249.3(MLH1):c.1669 G>A p.(Glu557Lys)	Frequency cero	Interpretation: Uncertain significance Algorithms developed to predict the effect of missense changes on protein structure and function (SIFT, PolyPhen-2, Align-	located in the regions of interaction with EXO1,PMS2/MLH3/PMS1 a change from a Glu to a Lys is a very large, and might well result in a change to	-	PM2 + BP4	-	**Cool**

		GVGD) all suggest that this variant is likely to be tolerated,	protein's function				
NM_000249.3(MLH1):c.1709A>G p.(Asn570Ser)	0.00002831/ *African 0.0002404 African Allele Frequency considered BS1 for CanVIG	1xLikely benign; 7x Uncertain significance Conflicting	A change from an Asn to a Ser is not a large one, and may or may not result in a change to the protein's function. Doesn't interact with a ligand, metal, DNA/RNAo r protein	Pathogenic computational verdict based on 9 pathogenic predictions	PP3 + BS1	-	**Cold**
NM_000249.3(MLH1):c.1939G>A p.(Val647Met)	0.0001061/ European (Finnish) 0.0005575 European Finnish Allele Frequency considered BS1 for CanVIG	Interpretation: Uncertain significance Algorithms developed to predict the effect of missense changes on protein structure and function (SIFT, PolyPhen-2, Align-GVGD) all suggest that this variant is likely to be tolerated	A change from a Val to a Met is not a large one and may or may not result in a change to the protein's function. located in the C terminus of MLH1 DNA mismatch repair protein		BP4 + BS1 Likely Benign		**Class 2**
NM_000249.3(MLH1):c.2009A>G p.(Lys670Arg)	No Frequency	Interpretation= Uncertain significance In silico analyses predict that the variant is tolerated	located in the C-terminus MLH1 DNA mismatch repair protein no interaction with a		PM2 + BP4	-	**Cool**

			ligand, DNA/RNA or protein				

Appendix 3

MSH6 VUS Variant Classification

Variant (Reference Sequence)	**gnomAD Mean allele frequency/Highest pop. Frequency**	**ClinVar**	**VarMap**	**VarSome**	**ACGS classification rules**	**INSIGHT MMR classification**	**VUS temperature**
NM_000179.2 (MSH6):c.-66C>G	Frequency cero	No Information found	-	-	PM2	-	**Cold**
NM_000179.2 (MSH6):c.59C>T p.(Ala20Val)	0.00009814/ European non-Finnish 0.0001938 considered BS1 for CanVIG	Information found with the version number NM_000179.3 1x Likely benign; 3 stars reviewed by expert panels Multifactorial likelihood analysis posterior probability <0.005 (0.028) Four of five in-silico tools predict a benign effect of the variant on protein function In vitro studies report experimental evidence evaluating an impact on protein function in which MMR activity in a cell free assay was 50%-90% of wild-type activity (Drost_2011, Drost_2020)	A change from an Ala to a Val is not a large one and may or may not result in a change to the protein's function. Not located in any functional domain	Benign computational verdict based on 9 benign predictions	BS1 + BS3 (Sup) + BP4 Likely benign	Class 2 Likely not pathogenic / little clinical significance	**Class 2**
NM_000179.2 (MSH6):c.73G>T p.(Ala25Ser)	0.0001462 / (Other) 0.0004242 considered BS1 for CanVIG	Information found it with the version number NM_000179.3 6x Likely benign & 6x Uncertain	Associated with the disease Hereditary non-polyposis	Benign computational verdict based on 11 benign	BP4 + BS1 = likely benign	N.I.	**Class 2**

		significance Conflicting At least two independent publications report experimental evidence that the variant is MMR proficient (Drost 2011, Houlleberghd 2017).	colorecta l cancer normal mismatc h repair activity no located in any functiona l domain	predictio ns			
NM_000179.2 (MSH6):c.87C >T p.(Arg29=)	0.000008427/ Latino 0.00002932	Interpretation: Likely benign It is predicted to be benign by multiple in silico algorithms, and/or has population frequency not consistent with disease.	-	Synonym ous variant (in transcript NM_000 179.3), not predicted splicing (variant is 174 BPs before the splice-site and not found in scSNV), and the position is not strongly conserve d (GERP+ + rejected substituti ons = - 4.1 is less than 5.5).	BP4 + BP7 Likely benign	-	**Class 2**
NM_000179.2 (MSH6):c.94 G>T p.(Gly32Cys)	0.00002622 / Other 0.0001452 considered BS1 for CanVIG	**Interpretation: Uncertain significance** Based on in silico analyzes doesn't alter protein function	A change from a Gly to a Cys is not a large one, and may not result in a change	-	BP4 + BS1 Likely benign	-	**Class 2**

