Cancer Chemoprevention with Aspirin: Enhancing Discovery of Candidates with Mismatch Repair Deficiency

Peter Sowter

A thesis submitted to Newcastle University
for the degree of Doctor of Philosophy

Institute of Genetic Medicine
Faculty of Medical Sciences
Newcastle University

September 2023

Abstract

Lynch syndrome (LS) is an inherited cancer syndrome arising from germline mutations in genes of the DNA mismatch repair (MMR) system. To optimise clinical management, the efficient diagnosis of LS individuals is essential, with known carriers of LS-associated gene defects benefiting from various practices including daily aspirin prophylaxis. Current guidelines in the UK advocate the use of two techniques in the screening for LS: MMR protein immunohistochemistry and DNA microsatellite instability (MSI) testing. However, several limitations compromise the efficacy of this clinical guidance, including the restricted tumour spectrum recommended for analysis, and the ambiguous genetic diagnoses returned by existing assays. This project aimed to enhance the existing techniques for detecting mismatch repair deficiency (MMRd) and LS individuals by addressing both these issues.

In considering the potential cancer spectrum for LS screening, I initially analysed 122 extracolonic cancer samples from LS gene carriers using a new MSI assay (the Newcastle Assay) to review MSI as a biomarker of MMR deficiency in these tumour types. An MSI-H classification was returned for the majority of tumours, including 80% (41/51) of those considered of the LS spectrum, but a comparatively low proportion of endometrial cancer (EC) samples (26/35 - 72%) were found with this phenotype. Further investigation of MSI in EC specifically involved the analysis of 363 well-characterised samples, from two external clinical trial cohorts, using the Newcastle Assay. In this study, the frequency of instability between the cohorts varied for samples with confirmed MMR deficiency, highlighting caveats when using MSI as a biomarker for MMRd in EC. With this approach also often failing to detect MMRd for which *MSH6*-deficiency is responsible, these findings support the prioritisation of IHC for LS screening of this tumour type as recently recommended by NICE, but also suggest that additional MSI testing could have clinical benefit.

For the diagnosis of LS, confirmation of a germline pathogenic MMR variant is required, but this is complicated for the MMR gene *PMS2* by the presence of multiple pseudogenes. To improve the analysis of this gene, I attempted to develop a sequencing-based assay using a combination of long-range PCR sequencing and Molecular Inversion Probe (MIP) technology. I established a MIP pool consisting of 42

exonic and 100 intronic probes to assess for sequence variants and CNVs/loss of heterozygosity respectively. In the review of samples for which *PMS2* mutation was confirmed, detection of variants by the exon-tiling component of this assay was demonstrated. Validated point mutations were accurately identified, with all but one of the 22 pathogenic variant calls returned from the 66/138 samples with verified *PMS2* changes. However, the proficiency of this assay for detecting copy-number variation and loss of heterozygosity remains to be analysed, curtailed by time constraints. Information derived from intronic SNPs has thus far been inconclusive, and in no samples has homozygosity or substantial deletions been indicated. Results from the assay did however suggest that LS patients may have mutations in both *PMS2* alleles, and these will require further investigation.

Both these studies endeavoured to further the guidance and techniques for LS screening, and subsequently the application of treatment practices such as aspirin prophylaxis. The results of MSI analysis in extracolonic tumour samples demonstrated the facility of this biomarker in various Lynch-spectrum cancers, with EC analysis highlighting conditions for its use. In contrast, development of a *PMS2* assay was less successful, and alternative long-read sequencing approaches are likely to supplant it. This suggests that continued development of the assay is likely not viable, although the complexity of the results highlight the need for accurate mutation screening of this MMR gene.

Acknowledgements

I would like to thank my project supervisors Professor Sir John Burn, Dr. Michael Jackson and Dr. Mauro Santibanez-Koref for furthering my development as a scientist, and maintaining the direction of my research. In particular, I thank John for providing the clinical perspective of this research, and for connecting me with collaborators. I thank Michael and Mauro for their routine supervision, meticulous feedback and, the latter, for his bioinformatic support. I also thank Cancer Research UK for funding this PhD, and all concurrent studies of the Catalyst collaboration.

Dr. Richard Gallon, I thank you for insightful contributions to my research, as well as the friendship and frequent reassurance you have afforded me. The Samwise to my Frodo, I feel incredibly fortunate to have had you as my postdoctoral colleague and support during this project.

I would like to thank all members of the CaPP3 clinical trial, both locally and nationally, for the samples and corresponding genetic information they have provided. From the Central Study Team in Newcastle, I give particular thanks to Christine Hayes, a mother figure during my studies, for her unwavering compassion and assistance in all aspects of this work, Dr Gillian Borthwick for coordinating research collaborations, and her diligent management of research ethics. I also thank Lynn Reed, Lynne Longstaff and Donna Job for addressing CaPP3 queries and Amy McAllister for organising meetings and conference attendance.

Collaborations with other research institutions, and the provision of external samples, were essential to my PhD. I express my gratitude to Dr. Heather Hampel and Paul Goodfellow, PhD of the Ohio State University, as well as Dr. Emma Crosbie of the University of Manchester and Dr. Neil Ryan for their cooperation and sourcing of suitable material.

Finally, I give the foremost thanks to my incredible grandparents for their eternal support and encouragement. My recently-departed grandad Kevin, your enthusiasm inspired me to undertake this studentship, and my grandma Wendy, your strength and guidance motivated me to complete it. Additional thanks to any other individual not previously named, including my amazing girlfriend Anna, without whom these last five years would not have been possible or as enjoyable.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

15-PGDH	15-hydroxyprostaglandin dehydrogenase
AA	Arachidonic acid
AH	Atypical hyperplasia
ARE	AU-rich element
BRCA	Breast cancer gene
BWA	Burrows-Wheeler aligner
CaPP	Cancer Prevention Programme
CMMRD	Constitutional mismatch repair deficiency
CNV	Copy-number variants
CRC	Colorectal cancer
CVD	Cardiovascular disease
DG151	Diagnostic Guidance 151
EC	Endometrial cancer
ERK	Extracellular-signal-regulated kinase
EXO1	Exonuclease 1
FAP	Familial adenomatous polyposis
FFPE	Formalin-fixed, paraffin-embedded
GATK	Genome Analysis Toolkit
GI	Gastrointestinal
HLA	Human leukocyte antigen
HNPCC	Hereditary non-polyposis colorectal cancer
HR	Hazard ratio
ICB	Immune checkpoint blockade
IDL	Insertion deletion
IHC	Immunohistochemistry
InSiGHT	the International Society for Gastrointestinal Hereditary Tumours
ITT	Intention-to-treat
KRAS	Kirsten rat sarcoma virus
LLS	Lynch-like syndrome
LRS	Long-read sequencing

LS	Lynch syndrome
MI	Myocardial infarction
MIP	Molecular inversion probe
MLH1	MutL-homolog 1
MLH2	MutL-homolog 2
MLH3	MutL-homolog 3
MMR	Mismatch repair
MMRd	Mismatch repair deficient
MMRp	Mismatch repair deficient
MQ	Mapping quality
MSH2	MutS-Homolog 2
MSH3	MutS-Homolog 3
MSH6	MutS-Homolog 6
MSI	Microsatellite instability
MSI-H	High levels of microsatellite instability
MSI-L	Low levels of microsatellite instability
MSS	Microsatellite stable
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NICE	National Institute for Clinical and Healthcare Excellence
NGS	Next-generation sequencing
NO	Nitric oxide
NSAID	Non-steroidal anti-inflammatory drug
OR	Odds ratio
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PD-1	Programmed cell death protein 1
PD-L1	Programmed death-ligand 1
PG	Prostaglandin
PGE_2	Prostaglandin E
PGH_2	Prostaglandin H
PGI_2	Prostaglandin I
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha

PLSD	Prospective Lynch Syndrome Database
PMS2	Post-meiotic segregation 2
RCT	Randomised control trial
RR	Relative risk
smMIP	Single-molecular molecular inversion probes
SNP	Single nucleotide polymorphism
TXA_2	Thromboxane A_2
VAF	Variant allele frequency
VEP	Variant effect predictor
VUS	Variant of unknown significance

Chapter 1. Introduction

1.1. Aspirin and its Therapeutic Uses

Acetylsalicylic acid, commonly known as aspirin, is a non-steroidal anti-inflammatory drug (NSAID) produced in an esterification reaction between salicylic acid and ethanoic anhydride. First synthesised by Felix Hoffman (under the supervision of Bayer chemist Arthur Eichengrun) and released in 1899, its administration for various therapeutic purposes has since become well established, although the underlying mechanisms by which it achieves these continue to be investigated. The potential use of aspirin as a targeted means of cancer prevention specifically was the primary driver of my research described hereafter.

1.1.1. Prevention of Cardiovascular Disease

Principal among the recent uses of aspirin is its role in the secondary prevention of cardiovascular disease (CVD), facilitated by its anti-platelet activity. Individuals who experience one or more CVD events, for example myocardial infarction (MI), are at a higher risk of subsequent events. These may however be prevented by aspirin therapy, as was first unequivocally demonstrated in the Second International Study of Infarct Survival (ISIS-2) trial (ISIS-2 Collaborative Group, 1988). This found low-dose aspirin, commenced immediately after MI and maintained for one month, to give significant reductions in non-fatal reinfarction and non-fatal stroke, with additional mortality benefits realised by a longer duration of treatment. Subsequent meta-analyses by the Antiplatelet Trialists' Collaboration (BMJ, 1994) and Antithrombotic Trialists' (ATT) Collaboration (BMJ, 2002), recognised similar protective effects with long-term aspirin use. In both studies, an approximate one-quarter decrease in vascular events was observed across high-risk patients subject to treatment, with equivalent efficacy between lower and higher dosing strategies. These findings, and further studies supporting them, explain why aspirin therapy is now the established preventative approach it is, and recommended by several organisations (Ittaman, VanWormer and Rezkalla, 2014).

The success of aspirin in the secondary prevention of CVD propagated a logical exploration of its use as a primary preventative measure. This has not been without

controversy however (Raber *et al.*, 2019), with different analyses returning contrasting verdicts on efficacy, while reporting potential, serious side effects. Initial trials across selected and non-selected populations indicated a reduction in both MI and stroke incidence, as well as a general, but not statistically significant, reduction in mortality, particularly in those at an increased risk of CVD (that is, a 10-year risk >10%) (Ittaman, VanWormer and Rezkalla, 2014; Murphy, Curneen and McEvoy, 2021). Although these benefits were accompanied by an increase in bleeding issues, such findings were consistent, and ultimately resulted in the first appearance of aspirin as a primary prevention in guidelines. More recent trials reporting neutral results, or even evidence indicative of net harm, have since questioned this use, and influenced a revision of said guidelines. While these now vary by country and professional organisation, current clinical recommendations generally advocate the selective prescription of aspirin therapy as a primary prevention in instances of significantly elevated CVD risk, with its use in low-risk populations having seen a reduction in support.

1.1.2. Mechanisms of CVD Prevention by Aspirin

Prostaglandin (PG) H-synthase, also known as COX, is a membrane-bound enzyme responsible for the biosynthesis of cyclic prostanoids from arachidonic acid (AA). In humans, such primarily exists in two distinct isoforms (Williams and DuBois, 1996), COX-1 and COX-2, which share significant homology (Vane and Botting, 2003) but differ in their tissue expression and selectivity (Kalgutkar *et al.*, 1998), the latter determined by a single amino acid difference in their catalytic centres (Gierse *et al.*, 1996). COX-1, constitutively expressed in the endoplasmic reticulum of most cells, functions in the production of homeostatic prostaglandins including prostaglandin E (PGE_2), necessary for the maintenance of gastrointestinal homeostasis, and thromboxane A_2 (TXA_2), involved in the regulation of platelet activity and aggregation (Patrono, Patrignani and Rodríguez, 2001; Capone *et al.*, 2007). In contrast, COX-2 is not generally present in most mammalian cells, but is inducible by inflammatory stimuli for a short-lived, significant generation of prostaglandins with vasodilatory and anti-aggregatory effects (Vane and Botting, 2003). A limited constitutive expression of COX-2 does however exist in discrete locations, contributing, among other roles, to the sustained production of vasoprotective prostacyclin (PGI_2) (O'Donnell, 2003) in endothelial cells, and protection of the gastric mucosa (Peskar, 2001).

Both COX-1 and COX-2 homodimers, consisting of 576 and 581 amino acids respectively (Simmons, Botting and Hla, 2004), possess cyclo-oxygenase and

peroxidase activity, and consequently serve the same purpose, catalysing the same rate-limiting stage in prostanoid synthesis: the conversion of AA to prostaglandin H (PGH_2) (Capone *et al.*, 2007). Their contrasting functions are therefore attributable to other biological differences between the two isozymes, for example the distinctive regulation of their expression (Capone *et al.*, 2007), or their requirement for different levels of AA to promote catalysis. Responses are also dictated by the type and levels of the specific G-protein-coupled receptors with which they interact in both an autocrine and paracrine manner (Narumiya and FitzGerald, 2001; Smyth *et al.*, 2009).

Aspirin achieves its primary cardioprotective effects through the irreversible inhibition of COX. In an example of suicide inactivation, the acetylation of specific serine residues (Ser-530 in COX-1, Ser-516 in COX-2) by aspirin essentially modifies, and prevents access to, the cyclo-oxygenase active site to disrupt substrate binding (Loll, Picot and Garavito, 1995). While this promotes COX-2 to convert AA to 15-R-hydroxyeicosatetraenoic acid instead of PGH_2, it completely deactivates COX-1, the isoform for which aspirin has an approximate 170-fold greater potency in inhibition (Vane and Botting, 2003). A resulting blockade of TXA_2 production in platelets inevitably reduces coagulation and thrombus formation, effects which potentially prevent CVD events.

Additional mechanisms have been proposed to further explain platelet inhibition by aspirin. These include the suppression of platelet activation by neutrophils, a process regulated by a nitric oxide (NO)/cGMP-dependent pathway (López-Farré *et al.*, 1995), and propagated by the increased NO production in endothelial cells that accompanies reduced prostacyclin synthesis. Moreover, its potential antioxidant character may also provide a means by which aspirin has clinical utility in the general treatment of CVD, possibly restricting the progression of atherosclerosis by preventing the oxidative modification of low-density lipoproteins (Steer *et al.*, 1997). Studies continue to investigate the pharmacodynamics of this salicylate, but it is probable that any reductions in the incidence of CVD issues are realised through a combination of the aforementioned effects and others, while doses as low as 75mg daily appears sufficient to deliver these.

1.1.3. Prevention of Cerebrovascular Disease

Aspirin therapy is also advocated for the treatment of cerebrovascular disorders, with its efficacy again varying with context. A meta-analysis of 12 randomised controlled trials (Rothwell et al., 2016) showed the use of low-dose aspirin as an acute therapy, administered within 48 hours of an ischemic stroke or transient ischemic attack significantly reduces the risk of recurrence. Other trials have further reported a decreased incidence of death or significant disability at four weeks of follow-up with such an approach, all importantly accompanied by no significant increase in the occurrence of haemorrhagic stroke (Chinese Acute Stroke Trial Collaborative Group, 1997; International Stroke Trial Collaborative Group, 1997). For the longer-term secondary prevention of cerebrovascular conditions, low-dose aspirin, alone or in combination with additional anti-platelet drugs, is also recommended, supported by the antiplatelet trialists review which reported highly significant decreases in the probability of non-fatal stroke with treatment (BMJ, 1994). The use of aspirin in primary intervention is however less positive, with several trials returning conflicting results on efficacy while describing increases in adverse effects (Peto *et al.*, 1988; Steering Committee of the Physicians' Health Study Research Group, 1989; Cote, 1995). Nevertheless, there is consensus on the benefit of low dose aspirin as a secondary preventive measure, reflected in clinical guidance (National Institute for Health and Care Excellence, 2023).

1.1.4. Reduction of Pain and Inflammation

Higher doses of aspirin realise its traditional analgesic and antipyretic effects primarily through the preferential inhibition of COX-2 (Kalgutkar *et al.*, 1998). This isoform is normally responsible for the production of pro-inflammatory PGI_2 and PGE_2 (Figure 1.1.), mediators which enhance nociception by lowering the activation threshold for the opening of neuronal sodium channels (Murata *et al.*, 1997). However, in the presence of aspirin, synthesis of both prostanoids is significantly reduced, and modified COX-2 yields lipoxins instead (Goel *et al.*, 2012), the majority of which are anti-inflammatory. This coincides with the metabolism of various polyunsaturated fatty acids to generate specialised pro-resolving mediators, compounds arising from lipoxygenase activity that further oppose inflammation (Serhan and Chiang, 2013; Romano *et al.*, 2015).

The presence of COX-1 and COX-2 mRNA in circulating inflammatory cells alternatively indicates that both isoforms may in fact contribute to acute inflammatory responses, with prostaglandins of the former produced in the initial phase, and

upregulation of the latter becoming the dominant pathway in the ensuing chronic phase (Smyth *et al.*, 2009).

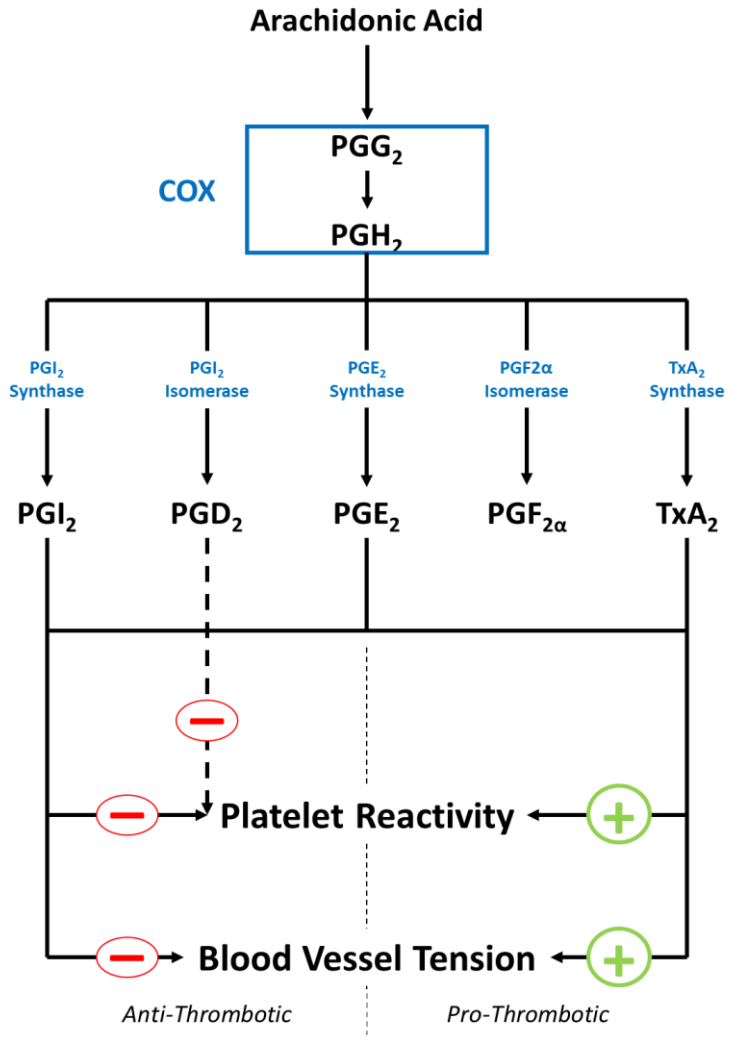

Figure 1.1. Schematic of the principal pathways responsible for prostanoid synthesis and effect in relation to platelets and blood vessels. Arachidonic Acid (AA), liberated from membrane phospholipids by phospholipase, is iteratively metabolised by the heterodimer prostaglandin (PG) H-synthase (COX). Through cyclo-oxygenase and peroxidase activities respectively, AA molecules are cyclised by the addition of a 15-hydroperoxy group to form PGG_2, before the hydroperoxyl group of such is reduced to ultimately yield PGH_2.Both intermediates, described as prostaglandin endoperoxidases, possess individual biological function, but are generally unstable compounds that primarily serves as substrates for subsequent catalysis. PGH_2 is ultimately converted to functional prostaglandins by specific synthases, the most relevant in a cardiovascular context being thromboxane A_2 (TXA_2), produced by thromboxane synthase, and PGI_2 (prostacyclin), produced by PGI-synthase.

1.1.5. Potential Side Effects of Aspirin Use

Notwithstanding its potential therapeutic benefits, aspirin use is associated with a number of adverse effects, many of which result from alterations in the PG environment. The first of these, in part caused by the same antiplatelet mechanisms supporting aspirin prophylaxis in CVD prevention, is the well-documented increased risk of bleeding, particularly of the gastrointestinal (GI) system. COX-1 is the predominant isoform of the gastric mucosa, constitutively expressed and implicated in the production of cytoprotective PGE_2 and PGI_2 (Patrignani *et al.*, 2011). When this enzyme is affected by aspirin, there follows an aberration of this GI cytoprotection, as well as the acknowledged inhibition of TXA_2-mediated platelet function (Patrono *et al.*, 2001). A recent meta-analysis of 11 randomised clinical trials highlighted the consequence of these changes, relating an increased incidence of major GI bleeding with therapy (Hazard Ratio (HR) = 1.56) in a context of low heterogeneity (Zheng and Roddick, 2019). The two effects are nevertheless disparate, with the former evidently dose dependent, increasing the probability of bleeding by propagating mucosal lesions, and the latter independent of dose, chiefly responsible for the increased risk of upper GI bleeding typically observed with antiplatelet therapies (Patrono, Patrignani and Rodríguez, 2001): this explains the aforementioned advocation of lower aspirin doses in CVD intervention.

Additional bleeding complications, again resulting from its antithrombotic effectiveness, seemingly accompany the use of aspirin, including an increased risk of cerebrovascular issues. The evaluation of 16 placebo-controlled trials related such an increase for haemorrhagic stroke in particular, with a reported relative risk (RR) of 1.84 and absolute risk of 12 incidents per 10 000 participants (He *et al.*, 1998). This was subsequently corroborated by the ATT Collaboration (BMJ, 2002), which returned a statistically significant 22% amelioration in stroke incidence. A recent meta-analysis of 13 randomised clinical trials described similar increases in the probability of intracranial bleeding, with an RR of 1.37, and more so subdural or extradural haemorrhage, for which an RR of 1.53 was found (Huang *et al.*, 2019). With aspirin therapy, the balance between preventing vascular occlusion and inducing bleeding ultimately depends on the absolute risk of both. Aspirin does not effect a generalised bleeding problem unless administered to individuals with inherent haemostatic defects, therefore the potential benefits exceed any risk in patients susceptible for CVD or cerebrovascular injury.

A general GI toxicity, related once more to the inhibition of COX-1, and the disrupted cytoprotection of PGE_2 in the gastric mucosa, further accompanies the use

of traditional NSAIDs including aspirin. These factors have been demonstrated to be responsible for mucosal insult throughout the GI tract, ranging from minor superficial lesions to more serious, but less frequent, ulcers (which may precipitate with coincidental bleeding or obstruction). Compared to placebo controls, a greater incidence of lesser GI symptoms (such as nausea and indigestion), peptic ulcers, and minor GI bleeding, have been described with aspirin administration (Awtry and Loscalzo, 2000). Furthermore, as is the case with general bleeding complications, the increased risk, frequency, and severity of issues appears to be dose-dependent, significantly elevated by higher strength therapies (UK-TIA, 1988; Hawkey, 1990; Henry *et al.*, 1996; Hernández-Díaz and Rodríguez, 2000). These effects are maintained even after the initial months of chronic treatment, but completely abates approximately two months following withdrawal (Patrono, Patrignani and Rodríguez, 2001).

On account of the constitutive expression of COX-2 in the human kidney, adverse renal effects have been described with aspirin use, especially in high-risk individuals (FitzGerald and Patrono, 2001). In both cortical and medullary tissues COX-2 is responsible for the synthesis of prostanoids, particularly PGE_2 and PGI_2, with vasodilatory and natriuretic functions (Yu *et al.*, 2010). Thus, in a situation of impaired cyclo-oxygenase activity, the resulting renal vasoconstriction and water retention may increase blood pressure, and compromise efforts for hypertension control. Consistent with a dosing threshold for COX-2 suppression (Patrignani, Filabozzi and Patrono, 1982), the risk of clinically significant effects does however appear relatively small with routinely prescribed aspirin doses (Hansson *et al.*, 1998), and important effects are only observed with higher doses of 1500mg daily (Riegger *et al.*, 1991). Nevertheless, the use of aspirin in patients with hypertension or impaired renal function requires expert assessment (FitzGerald and Patrono, 2001).

Finally, contrary to its foremost therapeutic use, there are suggestions that aspirin therapy may actually contribute to an increased risk of cardiovascular complications, attributable in part to the potential pro-thrombotic effects of COX-2 inhibition. Evidenced by primary literature, and supported by various meta-analyses (Coxib and Traditional NSAID Trialists' Collaboration, 2013), endothelial PGI_2 production in humans contributes to thromboresistance, and is primarily achieved by vascular COX-2 expression (Ricciotti *et al.*, 2013). Still, higher aspirin doses are accepted as inhibiting this COX isoform, and consequently reducing the production of PGI_2. Such reductions in intravascular PGI_2 specifically may reverse any existing

repression of atherosclerotic disease progression and platelet activation, and instead promote vasoconstriction and clotting (Smith, Araki and Lefer, 1980), effects that conceivably propagate CVD conditions. These theoretical concerns have however been proven unfounded, with research suggesting any anti-thrombotic effects on cardiovascular homeostasis (resulting from the aspirin-induced suppression of TXA_2 production) ultimately exceed any pro-atherogenic, pro-thrombotic effects from reduced PGI_2, especially at clinically relevant doses (Patrono *et al.*, 2001).

1.2. Aspirin in Cancer Prevention

In 1988, Kune et al described a case control study in Melbourne which indicated that people taking aspirin, or other NSAIDs, developed significantly fewer colorectal cancers (Kune, Kune and Watson, 1988). Evidence from several subsequent observational studies has related a reduced risk, and long-term incidence, of various cancer types with daily aspirin use (Jacobs et al., 2007; Algra and Rothwell, 2012; Rothwell, Wilson, et al., 2012). This was further supported by a secondary meta-analysis of eight randomised trials of aspirin in cardiovascular disease, analysis of which demonstrated a lower mortality rate, both during and after observation, from several prevalent cancers among those randomised to regular aspirin prophylaxis in the original trials (Rothwell *et al.*, 2011). Of particular interest is colorectal cancer (CRC), in which an approximate 27% reduced risk has been reported with this treatment (Bosetti *et al.*, 2012), resulting in specific studies investigating appropriate dosing strategies. Nevertheless, similar findings have also been disseminated for numerous other adenocarcinomas. The systematic use of 325mg aspirin has been shown to be associated with lower fatality from prostate cancer (Downer *et al.*, 2017), with post-diagnostic employment accompanied by an improved survival from this following diagnosis. Furthermore, a 39% reduction in breast cancer incidence is realised with aspirin prophylaxis (Fraser *et al.*, 2014), a relationship not affected by familial risk or the presence of BRCA mutations (Kehm *et al.*, 2019), while a comparable 20-34% decrease in ovarian cancer risk is apparent with an equivalent therapy (Trabert *et al.*, 2014). This apparent chemoprevention is seemingly not universal however, with no statistically significant association between aspirin therapy and the risk of pancreatic, endometrial, bladder and kidney cancer previously reported, all in coincidence with an increased risk of adverse bleeding events (Bosetti *et al.*,

2012). There are also suggestions any anti-cancer utility may not be realised for at least four years of administration (Burn *et al.*, 2008; Rothwell, Wilson, *et al.*, 2012), with several studies describing significant reductions in cancer mortality from five years onwards (Rothwell, Price, *et al.*, 2012), maintained in several cancer types at 20-year follow-up (Flossmann and Rothwell, 2007; Rothwell *et al.*, 2010, 2011). A review of five randomised trials, assessing prophylactic aspirin use for CVD prevention, explored how this delayed response may be the consequence of suppressed tumourigenesis or metastasis (Rothwell, Wilson, *et al.*, 2012). This found a lower risk of cancer with distant metastasis following allocation to aspirin (HR=0.64), as well as reductions in the development of metastasis (HR=0.45) and mortality (HR=0.50), all effects independent of age and gender. Be that as it may, the exploitation of aspirin for chemoprevention remains contentious and not a ubiquitous recommendation, with studies continuing to investigate this application and the mechanisms by which it may achieve any such effect.

Two large-scale randomised controlled trials (RCTs), with cancer as the primary endpoint, have revealed significant benefits of aspirin prophylaxis in chemoprevention, and provided the gold-standard evidence to guide clinical practice: both The Women's Health Study (Cook et al., 2013) and CAPP2 (Burn, Gerdes, *et al.*, 2011) are described in detail below.

1.2.1. Potential Mechanisms of Aspirin Chemoprevention

Aspirin delivers its foremost biological effects through the irreversible acetylation, and disrupted function, of COX enzymes as described previously. The importance of this inhibition, and the precise mechanism by which it also produces any chemopreventative success remains unclear, but both dependent and independent of COX have been theorised (Thun, Jacobs and Patrono, 2012).

In tumour tissue, aspirin has been shown to promote apoptosis (Rao and Reddy, 2004). Caused by the suppression of cyclo-oxygenase activity, and the subsequent absence of the fatty acid's metabolism, one manner in which it does this is through increases in AA which propagate the conversion of sphingomyelin to ceramide, an accepted mediator of apoptosis (Chan *et al.*, 1998). Alternative COX-1-centric theories relate to the prevention of platelet activation, which is known to be elevated in cancer patients, and believed to contribute to tumour progression and metastasis (Gay and Felding-Habermann, 2011). In contrast, the inhibition and compromised transcription of COX-2 (Xu *et al.*, 1999), resulting from aspirin

administration, decreases the production of pro-inflammatory prostaglandins and therefore prevents inflammation, an acknowledged risk factor for cancer.

In CRC specifically, COX-2 is known to be over-expressed (Williams, Smalley and DuBois, 1997), possibly attributed to increased levels of IL-1β, PGDF and TGF-β (Sciulli *et al.*, 2005). This gives an enhanced synthesis of PGE_2, a prostaglandin recognised as responsible for angiogenesis and a resistance to apoptosis, and has been pronounced as advancing several malignancies including those of the breast and lung (Dixon *et al.*, 2013). Interruption of this process through the aspirin-induced modification of COX-2 would feasibly obstruct any such carcinogenic effects, a theory supported by the improved regression of adenomas observed when PGE_2 levels are significantly repressed (Giardiello *et al.*, 2004). COX-2 acetylation also results in the downregulation of phosphatidylinositol 3-kinase signalling (Liao *et al.*, 2012), and the production of lipoxins as previously mentioned (Goel *et al.*, 2012), which oppose tumorigenic proliferation and angiogenesis (Garcia-Albeniz and Chan, 2011). For this multitude of reasons, and the evidence supporting them, it is believed that any chemoprevention by aspirin in CRC is at least partially owing to the disturbance of COX-2 function.

There is however increasing evidence that mechanisms independent of COX are involved in any aspirin-derived chemoprevention: principal among these is the prevention of NF-κB activation (Kopp and Ghosh, 1994). IκB-kinase, an enzyme responsible for regulating the activity of NF-κB, is accepted as directly interacting with aspirin, and its signalling pathway is known to contribute to angiogenesis and inflammation. Such provides a situation in which aspirin conceivably exerts its effects, but accounts on the veracity of this are conflicting. Several studies have explicitly reported the requirement for NF-κB signalling to promote apoptosis in CRC (Stark *et al.*, 2006), and it is possible that contrasting effects are experienced depending on the cell type and context. The interference with cell signalling is however a common theme in how aspirin may arrest cancer development, further attested to by the accepted disruption of extracellular-signal-regulated kinase (ERK) signalling by the salicylate. With this likely involving the restricted binding of c-Raf with Ras (Pan, Chang and Hung, 2008), limiting the ERK pathway may prevent various cellular processes involved in tumourigenesis, such as differentiation and proliferation, and is a concept similar to that observed with the concentration-dependent inhibition of the Wnt/β-catenin pathway (Bos *et al.*, 2006).

In addition to signalling-related explanations, perturbed mitochondrial function could also be involved in any observed chemoprevention. As well as inhibiting calcium transport (Núñez *et al.*, 2006), aspirin increases mitochondrial membrane permeability, allowing for the release of the haemoprotein cytochrome c and the ensuing activation of the caspase cascade (Dikshit *et al.*, 2006). While this ultimately results in apoptosis, in CRC specifically this has also been shown to bring about the proteolysis of transcription factors implicated in angiogenesis and proliferation, all consequences possibly in opposition to tumour development. Aspirin and its derivatives also seemingly affect other processes with potential to disrupt oncogenesis, such as glucose consumption, reduced by the inhibition of 6-phosphofructo-1-kinase (Spitz *et al.*, 2009), and cell growth, limited by the downregulation of c-Myc, Cyclin A and Cyclin D1 (Law *et al.*, 2000). Regardless of the specific mechanism, it appears from the preceding assessment that any COX-independent chemoprevention involves the direct or indirect manipulation of mitochondrial function, signalling and/or transcription factors, but these remain unconfirmed and requiring further elucidation.

1.2.2. Prevention of Colorectal Cancer

As previously referenced, the chemopreventative potential of aspirin is particularly evident in CRC, and multiple studies have been dedicated to exploring this specific use.

CRC is the third most prevalent cancer type globally, and fourth most common cause of cancer-related fatality, accounting for approximately 700 000 deaths annually (Mármol *et al.*, 2017). While its global distribution varies considerably, an increase of 60% in its global incidence is expected by 2030 (Arnold *et al.*, 2017), with rising rates in younger adults considered average risk possibly contributing to such a development. The implementation of regular screening programmes allows for the early detection of colorectal polyps before they progress, and improved treatment strategies have realised advances in survival rate. However, given the predictions of increasing CRC occurrence, and growing risk for specific demographics, alternative therapeutic options may be required in the future.

Since the first case control study indicative of CRC prevention with aspirin prophylaxis (Kune, Kune and Watson, 1988), several retrospective meta-analyses have been performed, in addition to RCTs, that support this concept. These include an assessment of 30 observational studies that reported a reduced risk of developing CRC with regular aspirin use (RR=0.73, 95% CI=0.67-0.79) (Bosetti *et al.*, 2012), as

well as a Danish case-control study that related similar reductions with continuous low-dose therapy for at least five years (Odds Ratio (OR)=0.73, 95% CI=0.54-0.99) (Friis *et al.*, 2015). Rothwell et al. further traced over 25 000 participants of eight major cardiovascular trials and found a significant reduction in deaths from gastrointestinal cancers with allocation to aspirin (HR=0.65, 95% CI=0·54-0·78) (Rothwell *et al.*, 2011). This finding was supported by that of the Women's Health Study, which originally reported no benefit from 10 years of alternate day 81mg aspirin versus placebo in 18000 women, but revealed upon post-trial review a significantly lower incidence of CRC among those randomised to aspirin (HR=0.80, 95% CI=0·67-0·97), with this effect emerging after a decade (Cook et al., 2013).

The CAPP2 RCT, with cancer as a primary endpoint, focussed on aspirin prophylaxis in Lynch syndrome specifically (discussed in detail in Section 1.3.). This trial used a higher sub-analgesic dose of 600mg of aspirin versus placebo for two to four years, with planned 10 year follow up, and revealed a reduction in CRC occurrence. Initially this was non-significant on an intention-to-treat (ITT) analysis (HR=0.63, 95% CI=0.35-1.13), but significant on per-protocol analysis following two years of intervention (HR=0.41, 95% CI=0.19-0.86) (Burn, Gerdes, *et al.*, 2011). The definitive, intended 10-year follow up also revealed a 50% reduction in the per-protocol group using the incidence rate ratio (HR=0.50, 95% CI=0.31-0.81), with this taking account of recurrent primary cancers and, crucially, a significant result on ITT (HR=0.65, 95% CI=0.43-0.97) (Burn *et al.*, 2020).

Recent analyses have explored the association of response to aspirin with the different molecular pathological subtypes of CRC (Amitay *et al.*, 2019). These found, with regular treatment, a reduced cancer risk associated with several genetic subtypes including tumours that are BRAF wild type (OR=0.67, 95% CI=0.58 - 0.78) or KRAS wild type (OR=0.68, 95% CI=0.58 - 0.80). Reductions appear less pronounced with longer-term therapy, and were mostly not statistically significant in BRAF-mutated and KRAS-mutated. Such findings are indicative of a variable response to chemopreventative aspirin prophylaxis depending on the molecular character of the CRC, and further suggests that any effects may be dependent on the mechanisms by which a given tumour develops and is maintained.

1.2.3. Importance of Cyclooxygenase Activity in CRC

While numerous genetic, epigenetic and inflammatory mechanisms have been implicated in CRC tumourigenesis, considerable evidence exists for the contribution of constitutive, aberrant COX-2 expression to cancerous developments (Dixon *et al.*, 2013). Through compromised prostaglandin biosynthesis, the irregular activity of this cyclooxygenase may affect colonic carcinogenesis at various stages, including an eventual involvement in hyperplasia, dysplasia and metastasis.

The COX-2-dependent generation of PGE_2, the most abundant prostaglandin in the disease (Rigas, Goldman and Levine, 1993), is seemingly crucial in human CRC (D Wang and DuBois, 2010), responsible for altering signal transduction pathways involved in processes such as proliferation, angiogenesis, apoptosis and cellular adhesion (Ferrandez, Prescott and Burt, 2003; D Wang and DuBois, 2010). COX-2 expression is typically controlled at the post-transcriptional level through the interaction of RNA-binding proteins with various sequence elements contained in the 3' untranslated region of its mRNA (Patrignani and Patrono, 2015). In particular, associations with a conserved AU-rich element (ARE) function to direct COX-2 mRNA for degradation, and subsequent inhibition, to maintain appropriate levels. However, in CRC the COX-2 ARE appears dysfunctional, resulting in an increased mRNA stability and the recognised upregulation of COX-2 gene expression observed in the tumour microenvironment (Dixon *et al.*, 2001). The findings may further be explained by the binding of the stabilising factor human antigen R, for which overexpression and cytoplasmic localisation in CRC has been found (Young *et al.*, 2009).

The involvement of COX-2 in colorectal tumorigenesis is evidenced by the efficacy of selective inhibitors in reducing disease risk (Steinbach *et al.*, 2000; Baron *et al.*, 2006; Bertagnolli *et al.*, 2006), particularly with long-term, high intensity administration (Friis *et al.*, 2015). In contrast, aspirin may realise chemopreventative effects on this cyclo-oxygenase activity upstream of the enzyme itself, arising through the aforementioned modulation of platelet activation. As platelet-derived mediators such as cytokines may enhance COX-2 expression in adjacent epithelia (Thun, Jacobs and Patrono, 2012; Dovizio *et al.*, 2013), the restriction of platelet function would logically prevent any upregulation, and ultimately contribute to the reduced CRC incidence observed with low-dose therapy. This theory is however unconfirmed, and its importance is to be demonstrated, but the identification of suitable patients for any such therapies is crucial given potential contraindications.

1.2.4. Biomarkers Informing the Use of Chemopreventive Aspirin Prophylaxis

The identification of biomarkers for the chemopreventative effects of aspirin is a necessary and continuing process, allowing for individualised assessments of the advantages and disadvantages of therapy.

The status of BRAF, a fundamental component in the RAS-proliferative signalling pathway, is seemingly predictive of response to aspirin prophylaxis, particularly in CRC. Regular aspirin use is associated with a lower cancer risk in those possessing the wild-type gene, as well as improved survival following diagnosis, while mutated forms are associated with a resistance to aspirin's anticancer capacity (Nishihara, Wu, *et al.*, 2013). The inverse situation appears to be the case for PIK3CA mutation status, with adjuvant use returning improved CRC survival in mutant but not wild-type cases (Liao *et al.*, 2012). A combination therapy of aspirin and PI3K pathway inhibitors has since been proposed for the targeted treatment of breast cancer (Henry *et al.*, 2017), with the greatest sensitivity to the former observed in the presence of both *PIK3CA* and *KRAS* mutations (Turturro *et al.*, 2016). Expression levels of 15-hydroxyprostaglandin dehydrogenase (15-PGDH), a metabolic antagonist of COX-2, may further represent a potential biomarker of benefit from aspirin chemoprevention. As demonstrated by a review of 15-PGDH mRNA levels in the normal mucosa of CRC patients, regular aspirin use realised a lower cancer incidence with higher 15-PGDH expression, indicative of interaction with an associated, active tumour suppressor pathway (Fink *et al.*, 2014). This contrasts the effect of aspirin observed relative to programmed death-ligand 1 (PD-L1) levels, where situations of reduced expression, and the consequences of such on immune checkpoint blockade, are associated with greater survival, albeit limited again to CRC (Hamada *et al.*, 2017).

Possible biomarkers indicative of the potential of aspirin use in CRC specifically have primarily been derived from the study of three large cohorts (Chan, 2009; Bastiaannet *et al.*, 2012). In the first instance these reported positive effects on outcome with COX-2 overexpression, a relatively common feature of colonic tumours, which was not realised with weak or absent expression. This predictive utility of COX-2 levels has however not been observed in other CRC studies (Reimers *et al.*, 2014), or comparative research in breast cancer (Holmes *et al.*, 2011). Any benefits of therapy appear further restricted to tumours expressing human leukocyte antigen (HLA) Class I antigens, and absent from those lacking such expression (Reimers *et al.*, 2014). This finding again requires corroboration, but is of potential significance given that loss of HLA expression is a common feature of CRCs.

There are conflicting opinions on the practicality of using PIK3CA levels as a therapeutic biomarker, for which the benefits of mutant status were previously mentioned. While earlier reports describe improved survival (Chan, 2009) and recurrence rates (Domingo *et al.*, 2013) with aspirin in the context of *PIK3CA* mutation, more recent studies contradict these findings, relating equivalent effects in wild-type situations (Bastiaannet *et al.*, 2012), as well as a lack of statistical significance in any advances seen with therapy (Reimers *et al.*, 2014; Kothari *et al.*, 2015). These reviews do nonetheless endorse previous assertions about the prognostic value of wild-type BRAF, showing an association between aspirin use and reduced CRC risk in such circumstances (Nishihara, Lochhead, *et al.*, 2013).

Substantial evidence currently exists supporting the prophylactic use of aspirin for chemoprevention, and understanding of how any benefits are achieved continues to develop through research dedicated to this. Nevertheless, given contradictory findings and the possibility of adverse side effects, this remains a limited practice, and decisions on its employment should be made on an individualised basis, considering a patient's absolute risk. One group for whom the benefits of aspirin therapy have been demonstrated to exceed any potential risk, however, are those with Lynch syndrome.

1.3. Lynch Syndrome

Lynch syndrome (LS) describes a hereditary condition arising from germline mutations in genes of the DNA mismatch repair (MMR) system. In accordance with Knudson's 'Two-Hit' hypothesis (Knudson, 1971), the subsequent loss of heterozygosity of an affected MMR gene results in MMR-deficiency (MMRd), compromising genomic stability and facilitating an increase in the frequency of further somatic mutations. LS individuals consequently experience a significant predisposition to, and early onset of, various cancer types (Table 1.1.), the most common being those of the colorectum and endometrium, with cumulative incidences of up to 57.1% and 48.9% by age 75 years respectively (Dominguez-Valentin *et al.*, 2020). The percentage of cases observed for other cancer types varies considerably, ultimately dependent on the underlying genetic mutation, but current estimates of LS prevalence in the general population are 1 in 279 (Win *et al.*, 2017).

The first clinical description of LS was provided by Alfred Scott Warthin in 1913, who documented cancer diagnoses in thirty-three members of the pedigree "Family G"

(Douglas *et al.*, 2005), indicative of an autosomal-dominant inheritance of elevated cancer risk. Subsequent studies of additional cancer families by Henry T. Lynch, including a further review of Family G (Lynch and Krush, 1971), resulted in the proposal that there was a cancer family syndrome, of which two types (Type I and Type II) exist based on their associated tumour spectra. With the emergence of efforts to identify the genes underlying hereditary traits, a research community developed using the first accepted term for LS, Hereditary Non-Polyposis Colorectal Cancer (HNPCC), to differentiate these families from Familial Adenomatous Polyposis (FAP). The name was adopted into clinical practice, but was challenged on the basis of being too restrictive, and an inelegant definition based on absence of another condition.

Given his prolonged efforts to establish the existence of the disorder, the eponym ‘Lynch syndrome’ was eventually proposed to replace HNPCC, recognising the spectrum of tumours involved in the condition, and unifying the subtypes of this by their shared aetiology. In 1993 the genetic cause of HNPCC was identified deduced from the analysis of polymorphic microsatellites and concurrent in vivo experiments in yeast with loss-of-function MMR gene mutations (Strand *et al.*, 1993). A causative relationship between HNPCC and MMR gene defects was ultimately confirmed through the observation of *MSH2* pathogenic variants segregating with instances of disease (Fishel *et al.*, 1993; Leach *et al.*, 1993), before additional MMR genes were found to be implicated (Bronner *et al.*, 1994; Nicolaides *et al.*, 1994; Miyaki *et al.*, 1997).

To optimise their clinical management, the proficient diagnosis of LS individuals is essential. Known carriers of LS-associated gene defects may benefit from various practices, including genetic counselling, regular tumour surveillance, personalised cancer treatment, and/or, as more recently demonstrated, daily aspirin prophylaxis.

	Organ	Cumulative Incidence at Age 75 Years (% (95% CI))			
		path_MLH1	***path_MSH2***	***path_MSH6***	***path_PMS2***
Colorectal Cancer	Colon	46.7 (39.2 – 54.3)	42.4 (32.9 – 51.9)	14.2 (3.1 – 25.4)	0
	Sigmoid and Rectum	11.8 (7.2 – 16.4)	18.3 (10.9 – 25.6)	4.6 (0.0 – 9.7)	0
Endometrial and Ovarian Cancer	Endometrium	42.7 (33.1 – 52.3)	56.7 (41.8 – 71.6)	46.2 (27.3 – 65.0)	26.4 (0.8 – 51.9)
	Ovaries	10.1 (4.8 – 15.4)	16.9 (5.7 – 28.0)	13.1 (0.0 – 31.2)	0
Upper Gastrointestinal Cancer	Stomach	7.1 (3.5 – 10.8)	7.7 (1.9 – 13.6)	5.3 (0.0 – 13.1)	0
	Duodenum	6.5 (2.7 – 10.2)	2.0 (0.1 – 4.0)	0	0
	Bile Duct and Gall Bladder	3.7 (1.3 - 6.2)	1.7 (0.0 - 5.1)	0	0
	Pancreas	6.2 (2.6 – 9.8)	0.5 (0.0 – 1.5)	1.4 (0.0 – 4.2)	0
Urinary Tract Cancer	Urinary Bladder	4.1 (1.5 – 6.7)	8.1 (2.8 – 13.3)	8.2 (0.0 – 16.9)	0
	Ureter and Kidney	4.6 (1.6 – 7.6)	17.8 (10.6 – 25.0)	3.0 (0.0 – 7.0)	0
Other LS and Possible LS Cancer	Breast	12.0 (6.7 – 17.3)	11.5 (4.6 – 18.4)	13.3 (2.2 – 24.4)	55.9 (0.0 – 100.0)
	Prostate	16.9 (8.5 – 25.3)	31.6 (11.7 – 51.5)	18.3 (0.0 – 44.4)	37.9 (0.0 – 95.9)
	Brain	1.0 (0.0-2.4)	5.3 (0.2 – 10.3)	1.4 (0.0 – 4.2)	0

Table 1.1. (Adapted from the Prospective Lynch syndrome Database) Cumulative incidences of individual Lynch and possible Lynch cancers, delineated by *path_MMR* variant, at 75 years of age. Incidence of breast and gynaecological cancers were calculated exclusively in females, while prostate cancer incidence was calculated in males; all other cancers comprise results from both genders combined.

1.3.1. Aspirin Chemoprevention in Lynch Syndrome

In 1993, Richard Kolodner, a yeast geneticist from Boston, approached the Newcastle research team in search of the DNA from families with HNPCC to test the theory that a MMR gene defect might underlie the condition, based on the observation that LS cancers have a high level of mutation. Kolodner's team were the first to identy a variant in hMSH2 segregating in one Northumberland family (Fishel *et al.*, 1993), and this was immediately introduced to clinical practice in Newcastle in 1994. The discovery of the genetic basis of HNPCC offered an alternative study population, with CRCs developed without a large number of adenomas in an older population more likely to comply with prolonged intervention. By 1997, funding and international consensus had been achieved to launch a second genetically-targeted factorial RCT using the same interventions in people with LS.

The CAPP1 RCT, with CAPP standing for Colorectal Adenoma/carcinoma Prevention Programme (later simplified to Cancer Prevention Programme - CaPP) tested the effect of aspirin (600mg/day) and/or resistant starch (RS) (30g/day) on disease progression (Burn, Bishop, *et al.*, 2011) in individuals with FAP. This found a non-significant reduction in both the size and number of polyps, precursors of colonic malignancy, in those assigned aspirin therapy. Furthermore, neither intervention was observed to significantly reduce polyp levels in the rectum and sigmoid colon after a median intervention period of 17 months (Aspirin RR = 0.77, 95% CI=0.54-1.10; RS RR=1.05, 95% CI=0.73-1.49).

The ensuing CaPP2 RCT randomised 1007 individuals diagnosed with LS in 43 centres across 16 countries. With CRC as the primary endpoint, this study intended to offer 600mg or placebo for two years, later extended to four years, and used a commercial resistant starch, Novelose, in place of the potato starch/Hylon 7 maize used in CAPP1. In total, 861 individuals were eligible for analysis in the aspirin limb of the trial (Burn *et al.*, 2020). Following a mean intervention period of 25 months, initial reports described no significant difference in the incidence of colorectal neoplasia with aspirin intervention (Burn *et al.*, 2008). However, long-term follow up was designed into this trial. When the first CaPP2 recruits reached the 10 year follow up period in 2009, a subsequent analysis of cancer incidence demonstrated a reduction in CRC in the aspirin arm which was significant on per-protocol analysis but did not reach significance on ITT analysis. Further review at the planned 10 years of follow-up for the entire study population found a significant reduction in CRC incidence was

recognised (HR=0.56, 95% CI=0.34-0.91), with comparable findings maintained in the corresponding ITT analysis (HR=0.65, 95% CI=0.43-0.97). The incidence rate ratio of this appraisal, accounting for all primary CRCs rather than just the first, also indicated a 50% reduction in CRC incidence in participants who took 600mg of aspirin for a period of two years.

The benefits of intervention seemingly develop approximately five years into treatment, consistent with the delayed effect of aspirin observed in other populations (Drew, Cao and Chan, 2016), and exhibit a 'legacy' effect, providing sustained CRC prevention in those subject to therapy for a finite duration. Of further potential note is the coincidental, if not statistically significant, reduction in endometrial cancer (EC) reported among female participants (HR=0.50, 95% CI=0.22-1.11), as well as the suggestion of enhanced effects in obese LS carriers (Movahedi *et al.*, 2015). Regardless of these tangential observations, the principal findings of this study ultimately support the recommendation to consider aspirin prophylaxis for the prevention of CRC at least in LS carriers, and indicate a possible opportunity in early adulthood for exposure that maximises the benefit-risk profile.

Serious adverse effects in both CAPP trials were rare, and no more common in participants assigned therapy versus placebo. However, due to the young mean age of recruits, reservations remain about the appropriate dose and treatment strategy for LS-associated chemoprevention in routine practice given the risk of side effects. The ongoing CaPP3 clinical trial (NCT02497820) is currently assessing the optimal aspirin dose to minimise complications, while returning equivalent chemoprevention as earlier reported. In their guidance on the use of aspirin in LS, the UK National Institute for Clinical and healthcare Excellence (NICE) noted in 2020 (DG151) the uncertainty around optimal dose, but still directed physicians to recommend aspirin to people with LS, thus exposing the continuing shortfall in LS detection. The incidence of 1 in 279 reported for LS by Win et al (2017), would indicate at least 100,000 adults in England with the condition, yet fewer than 10,000 are known to cancer genetic services (J. Burn Personal Communication). The need to improve the methods available for the recognition of LS in healthcare therefore exists.

1.3.1. Mismatch Repair Deficiency

DNA damage occurs, and progressively accumulates, with exposure to exogenous mutagenic chemicals and radiation, as well as endogenous reactive compounds produced in normal metabolism. It may also arise through the aberration of routine genetic processes, including replication and recombination, where the misincorporation of nucleotides during DNA synthesis, observed more frequently with lower-fidelity polymerases, produces base-base mismatches. If unrepaired, damage, and resulting genetic instability, may yield mutations that ultimately change a cellular phenotype and cause dysfunction, as well as facilitate the progression of disease. To prevent such deleterious effects and maintain genomic integrity, several mechanisms have evolved to recognise and correct instances of DNA corruption (Jackson and Bartek, 2009), including the mismatch repair system.

MMR is a highly conserved, ATP-dependent, multi-step process primarily responsible for the correction of mismatches and insertion-deletion loops generated by polymerase error during DNA replication. First characterised in Escherichia coli (Su and Modrich, 1986), MMR in prokaryotes is initiated through the detection of lesions by MutS, a homodimer with intrinsic ATPase activity. When bound to DNA, this physically interacts with, and recruits, a similar homodimeric 'DNA-clamp' MutL, with the resulting complex improving mismatch recognition, as well as coordinating the repair process through accessory enzymes (Kunkel and Erie, 2005). The revision of replication errors ultimately requires the newly synthesised DNA strand containing the given inaccuracy to be targeted for excision and replacement. This is achieved by the latent endonuclease activity of MutH, a restriction enzyme specific to this model organism. Recruited and activated by the MutS-MutL complex, MutH subsequently cleaves the DNA at hemi-methylated GATC sites located within proximity of the error, before engaging DNA helicase II and relevant exonucleases for the digestion of such (Kunkel and Erie, 2005). Repair is eventually completed by highly accurate DNA polymerase III and DNA ligase, which correctly resynthesise the removed sequence, and seal its terminal nick, respectively. As well as determining the process of prokaryotic MMR and the proteins involved, the early experiments on E. coli from which this mechanism was elucidated demonstrated further features important to the pathway. Primarily, repair appears strand specific, restricted to the newly synthesised sequence, and is executed in a bidirectional manner, dependent on the position of the incision introduced by MutH relative to the error. Furthermore, MMR has an extensive

substrate specificity, functioning for the correction of both single base mismatches and insertion-deletion discrepancies. Given the significant conservation of this process throughout evolution, this therefore provides a convenient model for eukaryotic MMR, and the basis from which an understanding of this could be developed.

Various similarities exist between the prokaryotic MMR system and that of eukaryotes, including its bidirectionality and nick-directed strand specificity. The function of hemi-methylated sequences in strand discrimination is however not a conserved feature, while the proteins implicated in the pathway also differs, with further variation depending on the type of the mismatch and the substrate requiring excision (Kunkel and Erie, 2005). Eukaryotic repair is initiated by the detection of, and binding to, a given error by heterodimeric complexes of MutS homologs. Of these there are two distinct forms with disparate protein compositions and precise functions: MutSα (MSH2-MSH6), primarily responsible for correcting single-base and 'short' insertion-deletion (IDL) discrepancies, and MutSβ (MSH2-MSH3), which preferentially repairs larger IDL mismatches up to 16 additional nucleotides in length (Li, 2003). The subsequent connection of both with Proliferating Cell Nuclear Antigen (PCNA) facilitates the recruitment of MutL, a heterodimer essential for coordinating the repair process of which three distinct homologs exist (Kunkel and Erie, 2005). While the function of MutLβ (MLH1-MLH2) in eukaryotes is currently unknown, MutLα (MLH1-PMS2) is responsible for the repair of mismatches, with MutLγ (MLH1-MLH3) apparently associated with crossing during meiotic recombination (Li, 2003; Santucci-Darmanin and Paquis-Flucklinger, 2003). Hydrolysis of the nascent, erroneous strand is eventually performed by exonuclease 1 (EXO1), inducted through interaction with MLH1, MSH2 and/or MSH3. This enzyme, essential for both 5'- and 3'-directed excision, can readily achieve the former in the presence of MutSα or MutSβ, but its functioning in catalysing the former requires the additional latent endonuclease activity of MutLα, stimulated by PCNA and replication factor C. As with its prokaryotic equivalent, repair concludes with DNA resynthesis, catalysed by a high-fidelity, aphidicolin-sensitive polymerase such as DNA polymerase δ (Longley, Pierce and Modrich, 1997).

In recent decades extensive research into the MMR proteins has been conducted, but ambiguity remains with regards to the precise nature of the process and potential redundancy within it, as well as the consequences of aberrations to the repair apparatus.

Dysfunction in the repair mechanism, resulting from MMR gene mutations, is termed mismatch repair deficiency (MMRd), and results in an increased spontaneous mutation rate. Autosomal dominant variants in *MLH1*, *MSH2*, *MSH6* and *PMS2* genes are responsible for Lynch syndrome (Lynch *et al.*, 2009), and the genes account for approximately 50%, 40% 7-10% and <5% of diagnosed cases respectively. Incidence however varies with age of onset, as well as the location and clinical stage of a given tumour (Pearlman *et al.*, 2017; Álvaro *et al.*, 2019; Ciardiello *et al.*, 2019). Rarely, deletions in the EPCAM gene, neighbouring *MSH2*, lead to loss of the *MSH2* promoter, and a CRC-only phenotype, while epigenetic constitutional silencing of *MLH1* also accounts for a small number of isolated LS cases (Hitchins *et al.*, 2011).

MMRd tumours are associated with particular features, as exemplified by the poor differentiation and a better stage-adjusted prognosis compared to mismatch repair-proficient (MMRp) tumours (Popat, Hubner and Houlston, 2005; He *et al.*, 2016). Failure to correct intragenic mutations in the context of defective repair may lead to the generation of frameshift peptides considered foreign by a host's immune system. This inherent immunogenicity results in an increased presence of tumour-infiltrating lymphocytes, and accounts for an enhanced response to immune checkpoint blockade therapies such as PD-1 inhibitors (Lee *et al.*, 2016). MMRd may thus be used as a prognostic tool as a predictive marker of treatment efficacy, and as a means of identifying patients possessing this may direct therapy decisions.

1.3.2. Microsatellite Instability

Microsatellites are tandemly repeated DNA sequences of 1-6bp, and can be subdivided into mono-, di-, tri-, tetra-, penta- and hexanucleotide repeats depending on the length of the recurring unit (Subramanian, Mishra and Singh, 2003; Ellegren, 2004). On account of their repetitive nature, microsatellites are highly mutable, with three proposed mutational mechanisms: (1) polymerase slippage during DNA replication creating IDLs that are stabilised by the repetitive sequence (Kornberg *et al.*, 1964), (2) homology-driven incorporation of retrotransposons (Nadir *et al.*, 1996), and (3) unequal crossing over in meiosis (Huang *et al.*, 2002). The degree of this mutability depends on several factors including the genomic context, the structure of the microsatellite, and the sequence surrounding it (Bacolla *et al.*, 2008; Chung *et al.*, 2010; Kelkar *et al.*, 2011). Taking these factors into account, in vivo (Strand *et al.*,

1993), in vitro (Schlötterer and Tautz, 1992) and *in silico* (Dieringer and Schlötterer, 2003) analyses have all shown that polymerase slippage is the best model for microsatellite mutation rate and, hence, it is accepted as the predominant mechanism (Fan and Chu, 2007). “Microsatellite instability” (MSI), an example of a mutational signature, is the term used to define this mutability.

MMRd presents with the molecular phenotype of exceptionally elevated MSI, a mutator phenotype which may increase the frequency of additional genome-destabilising mutations, and promote tumourigenesis through the development of frameshifts (Duval and Hamelin, 2002). This relationship, first correlated in tumours of HNPCC patients, has since been demonstrated in various malignancies (Boland *et al.*, 1998), and reported to compromise genes with potential effects on carcinogenesis (Duval and Hamelin, 2002; Li, 2008). In the context of defective repair, insertions, deletions and mismatches occurring through DNA metabolism fail to be detected and corrected, but accumulate. Increased MSI therefore behaves as a biomarker associated with MMRd, defining the subtype of tumours with this.

1.3.3. Clinical Criteria for the Detection of Lynch syndrome and Current Screening Techniques

For the identification of potential LS individuals consensus clinical guidelines were initially used, with the first of these, described as the Amsterdam criteria, being introduced in 1991 (Vasen *et al.*, 1991). Informally described as ‘3-2-1’ criteria, these stipulated a minimum of three family members diagnosed with CRC across two successive generations, and at least one of these cases occurring before the age of 50 years. These were however considered too restrictive, and in 1999 the Amsterdam criteria were extended to recognise the extensive spectrum of tumours associated with LS, as well as exclude instances of FAP (Vasen *et al.*, 1999). The coexisting Bethesda guidelines, first published in 1997, provided recommendations concerning who should receive additional tumour testing for LS, given the earlier determination of disease aetiology (Rodriguez-Bigas *et al.*, 1997). This advice described the necessity for genetic analysis in the following situations: a CRC diagnosis before the age of 50 years; the presence of synchronous or metachronous Lynch-associated malignancies (regardless of age); a CRC diagnosis before the age of 60 years with MSI-High histology; at least one first-degree relative with an LS-associated tumour, and/or a

CRC diagnosis with two or more first- or second-degree relatives with LS-associated tumours.

As with the Amsterdam Criteria, these too were subject to revision in 2004, with the improved specification considered more sensitive for the detection of LS risk (Umar *et al.*, 2004). Nevertheless, significant deficiencies were ultimately recognised in both clinical guidelines. In particular, half of germline-confirmed LS instances appear to not satisfy the amended Amsterdam criteria (Lynch and de la Chapelle, 2003), while the relatively low specificity of the Bethesda Guidelines compromises their clinical utility. Given this suboptimal performance of guidance-based approaches, and coinciding developments in the molecular understanding of LS, the identification of disease instances has since evolved from diagnoses based on clinical advice to tumour analysis with genetic confirmation.

In the diagnosis of LS two general approaches are currently employed, differentiated by a given patient's previous association to the disease. To identify candidates with potential pathogenic MMR gene variants, the molecular screening of tumour samples for evidence of MMRd is used, while direct germline testing is favoured for candidates with a personal or family history indicative of LS. Of the former, there has been considerable interest given its sensitivity and specificity in proband detection, and two standard techniques are now available for screening: MMR protein immunohistochemistry and DNA MSI testing. While both tests may be implemented independently (Shia, 2008; Zhang, 2008), the combination of the two maximises accuracy in detection, albeit with a higher individual cost per patient. Moreover, these assays may further be supplemented by the assessment of *MLH1* promoter methylation and/or *BRAF* V600E mutation for the identification of sporadic MMRd situations, in addition to next-generation somatic or germline testing for the confirmation of possible variants. As well as through LS, repair deficiency, and a resulting instability, may occur as a consequence of epigenetic silencing of the *MLH1* gene through hypermethylation of its promoter, a change associated with *BRAF* c.1799T>A, p.Val600Glu mutation (Weisenberger *et al.*, 2006) and observed in approximately 13% of sporadic CRCs (Jass, 2007). This is however mutually exclusive of the hereditary instability observed with LS, and therefore its identification can distinguish sporadic cases from familial.

Immunohistochemistry (IHC) involves the use of reliable antibodies for the MMR proteins to detect the loss of their expression through the absence of staining.

At a minimum, both MLH1 and MSH2 expression is typically assessed, with aberration of these reported responsible for approximately 70% of identified LS cases (Palomaki *et al.*, 2009), but evaluation of the four principal MMR genes is ultimately recommended (Cohen, Pritchard and Jarvik, 2019), recognising the heterodimeric nature of the repair system and the potential for inactivation without protein loss. In CRC detection, a sensitivity of 83-92% and specificity of 89% have previously been demonstrated for this approach (Shia, 2008; Evaluation of Genomic Applications in Practice and Prevention Working Group, 2009), and, contrasting the use of clinical criteria, has been shown to recognise 90% of LS patients (Lynch and de la Chapelle, 2003).

Exploiting the molecular phenotype of MMRd, MSI testing provides an alternative, PCR-based approach for the identification of LS, evaluating genomic variation at short, repetitive sequences. The most prevalent form of this technique involves five microsatellite loci (Boland *et al.*, 1998; Hendriks *et al.*, 2006; Murphy *et al.*, 2006; Patil *et al.*, 2012), with 30% instability, that is, variation in two of more of the subject markers, considered the threshold for designating the presence of elevated MSI (Bacher *et al.*, 2004). Less than 30% variation, affecting a single locus, is regarded as MSI-Low (MSI-L), while no instability returns an MSS classification. These instances of limited variation behave, and are clinically interpreted, similarly. Sensitivities of 55-96% have been reported with MSI testing for LS (Bacher *et al.*, 2004; Evaluation of Genomic Applications in Practice and Prevention Working Group, 2009), but caution is required with this approach as the majority of MSI-H malignancies are seemingly attributable to sporadic mutations. Furthermore, the system is based on the assumption that the large mononucleotide fragments are monomorphic, but polymorphisms have been recorded which can result in the incorrect reporting of an MSI-Low phenotype.

Assaying MSI typically involves the analysis of amplified fragments by capillary electrophoresis, with current approaches dependent on long MNRs (T. A. Boyle *et al.*, 2014). These markers have however been demonstrated unstable both *in vivo* and *in vitro*, with PCR-induced errors known as stutter peaks occurring which complicate subsequent phenotype interpretation and the visual assessment of fragment profiles. This means reporting relies on an expert analyst. The length of the repeats reviewed further means current sequencing technologies cannot be used, with capillary

electrophoresis being replaced as a result by high-throughput, short-read sequencing in diagnostic laboratories.

In an effort to develop an MSI-testing LS-screening procedure suitable for next generation sequencers, previous research by our laboratory group explored the prospect of employing short (7-12bp) mononucleotide repeats, in conjunction with Molecular Inversion Probe (MIP) technology for MSI classification (MIPs are explained in further detail in Section 2.5). Accurate MSI determination was shown using a series of 17 repeat sequences, in combination with a classifier trained by a novel Bayesian approach (Redford *et al.*, 2018). Consequently, an assay was developed of 24 mononucleotide repeat sequences, each associated with an informative polymorphism, to discern the frequency and allelic distribution of microsatellites in CRCs. Including testing by an independent laboratory, this technique has been demonstrated to be both 100% sensitive and specific in MSI status classification (Gallon *et al.*, 2020). Thusfar application of this method has been limited to CRC samples. However, as such has been developed for high-throughput analysis and is suitable for routine clinical use, it offers a potential cost-effective advance to the current screening process. Furthermore, the performance of the assay in constitutional mismatch repair deficiency (CMMRD) detection from peripheral blood leukocytes (Gallon *et al.*, 2020), as well as its use of non-coding markers demonstrated to experience equivalent MSI across cancer types (Cortes-Ciriano *et al.*, 2017), suggests this could also be employed for a similar purpose with extra-colonic LS-spectrum tumours.

Discordance between MSI and MMR IHC is a relatively frequent occurrence, but may result in the misinterpretation of one or both findings. However, disparate results can both be correct, with each representing the different cellular functions and biology analysed by the opposing screening strategies. For example, aberrant protein function, emanating from a missense pathogenic variant, without a corresponding loss of expression may realise an MSI-H tumour with normal IHC. Conversely, some pathogenic variants, particularly in *MSH6* and *PMS2*, have been reported to precipitate protein loss without a subsequent instability phenotype. Despite these caveats, the two techniques remain the advocated screening approaches for LS, either in combination or separately, and are included as the primary device in guidance for the testing of several tumour types.

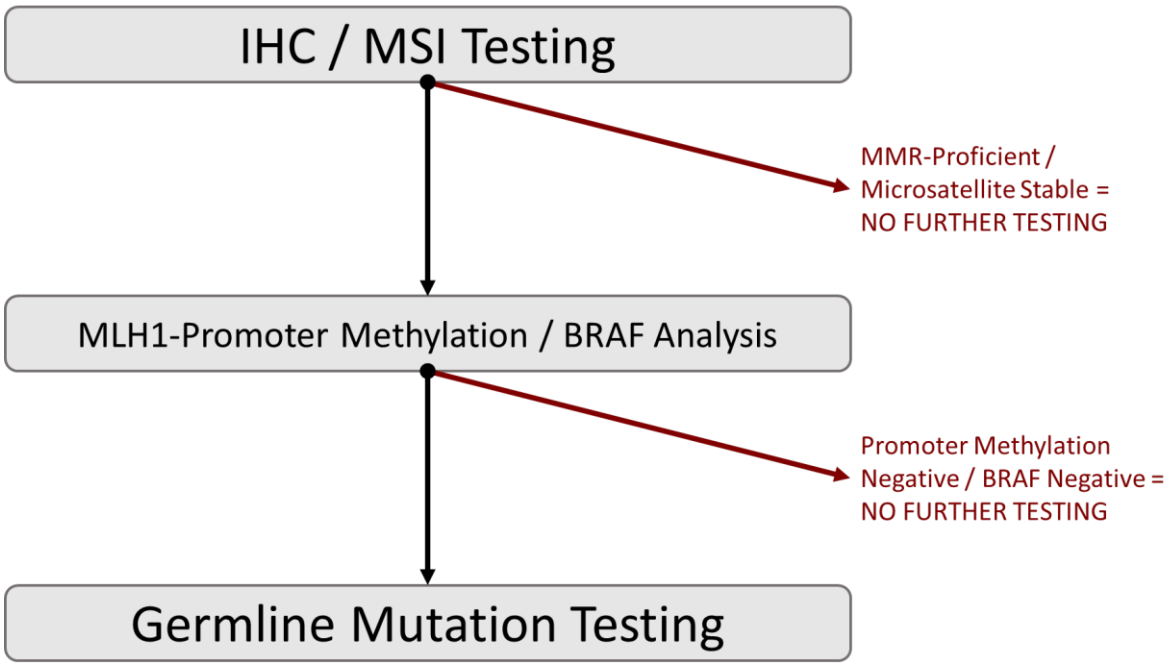

Figure 1.2. Established pipeline for Lynch syndrome screening. Initial assessment for MMRd by either IHC or MSI testing is proceeded by MLH1-Promoter Methylation and BRAF analysis to identify and disregard instances of sporadic-type tumours. In samples negative for these indicators of sporadic MMRd, genetic testing is ultimately performed to validate the presence of a germline mutation in one of the four principal MMR genes.

1.3.4. Screening Guidelines for Lynch syndrome

For the detection of LS, current guidelines from the UK National Institute of Health and Care Excellence (NICE), first published in 2017, advocate the screening of all CRC for LS, regardless of age, clinical presentation or family history. A standard analysis pipeline has been established for such a purpose (Figure 1.2.), with this approach supported by similar recommendations from the European Society of Medical Oncology (Balmaña *et al.*, 2013) and the American Society of Clinical Pathology (Stoffel *et al.*, 2015). Guidance follows considerable evidence demonstrating the advantages of molecular screening strategies compared with screening by clinical criteria, including an analysis of four CRC cohorts reporting improved sensitivity in LS carrier detection relative to the use of the Bethesda Guidelines (Moreira *et al.*, 2012). Additional assessments of cost-effectiveness, juxtaposing the cost of screening, and the subsequent cascade testing of relatives, with the potential benefit of intervention, further show its financial validity (Snowsill *et al.*, 2014). However, these analyses

assume the identification of LS probands and asymptomatic relatives will lead to decreases in cancer burden and costs through preventative measures and increased surveillance (Snowsill *et al.*, 2014), thus highlighting the necessity for a system with the capacity to offer testing to all candidate patients.

More recently, additional NICE guidance has been issued proposing that testing for LS be offered to all individuals diagnosed with EC (National Institute for Health and Care Excellence, 2020). This follows the advice of gynaecologists (Mills et al., 2014), and an international consensus meeting (Crosbie *et al.*, 2019), who advocated the extension of screening practices to this tumour type, in which approximately 3% of cancer cases were found attributable to LS in a systematic review (Ryan *et al.*, 2019). Contrary to the evidence-based recommendations of the Manchester International Consensus Group, which suggests the use of either foremost technique in an EC screening pipeline (Crosbie *et al.*, 2019), NICE guidelines encourage the use of IHC only in the first instance, proceeded by *MLH1* promoter hypermethylation and germline genetic testing as necessary (National Institute for Health and Care Excellence, 2020). Nevertheless, considering the observed frequency of MSI reported in EC with MMRd (Nieminen *et al.*, 2009), it is possible that either screening strategy has utility in this situation. Furthermore, given reports detailing an MSI prevalence of 46% (Therkildsen *et al.*, 2018) and 100% (Akbari *et al.*, 2017) for urothelial and ovarian cancers in an MMRd context respectively, it is conceivable that screening guidance should be extended further, including additional tumour types considered of the LS-spectrum in which comparable, detectable phenotypes as those seen in LS CRC may be reviewed.

In clinical practice, the use of LS screening has however had mixed success. Assuming a cohort in which 3% of individuals were disease carriers, the largest study evaluating universal LS screening reported a 2.2% detection rate, corresponding with the identification of approximately 73% of gene carriers (Heald *et al.*, 2013). In contrast, reports from the United States describe how only 43.1% of CRC patients aged 70 years or below received testing for MMR deficiency between 2010 and 2012 (Shaikh *et al.*, 2018), notwithstanding concurrent estimates the clinical services were aware of only 1.2% of gene carriers (Hampel and de la Chapelle, 2011). This echoes similar findings from the United Kingdom where, despite the promotion of guidance by the Royal College of Pathologists (2014), significant proportions of the health service were estimated to have neglected routine screening for LS (The Royal College of Pathologists, 2016). Explaining these observations, it appears various factors

compromise the implementation of LS testing, while issues exist with the current guidance: all require resolution to realise improvements in screening practices.

1.3.5. Issues with Current LS Screening Guidance

Although the collective understanding of the disease, and the diagnostic techniques to identify it, have developed considerably in recent decades, LS remains greatly underdiagnosed. Research estimates a prevalence in the general population of up to one in 279 (*MLH1* = one in 1946, *MSH2* = one in 2841, *MSH6* = one in 758, and *PMS2* = one in 714) (Win *et al.*, 2017), but the majority of these individuals are unaware of their cancer predisposition. For this, there are several possible explanations.

Despite the advice of multiple authorities, screening targets are not being realised (Shaikh *et al.*, 2018), inevitably resulting in an underdetection of LS gene carriers. A survey of clinicians associated with the American College of Gastroenterology described common reasons for this insufficiency, including prohibitive costs and the unavailability of genetic counselling or germline testing (Noll *et al.*, 2018), with such financial and/or resource limitations reflected in UK services (The Royal College of Pathologists, 2016). Failings in the follow-up of suspected LS patients compound these issues, with a lack of subsequent consultation following the detection of an MMRd tumour shown to correlate with lower levels of LS diagnosis (Brennan *et al.*, 2017). In different demographics the deployment of screening may also be restricted by additional complications, such as a limited access to clinical services, communication barriers or a negative perception of healthcare based on cultural beliefs (Kidambi *et al.*, 2016). Dedicated systems and testing programmes to negate such specific issues are advocated by healthcare professionals (Bombard *et al.*, 2017), and have been demonstrated to return high-quality screening of >90% of candidate patients irrespective of their situation (Kidambi *et al.*, 2016). Nevertheless, such endeavours require extensive, sustained funding and resources to be implemented successfully.

Education, or a lack thereof, at both the clinical and patient level provides a further obstacle in the implementation of LS screening. Numerous surveys of healthcare practitioners and students have reported a lack of LS awareness, or the guidelines related to it, with this confounded by the genetic aetiology and treatment options in those with knowledge of the disease (K. Frey *et al.*, 2012; Tan, Spurdle and

Obermair, 2014; Smith *et al.*, 2017). The same studies however state the most frequent barrier in the referral of suspected LS individuals for further analysis to be a lack of enthusiasm for such from the patient (Tan, Spurdle and Obermair, 2014). Attitudes to additional testing appear improved however with a detailed explanation of its potential benefits, coinciding with reduced patient anxiety (Hunter *et al.*, 2015), thus emphasising the requirement for high-quality patient information to improve the perception of screening and its subsequent acceptance.

