Abstract

Colorectal cancer is the one of the most frequent malignancies and accounts for approximately 3500 new cases each year in Norway. Colorectal cancer is a heterogeneous disorder and can be divided into three main groups: sporadic, familial, and hereditary. The most common hereditary syndrome is Lynch syndrome. It is caused by a mutation in one of the DNA mismatch repair genes: *MSH2*, *MSH6*, *MLH1*, and *PMS2* and follows the microsatellite instability pathway. Surveillance programs can help to reduce the cancer risk and mortality of these individuals. Currently, no entirely satisfactory diagnostic tool is available to identify Lynch syndrome patients. Clinical diagnostic criteria are not accurate enough and Sanger sequencing is too expensive and time-consuming to perform large-scale molecular screening analysis. To allow sequencing of more samples the Roche GS Junior System, a downscaled next-generation sequencing platform, was established at the Medical Genetics Laboratory, St. Olavs Hospital. In this master study efficient workflows were developed for massive parallel amplicon sequencing of the MMR genes *MSH2*, *MSH6*, *MLH1*, and *PMS2*. Thereby, the sample throughput was increased and the costs were reduced. Another aim of the master project was the performance of a pilot study to determine the suitability of tumor screening analysis for identification of hereditary and non-hereditary CRC subtypes. The results indicate that most individuals with Lynch syndrome can be identified using tumor analysis for selecting samples that have to be sequenced.

Acknowledgements

The present Master's thesis was conducted in the laboratory of the Medical Genetics department at the St. Olav's Hospital under supervision of Dr. Wenche Sjursen and Liss Anne S. Lavik.

I would like to thank Wenche Sjursen for giving me the opportunity to work on this interesting project. Thank you very much for your feedback, the support during writing process and your patience. It was good to have you as a supervisor for the Master's as well as Bachelor's thesis.

My special thanks go to Liss Anne S. Lavik who has provided great assistance during the daily laboratory work. Thank you for your advices, for help with reviewing my thesis, and that you always took the time to answer my questions.

Many thanks to Maren F. Hansen for your help during reviewing process and the good teamwork in the laboratory. Thanks for listening and your support. Wish you the best for your doctoral study.

Thanks to Trine Vold for the good teamwork and your help when I had question regarding your master thesis.

I would like to express my gratitude to the whole staff of the Medical Genetics laboratory, for their assistance in the laboratory, their kindness and patience especially regarding language problems. Thanks for the nice lunches and conversations.

Thanks to my parents and friends for their support and for always believing in me. In particular, I would like to thank Samita Wilson and Mayte Troncoso Abelleira for the cheerful as well as encouraging conversations. Thanks girls!

Trondheim, June 2013

List of Abbreviations

AM II	Amsterdam II
APS	Adenosine phosphosulfate
ATP	Adenosine triphosphate
AVA	Amplicon Variant Analyzer
BAC	Bacterial Artificial Chromosome
bp	Base pair
CCD	Charge-coupled Device
cDNA	Complementary DNA
CIN	Chromosomal instability
CRC	Colorectal cancer
EDTA	Ethylenediamine tetraacetic acid
emPCR	Emulsion PCR
FAP	Familial adenomatous polyposis
FGS	First generation sequencing
gDNA	Genomic DNA
HNPCC	Hereditary non-polyposis colorectal cancer
HPs	Homopolymer runs
IDL	Insertions/deletion loop
IHC	Immunohistochemistry
LS	Lynch syndrome
MC	Minimal coverage
$MgCl_2$	Magnesium chloride
MID	Multiplex Identifier
MMR	Mismatch repair
MPC	Magnetic Particle Concentrator
MSI	Microsatellite instability
MSI-H	High-frequency microsatellite instability
MSI-L	Low- frequency microsatellite instability

NaOH	Sodium hydroxide
NGS	Next generation sequencing
PMS2 CL	PMS2 C-terminal like pseudogene
PPi	Inorganic pyrophosphate
PTP	Pico TiterPlate
RBG	Revised Bethesda Guidelines
SBS	Sequencing by synthesis
SGS	Second generation sequencing
SMS	Single-molecule sequencing
TGS	Third generation sequencing
VUS	Variant of unknown significant

Table of Content

1 Introduction

1.2 Colorectal Cancer

The International Agency for Research on Cancer recorded a worldwide incidence of 1,235,108 and a mortality of 609,051 for colorectal cancer (CRC) in 2008[1]. In Norway, CRC is the most common malignancy after prostate and lung cancer in males and breast cancer in females, with about 3,500 new cases each year[2].

CRC can be classified into three groups: (1) sporadic CRC (~ 60%), these patients have no noticeable family history and no inherited gene mutation; (2) familial CRC (~ 30%), these patients have at least one close relative with CRC, however, it is not possible to identify a germline mutation or obvious pattern of inheritance; (3) hereditary CRC syndromes (~ 10%) these patients inherited a single gene mutation in a highly penetrant cancer susceptibility gene from an affected parent[3]. The absence or presence of multiple colorectal polyps is used for classification of hereditary CRC syndromes into two categories: non-polyposis and polyposis disorders[4]. The most common hereditary syndrome is Lynch syndrome (LS), also known as hereditary non-polyposis colorectal cancer (HNPCC) which affects 2% – 5% of all CRC patients[5]. Familial adenomatous polyposis (FAP), which is the most common polyposis disorder, affects less than 1% of all CRC patients[4,6].

Although the exact genetic mechanisms underlying hereditary and sporadic CRC differ, the CRC development can be divided into two distinct pathways; chromosomal instability (CIN) and microsatellite instability (MSI). CIN also known as microsatellite stability pathway can be found in about 85% of the CRC cases and is characterized by allelic losses, chromosomal amplifications and translocations[7]. These tumors arise more often in the “left colon”, distal to the splenic flexure; they have aneuploid DNA, show distinctive mutations for example in *K-Ras*, *APC*, and *p53*, are aggressive and have a poor prognosis[8,9]. One example for the CIN pathway is FAP[9]. In contrast, MSI is characterized by changes in the length of microsatellites[7]. Tumors with MSI are predominately located in the “right colon”, proximal to the splenic flexure; they have diploid DNA, distinctive mutations for example in *TGF-RII* and *BAX*, are indolent and a better prognosis[8,9]. LS is one example for the MSI pathway[7].

1.3 Lynch Syndrome

LS is inherited in an autosomal dominant way, affecting about 50% of the offspring[10]. With an estimated population incidence between 1:2,000 and 1:660, LS is one of the most frequent highly deleterious heritable disorders[11]. It is caused by a germline mutation in one of the four DNA mismatch repair (MMR) genes: *MLH1*, *MSH2*, *MSH6*, and *PMS2*[5]. In 2004 at the international meeting in Bethesda one agreed to use the term "Lynch syndrome" instead of "HNPCC" since the latter is partly misleading[12]. Contrary to the term HNPCC, colorectal polyps can be observed and extracolonic tumors are common[13].

1.3.1 Clinical Presentation

In addition to the risk for CRC, individuals with LS have an increased risk for malignancy at certain extracolonic sites; carcinoma of the endometrium, ovary, stomach, urinary tract, small bowel, brain, and hepatobiliary tract. The penetrance for these cancers depends on various factors, for instance MMR gene affected, type of mutation, and gender[10]. The penetrance is with 70% by age 70 highest for colorectal and endometrial cancer[14]. The life time risks for the other carcinomas are considerably lower[12].

CRC in LS shows several characteristic features. Several studies show that the average age of onset of CRC in LS (45-60 years[15,16]) is lower than in sporadic CRC (65 years[15]). In addition, CRCs in LS arise more frequently in the proximal colon and these tumors show an accelerated adenoma-carcinoma sequence[17]. Individuals with LS tend to have multiple tumors; synchronous and metachronous. The tumors are histologically characterized by poor differentiation, signet-cell features, mucin production, and lymphoid infiltration of tumor[15]. Despite the poorly differentiated histology, LS-related CRCs are associated with better survival rates when controlled for stage and age of sporadic CRC[4].

1.3.2 Diagnosis

LS patients are identified by a step-by-step approach: (1) clinical diagnostic criteria, (2) molecular tumor analysis, and (3) germline testing of DNA MMR genes. Currently, two sets of clinical diagnostic criteria exist. The Amsterdam II (AM II) criteria are used to detect possible cases of LS. In contrast, the Revised Bethesda Guidelines 2004 (RBG) are used to identify CRC patients that are in need for molecular tumor analyses (MSI) and/or immune-histochemistry (IHC) testing. In case MSI or loss of MMR expression is detected in the tumor samples, sequencing analysis of the MMR genes is performed to confirm the diagnosis. [12,18]

AM II criteria and RBG are given in appendix 6.1. AM II criteria are based on the presence of LS tumors in two close relatives of the proband. However, small family sizes, unknown family history, adoption, and incomplete medical records can make it difficult to fulfill this criterion[11,19]. In several studies it was shown that it is not possible to identify all LS patients by means of these clinical guidelines[12,20]. The AM II criteria have a high rate of false positive LS cases, meaning that AM II criteria identified a large proportion of possible LS cases that cannot be confirmed with molecular analysis[20]. In contrast, about 50% of the tumors that are likely to be related with LS could not be identified by means of the RBG[20].

In IHC antibodies directed against the four MMR proteins are used to test for the presence or absence of protein expression. Because the IHC results can indicate the causative gene defect, they are used to direct the germline testing of the MMR genes. In this way, time and cost consumption can be reduced. The weakness of IHC is that the antibodies might detect a fragment of a truncated protein and thereby giving rise to false negative results. [12,21]

To be able to detect MSI, numerous cells have to possess the same alteration in the microsatellites and this in turn is an indication of clonal expansion which is characteristic for neoplasms[22]. Depending on the number of microsatellite markers showing instability, tumors are categorized into high-frequency (MSI-H) and low-frequency (MSI-L) MSI. According to the U.S. National Cancer Institute tumors which show mutations in two or more microsatellite markers out of a group of five belong to the first category[23]. If only one microsatellite marker is mutated the tumor falls into the second category. Given that MSI also occurs in about 15% of sporadic CRC cases, MSI analysis alone is insufficient to identify LS patients[7]. In sporadic CRC MSI results from hypermethylation of the promoter of one of the MMR genes, most often *MLH1*[24]. To discriminate between sporadic CRCs and LS, methylation analysis and *BRAF* mutational testing is used, because hypermethylation and *BRAF* mutations can frequently be found in sporadic CRCs with MSI, but not in LS related tumors[10,25].

If tumor analysis indicates LS, mutational screening of the MMR genes via sequencing analysis is performed to identify the underlying pathogenic mutation. This analysis is hampered, because there are no consensus mutational “hot spots” in the MMR genes and a broad spectrum of truncating, frameshift and missense mutations are associated with LS[21]. In order to avoid confusion, families with strong evidence of MMR deficiency may be said to have LS, whereas families that fulfill the AMII criteria without evidence of a germline mutation in a MMR gene are said to have “familial colorectal cancer type X”[13].

1.3.3 Surveillance and Management

Jarvinen *et al.* showed that it is possible to reduce the risk of CRC and the overall mortality by approximately 65% in LS families with the aid of colonoscopy at three year intervals[26]. Since cancer development in LS can be relatively rapid, surveillance every 1-2 years is recommended[12]. For patients diagnosed with CRC full colectomy with ileorectal anastomosis is the method of choice[27]. Extracolonic malignancies of LS patients are treated like sporadic cases. In some cases prophylactic surgery which is the removal of an organ (high-risk and nonessential) without the presence of cancer is recommended. Also the choice of a suitable chemotherapeutic agent is important. In vitro studies showed that cells lacking MMR are resistant to several chemotherapeutic drugs which are normally used for CRC treatment[10,28].

1.3.4 Mismatch Repair System

LS is caused by deficient MMR activity which is due to a heterozygous germline mutation followed by somatic inactivation of the remaining wild-type allele of one of the four MMR genes. Distribution of pathogenic variants is 40% in *MSH2*, 39% in *MLH1*, 15% in *MSH6*, and 6% in *PMS2*[29]. There are several mutations that occur repeatedly in unrelated individuals with LS. This might be due to two reasons: (1) Genetics or other factors promote the mutation or (2) the mutation is a so called founder mutation[30]. In populations with a high percentage of founder mutations this knowledge can be used to facilitate genetic analysis. *De novo* mutations are uncommon[10].

The MMR pathway is highly conserved and primarily functions in the correction of base-base mismatches and insertions/deletion loops (IDL) which occur during DNA recombination and replication. The latter one occurs when the DNA polymerase complex “slips out of position” during replication of microsatellites[23]. These are repetitive DNA sequences which can be found in the whole genome and are build up of one to six base pairs that are repeated up to 100 times[13]. In eukaryotic cells MMR is also involved in DNA damage signaling, where it plays a role in cell cycle arrest and/or programmed cell death induced by some types of DNA damage. Moreover, MMR suppresses homeologous recombination. Depending on the type and size of the DNA replication error, different MMR genes are involved in the repair process. The MMR proteins work as heterodimers. For the correction of single base-base and IDL mismatches of 1 or 2 nucleotides a complex of MSH2 and MSH6 is formed. The MSH2 and MSH3 complex plays a role in the repair of larger IDL mispairs. The principle order of events in MMR for MutSα (MLH1-PMS2) is shown in Figure 1.1. [28]

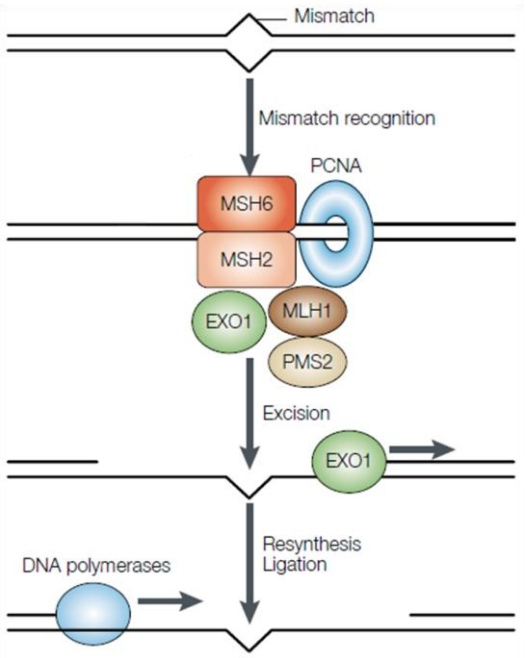

Figure 1.1. Order of events in human mismatch repair. MSH2-MSH6 heterodimer recognizes and binds to single base-base and small IDL mismatches. The next step is binding of the MLH1-PMS2 heterodimer and the formation of a complex. EXO1 exonuclease and PCNA are recruited to the complex. EXO1 is responsible for the removal of the mismatched base. The resulting gab is then filled in by DNA polymerase Polδ and the nick is ligated by DNA ligase I. Adapted from Martin and Scharff, 2002[31].

1.3.5 Microsatellite instability

A deficient MMR results in mutator or replication-error phenotype. In other words, inactivation of a MMR gene does not directly give rise to malignant transformation; however, mutations arising during DNA replication cannot be repaired properly anymore by the MMR system and accumulate[30]. As these mutations are predominantly deletions or insertions in microsatellites this type of genetic instability is called MSI[7].

If the mutated microsatellite is present in an intronic or noncoding sequence then the mutation presumable will have no effect. In the case the microsatellite is located in the coding region of a gene and this gene is important for regulation of cell growth or maintenance of genomic stability, the loss of the proper function can cause transformation to malignancy. Examples for genes affected by MSI are: (1) receptors for growth factors (insulin-like growth factor II receptor); (2) regulators of the cell cycle (*E2F4*); (3) regulators of apoptosis (*BAX*); and (3) mismatch repair genes (*MSH3*, *MSH6*). In contrast, the genes *p53*, *APC*, und *K-ras* are seldom affected by MSI[7]. [23]

1.3.6 *PMS2* Pseudogenes

Pseudogenes possess a high degree of sequence homology to a non-allelic functional gene, however they acquired an inactivating mutation and are non-functional. Pseudogenes are classified into nonprocessed and processed pseudogenes based on the gene duplication mechanism that led to their formation. In the majority of cases, nonprocessed pseudogenes arise by tandem duplication and are located in close proximity to their original gene. In these cases the promoter, upstream regulatory sequences, and all exons and introns might be duplicated. In contrast, processed pseudogenes arise by retrotransposition; where complementary DNA (cDNA) is made from a processed gene transcript, for instance mRNA, by a cellular reverse transcriptase and subsequently integrated into chromosomal DNA. This often results in interspersed gene families. These pseudogenes only contain exonic sequences and lack intronic and upstream promoter sequences. Therefore, they are normally not expressed. However, sometimes the cDNA is integrated adjacent to a promoter and functional gene products are produced. Currently, more than 12 000 pseudogenes are known in humans. It is believed that gene duplication is evolutionarily advantageous, since new functional gene variants can be created by this mechanism. In this connection, pseudogenes represent unsuccessful by-products. However, it has been observed that some pseudogenes function in gene regulation. [32]

The existence of several *PMS2* pseudogenes has been demonstrated by studies[33-38]. *PMS2* is located on chromosome 7 and has 15 exons. All 15 pseudogene loci are located in close proximity and are classified as nonprocessed[38]. It has been shown that several of these loci are transcribed and can hamper genomic DNA (gDNA) and cDNA mutation analyses[33,34,39]. 14 loci contain all or some of exons 1 – 5 and the PMS2 C-terminal like pseudogene (*PMS2 CL*) contains exons 9 and 11 – 15[38].

1.4 Sequencing Technologies

Sequencing technologies are used in both research and molecular diagnostics. Applications include de novo sequencing[40,41], detection of genome wide structural variations in the human population[42], cancer research[43], studies of inherited disorders[44], studies of complex human diseases[45], and mutation detection and carrier screening in hereditary cancer syndromes[46]. Since the development of the first sequencing methods in the late 1970s, the technologies were permanently improved and new ones developed. Until now, no standardized classification system exists for the various new platforms. Here the platforms will be divided into three groups (generations) based on their fundamental technology as described by Schadt *et. al.*, 2010[47]. The properties of an ideal sequencing method are high accuracy, easy handling, low running and acquisition costs, and low expenditure of time[48].

1.4.1 First Generation Sequencing Methods

The first generation sequencing (FGS) technologies, Maxam Gilbert and Sanger sequencing, are based on the separation of DNA fragments according to their size. These fragments are generated by either sequencing by synthesis (SBS) or degradation and they are identified by means of their end-labeling[47]. Sanger sequencing method is based on base-specific chain termination by incorporation of dideoxynucleotide triphosphates (ddNTPs) into the growing DNA chain. Four parallel reactions have to be conducted, each containing the four deoxynucleotide triphosphates (dNTPs; dTTP, dATP, dCTP, dGTP) and either ddTTP, ddATP, ddCTP, or ddGTP. For the purpose of detection one of the four dNTPs was radioactively labeled with ^{32}P. The resulting fragments were separated on a polyacrylamide gel and the sequence directly read from the pattern of bands (Figure 1.2A). [49]

Sanger sequencing was further developed and became the “gold standard”, as this method is easier to use and scale up than Maxam-Gilbert sequencing method[47]. Since 1990s automated fluorescence and capillary DNA sequencing instruments are available[50]. The usage of fluorescence dyes to label the base-specific reactions is advantageous, because all four reactions can be conducted in one reaction mixture (Figure 1.2B). With the Sanger 3730xl sequencer from Applied Biosystems read length between 400 and 900bp can be achieved[48].

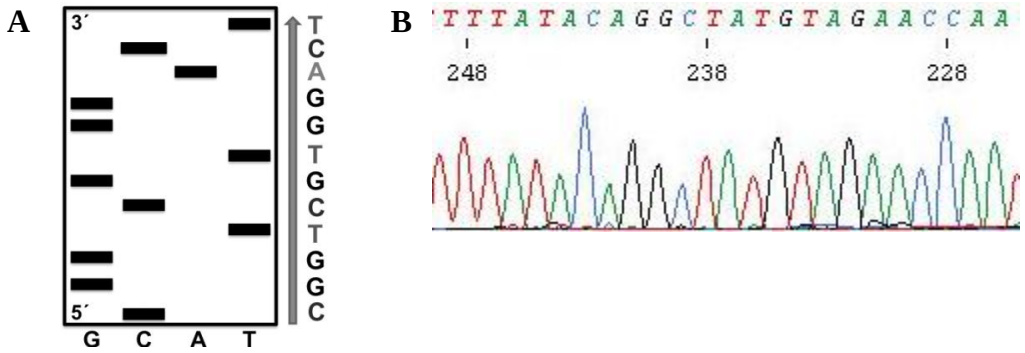

Figure 1.2. Sanger sequencing method. (A) Schematic representation of an autoradiograph used to obtain the sequence of a DNA fragment. Adapted from Anasagasti *et. al.,* 2012[51]. **(B)** Electropherogram obtained by SecScape. Sanger sequencing method using fluorescence dyes and capillary electrophoresis.

1.4.2 Second Generation Sequencing Technologies

Due to limitations in throughput and high costs of Sanger sequencing, new technologies were developed. To overcome these limitations, the new technologies focused on two areas: (1) Decrease of reaction volume needed for amplification and (2) increase of the number of amplifications which can be performed simultaneously. This miniaturization leads to a decreased sample and reagents consumption, shorter amplification times and increased sequencing throughput. [52]

In 2005, the first next generation sequencing (NGS) or second generation sequencing (SGS) technology was launched by Roche[48]. Since then a tremendous development has taken place to increase the throughput while simultaneously reducing the costs. Second generation systems are GS FLX+/GS Junior from Roche, Genome Analyzer/HiSeq 2000/MiSeq from Illumina, and SOLiD/ Ion PGM™ /Ion Proton™ from Life Technologies. All SGS technologies available have in common that fragmented DNA is anchored to a solid surface, amplified, and hundreds to thousands of identical strands are sequenced in parallel. The sequencing takes place in cycles which are repeated until the desired read length is achieved. The particular steps in a cycle depend on the SGS technology. In general the following steps are included in a cycle: (1) sequential addition of nucleotides and reagents, (2) incorporation of nucleotides into the growing DNA strand, in case the nucleotide is complementary to the template base, (3) termination of the incorporation reaction, (4) removal of unincorporated nucleotides and reagents, (5) identification of the incorporated nucleotide by scanning, and (6) preparation of the strand end for the next cycle, this is done by cleaving the terminating/inhibiting group and the fluorescent dye from the newly incorporated nucleotide and a washing step[53]. [47]

The requirement of PCR amplification in SGS has several negative consequences. Sample preparation is more complex and time-consuming. Amplification of some DNA fragments might be favored, resulting in an amplification bias. Furthermore, the DNA polymerase can incorporate the wrong nucleotide whereby leading to the introduction of errors in the DNA which serves as a template in the subsequent sequencing reaction. Another negative effect of sequencing clonal amplified fragments is the occurrence of dephasing, which results in an increased fluorescence noise, base-calling errors and decreased read length. Dephasing is observed when sequencing reactions of the identical templates do not occur synchronously, meaning that in some reactions multiple nucleotides were added in one cycle or extension of the template was incomplete. [47,53]

The read length of most of the platforms of the second generation is between 25bp and 250bp, depending on the particular platform and sequencing settings[54-57]. Longer read length can be obtained by the ion torrent sequencing technology from Life Technologies and the GS FLX+ System from Roche. Using these platforms the read length can be up to 400bp[58] and 1000bp[59], respectively.

The classification of the HeliScope™ Single Molecule Sequencer from Helicos is not straightforward. Although, this sequencer is the first commercially available platform that performs single-molecule sequencing (SMS) and does not require clonal amplification, it still relies on a "wash-and-scan" process. Therefore, this platform represents a junction between SGS and third-generation sequencing (TGS). [47]

1.4.3 Third Generation Sequencing Technologies

The platforms of the third generation are characterized by the ability to perform single-molecule sequencing (SMS) in a continuous manner. Thus, no halt is required between each sequencing cycle, which is in contrast to SGS. Thereby, read length and throughput can be increased, and time consumption of sample preparation as well as sequencing costs reduced. The various TGS platforms can be divided into three categories based on the measuring principle: (1) observation of individual DNA polymerase complexes while they synthesis a single DNA molecule, (2) detection of individual bases using nanopore technology, and (3) visualization of DNA molecules and identification of individual bases by sophisticated microscopic devices. [47]

1.5 Roche 454 Sequencing: GS Junior System

The SGS platform used in this master project is the GS Junior System from Roche which is based on the 454 Sequencing Technology. This downscaled platform has an average read length of 400 base pairs (bp) and provides the opportunity to sequence multiple amplicons of several patients in parallel. The workflow consists of four parts (Figure 1.3): (1) library preparation, (2) emulsion PCR (emPCR) amplification, (3) pyrosequencing, and (4) data processing and analysis. [60]

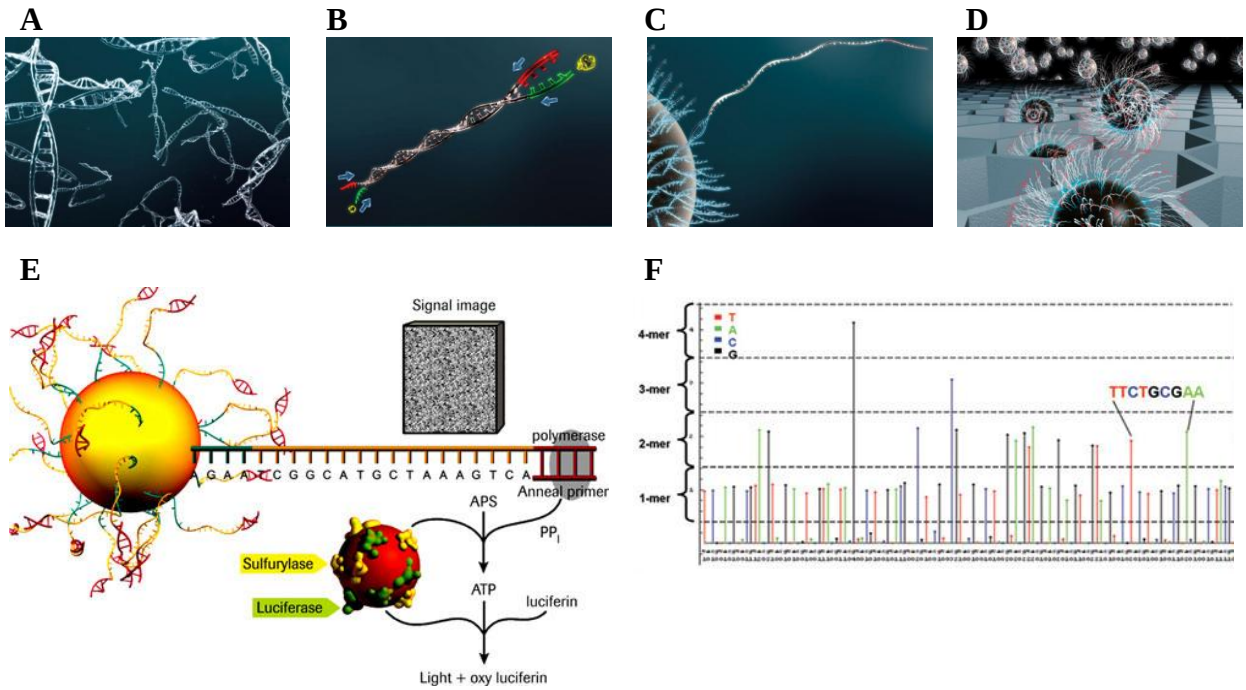

Figure 1.3. Overview of the 454 pyrosequencing technology. (A) The starting material such as PCR products, gDNA, cDNA, and Bacterial Artificial Chromosomes (BACs). **(B)** Library preparation: The Library Adaptors can either be ligated to the nucleic acid fragments or added by fusion primers in a PCR. The adaptors are needed during purification, quantitation, amplification, and sequencing. **(C)** Emulsion PCR: Clonal amplification of template DNA on a solid support. Preferably, only one single DNA species is attached to the DNA capture bead. **(D)** Loading of DNA capture beads carrying millions of clonally amplified DNA fragments onto Pico Titer Plate device. **(E)** ATP-sulfurylase/luciferase enzyme cascade. The steps are: (1) Incorporation of a complementary nucleotide into the growing DNA strand by DNA polymerase and release of inorganic pyrophosphate (PPi). (2) Generation of adenosine triphosphate (ATP) by ATP-sulfurylase using PPi and adenosine phosphosulfate (APS) as a substrate. (3) Usage of the ATP released in the second reaction to convert the substrate luciferin to oxyluciferin and light[61]. One bead represents one read. **(F)** Analysis of signal intensities to determine the sequence of the DNA fragments. The amount of photons is directly proportional to the number of nucleotides incorporated into the growing DNA strand. Adapted from the 454 Sequencing homepage[62].