			to the protein's function.				
NM_000179.2 (MSH6):c.229 C>T p.(Arg77Trp)	0.00002758/ Other 0.0005076 The Allele Frequency for the Highest Population (Other) is higher than 0.0001, numbers above or round this number are considered BS1 for CanVIG	Interpretation: Uncertain Significance this alteration is predicted to be tolerated by in silico analysis.	A change from an Arg to a Trp is a very large one, and might well result in a change to the protein's function.	Benign computational verdict based on 7 benign predictions	BP4 + BS1 Likely Benign	-	**Class 2**
NM_000179.2 (MSH6):C.335 A>G p.(Asn112Ser)	0.00002474/ African 0.00008010 *2 Observations in the Allele count in the African-American population	Information found with the version number NM_000179.3 Interpretation: Uncertain Significance In silico analyses supports that this variant doesn't alter protein structure/ function. .	Located in the PWWP domain and has an interaction with a protein	-	BP4	-	**Ice Cold**
NM_000179.2 (MSH6):c.628 G>A p.(Val 210 Ile)	Frequency cero	No information found	-	Benign computational verdict based on 9 benign predictions	PM2 + BP4	-	**Cool**
NM_000179.2 (MSH6):c.663 A>C p.(Glu221Asp)	0.0006911/ South Asian 0.001112 South Asian population Allele Frequency (17) higher than 0.0001, numbers above or round this number are considered BS1 for CanVIG	Information found with the version number NM_000179.3 5x Benign; 5x Likely benign ; 5xUncertain significance Conflicting Houlleberghs, Hellen, et al 2017 states that the variant is benign	The Glu residue at position 221 is very highly conserved. So, it's very likely to be important for the protein's	Benign computational verdict based on 10 benign predictions	BS1 + BP4 + BS3 Benign	-	**Class 1**

			function				
NM_000179.2 (MSH6):c.806 C>G p.(Thr269Ser)	0.00001195/ European (non-Finnish) 0.00002646	Interpretation: Uncertain Significance Terui Hiroko et al, 2013 states that the variant doesn't have an impact on MSH6	The Thr residue at position 269 is very highly conserved. So, it's very likely to be important for the protein's function. Not located in any functional domain.	Benign computational verdict based on 11 benign predictions	BP4 + BS3 = Likely Benign	-	**Class 2**
NM_000179.2 (MSH6):c.866 _867delinsAA p.(Gly289Glu)	Frequency cero	Interpretation: Conflicting interpretations of pathogenicity Likely benign(3); Uncertain significance(7) Terui Hiroko et al, 2013 states that the variant doesn't have an impact on MSH6 Algorithms developed to predict the effect of missense changes on protein structure and function (SIFT, PolyPhen-2, Align-GVGD), and an algorithm developed specifically for MSH6 (PMID: 23621914), suggest that this missense change is likely to be tolerated	-	-	BP4 + BS3 + PM2 Likely Benign	-	**Cool**
NM_000179.2 (MSH6):c.899 G>A p.(Arg300Gln)	0.00002476/ African 0.00004009 *Just 1	Interpretation: Uncertain Significance	The Arg residue at position 300 is	-	PM2_Sup	-	**Cold**

	observation (1 allele count) in the African population		very highly conserved. So, it's very likely to be important for the protein's function .				
NM_000179.2 (MSH6):c.926C>G p.(Ser309Cys)	0.0002759 / South Asian 0.002384 South Asian population Allele Frequency (17) higher than 0.0001, numbers above or round this number are considered BS1 for CanVIG	3x Likely benign; 4x Uncertain Significance Conflicting In silico analysis, which includes protein predictors and evolutionary conservation, supports a deleterious effect.	The Ser residue at position 309 is very highly conserved. So, it's very likely to be important for the protein's function not located in any functional domain	-	BS1 + PP3	-	**Cold**
NM_000179.2 (MSH6):c.1076G>C p.(Ser359Thr)	0.000003982 / European (non-Finnish) 0.000008812 *The total number of observations 1 (Allele count)	Information not found	-	Benign computational verdict based on 12 benign predictions	PM2_Sup + BP4	-	**Cold**
NM_000179.2 (MSH6):c.1224T>G p.(Pro408=)	Frequency cero	Interpretation: Likely benign Review status: 1 star, criteria provided, single submitter	-	Synonymous variant (in transcript NM_000179.3), not predicted splicing (not found in scSNV), and the position is not strongly	PM2 + BP7	-	**Cool**

