An Integrative DNA Sequencing Panel To Assess Mismatch Repair Deficiency

by

Leslie Elizabeth Oldfield

A thesis submitted in conformity with the requirements
for the degree of Master of Science

Department of Medical Biophysics
University of Toronto

An Integrative DNA Sequencing Panel To Assess Mismatch Repair Deficiency

Leslie Elizabeth Oldfield

Master of Science

Department of Medical Biophysics
University of Toronto

2018

Abstract

Detecting mismatch repair (MMR) deficiency requires serial testing of both germline and tumour DNA using several assays to determine the underlying mechanism of MMR gene disruption. We have created an integrated targeted panel (MultiMMR) that tests for multiple sources of genome variation from a single aliquot of tumour or normal DNA. We have profiled 11 genes related to MMR deficiency or hereditary cancer syndromes on 82 individuals. For each sample, we performed hybrid capture of a single DNA sequencing library constructed using methylated adapters for parallel bisulfite and conventional sequence analysis. MultiMMR recapitulated clinical testing in 22/24 cases and was able to explain the mechanism of MMR loss in an additional 28 patients. This study has shown the utility of integrated mutation, copy number, and methylation profiling to detect hereditary and somatic causes of MMR deficiency. MultiMMR amalgamates the current step-wise and complex clinical testing workflow into a single assay.

Acknowledgments

First and foremost, I would like to acknowledge my supervisor Dr. Trevor Pugh, for his constant mentorship and support over the course of the last few years. His dedication to translational and collaborative research has provided me with ample opportunity to engage in clinical research. I am grateful for his patience in helping me through challenging times and for his permission to explore interests of mine outside of my direct thesis work. I would also like to thank my supervisory committee, Dr. Steven Gallinger and Dr. Michael Hoffman, for their scientific support and feedback on this thesis work. It has been a privilege to be guided by exceptional scientists and mentors.

I would like to thank Dr. Sarah Ferguson, Dr. Alicia Tone, and Dr. Blaise Clarke for their role in conceiving, funding, and providing support in this thesis work. Thank you to the members of the Princess Margaret Genomics Centre (Julissa Tsao, Neil Winegarden, Zhibin Lu, and Carl Virtanen) for sequencing and providing raw data processing support.

A special thanks is warranted to all past and current members of the Pugh Lab. The constant support and camaraderie within our lab is something I will always cherish. Thank you Arnavaz Danesh, Jeff Bruce, Rene Quevedo, and Cindy Yang for computational support and Tiantian Li for laboratory support. Thank you to Danielle Croucher, Laura Richards, and Nina Wang for making the coolbicle a home for the past few years.

Most importantly, I would like to thank my family and friends. To my parents, thank you for providing me with endless love and for believing in me. To AAA, Laura, Vinny, and Jenna, I am so thankful I met all of you at Geneva Park. I am grateful to have shared so many incredible moments with all of you. Lastly, thank you Matt for your undying support and relentless encouragement and for keeping me fed and focused on the big picture. Thank you.

Table of Contents

Acknowledgments..... iii

Table of Contents..... iv

List of Figures and Tables..... vii

List of Abbreviations..... ix

Chapter 1 Introduction..... 1

1.1 Introduction to Mismatch Repair Deficiency..... 1

1.1.1 Mismatch Repair Pathway..... 1

1.1.2 Microsatellite Instability..... 4

1.2 Introduction to Knudson's Two-hit Hypothesis..... 6

1.3 Introduction to Somatic Mismatch Repair Deficiency..... 8

1.3.1 *MLH1* Promoter Methylation Testing..... 8

1.3.2 LLS Testing..... 9

1.4 Introduction to Lynch Syndrome (LS)..... 10

1.4.1 Diagnosis of LS..... 11

1.4.2 Treatment and Screening of LS..... 15

1.5 Introduction to Constitutional Mismatch Repair Deficiency (CMMRD)..... 16

1.5.1 Diagnosis and Treatment of CMMRD..... 17

1.6 Applications of NGS in Diagnostic Settings..... 18

1.7 Project Outline..... 19

1.7.1 Rationale and Hypothesis..... 19

1.7.2 Objectives..... 20

Chapter 2 Materials and Methods..... 20

2.1 Study Design..... 21

2.2 Panel Design..... 21

2.2.1 MultiMMR V1..... 21

2.2.2 MultiMMR V2..... 22

2.3 Study Cohort..... 24

2.4 Panel Library Preparation and Sequencing..... 24

2.5 Panel Conventional Sequence Analysis..... 27

2.6 Exome Sequencing Analysis..... 30

2.6.1 Library Preparation and Sequencing..... 30

2.6.2 Alignment, Processing, and Variant Calling..... 30

2.7 Variant Filtration and Interpretation..... 30

2.8 Microsatellite Instability Analysis..... 31

2.9 Methylation Sequence Analysis..... 31

Chapter 3 Results..... 33

3.1 Samples from 80 patients passed quality control……..... 33

3.2 MultiMMR recapitulates results from clinical testing..... 37

3.3 MultiMMR captures additional genome variation..... 40

3.4 MSI detected in targeted sequencing data..... 44

3.5 Promoter methylation found in a subset of MMR deficient cases..... 46

3.6 Between and within run reproducibility achieved on MultiMMR V2..... 49

3.7 Pseudogenes affect targeted sequence quality..... 49

Chapter 4 Discussion and Future Directions..... 51

4.1 Discussion..... 51

4.2 Future Directions..... 55

References..... 58

List of Figures and Tables

Figure 1-1 The three processes involved in mismatch repair

Figure 1-2 The spectrum of mismatch repair

Figure 1-3 Cancer Care Ontario Guidelines for screening CRC patients for LS

Figure 2-1 Laboratory workflow for the MultiMMR Panels

Figure 2-2 Bioinformatics workflow for the MultiMMR Panels

Figure 3-1 Quality control metrics across entire cohort

Figure 3-2 Overview of cohort results

Figure 3-3 Deep intronic *MSH2* variant missed by clinical testing

Figure 3-4 Microsatellite instability status and *MLH1* promoter methylation results

Figure 3-5 Visual comparison of *MLH1* methylation across two patients

Figure 3-6 Visual comparison of *MLH1* MS-MLPA probes across two patients

Figure 3-7 On and off-target regions captured within MultiMMR V2

Figure 4-1 Potential clinical utility of MultiMMR

Table 1-1 Mutation frequency and cancer risk

Table 1-2 Amsterdam and Bethesda Criteria for diagnosing Lynch Syndrome

Table 2-1 Genomic regions captured by MultiMMR V1 and V2

Table 2-2 Software parameters used in the bioinformatics analysis

Table 3-1 Sample cohort characteristics

Table 3-2 Results from the SeraCare SeraSeq Inherited DNA Mix

Table 3-3 Summary of 24 patients with complete clinical testing results

Table 3-4 Types of genome variation found across the cohort

List of Abbreviations

ADP	Adenosine Diphosphate
APC	Adenomatous Polyposis Coli
ATP	Adenosine Triphosphate
bp	Base pair
BRAF	B-raf Proto-oncogene Serine/Threonine-Protein Kinase
CAL	Café-Au-Lait
Chr	Chromosome
CMMRD	Constitutional Mismatch Repair Deficiency
CRC	Colorectal Cancer
dbSNP	Single Nucleotide Polymorphism database
DNA	Deoxyribonucleic Acid
EC	Endometrial Cancer
EGFR	Epidermal Growth Factor Receptor
EPCAM	Epithelial Cellular Adhesion Molecule
ExAC	Exome Aggregation Consortium
EXO1	Exonuclease 1
FAP	Familial Adenomatous Polyposis
FFPE	Formalin-fixed Paraffin-embedded
GATK	Genome Analysis Toolkit
HNPCC	Hereditary Non-Polyposis Colorectal Cancer
Hx	History

IDL	Insertion and Deletion Loop
IGV	Integrative Genomics Viewer
IHC	Immunohistochemistry
Indel	Insertion or deletion
InSiGHT	International Society for Gastrointestinal Hereditary Tumours
kbp	Kilo-base pair
LLS	Lynch-like Syndrome
LOH	Loss-Of-Heterozygosity
LS	Lynch Syndrome
Mb	Mega base pair
MLH1	MutL homolog 1
MLH3	MutL homolog 3
MMR	Mismatch Repair
MSH2	MutS homolog 2
MSH3	MutS homolog 3
MSH6	MutS homolog 6
MSI	Microsatellite Instability
MSI-H	Microsatellite Instability High
MSI-L	Microsatellite Instability Low
mSINGS	MSI Phenotype using NGS
MS-MLPA	Methylation Specific Multiplex Ligation-dependent Probe Amplification
MSS	Microsatellite Stable
MUTYH	MutY Homolog

NF-1	Neurofibromatosis Type 1
NGS	Next Generation Sequencing
PCNA	Proliferating Cell Nuclear Antigen
PCR	Polymerase Chain Reaction
PD-1	Programmed Cell Death Protein 1
PMS1	Postmeiotic Segregation Increased 1
PMS2	Postmeiotic Segregation Increased 2
PMS2CL	PMS2 C-Terminal Like Pseudogene
Pol δ	Polymerase delta
POLD1	DNA Polymerase Delta 1, Catalytic Subunit
POLE	DNA Polymerase Epsilon, Catalytic Subunit
PPAP	Polymerase Proofreading-Associated Polyposis
ROC	Receiver Operator Characteristic
RPA	Replication Protein A
SIR	Standardized Incidence Ratio
SMMRD	Somatic Mismatch Repair Deficiency
TCGA	The Cancer Genome Atlas
TGFBR2	Transforming Growth Factor-Beta Receptor Type 2
TP53	Tumour Protein p53
UTR	Untranslated Region
VUS	Variants of Uncertain (or Unknown) Significance

Chapter 1

Introduction

1.1 Introduction to Mismatch Repair Deficiency

1.1.1 Mismatch Repair Pathway

During DNA replication, nucleotide misincorporation can occur in the newly synthesized DNA strand. If left unrepaired these mutations can be sustained in dividing cells, alter the cellular phenotype, and cause disease[1]. To safeguard the integrity of the genome, cells have mismatch repair (MMR) machinery that can recognize post-replicative errors made by deoxyribonucleic acid (DNA) polymerases that have escaped proofreading mechanisms, and recruit repair machinery to fix the error. This system is able to recognize and repair two types of mismatches: base-base (also known as single-base) mismatches, and small insertion-and-deletion loops (IDLs)[2]. The MMR pathway is highly conserved and improves the fidelity of DNA replication by 50-1,000 fold by giving DNA polymerase a second chance at synthesizing the correct DNA strand[2]. Although this pathway remains an active area of research and has not yet been fully elucidated, it has become clear that there are three main processes involved in MMR: initiation of the MMR pathway, excision of the mismatch, and repair of the DNA strand containing the mismatch (Figure 1-1).

Initiation: As described below, the MMR pathway involves four primary proteins (MLH1, MSH2, MSH6, and PMS2) that form complexes to recognize mismatches within DNA. The MutS complex initiates MMR by recognizing and binding to the site of the mismatch. There are two MutS complexes, MutSα and MutSβ, which recognize different classes of mismatches. The most common MutS complex, MutSα, is comprised of a heterodimer formed between MSH2 (required for MMR) and MSH6. It is able to recognize base-base mismatches and IDLs that contain 1-2 extra-helical nucleotides[3]. MutSβ contains a heterodimer of MSH2 and MSH3 and can recognize larger IDLs (approximately 2-10 extra-helical nucleotides)[3]. Mismatch provoked ADP-ATP exchange converts MutSα, a complex containing ATPases, into sliding clamps that may move along the DNA to find the mismatch[4,5]. It is important to note that there are multiple models for MMR that stand in stark contrast to each other, with some models favouring

a sliding clamp that moves from the mismatch site to stabilize exonuclease 1 (EXO1) whereas others argue that MutS and MutL remain at the mismatch site[4]. Moreover, some models favour multiple sliding clamps where others believe that only a singular MutSα sliding clamp is present[4].

After recognizing and binding to the mismatch, a MutS complex may recruit a MutL complex to the site of the mismatch, MutL is the 'molecular matchmaker' that coordinates the assembly of other MMR machinery[5]. There are three MutL complexes, all involving MLH1: MutLα, MutLβ, and MutLγ. MLH1 heterodimerizes with PMS2, PMS1, or MLH3 to form MutLα, MutLβ, or MutLγ, respectively[6]. There is little known about the functions of these heterodimers[7]. Research has shown that MutLα, formed from MLH1 and PMS2, is the most common of the three and is required for MMR[4]. The other two complexes are rare and their role remains unclear. However, MutLγ has been shown *in vitro* to play a small role in assisting base-base and single-nucleotide IDL mismatches and to play a role in meiosis[6].

Excision and Repair: The MutSα sliding clamp (MSH2/MSH6 heterodimer) interacts with surrounding molecules differently, depending on the directionality of the excision. For 5'-excisions (in which the DNA break is located on the 5' side of the mismatch), the sliding clamp stimulates EXO1 on the DNA strand that contains the error. EXO1 catalyzes the excision and is responsible for degrading the area between the mismatch and the DNA break[5]. Simultaneously, replication protein A (RPA) stabilizes the parental single-stranded DNA. In 3'-excisions, the sliding clamp and the MutLα complex (MLH1/PMS2) combine with a proliferating cell nuclear antigen (PCNA) molecule that is bound to the nick site and activate EXO1[4]. EXO1 excises the mismatch. After EXO1 is inhibited, regardless of where the mismatch was located, Pol δ fills in the gap and DNA ligase I seals the nick[5]. This allows a second chance at producing an error-free strand[5]. If the MMR machinery becomes impaired through a mutation within one of the four main MMR genes (*MLH1, MSH2, MSH6,* or *PMS2*), the cell does not get a second chance at producing an error-free strand. The strand with the error may be sustained within dividing cells and lead to a phenomenon known as microsatellite instability (MSI).

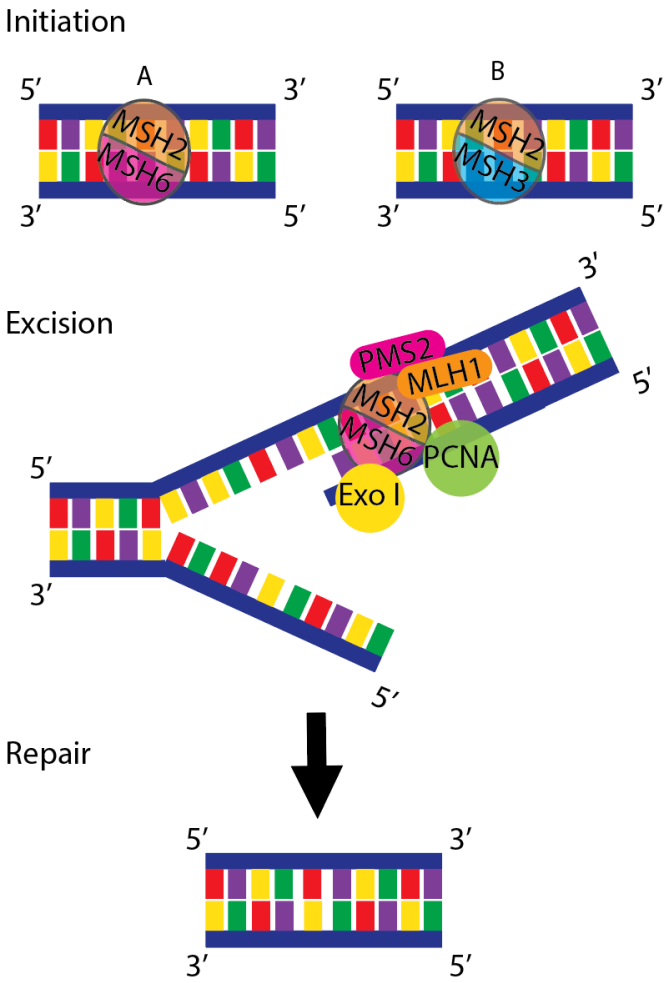


Figure 1-1: The three processes involved in mismatch repair
A simplified depiction of the three steps involved in mismatch repair: initiation, excision, and repair. *Initiation* of MMR occurs when the heterodimer MSH2/MSH6 or MSH2/MSH3 recognizes a mismatch and binds to the DNA. *Excision* begins after the PMS2/MLH1 complex is recruited to the mismatch site and coordinates other proteins to the site, such as EXO1 and PCNA. EXO1 is responsible for excising the area surrounding the mismatch and PCNA binds to the nick site. *Repair* of the DNA strand ends with DNA polymerase filling in the excision gap and DNA ligase sealing it.

1.1.2 Microsatellite Instability

A microsatellite is a repetitive sequence of short DNA motifs (e.g. TATATATA) that constitute up to 3% of the genome[8,9]. MSI is a condition of genetic hypermutability caused by the expansion or contraction of nucleotides from microsatellite tracts within a tumour. This results in novel alleles of varying lengths that differ from the germline DNA[10,11]. Typically, these abnormal alleles arise during replication when strand slippage occurs in microsatellite sites, forming mismatches and insertion-deletions loops, which can be corrected by MMR machinery. However, when the MMR machinery is impaired, these mismatches are not corrected and are sustained within the cell, resulting in expansions and contractions of microsatellites.

Evaluating MSI as a surrogate marker of MMR deficiency is often performed using Polymerase Chain Reaction (PCR)-based repeat sizing assays. In 1997 The National Cancer Institute proposed the Bethesda Panel, a set of 5 microsatellite markers to determine MSI status. The Bethesda Panel contains two mononucleotide markers, BAT25 and BAT26, and three dinucleotide markers D2S123, D5S346, and D17S250[12,13]. If two, one, or none of the five markers show instability (defined as at least two novel alleles), the tumour is classified as MSI-high (MSI-H), MSI-low (MSI-L), or microsatellite stable (MSS), respectively[12,13]. Recent literature has shown no statistical difference between MSI-L and MSS in terms of the number of gained microsatellite alleles in the tumour compared to normal tissue, promoting researchers to now categorize tumours as either MSI or MSS[10]. The Bethesda Panel is 80-91% sensitive among individuals with *MLH1* or *MSH2* mutations and 55-77% sensitive among individuals with *MSH6* or *PMS2*, with specificity around 90%[14]. Since the Bethesda Panel often fails to identify *MSH6*-deficient CRCs and the dinucleotide repeats perform poorly at identifying MSI-H tumours, many clinical laboratories now use a panel of five quasi-monomorphic mononucleotide markers [15]. Some laboratories use BAT-25, BAT-26, NR-21, NR-22, and NR-24 as the five markers of interest whereas others use the Promega Panel, a commercial panel that includes the above markers (with the exception of MONO-27 instead of NR-22) and two additional pentanucleotide markers (Penta C and Penta D)[16–18]. In contrast to these PCR-based methods, the increasing availability and ability to interrogate hundreds of sites simultaneously has resulted in multiple laboratories designing next-generation sequencing (NGS)-based MSI tests and

classification algorithms[10,19]. For example, a group at the University of Washington developed mSINGS, a method for classifying a tumour as MSS or MSI by comparing tumour/normal allele counts at user-defined MSI sites[20]. This group was able to stratify 324 tumour samples as MSS or MSI using between 15 to 2957 microsatellite loci[20].

The same University of Washington group used mSINGS to examine the presence of MSI across various cancer types[10]. Their study of the exomes of 5,930 individuals found tumours that display MSI in 14/18 cancer types tested[10]. However, MSI remains primarily associated with endometrial cancer (EC) and colorectal cancer (CRC) with approximately 30% of EC cases exhibiting MSI[10]. Despite the high prevalence of EC tumours exhibiting MSI, the relationship between MSI, EC, and clinical response remains unclear. For example, one study found that 15.6% of EC cases were MSI (all endometrioid subtype) and that no statistical difference in survival existed between patients with MSI-H versus MSS tumours[21]. In contrast, another study found worst outcomes in MSI tumours of EC patients compared to MSS tumours[22]. It should be noted that sample size was small in both studies (109 and 119, respectively). The Cancer Genome Atlas (TCGA) Consortium profiled 373 ECs and classified patients into four groups based on genomic characteristics, one being MSI[23]. However, survival between MSI and MSS tumours was not directly compared. Therefore further work is needed to understand the relationship between MSI and EC survival and treatment.

The relationship between MSI and CRC has been studied extensively as CRC is closely linked to MMR deficiency[17,24]. Many studies have shown that approximately 15% of CRCs display MSI[25]. Of this 15%, 3% can be attributed to Lynch Syndrome (LS) and the remaining 12% is primarily caused by somatic hypermethylation of the *MLH1* promoter[25]. However, this excludes unexplained cases where the individual has MSI and CRC, but lacks an inherited variant or *MLH1* somatic promoter methylation. It is estimated that approximately 70% of these patients have somatic, biallelic loss of MMR[11]. CRCs that exhibit MSI and MMR deficiency have a phenotype distinct from other CRCs, namely a tendency to arise in the proximal colon, are poorly differentiated, have a mucinous or signet ring appearance, and an abundance of tumour infiltrating lymphocytes[25,26]. MSI within tumours can be a predictive and prognostic biomarker with sporadic MSI tumours shown to have a better stage-adjusted survival compared

to MSS tumours[27]. In addition, one study showed that patients with MSI-H tumours responded poorly to 5-fluorouracil-based adjuvant chemotherapy compared to MSS tumours[28]. Moreover, new research has shown that MMR deficient tumours respond well to immune checkpoint blockade with anti-PD-1 antibodies[29,30].