1.5.1 Library Preparation

The first step of the sequencing process is the preparation of a DNA library. For this study, the analysis of the four MMR genes, targeted resequencing was chosen. In this approach only the regions of interest are sequenced and thereby the cost and time consumption are reduced. The term "amplicon" refers to any nucleic acid molecules generated by a nucleic amplification method performed in vitro [63]. Roche provides five experimental designs to prepare such an amplicon library: (1) Basic Amplicon sequencing, (2) Universal Tailed Amplicon sequencing, (3) Ligated Adaptors Amplicon sequencing, (4) Long Range PCR Amplicon sequencing, and (5) One-way Reads Amplicon sequencing. Library preparation is often the most time consuming and expensive part of the sequencing experiment and has therefore to be planned thoroughly. Universal Tailed Amplicon sequencing was chosen for this study. This approach allows, among other things, bidirectional sequencing of several samples in parallel and reduction of the number of primer pairs. [60]

Bidirectional sequencing is favored for detection of variants, as the forward and reverse reads, which represent two different sequencing reactions, are used for the independent verification of the validity of a variant. Furthermore, the accuracy of the basecalling is higher at the beginning of each read due to the effects of dephasing. Thereby, the more accurate sequencing information at the beginning of the forward and reverse reads complements each other. Accuracy of basecalling also depends on the environment of the particular sequencing reaction. It might be that some DNA regions cannot be sequenced in either forward or reverse direction, but in the opposite direction the sequencing might be performed successfully. Thus, bidirectional sequencing is used as it provides higher sequencing accuracy than unidirectional sequencing. [60]

When using the Universal Tailed Amplicon sequencing design, the library is prepared by two consecutive rounds of PCR and two sets of fusion primers (Figure 1.4). In the first PCR, primers containing both template-specific and universal sequences are used to amplify the regions of interest and to add universal tails. The directionality is preserved by usage of two different universal tails, Univ-A and Univ-B. These universal tails are targeted by another set of fusion primers in the second PCR. These primers consist of four parts: (1) Sequences complementary to Univ-A and Univ-B, respectively, (2) Multiplex Identifier (MID) barcode sequence, which allows identification of the samples, (3) key sequence which is used for signal calibration and read approval, and (4) 454 sequencing system primers A and B which allow directional sequencing of the sequencing target from 5' and 3' end. [60]

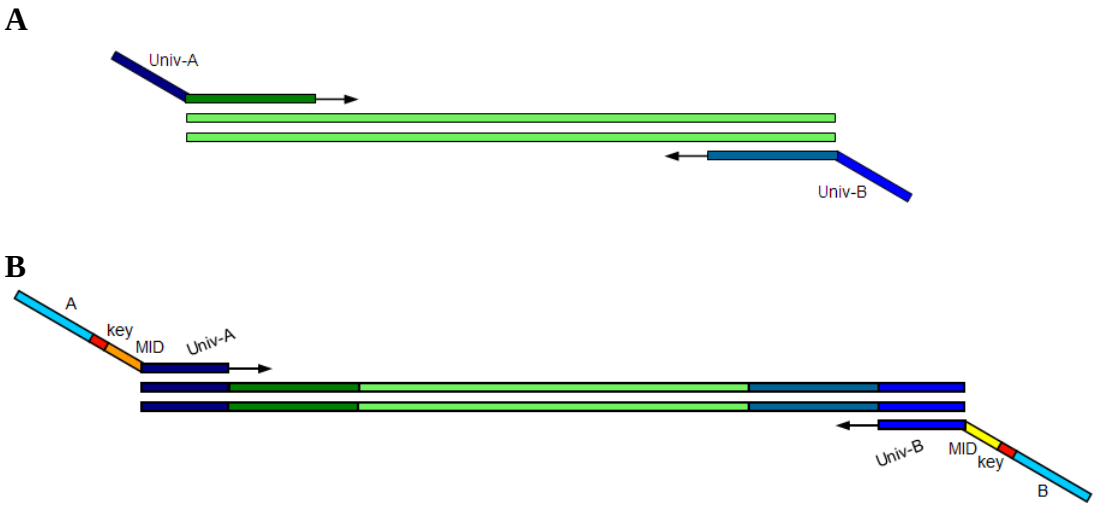

Figure 1.4. General structure of the fusion primers used for the library preparation according to the Universal Tailed Amplicon Library design from Roche. (A) First PCR, amplification of regions of interest and addition of universal tails, Univ-A and Univ-B. **(B)** Second PCR, addition of MID sequence, key sequence, and 454 Sequencing System primers, A and B, via targeting universal tails. [60]

1.5.2 emPCR Amplification

As already mentioned SGS technologies, including 454 sequencing, relies on sequencing of millions of copies of identical amplicons attached to a solid support. To optimize the sequencing process, hundreds to thousands of amplicons from a DNA library are sequenced in parallel. In order to allow for clonal amplification of such high numbers of amplicons in a cost and time efficient way, the 454 technology uses emPCR. This method combines the reliability and swiftness of PCR with the possibility to clonally amplify multiple samples simultaneously in one reaction tube without amplification bias and cross contamination. [63]

EmPCR is based on attachment of nucleic acids to capture beads which serve as a solid support and amplification of template DNAs in a heat-stable water-in-oil emulsion. One possibility to attach fragments to spherical beads is to cover the surface of the beads with numerous copies of a specific primer that binds complementary to a region of the template DNA and the amplification products. For that reason, amplicons of a DNA library have to contain constant regions at the 3' and 5' end that are complementary to the single primer species on the beads. In 454 sequencing these sequences are called system primer A and B (Figure 1.4). For amplification of nucleic acid all PCR components including template DNA, DNA capture beads, and PCR reagents (salts, dNTPS, primers, and polymerase) are added to the emulsion oil and emulsified. During this process aqueous microdroplets are formed. Only droplets containing all reagents required for PCR will take part in DNA amplification. [63]

The success of the emPCR is highly dependent on the starting material. It is essential to check the quality and quantity of the PCRs products before starting the emPCR procedure. It is to be ensured that the PCR products of the different DNA fragments have a uniform appearance and that no or only minor amounts of adapter dimers (e.g., ~ 90 bases) are visible on the gel. The determination of the exact DNA concentration is required to achieve an optimal mixing ratio of the template DNA and the DNA capture beads. For the pyrosequencing reaction it is crucial that one single DNA template is attached to a capture bead and clonally amplified. As distribution of DNA templates occurs by chance, different types of beads are present in the emulsion. The beads can have one or more DNA species or they can be empty. The droplets without template DNA will not participate in the amplification process. This type of beads will be removed in a purification step in which beads containing DNA are enriched. The number of beads containing more than one DNA species has to be kept as low as possible. Because they cannot be removed in the enrichment step and pyrosequencing signals originating from such beads cannot be evaluated and have to be filtered out. At the same time they will take up space on the Pico TiterPlate (PTP) device. The proportion of each bead population is dependent on the ratio of DNA templates to capture beads. DNA-to-bead ratios below 1 should be chosen to favor the generation of beads associated with no DNA and with only one DNA species. [63]

After emPCR the beads with the clonally amplified fragments have to be prepared in several steps for the subsequent pyrosequencing reaction. In order to make the DNA on the beads accessible the emulsion has to be broken. The amplification products are still bound to their particular DNA capture bead after this demulsification step. The next step is the enrichment of beads carrying DNA. Since the population of beads with no DNA represents the largest group and they do not have any positive value for the subsequent sequencing reaction, they have to be removed to increase the efficiency. Afterwards, the beads carrying the prepared single-stranded DNA fragments can be further processed and sequenced. [63]

1.5.3 Pyrosequencing

Pyrophosphate sequencing technology, which is based on sequencing by synthesis, was developed by Ronaghi and Nyrén in the mid 90s[64,65]. In contrast to Sanger sequencing, the nucleotide triphosphates are added sequentially to the reaction mixture and are not fluorescently labeled. The cycles of nucleotide addition are repeated until the desired read length is achieved. In case the introduced dNTP is complementary to the template strand, it will be incorporated and the DNA strand increases by one nucleotide or more nucleotides if a

homopolymer stretch is present. Both the type and number of nucleotides incorporated are determined by means of an ATP-sulfurylase/luciferase enzyme cascade (Figure 1.3E). The light signal is detected and quantified by a Charge-coupled Device (CCD) camera. The amount of photons generated is reported to be directly proportional for up to 8 nucleotides. [61]

The fundamental pyrosequencing principles are adopted and adjusted to high-throughput sequencing on PTP devices by 454 Sequencing. To decrease the amount of enzymes, that are needed for the sequencing process, to a minimum and thus reducing the costs, 454 Sequencing uses enzymes immobilized on solid microparticles. To meet the requirements of pyrosequencing in picoscale format, enzymes with specific properties had to be chosen. The enzymes used by Roche are more efficient and optimized than in standard pyrosequencing. Thereby, improved accuracy and longer read length can be achieved. In addition, the GS Junior Titanium Sequencing Kit contains PPiase beads to reduce noise signals by preventing the flow of PPi from one well into neighboring wells. [61]

1.5.4 Acquisition, Processing, and Analysis of Data

The 454 Sequencing System performs data handling in three main steps: (1) data acquisition, (2) data processing, and (3) data analysis. For each step particular analysis software is required. During data acquisition phase a CCD camera takes an image of the PTP surface during every nucleotide flow. The series of digital images is collected and stored by the GS Junior Sequencer software. These images give information about the amount of light emitted during a particular nucleotide flow; which is proportional to the number of nucleotides incorporated. [66]

The function of the data processing phase is conversion of raw image data to base-called results that can be used in the subsequent data analysis phase. Data processing is controlled by the GS Run Processor application and is divided into two steps; image processing and signal processing. Processes carried out during image processing include background subtraction, image normalization, identification of active wells, and extraction of raw signals for each flow from active wells. During signal processing well-level calculations are performed to generate well “flowgrams” and the basecalls of the amplicons in all the active wells. [66]

The last step is the data analysis. Roche offers three software programs to analyze the fully processed and “trimmed” read basecalls of a GS Junior run. In this study the GS Amplicon Variant Analyzer (AVA) software was used. The program compares the reads generated in the sequencing run to a reference sequence in order to detect and identify sequence variants. [66]

1.6 Aims

CRC can be sporadic, familial or hereditary. The identification of the subtype plays a crucial role in the choice of a suitable therapy. This master thesis, which is part of the research project "Colorectal Cancer in Central Norway; Identification of Hereditary and Non-Hereditary Subtypes", focuses on the most common hereditary CRC syndrome, known as Lynch syndrome. It is important to identify individuals with LS as early as possible and to include them in surveillance programs, genetic counseling and testing, because studies have shown that the overall mortality can be drastically reduced by such measures. At the moment no entirely satisfactory diagnostic tool is available. Clinical diagnostic criteria are not accurate enough and Sanger sequencing is too expensive and time-consuming to perform large-scale molecular screening analysis for all CRC patients.

The aim of this master study was to implement the GS Junior platform at the Department of Pathology and Medical Genetics, St. Olavs Hospital for sequencing analysis of the four MMR genes *MSH2*, *MSH6*, *MLH1*, and *PMS2*, which are involved in the pathogenesis of LS. The new sequencing workflow shall enable sequencing analysis of more patient samples in less time and with reduced costs. In addition, a pilot study was performed to evaluate the suitability of tumor screening analysis for identification of hereditary and non-hereditary CRC subtypes.

Sequencing of the MMR genes by the Sanger method is already well-established in the Medical Genetics laboratory. However, the transfer from Sanger sequencing to SGS technology in this instance the GS Junior System requires testing and validation of the new method. Within this study, two workflows for sequencing of *PMS2* and the four MMR genes in parallel were developed. Universal tailed amplicon sequencing was used for preparation of the amplicon libraries of both workflows. Preparation of the amplicon libraries were optimized thoroughly to obtain a uniform distribution of coverage. Thereby, the full capacity of the GS Junior sequencing platform can be exploited and the number of samples sequenced per run can be maximized. The properties, such as performance, cost and throughput of both Sanger sequencing and GS Junior system were compared. The sequencing results of the optimization of MMR genes library preparation were used for the pilot study. These results were compared with results of tumor analysis such as MSI, and *BRAF* mutation analysis and *MLH1* methylation.

2 Materials and Methods

2.1 Equipment

3130xl Genetic Analyzer with 36 cm capillary	Applied Biosystems, Foster City, USA
Agilent 2100 Bioanalyzer	Agilent Technologies, Inc., Santa Clara, CA, USA
Centrifuge 5415 R, Rotor F45-24-11	Eppendorf AG, Hamburg, Germany
Centrifuge 5430 v4.3, Rotor A-2-MTP	Eppendorf AG, Hamburg, Germany
Centrifuge 5810 R, Rotor A-4-81	Eppendorf AG, Hamburg, Germany
Dry block heating system QBD2	Grant Instruments Ltd, Shepreth, Cambridge, England
Dynal MPC®-2, magnetic particle concentrator	Dynal AS, Oslo, Norway
DYNAL® Bead Separations	Invitrogen Ltd, Paisley, UK
E-Gel® iBase™	Invitrogen Ltd, Paisley, UK
GeneAmp® PCR System 2700	Applied Biosystems, Foster City, USA
GeneAmp® PCR System 9700	Applied Biosystems, Foster City, USA
GS Junior	Roche Diagnostics GmbH, Mannheim, Germany
GS Junior emPCR Bead Counter	Roche Diagnostics GmbH, Mannheim, Germany
MICROLAB® STARlet	Hamilton Robotics AB, Kista, Sweden
MixMate, PCR 96	Eppendorf AG, Hamburg, Germany
NanoDrop 8000 UV-Vis Spectrophotometer	Thermo Scientific, Wilmington DE, USA
NanoDrop® ND-1000 Spectrophotometer	Thermo Scientific, Wilmington DE, USA
Rotator SB2	Stuart®, Bibby Scientific Ltd, Staffordshire, UK
ULTRA-TURRAX® Tube Drive control	IKA®-Werke GmbH & Co. KG, Staufen, Germany
VWR™ Galaxy Mini	VWR™ International

2.2 Consumables

E-Gel® Size Selected™ 2% Agarose, G6610-02	Invitrogen Ltd, Paisley, UK
MicroAmp™ Clear Adhesive Film, 4306311	Applied Biosystems, Foster City, USA
Special Lens Cleaning Tissue, 1019	Glaswarenfabrik Karl Hecht GmbH & Co KG – „Assistant", Sondheim, Germany
Thermo-Fast® 96, Non-Skirted, Abgene® PCR Plates	Thermo Scientific, Waltham, USA
Tissue, Zeiss Moistened Tissue, HC	Roche Diagnostics GmbH, Mannheim, Germany

2.3 Kits, Buffer, Solutions

310 Running Buffer, 10X, 402824	Applied Biosystems, Foster City, USA
3130 POP-7™ Performance Optimized Polymer, 4363785	Applied Biosystems, Foster City, USA
AccuPrime™ GC-Rich DNA Polymerase, 12337-016	Invitrogen Ltd, Paisley, UK
Agencourt® AMPure® XP beads, A63880	Beckman Coulter Inc., Brea, USA
Agilent DNA 7500 Reagents, 5067-1506	Agilent Technologies, Inc., Santa Clara, CA, USA
BigDye® Terminator v1.1/3.1, 5x Sequencing Buffer, 4339843	Applied Biosystems, Foster City, USA
BigDye® Terminator v3.1 Cycle Sequencing RR-100, 4336911	Applied Biosystems, Foster City, USA
BigDye® XTerminator™ Purification Kit, 4376487	Applied Biosystems, Foster City, USA
DNA Molecular Weight Marker VIII (19- 1114bp), 11 336 045 001	Roche Applied Science, Mannheim, Germany
Ecotainer® Aqua B. Braun	B. Braun, Melsungen, Germany
emPCR Kit – Bead Recovery Reagents, 05 996 490 001	Roche Applied Science, Mannheim, Germany
emPCR Kit – emPCR Reagents (Lib-A), 05 996 538 001	Roche Applied Science, Mannheim, Germany
emPCR Kit – Oil and Breakage Kit, 05 996 511 001	Roche Applied Science, Mannheim, Germany

illustra™ ExoStar™ 1-Step, US77720	GE Healthcare Ltd, Buckinghamshire, UK
SensiMix™ HRM Kit, QT805-05	Bioline Reagents Ltd, London, UK
Sequencing Kit – Packing Beads and Supplement CB, 05 996 597 001	Roche Applied Science, Mannheim, Germany
Sequencing Kit – PicoTiter Plate Kit, 05 996 619 001	Roche Applied Science, Mannheim, Germany
Sequencing Kit – Sequencing Buffers, 05 996 490 001	Roche Applied Science, Mannheim, Germany

2.4 Computer Programs

Amplicon Variant Analyzer v2.7	Roche Applied Science, Mannheim, Germany
GS Run Browser	Roche Applied Science, Mannheim, Germany
SeqScape v2.5	Applied Biosystems, Foster City, USA

2.5 Primer

Two categories of fusion primers were used in this master thesis (appendix 6.2): (1) Template-specific primer pairs for amplification of the coding and splice-site regions of *MSH2*, *MSH6*, *MLH1*, and *PMS2*. (2) Fusion primers for attachment of MID barcode, key sequence and 454 sequencing system primers; hereafter these primers are called MID primers. All primers were ordered from Eurogentec in Oslo, Norway.

All target specific primer pairs were designed by staff of the Department of Pathology and Medical Genetics. The laboratory performs genetic tests for the four MMR genes based on Sanger sequencing routinely. In the majority of cases, the same MMR gene specific primer sequences were used, as they were already tested and approved. Only in cases where the PCR products were not compatible with the new sequencing method new primers were designed. Such exclusion criteria were product size and formation of primer dimers which is facilitated by the attachment of universal tails. Attention was paid to the specifications from Roche. These state that the final PCR product length, i.e. amplicon length including universal tail, MID, key and 454 sequencing sequence, has to be between 200bp and 600bp and that the difference between the shortest and longest amplicon should be less than 150bp. However, the actual length of the amplicons analyzed in this study is between 283bp and 489bp, thus the difference between the amplicons is 206bp.

As universal tails, Univ-A and Univ-B, universal M13-tails (forward primer: cacgacgttgtaaaacgac and reverse primer: caggaaacagctatgacc) were used; as described by De Leeneer *et. al.* 2010[46]. The MID primer pairs were designed according to the Roche Guidelines for Amplicon Experimental Design, Universal Tailed Amplicon Sequencing[60].

2.6 Patient Material

gDNA isolated from blood samples was used to establish a sequencing protocol for the GS Junior sequencing system. The samples can be divided into two categories. For optimization of sequencing of *PMS2* DNA samples from 119 patients with a suspected mutation in *PMS2* were sequenced. For establishment of the MMR genes library preparation 32 samples from a research biobank were analyzed. 16 of these samples were already sequenced by Sanger sequencing. The biobank was founded in the context of the project "Colorectal cancer in Central-Norway; Identification of hereditary and non-hereditary subtypes". Biological material (blood, fresh frozen tumor- and normal tissue) from 362 patients with histopathologically confirmed cancers was collected in the period from January 2007 to June 2008[19,20]. From all patients in the study an informed consent was obtained. Various tests were already performed with the tumor samples, including: (1) mutations in BRAF exon 11 and 15, (2) methylation of the MLH1 promoter region, and (3) microsatellite instability (MSI).

gDNA was isolated from frozen ethylenediamine tetraacetic acid (EDTA) blood samples by the iPrep™ Pure Link™ gDNA Blood Kit from Invitrogen. This DNA isolation kit uses magnetic Dynabeads® MyOne™ SILANE for purification of gDNA from human blood[67]. Isolation was done according to the Invitrogen user manual[67]. DNA concentration and quality was measured by a NanoDrop® Spectrophotometer[68,69]. In order to facilitate the workflow and minimize variations in the singleplex PCRs and whereby obtaining as steady and comparable results as possible, all samples were diluted with molecular grade water to achieve the same DNA concentration.

2.7 Workflow of Sequencing of MMR Genes

Some of the amplicons contain homopolymer runs (HPs) longer than 8bp, which cannot be sequenced precisely by pyrosequencing. Therefore, both the GS Junior System and the Sanger method have to be used for sequencing of the MMR genes. *MLH1* exon 7 and 3' UTR are Sanger sequenced because of difficulties in optimization of singleplex PCR conditions. Figure 2.1 shows the MMR genes sequencing workflow. For SGS sequencing analysis of the MMR genes the GS Junior Titanium Series and bi-directional fusion primers were used. After data analysis the significance of the variants was evaluated by classifying them into five categories, where 1 is not pathogenic or of no clinical significance and 5 is definitely pathogenic[70]. A detailed description of the classification system is given in appendix 6.3. The results of the biobank samples were evaluated in regard to the aim of the CRC project, which is identification of hereditary and non-hereditary subtypes.

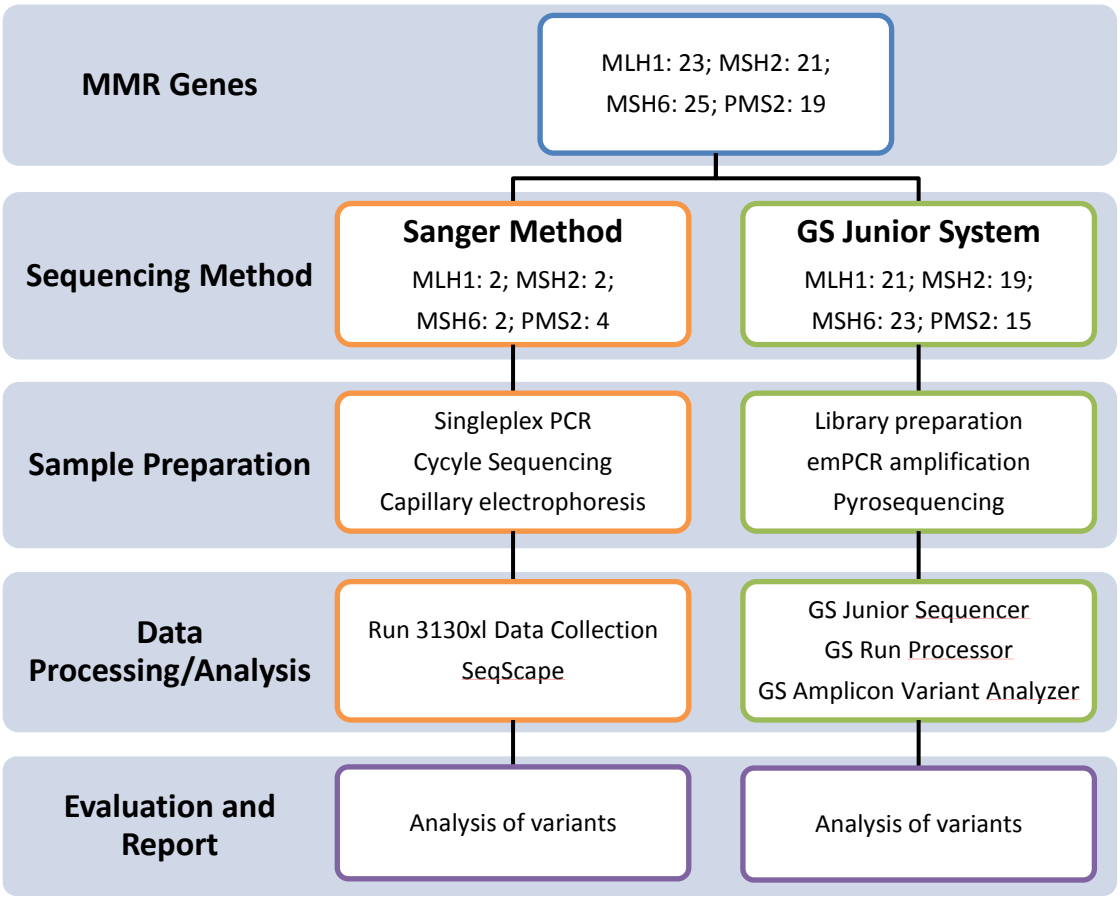

Figure 2.1. Workflow for sequencing of MMR genes. This overview indicates the number of amplicons sequenced by both methods.

2.8 DNA Sequencing using the GS Junior System

2.8.1 Amplicon Library Preparation

The library preparation is a crucial step in the sequencing protocol and has to be optimized thoroughly. This applies particularly to the singleplex PCRs and multiplexing step. For optimization of sequencing of *PMS2* and the MMR genes (*MSH2*, *MSH6*, *MLH1*, and *PMS2*) three and four GS Junior sequencing runs, respectively were performed. The number of samples pooled per amplicon library depends on the number of genes analyzed in parallel. When sequencing solely *PMS2*, 15 amplicons were sequenced per patient and 40 samples were pooled. When analyzing the four MMR genes, 78 amplicons were sequenced per patient and 8 samples were pooled. Optimization of MMR genes library preparation was based on the results of optimization the PMS2 library and a master thesis, in which sequencing of the MMR genes *MSH2*, *MSH6*, and *MLH1* was optimized[71].

Singleplex PCR

The coding and splice-site regions of the MMR genes were amplified by touchdown PCR. In touchdown PCR the initial annealing temperature, which is above the melting temperature of the primers, is gradually decreased during the first half of the PCR. During the second half this low annealing temperature is maintained. High annealing temperatures result in increased specificity of the PCR reaction and in this way the desired target sequence is amplified exclusively. Whereas, the lower annealing temperatures provide increased amplification efficiency and thereby high amounts of this desired product can be obtained. For the singleplex PCRs SensiMix™ HRM Kit from Bioline was used[72]. The exact volumes for one reaction mixture which were used for optimization of the PMS2 library preparation are given in Table 2.1. [73]

Table 2.1. Singleplex PCR reaction mixture for optimization of PMS2 library preparation.

Reagent	Concentration Stock Solution	Volume (for n = 1) [µL]
SensiMix HRM	2x	12.5
Molecular Grade Water		8.9
$MgCl_2$	50 mM	0.7
Sense Primer	10 µM	0.7
Antisense Primer	10 µM	0.7
Volume for Distribution		23.5
Sample (DNA Template)	20 ng	1.5
Total Volume		25.0

For optimization of the MMR genes library the pipetting robot, MICROLAB® STARlet, was used. To meet the specifications of the robot the PCR reaction mixture had to be adjusted (Table 2.2). The robot prepares a master mix containing SensiMix HRM, water, and magnesium chloride ($MgCl_2$) and adds the primer mixture and DNA.