				conserved (GERP++ rejected substitutions = -3.06 is less than 5.5)			
NM_000179.2 (MSH6):c.1402 C>T p.(Arg468Cys)	0.00001195 / European (non-Finnish) 0.00002643	Information found with the version number NM_000179.3 Interpretation: Uncertain significance Review status: 2 stars, criteria provided, multiple submitters, no conflicts reported in an individual with suspected Lynch syndrome	A change from an Arg to a Cys is a large one, and might well result in a change to the protein's function The Arg residue at position 468 is very highly conserved. So, it's very likely to be important for the protein's function located in the Connector and in the MSH2 interaction domain of MSH6 interacts with DNA/RNA	Pathogenic computational verdict based on 12 pathogenic predictions	PP3	-	**COLD**
NM_000179.2 (MSH6):c.1508C>G p.(Ser503Cys)	0.0006299 / European non-Finnish 0.001326	Interpretation: Likely benign Review status: 3	The Ser residue at position		**BS1 + BP4 + BS3** **Benign**	Likely not pathogenic/little clinical significance	**Class 1**

	European non-Finnish population Allele Frequency (17) higher than 0.0001, numbers above or round this number are considered BS1 for CanVIG	stars reviewed by expert panel Multifactorial likelihood analysis posterior probability 0.001 – 0.049 Publications have cited the variant in affected individuals with HNPCC or HNPCC related cancer. Multiple authors have classified the variant as "benign" and reported presence of MSH6 protein expression in tumors of CRC patients carrying this variant. Predicted to be benign by multiple in silico algorithms. According to Terui Hiroko et al, 2013 the variant showed normal expression of MSH6	503 is very highly conserved. So, it's very likely to be important for the protein's function The variant is located in the Connector and in the MSH2 interction domain of MSH6 the variant interacts with Mg				
NM_000179.2 (MSH6):c.1720T>A p.(Ser574Thr)	0.000007922 / European (non-Finnish) 0.00001765 * Just 2 Observations (Allele Count) in Total	Interpretation: Uncertain Significance Review status: 2 stars, criteria provided, multiple submitters, no conflicts	Variant located in the MutSII domain (in the connector and in the MSH2 interaction domain of MSH6)	Benign computational verdict based on 8 benign predictions	BP4		**Ice Cold**
NM_000179.2 (MSH6):c.1814C>G p.(Thr605Ser)	0.00005321 / European (non-Finnish) 0.00009314	Information found with the version number NM_000179.3 Interpretation: Uncertain Significance Review status: 2	Is located in the MutSII domain (Connector or domain of	-	PP3	-	**Cold**

		stars, criteria provided, multiple submitters, no conflicts . Algorithms developed to predict the effect of sequence changes on RNA splicing suggest that this variant may create or strengthen a splice site,	MSH6). Doesn't interact with a ligand, metal, DNA/RNA or protein				
NM_000179.2 (MSH6):c.1847 C>G p.(Ser616Cys)	0.00005318 / Latino 0.0004235 European non-Finnish population Allele Frequency (17) higher than 0.0001, numbers above or round this number are considered BS1 for CanVIG	Interpretation: 4x Likely benign; 4x Uncertain Siginificance Conflicting Algorithms developed to predict the effect of missense changes on protein structure and function output the following: SIFT: Tolerated; PolyPhen-2: Benign; Align-GVGD: Class C0.	The Ser residue at position 616 is very highly conserved. So, it's very likely to be important for the protein's function. Located in the MutSII domain (Connector domain of MSH6).	Benign computational verdict based on 9 benign predictions	BP4 + BS1 Likely benign	-	**Class 2**
NM_000179.2 (MSH6):c.1915G>A p.(Glu639Lys)	0.00001416 / (African/African – American) 0.00004019 * 1 Observation in the Allele Count for the (African/African-American)	Interpretation: Uncertain Significance Review status: 2 stars, criteria provided, multiple submitters, no conflicts Terui, Hiroko et al 2013 states that his variant has not impact on MSH6	A change from a Glu to a Lys is a very large one, and might well result in a change to the protein's function. The variant is located I	Pathogenic computational verdict based on 11 pathogenic predictions	PM2_Sup + PP3 + BS3	-	**Cool**