1.2 Introduction to Knudson's Two-hit Hypothesis

Alfred Knudson's study of the age-specific incidence of retinoblastoma, a cancer of the retina, lead him to postulate that two mutagenic events, or 'hits', were necessary for retinoblastoma to develop[31]. This idea is often referred to as the Knudson or "Two-hit" hypothesis and has shown to be true for many cancers that arise when tumour suppressors are inactivated. Two hits are needed to inactivate the gene, one hit to inactivate each allele. One hit is insufficient as the remaining allele is still functional and able to produce functional proteins. A second hit must occur in the second allele to result in complete inactivation.

There are multiple avenues to achieve two hits. Individuals can be born with only one functioning allele, meaning they inherited the first hit, which is often the case for Mendelian disorders. These individuals are at a greater risk of cancer or disease as they only need one more hit somewhere in the remaining functional allele to have complete inactivation of the protein. Therefore those with an inherited variant in a gene important for cancer growth and development have a risk of getting that 'second hit' at some point over the course of their life. Conversely, individuals can be unlucky and have two pathogenic hits occur in a disease-associated gene over their lifetime (somatic inactivation) or, in rare cases, can be born with two hits (one inherited from each parent).

When the MMR pathway loses the ability to safeguard the genome from replicative errors, the cells adapt a mutator phenotype and the rate of spontaneous mutation rises[2]. This has been associated with many cancers, primarily cancers of the colon and endometrium but also ovary, stomach, small intestine, hepatobiliary tract, urinary tract, brain, and skin[32]. Inactivation of the MMR pathway follows Knudson's hypothesis and can be inactivated by loss of one of the four MMR genes (*MLH1, MSH2, MSH6*, or *PMS2*) by a variety of mechanisms (Figure 1-2):

a) Somatic MMR deficiency (SMMRD): this grouping includes both individuals with *MLH1* promoter hypermethylation and individuals with 'Lynch-like Syndrome' (LLS), also known as unexplained MMR deficiency. LLS is the term used to denote cases where an individual has MMR deficiency but neither *MLH1* promoter methylation or a germline mutation are detected. Oftentimes these individuals have sporadic biallelic loss of a MMR gene.

b) Lynch Syndrome (LS): occurs when an individual inherits a pathogenic mutation from one parent in a MMR gene and therefore has a high lifetime risk of getting cancer as they were born with one hit.

c) Constitutional MMR deficiency (CMMRD): occurs when an individual inherits two pathogenic variants on differing alleles (one from each parent) of a MMR gene, resulting in complete loss of gene expression.

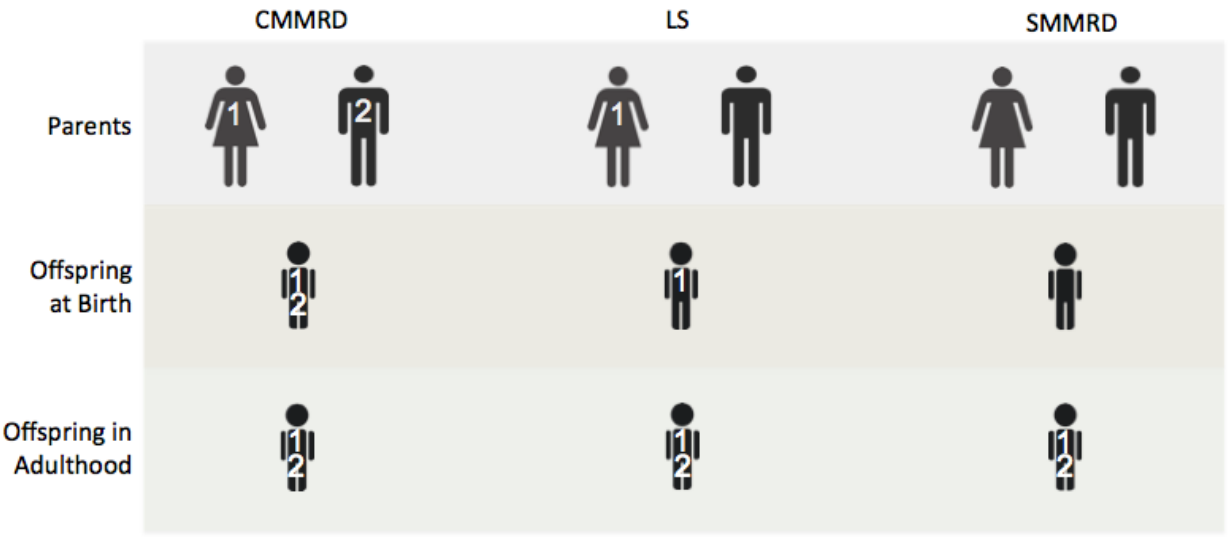


Figure 1-2: The spectrum of mismatch repair
This visualization depicts how each MMR deficiency syndrome is transmitted from the parent, to the child at birth, to adulthood. The numbers represent a hit, such as a variant. For example, in CMMRD, each parent has a variant in a MMR gene (depicted by 1 and 2) and the child inherits both variants from the parents resulting in CMMRD. For LS, one parent passes down a variant in a MMR gene that the child is born with (depicted by a 1). Over the course of the individual's life, they unfortunately acquire a mutation in the same gene (depicted by 2). In the case of SMMRD the child inherits no variants from either parent, but acquires two mutations in the same gene over the course of their lifetime (depicted by the 1 and 2 in adulthood).

1.3 Introduction to Somatic Mismatch Repair Deficiency

SMMRD arises when inactivation of a gene is caused by events that occur over an individual's lifetime and does not have a hereditary component. These events are tumour-specific and occur only in a subset of the individual's cells. There are two primary distinctions in SMMRD depending on whether promoter methylation-mediate silencing is detected. Hypermethylation of the *MLH1* promoter accounts for 10-12% of all MMR deficient CRCs and ECs. In *MLH1* deficient tumours, hypermethylation of the promoter is responsible for 80% and 89% of cases in CRC and EC, respectively[33,34]. Alternatively, the term Lynch-like syndrome (LLS) is used to refer to an individual that has protein loss of a MMR gene but does not have *MLH1* methylation or a germline variant. It is estimated that 70% of these individuals have somatic inactivation of a MMR gene[11]. The prevalence of LLS among CRC and EC patients is currently an area of active research. The Spanish EPICOLON study (n=1416) found 2.5% of their cohort had LLS and 3.9% of all patients in the Columbus LS (n=1066) study were found to have LLS[33,35]. SMMRD is diagnosed during screening of LS-suspected patients. Methylation testing is performed on suspected LS patients to determine if their disease is sporadic and somatic mutation testing is performed in an effort to explain MMR deficiency in cases when both methylation and germline mutations are not found.

1.3.1 *MLH1* Promoter Methylation Testing

The majority of *MLH1* deficiency is caused by somatic *MLH1* promoter methylation that results in epigenetic silencing of the gene. Oftentimes, Methylation Specific Multiplex Ligation-dependent Probe Amplification (MS-MLPA) testing is performed on the tumours of patients after MMR immunohistochemistry (IHC) has shown loss of MLH1 protein expression. MS-MLPA is a test that places several probes in the *MLH1* promoter region that contain recognition sites, approximately 4 bases in length, for *HhaI*, a methylation-sensitive endonuclease[36]. Probes with a methylated recognition site will produce a signal (as they are protected against *HhaI* digestion) whereas unmethylated samples will be digested by the *HhaI* enzyme and therefore cannot be amplified during PCR and will not produce a signal[36]. This test is

performed to determine promoter methylation status prior to performing an expensive germline test.

Often, *MLH1* promoter methylation is evaluated in combination with, or post, *BRAF* gene testing. The mutation valine to glutamic acid at residue 600 (p.V600E) in the *BRAF* gene is present in approximately 5-10% of all CRCs and is associated with sporadic origin of disease[11]. *BRAF* mutation testing can be both predictive and prognostic: it can predict resistance to *EGFR* therapies and is associated with worse outcome in MSS CRC[11]. Therefore knowing the *BRAF* status of patients adds value to the clinical management of the patient. One study showed 81% concordance between positive *BRAF* p.V600E mutation status and *MLH1* promoter methylation in 126 tumours, although the underlying mechanism between this correlation is not clear[37]. Very few studies have shown *BRAF* p.V600E mutations in LS[38], therefore it is often used after IHC has shown *MLH1* to be deficient to provide evidence for somatic origin of disease. Tumours with both the *BRAF* p.V600E mutation and *MLH1* promoter hypermethylation are almost always sporadic[39]. It should be noted though, that there have been very rare cases where a *BRAF* p.V600E mutation and MLH1 hypermethylation have been detected in a LS-related cancer, for example in the case of a heritable germline epimutation[40,41] or a second inactivating event[42,43]. One study showed that performing *BRAF* mutation testing in CRC cases with absent MMR protein via IHC resulted in a 40% reduction of patients needing genetic counselling[44]. Therefore, upfront *BRAF* testing can be a predictor of SMMRD. There is debate in the field as to whether performing *MLH1* methylation testing is necessary if the patient has already been found to have a *BRAF* p.V600E mutation[37].

1.3.2 LLS Testing

Individuals with LLS have mean age of onset and cancer risk rate that is in between LS and sporadic disease[35]. Families with LLS have a higher standardized incidence ratio (SIR) than those with sporadic CRC (SIR for LLS, 2.12, SIR for sporadic CRC, 0.48) but lower than families with confirmed LS (SIR for LS, 6.04)[35]. Patients with LLS are often counselled as though they do have LS, but more research is needed to further elucidate the best screening and surveillance protocols for these patients to ensure they are not over treated. Furthermore, it has been hypothesized that some cases within this group are in fact LS, but that current detection

methods fail to identify the germline mutation[35]. As detection methods become more sensitive, the ability to distinguish true LS patients from SMMRD will improve.

Clinical testing often stops after germline mutation testing, even if those results are inconclusive. Those without a germline mutation but whom are MSI and MMR deficient are binned as LLS, often without further tumour testing to confirm this suspicion. Somatic mutation testing of MMR genes is beginning to gain popularity in LLS cases as confirmation of somatic biallelic inactivation can reduce patient anxiety and affect patient management[33]. However, tumour based testing can be expensive and complex, costing approximately $3,000 per tumour through the University Health Network and requiring tissue to be sent across international borders (estimate provided by Cancer Care Ontario).

1.4 Introduction to Lynch Syndrome

Lynch syndrome (LS), previously referred to as hereditary nonpolyposis colorectal cancer (HNPCC), is the most common inherited cancer susceptibility syndrome with a population prevalence of 1 in 440[45]. While LS has been associated with various cancers, it is most commonly associated with CRC and EC, responsible for 2-5% of cases[46,47], and 2-4% of cases (9% of cases diagnosed earlier than 50 years of age)[48], respectively.

Inherited in an autosomal dominant manner, LS is caused by a loss of function germline mutation in one allele of a MMR gene (or a deletion in *EPCAM* resulting in epigenetic silencing of *MSH2*) and cancer can arise when the second, wild type, allele becomes mutated[49,50]. The second allele can be affected through a variety of mechanisms, including mutation, deletion, methylation, structural rearrangements, and loss-of-heterozygosity (LOH)[11]. Since individuals with LS have inherited one non-functional allele, they have a high lifetime risk of getting various cancers, including a 52-82% lifetime risk of CRC[51,52] and a 25-60% lifetime risk of EC[53,54]. These ranges in values occur because different MMR genes and types of variants confer different cancer susceptibility risks, as shown in Table 1-1. It should be noted that individuals with LS have an increased incidence rate of metachronous and synchronous CRC

with a second primary CRC developing in up to 50% of patients after 15 years; women with LS and EC are also at risk for developing these secondary cancers[55].

	MLH1		MSH2		MSH6		PMS2	
	Freq in LS	Cancer risk	Freq in LS	Cancer risk	Freq in LS	Cancer risk	Freq in LS	Cancer risk
CRC	32%	25%-51%	38%	27%-60%	10-20%	16%	<5%	11-19%
EC	24-40%	40–50%	50-66%	40–50%	10-13%	64–71%	<5%	12%

Table 1-1: Mutation frequency and cancer risk
The frequency (Freq) each MMR gene is mutated in a LS CRC or EC tumour and the cancer risk associated with germline variants in that particular MMR gene [48–51,53–56]. The MMR genes are mutated at varying frequencies and confer different CRC and EC cancer risk. For example, the frequency of *MSH6* variants is similar in both EC and CRC patients; however, the associated cancer risk varies significantly with a high penetrance found in EC patients.

1.4.1 Diagnosis of LS

Accurate identification of individuals with LS is crucial as it allows for life-saving surveillance protocols to be initiated for the individual. In 1990, the Amsterdam Criteria were established, and later revised in 1998 to the Amsterdam II Criteria to incorporate extracolonic tumours and to identify families likely to have LS using personal and family histories[32,60]. In addition, in 1997 the Bethesda Guidelines were created to identify patients that would benefit from tumour screening for MSI. These guidelines were revised in 2004 to include family history and to incorporate histology; this involved a standardized panel of 5 microsatellite markers, as mentioned previously (Bethesda Panel)[12,13]. Table 1-2 highlights the components of the Amsterdam Criteria and Bethesda Guidelines.

Amsterdam I	Amsterdam II	Revised Bethesda
1. Three or more relatives, one should be first-degree, with confirmed CRC diagnosis	1. Three or more relatives, one should be first-degree, with confirmed Lynch syndrome-related cancers	1. CRC diagnosed in patient who is <50 years of age
2. Two successive affected generations	2. Two successive affected generations	2. Presence of synchronous, metachronous CRC, or other LS-related tumour, regardless of age
3. One or more CRCs diagnosed before age 50	3. One or more CRCs diagnosed before age 50	3. CRC with MSI histology who <60 years of age
4. Exclusion of FAP	4. Exclusion of FAP	4. CRC diagnosed in 1 or more first-degree relatives with a LS-related tumour, one of the cancer being diagnosed <50 years of age
		5. CRC diagnosed in 2 or more first-or second-degree relatives with LS-related tumours, regardless of age

Table 1-2: Amsterdam and Bethesda Criteria for diagnosing Lynch Syndrome
The family history based guidelines for screening for LS. The Amsterdam I Criteria were the first established screening guidelines for LS. These were replaced by the Amsterdam II Criteria to include tumours other than CRC. In 1997 the Bethesda Guidelines, followed by the revised Bethesda Guidelines in 2004, replaced the Amsterdam Criteria. FAP: Familial Adenomatous Polyposis (a hereditary CRC syndrome).

Several studies have shown that the Amsterdam Criteria's and the Bethesda/Revised Bethesda Guidelines miss a significant fraction of LS patients, as many LS families do not meet these criteria. A population-based study of CRC patients showed that only 23% of individuals with a germline pathogenic variant in a MMR gene met the Amsterdam criteria[61]. Furthermore, another study found the Amsterdam II criteria sensitivity to be 87%, 62%, 38%, and 48% for those with a germline pathogenic variant in MLH1, MSH2, PMS2, and MSH6, respectively[62]. MSH6-associated LS tends to occur later in life, compared to MSH2 and MLH1 associated cancers, meaning the child may develop cancer prior to their parents, also resulting in a negative family history[63]. Therefore, relying solely on these established family history guidelines is not sufficient to diagnose all LS patients.

Many centers now engage in molecular testing for LS that begins with either MSI tumour testing using the Bethesda Panel, or IHC staining of the MMR proteins in tumour tissue. IHC staining has emerged as the preferred starting point for LS testing as it is less expensive to perform, more widely available, and indicates the MMR gene that likely harbours the germline variant, informing subsequent testing[64–66]. In addition, it has a sensitivity of 83% and a specificity of 89%[14]. Several organizations, including the National Comprehensive Cancer Network and the

American Society of Clinical Oncology, have recommended universal screening of all CRC patients for LS[67,68]. Cancer Care Ontario, the Government of Ontario's advisor on health care services related to cancer, published an executive summary in 2015 recommending LS testing on all individuals diagnosed with CRC at <70 years of age[69]. As shown in Figure 1-3, LS testing is serial and involves both histological and molecular testing. Screening for LS begins with an IHC screen of the 4 MMR proteins. If the individual has intact protein their testing ends unless there is a strong family history of disease, in which case they are referred to genetic counselling or tested for MSI.

Under the guidelines of Cancer Care Ontario, an individual with either loss of PMS2, MSH2, or MSH6 protein expression in their tumour is referred to genetic counselling and is offered germline variant testing that can include sequence and copy number profiling. Testing for tumours with MLH1 protein loss is more complex. In CRC cases, BRAF p.V600E mutation testing is performed after IHC in MLH1 deficient cases. Individuals with a somatic BRAF p.V600E mutation and a family history (or are younger than 50 years of age) are referred to genetic counselling, otherwise there is no further action required as the disease is likely sporadic. MLH1 methylation testing is performed on the tumour of those without a BRAF p.V600E mutation, as a secondary test for sporadic disease. No further testing is performed on MLH1 somatic promoter methylated individuals unless the individual with CRC is young (< 50 years of age), or has a strong family history. This is because an individual could have MLH1 somatic methylation in concert with a germline MLH1 pathogenic variant. Individuals with no methylation are offered germline variant testing as LS is hypothesized in these individuals. Genetic testing for LS in EC patients is similar, except BRAF testing is not completed; therefore patients with MLH1 IHC deficiency are tested directly for promoter methylation.

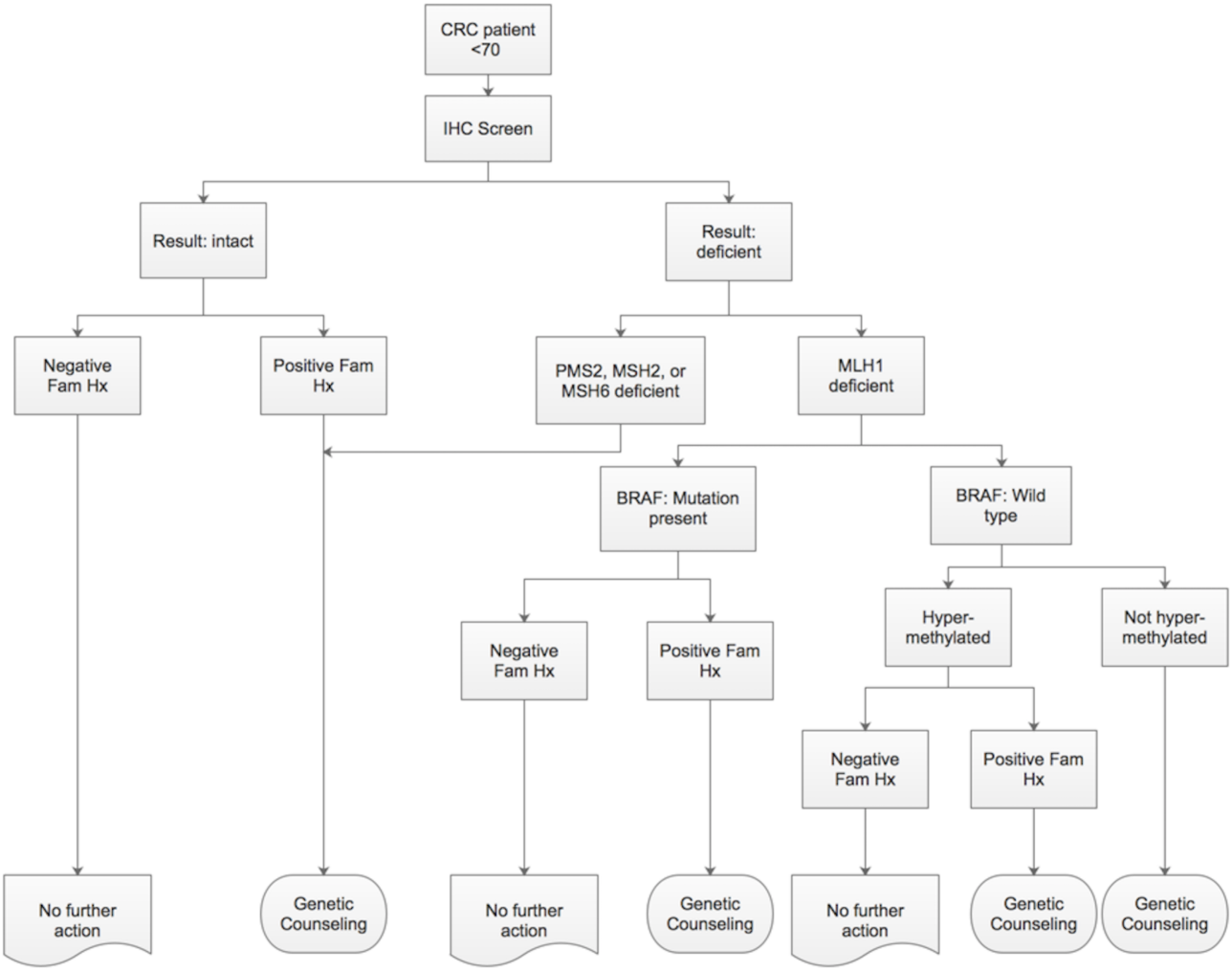


Figure 1-3: Cancer Care Ontario Guidelines for Screening CRC patients for LS
Cancer Care Ontario recommended screening guidelines for patients that have CRC diagnosed at less than 70 years of age. Screening commences with IHC testing of the four MMR proteins. Patients with intact MMR IHC and no family history are not sent for genetic counseling and all other patients are referred to genetic counseling and receive a cascade of testing dependent on the specific protein lost. Hx: history.