Table 2.2. Singleplex PCR reaction mixture for optimization of MMR genes library preparation.

Reagent	Concentration Stock Solution	Volume (for n = 1) [µL]
SensiMix HRM	2x	12.5
Molecular Grade Water		1.8
$MgCl_2$	50 mM	0.7
Volume for Distribution		15.0
Sense/antisense Primer	1.4 µM per primer	5.0
Sample (DNA Template)	6 ng	5.0
Total Volume		25.0

For amplification of the 88 fragments three touchdown PCR programs were used (Table 2.3). During the temperature gradient the annealing temperature was reduced by 0.5 °C every cycle. The PCR products were examined either on E-Gel SizeSelected Agarose Gels (2% or 4%) from Invitrogen or Bioanalyzer DNA 7500 Chips. This was done to verify the size and quality of the PCR product.

Table 2.3. Touchdown PCR programs. Annealing temperature range **(A)** 61 to 53 °C, **(B)** 64 to 56 °C, and **(C)** 66 to 62 °C.

A

Temp [°C]	Time [min]	Cycles
95	5	
95	0.5	16
61-53	1	
68	1	
95	0.5	24
53	1	
68	1	
68	5	
6	∞	

B

Temp [°C]	Time [min]	Cycles
95	5	
95	0.5	16
64-56	1	
68	1	
95	0.5	24
56	1	
68	1	
68	5	
6	∞	

C

Temp [°C]	Time [min]	Cycles
95	5	
95	0.5	8
66-62	1	
68	1	
95	0.5	32
62	1	
68	1	
68	5	
6	∞	

As the annealing temperature depends on the specific primer, the best suitable temperature range had to be determined experimentally (appendix 6.2.). To minimize the number of PCR programs only four annealing temperature ranges were tested: 61 – 53 °C, 64 – 56 °C, 66 – 58 °C, and 66 – 62 °C. Furthermore, various $MgCl_2$ concentrations (1 mM and 1.4 mM) and primer concentrations (0.28 µM, 0.2 µM, and 0.12 µM) were tested.

Multiplex PCR

For multiplexing different amounts of each singleplex PCR were mixed. The optimal volumes were determined experimentally (appendix 6.4). After multiplexing, each pool was diluted 1:100 with molecular grade water and 1 µL of this dilution was used per multiplex PCR. For the multiplex PCR AccuPrime™ GC-Rich DNA Polymerase from Invitrogen was used (Table 2.4)[74]. For preparation of the PMS2 and MMR genes library two and eight multiplex PCRs per sample were performed, respectively. All amplicons of one patient sample were labeled with the same MID barcode sequence, which was unique for this patient in the GS Junior run. Thereby, each patient was identified by the amplicon variant analyzer software. The multiplex PCR program is shown in Table 2.5.

Table 2.4. Reaction mixture of multiplex PCR using MID-primer pairs and AccuPrime™ GC-Rich DNA Polymerase from Invitrogen.

Reagent	Concentration Stock Solution	Volume (for n = 1) [µL]
Molecular grade water		17.1
AccuPrime™ GC-Rich Buffer A	5x	5.0
AccuPrime™ GC-Rich Polymerase	2.0 U/µL	0.5
Volume for Distribution		22.6
Sense MID-primer	10 µM	0.7
Antisense MID-primer	10 µM	0.7
1:100 diluted singleplex PCR mixture		1.0
Total Volume		25.0

Table 2.5. Multiplex PCR program. min, minutes; ∞, infinity.

Temperature [°C]	Time [min]	Cycles
95	5.0	
95	0.5	20
58	0.5	
72	1	
72	5	
6	∞	

To achieve the main goal, a uniform distribution of coverage, the optimization of the multiplexing step is of great importance. Three features had to be optimized: (1) amount of singleplex PCR product of each fragment, (2) composition of pools, and (3) optimal procedure for multiplexing. The amounts taken of each fragment were adjusted after each GS Junior run according to the coverage results obtained.

Purification and Amplicon Pooling

After multiplex PCR all reactions of one sample were mixed and purified using Agencourt® AMPure® XP - PCR Purification by Beckman Coulter. The purification step is required to remove salts, enzymes, unincorporated nucleotides and primers before emPCR amplification[75]. Everything was done according to Roche Amplicon Library Preparation Method Manual; 3.2.2 Library Purification for tubes[76].

The next step was the quality and quantity control. Agilent DNA 7500 chips were used to control the size and general distribution of the fragments in the library and the presence of small, unspecific fragments. Determination of DNA concentration by the NanoDrop spectrophotometer is necessary to ensure that the same amount of each sample is added to the amplicon pool. The calculation was done in two steps according to Roche Amplicon Library Preparation Method Manual (Figure 2.2)[76]. Afterwards, 10 µL of each of the diluted amplicon samples were mixed to prepare the amplicon pool. The pool was further diluted to 10^5 molecules/µL by preparing two consecutive 1:100 dilutions (mixing 10 µL of the amplicon pool with 990 µL molecular grade water).

A

$$Molecules/\mu L = \frac{Sample\ conc.\ [ng/\mu L] \times 6.022 \times 10^{23}}{656.6 \times 10^{9} \times average\ amplicon\ length\ [bp]}$$

B

$$Volume\ of\ TE\ [\mu L] = \frac{Molecules/\mu L}{10^{9}} - 1$$

Figure 2.2. Calculations for amplicon dilution. (A) Calculation of amplicon concentration in molecules/µL. Average amplicon length includes Univ-A/Univ-B, MID, key and universal tail A/B sequences. Average amplicon length PMS2 library: 385bp, MMR genes library: 371bp. **(B)** Calculation of amplicon dilution to achieve an end concentration of 1 x 10^9 molecules/µL.

For the three PMS2 GS Junior runs an amplicon pool containing 10^6 molecules/µL was prepared by adding 1 µL of the undiluted amplicon pool (10^9 molecules/µL) to 999 µL molecular grade water. However, in this way minor inaccuracies in pipetting can have significant impacts on the performance of the subsequent emPCR amplification. For that reason and to increase the volume of the amplicon pool used in the emPCR the amplicon pool is diluted to 10^5 molecules/µL.

2.8.2 emPCR Amplification Method – Lib-A

25 µL of the diluted amplicon library (10^5 molecules/µL) were added to Capture Beads A and B. This volume was determined experimentally. The emPCR amplification and bead recovery were done according to the Roche emPCR Amplification Method Manual – Lib-A[77]. GS Junior Titanium emPCR Kit emPCR Reagents (Lib-A), emPCR Kit Oil and Breaking Kit, Bead Recovery Reagents were used.

The collection of the enriched DNA beads, which is part of the DNA Library Bead Enrichment step, had to be slightly modified in order to obtain higher numbers of high quality beads. In this collection step the enriched DNA beads are separated from the paramagnetic enrichment beads by a sodium hydroxide (NaOH) treatment. Subsequently, the paramagnetic beads are removed with the aid of the Magnetic Particle Concentrator (MPC). Everything was done according to the Roche manual, expect of an additional separation step to remove the remaining paramagnetic beads from the enriched DNA beads: After transfer of the supernatant containing the enriched DNA beads into a tube, this tube was placed again into the MPC for about one minute and the supernatant transferred into a new tube. This supernatant was used for the sequencing procedure.

After bead enrichment, the amount of enriched beads was determined by the GS Junior Bead Counter. According to the manual, bead preparation has failed if more than 2 million enriched beads are counted and the whole emulsion process has to be repeated with a reduced amount of the amplicon library[77]. When between 500 000 and 2 000 000 beads were counted, 500 000 beads were used for pyrosequencing.

Practical experience showed that the recommended molecule-to-bead ratio of 2 is too high and that bead counts higher than 2 million are obtained. The actual number of molecules per bead can be calculated by an equation given in the Roche emPCR Amplification Method Manual – Lib-A[77] (Figure 2.3).

$$Molecules\ per\ bead = \frac{\mu L\ of\ amplicon\ library \times library\ concentration\ \ [molecules/\mu L]}{5\ million\ beads}$$

Figure 2.3. Calculation of molecule to bead ratio used in emPCR amplification. Calculation adapted from Roche emPCR Amplification Method Manual – Lib-A[77]. The volume of amplicon library (µL) refers to the amount per Capture Bead tube (Capture Beads A and B).

2.8.3 Pyrosequencing

The pyrosequencing was done according to the Roche Sequencing Method Manual[78]. The GS Junior Sequencing Kit Packing Beads and Supplement CB, PicoTiter Plate Kit, and Sequencing Buffers were used.

2.8.4 Analysis of Sequencing Data

The data obtained in the seven GS Junior runs were analyzed by the GS Junior Amplicon Variant Analyzer (AVA) software version 2.7. The variant/consensus parameters of the computation analysis are given in Table 2.6. Although, a variation of about 50% is expected for heterozygous variant, during the optimization process a minimum filter value (max. combined %) of 10% was chosen in the AVA software. The variant table generated by the AVA software was exported to excel and evaluated manually. This table includes the variant frequency (%) of the forward, reverse, and combined reads. Cross-sample comparison was applied to determine whether a variant is a true variant. New and/or potentially pathogenic variants were verified by Sanger sequencing.

Table 2.6. Variant/consensus parameters used for sequencing data analysis in AVA software.

Depth thresholds	Minimum read count 1 Minimum read percentage 0%
Directional support	Any
N-mer thresholding	Fixed (base calls)
Computation speed	CPUs 1

2.8.5 Statistical Analysis

The degree of uniformity of the coverage distribution of the GS Junior sequencing runs was evaluated by means of the spread correction factor. The spread correction factor can be determined in two steps Figure 2.4: (1) Calculation of fold difference to average coverage, and (2) plotting of the sample amplicon combination (SAC) in%, i.e. the fraction of amplicon in %, against the fold difference to average coverage. Spread correction factor F_{90} can be determined by setting the fraction of amplicons to a threshold of 90% and reading the value on the x-axis at which the histogram passes this threshold. For sake of accuracy and convenience, the Supplemental Tool S2 from De Leeneer *et al.*, 2011 was used for calculation of the spread correction factor[79].

A

$$Fold\ difference\ to\ average\ coverage\ = \frac{Average\ coverage}{Amplicon\ coverage}$$

B

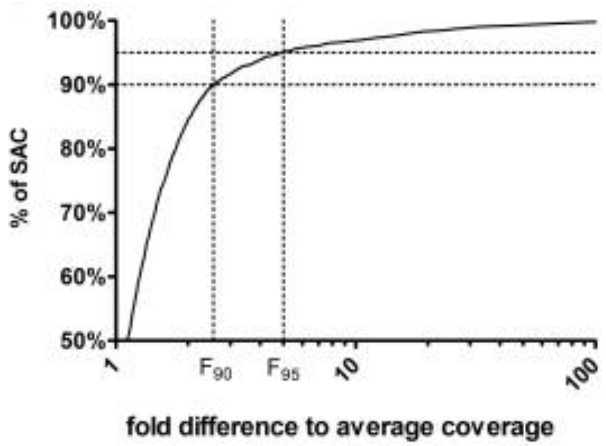

Figure 2.4. Determination of spread correction factor (F_{90} and F_{95}). (A) Calculation of the fold difference to average coverage. **(B)** Plotting of the sample amplicon combination (SAC) in % against the fold difference to average coverage. The spread correction factor is defined as value on the x-axis at which the histogram passes a given threshold. As indicated in the diagram, for determination of the spread correction factor F_{90} and F_{95} this threshold is set to 90% and 95%, respectively.

The lower the spread correction factor, the more uniform the distribution of coverage and the more samples can be screened in a run, while ensuring a sufficient coverage for detection of heterozygous variants. For calculation of capacity the Supplemental Tool S2 from De Leeneer *et al.* was used[79]. The settings were chosen according to the results of a study performed by De Leeneer *et al.* (Table 2.7)[46]. They showed that a coverage of 38 is required to detect a heterozygous variant with a probability of 99.9%, taking only variants with a frequency higher than 25% into account.

Table 2.7. Parameters used for calculation of capacity.

Parameter	Value
Required power to detect variants	99.900%
Threshold for sequence error filtering	25%
Required minimum coverage per SAC (MC)	38
Instrument read count Specifications	Number of mapped reads
Read count correction	100%
Number of amplicons per sample	PMS2 library: 15 MMR genes library: 78

2.9 DNA Sequencing using the Dideoxy Method

Sanger sequencing was performed for the following reasons: (1) As standard sequencing method for *MSH2* exons 2 and 5, *MSH6* exons 7 and 10, *MLH1* exon 7 and 3'UTR, and *PMS2* exons 7 due to HPs and for *PMS2* exons 13, 14, and 15 due to similarities of these exons with *PMS2 CL*. (2) For verification potential pathogenic variants. (3) For analysis of individual amplicons which were not evaluable by the GS Junior system, due to a too low coverage. The exons were amplified by singleplex PCRs using SensiMix™ HRM Kit and touchdown PCR (section 2.8.1).

2.9.1 Purification of PCR Products using Illustra™ ExoStar

PCR products were purified by Illustra™ ExoStar to prevent interference of the cycle sequencing reaction. ExoStar contains two enzymes. Exonuclease I degrades unincorporated primers and single-stranded DNA[80]. Alkaline Phosphatase dephosphorylates unconsumed dNTPs and deoxyribonucleoside 5' monophosphates which are released during the Exonuclease I reaction[80]. Purification was done according to the ExoStar protocol, except 2.5 µL PCR product and 1µL ExoStar were used[81].

2.9.2 Sequencing using BigDye v3.1

Cycle sequencing was performed using BigDye® Terminator v3.1 Cycle Sequencing Kit from Applied Biosystems[82]. The reaction mixture is given in Table 2.8. For each PCR product two cycle sequencing reactions were performed. For this universal sense and antisense sequencing primers targeting the universal tails Univ-A and Univ-B were used. For seven amplicons intrinsic sequencing primer had to be used: (1) due to HPs near the splicing sites (*PMS2* e7s; *MSH2* e2s, e5a; *MSH6* e7s, e10s) and (2) to reduce the sequencing product (*PMS2* e13a, e15a). The sequencing program is given in Table 2.9.

Table 2.8. Cycle sequencing reaction mixture using BigDye® Terminator v3.1 Cycle Sequencing Kit.

Reagent	Volume (for n =1) [µL]
Molecular Grade Water	5.3
5x BigDye® Terminator v1.1/3.1 Sequencing Buffer	2.0
Big Dye® Terminator v3.1 Cycle Sequencing RR-100	1.0
Primer	0.2
Volume for Distribution	8.5
PCR Product Purified with ExoStar	1.5
Total Volume	10.0

Table 2.9. Cycle sequencing program. sec, seconds; ∞, infinity.

Temperature [°C]	Time [sec]	Cycles
96	60	
96	10	26
55	5	
60	150	
6	∞	

2.9.3 Purification XT - BigDye® XTerminator™ Purification Kit

The cycle sequencing products were purified using BigDye® XTerminator™ Purification Kit. The kit contains XTerminator Solution to eliminate unincorporated dye terminators and free salts and SAM™ Solution to enhance BigDye XTerminator reagent performance and to stabilize the samples after purification. The clean-up was done according to the BigDye® XTerminator™ Purification Kit protocol. [83]

2.9.4 Sequencing using 3130xl Genetic Analyzer

Capillary electrophoresis was performed on 3130xl Genetic Analyzer using the following instrument settings: Method – 3130pop7_BDTv3_KB_BDXT; Analysis Instrument Protocol 1 – RapidSeq36_pop7_Z_V3BDTX_3sek. Sequencing results were evaluated with SeqScape v2.5.

2.10 Examination of PCR Products

PCR products were examined on both agarose gels and Agilent DNA 7500 chips. The E-Gels® from Invitrogen are bufferless agarose gels equipped with electrodes integrated in the gel matrix and ethidium bromide to visualize the DNA[84]. Everything was done according to the E-Gel® technical guide[84]. 2% agarose gels were used for the control of PCR products properties such as product size and presence of unspecific PCR products.

The Agilent 2100 Bioanalyzer which is a microfluidics-based platform was used for examination of singleplex and multiplex PCRs. Using Agilent DNA 7500 Kit DNA fragments between 100 and 7500bp can be separate according to size and quantified. Advantages of the Bioanalyzer compared to E-Gels® are automated quantification and increased sensitivity, whereby allowing the detection of low concentrations of unspecific PCR products and primer dimers. Everything was done according to the Agilent DNA 7500 and DNA 12000 Kit Quick Start Guide[85].

3 Results

In this study, seven GS Junior sequencing runs were performed to optimize the preparation procedure of two amplicon libraries for sequencing *PMS2* and the four MMR genes *MSH2, MSH6*, *MLH1*, and *PMS2* in parallel. The progress in optimization was evaluated by means of the spread correction factor, which is a measure for uniformity of coverage distribution in a sequencing run. In addition, differences in performance of the GS Junior sequencing and Sanger sequencing method were examined in terms of variant detection, running costs, and time consumption. The results obtained during optimization of the MMR genes library preparation were part of a pilot study, which aimed to evaluate the suitability of tumor screening analysis for identification of LS patients.

3.1 Optimization of Preparation of PMS2 Library

Within this master thesis project three GS Junior runs were performed to optimize the amplicon library preparation procedure for sequencing the MMR gene *PMS2*. Two factors were optimized in the PMS2 library preparation: (1) Conditions of singleplex PCRs and (2) volumes of singleplex PCR products used in multiplex PCRs. This was done to obtain a uniform distribution of coverage of all amplicons included in the GS Junior run and to exploit the capacity of the sequencing instrument at the best possible rate.

3.1.1 Optimization of Singleplex PCR Conditions

The quality of the singleplex PCR products is of great importance for the success of a GS Junior sequencing run. The optimal annealing temperature range of each of the fusion primer pairs was determined experimentally. During the optimization process two problems had to be solved. These were the presence of small unspecific fragments in some of the singleplex PCR reactions and the co-amplification of a PMS2 pseudogene.

Elimination of Primer Dimers

After each GS Junior run the number of reads that passed the image and signal processing steps of the GS Run Processor software, i.e. high quality reads, were recorded. In addition, the number of reads aligned to the reference sequence was determined by calculating the sum of the coverage of all amplicons. The passed and mapped reads of the four GS Junior runs can be found in Table 3.1. Comparison of these two values showed

that a considerable percentage of high quality reads could not be aligned by the AVA software in each of the three PMS2 GS Junior runs (46.5%, 30.5%, and 20.1%, resp.). In order to exploit the full capacity of the GS Junior sequencing instrument, the cause of this deviation had to be determined and resolved.

In each of the three runs small fragments, which are about 100bp smaller than the shortest amplicon, were detected by both the Agilent 2100 Bioanalyzer and the GS Run Browser (Figure 3.1). These small fragments represent either unspecific PCR products or primer dimers that are formed during singleplex or multiplex PCR reaction. In both cases the AVA software would not be able to align their sequences to the reference sequence.

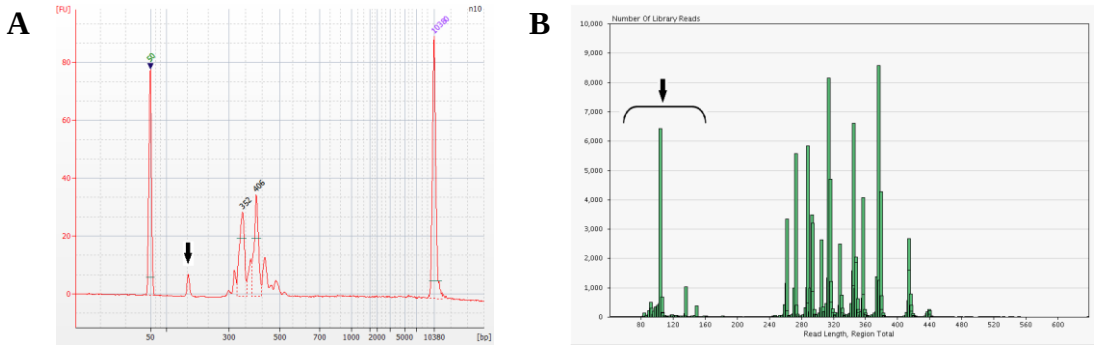

Figure 3.1. Illustration of fragment length of PMS2 amplicon pool. The small, unspecific fragments are indicated by black arrows **(A)** Exemplary amplicon length profile obtained by the Agilent 2100 Bioanalyzer. The electropherogram shows the amplicon pool of a 10 times diluted sample after multiplex PCR and Agencourt® AMPure® XP Purification. The Fragment size (bp) is given on the x-axis and the fluorescence is given on the y-axis, respectively. The analysis range of the Bioanalyzer DNA 7500 Chip is indicated by a lower marker (50bp) and an upper marker (10 380bp). **(B)** Amplicon length profile of GS Junior PMS2 run 3 obtained by the GS Run Browser software. This profile represents the total amplicon pool of the 39 patient samples included in the DNA library. The read length is given on the x-axis and the number of sequencing reads, i.e. the coverage, is given on the y-axis, respectively.

In order to find the cause and origin of the small fragments both singleplex and multiplex PCR products were analyzed on Bioanalyzer DNA 7500 Chips. In this way, the singleplex PCR products containing the small fragments were identified. It was observed that the length of these fragments was similar in all affected singleplex PCRs and that their length increased in the multiplex PCR amplification. The length detected in the singleplex and multiplex PCR products was about 70-90bp and 150bp, respectively. Comparison of the fragment length with the length of the fusion primer sets used in singleplex and multiplex PCR indicates that these short fragments are primer dimers. The fusion primers

used in the singleplex and multiplex PCRs are 35-45bp and 53-54bp, respectively. This second set of fusion primers contains sequences complementary to the Univ-A and Univ-B of the singleplex fusion primers. Without this universal tail sequence the primers have a length of 35bp. In the event, that the short fragments are formed during the singleplex PCR and extended during the multiplex PCR, i.e. that they can serve as a template for the second set of fusion primers, the fragments have to be 70bp longer after the multiplex PCR. This could be confirmed by the observations. Thus, in order to eliminate these smaller fragments the reaction conditions of the affected singleplex PCRs have to be optimized. The prevention of their formation is particularly important as PCR products larger than 100bp cannot be removed by the Agencourt AMPure XP Purification Kit[75].

In the optimization process reduced $MgCl_2$ concentrations, various primer concentrations, and increased annealing temperature ranges were tested. Finally, uniform $MgCl_2$ concentration (1.4 mM) and primer concentration (280 nM) were chosen for all singleplex PCR reactions. Three different annealing temperature ranges were selected (61-53 °C, 64-56 °C, and 66-62 °C). This facilitates the working routine, as all amplicons can be amplified efficiently with a minimal number of PCR programs.

Co-Amplification of PMS2 Pseudogenes

In the first PMS2 GS Junior sequencing run various variants, present in all samples with a frequency higher than 10%, were detected in fragment 11B Figure 3.2A. When comparing the sequence of *PMS2* exon 11 with the sequence of *PMS2 CL* it was observed that some variants are identical to the sequence differences of these two genes. This indicated that the primer pairs used for amplification of exon 11 are not specific enough to avoid co-amplification of the pseudogene *PMS2 CL*. To prevent co-amplification of *PMS2 CL* new primer pairs for fragment 11A and 11B were designed; taking the minor differences of these two genes better into consideration. It was not necessary to order new primers for fragment 11C and 11D. The new amplicon 11A antisense and the new amplicon 11B sense primer are directly located on the homopolymer repeat of eight adenosines that is present in *PMS2* but not in *PMS2 CL*. To illustrate the improvement the sequencing result using the new primer pair for fragment 11B are shown in Figure 3.2B as an example. The PCR conditions were optimized and the primer pair was successfully tested in the second PMS2 GS Junior run. By using this new primer pair it is possible to amplify fragment 11B of *PMS2* exclusively.

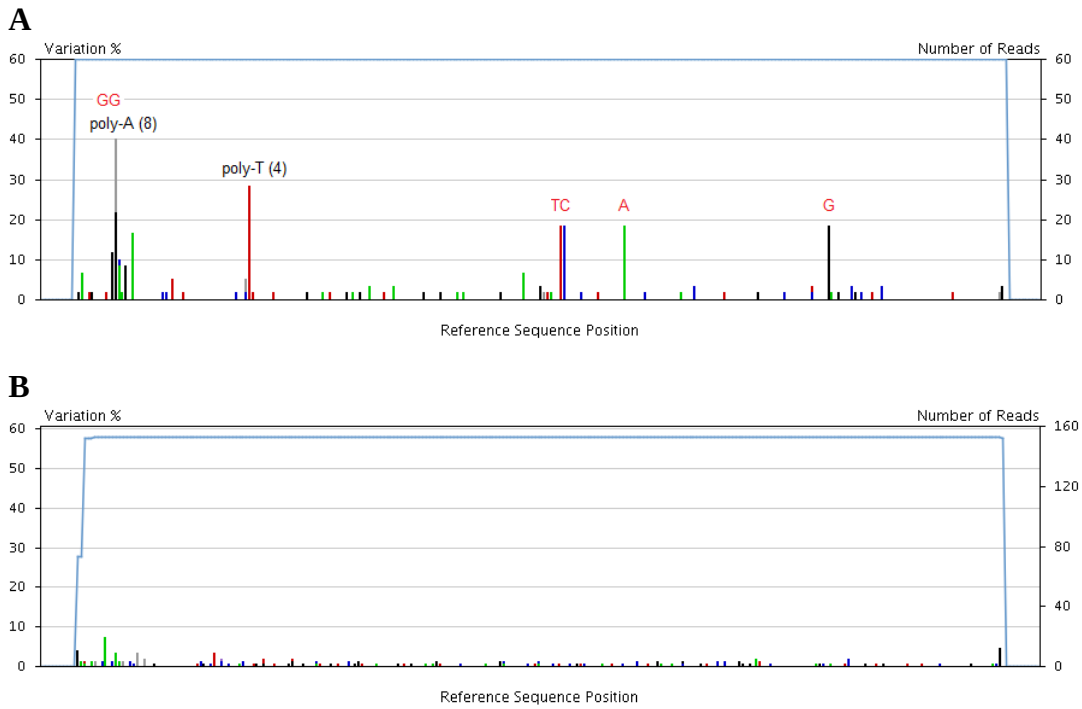

Figure 3.2. Comparison of GS Junior sequencing results of amplicon PMS2_11B before and after optimization of primer sequences and PCR conditions. Both figures show a variant frequency plot of the AVA software. **(A)** The plot shows six variants, marked in red, that are identical to regions of the PMS2 C-terminal like pseudogene (*PMS2 CL)*. The homopolymeric regions which are identical in PMS2 and *PMS2 CL* are marked in black. **(B)** The second plot shows the sequencing results with the newly designed primer pair. The new sense primer is directly located on the homopolymer stretch of eight As. Therefore, no incorrect base calls are detected for this homopolymer in the second run. The improved basecalling of the second homopolymer stretch of four Ts is most probably due to improved singleplex PCR conditions and optimized multiplexing. This primer pairs allows exclusive amplification of *PMS2* by exploiting the minor differences of the two genes.