In additional to logistical issues, the guidelines informing current screening practices are themselves not without inherent limitations, compromised by LS malignancies that do not present with the phenotypic features assessed in molecular analyses, or non-LS tumours that do. In CRC specifically, the possibility of sporadic MMR-proficient tumours developing in LS gene carriers has been described, with up to 6.9% of cases observed without deficiency through a variety of techniques (Gylling *et al.*, 2008; Yurgelun *et al.*, 2017; Hampel *et al.*, 2018; Porkka *et al.*, 2020). This may evidence the advancing belief that tumourigenesis in LS in a heterogeneous process, dependent on which MMR gene is affected in the germline. MMRd appears a less important contributor to tumour development in individuals with pathogenic *MSH6* and *PMS2* variants, with lower CRC incidence than observed with equivalent *MLH1* and *MSH2* defects (Dominguez-Valentin *et al.*, 2020), while fewer colorectal adenomas from pathogenic *MSH6* variant carriers are associated with repair deficiency (Ferreira *et al.*, 2009). Oncogenic mutations from pathogenic *PMS2* variant carriers are also not always associated with the mutational signature of MMRd (ten Broeke *et al.*, 2018), unlike those of the other MMR genes. Such examples of heterogeneity suggest that the frequency of deficiency in LS, and therefore the sensitivity of screening protocols, will vary by gene affected.

This accuracy is further influenced by the presentation of MMRd and *MLH1* promoter methylation negative tumours that present without germline pathogenic MMR variants, observed in approximately 13% of CRCs in a recent population-based study (Porkka *et al.*, 2019). In families with such a presentation, a significantly higher risk of CRC was found than that of families with sporadic disease, but lower that the risk with LS, thus explaining its description of Lynch-like syndrome (LLS) (Rodríguez–Soler *et al.*, 2013). Previous research initially found 52% of LLS instances attributable to double somatic pathogenic variants in either *MLH1* or *MSH2* (Mensenkamp *et al.*, 2014), but subsequent studies reported that approximately 73% of cases can be

explained by somatic deficiencies in any MMR gene (Pearlman *et al.*, 2019). It is possible the intermediate risk associated with LLS represents a heterogeneous population of unidentified LS patients, as well as sporadic cases and other inherited predisposition syndromes that have since developed MMRd independent of the causative germline variants (Carethers, 2014).

For a definitive diagnosis of LS the confirmation of a germline pathogenic MMR variant is ultimately required. This process may however be frustrated by several factors, including the presence of variants of unknown significance (VUS) in the subject MMR genes, or inadequacies of current technologies to recognise causative variants in these. In 2016, VUSs were realised to account for approximately 30% of MMR variants in the InSiGHT database (Peltomäki, 2016), requiring further investigation to confirm their effect. *In silico* analyses, which predict pathogenicity by combining clinical data and the theorised changes in protein sequence, may be used to prioritise variants for subsequent evaluation, before assays of MMR function, enabled by the conserved nature of the MMR system, are eventually employed for an assessment of a given variant's effects. These, and alternative, techniques are, however, time-consuming, expensive, and typically not available to diagnostic laboratories. For the processing and interpretation of existing MMR VUSs, allowing for their clinical utility, a specific, coordinated effort will therefore be required.

The assessment of *PMS2* presents particular challenges in the detection of pathogenic MMR variants, and exemplifies limitations in the existing techniques for this purpose. Located on chromosome 7 and consisting of 15 exons, the post-meiotic segregation 2 (*PMS2*) gene encodes the corresponding MMR protein, which contributes to repair through its latent endonucleolytic function (Kadyrov *et al.*, 2006). In LS patients, *PMS2* mutations have been described associated with an increased cumulative risk of approximately 13% for both EC and CRC (ten Broeke *et al.*, 2018), but the phenotypic penetrance of such variants appears relatively low in comparison to defects in the other MMR genes. While the precise reason for this is unknown, several hypotheses have been proposed. These include possible differences in the extent to which each protein contributes to MMR (Johnson *et al.*, 2010), or differences in the number of functions performed by each repair protein (Prolla *et al.*, 1998), as well as the potential for partial compensation by MLH3 (Chen *et al.*, 2005). A separate issue regards the relatively low detection rate of *PMS2* variants, with comparatively low numbers of heterozygous individuals identified. This is in conflict with the findings

of previous research concerning CMMRD, for which the biallelic inheritance of pathogenic *PMS2* variants accounts for a majority (58%) of cases (Wimmer *et al.*, 2014).

Analysis of *PMS2* is fundamentally complicated by the presence of 13 pseudogenes, situated on the same chromosome, which effectively contaminate the sequencing of the true gene sequence and reduce the sensitivity of mutation detection (Clendenning *et al.*, 2006). The largest of these, an approximate 100kb inverted duplication termed *PMS2CL*, shares ~98% homology with exons 9 and 11-15 of *PMS2*, while the remaining 12 each possess >90% identity to sequences in the 5' end of the gene. To circumvent sequencing issues, various techniques have been developed. The sequencing of cDNA from lymphocyte cultures, puromycin treated to inhibit the nonsense mediated decay of mRNA, has been shown to differentiate between *PMS2* and its pseudogenes, as well as detect splice variants and 'hybrid alleles', the latter of which may be responsible for ~10% of all *PMS2* alles in the general population (Etzler *et al.*, 2008). Alternatively, labour-intensive approaches using MLPA have been described allowing for the identification and localisation of copy-number variants (CNVs) (Herman *et al.*, 2018). Finally, Long-Range PCR, employing primers designed for regions with no homology to pseudogenes, or that target sites of deviation from pseudogene sequences, can be used for the specific amplification of the definite coding sequence (Clendenning *et al.*, 2006). Both approaches are not without limitation, primarily a lack of scalability, but the latter, with refinement, may provide a useful macro-sequencing option and the initial phase in improved *PMS2* testing.

An option for the further analysis of the amplification products of this Long-Range PCR involves MIP technology, previously described in the context of the Newcastle MSI Assay. With MIPs designed by the MIPgen computer programme according to critical performance parameters, gene tiling of exon regions is feasible to reveal positions of sequence variation. Moreover, combining Long-Range PCR with multiple MIPs targeting intronic polymorphisms may allow gene deletions to be precisely defined. Ultimately, this described screening pipeline would permit the detection of low-level variants with lesser penetrance from larger populations, as well as the improved validation of variants. Furthermore, this would increase the feasibility to assess broader, unselected populations for the frequency of compromised *PMS2*.

1.4. Summary and Aims

The chemopreventive potential of aspirin is clear, but adoption is limited due to the risk of adverse effects, especially in older people. One group for whom the benefits of aspirin prophylaxis exceed any potential negative side effects is those with Lynch syndrome. The anticancer utility of aspirin in this population has been demonstrated in the CAPP2 trial, and carriers should be advised to take aspirin to reduce their cancer risk, resulting in a clinical need to detect relevant mutation carriers.

Lynch syndrome is defined by mismatch repair deficiency, defects in the cellular repair mechanism that present with a molecular phenotype of microsatellite instability. Approaches exist for the identification of Lynch individuals, either assessing the presence of repair proteins or phenotypic evidence of their loss, with guidance evolving to advocate their use. However, several limitations compromise these clinical guidelines, including the restricted tumour spectrum analysed, while current recommendations fail to be implemented consistently and universally, possibly due, in part, to failings in the existing assays. Difficulties in the detection of mutations in *PMS2*, an MMR gene in which sequencing and variant detection is complicated by the presence of pseudogenes, epitomise such inadequacies, and highlight the need for improved practices to recognise those with LS, for whom therapeutic options like aspirin administration may be practicable.

The principle aim of this study is to improve the techniques for detecting MMRd and LS individuals, identifying potential candidates for prophylactic aspirin therapy. This can be further divided into the following series of objectives:

1 – Explore the utility of high-throughput MSI testing in extra-colonic tumours. This will involve expanding the use of the Newcastle MSI Assay to various cancer types, from which the ability to detect MSI will be determined, and, where possible, the frequency of such will subsequently be determined.

2 – Develop and assess a specific sequencing-based assay for the screening of PMS2. The calling of variants in this gene, as well as the detection of these in broader populations, would be enhanced by this.

Chapter 2. Materials and Methods

2.1. Ethical Approval for Research and Sample Availability

All analyses of LS gene carrier DNA were performed within the ethical framework of the CaPP3 clinical trial (ISRCTN16261285). This study, examining the optimal dose of prophylactic aspirin for chemoprevention in the context of Lynch, closed to recruitment in 2018, and will complete its five-year intervention phase in 2023. 1869 individuals with confirmed LS were randomised to 600mg, 300mg or 100mg daily in divided doses for two years blind, and a further open phase for three years. Blood samples for DNA extraction and serum storage were obtained from all participants at recruitment, and again after two and five years, with patient consent for use in research by Newcastle University. Such samples were coded upon enrolment, allowing for impartial analysis and deductions, but patient details are available upon request from the clinical team of CaPP3. Local approval for the processing and MSI analysis of samples was also in place following an ethical review by the Newcastle University Research Ethics Committee (REC Reference: 13/LO/1514).

Samples received from the Ohio State University Comprehensive Cancer Center and the University of Manchester were collected and analysed with patient consent and following local ethical approval.

2.2. DNA Samples

2.2.1. Samples for the Analysis of MSI in Extracolonic Tumours

81 independent extracolonic tumour samples, covering a range of LS and non-LS tumour types, were retrieved from the CaPP tissue resource in the form of formalin-fixed, paraffin-embedded (FFPE) material, as well as 43 Lynch CRC samples. While anonymised, limited patient details, including information of confirmed MMR genetic defects, were available for these upon request.

200 EC tumour DNA samples, divided in to two cohorts of 100 samples depending on purpose, were provided by Dr. Heather Hampel and Paul Goodfellow,

PhD of the Ohio State University Comprehensive Cancer Center. The first of these, designated for classifier training, consisted of unblinded material from patients prospectively registered in the Ohio Colorectal Cancer Prevention Initiative (OCCPI) (Hampel *et al.*, 2021), while the second, a designated validation cohort to which I was initially blinded, comprised samples of the Ohio Prevention and Treatment of Endometrial Cancer (OPTEC) Initiative (M. Levine *et al.*, 2021). For all, results of prior IHC and MSI analyses were ultimately available, as well as the conclusions of *MLH1* promoter methylation testing and germline genetic testing where such was undertaken.

In addition, FFPE material was provided for 191 ECs, previously involved in the Proportion of Endometrial Tumours Associated with Lynch syndrome (PETALS) prospective study (Ryan *et al.*, 2020), by Dr. Emma Crosbie of the University of Manchester. Delivered in two distinct cohorts, 95 of these constituted an assigned training group, while the remaining 96 realised the corresponding validation cohort, subject to preliminary blinding. Existing IHC and Promega MSI results for all were also eventually available, as well as other potentially useful information such as reports on tumour stage, grade and histopathology.

Details of all samples, relating gene defects and the results of previous analyses where available, are given in Appendix A.

2.2.2. Samples for the Development of a Sequencing-Based Assay for the Analysis of PMS2

138 DNA samples from the CaPP tissue resource, recovered during patient consultation at trial entry and randomisation (Year 0), were assessed during assay development and validation. These were distributed between MMR genes associated with Lynch syndrome, with 35 *MLH1*, 37 *MSH2* and 66 *PMS2* samples analysed.

For all material a dilution of 100ng/µl was initially produced, allowing for a consistency and ease in reaction preparation.

2.3. FFPE Tumour Tissue Curl Cutting

FFPE tumour blocks were cut on a microtome (Thermo Scientific HM325 Microm) to a thickness of 10µm, with six curls ultimately removed. As no histopathologist was available, tumour cell content analysis was not performed on this material before DNA extraction, with 1-3 untriaged tissue curls subsequently used for convenience.

2.4. DNA Extraction, Quantification and Dilution

The extraction of genomic DNA from FFPE tissue material was achieved using the QIAGEN (Hilden, Germany) GeneRead DNA FFPE kit in accordance with the manufacturer's protocol. Resulting concentrations for these templates, as well as associated amplicons, were quantified using the QuBit 2.0 Fluorometer and QuBit dsDNA BR kit (Invitrogen), following the manufacturer's guidelines, providing a ng/µl measure of DNA present.

For DNA dilution, ng/µl density values were converted to nanomolar concentration measures using the following equation, in which 660g/mol/bp represents the average molar mass of a single DNA base and N denotes the number of base pairs in the subject DNA molecule:

$$\text{Concentration(nM)} = \frac{\text{Density } \left(\frac{\text{ng}}{\mu\text{l}}\right) \text{x } 10^6}{660\text{g/mol/bp x N}}$$

Template and amplicon DNAs were ultimately diluted using 10mM Tris-Cl at pH8.5 buffer.

2.5. Single Molecule Molecular Inversion Probe Amplification Protocol

2.5.1. Marker Loci of the Newcastle smMIP-Based MSI Assay

The original marker panel of the Newcastle MSI Assay consists of 24 short, monomorphic mononucleotide repeats (7-12bp in length), selected from markers previously described (Redford *et al.*, 2018). Each is associated with a neighbouring single nucleotide polymorphism (SNP), facilitating the detection of allelic bias of deletions in heterozygous cases.

Markers of the Version 2 assay iteration were derived in a study of whole genome sequencing data, with characteristics indicative of improved MMRd detection. 62 markers were ultimately included in analysis with these generally longer than those of the original set (11-15bp).

All single-molecular Molecular Inversion Probes (smMIPs) targeting these markers, detailed in Appendix B, were synthesised by, and purchased from, Metabion GmbH (Planegg, Germany).

2.5.2. Probe Phosphorylation and Creation of smMIP Primer Pools

Following pooling and dilution to a final pool concentration of 2µM, 50µl of combined smMIPs was phosphorylated in a 100µl reaction volume with 1X T4 DNA Ligase buffer and 10U T4 Polynucleotide Kinase. Incubation at 37°C for 45 minutes then 80°C for 20 minutes produced a 1µM 5'-phosphorylated MIP pool which was subsequently diluted, using 10mM Tris HCl buffer, to give each individual smMIP at a concentration of 0.1nM (0.1fmol/µl). Probe pools were assessed prior to use through the amplification of control samples, to confirm MIP inclusion, successful phosphorylation and generation of sufficient amplicon product.

2.5.3. Target Capture and Amplification

Markers for MSI were amplified in multiplex from approximately 100ng of template DNA, observing an established smMIP-based protocol (Figure 2.1.) (Hiatt *et al.*, 2013). In a reaction volume of 10µl, smMIPs (each at a concentration of 0.1fmol) anneal to template DNA using 1x AmpLigase Reaction Buffer (Lucigen), and experience incubation at 98°C for 3 minutes, 85°C for 30 minutes, 60°C for 60 minutes, and 56°C for 120. Subsequent 'gap-fill' and ligation incorporates the target marker sequence

within circularised smMIPs using 1x AmpLigase Reaction Buffer (Lucigen), 5U AmpLigase DNA Ligase (Lucigen), 3.2U Herculase II Fusion DNA Polymerase (Agilent), 300pmol dNTPs (Agilent). This involves the a 10µl addition to each reaction, giving a total reaction volume of 20µl, and incubation at 56°C for 60 minutes and 72°C for 20 minutes. 1x AmpLigase Reaction, Buffer (Lucigen), 20U Exonuclease I (NEB) and 100U Exonuclease III (NEB) are then introduced into each reaction in a 3µl volume for the digestion of non-circularised smMIPs and template DNA, with this requiring incubation at 37°C for 60 minutes and 95°C for 2 minutes. Finally, for the amplification of the target sequence, 10µl of target capture reaction is combined with 1x Herculase II Reaction Buffer (Agilent), 1.25U Herculase II Fusion DNA Polymerase (Agilent), 6.25nmol dNTPs (Agilent), 6.25pmol MIP amplification forward primer and 6.25pmol of MIP amplification reverse primer, with the remaining target-capture volume stored at -20°C. This reaction involves initial incubation at 98°C for 2 minutes, followed by 30 cycles of 98°C for 15 seconds, 60°C for 30 seconds and 72°C for 30 seconds, before a closing 2 minutes at 72°C. Furthermore, the use of distinct smMIP amplification reverse primers for each individual sample in a given sequencing run, each incorporating a unique sample index sequence, allows for the eventual de-multiplexing of sequence reads.

Reaction products were, amplicons of 240-270bp in length, were analysed using 2.5% Agarose gel electrophoresis at 100mV for 60 minutes, or QIAxcel (QIAGEN) using programme AL420.

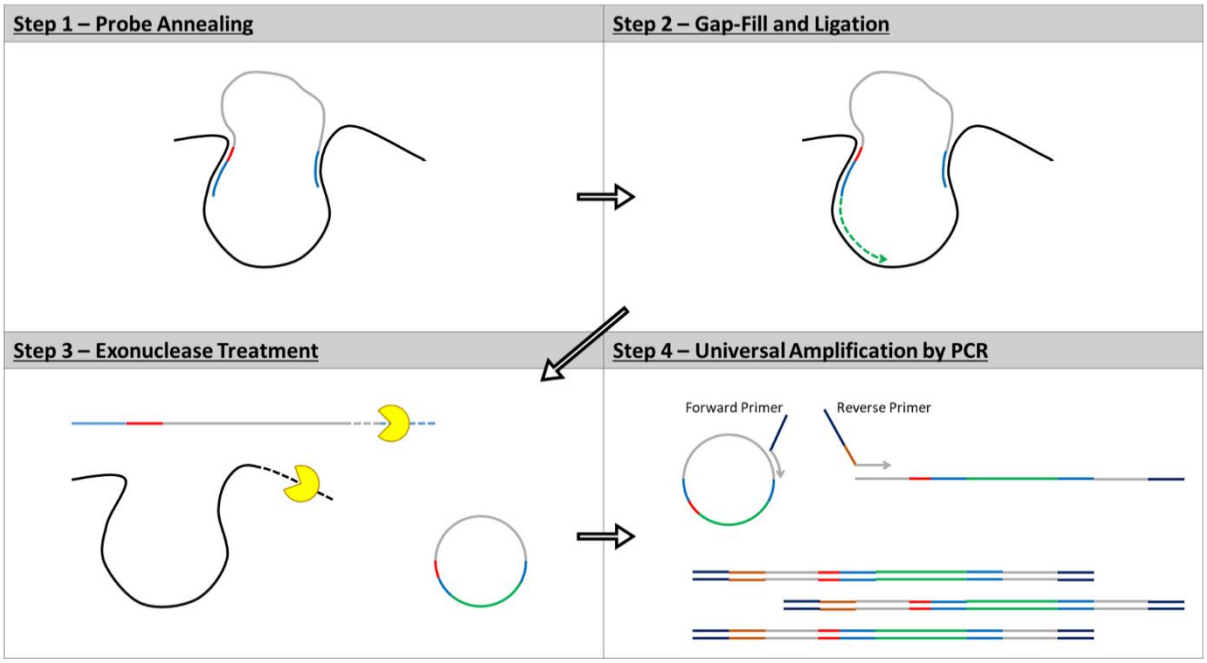

Figure 2.1. Schematic of the standard MIP protocol. Initial probe annealing occurs between the MIP targeting arms and complementary sequences in the region of interest. Extension by DNA polymerase between the two arms of the probe follows, using the sequence of interest as a template, before eventual (nick-sealing) ligation by DNA ligase to produce a circular molecule. Exonuclease treatment removes remaining linear DNA (including template and uncircularised probes), leaving amplicons that provide the template for subsequent sequence determination in this multiplex micro-sequencing approach.

2.5.4. Library Preparation and smMIP Amplicon Sequencing

In advance of sample preparation, every sequencing run was planned pursuant to a desired mean number of reads per marker per amplicon, the number of markers amplified per sample using the previously-explained MIP protocol, and the total sample number. This required the following equation:

$$\text{Reads/Marker/Sample} = 1.33 \times \frac{\text{Read Capacity}}{\text{Sample Number} \times \text{Marker Number}}$$

To account for the incidence of non-specific sequencing reads, a 1.33 correction factor was incorporated relating to the use of template DNA derived from FFPE material (such as that of the Manchester EC samples): an alternative value of 2.0 was used for templates from tumour tissue extractions such as those supplied by Ohio State University. Moreover, read capacity denotes the expected total reads resulting from the given MiSeq sequencing kit used. With a capacity of 25 million, Version 3 kits were selected for all EC processing given the number of samples analysed and the intended read depth of 2000 reads per marker.

Amplicons from smMIP amplification were purified using Agencourt AMPure XP Beads (Beckman Coulter) and their associated protocol, with this followed by dilution in 10mM Tris pH 8.5 and the pooling of equal volumes of 4nM amplicons to ultimately achieve a 4nM DNA sequencing library. These libraries were sequenced using the MiSeq platform (Illumina) in accordance with the manufacturer's protocol, employing GenerateFastq workflow, paired end sequencing and smMIP custom sequencing primers as previously established (Hiatt *et al.*, 2013). Finally, the unprocessed reads, contained in FastQ files, as well as sequencing run metrics, such as the proportion of base calls with quality >30, were extracted from basespace.illumina.com for subsequent interpretation and presentation.

2.6. Sequencing Read Analysis

2.6.1. Compilation of Marker Results

Using Burrows-Wheeler Aligner (BWA) v0.7.17 (Li and Durbin, 2010), unprocessed reads, extracted from FastQ files, were first aligned to the hg19 reference genome to generate. sam files. Marker loci were subsequently defined and, from the aforementioned .sam files, the microsatellite lengths and SNPs returned in both orientations for each marker were counted using custom R scripts, and recorded in marker result tables. Such summary tables were produced using custom R scripts developed by Dr. Mauro Santibanez-Koref (Institute of Genetic Medicine, Newcastle University), and provided the basis for several further analyses, including sample classification.

2.6.2. Read Counting and MSI Classification

As detailed by Redford et al. (2018), sample classification involved the assessment of sequencing reads from microsatellite markers for both deletions and the allelic bias of any such change. From these, a ratio of the posterior probabilities that any findings were a result of an MSI-high or MSS phenotype, a Bayes factor, were determined. Such could be represented as a score using the following equation:

$$score = log_{10} \frac{P(MMRd|O)}{P(MMRp|O)}$$

In the first instance, classifier parameters were defined using a training cohort of CRCs of known MSI status (Gallon *et al.*, 2020), before this was adjusted in a tumour-specific manner as required (described in Section 4.4.6.). Training and implementation of the MSI classifier again employed custom R scripts written by Dr. Santibanez-Koref, with Marker Result tables provided as input. Samples ultimately returning a score <0 were adjudged MSS, whereas an MSI-H classification was assigned to those with score >0.

2.6.3. Statistical Analysis and Data Presentation

All statistical analyses were performed using base R, version 1.0.143, unless specified otherwise. Scoring distributions were modelled using the ggpmisc package, while ggplot was employed to graphically plot data.

2.7. *PMS2* Assay Development

2.7.1. Long-Range PCR Amplification of PMS2

For the specific amplification of the sequence from the *PMS2* gene locus, long-range PCR using three published primer pairs was employed (Clendenning *et al.*, 2006), initially as per the diagnostic services protocol provided from Oslo Universitetssykehus. In summary, this approach involves combining 100ng of high-quality DNA template with 2.5µl of 10XLA PCR Buffer II (Clontech), 10nmol dNTPs (Clontech), 1.25U TaKaRa LA Taq (Clontech) and 5pmol of each primer of a given pair in a reaction volume of 25µl. Through the following PCR program, three amplicons, spanning all 15 exons of the gene, are generated by three separate reactions involving the different primer pairs in singleplex: incubation at 94°C for 1 minute, followed by 35 cycles of 94°C for 15 seconds, 65°C for 30 seconds and 68°C for 15 minutes, before a final 10 minutes at 72°C. Eventual reaction products, amplicons of 10-20kbp in length, were then reviewed using 0.6% Agarose gel electrophoresis at 75mV for 180 minutes.

Following purification (in accordance with the Agencourt™ AMPure™ XP Reagent purification protocol), the three amplicons for a given sample were pooled in a 1:1:1 ratio, calculated based on their relative sizes and concentration readings, for use in subsequent MIP amplification reactions.

2.7.2 Design of Molecular Inversion Probes and MIP Pool Generation

To produce optimised smMIP sequences for the complete coverage of the *PMS2* exons, the computer programme MIPGen (E. A. Boyle *et al.*, 2014) was employed with the following user-defined parameters: tag size (molecular barcode length) =4,4, minimum capture size=120, maximum capture size=160, and minimum logistics score=0.7. In instances of initial failed coverage, the manual selection of MIP sequences was performed from a default output file of all possible smMIPs using custom R scripts: a reduced minimum logistic score (≥0.5) and increased maximum capture size (≤180) were considered for this.

By equivalent means, a selection of additional MIPs was also produced to target common intronic SNPs of the *PMS2* sequence and indicate allelic loss. For each long-

range amplicon a sufficient number of MIPs are required to give confidence in power calculations for the probability of homozygosity from the implicated SNPs: re-designs were performed until such quantities were realised.

All 142 smMIPs ultimately selected (Appendix C), 42 covering all exons in a single-tiling approach and 100 capturing intronic SNPs, were synthesised by, and purchased from, Metabion GmbH (Planegg, Germany).

2.7.3. Single-Molecule Molecular Inversion Probe Amplification and Preparation for Sequencing

From approximately 50ng of template DNA, a pool of purified long-range PCR amplicons, the 142 aforementioned PMS2 positions were amplified in multiplex, following an appropriately-adapted version of the established smMIP-based protocol(Hiatt *et al.*, 2013). The reactions and thermocycler conditions of this process were as detailed previously for target capture and amplification in the MSI assay.

Reaction products, amplicons of 240-280bp in length, were analysed using 2.5% Agarose gel electrophoresis at 100mV for 60 minutes, or QIAxcel (QIAGEN) using programme AL420.

2.7.4 Sequencing Data Analysis and Variant Calling

In the initial processing and interpretation of sequence data, FastQ files were trimmed, with reference a file designating the target positions of the PMS2 Assay's MIPs, so as to remove targeting arm sequences from reads. The resulting files were subsequently analysed using BWA v0.7.17 (Li and Durbin, 2010), with the unprocessed reads contained within these aligned aligned to the hg19 reference genome to generate. sam files. With the positions of targets withing the PMS2 gene subsequently defined, the sequence data of each MIP, and therefore the variants contained within such, were subsequently reviewed in both orientations and recorded for subsequent analysis.

The calling of variants within sequence reads was achieved by custom scripts using the Genome Analysis Toolkit (GATK) module for the identification of recurrent deviations from the nominated hg19 reference genome. In this process, a 'Conservative' PCR indel model for variant detection was implemented within a 'Discovery' genotyping mode to ensure only true mutations were called as such, and not isolated PCR artefacts: a minimum threshold of 50 reads for a given variant was

also established for this purpose. Identified variants were compiled to a results table, before these were subsequently assessed by the Ensembl Variant Effect Predictor (VEP) pathogenicity predictor. In this process, requiring the installation of the VEP module, variants were analysed to predict their physiological effect, as well as compared against databases to determine if they had been previously reported. The results of this inquiry were integrate with those of variant calling to produce a complete table of variants analysed within the PCR products of a given sample.

All subsequent sequencing data analysis and representation was performed using base R, version 1.0.143, unless specified otherwise, while ggplot was employed to graphically plot data.

Chapter 3. MSI Detection in Extracolonic Tumours

3.1. Introduction

3.1.1. LS Detection in CRC

With an estimated prevalence in the general population of one in 300 (Win *et al.*, 2017), LS is a common hereditary cancer predisposition syndrome, responsible for approximately 3% of all CRC cases (Moreira *et al.*, 2012). LS gene carriers, once known to clinical services, may benefit from specific treatment practices including enhanced tumour surveillance, prophylactic surgery and/or daily aspirin use (Vasen *et al.*, 2013), while genetic counselling and the cascade testing of relatives to identify others affected may also be provided. However, the condition remains greatly under-reported, with diagnoses obstructed by various practical, social and financial factors (Refer to Section 1.3.5.): it follows that improving detection will therefore have clinical benefits.

In LS colonic tumourigenesis, the development of MMRd is understood to be an early event that precedes adenoma formation (Sekine *et al.*, 2017). A recent analysis of IHC data from LS mutation carriers substantiates this theory, with the presence of repair deficient foci reported in normal mucosa, adjacent to dysplastic adenoma tissue, in 76.7% of instances (Ahadova *et al.*, 2018). Resulting from the somatic inactivation of a second MMR allele, MMRd presents with the molecular phenotype of exceptionally high MSI, variations in the length of tandem nucleotide repeats throughout the genome at a higher-than-normal rate. The ability to screen for LS gene carriers in CRC using MSI has previously been shown (Table 3.1.), with detection rates in excess of 93%, and is clinically advantageous, with this demonstrated cost-effective primarily for new diagnoses before the age of 50 years, but up to the age of 70 years. Furthermore, screening may indicate patients for whom immunotherapy may be practicable, with this treatment shown to improve one-year overall survival rates following MSI-H tumour incidence up to 85% (Motta *et al.*, 2021).

The benefits of MMRd detection are reflected in the current recommendations of several authorities, which advocate the testing of all individuals diagnosed with CRC to identify potential instances of germline repair deficiency. However, at the outset of

my work such clinical advice was limited to this tumour type, despite the proposals of several reviews encouraging an expansion of testing to other LS-related malignancies, including sebaceous neoplasms (Everett *et al.*, 2014), urothelial (Ju *et al.*, 2018) and ovarian cancers (Takeda *et al.*, 2018), to further increase detection rates.

3.1.2. The Potential of LS Screening in Extracolonic Cancers

Other forms of cancer generally considered associated with LS, where an increased risk of occurrence in LS carriers is now established, include those of the endometrium, upper gastrointestinal tract (stomach, pancreas and small bowel), ovaries, sebaceous glands and urothelial tract (bladder, ureter and kidney), while the inclusion of breast and prostate cancers within the LS spectrum is contentious (Vasen *et al.*, 2013). Between these tumour types, estimates from the literature of the frequency of cases caused by LS varies, as shown in Table 3.1., with 33.3% of sebaceous adenomas, but only 0.4% of ovarian cancers, attributed to the condition. What is more, the cumulative incidence of cancer by the age of 75 years also differs, reported as high as 56.7% in EC (Møller *et al.*, 2017). No official guidance for MMRd testing of these cancers existed when this work commenced, with this potentially representing a missed opportunity for clinical intervention. Molecular testing, for the identification of repair deficiency, provides the foremost option to address this potential failing, avoiding the false-positive results that may accompany the use of family history in diagnosis, and with a demonstrated utility in CRC.

MSI detection has previously been used to analyse extracolonic cancers, but the frequency of the phenotype, although sufficiently high for observation, varies by tumour type (Table 3.2.). A further difference in the phenotype may also exist with regard to the distribution of allele size, with shorter mean shifts of three nucleotides reported in Lynch ECs compared to the six of CRCs when assessed by a multiplex PCR approach using five MNR markers (Wang *et al.*, 2017). Such findings regarding instability in extracolonic tumours are however questionable, realised in studies limited by the use of subjective, outdated approaches, or restrictive sample sizes. Nevertheless, given the aforementioned incidence of LS, and the shared fundamental biology of repair deficiency responsible for predisposition, it is conceivable that current screening guidance for LS could be extended to additional tumours of the Lynch

spectrum, and exploit the molecular phenotype of exceptionally high MSI with which, although inconsistent, they have been shown to present.

Table 3.1. Estimated frequency of LS in unselected patients across various tumour types. 'Unselected' describes the absence of patient selection by clinical features, such as family history, although in some studies germline MMR testing followed tumour MMRd testing.

Tumour Type	Frequency	Method	Reference
Colorectal Cancer	**426/14075 (3.0%, 95% CI=2.7 – 3.3%)**	**Total**	
	23/1066 (2.2%)	Germline MMR gene testing of patients with MMRd tumours	Hampel et al., 2005
	18/500 (3.6%)	Germline MMR gene testing of patients with MMRd tumours	Hampel et al., 2008
	312/10206 (3.1%	Germline MMR gene testing of 26% of patients	Moreira et al., 2012
	33/1058 (3.1%)	Germline MMR gene testing of all patients	Yurgelun et al., 2017
	12/419 (3.4%)	Germline MMR gene testing of all patients	Hampel et al., 2018
	28/826 (3.4%)	Germline MMR gene testing of all patients	Latham et al., 2019
Endometrial Cancer	**160/6518 (2.5%, 95% CI=2.1 – 2.9%)**	**Total**	
	145/5882 (2.5%)	Meta-analysis of germline MMR gene testing	Ryan et al., 2019
	6/111 (5.4%)	Germline MMR gene testing of all patients	Chao et al., 2019)
	9/525 (1.7%)	Germline MMR gene testing of all patients	Latham et al., 2019
Gastric Cancer	**3/308 (1.0%, 95% CI=0.2–2.8 %)**	**Total**	
	1/97 (1.0%)	Germline MMR gene testing of patients with MMRd tumours	Christakis et al., 2019
	2/211 (0.9%)	Germline MMR gene testing of all patients	Latham et al., 2019
Ovarian Cancer	**11/2892 (0.4%, 95% CI=0.4 – 0.7%)**	**Total**	
	9-64/1893 (0.5 – 3.4%)	Germline *MLH1*, *MSH2*, and *MSH6* gene testing of all patients.	Pal et al., 2012
	4/656 (0.6%)	Germline *MLH1*, *MSH2*, and *MSH6* gene testing of all patients.	Akbari et al., 2017
	0/343 (0.0%)	Germline MMR gene testing of all patients	Latham et al., 2019

Pancreatic Cancer	**27/2395 (1.1%, 95% CI=0.7 – 1.6%)**	**Total**	
	4/290 (1.4%)	Germline MMR gene testing of all patients	Grant et al., 2015
	5/249 (2.0%)	Germline MMR gene testing of all patients	Connor et al., 2017
	9/833 (1.1%)	Germline MMR gene testing of all patients	Hu et al., 2018
	1/199 (0.5%)	Germline MMR gene testing of patients with MMRd tumours	Christakis et al., 2019
	8/824 (1.0%)	Germline MMR gene testing of all patients	Latham et al., 2019
Small Bowel Cancer	**14/281 (5.0%, 95% CI=2.8 – 8.2%)**	**Total**	
	8/195 (4.1%)	Germline MMR gene testing of patients with MMRd tumours	Jun et al., 2017
	4/29 (13.8%)	Germline MMR gene testing of patients with MMRd tumours	Christakis et al., 2019
	2/57 (3.5%)	Germline MMR gene testing of all patients	Latham et al., 2019
Urothelial Cancer	**28/1003 (2.8%, 95% CI=1.9 – 4.0%)**	**Total**	
	13/551 (2.4%)	Germline MMR gene testing of all patients	Latham et al., 2019
Upper Tract Urothelial	6/194 (3.1%)	Germline MMR gene testing of patients with MMRd tumours	Harper et al., 2017
Upper Tract Urothelial	7/115 (6.1%)	Germline MMR gene testing of patients with MMRd tumours	Metcalfe et al., 2018
Upper Tract Urothelial	2/143 (1.4%)	Germline MMR gene testing of 28.6% of patients with MMRd tumours	Urakami et al., 2018
Sebaceous Adenoma	**87/261 (33.3%, 95% CI=27.6 – 39.4%)**	**Total**	
	11/24 (45.8%)	Germline MMR gene testing of all patients	Dandapani et al., 2011
	25/86 (29.1%)	Germline MMR gene testing of 67.4% of patients	Everett et al., 2014
	40/89 (44.9%)	Germline MMR gene testing of all patients	Roberts et al., 2014
	11/62 (17.7%)	Germline MMR gene testing of 18.9% of patients / 14.5% known genetic diagnoses	Schon et al., 2018
Breast Cancer	**11/3011 (0.4%, 95% CI=0.2 – 0.7%)**	**Total**	
	4/640 (0.6%)	Germline MMR gene testing of all patients	Davies et al., 2017

	7/2371 (0.3%)	Germline MMR gene testing of all patients	Latham et al., 2019
Prostate Cancer	**65/4655 (1.4%, 95% CI=1.1 – 1.8%)**	**Total**	
	7/1048 (0.7%)	Germline MMR gene testing of all patients	Latham et al., 2019
	58/3607 (1.6%)	Germline MMR gene testing of all patients	Nicolosi et al., 2019

Table 3.2. Estimated frequency of MMRd in different tumour types from LS gene carriers. Studies are included where MMRd was determined by an MSI-testing approach, and were the LS diagnosis was made independent of tumour MMR status.

Tumour Type	Frequency	Method	Reference
Colorectal Cancer	**128/135 (94.8%, 95% CI=89.6 – 97.9%)**	**Total**	
	46/48 (95.8%)	MSI by FLA (Bethesda panel)	Gylling et al., 2008
	28/29 (96.6%)	MSI analysis by unspecified method	Yurgelun et al., 2017
	54/58 (93.1%)	MSI by FLA of 5 MNRs (Promega) and NGS (mSINGS)	Hampel et al., 2018
Endometrial Cancer	**79/98 (80.6%, 95% CI=71.4 – 87.9%)**	**Total**	
	8/8 (100.0%)	MSI by FLA (adapted Bethesda panel)	Lu et al., 2007
	38/60 (63.3%)	MSI by FLA (Bethesda panel)	Gylling et al., 2008
	19/21 (90.5%)	MSI (unspecified)	Ring et al., 2016
	5/12 (41.7%)	MSI by FLA (adapted Bethesda panel)	Rubio et al., 2016)
	4/4 (100.0%)	MSI by FLA of 5 MNRs (Sinomdgene Co. Ltd., Beijing China)	Chao et al., 2019
Gastric Cancer	**35/39 (89.6%, 95% CI=75.8 –97.1 %)**	**Total**	
	7-15/15 (46.7–100.0%)	MSI by FLA of 7 diNRs	Aarnio et al., 1997
	13/13 (100.0%)	MSI by FLA (Bethesda panel)	Gylling et al., 2008
	4/4 (100.0%)	MSI by FLA (Promega or Bethesda panel)	Fornasarig et al., 2018

Ovarian Cancer	**23/24 (95.8%, 95% CI=78.9 – 99.9%)**	**Total**	
	19/20 (95.0%)	MSI by FLA of 2 MNRs	Niskakoski et al., 2013
	4/4 (100.0%)	MSI by FLA (Bethesda panel)	Akbari et al., 2017
Pancreatic Cancer	**13/17 (76.5%, 95% CI=50.1 – 93.2%)**	**Total**	
	3/3 (100.0%)	MSI by FLA (Bethesda panel)	Yamamoto et al., 2001
	3/5 (60.0%)	Mutational signatures confirmed by MSI by FLA (Promega)	Connor et al., 2017
	7/9 (77.8%)	MSI by NGS (MSIsensor), MSI by FLA of 5 MNRs	Hu et al., 2018
Small Bowel Cancer	**21/21 (100.0%, 95% CI=83.9 – 100.0%)**	**Total**	
	21/21 (100.0%)	MSI by FLA (adapted Bethesda panel)	Schulmann et al., 2005
Urothelial Cancer	**15/21 (71.4%, 95% CI=47.8 – 88.7%)**	**Total**	
Bladder Urothelial	3/5 (60.0%)	MSI by FLA (Bethesda panel)	Gylling et al., 2008
Bladder Urothelial	6/7 (85.7%)	MSI by FLA (adapted Bethesda panel)	van der Post et al., 2010
Upper Tract Urothelial	6/9 (66.7%)	MSI by FLA (Bethesda panel)	Gylling et al., 2008
Breast Cancer	**8/23 (34.8%, 95% CI=16.4 – 57.3%)**	**Total**	
	8/23 (34.8%)	MSI by FLA (Bethesda panel)	Lotsari et al., 2012
Prostate Cancer	**9/27 (33.3%, 95% CI=16.5 – 54.0%)**	**Total**	
	2/16 (12.5%)	MSI by FLA (Promega)	Dominguez-Valentin et al., 2016
	7/11 (63.6%)	MSI by FLA (Promega)	Antonarakis et al., 2019

3.1.3. Previous Analysis of Extracolonic Cancers by the Newcastle MSI Assay

Especially in tumour types with an expected lower frequency of LS and associated MSI, a calculation of the cost-to-benefit ratio may prohibit the wider deployment of a given assay for screening (if the expense exceeds any benefits so as to render this not practicable or financially justified. Therefore, to facilitate any extension to screening recommendations, an alternative, more sensitive technique to the time-consuming, subjective approaches currently in use may be advantageous, allowing for an increased capacity of analysis, and the potential resolution of inconsistent MSI detection across tumour types. Such may be afforded by the comparatively cheap, automatable Newcastle MSI Assay. Considering its potential advantages, a previous MRes project undertaken by Shaun Prior in 2018 explored the utility of this technique for the testing of LS extracolonic tumours, with a total of 74 samples analysed across a range of cancer types. In this study, 40/49 (82%) of LS-associated cancers specifically were classified MSI-H by the Newcastle Assay, with a consistency, and no significant difference, realised between tumour type (Table 3.3.). However, these findings, like those of earlier studies, were ultimately limited, with a relatively small sample size, and the absence of comparator sporadic and LS CRC populations: before justified conclusions on the practicality of using this approach in screening additional analysis was therefore necessary.

Endometrial	Ovarian	Sebaceous Adenoma	Urothelial	Breast	Other Skin
77	83	100	100	63	59
(27 / 35)	(5 / 6)	(5 / 5)	(3 / 3)	(5 / 8)	(10 / 17)

Table 3.3. Percentage MSI-H across various tumour types, all developing in the context of LS, as determined in analysis by the Newcastle MSI Assay as part of a previous MRes project.

3.2. Aims

In the following project, I aimed to investigate the potential for MSI detection in extracolonic tumours using a high-throughput, sequencing-based assay, continuing the work of a former MRes student outlined above. For this, there were several objectives:

- Determine the utility of MSI to detect LS tumours across an extensive cohort of LS extracolonic tumours.
  To be addressed by the analysis of additional extracolonic tumours, arising since the aforementioned MRes study, as well as those previously analysed.

- Further investigate potential differences in MSI-H frequency observed between different tumour types (as suggested by the preliminary, MRes project data).
  To be addressed by the analysis of comparator LS CRC and sporadic EC cohorts for which a certain frequency of Lynch is expected.

- Investigate the reproducibility of MSI detection in LS samples by this approach.
  To be addressed by performing the repeat analyses of several samples previously included in the MRes project.

3.3. Methods

The CaPP3 tissue resource, representing a range of cancer types, was established by a trial investigating the optimal dose of chemopreventative aspirin therapy, with all tumours in it confirmed to have developed in the context of germline MMR defects. For my study, 88 tumours, consisting of both colonic and extracolonic malignancies, were analysed, giving a total of 121 independent samples. In addition, 41 unselected EC samples retrieved from the Newcastle Biobank were also assessed. All samples were provided in the form of FFPE material, from which curls were removed and DNA subsequently extracted, all without preliminary triaging for tumour cell content (Detailed in Sections 2.3. and 2.4.). The resulting DNA was analysed using the Newcastle MSI Assay with the original 24 MNR marker panel and a CRC-trained classifier (Detailed in Section 2.5), before products generated through this were sequenced to an average read depth of 5000X (Detailed in Detailed in Section 2.5.4.) and analysed by custom scripting pipelines (Detailed in Section 2.6.).

3.4. Results

3.4.1. MSI is Detected Across a Variety of Tumour Types from LS Patients Using the Newcastle MSI Assay

To identify additional samples for analysis, supplementing those previously studied, the CaPP3 database was initially queried for tumours recently arising within the cohort. 88 suitable samples, 48 exclusive to this work, were ultimately selected and retrieved to explore the potential for MSI detection in extracolonic cancers using the Newcastle MSI Assay. FFPE curls were removed from these, before DNA was extracted, amplified, sequenced and analysed to assess instability (Detailed in Section 3.3.). In this analysis, a majority of samples were classified MSI-High (MSI-H), with 26/31 (84%) of the LS-spectrum tumours specifically receiving this classification (Table 3.4.).

Colorectal	Endometrial	Ovarian	Urothelial	Breast	Other Skin
98	81	83	100	60	78
(42 / 43)	(17 / 21)	(5 / 6)	(4 / 4)	(3 / 5)	(7 / 9)

Table 3.4. Percentage MSI-H by tumour type from analysis of CaPP3 LS gene carriers by the Newcastle MSI Assay.

Across the different extracolonic tumour types individually, the percentage of MSI-H samples is relatively high, and generally consistent, especially between LS spectrum tumours for which no statistically significant difference is observed, However, except for urothelial samples, these proportions are all lower than that returned for the comparator CRC population, albeit with the only significant difference existing between those and the EC group (p=0.0013, 95% CI=58-95%).

3.4.2. MSI Detection in Extracolonic Tumours Using the Newcastle MSI Assay is Reproducible

To determine the reproducibility of MSI classification by the Newcastle MSI Assay, the repeat analysis of several tumours, first analysed as part of the aforementioned MRes project, was performed using independent DNA extractions of these samples. Comparison of 40 samples assessed twice revealed a strong positive correlation, represented by a corresponding R value of 0.89, and a 95% concordance in classification (Figure 3.1.). Discordance in classification was observed for only two samples between the analyses, one being from an endometrial cancer and the other a breast cancer. Both initially received MSI-H designations with scores of 15.35 and 6.75 respectively, but were assigned Microsatellite-Stable (MSS) classifications in subsequent processing, albeit in relatively close proximity, within seven points, of the classification threshold. Nevertheless, given the agreement between these two distinct analyses, it is reasonable and justified to propose the merging of data for a summative evaluation of MSI detection from all extracolonic tumours studied.

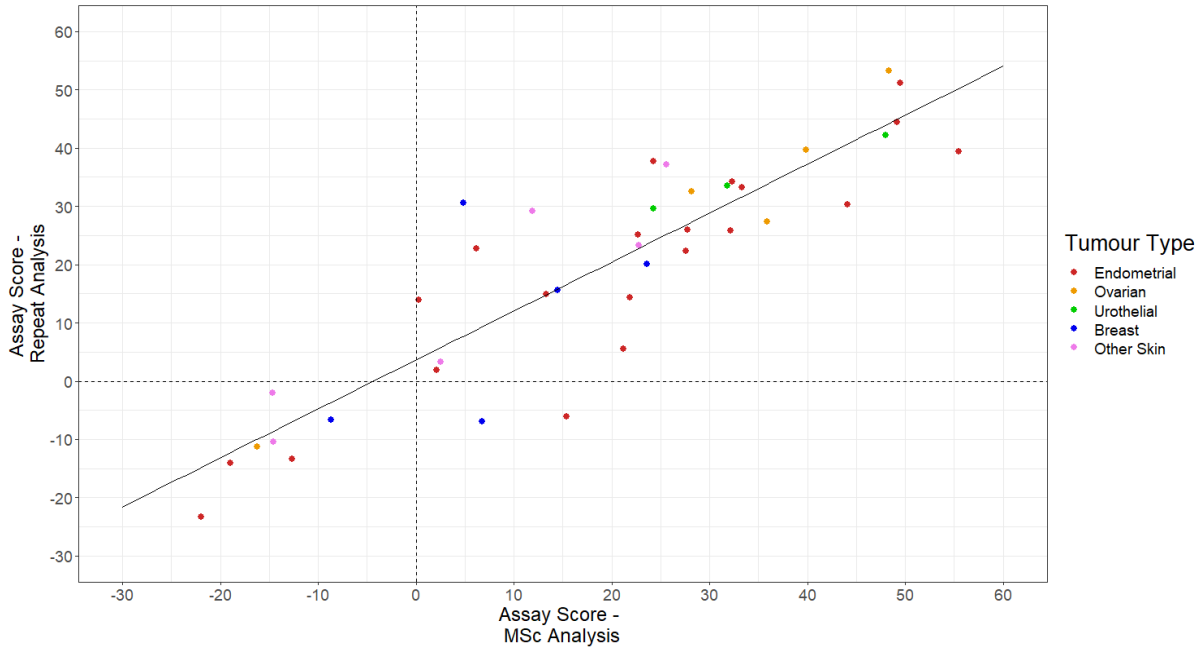

Figure 3.1. Comparison of scores returned from the repeat analysis of 40 extracolonic samples by the Newcastle MSI Assay. Samples represent a range of extracolonic malignancies, with 21 endometrial, five ovarian, three urothelial, five breast, and six non-sebaceous skin tumours interpreted.

3.4.3. MSI is not Ubiquitous in the Context of LS, with the Frequency Varying Significantly Between Tumour Types

The results above demonstrate an ability to detect MSI in extracolonic cancers using the Newcastle Assay, and evidence a broad consistency of this phenotype across extracolonic tumour types. However, they also indicate that, in the context of LS, the frequency of MSI is lower in several tumour types than that observed in CRCs, with a significant difference being returned for EC. MSI frequencies in these extracolonic cancers was investigated further by combining the data of both aforementioned analyses, with the distribution of scoring using the Newcastle MSI Assay assessed by tumour type in addition to the frequency of MSI-H classification.

With the amalgamated sequencing data for all 122 samples, a classification of MSI-H was assigned to the majority of cases (82%), including 80% of those considered of the Lynch spectrum (Table 3.5.). While the specific percentage MSI designations varies between LS tumour types, these remain relatively high, and generally comparable to the proportion returned in the accompanying CRC analysis.

Colorectal	Endometrial	Ovarian	Sebaceous Adenoma	Urothelial	Breast	Other Skin
98	74	86	100	100	50	65
(42 / 43)	(26 / 35)	(6 / 7)	(5 / 5)	(4 / 4)	(4 / 8)	(13 / 20)

Table 3.5. Combined percentage MSI-H across various tumour types from the analyses of CaPP3 LS gene carriers and sporadic ECs by the Newcastle MSI Assay.

The one exception to this among LS-spectrum cancers was again found with EC, where the percentage of MSI-H samples was significantly lower than that observed with CRCs ($p=8.12x10^{-8}$, 95% CI=57-86%). Further investigation into the distribution of scores for these samples also revealed a greater variation in scoring within EC compared to the accompanying CRC population in which generally higher scoring, and a sharper distribution profile, are observed (Figure 3.2.). This MSI-H percentage however significantly exceeds that observed for the collection of sporadic EC tumours, where genetic instability was identified in only 9.8% (4/41) of samples, reflected in a distinct scoring distribution.

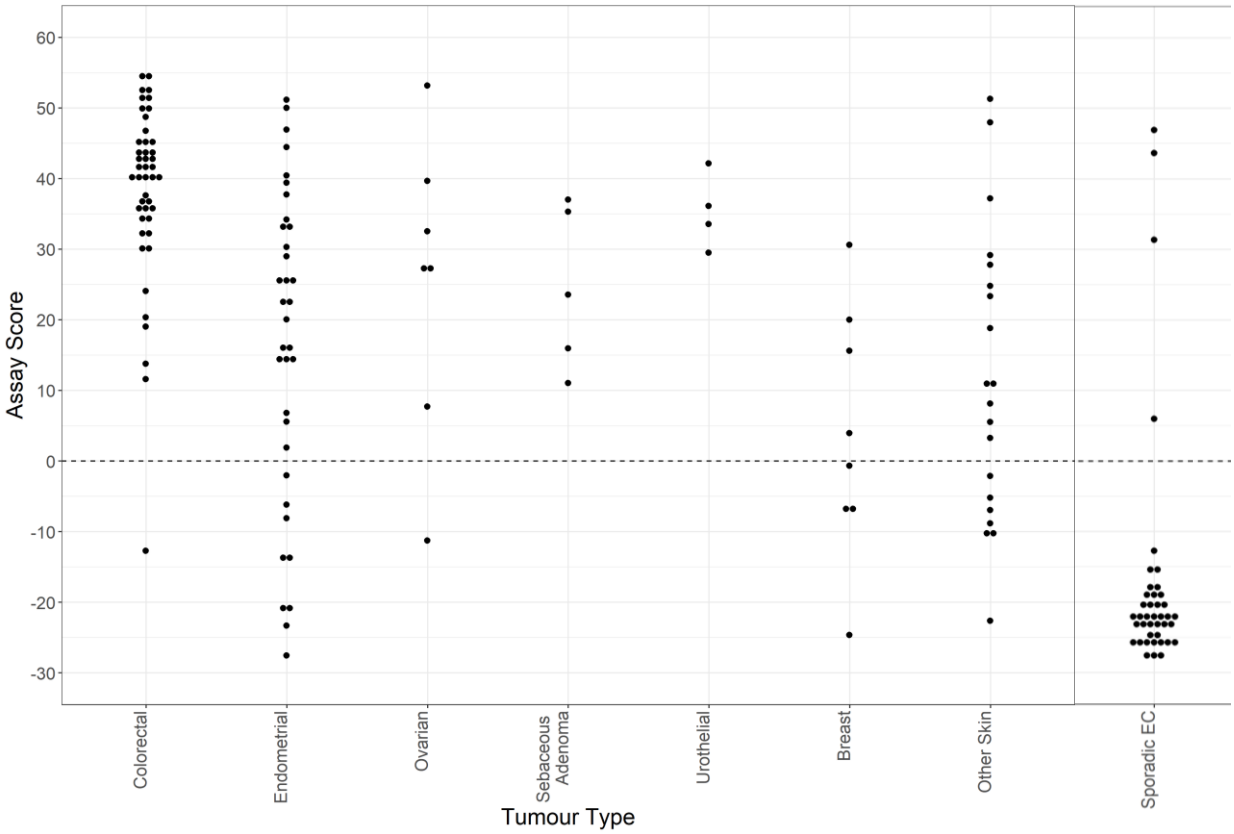

Figure 3.2. MSI Scores by tumour type from the combined analyses of CaPP3 LS gene carrier and sporadic EC tumours by the Newcastle MSI Assay. The dashed y-intercept at y=0 defines the classification, positive values above which result in an MSI-H designation, while negative values below such are designed MSS.

Previous analyses have established a variable contribution of defects in each MMR gene to the incidence of a given cancer in LS, and a gene-specific frequency of MSI-H in these (Møller *et al.*, 2018). Given this variation, I therefore analysed MSI scores with respect to the gene affected, with the results of this shown in Table 3.6. With reference to the underlying germline mutation defining cancer predisposition, the classifications of this study are generally consistent for each gene across the different tumour types, with no significant difference observed in MSI frequency between these. Exceptions to this exist in the percentage MSI-H detection between CRC and EC with either *MLH1* or *MSH6* deficiency, for which equivalent significant differences were calculated ($p=2.2x10^{-16}$, 95% CI (*MLH1*)=30-93%, 95% CI (*MSH6*)=29-96%). A notable difference is also found in the percentage MSI-H with *MSH2* defects again between CRC and EC (p=0.0067, 95% CI=52-94%), with not all instances of MSH2-

deficient EC classified as instable, but a 100% incidence of this in equivalent CRC samples. Such gene specific analyses of MSI frequency therefore reveal that the aforementioned differences in the prevalence of the phenotype between CRC and EC are a result of effects in all the major MMR genes, and not consequence of variations in only one.

	Colorectal	Endometrial	Ovarian	Sebaceous Adenoma	Urothelial	Breast	Other Skin
MLH1	100 (16 / 16)	67 (6 / 9)	100 (1 / 1)	-	-	100 (3 / 3)	60 (3 / 5)
MSH2	96 (22 / 23)	78 (14 / 18)	100 (3 / 3)	100 (5 / 5)	100 (4 / 4)	33 (1 / 3)	77 (10 / 13)
MSH6	100 (3 / 3)	71 (5 / 7)	100 (1 / 1)	-	-	0 (0 / 1)	0 (0 / 2)
PMS2	100 (1 / 1)	100 (1 / 1)	50 (1 / 2)	-	-	0 (0 / 1)	-

Table 3.6. Combined percentage MSI-H across various tumour types from analyses of CaPP3 LS gene carriers by the Newcastle MSI Assay, separated by confirmed underlying MMR gene defect.

3.5. Discussion

This study, investigating MSI detection using a sequencing-based technique, analysed a collection of extracolonic tumours, both of the accepted LS spectrum and not, in proportions similar to those observed in previous studies (Møller *et al.*, 2017) and representative of a typical LS population. For the detection of MMRd in LS screening, current diagnostic guidelines in the UK primarily recommends the testing of all individuals with CRC. However, the Newcastle MSI Assay returned interpretable scores, classifications of MSI and frequencies of this across the aforementioned range of tumour types, especially those considered associated with LS. Furthermore, the concordance between repeated analyses demonstrates the reproducibility of this approach, as well as the ubiquity in detection of the phenotype, in an LS context, in independent amplifications of the same samples. (Infrequent examples of discordance, occurring twice in this analysis, are likely due to technical variation or error, evidenced by the close proximity of scoring to the classification threshold, and may result from the absence of prior triaging of material for tumour cell content before DNA extraction.) The overall percentage of MSI-H results for each cancer is generally consistent with corresponding values reported in the literature, summarised in Table 3.5., while the proportions attributable to each separate gene defect within a given tumour type are also comparable to those previously reported (Møller *et al.*, 2018): both further indicate this population, although limited by sample number, to be representative of the actual LS situation. The availability of high throughput assays with a demonstrated capacity for MSI detection may, in principle, be applied to expedite developments in guidance. Furthermore, this may facilitate the testing of unselected populations of common tumours, such as prostate and breast cancer, in which MSI is relatively rare, and could negate any issues with the uptake of these as earlier described (Refer to Section 1.3.5.). However, for any such advances further consultation and economic evaluation will be required.

As discussed above, the frequencies of MSI classification, and the distribution of scores, by the Newcastle MSI Assay are generally consistent between tumour types, and comparable to the observations with CRC. However, three exceptions to this congruence exist, provided by endometrial, breast, and ‘other skin’ cancer. While the frequency of an MSI-H phenotype in both breast and ‘other skin’ tumours is significantly lower than observed in the comparator CRC population

(Breast: $p=1.9x10^{-5}$, 95% CI=0.16-0.84; Other Skin: $p=2.18x10^{-7}$, 95% CI=0.41-0.85), these represent tumour types not typically considered of the LS spectrum. EC, in contrast, is notable as an example of an acknowledged LS spectrum tumour in which the MSI-H designations are significantly fewer than returned for CRC. For these, scoring is also more variable with no clear dichotomy between MSS/MSI-H samples. This difference, and variation in scoring, may be biologically relevant, representing heterogeneous cell populations within a given sample, or a lesser contribution of MMRd to endometrial tumourigenesis. Alternatively, these results could be indicative of technical issues, illustrating the need to adapt the assay classifier by cancer to account for the specific biology of a given tumour type. The Newcastle MSI Assay, in its current iteration, employed a classifier with thresholds defined using high-quality CRC material, but it is reasonable to hypothesise that these parameters are not applicable for accurate classification across all tumour types, given the reported differences in allele length change. Tumour-specific classifier training would be necessary to negate this.

Of particular interest is the prevalence of EC readings designated MSS (26%), despite all samples originating from confirmed LS patients, and the use of a highly-sensitive assay. Consistent with existing literature, including a study in which 37% of EC samples were designated MSS using the Bethesda Panel (Gylling *et al.*, 2008), this shows that genetic instability is not exhibited ubiquitously in instance of deficiency, and that the presence of an MSS tumour does not exclude an LS diagnosis. While LS has been reported to be responsible for an equivalent proportion of malignancies in EC as CRC (Moreira *et al.*, 2012; Ryan *et al.*, 2019), the percentage of MSI-H in EC is significantly reduced from that of CRC in this study (p= 4.8x10-6, 95% CI= 58-87%), with a significantly lower incidence observed with germline defects in *MLH1*, *MSH2* and *MSH6*. As well as fundamentally questioning the importance of MSI in Lynch EC, this may also imply a different mechanism, independent of MMRd, by which tumours frequently develop in the condition. Alternatively, this could indicate a different way, independent of instability, in which MMRd may promote tumourigenesis, such as through the processes of homologous recombination or immunoglobulin class switching in which MMR genes are also implicated (Li, 2008). Nevertheless, conclusions regarding these EC findings are ultimately limited without the corresponding results of IHC and other analyses, with technical factors that cannot be

addressed here remaining a possible explanation for observations: resolution of this would require further investigation.

Classification of instability by the Newcastle MSI Assay is achieved through the use of a defined classifier, the accuracy of which is ultimately dependent on training using a comprehensive cross-section of samples, representative of all degrees of instability, which may be encountered in analysis. There is existing evidence in the literature that the amplitude of the MSI signal varies between tumour types, and is reduced in EC relative to CRC. As a result, use of a CRC-trained classifier may be inappropriate as previously mentioned, and could lead to MSI-H ECs being incorrectly designated MSS based on the properties of the CRCs used in training. This could however be addressed through cancer-specific classifier retraining, achieved through the use of well-characterised tumour cohorts.

Chapter 4. Assessment of MSI Detection by the Newcastle MSI Assay in Well-Characterised Endometrial Cancer Cohorts

4.1. Introduction

4.1.1. Endometrial Cancer: Epidemiology and Histopathology

With an estimated 417,000 new diagnoses in 2020, endometrial cancer (EC) is the second most common gynaecological malignancy globally, responsible for approaching 100,000 deaths per annum (Sung *et al.*, 2021). In the last three decades, the overall incidence of EC has increased globally, as has the number of EC-related deaths, with the greatest increases observed in the higher income countries of North America and Europe (Gu et al., 2021; Crosbie et al., 2022). Such developments are attributed to the differing prevalence of various risk factors, in age and obesity, with the frequency of diabetes and hypertension also considered contributory (Raglan *et al.*, 2019). Nevertheless, despite the aforementioned rises in disease, global mortality from EC has decreased by 15% over the same time period, effectively meaning more patients are both surviving and dying from the cancer. Currently, a given woman's lifetime risk of developing EC is approximately 3%, with a median age at diagnosis of 61 years. Diagnoses have, however, increased across all demographics, with a doubling of instances before the age of 40 years, which now represent approximately 4% of all cases (Matsuo *et al.*, 2021).

Traditionally, ECs have been classified into two general groups, Type I and Type II (Bokhman, 1983), with these differing in several regards such as epidemiology, histopathology and prognosis (Passarello et al., 2019). Type I tumours, also known as endometrioid adenocarcinomas (The American College of Obstetricians and Gynecologists, 2015), account for over 70% of diagnoses (Braun et al., 2016), but have a high overall five-year survival rate of 85% with low rates of recurrence (Morice *et al.*, 2016). Frequently associated with excess oestrogen stimulation, these cancers are generally well-differentiated, comparable to normal tissue, and are often confined to the uterus at the time of diagnosis, resulting in favourable prognoses (The American College of Obstetricians and Gynecologists, 2015; Passarello et al., 2019). In contrast, type II tumours, responsible for approximately 10% of disease instances (Sorosky,

2012), typically have a poor prognosis, with these accounting for 40% of EC-related deaths (The American College of Obstetricians and Gynecologists, 2015). These cancers, primarily of papillary serous or clear cell histological types (Passarello, Kurian and Villanueva, 2019), are generally considered more aggressive, with elevated levels of relapse and metastasis, as well as a lower five-year overall survival rate of 55% (The American College of Obstetricians and Gynecologists, 2015; Braun et al., 2016). More recently, an additional category of EC has been proposed in which genetic conditions are responsible for disease: approximately 10% of cases are currently considered attributed to this group (The American College of Obstetricians and Gynecologists, 2015).

4.1.2. Endometrial Cancer: Presentation and Diagnosis

EC most frequently presents early in disease progression with abnormal uterine bleeding, in particular postmenopausal bleeding (Clarke *et al.*, 2018). However, as a diagnostic criterion this is of limited value, with several other disorders producing this symptom, and only 5-10% of women with postmenopausal bleeding having underlying pathology (Crosbie *et al.*, 2022). The probability of postmenopausal bleeding as a consequence of EC before the age of 50 years is less than 1%, and only 3% by the age of 55 years, with this eventually increasing to approximately 25% in those older than 80 years (Gredmark *et al.*, 1995). Nevertheless, the timely investigation of all instances of postmenopausal bleeding is recommended to identify malignancy, particularly for individuals with risk factors for EC or hyperplasia such as polycystic ovaries, obesity, or past use of tamoxifen .

Similarly, an assessment of younger women with atypical uterine bleeding is also encouraged for the detection of EC, again with an emphasis on those with predisposing circumstances. Approximately 15% of all EC diagnoses present pre-menopause, with heavy, protracted and/or intermenstrual bleeding being typical presenting complaints, the latter being most predictive of disease (Pennant *et al.*, 2017). Still, as with postmenopausal bleeding, these are relatively common symptoms, and in only 0.3% of cases are these a result of endometrial malignancy (Crosbie *et al.*, 2022). The probability of EC presenting with bleeding abnormalities is commensurate with age and risk factors, and it is ultimately a consideration of these which determines the diagnostic strategy most suitable from a health-economic perspective (Feldman *et*

al., 1995; Gredmark *et al.*, 1995). However, it is also worth noting the frequency with which EC presents asymptomatically, occurring in 1-5% of disease instances: malignancy may be identified in these through abnormal cytology or as an incidental finding following hysterectomy (Passarello, Kurian and Villanueva, 2019).

Diagnosis of EC is primarily achieved through the histological examination of endometrial tissue, performed for individuals with endometrial pathology or a thickened endometrium on transvaginal ultrasound scan (Morrison *et al.*, 2022). In postmenopausal women with significant endometrial thickening (≥5mm), this invasive approach has been demonstrated to be effective for EC detection, with a 96.2% sensitivity observed in a systematic review of 1341 cases and 15,998 controls (Long *et al.*, 2020). However, the poor specificity of this, calculated at approximately 51.5%, means a significant proportion of women require additional analysis before endometrial pathology can be confirmed or rejected. The assessment of pre-menopausal women is similarly problematic, with transvaginal ultrasound less specific than in postmenopausal cases as a result of the cyclical fluctuations of endometrial thickness in healthy, reproductive-aged women (Crosbie *et al.*, 2022). In these situations, hysteroscopic analysis facilitates the direct sampling of suspicious lesions, and is advocated when focal endometrial pathology is detected by ultrasound, and/or for patients with recurrent symptoms suggestive of endometrial malignancy. The accuracy of diagnostic techniques, in particular endometrial biopsy, is ultimately greater in symptomatic and postmenopausal women and, despite their shortcomings, such approaches allow for the effective diagnosis of EC as opposed to atypical endometrial hyperplasia (Kerkar and Kaur, 2013).

4.1.3. Endometrial Cancer: Treatment

In the treatment of EC, surgery is the primary approach, with this typically involving total hysterectomy and bilateral salpingo-oophorectomy, as well as lymphadenectomy for selected patients with a high risk of nodal metastases (Matsuo *et al.*, 2021). For this, minimally invasive techniques such as laparoscopy are preferred, and have been shown to return comparable oncological outcomes to open surgery (Janda *et al.*, 2017). Surgical staging following tumour removal may be performed as part of the initial management in EC, with this affording the opportunity for prognostic stratification as well as the identification of patients who may benefit from adjuvant therapies such

as chemotherapy. Furthermore, in premenopausal women with apparent early-stage disease, this approach may allow for ovarian conservation, a practice demonstrated to be effective in negating the undesirable consequence of surgical menopause, without negatively effecting the probability of survival (Wright *et al.*, 2016). Surgery as described is ultimately a favoured strategy for treatment in the first instance for the aforementioned reasons among others, and, in early-stage, low-grade cancers with favourable prognoses, this alone may be curative (Matsuo *et al.*, 2021).

For women with extrauterine disease in preoperative assessment, prescription for surgery depends on several factors: the location of metastases; the prospect of complete cytoreduction; and, fundamentally, a patient's suitability for a given procedure. Retrospective meta-analyses of 672 patients with advanced or recurrent EC reported improved outcomes for patients subject to primary cytoreduction for advanced disease if debulking to no residual disease was achieved (Barlin, Puri and Bristow, 2010). Furthermore, cytoreductive surgery following chemotherapy corresponds with reduced perioperative morbidity, and equivalent survival, compared to upfront surgery with suboptimal resection, as well as improved outcomes compared with chemotherapy alone (Huang *et al.*, 2021). In support of surgery for recurrent EC, reports are limited, but the evidence that does exists indicate prolonged post-recurrence survival only if complete cytoreduction is realised (Scarabelli et al., 1998; Campagnutta et al., 2004; Bristow et al., 2006). However, such situations require additional considerations including the time from the original diagnosis, the location and extent of recurrence (resectability), and the general performance of the patient.

As an alternative therapeutic option, more recent studies have explored the use of immune checkpoint blockade (ICB) for the treatment of EC. Exploiting the presence of immune dysregulation in the disease, this has been employed as both a monotherapy, and in combination with other adjuvant therapies and/or targeted agents. Approximately 30% of primary ECs exhibit MSI (Bonneville *et al.*, 2017), while a further 13-30% of recurrent malignancies are found to be MSI-H or MMRd (Green, Feinberg and Makker, 2020). In these, mutation rates are elevated with neoantigens produced and presented that would otherwise be recognised as foreign for subsequent immunoediting and removal. However, such tumours also frequently demonstrate a capacity to evade immune regulation through the expression of inhibitory checkpoint molecules such as PD-1, and its complementary ligand PD-L1. The use of the PD-1 inhibitors dostarlimab and pembrolizumab in MSI-H ECs to counteract this has given

positive objective response rates (ORR) of 49% and 57% respectively in clinical trials, while PD-L1 inhibitors avelumab (ORR=27%) and durvalumab (ORR=43%) have also realised positive outcomes (Green, Feinberg and Makker, 2020). Studies regarding immune checkpoint blockade in combination with chemotherapy for the same purpose are continuing, with these following the suggestion of preclinical studies that chemotherapy may result in enhanced immune activation and cytotoxicity (Zitvogel, Kepp and Kroemer, 2011). Nevertheless, while these findings are promising and have been somewhat impactful, ICB has not delivered an overall improvement in survival from EC. For this treatment to be effectively deployed, and for an elucidation of the molecular mechanisms by which it operates, suitable candidates, that is, patients with MMRd, must be identified. One such population whom this may include are those with LS.

4.1.4. Identification of Lynch syndrome in Endometrial Cancer

For the identification of female LS gene carriers, NICE guidelines in the United Kingdom were extended in 2020 to advocate the testing of all EC instances (in addition to those of CRC) (National Institute for Health and Care Excellence, 2020). This follows the recommendations of several studies and collaborations, including those of the Manchester International Consensus Group, which have endeavoured to provide clinical direction for both the screening and management of EC in the context of LS (Crosbie *et al.*, 2019). However, unlike in the testing of CRC, the use of *BRAF* as a biomarker of MMR status is disregarded as mutations in this are so infrequent in EC, with *MLH1* promoter methylation considered a superior indicator of sporadic repair deficiency (Metcalf and Spurdle, 2014). What is more, the analysis of endometrial samples is advised, in the first instance, to be performed exclusively by IHC, with the recommended diagnostic pipeline not incorporating MSI testing.

The decision to exclude MSI analysis from clinical guidelines results in part from the lower sensitivity of this technique for MMRd detection in ECs, relative to IHC, observed in several studies. These include a review of 103 LS gene carriers, where MSI was determined by FLA of three mono- and three di-nucleotide repeats, in which instability was only detected in 41.66% of tumours from the 14 individuals with confirmed pathogenic Lynch mutations (Rubio *et al.*, 2016). However, other independent studies have, in contrast, described the suitability of using both IHC and

MSI analysis for the determination of MMR status in EC. A systematic evaluation of the approaches reported comparable sensitivity - 60.7–100% versus 41.7–100% - and specificity - 60.9–83.3% versus 69.2–89.9% - between IHC and MSI testing respectively across 13 studies of approximately 3500 participants (Stinton *et al.*, 2021). Furthermore, a relatively high concordance of approximately 94% between the two methods was also observed in a comparative study of FFPE material from 696 EC samples, with MSI as assessed by Promega MSI using a pentaplex panel of mononucleotide repeats (Stelloo *et al.*, 2017). Current recommendations are ultimately founded on studies using the existing MSI analysis techniques that suggest IHC is superior for the detection of repair deficiency in EC. However, the aforementioned references describe disparity in the identification of MMRd, and the concordance of IHC and MSI analysis results, a difference that may be resolved by the use of improved MSI testing. In addition, if the cost of both IHC and MSI testing were taken into consideration, the use of a more sensitive, and cheaper, practice for MSI review may have significant impact on the suitability of MSI testing strategies more widely. The Newcastle Assay may provide such an effective and convenient approach for this purpose, with this shown more sensitive to instability in the context of CMMRD (Gallon *et al.*, 2019), and capable of detecting the majority of LS cases in EC (El-Shakankery *et al.*, 2023).

My initial analysis of Lynch ECs using the Newcastle MSI Assay returned a detection rate for instability of approximately 74% (Refer to Section 3.4.3.). This finding is generally consistent with those reported in the literature for LS-derived material, while being realised using a technique which negates the subjective element of alternative approaches. However, the ability to draw conclusions from this study, with regards to the Newcastle Assay's capacity to detect repair deficiency in EC, is ultimately limited by the absence of supplementary information for the cohort analysed such as corresponding IHC classifications or MSI results using current methods. To thoroughly assess the use of this assay with ECs, its application with additional, well-characterised EC cohorts is required.

4.1.5. Sample Cohorts Suitable for Analysis by the Newcastle Assay

To further explore the ability of the Newcastle MSI Assay to accurately detect instances of MMRd in EC, additional, well-characterised samples were required. Two groups involved in clinical trials regarding EC management had large collections of tumour material and/or DNA samples, for which the results of previous molecular analyses, as well as extensive clinical data, is available. These samples were requested for analysis by the Newcastle Assay.

From Ohio State University, a cohort of 200 EC samples were provided in the form of purified DNA aliquots, all from tissue sections. The first 100 of these comprised patient material from individuals diagnosed with primary invasive EC between 2013 and 2016, all of whom were prospectively enrolled into the Ohio Colorectal Cancer Prevention Initiative, or OCCPI (results for which have previously been summarised (Hampel *et al.*, 2021). For all these samples, IHC and MSI typing was available, with the results of the latter generated through Promega analysis. The remaining 100 samples originated from women for whom a hysterectomy or diagnostic biopsy had confirmed a newly-diagnosed EC between 2017 and 2020, included in the OPTEC study of multigene panel testing for cancer susceptibility (M. D. Levine *et al.*, 2021). For all these samples, IHC and clinical NGS MSI typing were provided, while the results of Promega MSI testing was also available for 76 of the 100 cases.

A total of 191 EC samples were also received from the University of Manchester, representing a subset of the ~500 cases examined in the prospective PETALS study (Ryan *et al.*, 2020). These samples, provided in the form of FFPE tumour material, represent women recruited to the Manchester University NHS Foundation Trust gynaecological clinics between 2015 and 2017 following a diagnosis of EC or atypical hyperplasia. The results of previous IHC typing, performed on blocks with >70% tumour cell content, were available for this cohort, as well as the findings of MSI analysis by Promega MSI achieved using microdissection.

For all samples of the Ohio cohort, IHC analysis involved monoclonal antibodies for the four principal MMR proteins, with convincing staining in greater than 1% of cells, as well as equivocal or weak staining, considered ‘present’ and repair proficient. In assessment by the Promega MSI Analysis System, five monomorphic MNR markers were used (BAT-25, BAT-26, NR-21, NR-24, MONO-27), with tumours exhibiting instability in two or more of these designated MSI-H. These analyses are equivalent to those performed for the samples from the University of Manchester, with the same

MNR panel used for MSI assessment, and equivocal and ‘patchy’ staining of samples in IHC also considered reflective of repair proficiency. However, where the two cohorts differ is in the concordance between the results of these prior evaluations. While there is a 96.5% agreement between the findings of IHC and MSI testing for the samples of the Ohio cohort, the corresponding value for the samples from the University of Manchester is significantly lower at 74.6%.

Given this difference in the concordance of results, as well as the contrasting origins and pathological preparation of samples, it was therefore decided that the two cohorts would be analysed separately in the review of MMRd detection by the Newcastle MSI Assay.

4.2. Aims

In the following project, I aimed to investigate the potential utility of the Newcastle Assay for the analysis of endometrial cancer, and compare the performance of this to other commonly used methods, namely IHC and MSI. For this there were several objectives:

- Assess the ability of the Newcastle MSI Assay to detect repair-deficient endometrial tumours, including LS tumours.
  To be addressed through the analysis of EC samples provided by external collaborators as outlined above, and through comparison of results with previous IHC and MSI molecular analyses.

- Explore the potential for improvement in MMRd detection in EC using the Newcastle MSI Assay.
  To be addressed through additional analysis of EC samples using more sensitive markers, as well as through classifier re-training (using ECs rather than CRCs) for tumour-specific classification thresholds.

