DNA mismatch repair gene variant classification: evaluating the utility of somatic mutations and mismatch repair deficient colonic crypts and endometrial glands

Romy Walker¹,²*, Khalid Mahmood¹,²,³, Julia Como¹,², Mark Clendenning¹,², Jihoon E. Joo¹,², Peter Georgeson¹,², Sharelle Joseland¹,², Susan G. Preston¹,², Bernard J. Pope¹,²,³, James M. Chan¹,², Rachel Austin⁴, Jasmina Bojadzieva⁵, Ainsley Campbell⁵, Emma Edwards⁶, Margaret Gleeson⁷, Annabel Goodwin⁸,⁹, Marion T. Harris¹⁰, Emilia Ip¹¹, Judy Kirk⁷, Julia Mansour¹², Helen Marfan⁴, Cassandra Nichols¹³, Nicholas Pachter¹³,¹⁴,¹⁵, Abiramy Ragunathan⁷, Allan Spigelman⁸,¹⁶,¹⁷, Rachel Susman⁴, Michael Christie¹⁸,¹⁹, Mark A. Jenkins²,²⁰, Rish K. Pai²¹, Christophe Rosty¹,²,⁲²,²³, Finlay A. Macrae²⁴,²⁵,²⁶, Ingrid M. Winship²⁴,²⁶†, Daniel D. Buchanan¹,²,²⁴† for the ANGELS study

¹Colorectal Oncogenomics Group, Department of Clinical Pathology, Victorian Comprehensive Cancer Centre, The University of Melbourne, Melbourne, VIC, Australia
²University of Melbourne Centre for Cancer Research, Victorian Comprehensive Cancer Centre, Melbourne, VIC, Australia
³Melbourne Bioinformatics, The University of Melbourne, Melbourne, VIC, Australia
⁴Genetic Health Queensland, Royal Brisbane and Women's Hospital, Brisbane, QLD, Australia
⁵Clinical Genetics Unit, Austin Health, Melbourne, VIC, Australia
⁶Familial Cancer Service, Westmead Hospital, Sydney, NSW, Australia
⁷Hunter Family Cancer Service, Newcastle, NSW, Australia
⁸Cancer Genetics Department, Royal Prince Alfred Hospital, Camperdown, NSW, Australia
⁹University of Sydney, Sydney, NSW, Australia
¹⁰Monash Health Familial Cancer Centre, Clayton, VIC, Australia
¹¹Cancer Genetics service, Liverpool Hospital, Liverpool, NSW, Australia
¹²Tasmanian Clinical Genetics Service, Royal Hobart Hospital, Hobart, TAS, Australia
¹³Genetic Services of Western Australia, King Edward Memorial Hospital, Perth, WA, Australia
¹⁴Medical School, University of Western Australia, Perth, WA, Australia
¹⁵School of Medicine, Curtin University, Perth, WA, Australia
¹⁶St Vincent's Cancer Genetics Unit, Sydney, NSW, Australia
¹⁷Surgical Profassorial Unit, UNSW Clinical School of Clinical Medicine, Sydney, NSW, Australia
¹⁸Department of Medicine at Royal Melbourne Hospital, Melbourne Medical School, The University of Melbourne, Melbourne, VIC, Australia
¹⁹Department of Pathology, The Royal Melbourne Hospital, Melbourne, VIC, Australia
²⁰Centre for Epidemiology and Biostatistics, School of Population and Global Health, The University of Melbourne, Melbourne, VIC, Australia
²¹Department of Laboratory Medicine and Pathology, Mayo Clinic Arizona, Scottsdale, AZ, USA
²²Envoi Specialist Pathologists, Brisbane, QLD, Australia

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Simple Summary: Lynch syndrome is caused by germline pathogenic variants in the DNA mismatch repair (MMR) genes predisposing carriers to colorectal and endometrial cancer. Genetic testing for Lynch syndrome, in the form of multigene panel testing, frequently identifies variants of uncertain clinical significance (VUS). These VUS have limited clinical actionability and create uncertainty for patients and clinicians regarding their risk of cancer. In this study, we tested carriers of germline VUS for features consistent with Lynch syndrome, namely 1) tumor microsatellite instability/MMR-deficiency, 2) the presence of a somatic second hit in the MMR gene harboring the VUS by tumor sequencing and 3) the presence of MMR-deficiency in normal colonic mucosa crypts or normal endometrial glands. Our findings showed that microsatellite instability/MMR-deficiency status and somatic second hits were consistent with MMR variant classifications as determined by the ACMG/InSiGHT framework. In addition to this, the presence of MMR-deficient crypts/glands were consistent with pathogenic variant classification.