			the MutS_II domain (Connector domain of MSH6)				
NM_000179.2 (MSH6):c.1971G>C P.(Gln657His)	Frequency cero	Interpretation: Uncertain significance Review status: 2 stars, criteria provided, multiple submitters, no conflicts .	The Gln residue at position 657 is vey highly conserved. So, it's very likely to be important for the protein's function. The variant is located in the MutS_II domain (Connector domain of MSH6).	Benign computational verdict based on 8 benign predictions	PM2 + BP4	-	**Cool**
NM_000179.2 (MSH6):c.2076 A>C p.(Lys692Asn)	Frequency cero	Interpretation: Uncertain Significance	A change from a Lys to an Asn is a large one, and might potentially result in a change to the protein's function the Lys residue at position 692 is very highly conserved. So, it's very		PM2	-	**Cool**

			likely to be important for the protein's function Located in the MutS domain II (Connector domain)				
NM_000179.2 (MSH6):c.2195 G>A p.(Arg732Gln)	0.000003986 / South Asian 0.00003267 * Just 1 observation in the total Allele Count	Interpretation: Uncertain Significance Review Status: 2 stars, criteria provided, multiple submitters, no conflicts This variant is present in population databases (rs749746725, ExAC 0.006%)	This variant is located in the Connector domain of MSH6		PM2_Sup	-	**Cold**
NM_000179.2 (MSH6):c.2203C>A p.(Leu735Ile)	Frequency cero	Interpretation: Uncertain Significance Review status: 2 stars, criteria provided, multiple submitters, no conflicts this alteration is predicted to be tolerated by in silico analysis.	This variant s located in the lever domain.		PM2 + BP4	-	**Cool**
NM_000179.2 (MSH6):c.2511C>G p.(His837Gln)	0.00002416 / African 0.00006439 *Just 1 observation in the Allele count in African/African-American	Interpretation: Uncertain Significance Review status: 2 stars, criteria provided, multiple submitters, no conflicts	The His residue at position 837 is very highly conserved. So it's very likely to be importan	-	PM2_Sup	-	**Cold**

			t for the protein's function Located in the Lever domain of MSH6				
NM_000179.2 (MSH6):c.2713T>A p.(Leu905Met)	0.000003987 / European (non-Finnish) 0.000008829 * Just 1 observation in the Allele Count in Total	Variant found with the version number: NM_000179.3 Interpretation:Uncertain significance Review status: 2 stars, criteria provided, multiple submitters, no conflicts ExAc 0.002%	The Leu residue at position 905 is very highly conserved. So it's very likely to be important dor the protein's function The variant is located in the Lever domain		PM2_Sup	-	**Cold**
NM_000179.2 (MSH6):c.2961T>C p.(Thr987=)	Frequency cero	Interpretation: Likely benign; Review status 1 star for Invitae no evidence details		Synonymous variant (in transcript NM_000179.3), not predicted splicing (variant is 212 BPs before the splice-site and not found in scSNV), and the position is not strongly conserved (CSH phyloP100way = -0.255 is	PM2 + BP7 +	-	**Cool**

				less than 5). ClinVar classifies this variant as Likely Benign, rated 1 star, criteria provided, single submitter, with 1 submission.			
NM_000179.2 (MSH6):c.2962C>T p.(Arg988Cys)	0.00002012 / Latino 0.00003484 * 1 Observation of Allele Count in Latino population	Interpretation: Uncertain significance Review status: 2 stars, criteria provided, multiple submitters, no conflicts This variant is present in population databases (rs61753795, ExAC 0.009%).	A change from an Arg to a Cys is a very large one and might well result in a change to the protein's function. The Arg residue at position 988 is very highly conserved. So, it's very likely to be important for the protein's function. Located in the MutS family domain IV (Lever domain)	Pathogenic computational verdict based on 10 pathogenic predictions	PM2_Sup + PP3	-	**Cool**