1.4.2 Treatment and Screening of LS

After an individual has been diagnosed with LS, multiple different treatment options are available including having their primary tumour surgically removed, preventative surgery, and being enrolled in extensive surveillance programs. Treatment depends on the age, disease site and stage (if LS is diagnosed after cancer diagnosis). Individuals diagnosed with LS prior to any disease occurrence are enrolled into surveillance programs with the goal of detecting and treating disease early.

It is recommended that individuals with either LS, or individuals with a first-degree relative diagnosed with LS, begin CRC surveillance (colonoscopy) at age 20-25, or 2-5 years earlier than the youngest individual in the family diagnosed with CRC[70]. One study showed individuals that engaged in colonoscopic surveillance had a 72% decrease in mortality related to CRC[71]. However, some studies support delayed colonoscopy screening for individuals with heterozygous *MSH6* or *PMS2* variants, as their CRC risk is lower[56]. Conversely, surveillance protocols for EC are not as well established, but surveillance and prophylactic treatment have been shown to reduce morbidity and mortality[72–74]. The literature has conflicting viewpoints on the effectiveness of endometrial biopsies and transvaginal ultrasound investigations, although preliminary evidence has shown biopsies to improve the detection of premalignant tumours[64]. One set of guidelines recommend that women be offered annual pelvic examination and endometrial biopsy, beginning at age 30-35[70]. Interestingly, it has been noted that the literature reveals no significant survival benefit from endometrial surveillance[70]. This may be due in part to the fact that most LS patients with EC present at stage I and already have a high (~88%) survival rate, therefore detecting a significant decrease in death is challenging[70].

Most individuals with LS and detectable cancer have their primary tumours surgically removed using the method that is preferred for their specific cancer type, except in the case of CRC. Since metachronous cancers are common in CRC patients with LS, a full colectomy with ileorectal anastomosis is recommended compared to a segmental/partial colonic resection[75–77]. This is because partial resection of the colon leaves the individual susceptible to additional cancers arising on the remaining colon tissue. Preventative surgery is not recommended for CRC patients

since routine colonoscopies are effective, as noted above. However, women diagnosed with LS may opt for prophylactic removal of the uterus and ovaries after childbearing is complete[70].

Interestingly, studies of LS patients have shown that taking aspirin can significantly reduce CRC risk[78]. Adults given 600 mg of aspirin daily saw greater than 60% reduction in CRC incidence if they took the aspirin for at least two years[78]. In addition, immune checkpoint therapies have shown success in treating MSI CRCs with recent research showing that MSI CRCs may benefit from Pembrolizumab, an anti-PD-1 monoclonal antibody[30]. Within this study, 78% of patients had significant improvement[30]. In 2017 Pembrolizumab became the first tissue-agnostic drug approved by the United States Food and Drug Administration[79]. It is approved for both adult and paediatric cancer patients with either MSI-H or MMR deficient solid tumours that are metastatic or unresectable; these are patients who are progressing following treatment[79]. LS and CMMRD patients, regardless of their tumour-of-origin, may benefit from immune checkpoint inhibitors and this approval leads the way for personalizing cancer care.

1.5 Introduction to Constitutional Mismatch Repair Deficiency

CMMRD is caused by inherited biallelic variants in one of the MMR genes. Individuals with CMMRD inherit one pathogenic variant in a MMR gene from each parent and present with gastrointestinal polyposis and cancer, lesions on the skin called café-au-lait (CAL) macules, brain tumours, and haematological malignancies[80,81]. Gastrointestinal, brain, and haematological cancers present in childhood and if patients survive to adulthood, additional cancers often emerge[80]. One study examined a cohort of 24 CMMRD cases and looked for a gastrointestinal phenotype and found that almost 80% of the patients had gastrointestinal polyps and/or cancer with some having additional cancers, such as glioblastoma[81].

There are some stark differences in the phenotype of LS patients and CMMRD patients. While CRC is common in both diseases, with two-thirds of CMMRD patients initially presenting with CRC, the mean age of diagnosis is 16.4 and 45 years of age for CMMRD patients and LS patients, respectively[80,82]. Moreover, in contrast to LS, left-sided CRCs are more common in

patients with CMMRD and parental consanguinity has been found in greater than 50% of individuals with CMMRD[63,80].

Subsets of tumours from CMMRD patients have been termed 'ultramutated'; these tumours have a very high mutation rate (>100 coding mutations/Mb) with variants evenly distributed throughout the genome[83]. This is in stark contrast to most childhood cancers, which typically have less than 1 coding mutation/Mb[83]. Some research on these patients has shown that the tumours are MSS (although this is debated) and are driven by early somatic mutations in *POLE* and *POLD1* in combination with MMR deficiency; this results in a rate of disease progression greater than LS patients[63,83].

1.5.1 Diagnosis and Treatment of CMMRD

Accurate diagnosis of CMMRD remains an active area of research due to overlapping phenotypes with other diseases, such as neurofibromatosis type 1 (NF-1). Approximately 97% of CMMRD patients have CAL macules, which are more irregular than classic CALs, as do NF-1 patients[80]. This results in some patients being misdiagnosed as having NF-1; these misclassified patients may miss early detection of their tumours or receive suboptimal care[80].

Once NF-1, Familial Adenomatous Polyposis, and MUTYH-associated Polyposis have been ruled out, family and personal history in combination with IHC are used as starting places for CMMRD testing. If IHC exhibits loss of expression of a MMR protein in both tumour and normal cells, germline mutation testing is performed to confirm CMMRD diagnosis[11,80]. It should be noted that, unlike LS, MSI testing is not recommended as an initial screening test for CMMRD as it has been shown that different tumours exhibit different MSI genotypes[80].

While family history is often a starting place for CMMRD testing, family history can be negative despite an individual having CMMRD due to low cancer penetrance for *PMS2* LS cancers. The most common gene associated with CMMRD is *PMS2* with one study finding that 50% of CMMRD patients with gastrointestinal cancer had confirmed or suspected *PMS2* mutations[81]. Penetrance for LS-associated cancers is low for *PMS2* monoallelic cancers; this means that a

child may have a CMMRD-associated cancer but their parents do not have cancer and may be unaware of the LS status, resulting in a negative family history[63].

Since CMMRD is a very rare and newly described disease, evidence-based screening guidelines have yet to be formalized and long-term patient outcomes have not been assessed. There is no consensus on the optimal screening and surveillance guidelines for individuals with CMMRD; however, the International CMMRD Consortium has established surveillance protocols for patients based on the cancer location[63,80].

1.6 Applications of NGS in Diagnostic Settings

Most MMR deficiency testing is completed in a stepwise method, testing a single gene at a time until diagnosis is confirmed. Not only can this process be time consuming, but it also requires a complex testing algorithm (requiring up to 6 laboratory tests) and expertise in multiple areas[69,84]. The invention of NGS technologies has revolutionized genetic testing research with multigene panels being introduced into the clinical diagnostic setting. NGS technologies are being used in the screening, diagnosis, and clinical assessment of patients, with an increasing number of assays being validated and entering the clinic annually[85]. For example, in 2015 the National Comprehensive Cancer Control Network released guidelines that recommended the use of targeted NGS panel for the hereditary breast and ovarian cancer patients that tested negative for high penetrance genes[85]. Multiple commercial (such as GeneDx[86] and Invitae[87,88]) and research laboratories[89] have developed targeted NGS panels for hereditary cancer syndromes, including LS. However, these panels do not include examination of promoter methylation, which is responsible for 80% and 89% of *MLH1* deficiency in CRC and EC tumours, respectively[33,34]. Therefore, multiple genetic tests are still required to reach diagnosis.

Despite the technical and interpretation challenges associated with NGS, including issues surrounding coverage of repetitive regions, it provides the opportunity to perform large-scale multi-gene analyses. NGS technology allows for the evaluation of multiple genes simultaneously, increasing the likelihood of detecting the individual's disease-causing

mutation[85,90]. In cases where the phenotype is atypical, cancer risk may be unrecognized by testing a single gene[90]. Moreover, NGS provides several benefits over its predecessor (Sanger sequencing) including lower DNA input, the ability to scale and analyze many samples and loci at one time, and the ability to detect large indels[91]. As sensitivity and specificity increase, in combination with lowering prices, NGS technologies will lead the way in providing patients with personalized medicine.

1.7 Project Outline

1.7.1 Rationale and Hypothesis

Initial screening for LS usually involves testing the tumour tissue for either MSI or protein expression of the four MMR genes through IHC[92]. The more prevalent adoption of screening using MMR IHC on the tumour implicates both germline and somatic variants as the cause of MMR deficiency. Attributing the cause of IHC loss of a MMR gene in an individual to either CMMRD, LS, or SMMRD is critical for the patient and their family and often requires the serial testing of MMR genes with different tests in both tumour and germline DNA. Importantly, within each gene screened, multiple individual tests may be required as a gene can become non-functional through various types of genome variation, including: point mutation, small insertion or deletion, copy number change, loss-of-heterozygosity, structural rearrangement, and methylation of a gene promoter. This process can be time consuming and involve multiple independent, complex tests that are sometimes not available in a single laboratory and each consume DNA from often limited clinical tissues.

Therefore, I hypothesize that the current genetic testing workflow for LS can be amalgamated into a singular NGS panel. Further, I hypothesize that this test will be able to detect SMMRD and CMMRD.

1.7.2 Objectives

Aim 1: Design a targeted NGS panel that comprises the genes associated with MMR.

Aim 2: Generate MMR profiles on a cohort of EC and CRC patients.

Aim 3: Confirm clinical results and solve clinically unsolved cases.

Chapter 2

Methods

2.1 Study Design

Over the course of test development, we deployed two versions of the MultiMMR Panel. Both include probes targeting A) exonic, intronic, and untranslated regions (UTRs) of the four genes responsible for MMR deficiency (*MLH1, MSH2, MSH6,* and *PMS2*), B) 41 SNPs, three sex markers, and the gene *ABL* targeted by the complementary Sequenom iPLEX Pro Sample Identification Panel (Agena Bioscience, USA)[93], C) the exonic regions of 3 secondary genes associated with MMR deficiency (*BRAF, EPCAM,* MLH3), and D) exons of the 2 genes responsible for the main inherited CRC polyposis syndromes (*APC* and *MUTYH*) (Table 2-1). Since there are multiple *PMS2* pseudogenes*,* additional probes were specifically designed to ensure regions unique to *PMS2* were captured. The first version of the panel (termed: MultiMMR V1) was tested on tumour and matched-normal samples from 12 EC patients. The subsequent MultiMMR V2 was modified to improve sequencing coverage of MMR genes. This included increasing coverage of the MMR genes by adding probes in regions that were not captured in V1 and by placing redundant probes in areas that previously achieved low coverage. In addition, I targeted a published set of 125 MSI sites[89], added probes targeting every exon of *POLE* and *POLD1*, and removed probes targeting *PMS1*, *TGFBR2*, and *BRAF,* except for exon 15 which contains the *BRAF* p.V600E locus.

2.2 Panel Design

2.2.1 MultiMMR V1

MultiMMR V1 contained 195 kbp of the genome with 3,185 probes spanning the targeted regions. Probes were organized into categories with each supporting a different aspect of MMR. The first probeset (‘Identity Panel’) contained 41 SNPs, 3 sex markers (*AMELX/Y, ARSDX/Y*, and *TGIF2LX/Y*), and the *ABL* gene from the Sequenom iPLEX Pro Sample Identification Panel (Agena Bioscience, USA) to prevent sample misidentification. The second probeset, termed ‘Classic Lynch genes’ contained all exons, introns, and regulatory regions of the MMR genes and the intergenic region between *EPCAM* and *MSH2*. The third probeset, termed ‘Lynch associated genes’, contained the exons and UTRs of genes that have been reported in the

literature to be associated with LS but are not drivers in the disease: *PMS1, MLH3, BRAF, EPCAM*, and *TGFBR2*. The fourth probeset contained exons and UTRs of two genes associated with hereditary CRC syndromes (*APC* for Familial Adenomatous Polyposis and *MUTYH* for MUTYH-associated Polyposis). The fifth probeset contained 46 MSI sites taken from the Tandem Repeat Finder database with between 2-10 microsatellite sites probed on each chromosome. The final probeset contained redundant probes that overlapped the above probesets and the regions covered by the SureSelect MethylSeq Target Enrichment System (Illumina, USA), for methylation analysis.

2.2.2 MultiMMR V2

Although the overall purpose of each probeset remained the same, the contents of each probeset in V2, with the exception of the Identity Panel, were modified. In total, 1,981 and 52 additional probes were added to fill any regions missing coverage, within the Classic Lynch genes and Hereditary CRC genes probesets, respectively. *PMS1* and *TGFBR2* were removed and the exons and UTRs of *POLE* and *POLD1* were added. This change was made in response to recent literature that has shown individuals with *POLE* or *POLD1* mutations have a syndrome called Polymerase Proofreading-associated Polyposis (PPAP), which has some clinical features similar to LS[94]. Furthermore, recent literature has shown that *POLE* mutations hold prognostic value[95]. *PMS1* and *TGFBR2* were initially added to V1 because a few case reports noted LS families with variants in these genes, however we chose to remove both genes since they are not drivers in LS, have been reported in few case reports, and provide limited clinical utility (as interpretation of these variants is challenging). Exon 15 of *BRAF* (which contains the *BRAF* p.V600E locus) was retained but all other exons were removed. The probeset containing MSI loci was completely recreated. The new MSI probeset has coverage of 125 microsatellite loci: all sites from the ColoSeq panel (University of Washington, USA) that had greater than 90% coverage (n=113), the Bethesda Panel microsatellite loci (n=5), and other microsatellite loci used in clinical laboratories to supplement the Bethesda Panel (n=7).

MultiMMR V1		
Category	**Genes Included**	**Probes (Size)**
Identity panel	41 SNPs, *ABL*, 3 sex markers	235 (18 kbp)
Classic Lynch genes	Full length regulatory, coding, and intronic regions of *MLH1*, *MSH2*, *MSH6*, and *PMS2* and the space between *EPCAM* and *MSH2*	1600 (106 kbp)
Lynch associated genes	Coding exons and regulatory regions in *EPCAM, PMS1, TGFBR2, MLH3*, and *BRAF*	401 (26 kbp)
Hereditary CRC genes	Exons and regulatory regions of *APC* and *MUTYH*	257 (16 kbp)
MSI sites	46 MSI sites	107 (8 kbp)
MethylSeq regions	Regions overlapping the above 11 genes, extracted from the commercial Agilent SureSelect Methyl-Seq panel	585 (21 kbp)
		3185 (195 kbp)

MultiMMR V2		
Category	**Genes Included**	**Probes (Size)**
Identity panel	41 SNPs, *ABL*, 3 sex markers	233 (18 kbp)
Classic Lynch genes	Full length regulatory, coding, and intronic regions of *MLH1*, *MSH2*, *MSH6*, and *PMS2* and the space between *EPCAM* and *MSH2*. **Additional probes to enhance coverage.**	3581 (256 kbp)
Lynch associated genes	Coding exons and regulatory regions of *EPCAM, MLH3,* ***POLE, POLD1,*** and ***BRAF* exon 8**	757 (44 kbp)
Hereditary CRC genes	Exons and regulatory regions of *APC* and *MUTYH*. **Additional probes to enhance coverage.**	309 (19 kbp)
MSI sites	**125 MSI sites**	272 (24 kbp)
		5152 (361 kbp)

Table 2-1: Genomic regions captured by MultiMMR V1 and V2
Comparison of the regions included in MultiMMR V1 and V2. Additional probes were added to MultiMMR V2 to improve coverage of all genes, especially promoter regions. Two genes were removed (*PMS1* and *TGFBR2*) and two genes were added (*POLE* and *POLD1*). MSI sites were completely altered in MultiMMR V2 based on more recent literature. Bolded text refers to changes between V1 and V2. kbp: kilo-base pair.

2.3 Study Cohort

DNA from all patients were de-identified and each patient was given a case number that was prefixed with N to denote the normal DNA sample and T to represent tumour DNA. All patients tested on MultiMMR V1 had DNA extracted from formalin-fixed paraffin embedded (FFPE) endometrial tumour tissue. For the matching normal DNA, 9 patients had DNA extracted from adjacent FFPE normal tissue and 3 had DNA extracted from peripheral blood. All EC cases were MSI-H and were MMR-deficient by IHC. For MultiMMR V2, normal DNA was extracted from peripheral blood and tumour DNA was extracted from FFPE tissue in all samples except T_1277 and T_0163 where tumour DNA was extracted from fresh-frozen tissue. Furthermore, MMR IHC was completed on all cases in both V1 and V2, with the exception of the four suspected unaffected relatives of 1277 (N_1279, N_1280, N_1283, N_1284). Germline and tumour DNA from 10 cases had libraries remade and were re-sequenced to assess between-run reproducibility and to test for batch effects. Two of these cases were also tested for within-run reproducibility. In addition, we sequenced a commercially-available DNA control sample (3x) that contained eleven 1000 bp fragments containing a MMR mutation placed within the center and added to genomic DNA isolated from the GM24385 lymphocyte cell line (Seraseq Inherited Cancer DNA Mix, SeraCare). These eleven MMR mutations consist of two substitutions, six small indels (<10 bases), two medium length indels (> 10 bases but < 20 bases), and one long indel (> 20 bases) that are challenging to detect by conventional assays and NGS[96].

2.4 Panel Library Preparation and Sequencing

DNA from blood (normal control) was extracted using the MaXtract High Density (QIAGEN, Germany), QIAamp DNA Mini Kit (QIAGEN, Germany), or the PAXgene Blood DNA Kit (QIAGEN, Germany). DNA from tissue was extracted using the QIAamp DNA FFPE Tissue Kit (QIAGEN, Germany), or the Dneasy Blood and Tissue Kit (QIAGEN, Germany). After extraction, up to 500 ng of DNA was sheared into 200 bp fragments using an ultrasonicator (Covaris LE220, USA). Three cases were re-sequenced using 250 ng of normal and tumour DNA to briefly assess the feasibility of using lower DNA input. Genomic libraries were prepared using the SureSelect Methyl-Seq Target Enrichment kit for Illumina (Agilent, USA) in MultiMMR V1.

In V2, the KAPA Hyper Prep Kit (Kapa Biosystems, USA) reagents for genomic library preparation were used in combination with the SureSelect Methyl-Seq Methylated Adapters (Agilent, USA). Target capture was performed according to an optimized protocol, with an 18 hour hybridization and 1:10 diluted baits from the SureSelectXT Human Custom Panel (Agilent, USA) I designed, with 2x tiling density.

To enable profiling of limited clinical specimens, DNA was kept in one aliquot during shearing and hybrid capture (Figure 2-1). For each DNA sample, we performed hybrid capture of a single DNA sequencing library constructed using methylated adapters (SureSelect Methyl-Seq Target Enrichment, Agilent, USA) to allow for downstream bisulfite and conventional sequence analysis. Bisulfite conversion was performed using the EZ DNA Methylation-Lightning Kit (Zymo Research, USA). Following hybridization, the captured products were split into two aliquots with 33% of the volume used for conventional sequencing and 66% of the volume used for bisulfite-treatment and methylation profiling. Distinct indexes were added during PCR amplification to prevent patient misidentification. In MultiMMR V1, four unique libraries were created per patient: one bisulfite treated germline DNA, one untreated germline DNA, one bisulfite treated tumour, and one untreated tumour. In MultiMMR V2, three unique libraries were created per patient with germline bisulfite sequencing performed only on cases where the tumour was methylated. Patients with only normal DNA available had two libraries created: one untreated and one bisulfite treated.

For all samples, the size distribution of the final product was verified by TapeStation 2200 (Agilent, USA) and MiSeq sequencing (Illumina, USA) was used to re-balance libraries and examine sample quality. Captured libraries were pooled and sequenced on the Illumina NextSeq 500 at the Princess Margaret Genomics Centre (http://www.pmgenomics.ca/) using paired-end sequencing (2 x 150 base pairs).

