Acknowledgements

This project was conducted at the Department of Medical Genetics, St. Olavs Hospital, leading to a Master of Science in Molecular Medicine.

First of all, I would like to thank my supervisor Dr. Wenche Sjursen for giving me the opportunity to work with this exciting project. You have been an excellent mentor throughout the writing process, always available and helpful. I am very grateful for your valuable feedback and encouragement, and I admire your academic knowledge and positive attitude.

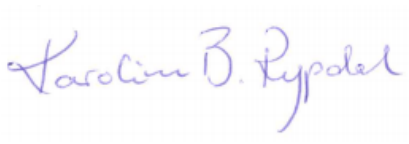
I wish to express my gratitude towards my co-supervisor Maren F. Hansen who has given me invaluable guidance throughout this whole process. Thank you for all assistance in the laboratory and with the data analysis, as well as for taking the time to review my thesis. You inspire me. I wish you the best of luck with your PhD, I know you will do great.

Big thanks to Jostein Johansen for helping out with bioinformatic analyses and data interpretations. I am very grateful for your time and assistance. I also wish to thank the kind ladies at the medical genetics lab, especially Margit, for guiding me in the laboratory and always answering my questions.

The amazing members of *Dendrittene* deserve special thanks. You guys have kept my motivation up and my mind sane. Thank you for all the lunches, conversations, adventures and special moments we have shared over the last two years. I wish all of you the best of luck.

Last but not least, I wish to express my appreciation of the two extraordinary men in my life. I am forever grateful to my loving dad, who always believes in me and supports me. You are my hero. To my dear Michael, you are magical. Thank you for all your feedback, love and support.

Trondheim, May 2015



Karoline Bjarnesdatter Rypdal

Abstract

Colorectal cancer (CRC) is the third most common malignancy worldwide, with over 1 million new cases annually. Around 30% of cases are believed to be familial with causal genetic alterations. However, fewer than 10% of cases are so far genetically explained. Lynch syndrome (LS) is the most common hereditary CRC syndrome, explaining around 5% of all cases. LS is caused by pathogenic germline mutations in one of four DNA mismatch repair (MMR) genes: *MLH1, MSH2, MSH6,* and *PMS2*. Extensive research is currently attempting to decipher the complete underlying genetic patterns of CRC, with the hope of improving cancer diagnosis and treatment.

In this project, a gene-panel consisting of 124 genes across 95 CRC patients was sequenced using NGS. The patients included in this study have a family history of CRC, but previous genetic testing has not identified causal mutations. The aim of the present Master of Science project was to search for potentially disease-causing pathogenic germline mutations within 22 MMR system-associated genes. Using *in silico* prediction tools, the deleterious potential of the detected variants was assessed. 20 candidate predisposition mutations were identified across 12 MMR-associated genes in 22 patients. Based on available information about the gene, mutation position, patient phenotype, findings in previous studies, and prediction tools, the identified variants are suggested to be involved in the development of cancer. To confirm these indications, further studies of the variants are needed. Segregation analysis will reveal familial inheritance and penetrance status of the mutations, and functional studies are required to elucidate the consequences on protein function.

This project has demonstrated the successful sequencing of a multipatient gene panel. The NGS approach is rapidly making an entrance in diagnostics, and this study has shown an efficient and reliable method of sequencing numerous genes and samples in parallel. To facilitate unambiguous data interpretation and standardized analysis further advances are needed before widespread routine application will be advantageous in the clinic.

Abbreviations

AC – Amsterdam criteria

AC-II – revised Amsterdam criteria

ATP – adenosine triphosphate

BG – Bethesda guidelines

bp – base pair

CRC – colorectal cancer

dbSNP – the single nucleotide polymorphism database

DNA – deoxyribonucleic acid

dNTP – deoxyribonucleoside triphosphate

EXO1 – exonuclease 1

EPCAM – epithelial cell adhesion molecule

FAP – familial adenomatous polyposis

HNPCC – hereditary non-polyposis colorectal cancer

IHC – immunohistochemistry

InSiGHT – international society for gastrointestinal hereditary tumours

LIG1 – ligase 1

LLS – lynch-like syndrome

LOVD – Leiden open variation database

LS – Lynch syndrome

MAF – minor allele frequency

MAP – MUTYH-associated polyposis

MMR – mismatch repair

MLH – human mutL homolog

MSH – human mutS homolog

MSI – microsatellite instability

MSI-H – microsatellite instability high

MSI-L – microsatellite instability low

MSS – microsatellite stable

MUTYH – human mutY homolog

NCBI – national centre for biotechnology information

NGS – next generation sequencing

Nt – nucleotide

PCNA – proliferating cell nuclear antigen

PCR – polymerase chain reaction

PMS – human postmeiotic segregation increased homolog

POLD – DNA polymerase δ

RFC – replication factor C

RPA – replication protein A

SNP – single nucleotide polymorphism

VUS – variant of uncertain significance

WGS – whole genome sequencing

XPG – xeroderma pigmentosum complementation group G

Table of contents

1. Introduction

1.1 Colorectal cancer

Colorectal cancer (CRC) is a heterogeneous disease that has become one of the most common malignancies worldwide. CRC is the third most diagnosed cancer in men, the second most diagnosed cancer in women, and globally there are over 1 million new cases each year. Taking both sexes into account, CRC has the fourth highest cancer mortality rate, following lung, liver, and stomach cancer, with an estimated 694,000 deaths annually. The highest rates of cases are found in Australia/New Zealand and Western Europe. The incidence of CRC is significantly higher in men than in women, with 746,000 and 614,000 cases respectively (1, 2). In Norway, CRC is the third most common malignancy with approximately 3,500 new cases each year (3).

The most important risk factor for developing CRC is family history. With no affected family members, the lifetime risk of an individual is 5-6%. If a first-degree relative is affected, the lifetime risk increases to 10-15%, and with a hereditary genetic syndrome, the risk ranges from 30-100% (4). CRC is classified into three groups, based on increasing cancer risk and hereditary influence: sporadic, familial, and hereditary CRC. The majority of CRC cases are categorised as sporadic (ca. 70%). Patients with sporadic cancer do not have a family history of cancer and no inherited pre-disposing mutations. Around 30% of cases are believed to be familial. Familial CRC is associated with genetic changes in high-, moderate-, and low-risk susceptibility genes, and the patients have two or more affected first-degree relatives. Characteristic of familial CRC is early age of onset, which indicates an inherited genetic component. However, in a large proportion of cases the specific causal germline mutations fail to be detected. Hereditary CRC is caused by inherited high-penetrant single-gene mutations in cancer susceptibility genes. The high-risk genetic mutations identified so far account for less than 10% of all CRC cases (5-7). An overview of the relative proportions and heterogeneity of identified hereditary CRC syndromes is shown in figure 1.1.

There are several genetically characterized inherited CRC syndromes. The most frequent syndromes today are Lynch Syndrome (LS), familial adenomatous polyposis (FAP), and *MUTYH*-associated polyposis (MAP) (4). MAP is a form of adenomatous polyposis, which results from bi-allelic mutations in the *MUTYH* gene. It is the first known polyposis syndrome with a recessive inheritance pattern, and it is characterised by a slight increase in the risk of

developing polyps/adenomas and CRC. The MUTYH protein is a base excision repair glycosylase, which repairs oxidative DNA damage. When the protein is non-functional, mutations in the DNA will accumulate and eventually lead to the development of cancer (6, 8). FAP is an autosomal dominant inherited disorder, characterised by the development of hundreds to thousands of colorectal adenomatous polyps during the second decade of life. It is the second most common inherited CRC syndrome, but accounts for less than 1% of all CRCs (6). If left untreated it results in almost complete penetrance of CRC by the age of 50. Classic FAP is caused by a high-penetrance germline mutation in the adenomatous polyposis coli (*APC*) gene, which is a tumour suppressor gene (6, 8, 9). Attenuated FAP (AFAP) is a less aggressive form of classic FAP, caused by mutations in other parts of the APC gene, with fewer polyps, later age of onset, and decreased cancer risk (8). The most frequent CRC syndrome is LS, accounting for up to 5% of all CRC cases (10).

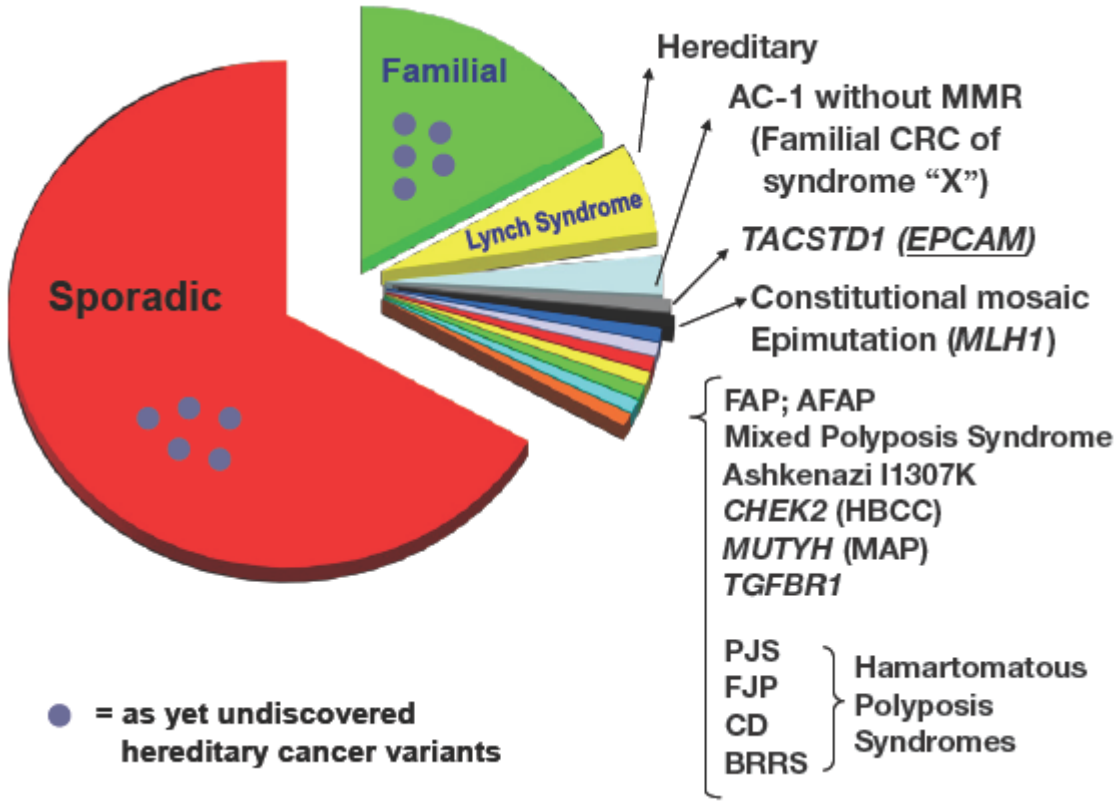


Figure 1.1. Figure adapted from Lynch, 2013. Pie chart depicting the heterogeneity of hereditary colorectal cancer syndromes. Abbreviations: AC-I, Amsterdam Criteria I; MMR, mismatch repair; FAP, familial adenomatous polyposis; AFAP, attenuated familial adenomatous polyposis; HBCC, hereditary breast and colorectal cancer; PJS, Peutz-Jeghers syndrome; FJP, familial juvenile polyposis; CD, Cowden's disease; BRRS, Bannayan-Ruvalcaba-Riley syndrome (11).

1.2 Lynch syndrome

Lynch Syndrome is an autosomal dominant disorder, named after Dr. Henry T. Lynch, who first described the syndrome in two large families in 1966 (12). The syndrome is also known as Hereditary Non-Polyposis Colorectal Cancer (HNPCC) (13). LS is the most common CRC-predisposing syndrome, with a significantly increased risk of colon cancer and other cancers, as well as an accelerated carcinogenesis. LS patients are subjected to an earlier age of cancer onset than the general population, especially CRC and endometrial cancers, with an average of 45 years, compared to 69 years in the general population (12).

Disease in patients with LS is caused by germline mutations or epimutations in one of four DNA mismatch repair (MMR) genes, *MLH1, MSH2, MSH6,* or *PMS2*. Mutations in the two former genes account for up to 90% of cases. In addition, germline deletions in the 3' end of the *EPCAM* gene, which inactivates the downstream *MSH2* gene through promoter methylation, occur in 1% of cases (5, 6). An inherited mutation in one of the four MMR genes, together with consecutive inactivation of the remaining wild-type allele by somatic events, leads to complete inactivation of the gene. This results in the inability to correct insertions/deletions and mismatches that occur during DNA replication, causing an accumulation of errors in the DNA (5). In addition to high-penetrance mutations in the MMR genes causing the disease, modifier mutations in other genes have been shown to affect the cancer risk in LS (14).

1.2.1. Microsatellite instability and immunohistochemistry

Inactivation of MMR often results in a strong mutator phenotype known as microsatellite instability (MSI). MSI is defined as a change, due to either insertions or deletions of any length, in repeating units within a microsatellite in a tumour (15). Microsatellites are simple repeated sequences that occur ubiquitously across the genome. Alterations in microsatellites within coding genes can increase tumour development and the risk of cancer. MSI is the hallmark of MMR deficiency, seen in most LS tumours and in a proportion of sporadic colorectal tumours. MSI tumours can be divided into groups based on whether certain genetic markers exhibit MSI. In MSI-high (MSI-H) tumours, over 30% of markers exhibit MSI, and in MSI-low (MSI-L) tumours, only a few markers exhibit MSI. In microsatellite stable (MSS) tumours, none of the markers exhibit MSI, and 60-70% of tumours overall fall into this category. Tumour phenotyping for MSI is a useful tool for identifying LS patients, because it

is associated with certain unique clinical and pathological characteristics. However, MSI tumours are not necessarily caused by heritable mutations (15, 16). Only 20-25% of MSI-H tumours are associated with germline mutations in the MMR genes, as MSI is sensitive, but not specific for LS (12). In sporadic CRCs there is a strong association between MSI-H tumours and loss of MLH1 protein expression. The explanation is generally silencing of the *MLH1* gene, due to somatic hypermethylation (17). Recently, CRC patients with MSI-H phenotype, but without detectable germline mutations or hypermethylation of the MMR genes have been diagnosed with the newly described Lynch-like syndrome (LLS) (18). The etiology of this disease is unknown and the clinical significance is uncertain. Immunohistochemistry (IHC) is a valuable supplement to MSI testing. IHC is a way of visualizing the presence of specific markers, such as proteins, through binding of antibodies to antigens in tissue. If the staining is interpreted correctly, IHC is extremely sensitive to MMR deficiency. This provides a certain way of knowing if a MMR gene is expressed in the tumour cells (12).

1.2.2 The Amsterdam Criteria and Bethesda Guidelines

In 1989, a set of criteria was proposed to provide uniform and standardized diagnostic principles for multicentre studies. These criteria were agreed upon by an international collaborative group of researchers in 1991, and became known as the Amsterdam Criteria (AC) for diagnosis of LS. The criteria were established before the identification of the MMR genes, and were later broadened to recognize the genetic component, and diagnostic role of extracolonic tumours. These became the AC-II in 1999. An additional set of guidelines was established in Bethesda in 1996, for the identification of colorectal tumours that should be tested for MSI, called the Bethesda Guidelines (BG) for identifying patients with HNPCC. These were revised in 2002, and are now known as the revised BG. Today, both the AC-II and the revised BG are used to identify families that are likely to have LS, before genetic testing is performed (12, 19). Each of the criteria must be fulfilled for the patient to be diagnosed with LS. Nonetheless, only about half of the patients that satisfy the AC are in fact diagnosed with LS. The AC-I and AC-II are shown in table 1.1, and the revised BG are shown in table 1.2.

Table 1.1. The Amsterdam Criteria. The Amsterdam Criteria (AC) I from 1991 to the left, and the revised AC-II from 1999 to the right (19, 20).

AC-I	AC-II
Three or more relatives have CRC	Three or more relatives have CRC or HNPCC associated cancer
At least one affected patient should be a first-degree relative	At least one affected patient should be a first-degree relative
Two or more successive generations should be affected	Two or more successive generations should be affected
Cancer in one or more cases is diagnosed before the age of 50 years	Cancer in one or more cases is diagnosed before the age of 50 years
	FAP should be excluded
	Tumours should be confirmed with histology

Table 1.2. Revised Bethesda guidelines for testing colorectal tumours of microsatellite instability (MSI) (17).

Tumours should be tested in the following situations:	
1	CRC is diagnosed in a patient before the age of 50 years
2	Synchronous, metachronous colorectal, or other HNPCC-associated tumours are present, regardless of age
3	CRC with high-frequency MSI (MSI-H) histology is diagnosed before the age of 60 years
4	CRC is diagnosed in one or more first-degree relatives with an HNPCC-related tumour, with one of the cancers diagnosed before the age of 50 years
5	CRC is diagnosed in two or more first- or second-degree relatives with HNPCC-related tumours, regardless of age

1.3 Familial CRC type X

Approximately 50% of all families with CRC that fulfil the AC are found to carry an inherited mutation in the MMR genes by genetic testing, and are consequently diagnosed with LS. However, around half of families that meet the AC do not have detectable genetic MMR mutations, and thus cannot be diagnosed with LS. These families are termed MMR-proficient HNPCC families with familial CRC type X (fCRC-X) (10, 21). Patients with fCRC-X have no identified mutations in the MMR genes, no tumour MSI, and no loss of IHC staining of the MMR proteins (6). The genetics underlying the cancer risk in these Lynch-like families remains elusive, but the symptoms point to the existence of unidentified CRC high- or moderate-risk loci. The AC indicate a strong familial aggregation, which makes it likely that certain fCRC-X cases are caused by high-penetrance mutations. Some of the familial risk may, however, be explained by co-inheritance of several low-risk loci, serving as predisposing factors that can interact with environmental factors. Such variants could mediate a risk too low to be detected in linkage analyses, and the low minor allele frequency (MAF) prevents them from being captured by genome-wide association studies (GWAS) (6, 10).

1.4 The mismatch repair system

DNA mismatch repair is a highly conserved biological pathway that maintains genomic stability. The primary objective of the MMR system is to eliminate and correct nucleotides that are incorrectly inserted or deleted during DNA synthesis. The system is additionally responsible for suppression of homologous recombination. The MMR genes code for proteins with various functions that aid in stabilization of the genome (22, 23).

The four MMR genes, *MLH1, MSH2, MSH6* or *PMS2,* were first identified in *E. coli,* where the protein complexes MutS and MutL are responsible for initiating mismatch repair. MutS recognizes the mismatch and MutL mediates downstream activities. Homologs of MutS and MutL were later identified in mammalian cells, existing as heterodimers. The human MutSα (hMutSα) complex consists of human MutS Homolog (MSH) 2 and 6. The complex recognizes base-base mismatches and insertions and deletions (indels), in addition to some larger indels. 80-90% of cellular MSH2 is bound in this heterodimer. The hMutSβ complex consists of MSH2 and MSH3, and recognizes mismatch indel loops of 2-10 nucleotides. The hMutLα complex consists of human MutL Homolog (MLH) 1 and human Postmeiotic Segregation Increased homolog (PMS) 2. This complex supports repair initiated by hMutSα

or hMutSβ. Approximately 90% of cellular MLH1 is bound in this heterodimer. An hMutLβ heterodimer, comprised of MLH1 and PMS1, has been isolated in humans, but its involvement in mismatch repair has not been confirmed. An hMutLγ heterodimer, comprised of MLH1 and MLH3, has been reported to confer some mismatch repair activity *in vitro* (22, 24).