			interacts with a protein				
NM_000179.2 (MSH6):c.3048T>C p.(Ala1016=)	Frequency cero	Interpretation: Likely benign Review status: 2 stars, criteria provided, multiple submitters, no conflicts It's predicted to be benign by multiple in silico algorithms		Synonymous variant (in transcript NM_000179.3), not predicted splicing (variant is 125 BPs before the splice-site and not found in scSNV), and the position is not strongly conserved (CSH phyloP100way = 0.045 is less than 5). -	PM2 + BP4 +BP7 Likely benign	-	**Cool**
NM_000179.2 (MSH6):c.3173-18T>C	0.00009196 / European (non-Finnish) 0.0001859 European non-Finnish population Allele Frequency (17) higher than 0.0001, numbers above or round this number are considered BS1 for CanVIG	Interpretation: Benign/Likely benign Review status: 2 stars, criteria provided, multiple submitters, no conflicts 4/4 computational tools predict no significant impact on normal splicing Two clinical diagnostic laboratories have submitted clinical-significance assessments for this variant to ClinVar after 2014 without evidence for independent evaluation	-		**BP4 + BS1 + BP6** **Likely benign**	-	**Class 2**

NM_000179.2 (MSH6):c.3173-5A>C	Frequency cero	No Information found			PM2	-	**Cool**
NM_000179.2 (MSH6):c.3203G>A p.(Arg1068Gln)	0.0001273 / African 0.0002804 European non-Finnish population Allele Frequency (17) higher than 0.0001, numbers above or round this number are considered BS1 for CanVIG	Interpretation: 1x Benign; 3x Likely benign; 5x Uncertain significance Conflicting	The Arg residue at position 1068 is very highly conserved. So, it's very likely to be important for the protein's function.	Benign computational verdict based on 10 benign predictions)	BP4 + BS1 Likely benign	-	**Class 2**
NM_000179.2 (MSH6):c.3217C>T p.(Pro1073Ser)	0.0004103 / Ashkenazi Jewish 0.007620 Ashkenazi Jewish population Allele Frequency (17) higher than 0.0001, numbers above or round this number are considered BS1 for CanVIG	Interpretation: Likely benign Review status 3 stars Multifactorial likelihood analysis is posterior probability 0.001 – 0.049	The Pro residue at position 1073 is very highly conserved. So, it's very likely to be important for the protein's function. Variant located in the Lever domain	Benign computational verdict based on 7 benign predictions	BP4 + BS1 Likely benign	Likely not pathogenic	**Class 2**
NM_000179.2 (MSH6):c.3259C>T p.(Pro1087Ser)	0.0001167 / European (non-Finnish) 0.0002477 European non-Finnish population Allele Frequency (17) higher than 0.0001, numbers above or round this number are considered BS1 for CanVIG	Information found with the version number NM_000179.3 4x Likely benign; 7x Uncertain significance Conflicting An in vitro cell free mismatch repair complementation assay demonstrated 88% repair efficiency, suggesting MSH6 Pro1087Ser is not	The Pro residue at position 1087 is very highly conserved. So, it's very likely to be important for the protein's function.	Pathogenic computational verdict based on 7 pathogenic predictions	PP3 + BS3 + BS1 Benign	-	**Class 1**

		pathogenic (Drost 2010)	located in the Lever domain				
NM_000179.2 (MSH6):c.3543 C>G p.(Asp1181Glu)	0.00003185 / European (non-Finnish) 0.00005286	Interpretation: Uncertain significance Review status: 2 stars, criteria provided, multiple submitters, no conflicts	the Asp residue at position 1181 is very highly conserved. So it's very likely to be important for the protein's function. Located in the Muts_V functional domain and located in the ATPase domain of MSH6	Pathogenic computational verdict based on 10 pathogenic predictions	PP3	-	**Cold**
NM_000179.2 (MSH6):c.3556G>C p.(Gly1186Arg)	0.000007968 / European (non-Finnish) 0.00001764 *2 observations in total in the allele count	Interpretation:Uncertain significance Review status: 2 stars, criteria provided, multiple submitters, no conflicts	A change from a Gly to an Arg is a very large one, and might well result in a change to the protein's function. The Gly residue at position 1186 is very highly conserved. So, it's very likely to be important for the protein's	Pathogenic computational verdict based on 13 pathogenic predictions from	PP3		**Cold**