4.3. Methods

For this study a total of 391 independent endometrial cancers were analysed across two separately-reviewed cohorts; 200 supplied by the University of Ohio, and a further 191 from the University of Manchester. Samples from both Ohio and Manchester were further divided into two data sets of equal size – a training set, for which existing IHC and MSI testing results were known, and a validation set, for which I was initially blinded to the information from these prior analyses. These data sets further contained similar frequencies of MMRp and MMRd EC tumours.

Samples originating in Ohio were presented in the form of purified DNA, each extracted from malignancies with confirmed pathogenic histology. In contrast, samples from Manchester were provided in the form of FFPE material, from which curls were removed and DNA subsequently extracted, all without any additional preparatory triaging for tumour cell content (Detailed in Sections 2.3. and 2.4.). All samples were initially assessed using the Newcastle MSI Assay with its original 24 MNR maker panel and the CRC-trained classifier. Products generated in PCR amplification by the Newcastle Assay were sequenced to an average read depth of 5000X (Detailed in Section 2.5.4.), and analysed by custom scripting pipelines (Detailed in Section 2.6.), before comparing the results to those of prior analyses. In instances of discordance with previous findings, specifically for material of the Manchester cohort, repeat analysis of samples was performed, with the reassessment of independent curls removed from FFPE blocks of interest to resolve disagreements. There followed the re-evaluation of all samples with a novel panel of 62 highly sensitive markers (Detailed in Section 2.5.), as well as the review of validation cohort samples following classifier re-training. (In classifier re-training, the designated EC training data sets were used to re-train the classifier of the Newcastle Assay, and the resulting parameters were used to classify the samples of their respective validation data sets.) Both subsequent re-analyses involved the same PCR amplification practices, and use of custom scripting for samples classification as was employed in the initial testing by the Newcastle Assay, facilitating a comparison of findings from the three inquiries.

4.4. Results

For an assessment of MMRd detection by the Newcastle Assay in EC, a total of 396 samples were provided from two external sources. As a result of the different forms in which these samples were provided, drop out during the analysis process was disparate between the cohorts, a summary of which is shown in Table 4.1.. From the University of Ohio, only four samples were lost in the processing pipeline, all at the sequencing/data quality control stage. In comparison, 24 samples from the University of Manchester failed to proceed through the Newcastle Assay pipeline, with six damaged in initial curl cutting and DNA extraction, and the remaining 18 returning insufficient amplification or sequencing output. Furthermore, a total of 42 samples from Manchester were also re-analysed to further investigate potential causes of a lower concordance between assays in this cohort: for this purpose separate curl removal and DNA extraction was performed.

As related in Table 4.2., analysis of the training and validation datasets from Ohio (by the Newcastle MSI Assay) returned a concordance with IHC of 96.88% and 95.00%, resulting in a combined concordance for the cohort 94.90%. This contrasts the same analysis of the Manchester samples, for which MSI analysis of the training dataset in Newcastle had an 85.00% concordance with IHC, and the validation dataset an 80.46% concordance, resulting in an 82.63% concordance for the cohort. Analysis of both cohorts with an expanded panel of 62 novel markers for MSI returned similar concordance levels with the samples from Ohio (96.70%), but a decreased concordance relative to IHC results for the Manchester material (76.70%), a trend that was repeated with analysis by the Newcastle Assay employing an EC-trained classifier. Use of the training datasets of each cohort to review their corresponding validation cohorts further revealed concordance levels similar to that of the initial analysis, with concordance levels relative to IHC of 94.19% and 79.09% for the Ohio and Manchester cohorts respectively. These results will be discussed in more detail in the following text.

Data Set		Blinded/ Unblinded	Samples Received	Exclusions				Samples Analysed	Samples Repeated
				Sample Damage	Insufficient DNA Extraction	Insufficient MIP Amplification	Failed QC		
Ohio	Training	Unblinded	100	-	-	3	1	96	-
	Validation	Blinded	100	-	-	-	-	100	-
	Merge	Unblinded at Analysis	200	-	-	3	1	196	-
Manchester	Training	Unblinded	96	2	-	13	1	80	24
	Validation	Blinded	95	3	1	4	-	87	18
	Merge	Unblinded at Analysis	191	5	1	17	1	167	42

Table 4.1. Summary of all EC samples provided to, and ultimately analysed by, the Newcastle Assay, with separation by their cohort of origin.

Sample Origin	Ohio						Manchester					
Data Set	Training	Validation	Merge	Merge	Validation	Merge	Training	Validation	Merge	Merge	Validation	Merge
Marker Panel	24 MNR	24 MNR	24 MNR	62 Marker	24 MNR	24 MNR	24 MNR	24 MNR	24 MNR	62 Marker	24 MNR	24 MNR
Classifier Training	CRC	CRC	CRC	CRC	EC	EC	CRC	CRC	CRC	CRC	EC	EC
% Concordance (Original IHC Vs. Original MSI)	95.83	97.00	96.43	96.15	96.51	96.43	71.25	81.61	76.65	73.30	81.40	79.04
% Concordance (Newcastle Assay Vs. IHC)	94.79	95.00	94.90	96.70	94.19	96.43	85.00	80.46	82.63	76.70	79.07	70.06
% Concordance (Newcastle Assay Vs. Original MSI)	96.88	98.00	97.45	98.35	97.67	97.96	83.75	82.76	83.23	85.23	86.05	64.67
Sensitivity and Specificity (Vs. IHC)	93.62 & 97.96	92.31 & 97.92	92.93 & 97.94	95.79 & 98.85	93.75 & 94.74	95.92 & 96.94	76.19 & 94.74	78.00 & 83.78	77.17 & 89.33	64.95 & 91.14	72.00 & 88.89	84.78 & 52.00

Table 4.2. Summary of all analyses of EC samples performed by the Newcastle Assay in its various iterations as indicated.

4.4.1. Accurate MSI Detection with the Newcastle MSI Assay in EC Samples from the University of Ohio

To determine the capacity for MSI detection in this tumour type, a cohort of 200 EC samples were provided by the University of Ohio, with this consisting of two distinct datasets depending on their application (Detailed in Section 2.2.1). All samples were provided in the form of purified DNA from which amplification was conducted, followed by sequencing and data analysis to assess instability in these collections of predominantly sporadic material. An example of MIP amplification is shown in Figure 4.1.. Amplicons of the expected size range (240-260bp) are observed in all samples apart from the negative control (Sample E8), with primer and dimer bands also visible (<100bp).

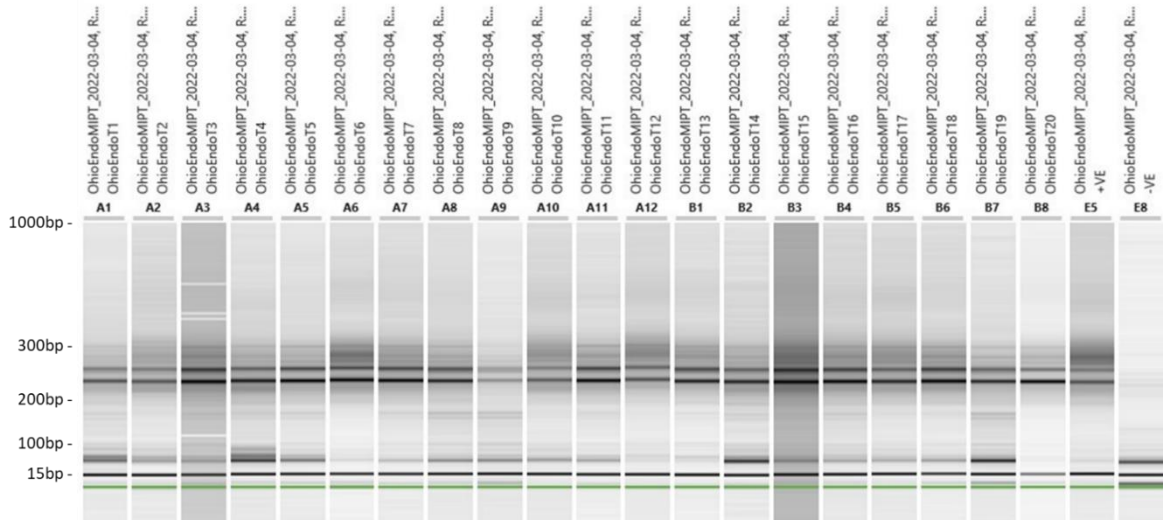

Figure 4.1. Exemplar Qiaxcel capillary electrophoresis image showing the results of MIP amplification from the purified tumour DNA of Ohio EC samples. Amplicons of approximately 240-260bp in length are expected from amplification using the 24 MNR MIP pool of the original iteration of the Newcastle Assay.

Overall, 96 of the 100 Ohio training set samples produced sufficient sequence data of the target read depth for review (Table 4.1.). In the analysis of these using the Newcastle Assay, with its 24 MNR marker panel and CRC-trained classifier, a high level agreement with the results of previous analyses was observed. Specifically, a 96.88% concordance with previous IHC findings was realised in this assessment, representing a negligible increase on the concordance between Ohio IHC and MSI testing (95.83% (p=0.80, 95% CI=0.91-0.99)) (Table 4.3.). IHC is susceptible to observer error, false negatives resulting from technical failure, and false positives where proteins are dysfunctional despite being present. Nevertheless, many consider IHC the 'gold-standard' for the detection of repair deficiency in EC, therefore the accuracy of our results was then calculated relative to classification by this approach. This afforded the Newcastle Assay a sensitivity of 93.62% and specificity of 97.96%, levels comparable to those obtained in the Promega MSI analysis of the same material (95.74% (p=0.45, 95% CI=0.82-0.97) and 97.96% (p=1.0, 95% CI=0.89-1.00) respectively.

		Concordance of Newcastle MSI Assay with Previous Ohio Assay Results				
		Both Ohio IHC and MSI Analyses	Ohio IHC Only	Ohio Promega MSI Only	Neither Ohio Analyses	% Concordance Ohio IHC and Ohio MSI
Ohio IHC	MMRp	98	-	2	-	98
		(48 / 49)		(1 / 49)		(48 / 49)
	MMRd	91	4	2	2	96
		(43 / 47)	(2 / 47)	(1 / 47)	(1 / 47)	(45 / 47)

Table 4.3. Percentage concordance of the Newcastle MSI Assay with previous analyses results from Ohio, separated by repair status as determined by IHC. Numbers of concordant results per group are shown in parentheses.

As little disagreement with previous analyses was seen in review of the training samples by the Newcastle Assay, the designated validation sample set was then amplified and sequenced as before, with all 100 samples successfully proceeding through the Newcastle Assay pipeline (Table 4.1.). The resulting scores and classifications of this process were sent to Ohio for unblinding to allow for a

comparison with earlier test findings. The test set returned similarly concordant results relative to previous analyses, with a concordance of 95.00% with existing IHC analysis being observed in this assessment, representing a negligible decrease on the concordance found between the original Ohio IHC and MSI testing results (97.00% (p=0.23, 95% CI=0.89-0.98)) (Table 4.4.).Furthermore, with respect to Ohio IHC findings, a sensitivity of 92.31% and specificity of 97.92% was realised for this sample set, levels again comparable to those observed in the original MSI analysis of this cohort (94.23% and 100.00% respectively) (Table and Summary Table), with the latter differing by a single call.

		Concordance of Newcastle MSI Assay with Previous Ohio Assay Results				
		Both Ohio IHC and MSI Analyses	Ohio IHC Only	Ohio MSI Only	Neither Ohio Analyses	% Concordance Ohio IHC and Ohio MSI
Ohio IHC	MMRp	98	-	-	2	100
		(47 / 48)			(1 / 48)	(48 / 48)
	MMRd	92	-	6	2	94
		(48 / 52)		(3 / 52)	(1 / 52)	(49 / 52)

Table 4.4. Percentage concordance of the Newcastle MSI Assay with previous analyses results from Ohio, separated by repair status as determined by IHC. Numbers of concordant results per group are shown in parentheses.

Given the similarity of results obtained for the Ohio sample sets separately (in regard to sensitivity, specificity and concordance with previous analyses), combination of the data from the two was considered justified to allow for summative review of the component samples. When combined, a sensitivity of 92.93% and specificity of 97.94% were returned for this analysis relative to the previous IHC observations of Ohio, with both values lower, but not significantly so, than the levels obtained in preceding MSI assessment (94.95% (p=0.35, 95% CI=0.86-0.97) and 98.97% (p=0.26, 95% CI=0.93-1.00) respectively).

Comparing the classifications returned by the Newcastle Assay with those of the original IHC results further reveals a 94.90% concordance between the two analyses, a not unexpected finding given the degree of agreement between the two

external results for these samples (96.43%). To investigate assay concordance further, the MSI score by the Newcastle Assay was plotted with respect to the nature of concordance with previous assay results as shown in Figure 4.2..

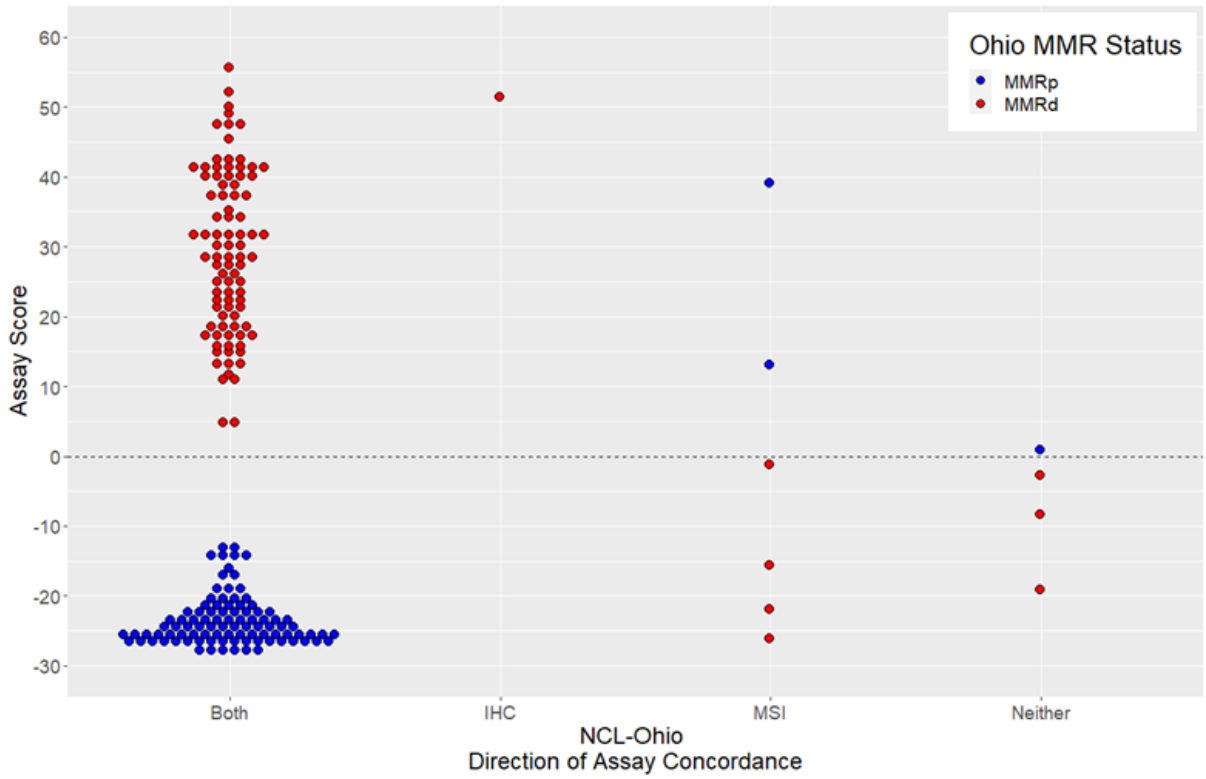

Figure 4.2. MSI Scores for all Ohio endometrial cancer samples when analysed by the Newcastle MSI Assay using its original marker panel and a CRC-trained classifier. Separation along the x-axis delineates the direction of concordance between the Newcastle Assay and previous analyses conducted by Ohio for the same samples, whereby the left-most column represents agreement with both prior IHC and MSI assessments, followed by agreement with Ohio IHC only, Ohio MSI testing only, and finally neither. The dashed y-intercept at y=0 defines the classification, positive values above which result in an MSI-H designation, while negative values below such are designed MSS. Finally, coloration is applied according to sample MMR status as determined by IHC.

This figure illustrates that, where all three assays are concordant, classification is clearly dichotomised, with only two MSI-H samples within 10 arbitrary points of the classification boundary. Furthermore, it also illustrates how the majority of situations in which the Newcastle Assay and IHC do not agree are repair deficient samples (by IHC) scoring below the classification threshold, with four of these also not detected by the original MSI analysis. The single instance in which the Newcastle Assay concurred with the previous IHC result only (Score=51.46) represents a sample with *MLH1/PMS2* deficiency, confirmed by Ohio upon reassessment of pathology slides, for which the

original MSI testing was seemingly insufficiently sensitive to identify instability. In contrast, six samples showed agreement between the outcomes of the Newcastle Assay and Promega MSI only, with four of these being cases of *MSH6* loss in which exceptionally-high levels of MSI are not detected: the remaining two samples were ascribed MSI-H classifications (Scores=13.14 and 39.17) despite MMR proficiency as assessed by IHC.

MLH1/PMS2-deficient samples account for three of the final four discordant cases, in which the Newcastle result differed from those of both IHC and Promega MSI. However, it is worth noting that two of these returned assay scores in close proximity to the classification threshold (Scores=-2.78 and -8.32). The same is true of the other sample for which similar discordance was observed, and an MSI-H status was given to a sample initially considered repair proficient. However, upon later review by Ohio, clonal loss of *MLH1* was recognised for this sample, a finding that may explain the earlier discordant result. Ultimately, in the analysis of this cohort, all three considered assays produced comparable results, with the concordance of any one of these with either of the other two being ~95% or greater. We can therefore conclude that the Newcastle Assay, with its original 24 MNR marker panel and default CRC-trained classifier, performs as well as Promega MSI for MMRd detection within the Ohio cohort relative to IHC, the advocated method for deficiency testing for this tumour type.

4.4.2. Concordance Between the Results of IHC and MSI Analyses is Significantly Lower in the University of Manchester EC Cohort

The second cohort of EC samples reanalysed by the Newcastle Assay consisted of 191 samples provided from the University of Manchester, with these again divided into two distinct data sets depending on their intended use for either classifier retraining or validation (Detailed in Section 2.2.1). However, unlike those from Ohio, considerable disagreement existed between the results of previous analyses for these samples, with a concordance of 74.6% between IHC and Promega MSI for the cohort. To establish if the Newcastle Assay can improve upon this relatively low concordance, these samples were assessed through the same pipeline as those from Ohio, with this also allowing for a determination of whether the results from analysis of the Ohio material are representative of the tumour type.

All samples were received from Manchester in the form of FFPE material from which curls were isolated. DNA was extracted, amplified, sequenced and analysed to evaluate instability (Detailed in Section 3.3.). An example of MIP amplification of this cohort is shown in the gel electropherogram of Figure 4.3.. Amplicons of the expected size range (240-260bp) are observed for all samples apart from one (highlighted) and the negative control, with primer and dimer bands also visible (<100bp).

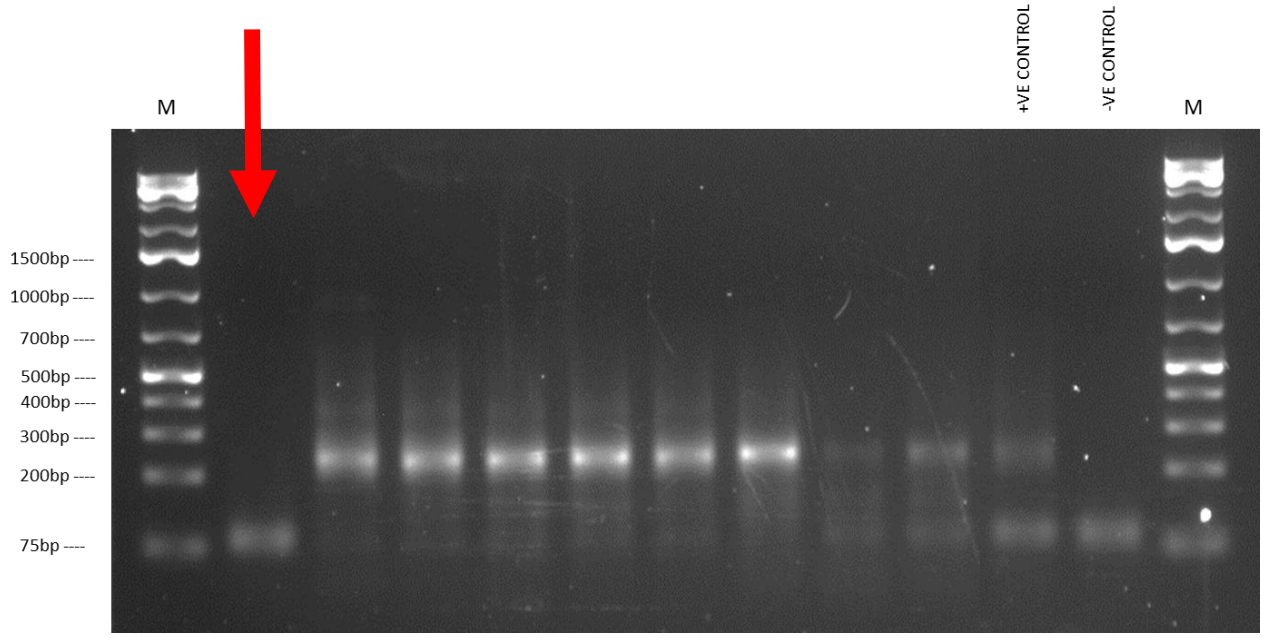

Figure 4.3. Exemplar gel electrophoresis image showing the results of MIP amplification from the extracted tumour DNA of Manchester EC samples. Amplicons of approximately 240-260bp in length are expected from amplification using the 24 MNR MIP pool of the original iteration of the Newcastle Assay. Failure of the amplification process is indicated by the arrow, with no banding present at a size corresponding with the expected products from this reaction.

In total, 80 of the 96 Manchester training set samples proceeded through the entire Newcastle MSI Assay pipeline, and produced sequence data of the target read depth for review (Table 4.1.). Analysis of these samples returned an 85.00% concordance with previous IHC findings (Table 4.5.), a significant increase on the value observed between the classifications of IHC and Promega MSI testing by Manchester (71.25% (p=0.0062, 95% CI=0.75-0.92)). Furthermore, relative to existing IHC results, this review by the Newcastle Assay produced a test sensitivity of 76.19% and specificity of 94.74% for the detection of repair deficiency with these samples. These levels are higher than those from Promega MSI analysis of the same material (54.76% and 89.47% respectively), although this difference is only significant with regards to sensitivity (p=0.005, 95% CI=0.61-0.88).

		Concordance of Newcastle MSI Assay with Previous Manchester Assay Results				
		Both Manchester IHC and MSI Analyses	Manchester IHC Only	Manchester Promega MSI Only	Neither Manchester Analysis	% Concordance Man. IHC and Man. MSI
Manchester IHC	MMRp	89	5	5	-	89
		(34 / 38)	(2 / 38)	(2 / 38)		(34 / 38)
	MMRd	52	24	21	2	55
		(22 / 42)	(10 / 42)	(9 / 42)	(1 / 42)	(23 / 42)

Table 4.5. Percentage concordance of the Newcastle MSI Assay with previous analyses results from Manchester, separated by repair status as determined by IHC. Numbers of concordant results per group are shown in parentheses.

Because of the lower concordance in the Manchester results, repeat analysis of discordant samples from the training set was undertaken to eliminate the possibility technical reasons for the disparity with IHC results observed (detailed in Section 4.4.4.). However, once these samples, and their corresponding MSI classification, had been verified, review of the designated validation set was performed. For this set, a total of 95 samples were initially available, but 87 of the collective were ultimately investigated by the Newcastle Assay (Table 4.1.). The material of these samples was sequenced and amplified as before, followed by the sharing of results with Manchester for unblinding and ultimately the comparison of test findings. In this analysis, a concordance of 80.46% with previous IHC results was returned (Table 4.6.), a value lower, but not significantly so, than the agreement between Manchester IHC and MSI testing of the same material (81.61% (p=0.78, 95% CI=0.71-0.88)). In addition, relative to previous IHC results, these findings represent a sensitivity of 78.00% in the identification of repair deficiency by the Newcastle Assay, a value higher than that observed in Promega MSI testing of the same material (68.00%), but not significantly so (p=0.17, 95% CI=0.64-0.88). The 83.78% specificity of this approach in MMRd detection is significantly lower than that of Promega MSI for these samples (100.00% (p=<2.2e-16, 95% CI=0.68-0.94)).

		Concordance of Newcastle MSI Assay with Previous Manchester Assay Results				
		Both Manchester IHC and MSI Analyses	Manchester IHC Only	Manchester Promega MSI Only	Neither Manchester Analysis	% Concordance Man. IHC and Man. MSI
Manchester IHC	MMRp	84	-	-	16	100
		(31 / 37)			(6 / 37)	(37 / 37)
	MMRd	64	14	18	4	68
		(32 / 50)	(7 / 50)	(9 / 50)	(2 / 50)	(34 / 50)

Table 4.6. Percentage concordance of the Newcastle MSI Assay with previous analyses results from Manchester, separated by repair status as determined by IHC. Numbers of concordant results per group are shown in parentheses.

In both training and validation sample sets, relatively low concordance levels were observed between previous IHC and Promega MSI results. Furthermore, the pattern of discordance of the Newcastle Assay with prior analyses was also similar in both sample sets, with most discordant samples being MMRd (20/24 in the training set, 17/23 in the validation set). To investigate the direction of concordance in more detail, and subsequently compare these results to those of the Ohio cohort, the data from the Manchester training and validation samples was therefore merged.

When integrated, an overall assay sensitivity of 77.17% and specificity of 89.33% were obtained for the Newcastle Assay relative to the previous IHC findings of Manchester, with the former significantly higher than the corresponding Promega MSI testing of the same material (61.96% (p=0.0025, 95% CI=0.67-0.85)), and the latter not significantly different (94.67% (p=0.063, 95% CI=0.80-0.95)). However, these values for both performance metrics are lower than those returned in the analysis of the Ohio samples, albeit only significantly so with respect to sensitivity (p=5.12e-9). This suggests that, for the detection of repair deficiency in EC, the Newcastle Assay is sufficiently specific, returning few false-positive classifications relative to existing IHC findings.

Relatively low concordance (76.65%) exists between the results of IHC and Promega MSI assessment for these samples of the Manchester cohort. Reviewing the concordance of the Newcastle Assay with these antecedent tests reveals an 82.63%

concordance with IHC and an 83.23% concordance with Promega MSI, both increases in the concordance observed between these prior analyses (76.65%), but significantly different only from the concordance with Promega MSI (p=0.044, 95% CI=0.77-0.89).These results are however both significantly lower than those returned in the same comparisons performed for the Ohio cohort , indicating the potential for contrasting concordance between EC cohorts. Furthermore, as high concordance is not observed specifically between the two MSI assays, this suggests our findings are not a result of MSI being less sensitive in this cohort, and that other factors must be responsible for the discordance.

To investigate the patterns of concordance within the Newcastle results further, the MSI scores by the Newcastle Assay were plotted, and separated by the nature of concordance with previous assay results: this is shown in Figure 4.4.. Unlike with the Ohio samples, this shows that where all three assay results are concordant, there is no clear separation between MSI-H and MSS classifications, with 31 samples returning an assay score between +10 and -10. Of further potential interest is the pattern of concordance observed, particularly when there is concordance between the Newcastle Assay and only one of the previous analyses by Manchester. When there is agreement between the Newcastle MSI classification and the previous IHC result only, these are generally instances of MMRd congruently detected by the Newcastle MSI Assay but not Promega MSI (17/19, Figure 4.4.). Conversely, where these is agreement with Promega MSI but not IHC, these are generally samples classified MMRd by IHC but MSS by both MSI testing approaches (18/20). Of the 17 samples considered MMRd by IHC and MSI-H by the Newcastle Assay, but not detected by Promega MSI, a majority (14/17) are defined by *MLH1/PMS2* loss (Scores=2.29 – 52.33). In contrast, of the 18 repair-deficient samples classified as MSS by both MSI testing approaches, a notable proportion (5/18) result from *MSH6* loss (Scores=-2.05 - -24.63), albeit with one of these in close proximity to the classification threshold. These patterns of concordance, and associated distributions of scoring, are consistent with the suggestion that sensitivity in MSI detection alone cannot account for these results of the Newcastle Assay. They also suggest that they are unlikely to be due to a phenomenon affecting the entire cohort, such as ‘noise’ introduced by DNA extraction from distinct FFPE curls. This directionality within the pattern of discordance ultimately indicates that there may be an underlying biological or technical reason for

the classification of EC samples from the Manchester cohort, and therefore the concordance of results between analyses.

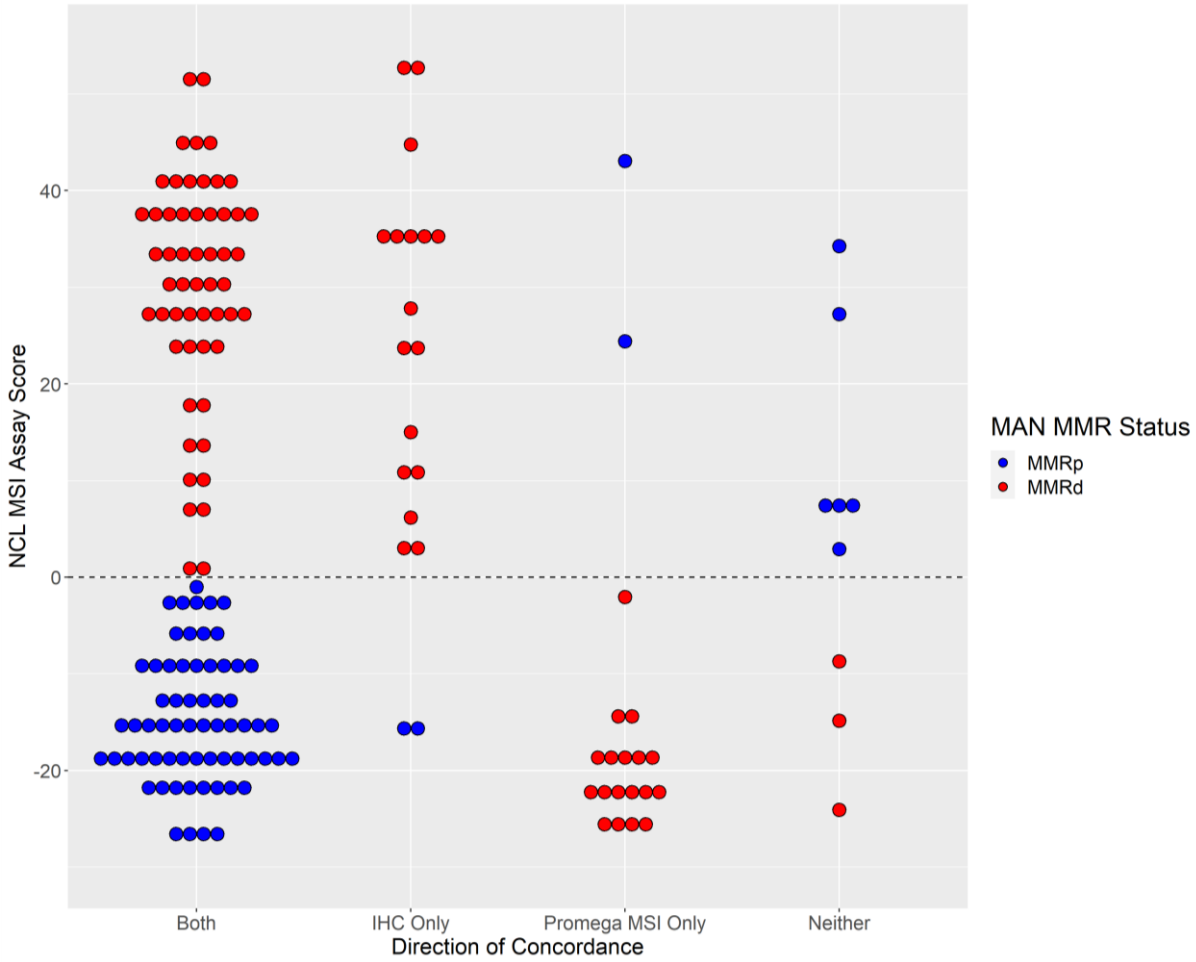

Figure 4.4. MSI Scores for all Manchester endometrial cancer samples when analysed by the Newcastle MSI Assay using its original marker panel and a CRC-trained classifier. Separation along the x-axis delineates the direction of concordance between the Newcastle Assay and previous analyses conducted by Manchester for the same samples, whereby the left-most column represents agreement with both prior IHC and MSI assessments, followed by agreement with Manchester IHC only, Manchester Promega MSI testing only, and finally neither. The dashed y-intercept at y=0 defines the classification threshold, positive values above which result in an MSI-H designation, while negative values below such are designed MSS. Finally, coloration is applied according to sample MMR status as determined by IHC.

4.4.3. MLH1 Methylation Analysis Suggests the Newcastle MSI Assay is More Sensitive than Promega MSI for the Detection of MMRd within the Manchester Cohort

Given the association of *MLH1* methylation with repair deficiency, whereby this modification of the *MLH1* promoter may result in sporadic MMRd and thus MSI, *MLH1* promoter methylation status was established for the majority of MMRd ECs as part of the Manchester LS detection pipeline. Consequently, it was possible to investigate the distribution of this methylation, and the quantitative scores for this, to provide independent confirmation of MMRd status, as well as consider how this relates to the findings in the initial analysis of the Manchester ECs using the Newcastle Assay.

Methylation information of varying detail was available for a total of 91 samples from the Manchester cohort. In the first instance, *MLH1* methylation classifications, that is, whether this was ‘methylated’ or ‘normal’, were extracted across the whole dataset, together with percentage methylation (as determined by the pyrosequencing approach of Reflex *MLH1*-methylation testing) which was only available for a subset of samples. The Newcastle MSI scores and the percentage of methylated bases were then compared for all samples were the data for both was available (N=36). A significant, moderate positive correlation was observed (Figure 4.5., ($R=0.49$, $p=0.0023$)), consistent with the known causative relationship between *MLH1* promoter methylation and microsatellite instability.

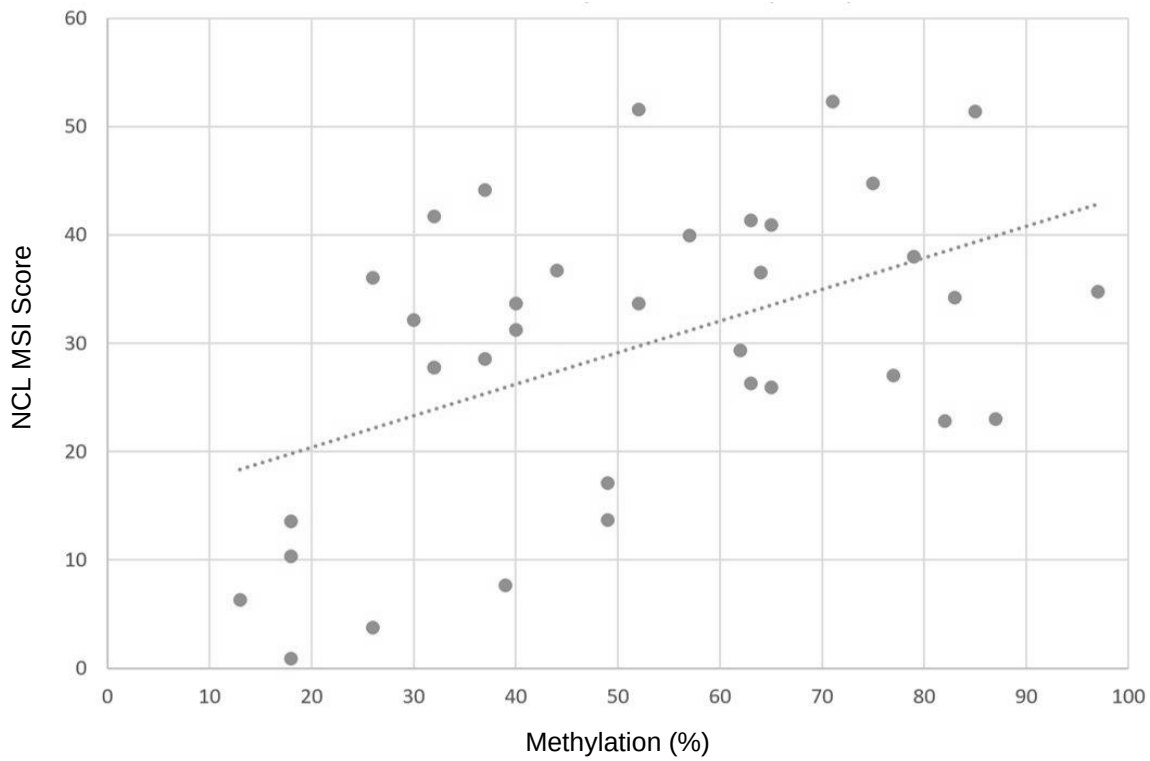

Figure 4.5. *MLH1* promoter hypermethylation versus tumour MSI score for Manchester endometrial cancer samples. Position along the x-axis denotes the percentage of methylated bases within the *MLH1* promoter of hypermethylated sporadic tumours (>10% methylated bases), while the y-axis relates the MSI score generated by the Newcastle Assay with a CRC-trained, 24 marker panel.

For a further 19 samples *MLH1* methylation status was known, but no quantitative result for this was recorded. Among all hypermethylated samples, all but one was MMRd by IHC (54/55), while 52 were MSI-H by the Newcastle Assay, and 42 were MSI-H by Promega MSI. The Newcastle MSI assay therefore fails to detect two hypermethylated MMRd samples, compared to 12 missed by Promega MSI. Considering this differing detection of methylation, I finally analysed the relationship between methylation status and the direction of concordance between assays. These results are presented in Figure 4.6.. As expected, among samples where all three assay results are concordant, there is a clear association between MMR and methylation status (54/65 MMRd samples with MLH1 loss are methylated). However, there is also a significant association between methylation status and the Newcastle MSI classification specifically among samples where the assay is concordant with only one of the original assays used (Fisher's Exact Test – p=0.0034). 11/13 samples which are MMRd by IHC and MSI-H by the Newcastle Assay are hypermethylated compared

to only 2/10 samples which are MMRd but MSS by both MSI assays, suggesting the Newcastle Assay is identifying hypermethylated/MMRd tumours which the Promega MSI assay is not. This can, at least partially, account for the increased sensitivity and concordance with IHC, of the Newcastle Assay relative to Promega MSI within this sample cohort. It also suggests that the Newcastle assay is both identifying MSI within MMRd samples which was not picked up by the Promega assay, and identifying some MSS samples erroneously classified as MMRd by IHC.

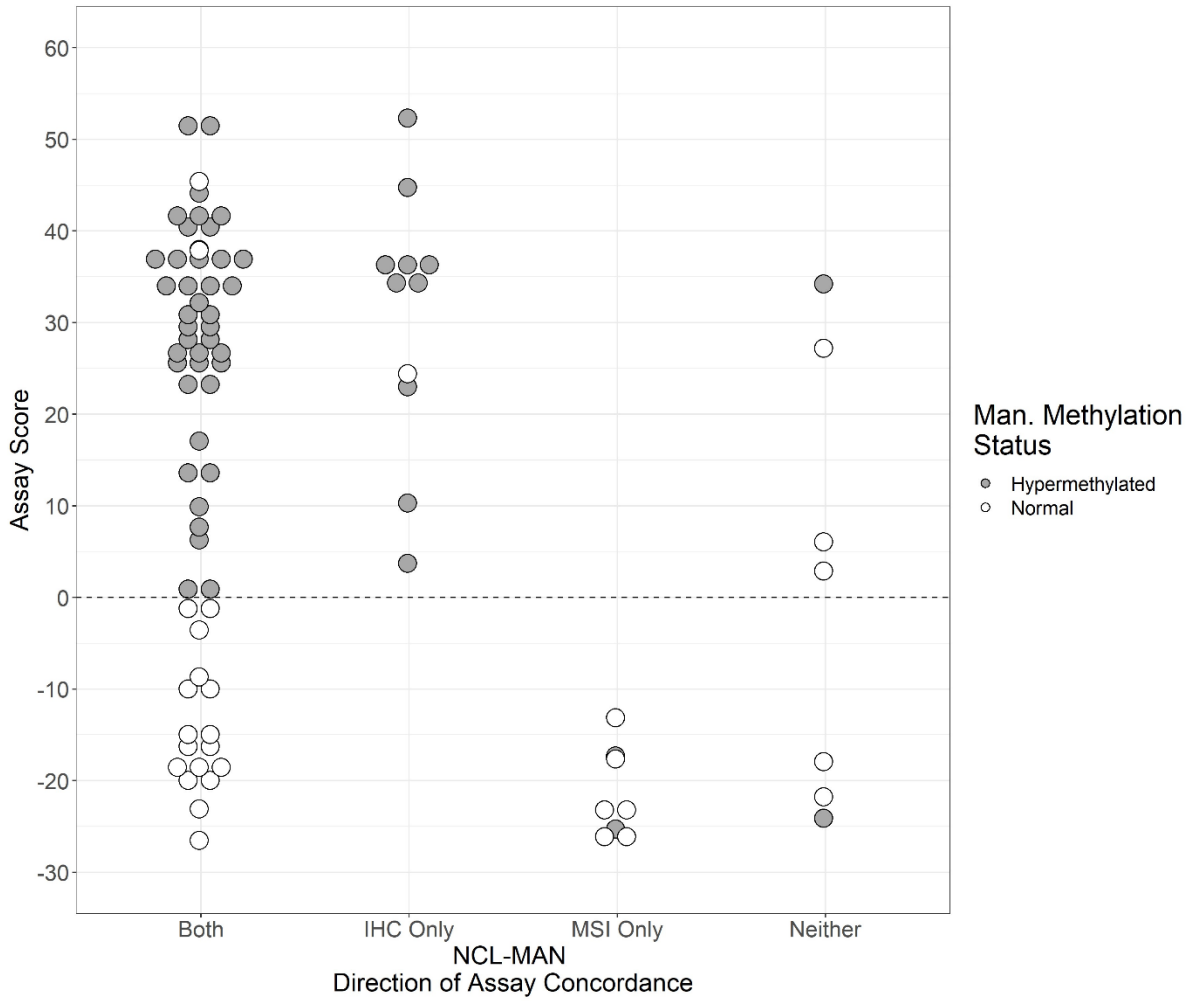

Figure 4.6. Methylation status of sporadic EC tumours versus the direction of concordance between the NCL MSI Assay scores and the original results of IHC and MSI testing. Separation along the x-axis delineates the direction of concordance between the Newcastle Assay and previous analyses conducted by Manchester for the same samples, whereby the left-most column represents agreement with both prior IHC and MSI assessments, followed by agreement with Manchester IHC only, Manchester Promega MSI testing only, and finally neither. The dashed y-intercept at y=0 defines the classification threshold, positive values above which result in an MSI-H designation, while negative values below such are designed MSS. Finally, filled points represent hypermethylated samples, while unfilled points denote those with no methylation.

4.4.4. MSI Determination Using the Newcastle Assay with Manchester EC Samples is Reproducible

To confirm the findings of the Newcastle Assay for the Manchester EC samples, specifically where disagreement with previous assays was found, repeat analysis of samples found to be discordant with one or both existing assay results was performed. This was undertaken both after the initial analysis of the test set to eliminate technical error as a contributing factor to the low discordance, and after the test set was unblinded, before the data of both sets was ultimately combined for a summative review. For this purpose, distinct curl removal, DNA extraction, amplification and sequencing was undertaken for each of the 47 originally-discordant samples, with this allowing for the rejection of technical error or variable tumour cell content between repeats as an explanation for results. All samples progressed through the same established assay pipeline as used in the initial review, with only five of these producing insufficient product in multiple MIP amplification attempts to proceed to sequencing. In the analysis of the remaining 42, 85.71% (36/42) returned the same result with regards to MSI classification between their first and second assessments, giving a notable correlation between repeats with an R value of 0.89 (Figure 4.7.).

Considering the six samples with conflicting classifications between analyses, four are MMRd samples confirmed by IHC (three with *MLH1/PMS2* loss and one with *PMS2* loss), with the remaining two being repair proficient by IHC. Only one of the aforementioned *MLH1/PMS2* samples has previously been assessed for methylation, and was found to be ‘normal’ in this regard, indicating that the methylation status of these samples does not impact upon this analysis. Of further note is that three of these samples, including two of the aforementioned MMRd instances (both *MLH1/PMS2*-deficient), had one of their two scores within five units of the established classification threshold (0). With it being unclear how small fluctuations in the variant allele frequencies (VAFs) of the microsatellites analysed may affect a score returned by the Newcastle Assay, this is important as it is suggests minor changes in instability may alter the classification of a given sample, and therefore the concordance of results between repeats.

Ultimately, for samples with contrasting classifications between their repeat analysis, a third independent analysis was conducted with a majority rule applied between all three assignments to give the final classification for those samples used in subsequent assessments and comparisons.

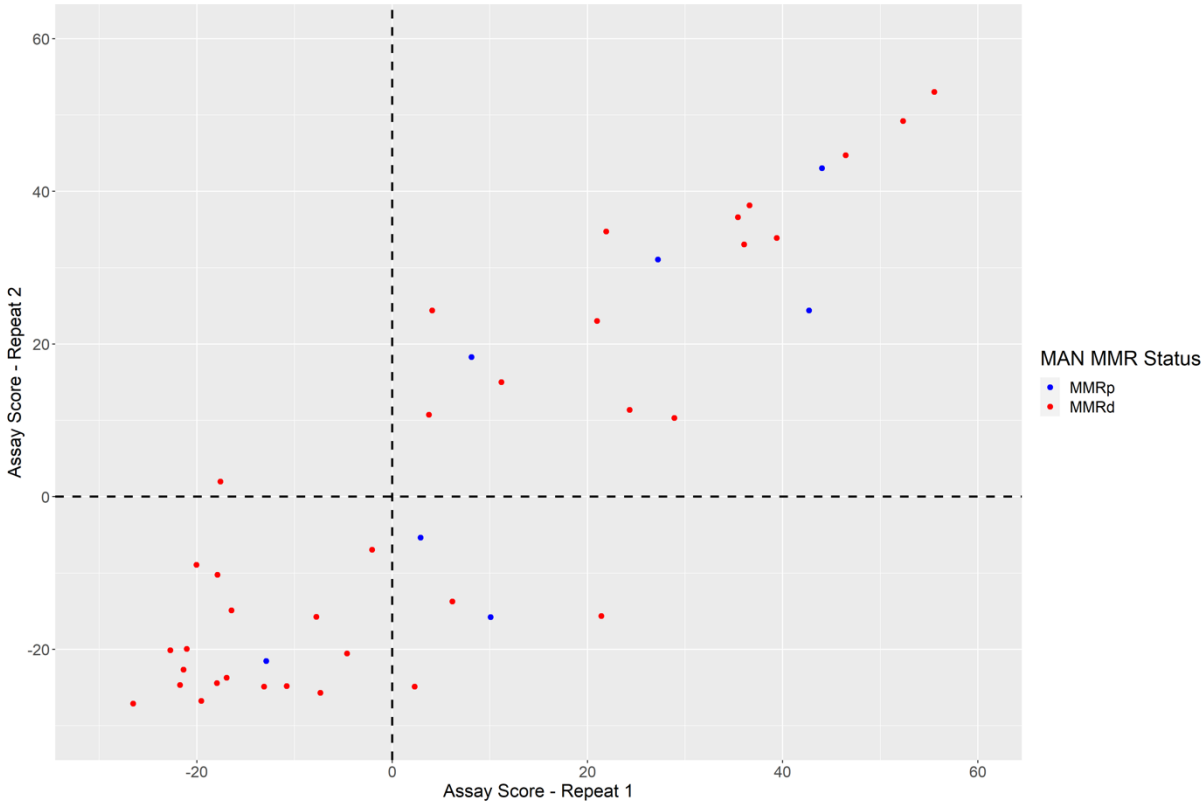

Figure 4.7. Comparison of MSI Scores for Manchester endometrial cancer samples analysed twice by the Newcastle MSI Assay using its original 24 MNR marker panel and a CRC-trained classifier. The dashed x-intercept at x=0 defines the classification threshold of the first analysis, while the dashed y-intercept at y=0 defines the classification threshold of the second analysis, positive values above which in both result in an MSI-H designation.

4.4.5. An Expanded Marker Panel Fails to Significantly Improve Concordance of the Newcastle Assay with Existing Analyses

Following the initial analyses of the Ohio and Manchester EC cohorts, there exists the question of why MMRd detection is so dissimilar between the two, and the sensitivity so significantly reduced in the latter. In the first instance, to determine if these findings were the result of the markers being used for assessment, and if concordance could be improved by increasing the sensitivity of the MSI assay, both Ohio and Manchester samples were analysed using a larger, and more sensitive marker panel. For this review, in place of the original 24 marker panel formerly used with this, an alternative selection of 62 markers was used (Detailed in Section 2.5.1.). Briefly, these markers have been demonstrated to be highly unstable within somatic tissues of CMMRD patients, and more discriminating than the original 24 MNR marker panel of the Newcastle Assay, producing a greater separation between MSI-H and MSS samples (Gallon *et al.*, 2023). This ultimately allowed for changes in MMRd detection in EC through use of an alternative marker panel to be determined, as well as deficiencies in the sensitivity of the original markers in this tumour type to be substantiated.

Of the 196 samples involved in the preceding analysis of the Ohio cohort, 182 were included in this analysis, with the other 14 having insufficient purified DNA material remaining to allow processing. Relative to existing IHC results, this assessment by the Newcastle Assay yielded a test sensitivity of 95.79% and specificity of 98.85% in the detection of repair deficiency, both slightly higher than the corresponding values from analysis with the original marker panel (92.93% (p=0.42, 95% CI=0.90-0.99) and 97.93% (p=1.00, 95% CI=0.94-1.00) respectively). Both values however exceed, or were equal to, the equivalent values from prior MSI evaluation of the same material, where a sensitivity and specificity of 94.74% and 98.85% respectively were observed. Of the four samples returning false-negative classifications, one represents a situation of *MLH1/PMS2* deficiency (Score= -64.52), while the other three were instances of *MSH6* loss (Scores= -43.62, -53.22 and -69.58) also not recognised in previous MSI testing. Overall, a 96.70% concordance with previous IHC classifications was returned (Table 4.7.), an increase, albeit insignificant, on that observed between the Ohio IHC and MSI testing results obtained for these samples (96.15% (p=0.85, 95% CI=0.93-0.99)).

This finding also represents a negligible increase on the concordance realised in the initial analysis of Ohio EC samples using the original 24 MNR marker panel

(94.90%, p=0.40). Nevertheless, these concordance results, and those describing assay accuracy, are effectively comparable to the results observed in the earlier analysis of the Ohio cohort with the original marker panel. No significant improvements in performance relative to existing IHC are realised by the use of the different marker panel, although such was not unexpected for this cohort given the high levels of concordance, sensitivity and specificity seen previously.

		Concordance of Newcastle MSI Assay with Previous Ohio Assay Results				
		Both Ohio IHC and MSI Analyses	Ohio IHC Only	Ohio MSI Only	Neither Ohio Analysis	% Concordance Ohio IHC and Ohio MSI
Ohio IHC	MMRp	99	-	1	-	99
		(86 / 87)		(1 / 87)		(86 / 87)
	MMRd	93	2	4	1	94
		(88 / 95)	(2 / 95)	(4 / 95)	(1 / 95)	(89 / 95)

Table 4.7. Percentage concordance of the Newcastle MSI Assay, employed with its 'Version Two' panel of 62 markers, with the results of previous analyses from Ohio, separated by repair situation as determined by IHC. Numbers of concordant results per group are shown in parentheses.

With 10 samples generating insufficient product in MIP amplification to proceed to sequencing, in addition to the five damaged in the curl removal process, 176 of the 191 EC samples initially supplied by Manchester were further assessed using the improved marker panel. Relative to existing IHC findings, this review by the Newcastle Assay achieved a test sensitivity of 64.95%, a value significantly lower than the corresponding metric from analysis with the original marker panel (77.17% (p=0.0073, 95% CI=0.55-0.74)), but still higher than the 56.70% sensitivity returned in Promega MSI testing of the same samples (p=0.12). This assessment further realised a specificity of 91.14%, an increase on the level obtained in the first summative study of the Manchester cohort (89.33% (p=0.72, 95% CI=0.83-0.96)), but decrease on that resulting from Promega MSI testing of the same samples (93.67% (p=0.35)). As with the original marker panel, both accuracy figures describe an inferior performance with

this cohort than observed with the samples from Ohio, being significantly different with regards to sensitivity (p=< 2.2e-16) and specificity (p=3.7e-5).

Overall, a 76.70% concordance with existing IHC appraisal was returned in this evaluation by the Newcastle Assay (Table 4.8.), a negligible increase on that found between the Ohio IHC and MSI testing results of the same samples (73.30% (p=0.35, 95% CI=0.70-0.83)). In realising these levels of concordance, the Newcastle Assay failed to identify 34 confirmed MMRd samples, 25 of which were also undetected by Promega MSI testing, with 21 false-negative classifications assigned to instances of *MLH1/PMS2* loss (Scores=-21.87 - -71.47), and eight to instances of *MSH6* loss (Scores=-9.29 - -72.72). Given the considerable discordance between previous assay results for this cohort, this disagreement from the Newcastle Assay is not unexpected. However, this result still represents an increase on the concordance observed with assessment using the original marker panel (82.36% p=0.46), albeit significantly inferior to the corresponding value derived in the analysis of the Ohio EC cohort (96.70% (p=< 2.2e-16)).

Ultimately, these findings fail to provide evidence that insufficient marker sensitivity is responsible for the previous levels of discordance, and inaccuracy relative to IHC observed for the Manchester cohort, or that MMRd detected in EC could be significantly improved by use of a more sensitive MSI marker panel. Seemingly, the default marker panel of the Newcastle Assay is sufficiently sensitive for this analysis of EC.

		Concordance of Newcastle MSI Assay with Previous Manchester Assay Results				
		Both Manchester IHC and MSI Analyses	Manchester IHC Only	Manchester Promega MSI Only	Neither Manchester Analysis	% Concordance Man. IHC and Man. MSI
Manchester IHC	MMRp	89	3	4	5	94
		(70 / 79)	(2 / 79)	(3 / 79)	(4 / 79)	(74 / 79)
	MMRd	51	14	29	6	61
		(49 / 97)	(14 / 97)	(28 / 97)	(6 / 97)	(59 / 97)

Table 4.8. Percentage concordance of the Newcastle MSI Assay, employed with its 'Version Two' panel of 62 markers, with the results of previous analyses from Manchester, separated by repair situation as determined by IHC. Numbers of concordant results per group are shown in parentheses.

4.4.6. Use of an EC-Trained Classifier Fails to Improve Sensitivity and Specificity in MMRd Determination by the Newcastle Assay

With the use of alternative markers explored, further efforts were made to assess the effects of classifier retraining on the detection of MMRd in EC by the Newcastle Assay. In its default arrangement, high quality CRC material is used in training the Newcastle Assay's classifier, and establishing classification thresholds. However, efforts were made to use EC samples for this purpose instead, with this theoretically instituting thresholds specific to the tumour type, and potentially accounting for EC-specific biology with regards to MSI. It was hypothesised that training with tissue appropriate samples could improve assay concordance, and thus performance, with IHC, the current 'gold-standard' approach for MMRd detection in EC. For this assessment, the designated training samples of each EC cohort, once reviewed by the Newcastle Assay, were used to train the assay's classifier, before classification of the corresponding validation cohort prior to its unblinding: analysis of the latter was ultimately performed following unblinding.

Considering the Ohio samples initially, high concordance with prior analyses, as well as sensitivity and specificity relative to IHC, had been observed in previous assessments by the Newcastle Assay, so little or no change in results was anticipated from classifier retraining. In this review, all 96 samples of the designated training set were used to classify 86 of the designated validation samples, with the remaining 14

of the latter set failing assignment by this novel approach. This analysis by the retrained Newcastle Assay gave a 94.19% concordance with existing IHC results (Table 4.9.), a level marginally lower than the concordance present between the Ohio IHC and MSI results for these samples (96.51% (p=0.23, 95% CI=0.87-0.98)). Only three samples returned false-negative results in this review, with all representing instances of *MSH6* deficiency also undetected in earlier MSI testing (Scores=-7.77, -13.06 and -14.99). Ultimately, this performance by the Newcastle Assay returned a sensitivity of 93.75% and specificity of 94.74% relative to previous IHC findings, neither significantly removed from the corresponding values generated in the analysis of the Ohio validation samples using a CRC-trained classifier. What is more, these values were comparable to those observed in previous MSI evaluation,

equal in regard to sensitivity, but lower in regard to specificity (100.00% (p=< 2.2e-16, 95% CI=0.82-0.99)). As expected, given the findings of prior analyses of this cohort, these performance metrics are high and comparable to those from previous assessment of the Ohio samples, with this cohort affording limited opportunity to observe improvements resulting from classifier retraining.

		Concordance of Newcastle MSI Assay with Previous Ohio Assay Results				
		Both Ohio IHC and MSI Analyses	Ohio IHC Only	Ohio MSI Only	Neither Ohio Analysis	% Concordance Ohio IHC and Ohio MSI
Ohio IHC	MMRp	95	-	-	5	100
		(36 / 38)			(2 / 38)	(38 / 38)
	MMRd	94	-	6	-	94
		(45 / 48)		(3 / 48)		(45 / 48)

Table 4.9. Percentage concordance of the Newcastle MSI Assay, executed with an EC-retrained classifier, with previous analyses results from Ohio, separated by repair situation as determined by IHC. Numbers of concordant results per group are shown in parentheses.

In the review of Manchester ECs by a retrained Newcastle Assay, all 86 validations samples were reclassified with a classifier developed using all 80 samples of the designated training dataset. This analysis returned a 79.07% concordance of Newcastle classifications with those of previous IHC (Table 4.10.), a reduction on the values observed in previous analysis of Manchester validation samples using a CRC-trained classifier (80.46% (p=0.79, 95% CI=0.67-0.87)), and between Manchester IHC and Promega MSI results for the same samples (81.40% (p=0.58, 95% CI=0.69-0.87)). The retrained Newcastle Assay returned false-negative assignments for a total of 14 samples, including four *MSH6*-deficient ECs (Scores=-2.24 - -4.49), with only three of these appropriately detected by Promega MSI testing. Ultimately, this performance by the Newcastle Assay returned a test sensitivity of 72.00% and specificity of 88.89% relative to previous IHC findings. The former represents a value lower than that seen in the earlier analysis using a CRC-trained classifier for this dataset (78.00% (p=0.31, 95% CI=0.58-0.84)), while the latter is negligibly higher than its corresponding metric from this (83.78% (p=0.5, 95% CI=0.74-0.97)). This level of sensitivity does however compare with the corresponding value from Promega MSI assessment of the same material (68.00% (p=0.65)), an evaluation that also returned 100% specificity in classification. As with the Ohio cohort, these findings, and the reduced Assay performance observed following retraining, are not entirely expected. These results provide no evidence that tissue specific classifier training has a significant impact upon assay performance in this instance.

		Concordance of Newcastle MSI Assay with Previous Manchester Assay Results				
		Both Manchester IHC and MSI Analyses	Manchester IHC Only	Manchester Promega MSI Only	Neither Manchester Analysis	% Concordance Man. IHC and Man. MSI
Manchester IHC	MMRp	89	-	-	11	100
		(32 / 36)			(4 / 36)	(36 / 36)
	MMRd	62	10	22	6	74
		(31 / 50)	(5 / 50)	(11 / 50)	(3 / 50)	(37 / 50)

Table 4.10. Percentage concordance of the Newcastle MSI Assay, executed with an EC-retrained classifier, with previous analyses results from Manchester, separated by repair situation as determined by IHC. Numbers of concordant results per group are shown in parentheses.

4.4.7. Cross-cohort Analysis Suggests Systematic Differences in Sample Quality

A final investigation into the effects of classifier retraining involved cross-cohort analysis, whereby the samples of Ohio were used to classify those from Manchester and vice versa. Through this evaluation, the effects of using a highly concordant sample set (Ohio) in classifier training, to classifier a cohort for which reduced test sensitivity and specificity has previously been returned, could be observed. Conversely, the effects on classification of using a discordant sample set (Manchester) to grade an otherwise concordant population could also be determined.

In the first instance, Ohio samples were used to retrain the Newcastle Assay's classifier for analysis of the Manchester EC samples, with all 196 of the former (from across both training and validation cohorts) used to review all 167 of the latter. This analysis returned a test sensitivity of 84.78% relative to existing IHC findings, a value moderately higher than that afforded in the initial analysis of this cohort (77.17% (p=0.083, 95% CI=0.76-0.91)), and a significant increase on that of Promega MSI testing of the same material (61.96% (p=2.47e-06)) However, the specificity of this test was only 52.00%, a significant decrease on the corresponding values from the initial analysis of the cohort (89.33% (p=4.688e-16, 95% CI=0.40-0.64)), and Promega MSI evaluation of this material (94.67% (p=< 2.2e-16)).

Considering the agreement of Newcastle classifications with those of previous analyses for the Manchester samples, use of the Ohio EC cohort in classifier development gave a concordance with IHC results of 70.06% (Table 4.11.). This value is significantly reduced from both that observed in the initial analysis of the Manchester cohort (82.63% (p=5.836e-05, 95% CI=0.63-0.77)), as well as the concordance between IHC and Promega MSI results for these samples (79.04% (p=0.0057)).

		Concordance of Newcastle MSI Assay with Previous Manchester Assay Results				
		Both Manchester IHC and MSI Analyses	Manchester IHC Only	Manchester Promega MSI Only	Neither Manchester Analysis	% Concordance Man. IHC and Man. MSI
Manchester IHC	MMRp	51	1	4	44	95
		(38 / 75)	(1 / 75)	(3 / 75)	(33 / 75)	(71 / 75)
	MMRd	60	25	12	2	66
		(55 / 92)	(23 / 92)	(12 / 92)	(2 / 92)	(61 / 92)

Table 4.11. Percentage concordance of the Newcastle MSI Assay, executed with a classifier trained using the Ohio EC cohort, with previous analyses results from Manchester, separated by repair situation as determined by IHC. Numbers of concordant results per group are shown in parentheses.

An additional notable consequence of retraining is realised in the average assay score assigned to both repair proficient and deficient samples, and the 'distance' of these from the classification threshold, compared to the assessment by the original Newcastle Assay pipeline. In the initial analysis of this cohort, with a CRC-trained classifier in operation, the average assay score of confirmed MMRd samples was 26.25. This increases significantly to 37.73 with use of the Ohio EC-trained classifier (p= 1.75e-05). A similar shift is observed for MMRp samples, for which the average scores also significantly increases from -14.83 to -0.71 (p= 9.75e-13) (Figure 4.8.).

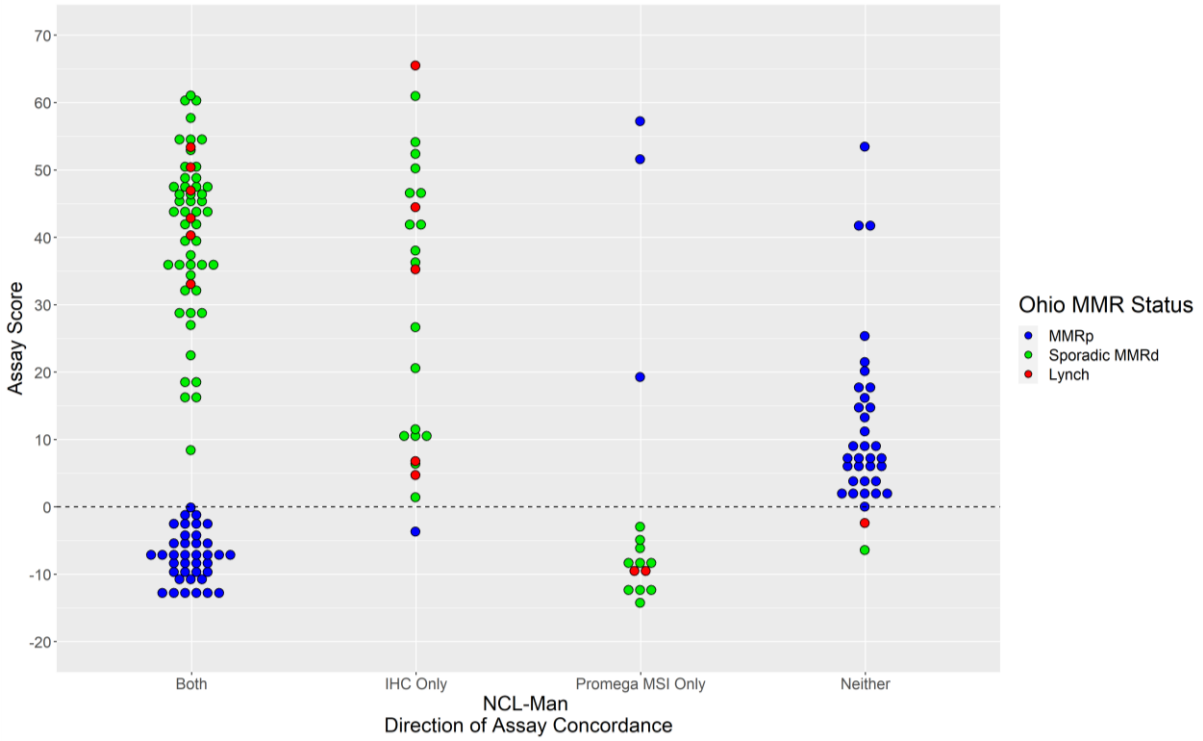

Figure 4.8. MSI Score Distributions for all Manchester endometrial cancer samples when analysed by the Newcastle MSI Assay using its original marker panel and classifier trained using the Ohio EC cohort. Separation along the x-axis delineates the direction of concordance between the Newcastle Assay and previous analyses conducted by Manchester for the same samples, whereby the left-most column represents agreement with both prior IHC and MSI assessments, followed by agreement with Manchester IHC only, Manchester Promega MSI testing only, and finally neither. The dashed y-intercept at y=0 defines the classification threshold, positive values above which result in an MSI-H designation, while negative values below such are designed MSS. Finally, coloration is applied according to sample MMR situation as determined by IHC.

Classifier retraining using the Manchester cohort for the analysis of the Ohio EC samples was not expected to realise significant changes in assay performance given the high sensitivity, specificity and concordance relative to IHC observed in prior analysis of this cohort by the Newcastle Assay. In this analysis a test sensitivity of 95.92% relative to previous IHC findings was returned, representing a small increase on that from the initial analysis of this cohort (92.03%, p=0.32, 95% CI=0.90-0.99), but negligible decrease on the value observed in MSI testing of the same samples (94.90%, p=0.82). Furthermore, this assessment gave a specificity in MMRd detection of 96.94%, lower than procured in Promega MSI testing of these samples (97.96%, p=0.46, 95% CI=0.91-0.99), but slightly higher than the levels seen in the initial

analysis of this cohort (97.94%, p=0.46). As anticipated, from these insignificant changes, it appears the use of the relatively discordant Manchester cohort in classifier retraining does not adjust classification thresholds sufficiently to affect assay performance, and compromise the classification of the highly concordant Ohio samples.

Considering the agreement of Newcastle classifications with those of previous analyses for the Ohio samples, use of the Manchester cohort in classifier training gave a concordance with previous IHC results of 96.43% (Table 4.12.). This value is a marginal increase on the concordance returned in the initial analysis of the Ohio cohort (94.90%, p=0.42, 95% CI=0.93-0.99), but equal to the agreement between the existing IHC and MSI results for these samples. Nevertheless, testament again to the high concordance observed between analyses for this cohort, significant changes in this regard were not expected.

		Direction of Concordance of Newcastle MSI Assay with Previous Ohio Assay Results				
		Both Ohio IHC and MSI Analyses	Ohio IHC Only	Ohio MSI Only	Neither Ohio Analysis	% Concordance Ohio IHC and Ohio MSI
Ohio IHC	MMRp	97	-	2	1	98
		(95 / 98)		(2 / 98)	(1 / 98)	(96 / 98)
	MMRd	94	2	3	1	95
		(92 / 98)	(2 / 98)	(3 / 98)	(1 / 98)	(93 / 98)

Table 4.12. Percentage concordance of the Newcastle MSI Assay, executed with a classifier trained using the Manchester EC cohort, with previous analyses results from Ohio, separated by repair situation as determined by IHC.

The consequences of this retraining are however evident in the average assay score given to both MMRp and MMRd samples, and their collective separation from the designated classification threshold (compared to the assessment by the original Newcastle Assay arrangement). In the first review of this cohort, when a CRC-trained classifier was in effect, the average assay score of verified repair-deficient samples was 28.26. However, use of a Manchester EC-trained classifier significantly reduced this to 19.25 (p=1.70e-05). The inverse is true of repair-proficient samples where the average score accompanying this classifier adjustment significantly increased from -24.60 to -5.85 (p=<2.2e-16) (Figure 4.9.). As with classifier retraining using the Ohio cohort, these results indicate a lower in classification thresholds, with such established further from the average MMRd score, but closer to the average MMRp score. Nevertheless, this has not adversely affected the overall accuracy of classification.

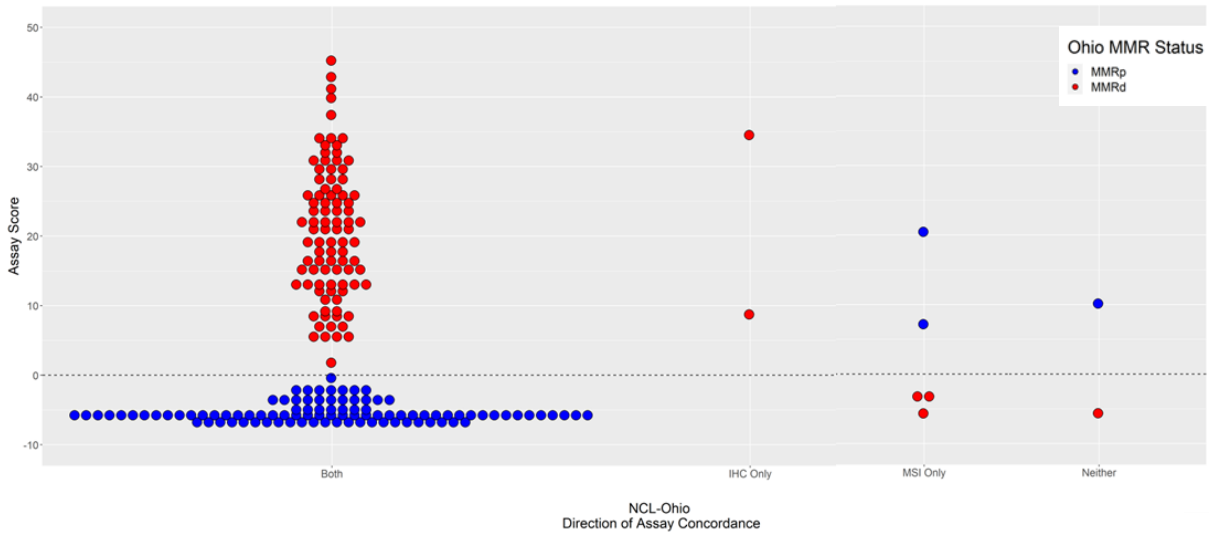

Figure 4.9. MSI Score Distributions for all Ohio endometrial cancer samples when analysed by the Newcastle MSI Assay using its original marker panel and classifier trained using the Manchester EC cohort. Separation along the x-axis delineates the direction of concordance between the Newcastle Assay and previous analyses conducted by Ohio for the same samples, whereby the left-most column represents agreement with both prior IHC and MSI assessments, followed by agreement with Ohio IHC only, Ohio MSI testing only, and finally neither. The dashed y-intercept at y=0 defines the classification threshold, positive values above which result in an MSI-H designation, while negative values below such are designed MSS. Finally, coloration is applied according to sample MMR situation as determined by IHC.

4.4.8. Contrasting LS Identification by the Newcastle Assay is Observed Between Ohio and Manchester EC Cohorts

A principal rationale for the use of IHC and MSI testing in EC is for the detection of LS, so the ability of different assays to detect tumours with this condition is of importance, and warrants separate review. In my analysis, individual consideration was therefore finally afforded to the detection of LS samples specifically within the Ohio and Manchester EC cohorts by the Newcastle Assay. From Ohio, a total of seven Lynch ECs were originally provided, with six of these progressing to analysis: the remaining sample, an instance of *MSH6* deficiency, repeatedly failed to generate sufficient product material in MIP amplification to proceed to sequencing. The original Newcastle Assay pipeline, with CRC-trained classifier and 24 MNR marker panel, detected five of these (83.33%) (Table 4.14.), failing to record a single sample with *MSH6* loss (Score=-1.11). This contrasts the percentage detection of sporadic MMRd instances for which 93.55% of cases were identified, albeit this difference is insignificant (p=0.33, 95% CI=0.36-1.00). Of further interest, the Newcastle Assay detected a single repair-deficient sample (with a pathogenic *MSH2* mutation confirmed by germline analysis) that IHC failed to identify: in effect this gives both assays equivalent performance for LS detection in this sample set. In ensuing analysis with the improved marker panel all LS cases were accurately classified, an achievement replicated in the subsequent classifier retraining assessments.

From Manchester, a total of 17 Lynch ECs were initially supplied, with 15 of these transitioning through the Newcastle Assay pipeline for eventual scoring analysis and classification. (The two samples failing to be processed represent, in one instance, a case of *MSH6/PMS2* deficiency for which the material was damaged in the precursory curl removal process, and in the other, a situation of *MSH6* deficiency for which inadequate product was realised in MIP amplification for ultimate sequencing.) The assessment of 14 of these 15 samples with the original Newcastle Assay pipeline saw eight detected and appropriately classified (57.14%) (Table 4.13.), with the six missed LS samples representing germline *MLH1/PMS2* deficiency (Scores=-20.83 and -22.81), *PMS2* deficiency (Score=-15.64) and, most frequently, *MSH6* deficiency (Scores=-14.83, -19.87 and -24.63). The percentage classification of LS samples with subsequent iterations of the Newcastle Assay varied slightly, but ultimately there was no significant improvement returned for this in most of these analyses. The one exception is observed following cross-cohort retraining, using the Ohio cohort for

classifier development, with 11/14 LS samples suitably identified, This represents a notable advanced on the corresponding values observed in the initial analysis of this cohort (p=0.17, 95% CI=0.49-0.95) and analysis using an improved marker panel (p=0.012), but comes at the expense of significantly less specific classifier conditions. In this final evaluation of LS cases specifically, the three samples missed represent an *MLH1/PMS2* deficient sample (Score=-9.83), and two examples of germline *MSH6* loss (Scores=-2.37 and -9.14).