3.1.2 Evaluation of Parameters of GS Junior Sequencing Runs

The coverage of the amplicons was determined after each run. Based on this data the volume of each amplicon was adjusted in the following multiplexing. This was done to obtain a uniform distribution of coverage, which is required to exploit the full capacity of the GS Junior instrument. This means, the number of reads obtained from the "less efficient" and "best performing" amplicons should be as similar as possible[46]. Evaluation of uniformity in coverage distribution was based on the spread correction factor. The lower the spread correction factor the more similar the number of reads across amplicons of the GS Junior run. Table 3.1 summarizes the characteristics of the three PMS2 GS Junior runs. Raw data of coverage are given in appendix 6.5. The sample capacity was calculated based on the number of mapped reads obtained in the particular GS Junior run.

Table 3.1. Overview of characteristic parameters of the three PMS2 GS Junior runs.

	Run 1	Run 2	Run 3
Samples	40	40	39
Amplicons	600	600	585
Passed reads	96 219	86 753	133 007
Mapped reads	51 508 (53.5%)	60 322 (69.5%)	106 301 (79.9%)
Min/avg/max coverage	0/86/467	0/100/510	0/182/772
Coverage SD	91	72	117
Variation coefficient	1.06	0.72	0.64
Spread corr. Factor 90%/95%	10.73/21.46	4.57/6.28	2.63/5.86
Sample capacity F90/F95	8/4	23/15	70/31
Amplicons with no coverage	79	1	40
Amplicons <38	252	111	71

When comparing the results of the three PMS2 GS Junior runs, it can be seen that a quality improvement was achieved after each optimization step. For instance the percentage of mapped reads is increasing and the spread correction factor is decreasing. Based on a study of De Leeneer *et al.* a minimum threshold of coverage of 38 was chosen[46]. With each optimization step the number of amplicons below this threshold is decreasing. The number of amplicons below a 38-fold coverage also contains the amplicons with no coverage. Sample capacity F90 and F95 provide information about how many samples can be pooled in one amplicon library, while taking into consideration that the percentage of amplicons not meeting the minimum threshold of coverage of 38 are maximal 10% and 5%, respectively. Based on the best spread correction factor (F90 2.63; F95 5.86), obtained in the third GS Junior run, 70 and 31 samples, respectively can be pooled and analyzed per GS Junior run.

Figure 3.3 presents the distribution of coverage of the three PMS2 runs graphically. The figure shows that a higher coverage of each amplicon was achieved after optimizing the singleplex PCR conditions and adjusting the volumes of singleplex PCR products used for multiplexing. Furthermore, it can be seen that distribution of coverage across the amplicons is still varying after two optimization steps and has to be further optimized.

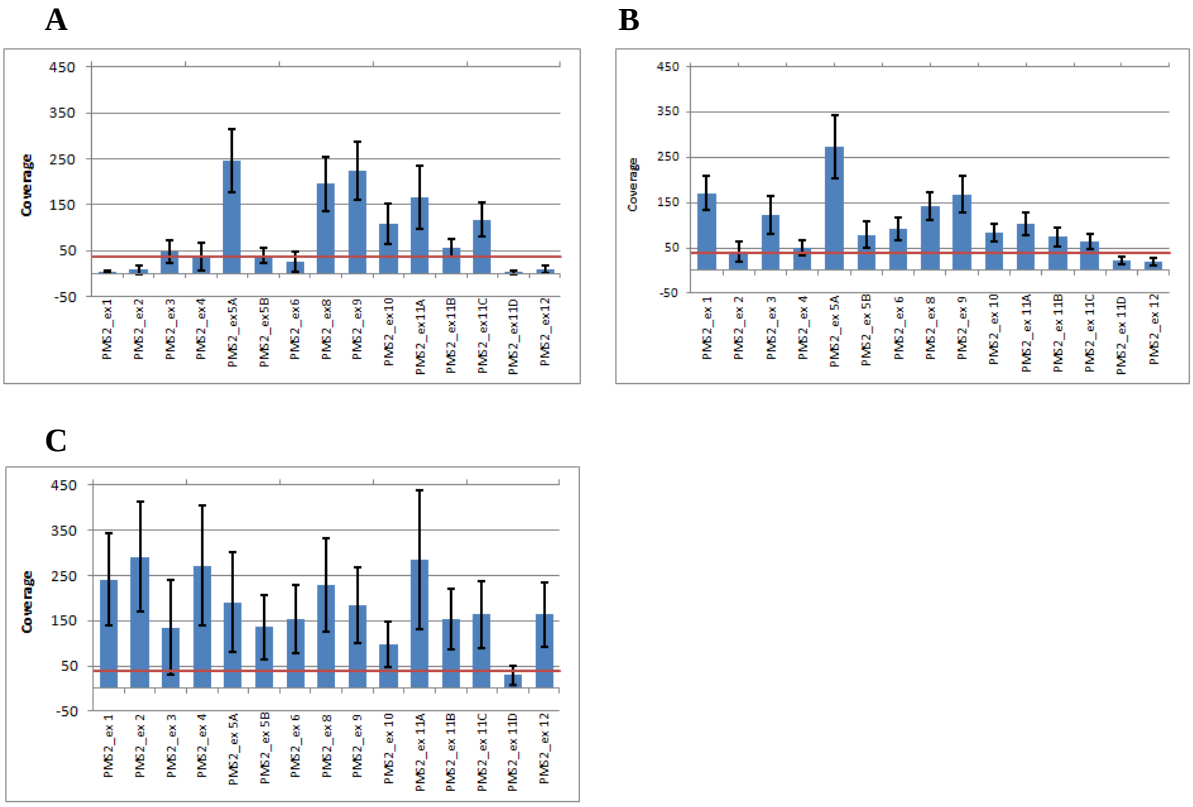

Figure 3.3. Distribution of coverage of the three PMS2 GS Junior sequencing runs. (A) – (C) Run 1 – 3. The 15 PMS2 amplicons are given on the x-axis and the coverage, i.e. the number of reads, is given on the y-axis. The average of coverage of each amplicon and the standard deviation are represented by the blue and black bars, respectively. The red line marks the minimum threshold of coverage of 38.

3.1.3 Variants Detected in *PMS2*

In total 754 variants were identified in the 119 patient samples (appendix 6.6). The majority of variants (94.2%) were classified as class 1, not pathogenic or of no clinical significance. 34 variants (4.5%) were classified as class 2, likely not pathogenic or of little clinical significance. And 10 (1.3%) variants were classified as class 3, variants of unknown significant (VUS). All variants of class 2 or higher were confirmed by a new sequencing analysis performed by the Sanger method. Eight of the 119 samples were completely resequenced by Sanger method. These results coincide with the sequencing results obtained by the GS Junior sequencing method.

3.2 Optimization of MMR Genes Library Preparation

Within this master project four GS Junior runs were performed to optimize the amplicon library preparation procedure for sequencing the MMR genes *MSH2*, *MSH6*, *MLH1*, and *PMS2*. In another master project this procedure was already optimized for sequencing of the MMR genes *MSH2*, *MSH6*, and *MLH1* with the GS Junior technology[71]. Optimization of MMR genes sequencing was based on the results of this master thesis and the experience gained during optimization of the PMS2 library preparation; the latter was done within this master thesis. Two factors were optimized: (1) Conditions of singleplex PCRs and (2) volumes of singleplex PCR products used in multiplex PCRs. This was done to obtain a uniform distribution of coverage of all amplicons included in the GS Junior run and to exploit the capacity of the sequencing instrument.

3.2.1 Optimization of Singleplex PCR Conditions

After each GS Junior sequencing run the number of passed and mapped reads was recorded (Table 3.2). Comparison of these two values shows that a considerable percentage of high quality reads could not be aligned by the AVA software in each of the four MMR genes GS Junior runs (16.1%, 24.4%, 24.0%, and 30.2%, resp.). The primer dimers seen during optimization of the PMS2 library preparation were also detected in these four MMR genes libraries (Figure 3.4).

The singleplex PCR conditions of the amplicons possessing primer dimers were further optimized by testing different annealing temperature ranges. In order to facilitate the workflow of routine analysis only three annealing temperature ranges were selected (61-53 °C, 64-56 °C, and 66-62 °C). By optimizing the annealing temperature the primer dimer formation was reduced in the individual singleplex PCRs. However, even after extensive singleplex optimization, primer dimers can still be detected in most of the multiplex PCR products by the Agilent 2100 Bioanalyzer. Although, the reaction conditions of the singleplex and multiplex PCRs as well as the workflow, except the multiplexing step, were not changed during the four MMR genes GS Junior runs a change in the percentage of unmapped reads was observed. Within these four runs the percentage of unmapped reads was almost doubled (run 1: 16.1%; run 4: 30.2%).

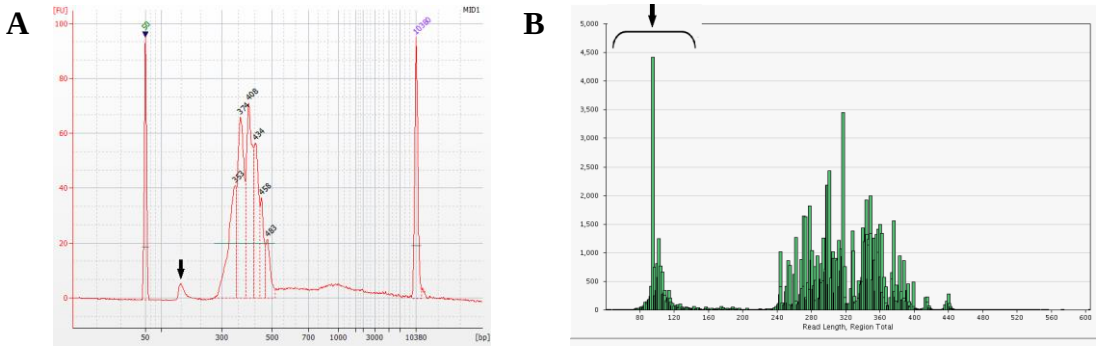

Figure 3.4. Illustration of fragment length of MMR genes amplicon pool. The small, unspecific fragments are indicated by black arrows. **(A)** Exemplary amplicon length profile obtained by the Agilent 2100 Bioanlyzer. The electropherogram shows the amplicon pool of a sample after multiplex PCR and Agencourt® AMPure® XP Purification. The Fragment size (bp) is given on the x-axis and the fluorescence is given on the y-axis, respectively. The analysis range of the Bioanalyzer DNA 7500 Chip is indicated by a lower marker (50bp) and an upper marker (10 380bp). **(B)** Amplicon length profile of GS Junior MMR genes run 2 obtained by the GS Run Browser software. This profile represents the total amplicon pool of the 8 samples included in the DNA library. The read length is given on the x-axis and the number of sequencing reads, i.e. the coverage, is given on the y-axis, respectively.

3.2.2 Evaluation of Parameters of GS Junior Sequencing Runs

A summary of the important characteristics of the four MMR genes GS Junior runs is given in Table 3.2. The sample capacity was calculated based on the number of mapped reads obtained in the particular GS Junior run. Raw data of coverage are given in appendix 6.7. Figure 3.5 presents the distribution of coverage of the first and fourth MMR genes runs graphically. This figure illustrates the improvement of the coverage distribution after the three optimization steps.

Table 3.2. Overview of characteristic parameters of the four MMR genes GS Junior sequencing runs.

	Run 1	Run 2	Run 3	Run 4
Samples	8	8	8	8
Amplicons	624	624	624	624
Passed reads	143 904	116 190	90 725	118 367
Mapped reads	120 796	87 873	68 973	82 598
Min/avg/max coverage	0/194/791	13/141/873	0/111/326	24/132/420
Coverage SD	108	87	51	53
Variation coefficient	0.56	0.62	0.46	0.40
Spread corr. Factor 90%/95%	2.23/2.85	2.07/2.47	1.94/2.35	1.84/2.21
Sample capacity F90/F95	18/14	14/11	11/9	15/12
Amplicons with no coverage	1	0	4	0
Amplicons <38	9	13	14	7

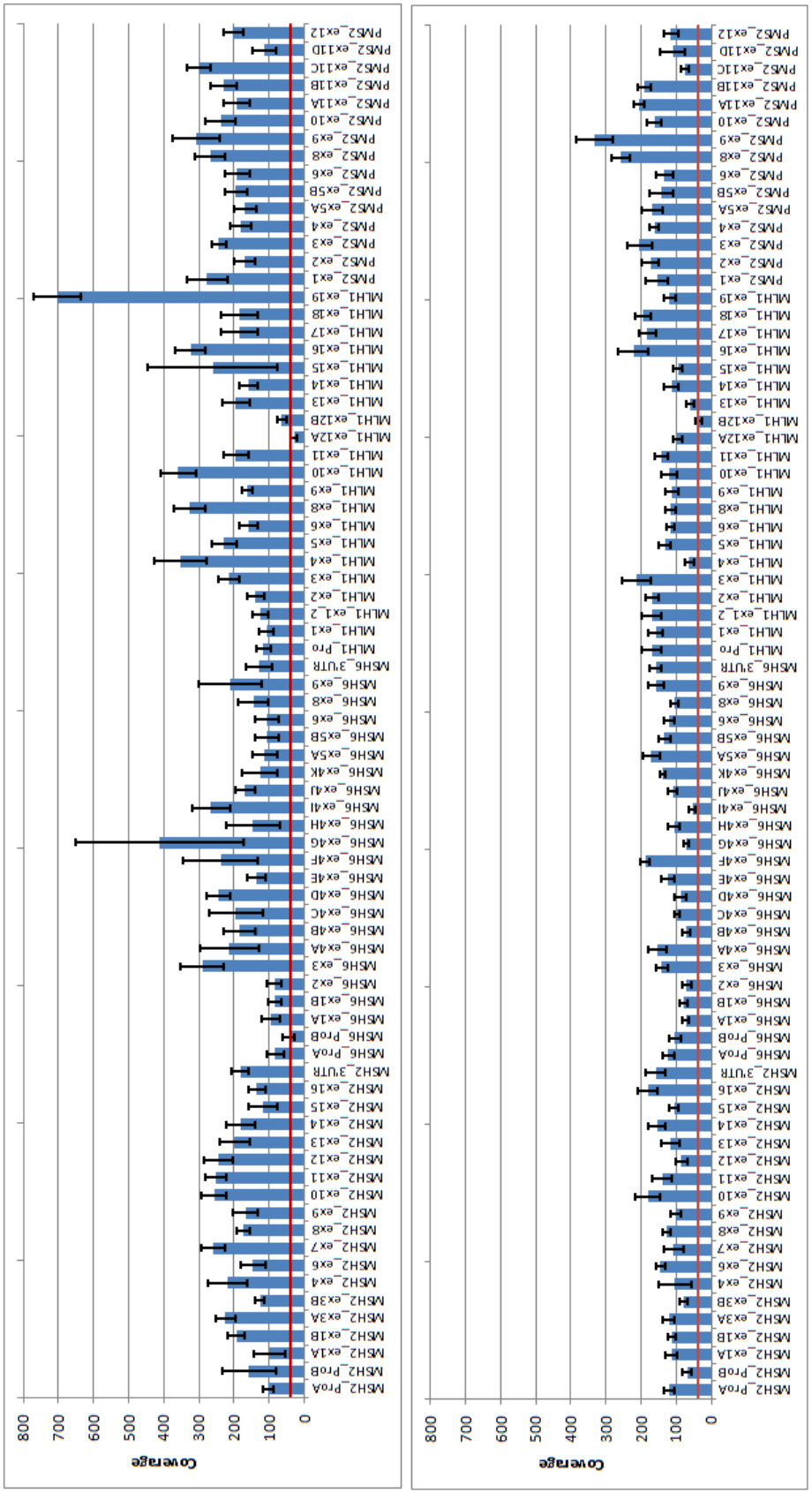

Figure 3.5. Distribution of coverage of the first and fourth MMR genes GS Junior sequencing runs. In both runs eight samples were pooled. The 78 amplicons of the MMR genes *MSH2, MSH6, MLH1,* and *PMS2* are given on the x-axis and the coverage is given on the y-axis. The average of coverage of each amplicon and the standard deviation are represented by the blue and black bars, respectively. The red line marks the minimum threshold of coverage of 38. (A) MMR genes GS Junior run 1. (B) MMR genes GS Junior run 4.

A more uniform distribution of coverage could be achieved by repeated optimization of the multiplexing. Based on the best spread correction factor (F90 1.84; F95 2.21), obtained in the four GS Junior run 15 and 12 samples, respectively can be pooled and analyzed per GS Junior run. Another indicator for an improved amplicon library preparation is the decreasing number of amplicons below the minimum threshold of coverage of 38.

1776 variants were detected in the four GS Junior runs when using a minimum threshold of 10% variant frequency. Only 367 (21%) of them were true variants. Heterozygote variants were detected in a range from 22%-63% of the reads (mean 47%). The detection range of homozygote variants was 94%-100% of the reads (mean 99%).

3.3 CRC-Biobank Pilot Study

In this pilot study, 32 patient samples from the CRC-biobank were analyzed for variants in the MMR genes *MSH2, MSH6, MLH1*, and *PMS2*. The biobank is part of the research project "CRC in Central Norway; Identification of Hereditary and Non-Hereditary Subtypes". The sequencing analysis was performed on blood samples. The samples were also used to optimize the preparation of the MMR genes library. In total 448 variants, consisting of 58 different variants, were identified in the 32 samples by GS Junior and Sanger sequencing (appendix 6.8). The majority of variants (94.2%) were classified as class 1. Variants of the other four classes were found less frequently: 2.5% in class 2; 2.0% in class 3; 0.4% in class 4; and 0.9% in class 5. All variants of class 2 or higher were confirmed by Sanger sequencing. 16 of the 32 samples were already sequenced by Sanger method. Comparison of the sequencing results showed no discrepancy.

Tumor analysis, including MSI-testing, *BRAF* mutational analysis, and *MLH1* methylation, was already performed for these samples. An overview of the tumor analysis and sequencing results is given in appendix 6.9. As described in the introduction (section 1.3.2) LS is indicated by microsatellite instable tumors (MSI-H) and a pathogenic variant in one of the four MMR genes. In contrast, an activating mutation in *BRAF* and a methylated *MLH1* promoter indicates a sporadic from of CRC. In six samples a definitely (class 5) or likely (class 4) pathogenic variant was found. None of these samples has a pathogenic *BRAF* mutation or methylation of the *MLH1* promoter. Four of these samples showed MSI in the tumor analysis, the two others were microsatellite stable (MSS). In three tumor samples methylation of *MLH1* was detected. These samples are microsatellite instable (MSI-H) and have a mutation in *BRAF* and no mutation in a MMR gene.

3.4 DNA-to-bead Ratio Used in emPCR

Tests showed that the recommended DNA-to-bead ratio of two molecules per bead is too high[77]. In this study a DNA-to-bead ratio of 0.5 was determined. This experimental determined ratio is in accordance with the calculations done by Berka *et al.*[63]. By such a low ratio the number of capture beads with more than one DNA species is kept low.

3.5 Comparison of Consumption of Time and Costs

To evaluate the benefit of the GS Junior system for analysis of the MMR genes the new technology was compared with Sanger sequencing method with regard to analysis costs and time. The calculations are based on the results obtained in the fourth MMR genes GS Junior run (section 3.2.2). In this run a spread correction factor (F_{95}) of 2.21 was achieved, that allows the sequencing of 12 samples in parallel. For calculating the sequencing preparation time, it is assumed that the singleplex PCRs are pipetted by a robot. The results show a clear benefit of the GS Junior system regarding consumption of costs and time (Table 3.3). It can be said that about two-thirds of costs and half of the time can be saved by using the new workflow for sequencing the four MMR genes of 12 patients per run.

Table 3.3. Comparison of consumption of costs and time of the GS Junior and Sanger sequencing method. The results refer to the sequencing of the four MMR genes of 12 patients in parallel.

	GS Junior*	**Sanger Sequencing**
Costs	**[Kr]**	**[Kr]**
Singleplex PCR reagents	7270	7270
Multiplex PCR reagents	660	-
Cycle Sequencing	8080	71090
96-well PCR plates	450	980
GS Junior Titanium Kits, Agilent 7500 DNA Kit	8010	-
Sum	**24470**	**79340**
Time	**[days]**	**[days]**
Sequencing preparation	6	7
Sequencing run	1 GS Junior instrument 10 h 3130xl Genetic Analyzer 15 h	8 3130xl Genetic Analyzer 132 h
Sequencing data analysis	3	7
Sum	**10**	**22**

* The calculations for the GS Junior sequencing method consider that 10 of the 88 fragments are analyzed by Sanger sequencing method.

4 Discussion

The study aimed to optimize massive parallel sequencing of the MMR genes *MSH2*, *MSH6*, *MLH1*, and *PMS2* using the GS Junior platform and to carry out a pilot study to determine the suitability of tumor screening analysis for identification of hereditary and non-hereditary CRC subtypes. The purpose of this new sequencing workflow is to replace Sanger sequencing, which is expensive and time-consuming. In this way, sequencing analysis of the four MMR genes can be offered to more CRC patients.

4.1 Choice of Experimental Design

Targeted sequencing was chosen as the aim is identification of variants in the MMR genes. Thereby, only information relevant for the experiment is produced and consumption of time and costs is reduced. There are multiple approaches to target regions of interest. Sequence capture, long range PCR amplicon sequencing, and universal tailed amplicon sequencing were shortlisted selection. Sequence capture methods physically isolate target DNA sequences from a pool of DNA molecules. However, it is not possible to discriminate between the sequence of interest and regions containing similar sequences such as common repeat regions, related genetic motifs and pseudogenes. For *PMS2* analysis it is important to differentiate between the region of interest and the *PMS2* pseudogenes. Therefore, amplicon experimental design and primers that are unique to the region of interest have to be used. For long range PCR amplicon sequencing amplicons longer than 1500bp are prepared to cover the sequence of interest. The disadvantage is that sequencing capacity is reduced by amplification of introns in addition to the coding and splice-site regions. [86]

Universal tailed amplicon sequencing was selected because it does not have these drawbacks and the number of primers can be reduced to a minimum. To reduce the number of PCRs for library preparation, an sequencing approach using two rounds of multiplex PCR, as described by De Leener *et al.*[46], was tested by another master student[71]. However, tests showed that it is difficult to optimize the conditions for the first round of multiplex PCRs. As it is crucial to work with highly optimized amplicon libraries to obtain uniform distribution of coverage, they were replaced by singleplex PCRs[46,71]. A drawback of the universal tail approach is that due to attachment of sequences required for amplification (universal tail), identification (MID sequence), and sequencing (key and 454 primers) about 25% of the maximal amplicon length (100bp) of each read is non-informative.

4.2 Optimization of MPS of MMR Genes

The experience gained during the performance of the seven GS Junior sequencing runs and evaluation of the data shows that several factors are important for the success of the sequencing analysis. These factors are a properly performed DNA library bead enrichment and thoroughly optimized amplicon library preparation procedure, which includes primer design, singleplex PCR conditions, multiplexing, and purification with AMPure beads.

4.2.1 Co-Amplification of *PMS2* Pseudogenes

There are 14 *PMS2* pseudogene loci that contain all or some of exons 1 – 5 and one pseudogene locus, *PMS2 CL*, that contains exons 9 and 11 – 15. In order to avoid co-amplification of the pseudogenes, the primers were designed in such a way that they take the minor differences of the loci into consideration. Thereby, it is possible to exclusively amplify the *PMS2* targets. As shown for fragment 11B (Figure 3.2), false priming to the pseudogenes can efficiently be prevented by using primers with mismatches at or near the 3' end[87]. This circumstance was also utilized for design of the sense and antisense primers of the other exons; in all these primers mismatches were located at the last or second last position of the 3' primer end. Since the sequences of intron 14 of *PMS2* and *PMS2 CL* near exon 15 are identical, no *PMS2* specific forward primer could be designed. In this case co-amplification of *PMS2 CL* is avoided by usage of a *PMS2* specific reverse primer. Because the combination of two primer sequences is required for a successful PCR amplification[87].

4.2.2 Exploitation of Throughput Capacity of GS Junior Platform

Roche guarantees 70 000 high quality reads when sequencing an amplicon library with the GS Junior platform[62]. In practice higher numbers of high quality reads, that passed the filters of the GS Run Processor, were obtained. However, a certain percentage of these reads could not be aligned to the reference sequence by the AVA software. Thereby, reducing the number of reads available for sequencing analysis. Thus, in order to exploit the capacity of the GS Junior system it is important to obtain high numbers of passed reads and to keep the difference between passed and mapped reads as low as possible.

In the three PMS2 GS Junior runs the percentage of mapped reads was increased from 53.5% to 79.9% by optimizing singleplex PCR conditions and thereby reducing primer dimer formation. The prevention of primer dimers is important because the Agencourt® AMPure® XP purification kit only allows the removal of fragments smaller than 100bp[75].

Although, the optimized singleplex PCR protocols were used for preparation of the MMR genes library, a decrease of mapped reads from 83.9% to 69.8% was observed in the four MMR genes GS Junior runs. This might indicate that the performance of the AMPure XP beads purification, i.e. the efficiency to remove small unwanted fragments from the amplicon library, is decreasing over time. According to information from Roche, the quality of the AMPure XP beads is greatly reduced by repeated vortexing. However, it was not possible to improve the purification efficiency with a new AMPure beads bottle. In addition, Roche suggested alternative protocols for purification of the amplicon library. Further tests have to be carried out to determine the optimal protocol for purification of the amplicon library.

In all four MMR genes runs the number of passes reads was above 70 000. It varied from 90 725 to 143 904. The difference can be explained by differences in the performance of the collection of the enriched DNA beads. In the third run, where the lowest number of passed reads was measured, this step was done according to the Roche manual[77]. While an extra cleaning step was performed in the other three runs, where more than 110 000 high quality reads were obtained. As explained in section 2.8.2, this additional step is used to remove the remaining paramagnetic beads from the collection of enriched DNA beads. It is important to remove as many paramagnetic enrichment beads as possible, because the enrichment beads can take up space on the PTP device. Thus, this extra step leads to a considerable reduction in the number of paramagnetic enrichment beads and it should be part of the standard procedure for collection of the enriched DNA beads.

4.2.3 Distribution of Coverage

For amplicon library preparation of PMS2 GS Junior run 1 equimolar pooling was used. In this run spread correction factors of 10.73 (F90) and 21.46 (F95) were obtained. The failure of achieving low spread correction factors by equimolarly pooling of singleplex PCRs might be due to local sequence characteristics and length bias during PCR amplification, where amplification of shorter fragments is favored over longer ones[88,89]. Dabney and Meyer showed that the median length of an amplicon library decreases with increasing number of PCR cycles[89]. This study demonstrated a strong correlation between the extent of decrease in length and the type of polymerase-buffer system used for amplification. According to their study the degree of length-bias in sequencing libraries can be minimized by choosing an appropriate polymerase-buffer system. Furthermore, they

showed that the length-bias is not enlarged by cycling into the plateau phase of PCR. This eases the planning of PCR programs used for amplification of amplicon libraries. The two polymerases used in this master study, SensiMix™ HRM from Bioline and AccuPrime™ GC-Rich from Invitrogen, were chosen because both allow high-specificity PCR amplification[72,74] and positive experience was gained during sequencing of the three MMR genes *MSH2*, *MSH6*, and *MLH1*[71]. Since both enzymes are not among the ten polymerases tested by Dabney and Meyer, no clear statement can be made about their performance in SGS analysis and the degree of length-bias introduced into the amplicon libraries. However, a tendency for a length-bias in favor of the short amplicons was seen in the first PMS2 GS Junior run and in three MSH2, MSH6, and MLH1 GS Junior runs[71], where equimolar pooling of singleplex PCRs was used.