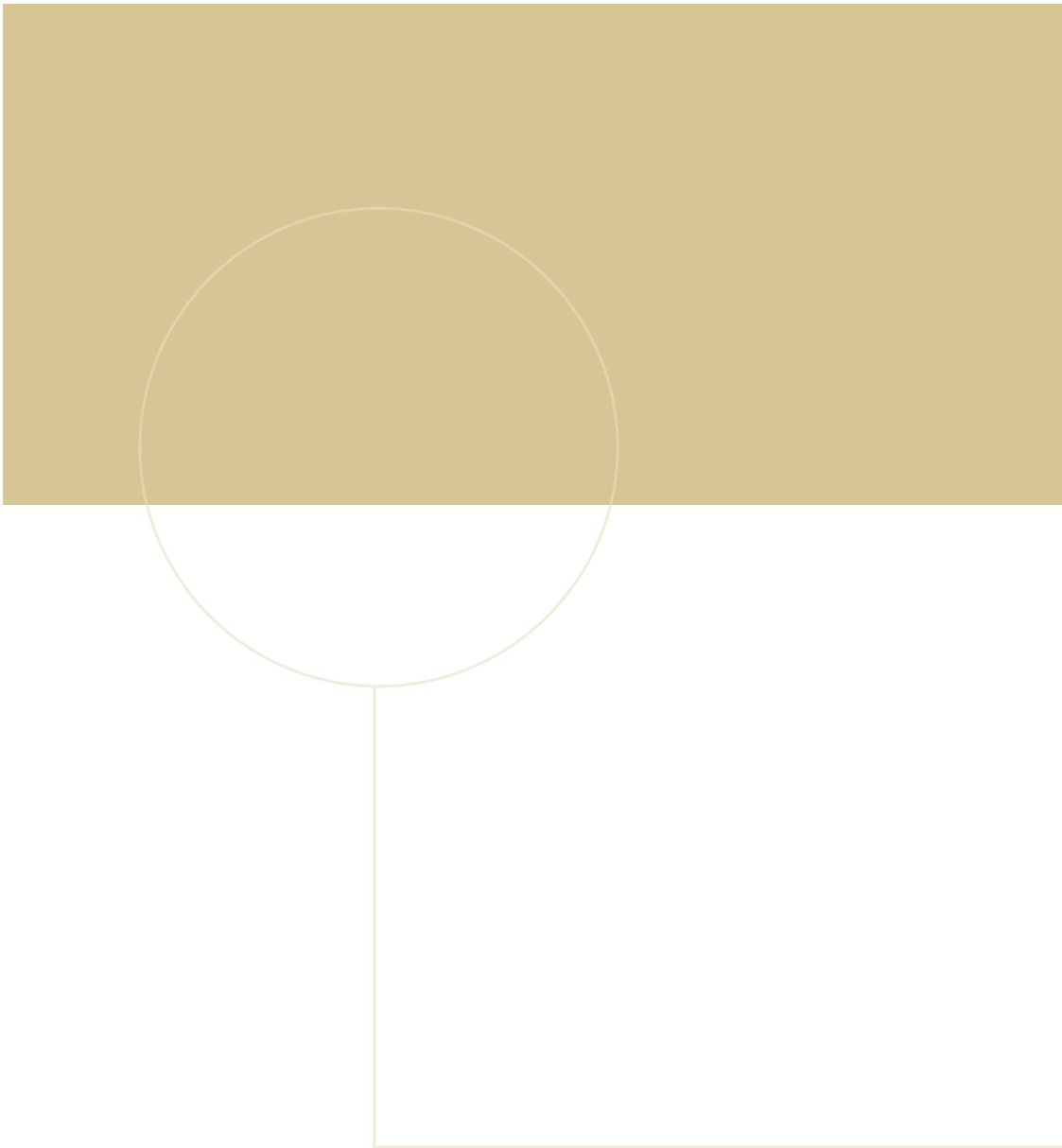
			function Variant located in the ATPase domain				
NM_000179.2 (MSH6):c.3727A>T p.(Thr1243Ser)	0.0002618 / South Asian 0.001143 South Asian population Allele Frequency (17) higher than 0.0001, numbers above or round this number are considered BS1 for CanVIG	4x Likely benign; 9x Uncertain significance. Conflicting	The Thr residue at position 1243 is very highly conserved. So, it's very likely to be important for the protein's function. Located in the MutS_V domain (ATPase domain)	Pathogenic computational verdict based on 7 pathogenic predictions	PP3 + BS1	-	**Cold**
NM_000179.2 (MSH6):c.3758T>A p.(Val1253Glu)	0.0001486 / European (Finnish) 0.0004379 European Finnish population Allele Frequency (17) higher than 0.0001, numbers above or round this number are considered BS1 for CanVIG	The information was found with the version number NM_000179.3 Interpretation: Uncertain significance Review status: 2 stars, criteria provided, multiple submitters, no conflicts	A change from a Val to a Glu is a large one, and might potentially result in a change to the protein's function. The Val residue at position 1253 is very highly conserved. So it's very likely to be important for the protein's function.	Pathogenic computational verdict based on 7 pathogenic predictions	PP3 + BS1	-	**Cold**