The MMR system not only consists of the four proteins mentioned above, rather a number of different proteins help mediate the repair. Another 18 protein-coding genes are considered to be included in the MMR pathway, namely *MLH3, MSH3, PMS1, POLD1-4, PCNA, RFC1-5, EXO1, RPA1-3,* and *LIG1* (23). These proteins make up the different steps in the MMR process, and are of utmost importance for maintenance of genomic stability. DNA polymerase delta (Pol δ) plays a major role in maintaining the genome, with many DNA-synthesis repair mechanisms. The primary function of the polymerase is replication of the lagging strand. It exhibits 5'→3' polymerase activity, in addition to 3'→5' proofreading exonuclease activity, which repair incorrectly inserted nucleotides (25). Mammalian Pol δ is composed of four subunits: the catalytic subunit p125 (encoded by *POLD1*), and the accessory subunits p50 (*POLD2*), p68 (*POLD3*), and p12 (*POLD4*). Pol δ cooperates closely with the processivity factor Proliferating Cell Nuclear Antigen (PCNA). PCNA is a cofactor of Pol δ during DNA synthesis and plays an important role in MMR. It binds to the DNA strand and acts as a scaffold, recruiting other proteins involved in MMR. It is called a eukaryotic replication sliding-clamp, and it strongly interacts with hMutSα and hMutSβ (25, 26). Replication Factor C (RFC) is a clamp loader enzyme, which loads PCNA onto the DNA helix, and is therefore crucial for assembly of the MMR apparatus. RFC and PCNA function together to regulate the directionality of excision and repair. RFC uses ATP hydrolysis to open and close PCNA onto primed sites used by DNA polymerases and repair factors. RFC is a heteropentamer, composed of five essential subunits: RFC1, RFC2, RFC3, RFC4, and RFC5 (26, 27). Exonuclease 1 (EXO1) belongs to the RAD2/XPG family of endo- and exonucleases, and is involved in repair, recombination, and replication of DNA. EXO1 plays a functional role in MMR by retaining a 5'→3' double stranded DNA exonuclease activity, and a 5' flap endonuclease activity. The exonuclease function is required for repair of base-base mismatches and single nucleotide indels. EXO1 interacts with, and stabilizes higher order components of the MMR system: MSH2, MLH1, and MSH3. The enzyme has also been shown to possess a cryptic 3'→5' double stranded DNA exonuclease activity, stimulated by PCNA (26, 28, 29). The single-stranded DNA-binding protein Replication Protein A (RPA) is

involved in mismatch repair by stimulating excision, stabilizing the DNA gap against endonuclease attack, and promoting repair synthesis of DNA. The protein consists of three subunits: RPA1, RPA2, and RPA3, each containing one or more DNA-binding domains (24, 30). For completion of excision repair, DNA joining events are required. DNA Ligase 1 (LIG1) is responsible for joining nicks in the DNA strand while cooperating with PCNA (31). A schematic overview of the MMR pathway components' functions is shown in figure 1.2.

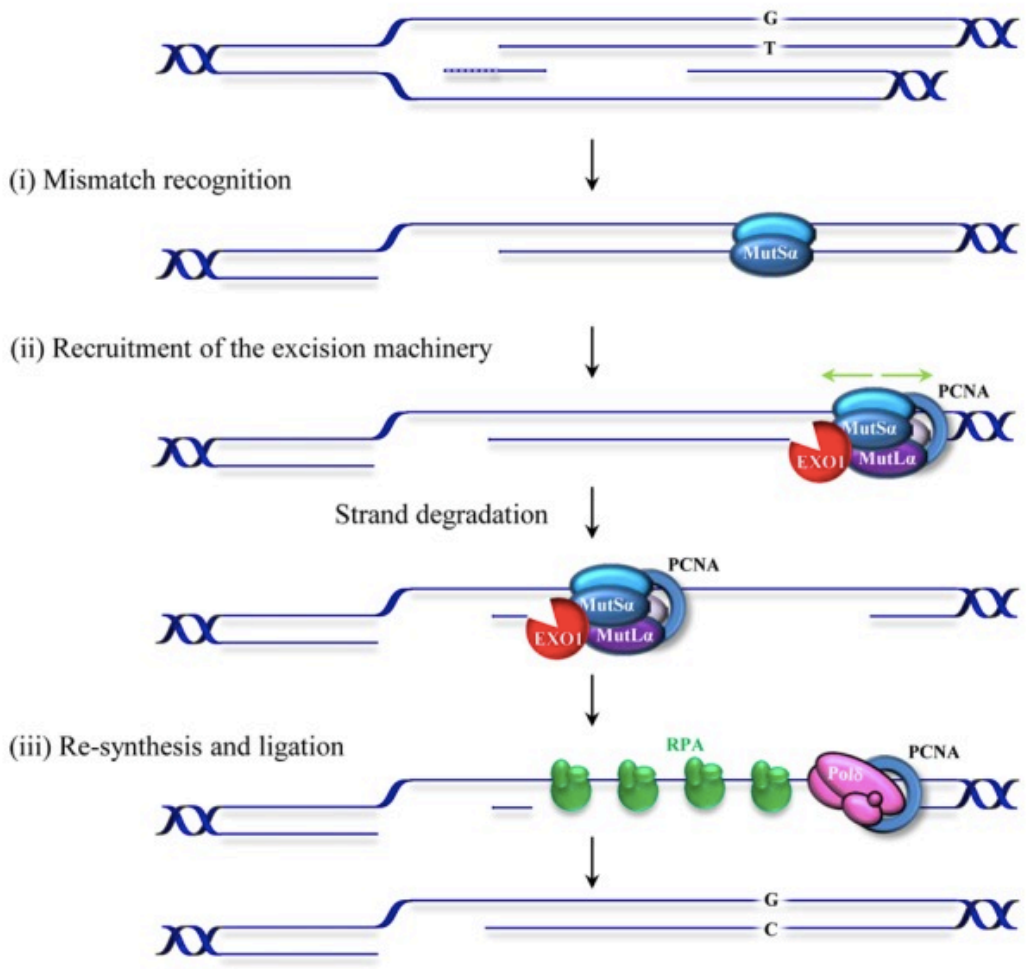


Figure 1.2. Components of the MMR pathway. Figure provided by Bak, 2014. After replication, the MSH2/MSH6 heterodimer, MutSα, recognizes a G/T mismatch on the leading strand and binds to it (i). MutSα recruits the MLH1/PMS2 heterodimer, MutLα, which can translocate in either direction. When it encounters a nick in the DNA strand, binding and loading of EXO1 by PCNA initiate degradation of the nicked strand (ii). The single-stranded gap is stabilized by RPA, filled by DNA polymerase, and sealed by LIG1 (iii). Small indels are corrected in the same way, initiated by the MSH2/MSH3 heterodimer MutSβ (32).

1.4.1 Pathogenicity of MMR pathway proteins

The four genes *MLH1, MSH2, MSH6* and *PMS2* are well-recognised CRC predisposition genes that are tested for mutations in diagnosis of LS. Inactivation of *MLH1* results in a broad spectrum of mismatches in the DNA, and LS is commonly caused by nonsense or missense mutations. The majority of identified germline mutations in *MLH1* do not inactivate the gene completely, rather result in reduced MMR efficiency of the protein, causing an elevated risk of CRC (33, 34). At least two mechanisms have been found to inactivate MLH1 function through amino acid substitution: lower expression levels of MLH1, caused by protein instability, and functional inactivation by structural alteration (35). The majority of *MLH1* variants that infer a functional defect on the protein have been located to one of two functional domains of the protein, the N-terminal ATPase domain and the C-terminal domain, responsible for interaction with PMS2 and other MMR components. Variants located between the two domains have been shown to alter or destabilize protein folding (34, 35). Mutations in *MLH1* and *MSH2* make up the majority of CRC cases. Mutations in *MSH2* are usually associated with impaired repair capability and decreased stability of the mutated protein. Mutations responsible for destabilization of the protein are generally located in the N-terminal region, while defects in ATP binding/hydrolysis are mostly observed in the C-terminal (36, 37). In 1997 it was demonstrated that mutations in *MSH6* could cause LS (38). Mutations in *MSH6* have been described in patients with MSI-L phenotypes, but more frequently it will exhibit a MSI-H phenotype. Mutations in *MSH6* cause cancer with high penetrance; however, the penetrance is reduced compared to mutations in *MLH1* and *MSH2*. Endometrial cancer is the most frequent manifestation among female *MSH6* mutation carriers (39, 40). A study that identified over 30 *MSH6* mutations in LS patients in 2002, concluded that *MSH6* analysis should be included for all patients suspected of LS, and that neither MSI nor IHC should be a selection criterion for *MSH6* analysis because of varying phenotypes (41). A study of 33 families with *MSH6* mutation from 2010 suggested that *MSH6* mutations are more common than previously assumed, and that the AC and BG are inadequate for identification of potential *MSH6* mutation carriers (42). According to MMR Genes Variant Database, up to 10% of LS cases can be explained by mutations in *MSH6* (43). The PMS2 protein was first associated with cancer in 1994, when a germline deletion in the *PMS2* gene was shown to be causative of LS (44). Subsequent studies identified several LS families with loss-of-function mutations in *PMS2,* and it was observed that the mean age of diagnosis in these families was higher than that of families with mutations in the other three MMR genes. A proposed explanation for this is that a functional MLH1-MLH3 protein can be formed in the absence of

PMS2, which exhibit a redundant MMR function (45-47). In recent studies, the risk of CRC from *PMS2* mutations has been shown to be similar to that reported of *MSH6* (48).

The role of *MSH3* in association with CRC is controversial. In its heterodimer form with MSH2, the MSH3 protein is required in MMR of insertion-deletion loops, however, the MSH3 and MSH6 heterodimer have partly redundant functions (49). Mutations in *MSH3* have been observed in connection with different cancers. A genetic screening of endometrial cancer cases in 1996 revealed a *MSH3* mutation that caused deficient MMR and MSI. After a wild-type chromosome encoding the MSH3 protein was introduced into the tumour cell line, the stability of some of the microsatellites was increased (50). It has been proposed that mutations in *MSH3* might enhance genomic instability and accelerate the accumulation of mutations in combination with mutations in other MMR genes. This was demonstrated in mouse models, by knocking out *MSH6* and *MSH3* (51-53). Mouse models have also shown that cells lacking MSH3 are defective in repair of indels, yet repair of base-base mismatches is functional and disease has a later age of onset. If the cells are depraved of both MSH3 and MSH6, the tumour development phenotype is identical to that of cells lacking MSH2 or MLH1, suggesting that MSH3 and MSH6 cooperate in tumour suppression (54).

Mutations in *MLH3* have been proposed to be involved in development of CRC because of the protein's close interaction with MLH1. Several studies have searched for germline mutations in the gene and the results have been ambiguous. While some studies claim to have discovered a strong association between *MLH3* mutations and cancer, others proclaim that the evidence is circumstantial at best (55, 56). A large screening of Dutch families with CRC in 2001 revealed germline mutations that were not prevalent among normal controls. They also reported that tumours from these cases could be both MSI positive and negative (56). A mouse study from 2002 presented that MLH3-deficient mice are MSS (57). A study of CRC patients from 2013 identified germline mutations in *MLH3*. The authors considered mutations in the gene to be modifying, contributing to disease in combination with other mutations (58).

Mutations in the *POLD1* gene have also been identified in cancer cases. The gene encodes the catalytic proofreading subunit of DNA polymerase δ, required for DNA synthesis repair. The proofreading exonuclease domain ensures the low mutation rate during DNA replication (59). A study from 2013 of two large families with predisposition to a variety of cancers identified a causal germline mutation in the exonuclease domain of *POLD1*. A third affected family was

later found to carry the same mutation. Tumours from all affected individuals were MSS. The findings suggested that decreased functionality of polymerase proofreading during replication could lead to an increased mutation rate, and consequently tumour development. Another *POLD1* missense mutation was found in a different CRC case in 2014, where no MMR defect was observed. The mutations are described as inherited in an autosomal dominant manner (60). Germline mutations in *POLD1* and *POLE* (the catalytic subunit of polymerase ε, responsible for leading strand synthesis) cause the CRC syndrome polymerase proofreading-associated polyposis (PPAP) (61). Segregation analysis has confirmed a dominant, high-penetrance predisposition to CRC, and the disease has a MSS phenotype. CRC-associated mutations in the other three POLD subunits have yet to be identified. POLD2 is an accessory subunit of the DNA polymerase δ complex, involved in interaction and stabilization of the other Pol δ subunits, PCNA, and RFC. The polypeptide shows high degree of conservation across species, suggesting an essential function (62). A study from 2010 found that the gene expression level of *POLD2* was upregulated in two subgroups of ovarian cancer, and proposed it as a potential biomarker for disease (63). POLD4 was described in 2000, and was the last subunit of the DNA polymerase δ complex to be identified (64). POLD4 is the smallest subunit of Pol δ, still, it is necessary for polymerase function and maintenance of genomic stability. The short transcript consists of 107 amino acids and is quite conserved across species (65).

The structure of the 5-protein clamp loader complex RFC, and its interaction with PCNA and DNA, was first reported in 2004 (66). The pentamer serves as a cofactor of DNA polymerase, and plays an essential role in DNA replication and repair. The subunits of RFC share structurally conserved domains, including an ATP-binding domain. Recently, the expression of the *RFC3* gene was shown to be upregulated in ovarian cancer. The study demonstrated that knock-out of *RFC3* inhibits cell proliferation and tumour cell growth, while overexpression increases cell growth and proliferation. The study established that *RFC3* has an oncogenic role, and the authors suggested the gene as a prognostic biomarker (67). A recent study also found expression of the subunit RFC4 to be significantly increased in CRC tissues compared to normal tissues. The high levels of RFC4 were associated with differentiation, tumour progression, and poor prognosis, compared to low levels of expression. The study found that loss of RFC4 suppressed CRC cell proliferation and induced cell cycle arrest (68). RFC4 has also been implicated in other cancers, and proposed as a possible target for development of cancer therapeutics (69).

The role of *EXO1* in CRC has been widely disputed. The gene has long been a suggested candidate for susceptibility to CRC, due to its function in the MMR pathway, yet its role in cancer predisposition has remained unclear. The human homologue to yeast *EXO1* was cloned in 1998 and the gene was proposed to be involved in the pathogenesis of LS, because of the protein's interaction with MSH2 (70). In 2001, 33 families with LS and 225 index patients suspected of LS were screened for germline mutations in *EXO1*. The results indicated an association of germline *EXO1* variants with LS, and encouraged further studies (71). A subsequent study revealed some of the identified mutations to be polymorphisms, present in healthy individuals, thereby undermining the relevance to LS (72). The gene has been extensively studied in *Saccharomyces cerevisiae* where mutations cause a weak mutator phenotype. This suggested that mutations in human *EXO1* could be less frequent or of a lower penetrance than that of the higher order MMR complexes. A study of mouse models from 2003 with inactivated *EXO1* revealed that silencing of the gene causes defective DNA MMR, much similar to the repair defects observed in *MSH6* mutant mice, which increase the risk of cancer (72, 73). A mouse model study from 2013 reported that EXO1's catalytic role is essential for DNA damage response and repair, chromosomal stability, and tumour suppression. It was also found that the structural role of EXO1 alone is critical for MMR (74). Another functional study from the same year determined that EXO1 preferentially repaired mismatches generated by DNA polymerase α. This study also reported that most replication errors made by Pol α could be repaired in the absence of EXO1 by other nucleases. Repair in absence of EXO1 was found to be dependent on the presence of MSH2 (75). Several studies have focused on determining the role of single amino acid changes in EXO1, and some variants have been shown to increase cancer risk, while others have not. The ambiguous results from these studies indicate that *EXO1* could be a moderate-risk gene predisposing to CRC. It would seem that the EXO1 protein has some overlapping functions with other nucleases. The effect of mutations in this gene alone may therefore not be very serious, but in combination with mutations in other genes, it could have a greater effect. The gene may also confer different risk in different populations (76, 77).

1.5 Next-generation sequencing

Next-generation sequencing (NGS) is one of the most significant advances in technology within biological science of the last 30 years. The term next-generation sequencing is used to describe the high throughput sequencing technology, able to sequence millions of different DNA templates in parallel. NGS has emerged as a practical method of obtaining patient-specific genetic data for targeted therapy in cancer medicine, and has replaced much of the popular Sanger sequencing method over the course of a few years. Due to the genetic aspect of the disease, medical researchers within the cancer field have been eager to utilize the technology (78, 79).

The Sanger sequencing method was in practice the only DNA sequencing method used for 30 years, since Fred Sanger and Alan R. Coulson published their paper on methodical rapid determination of DNA sequence in 1977. The Sanger technology is considered the gold standard in genetic sequencing and is widely adopted in laboratories around the world. Nevertheless, the semi-automatic technology has its limitations, both in throughput, scalability, speed, and resolution, making multigene panels laborious and expensive. The first-generation sequencing technology is too limited to meet the demands of today's complex genomic research. The Human Genome Project was sequenced using Sanger sequencing and the project took over 10 years, costing nearly three billion dollars. To overcome these barriers, new technology has been developed over the last ten years. In 2005, the sequencing-by-synthesis technology developed by 454 Life Sciences was launched, which caused rapid advancement of sequencing platforms into the second-generation. The following years the method was developed further, and over 100 research articles were published in less than two years. This lead to a revolution in sequencing technology: from the first-generation Sanger sequencing of a few DNA fragments, through second-generation platforms employing clonal amplification of DNA templates on a solid matrix and cycle sequencing, followed by third-generation platforms with single molecule polymerase chain reaction (PCR) free protocols and cycle-free chemistry, to the achievement of massive parallel sequencing (79-81).

The NGS technology enables researchers to study biological systems at a level that was not previously possible, making large-scale sequencing both accessible and practical (82). NGS has provided great benefits, with a method 50 times more throughput than Sanger sequencing and the costs reduced to $1/6^{th}$. Many laboratories are now considering it for routine diagnostic

uses and at is gradually being implemented in clinics worldwide. Sanger sequencing is still very useful, as it is a robust and reliable method of detecting DNA variations on a smaller scale. The NGS reaction enables deep sequencing of specifically targeted regions, exome sequencing and rapid whole genome sequencing (WGS). Exome sequencing is used to limit the sequencing to only include the coding regions of the genome. Targeted enrichment sequencing, where specific genes are selected for sequencing, allows the reaction to focus on particularly relevant areas, thereby increasing sensitivity and specificity of the test. NGS provides the ability to barcode and pool numerous samples together to find rare variants that are missed by other approaches, and still obtain high sequence coverage during a single run (78, 82, 83). For CRC patients with an unknown genetic cause of disease, NGS can be used to search for pathogenic mutations within a panel of candidate genes, instead of laborious testing of each gene, one-by-one. NGS has the potential of expanding the genetic repertoire used for detection and treatment of CRC. This will decrease genetic investigation time of the patients, leading to sooner diagnosis. Better understanding of CRC pathogenesis will also bridge a new era of personalized medicine for cancer patients (79).

1.6 Aim of project

LS is the most common predisposing colorectal cancer syndrome. Current diagnosis of LS is based on detection of mutations in one of four mismatch repair genes, *MLH1, MSH2, MSH6,* or *PMS2*. Many patients meet the AC and/or BG for diagnosis with LS, but do not have mutations identified in any of these four genes. Up to 30% of all CRC cases is caused by a hereditary genetic susceptibility, yet no more than 10% of cases have identified germline mutations. This means that around 20% of familial CRC cases are genetically unexplained. The genetic background of disease in these patients remains elusive, because further genetic testing is unavailable in the clinic. The specific segregation pattern shown in many of the affected families demonstrates that there are undiscovered high-risk mutations causing CRC. In other cases, there is a belief that risk can be explained by co-inheritance of multiple moderate- or low-penetrant variants (84). These variants can confer an increased risk of cancer in combination with more high-penetrant variants, or cause an additive risk along with other low-penetrant variants. Low-risk variants could also be contributing to disease together with environmental factors.

In this study, samples from 95 patients with CRC have been screened for causal mutations using NGS technology. Each patient fulfils the AC and/or BG, but clinical genetic screening of *MLH1, MSH2, MSH6,* and *PMS2* has failed to reveal the cause of disease. Our hypothesis is that pathogenic mutations exist somewhere within the 22 genes involved in the MMR pathway in these patients. MMR is crucial for genomic stability and defects in these proteins could quickly lead to an accumulation of mutations. The MMR system has a documented role in development of CRC, and it may be that predisposing variants in genes other than *MLH1, MSH2, MSH6,* and *PMS2* in the MMR pathway can be causative of CRC. By NGS of a gene panel the aim is to uncover candidate variants for the cause of disease in these patients. Identification of the underlying genetic mechanism for disease will contribute to the elucidation of fCRC-X. Disclosure of more susceptibility alleles to CRC is of tremendous importance for diagnosis and treatment of patients with hereditary syndromes. Knowledge of pathogenic variants will provide the patient and kindred the option of genetic testing and surveillance to detect CRC at an early stage. Elucidation of causative agents is also a step in the right direction towards cancer therapeutics and personalized medicine.

2. Materials and methods

2.1 Chosen genes and DNA samples

The patients in this study are selected males and females with a family history of CRC, and/or diagnosis with CRC at an early age. Each patient fulfils the AC and/or the BG. They have received genetic counselling and testing, yet no mutations have been detected through normal diagnostic screening. DNA samples from 95 patients have been used to sequence a gene panel consisting of 124 genes. Included in the sequencing reaction were the gene's exons, to get the protein-coding sequences, as well as ten base pairs of the upstream and downstream intronic flanking regions of each gene, to examine splice sites. The gene panel was sequenced in a single reaction on the Illumina HiSeq2500 Next-generation sequencing platform. Of the 124 genes, 22 MMR-related genes, *EXO1*, *LIG1*, *MLH1*, *MLH3*, *MSH2*, *MSH3*, *MSH6*, *PCNA*, *PMS1*, *PMS2*, *POLD1*, *POLD2*, *POLD3*, *POLD4*, *RFC1*, *RFC2*, *RFC3*, *RFC4*, *RFC5*, *RPA1*, *RPA2* and *RPA3*, have been studied in this master project. The further 102 genes included in the gene panel have been studied by co MSc student Ann-Therese Ali. A flow scheme of the complete process of this master's project is shown in figure 2.1.