Sample	Cohort	IHC	MSI	Original Assay Concordance	NCL MSI Score	NCL MSI Status	NCL LS Detection	Pathogenic MMR variant
ECT176	Ohio	MSH6 Loss	MSI-H	Y	13.1	MSI-H	Y	Not Available
ECT184	Ohio	MSH2 / MSH6 Loss	MSI-H	Y	15.7	MSI-H	Y	Not Available
ECV69	Ohio	MSH6 Loss	MSI-H	Y	17.7	MSI-H	Y	Not Available
ECV100	Ohio	MSH2 / MSH6 Loss	MSI-H	Y	16.9	MSI-H	Y	Not Available
ECT192	Ohio	MMR Proficient	MSI-H	N	39.2	MSI-H	Y	Not Available
ECT102	Ohio	MSH6 Loss	MSS	N	-1.1	MSS	N	Not Available
PET256	Man	MLH1/PMS2 Loss	MSI-H	Y	10.2	MSI-H	Y	MLH1 c.1409+1 G>C Exon 12
PET16 (M24859)	Man	MLH1/PMS2 Loss	MSI-H	Y	37.2	MSI-H	Y	MLH1 c.473delA p.(Asn158ThrfsTer2) Class 5
PET61	Man	MSH6/MSH2 Loss	MSI-H	Y	45.7	MSI-H	Y	MSH2 Exon 7 Deletion
PET215	Man	MSH6 Loss	MSI-H	Y	26.2	MSI-H	Y	MSH6 c.3004_3005delGG p.(Gly1002LeufsTer2) Class 5
PET173	Man	PMS2 Loss	MSI-H	Y	28.2	MSI-H	Y	PMS2 Del Exon 9-10
PET213	Man	MSH6 Loss	MSI-H	Y	30.3	MSI-H	Y	MSH6 c.2731C>T p(Arg911Ter)
PET255	Man	MSH6 Loss	MSS	N	11.4	MSI-H	Y	MSH6 c.2731C>T p.(Arg911Ter)
PET31	Man	MSH6 Loss	MSS	N	-19.1	MSS	N	MSH6 c.2731C>T p.(Arg911Ter)
PET128	Man	MSH6 Loss	MSI-L	N	-24.6	MSS	N	MSH6 c.3313G>T p.(Gly1105Ter)
ID5146 (BRC165)	Man	MSH6/MSH2 Loss	MSS	N	53.0	MSI-H	Y	MSH2 Exon 1 to Exon 8 Deletion
B000000882	Man	PMS2 Loss	MSS	N	-15.6	MSS	N	Homozygous PMS2 c.1500delC
PET241	Man	MSH6 Loss	MSI-H	Y	-14.8	MSS	N	MSH6 c.2731C>T p.(Arg911Ter)
ID66986 PREC08 (H15-4324)	Man	MLH1/PMS2 Loss	MSS	N	-20.8	MSS	N	MSH6 c.-118G>A p(Ala40Thr)
PET101	Man	MLH1/PMS2 Loss	MSS	N	-22.8	MSS	N	UV MSH2 c.2120G>A p.(Cys707Tyr) Class 4

Table 4.13. Details of the tumours with known pathogenic germline MMR gene mutations (LS/CMMRD) across both Ohio and Manchester cohorts. Results of prior IHC and MSI testing are shown alongside assay concordance, with the specifics of MMR gene variants only available for the Manchester cohort.

Previous work by our group has indicated that the mononucleotide mutation rate is lower in the blood of patients with constitutional *MSH6* deficiency compared to that of patients with equivalent deficiency in other MMR genes (Gallon *et al.*, 2023), while a reduced sensitivity of MSI analyses in detecting tumours lacking *MSH6* expression has also been reported (Gatius *et al.*, 2022). To investigate further our inadequacies in the detection of samples with germline *MSH6* loss, a comparison was performed of the VAFs for samples in which IHC had previously detected an MSH6 expression defect with those in which the loss of expression of other MMR genes had been found. For comparative purposes, I also included the data from samples that have lost the expression of both MSH2 and MSH6 concurrently. In these tumours, the primary defect probably involves MSH2, with the loss of MSH6 expression likely to result from the destabilisation of the MSH2/MSH6 heterodimeric complex. In this analysis, no single marker achieved statistical significance between the groups compared. However, as presented in Figure 4.10., markers in the samples with the isolated loss of MSH6 expression have lower variant allele frequencies than samples without MSH6 involvement ($p=6.1x10^{-15}$ (Ohio ECs); $1.7x10^{-12}$ (Man. ECs)). In contrast, the opposite is observed for samples with the lost expression of both MSH2 and MSH6 ($p=3.8x10^{-3}$ (Ohio ECs); $2.1x10^{-16}$ (Man. ECs)).

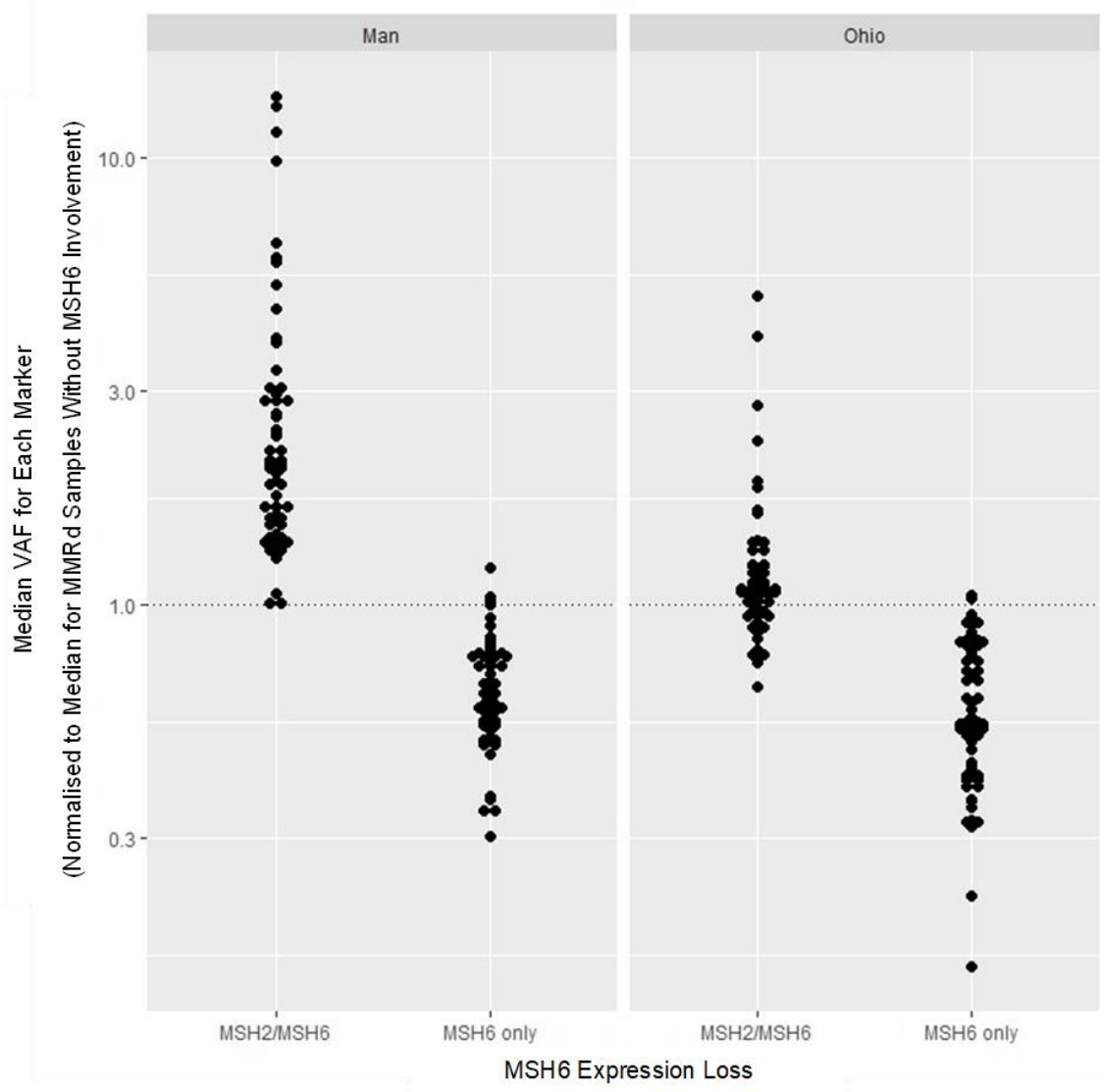

Figure 4.10. Comparison of the Median VAF for Each Marker of the Newcastle Assay Between EC Samples with Concurrent MSH2/MSH6 Loss and those with Isolated MSH6 Loss.

As the number of samples with isolated MSH6 expression defects, or individuals with LS, is relatively low across the EC cohorts analysed, it was supposed that these results may not reflect the general characteristics associated with this deficiency. Attention was therefore given to replicating this analysis using CRC samples. Figure 4.11. presents the comparison of VAFs for samples from patients with germline MSH6 defects to samples with germline defects in other MMR genes. Once again, a significant excess of markers with a lower median VAF is returned for samples with MHS6 defects ($p=4.5x10^{-7}$). While the difference in median VAFs is much less pronounced within CRCs, it does suggest that this is not a tumour specific

phenomenon, and that the differences observed for the ECs previously analysed do not reflect the use of an unrepresentative sample set.

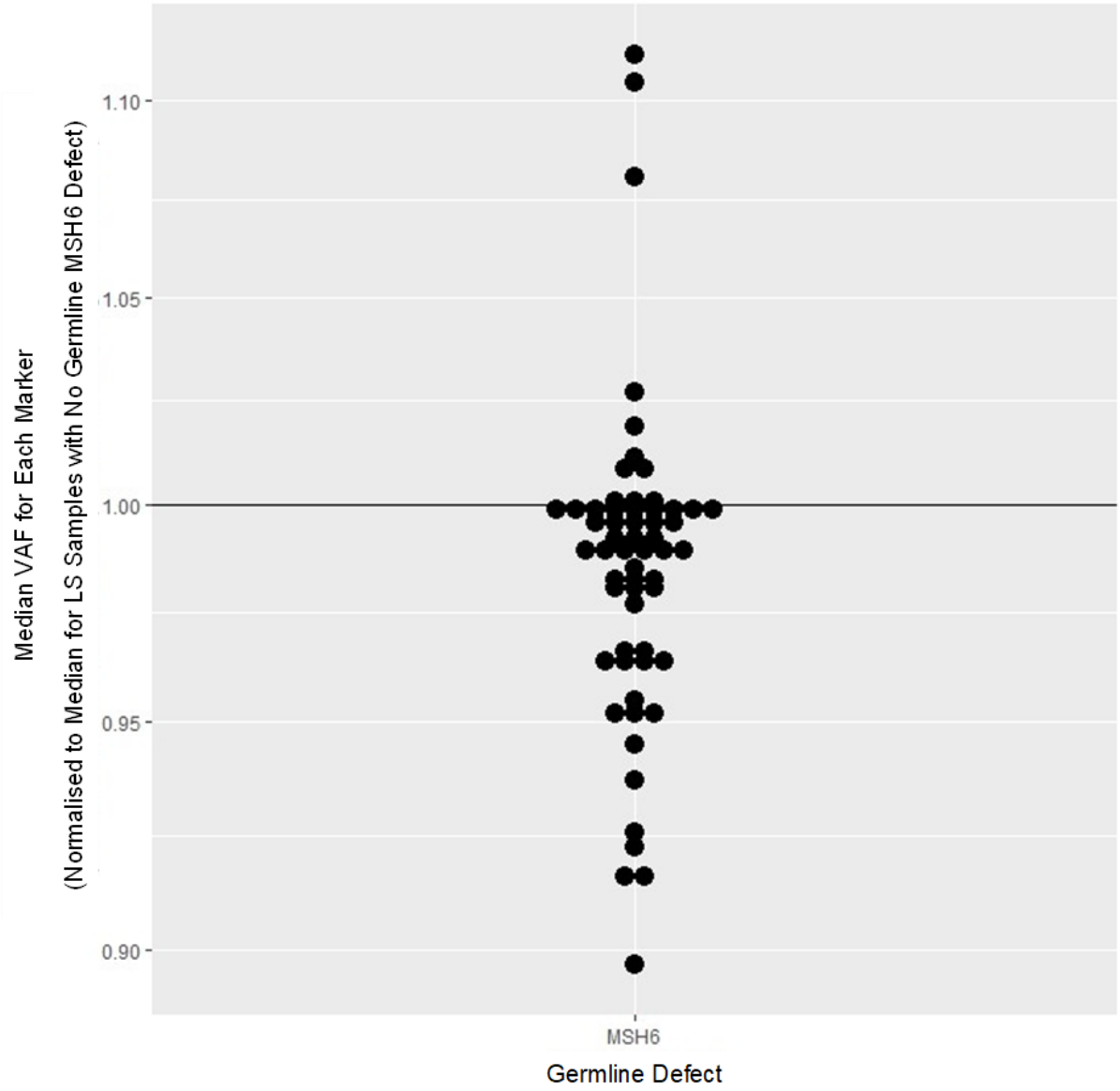

Figure 4.11. Distribution of the Median VAFs for Each Marker of the Newcastle Assay, Normalised to the Median for LS Samples with no Germline MSH6 Defect, for CRC Samples with Isolated MSH6 Loss.

4.5. Discussion

These results relate efforts to deploy the Newcastle Assay for the identification of MMRd in EC, using high levels of microsatellite instability as a surrogate for loss of MMR. In its original configuration, analysing the collective material provided by Ohio State University, this returned levels of MMRd detection comparable to that from Promega MSI testing of the same samples, as well as those observed in separate studies of other EC cohorts (McConechy *et al.*, 2015; Stelloo *et al.*, 2017). These results seemingly show MSI testing by the Newcastle Assay affords a highly-proficient approach, and an alternative to the more expensive and expert-dependent method of IHC for analysis of this tumour type, with a sensitivity and specificity of ~95% for triaged tumour material.

Concordance with previous IHC and MSI analyses performed in Ohio is relatively high (94.90% and 97.45% respectively), a non-significant deviation from the concordance between the two original results themselves for the same samples (96.43%). This is not unexpected, with high degrees of concordance to IHC having been reported when using MSI tests for MMRd detection in EC (McConechy *et al.*, 2015; Stelloo *et al.*, 2017). It stands to reason that, where one MSI assay is capable of performing with such precision to IHC, an assay for the same phenotype, particularly one with a higher sensitivity, would achieve equivalent levels of detection.

Nevertheless, instances of incorrect sample classification, and differing scoring distributions, exist for this cohort. Such evidences that, at least for this assay using these conditions, this is not a completely accurate biomarker for MMRd in EC. Attempts to correct misclassifications, by means of altering marker number or assay conditions, consistently returned comparable results to the initial analysis, but failed to completely eradicate conflicting calls relative to IHC findings. In the first instance, use of a higher sensitivity marker panel returned a marginally improved sensitivity and specificity, as well as concordance with the existing IHC findings. Subsequent classifier retraining returned similar observations with regards to specificity, but were accompanied by negligible decreases in sensitivity. However, given the high performance, and high levels of concordance, associated with the initial analysis, significant changes in classification were not expected with assay adjustments, and conclusions regarding the effect of various conditions of the assay on classification are ultimately limited. At least for this cohort, the established Newcastle Assay appears

sufficiently sensitive for the discrimination of repair-proficient and -deficient EC samples, and for high levels of MMRd detection.

Contrasting results were returned for the Manchester cohort of ECs, derived from an unselected patient population including material from patients recruited with early indications of endometrial abnormalities. Use of the Newcastle Assay with these samples returned a lower accuracy than observed for the Ohio samples, albeit an improvement on the concordance between Promega MSI and IHC obtained by Manchester. While this may relate to the samples represented, and the unselected nature of this referred cohort, it may also reflect the FFPE form in which these samples were provided, and how the Newcastle data was not produced from adjacent curls to those analysed by Manchester. Nevertheless, these results suggest that, at least for this cohort, the Newcastle Assay is more vulnerable to variations in instability, possibly through the assessment of loci that allow for earlier MSI detection.

Similar classification failings exist for the Manchester EC samples with regards to an inability to detect MSI in instances with confirmed MMR deficiencies as with the Ohio samples, albeit in greater number. This was ultimately not rectified by the use of an improved marker panel, where significant decreases in sensitivity and specificity are observed, accompanied by a marginal increase in the concordance with both existing analyses of Manchester. This suggests that the improved sensitivity of the novel marker panel is ineffective in enhancing discrimination between MMRp and MMRd samples in this cohort, and that the original markers, for the most part, are sufficiently sensitive for the detection of MSI in EC: it is not fundamentally a lack of sensitivity or assessing unsuitable loci that explains previous misclassifications in this cohort. Furthermore, as all markers were originally identified from the analysis of CRC material, it does not appear that different loci are affected, and a different marker panel is required, depending on tumour type.

Classifier retraining with ECs from this cohort similarly gave reduced sensitivity and specificity in sample classification, once more accompanied by a marginally increased concordance relative to the initial analysis, though to a lesser extent than observed with use of the alternative marker set. Not only does this imply that the tumour specificity of the classifier used is not responsible for previous inconsistencies, and that the use of analogous material for training is not required to improve classification (by accounting for the specific biology of the subject tumour type). It also suggests that the use of material of confirmed MMR status, and concordance between

MSI and IHC analysis, of one cancer type is suitable for the classification of another, and can return highly accurate readings for triaged data sets. Notwithstanding a marginally improved concordance in this situation, three *MSH6*-deficient tumours returned an MSS classification, again giving credence to a potential biological explanation for stability in these tumours despite confirmed MMRd.

Analysing the methylation situation of Manchester samples (for which this information was available) revealed a potential explanation for the discordance observed between Promega MSI and the Newcastle Assay in earlier analyses. Of these samples, a significant proportion were regarded as hypermethylated. Many of these were MSI-H by the Newcastle Assay, considerably more so than were detected by Promega MSI. The concordance of the former with existing IHC findings suggests that those identified by Newcastle are definite instances of repair deficiency missed by Promega MSI testing. Conceivably a result of lower levels of promoter methylation, this substantiates that the different MSI methods are not equivalent (either biologically or for technical reasons), and demonstrates the ability of the Newcastle Assay to detect deficiency where existing MSI techniques fail. The assay also identified a single hypermethylated samples that both Manchester IHC and Promega MSI failed to detect, further signifying its capacity to identify instability.

In general, where the Newcastle Assay differs from the results of IHC for the Manchester cohort, these are EC samples without hypermethylation where Promega MSI is also discordant, and/or frequently cases of *MSH6* deficiency. This again suggests there are situations were MSI is not a practicable biomarker for MMR deficiencies, and supports previous observations for when this discordance may be an issue (Gatius *et al.*, 2022). False positive situations, inappropriately designated MSI-H by both MSI assays, may however describe functional loss, and a resulting instable phenotype from defective DNA repair, without the loss of the protein motif recognised by IHC. These situations, where truncating or missense mutations in MMR genes have resulted in impaired function without complete protein loss, have previously been described, albeit in CRC (Salahshor *et al.*, 2001; Wahlberg *et al.*, 2002; McCarthy *et al.*, 2019). Incorrect classifications of this nature may alternatively be explained by mutations in POLE, a gene encoding the catalytic subunit of DNA polymerase epsilon, in which missense variants are associated with an ultramutated form of EC: instances of MSI-H with retained MMR protein expression have been reported in POLE exonuclease-domain variants (Stelloo *et al.*, 2017). Conversely, the

false-negative cases of MSH6 loss without exceptionally-high levels of MSI may represent situations in which the loss of the MMR gene is partially compensated for by the presence of MSH3. This MutS homologue, which dimerises with MSH2 to form a repair complex with specificity for larger indels, has been shown to provide some redundancy in the MMR system, and attenuate deleterious *MSH6* mutations (Acharya *et al.*, 1996; Risinger *et al.*, 1996). Inaccuracies in classification in either direction are likely not attributable to the presence of sub-clonal populations within a given tumour sample, as such have been shown not to have an effect on the concordance between MSI and IHC in EC, and therefore MSI assay sensitivity relative to IHC results (Stelloo *et al.*, 2017).

An interesting insight into the effect of differences in sample quality is afforded by the cross-cohort analysis undertaken with the two EC cohorts. Specifically, when the Ohio samples were used to retrain the Newcastle Assay's classifier for analysis of the EC samples, significant decreases in assay specificity were observed (in conjunction with moderate increases in assay sensitivity). Seemingly accounting for this reduction, at least in part, is the increase in MMRp samples being inaccurately classified as repair deficient, and further concordant with neither IHC nor Promega MSI testing (right-hand column of Figure). In using the Ohio samples, a highly concordant cohort in which marker stability and instability are distinct, in classifier training, the thresholds for individual markers to be classified as unstable may be established at relatively low VAFs. With this being the case, more markers for a given sample may exceed these thresholds, and therefore more samples may receive an MSI-H classification, including samples considered repair proficient by both IHC and Promega MSI. This notion is supported by the increase in average assay score observed for both MMRp and MMRd samples with the use of the Ohio cohort in classifier training. With the classification threshold established comparatively lower than with a CRC-trained classifier, such is theoretically closer to the mean MMRp score, and further from the mean MMRd score, with it therefore more likely repair proficient samples exceed this hypothetical parameter.

Particularly evident with the Manchester cohort, of further interest are the levels of Lynch detection specifically by the various iterations of the Newcastle Assay. In the first instance, this is significantly lower than the level of sporadic MMRd detection,

suggesting a difference in the biology of deficiency between LS and sporadic deficiency, and the respective importance of MSI. This is decreased further still through changes to the analysis pipeline by way of alterations in the marker panel used (questioning the effectiveness of these in detecting LS in EC). However, eventual assessment following classifier retraining does return an improved percentage detection. This indicates that, for accurate Lynch discovery by the Newcastle MSI Assay, classifier training is necessary, possibly to account for a tumour-type-specific biology which may also explain differential incidence between cancers (Møller *et al.*, 2018). This does however come at the expense of a reduced assay specificity, with classification thresholds sufficiently adjusted for the incorrect designation of a significant number of MSS samples, and so may not be a practicable approach. Nevertheless, it ultimately appears that there is a fundamental difference in the biology of sporadic MMRd and Lynch ECs, a proposition supported by the findings of pathological difference between the two situations, including an absence of MSI-tumour features in the latter (Broaddus et al., 2006; Mills et al., 2014; Sloan, Moskaluk and Mills, 2017). The question nevertheless remains as to why the different classifier context gives improved Lynch detection in ECs, and what the biological difference, or differences, may be that define these contrasting situations of repair deficiency.

Chapter 5. Assessing the Use of MIPs for the Detection of Pathogenic *PMS2* Variants

5.1. Introduction

5.1.1. PMS2 and the Penetrance of Mutations in this MMR Gene

Consisting of 15 exons situated in the p22 region of chromosome 7, the post-meiotic segregation 2 (*PMS2*) gene encodes a nuclear protein of the mismatch repair (MMR) system. This protein, which heterodimerises with MLH1 to establish the MutLα complex, is involved in the sequence-independent repair of mutations arising from aberrant DNA replication and recombination (Jiricny, 2006; Kunkel & Erie, 2005). PMS2 contributes to this repair through its latent endonucleolytic function (Kadyrov *et al.*, 2006), with it introducing nicks into DNA strands for their eventual excision and correction (van Oers *et al.*, 2010). As with other MMR proteins, in the absence of this activity, DNA repair malfunction arises.

Epidemiological studies estimate carriers of MMR pathogenic variants to be relatively common (up to 1 in 279 in the general population), and, among these, *PMS2* variants are the most prevalent with a frequency of 1 in 714 (Win *et al.*, 2017). In LS patients, *PMS2* mutations have been associated with an increased cumulative risk of approximately 13% for both endometrial and colorectal cancer (ten Broeke *et al.*, 2018), as well as an unusually high prevalence of breast, stomach and prostate cancers compared to MLH1 and MSH2 families (Kasela et al., 2019). However, in spite of MMR-deficiency being observed with *PMS2* mutations, the penetrance of these is relatively low in comparison to defects associated with alternative MMR genes, and *PMS2* is infrequently reported as mutated in LS (Talseth-Palmer *et al.*, 2010). A recent comprehensive analysis of the Prospective Lynch Syndrome Database (PLSD) has revealed a cumulative risk at 70 years of 3% for CRC, 13% for EC, and 3% for ovarian cancer in *PMS2* mutation carriers (Møller, 2020). This compares to the higher values associated with *MLH1* mutation (53%, 35% and 11% respectively), *MSH2* mutation (46%, 47% and 17% respectively) and *MSH6* mutation (12%, 41% and 11% respectively) in the same analysis. It should however be remembered that all cases in the PLSD are in receipt of regular, high quality surveillance colonoscopy with the

removal of polpys: it is possible that *PMS2* LS patients are more likely to develop their CRC in the context of a developed adenoma, and their removal prevents most cancers from developing (ten Broeke *et al.*, 2018),

While the precise reason for the lower penetrance outlined above is unknown, several hypotheses have been proposed. These include possible differences in the extent to which each protein contributes to the MMR process (Johnson *et al.*, 2010), differences in the number of functions performed by each protein (Prolla *et al.*, 1998), and the potential for partial compensation within the MMR system by MLH3 (Chen *et al.*, 2005). Regardless of the specific mechanism by which a reduced penetrance of *PMS2* mutations is realised, this reality has resulted in suggestions that *PMS2* should not be considered a predisposing gene for cancer in the same way as other MMR genes (Møller, 2020).

The distribution of MMR gene variants in constitutional mismatch repair deficiency (CMMRD) contrasts with that of LS, as a majority of cases are accounted for by biallelic *PMS2* mutations (Wimmer *et al.*, 2014). In a study of 146 patients with CMMRD, 58% possessed biallelic *PMS2* mutations, with the remaining 40% equally distributed between *MLH1/MSH2* and *MSH6* biallelic cases (Wimmer *et al.*, 2014). Paradoxically, mutation carriers with more than one malignancy were reported highest in those with biallelic *PMS2* deficiency (42%), but lowest with biallelic *MLH1/MSH2* deficiency (22%). The difference between situations for LS and CMMRD may partly evidence the reduced penetrance of heterozygous PMS2 mutations, with biallelic deficiency in the other MMR genes potentially lethal before presentation of a second cancer. Nevertheless, conclusions about the frequency and effects of *PMS2* variants are ultimately limited, compromised by an inability to accurately detect variants compared to other MMR genes.

5.1.2. Complications with the Detection of PMS2 Mutations

Of the 6658 unique, LS-causative variants currently detailed in the International Society for Gastrointestinal Hereditary Tumours (InSiGHT) database, only 14.7% are attributed to *PMS2*, compared to 27.9% for *MLH1*, 30.8% for *MSH2*, and 26.6% for *MSH6*. While this comparatively lower frequency of germline variants in *PMS2* may be explained by an absence of genetic testing for families with a lower penetrance of cancer (resulting from *PMS2* mutation) (Nakagawa et al., 2004; Senter et al., 2008;

Wolf et al., 2013), this may also be accounted for by complications in the analysis of this gene in clinical settings.

The assessment of *PMS2* is fundamentally complicated by the presence of 13 pseudogene fragments also situated on chromosome seven. A particular issue for short-read sequencing approaches, these effectively contaminate the analysis of the true coding sequence, and reduce the sensitivity of mutation detection (Clendenning *et al.*, 2006). The largest of these, as shown in Figure 5.1., is an approximate 100kb inverted duplication termed *PMS2CL* which shares ~98% sequence identity with exons 9 and 11-15 of the *PMS2* gene. The remaining 12 pseudogenes, also presented in Figure 5.1., each have a >90% identity to regions in the 5' end of the *PMS2* sequence. As a result, 6.7% of the gene, specifically Exon 1, is effectively a 'dead-zone' of sequencing due to ambiguous capture and read-alignment, while five other exons (~29% of the gene) are considered on an NGS high stringency problem list, and therefore difficult to analyse. In clinical testing, indiscriminate, short-read sequencing approaches may inadvertently amplify the aforementioned pseudogenes, covering variants in the actual gene sequence, and effectively reducing the sensitivity of mutation detection.

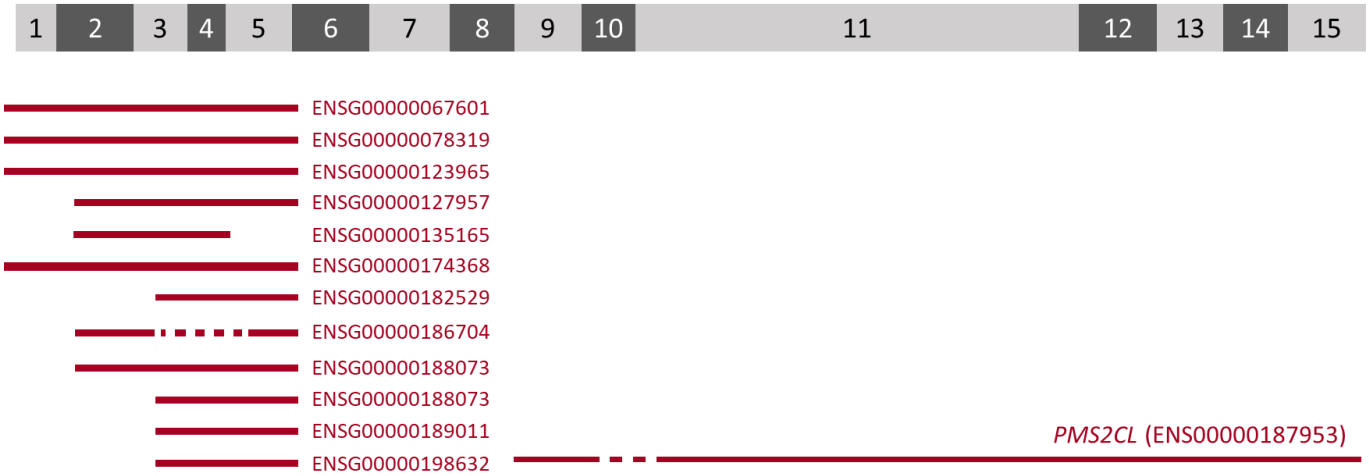

Figure 5.1. Schematic representation of the 13 pseudogenes associated with *PMS2*, mapped approximately to their regions of homology with the true coding sequence.

To negate the issues that complicate the sequencing of *PMS2*, various techniques have been developed. cDNA sequencing from lymphocyte cultures (puromycin-treated to inhibit the nonsense mediated decay of mRNA) has previously been shown to discriminate between *PMS2* and its pseudogenes, as well as detect splice variants and *PMS2/PMS2CL* 'hybrid alleles', the latter of which may be responsible for ~10% of all *PMS2* alleles in the European population (Etzler *et al.*, 2008). Approaches using MLPA have also been developed that enable the identification and localisation of copy number variants (CNVs), decreasing the number of cases requiring further mutation analysis. One study of 709 genomic DNA samples, including 17 known positives for PMS2 variation, used such a technique to return complete test sensitivity in mutation detection, with a 7% false-positive rate. (Herman *et al.*, 2018). The scalability of these methods is however limited, both by non-automatable protocols and their labour-intensive nature.

Long range PCR affords a further approach for the analysis of *PMS2*, providing a macro-sequencing option, or the initial phase in an improved testing pipeline for the gene. Using primers designed for regions with no/low sequence identity to pseudogenes, or that target sites of deviation from pseudogene sequences, this technique may be used for the preferential amplification of the *PMS2* coding sequence (Clendenning *et al.*, 2006). Alone, this is not considered wholly reliable for the accurate detection of large-scale deletions (Clendenning *et al.*, 2006), but, in combination with other techniques, the specific assessment of *PMS2* is possible. For example, in one analysis where the gene-specific products of long range PCR were reviewed by Sanger sequencing, 10 novel and 17 previously detected variants were identified from 30 colorectal and 11 endometrial cancer patients, including five novel pathogenic variants (Clendenning *et al.*, 2006). The capacity of long range PCR in tandem with MLPA has also been demonstrated, with 27 pathogenic mutations, including 10 large-scale deletions, identified in a cohort of 59 patients whose tumours exhibited isolated PMS2 loss by IHC (Vaughn *et al.*, 2010).

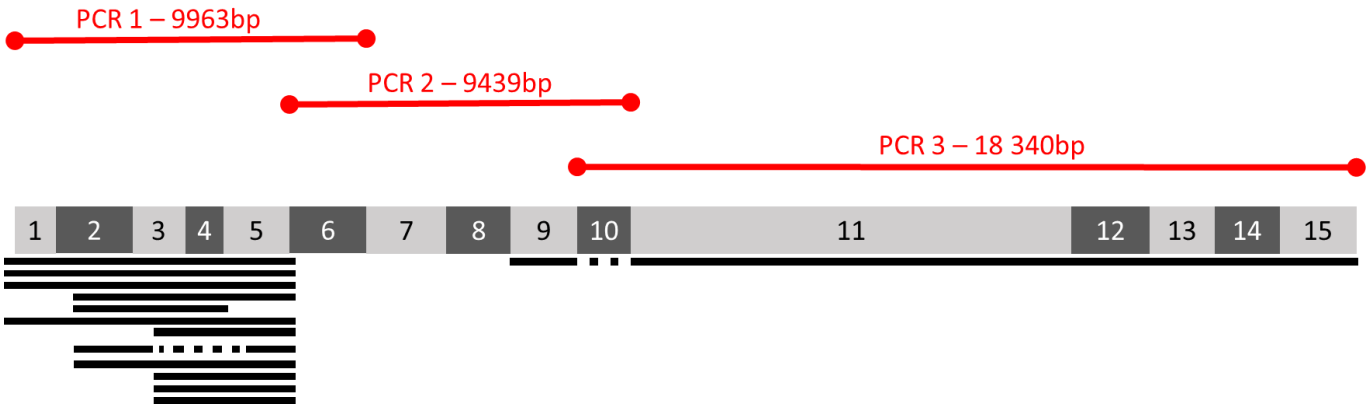

Figure 5.2. Schematic representation of the position and size of the three amplicon products resulting from long-range PCR using the published primer pairs of Clendenning et al. (2006).

An alternative option for the further analysis of the amplification products of long range PCR involves MIP technology, previously described in the context of the Newcastle MSI Assay. With MIPs designed according to critical performance parameters, gene tiling of exon regions is feasible to reveal positions of sequence variation. Moreover, combining Long-Range PCR with multiple MIPs targeting intronic polymorphisms may allow larger-scale gene deletions to be simultaneously identified. Ultimately, such a screening pipeline could permit the detection of low-level variants with lesser penetrance from larger populations, as well as the improved validation of variants. Furthermore, this would increase the feasibility of assessing larger, unselected populations for the frequency of compromised *PMS2*.

These analyses make it clear that, while mutations in *PMS2* are seemingly less penetrant that those in the other principal MMR genes, the comparatively lower frequency of variants reported in this gene among LS patients is also attributable to inadequacies in current screening techniques. As well as affecting the accuracy of any conclusions about *PMS2* in a research context, errors in mutation calling will also have implications for patients in terms of either unnecessary interventions (where a variant is incorrectly identified) or a lack of intervention (where a variant is missed). These failings would however be circumvented by an improved assay for the detection of *PMS2* variants. A MIP-based sequencing assay for the amplification of gene-specific

long range PCR products could be used for this purpose, and, by its nature, this would allow multiple classes of mutation, such as coding changes and deletions, to be identified.

5.2. Aims

In the work outlined in this chapter, I aimed to assess the potential application of MIP-based technology for the specific analysis of the *PMS2* gene. For this, there were several objectives:

- Create an exon-targeting MIP pool to identify single nucleotide variants (SNVs) affecting the *PMS2* coding sequence.
  To be addressed through the iterative analysis of MIP performance for the complete tiling of the true *PMS2* coding sequence.

- Create an intronic SNP-targeting MIP pool to identify copy number variants (CNVs) and a loss of heterozygosity.
  To be addressed through the iterative analysis of MIP performance for the amplification of target sequences within the *PMS2* introns.

- Assess the utility of a MIP-based assay for the detection of coding sequence variants in *PMS2*.
  To be addressed through the assessment of known *PMS2* variants, using a large cohort of confirmed *PMS2* carriers and controls.

5.3. Methods

For the analysis of samples using our sequencing-based approach, initial amplification of the *PMS2* gene specifically was required: this was achieved by long-range PCR (Detailed in Section 2.7.1.). Briefly, three published primer pairs were used to amplify the PMS2 gene locus of a given sample, with separate singleplex reactions involving each primer pair covering all 15 exons of the gene. Following purification, these amplicons were pooled in a 1:1:1 molar ratio, calculated based on their relative sizes and concentration readings, for use in subsequent MIP amplification reactions.

In the design of optimised smMIPs for the analysis of *PMS2* samples, the computer programme MIPGen (Boyle et al., 2014) was employed following user-defined parameters (Detailed in Section 2.7.2.). From the sequences returned by this, a selection of exon-targeting MIPs were initially tested in singleplex before pooling to assess the capacity for amplification with all combined. This same approach was used in the review of intron-targeting MIPs, for the determination of copy-number variation or a loss of heterozygosity in each amplicon analysed. In both instances, where insufficient MIP performance was realised, either in failed amplification or inadequate sequencing read depth, MIPs were re-designed and suitable replacements identified. Exon- and intron-targeting pools were eventually combined to establish an integrated pool of 142 MIPs, with 42 of the these covering the exons of the gene in a single-tiling approach, and 100 capturing intronic SNPs (Appendix C).

Samples were ultimately analysed by this combined MIP pool adapting the established sequencing protocol of Hiatt et al. (2013) and the Newcastle MSI Assay pipeline. All material for these was retrieved from the CaPP3 tissue resource in the form of purified DNA extracted from individuals with validated germline MMR defects. In total, 138 samples were reviewed in this process, 42 with confirmed *PMS2* mutations, and the remaining 96 with mutations in the other principal MMR genes (with the latter effectively behaving as a control population). Products generated in PCR amplification by the Newcastle Assay were sequenced to an average single read depth of 2000X per marker, and analysed by custom scripting pipelines (Detailed in Section 2.7.4.), before comparing the results to those of prior analyses.

5.4. Results

5.4.1. Modification of an Established Long-Range PCR Protocol

For the specific analysis of *PMS2* (in preference to its accompanying pseudogenes), precursory amplification across the gene's coding sequence is required. In this study, this was achieved by long-range PCR. Using primers designed specifically for target regions with no homology to, or significant deviation from, pseudogene sequences, three products can be generated in separate amplification reactions, together representing the true sequence of the gene for subsequent review (Figure 5.2.).

In the first instance, long-range PCR amplification was performed in accordance with the diagnostic services protocol of Oslo Universitetssykehus (Detailed in Section 2.7.1.). Among other conditions, this approach employs primers at a concentration of 10pmol to initiate amplification, and involves a total of 35 amplification cycles in product generation. Figure 5.3. presents the results when the protocol was used in conjunction with the aforementioned primer pairs for the specific amplification of *PMS2*. As shown for all three primer pairs, amplicons of an expected size were produced, with this seemingly the dominant product in all three reactions. However, as is also evident in each instance, with these reaction conditions there was off-target binding and the generation of superfluous amplification products (represented by 'smearing' and product banding at sizes other than calculated).

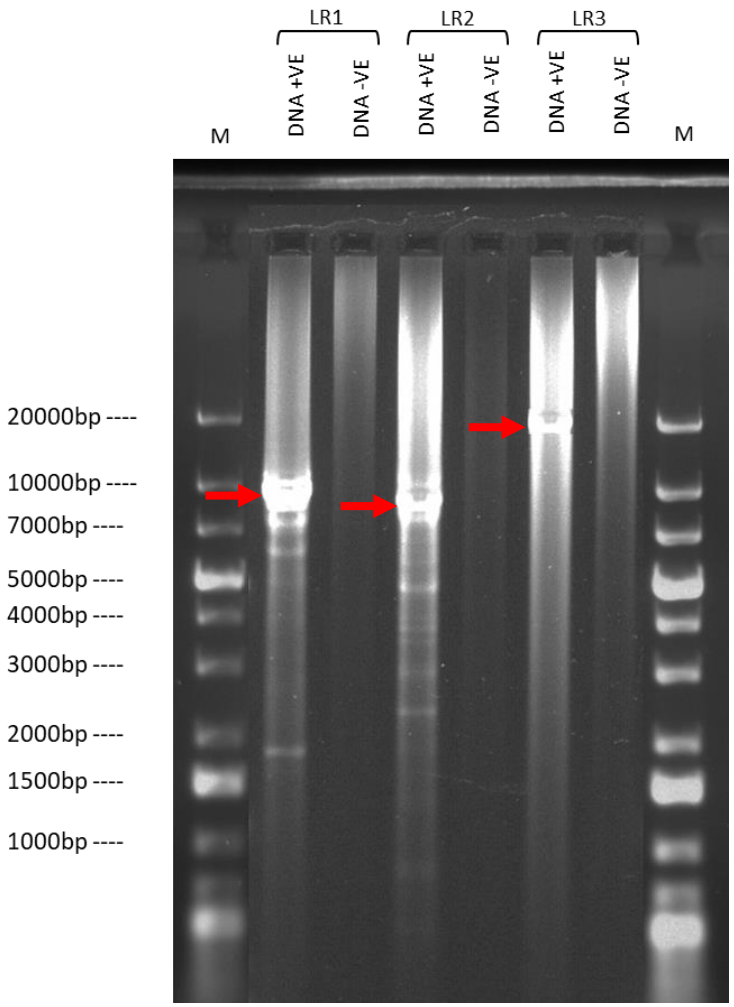

Figure 5.3. Gel electrophoresis image showing the results of long-range PCR amplification when amplifying the PMS2 coding sequence from cell line DNA using the diagnostic services protocol of Oslo Universitetssykehus. Amplicons of approximately 10kb in length are expected with primer pairs one and two, while amplicons of approximately 20kb are expected from amplification initiated by primer pair three: these are indicated by a red arrow for each amplicon.

As alternative amplification products may ultimately contaminate the eventual variant calling process in the analysis of *PMS2*, attention was subsequently given to adjusting the conditions for long-range PCR to eliminate these. Simultaneous efforts were made to reduce both the number of cycles involved in the amplification process, and the concentration of the primers used to initiate amplification (where possible), all without preventing the generation of intended products. The results of this experimentation are displayed in Figure 5.4.. Product banding of an expected size, without the presence of superfluous amplicons, was observed with 2.5pmol concentrations of primer for primer pairs one and two following as few as 20 amplification cycles. At the

equivalent cycle number, a primer concentration of 10pmol is required for primers of pair three, but, even at these levels, there was no artefact banding present. These amplification conditions and primer concentrations were adopted for the ensuing specific amplification of the *PMS2* gene from subject samples.

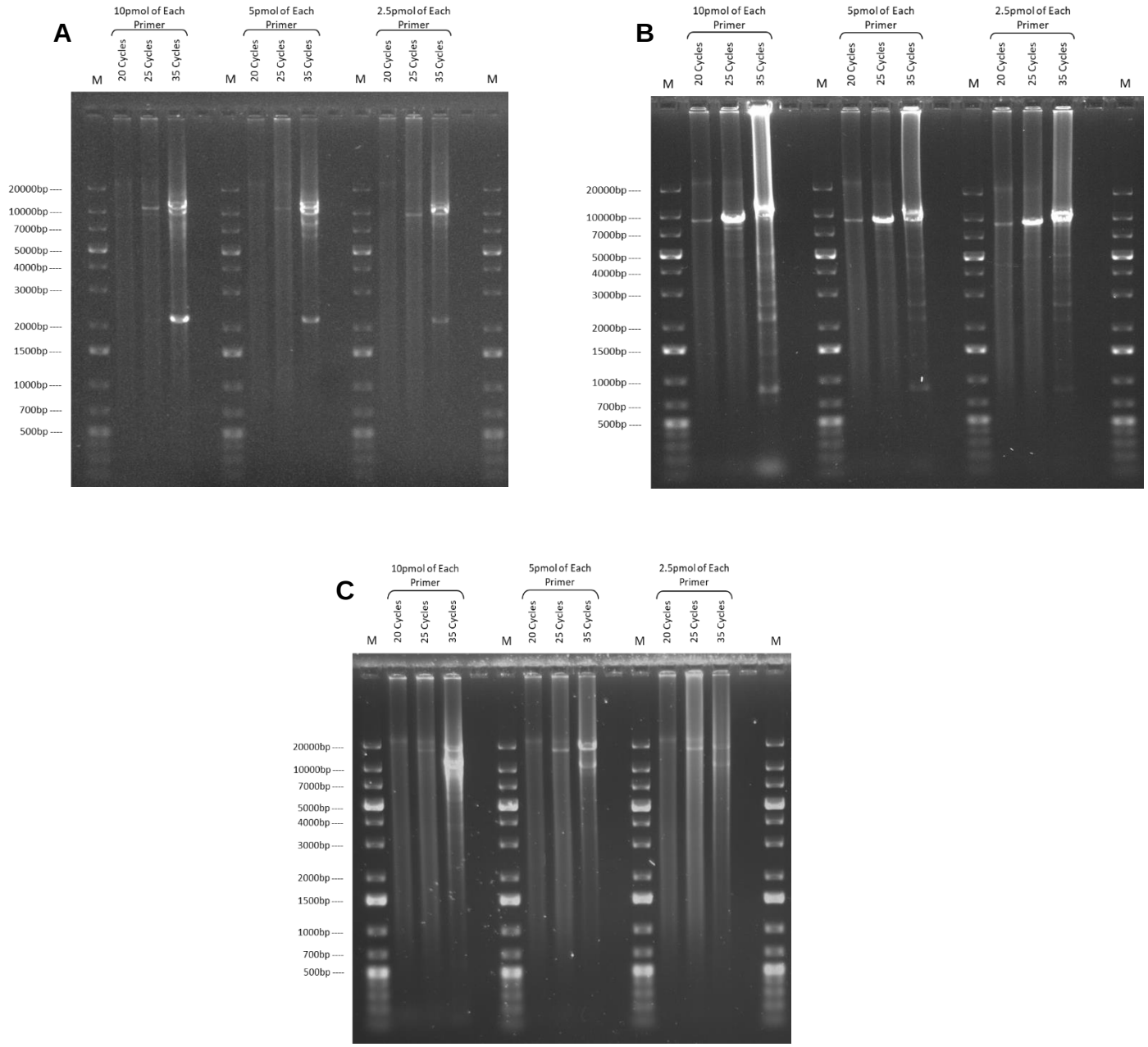

Figure 5.4. Gel electrophoresis images showing the results of long-range PCR amplification with different primer concentrations, and across a different number of amplification cycles, for primer pairs one (A), two (B) and three (C). Amplicons of approximately 10kb in length are expected with primer pairs one and two, while amplicons of approximately 20kb are expected from amplification initiated by primer pair three.

5.4.2. Development of PMS2 Exon-Tiling MIP Pool for Variant Detection

For the analysis of long-range amplicons representing *PMS2*, molecular inversion probe (MIP) technology was used. The design of MIPs specific to the *PMS2* coding sequence in the first instance was achieved through the use of the computer programme MIPGen (See Methods), with the sequences of the gene's 15 constituent exons supplied to this, and user-defined parameters for MIP performance specified. MIPs with the highest theoretical performance metrics (i.e. those expected to achieve sufficient amplification and sequencing read depth) were selected, with 42 MIPs ultimately considered sufficient for the complete coverage of the *PMS2* coding sequence in a single-tiling approach.

As a preliminary assessment of the ability of the identified MIPs to amplify from a long-range PCR amplicon template, ten of these with target regions across the *PMS2* coding sequence were selected, and the amplification achieved with them was explored. Using long-range PCR products from the amplification of K562 cell line genomic DNA, MIP amplification was performed in accordance with the established protocol of the Newcastle MSI Assay (Detailed in Section 2.5.), before the results of this were analysed by gel electrophoresis. The results of this amplification are presented in Figure 5.5.. For all ten MIPs, amplicons of an expected size (240-280bp) were observed in the absence of unexpected product banding, with primer and dimer bands also visible (<100bp). The intensity of this amplification does however vary from locus to locus, with amplification by MIP Exon15_0002 being especially faint, while amplification by MIP Exon10_0019 returns some off-target banding. Nevertheless, this assessment confirmed the capacity of MIPs, generated for this region by MIPGen, to amplify from the products of long-range PCR amplification, and therefore supported the continuation of this study with all probes for complete exon coverage.

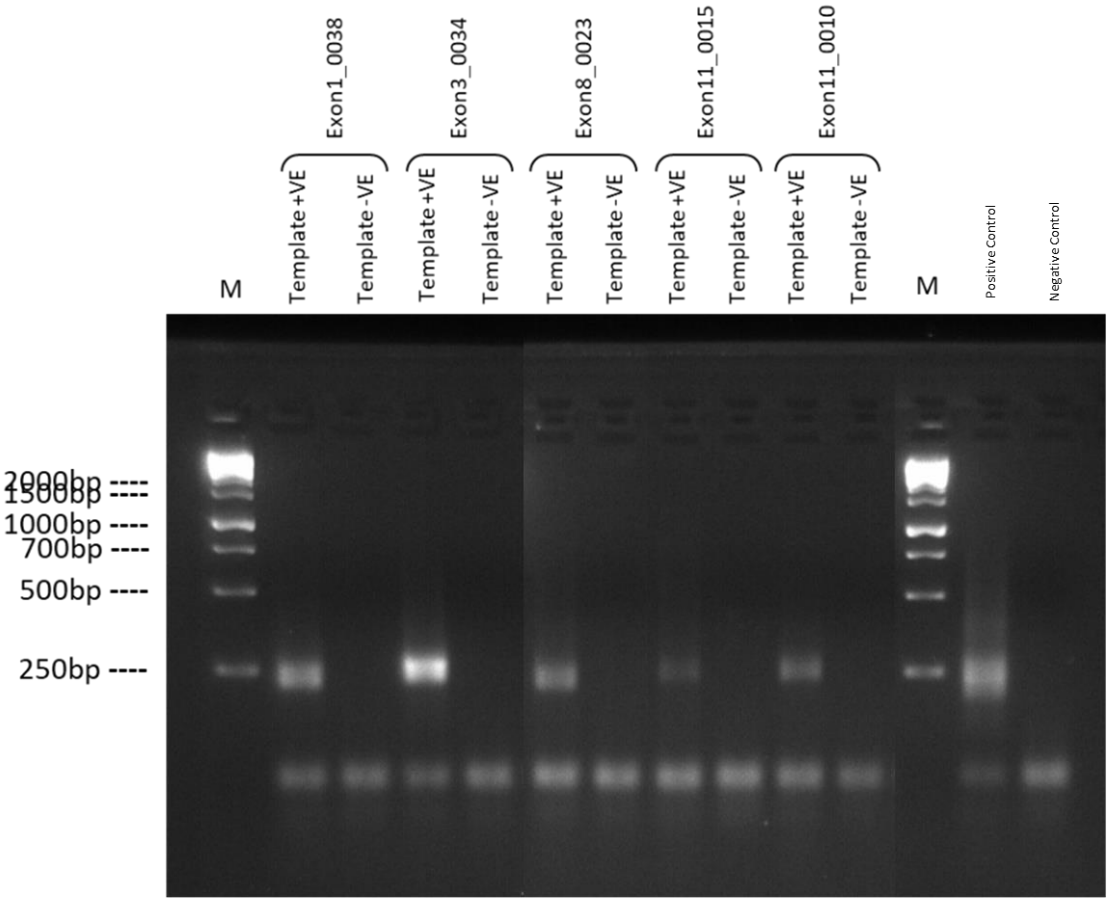

Figure 5.5. Gel electrophoresis image showing the results of MIP amplification by a selection of exon-targeting *PMS2* MIPs employed in singleplex. Amplicons of approximately 240-280bp in length are expected from amplification using these probes.

Given the success of MIP amplification from long-range PCR amplicons of *PMS2*, it was considered logical to subsequently proceed with the pooling of all 42 exon-targeting MIPs to produce a pool of probes covering the entire *PMS2* coding sequence. Initially, such a pool was established by combining equal volumes of all constituent MIPs into a single pool, with this then subjected to the necessary phosphorylation and dilution to prepare this for use in the MIP amplification process. Using this pool, amplification of the *PMS2* coding sequence was conducted from both long-range amplicon and genomic (cell line) DNA templates, with the results of this investigation analysed by gel electrophoresis, the output of which is shown in Figure 5.6.. Products of an expected size (240-280bp) were generated with amplification from both templates using this combined MIP pool, thus demonstrating that amplification was achieved by some, if not all, of the exon-targeting probes therein.

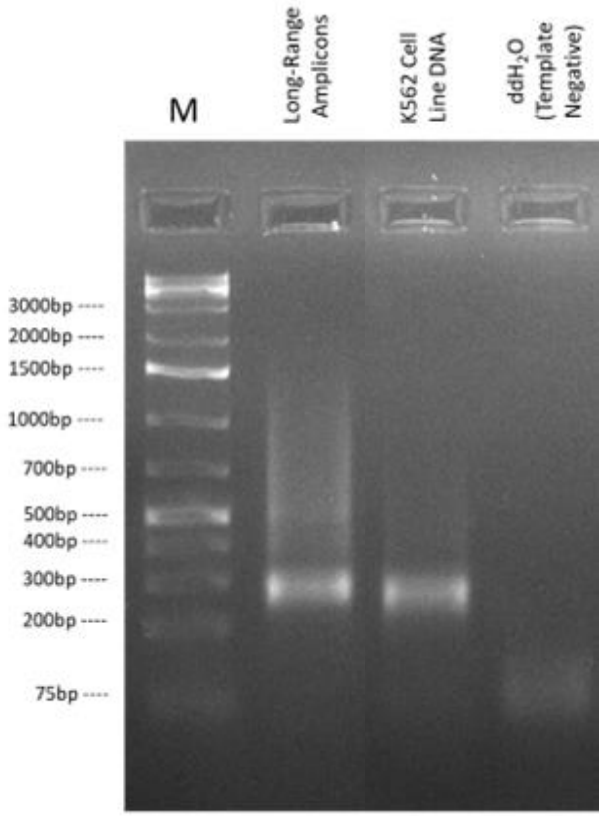

Figure 5.6. Gel electrophoresis image showing the results of amplification from various templates using the first iteration of an exon-tiling MIP pool for *PMS2*. Amplicons of approximately 240-280bp in length are expected from amplification using these

To assess the individual performance of MIPs within the pool produced, and their amplification of the *PMS2* coding sequence as intended, the sequencing of material produced in the prior amplification was conducted. The read numbers associated with each constituent probe of the pool from this process are presented in Figure 5.7 (A). Given that only the *PMS2* coding sequence is available as a template, the assignment of reads to MIPs interrogating long-range PCR amplicons indicated that the component probes of the pool are amplifying their targets as intended. However, the success of this amplification varied, and, in several instances, inadequate read numbers were returned for analysis.

With read balancing to follow, MIPs with a read count below 10% of the mean value across all probes were considered to have achieved an insufficient sequencing return. In total, five MIPs failed to meet this quality criterion with a sequencing depth below 200 reads, and were therefore selected for redesign using the MIPGen computer programme. Functional replacements (with a sufficiently-high logistic score)

were identified for these, and all were included in a second iteration of the exon-tiling MIP pool in place of their analogues. The performance of this novel pool was then tested as before to confirm coverage of the *PMS2* coding sequence and adequate performance by the replacement probes: the results of this are shown in Figure 5.7. (B). With both templates reviewed, complete coverage of the *PMS2* exons was achieved by the MIPs of this pool, albeit with variation once again in the performance of individual probes. In achieving this coverage, a greater range of read depths was returned with amplification from LR products (213-24235) than from gDNA template (153-3425), with this not associated with a specific amplicon. However, unlike in the previous analysis, when amplifying from long-range amplicon template, all MIPs returned a read depth exceeding 10% of the mean value across all probes, suggesting this pool is suitable to proceed with.

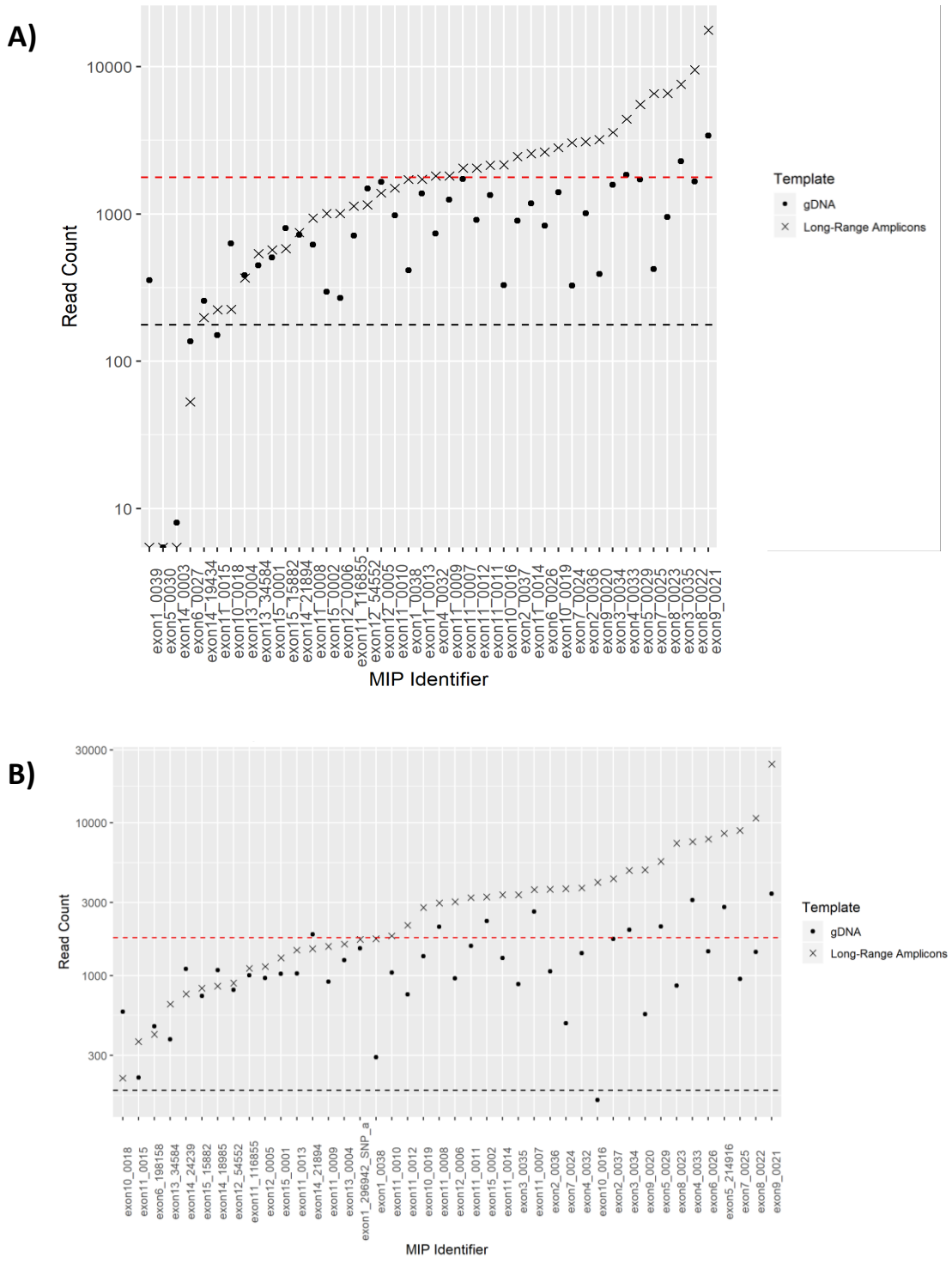

Figure 5.7. Comparison of the read counts by MIP returned from analysis of genomic DNA and long-range amplicon material by the first (A) and second (B) iterations of the *PMS2* exon-tiling MIP Pool. MIPs, and their corresponding read counts, are arranged in increasing order relative to their performance with long-range amplicon template. A red dashed line represents the median read value, and the black dashed line denotes 10% of this median, the minimum practicable threshold for rebalancing.

As a further adjustment in the development of the MIP pool, the aforementioned read balancing was performed in an attempt to normalise the distribution of sequencing reads across all component probes of the *PMS2* exon-tiling MIP pool. To achieve this, modified input volumes of each MIP were combined into a single mixture (Appendix C), with these values calculated with reference to the performance of individual probes in the prior analysis. As with the previous two, this third pool was used in MIP amplification from a long-range amplicon template, and the products were sequenced and analysed as before. The results of this investigation are related in Figure 5.8.. Once again, successful coverage of the entire PMS2 coding sequence was achieved using this exon-tiling pool, with sufficient read depths again returned across all positions for subsequent computational analysis. Figure 5.9. presents a further comparison between the range of scoring realised by the second and third iterations of the *PMS2* exon-tiling MIP pool. With a closer grouping of read counts across all probes, and a shorter range to this scoring, these findings show the effects and success of the read balancing process. This final pool was therefore used for all subsequent *PMS2* exon analysis.

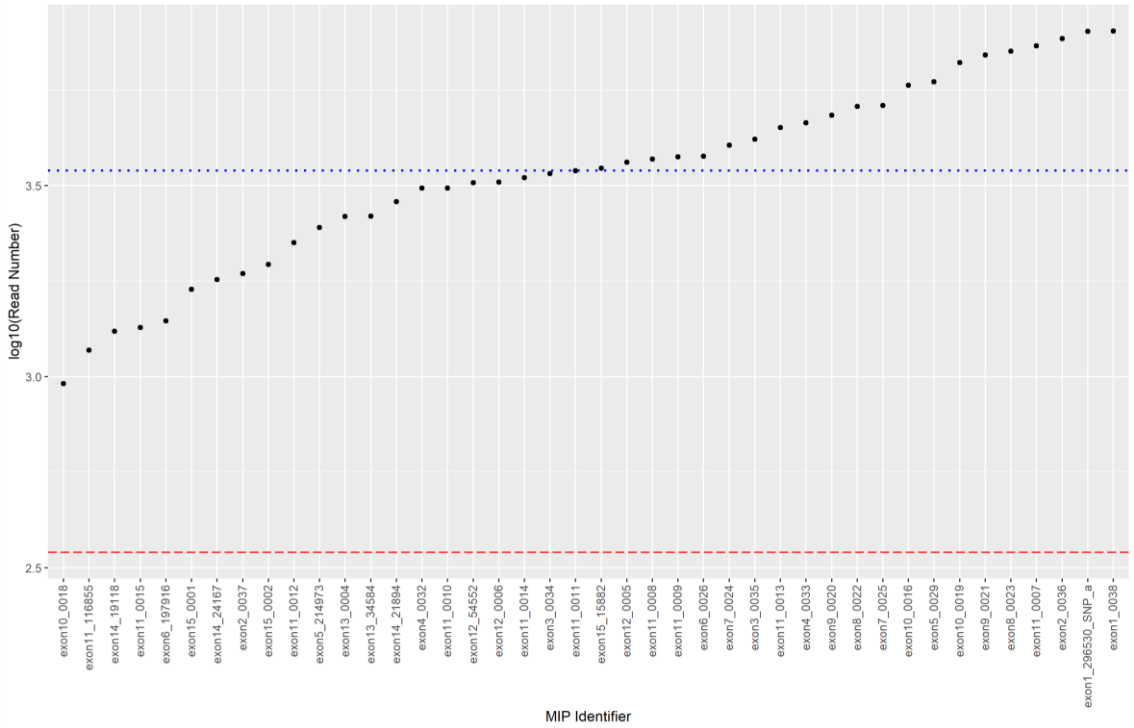

Figure 5.8. Read counts by MIP returned from analysis of long-range amplicon material by the third iteration of the *PMS2* exon-tiling MIP Pool. A blue dashed line represents the median read value, and the red dashed line denotes 10% of this median, the minimum practicable threshold for rebalancing.

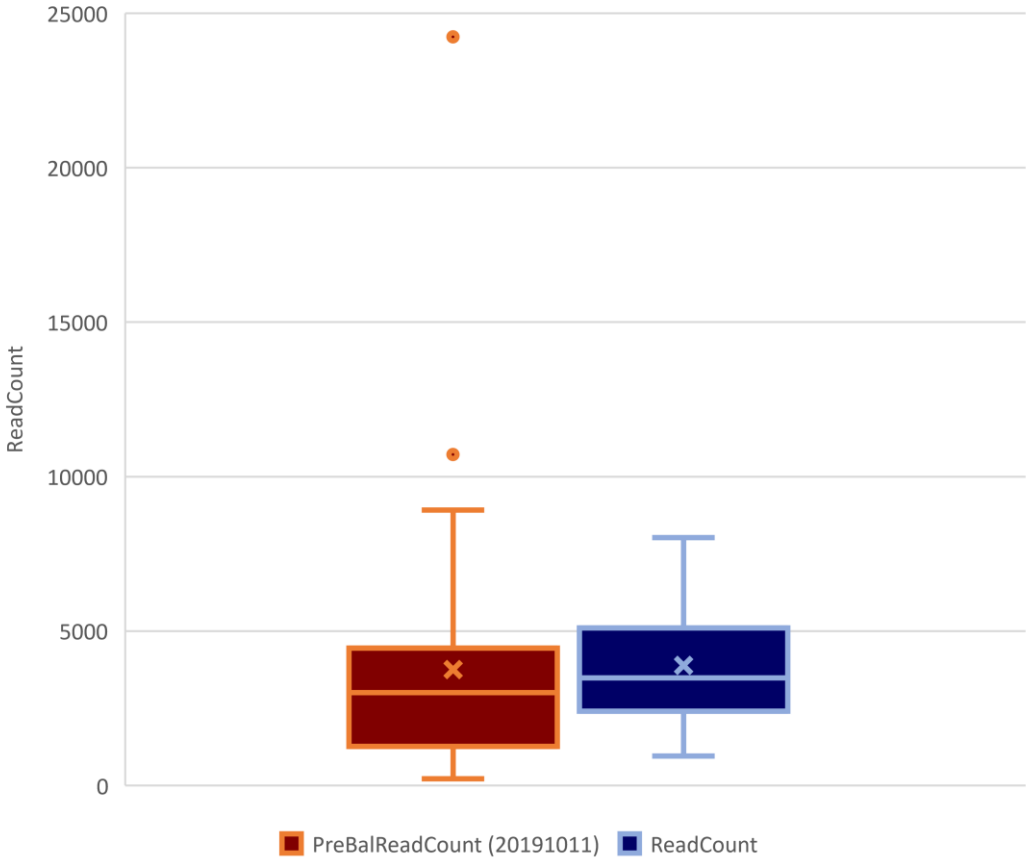

Figure 5.9. Comparison of the distribution of read counts returned from analysis of long-range amplicon material by the second (left) and third (right) iterations of the *PMS2* exon-tiling MIP Pool. For each box, the middle line represents the median read count across all probes reviewed, The bottom and top lines of each box define the lower and upper quartiles of each data set respectively, while the whiskers extend to indicate variability outside of these quartiles. A cross along each distribution indicates the mean read count across all probes in that review while outliers are shown as dots.

5.4.3. Initial Testing of PMS2 Exon-Tiling MIP Pool for Variant Detection

The ultimate ambition of this study was the development of an assay for the detection of pathogenic mutations in *PMS2*. To assess the ability of the *PMS2* exon-tiling MIP pool to detect variants within the coding sequence of this gene, a review of sample material was performed. In the first instance, assessment of both gDNA and LR products from a single sample with no *PMS2* mutation was analysed using the earliest version of a novel variant-calling pipeline, the results for which are presented in Figure 5.10.. Comparable MIP performance was returned with amplification directly from gDNA and from LR-PCR amplicon templates. However, in the former, more variants

(11), that is, deviations from the hg19 reference genome, were observed than in the latter (5). For some of these calls, an intermediate VAF was observed, with this potentially representing bias in PCR reactions and/or the amplification of other loci. Nevertheless, the additional calls with genomic DNA template possibly represent MIPs annealing to, and amplifying pseudogenes (and their associated variants), consistent with the chief advantage of the specific preparatory amplification of PMS2, essentially as a form of filtering.

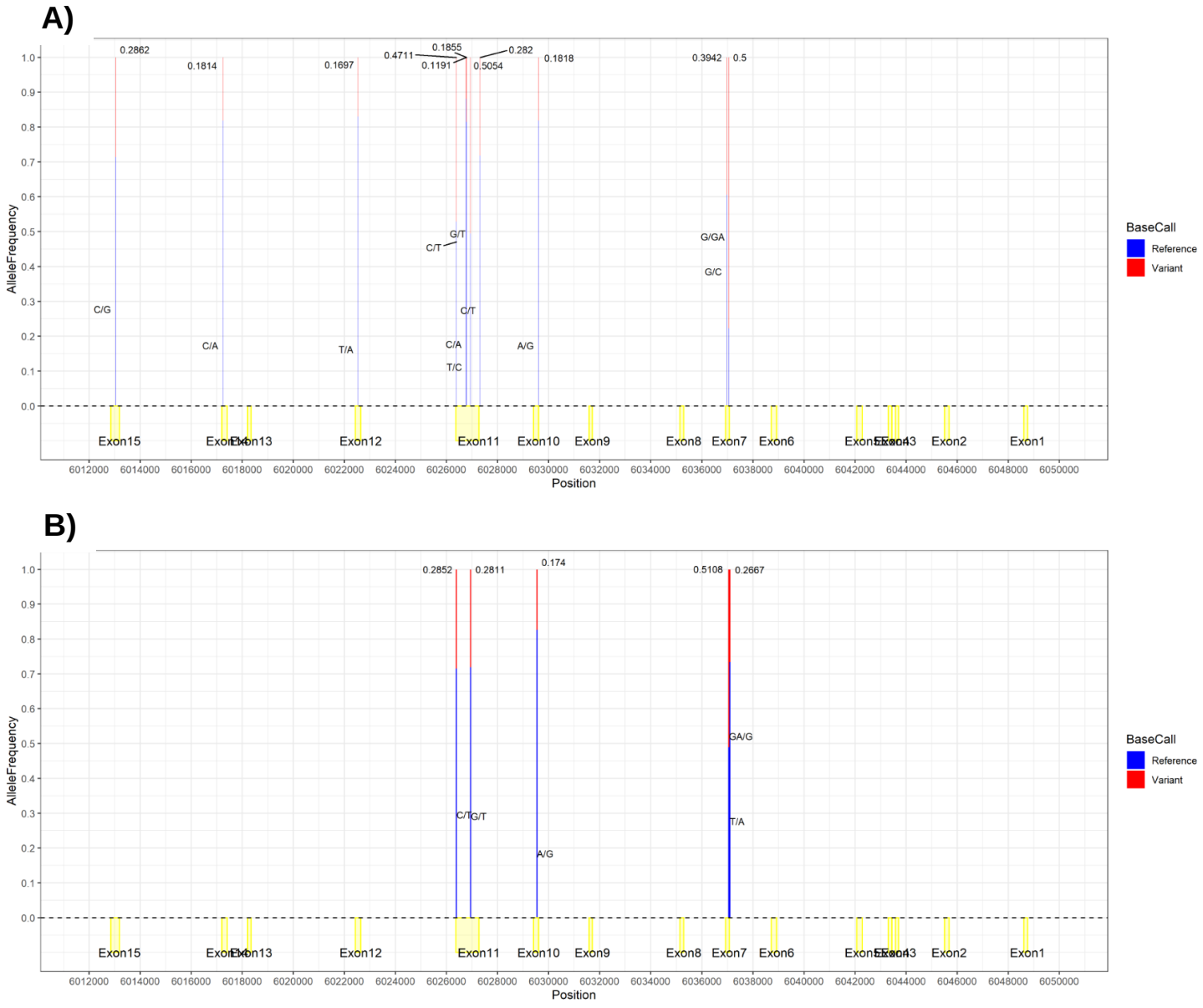

Figure 5.10. Comparison of variant calling following MIP amplification of both gDNA (A) and long-range amplicon (B) template by the *PMS2* exon-tiling MIP pool.

Given these positive results, the equivalent approach was subsequently used for the analysis of a further 12 controls and six LS samples with known pathogenic *PMS2* mutations (Details of which are given in Table 5.1.), to compare the output once more from the two templates, as well as determine the ability to detect previously confirmed variants. For these analyses, adjustments to the variant-calling pipeline were effected based on the earlier sample assessment, including a preliminary trimming of FastQ sequencing files (to remove data arising from MIP targeting arms). Such changes were accompanied by the implementation of filtering by the following established quality thresholds: Mapping Quality (MQ) ≥ 40; MQRankSum ≥ -12.5; QD ≥ 2; Depth (gt_DP) ≥ 100 (Van der Auwera *et al.*, 2013). In the resulting conditions, more variants were again returned with gDNA templates (96) versus long-range PCR amplicon templates (88), but, after this pipeline manipulation, the difference between the two templates was less disparate than previously observed with the non-LS sample (p-value=0.6059).

Considering the three samples with known pathogenic point mutations specifically, all of these were accurately detected by this approach. However, for two of these – Samples 01123 and 01130 – a homozygous allele state of the variant was reported. An example of this is shown in Figure 5.11.. While it is possible that these samples are genuinely homozygous, CMMRD situations, it is also conceivable that this represents the amplification of only one allele by LR-PCR, a situation that requires consideration. This could be caused by the presence of a deletion allele, or be an artefact resulting from the preferential amplification of one allele during the LR-PCR reaction. Other variants for these samples, given the information available, were of limited interest, with heterozygous variant allele frequencies (0.30-0.52) generally returned for these.

Pathogenic variations, particularly in the region of interest, were not identified in either of the two samples with an exon 10 deletion, while read counts for this exon were comparable to others within the same amplicon. At this stage, variant detection by this assay was relatively restricted, but it was envisaged that deletions such as this, as well as the aforementioned homozygosity of other variant calls, could be more easily resolved with the supplementary analyses of intronic SNPs.

Sample Identifier	Identified *PMS2* Variant
01135	c. 137 G>T (rs121434629)
01123	c. 137 G>T (rs121434629)
01130	c. 137 G>T (rs121434629)
01088	Exon 10 Deletion
01100	Exon 10 Deletion
01137	Exon 15 Dosage ≈0.8

Table 5.1. Details of the LS samples from the CaPP3 tissue resource used to investigate the potential of the novel pipeline for variant detection, along with their respective, pathogenic *PMS2* mutations.

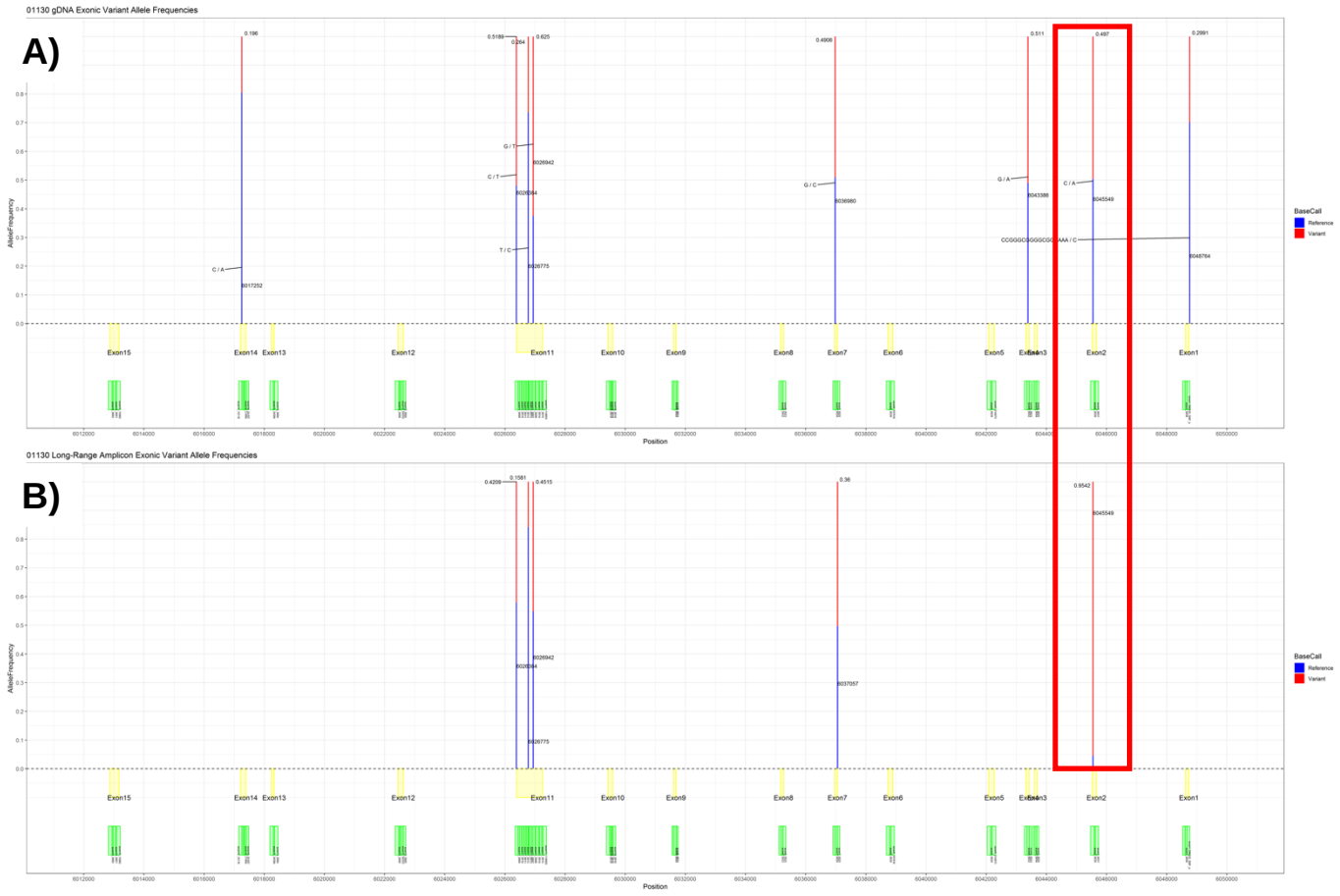

Figure 5.11. Comparison of variant calling in Sample 01130 following MIP amplification of both gDNA (A) and long-range amplicon (B) template by the *PMS2* exon-tiling MIP pool. The red box indicates the position of the confirmed pathogenic mutation in the *PMS2* gene of this sample.