			located in the MutS_V functional domain (ATPase domain)				
NM_000179.2(MSH6):c.3762_3764del p.(Glu1254del)	0.00002477 / African/ African-American 0.0001602 African-American population Allele Frequency (17) higher than 0.0001, numbers above or round this number are considered BS1 for CanVIG	Interpretation: Uncertain Significance This variant was not observed at a significant allele frequency in large population cohorts (Lek 2016). This deletion of a single Glutamic Acid amino acid is located in the ATPase domain (Warren 2007, Kansikas 2011) In silico analysis, which includes protein predictors and evolutionary conservation, supports a deleterious effect. This variant, c.3762_3764del, results in the deletion of 1 amino acid(s) of the MSH6 protein (p.Glu1254del), but otherwise preserves the integrity of the reading frame. This variant is present in population databases			PP3 + PM4_Sup + BS1	-	**Cool**
NM_000179.2(MSH6):c.3848_3850dup p.(Ile 1283 dup)	0.00002480 / European (non-Finnish) 0.00005432	Interpretation: Uncertain Significance Review status: 1 star, criteria provided, multiple submitters no conflicts This variant, c.3848_3850dupTTA, results in the		Protein coding length changes as a result of in frame variant in gene MSH6, and is not in a	PP3 + PM4_Sup	-	**Cool**

		insertion of 1 amino acid to the MSH6 protein (p.Ile1283dup), but otherwise preserves the integrity of the reading frame Variant observed in at least 2 individuals suspected to have Lynch syndrome Duplication is located within the ATPase domain In silico analysis support a deleterious effect ExAC 0.009%		repeat region.			
NM_000179.2 (MSH6):c.3961A>G p.(Arg1321Gly)	0.0001394 / Other 0.0002793 *2 Observations in the Higher Population allele frequency (Other)	I found it with the version number NM_000179.3 4x Likely benign; 8x Uncertain significance Conflicting Variant was also reported to co-occur with a MSH2 pathogenic deletion with mismatch repair deficient CRC.	A change from an Arg to a Gly is a large one, and might well result in a change to the protein's function. The Arg residue at position 1321 is very highly conserved. So it's very likely to be important for the protein's function Located in the MutS_V functional domain	Pathogenic computational verdict based on 8 pathogenic predictions	PP3 + BS1_Sup+	-	**Cold**

			(ATPase domain)				
NM_000179.2 (MSH6):c.4001+12_4001+15dup	0.0002685 / African/African-American 0.0007754	Interpretation: Benign/Likely benign Review status: 2 stars, criteria provided, multiple submitters, no conflicts Computational tools predict no significant impact on normal splicing Multiple clinical diagnostic laboratories have submitted clinical-significance assessments for this variant to ClinVar without evidence for independent evaluation. All laboratories classified the variant as benign/likely benign.	-		BS1_Sup + BP6 + BP4 Likely Benign	-	**Class 2**
NM_000179.2(MSH6):c.4002-10T>A	0.0002572 / Ashkenazi Jewish 0.001475	2x Benign ; 1x Likely benign ;8x Uncertain significance 4/5 computational tools predict no significant impact on normal splicing The observed variant frequency is approximately 1.8 fold of the estimated maximal expected allele frequency for a pathogenic variant in MSH6 causing Lynch Syndrome phenotype (0.00014), strongly suggesting that the variant is benign.	-		BS1 + BP4 Likely benign	-	**Class 2**

		.					
NM_000179.2 (MSH6):c.4030A>C p.(Thr1344Pro)	Frequency cero	**No information found**		Benign computational verdict based on 10 benign predictions	**PM2 + BP4**	-	**Cool**



NTNU

Norwegian University of
Science and Technology