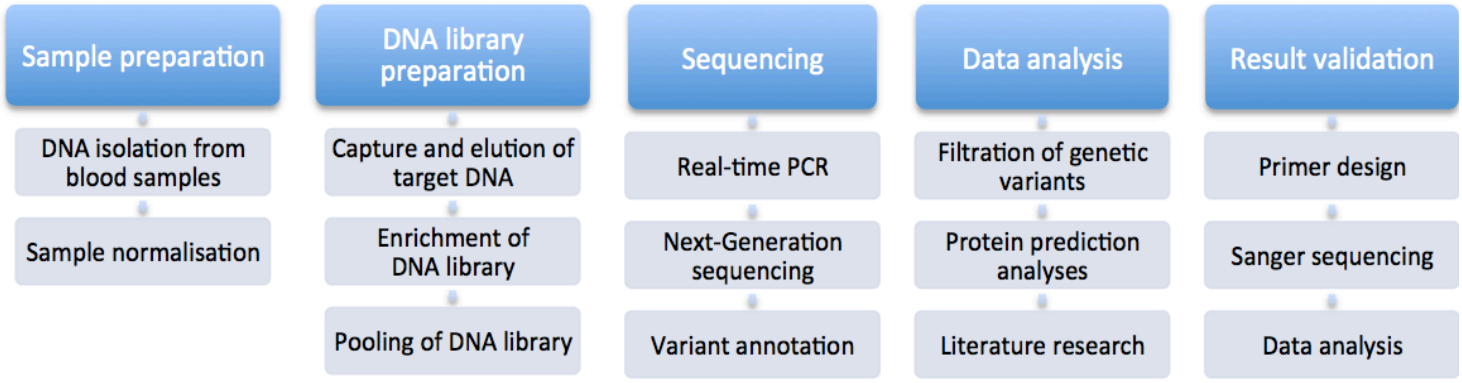


Figure 2.1. Flow scheme illustrating the methodical process of this master project.

2.2 Sample preparation

The biological material used in this project is gDNA isolated from ethylenediamine tetraacetic acid (EDTA) blood samples from the clinical laboratory at St. Olavs Hospital in Trondheim and the colorectal cancer research biobank in Mid-Norway. A total of 95 patient DNA samples were prepared for this project. The DNA was isolated using the iPrep™ PureLink™ gDNA Blood Kit from Invitrogen, in accordance with the Invitrogen user manual (85). The kit utilizes magnetic Dynabeads® MyOne™ SILANE to isolate gDNA from whole blood. The DNA concentration of the samples was measured using Nanodrop® ND-1000 spectrophotometer from Thermo Scientific, in accordance with the user manual (86). To confirm the Nanodrop measurements, the samples were measured again with Qubit 2.0 Fluorometer from Life Technologies, in accordance with the user manual (87).

2.3 DNA library preparation

A DNA library of 95 patient samples was prepared by target enrichment. Target enrichment can be performed using amplicon or capturing methods to amplify the DNA samples in a library for sequencing. In this experiment, the HaloPlex Target Enrichment System protocol for Illumina sequencing by Agilent Technologies was utilized, and the basic workflow is shown in figure 2.2. This is a selective circularisation-based, amplicon sequencing method. The technology is a further development of the selector probe principle, where genomic DNA is digested by specific restriction enzymes, generating short fragments. Restriction sites flank each target region, and eight restriction enzyme digests are performed in parallel, which capture all amplicons of interest in templates of 100-500 base pairs. Biotinylated oligonucleotide adaptors, specific to the Illumina NGS platform, hybridize to each end of a fragment, circularizing it. The fragment is immobilized and retrieved using magnetic streptavidin beads. The unbound DNA is washed away and the library is recovered by an elution step. The targeted DNA fragments are amplified by PCR and the PCR products are pooled to form a DNA library, which can be sequenced on the NGS platform. The amplicon method is popular within the clinic because of the simple workflow and short preparation time (83, 88, 89). The DNA enrichment was done according to the HaloPlex Target Enrichment System Protocol from Life Technologies (88).

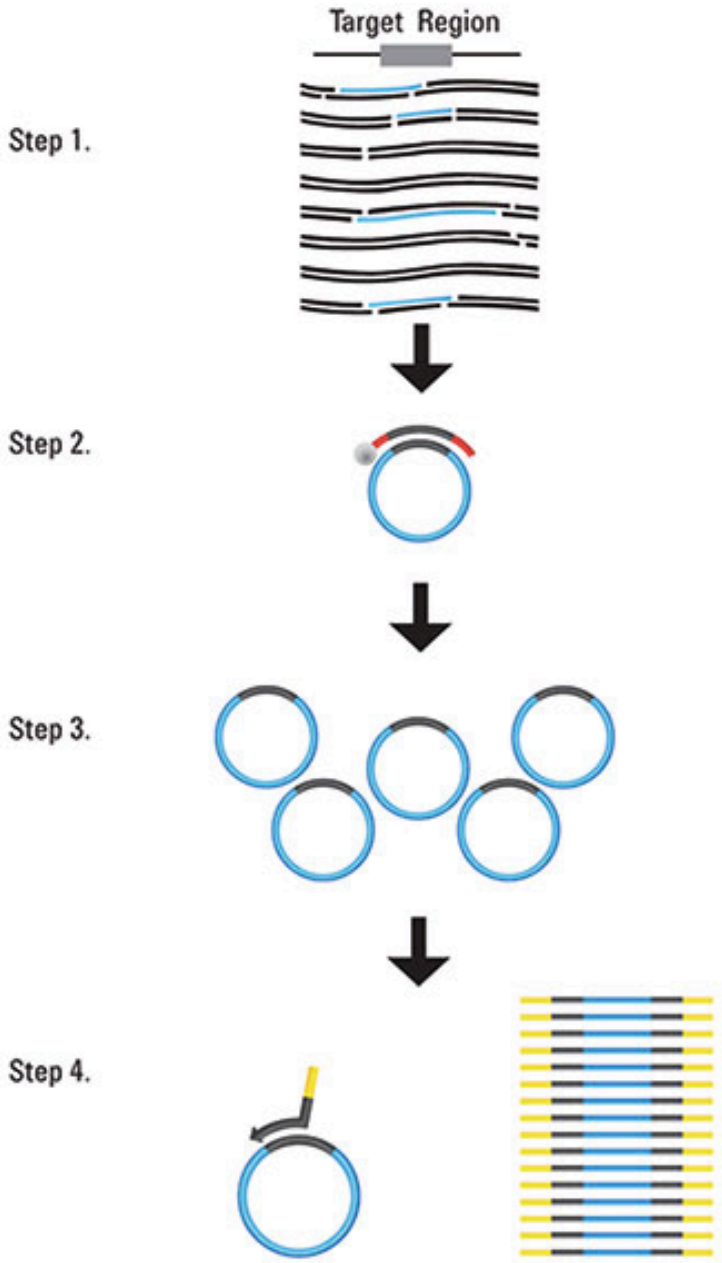


Figure 2.2. Schematic workflow of HaloPlex target enrichment. Figure provided by Agilent Technologies. Step 1: Digestion and denaturation of input gDNA containing target regions. Step 2: Hybridization of probe library, resulting in circularized gDNA fragments with incorporated indexes. Step 3: Capturing of biotinylated target DNA-probe hybrids using streptavidin-coated magnetic beads. Step 4: PCR amplification of targeted fragments, producing a target-enriched sample ready for sequencing (88).

2.3.1 Deviations from the HaloPlex Target Enrichment System Protocol

Step 1. Evaluation of DNA quality

The protocol suggests that the DNA size distribution is verified for all samples using gel electrophoresis, to ensure their quality and confirm that they are not degraded. The DNA size distribution of samples 1-10, and sample 86 was validated with gel electrophoresis. Samples 1-10 are the oldest DNA samples used in this study, collected from 2002 to 2004, and sample 86 is the newest, collected from 2014. The older the sample is, the more likely it becomes that the DNA will be degraded. It was therefore judged to be sufficient to test the oldest samples, and the newest for control.

Step 9. Validation of DNA enrichment

The DNA library was validated with the 2100 Bioanalyzer. Because of high DNA concentration, the samples were diluted 1:3 with Elution Buffer, containing Tris-HCl, and measured again. The result from the second measurement was used to pool all samples into one DNA library to a concentration of 3,62 ng/µL, validated on the bioanalyzer.

2.4 Sequencing

After pooling of the DNA library, the concentration of parallel dilutions of the library was measured using quantitative real-time PCR (qPCR). The concentration was used to dilute the DNA library to an optimal DNA concentration for the sequencing reaction. The DNA library was sequenced on the Illumina HiSeq2500 NGS platform in accordance with Illumina sequencing user manual (90).

2.4.1 Sequencing on the Illumina platform

The input material for NGS on the Illumina platform is double stranded (ds) DNA, converted into a sequencing library. Adaptors, consisting of synthetic DNA, which serve as primers for amplification and sequencing are ligated to the ends of the DNA fragments. The templates are immobilized by random binding to a flow cell for solid-phase amplification. Unlabelled nucleotides and enzyme are added to the reaction to initiate bridge PCR amplification. During bridge PCR, nucleotides are incorporated to build dsDNA bridges on the solid surface and then denatured to leave single stranded (ss) DNA templates anchored to the surface. Millions of micro sequencing reactions are carried out in parallel, creating up to 1000 identical copies

of each template molecule. Densities of ten million single-molecule clusters are achieved per square centimetre (78, 91).

Illumina utilizes the sequencing by synthesis (SBS) technology, in which four deoxynucleotides (dNTPs) labelled with fluorescent tags are used to sequence the DNA clusters in parallel. Fluorescently labelled nucleotides, primers, and polymerase are added to the reaction. A laser makes each cluster emit fluorescence and the first base is identified. For each cycle, a single dNTP is added and detected before the tags and reagents are removed again. The sequence content is read stepwise, nucleotide-by-nucleotide, and over multiple cycles the bases are recorded to build the linear sequence (78, 91). The Illumina sequencing process is shown in figures 2.3-2.5.

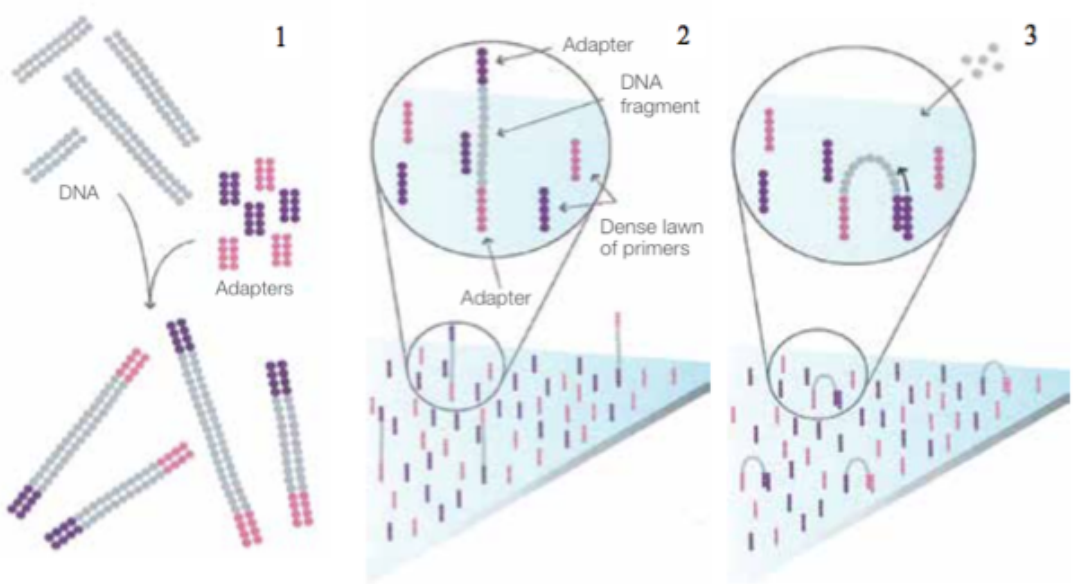


Figure 2.3. Illustrations provided by Illumina. Adaptors are ligated to the DNA templates (1), and single-stranded fragments are immobilized on a flow cell (2). Bridge PCR is initiated by adding enzyme and unlabelled nucleotides to the reaction. The adaptors bind to primers on the surface of the flow cell, and complimentary nucleotides hybridize to the template (3) (91).

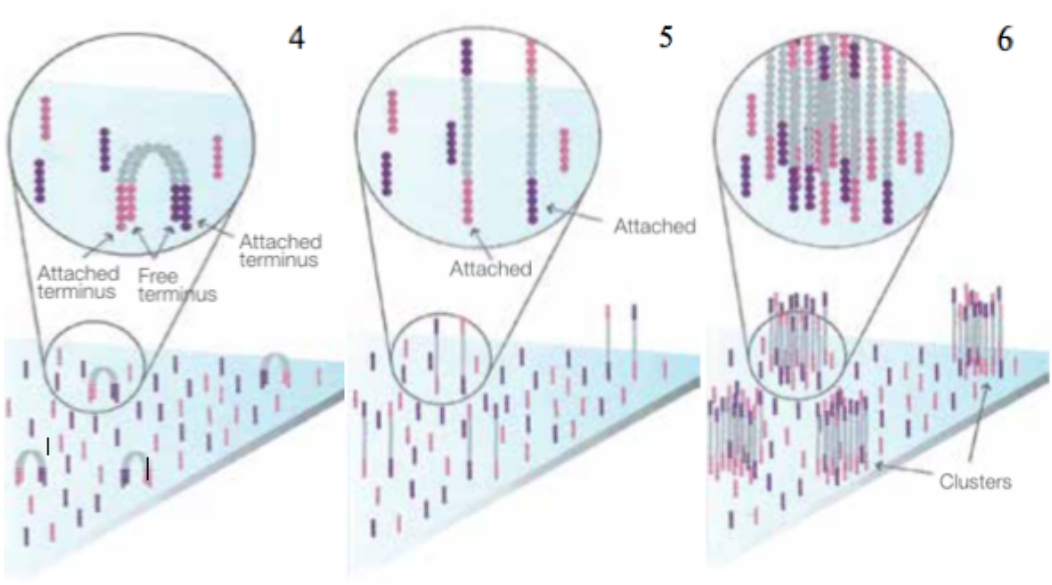


Figure 2.4. Illustrations provided by Illumina. The enzyme builds double-stranded bridges during bridge PCR (4). Denaturation leaves single-stranded templates anchored to the substrate (5). Millions of dense clusters of dsDNA are generated in each channel of the flow cell (6) (91).

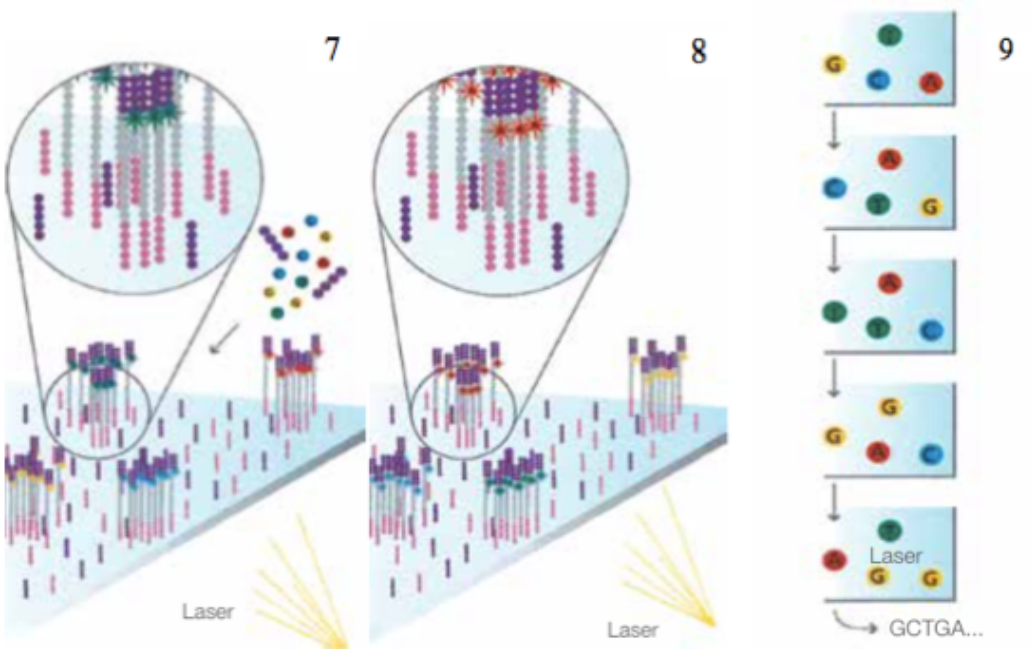


Figure 2.5. Illustrations provided by Illumina. The sequencing reaction is initiated by adding four fluorescently labelled deoxynucleotides (dNTPs), primers and DNA polymerase. A laser causes each cluster to emit fluorescence, and the first base is detected (7). In the next cycle the process is repeated with incorporation of four labelled dNTS, and the second base is identified (8). Continued cycle sequencing determines the base sequence in the DNA fragments, one base at a time (9) (91).

2.5 Data analysis

The initial analysis of the raw data from the sequencing reaction was performed by the Bioinformatics core facility at Norwegian University of Science and Technology (NTNU), where a data processing pipeline for the output was established. After variant annotation, the variants were filtrated on the basis of frequency in the population, effect on amino acid, and sequence reaction quality. The variants remaining after filtration were subjected to further analysis with protein prediction tools and literature research, to examine the pathogenic likelihood of the mutations.

2.5.1 Data analysis pipeline

NGS data analysis involves four principal steps: base calling, read alignment, variant calling, and variant annotation. The identification of the specific nucleotide in each sequencing read, base calling, is integrated into the Illumina platform software. The raw data output from the sequencing reaction is a large collection of sequence reads. For the reads to be aligned accurately, reference sequences are assigned to the data. Read alignment is done by correctly positioning the sequencing reads along the reference sequences. The human genome hg19 (UCSC assembly, February 2009) was used as reference in this project, and the alignment was done using the Burrows-Wheeler-Aligner (92). Comparison of the analysed sequences to the reference sequences enables detection of variations within the analysed genome. Identification of these variations is referred to as variant calling. In this project the Gene Analysis ToolKit (GATK) Best Practices Recommendations was used to call variants (93). The recommendations are a series of steps taking the aligned data sets to variants. The intermediate steps include local realignment around indels, recalibration of quality scores, and quality control of called variants. To analyse regions targeted in the sequencing experiment, a list of these regions was used as additional information when running GATK, to reduce running time of the pipeline. To define the limitations of the test, regions of low coverage were located. By tracking positions with absent data or an ambiguous call, areas of poor sequencing quality can be identified. After the variations have been identified, information of each detected variant is added for variant annotation (81). Variants were annotated using the annotation software ANNOVAR (94).

2.5.2 Filtration of variants and prediction tools

The variants were filtrated using the variant filtration tool Filtus version 0.99-91 (95). Filtus offers a statistical evaluation of shared variants across individuals and is well suited for Mendelian disease mapping. The genetic variants that were detected in the MMR-associated genes of the 95 patients were initially filtrated against 1000 Genomes Project with minor allele frequency (MAF) <0.01, to avoid common, benign variations. All synonymous variants were discarded, as these are considered to be harmless polymorphisms in the majority of incidences. Variants that did not pass the quality control filters established in the pipeline were removed, as these are likely to be false positives. The minimum sequencing coverage was set to 20X, because variant calls are more reliable as coverage increases, and low coverage increases the risk of both false negatives and false positives (81).