5.4.4. Development of PMS2 Intronic MIP Pool for Loss of Heterozygosity and Copy Number Variation Assessment

In addition to the detection of variants within the coding sequence, a review of copy number variation and/or the loss of heterozygosity was considered requisite of this novel *PMS2* assay. The design of MIPs to capture common intronic polymorphisms for this purpose was achieved once more through the use of the MIPGen computer programme, with numerous sequences incorporating SNPs across the entire gene provided to the software (in addition to the same user-defined parameters as used in the development of exon-tiling probes).This returned an output with multiple MIP sequences incorporating each provided SNP position within the *PMS2* introns, and, from these, the MIP with the highest theoretical performance metric (logistic score) for each polymorphism was selected. In total, 114 MIP sequences covering 139 SNPs were identified, with these giving cumulative probabilities of homozygosity sufficiently low across each amplicon for the recognition of larger-scale deletions and copy number variation.

To ascertain the ability of the previously identified, SNP-targeting MIPs to amplify from long-rang amplicon template, a sample of ten probes with targets across the PMS2 gene were selected, and the amplification achieved from these was explored. Using long-range PCR products from the amplification of cell line genomic DNA, MIP amplification was performed in accordance with the established protocol of the Newcastle MSI Assay (Detailed in Section 2.5.), before being analysed by gel electrophoresis. The results of this amplification are presented in Figure 5.12.. For seven of the nominated MIPs, obvious amplicons of an expected size (240-280bp) were observed in the absence of unexpected product banding, with primer and dimer bands also visible (<100bp). Demonstrating the capacity of these MIPs to amplify from the products of long-range PCR amplification, this inspired the continuation of this study with all probes for SNP assessment.

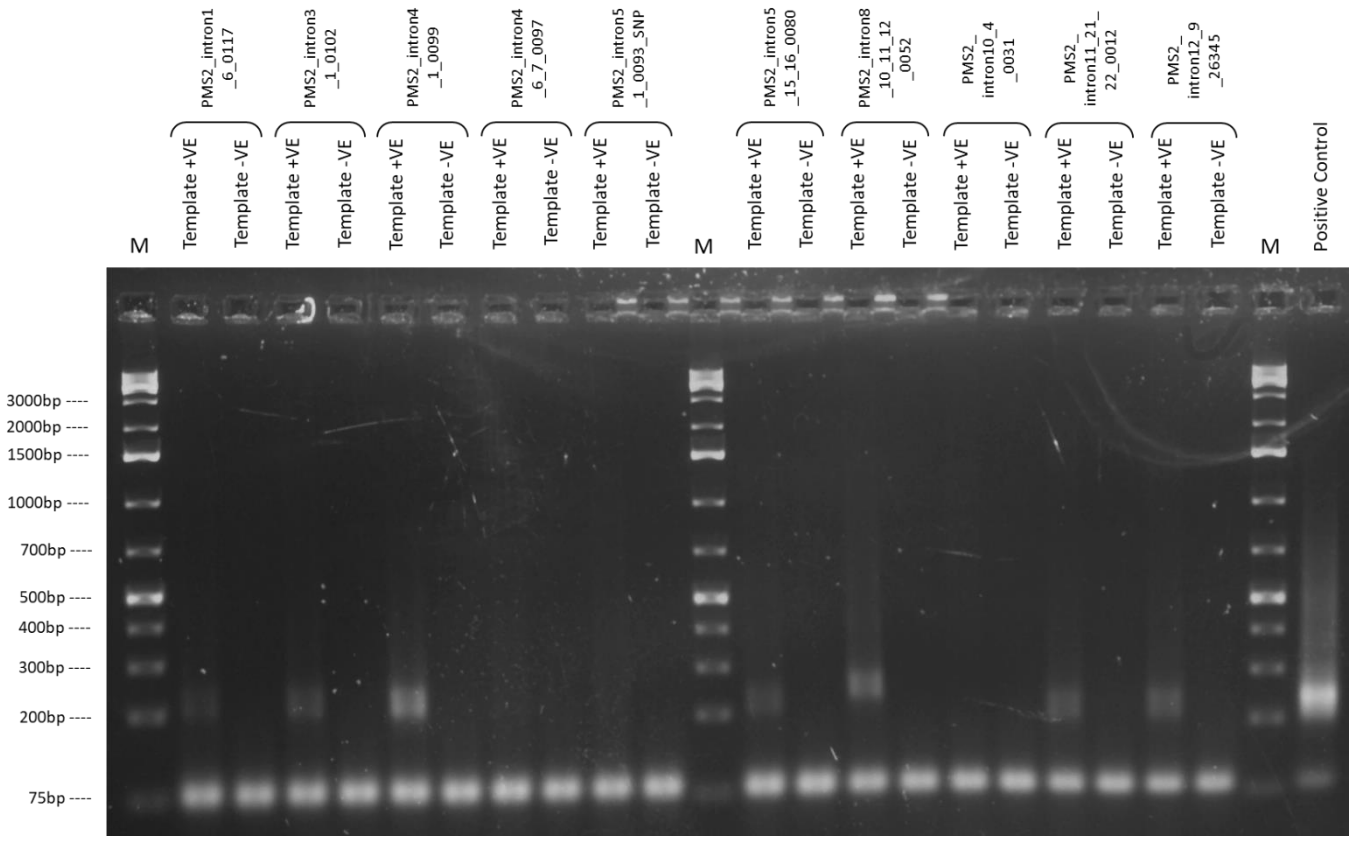

Figure 5.12. Gel electrophoresis image showing the results of MIP amplification by a selection of intron-targeting *PMS2* MIPs employed in singleplex. Amplicons of approximately 240-280bp in length are expected from amplification using these probes.

Given the results of the MIP amplification with a selection of SNP-targeting MIPs from long-range amplicons of *PMS2*, the pooling of all 114 probes identified previously for the review of polymorphisms was performed to establish a MIP pool for this purpose. Initially, as with the exon-tiling analogue, such a pool was produced by combining equal volumes of all constituent MIPs into a single aliquot, with this then subjected to the required phosphorylation and dilution to prepare this for use in the MIP amplification process. Using this pool, amplification of the selected common polymorphisms was conducted from both long-range amplicon and genomic (cell line) DNA templates, with the results of this investigation analysed by gel electrophoresis, the output of which is presented in Figure 5.13.. Products of an expected size (240-280bp) were generated with amplification from both templates using this combined MIP pool, thus demonstrating that amplification was achieved by some, if not all, of the SNP-capturing probes therein.

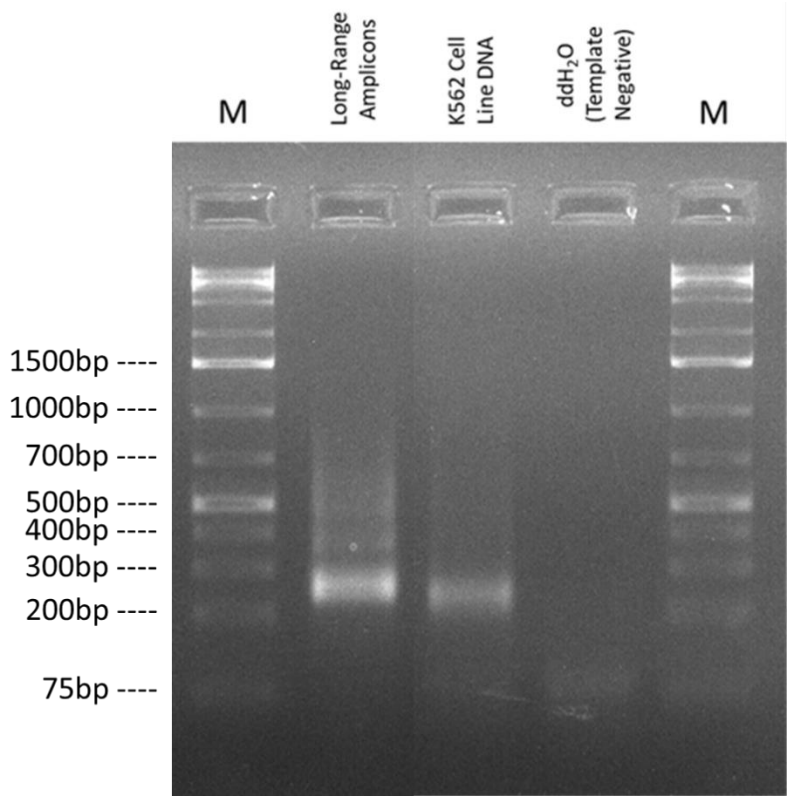

Figure 5.13. Gel electrophoresis image showing the results of amplification from various templates using the intron-targeting MIP pool for *PMS2*. Amplicons of approximately 240-280bp in length are expected from amplification using these probes.

To assess the individual performance of MIPs within the pool produced previously, and their amplification of *PMS2* intronic SNPs as intended, the sequencing of material produced in the prior amplification was performed. The read numbers associated with each constituent probe of the pool from this process are presented in Figure 5.14.. With only the *PMS2* gene sequence present within the LR-PCR amplified template, the assignment of reads to MIPs interrogating long-range PCR amplicons indicated that the component probes of the pool were amplifying their targets as envisaged. However, as with the earlier exonic analysis, the success of this amplification varied, and,in approximately 20% of markers (19/114) unsatisfactory read depth was returned for subsequent analysis.

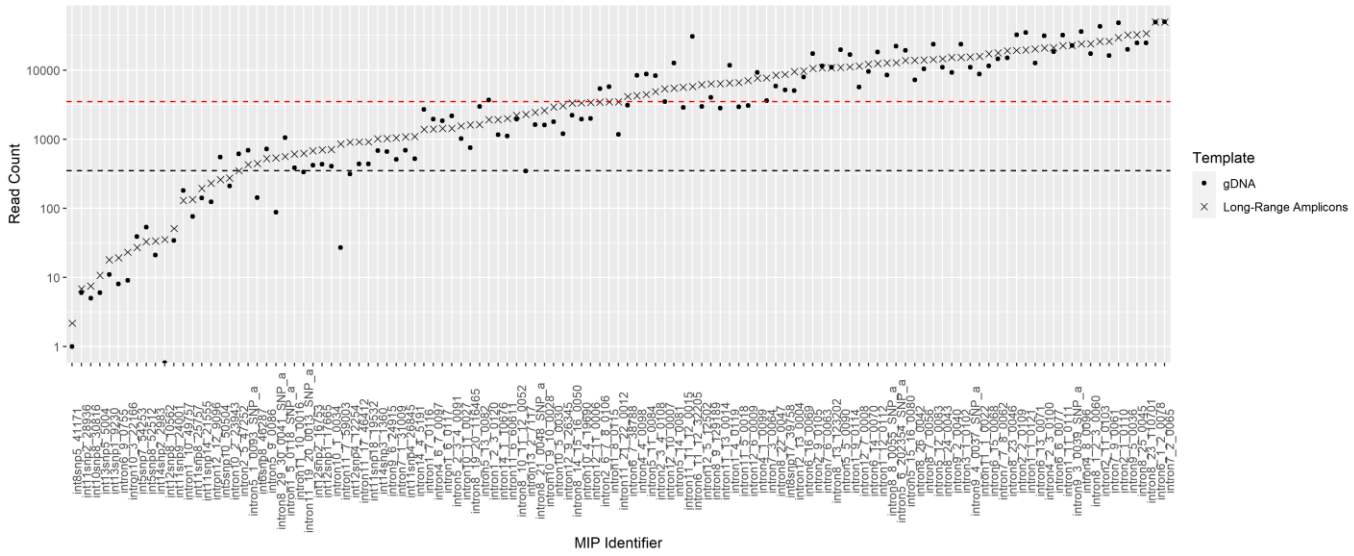

Figure 5.14. Comparison of the read counts by MIP returned from analysis of genomic DNA and long-range amplicon material by the first iteration of the *PMS2* intronic SNP-targeting MIP Pool. MIPs, and their corresponding read counts, are arranged in increasing order relative to their performance with long-range amplicon template. A red dashed line represents the median read value, and the black dashed line denotes 10% of this median, the minimum practicable threshold for rebalancing.

Power calculations revealed that there were sufficient MIP numbers across each long-range amplicon (for a satisfactory determination of a loss of heterozygosity) without the inclusion of probes that had given an inadequate read depth prior, therefore the redesign of functional replacements for these was not considered necessary for an initial assessment of the methodology. Instead, read balancing was performed for the remaining MIPs to normalise the distribution of sequencing reads across all component probes of the *PMS2* SNP-targeting pool. To achieve this, modified input volumes of each MIP were combined into a single volume, with these values calculated with reference to the performance of individual probes in the previous analysis. This novel pool was then used in MIP amplification from a long-range amplicon template, before the products of such were sequenced and analysed as before. The results of this experiment are related in Figure 5.15.. Once again, the successful capture of all SNP-containing sequences was achieved using this pool, with sufficient read depths again returned across all MIPs for subsequent computational analysis. Figure 5.16. presents a further comparison between the range of scoring realised by the first and second iterations of the intron-targeting MIP pool.

With a closer grouping of read counts across all probes, and a shorter range to this scoring, these findings characterise the effects of the read balancing exercise, and it was decided the pool assessed here would proceed to eventual use in the final *PMS2* MIP pool for sample analysis.

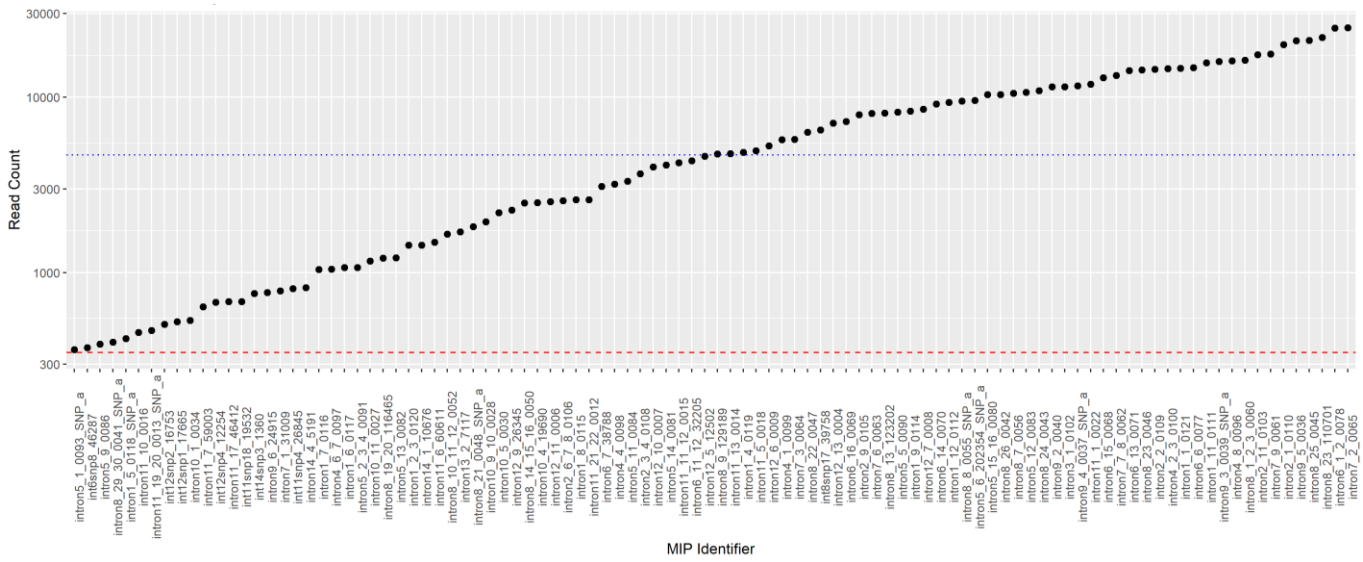

Figure 5.15. Read counts by MIP returned from analysis of long-range amplicon material by the rebalanced *PMS2* intronic SNP-targeting MIP Pool. A blue dashed line represents the median read value, and the red dashed line denotes 10% of this median, the minimum practicable threshold for rebalancing.

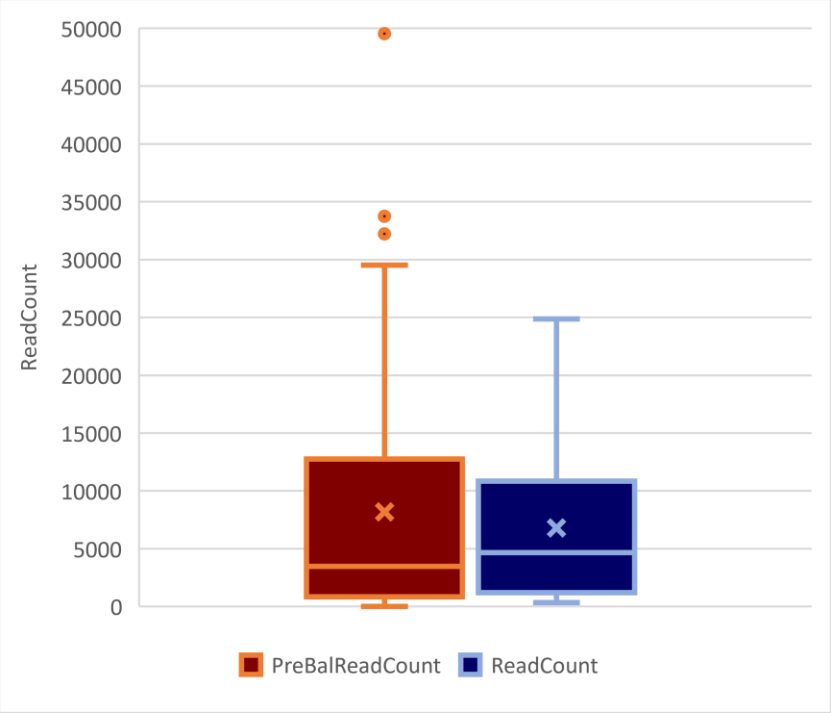

Figure 5.16. Comparison of the distribution of read counts returned from analysis of long-range amplicon material by the first (left) and second (right) iterations of the *PMS2* SNP-targeting MIP Pool. For each box, the middle line represents the median read count across all probes reviewed, The bottom and top lines of each box define the lower and upper quartiles of each data set respectively, while the whiskers extend to indicate variability outside of these quartiles. A cross along each distribution indicates the mean read count across all probes in that review while outliers are shown as dots.

5.4.5. Establishment and Testing of a Complete MIP Pool for PMS2 Analysis

With the separate MIP pools for exon tiling and SNP review established, and the amplification of their intended targets demonstrated, final efforts before sample analysis concerned the integration of the two pools, and an assessment of individual probe performance following this amalgamation. Volumes of each pool were combined relative to the number of probes contained in each, before the resulting mixture of 142 MIPs was used in MIP amplification (as per the protocol of the Newcastle MSI Assay) to interrogate both the genomic DNA and long-range amplicons of a control sample. The products of this amplification were subsequently sequenced and processed, returning the output presented in Figure 5.17.. In the assessment of both templates,

the successful capture of all intended sequences and polymorphisms of the *PMS2* gene was realised by the combined MIP pool. Furthermore, with a greater read depth consistently realised than with the products of amplification directly from genomic DNA, the individual probes contained in this obtained sufficient read depth when amplifying from long-range amplicon template for subsequent data analysis.

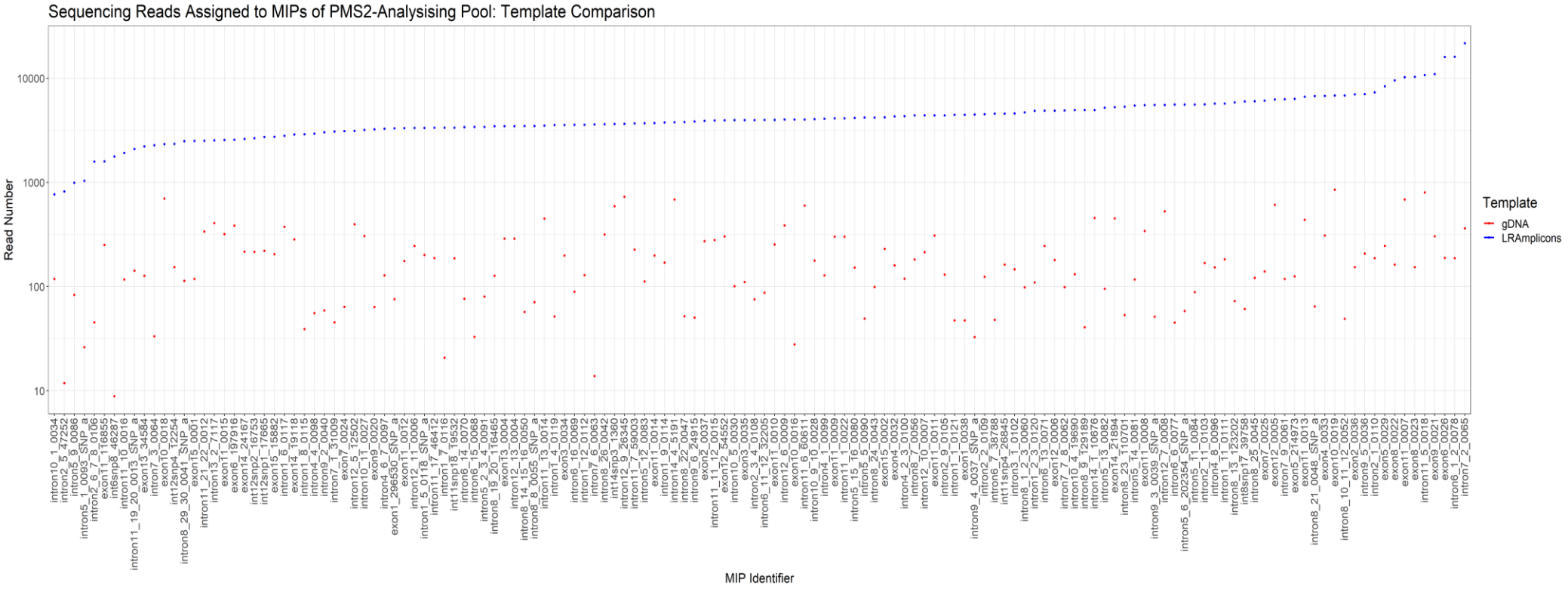

Figure 5.17. Comparison of the read counts by MIP returned from analysis of genomic DNA and long-range amplicon material by the combined PMS2 MIP pool for exon-tiling and SNP assessment. MIPs, and their corresponding read counts, are arranged in increasing order relative to their performance with long-range amplicon template.

5.4.6. Mutation Detection in PMS2 Mutation Carriers

Having generated balanced pools for both the exonic PMS2 sequence and informative SNPs spanning its genomic region, the analysis of sample material was ultimately performed. In preparation for this, samples for review were initially identified by querying the CaPP3 tissue database for tumours arising in individuals with a confirmed *PMS2* mutation. A total of 66 suitable samples were specified and retrieved for this investigation, as well as 72 controls with confirmed mutations exclusively in each of the other three principal MMR genes (Appendix D). All samples were procured from peripheral blood leukocytes in the form of purified genomic DNA, with long-range PCR executed for this material, and the products used as the template for MIP amplification with the previously established *PMS2* MIP pool. An example of the results of MIP amplification is shown in Figure 5.18., with amplicons of the expected size range (240-280bp) are observed in all samples apart from the negative control (Sample E8), with primer bands also visible in some channels (<100bp). MIP amplification was followed by sequencing and data analysis, with the latter primarily consisting of variant calling within the coding sequence and SNP calling, followed by the evaluation of identified mutations by pathogenicity predictors to consider their physiological consequence(s).

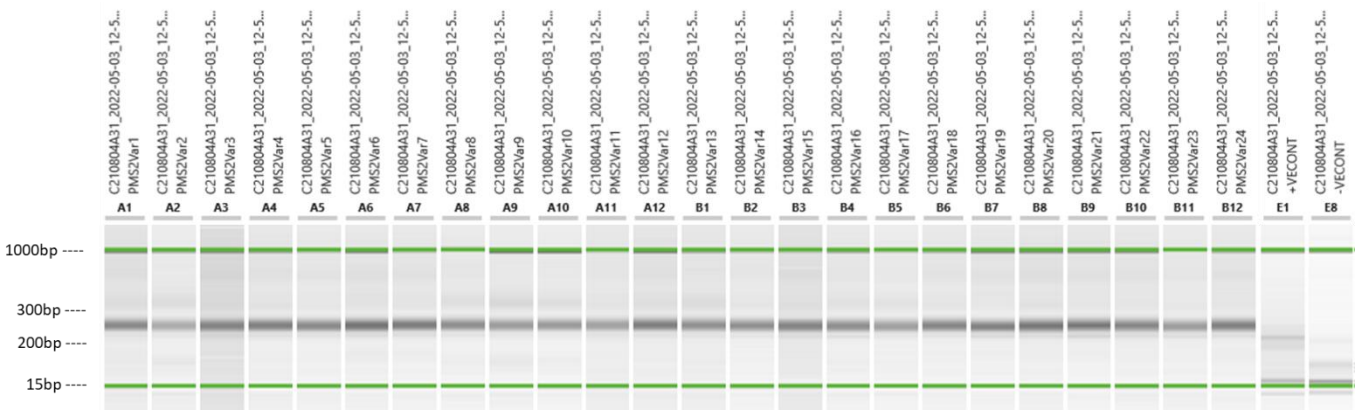

Figure 5.18. Exemplar Qiaxcel capillary electrophoresis image showing the results of MIP amplification from samples with confirmed *PMS2* mutations using the novel *PMS2* MIP pool. Amplicons of approximately 240-280bp in length are expected from amplification using these probes.

In variant calling, two stages of filtering were effectively applied. The first, included in the scripting of the pipeline, removed results based on the quality of the raw sequencing output, while the second was performed manually, intended to retain

variants with a ‘high’ impact, and moderately impactful SNVs with a ‘deleterious’ call by the SIFT pathogenicity predictor, or ‘probably/possibly damaging’ call by the PolyPhen predictor. Across all samples, a total of 187 distinct variants were returned following this processing (Appendix E), with more of these attributed to *PMS2* mutant samples (100) than the controls (87) as expected (p=0.067, 95% CI=0.46-0.61) .

Variant filtering identified 22 pathogenic mutations among the samples analysed, with 21 of these in *PMS2* mutation carriers (Table 5.2.). This includes 15 mutations considered ‘HIGH’ impact, either loss of function (2) or insertion/deletion (13), all reported previously as pathogenic in LS, and associated with a QUAL sequencing quality metric of >8000 in this analysis. Seven of the pathogenic mutations, all with QUAL readings >13000, were further classified as homozygous changes, two associated with a loss of function. These mutations, with VAFs of 0.94-0.99, involve three variants, two frameshift mutations at positions 6037019 and 6038885, the former a founder mutation, and, most frequently a missense mutation at position 6045549. The total of 22 pathogenic mutations does however represent fewer variants than expected of this cohort, especially given the 66 mutation carriers analysed, with this potentially associated with rearrangements and alternative genetic changes that would not be detected in the initial screening of this pipeline.

Sample	Mutated_ Gene	Uploaded_Variation	Location	Allele	Consequence	cDNA_Position	CDS_Position	Protein_ Position	Amino_A cids	Codons	IMPACT	BIOTYPE	SIFT	PolyPhen	EXON	VAF	QUAL
11001	PMS2	rs63750250	chr7:6026564-6026565	T	frameshift_variant	1937-1938	1831-1832	611	I/NX	att/aAtt	HIGH	protein_coding	-	-	11/15	0.376993166	16011.64
15012	PMS2	rs267608149	chr7:6036957-6036958	A	frameshift_variant,splice_region_variant	908-909	802-803	268	Y/LX	tac/tTac	HIGH	protein_coding	-	-	07/15	0.402799378	19259.64
17002	MLH1	rs267608149	chr7:6036957-6036958	A	frameshift_variant,splice_region_variant	908-909	802-803	268	Y/LX	tac/tTac	HIGH	protein_coding	-	-	07/15	0.435368755	20534.64
12005	PMS2	chr7_6037019_-/CTTCACACAC	chr7:6037018-6037019	CTTCACACAC	frameshift_variant	847-848	741-742	247-248	-/VCEX	-/GTGTGTGAAG	HIGH	protein_coding	-	-	07/15	0.314699793	5378.64
17032	PMS2	chr7_6037019_-/CTTCACACAC	chr7:6037018-6037019	CTTCACACAC	frameshift_variant	847-848	741-742	247-248	-/VCEX	-/GTGTGTGAAG	HIGH	protein_coding	-	-	07/15	0.438834951	8612.64
12015	PMS2	chr7_6037019_-/CTTCACACAC	chr7:6037018-6037019	CTTCACACAC	frameshift_variant	847-848	741-742	247-248	-/VCEX	-/GTGTGTGAAG	HIGH	protein_coding	-	-	07/15	0.623100304	32931.64
17033	PMS2	chr7_6037019_-/CTTCACACAC	chr7:6037018-6037019	CTTCACACAC	frameshift_variant	847-848	741-742	247-248	-/VCEX	-/GTGTGTGAAG	HIGH	protein_coding	-	-	07/15	0.629820051	9846.64
16056	PMS2	chr7_6037019_-/CTTCACACAC	chr7:6037018-6037019	CTTCACACAC	frameshift_variant	847-848	741-742	247-248	-/VCEX	-/GTGTGTGAAG	HIGH	protein_coding	-	-	07/15	0.677419355	8644.64
10063	PMS2	chr7_6037019_-/CTTCACACAC	chr7:6037018-6037019	CTTCACACAC	frameshift_variant	847-848	741-742	247-248	-/VCEX	-/GTGTGTGAAG	HIGH	protein_coding	-	-	07/15	0.685618729	8317.64
11017	PMS2	chr7_6037019_-/CTTCACACAC	chr7:6037018-6037019	CTTCACACAC	frameshift_variant	847-848	741-742	247-248	-/VCEX	-/GTGTGTGAAG	HIGH	protein_coding	-	-	07/15	0.737270876	29616.64
13064	PMS2	chr7_6037019_-/CTTCACACAC	chr7:6037018-6037019	CTTCACACAC	frameshift_variant	847-848	741-742	247-248	-/VCEX	-/GTGTGTGAAG	HIGH	protein_coding	-	-	07/15	0.93164557	15363.64
13082	PMS2	chr7_6037019_-/CTTCACACAC	chr7:6037018-6037019	CTTCACACAC	frameshift_variant	847-848	741-742	247-248	-/VCEX	-/GTGTGTGAAG	HIGH	protein_coding	-	-	07/15	0.987654321	13514.06
11015	PMS2	chr7_6038813_G/A	chr7:6038813	A	stop_gained	737	631	211	R/*	Cga/Tga	HIGH	protein_coding	-	-	06/15	0.243224217	15054.6
16006	PMS2	chr7_6038885_-/C	chr7:6038884-6038885	C	frameshift_variant	665-666	559-560	187	V/GX	gtc/gGtc	HIGH	protein_coding	-	-	06/15	0.981519507	42219.06
01135	PMS2	rs121434629	chr7:6045549	A	missense_variant	243	137	46	S/I	aGt/aTt	MODERATE	protein_coding	deleterious(0)	probably_damaging(1)	02/15	0.434036939	10300.6
12061	PMS2	rs121434629	chr7:6045549	A	missense_variant	243	137	46	S/I	aGt/aTt	MODERATE	protein_coding	deleterious(0)	probably_damaging(1)	02/15	0.536959554	12574.6
01123	PMS2	rs121434629	chr7:6045549	A	missense_variant	243	137	46	S/I	aGt/aTt	MODERATE	protein_coding	deleterious(0)	probably_damaging(1)	02/15	0.946428571	18666.03
06040	PMS2	rs121434629	chr7:6045549	A	missense_variant	243	137	46	S/I	aGt/aTt	MODERATE	protein_coding	deleterious(0)	probably_damaging(1)	02/15	0.964412811	26407.03
04122	PMS2	rs121434629	chr7:6045549	A	missense_variant	243	137	46	S/I	aGt/aTt	MODERATE	protein_coding	deleterious(0)	probably_damaging(1)	02/15	0.968879668	18413.03
18023	PMS2	rs121434629	chr7:6045549	A	missense_variant	243	137	46	S/I	aGt/aTt	MODERATE	protein_coding	deleterious(0)	probably_damaging(1)	02/15	0.983050847	26235.03
01130	PMS2	rs121434629	chr7:6045549	A	missense_variant	243	137	46	S/I	aGt/aTt	MODERATE	protein_coding	deleterious(0)	probably_damaging(1)	02/15	0.990532544	35410.03
16039	PMS2	chr7_6045613_G/A	chr7:6045613	A	stop_gained	179	73	25	Q/*	Cag/Tag	HIGH	protein_coding	-	-	02/15	0.403563129	19174.6

Table 5.2. Results of variant calling from the sequencing data of samples identified with pathogenic mutations in *PMS2*.
The single sample of a confirmed *MLH1* mutant carrier is highlighted in yellow, while homozygous VAFs are highlighted in red.

Among the remaining ~170 variants, seven had QUAL sequencing readings >10000 with VAFs consistent with heterozygosity (0.4-0.6) or homozygosity (>0.94). These were all identified within *PMS2* mutation carriers, and have previously been described as pathogenic (ClinVar). All other variant calls returned a QUAL score <3000, with intermediate VAF frequencies (0.1-0.4). A large proportion of these were found in multiple samples from both the *PMS2* and control populations (165 calls accounted for by 29 variants, with 79 being from *PMS2* carriers, 40 from *MLH1* carriers, 43 from *MSH2* carriers and three from *MSH6* carriers). This is consistent with these changes not being pathogenic, and the intermediate allele frequencies they exhibit suggests that these are either artefacts or low-quality calls, rather than true variants.

The further investigation of variants, in the absence of clinical information, involved a comparison of the calls returned from *PMS2* carriers with those from the control population. In preparation for this, the additional filtering of variants was applied so as to include those resulting in a loss of function or amino acid substitution, as well as changes considered deleterious by the SIFT and PolyPhen pathogenicity predictors. Duplicate variants, likely representing linked entries relating to the same lesion, were also removed from review, as were variants present in both control and PMS2 populations. Finally, as the variant QUAL scores are likely to be too relaxed (due to the high read depth and amplicon nature of the sequencing protocol), the VAFs and QUAL data for the remaining calls was compared between the *PMS2* and control datasets, with the expectation that pathogenic variants will be greatly enriched in the former. The results of this inquiry for the control samples is presented in Figure 5.19. (A), and for the *PMS2* carrier samples in Figure 5.19. (B). Almost all variants identified in the control population were clustered in a manner representing both low/intermediate VAFs and relatively low quality scores, with these therefore likely representing false-positive calls from artefacts of the analysis process. (The single sample not contained within this grouping corresponds with the aforementioned pathogenic mutation detected in the sample with MLH1 loss.) A similar cluster of calls with low VAF and sequencing quality was observed in the review of *PMS2* carriers. However, unlike the control population, there is also a population of variants with higher VAF and high QUAL scores, with these corresponding to the potentially pathogenic mutations already identified from in silico predictions alone. Reviewing these variants in the InSiGHT database confirmed their pathogenicity, as each has

been reported pathogenic or likely pathogenic previously, with one exception associated with moderate sequencing outputs.. Although a relatively rudimentary approach for considering the consequence of changes, this logical comparison reveals the difference between the *PMS2* carriers and controls, and supports the presence of pathogenic variants almost exclusively in samples with confirmed PMS2 loss. Nevertheless, in advancing this to the standards of a clinical test, further development would be required, and, given the number of low-quality calls and VAFs progressing through the pipeline, scores would need to be clinically defined.

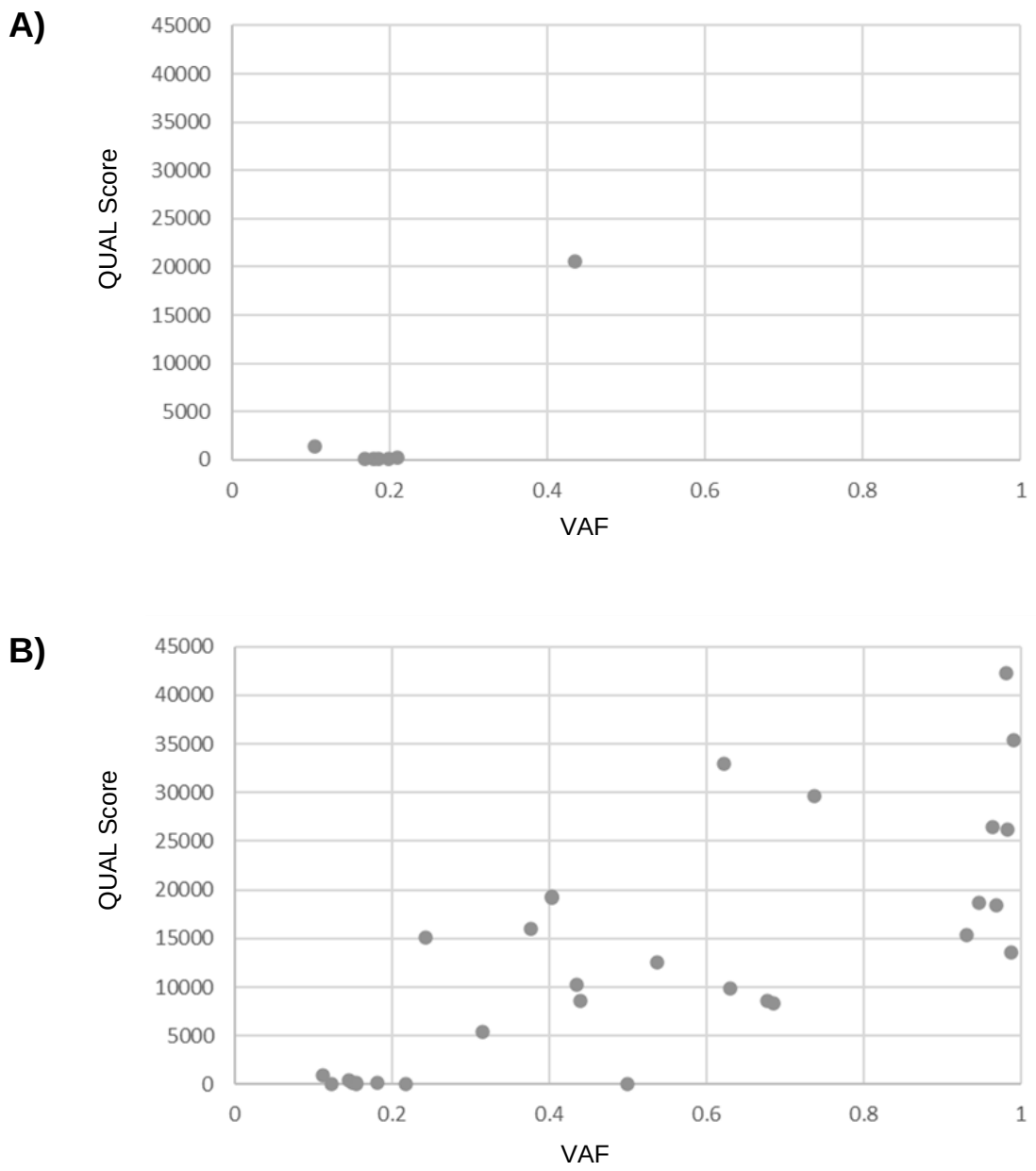

Figure 5.19. Comparison of QUAL scores and VAFs for variants identified by the MIP-based *PMS2* assay in control (A) and PMS2-mutant (B) populations.

5.5. Discussion

The results outlined here describe efforts to develop a sequencing-based assay, employing MIP technology, for the analysis of *PMS2*. In the first instance, it is worth acknowledging the relative success of this approach. In regard to the detection of SNVs, all three samples of the CaPP3 tissue resource with confirmed *PMS2* point mutations were accurately identified, in their correct genomic positions, by this novel assay. Furthermore, in the analysis of a larger cohort of confirmed *PMS2* carriers, more variants were distinguished from these than a corresponding cohort of control samples, with 21 of 22 pathogenic mutations identified in the *PMS2* mutants. Although in some samples homozygous calls were unexpectedly returned, these results evidence the ability of the exon-tiling component of this assay to detect variants in instances of mutation previously defined by other techniques and investigations.

The pipeline of this assay, and the output generated by it, demonstrates how, for techniques akin to this one, precursory specific amplification of the *PMS2* gene is required. The diagnostic problem presented by the *PMS2* pseudogenes, and the necessity for unequivocal amplification of *PMS2* before further analysis, has been outlined previously (Clendenning *et al.*, 2006). To prevent reads from pseudogenes contaminating the data generated from the true coding sequence, a variety of methods have frequently been used sequentially in mutation detection, with several approaches developed from unambiguous PMS2 review. One investigation in particular, in which a missense mutation (as well as other variations) was detected in *PMS2* by an NGS approach, explicitly described how traditional multiplex ligation-dependent probe amplification and Sanger sequencing approaches cannot distinguish whether exonic deletions are present in the coding sequence, especially in the 3' region of the *PMS2* gene which shares considerable homology to the *PMS2CL* pseudogene (Li *et al.*, 2015). In our study, the advantages of preliminary *PMS2* amplification were primarily evident during the development of the exon-tiling probe pool, where more variants were repeatedly returned with amplification directly from genomic DNA template compared with amplification from the long-range PCR products of preparatory *PMS2* amplification. With the additional calls likely attributable to the amplification of pseudogene sequences by MIPs, this shows how, without specific PMS2 amplification,

the analysis of artefacts may occur in place of the true coding sequence. Furthermore, this highlights why, when using such NGS methods, this combination is required.

The detection of *PMS2* variants by our assay was not exclusive to samples with confirmed *PMS2* mutations however, with a single pathogenic change detected in a control sample with a confirmed *MLH1* mutation (one in which no *PMS2* variation was therefore assumed). While atypical, this is not entirely unexpected, especially as the MLH1 and PMS2 proteins are binding partners within the human MMR system, forming the heterodimer MutLα responsible for the repair of DNA mismatches. Immunohistochemical MLH1 loss may be attributed to mutations in the *MLH1* gene - with the subsequent designation of a confirmed *MLH1* mutant - without genetic analysis, while it is actually an undetected *PMS2* mutation that results in this loss of expression through the absence of the MLH1 binding partner. Alternatively, the occurrence of concurrent *MLH1* and *PMS2* mutations may explain this finding, with such reported in up to ten samples in a previous study of the gene-specific alterations of 1057 MSI-H solid tumours (Salem *et al.*, 2020). With mutations in the former regarded more pathogenic given the acknowledged higher penetrance of such, *MLH1* variation may be regarded as explaining any perceived pathogenicity in instances of both *MLH1* and *PMS2* mutation, or only variations in this gene may be detected by approaches that do not discriminate between *PMS2* and its pseudogenes (and therefore fail to analyse the PMS2 coding sequence). These aforementioned issues, as evidence by the detection of a pathogenic mutation in a control sample in our study, may however be circumvented by approaches for specific *PMS2* review such as described in this chapter.

With only 21 pathogenic mutations returned from 66 *PMS2* carrier samples, the detection of variants by our assay is lower than expected. These results contrast those reported in the literature, with one study of 61 LS cases attributed to *PMS2* deficiency describing heterozygous mutations in 90.16% of cases, and homozygous mutations (including two deletions and four point mutations) in the remaining 9.83% (Senter *et al.*, 2008). Reasons for this discrepancy potentially include the analysis of pseudogene variants in older studies that inflate the number of mutations ascribed to the true *PMS2* gene. Alternatively, fewer mutations may have been detected/prescribed by our assay

as a result of the filtering applied in variant calling within the pipeline. The sequencing depth and quality of reads for true variants may have been insufficient by our approach so as to exceed the stipulated performance thresholds, and would therefore have been removed from the eventual analysis output. Alternatively, as the read depth by our approach is considerably higher than that of standard NGS, it is probable that the GATK QUAL thresholds applied in analysis were too lenient with this data. Manual filtering of data may also have removed affecting *PMS2* variants from our output, with mutations designated 'low impact' by pathogenicity predictors (based on previous experimentation and investigation by other research groups) withdrawn from analysis.

Nevertheless, in samples from confirmed *PMS2* variant carriers, 21 pathogenic mutations were identified, with several recurring across multiple samples. These include the common founder mutation rs121434629, responsible for the protein change Ser46Asn, reported previously in several studies (Senter *et al.*, 2008; Tomsic *et al.*, 2013), and the insertion of the sequence CTTCACACAC at position 6037019, described as responsible for a deleterious frameshift (Lagerstedt-Robinson *et al.*, 2016). The mutation rs267608149, present in two samples including the *MLH1* control, results in the frameshift Tyr268fs, bringing about a premature stop codon and absence of protein product (Senter *et al.*, 2008; Duzkale *et al.*, 2013; Rosty *et al.*, 2016). With all mutations associated with a history of findings and research describing their pathogenicity, their detection by our assay validates its capacity for accurate variant calling in the *PMS2* coding sequence.

This *PMS2* assay is not without limitations however, with these primarily resulting from the time-consuming nature of the approach (which consists of multiple stages, many proceeded by essential purification), and the requirement for high quality material from which to long-range PCR is performed. Further shortcomings seemingly exist with the detection of larger genetic changes, with questions therefore as to whether this pipeline offers sufficient advantages to current methodology. Originally, the analysis of SNPs was conceived as a convenient approach for identifying a complete drop out of a single amplicon allele. But the existence of a considerable number of variants with moderate VAFs (10-30%) implies that the PCR reactions involved are inefficient and prone to generating artefacts, making some SNP calls uncertain. In addition, the

presence of superfluous bands observed in some PCR reactions suggests that a direct sequencing approach would be more definitive, especially now accurate long-read methods are becoming available. While it may be feasible for read depth to be used as a surrogate for the deletion of specific exons or exon combinations, time constraints meant that this was not investigated within the samples with known deletions. However the broad range of primer efficiencies previously described indicates that this may not be straightforward to implement, or even possible.

Comparing it to existing techniques for *PMS2* evaluation, both our assay and modified MLPA approaches are capable of discriminating between the *PMS2* sequence and its pseudogenes. However, with reference DNAs containing equal quantities of *PMS2*- and *PMS2CL*-specific sequences, the latter is also able to reliably identify copy number alterations (Wernstedt *et al.*, 2012). Reflex workflows for variant discovery, involving hybrid-capture probes and filtering using gene-specific variants, have also been shown to return an analytical sensitivity and specificity of >99% specificity for the detection of SNVs and short indels in *PMS2* across 299 hereditary cancer screening samples, as well as >96% sensitivity and >99% specificity for the detection of copy-number variants (Gould *et al.*, 2018). Finally, long-range sequencing approaches such as Nanopore sequencing have been demonstrated to accurately confirm recurrent *PMS2* variants with a single read accuracy of 96.8% following pairwise alignment to the corresponding reference allele, all using a relatively time-efficient approach (Watson *et al.*, 2021). Our assay is inherently limited in its current configuration, and the juxtaposition of this with existing technologies for *PMS2* analysis queries the place for this as a prevalent testing option. It is worth noting that the development of such commenced in earnest before significant advances in many of the alternative approaches previously described, and improvements in our *PMS2* assay may ultimately further its capabilities. Whether these changes make it a viable testing option however, especially given the rate of advance of competing approaches, remains debatable, and the amount of development that would be required probably precludes further expenditure on this assay.

Should efforts be made in the future to further develop this assay, in the first instance investigations into the relatively low frequency of pathogenic mutations in confirmed *PMS2* mutants would be necessary, with this enabled by access to details for all the samples reviewed. At the time of analysis, specific mutation details, besides

the MMR gene affected, were unavailable for each sample of the final investigation, with a lack of access to clinical information which proved difficult to rescind. However, in the next year the CaPP3 clinical trial will be essentially ‘unlocked’, with the clinical data of all constituent samples becoming accessible. With knowledge of the specific variants confirmed responsible for pathogenicity in the sample analysed, it could be determined as to whether point mutations have been missed, or whether the 21 pathogenic changes identified are how many were present in this cohort of 66 *PMS2* carrier samples. It is possible pathogenicity in the remaining samples results not from smaller genetic mutations, but from other changes (such as copy-number variation) that are not detected by the initial analysis of this pipeline. However, it is also possible that many of the *PMS2* LS cases recruited to CaPP3 have been mis-classified due to errors in the short read sequencing approach.

Measures to assess samples for copy-number variants and a loss of heterozygosity involved the analysis of intronic SNP data in our *PMS2* assay. The application of this strategy was ultimately limited, with what information could be drawn from it indicating no evidence of homozygosity, including in the two *PMS2* samples with a confirmed exon 10 deletion. Data from SNP analysis, while instructive, was also of a lower quality than other sequencing data from this assay, with this confounding any conclusions that could be deduced from it. This situation is not exclusive to our assay, with a previous study of 8349 SNP-targeting MIPs finding imprecision in the calls returned for approximately 20% of the probes when analysing 104 DNA samples from healthy individuals (Biezuner *et al.*, 2022). Improvements in the review of intronic SNPs, to supplement the findings of exon analysis, may provide explanation for many of the remaining samples with confirmed *PMS2* mutations that did not present in this study. These changes may involve including more SNPs in this analysis, with the subsequent establishing of suitable thresholds for the classification of a loss of heterozygosity, as well as sequencing SNPs to a greater depth, so as to have more confidence in the data returned.

Improved SNP analysis may further allow for a more detailed review of pathogenic mutations returned with homozygous VAFs, six of which were identified (representing three variants) in the carrier investigation of this study. While it is possible that CMMRD hypomorphs exist to explain this phenomenon, as observed in previous studies of homozygous *PMS2* variants (L. Li et al., 2015; Gallon et al., 2019),

the presentation of these is limited, and these findings require further exploration. No concurrent loss of heterozygosity strongly suggests that such findings may be artefacts, particularly as we would expect to observe the same linked SNPs on both alleles for these samples (which is not the case). The results of amplification directly from the gDNA of samples with homozygous variant calls further submits this not to be the true situation for this material. This could be caused, for example, by the drop out of one allele in a specific PCR reaction, either due to incidental variation/problems with the LR-PCR, or as a result of a SNP within the probe sequence. However, as *PMS2* has multiple pseudogenes, localised gene conversion could also be responsible for the absence of any linkage between neighbouring SNPs. Given the wide variation reported in *PMS2* penetrance ,and the relatively high frequency of *PMS2* variants in the general population, this possibility requires consideration regardless. .To unambiguously investigate this further, the resequencing of samples could be undertaken using technologies such as Nanopore or Promethion, as long reads should capture both alleles (if two are available). If successful, these high fidelity, long-read sequencing approaches, which have only become available relatively recently, could also be applied to other PMS2 samples to confirm variants.

Continuation of this work, should it be deemed practicable and valuable, would involve development of the analyses previously described, and, while potentially convoluted, this should be achievable. Much of the work required to progress this study was not conducted given time constraints, with the review of *PMS2* and control samples only occurring towards the end of my PhD when this material became available and the analyses viable. However, the capacity of this test to assess the PMS2 coding sequence, and the benefits it offers, have been demonstrated, while samples have been identified by such that would be interesting to follow up with potential clinical implications.

Chapter 6. General Discussion and Future Work

6.1. Lynch Syndrome and its Detection

Lynch syndrome (LS) describes the most common inherited cancer syndrome, arising from autosomal-dominant germline mutations in genes of the DNA mismatch repair (MMR) system. LS individuals experience a significant predisposition to, and early onset of, various cancer types, the most common being those of the colorectum and endometrium, with cumulative incidences of up to 57.1% and 48.9% by age 75 years respectively despite surveillance (Dominguez-Valentin *et al.*, 2020). The percentage of cases observed for other cancer types varies considerably, ultimately dependent on the underlying genetic mutation, but current estimates of LS prevalence in the general population are 1 in 279 (Win *et al.*, 2017).

To optimise their clinical management, the efficient diagnosis of LS individuals is essential. Known carriers of LS-associated gene defects may benefit from various practices, including genetic counselling, regular tumour surveillance, personalised cancer treatment, and/or, as more recently demonstrated, daily aspirin prophylaxis. Current guidelines in the UK therefore advocate the screening of all CRC and EC cases for LS regardless of age, clinical presentation or family history, with two standard pathways available to that end. The first of these pathways, employed for the analysis of both colorectal and endometrial cancer, uses MMR protein IHC for LS detection (Figure 6.1.) and is usually completed with 7-10 days (NHS England, 2023). In contrast, the second pathway using DNA MSI testing exclusively for CRC analysis has a variable timescale for review depending on the specific assay used, with results typically communicated to the patient within four weeks (NHS England, 2023): this is presented in Figure 6.2.. At a cost of £202 per patient (as of 2016/17), the latter alone is cheaper than IHC alone (costing £210 per patient), but, when using MSI testing, all MSI-High tumours must also have *MLH1* methylation testing at an additional cost of £136, while IHC separates MLH1-deficient cases from MSH2-/MSH6-/PMS2-deficient cases in the first instance (and so *MLH1* methylation is ultimately performed on fewer samples overall) (Snowsill *et al.*, 2019). Both approaches have been shown to realise equivalent diagnostic performance (Dedeurwaerdere *et al.*, 2021).

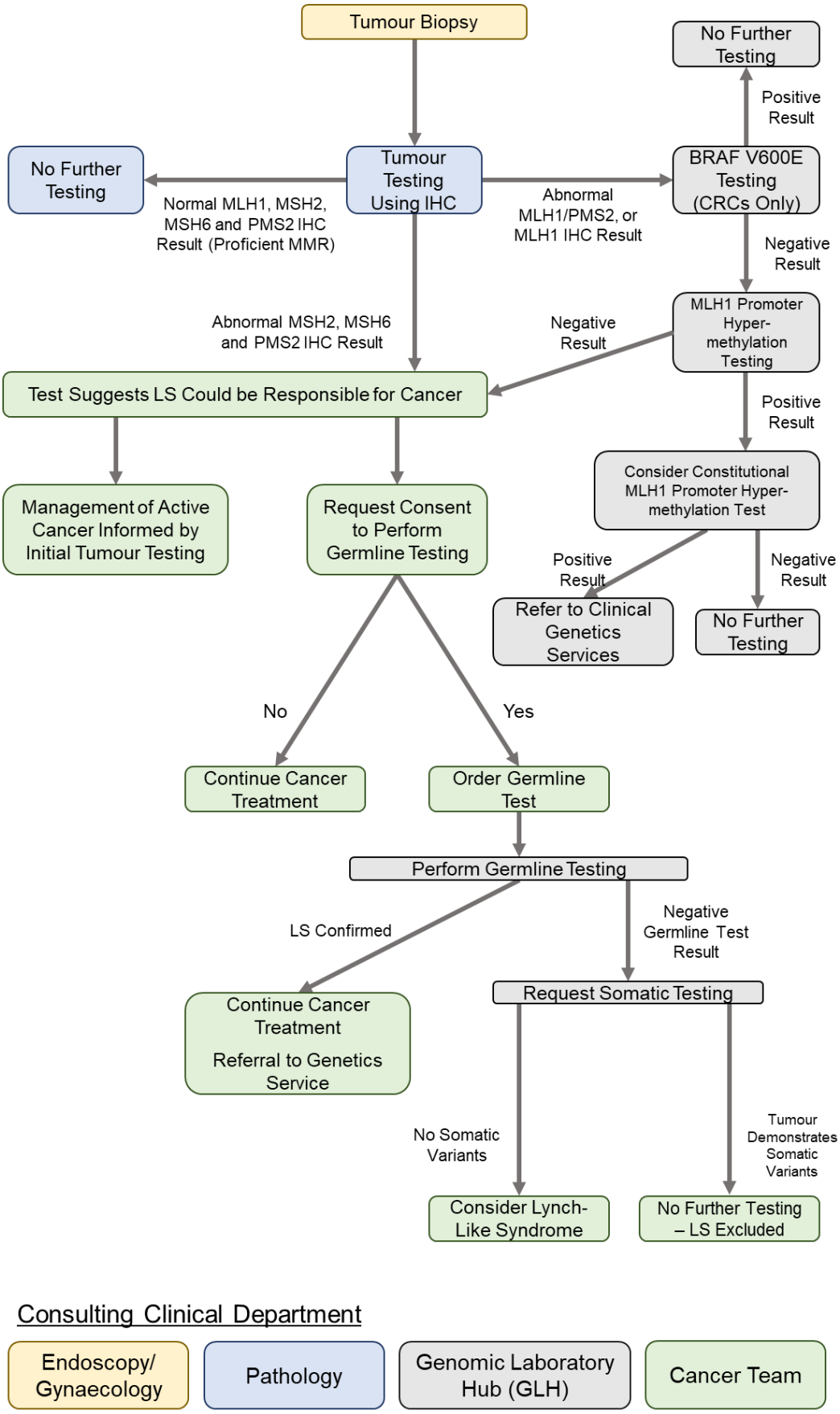

Figure 6.1. Standard LS testing pathway using MMR protein immunohistochemistry (IHC). (Adapted from NHS England (2023).

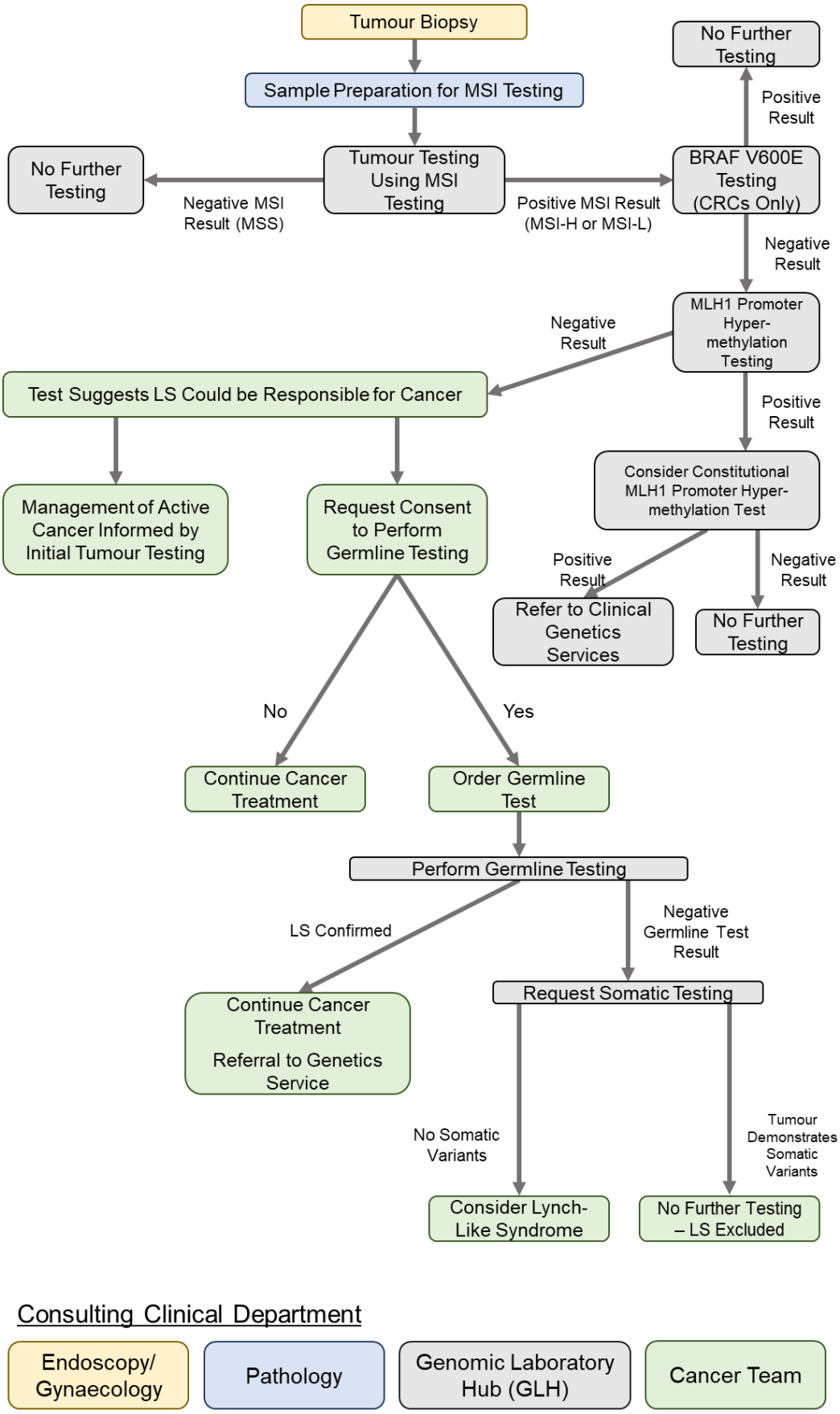

Figure 6.2. Standard LS testing pathway using DNA microsatellite instability (MSI) testing. (Adapted from NHS England (2023).

Several limitations compromise the current clinical guidelines however, including the restricted tumour spectrum analysed, while recommendations also fail to be implemented consistently and universally, possibly due, in part, to failings in the existing assays. Difficulties in the detection of mutations in *PMS2*, an MMR gene in which sequencing and variant detection is complicated by the presence of pseudogenes, epitomise such inadequacies, and highlight the need for improved practices to recognise those with LS, for whom therapeutic options like aspirin administration are now available.

Given the aforementioned limitations, the primary aim of this study was to improve the techniques for detecting MMRd and LS individuals, therefore advancing the identification of potential candidates for prophylactic aspirin therapy.

6.2. Use of a Sequencing-Based Assay for the Detection of Mismatch Repair Deficiency in Extracolonic Cancer

NICE Diagnostic Guidance 27, first published in 2017, advocates the testing of all CRCs for MMR deficiency to screen for LS, either by loss of MMR protein expression or the detection of exceptionally-high MSI in a given tumour.

There are however several other forms of cancer generally considered associated with LS, where an increased risk of occurrence in LS carriers is now established. These include those of the endometrium, upper gastrointestinal tract (stomach, pancreas and small bowel), ovaries, sebaceous glands and urothelial tract (bladder, ureter and kidney) (Vasen *et al.*, 2013). Between these tumour types, estimates from the literature of the frequency of cases caused by LS varies, while the cumulative incidence of cancer by the age of 75 years also differs. As a result of these inconsistencies, no official guidance for MMRd testing of these cancers existed when this work commenced, with this potentially representing a missed opportunity for clinical intervention.

In extracolonic cancers, MSI detection has previously been demonstrated, but the frequency of the phenotype, although sufficiently high for observation, varies by tumour type. The reported findings of instability in extracolonic tumours are however questionable, arising in studies limited by the use of subjective, outdated approaches, or restrictive sample sizes. Nevertheless, given the aforementioned incidence of LS,

and the shared fundamental biology of repair deficiency responsible for predisposition, it is conceivable that current screening guidance for LS could be extended to additional tumours of the Lynch spectrum, and exploit the molecular phenotype of exceptionally high MSI with which, although inconsistent, they have been shown to present.

My studies of extracolonic cancers primarily considered the ability of the Newcastle Assay to identify instances of MMRd, using high levels of microsatellite instability as a biomarker for a loss of protein function. This approach returned interpretable scores, classifications of MSI and frequencies of this across the aforementioned range of tumour types, especially those considered associated with LS. Furthermore, the concordance between repeated analyses evidenced the reproducibility of this analysis, as well as the ubiquity in detection of the phenotype in an LS context. The frequencies of MSI classification, and the distribution of scores, by the Newcastle MSI Assay were generally consistent between tumour types, and comparable to the observations with CRC as well as corresponding values reported in the literature. However, three exceptions to this congruence exist, provided by endometrial, breast, and ‘other skin’ cancer.

Of particular interest is the prevalence of EC readings designated MSS (26%), despite all samples originating from confirmed LS patients, and the use a highly-sensitive assay. Consistent with existing literature (Gylling *et al.*, 2008), this shows that genetic instability is not exhibited ubiquitously in instance of deficiency, and that the presence of an MSS tumour does not exclude an LS diagnosis. While questioning the involvement of MSI in Lynch EC, this may imply a different mechanism, independent of MMRd, by which tumours frequently develop in the condition. Alternatively, this could indicate a different way, independent of instability, in which MMRd may promote tumourigenesis, such as through the processes of homologous recombination or immunoglobulin class switching in which MMR genes are also implicated (Li, 2008). Nevertheless, conclusions from these EC findings were ultimately limited without the corresponding results of IHC and other analyses, and highlighted the need for further investigation.

6.3. Use of a Sequencing-Based Assay for the Detection of Mismatch Repair Deficiency in Endometrial Cancer

Since the start of my PhD, similar guidance for the screening of LS carriers, as issued for instance of CRC, have been communicated for all individuals diagnosed with EC, a tumour type in which approximately 3% of cancer cases were found attributable to LS (Ryan *et al.*, 2019). This follows the recommendations of gynaecologists and international collaborations, including those of the Manchester International Consensus Group, which intended to provide standardised clinical direction for both the screening and treatment of EC in the context of LS (Crosbie *et al.*, 2019). However, unlike in the testing of CRC, the analysis of endometrial samples is advised, in the first instance, to be performed exclusively by IHC, with the recommended diagnostic pipeline not incorporating MSI testing. The decision to exclude MSI analysis from clinical guidelines results in part from the reported lower sensitivity of this technique for MMRd detection in ECs, relative to IHC, observed in several studies using either Bethesda panel analysis or the Promega five marker assay (Rubio *et al.*, 2016). Nevertheless, the utility of this approach in CRC has been repeatedly demonstrated, with other independent studies describing the suitability of using both IHC and MSI analysis for the determination of MMR status in EC (Stelloo *et al.*, 2017; Stinton *et al.*, 2021). The absence of MSI testing from clinical guidelines may therefore be a result of disadvantages associated with the techniques available for this, such as their expense, complexity, or subjective nature.

My studies with EC primarily explored the capacity of the Newcastle Assay to identify instances of MMRd, using high levels of microsatellite instability as a surrogate for a loss of protein function. In the first instance, using samples provided by Ohio State University, levels of MMRd detection were returned comparable to those from Promega MSI testing of the same samples, as well as those observed in separate studies of other EC cohorts (McConechy *et al.*, 2015; Stelloo *et al.*, 2017). Furthermore, concordance with previous IHC and MSI reviews performed in Ohio was high (94.90% and 97.45% respectively), and not significantly different form the concordance between the results of the two techniques originally reported for the same samples (96.43%).These findings demonstrate the ability of the Newcastle Assay to detect repair deficiency, which it achieves as efficiently as the current standard for MSI testing (Promega), while offering the benefits of an automatable and scalable process without the requirement for expert interpretation of results.

Considering the wider application of MSI testing for the diagnosis of MMRd in EC, the high accuracy of detection observed in this study of the Ohio samples (with a

sensitivity and specificity of ~95% relative to previous IHC), and other analyses of this tumour type, evidence the feasibility of this approach. Moreover, in recognising repair-deficient samples missed by IHC, as occurred in the evaluation of material from the University of Ohio, there is an argument that MSI could be included in the initial screening of endometrial tumours for the maximum recognition of MMRd, as is the case in CRC analysis. The results of the Manchester samples, however, also reveal complexities that would be required in any expansion of clinical guidance, especially in appreciating the cost-effectiveness of this approach in a dual testing strategy with IHC.

6.4. Resolving the Varying Detection of Repair Deficiency in Endometrial Cancer by the Newcastle Assay

In contrast to the results from analysis of the Ohio samples, the review of ECs from the University of Manchester by the Newcastle Assay realised a test sensitivity of 77.17% and specificity of 89.33%. Furthermore, this assessment gave a concordance with previous IHC classifications of 82.63% which, while comparable to the 76.65% concordance between Promega MSI and IHC obtained in the original analysis of these samples, is significantly lower than the value observed in previous Ohio EC analysis. Consideration of the two EC cohorts analysed revealed few differences to explain the divergent results returned, with a similar stage distribution of the tumours tested, as well as similar methods by which previous MSI testing defined repair deficiency. However, in this appraisal two factors did emerge which could contribute to these findings of our analysis: how samples were prepared for assessment by the Newcastle Assay, and the grounds on which patients were identified for inclusion in the corresponding studies from which these samples were provided.

Considering sample preparation initially, material from the University of Ohio was provided in the form of purified DNA, in aliquots from which the previous assessment of these samples was achieved in Ohio. As a result, the same template was effectively tested in both prior analyses and those conducted in Newcastle, with these expected to yield equivalent results given the same genetic situation, and therefore genotype, is being reviewed. Conversely, analysis of tumours from the

University of Manchester required the preparation of fresh DNA aliquots, as samples were provided in the form of FFPE material. For that reason, the testing of samples in Manchester effectively used a different DNA template than that evaluated by the Newcastle Assay for the same samples. Inevitably, the possibility therefore exists that a different genetic context is being assessed in each analysis. Intratumoural genetic heterogeneity like this has previously been described, with one study reporting between zero and 8000 coding mutations to be heterogenous within the primary tumours they evaluated (Johnson *et al.*, 2014). In studying different sections of a given tumour, different situations of genetic instability in particular may be reviewed, with this potentially explaining any difference between the results of testing using different preparations from the same sample. Furthermore, the extraction of DNA for analysis in Newcastle was performed up to seven years after some of the original extractions from which the Manchester results were generated. In this time, the FFPE material of these samples will have naturally degraded, and the effect of this on the results of our assay has not been rigorously studied. While any impact on sample classification would be expected to be relatively minor, this could explain some of the inconsistent results, especially following re-training, where using more markers, or EC samples in classifier training, increased discordance relative to the original CRC-trained classifier.

An alternative explanation for the differing results observed between analysis of the Ohio and Manchester samples emanates from the different ways in which candidates were selected for the studies from which samples were recovered. For the Ohio cohort, samples represent individuals diagnosed with primary invasive EC, or individuals who had experienced a hysterectomy/diagnostic biopsy confirming a newly diagnosed EC, with all of these prospectively enrolled across two studies. In contrast, samples of the Manchester cohort originate from the PETALs study, with this comprising tumours from women diagnosed with EC or atypical hyperplasia (AH), a precancerous condition preceding endometrial tumourigenesis. In the latter, given the involvement of hyperplasic individuals, it is possible that samples were included before the appearance of tumour characteristics, such as the development of genetic instability like MSI. For a sufficient MSI signal to be available for analysis, a later stage of tumour maturation (than that afforded by AH) may be necessary, and testing tumours before this may explain the lower levels of MSI observed for the Manchester samples, many of which may not yet have progressed sufficiently for this to materialise.

A further difference between these two EC cohorts is afforded by the manner in which MMRd was defined by their respective studies. For the samples from the University of Ohio there was an unambiguous definition for this, with staining ≥1% considered MMRp on the assumption that repair-deficiency will be identified by MSI testing. (Given the importance of MMRd to LS detection and treatment, use of both methods has advantages, and both are standard of care in the US). In contrast, there was no clearly defined staining threshold for the classification of deficiency in the Manchester study, with 'patchy' IHC results being associated with both MMRd and MMRd, with classification subsequently devolved to the expert judgement of the pathologists, and repeated, if necessary, blinded to MSI status. This nuance provides a significant distinction between the two cohorts, and may explain some of the differences between the results produced for each, particularly those related to concordance with previous assay results.

As referenced earlier, a further issue highlighted primarily with the Manchester samples was the failure to detect instance in which *MSH6* loss is responsible for repair deficiency, and therefore MSI. The frequency of inherited *MSH6* mutation in EC has been reported on many occasions previously, with one study of 441 samples, unselected for age or family history, returning a prevalence for this of 1.6% (Goodfellow *et al.*, 2003). However, our investigation relates a failure to identify *MSH6*-deficient samples using MSI testing, with the loss of this protein seemingly not resulting in exceptionally-high levels of genetic instability as is observed with mutation of the other MMR genes. For this, there are various plausible explanations, including situations in which the loss of *MSH6* is partially compensated for by the presence of *MSH3*. This MutS homologue, which dimerises with *MSH2* to form a repair complex with specificity for larger indels, has been shown to provide some redundancy in the MMR system, and attenuate deleterious *MSH6* mutations (Acharya *et al.*, 1996; Risinger *et al.*, 1996). However, the analysis of MSI in tumours with isolated MSH6 loss reveals a generally lower signal of instability in *MSH6* carriers, with this potentially low enough in instances so as to prevent the detection of mutants by our assay (and likely other current MSI testing approaches). Complications, such as those presented by *MSH6* loss, represent circumstances that require attention in the future analysis of ECs before the ultimate recommendation of MSI testing clinically for this tumour type. Nevertheless, these issues represent a context in which MSI, while informative, is not

practicable for EC analysis, and therefore supports the use of IHC as the primary approach for MMRd diagnosis as per current clinical guidelines. They also give credence to the notion that considering LS as a single condition is out dated, with there actually being four distinct dominantly-inherited microsatellite-instable cancer syndromes attributed to each of the four traditional LS genes (Møller *et al.*, 2023). Should these have divergent or individual phenotypes, detection of all by the same technique may not be viable, a potential explanation for the lower detection of MSI with *MSH6*-loss specifically (if MSI is not a universal presentation with this flavour of LS).

Continuing the analysis of EC samples by the Newcastle MSI Assay, it is requisite to address several matters arising from this study thus far, to allow definitive conclusions about the facility of this approach, and the use of MSI as a biomarker in this tumour type, to be formed. The first, and arguably most important, of these is the necessary further investigation of the difference in results observed between the analysis of the two EC cohorts, with a determination of a reason for this. While potential explanations have been presented in Section 6.2., these remain speculation based on the literature and information known at the time of writing. Further discussion with the providers of this material will likely be required to dictate the direction of this research, but, in the first instance, additional clinical information would be convenient for the samples in which discrepancy with previous testing results, particularly IHC, was found. By identifying commonalities from this between discordant samples, the reason(s) for contradictory assay results may become apparent, with an explanation for a lack of instability in repair deficient samples (or vice versa) revealed. The provision of the initial sample aliquots (from which previous analyses were performed) from the University of Manchester could also help in establishing such an explanation, while presenting material from which analysis could be performed to more accurately compare the performance of Promega MSI and Newcastle MSI testing approaches. However, given the time since the Manchester study, it is possible that this original material may no longer be available, and therefore it may not be possible to resolve the aforementioned issues with these samples.

From the University of Ohio, more detailed information on the stage and grade of tumour samples could be informative, with this known in detail for the Manchester samples, but only summarised for the Ohio cohort. While this exercise would require

new ethical approval, access to these details may reveal patterns in the limited number of samples for which there was disagreement between the Newcastle Assay and previous Ohio analyses. Furthermore, in conjunction with the material from Manchester, this information could be used to regard the involvement of MSI in endometrial tumour development, to assess any relationship between MSI signal and tumour stage/grade, and subsequently whether instability occurs later in EC tumourigenesis as suspected.

6.5. Use of a Sequencing-Based Assay for the Analysis of *PMS2*

For an unqualified diagnosis of LS the confirmation of a germline pathogenic MMR variant is ultimately required. This process may however be complicated by several factors, including the inadequacies of current technologies to recognise causative variants. The assessment of *PMS2* exemplifies limitations in the existing techniques to accurately identify pathogenic MMR mutations. Located on chromosome 7 and consisting of 15 exons, the post-meiotic segregation 2 (*PMS2*) gene encodes the corresponding MMR protein, mutations in which have been described associated with an increased cumulative risk of approximately 13% for both EC and CRC (ten Broeke *et al.*, 2018). Analysis of *PMS2* is fundamentally complicated by the presence of 13 pseudogenes, situated on the same chromosome, which effectively contaminate the sequencing of the true gene sequence and reduce the sensitivity of mutation detection (Clendenning *et al.*, 2006). To circumvent sequencing issues, various techniques have been developed. However, these approaches are not without limitation, primarily a lack of scalability, but the latter, with refinement, may provide a useful macro-sequencing option and the initial phase in improved *PMS2* testing.

My studies regarding *PMS2* involved the development of a sequencing-based assay, using MIP technology to review long-range amplicons, for the specific analysis of the gene's sequence without contamination with pseudogene sequence. A MIP pool was established and balanced consisting of 42 exonic and 100 intronic probes to assess for sequence variants and CNVs/a loss of heterozygosity respectively, with this then used for sample analysis. In the review of samples for which *PMS2* mutation was confirmed, the capacity for the detection of variants by the exon-tiling component of this assay was demonstrated. Point mutations of the initial samples from the CaPP3 tissue resource were accurately identified by this approach, while more variant calls

were returned for the 66 samples with verified *PMS2* mutations than the 72 corresponding controls in the ultimate investigation of this study. Both results show that the exon-targeting MIPS of this assay are amplifying their intended targets, and, although findings are yet to be corroborated, that such has the facility to identify single-nucleotide and insertion-deletion variants. The detection of pathogenic mutations in particular further justifies the filtering incorporated in the interpretation of sequencing data by this pipeline, though confirmation of these relative to previous findings is again required.