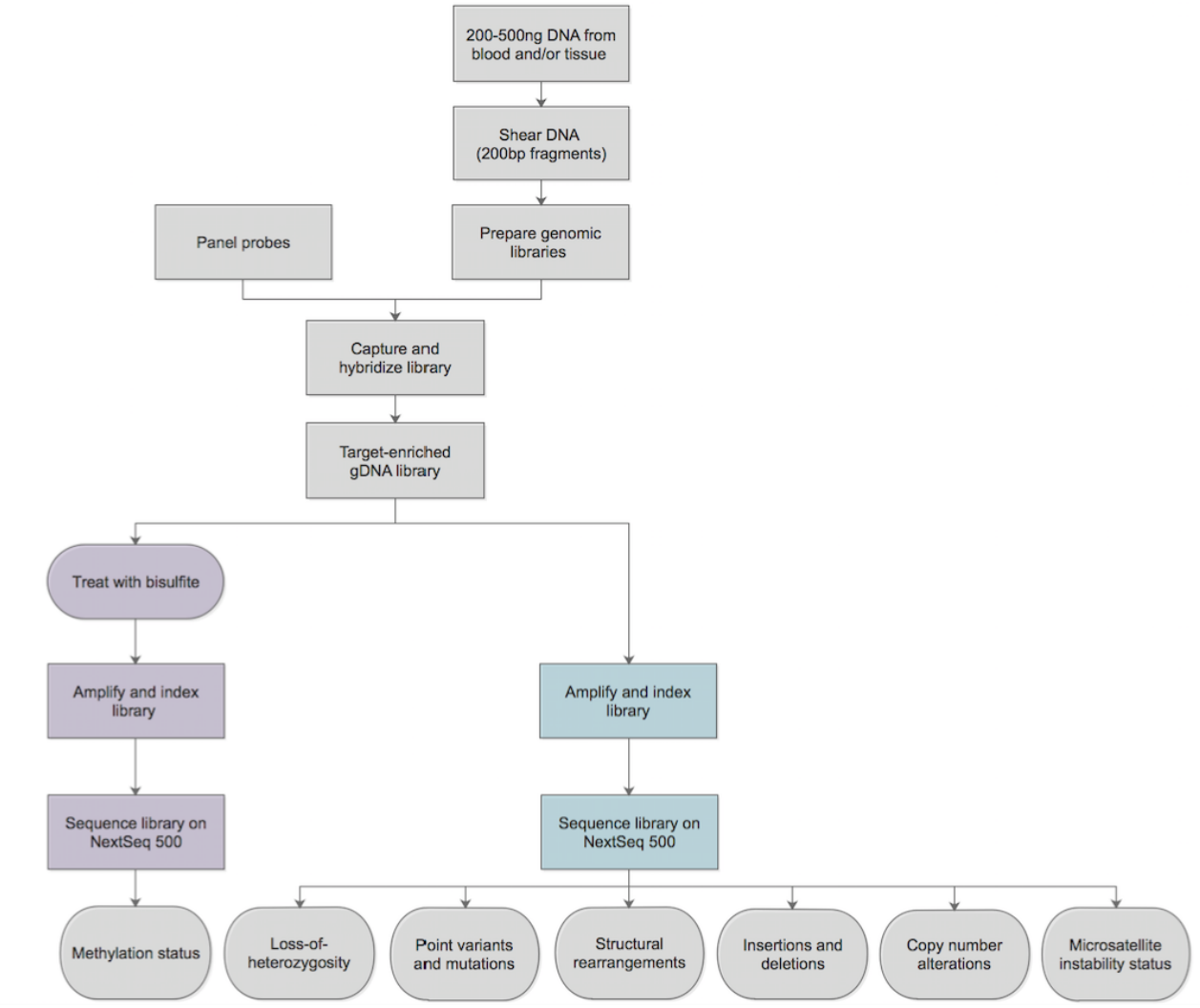


Figure 2-1: Laboratory workflow of the MultiMMR Panels
Overview of the laboratory workflow: from DNA extraction to the types of genome variation detected by MultiMMR V1 and V2. Purple: bisulfite arm; blue: conventional arm.

2.5 Panel Conventional Sequence Analysis

Raw sequencing reads were aligned to the human reference build (Illumina iGenomes UCSC hg19) using the Burrows-Wheeler Aligner (bwa) (version 0.7.12). After alignment, PCR duplicate reads were flagged using Picard's MarkDuplicates tool (version 1.140) and insertions and deletions (indels) were realigned using the Genome Atlas Toolkit (GATK) IndelRealigner (version 3.4-46). GATK's HaplotypeCaller was used for germline variant calling[97]. Somatic variant calling was performed using a union of MuTect[98], MuTect2[97], and Strelka[99]. The union of Strelka and MuTect2 was used for somatic insertion and deletion (indel) calling. Variant annotation was performed using Oncotator[100]. In addition, deletions/duplications and structural rearrangements were determined by our in-house tools, VisCap[101] and CluMP (available at: github.com/pughlab/CluMP), a tool that looks for clusters of reads with large insert sizes or soft-clipped bases supporting a common genomic breakpoint[102]. Loss-of-heterozygosity was determined by plotting the allelic fractions for all sequenced samples (Figure 2-2 and Table 2-2).

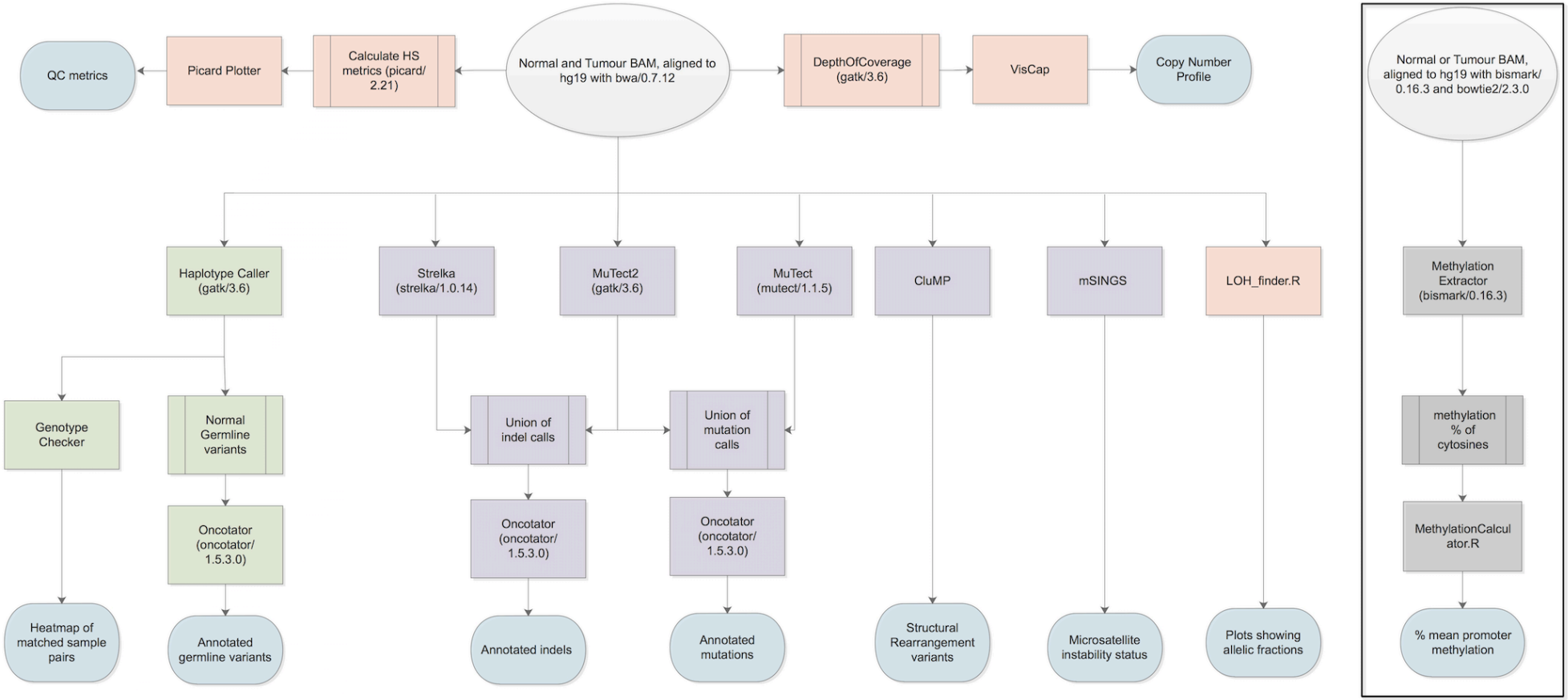


Figure 2-2: Bioinformatic workflow for the MultiMMR Panels
Overview of the bioinformatics tools used in the analysis of all samples. The green, purple, and salmon colours correspond to analysis completed on normal DNA, tumour DNA, or both normal and tumour DNA, respectively. Light blue corresponds to the final output from each branch. Boxes with two lines on the sides represent intermediate steps or tools used. The pipeline for bisulfite treated data is shown in the black box.

Tool Name	Purpose	Version	Additional Arguments
Picard Plotter	Asssess quality control	picard/2.2.1	
Depth of Coverage	Input to VisCap	gatk/3.6	
VisCap	Copy number		Ran normal samples against each other and removed all cases with deletions or gains, to improve panel of normals for tumour copy number calling
HaplotypeCaller	Variants	gatk/3.6	
Strelka	Point mutations and indels	strelka/1.0.14	
MuTect2	Point mutations and indels	gatk/3.6	max_alt_allele_in_normal_fraction= 0.01; max_alt_alleles_in_normal_count= 500; max_alt_alleles_in_normal_qscore_sum=20000
MuTect	Point mutations	mutect/1.1.5	downsampling_type=NONE
CluMP	Structural rearrangements	In-house script	
mSINGS	Detect MSI	Lastest version as of 11/30/17	Removed sites that have a standard deviation of 0; Removed loci positive in >10% of normal samples
LOH_finder.R	Visualize allelic fractions	In-house script	
Oncotator	Annotate variants	oncotator/1.5.3.0	output_format=TCGAMAF
Bismark	Alignment and extracting methylation values at each cytosine position	bismark/0.16.3	Bismark extractor arguments: -p, --no_overlap; --bedGraph; --comprensive
MethylationCalculator.R	Assess promoter methylation	In-house script	

Table 2-2: Software parameters used in the bioinformatics analysis
All bioinformatics tools, version number, and any additional arguments that deviate from default parameters. In-house scripts can be found at github.com/pughlab/CluMP or github.com/pughlab/MultiMMR.

2.6 Exome Sequence Analysis

2.6.1 Library Preparation and Sequencing

Whole exome sequencing was performed on 2 cases from MultiMMR V1. For each of these cases, 200 ng was extracted from both the tumour and the normal DNA and libraries were constructed following the protocol for the SureSelect XT Kit (Agilent, USA) with 750 ng of DNA used for input. Libraries were hybridized for 24 hours using baits from the SureSelect Human All Exon V5 + UTRs Kit (Agilent, USA). The Agilent Bio analyzer was used to verify the size distribution of the capture libraries. Cluster generation was performed on the Illumina cBot and then sequenced on the Illumina HiSeq 2500 instrument using 125 bp paired-end reads. Tumour and normal DNA samples were sequenced to achieve 250x and 50x coverage, respectively, at the Princess Margaret Genomics Centre.

2.6.2 Alignment, Processing, and Variant Calling

Raw sequencing reads were aligned and processed in the same manner as the targeted panel data. Variants were called using the same pipeline as the targeted panel with the addition of VarScan (version 2.3.8) to identify somatic indels and Sequenza (version 2.1.2) for copy-number calling, in addition to indel calling.

2.7 Variant Filtration and Interpretation

All variants with a population frequency >5% in the Exome Aggregation Consortium (ExAC, v0.3.1)[103] or Single Nucleotide Polymorphism (dbSNP, v146) databases were considered benign polymorphisms and removed from further analysis. In addition, variants were classified as benign if they were annotated as benign or likely benign in the ClinVar database (downloaded March 20th, 2017). To account for sequencing artefacts introduced as a result of formalin-induced DNA damage, which tends to occur at an allele frequency of less than 10%[104], only variants with an allele frequency greater than 10% were included. Reads supporting all candidate pathogenic variants were manually reviewed using the Integrative Genomics Viewer v2.3.91.

Variant interpretation was performed using existing variant classification frameworks. Germline variants were classified according to the 2015 American College of Medical Genetics and Genomics guidelines[105]. These guidelines group variants into one of five classification groups, based on the evidence available that supports the variant being either pathogenic or benign. In addition, all variants were queried on the International Society for Gastrointestinal Hereditary Tumours (InSiGHT) database. The InSiGHT database houses a similar 5-tiered variant classification system, which was used to confirm variant classification and interpretation[106].

2.8 MSI Analysis

MSI analysis was not performed successfully in MultiMMR V1 as the sites chosen could not discriminate MSI from MSS. All MSI sites were re-evaluated and altered in MultiMMR V2; 125 were covered. A validated tool, mSINGS, was used to distinguish MSI from MSS using the 125 microsatellite loci and defining the fraction of unstable microsatellite loci within the sample[20]. A panel of normal blood samples was used as a baseline. Each user-defined MSI locus was classified as stable or unstable based on a comparison of allele count between the baseline and a tumour sample at that particular site. The stable and unstable counts from each site were added to create a MSI score of the fraction of unstable loci within the sample. For the ROC analysis, the fraction of unstable loci was compared against the 33 tumours that underwent clinical MSI testing and all normal samples (with the exception of suspected CMMRD patients). All normal samples were classified as MSS and used in the ROC analysis, despite absent clinical testing results. A receiver-operator characteristic (ROC) analysis determined that 17% was the most discriminate threshold. Therefore, all samples with greater than 17% unstable loci (22/125) were classified as MSI.

2.9 Methylation Sequence Analysis

All tumour samples, and normal samples in cases without matching tumour, underwent bisulfite treatment. In cases where a tumour was classified as methylated, bisulfite treatment and sequencing was performed on the matched normal DNA to verify that the methylation was indeed somatic. Sequence reads from bisulfite-treated libraries were aligned to the human

reference genome (UCSC hg19) using the Bismark package (bismark, v0.16.3 and bowtie2, v2.3.0)**.** The Bismark package determines if a cytosine is methylated at each position within the panel. An in-house script parsed these positional files for each promoter CpG island of a MMR gene (determined by UCSC CpG islands) and calculated the percentage of cytosines methylated within that CpG island (script available at: github.com/pughlab/MultiMMR). If the mean methylation of all cytosines within the promoter regions was greater than 20%, as determined by an ROC analysis that included only tumours with clinical *MLH1* testing, the gene promoter was considered methylated.

To test our in-house methylation calculator, 19 DNA samples were sent for *MLH1* Methylation-Specific Multiplex Ligation-dependent Probe Amplification (MS-MLPA) using the MRC Holland ME011 Kit (Advanced Molecular Diagnostics Laboratory, Princess Margaret Cancer Centre). Methylation of each MMR gene promoter was manually reviewed using the Integrative Genomics Viewer v.2.3.91.

Chapter 3

Results

3.1 Samples from 80 patients passed quality control

In total, we sequenced DNA from 142 specimens (82 normal and 60 tumour samples) from 82 patients that collectively spanned the spectrum of MMR deficiency. To assess sequencing quality, I examined multiple quality control metrics from both picard (picard/2.21) and VisCap[101,107]. Sample quality was assessed by examination of total number of reads, mean target coverage, duplicate rate, number of unique molecules from picard, and the log2 interquartile range from VisCap (Figure 3-1). VisCap is a copy-number tool that measures sequencing quality by examining the spread of the log2 ratios of depth of coverage within a boxplot. A sample fails VisCap quality control if the boxplot whiskers extend past the theoretical log2 ratio of 0.58 (single-copy gain) and -1 (single-copy loss)[101]. Overall, samples failed quality control when the mean target coverage was less than 10x or the interquartile range from VisCap extended past the theoretical log2 ratio. In total, three samples were removed from our analysis: one normal sample (N_0001) failed VisCap quality control, one tumour sample (T_27302) failed both VisCap quality control and achieved a mean target coverage of only 1.88x, and one normal (N_27302) was removed because its matched tumour failed quality control. Ten samples received sequencing coverage below 30x (median coverage was 137x and 118x within MultiMMR V1 and V2, respectively) but remained in our cohort.

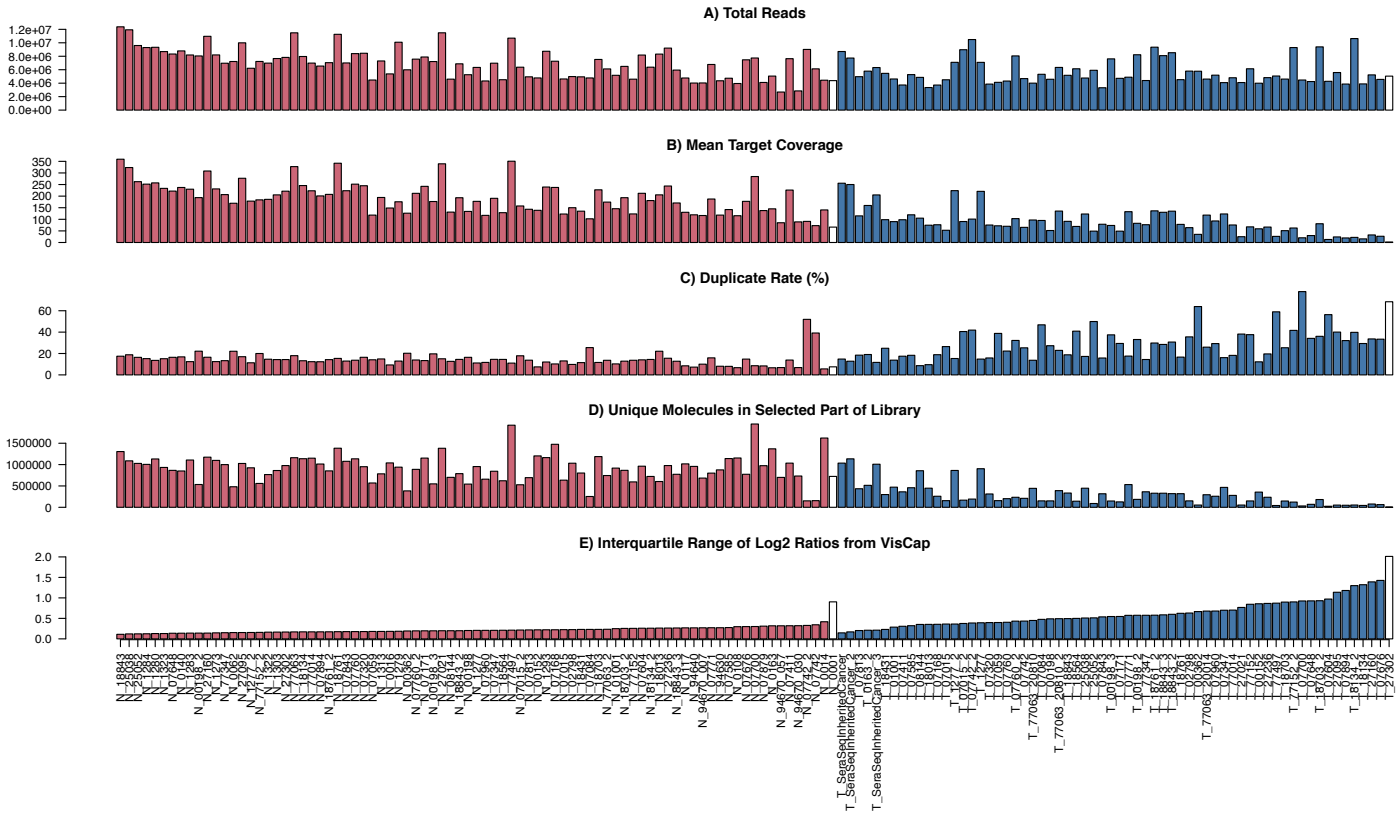


Figure 3-1: Quality control metrics across entire cohort
Quality control metrics extracted from picard/1.9.1 (A-D) and VisCap (E) for all samples. A) Total reads from the sequencer. B) Mean coverage within the region targeted. C) Percentage of reads marked as a duplicate. D) Number of unique molecules within the library. E) Interquartile range from of log2 ratios from VisCap. Pink: germline; blue: tumour; white: failed quality control.

Our sample cohort consisted 139 samples from 80 patients (80 normal, 59 tumours) that passed quality control (Table 3-1). Twelve patients with EC were tested on MultiMMR V1 (12 normal and 12 tumour samples) and the remaining 68 normal samples and 47 tumour samples were profiled on MultiMMR V2. Of these, 9 tumours were MSS and MMR intact. In 45 cases, MultiMMR Panel results were compared to corresponding clinical tests performed using gold standard assays (sequence analysis, MS-MLPA, and Multiplex Ligation-dependent Probe Amplification). Of these 45 patients, 24 received a definitive molecular diagnosis based on MMR loss. Clinical testing with a definite diagnosis refers to patients whose clinical testing was conclusive.