The variants that passed filtration in Filtus were subjected to further investigations with the mutation analysis software Alamut Visual (interactive Biosoftware, Rouen, France). The software enables the user to navigate at nucleotide level within the gene, and explore the predicted consequences of possible mutations. Alamut provides protein prediction via built in prediction tools, which enable evaluation of variant consequences at protein level. The protein prediction tools utilized through Alamut were Align-GVGD, SIFT, MutationTaster, and PolyPhen-2. Align-GVGD combines protein multiple sequence alignment with the biophysical properties of amino acids, to predict if a genetic missense substitution is neutral or deleterious. Input sequence data is subjected to graded classification, based on risk assessment estimates. The output is one of seven classes; C0, C15, C25, C35, C45, C55, and C65, with increasing pathogenicity risk. The risk estimate ranges from less than 0.90 in class C0, to 4.00 or higher in class C65 (96). Sorts Intolerant From Tolerant (SIFT) is a sequence homology-based tool that predicts whether an amino acid substitution will have a phenotypic effect. The program is based on conservation of positions in alignment of a protein family. The SIFT output is a score ranging from 0 to 1. The variant is predicted to be deleterious if the score is less than or equal to 0.05 (97). MutationTaster is a web application for evaluation of the deleterious potential of DNA sequence alterations. Information from biomedical databases is used to analyse evolutionary conservation, splice-site changes, protein features, and mRNA features, and to predict disease potential. The prediction is given as *disease causing* or *polymorphism*, with a p-value from 0 to 1, indicating the confidence of the prediction (98). Polymorphism Phenotyping version 2 (PolyPhen-2) predicts the functional effects of human nonsynonymous single nucleotide polymorphisms (SNPs). Physical

properties of amino acids and comparative analysis of sequences is used to predict the impact on human proteins. The prediction tool calculates the probability of a mutation being damaging, and scores it from 0 to 1, 0 being benign, 1 being probably damaging. The scoring system is based on pairs of false positive rate thresholds (99). The Grantham's distance is used to show how similar the properties of two amino acids are, to predict how well they can substitute each other. The distance between amino acid residues is calculated using a formula for combining properties that correlate best with residue substitution frequencies. Composition, polarity, and molecular volume are taken into account, and the score ranges from 0 to 215. A low score signifies that the physiochemical distance between the residues is small, and a high score signifies a great distance (100). The variants were researched within the Leiden Open Variation Database (LOVD) v.3.0, to identify previous observations of the variants (101). LOVD uses the five-tired International Society for Gastrointestinal Hereditary Tumours (InSiGHT) classification system for CRC. Identified variants are classified in one of five classes:

Class 1: not pathogenic
Class 2: likely not pathogenic
Class 3: uncertain pathogenicity
Class 4: likely pathogenic
Class 5: pathogenic.

Variants of class 3 have a pathogenic likelihood of 5-95% (102). Universal Protein Resource (UniProt), Prosite, and Conserved Domains Database (CDD) was used to assess the affected protein domains (103-105). dbSNP, a SNP database provided by National Center for Biotechnology Information (NCBI), was used to examine variants with a reference sequence (rs) number (106).

2.6 Result validation

The DNA variants that were predicted to be pathogenic by two or more prediction tools were selected for verification. These are the variants that continue to be the most likely candidates of predisposition to CRC. DNA from patient blood samples was isolated, and the variants were validated using Sanger sequencing PCR and the ABI 3130xl Genetic Analyzer from Applied Biosystems. Primers were designed using the Primer Designer™ tool from Life Technologies (107) and Primer-BLAST, an online primer designing tool from NCBI for finding specific primers (108). The primers used for each variant are shown in Appendix 1. For each variant, a DNA fragment of 200-600 nt containing the variant was sequenced. SeqScape software v3.0 from Applied Biosystems was used to analyse the data from the sequencing and interpret the results. The identified variants were confirmed to be true positives. Two additional variants, in *LIG1* and *RPA1*, which were excluded from the results because of low coverage, were confirmed to be false positives.

2.7 Materials

The Equipment, kits, and reagents used in this study are listed in table 2.1 and 2.2.

Table 2.1 Equipment (name, vendor, and ID) used in this study.

Description	Vendor	ID
Benchtop microcentrifuge, Galaxy Mini	VMR™ International	Cat: 93000-196
Benchtop plate centrifuge 5810R	Eppendorf	
Benchtop rotator FSR20	Grant Boekel	
2100 Bioanalyzer	Agilent Technologies	Cat: G2940CA
Buffer, 1x and 10x with EDTA		
Buffer 5x		
Capillary Array for 3730 DNA Analyzer	Life Technologies	
ABI DNA Analyser 3730	Applied Biosystems, Hitachi	Part.no. 625-0010
GeneFlash Bio Imaging system	SynGene	
iPrep™ purification instrument	Invitrogen	Cat: 10000
iPrep™ tubes	Invitrogen	
DynaMag-2 magnetic bead separator	Life Technologies	Cat: 12321D
Dynal MPC® Magnetic Particle Concentrator	Life Technologies	Cat: 120.27
MicroAmp™ optical adhesive film	Applied Biosystems®, Life Technologies	Cat: 4311971
MICROLAB STARLET Pre PCR robot	HAMILTON	
Multichannel pipettes (10 µL and 100 µL volume)	Biohit	
MS1 Minishaker vortexer	IKA®	
Finnpipette® Multichannel pipettes	ThermoScientific	
Nanodrop® ND-1000 spectrophotometer	ThermoScientific	
Pipetboy Comfort	IBS Integra Biosciences	
Plastic pipette, sterile, disposable	Sterilin	
Qubit 2.0 Fluorometer	Invitrogen™, Life Technologies	Cat: Q32866
Qubit assay tubes	Invitrogen™, Life Technologies	Cat: Q32856
Thermal cycler 2720	Applied Biosystems®, Life Technologies	Cat: 4359659
MixMate® vortex mixer with PCR 96 tube holder	Eppendorf	Cat: 5353000.014

Table 2.2 Kits and reagents (name, vendor, and ID) used in this study.

Description	Vendor	ID
Acetic acid solution 2 M		Lot: SLBH6779V
Agencourt AMPure® XP Kit	Beckman Coulter Genomics	Lot: 14060800
AmpliTaq Gold® 360 MasterMix	Applied Biosystems®, Life Technologies	Lot: 1405038
A'SAP PCR Cleanup Exonuclease 1	ArcticZymes	Lot: 1422
A'SAP PCR Cleanup Alkaline Phosphatase	ArcticZymes	Lot: 1422
BigDye® Terminator v3.1 Cycle sequencing kit	Applied Biosystems®, Life Technologies	Lot: 1405242
DNA Molecular weight Marker IV, 0.07 – 19.3 Kb		Lot: 11799634
E-Gel iBase™ 2% agarose	Invitrogen	Lot: B16074
Elution Buffer (EB)	Qiagen Gmblt	Lot: 145046057 Cat: 19086
Ethanol 100%, molecular biology grade		
Eurogentec universal M13 primers	NIMAGEN	Sense Lot: 5457061 Antisense Lot: 5457062
HaloPlex Target Enrichment System Kit	Agilent Technologies	Lot: 0006246792 Cat: 5190-5534
HCl, 0,3 M, for Nanodrop		
Herculase II Fusion Enzyme DNA Polymerase	Agilent Technologies	Lot: 0006212697 Cat: 600677-51
High Sensitivity DNA Chips	Agilent Technologies	Lot: SF04BK50
High Sensitivity DNA Kit	Agilent Technologies	Lot: 1420
iPrep™ PureLink™ gDNA Blood Kit	Invitrogen	Lot: 1603453
NaOH, molecular biology grade, 10 M		Lot: 1168043
Nuclease-free Water		
POP-7™ Polymer for 3730 DNA Analyzer	Life Technologies	
Qubit® dsDNA high sensitivity Assay Kit	Invitrogen	
SAM™ Solution	Applied Biosystems®, Life Technologies	Lot: 1412051
Tris-HCl, 10 mM, pH 8.0		
Tris, 10 mM, pH 7.5, for Nanodrop		
XTerminator™ Solution	Applied Biosystems®, Life Technologies	Lot: 1412057

3. Results

3.1 Sample preparation

The DNA concentration of 95 patient samples was measured using Nanodrop and Qubit, and the measurements from both instruments are listed in Appendix 2. A graphical illustration of how the measurements coincide is shown in figure 3.1.

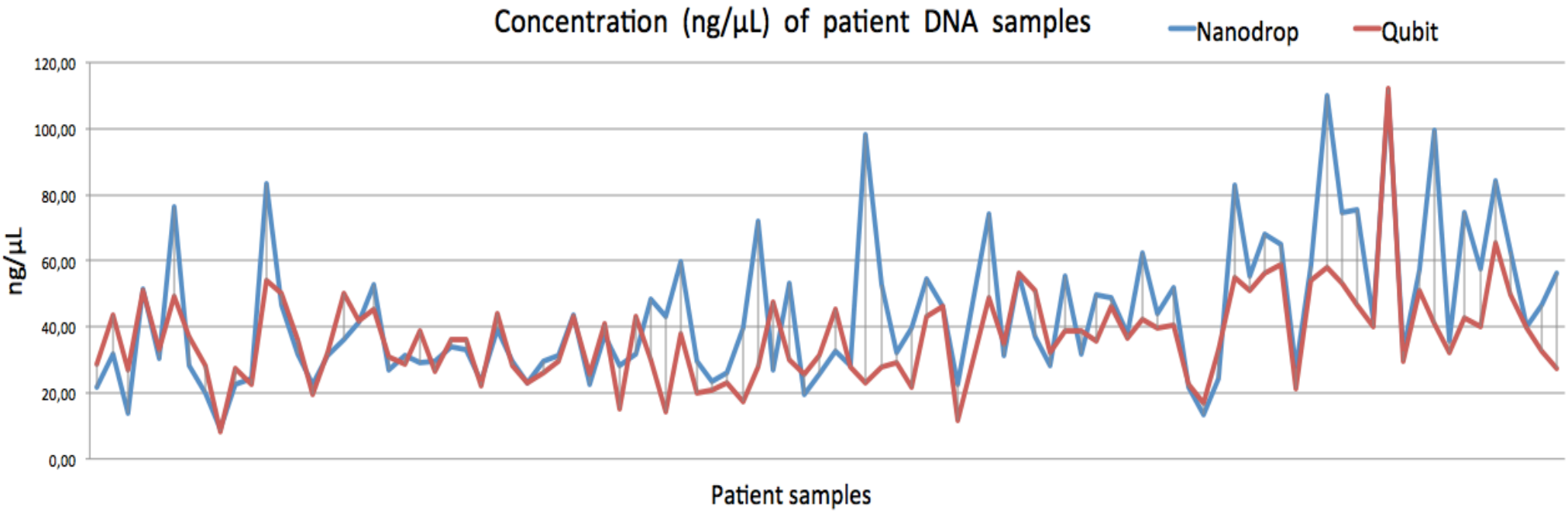


Figure 3.1. Graphical illustration of the DNA concentration measurements from 95 patient DNA samples, measured with Nanodrop (blue graph) and Qubit (red graph).

The concentration measurements show some degree of inconsistency between the two instruments, as is evident from the graph. The instruments are also somewhat inconsequent to their respective results, as the test was performed in parallel for each sample and the reported concentration is an average of this. To a large extent, the graphs show conformity; still, some of the measurements are quite divergent. This could indicate that the DNA was not equivalent in the sample solution, perhaps due to insufficient vortexing. The dilution of the samples was calculated based on an average of three parallel measurements on both instruments, and the average between them. Because of the questionable concentration results, it is not certain that all samples were diluted to the correct concentration. This should, however, have little effect on the sequencing results, as the DNA concentration of the samples in the library was calibrated again before pooling and sequencing.

3.2 DNA library preparation

The DNA concentration of each sample in the DNA library is listed in Appendix 3. The DNA concentration of the pooled library was measured to an average of 3,4 ng/µL. The DNA concentration was measured again before sequencing, using quantitative qPCR, to ensure the right conditions in the sequencing reaction. The DNA concentration of the pooled library and the Ct-value is shown in Appendix 3.

3.3 Sequencing

The average number of detected variants across the 22 MMR-associated genes, in each of the 95 patient samples, was 33. The number of detected variants in each sample can be found in Appendix 4. The number of variants is fairly even for all patients, the highest number of variants detected in a sample was 42, and the lowest number of variants detected in a sample was 26. The majority of these variants are synonymous single nucleotide substitutions. For each sample, an average of 11 genes were found to contain mutations. The mean coverage of the sequencing reaction was 258X, 13 times higher than 20X, which was set as the minimum acceptable coverage. On average, 86,72% of the target regions in all samples was successfully sequenced with coverage above 20X. All genes and exons were reported as covered, except for the first 250bp of *POLD4,* making up two and a half exons, which was poorly covered in the majority of the samples. This region is especially GC-rich, which could have made the probe enrichment of this region difficult. This is not unusual for promoter regions. Average coverage and % of target regions covered in each sample are shown in Appendix 4.

3.4 Data analysis

After sequencing, a total of 177 unique variants were identified in 21 of the 22 genes tested, across the 95 patients. No variants were detected in *RPA2*. The variants were subjected to filtration in Filtus, excluding variants with MAF >0.01 in 1000 Genomes Project and all synonymous variants, thus reducing the number of unique variants to 83, across 17 genes. For all variants, the quality of the sequencing reaction was inspected and variants of poor quality and low coverage were discarded, as these variants are likely to be false positives. The variants that passed these filters were inspected in Alamut software and subjected to protein prediction. At this stage, 39 variants across 13 genes were evaluated, and a complete table of variants with predictions is shown in Appendix 5. Variants classified as class one or two in LOVD were excluded because they are considered benign. Variants predicted to be pathogenic by two or more prediction tools were subjected to further evaluation. The result was 20 unique variants distributed across 12 genes among 22 of the 95 patients. For each step of filtration, the number of unique variants present in each gene amongst the 95 patients is shown in Table 3.1.

Table 3.1. The number of unique variants in each gene are shown before filtration and after each step of additive filtration. The variants are filtrated against synonymous variants, minor allele frequency (MAF) >0.01 in the 1000 Genomes project, poor sequencing quality and coverage <20X, and after protein prediction tools assessed the variants pathogenicity.

Gene	Filter				
	None	Synonyms	MAF >0.01	Quality / coverage	Protein prediction
EXO1	22	19	10	8	4
MSH3	15	12	7	2	1
MLH3	17	15	8	2	1
RPA1	14	9	8	0	0
PMS2	17	14	7	4	2
MSH6	20	13	11	5	1
RFC1	6	3	3	2	0
LIG1	8	3	1	0	0
RFC4	4	2	2	2	2
POLD1	13	7	5	4	2
POLD3	3	1	1	0	0
RPA3	3	2	2	0	0
POLD2	11	6	6	2	2
RFC2	2	0	0	0	0
RFC5	1	0	0	0	0
MSH2	8	4	3	3	2
MLH1	6	4	4	3	1
POLD4	4	4	4	1	1
PCNA	1	0	0	0	0
PMS1	1	1	0	0	0
RFC3	1	1	1	1	1
RPA2	0	0	0	0	0
Total variants	177	120	83	39	20

The 20 variants that were predicted by protein prediction tools to have deleterious consequences on their respective proteins were confirmed with Sanger sequencing, and are presented in Table 3.2, with nomenclature and predictions.

Table 3.2. Candidate risk-variants found in this project. The table lists affected gene with reference sequence, nomenclature of variants, patient number, and the predicted consequences of the mutations based on scores from the protein prediction tools, PolyPhen-2, Align-GVGD, SIFT, and MutationTaster, as well as the Grantham's distance between the residues.

Gene	Variant	Ref. seq. nr	Sample	Protein prediction score				
				PolyP[1]	GVGD[2]	SIFT[3]	MutT[4]	G.dist[5]
EXO1 NM_003686	c.745G>A p.Asp249Asn	rs61750993	6, 25, 43	0.946	C0	0.02	1	11
EXO1 NM_003686	c.1828A>G p.Ser610Gly	rs12122770	8	0.004	C0	0.04	0.739	29
EXO1 NM_003686	c.458C>T p.Ala153Val	N/A	28	1.000	C0	0.02	1	64
EXO1 NM_003686	c.82G>A p.Ala28Thr	N/A	30	1.000	C0	0.00	1	31
MLH1 NM_000249.3	c.739T>G p.Ser247Ala	rs63750948	14, 20	0.948	C0	0.02	1	49
MLH3 NM_001040108.1	c.885delG p.His296Thrfs*12	N/A	14	N/A	N/A	N/A	N/A	N/A
MSH2 NM_000251.2	c.138C>G p.His46Gln	rs33946261	9	0.994	C15	0.00	1	24
MSH2 NM_000251.2	c.815C>T p.Ala272Val	rs34136999	13	0.869	C0	0.01	1	64
MSH3 NM_002439.4	c.2336G>A p.Arg779His	rs199791286	36	1.000	C0	0.00	1	0
MSH6 NM_000179.2	c.2511C>G p.His837Gln	N/A	80	0.999	C15	0.00	1	24
PMS2 NM_000535.5	c.1004A>G p.Asn335Ser	rs200513014	24	1.000	C45	0.00	1	46
PMS2 NM_000535.5	c.137G>A p.Ser46Asn	rs121434629	76	0.996	C45	0.00	1	46
POLD1 NM_001256849.1	c.2861C>T p.Thr954Met	rs374016016	16	0.999	C0	0.01	1	59
POLD1 NM_001256849.1	c.3221G>A p.Arg1074Gln	N/A	2	1.000	C0	0.03	1	0
POLD2 NM_001127218.1	c.322C>A p.Leu108Met	N/A	33	1.000	C0	0.00	1	14
POLD2 NM_001127218.1	c.892G>A p.Glu298Lys	N/A	57	1.000	C0	0.74	0.986	34
POLD4 NM_001256870	c.79G>A p.Glu27Lys	rs200868910	38	0.001	C55	0.00	0.763 benign	56
RFC3 NM_002915	c.179A>G p.Tyr60Cys	rs377157774	74	1.000	C55	0.00	1	192
RFC4 NM_002916.3	c.181G>A p.Val61Met	rs151335809	73	0.999	C0	0.20	1	0
RFC4 NM_002916.3	c.365A>G p.Lys122Arg	N/A	43	0.967	C25	0.00	1	26

[1]PolyPhen-2 variant scores ranges from 0 to 1, with increasing pathogenicity risk.
[2]Align-GVGD classes ranges from C0 to C65, with increasing pathogenicity risk of the variant.
[3]SIFT scores ranges from 0 to 1. Variants with scores of 0 to 0.05 are predicted to be pathogenic.
[4]MutationTaster scores represent the p-value of the prediction. All variants are predicted to be pathogenic, unless when stated otherwise.
[5]Grantham's distance ranges from 0 to 215. The residues have higher similarity with decreasing scores.

3.4.1 Effect on protein and patient phenotypes

One frameshift mutation and 19 missense mutations have been identified and subjected to closer interrogation. The possible deleterious effects of the variants have been assessed, and their suggested pathogenicity has been proposed. An overview of the variants found in the four recognised CRC predisposition genes, describing how they might affect protein function and elicit pathogenicity, is shown in Table 3.3. An overview of the variants found in possibly new CRC predisposition genes, describing how they might affect protein function and elicit pathogenicity, is shown in Table 3.4.

Table 3.3. Variants identified in the four CRC predisposition genes, *MLH1, MSH2, MSH6,* and *PMS2,* which have been predicted to be pathogenic, with respective effects on protein and possible pathogenic consequences.

Protein variant	Gene	Effect on protein	Possible consequence
p.Ser247Ala	*MLH1*	Amino acid change within conserved interaction domain	Impaired MMR capability and reduced efficiency
p.His46Gln	*MSH2*	Change of conserved amino acid in DNA mismatch-binding domain	Protein unable to correctly bind mismatches, impairing MMR capability
p.Ala272Val	*MSH2*	Amino acid change in conserved helix motif of connector domain	Reduced mismatch binding affinity and partial skipping of exon 5
p.His837Gln	*MSH6*	Amino acid change within a signal-transducing lever domain	Reduced MMR efficiency by disrupted signalling between protein domains
p.Ser46Asn	*PMS2*	Change of conserved amino acid within ATP-binding domain	Impairment of ATP hydrolysis and MMR activity
p.Asn335Ser	*PMS2*	Amino acid change in conserved position	Altered protein function

Table 3.4. Variants identified in possibly novel CRC predisposition genes, which have been predicted to be pathogenic, with respective effects on protein and possible pathogenic consequences.

Protein variant	Gene	Effect on protein	Possible consequence
p.Arg779His	*MSH3*	Amino acid change in conserved position	Altered protein function
p.His296Thrfs*12	*MLH3*	Premature stop codon	Degraded mRNA resulting in lost protein
p.Arg1074Gln	*POLD1*	Amino acid change in conserved motif of zinc finger domain	Obstruction of correct formation of the polymerase complex
p.Thr954Met	*POLD1*	Amino acid change in conserved position	Reduced function of the DNA polymerase complex
p.Leu108Met	*POLD2*	Amino acid change within conserved area	Impaired formation and/or function of the polymerase complex
p.Glu298Lys	*POLD2*	Amino acid change in partly conserved position	Reduced function of the DNA polymerase complex
p.Glu27Lys	*POLD4*	Amino acid change in conserved position	Reduced function of the DNA polymerase complex
p.Asp249Asn	*EXO1*	Amino acid change within conserved MSH3 interaction domain	Reduced MMR capability
p.Ser610Gly	*EXO1*	Amino acid change within conserved MSH2 interaction domain	Reduced MMR capability
p.Ala153Val	*EXO1*	Amino acid change within conserved pentapeptide involved in a XPG-domain and MSH3 interaction	Obstruction of α-helix and impairment of DNA mismatch excision, reduced MMR capability
p.Ala28Thr	*EXO1*	Amino acid change in conserved position within XPG domain	Affected mismatch excision and reduced MMR capability
p.Tyr60Cys	*RFC3*	Substitution of a conserved and phosphorylated residue	Obstruction of Tyrosine kinase signalling and impaired protein function
p.Val61Met	*RFC4*	Amino acid change in conserved position	Impaired DNA excision and repair
p.Lys122Arg	*RFC4*	Amino acid change in conserved position	Impaired DNA excision and repair

Additional mutations in other genes have also been detected in the patients in this study, which have been identified in the clinic or found in the same gene panel by co MSc student Ann-Therese Ali. All mutations found in each patient that are likely to be pathogenic are shown in table 3.5. The patient tumours have been tested for MSI and loss of IHC staining of the four MMR proteins, MLH1, MSH2, MSH6, and PMS2, in the clinic. The IHC and MSI status of each sample is shown in Table 3.5

Table 3.5. All possible pathogenic variants that have been identified in this study, in the same gene panel, and in the clinic is shown for each patient. The IHC and MSI tumour status of each patient is listed.