In order to maximize the sequencing yield and the sample preparation efficiency, Roche recommends a length range of maximal 150bp for pooled amplicon libraries[60]. As the length range of the PMS2 library and MMR genes library is 178bp and 206bp, respectively this length-bias during PCR amplification might have impacts on distribution of coverage. To counterbalance this effect and possible bias due to local sequence characteristics the "weighted amplicon library mixture" approach was used. In this approach different ratio of singleplex PCR products are combined during amplicon library preparation. The ratio by which the most uniform coverage distribution can be obtained has to be determined experimentally. For this purpose the coverage of each amplicon was determined after each GS Junior run and the spread correction factor was calculated. This factor is a measure of the difference in coverage between "the least efficient" amplified fragment and "the best performing" fragment. The aim of adjusting the amounts of singleplex PCR products was to obtain similar coverage of all fragments and at the same time to avoid very high coverage which would be a waste of sequencing capacity[79]. In this way, low spread correction factors can be obtained and the number of patients analyzed per run can be increased. As seen in the GS Junior runs sequencing *PMS2* and the four MMR genes, the lowest spread correction factors were calculated for the last run after repeated optimization of the multiplexing step.

One important aspect during multiplex optimization was the development of the optimal procedure for preparation of the multiplex pools. The first GS Junior runs showed that it is difficult to obtain a uniform distribution of coverage by plain adjustment of the volumes of the singleplex PCRs. The volume adjustment of the fragments in a pool led to chances of the total volume of that particular pool. Consequently, as several pools had to be prepared

for each sample and the same amount of each pool was taken for the multiplex PCRs, the ratios within and between the different pools were changed. Thus, changing the volumes of the fragments in a pool without keeping the total volume stable had unforeseen impacts on the distribution of coverage. Therefore, a total volume for each pool that is dependent on the number of fragments included in the particular pool was defined. For convenience it was decided to multiply the number of fragments in a pool by the factor 10 to calculate the total volume. For example, a pool containing eight fragments will always have a total volume of 80 µL. This means, the total volume of a pool was independent from the amounts used of each fragment included in this pool and the volume difference was balanced with water.

4.2.4 Failure of Amplicons Detection

In five of the seven GS Junior runs performed in this study (appendix 6.5 and 6.7), amplicons with no coverage were recorded. The failures might be due to a poor singleplex PCR reaction or a mistake in multiplexing. In the PMS2 GS Junior runs, some samples showed a systematic failure of reads of all or the vast majority of amplicons. In PMS2 run 1, none of the amplicons of MID20 could be detected due to incorrect prepared MID20 primers dilutions. The failures in PMS2 run 3 are most probably due to a poor AMPure XP bead purification. Some amplicons of these samples were loaded on an agarose gel and singleplex PCR products were detected in all of them. However, only minor amounts of DNA were measured after AMPure XP purification by the NanoDrop spectrophotometer. Probably most of the failures are due to handling errors and can be avoid by thorough and accurate laboratory work. To minimize such errors a pipetting robot might be used for automation of the multiplexing and purification of the multiplex PCRs via AMPure XP beads. In the PMS2 run 3 a water control was included. As desired, no reads were recorded for this control.

4.3 Comparison of GS Junior and ABI3130XL Platform

Sanger sequencing on the ABI 3130XL platform, which is used at the moment for sequencing analysis of the four MMR genes, shall be replaced by 454 sequencing technology on the GS Junior platform. The aim of this conversion is to reduce analysis costs per patient and to increase throughput. At the same time, the new sequencing method has to be able to identify at least all variants detectable by Sanger method.

4.3.1 Detection of Variants in MMR Genes

To verify the correctness of the GS Junior sequencing results some of the samples were resequenced using Sanger method. In all cases the results were confirmed. This means, GS Junior platform performs as well as Sanger sequencing in terms of variant detection.

Theoretically, heterozygote variants should be detected with variant frequency of about 50%. In practice lower genotype calls are measured. The cause for this might be an imbalanced amplification of the two alleles of an amplicon, meaning that amplification of one allele is more efficient than of the other[88]. This might be caused by polymorphisms in or near the oligonucleotide priming sites[88]. In contrast to the literature, none of the amplicons in this study, where variants with low variation were observed, possesses known polymorphisms in or near the primer binding sites. It seems sequence characteristics nearby the variant play a role for basecalling efficiency. The variant with the lowest combined frequency is located in *PMS2* intron 12 between two short HPs (gtaaagttttaa) and is always detected with a frequency below 40%. Also other variants with a low combined frequency are repeatedly detected with lower frequencies when theoretically expected. Most often these variants are located in repetitive sequences and short HPs.

When analyzing the GS Junior sequencing data, the minimum threshold of the combined frequency (%) has to be chosen in such a way that false negative variants are avoided and the number of false positive variants is kept to a minimum. Initially, a combined frequency of 10% was used for data analysis in this study. This resulted in detection of 1409 (79%) false positive variants. When taking only variants with a frequency higher than 25% into account, as suggested by De Leeneer *et al.*, two true variants would have been missed in the four MMR genes runs[46]. Since this is not acceptable for diagnostic analysis a minimum threshold of 20% is suggested. The majority of the false positive variants were detected in homopolymeric regions.

4.3.2 Homopolymeric Regions

Correct basecalling in homopolymeric regions is proven difficult for pyrosequencing and it represents the major weakness of this SGS technology. According to the original description of the 454 sequencing technology, homopolymer stretches of up to eight nucleotides can be sequenced precisely[61]. However, problems with correct basecalling in homopolymeric repeats of four nucleotides were already reported[90]. This was also observed in the GS Junior sequencing results of this master project.

Nine of the 88 amplicons analyzed contain homopolymer repeats longer than 12bp. Eight of these amplicons are currently sequenced by Sanger method, as the results of pyrosequencing would not be sufficient for an accurate evaluation. Although, amplicon *MLH1* 12A contains a homopolymer stretch of 21 Ts it was analyzed on the GS Junior instrument. After optimization, more than 70 reverse reads can be measured, but less than 10 forward reads. As described in section 1.5.1, reads in opposite direction represent independent confirmation of the sequence analyzed and variant is considered to be real when it is detectable in both forward and reverse reads[60]. In contrast, many sequencing errors are only seen in one orientation. Thus, to obtain higher sequencing accuracy Sanger sequencing should be used as standard method for analysis of the amplicon *MLH1* 12A.

4.3.3 Consumption of Time

As expected, the usage of the GS Junior platform leads to a clear increase in sequencing throughput. Comparison of 454 sequencing and Sanger sequencing shows that the consumption of time can be reduced by 45% when sequencing the four MMR genes of 12 patients. The greatest reduction in time consumption (87.5%) is seen in the sequencing run itself. When using the ABI 3130XL Genetic Analyzer for Sanger sequencing maximally 2 x 96 cycle sequencing reactions can be sequenced per run. In order to sequence all amplicons of 12 patients 132 hours sequencing time are needed. In contrast, all 78 amplicons of each of the 12 patients can be sequenced in one single GS Junior sequencing run, which takes approximately 10 hours. Although, the AVA software is far from optimal for identification of variants in targeted genomic regions in diagnostic routines and it is more practical to export the variant frequency table to excel, 58% of time can be saved in the data analysis step. Only a minor improvement in time consumption could be achieved for the sequencing preparation procedure (13%). For preparation of the singleplex PCRs a pipetting robot from Hamilton is used. This reduces the time in the laboratory for both sequencing methods, Sanger and GS Junior, drastically. It is possible to reduce the time further by automation of more steps in the GS Junior workflow. For this purpose, Roche developed the REM e System, which is a liquid handling accessory and has to be installed on a pipetting robot[91]. Currently, this system can be used to fully automate the emulsion breakage, bead enrichment and sequence primer annealing procedures in the GS Junior System workflow[91]. Thereby, four hours of manual laboratory work can be saved and handling errors reduced.

4.3.4 Sequencing Analysis Costs

Calculation of the costs for kits and consumables for both GS Junior and Sanger sequencing showed that the aim, reduction of costs, could be achieved. The sequencing analysis of the four MMR genes per patient sample costs about 2 000 NOK and 6 600 NOK when using the combined GS Junior/Sanger sequencing approach and Sanger sequencing approach, respectively. This means, that the expenses can be reduced by about 70% when using the hybrid sequencing approach for sequencing of 12 patient samples per run. The saving is due to a considerable reduction of the sequencing costs. In contrast, the costs for PCR reagents and PCR plates are almost equal for both sequencing approaches.

4.4 Identification of Individuals with Lynch Syndrome

Early identification of individuals with LS is of great importance as surveillance programs can prevent development of advanced CRC and thereby mortality can be reduced[26,92]. In the pilot study the results of tumor analysis, including MSI, *BRAF* mutation, and *MLH1* methylation, were compared to the results of sequencing analysis of the four MMR genes *MSH2*, *MSH6*, *MLH1*, and *PMS2*. LS is characterized by a pathogenic variant in one of these MMR genes. The vast majority of LS-related tumors display a high level of MSI (MSI-H)[20]. However, MSI can also be found in about 15% of sporadic CRC cases[7]. In six samples a definitely pathogenic (class 5) or likely pathogenic (class 4) variant was detected; indicating LS. As expected for LS-associated tumors none of these had a pathogenic *BRAF* variant or abnormal methylation of *MLH1*. In contrast to the expectations, MSI was found only in four of the six samples. In one case this can be explained by the absence of tumor cells in the tissue sample due to radiotherapy. In the other case LS is caused by a splice-site mutation in *PMS2* (c.989-1T) that results in two abnormal transcripts; one transcript completely lacks the adjacent exon 10 and the other only the first 27 nucleotides of exon 10[93]. The lack of MSI in the tumor sample might be due to some remaining mismatch repair functionality of this truncated protein[93]. This *PMS2* mutation was also detected in another sample of the pilot study. However, this sample showed MSI as expected. Other factors such as modifier genes might play a role in the development of MSI in patients possessing this particular *PMS2* splice-site mutation. The results indicate that selection of samples for sequencing analysis of the MMR genes based on tumor analysis is suitable to identify most of the samples with LS. However, by this preselection of samples for mutational screening some LS cases will be missed.

Therefore, it is very important to take the age of onset and the family history, by means of AMII criteria and RBG, into account when selecting the samples for sequencing analysis of the MMR genes. In order to identify all individuals with LS, sequencing of all new CRC cases would be the best way. Currently, the cost-benefit ratio of such comprehensive screening programs is too low to be implemented into diagnostic routine. Furthermore, analysis of tumor samples for MSI and loss of MMR-protein expression by means of IHC may predict the response to chemotherapy and this is also relevant for sporadic cases of CRC[92]. In summary, it can be said that MSI analysis and/or IHC is recommended for all CRCs and LS-associated tumors[92,94].

4.5 Conclusion and Prospective Work

Within this master study two efficient workflows for massive parallel amplicon sequencing of *PMS2* and the four MMR genes (*MSH2*, *MSH6*, *MLH1*, and *PMS2*) were developed. By means of these workflows it is possible to reduce costs and working time drastically. Thereby, more samples can be sequenced in the Medical Genetics Laboratory. To decrease costs and time further multiplex PCR instead of singleplex PCR could be used for amplification of the coding and splice-site regions of the MMR genes. Such an approach was successfully established by De Leeneer *et al.* for sequencing of the breast cancer genes *BRCA1* and *BRCA2*[46]. In case of MMR genes sequencing, as performed in this master study, optimization of multiplex PCRs is hampered by the big differences of the shorted and longest amplicon in the DNA library. Probably, optimization will be too time-consuming to obtain a positive cost-benefit ratio. Automation of more steps in the sequencing workflow can further reduce working time as well as handling errors. In order to fully exploit the capacity of the 454 GS Junior sequencing platform an efficient protocol for the removal of the primer dimers has to be established. For this purpose the Medical Genetics Laboratory is in contact with Roche and is testing various strategies.

The results of the pilot study, in which 32 samples of a CRC biobank were sequenced, indicate that most individuals with LS can be identified by using tumor analysis for selecting samples that have to be sequenced. In order to make a clear statement the remaining 330 samples of the biobank have to be analyzed using the workflow developed in this thesis. This analysis will be part of another master student project.

5 References

1 Globocan 2008, International Agency for Research on Cancer. [2012-12-10] Available from: <http://globocan.iarc.fr/>

2 Cancer in Norway 2009, Cancer incidence, mortality, survival and prevalence in Norway. [2012-12-10] Available from: <http://www.kreftregisteret.no/Global/Cancer%20in%20Norway/2009/Cancer_in_Norway_2009_trykkversjonen_for_web.pdf>

3 Blanes, A. & Diaz-Cano, S. J. Complementary analysis of microsatellite tumor profile and mismatch repair defects in colorectal carcinomas. *World journal of gastroenterology : WJG* **12**, 5932-5940 (2006).

4 Strate, L. L. & Syngal, S. Hereditary colorectal cancer syndromes. *Cancer Causes Control* **16**, 201-213, doi:10.1007/s10552-004-3488-4 (2005).

5 Mecklin, J. P. The implications of genetics in colorectal cancer. *Ann. Oncol.* **19 Suppl 5**, v87-90, doi:10.1093/annonc/mdn318 (2008).

6 Gala, M. & Chung, D. C. Hereditary colon cancer syndromes. *Semin. Oncol.* **38**, 490-499, doi:10.1053/j.seminoncol.2011.05.003 (2011).

7 Soreide, K., Janssen, E. A., Soiland, H., Korner, H. & Baak, J. P. Microsatellite instability in colorectal cancer. *Br. J. Surg.* **93**, 395-406, doi:10.1002/bjs.5328 (2006).

8 Gervaz, P., Bucher, P. & Morel, P. Two colons-two cancers: paradigm shift and clinical implications. *J. Surg. Oncol.* **88**, 261-266, doi:10.1002/jso.20156 (2004).

9 Lynch, H. T. & Lynch, J. Lynch syndrome: genetics, natural history, genetic counseling, and prevention. *J. Clin. Oncol.* **18**, 19S-31S (2000).

10 Al-Sukhni, W., Aronson, M. & Gallinger, S. Hereditary colorectal cancer syndromes: familial adenomatous polyposis and lynch syndrome. *Surg. Clin. North Am.* **88**, 819-844, vii, doi:10.1016/j.suc.2008.04.012 (2008).

11 de la Chapelle, A. The incidence of Lynch syndrome. *Familial cancer* **4**, 233-237, doi:10.1007/s10689-004-5811-3 (2005).

12 Vasen, H. F. *et al.* Guidelines for the clinical management of Lynch syndrome (hereditary non-polyposis cancer). *J. Med. Genet.* **44**, 353-362, doi:10.1136/jmg.2007.048991 (2007).

13 Bellizzi, A. M. & Frankel, W. L. Colorectal cancer due to deficiency in DNA mismatch repair function: a review. *Adv. Anat. Pathol.* **16**, 405-417, doi:10.1097/PAP.0b013e3181bb6bdc (2009).

14 Stoffel, E. *et al.* Calculation of risk of colorectal and endometrial cancer among patients with Lynch syndrome. *Gastroenterology* **137**, 1621-1627, doi:10.1053/j.gastro.2009.07.039 (2009).

15 Lynch, H. T. *et al.* Phenotypic and genotypic heterogeneity in the Lynch syndrome: diagnostic, surveillance and management implications. *Eur. J. Hum. Genet.* **14**, 390-402, doi:10.1038/sj.ejhg.5201584 (2006).

16 Hampel, H. *et al.* Cancer risk in hereditary nonpolyposis colorectal cancer syndrome: later age of onset. *Gastroenterology* **129**, 415-421, doi:10.1016/j.gastro.2005.05.011 (2005).

17 Rijcken, F. E., Hollema, H. & Kleibeuker, J. H. Proximal adenomas in hereditary non-polyposis colorectal cancer are prone to rapid malignant transformation. *Gut* **50**, 382-386 (2002).

18 Julie, C. *et al.* Identification in daily practice of patients with Lynch syndrome (hereditary nonpolyposis colorectal cancer): revised Bethesda guidelines-based approach versus molecular screening. *Am. J. Gastroenterol.* **103**, 2825-2835; quiz 2836, doi:10.1111/j.1572-0241.2008.02084.x (2008).

19 Trano, G., Wasmuth, H. H., Sjursen, W., Hofsli, E. & Vatten, L. J. Awareness of heredity in colorectal cancer patients is insufficient among clinicians: a Norwegian population-based study. *Colorectal disease : the official journal of the Association of Coloproctology of Great Britain and Ireland* **11**, 456-461, doi:10.1111/j.1463-1318.2009.01830.x (2009).

20 Trano, G., Sjursen, W., Wasmuth, H. H., Hofsli, E. & Vatten, L. J. Performance of clinical guidelines compared with molecular tumour screening methods in identifying possible Lynch syndrome among colorectal cancer patients: a Norwegian population-based study. *Br. J. Cancer* **102**, 482-488, doi:10.1038/sj.bjc.6605509 (2010).

21 Tops, C. M., Wijnen, J. T. & Hes, F. J. Introduction to molecular and clinical genetics of colorectal cancer syndromes. *Best practice & research. Clinical gastroenterology* **23**, 127-146, doi:10.1016/j.bpg.2009.02.002 (2009).

22 de la Chapelle, A. Microsatellite instability. *N. Engl. J. Med.* **349**, 209-210, doi:10.1056/NEJMp038099 (2003).

23 Chung, D. C. & Rustgi, A. K. The hereditary nonpolyposis colorectal cancer syndrome: genetics and clinical implications. *Ann. Intern. Med.* **138**, 560-570 (2003).

24 Wheeler, J. M., Loukola, A., Aaltonen, L. A., Mortensen, N. J. & Bodmer, W. F. The role of hypermethylation of the hMLH1 promoter region in HNPCC versus MSI+ sporadic colorectal cancers. *J. Med. Genet.* **37**, 588-592 (2000).

25 Deng, G. *et al.* BRAF mutation is frequently present in sporadic colorectal cancer with methylated hMLH1, but not in hereditary nonpolyposis colorectal cancer. *Clin. Cancer Res.* **10**, 191-195 (2004).

26 Jarvinen, H. J. *et al.* Controlled 15-year trial on screening for colorectal cancer in families with hereditary nonpolyposis colorectal cancer. *Gastroenterology* **118**, 829-834 (2000).

27 Church, J. & Simmang, C. Practice parameters for the treatment of patients with dominantly inherited colorectal cancer (familial adenomatous polyposis and hereditary nonpolyposis colorectal cancer). *Dis. Colon Rectum* **46**, 1001-1012, doi:10.1097/01.dcr.0000080143.71778.af (2003).

28 Li, G. M. Mechanisms and functions of DNA mismatch repair. *Cell Res.* **18**, 85-98, doi:10.1038/cr.2007.115 (2008).

29 Woods, M. O. *et al.* A new variant database for mismatch repair genes associated with Lynch syndrome. *Hum. Mutat.* **28**, 669-673, doi:10.1002/humu.20502 (2007).

30 de la Chapelle, A. Genetic predisposition to colorectal cancer. *Nature reviews. Cancer* **4**, 769-780, doi:10.1038/nrc1453 (2004).

31 Martin, A. & Scharff, M. D. AID and mismatch repair in antibody diversification. *Nature reviews. Immunology* **2**, 605-614, doi:10.1038/nri858 (2002).

32 Strachan, T. & Read, A. Human Molecular Genetics. 4th edn, Garland Science, Taylor & Francis Group, LLC (2011).

33 Horii, A., Han, H. J., Sasaki, S., Shimada, M. & Nakamura, Y. Cloning, characterization and chromosomal assignment of the human genes homologous to yeast PMS1, a member of mismatch repair genes. *Biochem. Biophys. Res. Commun.* **204**, 1257-1264, doi:10.1006/bbrc.1994.2598 (1994).

34 Nicolaides, N. C. *et al.* Genomic organization of the human PMS2 gene family. *Genomics* **30**, 195-206, doi:10.1006/geno.1995.9885 (1995).

35 Osborne, L. R. *et al.* PMS2-related genes flank the rearrangement breakpoints associated with Williams syndrome and other diseases on human chromosome 7. *Genomics* **45**, 402-406, doi:10.1006/geno.1997.4923 (1997).

36 Chadwick, R. B., Meek, J. E., Prior, T. W., Peltomaki, P. & de La Chapelle, A. Polymorphisms in a pseudogene highly homologous to PMS2. *Hum. Mutat.* **16**, 530, doi:10.1002/1098-1004(200012)16:6<530::aid-humu15>3.0.co;2-6 (2000).

37 Valero, M. C., de Luis, O., Cruces, J. & Perez Jurado, L. A. Fine-scale comparative mapping of the human 7q11.23 region and the orthologous region on mouse chromosome 5G: the low-copy repeats that flank the Williams-Beuren syndrome deletion arose at breakpoint sites of an evolutionary inversion(s). *Genomics* **69**, 1-13, doi:10.1006/geno.2000.6312 (2000).

38 De Vos, M., Hayward, B. E., Picton, S., Sheridan, E. & Bonthron, D. T. Novel PMS2 pseudogenes can conceal recessive mutations causing a distinctive childhood cancer syndrome. *Am J Hum Genet* **74**, 954-964, doi:10.1086/420796 (2004).

39 Kondo, E., Horii, A. & Fukushige, S. The human PMS2L proteins do not interact with hMLH1, a major DNA mismatch repair protein. *J Biochem* **125**, 818-825 (1999).

40 Collins, F. S., Morgan, M. & Patrinos, A. The Human Genome Project: lessons from large-scale biology. *Science* **300**, 286-290, doi:10.1126/science.1084564 (2003).

41 Yalcin, B., Adams, D. J., Flint, J. & Keane, T. M. Next-generation sequencing of experimental mouse strains. *Mamm. Genome* **23**, 490-498, doi:10.1007/s00335-012-9402-6 (2012).

42 Abecasis, G. R. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061-1073, doi:10.1038/nature09534 (2010).

43 Meyerson, M., Gabriel, S. & Getz, G. Advances in understanding cancer genomes through second-generation sequencing. *Nature reviews. Genetics* **11**, 685-696, doi:10.1038/nrg2841 (2010).

44 Kuhlenbaumer, G., Hullmann, J. & Appenzeller, S. Novel genomic techniques open new avenues in the analysis of monogenic disorders. *Hum. Mutat.* **32**, 144-151, doi:10.1002/humu.21400 (2011).

45 Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937-948, doi:10.1038/ng.686 (2010).

46 De Leeneer, K. *et al.* Massive parallel amplicon sequencing of the breast cancer genes BRCA1 and BRCA2: opportunities, challenges, and limitations. *Hum. Mutat.* **32**, 335-344, doi:10.1002/humu.21428 (2011).

47 Schadt, E. E., Turner, S. & Kasarskis, A. A window into third-generation sequencing. *Hum. Mol. Genet.* **19**, R227-240, doi:10.1093/hmg/ddq416 (2010).

48 Liu, L. *et al.* Comparison of next-generation sequencing systems. *Journal of biomedicine & biotechnology* **2012**, 251364, doi:10.1155/2012/251364 (2012).

49 Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463-5467 (1977).

50 Pareek, C. S., Smoczynski, R. & Tretyn, A. Sequencing technologies and genome sequencing. *Journal of applied genetics* **52**, 413-435, doi:10.1007/s13353-011-0057-x (2011).

51 Anasagasti, A., Irigoyen, C., Barandika, O., Lopez de Munain, A. & Ruiz-Ederra, J. Current mutation discovery approaches in Retinitis Pigmentosa. *Vision Res.* **75**, 117-129, doi:10.1016/j.visres.2012.09.012 (2012).

52 Leamon, J. H., Lohman, K. L., Rothberg, J. M. & Weiner, M. P. Methods of Amplifying and Sequencing Nucleic Acids. US 8,158,359 B2. (2012).

53 Metzker, M. L. Sequencing technologies - the next generation. *Nature reviews. Genetics* **11**, 31-46, doi:10.1038/nrg2626 (2010).

54 Illumina, HiSeq 2500/1500, Specifications. [2013-02-07] Available from: <http://www.illumina.com/systems/hiseq_2500_1500/performance_specifications.ilmn>

55 Illumina, HiSeq 2000/1000, Performance & Specifications. [2013-02-07] Available from: <http://www.illumina.com/systems/hiseq_2000_1000/performance_specifications.ilmn> (

56 Illumina, Genome Analyzer IIx, Specifications. [2013-02-07] Available from: <http://www.illumina.com/systems/genome_analyzer_iix/performance_specifications.ilmn>

57 Illumina, MiSeq Personal Sequencer, Specifications. [2013-02-07] Available from: <http://www.illumina.com/systems/miseq/performance_specifications.ilmn>

58 Life TEchnologies, Ion PGM™ Sequencer. [2013-02-07] Available from: <https://tools.invitrogen.com/content/sfs/brochures/PGM-Specification-Sheet.pdf>

59 Roche, 454 Sequencing, GS FLX+ System. [2013-02-07] Available from: <http://454.com/products/gs-flx-system/index.asp>

60 GS Junior System, Guidelines for Amplicon Experimental Design. [2013-05-15] Available from: <http://454.com/downloads/my454/documentation/gs-junior/system-wide-documents/454SequencingSystemsGuidelinesForAmpliconExperimentalDesign_Nov2012.pdf>

61 Leamon, J. H. & Rothberg, J. M. Cramming more sequencing reactions onto microreactor chips. *Chem. Rev.* **107**, 3367-3376, doi:10.1021/cr068297s (2007).

62 Roche, 454 Sequencing Technology. [2013-05-31] Available from: <http://454.com/products/technology.asp>

63 Jan Berka, Y.-J. C., John H. Leamon, Steve Lefkowitz, Kenton L. Lohman, Vinod B. Makhijani, Jonathan M. Rothberg, Gary J. Sarkis, Maithreyan Srinivasan, Michael P. Weiner. Bead emulsion nucleic acid amplification. (2011).

64 Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlen, M. & Nyren, P. Real-time DNA sequencing using detection of pyrophosphate release. *Anal. Biochem.* **242**, 84-89, doi:10.1006/abio.1996.0432 (1996).

65 Ronaghi, M., Uhlen, M. & Nyren, P. A sequencing method based on real-time pyrophosphate. *Science* **281**, 363, 365 (1998).

66 GS Junior, 454 Sequencing System Software Manual, v2.7, General Overview and Data File Formats. [2013-05-31] Available from: <http://454.com/downloads/my454/documentation/gs-junior/software-manual/454SequencingSystemSoftwareManualv2.7-GeneralOverview_March2012.pdf>

67 Invitrogen, iPrep™ Pure Link™ gDNA Blood Kit, for Purification of gDNA from Human Blood using the iPrep™ Purification Instrument. [2013-06-01] Available from: <http://tools.invitrogen.com/content/sfs/manuals/iprep_bloodgDNA_man.pdf>

68 Thermo Scientific, NanoDrop® 1000 Spectrophotometer v.3.7 User's Manual. [2013-06-01] Available from: <http://memphys.dk/sites/default/files/files/basic_page/%3Cem%3EEdit%20Basic%20page%3C/em%3E%20UV-Vis%20spectroscopy/nd-1000-v3.7-users-manual-8.5x11.pdf>

69 Thermo Scientific, NanoDrop8000 Spectrophotometer V2.2 User Manual. [2013-06-01] Available from: <http://www.nanodrop.com/Library/nd-8000-v2.2%20-users-manual-8.5%20x%2011.pdf>

70 Plon, S. E. *et al.* Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum. Mutat.* **29**, 1282-1291, doi:10.1002/humu.20880 (2008).

71 Vold, T. Implementation of Massive Parallel Sequencing Technology in the Molecular Genetic Diagnostics of Lynch Syndrome; Universal tailed amplicon sequencing of the MMR genes MSH2, MSH6 and MLH1. Master's thesis, Norwegian University of Science and Technology, (2012).