Panel Version	Cancer Type	Paired or Unpaired[a]	Patients	Clinically Solved[b]
V1	Endometrial	Paired	12	2
V2	Endometrial	Paired	1*	0
V2	Colorectal	Paired	42	12
V2	Colorectal	Unpaired	5	0
V2	CMMRD	Paired	3	1
V2	CMMRD	Unpaired	17	9
V2	Synthetic DNA	Unpaired	1	1

Table 3-1: Overview of sample cohort
[a]Unpaired: patients with only germline sequencing completed
[b]Clinically solved: patients that received a definitive diagnosis from a genetic counselor via conventional molecular testing
*One patient has 2 tumours (1 CRC, 1 EC) sequenced alongside their normal DNA

MultiMMR V2 correctly identified all 11 mutations within the SeraCare Seraseq Inherited Cancer DNA Mix (Table 3-2). No pathogenic variants in the MMR genes of the nine MSS negative control cases were detected. Two cases from MultiMMR V1 underwent whole exome sequencing (007 and 048) and the same MMR variants were detected in the exome and panel data. In addition, three normal and three tumours from MultiMMR V1 (029, 047, and 120) were re-sequenced on the MultiMMR Panel V2 to confirm mutational calling, which was consistent. These three cases were tested on MultiMMR V2 with both 500 ng and 250 ng input and results were consistent regardless of DNA input. Henceforth, results from both panels are combined unless explicitly specified (Figure 3-2).

Gene	Variant	Attribute	Detected in Replicates
MLH1	NM_000249:c.1852_1854delAAG	Repeat region	Yes
MLH1	NM_000249:c.232_243delinsATGTAAGG	Medium indel, complex call	Yes
MSH2	NM_000251:c.942+3A>T	Neighbouring a homopolymer (25bp)	Yes
MSH2	NM_000251:c.1662-12_1677del	Long indel	Yes
MSH6	NM_000179:c.2056_2060delinsCTTCTACCTCAAAAA	Medium indel, complex call	Yes
MSH6	NM_000179:c.2308_2312delGGTAAinsT	Short indel, complex call	Yes
MSH6	NM_000179:c.2641delGinsAAAA	Short indel, complex call	Yes
MSH6	NM_000179:c.3163_3164insG	Short indel	Yes
PMS2	NM_000535:c.2444C>G	Non unique region	Yes
PMS2	NM_000535:c.2243_2246delAGAA	Not unique region	Yes
PMS2	NM_000535:c.861_864delACAG	Repeat region	Yes

Table 3-2: Results from the SeraCare SeraSeq Inherited DNA Mix
All 11 variants were detected in the synthetic DNA mix in three replicates.

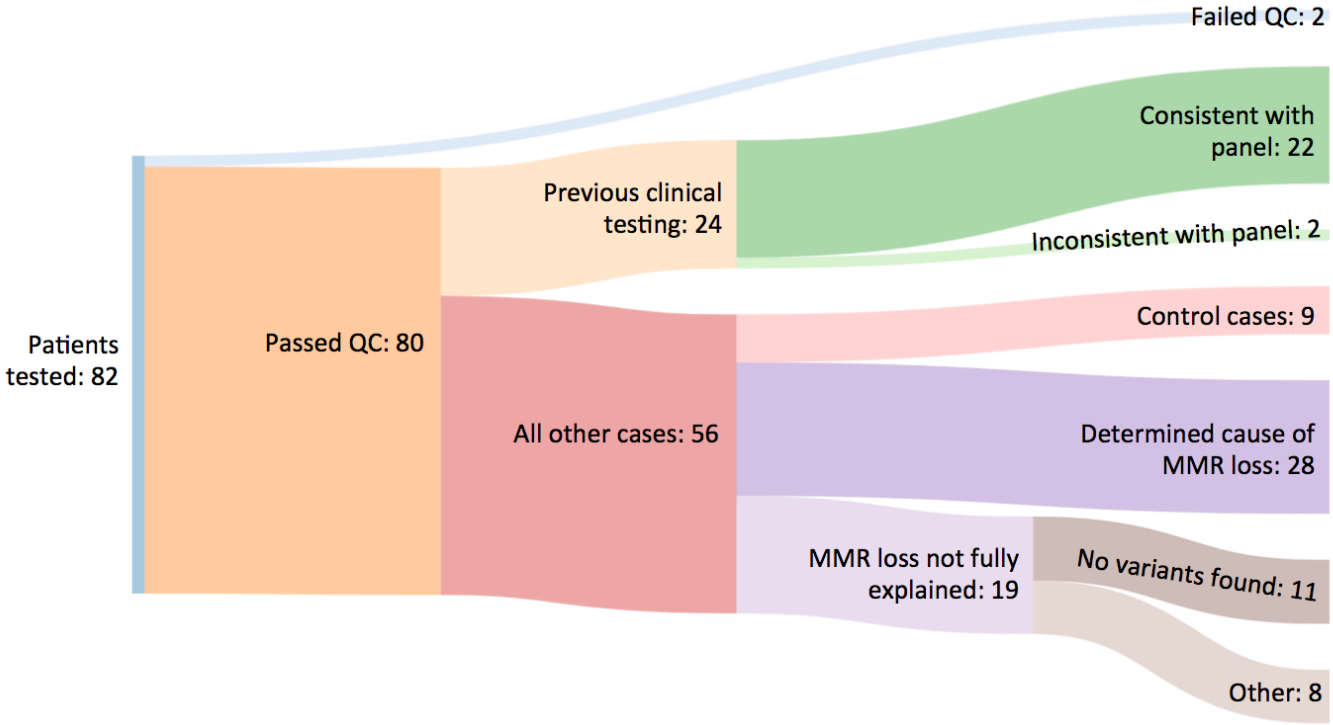


Figure 3-2: Overview of cohort results
Our cohort consisted of samples from 82 patients (one patient had metachronous tumours sequenced), with 80 passing quality control. After quality control, patients were split into two groups: those with definitive clinical testing (n=24) and those without (n=56). Within the cohort of patients without definitive testing, they were categorized by cases that acted as a control (MSS and MMR intact), those in which we could explain the cause of MMR deficiency (n=28) and those where we found no variants (n=11). 4/11 of the patients where no variants were found were suspected unaffected and 3/11 were suspected SMMRD and we only tested the normal tissue of these individuals. Cases where only one variant was found were classified as 'other' (n=8) since the second hit was not detected. We considered individuals with *BRAF* V600E mutations as solved, even if methylation of the *MLH1* promoter was not detected as clinical testing is often stopped after a positive *BRAF* mutation result.

3.2 MultiMMR recapitulates results from clinical testing

MultiMMR fully recapitulated the clinical testing result in 22 of 24 patients with a definitive explanation for MMR deficiency (2 EC, 12 CRC, 10 CMMRD; Table 3-3). In the ten CMMRD cases with clinical testing, our panel was consistent in 8/10 of the cases. In two inconsistent cases (Table 3-3, underlined), clinical testing found compound heterozygous hits in *PMS2* in both cases, involving a point mutation and a deletion of one exon. The panel found the point mutations but the two deletions (exon 14 and exon 10, NM_000535) were not identified. Our confidence to accurately call deletions in *PMS2* is restricted to deletions larger than one exon[101]. Furthermore high homology within exons 11-15 of *PMS2* hinders accurate deletion and variant calling. Many clinical labs are unable to confirm that deletions and mutations called in *PMS2* are not in fact from *PMS2CL*, a pseudogene.

In the two EC patients (098 and 103), clinical testing found that one individual had a germline *MSH2* intronic splice-site variant and the other had somatic *MLH1* promoter methylation. In addition to finding both of these variants, the MultiMMR Panel found an additional *MSH2* frameshift mutation (the second hit) in the tumour of the first individual. Of the 12 CRC patients with definitive clinical testing results, 6 were clinically diagnosed with LS. In all 6 LS cases, MultiMMR detected the known pathogenic germline variant as well as the compounding somatic mutation in the tumour. In the 6 sporadic CRC cases, clinical testing found somatic *BRAF* mutations (p.V600E) in two cases (00362 and 18431): I found the *BRAF* mutation in both and *MLH1* somatic promoter methylation in one. An additional three cases had *MLH1* somatic promoter methylation (07771, 18564, and 25038). Lastly, one case had somatic biallelic loss through two nonsense mutations in *MSH2* (77152).

Interestingly, three of these patients have known LS founder mutations (Table 3-3; denoted by #). Two patients have founder mutations specific to Canada: patient 01001 has a Newfoundland founder mutation in *MSH2* (c.942+3A>T, NM_000251) and patient 1303 has an Inuit founder mutation in *PMS2* (c.2002A>G, NM_000535). The *MSH2* founder mutation is located beside a microsatellite tract and thus often is hard to detect due to low-mapping quality. Patient 07585 has a European founder mutation in *PMS2* (p.137G>T, NM_000535). This patient has an additional

germline variant at the same amino acid site as the founder mutation (c.137G>A) resulting in CMMRD.

			Clinical and Molecular Results			Panel Explanation	
Case	**Paired/Unpaired**	**Tissue Type**	**IHC Loss**	**Clinical Germline Result**	**Clinical Somatic Result**	**Germline Result**	**Somatic Result**
098	Paired	Endometrial	MSH2/MSH6	*MSH2*:c.1661+5G>C		*MSH2*:c.1661+5G>C	***MSH2*:c.1463delTG(p.Leu488fs)**
103	Paired	Endometrial	MLH1/PMS2		*MLH1*: Methylated		*MLH1:*Methylated (33%)
0108	Unpaired	CMMRD	PMS2	*PMS2*:c.1261C>T(p.Arg421*) and *PMS2*:c.2531C>A(p.Pro844His)		*PMS2*:c.1261C>T(p.Arg421*) and *PMS2*:c.2531C>A(p.Pro844His)	
0117	Unpaired	CMMRD	PMS2	*PMS2:*exon 9-10 deletion		*PMS2*:exon 9-10 deletion	
0062	Unpaired	CMMRD	PMS2	*PMS2*:c.2458_2459insA(p.Thr820fs)		*PMS2*:c.2458_2459insA(p.Thr820fs)	
1273	Unpaired	CMMRD	PMS2	*PMS2*:c.1831_1832insA(p.Ile611fs)		*PMS2*:c.1831_1832insA(p.Ile611fs)	
1303	Unpaired	CMMRD	PMS2	*PMS2*:c.2002A>G(p.Ile668Val)#		*PMS2*:c.2002A>G(p.Ile668Val)	
1313	Unpaired	CMMRD	PMS2	*PMS2*:c.1738A>T(p.Lys580Ter)		*PMS2*:c.1738A>T(p.Lys580Ter)	
0140	Unpaired	CMMRD	PMS2	*PMS2*:c.209A>T(p.Asp70Val) and *PMS2*:exon 14 deletion		*PMS2*:c.209A>T(p.Asp70Val)	
1323	Unpaired	CMMRD	PMS2	*PMS2*:c.1831_1832insA(p.Ile611fs) and *PMS2*: exon 10 deletion		*PMS2*:c.1831_1832insA(p.Ile611fs)	
07585	Paired	CMMRD	PMS2	*PMS2*:c.137G>A(p.Ser46Asn) and *PMS2*:c.137G>T(p.Ser46Ile)#		*PMS2*:c.137G>A(p.Ser46Asn) and *PMS2*:c.137G>T(p.Ser46Ile)#	
07879	Unpaired	CMMRD	PMS2	*PMS2*:exon 14-15 deletion		*PMS2*:exon 14-15 deletion (NM_000535)	
00152	Paired	Colorectal	MSH6 Equivocal	*MSH6*:c.3219_3220delT(p.Met1074fs)		*MSH6*:c.3219_3220delT(p.Met1074fs)	***MSH6*:CN-LOH**
00362	Paired	Colorectal	MLH1 and MSH6		*BRAF*:c.1799T>A(p.Val600Glu)		*BRAF*:c.1799T>A(p.Val600Glu)
01001	Paired	Colorectal	MSH2	*MSH2*:c.942+3A>T(p.Val265_Q314del)#		*MSH2*:c.942+3A>T(p.Val265_Q314del)#	***MSH2*:c.1111G>T(p.Glu371*)**
07742	Paired	Colorectal	MSH2 and MSH6	*MSH2*:c.610_611insGAGA(p.-205fs)		*MSH2*:c.610_611insGAGA(p.-205fs)	***MSH2*:c.425C>G(p.Ser142*)**
07771	Paired	Colorectal	MLH1 and MSH6		*MLH1*:Methylated		*MLH1*:Methylated(48%)
07813	Paired	Colorectal	MLH1	*MLH1*:c.1039-1409del(p.T347xfs)		*MLH1*:Exon 12 deletion (NM_001258271)	***MLH1*:c.340A>C(p.Thr114Pro)**
07960	Paired	Colorectal	PMS2	*PMS2*:c.903+1G>A		*PMS2*:c.903+1G>A	***PMS2*:c.1939A>T(p.Lys647*)**
18431	Paired	Colorectal	MLH1		*BRAF*:c.1799T>A(p.Val600Glu)		***MLH1*:Methylated(37%)** and *BRAF*:c.1799T>A(p.Val600Glu)
18564	Paired	Colorectal	MLH1		*MLH1*:Methylated		*MLH1*:Methylated(22%)
77152	Paired	Colorectal	MSH2		*MSH2*:c. 1216C>T(p.Arg406*) and MSH2:c.1777C>T(p.Gln593*)		*MSH2*:c. 1216C>T(p.Arg406*) and MSH2:c.1777C>T(p.Gln593*)
25038	Paired	Colorectal	MLH1 and MSH2		*MLH1*:Methylated		*MLH1*:Methylated(42%)
27095	Paired	Colorectal	MLH1 and PMS2	*MLH1*:c.1989G>T(p.Glu663Asp)		*MLH1*:c.1989G>T(p.Glu663Asp)	***MLH1*:CN-LOH**

Table 3-3: Summary of 24 patients with complete clinical testing results

Summary of clinical testing and MultiMMR results for the 24 patients that underwent clinical testing with a definitive result. MultiMMR was consistent in 22/24 cases. The two inconsistent cases are underlined and founder mutations are denoted with a #. Bolded text refers to variation depicted by MultiMMR panel and not clinical testing. CN-LOH: copy-neutral loss-of-heterozygosity.

3.3 MultiMMR captures additional genome variation

The remaining 47 patients tested had incomplete clinical genetic testing; either clinical testing was not performed, is currently underway, or was inconclusive. MultiMMR was able to explain the loss of MMR in 28 of these patients, including 10 patients with EC, 2 with CMMRD, and the remaining 16 with CRC. Cases were deemed solved when both hits were detected, except when *BRAF* p.V600E mutations were present as a *BRAF* mutation alone can indicate somatic origin and halt diagnostic testing.

The panel could explain the loss of MMR within 10 ECs. Four cases had biallelic somatic inactivation of a MMR gene: two by compound frameshift deletion and copy-neutral loss-of-heterozygosity in *MSH2* and two with compound nonsense and frameshift mutations in *MSH6* and *MSH2*. *MLH1* somatic promoter methylation explained the *MLH1* loss in five patients consistent with lack of protein expression by IHC. The last individual (07015) underwent clinical testing in 2007 (sequencing and MLPA) and no variant was found, despite having IHC loss of MSH2. Re-testing of this individual on MultiMMR found a deep intronic germline *MSH2* variant[108] compounded by somatic *MSH2* loss-of-heterozygosity (Figure 3-3). This deep intronic *MSH2* variant (c.212-478T>G, NM_000251) was previously reported in the literature and creates a *de novo* exon with a stop codon at the end of the exon that is predicted to truncate the protein from 934 amino acids to 94[108]. Within the remaining 16 CRC cases, multiple sources of genome variation were found. These included 8 pathogenic mutations, 4 frameshift deletions, 1 copy-neutral loss-of-heterozygosity event, 2 copy-number deletions, 7 promoters with *MLH1* methylation, and 5 cases with somatic *BRAF* p.V600E mutations.

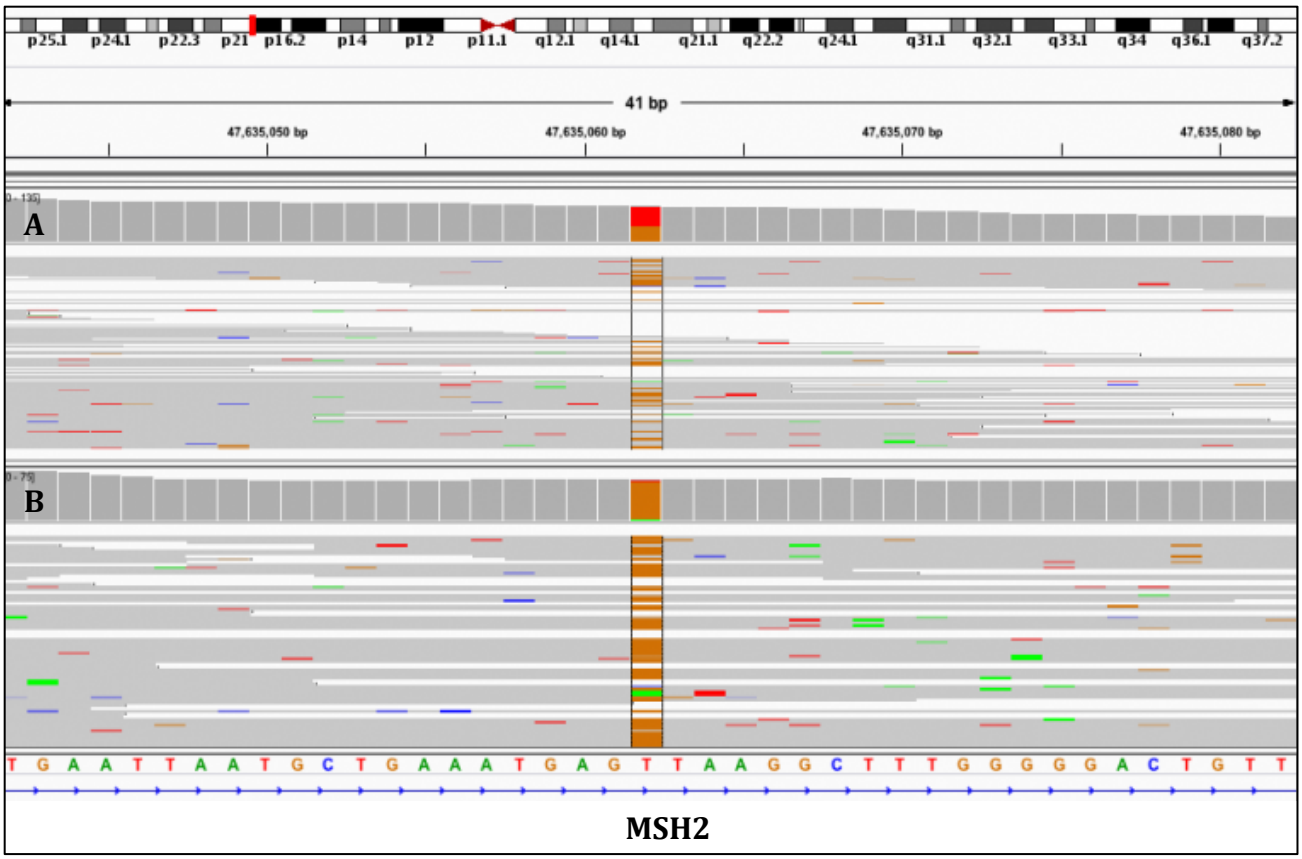
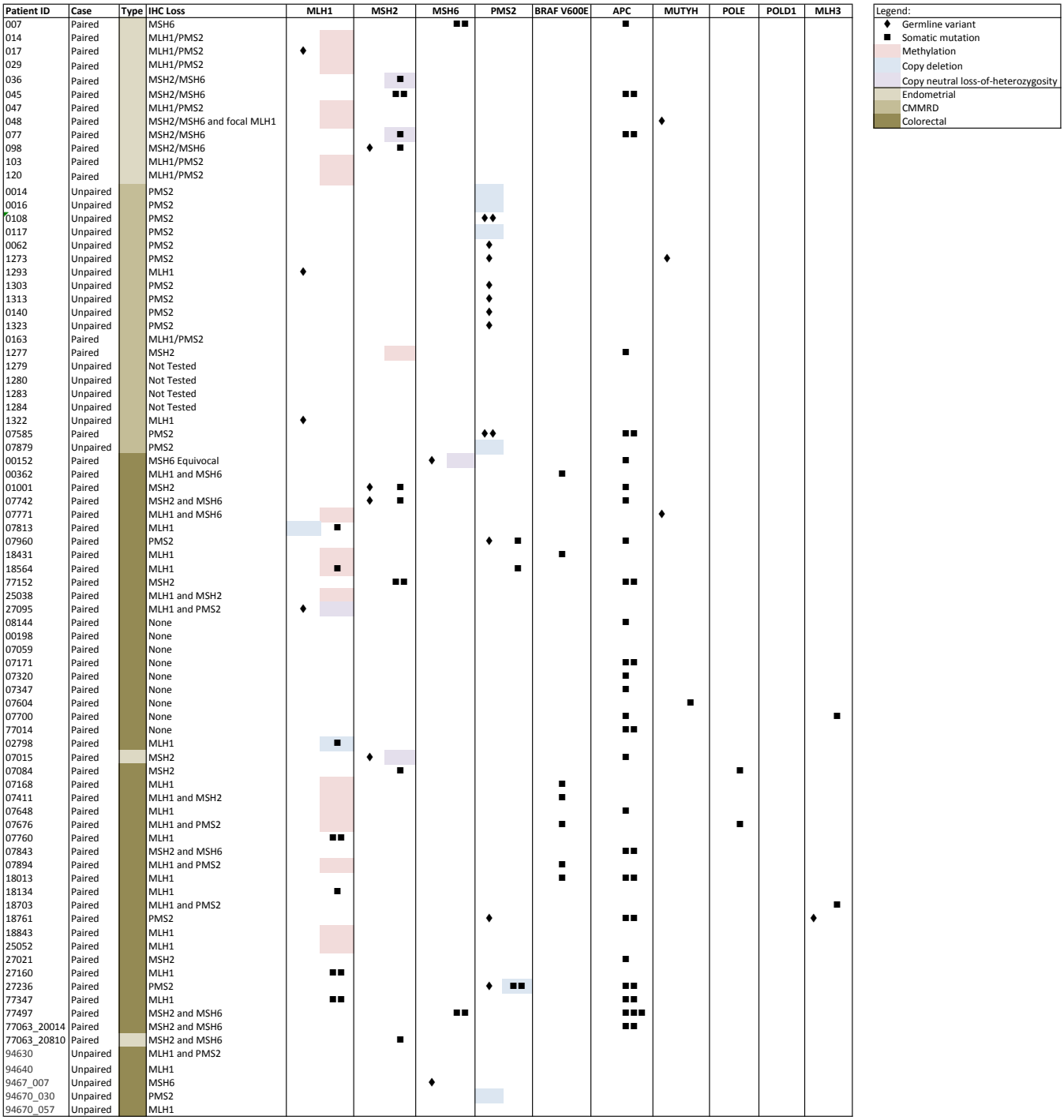


Figure 3-3: Deep intronic *MSH2* variant missed by clinical testing
An Integrative Genomics Viewer screenshot of the A) normal sequence reads and B) tumour sequence reads from patient 07015. The orange area shows a germline intronic variant present in both the normal and tumour DNA from the individual. Gray lines refer to sequence reads.