Sample	Variants	Gene	IHC staining	MSI status
2	p.Arg1074Gln p.Leu251Valfs*2	*POLD1* *FAM166A*	Normal	MSS
6	p.Asp249Asn	*EXO1*	Normal	MSI-L
8	p.Ser610Gly p.Pro431Leu	*EXO1* *TGFBR2*	MLH1/PMS2	MSI-H
9	p.His46Gln p.Pro751_Lys752insAsp	*MSH2* *BUB1B*	MSH2/MSH6	MSI-H
13	p.Ala272Val	*MSH2*	MSH2/MSH6	MSI-H
14	p.Ser247Ala p.His296Thrfs*12 p.Leu251Valfs*2	*MLH1* *MLH3* *FAM166A*	Normal	MSS
16	p.Thr954Met p.Gln502Hisfs*20	*POLD1* *MAML3*	MSH2/MSH6	MSS
20	p.Ser247Ala	*MLH1*	Normal	MSS
24	p.Asn335Ser p.Ile138Val	*PMS2* *OGG1*	Normal	MSS
25	p.Asp249Asn p.Ala257Ser c.-7C>T and c.-28C>T	*EXO1* *AXIN2* *MLH1*	N/A	N/A
28	p.Ala153Val p.Thr66Ilefs*35 p.Tyr458Phe p.Val427Phe	*EXO1* *GREM1* *POLE* *FANCD2*	Normal	MSS
30	p.Ala28Thr p.Val262Ala p.Arg69Pro	*EXO1* *BAMPR1A* *AKT1*	MLH1/PMS2	MSI-H
33	p.Leu108Met p.Ala126Val p.280_281del	*POLD2* *PTEN* *RAI1*	Normal	MSS
36	p.Arg779His p.T973M p.Cys36fs*	*MSH3* *LAMA5* *PIK3CA*	Normal	MSS
38	p.Glu27Lys p.Pro2178Ser	*POLD4* *NOTCH3*	Normal	MSS
43	p.Asp249Asn p.Lys122Arg p.Pro751_Lys752insAsp	*EXO1* *RFC4* *BUB1B*	Normal	MSS
57	p.Glu298Lys c.-93G>C	*POLD2* *PMS2*	Normal	MSS
73	p.Val61Met p.Cys503Ser	*RFC4* *FLCN*	N/A	N/A
74	p.Tyr60Cys	*RFC3*	MSH2/MSH6	MSI-H
76	p.Ser46Asn p.Gln502Hisfs*20	*PMS2* *MAML3*	MSH2/MSH6	MSI-H
80	p.His837Gln	*MSH6*	Normal + MLH1/PMS2	MSI-H

4. Discussion

4.1 Overview

In this project, possibly pathogenic DNA variants have been identified in 22 of the 95 patients participating in the study. Within 12 MMR associated genes, 20 unique variants have been detected. Some of the variants are localized within genes that are known to predispose to CRC, some are found within genes of which connection to CRC is widely disputed, while other variants are found in genes that are of unknown clinical significance to CRC. The genes in which mutations have been detected were researched to enlighten the potential consequences of each variant. The pathogenic probability of the variants has been contemplated based on the likely effects on protein function. In the following discussion, it will be argued that mutations predisposing to CRC have been identified in these genes. As of yet, none of the variants present evidence of being disease-causing agents in the CRC-patients, and remain variants of uncertain significance (VUS) until further elucidation can be done. These variants are predicted to cause disease by at least two *in silico* methods and pathogenic variants have not previously been detected in the patients. Consequently, their pathogenic status needs to be elaborated. The patients have varying phenotypes, and the majority of patients (18) are carrying more than one identified variant. This suggests that among the 20 variants there are mutations of various penetrance and risk, that could exhibit pathogenicity alone, or in combination with other variants.

4.2 Identified variants in established CRC predisposition genes

4.2.1 p.Ser247Ala, *MLH1*

The missense variant p.Ser247Ala was identified in *MLH1* of patients 14 and 20, who are siblings with CRC. The mutation changes a Serine to Alanine within a beta strand motif, in a domain involved in interaction with the MutS complexes. Serine is a small, polar residue, common in protein functional centres, which contains a reactive hydroxyl group, enabling it to form hydrogen bonds with polar substrates. Alanine is small, non-polar, and non-reactive (109). The Grantham's distance between the two residues is 49. The variant is novel, but a variant in the same position has been classified as class 5 with InSiGHT classification (102). This variant changes Serine to Proline, which like Alanine is small, non-polar, and non-reactive, rarely involved in functional sites (109). The change from Serine to Proline has been found to alter the quantity, subcellular localization, and MMR capability of MLH1, and reduce the repair efficiency to 17% (34). Other missense variants in close proximity have

been shown to alter protein folding and reduce MMR efficiency (33, 35). Two protein prediction tools have predicted this variant to be pathogenic. *MLH1* is a known high-penetrance CRC predisposition gene, which makes it likely that the mutation could be pathogenic. Especially considering the large physiochemical difference of the residues and the high conservation of the area. The pathogenicity of the similar p.Ser247Pro variant further supports the indication that this change should be pathogenic as well. However, the patient's tumours are MSS and show no loss of the four MMR proteins. This indicates that MMR is functioning normally, and that *MLH1* is expressed. In a study analysing the functional effects of mutations in *MLH1 in vitro,* MMR activities were impaired without reduced MLH1 levels in some missense mutations (35). This suggests that it is possible to positively detect MLH1 by IHC even if pathogenic variants are present. Nonetheless, if this were the case, the tumour phenotype would likely be MSI. The fact that two siblings affected with CRC both share the same variant indicates that it could be involved in disease development. However, the MSS tumours advocate against an MMR defect. Functional studies of the protein and segregation analysis of this family will be useful to determine how the protein is affected and whether the mutation is inherited in other affected family members. Until further studies can elucidate its role, the pathogenicity of the variant remains unknown.

4.2.2 p.His46Gln, *MSH2*

The missense variant p.His46Gln was identified in *MSH2* of patient 9. The variant changes a conserved Histidine to Glutamine in the end of a beta strand motif in the N-terminal DNA mismatch-binding domain of the protein. Histidine is considered a polar residue, but its unique chemical properties make any other residue a poor substitute. Histidine is ideal for protein active or binding sites because of the ease with which it transfers protons. Glutamine is polar and prefers to reside on the protein surface (109). Three studies report that the variant has been found in patients with CRC (110-112), and it is classified as class 3, based on InSiGHT classification. The amino acid position is conserved and part of a critical domain, indicating that it plays an important role in protein function. The variant could therefore impair MMR activity by e.g. alteration of MSH2 binding to DNA mismatches. Three prediction tools predict the change to be pathogenic. The patient's tumours are MSI-H and IHC staining shows loss of MSH2/MSH6. This suggests that either *MSH2* or *MSH6* has mutations that lead to loss of protein, which causes the other protein to disintegrate, when unable to form the MutSα heterodimer. The fact that mutations in *MSH2* are known to be high-penetrant indicates that the identified *MSH2* variant could be the explanation for disease

in this patient. A nonframeshift insertion in the BUB1B gene of this patient was also identified in the same gene panel. The *BUB1B* gene is responsible for mitotic checkpoints, and mutations in this gene have been detected in CRC cell lines (113). It is possible that this variant could be causing or contributing to disease. However, because of the gene's documented role in predisposition to CRC, the biggest risk of CRC is more likely inferred by the *MSH2* variant.

4.2.3 p.Ala272Val, *MSH2*

The missense variant p.Ala272Val was identified in *MSH2* of patient 13. The variant changes an Alanine to Valine in a 14-residue helix motif, located in the connector domain of the protein. The domain connects the DNA-binding N-terminal domain to the rest of the heterodimer, and is responsible for interactions and signalling between different protein elements (114). Both residues are small and non-reactive, with hydrophobic side chains, and are rarely involved in catalytic function. Valine is, however, branched at the Cβ carbon, giving it more bulkiness around the protein backbone. It therefore has a lot of difficulties adopting an α-helical conformation, which in this case could disrupt the helix (109). The Grantham's distance between the residues is 64. The mutation has been associated with CRC in several studies, and it is categorised as a class 3, based on InSiGHT classification. The studies report that the mutant protein shows slightly reduced mismatch binding affinity, and both functional and *in silico* analyses demonstrate that the mutation has an effect on splicing, resulting in partial skipping of exon five, with a 12% exon exclusion (37, 115, 116). This suggests that the mutation could be predisposing to CRC. Two prediction tools predict the change to be pathogenic. The patient has MSI-H tumours and IHC staining shows loss of MSH2/MSH6, indicating that the present mutation could be deleterious. As no other germline mutations have been found in this patient it is quite possible that p.Ala272Val is causal, and it should therefore be considered a high-risk candidate and subjected to further studies.

4.2.4 p.His837Gln, *MSH6*

The missense variant p.His837Gln was identified in *MSH6* of patient 80, which changes a Histidine to Glutamine in a single residue position connecting two helix motifs. Both residues are polar and frequently found in active or binding sites. Histidine has unique chemical properties that make it ideal for catalytic domains, while Glutamine prefers to reside on the protein surface (109). The position is part of a long α-helical lever domain, which is believed to acts as a signal transducer between the ATPase domain and DNA-binding domain (114). A

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 signalling between domains is disrupted (117). This suggests that the mutation could impair MMR activity. Four protein prediction tools predict the change to be pathogenic. The patient is a female with endometrial cancer, which is frequently observed in cases with *MSH6* mutations (40). The patient had two tumours; one was MSI-H with loss of MLH1 and PMS2, while the other was MSH-H with normal IHC staining. The tumour with loss of MLH1/PMS2 showed somatic hypermethylation of the *MLH1* promoter, which is the most common cause of sporadic loss of MLH1 (118). This phenotype suggests that the patient could have germline mutations increasing the risk of CRC, as well as somatic mutations in the tumour lacking MLH1/PMS2. The MSI-H phenotype, the established role of *MSH6* in predisposition to CRC, the presence of cancer in the endometrium, and the lack of other mutations makes this variant a CRC predisposition candidate, possibly conferring a high risk.

4.2.5 p.Ser46Asn, *PMS2s*

The missense variant p.Ser46Asn was identified in *PMS2* of patient 76. The variant changes a conserved Serine to Asparagine within a helix motif in the N-terminal ATP-binding domain of the protein. Serine is small and often resides within tight turns on the protein surface, where it forms hydrogen bonds with the protein backbone (109). Asparagine's larger side chain could in this case distort the helix motif. The mutation has not previously been reported, but a substitution in the same position, which changes Serine to Isoleucine, has been identified in a study of patients with CRC (119). This mutation is classified as class 4 with InSiGHT classification. As the residue is conserved and part of a specific motif, it is likely to be important for correct function of the ATPase domain. If so, the change could impair ATP hydrolysis. Four protein prediction tools predict the change to be pathogenic. The patient has MSI-H tumours and IHC shows loss of MSH2/MSH6, which indicates that disease is associated with mutations in one of these genes. However, no germline variants in *MSH2* or *MSH6* have been identified. This could indicate that the patient has somatic mutations in *MSH2*/*MSH6*, causing tumour development, or that the tumour is caused by heritable mutations and happen to have two somatic hits in *MSH2*/*MSH6*. The tumour should therefore be analysed for such mutations. Because of the familial history of cancer, the patient could also be a candidate of the recently reported LLS, found in patients with MSI-H phenotypes and familial cancer, but without identified mutations in the MMR genes. The presence of PMS2 in the tumours confirms that the gene is expressed; yet it does not prove that the

protein is functional. As the mutation is in a conserved position of a high-penetrant gene, it is likely that it could be involved in development of disease. A frameshift deletion of the *MAML3* gene was identified in the same patient, which functions as a transcriptional activator. It is possible that this mutation could be causing disease or contribute to the cancer risk. The p.Ser46Asn variant could also be causal of cancer, or increase the risk in combination with other variants.

4.2.6. p.Asn335Ser, *PMS2*

The missense variant p.Asn335Ser was identified in *PMS2* of patient 24. The variant changes an Asparagine to Serine in a conserved position located between the ATPase and nuclease domains of the protein. The function of this protein area is uncertain. Knowing that pathogenic mutations have been found distributed all across the protein, and that the residue is conserved, it is possible that the variant affects protein function, and consequently MMR function. Four prediction tools predict the change to be pathogenic. The tumours are MSS and IHC staining of MMR proteins is normal in the patient, indicating that MMR activity is normal. A missense mutation in the *OGG1* gene of the same patient was also identified in the gene panel. It is also possible that disease could be caused by this mutation, or that it contribute to a polygenic explanation for disease. As *PMS2* is a high-penetrant gene, well known to predispose to CRC, it is possible that p.Asn335Ser could cause a risk in itself, especially since all protein prediction tools predict it to be pathogenic. However, the MSS tumour and normal MMR protein staining contradicts this theory, making the clinical significance of the mutation uncertain.

4.3 Identified variants in potentially new predisposition genes

4.3.1 p.Arg779His, *MSH3*

The missense variant p.Arg779His was identified in *MSH3* of patient 36. The variant changes an Arginine to Histidine within a beta strand motif. Both residues are large, polar, often positively charged, and found in protein active or binding sites. However, Arginine will in most cases only substitute well with Lysine, and Histidine does not really substitute well with any other amino acid, due to its unique chemical properties (109). The position is highly conserved across species, and three prediction tools predict the change to be pathogenic. The tumours from this patient are MSS and IHC staining shows no loss of the four MMR proteins, indicating functional MMR. Two additional variants were also identified in this patient from the gene panel, a missense mutation in *LAMA5*, and a stop codon gain in the oncogene *PIK3CA*. These mutations alone could be causing the disease, or the p.Arg779His in *MSH3* could be the cause of disease. p.Arg779His could also be a modifying variant, segregating with mutations in other genes and contributing to the cancer risk.

4.3.2 p.His296Thrfs*12, *MLH3*

The frameshift deletion p.His296Thrfs*12 was identified in *MLH3* of patient 14. The deletion causes the formation of a premature stop codon twelve amino acids downstream, and the original 1453 residue polypeptide is reduced to 316 residues. With a stop codon inserted in the beginning of the gene, it is likely that the mRNA will be degraded, resulting in loss of MLH3. The same mutation was reported in a Swedish study from Karolinska Hospital in 2003, where the variant was found in a patient with colorectal cancer, a family member with colorectal cancer, a family member with endometrial cancer, and one of three unaffected relatives. The study did not find the mutation in any of the 96 controls or sporadic CRC patients. The authors proposed that the mutation could be associated with disease with reduced penetrance (120). A nonsynonymous substitution in *MLH1* and a frameshift insertion in FAM166A have also been identified in the same patient. There is a possibility that these mutations could contribute to disease together, especially as MLH1 and MLH3 interact with each other in formation of the MutLγ heterodimer. The patient has a sibling with CRC, patient 20, who only has the *MLH1* mutation. This could indicate that the mutation is a modifying mutation, contributing to the disease. However, IHC shows normal staining of MMR proteins, including MLH1. Both siblings have MSS tumours, suggesting that the cause of disease is not MMR related. Studies have shown that tumours from patients with *MLH3*

mutations could be both MSI and MSS, and that *MLH3*-deficient mice have shown to be MSS (56, 57). Consequently, it is possible for colorectal tumours associated with defects in MLH3 to be MSS and still cause disease, making this variant a candidate of CRC predisposition.

4.3.3 p.Thr954Met, *POLD1*

The missense variant p.Thr954Met was identified in *POLD1* of patient 16. The variant changes a conserved Threonine to Methionine. Threonine can be substituted with other polar amino acids, and resides both within the protein and on the protein surface. It forms hydrogen bonds with polar substrates and is quite common in functional centres. Methionine is non-polar, preferring to be buried in hydrophobic cores, and is rarely directly involved in protein function (109). As the position is conserved, the difference in structure and properties of the residues could distort protein folding or function. Three prediction tools predict the change to be pathogenic. Other studies have found *POLD1* mutations to be causal of CRC, but these mutations have been detected in the exonuclease domain of the protein. The present variant is not located within this catalytic domain, and the consequences are therefore uncertain. A frameshift deletion in the transcriptional activator *MAML3* was also identified in the patient, indicating that this variant also could be involved in cancer development. The tumours are MSS, which is consistent with the phenotype reported in other *POLD1* mutation patients. IHC staining show loss of MSH2/MSH6, but no germline *MSH2*/*MSH6* mutations have been identified. This could indicate somatic mutations leading to inactivation of these genes, although this should have inferred MSI. The tumour could therefore be caused by germline mutations and happen to have two somatic MSH2/MSH6 hits.

4.3.4 p.Arg1074Gln, *POLD1*

The missense variant p.Arg1074Gln was identified in *POLD1* of patient 2. The variant changes an Arginine to Glutamine in a highly conserved CysB protein motif, within the C-terminal zinc finger domain of POLD1 (121). Arginine is positively charged and prefers to reside on the protein surface. It is often involved in stabilizing salt-bridges and participates in active or binding sites, where it binds negatively charged substrates. Glutamine is smaller and uncharged, and therefore unable to mimic the role of Arginine (109). The CysB motif is a sequence of 19 residues that bind an iron-sulphur cluster, which is the cofactor required for formation of the polymerase complex. The variant is located between two metal-binding residues (122). This suggests that the mutation could obstruct correct formation of the polymerase complex. Three prediction tools predict the change to be pathogenic. The

patient's tumours are MSS, and IHC staining of MMR proteins is normal. This is consistent with the phenotype for disease-causing *POLD1* mutations, although these variants have all been detected in the exonuclease domain (59, 123). A frameshift insertion was identified in the *FAM166A* gene in the same patient, indicating that this mutation could be causative of disease as well. Seeing that the *POLD1* variant is conserved within a critical motif of a gene whose involvement in CRC has been documented, p.Arg1074Gln should be considered a candidate of disease-cause in this patient. However, little is known about mutations in the polymerase-part of the protein.

4.3.5 p.Leu108Met, *POLD2*

The missense variant p.Leu108Met was identified in *POLD2* of patient 33. The variant changes a Leucine to Methionine in a conserved protein position. Both residues are nonpolar, and prefer to be buried within hydrophobic cores. They are fairly non-reactive and seldom directly involved in protein function. Aliphatic residues tend to substitute each other well, however, unlike proper aliphatic residues, Methionine contains a sulphur atom in its side chain (109). The area is highly conserved, indicating a specific function, perhaps in scaffolding and interaction with other proteins, which is the subunit's primary function. It is therefore possible that the change could obstruct the correct formation and function of the polymerase complex. Three prediction tools predict the change to be pathogenic. The tumours of the patient are MSS and IHC staining is normal, suggesting that the cause of disease should be found outside of the four MMR genes. The patient was diagnosed with CRC at a young age, and no other mutations have been identified in this patient. These features points to p.Leu108Met as a possible candidate of CRC risk.

4.3.6 p.Glu298Lys, *POLD2*

The missense variant p.Glu298Lys was identified in *POLD2* of patient 57. The variant changes a Glutamate to Lysine within a short coil, between two beta strands. Glutamate is negatively charged and therefore prefers to reside on the protein surface. It is frequently involved in salt-bridges and protein active or binding sites. Lysine is positively charged and amphipathic, often involved in salt-bridges and active sites, performing the opposite role of Glutamate (109). Being only partly conserved, and not part of any known specific motif, it is possible that substitutions in this position are tolerated. However, the role of the protein area is not well understood, and the change in charge could affect the immediate protein surroundings. The protein is involved in both interaction with PCNA and the other POLD

subunits, as a scaffold. Two prediction tools predict the change to be pathogenic. The patient's tumours are MSS and IHC staining of the four MMR proteins is normal. A missense mutation in the untranslated N-terminal region of *PMS2* has been detected in the same patient by the clinic. This could have a deleterious effect on PMS2, however, IHC show no loss of PMS2 and the MSS phenotype suggests that MMR is functional. It is possible that the *POLD2* variant is contributing to the risk of cancer in the patient, through impairment of correct polymerase function. As of now, there is little evidence to claim that the variant is pathogenic; still it cannot be ruled out as a risk candidate.