72 Bioline, SensiMix™ HRM Kit. [2013-06-02] Available from: <http://www.bioline.com/documents/product_inserts/SensiMix%20HRM%20Kit.pdf>

73 McPherson, M. & Møller, S. *PCR*. 2nd edn, (Garland Science, Taylor & Francis Group, 2006).

74 Invitrogen, AccuPrime™ GC-Rich DNA Polymerase. [2013-06-02] Available from: <http://tools.invitrogen.com/content/sfs/manuals/12337.pdf>

75 Beckman Coulter, Agencourt® AMPure® XP - PCR Purification. [2013-06-01] Available from: <https://www.beckmancoulter.com/wsrportal/bibliography?docname=Protocol_000387v001.pdf>

76 Roche, Amplicon Library Preparation Method Manual. [2013-06-01] Available from: <http://454.com/downloads/my454/documentation/gs-junior/method-manuals/GSJuniorAmpliconLibraryPrepMethodManual_March2012.pdf>

77 Roche, emPCR Amplification Method Manual - Lib-A. [2013-06-01] Available from: <http://454.com/downloads/my454/documentation/gs-junior/method-manuals/GSJunioremPCRAmplificationMethodManualLib-A_March2012.pdf>

78 Roche, Sequencing Method Manual. [2013-06-01] Available from: <http://454.com/downloads/my454/documentation/gs-junior/method-manuals/GSJuniorSequencingManual_Jan2013.pdf>

79 De Leeneer, K. *et al.* Practical tools to implement massive parallel pyrosequencing of PCR products in next generation molecular diagnostics. *PloS one* **6**, e25531, doi:10.1371/journal.pone.0025531 (2011).

80 GE Healthcare Life Sciences, Illustra™ ExoStar. [2013-05-31] Available from: <https://www.gelifesciences.com/gehcls_images/GELS/Related%20Content/Files/1326706518989/litdoc29010724_20130129211715.pdf>

81 Ge Healthcare Life Sciences, Illustra™ ExoStar™ 1-Step. [2013-05-31] Available from: <http://193.218.17.133/ex/downloads/brochures/life_science/ge_illustra_exostar.pdf>

82 Applied Biosystems, BigDye® Terminator v3.1 Cycle Sequencing Kit - Protocol. [2013-05-27] Available from: <http://tools.invitrogen.com/content/sfs/manuals/cms_041329.pdf>

83 Applied Biosystems, BigDye® XTerminator™ Purification Kit - Protocol. [2013-05-27] Available from: <http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_042772.pdf>

84 Invitrogen, E-Gel® Technical Guide. [2013-06-01] Available from: <http://tools.invitrogen.com/content/sfs/manuals/egelguide_man.pdf>

85 Agilent Technology, DNA 7500 and DNA 12000 Kit Quick Start Guide. [2013-06-01] Available from: <http://www.uri.edu/research/gsc/docs/DNA7500_Guide.pdf>

86 Roche, GS Junior System Research Applications Guide. [2013-05-11] Available from: <http://454.com/downloads/my454/documentation/gs-junior/system-wide-documents/GSJuniorSystemResearchApplicationsGuide_March2012.pdf>

87 Mitsuhashi, M. Technical report: Part 2. Basic requirements for designing optimal PCR primers. *J. Clin. Lab. Anal.* **10**, 285-293, doi:10.1002/(sici)1098-2825(1996)10:5<285::aid-jcla9>3.0.co;2-7 (1996).

88 Harismendy, O. *et al.* Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome biology* **10**, R32, doi:10.1186/gb-2009-10-3-r32 (2009).

89 Dabney, J. & Meyer, M. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* **52**, 87-94, doi:10.2144/000113809 (2012).

90 Hert, D. G., Fredlake, C. P. & Barron, A. E. Advantages and limitations of next-generation sequencing technologies: a comparison of electrophoresis and non-electrophoresis methods. *Electrophoresis* **29**, 4618-4626, doi:10.1002/elps.200800456 (2008).

91 Roche, REM e System. [2013-05-31] Available from: <http://454.com/downloads/NIMBUSREMeGSJunior_ApplicationNote_Feb2013.pdf>

92 Vasen, H. F. *et al.* Recommendations to improve identification of hereditary and familial colorectal cancer in Europe. *Familial cancer* **9**, 109-115, doi:10.1007/s10689-009-9291-3 (2010).

93 Sjursen, W. *et al.* A homozygote splice site PMS2 mutation as cause of Turcot syndrome gives rise to two different abnormal transcripts. *Familial cancer* **8**, 179-186, doi:10.1007/s10689-008-9225-5 (2009).

94 Sjursen, W. *et al.* Current clinical criteria for Lynch syndrome are not sensitive enough to identify MSH6 mutation carriers. *J. Med. Genet.* **47**, 579-585, doi:10.1136/jmg.2010.077677 (2010).

95 Vasen, H. F., Watson, P., Mecklin, J. P. & Lynch, H. T. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology* **116**, 1453-1456 (1999).

96 Umar, A. *et al.* Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J. Natl. Cancer Inst.* **96**, 261-268 (2004).

6 Appendix

6.1 Clinical Diagnostic Criteria

Amsterdam Criteria II **(all criteria must be fulfilled)**[95]
There should be at least three relatives with an Lynch syndrome-associated cancer: CRC, cancer of the endometrium, small bowel, ureter, or renal pelvis
One should be a first-degree relative of the other two
At least two successive generations should be affected
At least one should be diagnosed before age 50
Familial adenomatous polyposis should be excluded in the CRC case(s) if any
Tumors should be verified by pathological examination
Revised Bethesda Guidelines (one of the criteria is sufficient to classify the patient/family)[96]
CRC diagnosed in a patient < 50 years
Presence of synchronous, metachronous colorectal or other Lynch-related tumours[1], regardless of age
CRC with MSI-H phenotype[2] diagnosed in a patient diagnosed aged < 60 years
Patient with CRC and a first-degree relative with a Lynch syndrome-related tumor, with one of the cancers diagnosed at age < 50 years
Patient with CRC with two or more first-degree or second-degree relatives with a Lynch syndrome-related tumor, regardless of age

[1]Lynch syndrome-related tumors: colorectal, endometrial, stomach, ovarian, pancreas, ureters, renal pelvis, biliary tract and brain tumors, sebaceous gland adenomas and keratoacanthomas, and carcinoma of the small bowel.

[2]Lymphocyte-infiltrating tumors, low grade or undifferentiated, Crohn's-like lymphocyte infiltration, the presence of mucin or signet cells in the tumors, and "cribriform growth pattern".

6.2 Fusion Primers for Singleplex and Multiplex PCR

The annealing temperature range refers to the annealing temperature used in the singleplex PCR amplification. Fragments that are solely sequenced by Sanger sequencing are marked in purple. Intrinsic sequencing primer used for Sanger sequencing are marked in blue.

Primer Name	Primer sequence: universal tail – target specific sequence	Temperature [°C]
GS-MSH2proA-s	CACGACGTTGTAAAACGAC-TCCTTCTGATGTTACTCCCATGC	64-56
GS-MSH2proA-as	CAGGAAACAGCTATGACC-GCATGCCTGCGCCTAGGTC	
GS-MSH2proB-s	CACGACGTTGTAAAACGAC-TTGCATACACCCCACCCAGG	64-56
GS-MSH2proB-as	CAGGAAACAGCTATGACC-AAGAAAATGCGCGACCCCAC	
GS-MSH2ex1A-s	CACGACGTTGTAAAACGAC-GGAGGCGGGAAACAGCTTAGTG	64-56
GS-MSH2ex1A-as	CAGGAAACAGCTATGACC-ACCCCCTGGGTCTTGAACACC	
GS-MSH2ex1B-s	CACGACGTTGTAAAACGAC-TCAACCAGGAGGTGAGGAGGTTTC	64-56
GS-MSH2ex1B-as	CAGGAAACAGCTATGACC-GGAAAGGAGCCGCGCCAC	
GS-MSH2ex2-s	CACGACGTTGTAAAACGAC-GAAGTCCAGCTAATACAGTGC	61-53
GS-MSH2ex2-as	CAGGAAACAGCTATGACC-ACTAAAACACAATTAAATTCTTCAC	
GS-MSH2ex2sek-s	CACGACGTTGTAAAACGACAATGTACTTTTTTTTTTTTTTAAG	-
GS-MSH2ex3A.s	CACGACGTTGTAAAACGAC-TTAAAGTATGTTCAAGAGTTTGTT	61-53
GS-MSH2ex3A.as	CAGGAAACAGCTATGACC-GGAGAACTGATCATTATCAGG	
GS-MSH2ex3B.s	CACGACGTTGTAAAACGAC-GTTGTGGGTGTTAAAATGTC	61-53
GS-MSH2ex3B.as	CAGGAAACAGCTATGACC-GAATCTCCTCTATCACTAGACTCA	
GS-MSH2ex4-s	CACGACGTTGTAAAACGAC-TTCTTATTCCTTTTCTCATAGTAG	61-53
GS-MSH2ex4-as	CAGGAAACAGCTATGACC-ATATTGTAATTCACATTTATAATCC	
GS-MSH2ex5-s	CACGACGTTGTAAAACGAC-GATCCAGTGGTATAGAAATCTTC	61-53
GS-MSH2ex5-as	CAGGAAACAGCTATGACC-CCATTCAACATTTTTAACCCT	
GS-MSH2ex5sek-as	CAGGAAACAGCTATGACC-TTTTTTTTTTTTTTTTACCTGAA	-
GS-MSH2ex6-s	CACGACGTTGTAAAACGAC-TTCACTAATGAGCTTGCCAT	64-56
GS-MSH2ex6-as	CAGGAAACAGCTATGACC-TGGGTAACTGCAGGTTACAT	
GS-MSH2ex7-s	CACGACGTTGTAAAACGAC-CTAAAATATTTTACATTAATTCAAG	61-53
GS-MSH2ex7-as	CAGGAAACAGCTATGACC-AAGTATATATTGTATGAGTTGAAGG	
GS-MSH2ex8-s	CACGACGTTGTAAAACGAC-GATTTGTATTCTGTAAAATGAGATC	61-53
GS-MSH2ex8-as	CAGGAAACAGCTATGACC-CCTTTGCTTTTTAAAAATAACTAC	
GS-MSH2ex9-s	CACGACGTTGTAAAACGAC-CCCATTATTTATAGGATTTTGTCA	61-53
GS-MSH2ex9-as	CAGGAAACAGCTATGACC-AATTATTCCAACCTCCAATG	
GS-MSH2ex10-s	CACGACGTTGTAAAACGAC-TGGTAGTAGGTATTTATGGAATAC	61-53
GS-MSH2ex10-as	CAGGAAACAGCTATGACC-TTAGGGAATTAATAAAGGGTT	
GS-MSH2ex11-s	CACGACGTTGTAAAACGAC-GATACTTTGGATATGTTTCACG	61-53
GS-MSH2ex11-as	CAGGAAACAGCTATGACC-TGACATTCAGAACATTATTAGTTC	
GS-MSH2ex12-s	CACGACGTTGTAAAACGAC-TTTATTATTCAGTATTCCTGTGTAC	61-53
GS-MSH2ex12-as	CAGGAAACAGCTATGACC-AACGTTACCCCCACAAAG	
GS-MSH2ex13-s	CACGACGTTGTAAAACGAC-CATTTATTAGTAGCAGAAAGAAGT	61-53
GS-MSH2ex13-as	CAGGAAACAGCTATGACC-ATACATTTCTATCTTCAAGGGAC	
GS-MSH2ex14-s	CACGACGTTGTAAAACGAC-ACCACATTTTATGTGATGGG	61-53
GS-MSH2ex14-as	CAGGAAACAGCTATGACC-CAAGTTCTGAATTTAGAGTACTCC	
GS-MSH2ex15-s	CACGACGTTGTAAAACGAC-TTGCTGTCTCTTCTCATGC	64-56
GS-MSH2ex15-as	CAGGAAACAGCTATGACC-TTCATCTTAGTGTCCTGTTTATG	
GS-MSH2ex16-s	CACGACGTTGTAAAACGAC-ATTTTAATTACTAATGGGACATTC	61-53
GS-MSH2ex16-as	CAGGAAACAGCTATGACC-TCAATATTACCTTCATTCCATTAC	
GS-MSH2_3'UTR-s	CACGACGTTGTAAAACGAC-TTTCACGAATAAAAGTTACTACG	61-53
GS-MSH2_3'UTR-as	CAGGAAACAGCTATGACC-TGCGAAGAACTACAATGC	

Primer Name	Primer sequence: universal tail – target specific sequence	Temperature [°C]
GS-MSH6proA-s	CACGACGTTGTAAAACGAC-CCGCTTCCGCTCCAGAGAGG	64-56
GS-MSH6proA-as	CAGGAAACAGCTATGACC-GCTGGCACACTGGTGGGTAGG	
GS-MSH6proB-s	CACGACGTTGTAAAACGAC-CACCGCCAGCGTGCCAG	64-56
GS-MSH6proB-as	CAGGAAACAGCTATGACC-GCTGTACAGGGTGCTCTGTCGC	
GS-MSH6ex1A-s	CACGACGTTGTAAAACGAC-TCCGTCCGACAGAACGGTTG	64-56
GS-MSH6ex1A-as	CAGGAAACAGCTATGACC-CGTTGAGGTTCTTCGCCTTGG	
GS-MSH6ex1B-s	CACGACGTTGTAAAACGAC-GCTGAGTGATGCCAACAAGGCCTC	64-56
GS-MSH6ex1B-as	CAGGAAACAGCTATGACC-CCAACCCCCTGTGCGAGCCTC	
GS-MSH6ex2-s	CACGACGTTGTAAAACGAC-ACTAAGTTATGTATTTCCTTTTGG	64-56
GS-MSH6ex2-as	CAGGAAACAGCTATGACC-ACAAACACACACACATGGC	
GS-MSH6ex3-s	CACGACGTTGTAAAACGAC-CCCTTATTGTTTATAAATACATTTC	61-53
GS-MSH6ex3-as	CAGGAAACAGCTATGACC-CACCCTAACATAAATAACAACTG	
GS-MSH6ex4A-s	CACGACGTTGTAAAACGAC-AAAAATCATAAGTTGAACTGTCT	61-53
GS-MSH6ex4A-as	CAGGAAACAGCTATGACC-CTTCCTCCTTAGTGTCTGG	
GS-MSH6ex4B-s	CACGACGTTGTAAAACGAC-GATTCTGAGAGTGACATTGGT	61-53
GS-MSH6ex4B-as	CAGGAAACAGCTATGACC-CATCACCACCTCCACTAAC	
GS-MSH6ex4C-s	CACGACGTTGTAAAACGAC-AATTCTGAATCCCAAGCC	61-53
GS-MSH6ex4C-as	CAGGAAACAGCTATGACC-ATCCATGTGGTACAGCTCA	
GS-MSH6ex4D-s	CACGACGTTGTAAAACGAC-ACTTTGATCTTGTCATCTGTTAC	61-53
GS-MSH6ex4D-as	CAGGAAACAGCTATGACC-GTACCCTTGGTAATGATCCTA	
GS-MSH6ex4E-s	CACGACGTTGTAAAACGAC-GCACATATATCCAAGTATGATAGAG	61-53
GS-MSH6ex4E-as	CAGGAAACAGCTATGACC-AGTTTCCTTTGAGAGATTTCC	
GS-MSH6ex4F-s	CACGACGTTGTAAAACGAC-AGGACTCTAGTGGCACACTATC	61-53
GS-MSH6ex4F-as	CAGGAAACAGCTATGACC-CACTTTTCTCTCCTGGTGTC	
GS-MSH6ex4G-s	CACGACGTTGTAAAACGAC-CTTAAAGGTATGACTTCAGAGTCT	61-53
GS-MSH6ex4G-as	CAGGAAACAGCTATGACC-GTAGGGTTCCTTCAGTAGAAC	
GS-MSH6ex4H-s	CACGACGTTGTAAAACGAC-AGATGCAGTGACATTAAACAAC	61-53
GS-MSH6ex4H-as	CAGGAAACAGCTATGACC-TCTGACTCTTCAGGGGAGAC	
GS-MSH6ex4I-s	CACGACGTTGTAAAACGAC-AAGAAGCTTCCAGATCTTGAG	61-53
GS-MSH6ex4I-as	CAGGAAACAGCTATGACC-CAATTCTACAGTCAAATCAGGA	
GS-MSH6ex4Js	CACGACGTTGTAAAACGAC-AGCAGGTCATCTCTCTGCA	61-53
GS-MSH6ex4Jas	CAGGAAACAGCTATGACC-GAATTTCCAGCTGGTAACG	
GS-MSH6ex4K-s	CACGACGTTGTAAAACGAC-TTGGCTGTAGGACCATAGTC	61-53
GS-MSH6ex4K-as	CAGGAAACAGCTATGACC-GGCAAACAGCACTACTTATC	
GS-MSH6ex5A-s	CACGACGTTGTAAAACGAC-AAACACTTAGGCTGATAAAACC	61-53
GS-MSH6ex5A-as	CAGGAAACAGCTATGACC-ACACAATAGGCTTTGCCAT	
GS-MSH6ex5B-s	CACGACGTTGTAAAACGAC-TGGAGATGATTTTATTCCTAATG	61-53
GS-MSH6ex5B-as	CAGGAAACAGCTATGACC-TCCTATTAAGTCACTGGCTGA	
GS-MSH6ex6-s	CACGACGTTGTAAAACGAC-ACTGTTACTACCAGTCATAAAAGAC	61-53
GS-MSH6ex6-as	CAGGAAACAGCTATGACC-ATGACTGAATGAGAACTTAAGTG	
GS-MSH6ex7-s	CACGACGTTGTAAAACGAC-GATGAATTTATGTAATATGATTTGC	61-53
GS-MSH6ex7-as	CAGGAAACAGCTATGACC-TAGTCTTCAAATGAGAAGTTTAATG	
GS-MSH6ex7sek-s	CACGACGTTGTAAAACGACCATTTGTGATTTTTTTTTTTTTAAG	-
GS-MSH6ex8-s	CACGACGTTGTAAAACGAC-TCCTTTGAGTTACTTCCTTATGC	64-56
GS-MSH6ex8-as	CAGGAAACAGCTATGACC-TCTCAAAAAACCGAATTTGTG	
GS-MSH6ex9-s	CACGACGTTGTAAAACGAC-GCACATGTATCGCTAATATTTTTC	61-53
GS-MSH6ex9-as	CAGGAAACAGCTATGACC-ATCCCTTCCCCTTTTACTG	
GS-MSH6ex10-s	CACGACGTTGTAAAACGAC-AACTAACTGACCTTAAGTTTCAAAG	61-53
GS-MSH6ex10-as	CAGGAAACAGCTATGACC-ATTTACCACCTTTGTCAGAAGTC	
GS-MSH6ex10sek-s	CACGACGTTGTAAAACGACTTTTTTTTTTTTTTTTAATTTTAAG	-
GS-MSH6_3'UTR-s	CACGACGTTGTAAAACGAC-AACTGTAGATGCTGAAGCTGTC	61-53
GS-MSH6_3'UTR-as	CAGGAAACAGCTATGACC-TGGGTATAAAACAGCCTGAA	

Primer Name	Primer sequence: universal tail – target specific sequence	Temperature [°C]
MLH1pro-s	CACGACGTTGTAAAACGAC-GCTCCTAAAAACGAACCAATAG	64-56
MLH1pro-as	CAGGAAACAGCTATGACC-AGTGCCTTCAGCCAATCAC	
MLH1ex1-s	CACGACGTTGTAAAACGAC-ACGTGAGCACGAGGCACTG	64-56
MLH1ex1-as	CAGGAAACAGCTATGACC-CGGCCCGTTAAGTCGTAGC	
MLH1ex1_2-s	CACGACGTTGTAAAACGAC-GGCTCACTTAAGGGCTACG	66-62
MLH1ex1_2-as	CAGGAAACAGCTATGACC-AGTGGCTTCCTTACTTAGTTAACG	
MLH1ex2-s	CACGACGTTGTAAAACGAC-ATGTACATTAGAGTAGTTGCAGAC	61-53
MLH1ex2-as	CAGGAAACAGCTATGACC-GAACAGAGAAAGGTCCTGAC	
MLH1ex3-s	CACGACGTTGTAAAACGAC-AGAGATTTGGAAAAATGAGTAAC	61-53
MLH1ex3-as	CAGGAAACAGCTATGACC-ACTAACAAATGACAGACAATGTC	
MLH1ex4-s	CACGACGTTGTAAAACGAC-AAATGGAAGCAGCAGTTC	61-53
MLH1ex4-as	CAGGAAACAGCTATGACC-TGAGACCTAGGCAAAAAATAC	
MLH1ex5-s	CACGACGTTGTAAAACGAC-TTCCCCTTGGGATTAGTAT	61-53
MLH1ex5-as	CAGGAAACAGCTATGACC-ATTTATACAAACAAAGCTTCAAC	
MLH1ex6-s	CACGACGTTGTAAAACGAC-TTTTCAAGTACTTCTATGAATTTAC	61-53
MLH1ex6-as	CAGGAAACAGCTATGACC-GATGACAAATCTCAGAGACC	
MLH1ex7-s	CACGACGTTGTAAAACGAC-GCTCTGACATCTAGTGTGTGTT	61-53
MLH1ex7-as	CAGGAAACAGCTATGACC-ATCATAACCTTATCTCCACCA	
MLH1ex8-s	CACGACGTTGTAAAACGAC-CAATAAATCCTTGTGTCTTCTG	61-53
MLH1ex8-as	CAGGAAACAGCTATGACCA-ATGTGATGGAATGATAAACC	
MLH1ex9-s	CACGACGTTGTAAAACGAC-TTAGTTTATGGGAAGGAACC	61-53
MLH1ex9-as	CAGGAAACAGCTATGACC-GGTGTTTCCTGTGAGTGG	
MLH1ex10-s	CACGACGTTGTAAAACGAC-AATGTACACCTGTGACCTCAC	61-53
MLH1ex10-as	CAGGAAACAGCTATGACC-GAGAGCCTGATAGAACATCTG	
MLH1ex11-s	CACGACGTTGTAAAACGAC-CATATGTGGGCTTTTTCTCC	64-56
MLH1ex11-as	CAGGAAACAGCTATGACC-AAATCTGGGCTCTCACGTC	
MLH1ex12A-s	CACGACGTTGTAAAACGAC-TCTTTCTTAGTACTGCTCCATTTG	61-53
MLH1ex12A-as	CAGGAAACAGCTATGACC-CCGGGAATCTGTACGAAC	
MLH1ex12B-s	CACGACGTTGTAAAACGAC-GATAAGGTCTATGCCCACC	61-53
MLH1ex12B-as	CAGGAAACAGCTATGACC-TTTATTACAGAATAAAGGAGGTAGG	
MLH1ex13-s	CACGACGTTGTAAAACGAC-GCTCCTCCAAAATGCAAC	61-53
MLH1ex13-as	CAGGAAACAGCTATGACC-TTGAGGCCCTATGCATC	
MLH1ex14-s	CACGACGTTGTAAAACGAC-GATTCTATTACTTACCTGTTTTTTGG	61-53
MLH1ex14-as	CAGGAAACAGCTATGACC-TAGCTTTTGTGCCTGTGCTC	
MLH1ex15-s	CACGACGTTGTAAAACGAC-CAACTGGTTGTATCTCAAGC	61-53
MLH1ex15-as	CAGGAAACAGCTATGACC-CTACTATTTTCAGAAACGATCAG	
MLH1ex16-s	CACGACGTTGTAAAACGAC-CTCCGTTAAAGCTTGCTC	61-53
MLH1ex16-as	CAGGAAACAGCTATGACC-AATTTTATTTGAAGAATACAACAG	
MLH1ex17-s	CACGACGTTGTAAAACGAC-AGTAACGTGGTCACCCAG	61-53
MLH1ex17-as	CAGGAAACAGCTATGACC-ACATGCATGTACCGAAATG	
MLH1ex18-s	CACGACGTTGTAAAACGAC-GTAGTCTGTGATCTCCGTTTAG	61-53
MLH1ex18-as	CAGGAAACAGCTATGACC-TTGTATGAGGTCCTGTCCTA	
MLH1ex19-s	CACGACGTTGTAAAACGAC-TGTATGTTGGGATGCAAAC	61-53
MLH1ex19-as	CAGGAAACAGCTATGACC-AACACTTTGTATCGGAATACAG	
MLH1_3'UTR-s	CACGACGTTGTAAAACGAC-CTAAACATTTCACAGAAGATGG	61-53
MLH1_3'UTR-as	CAGGAAACAGCTATGACC-TTTTGGCATCTGAACTGAC	

Primer Name	Primer sequence: universal tail – target specific sequence	Temperature [°C]
GS-PMS2e1S	CACGACGTTGTAAAACGAC-TCTTTGACGTCACGAAGTCG	66-62
GS-PMS2e1AS	CAGGAAACAGCTATGACC-GTTGGAATGCCGTGGGTC	
GS-PMS2e2s	CACGACGTTGTAAAACGAC-TGAAAATATATAGCTGAATACTTGA	61-53
GS-PMS2e2a	CAGGAAACAGCTATGACC-CACATAATAGGTGCTAACTTC	
GS-PMS2e3s	CACGACGTTGTAAAACGAC-GGGTCCGTTTTTAATAAATATG	61-53
GS-PMS2e3a	CAGGAAACAGCTATGACC-AAGACAGTGTTACTCAAAATTCTG	
GS-PMS2e4s	CACGACGTTGTAAAACGAC-ACACTGTCTTGGGAAATG	64-56
GS-PMS2e4a	CAGGAAACAGCTATGACC-CAATTAATTTTCAGAGAGGTTTC	
GS-PMS2e5A-s	CACGACGTTGTAAAACGAC-CCTCAACCATTTAGATCTTGA	64-56
GS-PMS2e5A-a	CAGGAAACAGCTATGACC-TCTGGATAATTTTCCCATTG	
GS-PMS2e5B-s	CACGACGTTGTAAAACGAC-AGGTTGGAACTCGACTGA	64-56
GS-PMS2e5B-a	CAGGAAACAGCTATGACC-CAATAAAGCATTTCTCAATAAT	
GS-PMS2e6s	CACGACGTTGTAAAACGAC-GGATGTTGTAACTTGAGCTGT	64-56
GS-PMS2e6a	CAGGAAACAGCTATGACC-CCCGCTATAATCACTAGAGC	
T-PMS2e7s	CACGACGTTGTAAAACGAC-GTCCACTCTGTCTTTATTAG	61-53
T-PMS2e7a	CAGGAAACAGCTATGACC-AGCTCTCAGGATAAAATGTTC	
T-PMS2e7s-sek	CACGACGTTGTAAAACGAC-TTTTTTTTTTTTCAGTTGC	-
GS-PMS2e8s	CACGACGTTGTAAAACGAC-TAATCCCTTTCACTCTGG	61-53
GS-PMS2e8a	CAGGAAACAGCTATGACC-ACGTAAACTGCCTATTATCAG	
GS-PMS2e9s	CACGACGTTGTAAAACGAC-GGGGCTGGGAACATTTGTC	61-53
GS-PMS2e9a	CAGGAAACAGCTATGACC-ATAGCAGAGCTGTAGAATTTC	
GS-PMS2e10s	CACGACGTTGTAAAACGAC-TTTAAACATAATAAATATGTTTTCTT	61-53
GS-PMS2e10a	CAGGAAACAGCTATGACC-GGAAACACATTAGCTAAAAGC	
GS-PMS2e11A-s	CACGACGTTGTAAAACGAC-ATAGTTTTATTTTGAACATTGAACTG	64-56
GS-PMS2e11A-AS	CAGGAAACAGCTATGACC-CTGGAAATGGACACGTCT	
GS-PMS2e11B-s	CACGACGTTGTAAAACGAC-TCATTAAGGACTGGAGAAGAAAAAAA	66-62
GS-PMS2e11B-a	CAGGAAACAGCTATGACC-CACGGAAGTGCTGCCGT	
GS-PMS2e11C-s	CACGACGTTGTAAAACGAC-GAAAGAGGCAGTGAGTTCCAGTCAC	66-62
GS-PMS2e11C-a	CAGGAAACAGCTATGACC-CGCTTTGTGTTTGGGGTTGC	
GS-PMS2e11D-s	CACGACGTTGTAAAACGAC-GATACCGGATGTAAATTTCG	64-56
GS-PMS2e11D-a	CAGGAAACAGCTATGACC-AAATTTTAGATAAAAAGAGAAAAAGT	
GS-PMS2e12s	CACGACGTTGTAAAACGAC-GTATTGTTTGACTTTTTTTTTATTAC	64-56
GS-PMS2e12a	CAGGAAACAGCTATGACC-TCAAACTCCTGGCCTCTT	
T-PMS2e13s	CACGACGTTGTAAAACGAC-ACTTAGCTGAGTAGTGTTGTTATTTG	61-53
T-PMS2e13a	CAGGAAACAGCTATGACC-AACACCTGAAAGAGAGGAAACT	
T-PMS2e13a-sek	CAGGAAACAGCTATGACC-GGCTGGTCTCAAACTCCTGAC	-
T-PMS2e14s	CACGACGTTGTAAAACGAC-TGTTAGCCAGGATGGTCTGT	64-56
T-PMS2e14a	CAGGAAACAGCTATGACC-GAGTTCAAGGTCACAGAGAACG	
T-PMS2e15s	CACGACGTTGTAAAACGAC-CAAAAAACTACTAAAACGTTGAACC	64-56
T-PMS2e15a	CAGGAAACAGCTATGACC-TGTTTTTTGAGACACAGTCTTGT	
T-PMS2e15a-sek	CAGGAAACAGCTATGACC-TTCATTTTAAAACAAAAAAGGTTAG	-