Abstract: Germline pathogenic variants in the DNA mismatch repair (MMR) genes (Lynch syndrome) predispose to colorectal (CRC) and endometrial (EC) cancer. Lynch syndrome specific tumor features were evaluated for their ability to support the ACMG/InSiGHT framework in classifying variants of uncertain clinical significance (VUS) in the MMR genes. Twenty-eight CRC or EC tumors from 25 VUS carriers (6xMLH1, 9xMSH2, 6xMSH6, 4xPMS2), underwent targeted tumor sequencing for the presence of microsatellite instability/MMR-deficiency (MSI-H/dMMR) status and identification of a somatic MMR mutation (second hit). Immunohistochemical testing for the presence of dMMR crypts/glands in normal tissue was also performed. The ACMG/InSiGHT framework reclassified 7/25 (28%) VUS to likely pathogenic (LP), three (12%) to benign/likely benign, and 15 (60%) VUS remained unchanged. For the seven re-classified LP variants comprising nine tumors, tumor sequencing confirmed MSI-H/dMMR (8/9, 88.9%) and a second hit (7/9, 77.8%). Of these LP reclassified variants where normal tissue was available, the presence of a dMMR crypt/gland was found in 2/4 (50%). Furthermore, a dMMR endometrial gland in a carrier of an MSH2 exon 1-6 duplication provided further support for upgrade of this VUS to LP. Our study confirmed that identifying these Lynch syndrome features can improve MMR variant classification, enabling optimal clinical care.
Keywords: Lynch syndrome, DNA mismatch repair gene variant classification, DNA mismatch repair deficient crypts/glands, colorectal cancer, endometrial cancer, variant of uncertain significance, DNA mismatch repair gene somatic mutations
1. Introduction

Lynch syndrome is the most common hereditary cancer predisposition syndrome with an estimated carrier frequency in the population of 1 in 280 [1] and a prevalence of up to 5% in colorectal cancer (CRC) or endometrial cancer (EC) affected people [2–5]. Lynch syndrome is caused by germline pathogenic variants in one of the DNA mismatch repair (MMR) genes, MLH1, MSH2, MSH6, PMS2 [6] or by deletions in the 3’ end of the EPCAM gene leading to transcriptional silencing of MSH2 [7]. People with Lynch syndrome have an increased risk of not only CRC and EC but also of cancers of the ovaries, stomach, duodenum, bile duct, gall bladder, pancreas, urinary bladder, ureter, kidney, breast, prostate and brain tumors [8]. Current estimates of penetrance for CRC vary by gene and sex but on average approximately one-third to one-half of germline MMR pathogenic variant carriers will be diagnosed with CRC by the age of 70 [9,10]. Once identified, carriers of MMR gene pathogenic variants can be offered screening via colonoscopy with polypectomy and other opportunities to prevent cancer development or the ability to diagnose Lynch syndrome cancers at an early curable stage [11].

The current diagnostic approach to identify carriers of pathogenic MMR variants involves immunohistochemical (IHC) testing of CRC or EC tumors for loss of MMR protein expression followed by germline multigene panel testing of the MMR genes [12]. A recurring outcome from genetic tests is the identification of germline variants of uncertain clinical significance (VUS) in one of the MMR genes, reported to occur in 6% of the cases undergoing testing for Lynch syndrome [13]. Uncertainty regarding the pathogenicity of a VUS impacts clinical management, as carriers of pathogenic variants receive more intensive clinical care including screening such as colonoscopy, and choice/timing of possible risk reducing surgeries, than carriers of benign variants [14]. The American College of Medical Genetics and Genomics (ACMG) has developed standards and guidelines for the interpretation of sequence variants identified in Mendelian disorders [15]. Here, the recommendation is to classify variants into five categories (class 1–5) based on available evidence deriving from variant prevalence in the population, in silico effect prediction, functional assays and segregation data, amongst other data sources [15]. Comparably, the International Society for Gastrointestinal Hereditary Tumours (InSiGHT) working group determined InSiGHT criteria to aid with MMR variant classification (https://www.insight-group.org/criteria/, last accessed date: 31st of May 2023). In this Bayesian based analysis, variant pathogenicity probabilities derive from tumor characteristics and predetermined combinations of evidence types to predict the variant pathogenicity likelihood [16].

The tumor characteristics used in the InSiGHT criteria to guide MMR variant classification are features commonly observed in individuals diagnosed with Lynch syndrome, such as high levels of microsatellite instability (MSI-H) and MMR-deficiency (dMMR) [17], as determined by MSI polymerase chain reaction (MSI-PCR) and IHC assays, respectively. MSI-PCR and IHC each have their own limitations leading to false positive or false negative testing results [18,19] which impedes accurate classification of MMR variants. MSI-H/dMMR tumors develop in Lynch syndrome when one allele in an MMR gene becomes inactivated by a germline...
pathogenic variant while the other allele becomes inactivated by either a somatic mutation or by loss of heterozygosity (LOH). This acquisition of a somatic second hit, as described by the Knudson two-hit hypothesis [20], results in complete loss of MMR function. As next-generation sequencing (NGS) becomes more widely adopted for precision oncology and diagnostic purposes, the ability to accurately determine MSI-H/dMMR status [21,22] and identify MMR gene somatic mutations/LOH [23] using this methodology is becoming increasingly attractive as a streamlined approach to diagnosing Lynch syndrome. While the strength of both the ACMG guidelines and InSiGHT criteria are the ability to draw on multiple data sources, these may not always be accessible in clinical settings [16]. Therefore, despite important advances, classification of germline variants in the MMR genes remains challenging.