Beyond variant calling by its exon-targeting component, the proficiency of this assay, particularly for the detection of copy-number variation and the loss of heterozygosity, remains to be analysed, curtailed due to time constraints. Information derived from MIPs targeting intronic SNPs has thusfar been inconclusive, and in no samples, including the limited number with confirmed larger deletions, has homozygosity or substantial deletions been indicated. Difficulties in amplifying intronic SNPs using MIP technology is not unique to our study, with a previous investigation using 8349 SNP-targeting MIPs finding imprecision in the calls returned for approximately 20% of the probes (Biezuner *et al.*, 2022). At present, similar shortcomings of our *PMS2* assay remain to be determined, but extensive variation in VAF frequencies, and variation in read counts, suggests this would be difficult to achieve.

6.6. The Current Perspective for *PMS2* Testing

At the outset of my PhD studies, various options were available for the specific interrogation of *PMS2*, with several involving similar technologies and/or rationale as employed in our assay. These approaches include cDNA sequencing from lymphocyte cultures (puromycin-treated to inhibit the nonsense mediated decay of mRNA), which has previously been shown to discriminate between *PMS2* and its pseudogenes, as well as detect splice variants and *PMS2/PMS2CL* ‘hybrid alleles’ (Etzler *et al.*, 2008). Systems using MLPA have also been developed that enable the identification and localisation of CNVs, decreasing the number of cases requiring further mutation analysis (Herman *et al.*, 2018). Despite their advantages, for each of these techniques

there are limitations, primarily in their scalability, with both constrained by non-automatable protocols and their labour-intensive nature. Given such, there existed the necessity to develop alternative assays which, while being specific for *PMS2*, were not limited in the same manner.

During the course of my PhD however, several methodologies were developed that allow for the specific analysis of *PMS2*, with advantages so as to supersede our assay. Principal among these is long-read sequencing (LRS), such as is offered by the platforms Nanopore and PromethION. In this process, the direct, real-time analysis of 'long' DNA fragments is performed without the requirement for preparatory PCR amplification or chemical labelling of a sample, with this therefore a relatively low-cost and quick sequencing option. Knowing the genetic coordinates of *PMS2*, this approach may be used for the specific review of the gene's sequence, with the data from this analysed by coding pipelines, similar to those of our assay, for variant calling and pathogenicity prediction. A recent study demonstrated the capacity of this technology for the accurate confirmation of a recurrent *PMS2* insertion-deletion variant, with 100% sequence identity returned in review of the generated consensus nanopore sequence, albeit for a single sample (Watson *et al.*, 2021). An alternative technique for variant discovery specifically is afforded by reflex workflows, which involve hybrid-capture probes and filtering using gene-specific variants. This approach has been shown to return an analytical sensitivity and specificity of >99% specificity for the detection of SNVs and short indels in *PMS2*, as well as >96% sensitivity and >99% specificity for the detection of copy-number variants (Gould *et al.*, 2018).

While neither aforementioned methodology is perfect, both remain relatively novel technologies with numerous advantages, and, if they continue to develop at the rate they have in recent years, the need for *PMS2* assays such as this, with its numerous limitations in its current form, may be negated.

6.7. Issues with Our Sequencing-Based Approach to *PMS2* Analysis and Continuation of this Assay's Development

As with EC analysis by the Newcastle Assay, the continuation of my work regarding the development of a sequencing based *PMS2* assay would primarily involve

addressing two matters, with the first of these being the confirmation of *PMS2* mutations identified in variant calling thus far. As previously described, in the final analysis of samples with confirmed *PMS2* mutations, numerous variants were called by our assay that could not be fully verified, given limited access to the specific clinical data for this material. As a result of this, the accuracy of our *PMS2* assay for variant detection could not be determined, and results remain the speculation of our approach. Fortunately, in the next year the CaPP3 clinical trial will be essentially 'unlocked', with the clinical data of all constituent samples becoming accessible and available for comparison to our findings. The complete clinical details for such material is necessary in the interpretation of long-read NGS data, and will allow for putative homozygotes, with variant calls unexpectedly considered homozygous, identified by our assay to be followed up. Furthermore, with knowledge of the specific variants confirmed responsible for pathogenicity in the samples analysed, it could be established as to whether the accurate detection of exonic variants by our assay was achieved. This is particularly important for the 15 mutations considered pathogenic, following variant calling, by the integrated pathogenicity predictors of a pipeline, with such effectively defining the utility of our assay for this purpose.

Access to the aforementioned clinical data would also resolve the lower-than-expected rate of pathogenic mutations in the samples analysed. With 66 samples considered for which deleterious variants have been reported, more than 15 pathogenic SNVs would be anticipated so as to explain this situation. It is possible pathogenicity in the remaining samples results not from smaller genetic mutations, but from other changes (such as copy-number variation) that are not detected by the initial analysis of this pipeline. However, the literature suggests this would not be the case, with one review of 197 *PMS2* samples from the InSiGHT database describing missense and nonsense/frameshift changes in 62% and 4% of instances respectively (Peltomäki, 2016). By having access to patient records including mutation data, it could be determined as to whether point mutations have been missed by our assay, highlighting a fundamental failure of such in its current arrangement, or whether the pathogenic changes in PMS2 originally reported in this group of 66 recruits reflect sequencing errors at original diagnosis due to pseudogene interference.

Should this study continue, the second matter to attend to, referred to earlier, is the apparent failure of the current iteration of the assay to detect copy-number variants and/or a loss of heterozygosity. Described previously, MIPs targeting intronic SNPs,

from which information is derived for calculations, are used in tandem by our *PMS2* assay to detect larger changes or deletions in the gene. Findings using this approach have been related, with the power calculations based on linkage disequilibrium for samples with a loss of heterozygosity not returning this result, albeit in a time-limited review.

With a considerable proportion of pathogenic mutations in the gene related to copy-number variants or a loss of heterozygosity, it is necessary for any *PMS2* assay to identify such situations, and, for that reason, changes to our assay are required to rectify this current failing. Involving more MIPs to target more SNPs may offer a solution, with this providing more information in subsequent power calculations, and, from such, the definition of more accurate thresholds for a designation of homozygosity. Alternatively, the sequence depth to which SNP sequences are analysed could be increased, giving more coverage of proposed variants, and more confidence in the data retrieved from them for calculations. Exploring potential improvements to this assay would not be a trivial exercise however, with the design and testing of SNPs a convoluted process, and analysis for the establishment of classification thresholds requiring considerable time and data.

Given several methodologies have now been developed that rapidly achieved variant calling with considerable accuracy, whether work on our assay continues is undecided, and will ultimately be at the judgement of whomever may wish to pursue this. Ultimately, this would depend on whether the advantages of a refined version of this assay, with improved SNP analysis in particular, offers significant benefits over analogous approaches.

Chapter 7. Appendices

7.1. Appendix A: Endometrial Cancer Samples from External Providers

Sample	MSI Status	IHC	*MLH1* Methylation	Source	Dataset
882BRC	MSS	PMS2 Loss	Normal	Manchester	Training
PET16	MSI-H	MLH1 / PMS2 Loss	Normal	Manchester	Training
PET25	MSI-H	MSH6/PMS2 Loss	Not Tested	Manchester	Training
165BRC	MSS	MSH6/MSH2 Loss	Not Tested	Manchester	Training
PREC08	MSS	MLH1 / PMS2 Loss	Normal	Manchester	Training
PET31	MSS	MSH6 Loss	Not Tested	Manchester	Training
PET255	MSS	MSH6 Loss	Not Tested	Manchester	Training
PET128	MSI-L	MSH6 Loss	Not Tested	Manchester	Training
PET101	MSS	MLH1 / PMS2 Loss	Normal	Manchester	Training
PET87	MSS	MLH1 Loss	Normal	Manchester	Training
PET94	MSS	Patchy IHC Loss	Hypermethylated	Manchester	Training
PET117	MSS	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
PET119	MSS	MLH1 / PMS2 Loss	Normal	Manchester	Training
PET179	MSS	MLH1 / PMS2 Loss	Normal	Manchester	Training
PET242	MSS	MSH6 Loss	Not Tested	Manchester	Training
PET21	MSS	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
B15-0017632A	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
H16-00138B10	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
H16-00243B5	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
H16-02557B4 - PREC19	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
H16-2758 B8	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
H1601687B8 - PREC17	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
H17-0385	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
M8610/15 B	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
M16208/13	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
M4061/1416	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
MET001/02	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
MET006/02	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
MET014/02	MSI-H	MLH1 / PMS2 Loss	Normal	Manchester	Training
MET015/02	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
MET017/02	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
MIR 22	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
MIR 31	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
MO5266	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
PRE010	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
PREC14	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
RB002/02	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
RB006/T1	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
RB01/02	MSI-H	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
44BRC	MSI-H	No Loss	Not Tested	Manchester	Training
PET138	MSI-H	No Loss	Not Tested	Manchester	Training
H170027 B46	MSS	MLH1 / PMS2 Loss	Normal	Manchester	Training
PET258	MSS	MLH1 / PMS2 Loss	Normal	Manchester	Training
237BRC	MSI-H	No Loss	Not Tested	Manchester	Training
H16-0073 A12	MSI-H	No Loss	Not Tested	Manchester	Training
PET277	MSI-H	No Loss	Not Tested	Manchester	Training
416BRC	MSI-H	No Loss	Not Tested	Manchester	Training
M10878/13	MSI-H	No Loss	Not Tested	Manchester	Training
1510995A	MSS	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
417BRC	MSS	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training

Sample	MSI Status	IHC	*MLH1* Methylation	Source	Dataset
H15-03999-B3	MSS	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
M12415/13	MSS	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
M28666	MSS	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
PRE001	MSS	MLH1 / PMS2 Loss	Hypermethylated	Manchester	Training
RB021/02	MSS	MLH1 Loss	Hypermethylated	Manchester	Training
RB008/02	MSS	MLH1 / PMS2 Loss	NA	Manchester	Training
PET169	MSS	No Loss	Not Tested	Manchester	Training
PET175	MSS	No Loss	Not Tested	Manchester	Training
PET181	MSS	No Loss	Not Tested	Manchester	Training
PET239	MSS	No Loss	Not Tested	Manchester	Training
PET240	MSS	No Loss	Not Tested	Manchester	Training
PET243	MSS	No Loss	Not Tested	Manchester	Training
PET244	MSS	No Loss	Not Tested	Manchester	Training
PET246	MSS	No Loss	Not Tested	Manchester	Training
PET253	MSS	No Loss	Not Tested	Manchester	Training
PET254	MSS	No Loss	Not Tested	Manchester	Training
PET257	MSS	No Loss	Not Tested	Manchester	Training
PET259	MSS	No Loss	Not Tested	Manchester	Training
PET261	MSS	No Loss	Not Tested	Manchester	Training
PET264	MSS	No Loss	Not Tested	Manchester	Training
PET265	MSS	No Loss	Not Tested	Manchester	Training
PET266	MSS	No Loss	Normal	Manchester	Training
PET267	MSS	No Loss	Not Tested	Manchester	Training
PET270	MSS	No Loss	Normal	Manchester	Training
PET271	MSS	No Loss	Normal	Manchester	Training
PET273	MSS	No Loss	Not Tested	Manchester	Training
PET274	MSS	No Loss	Not Tested	Manchester	Training
PET275	MSS	No Loss	Not Tested	Manchester	Training
PET280	MSS	No Loss	Not Tested	Manchester	Training
PET281	MSS	No Loss	Not Tested	Manchester	Training
PET282	MSS	No Loss	Not Tested	Manchester	Training
PET283	MSS	No Loss	Not Tested	Manchester	Training
1205BRC	MSS	No Loss	Not Tested	Manchester	Training
150BRC	MSS	No Loss	Not Tested	Manchester	Training
178BRC	MSS	No Loss	Not Tested	Manchester	Training
301BRC	MSS	No Loss	Normal	Manchester	Training
307BRC	MSS	No Loss	Not Tested	Manchester	Training
308BRC	MSS	No Loss	Not Tested	Manchester	Training
349BRC	MSS	No Loss	Not Tested	Manchester	Training
358BRC	MSS	No Loss	Not Tested	Manchester	Training
381BRC	MSS	No Loss	Not Tested	Manchester	Training
3BRC	MSS	No Loss	Not Tested	Manchester	Training
415BRC	MSS	No Loss	Not Tested	Manchester	Training
431BRC	MSS	No Loss	Not Tested	Manchester	Training
434BRC	MSS	No Loss	Not Tested	Manchester	Training
437BRC	MSS	No Loss	Not Tested	Manchester	Training
4BRC	MSS	No Loss	Normal	Manchester	Training
503BRC	MSS	No Loss	Not Tested	Manchester	Training
527BRC	MSS	No Loss	Not Tested	Manchester	Training
191BRC	MSS	No Loss	Not Tested	Manchester	Training

Sample	MSI Status	IHC	*MLH1* Methylation	Source	Dataset
215BRC	MSS	No Loss	Not Tested	Manchester	Validation
RB015	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET120	MSI-L	MLH1/PMS2 Loss	Normal	Manchester	Validation
293BRC	MSS	No Loss	Not Tested	Manchester	Validation
299BRC	MSS	No Loss	Not Tested	Manchester	Validation
RB020	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET247	MSS	No Loss	Not Tested	Manchester	Validation
PET27	MSS	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET215	MSI-H	MHS6 Loss	Not Tested	Manchester	Validation
PET249	MSS	No Loss	Not Tested	Manchester	Validation
PET19	MSS	No Loss	Not Tested	Manchester	Validation
PET24	MSS	No Loss	Not Tested	Manchester	Validation
PET269	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET20	MSS	No Loss	Not Tested	Manchester	Validation
PET256	MSI-H	MLH1/PMS2 Loss	Normal	Manchester	Validation
PET276	MSI-L	MLH1/PMS2 Loss	Normal	Manchester	Validation
PET245	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET59	MSS	No Loss	Not Tested	Manchester	Validation
PET74	MSS	No Loss	Not Tested	Manchester	Validation
PET85	MSS	No Loss	Not Tested	Manchester	Validation
PET163	MSS	No Loss	Not Tested	Manchester	Validation
PET159	MSS	No Loss	Not Tested	Manchester	Validation
PET194	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET50	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PRE011	MSS	MHS6 Loss	Not Tested	Manchester	Validation
RB022	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET42	MSS	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET218	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET98	MSS	No Loss	Not Tested	Manchester	Validation
PET130	MSS	No Loss	Not Tested	Manchester	Validation
PET134	MSS	No Loss	Not Tested	Manchester	Validation
PET145	MSS	No Loss	Not Tested	Manchester	Validation
PET155	MSS	No Loss	Not Tested	Manchester	Validation
H19524-16	MSS	No Loss	Normal	Manchester	Validation
MET29	MSS	No Loss	Normal	Manchester	Validation
PET93	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET109	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET135	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET103	MSS	No Loss	Normal	Manchester	Validation
PET118	MSS	No Loss	Normal	Manchester	Validation
PET125	MSS	No Loss	Normal	Manchester	Validation
PET127	MSS	No Loss	Normal	Manchester	Validation
PET129	MSS	No Loss	Normal	Manchester	Validation
PET136	MSS	No Loss	Normal	Manchester	Validation
PET241	MSI-H	MHS6 Loss	Not Tested	Manchester	Validation
PET191	MSS	MLH1/PMS2 Loss	Not Tested	Manchester	Validation
PET158	MSI-L	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET173	MSI-H	PMS2 Loss	Not Tested	Manchester	Validation
PET62	MSI-H	MHS6 Loss	Not Tested	Manchester	Validation
PET171	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation

Sample	MSI Status	IHC	*MLH1* Methylation	Source	Dataset
PET139	MSS	No Loss	Normal	Manchester	Validation
PET192	MSS	No Loss	Normal	Manchester	Validation
PET278	MSS	No Loss	Normal	Manchester	Validation
PET200	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET115	MSS	MLH1/PMS2 Loss	Normal	Manchester	Validation
PET153	MSS	MLH1/PMS2 Loss	Normal	Manchester	Validation
PET2	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET6	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET8	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET68	MSI-H	MSH2/MSH6 Loss	Not Tested	Manchester	Validation
PET110	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET143	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET64	MSS	No Loss	Not Tested	Manchester	Validation
PET67	MSS	No Loss	Not Tested	Manchester	Validation
PET83	MSS	No Loss	Normal	Manchester	Validation
PET91	MSS	No Loss	Normal	Manchester	Validation
PET66	MSS	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET78	MSS	MLH1/PMS2 Loss	Normal	Manchester	Validation
PET193	MSS	MLH1/PMS2 Loss	Not Tested	Manchester	Validation
PET18	MSS	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET61	MSI-H	MSH2/MSH6 Loss	Not Tested	Manchester	Validation
PET35	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET51	MSS	No Loss	Normal	Manchester	Validation
PET182	MSS	MHS6 Loss	Not Tested	Manchester	Validation
PET144	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET165	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET186	MSI-H	MLH1/PMS2 Loss	Normal	Manchester	Validation
PET197	MSI-H	MSH2/MSH6 Loss	Not Tested	Manchester	Validation
PET204	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET235	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET272	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET285	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET213	MSI-H	MLH1 Loss	Not Tested	Manchester	Validation
PET79	MSS	MLH1 Loss	Normal	Manchester	Validation
PET286	MSS	No Loss	Normal	Manchester	Validation
PET5	MSS	No Loss	Normal	Manchester	Validation
PET7	MSS	No Loss	Normal	Manchester	Validation
PET11	MSS	No Loss	Normal	Manchester	Validation
PET37	MSS	No Loss	Normal	Manchester	Validation
PET40	MSS	No Loss	Normal	Manchester	Validation
PET52	MSS	No Loss	Normal	Manchester	Validation
PET157	MSS	MHS6 Loss	Not Tested	Manchester	Validation
PET150	MSS	MHS6 Loss	Not Tested	Manchester	Validation
PET96	MSS	MHS6 Loss	Not Tested	Manchester	Validation
PET122	MSI-H	MSH2/MSH6 Loss	Not Tested	Manchester	Validation
PET58	MSS	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation
PET65	MSI-H	MLH1/PMS2 Loss	Hypermethylated	Manchester	Validation

Sample	MSI Status	IHC	*MLH1* Methylation	Source	Dataset
JB1	MSS	No Loss	Not Tested	Ohio	Validation
JB2	MSS	No Loss	Present (OSU Rev)	Ohio	Validation
JB3	MSS	No Loss	Not Tested	Ohio	Validation
JB4	MSS	No Loss	Not Tested	Ohio	Validation
JB5	MSS	No Loss	Not Tested	Ohio	Validation
JB6	MSS	No Loss	Not Tested	Ohio	Validation
JB7	MSS	No Loss	Not Tested	Ohio	Validation
JB8	MSS	No Loss	Not Tested	Ohio	Validation
JB9	MSS	No Loss	Not Tested	Ohio	Validation
JB10	MSS	No Loss	Not Tested	Ohio	Validation
JB11	MSS	No Loss	Not Tested	Ohio	Validation
JB12	MSS	No Loss	Not Tested	Ohio	Validation
JB13	MSS	No Loss	Not Tested	Ohio	Validation
JB14	MSS	MSH6 Loss	Not Tested	Ohio	Validation
JB15	MSS	MSH6 Loss	Not Tested	Ohio	Validation
JB16	MSS	No Loss	Not Tested	Ohio	Validation
JB17	MSS	No Loss	Not Tested	Ohio	Validation
JB18	MSS	No Loss	Not Tested	Ohio	Validation
JB19	MSS	No Loss	Not Tested	Ohio	Validation
JB20	MSS	No Loss	Not Tested	Ohio	Validation
JB21	MSS	No Loss	Not Tested	Ohio	Validation
JB22	MSS	No Loss	Not Tested	Ohio	Validation
JB23	MSS	No Loss	Not Tested	Ohio	Validation
JB24	MSS	No Loss	Not Tested	Ohio	Validation
JB25	MSS	No Loss	Not Tested	Ohio	Validation
JB26	MSS	MSH6 Loss	Not Tested	Ohio	Validation
JB27	MSS	No Loss	Not Tested	Ohio	Validation
JB28	MSS	No Loss	Not Tested	Ohio	Validation
JB29	MSS	No Loss	Not Tested	Ohio	Validation
JB30	MSS	No Loss	Not Tested	Ohio	Validation
JB31	MSS	No Loss	Not Tested	Ohio	Validation
JB32	MSS	No Loss	Not Tested	Ohio	Validation
JB33	MSS	No Loss	Not Tested	Ohio	Validation
JB34	MSS	No Loss	Not Tested	Ohio	Validation
JB35	MSS	No Loss	Not Tested	Ohio	Validation
JB36	MSS	No Loss	Not Tested	Ohio	Validation
JB37	MSS	No Loss	Not Tested	Ohio	Validation
JB38	MSS	No Loss	Not Tested	Ohio	Validation
JB39	MSS	No Loss	Not Tested	Ohio	Validation
JB40	MSS	No Loss	Not Tested	Ohio	Validation
JB41	MSS	No Loss	Not Tested	Ohio	Validation
JB42	MSS	No Loss	Not Tested	Ohio	Validation
JB43	MSS	No Loss	Not Tested	Ohio	Validation
JB44	MSS	No Loss	Not Tested	Ohio	Validation
JB45	MSS	No Loss	Not Tested	Ohio	Validation
JB46	MSI-H	MLH1 / PMS2 Loss	Present (OSU Rev)	Ohio	Validation
JB47	MSS	No Loss	Not Tested	Ohio	Validation
JB48	MSS	No Loss	Not Tested	Ohio	Validation
JB49	MSS	No Loss	Not Tested	Ohio	Validation
JB50	MSS	No Loss	Not Tested	Ohio	Validation

Sample	MSI Status	IHC	*MLH1* Methylation	Source	Dataset
JB51	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB52	MSI-H	MSH6 Loss	Absent (OSU Rev)	Ohio	Validation
JB53	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB54	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB55	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB56	MSI-H	MSH6 Loss	Not Tested	Ohio	Validation
JB57	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB58	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB59	MSI-H	MSH2 / MSH6 Loss	Absent (OSU Rev)	Ohio	Validation
JB60	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB61	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB62	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB63	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB64	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB65	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB66	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB67	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB68	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB69	MSI-H	MSH6 Loss	Present (OSU Rev)	Ohio	Validation
JB70	MSI-H	MSH2 / MSH6 Loss	Not Tested	Ohio	Validation
JB71	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB72	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB73	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB74	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB75	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB76	MSI-H	MLH1 / PMS2 Loss	Present (OSU Rev - Formerly Absent)	Ohio	Validation
JB77	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB78	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB79	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB80	MSS	No Loss	Not Tested	Ohio	Validation
JB81	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB82	MSS	No Loss	Not Tested	Ohio	Validation
JB83	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB84	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB85	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB86	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB87	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB88	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB89	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB90	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB91	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB92	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB93	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB94	MSI-H	MSH2 / MSH6 Loss	Not Tested	Ohio	Validation
JB95	MSI-H	MLH1 / PMS2 Loss	Present (OSU Rev)	Ohio	Validation
JB96	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB97	MSI-H	MLH1 / PMS2 Loss	Present (OSU Rev - Formerly Absent)	Ohio	Validation
JB98	MSI-H	MSH2 / MSH6 Loss	Not Tested	Ohio	Validation
JB99	MSI-H	MLH1 / PMS2 Loss	Present	Ohio	Validation
JB100	MSI-H	MSH2 / MSH6 Loss	Not Tested	Ohio	Validation

Sample	MSI Status	IHC	*MLH1* Methylation	Source	Dataset
JB101	MSS	No Loss	NA	Ohio	Training
JB102	MSS	MSH6 Loss	NA	Ohio	Training
JB103	MSS	MLH1 / PMS2 Loss	NA	Ohio	Training
JB104	MSS	No Loss	NA	Ohio	Training
JB105	MSS	No Loss	NA	Ohio	Training
JB106	MSS	No Loss	NA	Ohio	Training
JB107	MSS	No Loss	NA	Ohio	Training
JB108	MSS	No Loss	NA	Ohio	Training
JB109	MSS	No Loss	NA	Ohio	Training
JB110	MSS	No Loss	NA	Ohio	Training
JB111	MSS	No Loss	NA	Ohio	Training
JB112	MSS	No Loss	NA	Ohio	Training
JB113	MSS	No Loss	NA	Ohio	Training
JB114	MSS	No Loss	NA	Ohio	Training
JB115	MSS	No Loss	NA	Ohio	Training
JB116	MSS	No Loss	NA	Ohio	Training
JB117	MSS	No Loss	NA	Ohio	Training
JB118	MSS	No Loss	NA	Ohio	Training
JB119	MSS	No Loss	NA	Ohio	Training
JB120	MSS	No Loss	NA	Ohio	Training
JB121	MSS	No Loss	NA	Ohio	Training
JB122	MSS	No Loss	NA	Ohio	Training
JB123	MSS	No Loss	NA	Ohio	Training
JB124	MSS	No Loss	NA	Ohio	Training
JB125	MSS	No Loss	NA	Ohio	Training
JB126	MSS	No Loss	NA	Ohio	Training
JB127	MSS	No Loss	NA	Ohio	Training
JB128	MSS	No Loss	NA	Ohio	Training
JB129	MSS	No Loss	NA	Ohio	Training
JB130	MSS	No Loss	NA	Ohio	Training
JB131	MSS	No Loss	NA	Ohio	Training
JB132	MSS	No Loss	NA	Ohio	Training
JB133	MSS	No Loss	NA	Ohio	Training
JB134	MSS	No Loss	NA	Ohio	Training
JB135	MSS	No Loss	NA	Ohio	Training
JB136	MSS	No Loss	NA	Ohio	Training
JB137	MSS	No Loss	NA	Ohio	Training
JB138	MSS	No Loss	NA	Ohio	Training
JB139	MSS	No Loss	NA	Ohio	Training
JB140	MSS	No Loss	NA	Ohio	Training
JB141	MSS	No Loss	NA	Ohio	Training
JB142	MSS	No Loss	NA	Ohio	Training
JB143	MSS	No Loss	NA	Ohio	Training
JB144	MSS	No Loss	NA	Ohio	Training
JB145	MSS	No Loss	NA	Ohio	Training
JB146	MSS	No Loss	NA	Ohio	Training
JB147	MSS	No Loss	NA	Ohio	Training
JB148	MSS	No Loss	NA	Ohio	Training
JB149	MSS	No Loss	NA	Ohio	Training
JB150	MSS	No Loss	NA	Ohio	Training

Sample	MSI Status	IHC	*MLH1* Methylation	Source	Dataset
JB151	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB152	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB153	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB154	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB155	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB156	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB157	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB158	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB159	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB160	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB161	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB162	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB163	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB164	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB165	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB166	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB167	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB168	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB169	MSI-H	MSH2 / MSH6 Loss	NA	Ohio	Training
JB170	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB171	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB172	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB173	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB174	MSI-H	MSH6 Loss	NA	Ohio	Training
JB175	MSI-H	MSH2 / MSH6 Loss	NA	Ohio	Training
JB176	MSI-H	MSH6 Loss	NA	Ohio	Training
JB177	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB178	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB179	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB180	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB181	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB182	MSI-H	MSH2 / MSH6 Loss	NA	Ohio	Training
JB183	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB184	MSI-H	MSH2 / MSH6 Loss	NA	Ohio	Training
JB185	MSI-H	MSH2 / MSH6 Loss	NA	Ohio	Training
JB186	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB187	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB188	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB189	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB190	MSI-H	No Loss	NA	Ohio	Training
JB191	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB192	MSI-H	No Loss	NA	Ohio	Training
JB193	MSI-H	MSH2 / MSH6 Loss	NA	Ohio	Training
JB194	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB195	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB196	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB197	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB198	MSI-H	PMS2 Loss	NA	Ohio	Training
JB199	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training
JB200	MSI-H	MLH1 / PMS2 Loss	NA	Ohio	Training

Table 7.1. Details of the endometrial cancer samples from external providers analysed by the Newcastle Assay. For Ohio samples, MSI was determined by a combination of Promega MSI Analysis System v1.2 and computer learning approaches, while MSI status for Manchester samples was assessed by Promega MSI Analysis System v1.2 exclusively. *MLH1* methylation analysis for both cohorts was achieved by pyrosequencing (PSQ 96MA). Samples designated of a training dataset were used in eventual classifier retraining for classification of the corresponding validation dataset for their cohort.

7.2. Appendix B: Marker Loci of the Newcastle MSI Assay

Assay Version	Marker	Chromosome	MNR Start	MNR End	MNR Length	SNP Variant	SNP Variant Position
1	DEPDC2	8	68926683	68926690	8	rs4610727	68926700
1	GM01	11	28894429	28894438	10	rs7951012	28894411
1	GM07	7	93085748	93085758	11	rs2283006	93085722
1	GM09	20	6836977	6836984	8	rs6038623	6836952
1	GM11	5	166099891	166099899	9	rs347435	166099902
					9	rs72817807	166099948
1	GM14	3	177328818	177328828	11	rs6804861	177328829
1	GM17	11	95551111	95551119	9	rs666398	95551136
1	GM22	14	43401010	43401019	10	rs17113692	43400964
1	GM26	14	49584751	49584760	10	rs11628435	49584720
1	GM29	3	70905560	70905569	10	rs2687195	70905581
1	IM16	18	1108767	1108775	9	rs4392141	1108738
					9	rs59912715	1108746
					9	rs73367791	1108784
1	IM49	3	56682066	56682077	12	rs7642389	56682093
1	LR10	1	81591388	81591396	9	rs1768398	81591398
					9	rs1768397	81591415
1	LR11	2	217217871	217217881	11	rs13011054	217217857
					11	rs16855951	217217913
1	LR17	14	55603031	55603040	10	rs79618905	55603041
					10	rs77482253	55603042
					10	rs1009978	55603061
					10	rs1009977	55603002
1	LR20	1	64029634	64029641	8	rs217474	64029606
1	LR24	1	153779429	153779437	9	rs1127091	153779412
1	LR36	4	98999723	98999734	12	rs17550217	98999699
1	LR40	2	13447470	13447478	9	rs6432372	13447484
1	LR44	10	99898286	99898297	12	rs7905384	99898268
					12	rs7905388	99898281
1	LR46	20	10660085	10660092	8	rs6040079	10660063
1	LR48	12	77988097	77988107	11	rs11105832	77988123
1	LR49	15	93619048	93619054	7	rs12903384	93619037
1	LR52	16	63861441	63861452	12	rs2434849	63861437
2	AKMmono01v2	1	163944658	163944669	12	rs12034420	163944712
2	AKMmono02	1	231388450	231388461	12	N/A	231388464
2	AKMmono03	2	20514738	20514750	13	N/A	20514753
2	AKMmono04	2	48427470	48427482	13	N/A	48427485
2	AKMmono05	3	47153240	47153252	13	N/A	47153255
2	AKMmono06	4	31266124	31266138	15	N/A	31266141
2	AKMmono07	5	14855958	14855970	13	N/A	14855973
2	AKMmono08v2	6	119660766	119660776	11	rs195082	119660876
2	AKMmono10v2	7	8110837	8110847	11	rs10486207	8110780
2	AKMmono11	1	188783747	188783758	12	N/A	188783761
2	AKMmono12	1	220008901	220008912	12	N/A	220008915
2	AKMmono13	2	95829922	95829933	12	N/A	95829936
2	AKMmono14	2	115319229	115319239	11	N/A	115319242
2	AKMmono16	2	146400083	146400094	12	N/A	146400097
2	AKMmono17v2	8	59843890	59843900	11	rs7834158	59843813
2	AKMmono22	9	104829958	104829971	14	N/A	104829974

Assay Version	Marker	Chromosome	MNR Start	MNR End	MNR Length	SNP Variant	SNP Variant Position
2	EJmono01	7	118718316	118718326	11	N/A	118718329
2	EJmono02	8	28204461	28204473	13	N/A	28204476
2	EJmono03	8	53511204	53511214	11	N/A	53511217
2	EJmono04	9	8807599	8807613	15	N/A	8807616
2	EJmono05	9	28483545	28483555	11	N/A	28483558
2	EJmono06v2	9	84371946	84371961	16	rs1007995	84371927
2	EJmono12	3	65336135	65336146	12	N/A	65336149
2	EJmono13v2	5	4294954	4294964	11	rs16873198	4295050
2	EJmono14v2	5	103200293	103200303	11	rs2562279	103200197
2	EJmono16	10	57222859	57222871	13	N/A	57222874
2	EJmono21v2	18	7068797	7068807	11	rs7234998	7068724
2	HGtetra23ms2	10	70692828	70692841	14	N/A	70692844
2	LMmono01	14	89137211	89137221	11	N/A	89137224
2	LMmono03	17	58399233	58399246	14	N/A	58399249
2	LMmono04v2	18	53192468	53192480	13	rs10401120	53192498
2	LMmono05v2	20	33745137	33745149	13	rs6088734	33745046
2	LMmono07	15	55347997	55348009	13	N/A	55348012
2	LMmono08	15	73175070	73175080	11	N/A	73175083
2	LMmono09	19	49540384	49540397	14	N/A	49540400
2	LMmono10v2	20	13814245	13814257	13	rs10485769	13814346
2	LMmono12	16	80521284	80521296	13	N/A	80521299
2	LMmono16	1	150737578	150737592	15	N/A	150737595
2	MSJcom06ms1	10	105386364	105386376	13	N/A	105386399
2	MSJcom06ms2	10	105386384	105386396	13	N/A	105386399
2	MSJmono10	1	219386785	219386796	12	N/A	219386799
2	MSJmono11	1	235744815	235744833	19	N/A	235744836
2	MSJmono15	16	50022767	50022779	13	N/A	50022782
2	MSJmono17	16	34332376	34332396	21	N/A	34332399
2	MSJmono19ms1	12	80004543	80004567	25	N/A	80004570
2	MSJmono19ms2	12	80004578	80004588	11	N/A	80004591
2	MSJmono20	8	142928089	142928101	13	N/A	142928104
2	MSJmono22	2	70463335	70463347	13	N/A	70463350
2	MSJmono23v2	6	126371845	126371857	13	rs4897168	126371788
2	MSJmono26	14	104291032	104291045	14	N/A	104291048
2	MSJmono27	10	21613158	21613189	32	N/A	21613192
2	MSJmono30v2	4	140348882	140348894	13	rs13136124	140348815
2	MSJmono32	4	75787356	75787368	13	rs7700246	75787289
2	MSJmono36	15	56017792	56017804	13	N/A	56017807
2	MSJmono37	14	64159174	64159185	12	N/A	64159188
2	MSJmono38	13	59791883	59791894	12	N/A	59791897
2	MSJmono39	15	68400135	68400145	11	N/A	68400148
2	MSJmono40	12	29803003	29803033	31	N/A	29803036
2	MSJmono41	12	50382596	50382615	20	N/A	50382618
2	MSJmono44	4	1911366	1911381	16	N/A	1911384
2	MSJmono45	6	14647998	14648022	25	N/A	14648025
2	MSJmono46	19	38154539	38154551	13	N/A	38154554

Table 7.2. Marker Loci of the smMIP, sequencing-based Newcastle MSI Assay. Loci are defined by chromosomal coordinates using the reference genome hg19, while mononucleotide repeats (MNRs) are specified by the chromosomal position at which the repeat starts and ends.

7.3. Appendix C: MSI Classifier Script

```
#! /bin/bash
#$ -cwd
#$ -j y
#SBATCH --array=1-n

######################
# Load Rocket Modules #
######################
module load BWA/0.7.17-foss-2017b
module load R/3.6.0-foss-2019a

#########################
# Set Inputs and Directories #
#########################
RefGenome="/mnt/nfs/home/b7000371/WORKING_DATA/human_g1k_v37.fasta"

RunID=""

sampleData="SampleData/$RunID.csv"
fastqDir="fastq/$RunID/"
samDir="sam/$RunID/"
variantsDir="VariantsData/$RunID/"
countsDir="CountsData/$RunID/"
logsDir="Rlogs/$RunID/"

##############################
# Part 1 Align Reads From .fastq #
##############################
if [ ! -d $samDir ];
```

```
then
 mkdir $samDir
fi

fastqNames=( $(Rscript SourceFiles/Rscripts/extractNames.r $sampleData "fastq") )
fastqNames=${fastqNames[@]:1}
fastqNames=`sed -e 's/^"//' -e 's/"$//' <<<$fastqNames`
fastqNames=( $fastqNames )

Indx=`expr $SLURM_ARRAY_TASK_ID - 1`
fastqName=${fastqNames[$Indx]}

READ_FILE1="$fastqDir/${fastqName}_L001_R1_001.fastq.gz"
READ_FILE2="$fastqDir/${fastqName}_L001_R2_001.fastq.gz"
SAM_FILE="$samDir/${fastqName}.sam"
echo $RefGenome $READ_FILE1 $READ_FILE2 $SAM_FILE
bwa mem $RefGenome $READ_FILE1 $READ_FILE2 > $SAM_FILE

################################################
# Part 2 Generate Variants and Counts Data From .sam #
################################################
if [ ! -d $logsDir ];
then
 mkdir $logsDir
fi

if [ ! -d $variantsDir ];
then
 mkdir $variantsDir
fi
```

```
MARKER_DATA="SourceFiles/Markers/MNRandBRAFloci.csv"
VARIANTS_FILE="$variantsDir/$fastqName.R"
LOG_FILE1="$logsDir/$fastqName.variants.log"

PARAMETERS1="$SAM_FILE $MARKER_DATA $VARIANTS_FILE"
echo $PARAMETERS1
R CMD BATCH "--args $PARAMETERS1" SourceFiles/Rcm/exploreMNR3.Rcm
$LOG_FILE1

if [ ! -d $countsDir ];
then
 mkdir $countsDir
fi

AMPLICON_DATA="SourceFiles/Markers/MNRandBRAFamplicons.csv"
COUNTS_FILE="$countsDir/$fastqName.csv"
LOG_FILE2="$logsDir/$fastqName.counts.log"

PARAMETERS2="$SAM_FILE $MARKER_DATA $AMPLICON_DATA 1 200000
$COUNTS_FILE"
echo $PARAMETERS2
R CMD BATCH "--args $PARAMETERS2" SourceFiles/Rcm/detTagCoverage.Rcm
$LOG_FILE2

##########################################
# Part 3 Classify Samples From .R Variants Data #
##########################################

OUT_FILE="Output/$RunID.csv"
LOG_FILE3="$logsDir/classification.log"

PARAMETERS3="$OUT_FILE $sampleData $variantsDir $countsDir
$MARKER_DATA"
```

```
echo $PARAMETERS3

R CMD BATCH --no-save "--args $PARAMETERS3"
SourceFiles/Rcm/classification.Rcm $LOG_FILE3
```

Figure 7.1. Example MSI Classifier Script of the Newcastle Assay. Script interprets input fastq sequencing files with reference to a designated marker file, generating counts of marker length for each component locus, before collating these data for the classification of each sample and ultimately combing all classifications into a single .csv file for the given sequencing run.

7.4. Appendix D: MSI Classifier Script Output

Sample ID54155 – MSS

```
Result <-
list(DEPDC2_SNP1 = structure(c(0L, 0L, 0L, 5L, 0L, 0L, 72L, 1L,
1L, 10437L, 32L, 25L, 0L, 0L, 0L, 1L, 0L, 0L, 62L, 0L, 0L, 9868L,
30L, 44L, 3L, 0L, 0L, 313L, 1L, 0L), .Dim = 6:5, .Dimnames =
structure(list(
    c("-1", "-2", "-3", "0", "1", NA), c("A", "C", "G", "T",
    NA)), .Names = c("", "")), class = "table"), GM01_SNP1 =
structure(c(1L,
0L, 11L, 0L, 0L, 153L, 1L, 9476L, 34L, 13L, 5L, 0L, 114L, 1L,
0L), .Dim = c(5L, 3L), .Dimnames = structure(list(c("-1", "-2",
"0", "1", NA), c("A", "G", NA)), .Names = c("", "")), class = "table"),
    GM07_SNP1 = structure(c(5L, 0L, 41L, 1L, 0L, 0L, 1L, 0L,
    0L, 1L, 0L, 0L, 0L, 0L, 341L, 7L, 7263L, 218L, 3L, 1L, 37L,
    0L, 0L, 1L, 0L, 0L, 0L, 0L, 8L, 1L, 197L, 5L, 0L, 0L, 0L), .Dim =
c(7L,
    5L), .Dimnames = structure(list(c("-1", "-2", "0", "1", "2",
    "3", NA), c("A", "C", "G", "T", NA)), .Names = c("", "")), class =
"table"),
    GM09_SNP1 = structure(c(47L, 10277L, 13L, 4L, 48L, 9819L,
    12L, 3L, 0L, 4L, 0L, 0L, 2L, 212L, 0L, 0L), .Dim = c(4L,
    4L), .Dimnames = structure(list(c("-1", "0", "1", NA), c("A",
    "G", "T", NA)), .Names = c("", "")), class = "table"), GM11_SNP1 =
structure(c(0L,
    6L, 0L, 0L, 31L, 5196L, 22L, 3L, 0L, 65L, 1L, 0L), .Dim = 4:3,
.Dimnames = structure(list(
        c("-1", "0", "1", NA), c("A", "G", NA)), .Names = c("",
    "")), class = "table"), GM11_SNP2 = structure(c(31L, 5216L,
    23L, 4L, 0L, 1L, 0L, 0L, 0L, 50L, 0L, 0L), .Dim = 4:3, .Dimnames =
structure(list(
        c("-1", "0", "1", NA), c("A", "N", NA)), .Names = c("",
    "")), class = "table"), GM14_SNP1 = structure(c(0L, 0L, 0L,
    2L, 0L, 0L, 0L, 0L, 356L, 11L, 1L, 11773L, 90L, 1L, 1L, 35L,
    0L, 0L, 0L, 3L, 0L, 0L, 0L, 0L, 2L, 0L, 0L, 172L, 1L, 0L,
    0L, 0L), .Dim = c(8L, 4L), .Dimnames = structure(list(c("-1",
    "-2", "-3", "0", "1", "2", "3", NA), c("A", "C", "T", NA)), .Names
= c("",
    "")), class = "table"), GM17_SNP1 = structure(c(0L, 2L, 0L,
    0L, 0L, 67L, 5590L, 19L, 0L, 4L, 65L, 6245L, 14L, 0L, 4L,
    2L, 121L, 0L, 1L, 0L), .Dim = 5:4, .Dimnames = structure(list(
        c("-1", "0", "1", "2", NA), c("A", "C", "T", NA)), .Names =
c("",
    "")), class = "table"), GM22_SNP1 = structure(c(72L, 6L,
    5435L, 23L, 0L, 5L, 47L, 4L, 4013L, 19L, 1L, 3L, 2L, 0L,
    194L, 2L, 0L, 0L), .Dim = c(6L, 3L), .Dimnames = structure(list(
        c("-1", "-2", "0", "1", "3", NA), c("A", "G", NA)), .Names =
c("",
    "")), class = "table"), GM26_SNP1 = structure(c(38L, 2L,
    1L, 5756L, 15L, 5L, 0L, 0L, 0L, 1L, 0L, 0L, 2L, 0L, 0L, 55L,
    0L, 0L), .Dim = c(6L, 3L), .Dimnames = structure(list(c("-1",
    "-2", "-3", "0", "1", NA), c("A", "G", NA)), .Names = c("",
    "")), class = "table"), GM29_SNP1 = structure(c(97L, 1L,
    1L, 3084L, 12L, 3L, 57L, 0L, 0L, 4424L, 16L, 3L, 3L, 0L,
    0L, 76L, 0L, 0L), .Dim = c(6L, 3L), .Dimnames = structure(list(
```

```
        c("-1", "-2", "-3", "0", "1", NA), c("A", "G", NA)), .Names =
c("",
    "")), class = "table"), IM16_SNP1 = structure(c(28L, 1L,
    0L, 3012L, 45L, 1L, 1L, 27L, 1L, 1L, 4525L, 36L, 1L, 2L,
    0L, 0L, 0L, 1L, 0L, 0L, 0L, 0L, 0L, 0L, 128L, 10L, 0L, 0L
    ), .Dim = c(7L, 4L), .Dimnames = structure(list(c("-1", "-2",
    "-4", "0", "1", "2", NA), c("A", "G", "T", NA)), .Names = c("",
    "")), class = "table"), IM16_SNP2 = structure(c(0L, 0L, 0L,
    1L, 0L, 0L, 0L, 27L, 1L, 1L, 4501L, 38L, 1L, 2L, 28L, 1L,
    0L, 2993L, 43L, 1L, 1L, 0L, 0L, 0L, 2L, 0L, 0L, 0L, 0L, 0L,
    0L, 169L, 10L, 0L, 0L), .Dim = c(7L, 5L), .Dimnames =
structure(list(
        c("-1", "-2", "-4", "0", "1", "2", NA), c("A", "C", "G",
        "N", NA)), .Names = c("", "")), class = "table"), IM16_SNP3 =
structure(c(55L,
    2L, 1L, 7608L, 90L, 2L, 3L, 0L, 0L, 0L, 1L, 0L, 0L, 0L, 0L,
    0L, 0L, 57L, 1L, 0L, 0L), .Dim = c(7L, 3L), .Dimnames =
structure(list(
        c("-1", "-2", "-4", "0", "1", "2", NA), c("C", "T", NA
        )), .Names = c("", "")), class = "table"), IM49_SNP1 =
structure(c(203L,
    3L, 5811L, 103L, 2L, 15L, 112L, 3L, 4064L, 84L, 4L, 6L, 0L,
    0L, 1L, 0L, 0L, 0L, 3L, 1L, 397L, 9L, 0L, 0L), .Dim = c(6L,
    4L), .Dimnames = structure(list(c("-1", "-2", "0", "1", "2",
    NA), c("C", "G", "T", NA)), .Names = c("", "")), class = "table"),
    LR10_SNP1 = structure(c(0L, 0L, 1L, 0L, 0L, 52L, 2L, 3046L,
    15L, 2L, 71L, 1L, 3330L, 5L, 1L, 2L, 0L, 91L, 3L, 0L), .Dim = 5:4,
.Dimnames = structure(list(
        c("-1", "-2", "0", "1", NA), c("A", "C", "T", NA)), .Names =
c("",
    "")), class = "table"), LR10_SNP2 = structure(c(54L, 2L,
    3035L, 13L, 1L, 70L, 1L, 3339L, 8L, 1L, 1L, 0L, 94L, 2L,
    0L), .Dim = c(5L, 3L), .Dimnames = structure(list(c("-1",
    "-2", "0", "1", NA), c("C", "T", NA)), .Names = c("", "")), class =
"table"),
    LR11_SNP1 = structure(c(0L, 0L, 0L, 1L, 0L, 0L, 224L, 3L,
    1L, 8686L, 67L, 12L, 0L, 0L, 0L, 4L, 0L, 0L, 6L, 0L, 0L,
    335L, 5L, 0L), .Dim = c(6L, 4L), .Dimnames = structure(list(
        c("-1", "-2", "-4", "0", "1", NA), c("A", "C", "T", NA
        )), .Names = c("", "")), class = "table"), LR11_SNP2 =
structure(c(0L,
    0L, 0L, 1L, 0L, 0L, 0L, 0L, 0L, 2L, 0L, 0L, 224L, 2L, 1L,
    8693L, 66L, 13L, 0L, 0L, 0L, 1L, 0L, 0L, 6L, 1L, 0L, 329L,
    6L, 0L), .Dim = 6:5, .Dimnames = structure(list(c("-1", "-2",
    "-4", "0", "1", NA), c("A", "C", "G", "T", NA)), .Names = c("",
    "")), class = "table"), LR17_SNP1 = structure(c(245L, 11L,
    6411L, 149L, 35L, 0L, 0L, 3L, 0L, 0L, 9L, 0L, 582L, 3L, 0L
    ), .Dim = c(5L, 3L), .Dimnames = structure(list(c("-1", "-2",
    "0", "1", NA), c("A", "G", NA)), .Names = c("", "")), class =
"table"),
    LR17_SNP2 = structure(c(0L, 0L, 5L, 0L, 0L, 250L, 11L, 6510L,
    142L, 19L, 4L, 0L, 481L, 10L, 0L), .Dim = c(5L, 3L), .Dimnames =
structure(list(
        c("-1", "-2", "0", "1", NA), c("A", "G", NA)), .Names = c("",
    "")), class = "table"), LR17_SNP3 = structure(c(0L, 0L, 1L,
    0L, 0L, 128L, 10L, 2918L, 85L, 9L, 113L, 1L, 2366L, 29L,
    4L, 13L, 0L, 1711L, 38L, 0L), .Dim = 5:4, .Dimnames =
structure(list(
```

```
        c("-1", "-2", "0", "1", NA), c("A", "C", "G", NA)), .Names =
c("",
    "")), class = "table"), LR17_SNP4 = structure(c(122L, 1L,
    3633L, 55L, 16L, 131L, 9L, 3213L, 95L, 23L, 1L, 1L, 150L,
    2L, 0L), .Dim = c(5L, 3L), .Dimnames = structure(list(c("-1",
    "-2", "0", "1", NA), c("G", "T", NA)), .Names = c("", "")), class =
"table"),
    LR20_SNP1 = structure(c(4L, 3959L, 11L, 1L, 0L, 1L, 0L, 0L,
    6L, 3938L, 9L, 3L, 0L, 47L, 0L, 0L), .Dim = c(4L, 4L), .Dimnames =
structure(list(
        c("-1", "0", "1", NA), c("A", "C", "G", NA)), .Names = c("",
    "")), class = "table"), LR24_SNP1 = structure(c(0L, 0L, 5L,
    0L, 0L, 2L, 1L, 168L, 0L, 0L, 91L, 2L, 22659L, 36L, 10L,
    0L, 1L, 249L, 0L, 0L), .Dim = 5:4, .Dimnames = structure(list(
        c("-1", "-2", "0", "1", NA), c("-", "A", "G", NA)), .Names =
c("",
    "")), class = "table"), LR36_SNP1 = structure(c(0L, 0L, 0L,
    18L, 0L, 0L, 0L, 0L, 0L, 0L, 1L, 0L, 0L, 0L, 0L, 0L, 0L,
    2L, 0L, 0L, 0L, 407L, 21L, 1L, 9646L, 69L, 2L, 39L, 14L,
    1L, 0L, 571L, 4L, 0L, 0L), .Dim = c(7L, 5L), .Dimnames =
structure(list(
        c("-1", "-2", "-4", "0", "1", "2", NA), c("A", "C", "G",
        "T", NA)), .Names = c("", "")), class = "table"), LR40_SNP1 =
structure(c(1L,
    69L, 0L, 0L, 0L, 67L, 6842L, 16L, 1L, 3L, 0L, 1L, 0L, 0L,
    0L, 1L, 52L, 0L, 0L, 0L), .Dim = 5:4, .Dimnames = structure(list(
        c("-1", "0", "1", "2", NA), c("A", "C", "T", NA)), .Names =
c("",
    "")), class = "table"), LR44_SNP1 = structure(c(543L, 32L,
    2L, 7100L, 104L, 2L, 29L, 4L, 1L, 0L, 164L, 1L, 0L, 2L, 8L,
    0L, 0L, 274L, 5L, 0L, 0L), .Dim = c(7L, 3L), .Dimnames =
structure(list(
        c("-1", "-2", "-3", "0", "1", "2", NA), c("C", "T", NA
        )), .Names = c("", "")), class = "table"), LR44_SNP2 =
structure(c(0L,
    0L, 0L, 42L, 0L, 0L, 0L, 530L, 32L, 2L, 7210L, 106L, 2L,
    21L, 3L, 1L, 0L, 33L, 1L, 0L, 0L, 22L, 0L, 0L, 253L, 3L,
    0L, 0L), .Dim = c(7L, 4L), .Dimnames = structure(list(c("-1",
    "-2", "-3", "0", "1", "2", NA), c("-", "C", "T", NA)), .Names =
c("",
    "")), class = "table"), LR46_SNP1 = structure(c(0L, 1L, 0L,
    0L, 15L, 4339L, 5L, 2L, 0L, 1L, 0L, 0L, 0L, 42L, 0L, 0L), .Dim =
c(4L,
    4L), .Dimnames = structure(list(c("-1", "0", "1", NA), c("A",
    "G", "T", NA)), .Names = c("", "")), class = "table"), LR48_SNP1 =
structure(c(0L,
    0L, 1L, 0L, 0L, 161L, 2L, 6256L, 77L, 10L, 1L, 0L, 207L,
    1L, 0L), .Dim = c(5L, 3L), .Dimnames = structure(list(c("-1",
    "-2", "0", "1", NA), c("A", "C", NA)), .Names = c("", "")), class =
"table"),
    LR49_SNP1 = structure(c(23L, 5469L, 4L, 1L, 0L, 1L, 0L, 0L,
    6L, 4661L, 2L, 0L, 0L, 152L, 0L, 0L), .Dim = c(4L, 4L), .Dimnames =
structure(list(
        c("-1", "0", "1", NA), c("A", "C", "G", NA)), .Names = c("",
    "")), class = "table"), LR52_SNP1 = structure(c(1L, 0L, 0L,
    2L, 0L, 0L, 736L, 53L, 3L, 5759L, 32L, 92L, 34L, 0L, 0L,
    527L, 4L, 0L), .Dim = c(6L, 3L), .Dimnames = structure(list(
        c("-1", "-2", "-3", "0", "1", NA), c("C", "T", NA)), .Names =
```

```
c("",
    "")), class = "table"), BRAF_SNP1 = structure(c(12539L, 2L,
    2L, 53L), .Dim = c(1L, 4L), .Dimnames = structure(list("0",
        c("A", "G", "T", NA)), .Names = c("", "")), class = "table"))
```

Sample ID21269 – MSI-H

```
Result <-
list(DEPDC2_SNP1 = structure(c(119L, 1L, 9771L, 32L, 0L, 29L,
0L, 0L, 1L, 0L, 0L, 0L, 65L, 0L, 7982L, 902L, 3L, 698L, 1L, 0L,
336L, 10L, 0L, 0L), .Dim = c(6L, 4L), .Dimnames = structure(list(
    c("-1", "-2", "0", "1", "2", NA), c("C", "G", "T", NA)), .Names =
c("",
"")), class = "table"), GM01_SNP1 = structure(c(1L, 0L, 122L,
0L, 0L, 0L, 353L, 1L, 9127L, 48L, 1L, 14L, 4L, 0L, 111L, 1L,
0L, 0L), .Dim = c(6L, 3L), .Dimnames = structure(list(c("-1",
"-2", "0", "1", "2", NA), c("A", "G", NA)), .Names = c("", ""
)), class = "table"), GM07_SNP1 = structure(c(0L, 0L, 0L, 1L,
0L, 0L, 0L, 1009L, 19L, 1L, 2356L, 71L, 1L, 46L, 1373L, 26L,
0L, 2519L, 89L, 2L, 60L, 0L, 0L, 0L, 0L, 0L, 0L, 1L, 20L, 4L,
0L, 150L, 3L, 0L, 0L), .Dim = c(7L, 5L), .Dimnames = structure(list(
    c("-1", "-2", "-3", "0", "1", "2", NA), c("-", "A", "G",
    "T", NA)), .Names = c("", "")), class = "table"), GM09_SNP1 =
structure(c(290L,
9224L, 15L, 6L, 65L, 10324L, 43L, 3L, 0L, 10L, 0L, 0L, 2L, 181L,
1L, 0L), .Dim = c(4L, 4L), .Dimnames = structure(list(c("-1",
"0", "1", NA), c("A", "G", "T", NA)), .Names = c("", "")), class =
"table"),
    GM11_SNP1 = structure(c(128L, 5995L, 16L, 2L, 0L, 7L, 0L,
    0L, 0L, 2L, 0L, 0L, 3L, 54L, 1L, 0L), .Dim = c(4L, 4L), .Dimnames =
structure(list(
        c("-1", "0", "1", NA), c("A", "G", "T", NA)), .Names = c("",
    "")), class = "table"), GM11_SNP2 = structure(c(128L, 6015L,
    16L, 5L, 0L, 1L, 0L, 0L, 3L, 42L, 1L, 0L), .Dim = 4:3, .Dimnames =
structure(list(
        c("-1", "0", "1", NA), c("A", "T", NA)), .Names = c("",
    "")), class = "table"), GM14_SNP1 = structure(c(885L, 393L,
    3L, 8440L, 86L, 31L, 1L, 0L, 0L, 9L, 0L, 0L, 2L, 0L, 0L,
    146L, 0L, 0L), .Dim = c(6L, 3L), .Dimnames = structure(list(
        c("-1", "-2", "-3", "0", "1", NA), c("C", "T", NA)), .Names =
c("",
    "")), class = "table"), GM17_SNP1 = structure(c(56L, 0L,
    4443L, 13L, 1L, 121L, 1L, 4169L, 8L, 1L, 0L, 0L, 82L, 0L,
    0L), .Dim = c(5L, 3L), .Dimnames = structure(list(c("-1",
    "-2", "0", "1", NA), c("C", "T", NA)), .Names = c("", "")), class =
"table"),
    GM22_SNP1 = structure(c(739L, 9L, 8026L, 110L, 21L, 0L, 0L,
    5L, 0L, 0L, 6L, 0L, 120L, 1L, 0L), .Dim = c(5L, 3L), .Dimnames =
structure(list(
        c("-1", "-2", "0", "1", NA), c("A", "G", NA)), .Names = c("",
    "")), class = "table"), GM26_SNP1 = structure(c(302L, 1L,
    5418L, 14L, 10L, 0L, 0L, 1L, 0L, 0L, 1L, 0L, 52L, 0L, 0L), .Dim =
c(5L,
    3L), .Dimnames = structure(list(c("-1", "-2", "0", "1", NA
    ), c("A", "G", NA)), .Names = c("", "")), class = "table"),
```

```
    GM29_SNP1 = structure(c(336L, 1L, 5330L, 9L, 3L, 0L, 0L,
    3L, 0L, 0L, 0L, 0L, 1L, 0L, 0L, 3L, 0L, 46L, 0L, 0L), .Dim = 5:4,
.Dimnames = structure(list(
        c("-1", "-2", "0", "1", NA), c("A", "G", "T", NA)), .Names =
c("",
    "")), class = "table"), IM16_SNP1 = structure(c(0L, 0L, 7L,
    0L, 0L, 0L, 170L, 1L, 6346L, 103L, 4L, 6L, 0L, 0L, 2L, 0L,
    0L, 0L, 0L, 0L, 105L, 11L, 0L, 0L), .Dim = c(6L, 4L), .Dimnames =
structure(list(
        c("-1", "-2", "0", "1", "2", NA), c("A", "G", "T", NA
        )), .Names = c("", "")), class = "table"), IM16_SNP2 =
structure(c(0L,
    0L, 1L, 0L, 0L, 0L, 167L, 1L, 6319L, 105L, 4L, 8L, 0L, 0L,
    1L, 0L, 0L, 0L, 0L, 0L, 2L, 0L, 0L, 0L, 0L, 0L, 1L, 0L, 0L,
    0L, 3L, 0L, 136L, 9L, 0L, 0L), .Dim = c(6L, 6L), .Dimnames =
structure(list(
        c("-1", "-2", "0", "1", "2", NA), c("A", "C", "G", "N",
        "T", NA)), .Names = c("", "")), class = "table"), IM16_SNP3 =
structure(c(169L,
    1L, 6409L, 113L, 4L, 6L, 0L, 0L, 1L, 0L, 0L, 0L, 1L, 0L,
    50L, 1L, 0L, 0L), .Dim = c(6L, 3L), .Dimnames = structure(list(
        c("-1", "-2", "0", "1", "2", NA), c("C", "G", NA)), .Names =
c("",
    "")), class = "table"), IM49_SNP1 = structure(c(1331L, 103L,
    1L, 6719L, 108L, 1L, 49L, 152L, 2L, 1L, 2045L, 36L, 0L, 7L,
    1L, 0L, 0L, 1L, 0L, 0L, 0L, 13L, 1L, 0L, 320L, 3L, 0L, 0L
    ), .Dim = c(7L, 4L), .Dimnames = structure(list(c("-1", "-2",
    "-3", "0", "1", "2", NA), c("C", "G", "T", NA)), .Names = c("",
    "")), class = "table"), LR10_SNP1 = structure(c(0L, 0L, 1L,
    0L, 0L, 107L, 1L, 3391L, 13L, 6L, 72L, 1L, 5237L, 20L, 3L,
    2L, 0L, 103L, 0L, 0L), .Dim = 5:4, .Dimnames = structure(list(
        c("-1", "-2", "0", "1", NA), c("A", "C", "T", NA)), .Names =
c("",
    "")), class = "table"), LR10_SNP2 = structure(c(1L, 0L, 0L,
    0L, 0L, 104L, 1L, 3384L, 12L, 6L, 73L, 1L, 5225L, 16L, 3L,
    3L, 0L, 123L, 5L, 0L), .Dim = 5:4, .Dimnames = structure(list(
        c("-1", "-2", "0", "1", NA), c("A", "C", "T", NA)), .Names =
c("",
    "")), class = "table"), LR11_SNP1 = structure(c(1440L, 23L,
    1L, 2319L, 12L, 29L, 1015L, 272L, 2L, 2150L, 14L, 19L, 0L,
    0L, 0L, 1L, 0L, 0L, 26L, 3L, 0L, 167L, 1L, 0L), .Dim = c(6L,
    4L), .Dimnames = structure(list(c("-1", "-2", "-3", "0",
    "1", NA), c("A", "C", "G", NA)), .Names = c("", "")), class =
"table"),
    LR11_SNP2 = structure(c(0L, 0L, 0L, 1L, 0L, 0L, 1415L, 23L,
    1L, 2248L, 10L, 29L, 1017L, 273L, 2L, 2182L, 15L, 19L, 1L,
    0L, 0L, 0L, 0L, 0L, 48L, 2L, 0L, 206L, 2L, 0L), .Dim = 6:5,
.Dimnames = structure(list(
        c("-1", "-2", "-3", "0", "1", NA), c("A", "C", "G", "T",
        NA)), .Names = c("", "")), class = "table"), LR17_SNP1 =
structure(c(1882L,
    35L, 2L, 5104L, 84L, 3L, 48L, 3L, 0L, 0L, 1L, 0L, 0L, 0L,
    121L, 0L, 0L, 441L, 1L, 0L, 0L), .Dim = c(7L, 3L), .Dimnames =
structure(list(
        c("-1", "-2", "-3", "0", "1", "2", NA), c("A", "G", NA
        )), .Names = c("", "")), class = "table"), LR17_SNP2 =
structure(c(0L,
    0L, 0L, 2L, 0L, 0L, 0L, 0L, 0L, 0L, 1L, 0L, 0L, 0L, 0L, 0L,
```

```
    0L, 1L, 0L, 0L, 0L, 1948L, 34L, 2L, 5138L, 80L, 3L, 27L,
    1L, 0L, 0L, 1L, 0L, 0L, 0L, 57L, 1L, 0L, 403L, 5L, 0L, 0L
    ), .Dim = 7:6, .Dimnames = structure(list(c("-1", "-2", "-3",
    "0", "1", "2", NA), c("-", "A", "C", "G", "T", NA)), .Names = c("",
    "")), class = "table"), LR17_SNP3 = structure(c(0L, 0L, 0L,
    1L, 0L, 0L, 0L, 1736L, 30L, 2L, 2435L, 42L, 1L, 30L, 191L,
    5L, 0L, 1741L, 24L, 1L, 2L, 79L, 0L, 0L, 1369L, 19L, 1L,
    0L), .Dim = c(7L, 4L), .Dimnames = structure(list(c("-1",
    "-2", "-3", "0", "1", "2", NA), c("A", "C", "G", NA)), .Names =
c("",
    "")), class = "table"), LR17_SNP4 = structure(c(206L, 5L,
    0L, 2617L, 34L, 2L, 12L, 1787L, 30L, 2L, 2799L, 50L, 1L,
    46L, 13L, 0L, 0L, 130L, 1L, 0L, 0L), .Dim = c(7L, 3L), .Dimnames =
structure(list(
        c("-1", "-2", "-3", "0", "1", "2", NA), c("G", "T", NA
        )), .Names = c("", "")), class = "table"), LR20_SNP1 =
structure(c(0L,
    0L, 6L, 0L, 0L, 1065L, 2L, 7750L, 9L, 8L, 10L, 0L, 74L, 1L,
    0L), .Dim = c(5L, 3L), .Dimnames = structure(list(c("-1",
    "-2", "0", "1", NA), c("A", "G", NA)), .Names = c("", "")), class =
"table"),
    LR24_SNP1 = structure(c(690L, 1L, 0L, 8221L, 18L, 7L, 492L,
    2L, 1L, 8014L, 22L, 9L, 3L, 0L, 0L, 140L, 1L, 0L), .Dim = c(6L,
    3L), .Dimnames = structure(list(c("-1", "-2", "-3", "0",
    "1", NA), c("A", "G", NA)), .Names = c("", "")), class = "table"),
    LR36_SNP1 = structure(c(0L, 0L, 0L, 0L, 1L, 0L, 0L, 0L, 887L,
    1543L, 16L, 2L, 8086L, 77L, 1L, 83L, 0L, 0L, 0L, 0L, 1L,
    0L, 0L, 0L, 1L, 0L, 0L, 0L, 15L, 0L, 0L, 0L, 6L, 6L, 0L,
    0L, 180L, 1L, 0L, 0L), .Dim = c(8L, 5L), .Dimnames =
structure(list(
        c("-1", "-2", "-3", "-4", "0", "1", "2", NA), c("-",
        "A", "G", "T", NA)), .Names = c("", "")), class = "table"),
    LR40_SNP1 = structure(c(0L, 0L, 1L, 0L, 0L, 0L, 202L, 1L,
    5808L, 15L, 1L, 6L, 0L, 0L, 1L, 0L, 0L, 0L, 1L, 0L, 32L,
    1L, 0L, 0L), .Dim = c(6L, 4L), .Dimnames = structure(list(
        c("-1", "-3", "0", "1", "2", NA), c("A", "C", "T", NA
        )), .Names = c("", "")), class = "table"), LR44_SNP1 =
structure(c(0L,
    0L, 1L, 0L, 0L, 1L, 0L, 0L, 0L, 0L, 222L, 693L, 158L, 0L,
    0L, 1928L, 16L, 0L, 0L, 13L, 921L, 1741L, 124L, 3L, 1L, 2826L,
    49L, 1L, 1L, 37L, 15L, 27L, 4L, 0L, 0L, 134L, 2L, 0L, 0L,
    0L), .Dim = c(10L, 4L), .Dimnames = structure(list(c("-1",
    "-2", "-3", "-4", "-5", "0", "1", "2", "3", NA), c("A", "C",
    "T", NA)), .Names = c("", "")), class = "table"), LR44_SNP2 =
structure(c(0L,
    0L, 0L, 0L, 0L, 64L, 0L, 0L, 0L, 0L, 0L, 1L, 0L, 0L, 0L,
    3L, 0L, 0L, 0L, 0L, 221L, 688L, 159L, 0L, 0L, 1877L, 16L,
    0L, 0L, 8L, 930L, 1754L, 127L, 3L, 1L, 2873L, 50L, 1L, 1L,
    41L, 7L, 18L, 1L, 0L, 0L, 72L, 1L, 0L, 0L, 0L), .Dim = c(10L,
    5L), .Dimnames = structure(list(c("-1", "-2", "-3", "-4",
    "-5", "0", "1", "2", "3", NA), c("-", "A", "C", "T", NA)), .Names =
c("",
    "")), class = "table"), LR46_SNP1 = structure(c(0L, 69L,
    0L, 0L, 7L, 4494L, 9L, 1L, 0L, 4L, 0L, 0L, 0L, 53L, 0L, 0L
    ), .Dim = c(4L, 4L), .Dimnames = structure(list(c("-1", "0",
    "1", NA), c("A", "G", "T", NA)), .Names = c("", "")), class =
"table"),
    LR48_SNP1 = structure(c(648L, 125L, 4988L, 80L, 14L, 0L,
```

```
    0L, 1L, 0L, 0L, 4L, 0L, 181L, 1L, 0L), .Dim = c(5L, 3L), .Dimnames
= structure(list(
        c("-1", "-2", "0", "1", NA), c("C", "T", NA)), .Names = c("",
    "")), class = "table"), LR49_SNP1 = structure(c(0L, 18L,
    0L, 0L, 34L, 8576L, 5L, 2L, 0L, 75L, 0L, 0L), .Dim = 4:3, .Dimnames
= structure(list(
        c("-1", "0", "1", NA), c("A", "G", NA)), .Names = c("",
    "")), class = "table"), LR52_SNP1 = structure(c(1L, 0L, 0L,
    1L, 2L, 0L, 0L, 1L, 537L, 626L, 376L, 0L, 3572L, 26L, 1L,
    87L, 23L, 5L, 0L, 0L, 360L, 3L, 0L, 0L), .Dim = c(8L, 3L),
.Dimnames = structure(list(
        c("-1", "-2", "-3", "-4", "0", "1", "2", NA), c("C",
        "T", NA)), .Names = c("", "")), class = "table"), BRAF_SNP1 =
structure(c(13226L,
    2L, 2L, 62L), .Dim = c(1L, 4L), .Dimnames = structure(list(
        "0", c("A", "G", "T", NA)), .Names = c("", "")), class =
"table"))
```

Figure 7.2. Example Output of the Newcastle MSI Assay. Each variant file contains the read counts for the different lengths of each microsatellite marker in a given sample. Shown are examples of a clear MSS (ID54155 – Newcastle Assay Score = -25.2) and clear MSI-H sample (ID21269 – Newcastle Assay Score = +45.4).

7.5. Appendix E: MIPs of the *PMS2* Assay

MIP ID	MIP Start Position	MIP Stop Position	Capture Size	Probe Strand	MIP Sequence	Logistic Score
exon1_296530_SNP_a	6048631	6048745	155	-	GAGGTGAGCGGGGCTCGCAGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGGAAAGGGTGGAGCACAACG	0.736
exon1_0038	6048551	6048655	145	+	CAACACCCGATCCGCCTCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCTCGGCCATGTTCCCCCCATTT	0.871
exon2_0037	6045603	6045712	150	+	CATCCTGATTTTAACTGTGGGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNACTCAGTACCACCTGCCCA	0.909
exon2_0036	6045493	6045612	160	+	GATGGACTGACTTCCGATCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCAACAACATTCACAGATCATT	0.794
exon3_0035	6043630	6043729	140	+	CCCATGCTATCAGTTTTTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCCTTCGAAGTTTTCTTCTTCTA	0.741
exon3_0034	6043573	6043657	125	-	AATAATTGGGTCACATGTCTCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGGACTATGGAGTGGATCTT	0.857
exon4_0033	6043364	6043478	155	+	TATACATGATATCTAGTAACTGGCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTCCCCCCGAAAGCCAA	0.825
exon4_0032	6043291	6043410	160	-	CCCCTTAGAGAAACCTCTCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTCTTTCAGCTCTGAAACATCA	0.905
exon5_214973	6042172	6042291	160	+	GTGTTCAGTGAGAGACCCATGATNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTGGTCCCTCTGGGGCGG	0.911
exon5_0029	6042054	6042173	160	+	GTAGGGGGTTTTCTGGATANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGCTCATGTGCATTAACCAATA	0.857
exon6_197916	6038807	6038926	160	-	CTGTGGTATGCACAGGTGGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTTAAAGTAATGTAGAATATTG	0.672
exon6_0026	6038709	6038823	155	-	TATTGCTCTAGTGATTATAGCGGGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGCACCAATCAGCTTGG	0.791
exon7_0025	6036995	6037114	160	-	GGTTTGAGCTGTTCCGATGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGCTTGACTAAGAAAGAAAACA	0.949
exon7_0024	6036927	6037041	155	+	TGAGGCTTTGCAACTGAANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGCTCTCAGGATAAAATGTTCAA	0.707
exon8_0023	6035205	6035319	155	-	GTTTTTCTTTATCAACCGGCGGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNAATGTGCCATGTGATCGT	0.904
exon8_0022	6035135	6035229	135	+	CTCCATGCGTGCATTGTGAAANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNAAAAGTCAAAGGCATAAAG	0.892
exon9_0021	6031629	6031738	150	+	CCCCCGCATTCTAACAACNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNAATCAACAGAAATGTTAAGAAC	0.877
exon9_0020	6031574	6031693	160	-	GTGACTAATGCCTCAAAGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCTTTTCTACTCCTTGTATTTTG	0.769
exon10_0016	6029567	6029656	130	+	CAGACAGCGTCTCACTCTGTCGCCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGCCTTTTATCTGGAGT	0.891
exon10_0019	6029500	6029579	120	-	TGATAGGAATGTTTGATAGTGATGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTTGCCTTAGAATGCGT	0.766
exon10_0018	6029399	6029513	155	+	ACTGCCAACAAAAGCTTTTCCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNACACATTAGCTAAAAGCTT	0.726
exon11_116855	6027240	6027359	160	+	CACAGATGAACACAGTTCANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCCAAATCCGCTGCATGCATTT	0.839
exon11_0015	6027124	6027243	160	-	TCTCTTCGTCACACAACANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCCACGTTTGCTTAGGTAACTTA	0.840
exon11_0014	6027014	6027133	160	+	CGCAGTCTGGAAATGGACACGTCTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGCCTTTGTCAGAGATG	0.923
exon11_0008	6026904	6027023	160	-	GCACTTCCGTGGATTCTGAGGGGTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNATGCTGTCTTCTAGCA	0.883
exon11_0007	6026824	6026918	135	-	GGGGACAGGGGCTCGCAGGAANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNAGAGCGGAGGTGGAGAAGG	0.848
exon11_0013	6026754	6026863	150	-	ATGTGGACTGCCATTCAAACCAGGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTCAGCATCCCAGACAC	0.827
exon11_0012	6026659	6026763	145	+	GTCGTCAGTTTTAGGCGCTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGGAAAGAATTTCTTCTTTTTT	0.712
exon11_0011	6026569	6026668	140	+	GTGTTTGGGGTTGCGAGANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGCACAACTTTCTTATTAATTTT	0.789
exon11_0010	6026459	6026578	160	-	GAAGTTTAGGGCAAAGATTTGTCCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCAGCCTCTCAGGTTGA	0.882
exon11_0009	6026359	6026468	150	+	GTTCCCCTTCACTTTGCTGTGCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNAATAAAAATTTTAGATAA	0.259
exon12_0005	6022563	6022682	160	+	TAGAACACTGTAATAAAAAAAAAGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNATCCTCATTCAGTTTG	0.060
exon12_54552	6022488	6022607	160	-	GCACACCGTGCTCCAGGGGCANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGCAGTAAAACGATGTTTGC	0.922
exon12_0006	6022373	6022492	160	-	CCAGGAGTTTGAGACCAGCCCGGANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTAACTTCGAGATGCTG	0.779
exon13_0004	6018312	6018431	160	+	AAATAACAACACAAATAACAACACNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCATTAACAGCAGTTAA	0.433
exon13_34584	6018209	6018323	155	-	CTATAGCGGCTGGGTGTGGTGGCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTTGTGTTTTTCAGACCT	0.790
exon14_24167	6017361	6017460	140	+	ACATCTGAGGCCGGGCGTGGTGGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTACTAGTTGGCAAGGAA	0.946
exon14_21894	6017272	6017381	150	-	GCCGGCCTTCCCGAGTCAANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTTGTATTCTTTTAGCTCCAGT	0.811
exon14_19118	6017176	6017295	160	-	GCAGTCGAGGAAGGTCTCAGCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGATGAACTGATCTTCATGC	0.795
exon15_15882	6013072	6013191	160	-	ATGAGACACATCGCCAACCTGGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCTACTAAAACGTTGAACC	0.856
exon15_0001	6012959	6013078	160	-	CCTTTTTTGTTTTAAAATGAACCTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNACTGTCCCCATGGAAG	0.551
exon15_0002	6012840	6012959	160	-	ATGTTTGCATGCGCTCGTGTGTGTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGACAGAGTCTTCACTA	0.893

Table 7.3. Exon-targeting probes of the smMIP, sequencing-based *PMS2* Assay. MIP capture positions are specified chromosomal coordinates using the reference genome hg19.