Primer Name	Primer sequence: Primer A/B – key – MID – universal tail
MID1-S	CGTATCGCCTCCCTCGCGCCA-TCAG-ACGAGTGCGT-CACGACGTTGTAAAACGAC
MID1-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-ACGAGTGCGT-CAGGAAACAGCTATGACC
MID2-S	CGTATCGCCTCCCTCGCGCCA-TCAG-ACGCTCGACA-CACGACGTTGTAAAACGAC
MID2-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-ACGCTCGACA-CAGGAAACAGCTATGACC
MID3-S	CGTATCGCCTCCCTCGCGCCA-TCAG-AGACGCACTC-CACGACGTTGTAAAACGAC
MID3-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-AGACGCACTC-CAGGAAACAGCTATGACC
MID4-S	CGTATCGCCTCCCTCGCGCCA-TCAG-AGCACTGTAG-CACGACGTTGTAAAACGAC
MID4-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-AGCACTGTAG-CAGGAAACAGCTATGACC
MID5-S	CGTATCGCCTCCCTCGCGCCA-TCAG-ATCAGACACG-CACGACGTTGTAAAACGAC
MID5-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-ATCAGACACG-CAGGAAACAGCTATGACC
MID6-S	CGTATCGCCTCCCTCGCGCCA-TCAG-ATATCGCGAG-CACGACGTTGTAAAACGAC
MID6-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-ATATCGCGAG-CAGGAAACAGCTATGACC
MID7-S	CGTATCGCCTCCCTCGCGCCA-TCAG-CGTGTCTCTA-CACGACGTTGTAAAACGAC
MID7-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-CGTGTCTCTA-CAGGAAACAGCTATGACC
MID8-S	CGTATCGCCTCCCTCGCGCCA-TCAG-CTCGCGTGTC-CACGACGTTGTAAAACGAC
MID8-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-CTCGCGTGTC-CAGGAAACAGCTATGACC
MID9-S	CGTATCGCCTCCCTCGCGCCA-TCAG-TAGTATCAGC-CACGACGTTGTAAAACGAC
MID9-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-TAGTATCAGC-CAGGAAACAGCTATGACC
MID10-S	CGTATCGCCTCCCTCGCGCCA-TCAG-TCTCTATGCG-CACGACGTTGTAAAACGAC
MID10-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-TCTCTATGCG-CAGGAAACAGCTATGACC
MID11-S	CGTATCGCCTCCCTCGCGCCA-TCAG-TGATACGTCT-CACGACGTTGTAAAACGAC
MID11-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-TGATACGTCT-CAGGAAACAGCTATGACC
MID12-S	CGTATCGCCTCCCTCGCGCCA-TCAG-TACTGAGCTA-CACGACGTTGTAAAACGAC
MID12-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-TACTGAGCTA-CAGGAAACAGCTATGACC
MID13-S	CGTATCGCCTCCCTCGCGCCA-TCAG-CATAGTAGTG-CACGACGTTGTAAAACGAC
MID13-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-CATAGTAGTG-CAGGAAACAGCTATGACC
MID14-S	CGTATCGCCTCCCTCGCGCCA-TCAG-CGAGAGATAC-CACGACGTTGTAAAACGAC
MID14-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-CGAGAGATAC-CAGGAAACAGCTATGACC
MID15-S	CGTATCGCCTCCCTCGCGCCA-TCAG-ATACGACGTA-CACGACGTTGTAAAACGAC
MID15-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-ATACGACGTA-CAGGAAACAGCTATGACC
MID16-S	CGTATCGCCTCCCTCGCGCCA-TCAG-TCACGTACTA-CACGACGTTGTAAAACGAC
MID16-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-TCACGTACTA-CAGGAAACAGCTATGACC
MID17-S	CGTATCGCCTCCCTCGCGCCA-TCAG-CGTCTAGTAC-CACGACGTTGTAAAACGAC
MID17-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-CGTCTAGTAC-CAGGAAACAGCTATGACC
MID18-S	CGTATCGCCTCCCTCGCGCCA-TCAG-TCTACGTAGC-CACGACGTTGTAAAACGAC
MID18-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-TCTACGTAGC-CAGGAAACAGCTATGACC
MID19-S	CGTATCGCCTCCCTCGCGCCA-TCAG-TGTACTACTC-CACGACGTTGTAAAACGAC
MID19-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-TGTACTACTC-CAGGAAACAGCTATGACC
MID20-S	CGTATCGCCTCCCTCGCGCCA-TCAG-ACGACTACAG-CACGACGTTGTAAAACGAC
MID20-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-ACGACTACAG-CAGGAAACAGCTATGACC
MID21-S	CGTATCGCCTCCCTCGCGCCA-TCAG-CGTAGACTAG-CACGACGTTGTAAAACGAC
MID21-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-CGTAGACTAG-CAGGAAACAGCTATGACC
MID22-S	CGTATCGCCTCCCTCGCGCCA-TCAG-TACGAGTATG-CACGACGTTGTAAAACGAC
MID22-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-TACGAGTATG-CAGGAAACAGCTATGACC
MID23-S	CGTATCGCCTCCCTCGCGCCA-TCAG-TACTCTCGTG-CACGACGTTGTAAAACGAC
MID23-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-TACTCTCGTG-CAGGAAACAGCTATGACC
MID24-S	CGTATCGCCTCCCTCGCGCCA-TCAG-TAGAGACGAG-CACGACGTTGTAAAACGAC
MID24-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-TAGAGACGAG-CAGGAAACAGCTATGACC
MID25-S	CGTATCGCCTCCCTCGCGCCA-TCAG-TCGTCGCTCG-CACGACGTTGTAAAACGAC
MID25-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-TCGTCGCTCG-CAGGAAACAGCTATGACC

Primer Name	Primer sequence: Primer A/B – key – MID – universal tail
MID26-S	CGTATCGCCTCCCTCGCGCCA-TCAG-ACATACGCGT-CACGACGTTGTAAAACGAC
MID26-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-ACATACGCGT-CAGGAAACAGCTATGACC
MID27-S	CGTATCGCCTCCCTCGCGCCA-TCAG-ACGCGAGTAT-CACGACGTTGTAAAACGAC
MID27-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-ACGCGAGTAT-CAGGAAACAGCTATGACC
MID28-S	CGTATCGCCTCCCTCGCGCCA-TCAG-ACTACTATGT-CACGACGTTGTAAAACGAC
MID28-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-ACTACTATGT-CAGGAAACAGCTATGACC
MID29-S	CGTATCGCCTCCCTCGCGCCA-TCAG-ACTGTACAGT-CACGACGTTGTAAAACGAC
MID29-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-ACTGTACAGT-CAGGAAACAGCTATGACC
MID30-S	CGTATCGCCTCCCTCGCGCCA-TCAG-AGACTATACT-CACGACGTTGTAAAACGAC
MID30-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-AGACTATACT-CAGGAAACAGCTATGACC
MID31-S	CGTATCGCCTCCCTCGCGCCA-TCAG-AGCGTCGTCT-CACGACGTTGTAAAACGAC
MID31-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-AGCGTCGTCT-CAGGAAACAGCTATGACC
MID32-S	CGTATCGCCTCCCTCGCGCCA-TCAG-AGTACGCTAT-CACGACGTTGTAAAACGAC
MID32-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-AGTACGCTAT-CAGGAAACAGCTATGACC
MID33-S	CGTATCGCCTCCCTCGCGCCA-TCAG-ATAGAGTACT-CACGACGTTGTAAAACGAC
MID33-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-ATAGAGTACT-CAGGAAACAGCTATGACC
MID34-S	CGTATCGCCTCCCTCGCGCCA-TCAG-CACGCTACGT-CACGACGTTGTAAAACGAC
MID34-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-CACGCTACGT-CAGGAAACAGCTATGACC
MID35-S	CGTATCGCCTCCCTCGCGCCA-TCAG-CAGTAGACGT-CACGACGTTGTAAAACGAC
MID35-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-CAGTAGACGT-CAGGAAACAGCTATGACC
MID36-S	CGTATCGCCTCCCTCGCGCCA-TCAG-CGACGTGACT-CACGACGTTGTAAAACGAC
MID36-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-CGACGTGACT-CAGGAAACAGCTATGACC
MID37-S	CGTATCGCCTCCCTCGCGCCA-TCAG-TACACACACT-CACGACGTTGTAAAACGAC
MID37-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-TACACACACT-CAGGAAACAGCTATGACC
MID38-S	CGTATCGCCTCCCTCGCGCCA-TCAG-TACACGTGAT-CACGACGTTGTAAAACGAC
MID38-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-TACACGTGAT-CAGGAAACAGCTATGACC
MID39-S	CGTATCGCCTCCCTCGCGCCA-TCAG-TACAGATCGT-CACGACGTTGTAAAACGAC
MID39-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-TACAGATCGT-CAGGAAACAGCTATGACC
MID40-S	CGTATCGCCTCCCTCGCGCCA-TCAG-TACGCTGTCT-CACGACGTTGTAAAACGAC
MID40-AS	CTATGCGCCTTGCCAGCCCGC-TCAG-TACGCTGTCT-CAGGAAACAGCTATGACC

6.3 Variant Classification System

Class	Description	Probability of Being Pathogenic
5	Definitely pathogenic	>0.99
4	Likely pathogenic	0.95-0.99
3	Uncertain	0.05-0.949
2	Likely not pathogenic or of little clinical significance	0.001-0.049
1	Not pathogenic or of no clinical significance	<0.001

Adapted from Plon *et al.*, 2008[70].

6.4 Optimized Multiplexing Pools

Amp, Amplicon.

Pool 1 (MSH2)		Pool 2 (MSH2)		Pool 3 (MSH6)		Pool 4 (MSH6)		Pool 5 (MLH1)		Pool 6 (MLH1)		Pool 7 (PMS2)		Pool 8 (PMS2)	
Amp	**Vol [µL]**	**Amp**	**Vol [µL]**	**Amp**	**Vol [µL]**	**Amp**	**Vol [µL]**	**Amp**	**Vol [µL]**	**Amp**	**Vol [µL]**	**Amp**	**Vol [µL]**	**Amp**	**Vol [µL]**
ProA	2,5	Ex3A	7,5	ProA	7,5	Ex4F	5	Pro	2,5	Ex4	1,5	Ex1	1	Ex4	1
ProB	5	Ex3B	5	ProB	20	Ex4G	1,5	Ex1	2,5	Ex5	3,5	Ex2	1,5	Ex5A	1
Ex1A	2,5	Ex4	7,5	Ex1A	5	Ex4H	2,5	Ex1_2	2,5	Ex6	2,5	Ex3	2,5	Ex5B	3,5
Ex1B	3,5	Ex7	15	Ex1B	7,5	Ex4I	1,5	Ex2	1	Ex8	5	Ex8	1	Ex6	3,5
Ex6	1,5	Ex8	7,5	Ex2	2,5	Ex4J	2,5	Ex3	2,5	Ex9	10	Ex9	1	Ex11A	1
Ex14	5	Ex9	1,5	Ex3	5	Ex4K	15	Ex16	2,5	Ex10	3,5	Ex10	1	Ex11D	20
Ex15	1,5	Ex10	7,5	Ex4A	5	Ex5A	15	Ex17	1	Ex12A	20	Ex11B	3,5	Ex12	7,5
Ex16	1,5	Ex11	10	Ex4B	5	Ex5B	5	Ex18	1,5	Ex12B	20	Ex11C	1	Water	32,5
3`UTR	12,5	Ex12	5	Ex4C	3,5	Ex6	1,5	Ex19	1	Ex13	1,5	Water	67,5	Total	70
Water	54,5	Ex13	5	Ex4D	3,5	Ex9	5	Ex11	1,5	Ex14	2,5	Total	80		
Total	90	Water	28,5	Ex4E	7,5	3`UTR	20	Water	81,5	Ex15	1,5				
		Total	100	Ex8	2,5	Water	35,5	Total	100	Water	38,5				
				Water	45,5	Total	110			Total	110				
				Total	120										

6.5 Coverage Optimization of *PMS2* Library

Amplicons (Amp) are given in the columns and samples (MID1 – 40) are given in the row.

PMS2 - GS Junior run 1:

Amp/ Sample	1	2	3	4	5A	5B	6	8	9	10	11A	11B	11C	11D	12
MID1	3	11	27	51	164	47	19	156	195	63	96	38	85	0	4
MID2	6	0	45	51	286	54	41	242	255	111	177	81	133	3	19
MID3	0	14	31	54	173	37	38	199	242	108	115	43	93	0	11
MID4	0	3	37	64	230	31	8	177	226	59	293	59	98	0	9
MID5	5	13	28	69	221	50	23	196	251	131	123	36	89	3	4
MID6	5	0	32	102	202	67	15	125	126	19	145	63	109	4	12
MID7	11	0	45	169	360	78	22	187	232	139	152	80	149	0	10
MID8	0	11	27	63	223	30	29	151	162	46	134	60	128	3	5
MID9	3	0	20	57	327	65	51	176	207	103	148	72	177	0	19
MID10	3	13	68	6	230	44	40	214	245	165	148	46	77	5	20
MID11	0	0	23	0	247	9	9	147	212	48	189	57	123	0	3
MID12	3	4	29	8	205	42	19	167	244	123	50	45	75	0	8
MID13	7	0	22	7	185	31	8	130	152	107	158	47	83	0	7
MID14	3	0	11	11	258	45	7	117	130	68	173	59	110	4	0
MID15	0	0	25	16	235	34	12	175	207	72	136	47	124	5	8
MID16	6	0	23	8	184	32	22	141	153	83	93	79	112	3	0
MID17	6	15	33	55	185	37	46	224	192	91	92	44	129	4	7
MID18	3	0	36	47	290	55	34	207	183	75	130	64	156	3	4
MID19	4	0	27	21	249	51	14	116	145	62	140	58	136	3	0
MID20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MID21	10	0	57	48	467	65	43	274	266	118	310	104	233	11	6
MID22	4	17	65	14	246	23	27	230	268	98	157	48	130	0	3
MID23	9	32	60	30	300	44	67	289	279	170	161	58	118	13	24
MID24	5	0	54	33	281	40	54	124	238	93	152	121	154	4	19
MID25	10	21	66	58	224	54	44	202	224	132	122	53	126	4	16
MID26	3	23	86	41	210	40	47	209	268	126	135	45	90	5	23
MID27	5	17	71	17	261	46	42	232	260	110	140	58	113	6	11
MID28	8	21	96	38	289	58	53	275	258	179	192	48	132	4	15
MID29	0	41	76	35	270	54	40	191	251	130	157	60	120	5	10
MID30	6	0	49	51	336	33	29	186	167	106	179	59	117	0	17
MID31	3	17	70	49	247	65	69	256	274	133	142	70	140	0	17
MID32	10	0	68	32	216	25	68	182	162	80	123	58	111	3	10
MID33	7	0	42	4	271	19	0	232	275	77	205	70	192	0	13
MID34	5	0	67	23	271	33	0	243	257	150	203	50	130	0	15
MID35	0	15	41	20	213	32	0	268	289	177	270	46	123	0	0
MID36	3	22	80	33	253	13	0	305	339	171	325	46	101	0	0
MID37	0	0	40	13	233	14	0	17	245	89	288	40	82	0	0
MID38	0	0	57	12	247	22	0	208	263	159	292	44	91	0	0
MID39	3	0	83	31	280	55	0	245	289	177	219	58	120	0	0
MID40	0	22	108	22	233	22	0	266	335	146	186	58	80	5	14

PMS2 - GS Junior run 2:

Amp/ Sample	1	2	3	4	5A	5B	6	8	9	10	11A	11B	11C	11D	12
MID1	227	52	86	38	263	102	85	153	130	106	101	94	75	31	21
MID2	143	18	73	30	189	66	60	150	137	83	75	68	66	24	11
MID3	155	19	75	29	186	78	69	135	130	79	99	69	64	27	9
MID4	181	25	98	19	225	43	49	122	161	74	101	76	79	16	19
MID5	175	19	93	28	205	32	61	101	148	67	95	37	85	12	20
MID6	150	27	83	30	167	59	52	139	112	73	68	76	80	22	15
MID7	193	33	89	27	239	51	100	118	145	93	98	63	70	25	19
MID8	141	11	79	35	157	76	57	117	121	87	67	56	75	40	11
MID9	288	23	166	40	371	74	85	262	283	90	148	133	111	11	25
MID10	129	47	175	44	231	90	85	152	155	58	86	82	54	35	12
MID11	165	51	78	45	237	109	79	148	139	59	89	79	67	20	18
MID12	149	76	84	46	294	101	91	132	133	59	95	63	67	26	16
MID13	163	57	97	40	274	110	90	149	153	78	128	64	91	35	19
MID14	163	50	81	31	284	73	74	107	120	53	102	63	53	24	19
MID15	122	25	53	31	265	129	51	112	109	48	66	42	54	27	20
MID16	147	14	81	40	175	74	54	148	132	63	53	63	60	16	0
MID17	249	22	108	64	300	39	84	155	199	66	124	129	102	16	19
MID18	186	16	94	85	285	57	75	147	197	71	101	100	75	13	10
MID19	179	16	79	51	237	40	65	131	145	78	89	69	66	14	9
MID20	204	38	75	57	327	69	105	108	166	72	138	92	81	24	19
MID21	191	12	98	74	225	58	66	165	175	108	88	82	58	19	12
MID22	128	19	137	62	284	65	129	115	160	89	87	46	50	29	19
MID23	146	22	163	56	184	36	96	131	153	89	71	62	41	17	16
MID24	135	38	136	47	323	76	124	128	169	94	120	71	57	24	23
MID25	205	105	169	50	299	87	144	150	179	103	127	87	70	20	21
MID26	185	54	175	50	231	56	96	140	188	94	96	91	52	17	25
MID27	165	40	168	62	231	47	92	164	205	78	113	87	71	6	28
MID28	164	50	185	47	374	43	100	155	227	119	109	92	48	11	11
MID29	163	70	158	55	379	86	98	141	188	95	147	62	57	14	22
MID30	162	94	136	49	286	80	117	127	154	103	104	49	47	14	25
MID31	246	48	161	93	368	96	128	137	180	114	128	87	88	19	35
MID32	143	32	145	53	282	49	103	131	189	85	92	65	63	20	24
MID33	215	79	133	87	510	162	123	219	243	110	157	126	72	34	40
MID34	129	65	144	62	363	129	114	136	148	81	147	55	27	27	30
MID35	172	58	138	62	290	115	103	166	183	74	112	62	53	22	13
MID36	186	34	144	44	226	89	102	124	148	93	83	71	63	23	13
MID37	160	76	174	76	289	123	125	157	198	102	120	74	61	30	22
MID38	141	62	134	47	287	100	106	105	143	67	117	65	28	29	30
MID39	108	39	227	50	323	86	122	172	287	137	117	60	42	34	34
MID40	177	35	163	72	266	76	114	141	181	91	86	64	59	22	12

PMS2 - GS Junior run 3:

Amp/ Sample	1	2	3	4	5A	5B	6	8	9	10	11A	11B	11C	11D	12
MID1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MID2	187	247	136	220	148	113	123	170	168	98	183	113	116	51	139
MID3	148	216	116	211	122	95	116	152	131	80	169	99	104	21	109
MID4	198	297	131	254	180	135	157	240	158	105	243	123	152	47	153
MID5	197	262	127	196	132	105	121	212	151	90	174	125	120	49	122
MID6	194	332	161	359	255	205	191	199	178	137	291	157	151	65	223
MID7	167	256	134	321	208	174	182	190	149	109	278	97	100	77	191
MID8	179	243	147	259	170	129	146	201	128	99	233	129	101	53	167
MID9	248	364	188	287	207	211	195	239	214	145	371	180	212	69	187
MID10	200	324	294	204	103	94	129	227	218	160	180	146	183	31	164
MID11	186	300	137	185	161	120	126	173	181	104	248	154	163	23	145
MID12	301	443	185	260	183	179	134	332	294	184	281	183	205	46	219
MID13	287	429	206	239	180	121	134	294	329	182	233	232	257	38	176
MID14	130	204	102	139	85	84	80	116	115	60	102	107	118	18	84
MID15	205	282	129	270	211	215	154	245	180	146	297	166	170	64	192
MID16	187	231	124	240	182	182	164	183	136	103	280	126	183	44	166
MID17	149	269	86	371	271	148	218	145	117	58	409	126	111	51	280
MID18	225	279	112	184	116	96	98	186	153	70	140	148	123	17	134
MID19	161	217	92	220	178	89	132	160	104	69	212	102	93	17	121
MID20	308	332	26	242	188	43	113	278	199	61	201	234	237	8	98
MID21	181	328	134	464	234	169	263	214	141	87	353	142	143	27	233
MID22	356	0	0	0	0	0	0	0	6	0	0	3	0	0	0
MID23	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0
MID24	250	278	30	315	189	76	117	179	145	35	286	143	172	4	121
MID25	369	373	28	305	225	210	105	296	173	70	351	232	248	15	174
MID26	357	337	38	376	226	220	136	291	212	51	359	197	231	0	176
MID27	200	485	479	228	147	153	252	379	259	193	215	164	127	20	219
MID28	236	422	375	276	179	164	206	419	365	207	250	190	185	27	178
MID29	284	118	141	260	325	46	341	137	255	59	380	218	228	3	210
MID30	375	452	165	459	292	235	281	327	307	102	474	213	284	19	270
MID31	0	0	0	0	0	0	5	0	0	0	3	0	0	0	0
MID32	286	236	198	222	175	105	199	237	222	100	256	183	208	22	156
MID33	342	292	66	275	212	168	124	272	207	91	322	141	203	34	164
MID34	190	339	289	219	108	112	219	247	192	132	259	118	149	47	241
MID35	223	356	345	221	209	117	293	285	177	153	330	118	139	42	306
MID36	428	444	96	345	286	154	139	318	275	108	442	227	225	12	174
MID37	290	249	121	577	448	170	191	388	273	125	637	171	150	23	145
MID38	494	543	65	662	535	360	210	457	322	122	772	339	338	49	277
MID39	370	377	42	446	364	169	112	313	212	88	502	223	249	25	156
MID40	315	210	26	281	4	115	79	228	152	44	368	223	196	18	112

6.6 Variants found in PMS2 GS Junior Runs

Variants found in the amplicons solely analyzed by GS Junior sequencing or Sanger sequencing are indicated by the word "GS Junior" and "Sanger", respectively. HET, heterozygous; HOM, homozygous.

DNA	Protein	Class	HET	HOM	Agreement with Sanger
c.-195T>C		1	3	1	yes
c.-154C>G		1	29	0	yes
c.-93G>C		3	1	0	yes
c.23+10G>C		2	3	0	yes
c.52A>G	p.Ile18Val	3	1	0	yes
c.59G>A	p.Arg20Gln	1	17	3	yes
c.86G>C	p.Gly29Ala	3	1	0	yes
c.251-72A>G		1	7	0	yes
c.251-20T>G		3	2	0	yes
c.255G>C	c.Leu85Leu	3	1	0	yes
c.288C>T	p.Ala96Ala	1	7	0	GS Junior
c.705+17A>G		1	56	15	yes
c.780C>G	p.Ser260Ser	1	33	82	Sanger
c.1145-116C>T		2	1	0	yes
c.1408C>T	p.Pro470Ser	1	59	11	GS Junior
c.1454C>A	p.Thr485Lys	1	14	1	GS Junior
c.1531A>G	p.Thr511Ala	1	12	0	GS Junior
c.1569C>G	p.Ser523Ser	2	1	0	yes
c.1621G>A	p.Glu541Lys	1	18	2	yes
c.1688G>T	p.Arg563Leu	2	5	0	yes
c.1789A>T	p.Thr597Ser	2	3	1	yes
c.1866G>A	p.Met622Ile	2	11	0	yes
c.2006+6G>A		1	11	1	GS Junior
c.2007-7C>T		1	18	4	yes
c.2007-4G>A		1	31	1	yes
c.2275+117G>C		3	2	0	Sanger
c.2275+126G>C		2	6	3	Sanger
c.2275+169G>C		1	5	0	Sanger
c.2275+169delG		1	4	0	Sanger
c.2275+234G>C		3	1	0	Sanger
c.2276-135T>C		1	28	3	Sanger
c.2445+30A>G		3	1	0	Sanger
c.2466T>C	p.Leu822Leu	1	53	0	Sanger
c.2570G>C	p.Gly857Ala	1	40	0	Sanger
c.2589+17G>A		1	24	0	Sanger
c.2589+92insA		1	110	7	Sanger
Gene conversion exon 13-14		3	5	0	Sanger

6.7 Coverage Optimization of MMR Genes Library

Amplicons (Amp) are given in the columns and samples (MID1 – 40) are given in the row.