4.3.7 p.Glu27Lys, *POLD4*

The missense variant p.Glu27Lys was identified in *POLD4* of patient 38. The variant changes a conserved Glutamate to Lysine in the N-terminal region of the protein. The two residues serve similar functions, but are of opposite charge. The short, conserved sequence indicates that the subunit is vulnerable to changes. Two protein prediction tools predict the variant to be pathogenic. The tumours of the patient are MSS and IHC staining of MMR proteins is normal. This is consistent with other pathogenic *POLD* mutations. The patient also carries a missense mutation in the *NOTCH3* gene from the gene panel. The NOTCH3 protein functions as a tumour suppressor and could be involved in disease development (124). This could also suggest a polygenetic cause of disease, where p.Glu27Lys is implicated. There is too little evidence to claim that the variant is a risk candidate of CRC. However, given the possible obstruction of correct formation of the polymerase complex, the variant may cause increased risk of CRC in combination with other mutations.

4.3.8 p.Asp249Asn, *EXO1*

The missense variant p.Asp249Asn was identified in *EXO1* of patient 6, 25 and 43. The variant changes an Aspartate to Asparagine within a protein domain responsible for interaction with MSH3. The residues are quite similar in structure, and substitute each other well. They prefer to reside on the protein surface and are frequently involved in protein active or binding sites (109). However, the negative charge of Aspartate could alter interaction with MSH3 e.g. through binding affinity. Two protein prediction tools predict the change to be pathogenic. Patient 6 has MSI-L tumours and normal IHC staining of MMR proteins, indicating a weak mutator MMR phenotype. No other mutation was found in this patient. MSI and IHC analyses were not performed on tumours from patient 25. This patient carried two additional class 3 missense mutations in the untranslated C-terminal region of *MLH1*,

detected by the clinic, and a missense mutation in *AXIN2*, identified in the same gene panel. Mutations in *AXIN2* have been reported in several CRC cases (125). Patient 43 has MSS tumours and normal IHC staining of MMR proteins. This patient has an additional missense mutation in *RFC4*, and a nonframeshift insertion in the *BUB1B* from the same gene panel. The *BUB1B* gene is responsible for mitotic checkpoints, and has been associated with sporadic CRC (113). The normal IHC staining of the four MMR proteins suggests that disease is caused by defects in other genes, but this only rules out mutations that prohibit protein expression of the four MMR genes. The presence of other mutations in addition to p.Asp249Asn in two of the patients could indicate a polygenetic susceptibility to CRC. A co-inheritance of several moderate- or low-risk genes could be a possibility for the explanation of disease in these cases. The involvement of *EXO1* in cancer is widely disputed, however, it could seem that mutations in *EXO1* are low-penetrant or modifying of CRC development.

4.3.9 p.Ser610Gly, *EXO1*

The missense variant p.Ser610Gly was identified in *EXO1* of patient 8. The variant changes a Serine to Glycine within a protein domain responsible for interaction with MSH2. Serine is small, polar, and quite common in protein functional centres. Glycine is nonpolar and miniscule; the side chain only consists of a hydrogen atom, providing a lot of conformational flexibility. The two amino acids generally substitute each other well (109). The variant is located within a conserved interaction domain, indicating that a change in this position could alter binding affinity. Two protein prediction tools predict the change to be pathogenic. The patient has MSI-H tumours and IHC staining shows loss of MLH1. This suggests that the cause of CRC should be found within the high-penetrant *MLH1* gene. However, the patient has no likely pathogenic germline mutations in *MLH1*, and no somatic methylation of the promoter. The patient has a missense mutation in *TGFBR2*, identified in the same gene panel, which could be contributing to disease. The pathogenic effect of the p.Ser610Gly variant is not obvious, but it is possible that it causes, or contribute to, an elevated cancer risk, possibly in addition to the *TGFBR2* variant and other unidentified variants.

4.3.10 p.Ala153Val, *EXO1*

The missense variant p.Ala153Val was identified in *EXO1* of patient 28. The variant changes an Alanine to Valine within a conserved helix motif in a domain responsible for interaction with MSH3. Both residues are small, non-polar, and fairly non-reactive, yet Alanine will often play a role in substrate recognition and specificity. Valine is branched at the Cβ carbon, giving it a lot more bulkiness around the protein backbone than Alanine. The most pronounced effect of this is that Valine has difficulties with adopting a α-helical conformation, and could therefore disrupt the helix motif (109). The protein position is highly conserved, indicating that the protein tolerates changes in this position poorly, and that it could be an active component of the interaction domain. The change is also located within a Xeroderma Pigmentosum complementation Group G (XPG) domain, which is characteristic of the RAD2 protein family, including various nucleases. The XPG domain has two highly conserved regions: the first is located near the N-terminal, and the other is an internal region, an I-domain, spanning about 140 residues. The I-domain contains a highly conserved core of 27 amino acids, including a conserved pentapeptide: E-A-[DE]-A-[QS]. The identified variant is the second Alanine in this structure, and therefore likely involved in the catalytic mechanism of DNA excision repair in XPG (126). This suggests that the change could be obstructing this function. Three prediction tools predict the change to be pathogenic. The patient has MSS tumours and IHC staining of the MMR proteins is normal. The patient was young at diagnosis, indicating high penetrance. Three additional variants were detected in the patient from the same gene panel, a frameshift insertion in *GREM1,* a missense mutation in *FANCD2,* and a missense mutation in *POLE*. Mutations in *GREM1* and *FANCD2* have been linked to cancer development in previous studies, and mutations in *POLE* are associated with disrupted proofreading activity of the DNA polymerase, resulting in CRC (123, 127, 128). It would seem that p.Ala153Val disrupts a critical function of the EXO1 protein, and the presence of several mutations in this patient suggests that disease could have a polygenic explanation. The *POLE* variant has been shown to be the primary disease-causing agent in this family, with several members affected by CRC (129). This patient has the most severe phenotype in the family, and is the only patient who carries the *EXO1* variant. This suggests that the *POLE* variant and the *EXO1* variant could have a combined effect on development of disease, with p.Ala153Val worsening the condition.

4.3.11 p.Ala28Thr, *EXO1*

The missense variant p.Ala28Thr was identified in *EXO1* of patient 30. The variant changes a conserved Alanine to Threonine within a beta strand of six residues. Alanine is small, non-polar, and non-reactive, while Threonine is polar, branched at the Cβ carbon, and quite reactive (109). The different properties of the two amino acids make them poor substitutes of each other. The amino acid motif is part of the highly conserved N-terminal XPG domain, which together with the XPG-I domain makes up the characteristic regions of RAD2 nucleases (126). As the domain is important for function, the variant could affect the excision of mismatches. Three protein prediction tools predict the change to be pathogenic. The patient has a MSI-H tumour and IHC staining shows loss of MLH1/PMS2, indicating that these genes should be involved in the pathogenicity. No germline *MLH1* or *PMS2* mutations have been identified, which means that somatic mutations could be causing the loss of protein expression. A missense mutation in the *BMPR1A* gene and a missense mutation in the *AKT1* gene from the same gene panel were also identified in this patient. Both genes have been associated with CRC (130, 131). These findings indicate that several variants could be candidates of causal mutations, or there could be a polygenetic explanation for disease. As the p.Ala28Thr variant is located within a domain important for protein function, it is likely that it could be a possible risk-variant.

4.3.12 p.Tyr60Cys, *RFC3*

The missense variant p.Tyr60Cys was identified in *RFC3* of patient 74. The variant changes a Tyrosine to Cysteine in a conserved position close to the N-terminus. Tyrosine is large, aromatic, and partially hydrophobic, preferring 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. According to PhosphoSitePlus® the Tyrosine residue in this position is phosphorylated. This is a common condition of Tyrosines within intracellular proteins, where Tyrosine kinases attach a phosphate to their side chains as part of a signal transduction process. The enzymes that catalyse this reaction are highly specific, and do not work on other phosphorylated residues (109). This means that the phosphorylated residue could be important for signalling, and that the change to Cytosine could disrupt this signalling, thereby altering protein function. Four prediction tools predict the change to be pathogenic, and the Grantham's distance between the residues is 192. This is a significantly high score, indicating a deleterious change. The tumours are MSI-H and IHC staining shows loss of MSH2 and MSH6. This could indicate that the cause of disease should be found within

these MMR genes, yet no such mutations were detected. It is possible that the loss of these proteins is caused by somatic mutations. These mutations could be the cause of tumour development, but the tumour could also be caused by germline mutations and happen to contain somatic mutations as well. There is also a possibility that the mutation in RFC could make other genes more prone to mutate. No other germline mutations than p.Tyr60Cys have been found in this patient, making it the most likely risk-candidate yet. However, the MSI-H phenotype indicates a MMR defect that has not been located.

4.3.13 p.Val61Met, *RFC4*

The missense variant p.Val61Met was identified in *RFC4* of patient 73. The variant changes a Valine to Methionine in a conserved protein position. Both residues are hydrophobic and non-reactive, but Valine is branched at the Cβ carbon, creating bulkiness around the protein backbone. Because the residues are known to substitute each other, the change could very well be tolerated. Still, the position is highly conserved across species, indicating that correct residue is important. Two prediction tools predict the change to be pathogenic. MSI and IHC have not been performed for the patient. A missense mutation in the *FLCN* gene from the gene panel was also detected in the patient. FLCN is a tumour suppressor that has been linked to familial CRC (132). There is little evidence to claim that the *RFC4* variant is a pathogenic mutation, but it could be a modifying or low-risk variant, contributing to disease in combination with other risk variants.

4.3.14 p.Lys122Arg, *RFC4*

The missense variant p.Lys122Arg was identified in *RFC4* of patient 43. The variant changes a highly conserved Lysine to Arginine. The residues are both large, positively charged, and substitute each other well. They play important roles in structure and active sites, and prefer to reside on the protein surface. However, the change between Lysine and Arginine is not always neutral, as Arginine is able to form multiple hydrogen bonds (109). Although the residue similarity indicates that the change is tolerated, the high conservation of the area suggests that the correct residue could be necessary. Four prediction tools predict the change to be pathogenic. The patient has MSS tumours, normal IHC staining of the fourMMR proteins, and was young at the age of diagnosis, indicating a high-penetrant cause. The patient also has the p.Asp249Asn mutation in *EXO1*, and a nonframeshift insertion in the *BUB1B* gene, from the same gene panel. This suggests that several moderate- or low-risk variants,

outside of the four high-penetrant MMR genes, i.e. p.Lys122Arg, could be the explanation for disease in this case.

4.4 Result review

The identified variants in *MLH1, MSH2*, *MSH6*, and *PMS2* have been found in genes of which predisposition to CRC is well established. Most of the variants identified in these genes could be potential high-risk variants for disease in the patients who carries them. The p.Asn335Ser variant in *PMS2* are more uncertain, seeing that the phenotype does not coincide with a causal MMR defect. The identified mutations in *MSH3*, *MLH3, POLD1*, *POLD2*, *POLD4, EXO1, RFC3* and *RFC4* have been found in genes whose role in CRC is to a large extent unknown or disputed. Some of these variants could be candidates of high-risk mutations, causing tumour development and CRC in the patients. However, the majority of these variants are more likely to be candidates of modifying mutations that contribute to the pathogenic effect of more penetrant mutations, or low-risk variants, possibly co-inherited with other variants, creating an additive, pathogenic effect. Some of these variants could also be harmless, not contributing to disease at all. A large proportion of the variants have been found in addition to mutations in other genes, some of which could be more penetrant. Especially variants in the *EXO1* gene seem to be present in combination with mutations in other genes linked to cancer. This suggests that defective EXO1 could have a modifying effect on pathogenicity when other proteins important for genomic stability are affected. To elucidate the consequences of mutations in *EXO1,* functional studies and further studies of the families will be very useful. Variants such as p.Val61Met in *RFC4* and p.Glu298Lys in *POLD2* are predicted to be pathogenic, yet show no obvious sign of pathogenicity, other than that no other disease-explanatory mutations have been identified. These variants need to be elucidated further before any reasonable speculation of their pathogenicity can be done. The propositions made here, of the variants' pathogenicity, are not definite conclusions, rather suggestions based on circumstantial evidence.

4.5 Further work

Even though the predictions are not proof of pathogenicity of the variants, they are candidates of disease-risk in the 22 patients they are detected in. Still, the majority of patients in this study do not yet have any genetic explanation for their disease. Many of the patients have MSI tumours, which suggests that there could exist differences in MMR capability between individuals in the population, due to common polymorphisms. Such variants could exhibit weak penetrance and still predispose to cancer. If this is the case, then elucidation of such variants will facilitate development of new preventive strategies (33). It is also possible that disease-causing variants in MMR associated genes have gone undetected in this study. For instance, all synonymous variants were filtrated out because they are likely to be functionally neutral. However, synonymous variants have been shown to affect splicing, thereby altering protein structure and function (133). These would be variants of low frequency in the population.

To determine if the variants detected in this study are causative of CRC, further investigations are needed. There is a possibility that some of the tumours could host somatic mutations, resulting from Knudsons two-hit hypothesis, in which case there is a chance that the tumour could be sporadic. However, the tumour could still be caused by germline mutations, and happen to contain two somatic mutations as well (134). It is also possible that some germline mutations could make other genes more prone to mutate. Somatic mutations in *MLH1* and *MSH2* have been shown to be a frequent cause of MMR deficiency in MSI-H tumours (135). Additionally, it is possible for such tumours to arise from LLS. In the cases where tumours show loss of MMR proteins and no germline MMR mutations can be identified, the tumour genes should be examined for two-hit somatic mutations and loss of heterozygosity. The patients in this study have a family history of cancer, so if somatic mutations were confirmed, it should be investigated if this really is the cause of cancer.

Segregation analysis of the families should be performed to see if any of the observed mutations segregate with disease, which will be a good indication that the mutation is pathogenic. Frequent surveillance of families with a high burden of CRCs has been proven to significantly reduce mortality. However, it can be challenging to correlate a variant with pathogenic effect, due to small family size, unavailable clinical samples, or ethical issues. Determination of the relation between sick and healthy individuals who share the variant in

question, will also aid in the elucidation of penetrance. Additionally, segregation analysis is important to determine co-inheritance of variants where this is suspected.

The pathogenic significance of novel variants, especially missense mutations, is not easily evaluated without functional studies. It is not always obvious from *in silico* analyses if a mutation will alter or impair protein activity. Even if it is clear that the variant will have a deleterious effect on protein, there is a possibility that other proteins with redundant functions can replace its cellular role. Examination of the operative effect of specific variants on protein and cellular function is needed for evidence. Knock-out studies with introduced mutagenesis in animal models are used as an efficient mode of determining functionality *in vivo*. Novel variants that are detected through gene sequencing should be interpreted with caution until more complete analyses, including segregation analysis and functional studies, are available. It is important that researchers who do conduct such studies on *de novo* genomic variations, publish the results, to facilitate international information sharing, enabling sooner breakthroughs in medicine.

4.6 Next-generation sequencing with gene panels

NGS technology is rapidly being introduced in hospital clinics, and gradually replacing traditional diagnostic technology for genetic disorders. The technology enables examination of a large amount of genes and patients in a single test, superseding time- and cost-consuming gene-by-gene approaches. The possible impact and various applications of NGS in clinical laboratories have received excellent reviews, and the technology is successfully used today in many areas of clinical genetic testing (136, 137). The technology is gaining increasing acceptance as a diagnostic tool, as it is capable of replacing most other molecular diagnostic technologies, and pioneer laboratories have already implemented NGS-based gene panels in a diagnostic setting (138, 139). The latest NGS technology shows great promise in enabling advantageous cost efficiency, and strategies that provide low-cost solutions for NGS-based testing in clinical laboratories have been developed (139, 140). Therefore, the biggest challenge for diagnostic implementation of NGS is not the technology in itself, rather the aspects that follow sequencing, in particular correct gene and variant annotation, bioinformatic and statistical analyses, and unambiguous interpretation of the data. Ethical issues such as informed consent, genetic counselling of patients, and processing of personal data need also be addressed. Reliable interpretations of the many novel variants that are

detected through NGS require experience and expertise, and it is crucial that this is established before the technology reaches the clinic on a large scale. *In silico* functional prediction tools are widely available, yet the validity of computational prediction algorithms must be improved to the point at which their implementation in the clinical setting can be executed with confidence (140, 141). This project has demonstrated a successful NGS experiment in which a gene panel consisting of 124 genes across 95 patients have been sequenced. Assessment and interpretation of variants from 22 MMR-associated genes in this panel have revealed 20 novel, possibly pathogenic variants, which could provide intelligence to the cause of disease in the patients. Implementation of this process in a clinical setting will provide the patients with better, more suitable diagnosis, possibly influencing the individual therapeutic strategy.

5. Conclusion

Next-generation sequencing of a gene panel, including 22 genes involved in the human mismatch repair (MMR) system, has revealed 20 novel variants that could be associated with risk of cancer in 12 patients suffering from familial colorectal cancer (CRC). Six of the identified variants have been found within the high-penetrant predisposition genes: *MLH1*, *MSH2*, *MSH6*, and *PMS2*, well known to cause CRC. An additional 14 variants have been identified within genes whose involvement in CRC is disputed or unknown: *MSH3*, *MLH3*, *POLD1*, *POLD2*, *POLD4*, *EXO1*, *RFC3*, and *RFC4*. These variants have been suggested to confer a predisposition to CRC, possibly in combination with other risk-variants. Several methods of pathogenic prediction indicate that many of these variants could be involved in disease development, and they should therefore be examined further to elucidate the involvement in CRC. The patients in this study have previously been tested for recognized hereditary cancer syndromes, without positive results. To determine the true pathogenic role of the detected variants, both segregation analysis and functional studies are needed.

The project has demonstrated the successful sequencing of a gene panel consisting of 124 genes across 95 samples with NGS. The technology shows great potential for application of gene panel NGS in diagnostic use. Correct variant annotation, unambiguous data interpretation, and safe ethical standards are needed for this to be routine practice in hospitals.

6. Literature

1. GLOBOCAN. GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012. World Health Organization: IARC; 2012.
2. Binefa G, Rodriguez-Moranta F, Teule A, Medina-Hayas M. Colorectal cancer: from prevention to personalized medicine. World J Gastroenterol. 2014;20(22):6786-808.
3. Bowel Cancer Screening in Norway – a pilot study [Internet]. Giske Ursin. 2012 [cited 10/7/2014]. Available from: http://www.kreftregisteret.no/en/Cancer-prevention/Screening-for-colorectal-cancer/.
4. DeRycke MS, Gunawardena SR, Middha S, Asmann YW, Schaid DJ, McDonnell SK, et al. Identification of novel variants in colorectal cancer families by high-throughput exome sequencing. Cancer Epidemiol Biomarkers Prev. 2013;22(7):1239-51.
5. Patel SG, Ahnen DJ. Familial colon cancer syndromes: an update of a rapidly evolving field. Current gastroenterology reports. 2012;14(5):428-38.
6. Valle L. Genetic predisposition to colorectal cancer: Where we stand and future perspectives. World J Gastroenterol. 2014;20(29):9828-49.
7. Blanes A, Diaz-Cano SJ. Complementary analysis of microsatellite tumor profile and mismatch repair defects in colorectal carcinomas. World J Gastroenterol. 2006;12(37):5932-40.
8. Half E, Bercovich D, Rozen P. Familial adenomatous polyposis. Orphanet J Rare Dis. 2009;4:22.
9. Ku CS, Cooper DN, Wu M, Roukos DH, Pawitan Y, Soong R, et al. Gene discovery in familial cancer syndromes by exome sequencing: prospects for the elucidation of familial colorectal cancer type X. Mod Pathol. 2012;25(8):1055-68.
10. van Wezel T, Middeldorp A, Wijnen JT, Morreau H. A review of the genetic background and tumour profiling in familial colorectal cancer. Mutagenesis. 2012;27(2):239-45.
11. Lynch HT, Shaw TG. Practical genetics of colorectal cancer. Chinese clinical oncology. 2013;2(2):12.
12. Lynch HT, Lynch PM, Lanspa SJ, Snyder CL, Lynch JF, Boland CR. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. Clin Genet. 2009;76(1):1-18.
13. Vasen HF, Moslein G, Alonso A, Bernstein I, Bertario L, Blanco I, et al. Guidelines for the clinical management of Lynch syndrome (hereditary non-polyposis cancer). J Med Genet. 2007;44(6):353-62.
14. Talseth-Palmer BA, Brenne IS, Ashton KA, Evans TJ, McPhillips M, Groombridge C, et al. Colorectal cancer susceptibility loci on chromosome 8q23.3 and 11q23.1 as modifiers for disease expression in Lynch syndrome. J Med Genet. 2011;48(4):279-84.
15. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, et al. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res. 1998;58(22):5248-57.
16. Yamamoto H, Imai K. Microsatellite instability: an update. Arch Toxicol. 2015.
17. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Ruschoff 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.
18. Kang SY, Park CK, Chang DK, Kim JW, Son HJ, Cho YB, et al. Lynch-like syndrome: characterization and comparison with EPCAM deletion carriers. Int J Cancer. 2015;136(7):1568-78.