MIP ID	MIP Start Position	MIP Stop Position	Capture Size	Probe Strand	MIP Sequence	Logistic Score	SNP Variant	SNP Position
intron1_1_0121	6048552	6048631	120	+	AGCTCTCAGCTCGCTCCANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTCGGCCATGTTCCCCCCATTTC	0.878	rs3735295	6048556
intron1_2_3_0120	6048333	6048412	120	-	GTGGACTTGAAGCGGGAGCGTGAANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTTCGGGGTCCTGGGCT	0.843	rs547956577	6048367
							rs12702466	6048359
intron1_4_0119	6048112	6048191	120	+	GCAACGCAGCGAAATCCCCGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTTAAAAAGGCCAGGCACGGT	0.887	rs7803118	6048186
intron1_5_0118_SNP_a	6047557	6047636	120	-	ATCGTGCCACTGCAGTCCGGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTAACTGGGCATGATGGCGGG	0.821	rs111277664	6047576
intron1_6_0117	6047057	6047136	120	-	CTACAATATAGGGGATGGCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTTTGGGCAGAGGAGGAAGAGG	0.937	rs7799214	6047131
intron1_7_0116	6046947	6047026	120	+	GACCGCAACCAGCCATCCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTTAGAAGAGATAGGGTTTCACC	0.862	rs113279793	6046990
intron1_8_0115	6046709	6046788	120	-	TGATGAAGGGAAAGGATTAGAGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCCATTGCCCTCTAGCATA	0.723	rs7797466	6046783
intron1_9_0114	6046558	6046637	120	-	GCTCTGGTGATAGAGTGGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCCCAATGATGACTAAATTTTAA	0.723	rs139778605	6046627
intron1_11_0111	6045896	6045975	120	+	CTTCTTTCTTGAAAAGTGAAAGCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTTCAACAGATCCTTACT	0.616	rs12702465	6045934
intron1_12_0112	6045826	6045905	120	+	ACTAAGATCCACATGGAGAAAACNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCCAGAAATAGAAACACT	0.694	rs1860462	6045881
intron2_1_0110	6045402	6045481	120	-	GTTGAAGTTAGCACCTATTATGTGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCAAGAAATGATCTGTG	0.676	NA	6045467
intron2_2_0109	6045347	6045426	120	-	ACAGCTTTTATTGTGTTCTTACGGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGACTATCAAGTGTTGA	0.576	rs62456183	6045385
intron2_3_4_0108	6045181	6045260	120	-	ACCTGGCCAACATGGTGAGACCTTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTTGCAGAACAGCTGGG	0.890	rs35657389	6045220
							rs34260367	6045192
intron2_5_47252	6044577	6044656	120	-	GTGACGAGCAAAACTCCATCTCAANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTCTCAGCTCCTCAGGA	0.988	rs12538339	6044650
intron2_6_7_8_0106	6044455	6044534	120	-	GAGGAAATTAGGGGATAGGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCAAAAGAAAAGAAAAGAATAA	0.546	rs148674986	6044528
							rs6978310	6044521
							rs6954766	6044515
intron2_9_0105	6044398	6044477	120	+	GTTCTATCTCATAAAGTTATTCTGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCAGACTTCCTGGATCC	0.581	rs77314273	6044428
intron2_11_0103	6043737	6043816	120	-	GCATGGGTCCGTTTTTAATNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCAATGAAATATACAAAATGAC	0.637	rs12538294	6043797
intron3_1_0102	6043442	6043521	120	+	ATTTCCCAAGACAGTGTTACTCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCTGAAAGAGAGTGTAAAG	0.859	rs117831773	6043495
intron4_1_0099	6043069	6043148	120	+	ATTGCTCACATTTCAGAAGTACTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCATTTATTATCTACATG	0.067	rs6463526	6043112
intron4_2_3_0100	6043009	6043088	120	+	ACTGTTTTAATAAGGGTAACCATCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNAATAGTGCTTTGGTGA	0.502	rs12537608	6043065
							rs6463525	6043057
intron4_4_0098	6042925	6043004	120	-	TAGACAGAGTCTCTCTTTGTCACCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTAGGGAAACATTTCAC	0.080	rs7788051	6042930
intron4_6_7_0097	6042559	6042638	120	-	TGTGAATAAATAAATCTTAAGATCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCCACGCTCTGCCCTAC	0.176	rs118162198	6042635
							rs10263455	6042563
intron4_8_0096	6042483	6042562	120	-	GCAAAGCATGTTTAGTGCTCTCTGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTAACAAAACAAATCTG	0.812	rs11769380	6042522
intron5_1_0093_SNP_a	6041812	6041901	130	+	AGAAAAAAGTAAATCATCCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGCTGGAACTACAGGTACATGT	0.459	rs6953340	6041897
intron5_2_3_4_0091	6041815	6041894	120	-	GTGACATGTACCTGTAGTTCCAGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNACTTTTTTCTTTTCATC	0.771	rs10951973	6041836
							rs10951972	6041818
							rs114887050	6041812
intron5_5_0090	6041587	6041666	120	-	GGTGGCGCGATCTTGGTTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNATCCTGTCTGGAAAAATATATA	0.765	rs2345057	6041653
intron5_6_202354_SNP_a	6041500	6041579	120	-	GGCGTGTACCACCACAGCTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCCTGGCTGGAGTGCGGTGGCG	0.785	rs12702464	6041506
intron5_9_0086	6040625	6040704	120	+	TAGATTATGAAAAACACACATACNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTGAGAAGACTCCATCTA	0.311	rs2345058	6040631
intron5_11_0084	6039882	6039961	120	+	GCCACTACATCCCAGCCTAGGTGANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNATGGTAGCACGGGCCT	0.859	rs13245536	6039941
intron5_12_0083	6039491	6039570	120	+	CTCTATGAACAAACAGAAACAACCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNAATCTCTCCTTTTGAA	0.801	rs58026649	6039529
intron5_13_0082	6039326	6039405	120	+	GACAGAGTGAGGCCCTATCTCTANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCACAGCTACTGAGGAGG	0.880	rs79245786	6039384
intron5_14_0081	6039199	6039278	120	-	CCTCCTGCCCTCAGCCTCCTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCATGCCATGACGCCCGGCTA	0.787	rs183113848	6039266
intron5_15_16_0080	6039057	6039136	120	+	GGTAAAACAACTAAAAAGTACAGCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCTGTAAAGAGGTGTCT	0.718	rs2345060	6039129
							rs7784177	6039080

MIP ID	MIP Start Position	MIP Stop Position	Capture Size	Probe Strand	MIP Sequence	Logistic Score	SNP Variant	SNP Position
intron6_1_2_0078	6038666	6038745	120	+	TGCCCAAACACAGAGCCGATATTTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTTTATTCTCCATTCTA	0.787	rs62456182	6038722
							NA	6038703
intron6_6_0077	6038252	6038331	120	+	AAAAAAACCGGCCGGGCTTGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGGAGAATTGCTTCAACGCGG	0.917	rs146370552	6038295
intron6_7_38788	6038146	6038225	120	-	CATCTTGTTGGCCAGACTGGCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGAAGCAATTCTCCTACCTT	0.919	rs12702463	6038179
int6snp8_46287	6038023	6038102	120	+	CACTAGAGGTCAGGAGTTCGAGGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGGGCCTTGTGATAGGCA	0.944	rs73331331	6038045
intron6_11_12_32205	6037651	6037730	120	+	GGGATGGTGTCTTGCTCTCTCGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGTATACGAGTGTCTCTGG	0.979	rs74189692	6037723
							rs75745061	6037715
intron6_13_0071	6037597	6037676	120	-	TTATTGTGGGCAATGTGTGGGTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCAAAAATTGATGTTTGAG	0.777	rs73331330	6037641
intron6_14_0070	6037463	6037542	120	-	ATGTGTTTGAACCACCCGGCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGGTACTTGTTTCATCTTTCT	0.915	rs76984954	6037509
intron6_15_0068	6037186	6037265	120	+	CGTGGGTTCAAGCAATCCTCCTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGAATCTTTTGTTTTGTTT	0.831	rs59790240	6037232
intron6_16_0069	6037153	6037232	120	-	CGCTTGTAATGTCAATAGCTTGACNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNAGGTATGATTGAGCCA	0.911	rs10267842	6037168
intron7_1_31009	6036781	6036860	120	+	GACACGAAACTATTAGCCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTGCTTGAACTCAGGAGGTAGAG	0.935	rs7793254	6036802
intron7_2_0065	6036441	6036520	120	-	GGTGGTTTTCTGTTGAGCAGGTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTATAAAACTGTGAGCAAT	0.904	rs12112229	6036515
intron7_3_0064	6036172	6036251	120	+	ATAGATAGATATCTTTACAGGACCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTCACGGCACTGCACTG	0.824	rs10235277	6036245
intron7_6_0063	6035682	6035761	120	-	GGTTTTGAAAGGGGAGGGANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGTGAAGGTGGTGTTCTTCTGG	0.959	rs115360158	6035756
intron7_7_8_0062	6035462	6035541	120	-	AGAATATTCCAGAAGGTTTGGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGAGGCTGAGGATTTATCAG	0.785	rs74448798	6035536
							rs2286680	6035507
intron7_9_0061	6035358	6035437	120	+	GAACCAAACCTTCTGGAATATTCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNAAAGTGAAATGAAAACA	0.739	rs2286681	6035428
intron8_1_2_3_0060	6035007	6035126	160	+	GCATAAAGAACAAACTAACACAANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTCTGAAGCTGGGGGCCA	0.878	rs3815383	6035081
							rs12534423	6035065
							rs12702462	6035021
intron8_7_0056	6034599	6034678	120	+	ATTTTTTTGAAACACAGTCTCGCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCCGAAAAACAATGGTTC	0.820	rs2345061	6034670
intron8_8_0055_SNP_a	6034563	6034642	120	+	GCAACAAGGTAAAGTAGTAGAAAANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTGTTGAATATCTTCTG	0.607	rs115227599	6034583
intron8_9_129189	6034397	6034486	130	+	CAGCCTTGCCTCAGAATACNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCAGCCTGGGTGACAAACAAGA	0.946	rs114014535	6034400
intron8_10_11_12_0052	6034272	6034386	155	-	CCACGCCTGGTTAATTTTTTTTTTGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNACGGAGTCTTGTTTGT	0.518	rs7798762	6034346
							rs191539307	6034324
							rs12540010	6034293
intron8_13_123202	6034153	6034247	135	-	AATTACAGGTGTGAGCCACCGCGCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGGTTAATTTTTTTTTG	0.645	rs111515162	6034203
intron8_14_15_16_0050	6034106	6034200	135	+	AAAACCATCCTGGCTAACAAGGTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTCCTGATAAACCCATAA	0.849	rs144186083	6034170
							rs7786901	6034157
							rs115649412	6034117
int8snp17_39758	6033843	6033922	120	-	GACGGAATTTCACCATGTGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNACCTCCCGAGTTCAAGCGATT	0.842	rs79159932	6033883
intron8_19_20_116465	6033448	6033527	120	+	GAAACCATGTAATCAGTCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTGCTTGCAGTGAGCAGAGATCA	0.912	rs180914892	6033480
							rs12702461	6033469
intron8_21_0048_SNP_a	6033222	6033301	120	+	GCCAATGTGGTGAAACCCCGTANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNAAACCATGGGCCGGGCGC	0.771	rs185403963	6033264
intron8_22_0047	6033063	6033142	120	+	GCCCTTATTTTCCTAGGAAAATAGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTTACTGAGGTACAATG	0.879	rs113516018	6033095
intron8_23_110701	6032703	6032782	120	+	CCTGAGCTCAGGAGTTCTTGACTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNACAAGCAAACTCTAAAG	0.871	rs12702460	6032748
intron8_24_0043	6032509	6032588	120	-	GCATTTCACTAGAATTTCTCATAGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNAATGTGATATTTTGGT	0.460	rs112099982	6032511
intron8_25_0045	6032429	6032508	120	-	CACTTTGAGATTGTGAGGAAACTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNACATTTAAGCTGCACTT	0.733	rs77319935	6032506
intron8_26_0042	6032281	6032360	120	+	GAAACTATCCCCACAATCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCATTTTTTTAACAGCTTTACTG	0.717	rs140551329	6032331
intron8_29_30_0041_SNP_a	6031951	6032030	120	+	ACCTCTGCCTCCGGGGCTCAAGTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGATTACAGGCAGGATTT	0.818	rs115643697	6031964
							rs7776504	6031957
intron9_2_0040	6030677	6030756	120	+	CCAAACTATAACCATGCATTGGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTAATAAAGATGCAGATAA	0.044	rs55842018	6030722
intron9_3_0039_SNP_a	6030584	6030663	120	-	ACCTCCGATTCCTGAGTTCAAGCGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTCATTATCTGCATCTT	0.872	rs117279697	6030588

MIP ID	MIP Start Position	MIP Stop Position	Capture Size	Probe Strand	MIP Sequence	Logistic Score	SNP Variant	SNP Position
intron9_4_0037_SNP_a	6030427	6030506	120	+	ATGGTGGCAGGTGCCTGTAATCCGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGAGACGAACAGATAAG	0.980	rs77283699	6030437
intron9_5_0036	6030226	6030305	120	+	CACTGACCCAGTAATTCTACTCCCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTAGCTTAACAATTATA	0.705	rs62456178	6030287
intron9_6_24915	6029774	6029853	120	+	GTCTCGAACTCACAACCTCANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCTCAGCCTCCGAAGTAGCTA	0.976	rs79815075	6029842
intron10_1_0034	6029079	6029158	120	+	AAAAAACTAGAGGTACTTGGAGGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCTCAAAAAAACAGACAT	0.798	rs112688886	6029145
intron10_4_19690	6028714	6028793	120	-	GTCAGGAGATCGAGACTATCTANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGGGATTTCCTGACTTCTA	0.720	rs12536167	6028768
intron10_5_0030	6028696	6028775	120	-	TCTAACACGGTGAAACTCTGTCTCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNAGTAGAATTTAGGCCA	0.561	rs112867177	6028696
intron10_9_10_0028	6028130	6028209	120	-	CCTGCCTTCCAGGTTCAAGTGATNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNACGAGCATAGATAGGCA	0.870	rs78549569	6028190
							rs139380339	6028187
intron10_11_0027	6027350	6027429	120	+	TATCCCAGCACTTTAGGAGGCCANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCAAAAGAGGAGATCCAC	0.956	rs111255573	6027367
intron11_1_0022	6026367	6026446	120	+	GCCCTAAACTTCCTGTAATTCTGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTTTAGATAAAAAGAGAA	0.546	rs111905775	6026384
int11snp4_26845	6025915	6025994	120	+	GCCACCACACCTAGCCTCAAANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNAGGTCTCCCTATGTTGTCC	0.986	rs144792821	6025980
intron11_5_0018	6025859	6025938	120	+	CAAGTGATCCTCCCTCCTTGGCTTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGCCGGAATTACAACTG	0.848	rs143580677	6025894
intron11_6_60611	6025804	6025883	120	+	ATTGTAGAGACGAGGTCTCCCTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCCATGTGTTTGGAGTCTG	0.840	rs62456177	6025845
intron11_7_59003	6025676	6025755	120	+	GCTGGTGCTGGCCTAAAAGACNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGCCAGCACAAATACCACCT	0.922	rs201018511	6025703
intron11_10_0016	6024754	6024833	120	-	CTGTCTTAAAAAAAAAAGTCCCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGGGGCTGAGGCAGGAGGC	0.509	rs141374948	6024794
intron11_11_12_0015	6023968	6024047	120	+	TTTCCGCAGTATCAGCGCGGTGATNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNAATCTGGACAAAAGAC	0.923	rs11978631	6024038
							rs6964944	6024035
intron11_13_0014	6023877	6023956	120	-	AGAAGGATTGATATCCAGCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGGTGGTGTCTGTCTTTTGTCC	0.917	rs11972027	6023951
intron11_17_46412	6022955	6023034	120	-	GCCTGACCAACATGGTGGAACTCTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTGGGATGGTTAAGCCG	0.880	rs55782426	6023033
int11snp18_19532	6022955	6023034	120	-	GCCTGACCAACATGGTGGAACTCTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTGGGATGGTTAAGCCG	0.880	rs28462984	6023020
intron11_19_20_0013_SNP_a	6022711	6022790	120	+	TACAGTGCAATGGCACAATCTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGTAATAAAAAAAAAGTCAA	0.597	rs144069993	6022785
							rs2711201	6022773
intron11_21_22_0012	6022618	6022697	120	-	CGATGTTTGCAGAAATGGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTTGTATTGTTTGACTTTTTTTT	0.230	rs55954143	6022629
							rs1805326	6022626
int12snp1_17665	6022190	6022269	120	+	CTAGATATTTTTTATTTTTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCGTTGGCCAGGCTGGTCTCCA	0.250	rs556171361	6022260
int12snp2_16753	6022172	6022251	120	+	GGCACCGCGCCTGGCCAACTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGAGGCGGGGTTTCACCACGT	0.717	rs58032887	6022178
int12snp4_12254	6021557	6021636	120	-	GTTACCATTTATTCAGTGCCTAGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGAGACCTAATAGTTTAT	0.431	rs79192116	6021626
intron12_5_12502	6020813	6020892	120	-	CTCTCTCCTCGGCCAAACNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCAGAGTGAGACAGTGAGACTCT	0.977	rs187757184	6020865
intron12_6_0009	6020709	6020788	120	+	ATTCCAGTTTGGCCGAGGAGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNCGAAAATCCATGACCTCTTC	0.918	rs118178491	6020774
intron12_7_0008	6020524	6020603	120	+	ACACAGGTTCAGTGGTACANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGCATTCCACAGGAGAGTACTC	0.931	rs113937567	6020598
intron12_9_26345	6020019	6020098	120	+	ACAAGACCTCAGCCGGGCANNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTTGAAGTGTGGTCCATCTGGG	0.967	rs62454752	6020084
intron12_10_0007	6019298	6019377	120	+	CTGGGCTCAAGCGATCCTTCCACTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTCCTGGATTTACTGTT	0.567	rs372985395	6019340
intron12_11_0006	6019186	6019265	120	-	CTAACCCAGTGTTTCAAATTCNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGGAGGAGAGAGGGATGGAG	0.929	rs2692544	6019224
intron12_13_0004	6018391	6018470	120	+	CACAAAACTTCCTGAGAAGTTCCTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTAGCTTTACAGCAGAA	0.890	rs75973354	6018444
intron13_2_7117	6017801	6017880	120	+	AAGACAAGAATCTATAGTTCTGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNGGAAACTCACAAAATGCT	0.481	rs116623447	6017850
intron14_1_10676	6017078	6017157	120	-	CGTCGTAGCTGAGCATCTGTGNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTGCAGTCGAGGAAGGTCTC	0.936	rs562356357	6017084
int14snp3_1360	6016391	6016470	120	-	GTTGACTCCAACGCAGTTAGCGTTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNTCTCCTTTTGCCCCTT	0.944	rs377628616	6016417
intron14_4_5191	6013280	6013359	120	+	CAGGCTCCTTGTGGCTCCTNNNNCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNAAGGCTGGACAAGATTACAGC	0.846	rs527425812	6013291

Table 7.4. SNP-targeting probes of the smMIP, sequencing-based *PMS2* Assay. MIP capture positions are specified chromosomal coordinates using the reference genome hg19. Where a single probe captures multiple SNPs, the rs numbers and positions of each are detailed together with the one MIP ID.

7.6. Appendix F: Samples Analysed by the *PMS2* Assay

Sample	Defective MMR Gene
01088	PMS2
01100	PMS2
01123	PMS2
01130	PMS2
01135	PMS2
01137	PMS2
02001	PMS2
02008	PMS2
04002	PMS2
04058	PMS2
04093	PMS2
04112	PMS2
04122	PMS2
04130	PMS2
06018	PMS2
06022	PMS2
06030	PMS2
06040	PMS2
10001	PMS2
10013	PMS2
10017	PMS2
10051	PMS2
10056	PMS2
10057	PMS2
10058	PMS2
10059	PMS2
10063	PMS2
10080	PMS2
10081	PMS2
11001	PMS2
11015	PMS2
11017	PMS2
12002	PMS2
12005	PMS2

Sample	Defective MMR Gene
12015	PMS2
12051	PMS2
12057	PMS2
12061	PMS2
16006	PMS2
16008	PMS2
16009	PMS2
16012	PMS2
13064	PMS2
13079	PMS2
13082	PMS2
13090	PMS2
14003	PMS2
14014	PMS2
14024	PMS2
15012	PMS2
16013	PMS2
16014	PMS2
16039	PMS2
16056	PMS2
16065	PMS2
17005	PMS2
17032	PMS2
17033	PMS2
17066	PMS2
17074	PMS2
17075	PMS2
17076	PMS2
17088	PMS2
17106	PMS2
17116	PMS2
18023	PMS2
02006	MLH1
03005	MLH1

Sample	Defective MMR Gene
03006	MLH1
04007	MLH1
04137	MLH1
05002	MLH1
05005	MLH1
06003	MLH1
06004	MLH1
07037	MLH1
08009	MLH1
09011	MLH1
10011	MLH1
10027	MLH1
11006	MLH1
11026	MLH1
12009	MLH1
12024	MLH1
13002	MLH1
14001	MLH1
14007	MLH1
15031	MLH1
16017	MLH1
16021	MLH1
17002	MLH1
17009	MLH1
17015	MLH1
18005	MLH1
18013	MLH1
19001	MLH1
19004	MLH1
21002	MLH1
23003	MLH1
23007	MLH1
25117	MLH1
02007	MSH2
02010	MSH2

Sample	Defective MMR Gene
03001	MSH2
03002	MSH2
04006	MSH2
04008	MSH2
05001	MSH2
05003	MSH2
06007	MSH2
06009	MSH2
07002	MSH2
07003	MSH2
25112	MSH2
09002	MSH2
09004	MSH2
10003	MSH2
10006	MSH2
11003	MSH2
11004	MSH2
12003	MSH2
12006	MSH2
13001	MSH2
14006	MSH2
15002	MSH2
15003	MSH2
16004	MSH2
16007	MSH2
17003	MSH2
17004	MSH2
18001	MSH2
18002	MSH2
19014	MSH2
19017	MSH2
21003	MSH2
23004	MSH2
25103	MSH2
25111	MSH2

Table 7.5. Samples from the CaPP3 tissue resource analysed by the MIP-based *PMS2* assay. In total, 66 samples with a confirmed *PMS2* mutation were analysed, as well as 72 controls (35 with an *MLH1* mutation and 37 with an *MSH2* Mutation).

7.7. Appendix G: Variants Identified by the PMS2 Assay

Sample	Defective MMR Gene	Uploaded Variation	Location	Allele	Consequence	cDNA Position	CDS Position	Protein Position	Amino Acids	Codons	Impact	BioType	SIFT	PolyPhen	EXON	VAF	QUAL
11001	PMS2	rs63750250	chr7:6026564-6026565	T	frameshift_variant	1937-1938	1831-1832	611	I/NX	att/aAtt	HIGH	protein_coding	-	-	11/15	0.376993166	16011.64
15012	PMS2	rs267608149	chr7:6036957-6036958	A	frameshift_variant,splice_region_variant	908-909	802-803	268	Y/LX	tac/tTac	HIGH	protein_coding	-	-	7/15	0.402799378	19259.64
17002	MLH1	rs267608149	chr7:6036957-6036958	A	frameshift_variant,splice_region_variant	908-909	802-803	268	Y/LX	tac/tTac	HIGH	protein_coding	-	-	7/15	0.435368755	20534.64
12005	PMS2	chr7_6037019_-/CTTCACACAC	chr7:6037018-6037019	CTTCACACAC	frameshift_variant	847-848	741-742	247-248	-/VCEX	-/GTGTGTGAAG	HIGH	protein_coding	-	-	7/15	0.314699793	5378.64
17032	PMS2	chr7_6037019_-/CTTCACACAC	chr7:6037018-6037019	CTTCACACAC	frameshift_variant	847-848	741-742	247-248	-/VCEX	-/GTGTGTGAAG	HIGH	protein_coding	-	-	7/15	0.438834951	8612.64
12015	PMS2	chr7_6037019_-/CTTCACACAC	chr7:6037018-6037019	CTTCACACAC	frameshift_variant	847-848	741-742	247-248	-/VCEX	-/GTGTGTGAAG	HIGH	protein_coding	-	-	7/15	0.623100304	32931.64
17033	PMS2	chr7_6037019_-/CTTCACACAC	chr7:6037018-6037019	CTTCACACAC	frameshift_variant	847-848	741-742	247-248	-/VCEX	-/GTGTGTGAAG	HIGH	protein_coding	-	-	7/15	0.629820051	9846.64
16056	PMS2	chr7_6037019_-/CTTCACACAC	chr7:6037018-6037019	CTTCACACAC	frameshift_variant	847-848	741-742	247-248	-/VCEX	-/GTGTGTGAAG	HIGH	protein_coding	-	-	7/15	0.677419355	8644.64
10063	PMS2	chr7_6037019_-/CTTCACACAC	chr7:6037018-6037019	CTTCACACAC	frameshift_variant	847-848	741-742	247-248	-/VCEX	-/GTGTGTGAAG	HIGH	protein_coding	-	-	7/15	0.685618729	8317.64
11017	PMS2	chr7_6037019_-/CTTCACACAC	chr7:6037018-6037019	CTTCACACAC	frameshift_variant	847-848	741-742	247-248	-/VCEX	-/GTGTGTGAAG	HIGH	protein_coding	-	-	7/15	0.737270876	29616.64
13064	PMS2	chr7_6037019_-/CTTCACACAC	chr7:6037018-6037019	CTTCACACAC	frameshift_variant	847-848	741-742	247-248	-/VCEX	-/GTGTGTGAAG	HIGH	protein_coding	-	-	7/15	0.93164557	15363.64
13082	PMS2	chr7_6037019_-/CTTCACACAC	chr7:6037018-6037019	CTTCACACAC	frameshift_variant	847-848	741-742	247-248	-/VCEX	-/GTGTGTGAAG	HIGH	protein_coding	-	-	7/15	0.987654321	13514.06
11015	PMS2	chr7_6038813_G/A	chr7:6038813	A	stop_gained	737	631	211	R/*	Cga/Tga	HIGH	protein_coding	-	-	6/15	0.243224217	15054.6
16006	PMS2	chr7_6038885_-/C	chr7:6038884-6038885	C	frameshift_variant	665-666	559-560	187	V/GX	gtc/gGtc	HIGH	protein_coding	-	-	6/15	0.981519507	42219.06
01135	PMS2	rs121434629	chr7:6045549	A	missense_variant	243	137	46	S/I	aGt/aTt	MODERATE	protein_coding	deleterious(0)	probably_damaging(1)	2/15	0.434036939	10300.6
12061	PMS2	rs121434629	chr7:6045549	A	missense_variant	243	137	46	S/I	aGt/aTt	MODERATE	protein_coding	deleterious(0)	probably_damaging(1)	2/15	0.536959554	12574.6
01123	PMS2	rs121434629	chr7:6045549	A	missense_variant	243	137	46	S/I	aGt/aTt	MODERATE	protein_coding	deleterious(0)	probably_damaging(1)	2/15	0.946428571	18666.03
06040	PMS2	rs121434629	chr7:6045549	A	missense_variant	243	137	46	S/I	aGt/aTt	MODERATE	protein_coding	deleterious(0)	probably_damaging(1)	2/15	0.964412811	26407.03
04122	PMS2	rs121434629	chr7:6045549	A	missense_variant	243	137	46	S/I	aGt/aTt	MODERATE	protein_coding	deleterious(0)	probably_damaging(1)	2/15	0.968879668	18413.03
18023	PMS2	rs121434629	chr7:6045549	A	missense_variant	243	137	46	S/I	aGt/aTt	MODERATE	protein_coding	deleterious(0)	probably_damaging(1)	2/15	0.983050847	26235.03
01130	PMS2	rs121434629	chr7:6045549	A	missense_variant	243	137	46	S/I	aGt/aTt	MODERATE	protein_coding	deleterious(0)	probably_damaging(1)	2/15	0.990532544	35410.03
16039	PMS2	chr7_6045613_G/A	chr7:6045613	A	stop_gained	179	73	25	Q/*	Cag/Tag	HIGH	protein_coding	-	-	2/15	0.403563129	19174.6
16014	PMS2	chr7_6013094_T/C	chr7:6013094	C	missense_variant	2631	2525	842	N/S	aAc/aGc	MODERATE	protein_coding	deleterious(0.01)	probably_damaging(0.911)	15/15	0.153846154	43.6
06003	MLH1	chr7_6013095_T/A	chr7:6013095	A	missense_variant	2630	2524	842	N/Y	Aac/Tac	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.993)	15/15	0.169312169	129.6
14006	MSH2	chr7_6013095_T/A	chr7:6013095	A	missense_variant	2630	2524	842	N/Y	Aac/Tac	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.993)	15/15	0.14893617	76.6
18005	MLH1	chr7_6013095_T/C	chr7:6013095	C	missense_variant	2630	2524	842	N/D	Aac/Gac	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.987)	15/15	0.209580838	232.6
04112	PMS2	chr7_6013098_A/G	chr7:6013098	G	missense_variant	2627	2521	841	W/R	Tgg/Cgg	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.997)	15/15	0.217647059	89.6
16065	PMS2	chr7_6013155_G/A	chr7:6013155	A	missense_variant	2570	2464	822	L/F	Ctt/Ttt	MODERATE	protein_coding	deleterious(0)	probably_damaging(1)	15/15	0.5	29.6
25117	MLH1	chr7_6017283_G/-	chr7:6017283	-	frameshift_variant	2487	2381	794	P/X	cCt/ct	HIGH	protein_coding	-	-	14/15	0.333333333	19.6
16039	PMS2	chr7_6017344_T/C	chr7:6017344	C	missense_variant	2426	2320	774	K/E	Aaa/Gaa	MODERATE	protein_coding	deleterious(0)	possibly_damaging(0.823)	14/15	0.666666667	32.61
06030	PMS2	chr7_6022507_T/A	chr7:6022507	A	missense_variant	2228	2122	708	N/Y	Aac/Tac	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.996)	12/15	0.204545455	28.6
12003	MSH2	chr7_6022507_T/A	chr7:6022507	A	missense_variant	2228	2122	708	N/Y	Aac/Tac	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.996)	12/15	0.185714286	28.6
01100	PMS2	chr7_6022509_T/A	chr7:6022509	A	missense_variant	2226	2120	707	Y/F	tAt/tTt	MODERATE	protein_coding	tolerated(0.12)	possibly_damaging(0.672)	12/15	0.163636364	133.6
05003	MSH2	chr7_6022509_T/C	chr7:6022509	C	missense_variant	2226	2120	707	Y/C	tAt/tGt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.997)	12/15	0.155642023	267.6
06003	MLH1	chr7_6022509_T/C	chr7:6022509	C	missense_variant	2226	2120	707	Y/C	tAt/tGt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.997)	12/15	0.148080439	208.6
07003	MSH2	chr7_6022509_T/C	chr7:6022509	C	missense_variant	2226	2120	707	Y/C	tAt/tGt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.997)	12/15	0.256862745	1128.6
12015	PMS2	chr7_6022509_T/C	chr7:6022509	C	missense_variant	2226	2120	707	Y/C	tAt/tGt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.997)	12/15	0.194711538	519.6
12051	PMS2	chr7_6022509_T/C	chr7:6022509	C	missense_variant	2226	2120	707	Y/C	tAt/tGt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.997)	12/15	0.450980392	1457.6
12003	MSH2	chr7_6022510_A/T	chr7:6022510	T	missense_variant	2225	2119	707	Y/N	Tat/Aat	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.991)	12/15	0.185714286	31.6
23007	MLH1	chr7_6022510_A/G	chr7:6022510	G	missense_variant	2225	2119	707	Y/H	Tat/Cat	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.936)	12/15	0.179591837	124.6
14006	MSH2	chr7_6022512_T/C	chr7:6022512	C	missense_variant	2223	2117	706	K/R	aAg/aGg	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.998)	12/15	0.376068376	379.6
23007	MLH1	chr7_6022512_T/C	chr7:6022512	C	missense_variant	2223	2117	706	K/R	aAg/aGg	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.998)	12/15	0.198924731	125.6
01123	PMS2	chr7_6022568_T/C	chr7:6022568	C	missense_variant	2167	2061	687	I/M	atA/atG	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.997)	12/15	0.125	13.59
06007	MSH2	chr7_6022596_A/G	chr7:6022596	G	missense_variant	2139	2033	678	I/T	aTc/aCc	MODERATE	protein_coding	deleterious(0)	possibly_damaging(0.864)	12/15	0.241975309	616.6
01088	PMS2	chr7_6022599_T/C	chr7:6022599	C	missense_variant	2136	2030	677	E/G	gAa/gGa	MODERATE	protein_coding	deleterious(0)	possibly_damaging(0.83)	12/15	0.148148148	32.6
11017	PMS2	chr7_6022599_T/C	chr7:6022599	C	missense_variant	2136	2030	677	E/G	gAa/gGa	MODERATE	protein_coding	deleterious(0)	possibly_damaging(0.83)	12/15	0.186813187	136.6
12057	PMS2	chr7_6022599_T/A	chr7:6022599	A	missense_variant	2136	2030	677	E/V	gAa/gTa	MODERATE	protein_coding	deleterious(0)	possibly_damaging(0.696)	12/15	0.244285074	1287.6
13002	MLH1	chr7_6022599_T/A	chr7:6022599	A	missense_variant	2136	2030	677	E/V	gAa/gTa	MODERATE	protein_coding	deleterious(0)	possibly_damaging(0.696)	12/15	0.201834862	170.6
14001	MLH1	chr7_6022599_T/A	chr7:6022599	A	missense_variant	2136	2030	677	E/V	gAa/gTa	MODERATE	protein_coding	deleterious(0)	possibly_damaging(0.696)	12/15	0.188034188	119.6
14014	PMS2	chr7_6022599_T/A	chr7:6022599	A	missense_variant	2136	2030	677	E/V	gAa/gTa	MODERATE	protein_coding	deleterious(0)	possibly_damaging(0.696)	12/15	0.142493639	46.6
16065	PMS2	chr7_6022599_T/A	chr7:6022599	A	missense_variant	2136	2030	677	E/V	gAa/gTa	MODERATE	protein_coding	deleterious(0)	possibly_damaging(0.696)	12/15	0.150717703	50.6
17075	PMS2	chr7_6022599_T/A	chr7:6022599	A	missense_variant	2136	2030	677	E/V	gAa/gTa	MODERATE	protein_coding	deleterious(0)	possibly_damaging(0.696)	12/15	0.215517241	200.6
17116	PMS2	chr7_6022599_T/A	chr7:6022599	A	missense_variant	2136	2030	677	E/V	gAa/gTa	MODERATE	protein_coding	deleterious(0)	possibly_damaging(0.696)	12/15	0.232758621	262.6
21003	MSH2	chr7_6022599_T/A	chr7:6022599	A	missense_variant	2136	2030	677	E/V	gAa/gTa	MODERATE	protein_coding	deleterious(0)	possibly_damaging(0.696)	12/15	0.1875	75.6
23004	MSH2	chr7_6022599_T/A	chr7:6022599	A	missense_variant	2136	2030	677	E/V	gAa/gTa	MODERATE	protein_coding	deleterious(0)	possibly_damaging(0.696)	12/15	0.173396675	161.6
25111	MSH2	chr7_6022599_T/A	chr7:6022599	A	missense_variant	2136	2030	677	E/V	gAa/gTa	MODERATE	protein_coding	deleterious(0)	possibly_damaging(0.696)	12/15	0.171122995	62.6
12003	MSH2	chr7_6026433_C/A	chr7:6026433	A	stop_gained	2069	1963	655	G/*	Gga/Tga	HIGH	protein_coding	-	-	11/15	0.237077342	1035.6
12057	PMS2	chr7_6026433_C/A	chr7:6026433	A	stop_gained	2069	1963	655	G/*	Gga/Tga	HIGH	protein_coding	-	-	11/15	0.244285074	1287.6
16014	PMS2	chr7_6026433_C/A	chr7:6026433	A	stop_gained	2069	1963	655	G/*	Gga/Tga	HIGH	protein_coding	-	-	11/15	0.242294521	1013.6

Sample	Defective MMR Gene	Uploaded Variation	Location	Allele	Consequence	cDNA_position	CDS_position	Protein_position	Amino_acids	Codons	IMPACT	BIOTYPE	SIFT	PolyPhen	EXON	VAF	QUAL
17002	MLH1	chr7_6026433_C/A	chr7:6026433	A	stop_gained	2069	1963	655	G/*	Gga/Tga	HIGH	protein_coding	-	-	11/15	0.209192906	322.6
17003	MSH2	chr7_6026433_C/A	chr7:6026433	A	stop_gained	2069	1963	655	G/*	Gga/Tga	HIGH	protein_coding	-	-	11/15	0.278366112	2054.6
17004	MSH2	chr7_6026433_C/A	chr7:6026433	A	stop_gained	2069	1963	655	G/*	Gga/Tga	HIGH	protein_coding	-	-	11/15	0.226868327	571.6
17032	PMS2	chr7_6026433_C/A	chr7:6026433	A	stop_gained	2069	1963	655	G/*	Gga/Tga	HIGH	protein_coding	-	-	11/15	0.217495987	607.6
17033	PMS2	chr7_6026433_C/A	chr7:6026433	A	stop_gained	2069	1963	655	G/*	Gga/Tga	HIGH	protein_coding	-	-	11/15	0.26875	1579.6
17088	PMS2	chr7_6026433_C/A	chr7:6026433	A	stop_gained	2069	1963	655	G/*	Gga/Tga	HIGH	protein_coding	-	-	11/15	0.194726166	30.6
18001	MSH2	chr7_6026433_C/A	chr7:6026433	A	stop_gained	2069	1963	655	G/*	Gga/Tga	HIGH	protein_coding	-	-	11/15	0.268973214	1563.6
18023	PMS2	chr7_6026433_C/A	chr7:6026433	A	stop_gained	2069	1963	655	G/*	Gga/Tga	HIGH	protein_coding	-	-	11/15	0.218370884	847.6
25103	MSH2	chr7_6026433_C/A	chr7:6026433	A	stop_gained	2069	1963	655	G/*	Gga/Tga	HIGH	protein_coding	-	-	11/15	0.208129469	291.6
03001	MSH2	rs63751028	chr7:6026877	A	missense_variant	1625	1519	507	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.726)	11/15	0.12125	350.6
18013	MLH1	rs63751028	chr7:6026877	A	missense_variant	1625	1519	507	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.726)	11/15	0.142857143	236.6
11001	PMS2	chr7_6026880_-/AGTCAT	chr7:6026879-6026880	AGTCAT	inframe_insertion	1622-1623	1516-1517	506	F/YDF	ttc/tATGACTtc	MODERATE	protein_coding	-	-	11/15	0.111597374	962.64
21003	MSH2	chr7_6026880_-/AGTACT	chr7:6026879-6026880	AGTACT	stop_gained	1622-1623	1516-1517	506	F/*YF	ttc/tAGTACTtc	HIGH	protein_coding	-	-	11/15	0.105182927	1398.64
11015	PMS2	chr7_6027104_T/A	chr7:6027104	A	missense_variant	1398	1292	431	E/V	gAg/gTg	MODERATE	protein_coding	deleterious(0.05)	possibly_damaging(0.451)	11/15	0.193009119	3314.6
15002	MSH2	chr7_6027104_T/A	chr7:6027104	A	missense_variant	1398	1292	431	E/V	gAg/gTg	MODERATE	protein_coding	deleterious(0.05)	possibly_damaging(0.451)	11/15	0.206438632	7317.6
16056	PMS2	chr7_6027104_T/A	chr7:6027104	A	missense_variant	1398	1292	431	E/V	gAg/gTg	MODERATE	protein_coding	deleterious(0.05)	possibly_damaging(0.451)	11/15	0.310225303	2813.6
17004	MSH2	chr7_6027104_T/A	chr7:6027104	A	missense_variant	1398	1292	431	E/V	gAg/gTg	MODERATE	protein_coding	deleterious(0.05)	possibly_damaging(0.451)	11/15	0.199314188	1552.6
17015	MLH1	chr7_6027104_T/A	chr7:6027104	A	missense_variant	1398	1292	431	E/V	gAg/gTg	MODERATE	protein_coding	deleterious(0.05)	possibly_damaging(0.451)	11/15	0.131055901	1039.6
01130	PMS2	chr7_6029547_A/G	chr7:6029547	G	missense_variant	1134	1028	343	I/T	aTt/aCt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.993)	10/15	0.179649123	2312.6
04137	MLH1	chr7_6029547_A/G	chr7:6029547	G	missense_variant	1134	1028	343	I/T	aTt/aCt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.993)	10/15	0.13437058	931.6
05005	MLH1	chr7_6029547_A/G	chr7:6029547	G	missense_variant	1134	1028	343	I/T	aTt/aCt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.993)	10/15	0.137457045	471.6
10013	PMS2	chr7_6029547_A/G	chr7:6029547	G	missense_variant	1134	1028	343	I/T	aTt/aCt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.993)	10/15	0.141737892	869.6
12009	MLH1	chr7_6029547_A/G	chr7:6029547	G	missense_variant	1134	1028	343	I/T	aTt/aCt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.993)	10/15	0.127615063	47.6
12024	MLH1	chr7_6029547_A/G	chr7:6029547	G	missense_variant	1134	1028	343	I/T	aTt/aCt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.993)	10/15	0.146282974	618.6
15031	MLH1	chr7_6029547_A/G	chr7:6029547	G	missense_variant	1134	1028	343	I/T	aTt/aCt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.993)	10/15	0.154135338	533.6
16021	MLH1	chr7_6029547_A/G	chr7:6029547	G	missense_variant	1134	1028	343	I/T	aTt/aCt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.993)	10/15	0.116620112	30.6
16039	PMS2	chr7_6029547_A/G	chr7:6029547	G	missense_variant	1134	1028	343	I/T	aTt/aCt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.993)	8/15	0.140156454	42.6
17075	PMS2	chr7_6029547_A/G	chr7:6029547	G	missense_variant	1134	1028	343	I/T	aTt/aCt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.993)	10/15	0.172037138	3522.6
18023	PMS2	chr7_6029547_A/G	chr7:6029547	G	missense_variant	1134	1028	343	I/T	aTt/aCt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.993)	10/15	0.148357871	1407.6
01123	PMS2	chr7_6031604_C/G	chr7:6031604	G	missense_variant,splice_region_variant	1094	988	330	E/Q	Gaa/Caa	MODERATE	protein_coding	deleterious(0.02)	possibly_damaging(0.839)	7/15	0.170053476	40.64
01130	PMS2	chr7_6031604_C/G	chr7:6031604	G	missense_variant,splice_region_variant	1094	988	330	E/Q	Gaa/Caa	MODERATE	protein_coding	deleterious(0.02)	possibly_damaging(0.839)	7/15	0.166666667	815.6
06004	MLH1	chr7_6031604_C/T	chr7:6031604	T	missense_variant,splice_region_variant	1094	988	330	E/K	Gaa/Aaa	MODERATE	protein_coding	deleterious(0.02)	possibly_damaging(0.72)	7/15	0.169242658	190.6
08009	MLH1	chr7_6031604_C/G	chr7:6031604	G	missense_variant,splice_region_variant	1094	988	330	E/Q	Gaa/Caa	MODERATE	protein_coding	deleterious(0.02)	possibly_damaging(0.839)	7/15	0.166257669	572.64
10056	PMS2	chr7_6031604_C/T	chr7:6031604	T	missense_variant,splice_region_variant	1094	988	330	E/K	Gaa/Aaa	MODERATE	protein_coding	deleterious(0.02)	possibly_damaging(0.72)	7/15	0.16539924	10.6
10063	PMS2	chr7_6031604_C/G	chr7:6031604	G	missense_variant,splice_region_variant	1094	988	330	E/Q	Gaa/Caa	MODERATE	protein_coding	deleterious(0.02)	possibly_damaging(0.839)	7/15	0.157170923	637.6
10080	PMS2	chr7_6031604_C/T	chr7:6031604	T	missense_variant,splice_region_variant	1094	988	330	E/K	Gaa/Aaa	MODERATE	protein_coding	deleterious(0.02)	possibly_damaging(0.72)	7/15	0.25253664	1728.6
11015	PMS2	chr7_6031604_C/G	chr7:6031604	G	missense_variant,splice_region_variant	1094	988	330	E/Q	Gaa/Caa	MODERATE	protein_coding	deleterious(0.02)	possibly_damaging(0.839)	7/15	0.175799087	1252.6
12009	MLH1	chr7_6031604_C/T	chr7:6031604	T	missense_variant,splice_region_variant	1094	988	330	E/K	Gaa/Aaa	MODERATE	protein_coding	deleterious(0.02)	possibly_damaging(0.72)	7/15	0.195612431	885.6
12061	PMS2	chr7_6031604_C/T	chr7:6031604	T	missense_variant,splice_region_variant	1094	988	330	E/K	Gaa/Aaa	MODERATE	protein_coding	deleterious(0.02)	possibly_damaging(0.72)	7/15	0.177836412	147.6
13064	PMS2	chr7_6031604_C/T	chr7:6031604	T	missense_variant,splice_region_variant	1094	988	330	E/K	Gaa/Aaa	MODERATE	protein_coding	deleterious(0.02)	possibly_damaging(0.72)	7/15	0.169861555	137.6
13079	PMS2	chr7_6031604_C/G	chr7:6031604	G	missense_variant,splice_region_variant	1094	988	330	E/Q	Gaa/Caa	MODERATE	protein_coding	deleterious(0.02)	possibly_damaging(0.839)	7/15	0.15551082	168.6
13082	PMS2	chr7_6031604_C/G	chr7:6031604	G	missense_variant,splice_region_variant	1094	988	330	E/Q	Gaa/Caa	MODERATE	protein_coding	deleterious(0.02)	possibly_damaging(0.839)	7/15	0.14676259	75.6
14006	MSH2	chr7_6031604_C/A	chr7:6031604	A	stop_gained,splice_region_variant	1094	988	330	E/*	Gaa/Taa	HIGH	protein_coding	-	-	7/15	0.136823319	328.6
16021	MLH1	chr7_6031604_C/T	chr7:6031604	T	missense_variant,splice_region_variant	1094	988	330	E/K	Gaa/Aaa	MODERATE	protein_coding	deleterious(0.02)	possibly_damaging(0.72)	7/15	0.178696566	315.6
17009	MLH1	chr7_6031604_C/G	chr7:6031604	G	missense_variant,splice_region_variant	1094	988	330	E/Q	Gaa/Caa	MODERATE	protein_coding	deleterious(0.02)	possibly_damaging(0.839)	7/15	0.16851595	977.6
18013	MLH1	chr7_6031604_C/G	chr7:6031604	G	missense_variant,splice_region_variant	1094	988	330	E/Q	Gaa/Caa	MODERATE	protein_coding	deleterious(0.02)	possibly_damaging(0.839)	9/15	0.168704156	398.6
01137	PMS2	chr7_6035181_G/C	chr7:6035181	C	missense_variant	993	887	296	P/R	cCt/cGt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.996)	8/15	0.145746579	510.6
02007	MSH2	chr7_6035181_G/T	chr7:6035181	T	missense_variant	993	887	296	P/H	cCt/cAt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.996)	8/15	0.172477064	3548.6
03001	MSH2	chr7_6035181_G/A	chr7:6035181	A	missense_variant	993	887	296	P/L	cCt/cTt	MODERATE	protein_coding	deleterious(0.01)	possibly_damaging(0.703)	8/15	0.228623408	2134.6
10003	MSH2	chr7_6035181_G/T	chr7:6035181	T	missense_variant	993	887	296	P/H	cCt/cAt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.996)	8/15	0.164823009	3657.6
10058	PMS2	chr7_6035181_G/T	chr7:6035181	T	missense_variant	993	887	296	P/H	cCt/cAt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.996)	8/15	0.117333333	1975.6
10059	PMS2	chr7_6035181_G/T	chr7:6035181	T	missense_variant	993	887	296	P/H	cCt/cAt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.996)	8/15	0.155733029	3418.6
10081	PMS2	chr7_6035181_G/A	chr7:6035181	A	missense_variant	993	887	296	P/L	cCt/cTt	MODERATE	protein_coding	deleterious(0.01)	possibly_damaging(0.703)	8/15	0.133008921	948.6
11003	MSH2	chr7_6035181_G/A	chr7:6035181	A	missense_variant	993	887	296	P/L	cCt/cTt	MODERATE	protein_coding	deleterious(0.01)	possibly_damaging(0.703)	8/15	0.148464827	282.6
11015	PMS2	chr7_6035181_G/T	chr7:6035181	T	missense_variant	993	887	296	P/H	cCt/cAt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.996)	8/15	0.168886199	3578.6
12002	PMS2	chr7_6035181_G/A	chr7:6035181	A	missense_variant	993	887	296	P/L	cCt/cTt	MODERATE	protein_coding	deleterious(0.01)	possibly_damaging(0.703)	8/15	0.266167461	3407.6
12051	PMS2	chr7_6035181_G/C	chr7:6035181	C	missense_variant	993	887	296	P/R	cCt/cGt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.996)	8/15	0.139205842	659.6
13001	MSH2	chr7_6035181_G/A	chr7:6035181	A	missense_variant	993	887	296	P/L	cCt/cTt	MODERATE	protein_coding	deleterious(0.01)	possibly_damaging(0.703)	8/15	0.199771689	1680.6
13079	PMS2	chr7_6035181_G/C	chr7:6035181	C	missense_variant	993	887	296	P/R	cCt/cGt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.996)	8/15	0.165517241	2207.6

Sample	Defective MMR Gene	Uploaded Variation	Location	Allele	Consequence	cDNA Position	CDS Position	Protein Position	Amino Acids	Codons	Impact	BioType	SIFT	PolyPhen	EXON	VAF	QUAL
14003	PMS2	chr7_6035181_G/C	chr7:6035181	C	missense_variant	993	887	296	P/R	cCt/cGt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.996)	8/15	0.139365918	839.6
15002	MSH2	chr7_6035181_G/C	chr7:6035181	C	missense_variant	993	887	296	P/R	cCt/cGt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.996)	8/15	0.160994764	1493.6
16004	MSH2	chr7_6035181_G/C	chr7:6035181	C	missense_variant	993	887	296	P/R	cCt/cGt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.996)	8/15	0.167553191	2101.6
16007	MSH2	chr7_6035181_G/C	chr7:6035181	C	missense_variant	993	887	296	P/R	cCt/cGt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.996)	8/15	0.15302267	1562.6
16013	PMS2	chr7_6035181_G/C	chr7:6035181	C	missense_variant	993	887	296	P/R	cCt/cGt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.996)	8/15	0.17672682	2666.6
16021	MLH1	chr7_6035181_G/A	chr7:6035181	A	missense_variant	993	887	296	P/L	cCt/cTt	MODERATE	protein_coding	deleterious(0.01)	possibly_damaging(0.703)	8/15	0.191834452	2121.6
16039	PMS2	chr7_6035181_G/T	chr7:6035181	T	missense_variant	993	887	296	P/H	cCt/cAt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.996)	8/15	0.165693043	4149.6
16065	PMS2	chr7_6035181_G/C	chr7:6035181	C	missense_variant	993	887	296	P/R	cCt/cGt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.996)	8/15	0.164600551	2070.6
17005	PMS2	chr7_6035181_G/T	chr7:6035181	T	missense_variant	993	887	296	P/H	cCt/cAt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.996)	8/15	0.163333333	3638.6
18002	MSH6	chr7_6035181_G/-	chr7:6035181	-	frameshift_variant	993	887	296	P/X	cCt/ct	HIGH	protein_coding	-	-	8/15	0.154180239	329.6
18013	MLH1	chr7_6035181_G/-	chr7:6035181	-	frameshift_variant	993	887	296	P/X	cCt/ct	HIGH	protein_coding	-	-	8/15	0.13613685	272.6
18013	MLH1	chr7_6035181_G/C	chr7:6035181	C	missense_variant	993	887	296	P/R	cCt/cGt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.996)	8/15	0.13613685	272.6
25117	MLH1	chr7_6035181_G/T	chr7:6035181	T	missense_variant	993	887	296	P/H	cCt/cAt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.996)	8/15	0.156156156	1904.6
05002	MLH1	rs375553553	chr7:6035182	A	missense_variant	992	886	296	P/S	Cct/Tct	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.947)	8/15	0.126918536	344.6
10006	MSH2	rs375553553	chr7:6035182	A	missense_variant	992	886	296	P/S	Cct/Tct	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.947)	8/15	0.125631313	410.6
10011	MLH1	rs375553553	chr7:6035182	A	missense_variant	992	886	296	P/S	Cct/Tct	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.947)	8/15	0.137757948	315.6
10051	PMS2	rs375553553	chr7:6035182	A	missense_variant	992	886	296	P/S	Cct/Tct	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.947)	8/15	0.134146341	342.6
10080	PMS2	rs375553553	chr7:6035182	A	missense_variant	992	886	296	P/S	Cct/Tct	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.947)	8/15	0.137148047	467.6
15003	MSH2	rs375553553	chr7:6035182	A	missense_variant	992	886	296	P/S	Cct/Tct	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.947)	8/15	0.135746606	166.6
12002	PMS2	rs267608149	chr7:6036957-6036958	A	frameshift_variant,splice_region_variant	908-909	802-803	268	Y/LX	tac/tTac	HIGH	protein_coding	-	-	7/15	0.181818182	163.64
17116	PMS2	chr7_6038772_C/A	chr7:6038772	A	missense_variant	778	672	224	K/N	aaG/aaT	MODERATE	protein_coding	deleterious(0.01)	possibly_damaging(0.45)	6/15	0.198026316	2474.6
02007	MSH2	chr7_6038785_C/A	chr7:6038785	A	missense_variant	765	659	220	S/I	aGc/aTc	MODERATE	protein_coding	deleterious(0.01)	possibly_damaging(0.735)	6/15	0.173562059	2740.6
03002	MSH2	chr7_6038785_C/A	chr7:6038785	A	missense_variant	765	659	220	S/I	aGc/aTc	MODERATE	protein_coding	deleterious(0.01)	possibly_damaging(0.735)	6/15	0.144578313	1955.6
07002	MSH2	chr7_6038785_C/A	chr7:6038785	A	missense_variant	765	659	220	S/I	aGc/aTc	MODERATE	protein_coding	deleterious(0.01)	possibly_damaging(0.735)	6/15	0.113684211	1460.6
07003	MSH2	chr7_6038785_C/A	chr7:6038785	A	missense_variant	765	659	220	S/I	aGc/aTc	MODERATE	protein_coding	deleterious(0.01)	possibly_damaging(0.735)	6/15	0.150684932	1733.6
11001	PMS2	chr7_6038785_C/A	chr7:6038785	A	missense_variant	765	659	220	S/I	aGc/aTc	MODERATE	protein_coding	deleterious(0.01)	possibly_damaging(0.735)	6/15	0.14870181	2215.6
13079	PMS2	chr7_6038785_C/A	chr7:6038785	A	missense_variant	765	659	220	S/I	aGc/aTc	MODERATE	protein_coding	deleterious(0.01)	possibly_damaging(0.735)	6/15	0.184897025	202.6
13090	PMS2	chr7_6038785_C/A	chr7:6038785	A	missense_variant	765	659	220	S/I	aGc/aTc	MODERATE	protein_coding	deleterious(0.01)	possibly_damaging(0.735)	6/15	0.147673635	2859.6
14024	PMS2	chr7_6038785_C/A	chr7:6038785	A	missense_variant	765	659	220	S/I	aGc/aTc	MODERATE	protein_coding	deleterious(0.01)	possibly_damaging(0.735)	6/15	0.101488095	1015.6
17002	MLH1	chr7_6038785_C/A	chr7:6038785	A	missense_variant	765	659	220	S/I	aGc/aTc	MODERATE	protein_coding	deleterious(0.01)	possibly_damaging(0.735)	6/15	0.193151888	3714.6
18005	MLH1	chr7_6038785_C/A	chr7:6038785	A	missense_variant	765	659	220	S/I	aGc/aTc	MODERATE	protein_coding	deleterious(0.01)	possibly_damaging(0.735)	6/15	0.158634538	240.6
21003	MSH2	chr7_6038785_C/A	chr7:6038785	A	missense_variant	765	659	220	S/I	aGc/aTc	MODERATE	protein_coding	deleterious(0.01)	possibly_damaging(0.735)	6/15	0.134974533	1206.6
23004	MSH2	chr7_6038785_C/A	chr7:6038785	A	missense_variant	765	659	220	S/I	aGc/aTc	MODERATE	protein_coding	deleterious(0.01)	possibly_damaging(0.735)	6/15	0.16579537	938.6
01135	PMS2	chr7_6038786_T/A	chr7:6038786	A	missense_variant	764	658	220	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.854)	6/15	0.106124002	333.6
02001	PMS2	chr7_6038786_T/A	chr7:6038786	A	missense_variant	764	658	220	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.854)	6/15	0.128219805	132.6
02010	MSH2	chr7_6038786_T/A	chr7:6038786	A	missense_variant	764	658	220	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.854)	6/15	0.209472982	3591.6
04002	PMS2	chr7_6038786_T/A	chr7:6038786	A	missense_variant	764	658	220	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.854)	6/15	0.159679183	3585.6
04006	PMS2	chr7_6038786_T/G	chr7:6038786	G	missense_variant	764	658	220	S/R	Agc/Cgc	MODERATE	protein_coding	deleterious(0.04)	possibly_damaging(0.723)	6/15	0.1307393	162.6
06030	PMS2	chr7_6038786_T/A	chr7:6038786	A	missense_variant	764	658	220	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.854)	6/15	0.138433515	1577.6
07002	MSH2	chr7_6038786_T/G	chr7:6038786	G	missense_variant	764	658	220	S/R	Agc/Cgc	MODERATE	protein_coding	deleterious(0.04)	possibly_damaging(0.723)	6/15	0.099794942	631.6
10011	MLH1	chr7_6038786_T/A	chr7:6038786	A	missense_variant	764	658	220	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.854)	6/15	0.150844278	3504.6
10057	PMS2	chr7_6038786_T/A	chr7:6038786	A	missense_variant	764	658	220	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.854)	6/15	0.203032301	1800.6
10058	PMS2	chr7_6038786_T/A	chr7:6038786	A	missense_variant	764	658	220	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.854)	6/15	0.118834081	796.6
10059	PMS2	chr7_6038786_T/A	chr7:6038786	A	missense_variant	764	658	220	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.854)	6/15	0.148737138	3057.6
11004	MSH2	chr7_6038786_T/A	chr7:6038786	A	missense_variant	764	658	220	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.854)	6/15	0.219541616	750.6
14006	MSH2	chr7_6038786_T/A	chr7:6038786	A	missense_variant	764	658	220	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.854)	6/15	0.205717837	3536.6
14007	MLH1	chr7_6038786_T/A	chr7:6038786	A	missense_variant	764	658	220	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.854)	6/15	0.246666667	5274.6
14014	PMS2	chr7_6038786_T/A	chr7:6038786	A	missense_variant	764	658	220	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.854)	6/15	0.165680473	3145.6
15031	MLH1	chr7_6038786_T/A	chr7:6038786	A	missense_variant	764	658	220	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.854)	6/15	0.154922001	1322.6
16021	MLH1	chr7_6038786_T/A	chr7:6038786	A	missense_variant	764	658	220	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.854)	6/15	0.217232376	132.6
17005	PMS2	chr7_6038786_T/A	chr7:6038786	A	missense_variant	764	658	220	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.854)	6/15	0.168670886	4627.6
17116	PMS2	chr7_6038786_T/A	chr7:6038786	A	missense_variant	764	658	220	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.854)	6/15	0.238369305	7719.6
18002	MSH6	chr7_6038786_T/A	chr7:6038786	A	missense_variant	764	658	220	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.854)	6/15	0.142307692	1611.6
21002	MLH1	chr7_6038786_T/A	chr7:6038786	A	missense_variant	764	658	220	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.854)	6/15	0.255487805	7070.6
25117	MLH1	chr7_6038786_T/A	chr7:6038786	A	missense_variant	764	658	220	S/C	Agc/Tgc	MODERATE	protein_coding	deleterious(0.03)	possibly_damaging(0.854)	6/15	0.271213323	1649.6
13082	PMS2	chr7_6043605_A/G	chr7:6043605	G	missense_variant,splice_region_variant	354	248	83	L/S	tTa/tCa	MODERATE	protein_coding	deleterious(0)	probably_damaging(1)	3/15	0.123267688	51.6
16012	PMS2	chr7_6043605_A/G	chr7:6043605	G	missense_variant,splice_region_variant	354	248	83	L/S	tTa/tCa	MODERATE	protein_coding	deleterious(0)	probably_damaging(1)	3/15	0.144046628	382.6
03001	MSH2	chr7_6043606_A/T	chr7:6043606	T	missense_variant	353	247	83	L/I	Tta/Ata	MODERATE	protein_coding	deleterious(0.02)	probably_damaging(0.989)	3/15	0.153443766	1151.6
04112	PMS2	chr7_6043606_A/T	chr7:6043606	T	missense_variant	353	247	83	L/I	Tta/Ata	MODERATE	protein_coding	deleterious(0.02)	probably_damaging(0.989)	3/15	0.130742049	169.6

Sample	Defective MMR Gene	Uploaded Variation	Location	Allele	Consequence	cDNA Position	CDS Position	Protein Position	Amino Acids	Codons	Impact	BioType	SIFT	PolyPhen	EXON	VAF	QUAL
07002	MSH2	chr7_6043606_A/T	chr7:6043606	T	missense_variant	353	247	83	L/I	Tta/Ata	MODERATE	protein_coding	deleterious(0.02)	probably_damaging(0.989)	3/15	0.129349967	612.6
11026	MLH1	chr7_6043606_A/T	chr7:6043606	T	missense_variant	353	247	83	L/I	Tta/Ata	MODERATE	protein_coding	deleterious(0.02)	probably_damaging(0.989)	3/15	0.153179191	226.6
13090	PMS2	chr7_6043606_A/T	chr7:6043606	T	missense_variant	353	247	83	L/I	Tta/Ata	MODERATE	protein_coding	deleterious(0.02)	probably_damaging(0.989)	3/15	0.126847291	1049.6
15031	MLH1	chr7_6043606_A/T	chr7:6043606	T	missense_variant	353	247	83	L/I	Tta/Ata	MODERATE	protein_coding	deleterious(0.02)	probably_damaging(0.989)	3/15	0.130935252	599.6
04093	PMS2	chr7_6045582_A/G	chr7:6045582	G	missense_variant	210	104	35	L/P	cTa/cCa	MODERATE	protein_coding	deleterious(0)	probably_damaging(1)	2/15	0.149300156	171.6
12061	PMS2	chr7_6045582_A/G	chr7:6045582	G	missense_variant	210	104	35	L/P	cTa/cCa	MODERATE	protein_coding	deleterious(0)	probably_damaging(1)	2/15	0.153687112	228.6
16056	PMS2	rs63750123	chr7:6045634	C	missense_variant	158	52	18	I/V	Att/Gtt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.994)	2/15	0.22817354	7232.6
18001	MSH2	rs63750123	chr7:6045634	C	missense_variant	158	52	18	I/V	Att/Gtt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.994)	2/15	0.413114754	5631.6
18002	MSH6	rs63750123	chr7:6045634	C	missense_variant	158	52	18	I/V	Att/Gtt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.994)	2/15	0.347343841	15979.6
18013	MLH1	rs63750123	chr7:6045634	C	missense_variant	158	52	18	I/V	Att/Gtt	MODERATE	protein_coding	deleterious(0)	probably_damaging(0.994)	2/15	0.341949381	13919.6

Table 7.6. Details of the variants identified by the MIP-based *PMS2* assay across all samples analysed. Variants are ordered initially by their predicted impact, followed by their cDNA position within the *PMS2* gene. Where an identified variant has previously been reported, the corresponding rs value is detailed under 'Uploaded Variation', with the position of novel variants alternatively related therein.

Chapter 8. References

Aarnio, M. *et al.* (1997) 'Features of gastric cancer in hereditary non-polyposis colorectal cancer syndrome', *International Journal of Cancer*, 74(5), pp. 551–555. Available at: https://doi.org/10.1002/(SICI)1097-0215(19971021)74:5<551::AID-IJC13>3.0.CO;2-9.

Acharya, S. *et al.* (1996) 'hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6.', *Proceedings of the National Academy of Sciences of the United States of America*, 93(24), pp. 13629–34. Available at: https://doi.org/10.1073/pnas.93.24.13629.

Ahadova, A. *et al.* (2018) 'Three molecular pathways model colorectal carcinogenesis in Lynch syndrome', *International Journal of Cancer*, 143(1), pp. 139–150. Available at: https://doi.org/10.1002/ijc.31300.

Akbari, M.R. *et al.* (2017) 'Correlation between germline mutations in MMR genes and microsatellite instability in ovarian cancer specimens', *Familial Cancer*, 16(3), pp. 351–355. Available at: https://doi.org/10.1007/s10689-017-9973-1.

Algra, A.M. and Rothwell, P.M. (2012) 'Effects of regular aspirin on long-term cancer incidence and metastasis: a systematic comparison of evidence from observational studies versus randomised trials', *The Lancet Oncology*, 13(5), pp. 518–527. Available at: https://doi.org/10.1016/S1470-2045(12)70112-2.

Álvaro, E. *et al.* (2019) 'Clinical and Molecular Comparative Study of Colorectal Cancer Based on Age-of-onset and Tumor Location: Two Main Criteria for Subclassifying Colorectal Cancer', *International Journal of Molecular Sciences*, 20(4), p. 968. Available at: https://doi.org/10.3390/ijms20040968.

Amitay, E.L. *et al.* (2019) 'Association of Aspirin and Nonsteroidal Anti-Inflammatory Drugs With Colorectal Cancer Risk by Molecular Subtypes', *JNCI: Journal of the National Cancer Institute*, 111(5), pp. 475–483. Available at: https://doi.org/10.1093/jnci/djy170.

Antonarakis, E.S. *et al.* (2019) 'Clinical Features and Therapeutic Outcomes in Men with Advanced Prostate Cancer and DNA Mismatch Repair Gene Mutations', *European Urology*, 75(3), pp. 378–382. Available at: https://doi.org/10.1016/j.eururo.2018.10.009.

Arnold, M. *et al.* (2017) 'Global patterns and trends in colorectal cancer incidence and mortality', *Gut*, 66(4), pp. 683–691. Available at: https://doi.org/10.1136/gutjnl-2015-310912.

Van der Auwera, G.A. *et al.* (2013) 'From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline', *Current Protocols in Bioinformatics*, 43(1). Available at: https://doi.org/10.1002/0471250953.bi1110s43.

Awtry, E.H. and Loscalzo, J. (2000) 'Aspirin', *Circulation*, 101(10), pp. 1206–1218. Available at: https://doi.org/10.1161/01.CIR.101.10.1206.

Bacher, J.W. *et al.* (2004) 'Development of a Fluorescent Multiplex Assay for Detection of MSI-High Tumors', *Disease Markers*, 20(4–5), pp. 237–250. Available at: https://doi.org/10.1155/2004/136734.

Bacolla, A. *et al.* (2008) 'Abundance and length of simple repeats in vertebrate genomes are determined by their structural properties', *Genome Research*, 18(10), pp. 1545–1553. Available at: https://doi.org/10.1101/gr.078303.108.

Balmaña, J. *et al.* (2013) 'Familial risk-colorectal cancer: ESMO Clinical Practice Guidelines', *Annals of Oncology*, 24, pp. vi73–vi80. Available at: https://doi.org/10.1093/annonc/mdt209.

Barlin, J.N., Puri, I. and Bristow, R.E. (2010) 'Cytoreductive surgery for advanced or recurrent endometrial cancer: A meta-analysis', *Gynecologic Oncology*, 118(1), pp. 14–18. Available at: https://doi.org/10.1016/j.ygyno.2010.04.005.

Baron, J.A. *et al.* (2006) 'A Randomized Trial of Rofecoxib for the Chemoprevention of Colorectal Adenomas', *Gastroenterology*, 131(6), pp. 1674–1682. Available at: https://doi.org/10.1053/j.gastro.2006.08.079.

Bastiaannet, E. *et al.* (2012) 'Use of Aspirin postdiagnosis improves survival for colon cancer patients', *British Journal of Cancer*, 106(9), pp. 1564–1570. Available at: https://doi.org/10.1038/bjc.2012.101.

Bertagnolli, M.M. *et al.* (2006) 'Celecoxib for the Prevention of Sporadic Colorectal Adenomas', *New England Journal of Medicine*, 355(9), pp. 873–884. Available at: https://doi.org/10.1056/NEJMoa061355.

Biezuner, T. *et al.* (2022) 'An improved molecular inversion probe based targeted sequencing approach for low variant allele frequency', *NAR Genomics and Bioinformatics*, 4(1). Available at: https://doi.org/10.1093/nargab/lqab125.

BMJ (1994) 'Collaborative overview of randomised trials of antiplatelet therapy Prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients', *BMJ*, 308(6921), pp. 81–106. Available at: https://doi.org/10.1136/bmj.308.6921.81.

BMJ (2002) 'Collaborative meta-analysis of randomised trials of antiplatelet therapy for prevention of death, myocardial infarction, and stroke in high risk patients', *BMJ*, 324(7329), pp. 71–86. Available at: https://doi.org/10.1136/bmj.324.7329.71.

Bokhman, J. v. (1983) 'Two pathogenetic types of endometrial carcinoma', *Gynecologic Oncology*, 15(1), pp. 10–17. Available at: https://doi.org/10.1016/0090-8258(83)90111-7.

Boland, C.R. *et al.* (1998) 'A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer.', *Cancer research*, 58(22), pp. 5248–57.

Bombard, Y. *et al.* (2017) ‘Universal tumor screening for Lynch syndrome: health-care providers’ perspectives’, *Genetics in Medicine*, 19(5), pp. 568–574. Available at: https://doi.org/10.1038/gim.2016.150.

Bonneville, R. *et al.* (2017) ‘Landscape of Microsatellite Instability Across 39 Cancer Types’, *JCO Precision Oncology*, (1), pp. 1–15. Available at: https://doi.org/10.1200/PO.17.00073.

Bos, C.L. *et al.* (2006) ‘Effect of aspirin on the Wnt/β-catenin pathway is mediated via protein phosphatase 2A’, *Oncogene*, 25(49), pp. 6447–6456. Available at: https://doi.org/10.1038/sj.onc.1209658.

Bosetti, C. *et al.* (2012) ‘Aspirin and cancer risk: a quantitative review to 2011’, *Annals of Oncology*, 23(6), pp. 1403–1415. Available at: https://doi.org/10.1093/annonc/mds113.

Boyle, E.A. *et al.* (2014) ‘MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing’, *Bioinformatics*, 30(18), pp. 2670–2672. Available at: https://doi.org/10.1093/bioinformatics/btu353.

Boyle, T.A. *et al.* (2014) ‘Summary of Microsatellite Instability Test Results From Laboratories Participating in Proficiency Surveys: Proficiency Survey Results From 2005 to 2012’, *Archives of Pathology & Laboratory Medicine*, 138(3), pp. 363–370. Available at: https://doi.org/10.5858/arpa.2013-0159-CP.

Braun, M.M., Overbeek-Wager, E.A. and Grumbo, R.J. (2016) ‘Diagnosis and Management of Endometrial Cancer.’, *American family physician*, 93(6), pp. 468–74.

Brennan, B. *et al.* (2017) ‘Universal molecular screening does not effectively detect Lynch syndrome in clinical practice’, *Therapeutic Advances in Gastroenterology*, 10(4), pp. 361–371. Available at: https://doi.org/10.1177/1756283X17690990.

Bristow, R.E. *et al.* (2006) ‘Salvage cytoreductive surgery for recurrent endometrial cancer’, *Gynecologic Oncology*, 103(1), pp. 281–287. Available at: https://doi.org/10.1016/j.ygyno.2006.03.011.

Broaddus, R.R. *et al.* (2006) ‘Pathologic features of endometrial carcinoma associated with HNPCC’, *Cancer*, 106(1), pp. 87–94. Available at: https://doi.org/10.1002/cncr.21560.

ten Broeke, S.W. *et al.* (2018) ‘Cancer Risks for *PMS2* -Associated Lynch Syndrome’, *Journal of Clinical Oncology*, 36(29), pp. 2961–2968. Available at: https://doi.org/10.1200/JCO.2018.78.4777.

Bronner, C.E. *et al.* (1994) ‘Mutation in the DNA mismatch repair gene homologue hMLH 1 is associated with hereditary non-polyposis colon cancer’, *Nature*, 368(6468), pp. 258–261. Available at: https://doi.org/10.1038/368258a0.

Burn, J. *et al.* (2008) ‘Effect of Aspirin or Resistant Starch on Colorectal Neoplasia in the Lynch Syndrome’, *New England Journal of Medicine*, 359(24), pp. 2567–2578. Available at: https://doi.org/10.1056/NEJMoa0801297.

Burn, J., Bishop, D.T., *et al.* (2011) 'A Randomized Placebo-Controlled Prevention Trial of Aspirin and/or Resistant Starch in Young People with Familial Adenomatous Polyposis', *Cancer Prevention Research*, 4(5), pp. 655–665. Available at: https://doi.org/10.1158/1940-6207.CAPR-11-0106.

Burn, J., Gerdes, A.-M., *et al.* (2011) 'Long-term effect of aspirin on cancer risk in carriers of hereditary colorectal cancer: an analysis from the CAPP2 randomised controlled trial', *The Lancet*, 378(9809), pp. 2081–2087. Available at: https://doi.org/10.1016/S0140-6736(11)61049-0.

Burn, J. *et al.* (2020) 'Cancer prevention with aspirin in hereditary colorectal cancer (Lynch syndrome), 10-year follow-up and registry-based 20-year data in the CAPP2 study: a double-blind, randomised, placebo-controlled trial', *The Lancet*, 395(10240), pp. 1855–1863. Available at: https://doi.org/10.1016/S0140-6736(20)30366-4.

Campagnutta, E. *et al.* (2004) 'Surgical treatment of recurrent endometrial carcinoma', *Cancer*, 100(1), pp. 89–96. Available at: https://doi.org/10.1002/cncr.11868.

Capone, M.L. *et al.* (2007) 'Pharmacodynamic of cyclooxygenase inhibitors in humans', *Prostaglandins & Other Lipid Mediators*, 82(1–4), pp. 85–94. Available at: https://doi.org/10.1016/j.prostaglandins.2006.05.019.

Carethers, J.M. (2014) 'Differentiating Lynch-Like From Lynch Syndrome', *Gastroenterology*, 146(3), pp. 602–604. Available at: https://doi.org/10.1053/j.gastro.2014.01.041.

Chan, A.T. (2009) 'Aspirin Use and Survival After Diagnosis of Colorectal Cancer', *JAMA*, 302(6), p. 649. Available at: https://doi.org/10.1001/jama.2009.1112.

Chan, T.A. *et al.* (1998) 'Mechanisms underlying nonsteroidal antiinflammatory drug-mediated apoptosis', *Proceedings of the National Academy of Sciences*, 95(2), pp. 681–686. Available at: https://doi.org/10.1073/pnas.95.2.681.

Chao, X. *et al.* (2019) 'Comparison of screening strategies for Lynch syndrome in patients with newly diagnosed endometrial cancer: a prospective cohort study in China', *Cancer Communications*, 39(1), p. 42. Available at: https://doi.org/10.1186/s40880-019-0388-2.

Chen, P.-C. *et al.* (2005) 'Contributions by MutL Homologues *Mlh3* and *Pms2* to DNA Mismatch Repair and Tumor Suppression in the Mouse', *Cancer Research*, 65(19), pp. 8662–8670. Available at: https://doi.org/10.1158/0008-5472.CAN-05-0742.

Chinese Acute Stroke Trial Collaborative Group (1997) 'CAST: randomised placebo-controlled trial of early aspirin use in 20,000 patients with acute ischaemic stroke. CAST (Chinese Acute Stroke Trial) Collaborative Group.', *Lancet (London, England)*, 349(9066), pp. 1641–9.