MMR Genes - GS Junior Run 1:

Amplicon	MID1	MID2	MID3	MID4	MID5	MID6	MID7	MID8
MSH2_ProA	127	89	113	109	106	102	79	96
MSH2_ProB	92	84	128	204	215	300	131	99
MSH2_ex1A	143	89	133	133	113	84	101	0
MSH2_ex1B	213	197	184	213	219	166	201	150
MSH2_ex3A	269	196	233	245	211	189	211	246
MSH2_ex3B	110	113	139	140	139	112	126	128
MSH2_ex4	218	202	271	307	184	179	138	252
MSH2_ex6	213	120	146	191	124	110	126	132
MSH2_ex7	298	226	286	308	231	214	253	258
MSH2_ex8	206	154	172	176	164	148	171	194
MSH2_ex9	146	147	179	233	181	146	183	122
MSH2_ex10	220	277	312	281	270	240	209	247
MSH2_ex11	281	187	271	263	261	272	239	245
MSH2_ex12	317	221	284	274	214	210	240	201
MSH2_ex13	184	146	204	230	246	156	257	163
MSH2_ex14	259	128	197	190	184	153	185	139
MSH2_ex15	213	87	86	102	123	118	106	108
MSH2_ex16	170	114	113	139	172	123	131	114
MSH2_3'UTR	188	154	182	221	182	154	171	205
MSH6_ProA	38	74	88	113	107	101	73	71
MSH6_ProB	41	45	33	73	55	44	12	46
MSH6_ex1A	43	88	94	130	105	100	104	95
MSH6_ex1B	56	77	83	123	87	78	77	90
MSH6_ex2	66	58	86	82	102	85	123	75
MSH6_ex3	155	261	311	367	312	309	282	326
MSH6_ex4A	327	182	289	163	127	139	322	153
MSH6_ex4B	280	164	193	189	197	157	141	144
MSH6_ex4C	377	139	155	195	182	202	167	145
MSH6_ex4D	268	223	246	274	286	255	219	187
MSH6_ex4E	83	117	132	171	147	155	132	146
MSH6_ex4F	484	178	183	219	198	195	283	165
MSH6_ex4G	163	610	431	465	546	782	161	133
MSH6_ex4H	146	201	302	141	100	48	104	121
MSH6_ex4I	303	205	200	305	219	350	274	274
MSH6_ex4J	187	147	143	181	213	188	160	123
MSH6_ex4K	186	57	69	127	128	88	173	179
MSH6_ex5A	158	63	80	135	128	83	106	141
MSH6_ex5B	142	60	71	125	105	79	140	125
MSH6_ex6	135	63	78	153	134	78	122	94
MSH6_ex8	201	94	80	168	164	136	179	138
MSH6_ex9	357	120	81	251	234	147	256	232
MSH6_3'UTR	195	87	89	149	128	98	165	116

Amplicon	MID1	MID2	MID3	MID4	MID5	MID6	MID7	MID8
MLH1_Pro	123	94	117	148	140	115	95	96
MLH1_ex1	91	79	120	138	103	135	89	107
MLH1_ex1_2	137	141	169	102	109	100	127	110
MLH1_ex2	141	106	158	162	122	171	122	124
MLH1_ex3	217	160	216	237	192	254	231	213
MLH1_ex4	298	378	309	467	442	309	363	251
MLH1_ex5	283	219	207	264	184	250	227	188
MLH1_ex6	172	123	171	200	161	155	164	120
MLH1_ex8	380	278	377	352	256	333	309	312
MLH1_ex9	161	155	162	193	148	138	166	168
MLH1_ex10	412	272	377	434	344	367	336	323
MLH1_ex11	273	176	198	182	207	181	187	155
MLH1_ex12A	18	21	22	25	37	26	41	38
MLH1_ex12B	68	56	50	55	77	87	67	55
MLH1_ex13	288	185	186	190	204	176	169	161
MLH1_ex14	151	116	178	192	157	131	184	156
MLH1_ex15	178	191	226	711	233	194	201	145
MLH1_ex16	344	284	292	307	373	379	354	262
MLH1_ex17	261	173	197	200	232	182	114	107
MLH1_ex18	303	131	185	184	182	198	154	138
MLH1_ex19	737	603	791	668	774	728	705	624
PMS2_ex1	205	223	302	331	318	216	354	264
PMS2_ex2	200	142	195	216	175	151	138	148
PMS2_ex3	260	218	244	263	268	241	211	236
PMS2_ex4	191	161	212	210	175	168	205	128
PMS2_ex5A	176	99	181	194	194	168	179	150
PMS2_ex5B	256	192	206	214	170	164	186	164
PMS2_ex6	197	175	196	240	187	199	219	117
PMS2_ex8	344	217	245	311	281	257	240	249
PMS2_ex9	331	261	325	426	347	305	245	221
PMS2_ex10	225	191	257	314	242	253	248	172
PMS2_ex11A	209	176	206	243	185	174	220	116
PMS2_ex11B	271	175	252	279	230	218	218	186
PMS2_ex11C	349	269	258	322	334	290	303	273
PMS2_ex11D	124	110	72	78	111	103	178	138
PMS2_ex12	260	174	183	186	191	183	232	201

MMR Genes - GS Junior Run 2:

Amplicon	MID1	MID2	MID3	MID4	MID5	MID6	MID7	MID8
MSH2_ProA	121	83	143	111	118	106	140	148
MSH2_ProB	58	65	82	109	70	97	68	61
MSH2_ex1A	85	92	117	135	107	102	89	99
MSH2_ex1B	121	125	120	122	148	119	116	128
MSH2_ex3A	200	133	113	151	118	129	171	134
MSH2_ex3B	83	58	83	78	60	72	78	66
MSH2_ex4	134	119	123	130	111	93	134	107
MSH2_ex6	132	134	162	178	156	187	140	112
MSH2_ex7	24	227	119	186	146	149	169	212
MSH2_ex8	165	126	126	124	134	130	146	111
MSH2_ex9	102	76	157	125	143	91	143	116
MSH2_ex10	185	148	152	125	116	190	159	157
MSH2_ex11	208	133	158	189	176	161	218	213
MSH2_ex12	117	91	77	138	159	124	135	124
MSH2_ex13	118	121	78	114	137	89	127	103
MSH2_ex14	137	127	133	185	175	181	179	168
MSH2_ex15	117	82	133	117	92	89	136	79
MSH2_ex16	146	120	157	166	174	138	198	136
MSH2_3'UTR	190	210	214	175	190	175	224	230
MSH6_ProA	124	136	131	153	156	113	145	135
MSH6_ProB	115	139	159	151	168	152	135	153
MSH6_ex1A	97	70	68	71	71	79	65	110
MSH6_ex1B	87	90	98	128	105	87	105	101
MSH6_ex2	83	87	95	87	92	70	72	72
MSH6_ex3	159	169	142	177	143	148	168	130
MSH6_ex4A	163	158	137	164	139	173	180	170
MSH6_ex4B	117	107	126	95	88	99	100	88
MSH6_ex4C	103	107	101	112	110	95	128	91
MSH6_ex4D	116	105	185	112	115	93	105	100
MSH6_ex4E	165	120	120	159	117	141	139	108
MSH6_ex4F	167	151	143	105	92	269	243	181
MSH6_ex4G	46	80	122	121	232	150	198	61
MSH6_ex4H	132	109	212	99	469	235	255	90
MSH6_ex4I	146	112	309	712	153	341	354	95
MSH6_ex4J	75	91	187	68	179	205	93	128
MSH6_ex4K	174	118	71	73	51	58	67	125
MSH6_ex5A	143	106	62	67	57	36	59	152
MSH6_ex5B	110	76	74	61	56	42	60	121
MSH6_ex6	91	102	45	67	45	34	65	89
MSH6_ex8	87	79	17	96	98	81	88	71
MSH6_ex9	178	155	142	96	102	50	81	195
MSH6_3'UTR	228	157	123	129	89	67	152	185

Amplicon	MID1	MID2	MID3	MID4	MID5	MID6	MID7	MID8
MLH1_Pro	90	50	74	127	118	141	161	145
MLH1_ex1	92	59	68	98	115	120	136	173
MLH1_ex1_2	248	82	112	105	117	152	144	147
MLH1_ex2	311	544	218	143	223	205	290	206
MLH1_ex3	137	92	82	158	174	200	206	336
MLH1_ex4	116	390	195	610	353	389	187	75
MLH1_ex5	149	64	98	70	92	82	103	108
MLH1_ex6	90	68	76	76	95	82	87	56
MLH1_ex8	111	85	95	101	110	105	105	94
MLH1_ex9	118	102	134	99	99	82	107	120
MLH1_ex10	137	86	107	75	88	88	108	107
MLH1_ex11	148	290	221	261	113	130	306	123
MLH1_ex12A	72	53	46	48	52	61	58	60
MLH1_ex12B	26	24	58	28	25	30	28	32
MLH1_ex13	109	213	61	69	57	67	73	72
MLH1_ex14	118	154	143	84	111	96	168	100
MLH1_ex15	78	41	95	57	66	53	82	68
MLH1_ex16	165	69	114	151	169	230	230	243
MLH1_ex17	198	619	818	873	736	377	357	223
MLH1_ex18	298	63	164	151	174	212	137	221
MLH1_ex19	320	13	161	81	120	111	161	320
PMS2_ex1	163	196	202	149	170	184	144	141
PMS2_ex2	228	166	215	198	158	198	192	228
PMS2_ex3	198	201	239	199	228	215	243	184
PMS2_ex4	140	128	121	158	154	112	133	132
PMS2_ex5A	114	132	141	171	148	167	166	123
PMS2_ex5B	120	84	102	112	77	79	158	112
PMS2_ex6	167	131	165	189	175	213	191	156
PMS2_ex8	286	138	194	215	214	240	268	224
PMS2_ex9	192	193	282	291	269	268	274	251
PMS2_ex10	135	101	161	135	152	129	146	138
PMS2_ex11A	120	115	177	192	148	151	143	111
PMS2_ex11B	240	222	281	231	232	213	243	229
PMS2_ex11C	88	71	17	81	80	121	78	64
PMS2_ex11D	107	158	80	61	111	134	151	139
PMS2_ex12	90	87	94	83	98	130	132	85

MMR Genes - GS Junior Run 3:

Amplicon	MID1	MID2	MID3	MID4	MID5	MID6	MID7	MID8
MSH2_ProA	82	74	89	94	113	101	109	77
MSH2_ProB	54	72	58	72	55	64	79	37
MSH2_ex1A	79	114	73	108	89	48	88	55
MSH2_ex1B	86	88	88	111	124	112	101	89
MSH2_ex3A	72	111	110	132	103	107	133	55
MSH2_ex3B	61	66	72	56	67	72	80	41
MSH2_ex4	57	76	72	47	103	63	72	64
MSH2_ex6	134	29	137	124	114	121	171	85
MSH2_ex7	70	110	117	128	103	104	101	78
MSH2_ex8	63	99	77	95	85	89	101	87
MSH2_ex9	75	76	73	92	103	94	110	62
MSH2_ex10	103	119	154	155	133	153	174	91
MSH2_ex11	142	119	144	126	138	129	125	97
MSH2_ex12	66	81	76	118	86	114	98	64
MSH2_ex13	64	88	70	101	120	80	106	98
MSH2_ex14	141	172	148	176	177	170	186	121
MSH2_ex15	96	83	83	104	109	90	109	79
MSH2_ex16	124	124	135	192	112	143	125	91
MSH2_3'UTR	123	166	152	155	149	134	179	120
MSH6_ProA	68	108	99	59	71	72	89	69
MSH6_ProB	58	104	78	54	70	14	87	64
MSH6_ex1A	67	59	73	68	50	46	76	55
MSH6_ex1B	66	70	57	57	67	64	64	62
MSH6_ex2	69	46	77	68	55	53	94	32
MSH6_ex3	118	129	111	128	120	112	147	75
MSH6_ex4A	101	127	91	125	117	102	186	127
MSH6_ex4B	83	81	83	100	86	87	86	67
MSH6_ex4C	88	94	106	93	79	92	127	67
MSH6_ex4D	86	85	102	90	83	66	108	38
MSH6_ex4E	138	125	106	120	143	104	153	78
MSH6_ex4F	151	144	126	141	148	126	177	99
MSH6_ex4G	47	59	50	84	76	71	80	55
MSH6_ex4H	85	101	113	120	90	91	126	57
MSH6_ex4I	57	50	56	42	45	39	57	39
MSH6_ex4J	107	93	83	103	70	101	89	68
MSH6_ex4K	138	158	142	160	145	171	174	112
MSH6_ex5A	97	66	60	88	50	62	75	49
MSH6_ex5B	101	102	117	109	78	96	119	66
MSH6_ex6	71	116	98	101	98	85	106	61
MSH6_ex8	85	74	0	79	73	75	0	63
MSH6_ex9	127	159	148	154	127	104	166	118
MSH6_3'UTR	115	181	117	152	156	156	155	128

Amplicon	MID1	MID2	MID3	MID4	MID5	MID6	MID7	MID8
MLH1_Pro	136	142	131	155	116	186	142	118
MLH1_ex1	123	122	128	130	128	138	128	111
MLH1_ex1_2	179	117	110	126	115	0	181	89
MLH1_ex2	128	85	123	125	116	131	171	75
MLH1_ex3	230	218	242	262	281	278	319	208
MLH1_ex4	42	51	57	59	64	50	61	44
MLH1_ex5	100	98	82	92	94	99	122	79
MLH1_ex6	62	63	93	91	76	81	66	41
MLH1_ex8	94	106	106	86	87	90	92	52
MLH1_ex9	56	69	91	100	102	92	97	70
MLH1_ex10	97	95	87	106	74	77	111	71
MLH1_ex11	90	99	102	142	95	103	109	81
MLH1_ex12A	83	87	84	89	99	75	79	58
MLH1_ex12B	32	46	43	25	39	33	42	34
MLH1_ex13	36	54	43	57	62	43	49	44
MLH1_ex14	78	67	86	84	89	91	98	69
MLH1_ex15	59	43	62	77	92	65	65	33
MLH1_ex16	261	228	228	229	219	244	275	155
MLH1_ex17	160	110	139	151	118	137	121	107
MLH1_ex18	204	210	235	249	224	241	262	153
MLH1_ex19	92	107	107	117	109	79	117	60
PMS2_ex1	216	127	163	303	171	132	0	87
PMS2_ex2	134	181	185	169	155	132	180	146
PMS2_ex3	130	151	151	200	173	132	248	129
PMS2_ex4	103	106	106	98	96	92	118	94
PMS2_ex5A	119	85	119	170	118	156	138	100
PMS2_ex5B	166	161	107	123	85	160	152	90
PMS2_ex6	128	144	179	145	161	177	183	117
PMS2_ex8	166	180	188	221	204	117	295	153
PMS2_ex9	243	255	232	270	257	219	326	175
PMS2_ex10	120	115	111	134	125	121	198	88
PMS2_ex11A	127	130	184	191	150	188	193	109
PMS2_ex11B	153	199	173	190	166	167	237	132
PMS2_ex11C	102	84	108	75	76	62	111	70
PMS2_ex11D	116	132	110	141	219	73	148	135
PMS2_ex12	69	111	102	129	105	122	105	94

MMR Genes - GS Junior Run 4:

Amplicon	MID1	MID2	MID3	MID4	MID5	MID6	MID7	MID8
MSH2_ProA	122	87	129	115	124	123	143	113
MSH2_ProB	63	72	58	83	64	91	70	49
MSH2_ex1A	130	98	105	97	135	130	121	92
MSH2_ex1B	108	128	111	122	101	106	121	100
MSH2_ex3A	113	112	143	123	94	110	143	123
MSH2_ex3B	78	78	82	75	71	67	107	71
MSH2_ex4	89	92	104	81	86	125	208	51
MSH2_ex6	132	138	160	143	123	150	149	158
MSH2_ex7	79	123	100	132	81	105	154	83
MSH2_ex8	120	115	119	134	128	133	144	112
MSH2_ex9	103	96	95	109	84	126	106	86
MSH2_ex10	150	203	137	177	185	180	248	168
MSH2_ex11	107	128	139	109	126	161	174	175
MSH2_ex12	75	83	96	96	57	82	112	76
MSH2_ex13	87	100	131	81	117	121	159	131
MSH2_ex14	162	139	168	137	136	163	201	124
MSH2_ex15	112	91	111	106	86	101	128	109
MSH2_ex16	163	148	191	214	157	208	214	155
MSH2_3'UTR	132	141	133	191	163	153	152	205
MSH6_ProA	129	93	134	119	103	136	144	117
MSH6_ProB	85	122	108	104	87	124	118	79
MSH6_ex1A	76	60	84	64	75	71	84	62
MSH6_ex1B	95	81	84	66	60	79	92	79
MSH6_ex2	57	71	63	78	50	77	92	74
MSH6_ex3	155	138	131	152	115	152	164	125
MSH6_ex4A	149	120	183	185	152	124	178	129
MSH6_ex4B	65	65	77	67	53	76	90	73
MSH6_ex4C	97	95	101	106	89	86	111	91
MSH6_ex4D	75	79	71	121	86	78	81	101
MSH6_ex4E	147	100	120	123	110	117	154	114
MSH6_ex4F	191	199	175	205	195	163	184	191
MSH6_ex4G	73	66	65	58	68	78	80	76
MSH6_ex4H	87	97	105	97	123	109	136	89
MSH6_ex4I	59	43	50	49	45	61	73	50
MSH6_ex4J	101	103	101	102	103	113	140	109
MSH6_ex4K	140	134	144	147	123	133	141	142
MSH6_ex5A	223	160	165	187	144	161	166	159
MSH6_ex5B	141	115	155	137	119	116	156	132
MSH6_ex6	125	108	120	146	105	114	132	110
MSH6_ex8	101	115	98	100	93	102	131	97
MSH6_ex9	163	152	163	179	151	140	187	122
MSH6_3'UTR	148	154	151	141	141	178	178	173

Amplicon	MID1	MID2	MID3	MID4	MID5	MID6	MID7	MID8
MLH1_Pro	149	126	180	194	150	168	213	177
MLH1_ex1	157	154	168	183	128	130	185	163
MLH1_ex1_2	200	163	171	187	164	147	205	123
MLH1_ex2	155	165	153	165	178	144	192	190
MLH1_ex3	178	159	234	210	203	236	291	191
MLH1_ex4	55	45	57	58	62	65	86	76
MLH1_ex5	117	109	141	133	116	143	159	142
MLH1_ex6	115	106	132	124	101	106	114	134
MLH1_ex8	112	101	101	146	118	115	119	114
MLH1_ex9	84	95	105	112	107	121	148	125
MLH1_ex10	145	94	143	128	102	126	130	91
MLH1_ex11	113	148	125	168	148	147	157	123
MLH1_ex12A	95	79	74	98	104	111	106	99
MLH1_ex12B	54	31	34	36	24	31	33	34
MLH1_ex13	70	38	60	71	53	56	64	71
MLH1_ex14	112	86	118	124	120	114	146	89
MLH1_ex15	93	90	100	89	91	92	122	85
MLH1_ex16	204	175	195	275	224	185	292	225
MLH1_ex17	207	150	152	198	178	201	198	168
MLH1_ex18	180	195	211	194	156	190	232	209
MLH1_ex19	109	101	94	128	128	117	136	142
PMS2_ex1	150	191	91	182	145	146	179	151
PMS2_ex2	150	174	213	208	156	156	163	168
PMS2_ex3	150	173	207	224	194	235	261	193
PMS2_ex4	145	179	169	170	144	176	163	153
PMS2_ex5A	145	145	162	152	161	179	240	159
PMS2_ex5B	195	106	109	180	119	143	129	152
PMS2_ex6	97	113	151	173	107	143	141	141
PMS2_ex8	282	206	256	288	270	254	262	245
PMS2_ex9	335	258	352	325	332	363	420	266
PMS2_ex10	158	141	162	162	134	160	204	167
PMS2_ex11A	180	198	195	222	200	215	220	219
PMS2_ex11B	171	192	205	210	167	166	205	211
PMS2_ex11C	93	61	62	83	73	76	84	74
PMS2_ex11D	146	112	73	78	56	143	147	121
PMS2_ex12	86	110	97	138	101	120	149	115

6.8 Variants found MMR Genes GS Junior Runs

Variants found in the amplicons solely analyzed by GS Junior sequencing or Sanger sequencing are indicated by the word "GS Junior" and "Sanger", respectively. HET, heterozygous; HOM, homozygous.

[1] r[=]+[1662_1759del]
[2] r[=]+[989_1144del,989_1050del]

Gene	DNA	Protein	Class	HET	HOM	Agreement with Sanger
MLH1	c.-93G>A		1	9	1	yes
MLH1	c.1-7C>T		3	1	0	yes
MLH1	c.1-28A>G		3	1	0	yes
MLH1	c.655A>G	p.Ile219Val	1	15	4	yes
MLH1	c.1558+14G>A		1	2	0	yes
MLH1	c.1668-19A>G		1	15	8	yes
MLH1	c.1852_1853delAAinsGC	p.Lys618Ala	2	1	0	yes
MLH1	c.1959G>T	p.Leu653Leu	2	2	0	yes
MLH1	c.*35_37delCTT		2	2	0	Sanger
MSH2	c.-118T>C		1	8	1	yes
MSH2	c.211+9C>G		1	15	10	yes
MSH2	c.965G>A	P.Gly322Asp	2	2	0	yes
MSH2	c.1511-9A>T		1	7	2	yes
MSH2	c.1661+12G>A		1	18	3	yes
MSH2	c.1666T>C	p.Leu556Leu	1	1	0	yes
MSH2	c.1759G>C [1]	p.Gly587Arg	5	1	0	yes
MSH2	c.1786_1788delAAT	p.Asn596del	4	1	0	yes
MSH2	c.2006-6T>C		1	7	1	yes
MSH6	c.-159C>T		1	10	1	yes
MSH6	c.116G>A	p.Gly39Glu	1	3	0	yes
MSH6	c.186C>A	p.Arg62Arg	1	11	0	yes
MSH6	c.260+22C>G		1	11	0	yes
MSH6	c.276A>G	p.Pro92Pro	1	11	0	yes
MSH6	c.540T>C	p.Asp180Asp	1	15	4	yes
MSH6	c.628-56C>T		1	10	1	yes
MSH6	c.642C>T	p.Tyr214Tyr	1	11	1	yes
MSH6	c.1186C>G	p.Leu396Val	2	3	0	yes
MSH6	c.2633T>C	p.Val878Ala	3	1	0	yes
MSH6	c.3261dupC	p. Phe1088LFsX5	5	1	0	yes
MSH6	c.3438+13dupT		3	1	0	yes
MSH6	c.3438+14A>T		1	16	3	yes
MSH6	c.3439-16C>T		3	1	0	yes
MSH6	c.*85T>A		3	2	0	yes

Gene	DNA	Protein	Class	HET	HOM	Agreement with Sanger
PMS2	c.-154C>G		1	6	1	yes
PMS2	c.52A>G	p.Ile18Val	3	1	0	yes
PMS2	c.59G>A	p.Arg20Gln	1	6	0	yes
PMS2	c.288C>T	p.Ala96Ala	1	6	0	yes
PMS2	c.251-72A>G		1	2	0	GS Junior
PMS2	c.705+17A>G		1	16	4	yes
PMS2	c.780C>G	p.Ser260Ser	1	10	16	Sanger
PMS2	c.823C>T	p.Gln275X	4	1	0	yes
PMS2	c.989-1G>T [2]		5	2	0	yes
PMS2	c.1408C>T	p.Pro470Ser	1	17	3	yes
PMS2	c.1437C>G	p.His479Gln	3	1	0	yes
PMS2	c.1454C>A	p.Thr485Lys	1	7	0	yes
PMS2	c.1531A>G	p.Thr511Ala	1	3	1	yes
PMS2	c.1621G>A	p.Glu541Lys	1	10	1	yes
PMS2	c.1688G>T	p.Arg563Leu	2	1	0	yes
PMS2	c.1866G>A	p.Met622Ile	2	1	0	GS Junior
PMS2	c.2006+6G>A		1	6	1	yes
PMS2	c.2007-4G>A		1	7	0	yes
PMS2	c.2007-7C>T		1	6	0	yes
PMS2	c.2275+169G>C		1	4	1	Sanger
PMS2	c.2276-135T>C		1	7	1	Sanger
PMS2	c.2466T>C	p.Leu822Leu	1	12	0	Sanger
PMS2	c.2570G>C	p.Gly857Ala	1	3	0	Sanger
PMS2	c.2589+17G>C		1	5	0	Sanger
PMS2	c.2589+92insA		1	24	0	Sanger

6.9 CRC Pilot Study

For blood analysis, the highest class detected in the four MMR genes is given. For two patients more than one tumor was analyzed, AMG/0003 and JH/0094. ID, initials and sample number; F, female; M, male; MSI, microsatellite instability analysis; MSI-H, high-frequency microsatellite instability; MSS, microsatellite stable; WT, wild type; MUT, pathogenic mutation detected; NA, not analyzed; um, unmethylated; pm, partial methylated; m, methylated.

[1] tissue samples used for analysis did not contain tumor cells due to radiotherapy

Patient Information			Tumor Analysis				Blood Analysis			
ID	Sex	Age	MSI	*BRAF* ex11	*BRAF* ex15	*MLH1*	*MSH2*	*MSH6*	*MLH1*	*PMS2*
JH/0001	F	81	MSS	WT	WT	um	1	1	1	1
AMB/0002	F	81	MSS	WT	WT	um	1	1	1	3
AMG/0003	M	81	MSS	WT	WT	um	1	1	2	1
			MSS	WT	WT	um				
GDT/0005	M	64	MSS	WT	WT	um	1	1	3	1
TL/0007	F	50	MSS	NA	WT	um	1	3	1	1
KPH/0009	M	81	MSS	NA	WT	um	1	3	1	1
PAK/0011	M	81	MSS	WT	WT	um	1	2	1	1
IAT/0012	F	41	MSS	NA	WT	um	0	3	2	1
KS/0014	M	85	MSS	NA	WT	um	1	0	2	1
REMB/0015	F	79	MSS	NA	WT	pm	0	1	1	3
OH/0017	F	75	MSS	NA	WT	um	1	2	1	5
BS/0018	F	66	MSS	NA	WT	um	1	1	1	1
GWG/0019	M	46	MSS	NA	MUT	um	1	1	1	1
BJ/0025	M	73	MSI-H	WT	MUT	m	0	3	1	1
EME/0026	F	82	MSS	WT	WT	um	2	1	1	1
PEE/0028	M	56	MSS	NA	WT	um	1	1	1	1
MLS/0029	F	87	MSI-H	WT	MUT	m	1	1	1	1
KG/0070	M	50	MSI-H	WT	WT	um	1	1	1	5
JH/0094	F	82	MSS	WT	MUT	um	0	1	1	1
			MSI-H	WT	MUT	m				
			MSI-H	NA	WT	um				
			MSS	NA	WT	um				
SB/0135	F	76	MSS	NA	WT	um	1	1	1	1
MLSH/0147	F	47	MSI-H	NA	WT	um	5	1	1	1
MS/0158	F	70	MSI-H	NA	WT	um	1	1	1	2
JSK/0220	M	55	MSI-H	NA	WT	um	1	1	1	1
HAK/0235	M	81	MSS	NA	WT	um	2	0	0	1
KPR/0243	M	64	MSI-H	NA	WT	um	1	3	1	1
AS/0347	F	72	MSI-H	NA	WT	um	0	0	1	1
RMA/0392	M	58	MSI-H	NA	WT	um	0	5	1	1
AA/0432	M	60	MSS[1]	NA	WT	um	4	1	1	1
SS/446	M	68	MSI-H	NA	WT	um	1	1	0	1
OE/0472	M	75	MSI-H	NA	WT	pm	1	2	2	1
AR/475	M	72	MSI-H	WT	WT	um	1	1	1	1
JS/0495	M	88	MSI-H	NA	WT	um	0	1	1	4