19. Umar A, Risinger JI, Hawk ET, Barrett JC. Testing guidelines for hereditary non-polyposis colorectal cancer. Nat Rev Cancer. 2004;4(2):153-8.
20. Vasen HF, Mecklin JP, Khan PM, Lynch HT. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Dis Colon Rectum. 1991;34(5):424-5.
21. Lindor NM, Rabe K, Petersen GM, Haile R, Casey G, Baron J, et al. Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X. JAMA. 2005;293(16):1979-85.
22. Silva FC, Valentin MD, Ferreira Fde O, Carraro DM, Rossi BM. Mismatch repair genes in Lynch syndrome: a review. Sao Paulo Med J. 2009;127(1):46-51.
23. Li GM. Mechanisms and functions of DNA mismatch repair. Cell Res. 2008;18(1):85-98.
24. Modrich P. Mechanisms in eukaryotic mismatch repair. J Biol Chem. 2006;281(41):30305-9.
25. Prindle MJ, Loeb LA. DNA polymerase delta in DNA replication and genome maintenance. Environ Mol Mutagen. 2012;53(9):666-82.
26. Iyer RR, Pluciennik A, Burdett V, Modrich PL. DNA mismatch repair: functions and mechanisms. Chem Rev. 2006;106(2):302-23.
27. Yao NY, O'Donnell M. The RFC clamp loader: structure and function. Subcell Biochem. 2012;62:259-79.
28. Kolodner RD, Marsischky GT. Eukaryotic DNA mismatch repair. Curr Opin Genet Dev. 1999;9(1):89-96.
29. Lee Bi BI, Nguyen LH, Barsky D, Fernandes M, Wilson DM, 3rd. Molecular interactions of human Exo1 with DNA. Nucleic Acids Res. 2002;30(4):942-9.
30. Chen R, Wold MS. Replication protein A: Single-stranded DNA's first responder: Dynamic DNA-interactions allow replication protein A to direct single-strand DNA intermediates into different pathways for synthesis or repair. Bioessays. 2014.
31. Tomkinson AE, Mackey ZB. Structure and function of mammalian DNA ligases. Mutat Res. 1998;407(1):1-9.
32. Bak ST, Sakellariou D, Pena-Diaz J. The dual nature of mismatch repair as antimutator and mutator: for better or for worse. Frontiers in genetics. 2014;5:287.
33. Ellison AR, Lofing J, Bitter GA. Human MutL homolog (MLH1) function in DNA mismatch repair: a prospective screen for missense mutations in the ATPase domain. Nucleic Acids Res. 2004;32(18):5321-38.
34. Raevaara TE, Korhonen MK, Lohi H, Hampel H, Lynch E, Lonnqvist KE, et al. Functional significance and clinical phenotype of nontruncating mismatch repair variants of MLH1. Gastroenterology. 2005;129(2):537-49.
35. Takahashi M, Shimodaira H, Andreutti-Zaugg C, Iggo R, Kolodner RD, Ishioka C. Functional analysis of human MLH1 variants using yeast and in vitro mismatch repair assays. Cancer Res. 2007;67(10):4595-604.
36. 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.
37. Ollila S, Dermadi Bebek D, Jiricny J, Nystrom M. Mechanisms of pathogenicity in human MSH2 missense mutants. Hum Mutat. 2008;29(11):1355-63.
38. Akiyama Y, Sato H, Yamada T, Nagasaki H, Tsuchiya A, Abe R, et al. Germ-line mutation of the hMSH6/GTBP gene in an atypical hereditary nonpolyposis colorectal cancer kindred. Cancer Res. 1997;57(18):3920-3.

39. Wu Y, Berends MJ, Mensink RG, Kempinga C, Sijmons RH, van Der Zee AG, et al. Association of hereditary nonpolyposis colorectal cancer-related tumors displaying low microsatellite instability with MSH6 germline mutations. Am J Hum Genet. 1999;65(5):1291-8.
40. Wagner A, Hendriks Y, Meijers-Heijboer EJ, de Leeuw WJ, Morreau H, Hofstra R, et al. Atypical HNPCC owing to MSH6 germline mutations: analysis of a large Dutch pedigree. J Med Genet. 2001;38(5):318-22.
41. Berends MJ, Wu Y, Sijmons RH, Mensink RG, van der Sluis T, Hordijk-Hos JM, et al. Molecular and clinical characteristics of MSH6 variants: an analysis of 25 index carriers of a germline variant. Am J Hum Genet. 2002;70(1):26-37.
42. Sjursen W, Haukanes BI, Grindedal EM, Aarset H, Stormorken A, Engebretsen LF, et al. Current clinical criteria for Lynch syndrome are not sensitive enough to identify MSH6 mutation carriers. J Med Genet. 2010;47(9):579-85.
43. Ou J, Niessen RC, Vonk J, Westers H, Hofstra RM, Sijmons RH. A database to support the interpretation of human mismatch repair gene variants. Hum Mutat. 2008;29(11):1337-41.
44. Nicolaides NC, Papadopoulos N, Liu B, Wei YF, Carter KC, Ruben SM, et al. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. Nature. 1994;371(6492):75-80.
45. Thompson E, Meldrum CJ, Crooks R, McPhillips M, Thomas L, Spigelman AD, et al. Hereditary non-polyposis colorectal cancer and the role of hPMS2 and hEXO1 mutations. Clin Genet. 2004;65(3):215-25.
46. Worthley DL, Walsh MD, Barker M, Ruszkiewicz A, Bennett G, Phillips K, et al. Familial mutations in PMS2 can cause autosomal dominant hereditary nonpolyposis colorectal cancer. Gastroenterology. 2005;128(5):1431-6.
47. Hendriks YM, Jagmohan-Changur S, van der Klift HM, Morreau H, van Puijenbroek M, Tops C, et al. Heterozygous mutations in PMS2 cause hereditary nonpolyposis colorectal carcinoma (Lynch syndrome). Gastroenterology. 2006;130(2):312-22.
48. ten Broeke SW, Brohet RM, Tops CM, van der Klift HM, Velthuizen ME, Bernstein I, et al. Lynch syndrome caused by germline PMS2 mutations: delineating the cancer risk. J Clin Oncol. 2015;33(4):319-25.
49. Peltomaki P. Deficient DNA mismatch repair: a common etiologic factor for colon cancer. Hum Mol Genet. 2001;10(7):735-40.
50. Risinger JI, Umar A, Boyd J, Berchuck A, Kunkel TA, Barrett JC. Mutation of MSH3 in endometrial cancer and evidence for its functional role in heteroduplex repair. Nat Genet. 1996;14(1):102-5.
51. Yin J, Kong D, Wang S, Zou TT, Souza RF, Smolinski KN, et al. Mutation of hMSH3 and hMSH6 mismatch repair genes in genetically unstable human colorectal and gastric carcinomas. Hum Mutat. 1997;10(6):474-8.
52. Akiyama Y, Tsubouchi N, Yuasa Y. Frequent somatic mutations of hMSH3 with reference to microsatellite instability in hereditary nonpolyposis colorectal cancers. Biochem Biophys Res Commun. 1997;236(2):248-52.
53. de Wind N, Dekker M, Claij N, Jansen L, van Klink Y, Radman M, et al. HNPCC-like cancer predisposition in mice through simultaneous loss of Msh3 and Msh6 mismatch-repair protein functions. Nat Genet. 1999;23(3):359-62.
54. Edelmann W, Umar A, Yang K, Heyer J, Kucherlapati M, Lia M, et al. The DNA mismatch repair genes Msh3 and Msh6 cooperate in intestinal tumor suppression. Cancer Res. 2000;60(4):803-7.

55. Lipkin SM, Wang V, Stoler DL, Anderson GR, Kirsch I, Hadley D, et al. Germline and somatic mutation analyses in the DNA mismatch repair gene MLH3: Evidence for somatic mutation in colorectal cancers. Hum Mutat. 2001;17(5):389-96.
56. Wu Y, Berends MJ, Sijmons RH, Mensink RG, Verlind E, Kooi KA, et al. A role for MLH3 in hereditary nonpolyposis colorectal cancer. Nat Genet. 2001;29(2):137-8.
57. Lipkin SM, Moens PB, Wang V, Lenzi M, Shanmugarajah D, Gilgeous A, et al. Meiotic arrest and aneuploidy in MLH3-deficient mice. Nat Genet. 2002;31(4):385-90.
58. Smith CG, Naven M, Harris R, Colley J, West H, Li N, et al. Exome resequencing identifies potential tumor-suppressor genes that predispose to colorectal cancer. Hum Mutat. 2013;34(7):1026-34.
59. Heitzer E, Tomlinson I. Replicative DNA polymerase mutations in cancer. Curr Opin Genet Dev. 2014;24:107-13.
60. Valle L, Hernandez-Illan E, Bellido F, Aiza G, Castillejo A, Castillejo MI, et al. New insights into POLE and POLD1 germline mutations in familial colorectal cancer and polyposis. Hum Mol Genet. 2014;23(13):3506-12.
61. Church JM. Polymerase proofreading-associated polyposis: a new, dominantly inherited syndrome of hereditary colorectal cancer predisposition. Dis Colon Rectum. 2014;57(3):396-7.
62. Zhang J, Tan CK, McMullen B, Downey KM, So AG. Cloning of the cDNAs for the small subunits of bovine and human DNA polymerase delta and chromosomal location of the human gene (POLD2). Genomics. 1995;29(1):179-86.
63. Elgaaen BV, Haug KB, Wang J, Olstad OK, Fortunati D, Onsrud M, et al. POLD2 and KSP37 (FGFBP2) correlate strongly with histology, stage and outcome in ovarian carcinomas. PLoS One. 2010;5(11):e13837.
64. Liu L, Mo J, Rodriguez-Belmonte EM, Lee MY. Identification of a fourth subunit of mammalian DNA polymerase delta. J Biol Chem. 2000;275(25):18739-44.
65. Huang QM, Akashi T, Masuda Y, Kamiya K, Takahashi T, Suzuki M. Roles of POLD4, smallest subunit of DNA polymerase delta, in nuclear structures and genomic stability of human cells. Biochem Biophys Res Commun. 2010;391(1):542-6.
66. Bowman GD, O'Donnell M, Kuriyan J. Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. Nature. 2004;429(6993):724-30.
67. Shen H, Cai M, Zhao S, Wang H, Li M, Yao S, et al. Overexpression of RFC3 is correlated with ovarian tumor development and poor prognosis. Tumour Biol. 2014;35(10):10259-66.
68. Xiang J, Fang L, Luo Y, Yang Z, Liao Y, Cui J, et al. Levels of human replication factor C4, a clamp loader, correlate with tumor progression and predict the prognosis for colorectal cancer. J Transl Med. 2014;12(1):320.
69. Arai M, Kondoh N, Imazeki N, Hada A, Hatsuse K, Matsubara O, et al. The knockdown of endogenous replication factor C4 decreases the growth and enhances the chemosensitivity of hepatocellular carcinoma cells. Liver international : official journal of the International Association for the Study of the Liver. 2009;29(1):55-62.
70. Schmutte C, Marinescu RC, Sadoff MM, Guerrette S, Overhauser J, Fishel R. Human exonuclease I interacts with the mismatch repair protein hMSH2. Cancer Res. 1998;58(20):4537-42.
71. Wu Y, Berends MJ, Post JG, Mensink RG, Verlind E, Van Der Sluis T, et al. Germline mutations of EXO1 gene in patients with hereditary nonpolyposis colorectal cancer (HNPCC) and atypical HNPCC forms. Gastroenterology. 2001;120(7):1580-7.

72. Jagmohan-Changur S, Poikonen T, Vilkki S, Launonen V, Wikman F, Orntoft TF, et al. EXO1 variants occur commonly in normal population: evidence against a role in hereditary nonpolyposis colorectal cancer. Cancer Res. 2003;63(1):154-8.
73. Wei K, Clark AB, Wong E, Kane MF, Mazur DJ, Parris T, et al. Inactivation of Exonuclease 1 in mice results in DNA mismatch repair defects, increased cancer susceptibility, and male and female sterility. Genes Dev. 2003;17(5):603-14.
74. Schaetzlein S, Chahwan R, Avdievich E, Roa S, Wei K, Eoff RL, et al. Mammalian Exo1 encodes both structural and catalytic functions that play distinct roles in essential biological processes. Proc Natl Acad Sci U S A. 2013;110(27):E2470-9.
75. Liberti SE, Larrea AA, Kunkel TA. Exonuclease 1 preferentially repairs mismatches generated by DNA polymerase alpha. DNA repair. 2013;12(2):92-6.
76. Kabzinski J, Przybylowska K, Mik M, Sygut A, Dziki L, Dziki A, et al. Association of polymorphism of Lys589Glu Exo1 gene with the risk of colorectal cancer in the Polish population. Pol Przegl Chir. 2014;86(8):370-3.
77. Akbari Z, Taleghani MY, Aghdaei HA, Mohebbi SR, Haghighi MM, Vahedi M, et al. Single Nucleotide Polymorphism (K589E) of the EXO1 Gene: Association with Colorectal Cancer Susceptibility and Clinicopathological Features. Gastroenterology & Hepatology: Open Access. 2014;1(3).
78. Rizzo JM, Buck MJ. Key principles and clinical applications of "next-generation" DNA sequencing. Cancer Prev Res (Phila). 2012;5(7):887-900.
79. Meldrum C, Doyle MA, Tothill RW. Next-generation sequencing for cancer diagnostics: a practical perspective. Clin Biochem Rev. 2011;32(4):177-95.
80. Schuster SC. Next-generation sequencing transforms today's biology. Nat Meth. 2008;5(1):16-8.
81. Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, et al. ACMG clinical laboratory standards for next-generation sequencing. Genet Med. 2013;15(9):733-47.
82. Illumina. An Introduction to Next-Generation Sequencing Technology http://www.illumina.com: illumina inc; 2013 [cited 2015 4/18/15]. Available from: http://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf.
83. Hagemann IS, Cottrell CE, Lockwood CM. Design of targeted, capture-based, next generation sequencing tests for precision cancer therapy. Cancer Genet. 2013;206(12):420-31.
84. Picelli S, Zajac P, Zhou XL, Edler D, Lenander C, Dalen J, et al. Common variants in human CRC genes as low-risk alleles. Eur J Cancer. 2010;46(6):1041-8.
85. LifeTechnologies. iPrepTM PureLink® gDNA Blood Kit
For purification of gDNA from human blood using the iPrepTM Purification Instrument. Life Technologies; 2012.
86. ThermoScientific. NanoDrop 1000 Spectrophotometer V3.7 User's Manual. 2008.
87. Invitrogen™. Qubit® 2.0 Fluorometer. Life Technologies™ 2010. p. 13-9.
88. Agilent. HaloPlex Target Enrichment System For Illumina Sequencing Protocol. Agilent Technologies; 2013.
89. Berglund EC, Lindqvist CM, Hayat S, Overnas E, Henriksson N, Nordlund J, et al. Accurate detection of subclonal single nucleotide variants in whole genome amplified and pooled cancer samples using HaloPlex target enrichment. BMC Genomics. 2013;14:856.
90. Illumina. HiSeq® 2500 System
User Guide. D ed: Illumina Inc; 2014.

91. Illumina. Illumina Sequencing Technology. In: Inc I, editor. http://www.Illumina.com: Illumina Inc; 2010.
92. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60.
93. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491-8.
94. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
95. Vigeland MD. Filtus 2015. 0.99-91:[Available from: http://folk.uio.no/magnusv/filtus.html.
96. Tavtigian SV, Greenblatt MS, Lesueur F, Byrnes GB. In silico analysis of missense substitutions using sequence-alignment based methods. Hum Mutat. 2008;29(11):1327-36.
97. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073-81.
98. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nature methods. 2010;7(8):575-6.
99. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nature methods. 2010;7(4):248-9.
100. Grantham R. Amino acid difference formula to help explain protein evolution. Science. 1974;185(4154):862-4.
101. LOVD [Internet]. Leiden Open Variant Database. 2015 [cited 11. March 2015]. Available from: http://www.lovd.nl/3.0/home.
102. Thompson BA, Spurdle AB, Plazzer JP, Greenblatt MS, Akagi K, Al-Mulla F, et al. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nat Genet. 2014;46(2):107-15.
103. UniProt [Internet]. 2015 [cited 11. March 2015]. Available from: http://www.uniprot.org/.
104. Prosite [Internet]. SIB Swiss Institute of Bioinformatics. Available from: http://prosite.expasy.org/.
105. Conserved Domains [Internet]. National Center for Biotechnology Information. Available from: http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml.
106. dbSNP Short Genetic Variations [Internet]. [cited 3/20/2015]. Available from: http://www.ncbi.nlm.nih.gov/SNP/.
107. LifeTechnologies. Primer Designer Tool Lifetechnologies.com: Thermo Fisher Scientific Inc.; 2015 [cited 2015]. Available from: https://www.lifetechnologies.com/no/en/home/life-science/sequencing/sanger-sequencing/pre-designed-primers-pcr-sanger-sequencing.html?icid=fr-primerdes-main.
108. NCBI. Primer-BLAST NCBI [cited 2015 15.03.2015]. Available from: http://www.ncbi.nlm.nih.gov/tools/primer-blast/.
109. Barnes MR, Gray IC, Southan C, Semple CA, Blake JA, Eppig J, et al. Bioinformatics for Geneticists: John Wiley & Sons, Ltd; 2003. 408 p.
110. Bubb VJ, Curtis LJ, Cunningham C, Dunlop MG, Carothers AD, Morris RG, et al. Microsatellite instability and the role of hMSH2 in sporadic colorectalcancer. Oncogene. 1996;12(12):2641-9.

111. Barnetson RA, Cartwright N, van Vliet A, Haq N, Drew K, Farrington S, et al. Classification of ambiguous mutations in DNA mismatch repair genes identified in a population-based study of colorectal cancer. Hum Mutat. 2008;29(3):367-74.
112. Fearnhead NS, Wilding JL, Winney B, Tonks S, Bartlett S, Bicknell DC, et al. Multiple rare variants in different genes account for multifactorial inherited susceptibility to colorectal adenomas. Proc Natl Acad Sci U S A. 2004;101(45):15992-7.
113. Cahill DP, Lengauer C, Yu J, Riggins GJ, Willson JK, Markowitz SD, et al. Mutations of mitotic checkpoint genes in human cancers. Nature. 1998;392(6673):300-3.
114. Warren JJ, Pohlhaus TJ, Changela A, Iyer RR, Modrich PL, Beese LS. Structure of the human MutSalpha DNA lesion recognition complex. Mol Cell. 2007;26(4):579-92.
115. 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.
116. Lastella P, Surdo NC, Resta N, Guanti G, Stella A. In silico and in vivo splicing analysis of MLH1 and MSH2 missense mutations shows exon- and tissue-specific effects. BMC Genomics. 2006;7:243.
117. 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.
118. Simpkins SB, Bocker T, Swisher EM, Mutch DG, Gersell DJ, Kovatich AJ, et al. MLH1 promoter methylation and gene silencing is the primary cause of microsatellite instability in sporadic endometrial cancers. Hum Mol Genet. 1999;8(4):661-6.
119. Nakagawa H, Lockman JC, Frankel WL, Hampel H, Steenblock K, Burgart LJ, et al. Mismatch repair gene PMS2: disease-causing germline mutations are frequent in patients whose tumors stain negative for PMS2 protein, but paralogous genes obscure mutation detection and interpretation. Cancer Res. 2004;64(14):4721-7.
120. Liu HX, Zhou XL, Liu T, Werelius B, Lindmark G, Dahl N, et al. The role of hMLH3 in familial colorectal cancer. Cancer Res. 2003;63(8):1894-9.
121. Yoshida R, Miyashita K, Inoue M, Shimamoto A, Yan Z, Egashira A, et al. Concurrent genetic alterations in DNA polymerase proofreading and mismatch repair in human colorectal cancer. Eur J Hum Genet. 2011;19(3):320-5.
122. Chung DW, Zhang JA, Tan CK, Davie EW, So AG, Downey KM. Primary structure of the catalytic subunit of human DNA polymerase delta and chromosomal location of the gene. Proc Natl Acad Sci U S A. 1991;88(24):11197-201.
123. Palles C, Cazier JB, Howarth KM, Domingo E, Jones AM, Broderick P, et al. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. Nat Genet. 2013;45(2):136-44.
124. Cui H, Kong Y, Xu M, Zhang H. Notch3 functions as a tumor suppressor by controlling cellular senescence. Cancer Res. 2013;73(11):3451-9.
125. Liu W, Dong X, Mai M, Seelan RS, Taniguchi K, Krishnadath KK, et al. Mutations in AXIN2 cause colorectal cancer with defective mismatch repair by activating beta-catenin/TCF signalling. Nat Genet. 2000;26(2):146-7.
126. XPG protein signatures [Internet]. ExPASy. 2015 [cited 11. March 2015]. Available from: http://prosite.expasy.org/cgi-bin/prosite/prosite-search-ac?PDOC00658 - ref2.
127. Jaeger E, Leedham S, Lewis A, Segditsas S, Becker M, Cuadrado PR, et al. Hereditary mixed polyposis syndrome is caused by a 40-kb upstream duplication that leads to increased and ectopic expression of the BMP antagonist GREM1. Nat Genet. 2012;44(6):699-703.