Both suspected-CMMRD patients, 1293 and 1322, had loss of *MLH1* with no clinical testing performed on them. I found an *MLH1* homozygous and heterozygous splice site variant in blood samples from 1293 and 1322, respectively. Since the *MLH1* variant in 1322 is heterozygous, it indicates that this individual has LS and was misdiagnosed as CMMRD likely due to a strong family history of disease. Review of clinical and IHC staining confirmed this finding.

In the remaining 19 patients where loss of MMR was not fully explained, mutations were found in 9 cases but the second-hit, which would explain the protein loss, was not found. Of the remaining 11 cases, three were suspected SMMRD and no tumour DNA was available for sequencing. Germline sequencing found no pathogenic aberrations, consistent with clinical observation. Interestingly, four cases (07843, 18703, 27021, and 77063_20014) had IHC loss of a MMR protein but no germline variants or somatic mutations were found. Four relatives of suspected CMMRD patient 1277 had their germline DNA sequenced and no pathogenic variants were found in any of the family members. It should be noted that no structural rearrangements were found within our cohort. One patient (77063) had their normal DNA and two tumours (one EC and one CRC) sequenced. No germline mutations were found in this individual; however, the endometrial tumour (77063_20810) had a *MSH2* frameshift deletion while no variants were found within the colorectal tumour (77063_20014).

While our analysis was focused on MMR genes, variants were found in the additional genes added to MultiMMR: *POLE, POLD1, MLH3, MUTYH,* and *APC* (Table 3-4). Many genome variants found within these genes were classified as variants of uncertain significance (VUS), with a few known pathogenic mutations detected. A germline *MUTYH* hotspot mutation (c.536A>G; p.Tyr179Cys) associated with *MUTYH*-associated Polyposis was present in the germline of one EC (048) and one CRC (07771)[109,110]. Interestingly, in both samples this variant was heterozygous and no second-hit was found. Many CRC tumours within the cohort carried either known somatic pathogenic mutations (i.e. c.637C>T; p.Arg213Ter) or somatic VUS mutations within *APC*, alongside variants in MMR.



Patient ID	Case	Type	IHC Loss	MLH1	MSH2	MSH6	PMS2	BRAF V600E	APC	MUTYH	POLE	POLD1	MLH3
007	Paired		MSH6			■■			■				
014	Paired		MLH1/PMS2										
017	Paired		MLH1/PMS2	♦									
029	Paired		MLH1/PMS2										
036	Paired		MSH2/MSH6		■								
045	Paired		MSH2/MSH6		■■				■■				
047	Paired		MLH1/PMS2										
048	Paired		MSH2/MSH6 and focal MLH1							♦			
077	Paired		MSH2/MSH6		■				■■				
098	Paired		MSH2/MSH6		♦ ■								
103	Paired		MLH1/PMS2										
120	Paired		MLH1/PMS2										
0014	Unpaired		PMS2										
0016	Unpaired		PMS2										
0108	Unpaired		PMS2				♦♦						
0117	Unpaired		PMS2										
0062	Unpaired		PMS2				♦						
1273	Unpaired		PMS2				♦			♦			
1293	Unpaired		MLH1	♦									
1303	Unpaired		PMS2				♦						
1313	Unpaired		PMS2				♦						
0140	Unpaired		PMS2				♦						
1323	Unpaired		PMS2				♦						
0163	Paired		MLH1/PMS2										
1277	Paired		MSH2						■				
1279	Unpaired		Not Tested										
1280	Unpaired		Not Tested										
1283	Unpaired		Not Tested										
1284	Unpaired		Not Tested										
1322	Unpaired		MLH1	♦									
07585	Paired		PMS2				♦♦		■■				
07879	Unpaired		PMS2										
00152	Paired		MSH6 Equivocal			♦			■				
00362	Paired		MLH1 and MSH6					■					
01001	Paired		MSH2		♦ ■				■				
07742	Paired		MSH2 and MSH6		♦ ■				■				
07771	Paired		MLH1 and MSH6							♦			
07813	Paired		MLH1	■									
07960	Paired		PMS2				♦ ■		■				
18431	Paired		MLH1					■					
18564	Paired		MLH1	■			■						
77152	Paired		MSH2		■■				■■				
25038	Paired		MLH1 and MSH2										
27095	Paired		MLH1 and PMS2	♦									
08144	Paired		None						■				
00198	Paired		None										
07059	Paired		None										
07171	Paired		None						■■				
07320	Paired		None						■				
07347	Paired		None						■				
07604	Paired		None							■			
07700	Paired		None						■				■
77014	Paired		None						■■				
02798	Paired		MLH1	■									
07015	Paired		MSH2		♦				■				
07084	Paired		MSH2		■						■		
07168	Paired		MLH1					■					
07411	Paired		MLH1 and MSH2					■					
07648	Paired		MLH1						■				
07676	Paired		MLH1 and PMS2					■			■		
07760	Paired		MLH1	■■									
07843	Paired		MSH2 and MSH6						■■				
07894	Paired		MLH1 and PMS2					■					
18013	Paired		MLH1					■	■■				
18134	Paired		MLH1	■									
18703	Paired		MLH1 and PMS2										■
18761	Paired		PMS2				♦		■■				♦
18843	Paired		MLH1										
25052	Paired		MLH1										
27021	Paired		MSH2						■				
27160	Paired		MLH1	■■									
27236	Paired		PMS2				♦ ■■		■■				
77347	Paired		MLH1	■■					■■				
77497	Paired		MSH2 and MSH6			■■			■■■				
77063_20014	Paired		MSH2 and MSH6						■■				
77063_20810	Paired		MSH2 and MSH6		■								
94630	Unpaired		MLH1 and PMS2										
94640	Unpaired		MLH1										
9467_007	Unpaired		MSH6			♦							
94670_030	Unpaired		PMS2										
94670_057	Unpaired		MLH1										

Table 3-4: Types of genome variation found across the cohort

This table depicts the various types of genome variation (denoted in the top half of the legend) within the various tumour types (bottom half of legend) within our cohort. Only variants with greater than 10x coverage and with a variant interpretation of likely pathogenic or pathogenic were included in this table. In addition, our cohort consisted of 2 samples with germline VUS variants in *POLE* and 3 samples with somatic VUS *POLE* mutations. In addition, somatic *POLD1* mutations classified as VUS were found in 2 samples.

3.4 MSI detected in targeted sequence data

The mSINGS algorithm was used to determine MSI status in the MultiMMR Panel V2 (Figure 3-4A)[20]. A receiver-operator characteristic (ROC) analysis on our cohort showed the maximum sensitivity and specificity at a threshold of 17% (22/125) (Figure 3-4B). In cases with clinical MSI testing, the mSINGS algorithm accurately detected MSI status in 32/33 cases using 125 loci (sensitivity: 100%, specificity: 96%). The normal samples and MSS MMR intact tumours were used as a negative control and the mSINGS algorithm did not classify any of these samples as MSI. All tumours without MSI clinical testing, except two (T_18013 and T_18703), were classified as MSI. Interestingly, T_18703 had MMR loss of *MLH1* and *PMS2* by IHC but no germline or somatic variants were detected.

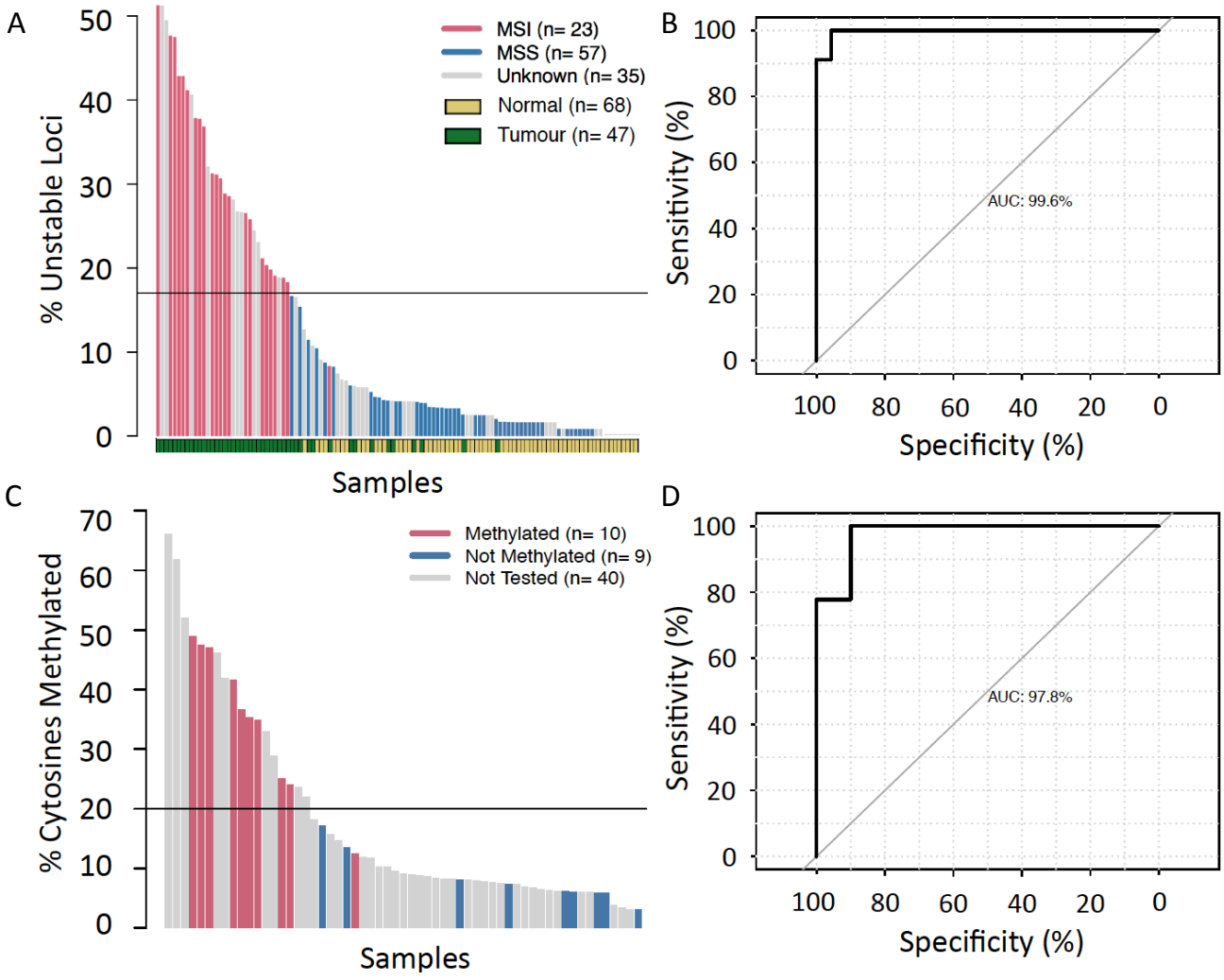


Figure 3-4: Microsatellite instability status and *MLH1* promoter methylation results
A) Fraction of unstable loci from all 115 samples tested in MultiMMR V2, determined by mSINGS. B) Receiver operator characteristic (ROC) curve of mSINGS fraction of unstable loci within tumour samples that underwent clinical MSI testing. Germline samples from all LS, SMMRD, and MSS tumour intact were assumed MSS (n= 48). C) Percentage of cytosines methylated within the *MLH1* promoter (defined by the CpG island location) of all tumours (n=59). D) ROC curve of in-house methylation script for the promoter of *MLH1* compared to the current gold-standard *MLH1* MS-MLPA.

3.5 Promoter methylation found in a subset of MMR deficient cases

Nineteen tumours had somatic *MLH1* promoter methylation and one suspected CMMRD patient had somatic *MSH2* promoter methylation, using a threshold of 20% (ROC analysis: sensitivity of 100% and a specificity of 90%). A subset of methylated and unmethylated tumours (n=19) were sent for clinical *MLH1* MS-MLPA testing. The panel's classification had 95% concordance with MS-MLPA, correctly classifying 18/19 (Figure 3-4C; Figure 3-4D). In the inconsistent case (T_18013), our method determined the tumour to have 13% promoter methylation, and MS-MLPA found 34% methylation (with two probes positive). Promoter *MLH1* methylation did not appear to be 34% by manual, visual inspection (Figure 3-5). However, the MS-MLPA probe locations appear methylated (Figure 3-6). For the 19 tumours with *MLH1* promoter somatic methylation, the normal DNA for these cases were bisulfite treated and analyzed; no germline *MLH1* or *MSH2* methylation was found.

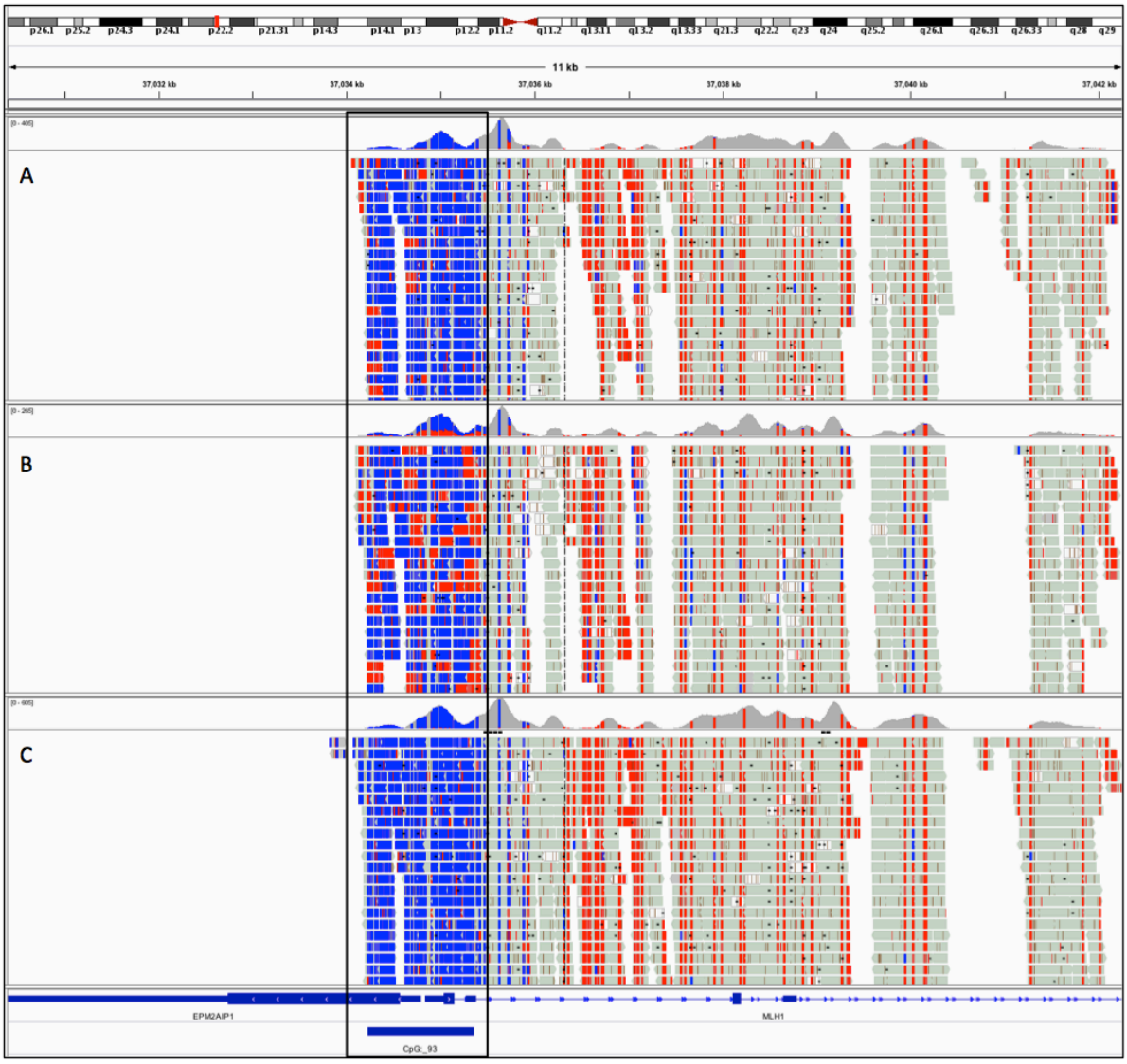


Figure 3-5: Visual comparison of *MLH1* methylation across two patients
Integrative Genomics Viewer display of the promoter of *MLH1* (black rectangle) in two individuals with somatic *MLH1* promoter methylation by MS-MLPA. We classified the promoter region as chr3:37034616-37036342 (hg19). Red: methylated cytosines; blue: unmethylated cytosines; gray: sequencing reads and nucleotides that are not cytosine. Wiggle track depicts sequencing coverage across the region. A) Tumour from 18013 that is inconsistent between our method and MS-MLPA. MS-MLPA classified the tumour as 34% methylated and our algorithm classified this tumour as unmethylated, with only 13% methylated. B) Tumour from 07168, found to be methylated by both our algorithm and MS-MLPA. C) Unmethylated normal DNA from 07168.