Christakis, A.G. *et al.* (2019) 'Targeted Cancer Next-Generation Sequencing as a Primary Screening Tool for Microsatellite Instability and Lynch Syndrome in Upper

Gastrointestinal Tract Cancers', *Cancer Epidemiology, Biomarkers & Prevention*, 28(7), pp. 1246–1251. Available at: https://doi.org/10.1158/1055-9965.EPI-18-1250.

Chung, H. *et al.* (2010) 'Both microsatellite length and sequence context determine frameshift mutation rates in defective DNA mismatch repair', *Human Molecular Genetics*, 19(13), pp. 2638–2647. Available at: https://doi.org/10.1093/hmg/ddq151.

Ciardiello, D. *et al.* (2019) 'Immunotherapy of colorectal cancer: Challenges for therapeutic efficacy', *Cancer Treatment Reviews*, 76, pp. 22–32. Available at: https://doi.org/10.1016/j.ctrv.2019.04.003.

Clarke, M.A. *et al.* (2018) 'Association of Endometrial Cancer Risk With Postmenopausal Bleeding in Women', *JAMA Internal Medicine*, 178(9), p. 1210. Available at: https://doi.org/10.1001/jamainternmed.2018.2820.

Clendenning, M. *et al.* (2006) 'Long-range PCR facilitates the identification ofPMS2-specific mutations', *Human Mutation*, 27(5), pp. 490–495. Available at: https://doi.org/10.1002/humu.20318.

Cohen, S.A., Pritchard, C.C. and Jarvik, G.P. (2019) 'Lynch Syndrome: From Screening to Diagnosis to Treatment in the Era of Modern Molecular Oncology', *Annual Review of Genomics and Human Genetics*, 20(1), pp. 293–307. Available at: https://doi.org/10.1146/annurev-genom-083118-015406.

Connor, A.A. *et al.* (2017) 'Association of Distinct Mutational Signatures With Correlates of Increased Immune Activity in Pancreatic Ductal Adenocarcinoma', *JAMA Oncology*, 3(6), p. 774. Available at: https://doi.org/10.1001/jamaoncol.2016.3916.

Cook, Nancy R. *et al.* (2013) 'Alternate-Day, Low-Dose Aspirin and Cancer Risk: Long-Term Observational Follow-up of a Randomized Trial', *Annals of Internal Medicine*, 159(2), p. 77. Available at: https://doi.org/10.7326/0003-4819-159-2-201307160-00002.

Cook, Nancy R *et al.* (2013) 'Alternate-day, low-dose aspirin and cancer risk: long-term observational follow-up of a randomized trial.', *Annals of internal medicine*, 159(2), pp. 77–85. Available at: https://doi.org/10.7326/0003-4819-159-2-201307160-00002.

Cortes-Ciriano, I. *et al.* (2017) 'A molecular portrait of microsatellite instability across multiple cancers', *Nature Communications*, 8(1), p. 15180. Available at: https://doi.org/10.1038/ncomms15180.

Cote, R. (1995) 'Lack of Effect of Aspirin in Asymptomatic Patients with Carotid Bruits and Substantial Carotid Narrowing', *Annals of Internal Medicine*, 123(9), p. 649. Available at: https://doi.org/10.7326/0003-4819-123-9-199511010-00002.

Coxib and Traditional NSAID Trialists' Collaboration (2013) 'Vascular and upper gastrointestinal effects of non-steroidal anti-inflammatory drugs: meta-analyses of individual participant data from randomised trials', *The Lancet*, 382(9894), pp. 769–779. Available at: https://doi.org/10.1016/S0140-6736(13)60900-9.

Crosbie, E.J. *et al.* (2019) 'The Manchester International Consensus Group recommendations for the management of gynecological cancers in Lynch syndrome', *Genetics in Medicine*, 21(10), pp. 2390–2400. Available at: https://doi.org/10.1038/s41436-019-0489-y.

Crosbie, E.J. *et al.* (2022) 'Endometrial cancer', *The Lancet*, 399(10333), pp. 1412–1428. Available at: https://doi.org/10.1016/S0140-6736(22)00323-3.

Dandapani, M. *et al.* (2011) 'Sebaceous neoplasms in Lynch syndrome', *Hereditary Cancer in Clinical Practice*, 9(Suppl 1), p. P7. Available at: https://doi.org/10.1186/1897-4287-9-S1-P7.

Davies, H. *et al.* (2017) 'Whole-Genome Sequencing Reveals Breast Cancers with Mismatch Repair Deficiency', *Cancer Research*, 77(18), pp. 4755–4762. Available at: https://doi.org/10.1158/0008-5472.CAN-17-1083.

Dedeurwaerdere, F. *et al.* (2021) 'Comparison of microsatellite instability detection by immunohistochemistry and molecular techniques in colorectal and endometrial cancer', *Scientific Reports* 11, Available at: https://doi.org/10.1038/s41598-021-91974-x

Dieringer, D. and Schlötterer, C. (2003) '<scp>microsatellite analyser</scp> (MSA): a platform independent analysis tool for large microsatellite data sets', *Molecular Ecology Notes*, 3(1), pp. 167–169. Available at: https://doi.org/10.1046/j.1471-8286.2003.00351.x.

Dikshit, P. *et al.* (2006) 'Aspirin Induces Apoptosis through the Inhibition of Proteasome Function', *Journal of Biological Chemistry*, 281(39), pp. 29228–29235. Available at: https://doi.org/10.1074/jbc.M602629200.

Dixon, D.A. *et al.* (2001) 'Altered expression of the mRNA stability factor HuR promotes cyclooxygenase-2 expression in colon cancer cells', *Journal of Clinical Investigation*, 108(11), pp. 1657–1665. Available at: https://doi.org/10.1172/JCI12973.

Dixon, D.A. *et al.* (2013) 'Mechanistic Aspects of COX-2 Expression in Colorectal Neoplasia', in, pp. 7–37. Available at: https://doi.org/10.1007/978-3-642-30331-9_2.

Domingo, E. *et al.* (2013) 'Evaluation of *PIK3CA* Mutation As a Predictor of Benefit From Nonsteroidal Anti-Inflammatory Drug Therapy in Colorectal Cancer', *Journal of Clinical Oncology*, 31(34), pp. 4297–4305. Available at: https://doi.org/10.1200/JCO.2013.50.0322.

Dominguez-Valentin, M. *et al.* (2016) 'Frequent mismatch-repair defects link prostate cancer to Lynch syndrome', *BMC Urology*, 16(1), p. 15. Available at: https://doi.org/10.1186/s12894-016-0130-1.

Dominguez-Valentin, M. *et al.* (2020) 'Cancer risks by gene, age, and gender in 6350 carriers of pathogenic mismatch repair variants: findings from the Prospective Lynch Syndrome Database', *Genetics in Medicine*, 22(1), pp. 15–25. Available at: https://doi.org/10.1038/s41436-019-0596-9.

Douglas, J.A. *et al.* (2005) ‘History and Molecular Genetics of Lynch Syndrome in Family G’, *JAMA*, 294(17), p. 2195. Available at: https://doi.org/10.1001/jama.294.17.2195.

Dovizio, M. *et al.* (2013) ‘Pharmacological Inhibition of Platelet-Tumor Cell Cross-Talk Prevents Platelet-Induced Overexpression of Cyclooxygenase-2 in HT29 Human Colon Carcinoma Cells’, *Molecular Pharmacology*, 84(1), pp. 25–40. Available at: https://doi.org/10.1124/mol.113.084988.

Downer, M.K. *et al.* (2017) ‘Regular Aspirin Use and the Risk of Lethal Prostate Cancer in the Physicians’ Health Study’, *European Urology*, 72(5), pp. 821–827. Available at: https://doi.org/10.1016/j.eururo.2017.01.044.

Drew, D.A., Cao, Y. and Chan, A.T. (2016) ‘Aspirin and colorectal cancer: the promise of precision chemoprevention’, *Nature Reviews Cancer*, 16(3), pp. 173–186. Available at: https://doi.org/10.1038/nrc.2016.4.

Duval, A. and Hamelin, R. (2002) ‘Mutations at coding repeat sequences in mismatch repair-deficient human cancers: toward a new concept of target genes for instability.’, *Cancer research*, 62(9), pp. 2447–54.

Duzkale, H. *et al.* (2013) ‘A systematic approach to assessing the clinical significance of genetic variants’, *Clinical Genetics*, 84(5), pp. 453–463. Available at: https://doi.org/10.1111/cge.12257.

Ellegren, H. (2004) ‘Microsatellites: simple sequences with complex evolution’, *Nature Reviews Genetics*, 5(6), pp. 435–445. Available at: https://doi.org/10.1038/nrg1348.

El-Shakankery, K.H. *et al.* (2023) ‘Ensuring no patient is lost along the way – a single-centre experience of the clinical genetics referral pathways for Lynch syndrome’, *Colorectal Disease*, 25(5), pp. 838–842. Available at: https://doi.org/10.1111/codi.16496.

Etzler, J. *et al.* (2008) ‘RNA-based mutation analysis identifies an unusual MSH6 splicing defect and circumvents PMS2 pseudogene interference’, *Human Mutation*, 29(2), pp. 299–305. Available at: https://doi.org/10.1002/humu.20657.

Evaluation of Genomic Applications in Practice and Prevention Working Group (2009) ‘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’, *Genetics in Medicine*, 11(1), pp. 35–41. Available at: https://doi.org/10.1097/GIM.0b013e31818fa2ff.

Everett, J.N. *et al.* (2014) ‘Screening for Germline Mismatch Repair Mutations Following Diagnosis of Sebaceous Neoplasm’, *JAMA Dermatology*, 150(12), p. 1315. Available at: https://doi.org/10.1001/jamadermatol.2014.1217.

Fan, H. and Chu, J.-Y. (2007) ‘A Brief Review of Short Tandem Repeat Mutation’, *Genomics, Proteomics & Bioinformatics*, 5(1), pp. 7–14. Available at: https://doi.org/10.1016/S1672-0229(07)60009-6.

Feldman, S. *et al.* (1995) ‘Predicting Endometrial Cancer among Older Women Who Present with Abnormal Vaginal Bleeding’, *Gynecologic Oncology*, 56(3), pp. 376–381. Available at: https://doi.org/10.1006/gyno.1995.1066.

Ferrandez, A., Prescott, S. and Burt, R. (2003) ‘COX-2 and Colorectal Cancer’, *Current Pharmaceutical Design*, 9(27), pp. 2229–2251. Available at: https://doi.org/10.2174/1381612033454036.

Ferreira, A.M. *et al.* (2009) ‘Mononucleotide precedes dinucleotide repeat instability during colorectal tumour development in Lynch syndrome patients’, *The Journal of Pathology*, 219(1), pp. 96–102. Available at: https://doi.org/10.1002/path.2573.

Fink, S.P. *et al.* (2014) ‘Aspirin and the Risk of Colorectal Cancer in Relation to the Expression of 15-Hydroxyprostaglandin Dehydrogenase (*HPGD*)’, *Science Translational Medicine*, 6(233). Available at: https://doi.org/10.1126/scitranslmed.3008481.

Fishel, R. *et al.* (1993) ‘The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer’, *Cell*, 75(5), pp. 1027–1038. Available at: https://doi.org/10.1016/0092-8674(93)90546-3.

FitzGerald, G.A. and Patrono, C. (2001) ‘The Coxibs, Selective Inhibitors of Cyclooxygenase-2’, *New England Journal of Medicine*, 345(6), pp. 433–442. Available at: https://doi.org/10.1056/NEJM200108093450607.

Flossmann, E. and Rothwell, P.M. (2007) ‘Effect of aspirin on long-term risk of colorectal cancer: consistent evidence from randomised and observational studies’, *The Lancet*, 369(9573), pp. 1603–1613. Available at: https://doi.org/10.1016/S0140-6736(07)60747-8.

Fornasarig, M. *et al.* (2018) ‘Molecular and Pathological Features of Gastric Cancer in Lynch Syndrome and Familial Adenomatous Polyposis’, *International Journal of Molecular Sciences*, 19(6), p. 1682. Available at: https://doi.org/10.3390/ijms19061682.

Fraser, D.M. *et al.* (2014) ‘Aspirin use and survival after the diagnosis of breast cancer: a population-based cohort study’, *British Journal of Cancer*, 111(3), pp. 623–627. Available at: https://doi.org/10.1038/bjc.2014.264.

Friis, S. *et al.* (2015) ‘Low-Dose Aspirin or Nonsteroidal Anti-inflammatory Drug Use and Colorectal Cancer Risk’, *Annals of Internal Medicine*, 163(5), pp. 347–355. Available at: https://doi.org/10.7326/M15-0039.

Gallon, R. *et al.* (2019) ‘A sensitive and scalable microsatellite instability assay to diagnose constitutional mismatch repair deficiency by sequencing of peripheral blood leukocytes’, *Human Mutation*, 40(5), pp. 649–655. Available at: https://doi.org/10.1002/humu.23721.

Gallon, R. *et al.* (2020) ‘Sequencing-based microsatellite instability testing using as few as six markers for high-throughput clinical diagnostics’, *Human Mutation*, 41(1), pp. 332–341. Available at: https://doi.org/10.1002/humu.23906.

Gallon, R. *et al.* (2023) ‘Constitutional Microsatellite Instability, Genotype, and Phenotype Correlations in Constitutional Mismatch Repair Deficiency’, *Gastroenterology*, 164(4), pp. 579-592.e8. Available at: https://doi.org/10.1053/j.gastro.2022.12.017.

Garcia-Albeniz, X. and Chan, A.T. (2011) ‘Aspirin for the prevention of colorectal cancer’, *Best Practice & Research Clinical Gastroenterology*, 25(4–5), pp. 461–472. Available at: https://doi.org/10.1016/j.bpg.2011.10.015.

Gatius, S. *et al.* (2022) ‘Comparison of the Idylla™ MSI assay with the Promega™ MSI Analysis System and immunohistochemistry on formalin-fixed paraffin-embedded tissue of endometrial carcinoma: results from an international, multicenter study.’, *Virchows Archiv : an international journal of pathology*, 480(5), pp. 1031–1039. Available at: https://doi.org/10.1007/s00428-022-03291-x.

Gay, L.J. and Felding-Habermann, B. (2011) ‘Contribution of platelets to tumour metastasis’, *Nature Reviews Cancer*, 11(2), pp. 123–134. Available at: https://doi.org/10.1038/nrc3004.

Giardiello, F.M. *et al.* (2004) ‘Prostanoids, ornithine decarboxylase, and polyamines in primary chemoprevention of familial adenomatous polyposis’, *Gastroenterology*, 126(2), pp. 425–431. Available at: https://doi.org/10.1053/j.gastro.2003.11.013.

Gierse, J.K. *et al.* (1996) ‘A Single Amino Acid Difference between Cyclooxygenase-1 (COX-1) and −2 (COX-2) Reverses the Selectivity of COX-2 Specific Inhibitors’, *Journal of Biological Chemistry*, 271(26), pp. 15810–15814. Available at: https://doi.org/10.1074/jbc.271.26.15810.

Goel, A. *et al.* (2012) ‘Pharmacokinetic Solubility and Dissolution Profile of Antiarrythmic Drugs’, *International Journal of Pharma Professional’s Research (IJPPR)*, 3(1), pp. 508–517.

Goodfellow, P.J. *et al.* (2003) ‘Prevalence of defective DNA mismatch repair and MSH6 mutation in an unselected series of endometrial cancers’, *Proceedings of the National Academy of Sciences*, 100(10), pp. 5908–5913. Available at: https://doi.org/10.1073/pnas.1030231100.

Gould, G.M. *et al.* (2018) ‘Detecting clinically actionable variants in the 3′ exons of PMS2 via a reflex workflow based on equivalent hybrid capture of the gene and its pseudogene’, *BMC Medical Genetics*, 19(1), p. 176. Available at: https://doi.org/10.1186/s12881-018-0691-9.

Grant, R.C. *et al.* (2015) ‘Prevalence of Germline Mutations in Cancer Predisposition Genes in Patients With Pancreatic Cancer’, *Gastroenterology*, 148(3), pp. 556–564. Available at: https://doi.org/10.1053/j.gastro.2014.11.042.

Gredmark, T. *et al.* (1995) ‘Histopathological findings in women with postmenopausal bleeding’, *BJOG: An International Journal of Obstetrics and Gynaecology*, 102(2), pp. 133–136. Available at: https://doi.org/10.1111/j.1471-0528.1995.tb09066.x.

Green, A.K., Feinberg, J. and Makker, V. (2020) ‘A Review of Immune Checkpoint Blockade Therapy in Endometrial Cancer.’, *American Society of Clinical Oncology*

educational book. American Society of Clinical Oncology. Annual Meeting, 40, pp. 1–7. Available at: https://doi.org/10.1200/EDBK_280503.

Gu, B. *et al.* (2021) 'Variations in incidence and mortality rates of endometrial cancer at the global, regional, and national levels, 1990–2019', *Gynecologic Oncology*, 161(2), pp. 573–580. Available at: https://doi.org/10.1016/j.ygyno.2021.01.036.

Gylling, A.H.S. *et al.* (2008) 'Differential cancer predisposition in Lynch syndrome: insights from molecular analysis of brain and urinary tract tumors', *Carcinogenesis*, 29(7), pp. 1351–1359. Available at: https://doi.org/10.1093/carcin/bgn133.

Hamada, T. *et al.* (2017) 'Aspirin Use and Colorectal Cancer Survival According to Tumor CD274 (Programmed Cell Death 1 Ligand 1) Expression Status', *Journal of Clinical Oncology*, 35(16), pp. 1836–1844. Available at: https://doi.org/10.1200/JCO.2016.70.7547.

Hampel, H. *et al.* (2005) 'Screening for the Lynch Syndrome (Hereditary Nonpolyposis Colorectal Cancer)', *New England Journal of Medicine*, 352(18), pp. 1851–1860. Available at: https://doi.org/10.1056/NEJMoa043146.

Hampel, H. *et al.* (2008) 'Feasibility of Screening for Lynch Syndrome Among Patients With Colorectal Cancer', *Journal of Clinical Oncology*, 26(35), pp. 5783–5788. Available at: https://doi.org/10.1200/JCO.2008.17.5950.

Hampel, H. *et al.* (2018) 'Assessment of Tumor Sequencing as a Replacement for Lynch Syndrome Screening and Current Molecular Tests for Patients With Colorectal Cancer', *JAMA Oncology*, 4(6), p. 806. Available at: https://doi.org/10.1001/jamaoncol.2018.0104.

Hampel, H. *et al.* (2021) 'Double somatic mismatch repair gene pathogenic variants as common as Lynch syndrome among endometrial cancer patients', *Gynecologic Oncology*, 160(1), pp. 161–168. Available at: https://doi.org/10.1016/j.ygyno.2020.10.012.

Hampel, H. and de la Chapelle, A. (2011) 'The Search for Unaffected Individuals with Lynch Syndrome: Do the Ends Justify the Means?', *Cancer Prevention Research*, 4(1), pp. 1–5. Available at: https://doi.org/10.1158/1940-6207.CAPR-10-0345.

Hansson, L. *et al.* (1998) 'Effects of intensive blood-pressure lowering and low-dose aspirin in patients with hypertension: principal results of the Hypertension Optimal Treatment (HOT) randomised trial', *The Lancet*, 351(9118), pp. 1755–1762. Available at: https://doi.org/10.1016/S0140-6736(98)04311-6.

Harper, H.L. *et al.* (2017) 'Upper tract urothelial carcinomas: frequency of association with mismatch repair protein loss and lynch syndrome', *Modern Pathology*, 30(1), pp. 146–156. Available at: https://doi.org/10.1038/modpathol.2016.171.

Hawkey, C.J. (1990) 'Non-steroidal anti-inflammatory drugs and peptic ulcers.', *BMJ*, 300(6720), pp. 278–284. Available at: https://doi.org/10.1136/bmj.300.6720.278.

He, E.Y. *et al.* (2016) 'The Impact of Mismatch Repair Status in Colorectal Cancer on the Decision to Treat With Adjuvant Chemotherapy: An Australian Population-Based

Multicenter Study', *The Oncologist*, 21(5), pp. 618–625. Available at: https://doi.org/10.1634/theoncologist.2015-0530.

He, J. *et al.* (1998) 'Aspirin and Risk of Hemorrhagic Stroke', *JAMA*, 280(22), p. 1930. Available at: https://doi.org/10.1001/jama.280.22.1930.

Heald, B. *et al.* (2013) 'Implementation of Universal Microsatellite Instability and Immunohistochemistry Screening for Diagnosing Lynch Syndrome in a Large Academic Medical Center', *Journal of Clinical Oncology*, 31(10), pp. 1336–1340. Available at: https://doi.org/10.1200/JCO.2012.45.1674.

Hendriks, Y.M.C. *et al.* (2006) 'Diagnostic Approach and Management of Lynch Syndrome (Hereditary Nonpolyposis Colorectal Carcinoma): A Guide for Clinicians', *CA: A Cancer Journal for Clinicians*, 56(4), pp. 213–225. Available at: https://doi.org/10.3322/canjclin.56.4.213.

Henry, D. *et al.* (1996) 'Variability in risk of gastrointestinal complications with individual non-steroidal anti-inflammatory drugs: results of a collaborative meta-analysis', *BMJ*, 312(7046), pp. 1563–1566. Available at: https://doi.org/10.1136/bmj.312.7046.1563.

Henry, W.S. *et al.* (2017) 'Aspirin Suppresses Growth in PI3K-Mutant Breast Cancer by Activating AMPK and Inhibiting mTORC1 Signaling', *Cancer Research*, 77(3), pp. 790–801. Available at: https://doi.org/10.1158/0008-5472.CAN-16-2400.

Herman, D.S. *et al.* (2018) 'Efficient Detection of Copy Number Mutations in PMS2 Exons with a Close Homolog', *The Journal of Molecular Diagnostics*, 20(4), pp. 512–521. Available at: https://doi.org/10.1016/j.jmoldx.2018.03.010.

Hernández-Díaz, S. and Rodríguez, L.A.G. (2000) 'Association Between Nonsteroidal Anti-inflammatory Drugs and Upper Gastrointestinal Tract Bleeding/Perforation', *Archives of Internal Medicine*, 160(14), p. 2093. Available at: https://doi.org/10.1001/archinte.160.14.2093.

Hiatt, J.B. *et al.* (2013) 'Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation', *Genome Research*, 23(5), pp. 843–854. Available at: https://doi.org/10.1101/gr.147686.112.

Hitchins, M.P. *et al.* (2011) 'Dominantly Inherited Constitutional Epigenetic Silencing of MLH1 in a Cancer-Affected Family Is Linked to a Single Nucleotide Variant within the 5′UTR', *Cancer Cell*, 20(2), pp. 200–213. Available at: https://doi.org/10.1016/j.ccr.2011.07.003.

Holmes, M.D. *et al.* (2011) 'COX-2 expression predicts worse breast cancer prognosis and does not modify the association with aspirin', *Breast Cancer Research and Treatment*, 130(2), pp. 657–662. Available at: https://doi.org/10.1007/s10549-011-1651-7.

Hu, Z.I. *et al.* (2018) 'Evaluating Mismatch Repair Deficiency in Pancreatic Adenocarcinoma: Challenges and Recommendations', *Clinical Cancer Research*, 24(6), pp. 1326–1336. Available at: https://doi.org/10.1158/1078-0432.CCR-17-3099.

Huang, A.B. *et al.* (2021) ‘Neoadjuvant chemotherapy for advanced stage endometrial cancer: A systematic review’, *Gynecologic Oncology Reports*, 38, p. 100887. Available at: https://doi.org/10.1016/j.gore.2021.100887.

Huang, Q.-Y. *et al.* (2002) ‘Mutation Patterns at Dinucleotide Microsatellite Loci in Humans’, *The American Journal of Human Genetics*, 70(3), pp. 625–634. Available at: https://doi.org/10.1086/338997.

Huang, W.-Y. *et al.* (2019) ‘Frequency of Intracranial Hemorrhage With Low-Dose Aspirin in Individuals Without Symptomatic Cardiovascular Disease’, *JAMA Neurology*, 76(8), p. 906. Available at: https://doi.org/10.1001/jamaneurol.2019.1120.

Hunter, J.E. *et al.* (2015) ‘Universal tumor screening for Lynch syndrome: Assessment of the perspectives of patients with colorectal cancer regarding benefits and barriers’, *Cancer*, 121(18), pp. 3281–3289. Available at: https://doi.org/10.1002/cncr.29470.

International Stroke Trial Collaborative Group (1997) ‘The International Stroke Trial (IST): a randomised trial of aspirin, subcutaneous heparin, both, or neither among 19435 patients with acute ischaemic stroke. International Stroke Trial Collaborative Group.’, *Lancet (London, England)*, 349(9065), pp. 1569–81.

ISIS-2 Collaborative Group (1988) ‘Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group.’, *Lancet (London, England)*, 2(8607), pp. 349–60.

Ittaman, S. v., VanWormer, J.J. and Rezkalla, S.H. (2014) ‘The Role of Aspirin in the Prevention of Cardiovascular Disease’, *Clinical Medicine & Research*, 12(3–4), pp. 147–154. Available at: https://doi.org/10.3121/cmr.2013.1197.

Jackson, S.P. and Bartek, J. (2009) ‘The DNA-damage response in human biology and disease’, *Nature*, 461(7267), pp. 1071–1078. Available at: https://doi.org/10.1038/nature08467.

Jacobs, E.J. *et al.* (2007) ‘A Large Cohort Study of Long-Term Daily Use of Adult-Strength Aspirin and Cancer Incidence’, *JNCI Journal of the National Cancer Institute*, 99(8), pp. 608–615. Available at: https://doi.org/10.1093/jnci/djk132.

Janda, M. *et al.* (2017) ‘Effect of Total Laparoscopic Hysterectomy vs Total Abdominal Hysterectomy on Disease-Free Survival Among Women With Stage I Endometrial Cancer’, *JAMA*, 317(12), p. 1224. Available at: https://doi.org/10.1001/jama.2017.2068.

Jass, J.R. (2007) ‘Classification of colorectal cancer based on correlation of clinical, morphological and molecular features’, *Histopathology*, 50(1), pp. 113–130. Available at: https://doi.org/10.1111/j.1365-2559.2006.02549.x.

Jiricny, J. (2006) ‘The multifaceted mismatch-repair system’, *Nature Reviews Molecular Cell Biology*, 7(5), pp. 335–346. Available at: https://doi.org/10.1038/nrm1907.

Johnson, B.E. *et al.* (2014) 'Mutational Analysis Reveals the Origin and Therapy-Driven Evolution of Recurrent Glioma', *Science*, 343(6167), pp. 189–193. Available at: https://doi.org/10.1126/science.1239947.

Johnson, J.R. *et al.* (2010) 'Conservation of functional asymmetry in the mammalian MutLα ATPase', *DNA Repair*, 9(11), pp. 1209–1213. Available at: https://doi.org/10.1016/j.dnarep.2010.08.006.

Ju, J.Y. *et al.* (2018) 'Universal Lynch Syndrome Screening Should be Performed in All Upper Tract Urothelial Carcinomas', *American Journal of Surgical Pathology*, 42(11), pp. 1549–1555. Available at: https://doi.org/10.1097/PAS.0000000000001141.

Jun, S.-Y. *et al.* (2017) 'Lynch syndrome-related small intestinal adenocarcinomas', *Oncotarget*, 8(13), pp. 21483–21500. Available at: https://doi.org/10.18632/oncotarget.15277.

K. Frey, M. *et al.* (2012) 'Lynch Syndrome: Awareness among Medical Students at a United States Medical School', *Current Women's Health Reviews*, 8(3), pp. 242–247. Available at: https://doi.org/10.2174/157340412803760667.

Kadyrov, F.A. *et al.* (2006) 'Endonucleolytic Function of MutLα in Human Mismatch Repair', *Cell*, 126(2), pp. 297–308. Available at: https://doi.org/10.1016/j.cell.2006.05.039.

Kalgutkar, A.S. *et al.* (1998) 'Covalent Modification of Cyclooxygenase-2 (COX-2) by 2-Acetoxyphenyl Alkyl Sulfides, a New Class of Selective COX-2 Inactivators', *Journal of Medicinal Chemistry*, 41(24), pp. 4800–4818. Available at: https://doi.org/10.1021/jm980303s.

Kasela, M., Nyström, M. and Kansikas, M. (2019) '*PMS2* expression decrease causes severe problems in mismatch repair', *Human Mutation*, p. humu.23756. Available at: https://doi.org/10.1002/humu.23756.

Kehm, R.D. *et al.* (2019) 'Regular use of aspirin and other non-steroidal anti-inflammatory drugs and breast cancer risk for women at familial or genetic risk: a cohort study', *Breast Cancer Research*, 21(1), p. 52. Available at: https://doi.org/10.1186/s13058-019-1135-y.

Kelkar, Y.D. *et al.* (2011) 'A matter of life or death: How microsatellites emerge in and vanish from the human genome', *Genome Research*, 21(12), pp. 2038–2048. Available at: https://doi.org/10.1101/gr.122937.111.

Kerkar, R. and Kaur, S. (2013) 'Hysteroscopy in Endometrial Cancer', in *Textbook of Hysteroscopy*. Jaypee Brothers Medical Publishers (P) Ltd., pp. 215–215. Available at: https://doi.org/10.5005/jp/books/11731_18.

Kidambi, T.D. *et al.* (2016) 'Successful implementation of Lynch syndrome screening in a safety net institution', *Journal of Community Genetics*, 7(3), pp. 255–260. Available at: https://doi.org/10.1007/s12687-016-0270-8.

Knudson, A.G. (1971) ‘Mutation and Cancer: Statistical Study of Retinoblastoma’, *Proceedings of the National Academy of Sciences*, 68(4), pp. 820–823. Available at: https://doi.org/10.1073/pnas.68.4.820.

Kopp, E. and Ghosh, S. (1994) ‘Inhibition of NF-κB by Sodium Salicylate and Aspirin’, *Science*, 265(5174), pp. 956–959. Available at: https://doi.org/10.1126/science.8052854.

Kornberg, A. *et al.* (1964) ‘ENZYMATIC SYNTHESIS OF DEOXYRIBONUCLEIC ACID, XVI. OLIGONUCLEOTIDES AS TEMPLATES AND THE MECHANISM OF THEIR REPLICATION’, *Proceedings of the National Academy of Sciences*, 51(2), pp. 315–323. Available at: https://doi.org/10.1073/pnas.51.2.315.

Kothari, N. *et al.* (2015) ‘Impact of regular aspirin use on overall and cancer-specific survival in patients with colorectal cancer harboring a PIK3CA mutation’, *Acta Oncologica*, 54(4), pp. 487–492. Available at: https://doi.org/10.3109/0284186X.2014.990158.

Kune, G.A., Kune, S. and Watson, L.F. (1988) ‘Colorectal cancer risk, chronic illnesses, operations, and medications: case control results from the Melbourne Colorectal Cancer Study.’, *Cancer research*, 48(15), pp. 4399–404.

Kunkel, T.A. and Erie, D.A. (2005) ‘DNA MISMATCH REPAIR’, *Annual Review of Biochemistry*, 74(1), pp. 681–710. Available at: https://doi.org/10.1146/annurev.biochem.74.082803.133243.

Lagerstedt-Robinson, K. *et al.* (2016) ‘Mismatch repair gene mutation spectrum in the Swedish Lynch syndrome population’, *Oncology Reports*, 36(5), pp. 2823–2835. Available at: https://doi.org/10.3892/or.2016.5060.

Latham, A. *et al.* (2019) ‘Microsatellite Instability Is Associated With the Presence of Lynch Syndrome Pan-Cancer’, *Journal of Clinical Oncology*, 37(4), pp. 286–295. Available at: https://doi.org/10.1200/JCO.18.00283.

Law, B.K. *et al.* (2000) ‘Salicylate-induced Growth Arrest Is Associated with Inhibition of p70s6k and Down-regulation of c-Myc, Cyclin D1, Cyclin A, and Proliferating Cell Nuclear Antigen’, *Journal of Biological Chemistry*, 275(49), pp. 38261–38267. Available at: https://doi.org/10.1074/jbc.M005545200.

Leach, F.S. *et al.* (1993) ‘Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer’, *Cell*, 75(6), pp. 1215–1225. Available at: https://doi.org/10.1016/0092-8674(93)90330-S.

Lee, V. *et al.* (2016) ‘Mismatch Repair Deficiency and Response to Immune Checkpoint Blockade’, *The Oncologist*, 21(10), pp. 1200–1211. Available at: https://doi.org/10.1634/theoncologist.2016-0046.

Levine, M. *et al.* (2021) ‘Upfront Multigene Panel Testing for Cancer Susceptibility in Newly Diagnosed Endometrial Cancer Patients: A Prospective Cohort Study’, *SSRN Electronic Journal* [Preprint]. Available at: https://doi.org/10.2139/ssrn.3820515.

Levine, M.D. *et al.* (2021) ‘Up-Front Multigene Panel Testing for Cancer Susceptibility in Patients With Newly Diagnosed Endometrial Cancer: A Multicenter Prospective Study’, *JCO Precision Oncology*, (5), pp. 1588–1602. Available at: https://doi.org/10.1200/PO.21.00249.

Li, G.-M. (2003) ‘DNA mismatch repair and cancer’, *Frontiers in Bioscience*, 8(4), p. 1121. Available at: https://doi.org/10.2741/1121.

Li, G.-M. (2008) ‘Mechanisms and functions of DNA mismatch repair’, *Cell Research*, 18(1), pp. 85–98. Available at: https://doi.org/10.1038/cr.2007.115.

Li, H. and Durbin, R. (2010) ‘Fast and accurate long-read alignment with Burrows–Wheeler transform’, *Bioinformatics*, 26(5), pp. 589–595. Available at: https://doi.org/10.1093/bioinformatics/btp698.

Li, J. *et al.* (2015) ‘A Comprehensive Strategy for Accurate Mutation Detection of the Highly Homologous PMS2’, *The Journal of Molecular Diagnostics*, 17(5), pp. 545–553. Available at: https://doi.org/10.1016/j.jmoldx.2015.04.001.

Liao, X. *et al.* (2012) ‘Aspirin Use, Tumor *PIK3CA* Mutation, and Colorectal-Cancer Survival’, *New England Journal of Medicine*, 367(17), pp. 1596–1606. Available at: https://doi.org/10.1056/NEJMoa1207756.

Loll, P.J., Picot, D. and Garavito, R.M. (1995) ‘The structural basis of aspirin activity inferred from the crystal structure of inactivated prostaglandin H2 synthase’, *Nature Structural Biology*, 2(8), pp. 637–643. Available at: https://doi.org/10.1038/nsb0895-637.

Long, B. *et al.* (2020) ‘Ultrasound detection of endometrial cancer in women with postmenopausal bleeding: Systematic review and meta-analysis’, *Gynecologic Oncology*, 157(3), pp. 624–633. Available at: https://doi.org/10.1016/j.ygyno.2020.01.032.

Longley, M.J., Pierce, A.J. and Modrich, P. (1997) ‘DNA Polymerase δ Is Required for Human Mismatch Repair in Vitro’, *Journal of Biological Chemistry*, 272(16), pp. 10917–10921. Available at: https://doi.org/10.1074/jbc.272.16.10917.

López-Farré, A. *et al.* (1995) ‘Effects of Aspirin on Platelet-Neutrophil Interactions’, *Circulation*, 91(7), pp. 2080–2088. Available at: https://doi.org/10.1161/01.CIR.91.7.2080.

Lotsari, J.E. *et al.* (2012) ‘Breast carcinoma and Lynch syndrome: molecular analysis of tumors arising in mutation carriers, non-carriers, and sporadic cases’, *Breast Cancer Research*, 14(3), p. R90. Available at: https://doi.org/10.1186/bcr3205.

Lu, K.H. *et al.* (2007) ‘Prospective Determination of Prevalence of Lynch Syndrome in Young Women With Endometrial Cancer’, *Journal of Clinical Oncology*, 25(33), pp. 5158–5164. Available at: https://doi.org/10.1200/JCO.2007.10.8597.

Lynch, H. *et al.* (2009) ‘Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications’, *Clinical Genetics*, 76(1), pp. 1–18. Available at: https://doi.org/10.1111/j.1399-0004.2009.01230.x.

Lynch, H.T. and Krush, A.J. (1971) ‘Cancer family “G” revisited: 1895-1970’, *Cancer*, 27(6), pp. 1505–1511. Available at: https://doi.org/10.1002/1097-0142(197106)27:6<1505::AID-CNCR2820270635>3.0.CO;2-L.

Lynch, H.T. and de la Chapelle, A. (2003) ‘Hereditary Colorectal Cancer’, *New England Journal of Medicine*, 348(10), pp. 919–932. Available at: https://doi.org/10.1056/NEJMra012242.

Mármol, I. *et al.* (2017) ‘Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer’, *International Journal of Molecular Sciences*, 18(1), p. 197. Available at: https://doi.org/10.3390/ijms18010197.

Matsuo, K. *et al.* (2021) ‘Ovarian conservation for young women with early-stage, low-grade endometrial cancer: a 2-step schema’, *American Journal of Obstetrics and Gynecology*, 224(6), pp. 574–584. Available at: https://doi.org/10.1016/j.ajog.2020.12.1213.

McCarthy, A.J. *et al.* (2019) ‘Heterogenous loss of mismatch repair (MMR) protein expression: a challenge for immunohistochemical interpretation and microsatellite instability (MSI) evaluation’, *The Journal of Pathology: Clinical Research*, 5(2), pp. 115–129. Available at: https://doi.org/10.1002/cjp2.120.

McConechy, M.K. *et al.* (2015) ‘Detection of DNA mismatch repair (MMR) deficiencies by immunohistochemistry can effectively diagnose the microsatellite instability (MSI) phenotype in endometrial carcinomas’, *Gynecologic Oncology*, 137(2), pp. 306–310. Available at: https://doi.org/10.1016/j.ygyno.2015.01.541.

Mensenkamp, A.R. *et al.* (2014) ‘Somatic mutations in MLH1 and MSH2 are a frequent cause of mismatch-repair deficiency in Lynch syndrome-like tumors.’, *Gastroenterology*, 146(3), pp. 643-646.e8. Available at: https://doi.org/10.1053/j.gastro.2013.12.002.

Metcalf, A.M. and Spurdle, A.B. (2014) ‘Endometrial tumour BRAF mutations and MLH1 promoter methylation as predictors of germline mismatch repair gene mutation status: a literature review’, *Familial Cancer*, 13(1), pp. 1–12. Available at: https://doi.org/10.1007/s10689-013-9671-6.

Metcalfe, M.J. *et al.* (2018) ‘Universal Point of Care Testing for Lynch Syndrome in Patients with Upper Tract Urothelial Carcinoma’, *Journal of Urology*, 199(1), pp. 60–65. Available at: https://doi.org/10.1016/j.juro.2017.08.002.

Mills, A.M. *et al.* (2014a) ‘Lynch Syndrome Screening Should Be Considered for All Patients With Newly Diagnosed Endometrial Cancer’, *American Journal of Surgical Pathology*, 38(11), pp. 1501–1509. Available at: https://doi.org/10.1097/PAS.0000000000000321.

Mills, A.M. *et al.* (2014b) ‘Lynch Syndrome Screening Should Be Considered for All Patients With Newly Diagnosed Endometrial Cancer’, *American Journal of Surgical Pathology*, 38(11), pp. 1501–1509. Available at: https://doi.org/10.1097/PAS.0000000000000321.

Miyaki, M. *et al.* (1997) ‘Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer’, *Nature Genetics*, 17(3), pp. 271–272. Available at: https://doi.org/10.1038/ng1197-271.

Møller, P. *et al.* (2017) ‘Cancer incidence and survival in Lynch syndrome patients receiving colonoscopic and gynaecological surveillance: first report from the prospective Lynch syndrome database’, *Gut*, 66(3), pp. 464–472. Available at: https://doi.org/10.1136/gutjnl-2015-309675.

Møller, P. *et al.* (2018) ‘Cancer risk and survival in *path_MMR* carriers by gene and gender up to 75 years of age: a report from the Prospective Lynch Syndrome Database’, *Gut*, 67(7), pp. 1306–1316. Available at: https://doi.org/10.1136/gutjnl-2017-314057.

Møller, P. (2020) ‘The Prospective Lynch Syndrome Database reports enable evidence-based personal precision health care’, *Hereditary Cancer in Clinical Practice*, 18(1), p. 6. Available at: https://doi.org/10.1186/s13053-020-0138-0.

Moreira, L. *et al.* (2012) ‘Identification of Lynch Syndrome Among Patients With Colorectal Cancer’, *JAMA*, 308(15), p. 1555. Available at: https://doi.org/10.1001/jama.2012.13088.

Morice, P. *et al.* (2016) ‘Endometrial cancer’, *The Lancet*, 387(10023), pp. 1094–1108. Available at: https://doi.org/10.1016/S0140-6736(15)00130-0.

Morrison, J. *et al.* (2022) ‘British Gynaecological Cancer Society (BGCS) uterine cancer guidelines: Recommendations for practice’, *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 270, pp. 50–89. Available at: https://doi.org/10.1016/j.ejogrb.2021.11.423.

Motta, R. *et al.* (2021) ‘Immunotherapy in microsatellite instability metastatic colorectal cancer: Current status and future perspectives.’, *Journal of clinical and translational research*, 7(4), pp. 511–522.

Movahedi, M. *et al.* (2015) ‘Obesity, Aspirin, and Risk of Colorectal Cancer in Carriers of Hereditary Colorectal Cancer: A Prospective Investigation in the CAPP2 Study’, *Journal of Clinical Oncology*, 33(31), pp. 3591–3597. Available at: https://doi.org/10.1200/JCO.2014.58.9952.

Murata, T. *et al.* (1997) ‘Altered pain perception and inflammatory response in mice lacking prostacyclin receptor’, *Nature*, 388(6643), pp. 678–682. Available at: https://doi.org/10.1038/41780.

Murphy, E., Curneen, J.M.G. and McEvoy, J.W. (2021) ‘Aspirin in the Modern Era of Cardiovascular Disease Prevention’, *Methodist DeBakey Cardiovascular Journal*, 17(4), pp. 36–47. Available at: https://doi.org/10.14797/mdcvj.293.

Murphy, K.M. *et al.* (2006) ‘Comparison of the Microsatellite Instability Analysis System and the Bethesda Panel for the Determination of Microsatellite Instability in Colorectal Cancers’, *The Journal of Molecular Diagnostics*, 8(3), pp. 305–311. Available at: https://doi.org/10.2353/jmoldx.2006.050092.

Nadir, E. *et al.* (1996) 'Microsatellite spreading in the human genome: evolutionary mechanisms and structural implications.', *Proceedings of the National Academy of Sciences*, 93(13), pp. 6470–6475. Available at: https://doi.org/10.1073/pnas.93.13.6470.

Nakagawa, H. *et al.* (2004) 'Mismatch Repair Gene ***PMS2*** ', *Cancer Research*, 64(14), pp. 4721–4727. Available at: https://doi.org/10.1158/0008-5472.CAN-03-2879.

Narumiya, S. and FitzGerald, G.A. (2001) 'Genetic and pharmacological analysis of prostanoid receptor function', *Journal of Clinical Investigation*, 108(1), pp. 25–30. Available at: https://doi.org/10.1172/JCI13455.

Nicolaides, N.C. *et al.* (1994) 'Mutations of two P/WS homologues in hereditary nonpolyposis colon cancer', *Nature*, 371(6492), pp. 75–80. Available at: https://doi.org/10.1038/371075a0.

Nicolosi, P. *et al.* (2019) 'Prevalence of Germline Variants in Prostate Cancer and Implications for Current Genetic Testing Guidelines', *JAMA Oncology*, 5(4), p. 523. Available at: https://doi.org/10.1001/jamaoncol.2018.6760.

Nieminen, T.T. *et al.* (2009) 'Molecular Analysis of Endometrial Tumorigenesis: Importance of Complex Hyperplasia Regardless of Atypia', *Clinical Cancer Research*, 15(18), pp. 5772–5783. Available at: https://doi.org/10.1158/1078-0432.CCR-09-0506.

Nishihara, R., Lochhead, P., *et al.* (2013) 'Aspirin Use and Risk of Colorectal Cancer According to BRAF Mutation Status', *JAMA*, 309(24), p. 2563. Available at: https://doi.org/10.1001/jama.2013.6599.

Nishihara, R., Wu, K., *et al.* (2013) 'Long-Term Colorectal-Cancer Incidence and Mortality after Lower Endoscopy', *New England Journal of Medicine*, 369(12), pp. 1095–1105. Available at: https://doi.org/10.1056/NEJMoa1301969.

Niskakoski, A. *et al.* (2013) 'Distinct molecular profiles in Lynch syndrome-associated and sporadic ovarian carcinomas', *International Journal of Cancer*, p. n/a-n/a. Available at: https://doi.org/10.1002/ijc.28287.

Noll, A. *et al.* (2018) 'Barriers to Lynch Syndrome Testing and Preoperative Result Availability in Earlyonset Colorectal Cancer: A National Physician Survey Study', *Clinical and Translational Gastroenterology*, 9(9), p. e185. Available at: https://doi.org/10.1038/s41424-018-0047-y.

Núñez, L. *et al.* (2006) 'Cell proliferation depends on mitochondrial Ca $^{2+}$ uptake: inhibition by salicylate', *The Journal of Physiology*, 571(1), pp. 57–73. Available at: https://doi.org/10.1113/jphysiol.2005.100586.

O'Donnell, V.B. (2003) 'Free Radicals and Lipid Signaling in Endothelial Cells', *Antioxidants & Redox Signaling*, 5(2), pp. 195–203. Available at: https://doi.org/10.1089/152308603764816550.

van Oers, J.M.M. *et al.* (2010) ‘PMS2 endonuclease activity has distinct biological functions and is essential for genome maintenance.’, *Proceedings of the National Academy of Sciences of the United States of America*, 107(30), pp. 13384–9. Available at: https://doi.org/10.1073/pnas.1008589107.

Pal, T. *et al.* (2012) ‘Frequency of mutations in mismatch repair genes in a population-based study of women with ovarian cancer’, *British Journal of Cancer*, 107(10), pp. 1783–1790. Available at: https://doi.org/10.1038/bjc.2012.452.

Palomaki, G.E. *et al.* (2009) ‘EGAPP supplementary evidence review: DNA testing strategies aimed at reducing morbidity and mortality from Lynch syndrome’, *Genetics in Medicine*, 11(1), pp. 42–65. Available at: https://doi.org/10.1097/GIM.0b013e31818fa2db.

Pan, M.-R., Chang, H.-C. and Hung, W.-C. (2008) ‘Non-steroidal anti-inflammatory drugs suppress the ERK signaling pathway via block of Ras/c-Raf interaction and activation of MAP kinase phosphatases’, *Cellular Signalling*, 20(6), pp. 1134–1141. Available at: https://doi.org/10.1016/j.cellsig.2008.02.004.

Passarello, K., Kurian, S. and Villanueva, V. (2019) ‘Endometrial Cancer: An Overview of Pathophysiology, Management, and Care’, *Seminars in Oncology Nursing*, 35(2), pp. 157–165. Available at: https://doi.org/10.1016/j.soncn.2019.02.002.

Patil, D.T. *et al.* (2012) ‘A Five-marker Panel in a Multiplex PCR Accurately Detects Microsatellite Instability-high Colorectal Tumors Without Control DNA’, *Diagnostic Molecular Pathology*, 21(3), pp. 127–133. Available at: https://doi.org/10.1097/PDM.0b013e3182461cc3.

Patrignani, P. *et al.* (2011) ‘Managing the adverse effects of nonsteroidal anti-inflammatory drugs’, *Expert Review of Clinical Pharmacology*, 4(5), pp. 605–621. Available at: https://doi.org/10.1586/ecp.11.36.

Patrignani, P., Filabozzi, P. and Patrono, C. (1982) ‘Selective Cumulative Inhibition of Platelet Thromboxane Production by Low-dose Aspirin in Healthy Subjects’, *Journal of Clinical Investigation*, 69(6), pp. 1366–1372. Available at: https://doi.org/10.1172/JCI110576.

Patrignani, P. and Patrono, C. (2015) ‘Cyclooxygenase inhibitors: From pharmacology to clinical read-outs’, *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1851(4), pp. 422–432. Available at: https://doi.org/10.1016/j.bbalip.2014.09.016.

Patrono, C. *et al.* (2001) ‘Platelet-Active Drugs’, *Chest*, 119(1), pp. 39S-63S. Available at: https://doi.org/10.1378/chest.119.1_suppl.39S.

Patrono, C., Patrignani, P. and Rodríguez, L.A.G. (2001) ‘Cyclooxygenase-selective inhibition of prostanoid formation: transducing biochemical selectivity into clinical read-outs’, *Journal of Clinical Investigation*, 108(1), pp. 7–13. Available at: https://doi.org/10.1172/JCI200113418.

Pearlman, R. *et al.* (2017) ‘Prevalence and Spectrum of Germline Cancer Susceptibility Gene Mutations Among Patients With Early-Onset Colorectal Cancer’, *JAMA Oncology*, 3(4), p. 464. Available at: https://doi.org/10.1001/jamaoncol.2016.5194.

Pearlman, R. *et al.* (2019) ‘Clinical characteristics of patients with colorectal cancer with double somatic mismatch repair mutations compared with Lynch syndrome’, *Journal of Medical Genetics*, 56(7), pp. 462–470. Available at: https://doi.org/10.1136/jmedgenet-2018-105698.

Peltomäki, P. (2016) ‘Update on Lynch syndrome genomics’, *Familial Cancer*, 15(3), pp. 385–393. Available at: https://doi.org/10.1007/s10689-016-9882-8.

Pennant, M. *et al.* (2017) ‘Premenopausal abnormal uterine bleeding and risk of endometrial cancer’, *BJOG: An International Journal of Obstetrics & Gynaecology*, 124(3), pp. 404–411. Available at: https://doi.org/10.1111/1471-0528.14385.

Peskar, B.M. (2001) ‘Role of cyclooxygenase isoforms in gastric mucosal defence’, *Journal of Physiology-Paris*, 95(1–6), pp. 3–9. Available at: https://doi.org/10.1016/S0928-4257(01)00003-1.

Peto, R. *et al.* (1988) ‘Randomised trial of prophylactic daily aspirin in British male doctors’, *BMJ*, 296(6618), pp. 313–316. Available at: https://doi.org/10.1136/bmj.296.6618.313.

Popat, S., Hubner, R. and Houlston, R.S. (2005) ‘Systematic Review of Microsatellite Instability and Colorectal Cancer Prognosis’, *Journal of Clinical Oncology*, 23(3), pp. 609–618. Available at: https://doi.org/10.1200/JCO.2005.01.086.

Porkka, N. *et al.* (2019) ‘Epidemiological, clinical and molecular characterization of Lynch-like syndrome: A population-based study’, *International Journal of Cancer*, 145(1), pp. 87–98. Available at: https://doi.org/10.1002/ijc.32085.

Porkka, N.K. *et al.* (2020) ‘Does breast carcinoma belong to the Lynch syndrome tumor spectrum? – Somatic mutational profiles vs. ovarian and colorectal carcinomas’, *Oncotarget*, 11(14), pp. 1244–1256. Available at: https://doi.org/10.18632/oncotarget.27538.

van der Post, R.S. *et al.* (2010) ‘Risk of urothelial bladder cancer in Lynch syndrome is increased, in particular among MSH2 mutation carriers’, *Journal of Medical Genetics*, 47(7), pp. 464–470. Available at: https://doi.org/10.1136/jmg.2010.076992.

Prolla, T.A. *et al.* (1998) ‘Tumour susceptibility and spontaneous mutation in mice deficient in Mlh1, Pms1 and Pms2 DMA mismatch repair’, *Nature Genetics*, 18(3), pp. 276–279. Available at: https://doi.org/10.1038/ng0398-276.

Raber, I. *et al.* (2019) ‘The rise and fall of aspirin in the primary prevention of cardiovascular disease’, *The Lancet*, 393(10186), pp. 2155–2167. Available at: https://doi.org/10.1016/S0140-6736(19)30541-0.

Raglan, O. *et al.* (2019) ‘Risk factors for endometrial cancer: An umbrella review of the literature’, *International Journal of Cancer*, 145(7), pp. 1719–1730. Available at: https://doi.org/10.1002/ijc.31961.

Rao, C. and Reddy, B. (2004) ‘NSAIDs and Chemoprevention’, *Current Cancer Drug Targets*, 4(1), pp. 29–42. Available at: https://doi.org/10.2174/1568009043481632.

Redford, L. *et al.* (2018) ‘A novel panel of short mononucleotide repeats linked to informative polymorphisms enabling effective high volume low cost discrimination between mismatch repair deficient and proficient tumours’, *PLOS ONE*, 13(8), p. e0203052. Available at: https://doi.org/10.1371/journal.pone.0203052.

Reimers, M.S. *et al.* (2014) ‘Expression of HLA Class I Antigen, Aspirin Use, and Survival After a Diagnosis of Colon Cancer’, *JAMA Internal Medicine*, 174(5), p. 732. Available at: https://doi.org/10.1001/jamainternmed.2014.511.

Ricciotti, E. *et al.* (2013) ‘COX-2, the dominant source of prostacyclin’, *Proceedings of the National Academy of Sciences*, 110(3). Available at: https://doi.org/10.1073/pnas.1219073110.

Riegger, G.A. *et al.* (1991) ‘Effects of acetylsalicylic acid on renal function in patients with chronic heart failure.’, *The American journal of medicine*, 90(5), pp. 571–5.

Rigas, B., Goldman, I.S. and Levine, L. (1993) ‘Altered eicosanoid levels in human colon cancer.’, *The Journal of laboratory and clinical medicine*, 122(5), pp. 518–23.

Ring, K.L. *et al.* (2016) ‘Germline multi-gene hereditary cancer panel testing in an unselected endometrial cancer cohort’, *Modern Pathology*, 29(11), pp. 1381–1389. Available at: https://doi.org/10.1038/modpathol.2016.135.

Risinger, J.I. *et al.* (1996) ‘Mutation of MSH3 in endometrial cancer and evidence for its functional role in heteroduplex repair’, *Nature Genetics*, 14(1), pp. 102–105. Available at: https://doi.org/10.1038/ng0996-102.

Roberts, M.E. *et al.* (2014) ‘A clinical scoring system to identify patients with sebaceous neoplasms at risk for the Muir–Torre variant of Lynch syndrome’, *Genetics in Medicine*, 16(9), pp. 711–716. Available at: https://doi.org/10.1038/gim.2014.19.

Rodriguez-Bigas, M.A. *et al.* (1997) ‘A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: Meeting Highlights and Bethesda Guidelines’, *JNCI Journal of the National Cancer Institute*, 89(23), pp. 1758–1762. Available at: https://doi.org/10.1093/jnci/89.23.1758.

Rodríguez–Soler, M. *et al.* (2013) ‘Risk of Cancer in Cases of Suspected Lynch Syndrome Without Germline Mutation’, *Gastroenterology*, 144(5), pp. 926-932.e1. Available at: https://doi.org/10.1053/j.gastro.2013.01.044.

Romano, M. *et al.* (2015) ‘Lipoxins and aspirin-triggered lipoxins in resolution of inflammation’, *European Journal of Pharmacology*, 760, pp. 49–63. Available at: https://doi.org/10.1016/j.ejphar.2015.03.083.

Rosty, C. *et al.* (2016) 'Germline mutations in *PMS2* and *MLH1* in individuals with solitary loss of PMS2 expression in colorectal carcinomas from the Colon Cancer Family Registry Cohort', *BMJ Open*, 6(2), p. e010293. Available at: https://doi.org/10.1136/bmjopen-2015-010293.

Rothwell, P.M. *et al.* (2010) 'Long-term effect of aspirin on colorectal cancer incidence and mortality: 20-year follow-up of five randomised trials', *The Lancet*, 376(9754), pp. 1741–1750. Available at: https://doi.org/10.1016/S0140-6736(10)61543-7.

Rothwell, P.M. *et al.* (2011) 'Effect of daily aspirin on long-term risk of death due to cancer: analysis of individual patient data from randomised trials', *The Lancet*, 377(9759), pp. 31–41. Available at: https://doi.org/10.1016/S0140-6736(10)62110-1.

Rothwell, P.M., Wilson, M., *et al.* (2012) 'Effect of daily aspirin on risk of cancer metastasis: a study of incident cancers during randomised controlled trials', *The Lancet*, 379(9826), pp. 1591–1601. Available at: https://doi.org/10.1016/S0140-6736(12)60209-8.

Rothwell, P.M., Price, J.F., *et al.* (2012) 'Short-term effects of daily aspirin on cancer incidence, mortality, and non-vascular death: analysis of the time course of risks and benefits in 51 randomised controlled trials', *The Lancet*, 379(9826), pp. 1602–1612. Available at: https://doi.org/10.1016/S0140-6736(11)61720-0.

Rothwell, P.M. *et al.* (2016) 'Effects of aspirin on risk and severity of early recurrent stroke after transient ischaemic attack and ischaemic stroke: time-course analysis of randomised trials', *The Lancet*, 388(10042), pp. 365–375. Available at: https://doi.org/10.1016/S0140-6736(16)30468-8.

Rubio, I. *et al.* (2016) 'Analysis of Lynch Syndrome Mismatch Repair Genes in Women with Endometrial Cancer', *Oncology*, 91(3), pp. 171–176. Available at: https://doi.org/10.1159/000447972.

Ryan, N.A.J. *et al.* (2019) 'The proportion of endometrial cancers associated with Lynch syndrome: a systematic review of the literature and meta-analysis', *Genetics in Medicine*, 21(10), pp. 2167–2180. Available at: https://doi.org/10.1038/s41436-019-0536-8.

Ryan, N.A.J. *et al.* (2020) 'The proportion of endometrial tumours associated with Lynch syndrome (PETALS): A prospective cross-sectional study', *PLOS Medicine*, 17(9), p. e1003263. Available at: https://doi.org/10.1371/journal.pmed.1003263.

Salahshor, S. *et al.* (2001) 'Microsatellite Instability and hMLH1 and hMSH2 Expression Analysis in Familial and Sporadic Colorectal Cancer', *Laboratory Investigation*, 81(4), pp. 535–541. Available at: https://doi.org/10.1038/labinvest.3780262.

Salem, M.E. *et al.* (2020) 'Relationship between <scp>MLH1</scp> , <scp>PMS2</scp> , <scp>MSH2</scp> and <scp>MSH6</scp> gene-specific alterations and tumor mutational burden in 1057 microsatellite instability-high solid

tumors', *International Journal of Cancer*, 147(10), pp. 2948–2956. Available at: https://doi.org/10.1002/ijc.33115.

Santucci-Darmanin, S. and Paquis-Flucklinger, V. (2003) 'Les homologues de MutS et de MutL au cours de la méiose chez les mammifères', *médecine/sciences*, 19(1), pp. 85–91. Available at: https://doi.org/10.1051/medsci/200319185.

Scarabelli, C. *et al.* (1998) 'Maximal Cytoreductive Surgery as a Reasonable Therapeutic Alternative for Recurrent Endometrial Carcinoma', *Gynecologic Oncology*, 70(1), pp. 90–93. Available at: https://doi.org/10.1006/gyno.1998.5017.

Schlötterer, C. and Tautz, D. (1992) 'Slippage synthesis of simple sequence DNA', *Nucleic Acids Research*, 20(2), pp. 211–215. Available at: https://doi.org/10.1093/nar/20.2.211.

Schon, K. *et al.* (2018) 'Evaluation of universal immunohistochemical screening of sebaceous neoplasms in a service setting', *Clinical and Experimental Dermatology*, 43(4), pp. 410–415. Available at: https://doi.org/10.1111/ced.13359.

Schulmann, K. *et al.* (2005) 'HNPCC-associated small bowel cancer: Clinical and molecular characteristics', *Gastroenterology*, 128(3), pp. 590–599. Available at: https://doi.org/10.1053/j.gastro.2004.12.051.

Sciulli, M.G. *et al.* (2005) 'Platelet activation in patients with colorectal cancer', *Prostaglandins, Leukotrienes and Essential Fatty Acids*, 72(2), pp. 79–83. Available at: https://doi.org/10.1016/j.plefa.2004.10.006.

Sekine, S. *et al.* (2017) 'Mismatch repair deficiency commonly precedes adenoma formation in Lynch Syndrome-Associated colorectal tumorigenesis', *Modern Pathology*, 30(8), pp. 1144–1151. Available at: https://doi.org/10.1038/modpathol.2017.39.

Senter, L. *et al.* (2008) 'The Clinical Phenotype of Lynch Syndrome Due to Germ-Line PMS2 Mutations', *Gastroenterology*, 135(2), pp. 419-428.e1. Available at: https://doi.org/10.1053/j.gastro.2008.04.026.

Serhan, C.N. and Chiang, N. (2013) 'Resolution phase lipid mediators of inflammation: agonists of resolution', *Current Opinion in Pharmacology*, 13(4), pp. 632–640. Available at: https://doi.org/10.1016/j.coph.2013.05.012.

Shaikh, T. *et al.* (2018) 'Mismatch Repair Deficiency Testing in Patients With Colorectal Cancer and Nonadherence to Testing Guidelines in Young Adults', *JAMA Oncology*, 4(2), p. e173580. Available at: https://doi.org/10.1001/jamaoncol.2017.3580.

Shia, J. (2008) 'Immunohistochemistry versus Microsatellite Instability Testing For Screening Colorectal Cancer Patients at Risk For Hereditary Nonpolyposis Colorectal Cancer Syndrome', *The Journal of Molecular Diagnostics*, 10(4), pp. 293–300. Available at: https://doi.org/10.2353/jmoldx.2008.080031.

Simmons, D.L., Botting, R.M. and Hla, T. (2004) 'Cyclooxygenase Isozymes: The Biology of Prostaglandin Synthesis and Inhibition', *Pharmacological Reviews*, 56(3), pp. 387–437. Available at: https://doi.org/10.1124/pr.56.3.3.

Sloan, E.A., Moskaluk, C.A. and Mills, A.M. (2017) 'Mucinous Differentiation With Tumor Infiltrating Lymphocytes Is a Feature of Sporadically Methylated Endometrial Carcinomas', *International Journal of Gynecological Pathology*, 36(3), pp. 205–216. Available at: https://doi.org/10.1097/PGP.0000000000000315.

Smith, J.B., Araki, H. and Lefer, A.M. (1980) 'Thromboxane A2, prostacyclin and aspirin: effects on vascular tone and platelet aggregation.', *Circulation*, 62(6 Pt 2), pp. V19-25.

Smith, S.G. *et al.* (2017) 'General practitioner attitudes towards prescribing aspirin to carriers of Lynch Syndrome: findings from a national survey', *Familial Cancer*, 16(4), pp. 509–516. Available at: https://doi.org/10.1007/s10689-017-9986-9.

Smyth, E.M. *et al.* (2009) 'Prostanoids in health and disease', *Journal of Lipid Research*, 50, pp. S423–S428. Available at: https://doi.org/10.1194/jlr.R800094-JLR200.

Snowsill, T. *et al.* (2014) 'A systematic review and economic evaluation of diagnostic strategies for Lynch syndrome', *Health Technology Assessment*, 18(58), pp. 1–406. Available at: https://doi.org/10.3310/hta18580.

Sorosky, J.I. (2012) 'Endometrial Cancer', *Obstetrics & Gynecology*, 120(2, Part 1), pp. 383–397. Available at: https://doi.org/10.1097/AOG.0b013e3182605bf1.

Spitz, G.A. *et al.* (2009) 'Acetylsalicylic acid and salicylic acid decrease tumor cell viability and glucose metabolism modulating 6-phosphofructo-1-kinase structure and activity', *Biochemical Pharmacology*, 77(1), pp. 46–53. Available at: https://doi.org/10.1016/j.bcp.2008.09.020.

Stark, L.A. *et al.* (2006) 'Aspirin activates the NF- B signalling pathway and induces apoptosis in intestinal neoplasia in two in vivo models of human colorectal cancer', *Carcinogenesis*, 28(5), pp. 968–976. Available at: https://doi.org/10.1093/carcin/bgl220.

Steer, K.A. *et al.* (1997) 'Aspirin protects low density lipoprotein from oxidative modification.', *Heart*, 77(4), pp. 333–337. Available at: https://doi.org/10.1136/hrt.77.4.333.

Steering Committee of the Physicians' Health Study Research Group (1989) 'Final Report on the Aspirin Component of the Ongoing Physicians' Health Study', *New England Journal of Medicine*, 321(3), pp. 129–135. Available at: https://doi.org/10.1056/NEJM198907203210301.

Steinbach, G. *et al.* (2000) 'The Effect of Celecoxib, a Cyclooxygenase-2 Inhibitor, in Familial Adenomatous Polyposis', *New England Journal of Medicine*, 342(26), pp. 1946–1952. Available at: https://doi.org/10.1056/NEJM200006293422603.

Stelloo, E. *et al.* (2017) ‘Practical guidance for mismatch repair-deficiency testing in endometrial cancer’, *Annals of Oncology*, 28(1), pp. 96–102. Available at: https://doi.org/10.1093/annonc/mdw542.

Stinton, C. *et al.* (2021) ‘Testing for lynch syndrome in people with endometrial cancer using immunohistochemistry and microsatellite instability-based testing strategies – A systematic review of test accuracy’, *Gynecologic Oncology*, 160(1), pp. 148–160. Available at: https://doi.org/10.1016/j.ygyno.2020.10.003.

Stoffel, E.M. *et al.* (2015) ‘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’, *Journal of Clinical Oncology*, 33(2), pp. 209–217. Available at: https://doi.org/10.1200/JCO.2014.58.1322.

Strand, M. *et al.* (1993) ‘Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair’, *Nature*, 365(6443), pp. 274–276. Available at: https://doi.org/10.1038/365274a0.

Su, S.S. and Modrich, P. (1986) ‘Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs.’, *Proceedings of the National Academy of Sciences*, 83(14), pp. 5057–5061. Available at: https://doi.org/10.1073/pnas.83.14.5057.

Subramanian, S., Mishra, R.K. and Singh, L. (2003) ‘Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions’, *Genome Biology*, 4(2), p. R13. Available at: https://doi.org/10.1186/gb-2003-4-2-r13.

Sung, H. *et al.* (2021) ‘Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries’, *CA: A Cancer Journal for Clinicians*, 71(3), pp. 209–249. Available at: https://doi.org/10.3322/caac.21660.

Takeda, T. *et al.* (2018) ‘Screening for Lynch syndrome using risk assessment criteria in patients with ovarian cancer’, *Journal of Gynecologic Oncology*, 29(3). Available at: https://doi.org/10.3802/jgo.2018.29.e29.

Talseth-Palmer, B.A. *et al.* (2010) ‘MSH6 and PMS2 mutation positive Australian Lynch syndrome families: novel mutations, cancer risk and age of diagnosis of colorectal cancer’, *Hereditary Cancer in Clinical Practice*, 8(1), p. 5. Available at: https://doi.org/10.1186/1897-4287-8-5.

Tan, Y., Spurdle, A. and Obermair, A. (2014) ‘Knowledge, Attitudes and Referral Patterns of Lynch Syndrome: A Survey of Clinicians in Australia’, *Journal of Personalized Medicine*, 4(2), pp. 218–244. Available at: https://doi.org/10.3390/jpm4020218.

The American College of Obstetricians and Gynecologists (2015) ‘Practice Bulletin No. 149: Endometrial Cancer’, *Obstetrics & Gynecology*, 125(4), pp. 1006–1026. Available at: https://doi.org/10.1097/01.AOG.0000462977.61229.de.

Therkildsen, C. *et al.* (2018) 'Molecular subtype classification of urothelial carcinoma in Lynch syndrome', *Molecular Oncology*, 12(8), pp. 1286–1295. Available at: https://doi.org/10.1002/1878-0261.12325.

Thun, M.J., Jacobs, E.J. and Patrono, C. (2012) 'The role of aspirin in cancer prevention', *Nature Reviews Clinical Oncology*, 9(5), pp. 259–267. Available at: https://doi.org/10.1038/nrclinonc.2011.199.

Tomsic, J. *et al.* (2013) 'Recurrent and founder mutations in the PMS2 gene.', *Clinical genetics*, 83(3), pp. 238–43. Available at: https://doi.org/10.1111/j.1399-0004.2012.01898.x.

Trabert, B. *et al.* (2014) 'Aspirin, Nonaspirin Nonsteroidal Anti-inflammatory Drug, and Acetaminophen Use and Risk of Invasive Epithelial Ovarian Cancer: A Pooled Analysis in the Ovarian Cancer Association Consortium', *JNCI Journal of the National Cancer Institute*, 106(2), pp. djt431–djt431. Available at: https://doi.org/10.1093/jnci/djt431.

Turturro, S.B. *et al.* (2016) 'Mutations in PIK3CA sensitize breast cancer cells to physiologic levels of aspirin', *Breast Cancer Research and Treatment*, 156(1), pp. 33–43. Available at: https://doi.org/10.1007/s10549-016-3729-8.

UK-TIA (1988) 'United Kingdom transient ischaemic attack (UK-TIA) aspirin trial: interim results. UK-TIA Study Group.', *British medical journal (Clinical research ed.)*, 296(6618), pp. 316–20.

Umar, A. *et al.* (2004) 'Revised Bethesda Guidelines for Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndrome) and Microsatellite Instability', *JNCI Journal of the National Cancer Institute*, 96(4), pp. 261–268. Available at: https://doi.org/10.1093/jnci/djh034.

Urakami, S. *et al.* (2018) 'Clinicopathological characteristics of patients with upper urinary tract urothelial cancer with loss of immunohistochemical expression of the DNA mismatch repair proteins in universal screening', *International Journal of Urology*, 25(2), pp. 151–156. Available at: https://doi.org/10.1111/iju.13481.

Vane, J.R. and Botting, R.M. (2003) 'The mechanism of action of aspirin', *Thrombosis Research*, 110(5–6), pp. 255–258. Available at: https://doi.org/10.1016/S0049-3848(03)00379-7.

Vasen, H.F. *et al.* (1999) 'New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC.', *Gastroenterology*, 116(6), pp. 1453–6. Available at: https://doi.org/10.1016/s0016-5085(99)70510-x.

Vasen, H.F.A. *et al.* (1991) 'The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC)', *Diseases of the Colon & Rectum*, 34(5), pp. 424–425. Available at: https://doi.org/10.1007/BF02053699.

Vasen, H.F.A. *et al.* (2013) 'Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts', *Gut*, 62(6), pp. 812–823. Available at: https://doi.org/10.1136/gutjnl-2012-304356.

Vaughn, C.P. *et al.* (2010) 'Clinical analysis of *PMS2* : mutation detection and avoidance of pseudogenes', *Human Mutation*, p. n/a-n/a. Available at: https://doi.org/10.1002/humu.21230.

Wahlberg, S.S. *et al.* (2002) 'Evaluation of microsatellite instability and immunohistochemistry for the prediction of germ-line MSH2 and MLH1 mutations in hereditary nonpolyposis colon cancer families.', *Cancer research*, 62(12), pp. 3485–92.

Wang, Dingzhi and DuBois, R.N. (2010) 'Eicosanoids and cancer', *Nature Reviews Cancer*, 10(3), pp. 181–193. Available at: https://doi.org/10.1038/nrc2809.

Wang, D and DuBois, R.N. (2010) 'The role of COX-2 in intestinal inflammation and colorectal cancer', *Oncogene*, 29(6), pp. 781–788. Available at: https://doi.org/10.1038/onc.2009.421.

Wang, Y. *et al.* (2017) 'Differences in Microsatellite Instability Profiles between Endometrioid and Colorectal Cancers: A Potential Cause for False-Negative Results?', *The Journal of molecular diagnostics : JMD*, 19(1), pp. 57–64. Available at: https://doi.org/10.1016/j.jmoldx.2016.07.008.

Watson, C.M. *et al.* (2021) 'Long-read nanopore sequencing enables accurate confirmation of a recurrent PMS2 insertion–deletion variant located in a region of complex genomic architecture', *Cancer Genetics*, 256–257, pp. 122–126. Available at: https://doi.org/10.1016/j.cancergen.2021.05.012.

Weisenberger, D.J. *et al.* (2006) 'CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer', *Nature Genetics*, 38(7), pp. 787–793. Available at: https://doi.org/10.1038/ng1834.

Wernstedt, A. *et al.* (2012) 'Improved multiplex ligation-dependent probe amplification analysis identifies a deleterious *PMS2* allele generated by recombination with crossover between *PMS2* and *PMS2CL*', *Genes, Chromosomes and Cancer*, 51(9), pp. 819–831. Available at: https://doi.org/10.1002/gcc.21966.

Williams, C.S. and DuBois, R.N. (1996) 'Prostaglandin endoperoxide synthase: why two isoforms?', *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 270(3), pp. G393–G400. Available at: https://doi.org/10.1152/ajpgi.1996.270.3.G393.

Williams, C.S., Smalley, W. and DuBois, R.N. (1997) 'Aspirin use and potential mechanisms for colorectal cancer prevention.', *Journal of Clinical Investigation*, 100(6), pp. 1325–1329. Available at: https://doi.org/10.1172/JCI119651.

Wimmer, K. *et al.* (2014) 'Diagnostic criteria for constitutional mismatch repair deficiency syndrome: suggestions of the European consortium "Care for CMMRD" (C4CMMRD)', *Journal of Medical Genetics*, 51(6), pp. 355–365. Available at: https://doi.org/10.1136/jmedgenet-2014-102284.

Win, A.K. *et al.* (2017) ‘Prevalence and Penetrance of Major Genes and Polygenes for Colorectal Cancer’, *Cancer Epidemiology, Biomarkers & Prevention*, 26(3), pp. 404–412. Available at: https://doi.org/10.1158/1055-9965.EPI-16-0693.

Wolf, A.I., Buchanan, A.H. and Farkas, L.M. (2013) ‘Historical review of Lynch syndrome’, *Journal of Coloproctology (Rio de Janeiro)*, 33(2), pp. 95–110. Available at: https://doi.org/10.1590/S2237-93632013000200010.

Wright, J.D. *et al.* (2016) ‘Utilization and Outcomes of Ovarian Conservation in Premenopausal Women With Endometrial Cancer’, *Obstetrics & Gynecology*, 127(1), pp. 101–108. Available at: https://doi.org/10.1097/AOG.0000000000001181.

Xu, X.-M. *et al.* (1999) ‘Suppression of inducible cyclooxygenase 2 gene transcription by aspirin and sodium salicylate’, *Proceedings of the National Academy of Sciences*, 96(9), pp. 5292–5297. Available at: https://doi.org/10.1073/pnas.96.9.5292.

Yamamoto, H. *et al.* (2001) ‘Genetic and clinical features of human pancreatic ductal adenocarcinomas with widespread microsatellite instability.’, *Cancer research*, 61(7), pp. 3139–44.

Young, L.E. *et al.* (2009) ‘The mRNA Binding Proteins HuR and Tristetraprolin Regulate Cyclooxygenase 2 Expression During Colon Carcinogenesis’, *Gastroenterology*, 136(5), pp. 1669–1679. Available at: https://doi.org/10.1053/j.gastro.2009.01.010.

Yu, Y. *et al.* (2010) ‘Cyclooxygenase-2–Dependent Prostacyclin Formation and Blood Pressure Homeostasis’, *Circulation Research*, 106(2), pp. 337–345. Available at: https://doi.org/10.1161/CIRCRESAHA.109.204529.

Yurgelun, M.B. *et al.* (2017) ‘Cancer Susceptibility Gene Mutations in Individuals With Colorectal Cancer’, *Journal of Clinical Oncology*, 35(10), pp. 1086–1095. Available at: https://doi.org/10.1200/JCO.2016.71.0012.

Zhang, L. (2008) ‘Immunohistochemistry versus Microsatellite Instability Testing for Screening Colorectal Cancer Patients at Risk for Hereditary Nonpolyposis Colorectal Cancer Syndrome’, *The Journal of Molecular Diagnostics*, 10(4), pp. 301–307. Available at: https://doi.org/10.2353/jmoldx.2008.080062.

Zheng, S.L. and Roddick, A.J. (2019) ‘Association of Aspirin Use for Primary Prevention With Cardiovascular Events and Bleeding Events’, *JAMA*, 321(3), p. 277. Available at: https://doi.org/10.1001/jama.2018.20578.

Zitvogel, L., Kepp, O. and Kroemer, G. (2011) ‘Immune parameters affecting the efficacy of chemotherapeutic regimens’, *Nature Reviews Clinical Oncology*, 8(3), pp. 151–160. Available at: https://doi.org/10.1038/nrclinonc.2010.223.