128. Langevin F, Crossan GP, Rosado IV, Arends MJ, Patel KJ. Fancd2 counteracts the toxic effects of naturally produced aldehydes in mice. Nature. 2011;475(7354):53-8.
129. Hansen MF, Johansen J, Bjornevoll I, Sylvander AE, Steinsbekk KS, Saetrom P, et al. A novel POLE mutation associated with cancers of colon, pancreas, ovaries and small intestine. Fam Cancer. 2015.
130. Carpten JD, Faber AL, Horn C, Donoho GP, Briggs SL, Robbins CM, et al. A transforming mutation in the pleckstrin homology domain of AKT1 in cancer. Nature. 2007;448(7152):439-44.
131. Delnatte C, Sanlaville D, Mougenot JF, Vermeesch JR, Houdayer C, Blois MC, et al. Contiguous gene deletion within chromosome arm 10q is associated with juvenile polyposis of infancy, reflecting cooperation between the BMPR1A and PTEN tumor-suppressor genes. Am J Hum Genet. 2006;78(6):1066-74.
132. Nahorski MS, Lim DH, Martin L, Gille JJ, McKay K, Rehal PK, et al. Investigation of the Birt-Hogg-Dube tumour suppressor gene (FLCN) in familial and sporadic colorectal cancer. J Med Genet. 2010;47(6):385-90.
133. Pagani F, Raponi M, Baralle FE. Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. Proc Natl Acad Sci U S A. 2005;102(18):6368-72.
134. Berger AH, Knudson AG, Pandolfi PP. A continuum model for tumour suppression. Nature. 2011;476(7359):163-9.
135. Mensenkamp AR, Vogelaar IP, van Zelst-Stams WA, Goossens M, Ouchene H, Hendriks-Cornelissen SJ, et al. Somatic mutations in MLH1 and MSH2 are a frequent cause of mismatch-repair deficiency in Lynch syndrome-like tumors. Gastroenterology. 2014;146(3):643-6.e8.
136. Mardis ER. The impact of next-generation sequencing technology on genetics. Trends Genet. 2008;24(3):133-41.
137. Bras J, Guerreiro R, Hardy J. Use of next-generation sequencing and other whole-genome strategies to dissect neurological disease. Nat Rev Neurosci. 2012;13(7):453-64.
138. Worthey EA, Mayer AN, Syverson GD, Helbling D, Bonacci BB, Decker B, et al. Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. Genet Med. 2011;13(3):255-62.
139. Yohe S, Hauge A, Bunjer K, Kemmer T, Bower M, Schomaker M, et al. Clinical validation of targeted next-generation sequencing for inherited disorders. Arch Pathol Lab Med. 2015;139(2):204-10.
140. Vrijenhoek T, Kraaijeveld K, Elferink M, de Ligt J, Kranendonk E, Santen G, et al. Next-generation sequencing-based genome diagnostics across clinical genetics centers: implementation choices and their effects. Eur J Hum Genet. 2015.
141. Katsanis SH, Katsanis N. Molecular genetic testing and the future of clinical genomics. Nature reviews Genetics. 2013;14(6):415-26.

Appendices

Appendix 1: Primers used for variant validation

The designed primers used for verification of each variant with Sanger sequencing are shown.

Variant	Forward Primer with universal tail	Reverse Primer with universal tail
EXO1 c.745G>A	CACGACGTTGTAAAACGAC-CTAGGAATGTGCAGACAGCTTGG	CAGGAAACAGCTATGACC-TGCCTCAGTCATTTGCTCCTT
EXO1 c.1828A>G	CACGACGTTGTAAAACGAC-ATTTTGAATCTTGACACCCCTTGAGAA	CAGGAAACAGCTATGACC-TGCTGCAATGCTGTGCTTG
EXO1 c.82G>A	CACGACGTTGTAAAACGAC-ACTCCAAGCTTTCTCTTCATTTGTC	CAGGAAACAGCTATGACC-AATAGTCTTTGTCCATCAGTGTCTTG
POLD1 c.2861C>T	CACGACGTTGTAAAACGAC-GACCTGAGAGCCCTACAGACT	CAGGAAACAGCTATGACC-GTGGGTGTCTAGGATCTGGG
POLD1 c.3221G>A	CACGACGTTGTAAAACGAC-ACAGGTGATATACGGCCAGC	CAGGAAACAGCTATGACC-CCAGAGACCACAAAGCACCA
POLD2 c.322C>A	CACGACGTTGTAAAACGAC-AGTCCATTGTCTCTGAACCTGCTT	CAGGAAACAGCTATGACC-CCACACCCAGAGTCCCTGGTC
POLD2 c.892G>A	CACGACGTTGTAAAACGAC-AGATGCTAGGTGGGCCTCTGC	CAGGAAACAGCTATGACC-AATACCTCACCAAGAAAACCCAGGC
POLD4 c.79G>A	CACGACGTTGTAAAACGAC-GGGAGGACAGGCCACCTTTC	CAGGAAACAGCTATGACC-CCCCTCACCCTACACTGCCTA
MLH3 c.885delG	CACGACGTTGTAAAACGAC-TGGCTCCATGCACACATCAT	CAGGAAACAGCTATGACC-CCTATTTTACCAGCTTCCTGTAAGG
MSH3 c.2336G>A	CACGACGTTGTAAAACGAC-CACCAATGTATCTACCTCCCAGCTTTG	CAGGAAACAGCTATGACC-TGGTGGGTCATGAGCTGTAATCTCT
RFC3 c.179A>G	CACGACGTTGTAAAACGAC-AAATTGGGTGTATCTACTCATGAACTG	CAGGAAACAGCTATGACC-CAGGGCCTCAAAGTGAAATGC
RFC4 c.181G>A	CACGACGTTGTAAAACGAC-GGAATCTCTGGCTCTTCTTTATAGCAC	CAGGAAACAGCTATGACC-AACCTCTTTCCTCCTTCTCACCC
RFC4 c.365A>G	CACGACGTTGTAAAACGAC-ATGGAGGAGGGAGGGCAAAT	CAGGAAACAGCTATGACC-GTCTGCAAAGTGTGCAAATGT
LIG1 c.2447T>G	CACGACGTTGTAAAACGAC-CCAAGGTCCCCTCCCTACTCTG	CAGGAAACAGCTATGACC-TTCATTTCTTTCCATCTCCCATTCCT
RPA1 c.1083G>T	CACGACGTTGTAAAACGAC-TCAGTGCTTGCCGTTCATCA	CAGGAAACAGCTATGACC-CCCCATCCCAGCACTTACAT

Appendix 2: DNA concentration of DNA samples

The DNA concentration for all patient samples, measured with Nanodrop and Qubit is shown.

Sample number	Nanodrop (ng/µL)	(Qubit ng/µL)
1	21,54	28,9
2	31,59	43,4
3	13,92	27,0
4	51,48	51,0
5	30,43	32,9
6	76,41	49,3
7	28,07	37,0
8	20,10	28,4
9	8,89	8,2
10	22,46	27,2
11	24,44	22,7
12	83,29	54,0
13	46,70	50,0
14	31,67	36,2
15	22,37	19,6
16	31,16	32,3
17	36,01	50,0
18	41,41	41,8
19	52,78	45,2
20	27,03	30,8
21	31,30	28,8
22	29,10	38,7
23	29,49	26,6
24	33,99	36,2
25	33,02	36,3
26	23,58	22,2
27	39,30	44,1
28	29,73	28,3
29	23,19	22,8
30	29,66	26,2
31	31,22	29,7

32	43,44	43,2
33	22,73	25,5
34	37,22	41,1
35	28,28	15,3
36	31,79	43,1
37	48,29	30,0
38	43,29	14,2
39	59,52	37,8
40	29,63	19,8
41	23,53	20,8
42	26,12	22,8
43	39,51	17,3
44	71,92	27,9
45	27,06	47,3
46	53,19	30,2
47	19,40	25,7
48	25,59	31,4
49	32,73	45,2
50	28,12	27,8
51	98,33	23,1
52	52,57	27,6
53	32,06	29,2
54	39,64	21,8
55	54,57	43,3
56	46,12	46,2
57	22,72	11,6
58	N/A	N/A
59	74,04	48,6
60	31,16	35,0
61	55,81	56,0
62	36,77	51,0
63	28,35	32,3
64	55,30	38,7
65	31,53	38,7
66	49,70	35,5

67	48,63	46,1
68	38,00	36,4
69	62,23	42,3
70	43,86	39,7
71	51,78	40,3
72	21,75	22,6
73	13,54	17,1
74	24,29	33,1
75	83,05	55,0
76	55,39	51,0
77	67,96	56,0
78	64,76	59,0
79	27,42	21,1
80	58,20	54,0
81	110,08	58,0
82	74,62	53,0
83	75,64	46,4
84	42,41	40,0
85	110,38	112,0
86	32,51	29,4
87	56,91	51,0
88	99,49	40,9
89	35,64	32,0
90	74,65	42,8
91	57,37	40,2
92	84,39	65,4
93	62,71	49,5
94	40,19	39,7
95	46,75	32,6
96	56,14	27,2

Appendix 3: DNA concentration of DNA library

The DNA concentration of all samples in the DNA library, measured with Bioanalyzer is shown.

Sample	DNA concentration (ng/µL)
1	4,53
2	5,56
3	3,18
4	4,28
5	4,45
6	9,42
7	4,55
8	1,08
9	5,06
10	2,55
11	3,70
12	7,66
13	3,40
14	3,42
15	2,58
16	3,34
17	5,32
18	5,93
19	3,17
20	4,03
21	2,24
22	4,11
23	5,55
24	3,89
25	3,17
26	2,54
27	2,23
28	3,24
29	3,16
30	3,62
31	5,19

32	5,76
33	1,61
34	4,31
35	4,30
36	2,44
37	4,67
38	5,16
39	2,84
40	3,90
41	3,31
42	4,22
43	7,30
44	4,26
45	2,76
46	3,94
47	4,09
48	4,99
49	3,29
50	4,09
51	4,72
52	4,84
53	4,44
54	3,53
55	5,02
56	6,40
57	4,01
58	3,69
59	6,63
60	3,17
61	3,95
62	2,74
63	3,26
64	5,27
65	4,68
66	5,94

67	1,30
68	3,99
69	4,60
70	5,02
71	4,22
72	2,98
73	2,4
74	3,23
75	3,64
76	5,96
77	3,28
78	3,97
79	3,67
80	4,40
81	6,86
82	3,43
83	5,45
84	3,48
85	2,77
86	2,30
87	3,55
88	3,78
89	4,02
90	4,29
91	6,81
92	2,78
93	3,85
94	4,26
95	3,55
96	3,01

Appendix 3 continues

The DNA concentration of parallel dilutions of the pooled DNA library, measured with Bioanalyzer is shown.

Dilution	Parallel 1 (ng/µL)	Parallel 2 (ng/µL)	Parallel 3 (ng/µL)	Average (ng/µL)
1:1	3,07	3,50	3,76	3,44
1:10	0,47	0,32	0,47	0,42
1:20	0,09	0,10	0,18	0,12

The cycle threshold (Ct) values and final DNA concentration DNA library, measured with qPCR are shown.

Dilution	Mean Ct value	nM
1:2000	10,60	20,06
1:4000	11,50	23,81
1:8000	12,59	24,57

Appendix 4: Number of detected variants and coverage

The average number of variants, average coverage, and % of target region covered for each sample are shown.

Sample	Detected variants	Average coverage	Target covered (%)
1	26	218	86
2	40	155	83
3	32	284	86
4	36	316	87
5	29	226	85
6	36	223	87
7	31	248	86
8	32	239	87
9	38	219	87
10	28	378	89
11	34	263	86
12	33	238	87
13	33	272	87
14	30	303	87
15	30	294	85
16	36	255	87
17	33	213	86
18	37	202	85
19	39	280	87
20	37	282	88
21	32	322	88
22	35	270	89
23	31	245	87
24	39	270	87
25	32	203	84
26	30	335	89
27	29	267	84
28	32	245	86
29	37	295	86
30	32	291	87
31	30	226	85

32	37	119	81
33	25	453	90
34	30	248	87
35	29	260	87
36	38	234	86
37	36	230	86
38	29	482	91
39	32	261	88
40	28	302	89
41	39	273	88
42	31	313	90
43	35	168	85
44	36	272	89
45	28	242	86
46	34	260	87
47	30	227	87
48	34	158	84
49	27	226	86
50	30	210	86
51	33	184	83
52	32	241	86
53	32	186	83
54	32	223	85
55	34	238	87
56	35	206	86
57	38	207	85
58 (control)	N/A	N/A	N/A
59	36	278	88
60	32	250	86
61	30	283	87
62	29	332	90
63	34	248	88
64	36	199	86
65	36	195	86
66	34	231	87

67	27	430	91
68	34	238	87
69	34	259	89
70	34	214	87
71	37	242	86
72	28	308	88
73	26	289	88
74	31	281	87
75	33	362	88
76	39	168	84
77	37	211	86
78	42	235	86
79	31	234	86
80	38	192	84
81	29	285	87
82	35	316	89
83	41	238	86
84	35	274	87
85	35	272	88
86	25	258	87
87	30	317	88
88	30	291	88
89	28	218	87
90	29	244	87
91	31	226	86
92	35	243	87
93	33	303	89
94	45	235	87
95	30	299	89
96	36	329	88

Appendix 5: Detected variants with predictions

All detected variants that were analysed with prediction tools are shown with respective predictions.

Gene	Variant nomenclature	Ref. seq. nr	Sample nr	Polyphen-2	Align GVGD	SIFT	Mutation Taster
EXO1 NM_003686	c.1265G>A p.Ser422Asn Exon 11	rs148510810	88, 86	Benign (0.003)	Class C0	Tolerated (1.00)	Polymorphism
EXO1 NM_003686	c.745G>A p.Asp249Asn Exon 8	rs61750993	6, 25, 43	Possibly damaging (0.946)	Class C0	Deleterious (0.02)	Disease causing
EXO1 NM_003686	c.1828A>G p.Ser610Gly Exon 13	rs12122770	8	Benign (0.004)	Class C0	Deleterious (0.04)	Disease causing
EXO1 NM_003686	c.1670G>A p.Arg557His Exon 13	rs143800705	33, 46	Benign (0.000)	Class C0	Tolerated (0.18)	Polymorphism
EXO1 NM_003686	c.1918C>G p.Pro640Ala Exon 13	rs61736331	42	Benign (0.001)	Class C0	Tolerated (0.88)	Polymorphism
EXO1 NM_003686	c.1378G>C p.Val460Leu Exon 12	rs4149966	63, 67	Benign (0.000)	Class C0	Tolerated (0.65)	Polymorphism
EXO1 NM_003686	c.458C>T p.Ala153Val Exon 7	N/A	28	Probably damaging (1.000)	Class C0	Deleterious (0.02)	Disease causing
EXO1 NM_003686	c.82G>A p.Ala28Thr Exon 4	N/A	30	Probably damaging (1.000)	Class C0	Deleterious (0.00)	Disease causing
MLH1 NM_000249.3	c.2066A>G p.Gln689Arg Exon 18	rs63750702	8, 10	Benign (0.000)	Class C0	Tolerated (0.44)	Disease causing
MLH1 NM_000249.3	c.739T>G p.Ser247Ala Exon 9	rs63750948	14, 20	Possibly damaging (0.948)	Class C0	Deleterious (0.02)	Disease causing
MLH1 NM_000249.3	c.1379A>C p.Glu460Ala Exon 12	rs202038499	42	Benign (0.000)	Class C0	Tolerated (0.46)	Polymorphism
MLH3 NM_001040108	c.2425A>G p.Met809Val Exon 2	rs61752722	2	Benign (0.000)	Class C0	Tolerated (0.30)	Polymorphism
MLH3 NM_001040108	c.885delG p.His296Thrfs*12 Exon 2	N/A	14	N/A	N/A	N/A	N/A
MSH2 NM_000251	c.1321A>C p.T441P Exon 8	N/A	53	Benign (0.011)	Class C0	Tolerated (0.35)	Polymorphism
MSH2 NM_000251	c.138C>G p.His46Gln Exon 1b	rs33946261	9	Probably damaging (0.994)	Class C15	Deleterious (0.00)	Disease causing
MSH2 NM_000251	c.815C>T p.Ala272Val Exon 5	rs34136999	13	Possibly damaging (0.869)	Class C0	Deleterious (0.01)	Disease causing
MSH3 NM_002439	c.2336G>A p.Arg779His Exon 17	rs199791286	36	Probably damaging (1.000)	Class C0	Deleterious (0.00)	Disease causing
MSH3	c.2228A>G	N/A	55	Benign	Class	Tolerated	Polymorphism

NM_002439	p.Gln743Arg Exon 15			(0.002)	C0	(0.20)	
MSH6 NM_000179.2	c.3986C>T p.Ser1329Leu Exon 9	rs199594809	71	Benign (0.019)	Class C0	Tolerated (0.17)	Disease causing
MSH6 NM_000179.2	c.1186C>G p.Leu396Val Exon 4	rs2020908	93	Possibly damaging (0.530)	Class C0	Tolerated (0.11)	Disease causing
MSH6 NM_000179.2	c.1720T>A p.Ser574Thr Exon 4	N/A	13	Benign (0.008)	Class C0	Tolerated (0.88)	Polymorphism
MSH6 NM_000179.2	c.2511C>G p.His837Gln Exon 4	N/A	80	Probably damaging (0.999)	Class C15	Deleterious (0.00)	Disease causing
MSH6 NM_000179.2	c.3261dup Exon 5	rs267608087	41	N/A	N/A	N/A	N/A
PMS2 NM_000535	c.52A>G p.Ile18Val Exon 2	rs63750123	22, 34	Probably damaging (0.998)	Class C25	Deleterious (0.00)	Disease causing
PMS2 NM_000535	c.1789A>T p.Thr597Ser Exon 11	rs1805318	23, 42	Benign (0.004)	Class C0	Tolerated (0.78)	Polymorphism
PMS2 NM_000535	c.1004A>G p.Asn335Ser Exon 10	rs200513014	24	Probably damaging (1.000)	Class C45	Deleterious (0.00)	Disease causing
PMS2 NM_000535	c.137G>A p.Ser46Asn Exon 2	rs121434629	76	Probably damaging (0.996)	Class C45	Deleterious (0.00)	Disease causing
POLD1 NM_001256849	c.433G>A p.Ala145Thr Exon 4	rs137953986	3	Benign (0.068)	Class C0	Tolerated (0.43)	Disease causing
POLD1 NM_001256849	c.2861C>T p.Thr954Met Exon 23	rs374016016	16	Probably damaging (0.999)	Class C0	Deleterious (0.01)	Disease causing
POLD1 NM_001256849	c.80A>T p.Asp27Val Exon 2	rs150066950	90	Benign (0.028)	Class C0	Tolerated (0.06)	Disease causing
POLD1 NM_001256849	c.3221G>A p.Arg1074Gln Exon 27	N/A	2	Probably damaging (1.000)	Class C0	Deleterious (0.03)	Disease causing
POLD2 NM_001127218	c.322C>A p.Leu108Met Exon 3	N/A	33	Probably Damaging (1.000)	Class C0	Deleterious (0.00)	Disease causing
POLD2 NM_001127218	c.892G>A p.Glu298Lys Exon 8	N/A	57	Probably Damaging (1.000)	Class C0	Tolerated (0.74)	Disease causing
POLD4 NM_001256870	c.79G>A p.Glu27Lys Exon 1	rs200868910	38	Benign (0.001)	Class C55	Deleterious (0.00)	Polymorphism
RFC1 NM_001204747	c.1909G>A p.Gly637Ser Exon 14	rs147227437	55	Benign (0.038)	Class 0	Tolerated (0.55)	Disease causing
RFC1 NM_001204747	c.53G>A p.Ser18Asn Exon 1	rs61759896	79, 80	Benign (0.000)	Class C0	Tolerated (0.48)	Polymorphism
RFC3 NM_002915	c.179A>G p.Tyr60Cys Exon 2	rs377157774	74	Probably damaging (1.000)	Class C55	Deleterious (0.00)	Disease causing

RFC4 NM_002916	c.181G>A p.Val61Met Exon 3	rs151335809	73	Probably damaging (0.999)	Class C0	Tolerated (0.20)	Disease causing
RFC4 NM_002916	c.365A>G p.Lys122Arg Exon 5	N/A	43	Probably Damaging (0.967)	Class C25	Deleterious (0.00)	Disease causing