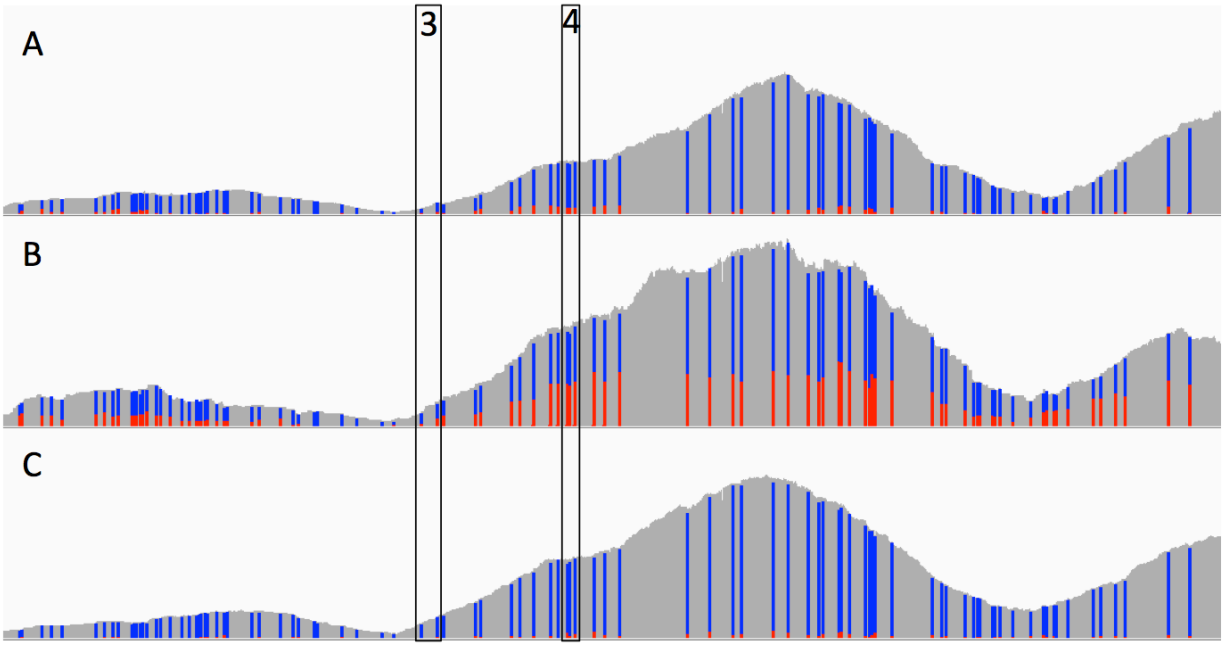


Figure 3-6: Visual comparison of *MLH1* MS-MLPA probes across two patients
This Integrative Genomics Viewer screenshot examines the wiggle track from the Figure 3-4 samples for two MS-MLPA probe sites (depicted as 3 and 4). These probe sites called A) tumour of 18013 and B) tumour of 07168 methylated and C) normal DNA from 07168 unmethylated. From this zoomed in image, it is clear to see that cytosine's outside the probe areas are methylated in track B more so than track A. Almost all cytosines in track B are methylated, whereas fewer cytosines are methylated in track A. Moreover, visual inspection of methylation shows that the percentage of methylation (length of red bars) at each site in track A is in between track B and track C.

3.6 Between and within run reproducibility achieved on MultiMMR V2

To test between-run reproducibility, 10 germline-tumour pairs had new libraries created and were sequenced on the same sequencer at a different period of time. Within-run reproducibility was examined with two of these patients, in which two independent libraries were created for the same patient's germline and tumour DNA and they were sequenced on the same lane of the sequencer, at the same time. In addition, the Seraseq Inherited Cancer DNA Mix was used to confirm between-and-within run reproducibility. Concordance of pathogenic variants was found in all cases except one (case: 18134), where a somatic pathogenic *MLH1* splice site variant (c.1668-1G>A) was found in the original sequencing run but not the re-run. In the original sequencing run, the *MLH1* splice site variant was found with 20x coverage and 8 reads supporting the G>A substitution. Manual inspection found this variant in the re-run, with 9x coverage and 3 reads supporting the variant. Low coverage in this region is due to nearby repetitive elements. The re-run sample achieved less than 10x coverage, explaining why this variant was missed.

3.7 Pseudogenes affect targeted sequence quality

MultiMMR V1 and MultiMMR V2 achieved a median coverage of 137x and 118x, respectively. On-target rate was 13% in MultiMMR V1 and 8% in MultiMMR V2. Manual examination of one tumour (T_08144) showed over 33,170 bases achieving greater than 100x coverage despite being untargeted by our panel (Figure 3-7). 3,284 and 27,484 of these bases corresponded to homologous regions within *MLH1* and *PMS2*, respectively. This is unsurprising given the challenges in mapping *PMS2* to the human genome. However, these homologous regions only accounted for approximately 0.5% of off-target rate within this sample. In addition, over 1000 off-target bases corresponded to microsatellites unintentionally captured. For example, one string of T's within T_08144 measured approximately 400 bases in length with some bases achieving greater than 55,000x coverage.

Figure 3-7: On and off-target regions captured within MultiMMR
Depth of coverage results from tumour 08144 (sequenced on MultiMMR V2) where each plot shows the on-target (shown in green; TRUE) and off target (shown in coral; FALSE) for a chromosome. The y-axis shows the number of reads supporting a specific genome coordinate, whereas the x-axis shows the start coordinate for the region. Each plot highlights a chromosome. It is clear that regions not targeted by the panel are being captured and sequenced.

Chapter 4

Discussion and Future Directions

4.1 Discussion

It is well known that LS testing can be stepwise and inconsistent[111], requiring multiple different tests and expertise. This can result in patients receiving suboptimal care[111]. Interest in universal tumour testing of CRC and EC has gained popularity; in 2009 the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) issued a recommendation that all individuals diagnosed with CRC be tested for LS[112]. Implementation of universal tumour testing would significantly increase the number of individuals tested for LS; this could be challenging in settings where testing is not standardized, and is complex. Prior to implementation of universal tumour testing, it is important to critically examine LS testing to optimize and standardize the process. A one-step testing method would simplify the diagnostic workflow for LS. Moreover, MultiMMR centers on clinical utility and only includes genes important for MMR that are clinically actionable. To our knowledge this study is the first to perform parallel bisulfite and conventional sequencing after keeping the normal or tumour DNA in one aliquot during shearing and hybrid capture. Moreover, this approach results in lower DNA loss, preserving precious clinical specimens.

Increasingly, NGS tests are being employed within clinical laboratories. This study has shown the ability to use hybrid capture sequencing to detect point mutations, insertions and deletions, copy number alterations, structural rearrangements, promoter methylation, and MSI. Ten different workflows from seven different laboratories across the United States tested the SeraCare Inherited DNA mix and only six labs found all eleven MMR variants[96]. Our panel has been able to detect all eleven variants in multiple sequencing runs, providing confidence in our bioinformatics pipeline. Moreover, the ability to detect additional pathogenic genome variation in other hereditary cancer syndromes speaks to the power of a multigene panel. Causes of protein expression loss by IHC have been found using the MultiMMR Panel, despite the low-targeted sequencing coverage our data received. Sequencing microsatellite loci and *PMS2* may explain some of the off-target rate, as MSI loci can cause DNA slippage during sequencing and

PMS2 is homologous with other regions of the genome: many of these pseudogenes have shown up in our data.

While IHC is invaluable in the screening for potential LS, additional information can be gleaned from examining all MMR genes for genome variation. For example, in one case (048), pathological examination found *MSH2* protein expression loss that was inconsistent with the somatic *MLH1* promoter methylation detected by MultiMMR. Re-examination of the IHC found focal MLH1 protein loss in addition to MSH2 loss, suggesting we received tissue enriched for this subclone. This case is currently undergoing additional examination, but this vignette highlights the value in examining multiple MMR genes and the role a comprehensive targeted panel can play in understanding the cause of MMR deficiency within a complex, multi-MMR deficiency case. Understanding focal MMR loss may have clinical relevance as the focal MMR may represent either a subclone of the primary or a separate tumour. Moreover, clinical testing workflows for LS are moving away from MSI testing in favour of MMR IHC testing, which gives information on the specific protein lost. However, recent literature has shown that MSI status provides predictive and prognostic value, being able to determine both the cause of MMR protein loss and MSI status in a singular assay is useful.

Currently MS-MLPA is the gold-standard for determining *MLH1* promoter methylation. The MS-MLPA kit examines 6 probes in the *MLH1* promoter region that contain recognition sites, approximately 4 bases in length, for a methylation-sensitive enzyme. Probes with a methylated recognition site will produce a signal[36]. In contrast, our method considers every cytosine within the promoter and provides global measurements of full promoter methylation status compared to examination of just a few sites. In our set of 19 cases with matched panel/MS-MLPA data, I encountered a case whereby the specific sites targeted by MS-MLPA have greater frequency of methylated reads compared all other CpGs in the promoter, as read by our panel (Figure 3-6). Since the specific recognition sites were methylated, the sample was classified as methylated. However, the majority of cytosines within the promoter were not methylated. This data is still preliminary and sample size is too small, but it would be interesting to further analyze samples where MS-MLPA and sequencing-based methylation profiling are discrepant. I hypothesize that sequencing-based methylation profiling may provide a more accurate read out

of promoter methylation. As NGS becomes more prevalent within the clinical laboratory, methylation testing may transition from MS-MLPA to sequenced-based testing to allow for amalgamation of multiple tests.

Multiple research groups have shown that individuals with an unknown cause of MMR deficiency (LLS patients) have an intermediate cancer risk[113]. The cause of MMR deficiency within these patients could be due to a germline mutation that was not identified, mosaicism, or biallelic somatic inactivation. Currently, no guidelines exist for the screening of these patients with some placed on LS surveillance[113]. Our panel has been shown to detect biallelic somatic inactivation, but cancer risk associated with biallelic somatic inactivation has not been investigated. The intermediate cancer risk associated with LLS could be a result of undetected germline mutations skewing the data. More research on the cancer risk associated with somatic biallelic inactivation would allow clinicians to better manage these patients.

In conclusion, we have developed a targeted sequencing panel that allows for parallel bisulfite and conventional sequencing of normal and tumour DNA to determine if an individual has LS, CMMRD, or SMMRD. The MultiMMR Panel contains genes associated with hereditary CRC syndromes that have clinical phenotypes that overlap LS, reducing the need for patients to undergo multiple genetic testing protocols to receive a diagnosis. Given that this study was a proof-of-principle study, external validation using a large prospective cohort is warranted. Moreover, improving off-target would decrease sequencing costs increasing the likelihood of the panel being translated to clinical laboratories. Our approach has particular value for universal LS screening, where every individual could be screened on MultiMMR for LS (Figure 4-1). Not only could LS and CMMRD be detected, but promoter methylation and somatic biallelic inactivation is examined and could explain the cause of MMR deficiency. Having SMMRD detected upfront can provide peace to patients and health care professions, avoiding sequential testing. Moreover, testing somatic cancers on MultiMMR can guide treatment with immunotherapy and other emerging treatments. Cases that remain unsolved by the panel are perfect candidates for research as additional drivers for LS or LLS may exist. Improving cancer screening through our panel and exploring cases with unknown aetiology is crucial to preventing future cancers and to better understanding LS and LLS.

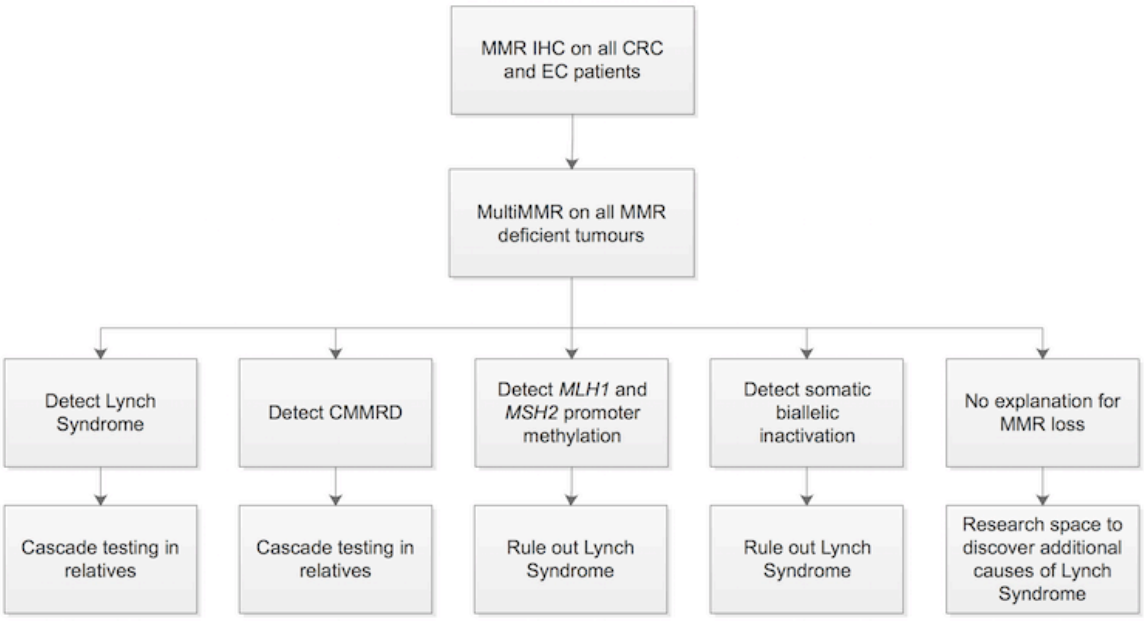


Figure 4-1: Potential clinical utility of the MultiMMR Panel
This figure depicts the clinical utility of the MultiMMR Panel in a universal tumour-testing model.

4.2 Future Directions

In the future we plan to further optimize the panel, perform external validation, explore unsolved cases, and implement MultiMMR within a clinical laboratory. Several cases within our cohort remained unsolved, meaning clinical testing and the panel failed to detect a pathogenic germline variant or somatic mutation in a MMR gene, despite having IHC loss of one of the MMR proteins. This is especially worrisome in cases that have a strong family history of disease. These individuals may have LS but current technology is unable to detect the pathogenic variant or MultiMMR could miss a variant due to low coverage in repetitive regions. There may be an additional gene or genes associated with LS that have yet to be discovered. A deletion in *EPCAM*, a gene upstream of *MSH2*, is able to silence the *MSH2* gene and was only recently discovered; therefore it's logical to believe that there may be additional drivers of LS[114]. The first step to solve these cases involves improving sequence coverage (sequencing deeper) and testing additional copy number and structural rearrangement algorithms, as these may detect additional germline or somatic copy number aberrations missed by the singular copy number and structural rearrangement tool used. Whole exome or whole genome sequencing in combination with whole transcriptome sequencing and epigenetic profiling could be performed on cases with unknown etiology to identify candidate genes associated with the LS phenotype. Genetic contributors to MMR deficiency that are not included on the MultiMMR Panel may exist. MultiMMR is an effective filter to discover these interesting cases that warrant further genome-wide profiling.

A third iteration of the panel has been created and is currently undergoing validation. This new panel contains all regions covered on MultiMMR V2 with the addition of tumor protein 53 (*TP53*). We believe that the addition of *TP53* will allow us to recapitulate the four prognostic groups for EC identified by the TCGA[23,115]. One group assessed multiple different surrogate assays with the goal of replicating the TCGA classification using a simpler, molecular-based assay[23,115]. This group found that MMR IHC followed by *POLE* testing and p53 IHC is able to recapitulate the TCGA four prognostic groupings[23,115]. MultiMMR includes all the MMR genes and *POLE*, therefore our panel can replace the MMR IHC and *POLE* testing. By adding the *TP53* gene we believe we can replace the p53 IHC and thus recapitulate the four prognostic

groups. Adding *TP53* improves the clinical utility of our panel and can be used by clinicians to tailor patient care based on patient prognosis. Conversely, we are considering adding *MSH3* and *NTHL1* to expand our panel's coverage of CRC polyposis-related syndromes[116]. This may be considered as our current panel includes only the genes responsible for the two most common polyposis-related syndromes (*APC* and *MUTYH*). Moreover, the addition of molecular barcoding will be considered to improve the identification of subclonal somatic mutations with low-allele frequency. After re-design and optimization and familiarization, validation with a minimum of 59 samples is recommended prior to clinical implementation, to achieve 95% confidence and 95% reliability[117]. Therefore, testing a minimum of 59 samples with varying types of genome variation and MMR deficiency will be performed to determine the accuracy, precision, analytical sensitivity and specificity[117].

As mentioned above, the panel suffered from a high-off target rate (~92%). Preliminary analysis has shown that 0.5% of this off-target rate is caused by capture of pseudogenes related to *PMS2* or non-unique sequences of *MLH1*. This preliminary analysis only examined off target regions within greater than 100x sequencing depth and only examined homology within two genes in one sample. I hypothesize that more than 0.5% of the off-target rate is explained by homology. Further examination of the data generated is needed. In concert, more research and methods testing should be performed to enhance detection of *PMS2* deletions and to differentiate *PMS2* mutations from its pseudogenes, namely *PMS2CL*. Being able to accurately detect *PMS2* mutations has clinical value considering approximately 50% of CMMRD cases are caused by biallelic PMS2 loss. Moreover, diagnosis of CMMRD is an active area of research and this study shows potential utility in using NGS panels although further study is needed to validate the sensitivity and specificity of this panel, particularly in detecting *PMS2* mutations.

Ultimately, the future direction of this project is to further refine our analysis pipeline and improve on-target rate to allow for translation of this integrative panel into the clinic and discover additional causes of LS. In collaboration with EC clinicians we plan to use MultiMMR to identify LLS patients and to match MultiMMR result with survival data to develop clinical guidelines for the care of these patients. With universal LS screening on the horizon, I see the MultiMMR Panel as a tool clinicians and genetic counsellors could use to detect LS, SMMRD,

CMMRD, *BRAF* status, methylation status, and MSI status from a singular assay (Figure 4-1). Moreover, researchers interested in discovering new causes of LS or CMMRD would benefit from the MultiMMR Panel to rule out known causes of MMR deficiency and perform extensive examination of a cohort of cases that remain unsolved by the panel.

References

[1] Li G-M. Mechanisms and functions of DNA mismatch repair. Cell Res 2008;18:85–98.

[2] Hsieh P, Yamane K. DNA mismatch repair: molecular mechanism, cancer, and ageing. Mech Ageing Dev 2008;129:391–407.

[3] Hopfner KP, Tainer JA. DNA mismatch repair: the hands of a genome guardian. Structure 2000;8:R237–41.

[4] Hsieh P, Zhang Y. The Devil is in the details for DNA mismatch repair. Proc Natl Acad Sci U S A 2017;114:3552–4.

[5] Jiricny J. The multifaceted mismatch-repair system. Nat Rev Mol Cell Biol 2006;7:335–46.

[6] Modrich P. Mechanisms in eukaryotic mismatch repair. J Biol Chem 2006;281:30305–9.

[7] Fishel R. Mismatch Repair. J Biol Chem 2015. doi:10.1074/jbc.R115.660142.

[8] Subramanian S, Mishra RK, Singh L. Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions. Genome Biol 2003;4:R13.

[9] Bagshaw ATM. Functional Mechanisms of Microsatellite DNA in Eukaryotic Genomes. Genome Biol Evol 2017;9:2428–43.

[10] Hause RJ, Pritchard CC, Shendure J, Salipante SJ. Classification and characterization of microsatellite instability across 18 cancer types. Nat Med 2016;22:1342–50.

[11] Chen W, Swanson BJ, Frankel WL. Molecular genetics of microsatellite-unstable colorectal cancer for pathologists. Diagn Pathol 2017;12:24.

[12] Rodriguez-Bigas MA, Boland CR, Hamilton SR, Henson DE, Jass JR, Khan PM, et al. A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines. J Natl Cancer Inst 1997;89:1758–62.

[13] Umar A, Boland CR, Terdiman JP, Syngal S, Chapelle A de la, Rüschoff J, et al. Revised Bethesda Guidelines for Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndrome) and Microsatellite Instability. J Natl Cancer Inst 2004;96:261–8.

[14] Palomaki GE, McClain MR, Melillo S, Hampel HL, Thibodeau SN. EGAPP supplementary evidence review: DNA testing strategies aimed at reducing morbidity and mortality from Lynch syndrome. Genet Med 2009;11:42–65.

[15] Takehara Y, Nagasaka T, Nyuya A, Haruma T, Haraga J, Mori Y, et al. Accuracy of four mononucleotide-repeat markers for the identification of DNA mismatch-repair deficiency in solid tumors. J Transl Med 2018;16:5.

[16] Buhard O, Suraweera N, Lectard A, Duval A, Hamelin R. Quasimonomorphic mononucleotide repeats for high-level microsatellite instability analysis. Dis Markers 2004;20:251–7.

[17] Murphy KM, Zhang S, Geiger T, Hafez MJ, Bacher J, Berg KD, et al. Comparison of the microsatellite instability analysis system and the Bethesda panel for the determination of microsatellite instability in colorectal cancers. J Mol Diagn 2006;8:305–11.

[18] Promega Corporation. Improving Genetic Analysis: From FFPE Tissue DNA Extraction to MSI Analysis n.d.

[19] Kim T-M, Laird PW, Park PJ. The landscape of microsatellite instability in colorectal and endometrial cancer genomes. Cell 2013;155:858–68.

[20] Salipante SJ, Scroggins SM, Hampel HL, Turner EH, Pritchard CC. Microsatellite instability detection by next generation sequencing. Clin Chem 2014;60:1192–9.

[21] Kanopienė D, Smailytė G, Vidugirienė J, Bacher J. Impact of microsatellite instability on survival of endometrial cancer patients. Medicina 2014;50:216–21.

[22] Arabi H, Guan H, Kumar S, Cote M, Bandyopadhyay S, Bryant C, et al. Impact of microsatellite instability (MSI) on survival in high grade endometrial carcinoma. Gynecol Oncol 2009;113:153–8.

[23] Levine DA, The Cancer Genome Atlas Research Network. Integrated genomic characterization of endometrial carcinoma. Nature 2013;497:67.

[24] Zhang L. Immunohistochemistry versus microsatellite instability testing for screening colorectal cancer patients at risk for hereditary nonpolyposis colorectal cancer syndrome. Part II. The utility of microsatellite instability testing. J Mol Diagn 2008;10:301–7.

[25] Boland CR, Goel A. Microsatellite instability in colorectal cancer. Gastroenterology 2010;138:2073–87.e3.

[26] Kawakami H, Zaanan A, Sinicrope FA. Microsatellite instability testing and its role in the management of colorectal cancer. Curr Treat Options Oncol 2015;16:30.

[27] Des Guetz G, Schischmanoff O, Nicolas P, Perret G-Y, Morere J-F, Uzzan B. Does microsatellite instability predict the efficacy of adjuvant chemotherapy in colorectal cancer? A systematic review with meta-analysis. Eur J Cancer 2009;45:1890–6.

[28] Ribic CM, Sargent DJ, Moore MJ, Thibodeau SN, French AJ, Goldberg RM, et al. Tumor Microsatellite-Instability Status as a Predictor of Benefit from Fluorouracil-Based Adjuvant Chemotherapy for Colon Cancer. N Engl J Med 2003;349:247–57.

[29] Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science 2017;357:409–13.

[30] Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N Engl J Med 2015;372:2509–20.

[31] Knudson AG Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A 1971;68:820–3.

[32] Vasen HF, Watson P, Mecklin JP, Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. Gastroenterology 1999;116:1453–6.

[33] Haraldsdottir S, Hampel H, Tomsic J, Frankel WL, Pearlman R, de la Chapelle A, et al. Colon and endometrial cancers with mismatch repair deficiency can arise from somatic, rather than germline, mutations. Gastroenterology 2014;147:1308–16.e1.

[34] Buchanan DD, Tan YY, Walsh MD, Clendenning M, Metcalf AM, Ferguson K, et al. Tumor mismatch repair immunohistochemistry and DNA MLH1 methylation testing of patients with endometrial cancer diagnosed at age younger than 60 years optimizes triage for population-level germline mismatch repair gene mutation testing. J Clin Oncol 2014;32:90–100.

[35] Rodríguez-Soler M, Pérez-Carbonell L, Guarinos C, Zapater P, Castillejo A, Barberá VM, et al. Risk of cancer in cases of suspected lynch syndrome without germline mutation. Gastroenterology 2013;144:926–32.e1; quiz e13–4.

[36] Nygren AOH, Ameziane N, Duarte HMB, Vijzelaar RNCP, Waisfisz Q, Hess CJ, et al. Methylation-specific MLPA (MS-MLPA): simultaneous detection of CpG methylation and copy number changes of up to 40 sequences. Nucleic Acids Res 2005;33:e128.

[37] Adar T, Rodgers LH, Shannon KM, Yoshida M, Ma T, Mattia A, et al. A tailored approach to BRAF and MLH1 methylation testing in a universal screening program for Lynch syndrome. Mod Pathol 2017;30:440–7.

[38] Walsh MD, Buchanan DD, Walters R, Roberts A, Arnold S, McKeone D, et al. Analysis of families with Lynch syndrome complicated by advanced serrated neoplasia: the importance of pathology review and pedigree analysis. Fam Cancer 2009;8:313–23.

[39] Parsons MT, Buchanan DD, Thompson B, Young JP, Spurdle AB. Correlation of tumour BRAF mutations and MLH1 methylation with germline mismatch repair (MMR) gene mutation status: a literature review assessing utility of tumour features for MMR variant classification. J Med Genet 2012;49:151–7.

[40] Hitchins MP, Wong JJL, Suthers G, Suter CM, Martin DIK, Hawkins NJ, et al. Inheritance of a cancer-associated MLH1 germ-line epimutation. N Engl J Med 2007;356:697–705.

[41] Gazzoli I, Loda M, Garber J, Syngal S, Kolodner RD. A hereditary nonpolyposis colorectal carcinoma case associated with hypermethylation of the MLH1 gene in normal tissue and loss of heterozygosity of the unmethylated allele in the resulting microsatellite instability-high tumor. Cancer Res 2002;62:3925–8.

[42] Ollikainen M, Hannelius U, Lindgren CM, Abdel-Rahman WM, Kere J, Peltomäki P. Mechanisms of inactivation of MLH1 in hereditary nonpolyposis colorectal carcinoma: a novel approach. Oncogene 2007;26:4541–9.

[43] Wheeler JM, Loukola A, Aaltonen LA, Mortensen NJ, Bodmer WF. The role of hypermethylation of the hMLH1 promoter region in HNPCC versus MSI+ sporadic colorectal cancers. J Med Genet 2000;37:588–92.

[44] Jin M, Hampel H, Zhou X, Schunemann L, Yearsley M, Frankel WL. BRAF V600E mutation analysis simplifies the testing algorithm for Lynch syndrome. Am J Clin Pathol 2013;140:177–83.

[45] Dinh TA, Rosner BI, Atwood JC, Boland CR, Syngal S, Vasen HFA, et al. Health benefits and cost-effectiveness of primary genetic screening for Lynch syndrome in the general population. Cancer Prev Res 2011;4:9–22.

[46] Lynch HT, de la Chapelle A. Hereditary colorectal cancer. N Engl J Med 2003;348:919–32.

[47] Aaltonen L, Johns L, Järvinen H, Mecklin J-P, Houlston R. Explaining the familial colorectal cancer risk associated with mismatch repair (MMR)-deficient and MMR-stable tumors. Clin Cancer Res 2007;13:356–61.

[48] Lu KH, Schorge JO, Rodabaugh KJ, Daniels MS, Sun CC, Soliman PT, et al. Prospective determination of prevalence of lynch syndrome in young women with endometrial cancer. J Clin Oncol 2007;25:5158–64.

[49] Kovacs ME, Papp J, Szentirmay Z, Otto S, Olah E. Deletions removing the last exon of TACSTD1 constitute a distinct class of mutations predisposing to Lynch syndrome. Hum Mutat 2009;30:197–203.

[50] Ligtenberg MJL, Kuiper RP, Chan TL, Goossens M, Hebeda KM, Voorendt M, et al. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. Nat Genet 2009;41:112–7.

[51] Lynch HT, Smyrk TC, Watson P, Lanspa SJ, Lynch JF, Lynch PM, et al. Genetics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: an updated review. Gastroenterology 1993;104:1535–49.

[52] Baglietto L, Lindor NM, Dowty JG, White DM, Wagner A, Gomez Garcia EB, et al. Risks of Lynch syndrome cancers for MSH6 mutation carriers. J Natl Cancer Inst 2010;102:193–201.

[53] MecKlin J-P, Järvinen HJ. Tumor spectrum in cancer family syndrome (hereditary nonpolyposis colorectal cancer). Cancer 1991;68:1109–12.

[54] Wijnen J, de Leeuw W, Vasen H, van der Klift H, Møller P, Stormorken A, et al. Familial endometrial cancer in female carriers of MSH6 germline mutations. Nat Genet 1999;23:142–4.

[55] Wang Y, Wang Y, Li J, Cragun J, Hatch K, Chambers SK, et al. Lynch syndrome related endometrial cancer: clinical significance beyond the endometrium. J Hematol Oncol 2013;6:22.

[56] Senter L, Clendenning M, Sotamaa K, Hampel H, Green J, Potter JD, et al. The clinical phenotype of Lynch syndrome due to germ-line PMS2 mutations. Gastroenterology 2008;135:419–28.

[57] Dowty JG, Win AK, Buchanan DD, Lindor NM, Macrae FA, Clendenning M, et al. Cancer risks for MLH1 and MSH2 mutation carriers. Hum Mutat 2013;34:490–7.

[58] Bonadona V, Bonaïti B, Olschwang S, Grandjouan S, Huiart L, Longy M, et al. Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. JAMA 2011;305:2304–10.

[59] Hampel H, Frankel W, Panescu J, Lockman J, Sotamaa K, Fix D, et al. Screening for Lynch syndrome (hereditary nonpolyposis colorectal cancer) among endometrial cancer patients. Cancer Res 2006;66:7810–7.

[60] Vasen HF, Mecklin JP, Khan PM, Lynch HT. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Dis Colon Rectum 1991;34:424–5.

[61] Hampel H, Stephens JA, Pukkala E, Sankila R, Aaltonen LA, Mecklin J-P, et al. Cancer risk in hereditary nonpolyposis colorectal cancer syndrome: later age of onset. Gastroenterology 2005;129:415–21.

[62] Sjursen W, Haukanes BI, Grindedal EM, Aarset H, Stormorken A, Engebretsen LF, et al. Current clinical criteria for Lynch syndrome are not sensitive enough to identify MSH6 mutation carriers. J Med Genet 2010;47:579–85.

[63] Durno C, Boland CR, Cohen S, Dominitz JA, Giardiello FM, Johnson DA, et al. Recommendations on Surveillance and Management of Biallelic Mismatch Repair Deficiency (BMMRD) Syndrome: A Consensus Statement by the US Multi-Society Task Force on Colorectal Cancer. Gastroenterology 2017;152:1605–14.

[64] Vasen HFA, Blanco I, Aktan-Collan K, Gopie JP, Alonso A, Aretz S, et al. Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. Gut 2013;62:812–23.

[65] Zhang X, Li J. Era of universal testing of microsatellite instability in colorectal cancer. World J Gastrointest Oncol 2013;5:12–9.

[66] Kidambi TD, Blanco A, Myers M, Conrad P, Loranger K, Terdiman JP. Selective Versus Universal Screening for Lynch Syndrome: A Six-Year Clinical Experience. Dig Dis Sci 2015;60:2463–9.

[67] Hampel H. NCCN increases the emphasis on genetic/familial high-risk assessment in colorectal cancer. J Natl Compr Canc Netw 2014;12:829–31.

[68] Stoffel EM, Mangu PB, Gruber SB, Hamilton SR, Kalady MF, Lau MWY, et al. Hereditary colorectal cancer syndromes: American Society of Clinical Oncology Clinical Practice Guideline endorsement of the familial risk-colorectal cancer: European Society for Medical Oncology Clinical Practice Guidelines. J Clin Oncol 2015;33:209–17.

[69] A. Pollett, J. Brown, M. Aronson, B. Clark, N. Baxter, E. Tomiak. Screening for Lynch Syndrome by Immunohistochemistry, BRAF Mutations Analysis, and MLH1 Promoter Methylation Analysis for Patients in Ontario with Colorectal or Endometrial Cancers 2015.

[70] Giardiello FM, Allen JI, Axilbund JE, Boland CR, Burke CA, Burt RW, et al. Guidelines on genetic evaluation and management of Lynch syndrome: a consensus statement by the US Multi-society Task Force on colorectal cancer. Am J Gastroenterol 2014;109:1159–79.

[71] Dove-Edwin I, Sasieni P, Adams J, Thomas HJW. Prevention of colorectal cancer by colonoscopic surveillance in individuals with a family history of colorectal cancer: 16 year, prospective, follow-up study. BMJ 2005;331:1047.

[72] de Jong AE, Hendriks YMC, Kleibeuker JH, de Boer SY, Cats A, Griffioen G, et al. Decrease in mortality in Lynch syndrome families because of surveillance. Gastroenterology 2006;130:665–71.

[73] Schmeler KM, Lynch HT, Chen L-M, Munsell MF, Soliman PT, Clark MB, et al. Prophylactic Surgery to Reduce the Risk of Gynecologic Cancers in the Lynch Syndrome. N Engl J Med 2006;354:261–9.

[74] Stupart DA, Goldberg PA, Algar U, Ramesar R. Surveillance colonoscopy improves survival in a cohort of subjects with a single mismatch repair gene mutation. Colorectal Dis 2009;11:126–30.

[75] Lynch HT, Watson P, Kriegler M, Lynch JF, Lanspa SJ, Marcus J, et al. Differential diagnosis of hereditary nonpolyposis colorectal cancer (Lynch syndrome I and Lynch syndrome II). Dis Colon Rectum 1988;31:372–7.

[76] Aarnio M, Mecklin JP, Aaltonen LA, Nyström-Lahti M, Järvinen HJ. Life-time risk of different cancers in hereditary non-polyposis colorectal cancer (HNPCC) syndrome. Int J Cancer 1995;64:430–3.

[77] Church J, Simmang C, Standards Task Force, American Society of Colon and Rectal Surgeons, Collaborative Group of the Americas on Inherited Colorectal Cancer and the Standards Committee of The American Society of Colon and Rectal Surgeons. Practice parameters for the treatment of patients with dominantly inherited colorectal cancer (familial adenomatous polyposis and hereditary nonpolyposis colorectal cancer). Dis Colon Rectum 2003;46:1001–12.

[78] Burn J, Gerdes A-M, Macrae F, Mecklin J-P, Moeslein G, Olschwang S, et al. Long-term effect of aspirin on cancer risk in carriers of hereditary colorectal cancer: an analysis from the CAPP2 randomised controlled trial. Lancet 2011;378:2081–7.

[79] FDA. FDA approves first cancer treatment for any solid tumor with a specific genetic feature 2017. https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm560167.htm.

[80] Durno CA, Sherman PM, Aronson M, Malkin D, Hawkins C, Bakry D, et al. Phenotypic and genotypic characterisation of biallelic mismatch repair deficiency (BMMR-D) syndrome. Eur J Cancer 2015;51:977–83.

[81] Aronson M, Gallinger S, Cohen Z, Cohen S, Dvir R, Elhasid R, et al. Gastrointestinal Findings in the Largest Series of Patients With Hereditary Biallelic Mismatch Repair Deficiency Syndrome: Report from the International Consortium. Am J Gastroenterol 2016;111:275–84.

[82] Lynch HT, Lynch PM, Lanspa SJ, Snyder CL, Lynch JF, Boland CR. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. Clin Genet 2009;76:1–18.

[83] Shlien A, Campbell BB, de Borja R, Alexandrov LB, Merico D, Wedge D, et al. Combined hereditary and somatic mutations of replication error repair genes result in rapid onset of ultra-hypermutated cancers. Nat Genet 2015;47:257–62.

[84] Hegde M, Ferber M, Mao R, Samowitz W, Ganguly A, Working Group of the American College of Medical Genetics and Genomics (ACMG) Laboratory Quality Assurance Committee. ACMG technical standards and guidelines for genetic testing for inherited colorectal cancer (Lynch syndrome, familial adenomatous polyposis, and MYH-associated polyposis). Genet Med 2014;16:101–16.

[85] Shen T, Pajaro-Van de Stadt SH, Yeat NC, Lin JC-H. Clinical applications of next generation sequencing in cancer: from panels, to exomes, to genomes. Front Genet 2015;6:215.

[86] Susswein LR, Marshall ML, Nusbaum R, Vogel Postula KJ, Weissman SM, Yackowski L, et al. Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. Genet Med 2016;18:823–32.

[87] Desmond A, Kurian AW, Gabree M, Mills MA, Anderson MJ, Kobayashi Y, et al. Clinical Actionability of Multigene Panel Testing for Hereditary Breast and Ovarian Cancer Risk Assessment. JAMA Oncol 2015;1:943–51.

[88] Cohen SA, Tan CA, Bisson R. An Individual with Both MUTYH-Associated Polyposis and Lynch Syndrome Identified by Multi-Gene Hereditary Cancer Panel Testing: A Case Report. Front Genet 2016;7:36.

[89] Pritchard CC, Smith C, Salipante SJ, Lee MK, Thornton AM, Nord AS, et al. ColoSeq provides comprehensive lynch and polyposis syndrome mutational analysis using massively parallel sequencing. J Mol Diagn 2012;14:357–66.

[90] LaDuca H, Stuenkel AJ, Dolinsky JS, Keiles S, Tandy S, Pesaran T, et al. Utilization of multigene panels in hereditary cancer predisposition testing: analysis of more than 2,000 patients. Genet Med 2014;16:830–7.

[91] Chiang JP-W, Lamey T, McLaren T, Thompson JA, Montgomery H, De Roach J. Progress and prospects of next-generation sequencing testing for inherited retinal dystrophy. Expert Rev Mol Diagn 2015;15:1269–75.

[92] Hampel H, Frankel WL, Martin E, Arnold M, Khanduja K, Kuebler P, et al. Screening for the Lynch Syndrome (Hereditary Nonpolyposis Colorectal Cancer). N Engl J Med 2005;352:1851–60.

[93] Sanchez JJ, Phillips C, Børsting C, Balogh K, Bogus M, Fondevila M, et al. A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis 2006;27:1713–24.

[94] Carethers JM, Stoffel EM. Lynch syndrome and Lynch syndrome mimics: The growing complex landscape of hereditary colon cancer. World J Gastroenterol 2015;21:9253–61.

[95] Campbell BB, Light N, Fabrizio D, Zatzman M, Fuligni F, de Borja R, et al. Comprehensive Analysis of Hypermutation in Human Cancer. Cell 2017;171:1042–56.e10.

[96] Stephen E Lincoln, Justin M Zook, Shimul Chowdhury, Shazia Mahamdallie, Andrew Fellowes, Eric W Klee, Rebecca Truty, Catherine Huang, Farol L Tomson, Megan H Cleveland, Peter M Vallone, Yan Ding, Sheila Seal, Wasanthi DeSilva, Russell K Garlick, Marc Salit, Nazneen Rahman, Stephen F Kingsmore, Swaroop Aradhya, Robert L Nussbaum, Matthew J Ferber, Brian H Shirts. An interlaboratory study of complex variant detection. bioRxiv 2017.

[97] DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 2011;43:491.

[98] Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 2013;31:213.

[99] Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. Bioinformatics 2012;28:1811–7.

[100] Ramos AH, Lichtenstein L, Gupta M, Lawrence MS, Pugh TJ, Saksena G, et al. Oncotator: cancer variant annotation tool. Hum Mutat 2015;36:E2423–9.

[101] Pugh TJ, Amr SS, Bowser MJ, Gowrisankar S, Hynes E, Mahanta LM, et al. VisCap: inference and visualization of germ-line copy-number variants from targeted clinical sequencing data. Genet Med 2016;18:712–9.

[102] Chow S. Targeted Capture and Sequencing of Immunoglobulin Rearrangements in Multiple Myeloma to Enable Detection of Minimal Residual Disease. Masters of Science. University of Toronto, 2017.

[103] Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285.

[104] Wong SQ, Li J, Tan AY-C, Vedururu R, Pang J-MB, Do H, et al. Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. BMC Med Genomics 2014;7:23.

[105] Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17:405–24.

[106] Thompson BA, Spurdle AB, Plazzer J-P, Greenblatt MS, Akagi K, Al-Mulla F, et al. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. Nat Genet 2014;46:107–15.

[107] Broad Institute. Picard Tools. Picard n.d. http://broadinstitute.github.io/picard/.

[108] Clendenning M, Buchanan DD, Walsh MD, Nagler B, Rosty C, Thompson B, et al. Mutation deep within an intron of MSH2 causes Lynch syndrome. Fam Cancer 2011;10:297–301.

[109] Sieber OM, Lipton L, Crabtree M, Heinimann K, Fidalgo P, Phillips RKS, et al. Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH. N Engl J Med 2003;348:791–9.

[110] Cleary SP, Cotterchio M, Jenkins MA, Kim H, Bristow R, Green R, et al. Germline MutY human homologue mutations and colorectal cancer: a multisite case-control study. Gastroenterology 2009;136:1251–60.

[111] Bombard Y, Rozmovits L, Sorvari A, Daly C, Carroll JC, Kennedy E, et al. Universal tumor screening for Lynch syndrome: health-care providers' perspectives. Genet Med 2016;19:568.

[112] Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. Recommendations from the EGAPP Working Group: genetic testing strategies in newly diagnosed individuals with colorectal cancer aimed at reducing morbidity and mortality from Lynch syndrome in relatives. Genet Med 2009;11:35–41.

[113] Buchanan DD, Rosty C, Clendenning M, Spurdle AB, Win AK. Clinical problems of colorectal cancer and endometrial cancer cases with unknown cause of tumor mismatch repair deficiency (suspected Lynch syndrome). Appl Clin Genet 2014;7:183–93.

[114] Guarinos C, Castillejo A, Barberá V-M, Pérez-Carbonell L, Sánchez-Heras A-B, Segura A, et al. EPCAM germ line deletions as causes of Lynch syndrome in Spanish patients. J Mol Diagn 2010;12:765–70.

[115] Talhouk A, McConechy MK, Leung S, Li-Chang HH, Kwon JS, Melnyk N, et al. A clinically applicable molecular-based classification for endometrial cancers. Br J Cancer 2015;113:299–310.

[116] Adam R, Spier I, Zhao B, Kloth M, Marquez J, Hinrichsen I, et al. Exome Sequencing Identifies Biallelic MSH3 Germline Mutations as a Recessive Subtype of Colorectal Adenomatous Polyposis. Am J Hum Genet 2016;99:337–51.

[117] Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J, et al. Guidelines for Validation of Next-Generation Sequencing-Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. J Mol Diagn 2017;19:341–65.