A novel finding in Lynch syndrome has been the identification of dMMR in morphologically normal colonic crypts [24–27] or endometrial glands [28,29]. dMMR crypts/glands are specific to people with Lynch syndrome, observed only in carriers and not in people with sporadic MSI-H/dMMR tumors [27]. In people with Lynch syndrome the acquisition of the somatic second hit in normal tissue results in biallelic inactivation of the MMR gene, which is evidenced by loss of MMR protein expression as detected by IHC [24,27], representing the initiation of tumorigenesis. Therefore, the presence of a dMMR crypt/gland is a strong indicator of a germline MMR pathogenic variant. To date, this unique characteristic has not yet been investigated for its potential application in MMR gene variant classification approaches. To be most inclusive of current variant classification parameters, we applied a combination of the ACMG guidelines integrated with the InSiGHT criteria, hereon referred to as the ACMG/InSiGHT framework, for classification of variants in the MMR genes. MMR gene variant classification has undergone substantial progress in the last decade through incorporation of features unique to the Lynch syndrome phenotype. With the evolution of NGS-based diagnostics, including for Lynch syndrome [30,31], the shift from tumor IHC to tumor NGS to determine MSI-H/dMMR status is gaining support. Additionally, evidence of the presence of dMMR crypts/glands has the potential to inform MMR gene variant classification within the ACMG/InSiGHT approach. Thus, in this study, we investigated the role of 1) NGS-based MSI-H/dMMR status using an additive feature combination approach as described previously [22], 2) the presence of a somatic second hit (single mutation or LOH) in the MMR gene harboring the VUS and 3) the presence of a dMMR crypt/gland in normal colonic or endometrial tissue, in classifying 25 MMR VUS and compared this to their classification status derived from the ACMG/InSiGHT framework. Determination of additional features or approaches to support MMR gene variant classification will improve the diagnosis of Lynch syndrome and the precision prevention of cancer in carriers.

2. Materials and Methods

2.1 Patient Cohort

Participants were men and women diagnosed with primary CRC or women diagnosed with EC (n=25) who were identified by clinical genetic testing to carry a germline MMR gene VUS as
defined by the clinical testing report and subsequently referred by one of the Family Cancer Clinics across Australia to the ANGELS study ("Applying Novel Genomic approaches to Early-onset and suspected Lynch Syndrome colorectal and endometrial cancers") between 2018 and 2022 [32] (test group, n=25 carriers who developed n=28 tumors, Table 1). Cancer-affected relatives were recruited where possible to investigate segregation of the VUS. The patient IDs used in this manuscript are not known to anyone outside the research group and cannot be used to identify study subjects.

<table>
<thead>
<tr>
<th>Family</th>
<th>Count</th>
<th>Gene Variant</th>
<th>Carrier Status</th>
<th>Tumors</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSH2</td>
<td>25</td>
<td>Leu223 Pro</td>
<td>Carrier - tumor 1</td>
<td>Skin (melaoma) 41-45</td>
</tr>
<tr>
<td>MSH6</td>
<td>4</td>
<td>Asp943 Tyr</td>
<td>Carrier - tumor 2</td>
<td>CRC 41-45</td>
</tr>
<tr>
<td>MLH1</td>
<td>2</td>
<td>Val180del</td>
<td>Carrier - tumor 2</td>
<td>CRC 31-35</td>
</tr>
<tr>
<td>PMS2</td>
<td>6</td>
<td>Glu81 Lys</td>
<td>Carrier - tumor 3</td>
<td>Brain Glioblastoma 51-55</td>
</tr>
<tr>
<td>MSH2</td>
<td>13</td>
<td>Lys110Gln</td>
<td>Carrier - tumor 1</td>
<td>CRC 76-80</td>
</tr>
<tr>
<td>MSH2</td>
<td>8</td>
<td>Arg621Leu</td>
<td>Carrier - tumor 2</td>
<td>CRC 76-80</td>
</tr>
<tr>
<td>MSH6</td>
<td>24</td>
<td>Tyr313Cys</td>
<td>Carrier - tumor 3</td>
<td>CRC 41-45</td>
</tr>
<tr>
<td>MLH1</td>
<td>16</td>
<td>Val180del</td>
<td>Carrier - tumor 2</td>
<td>CRC 31-35</td>
</tr>
<tr>
<td>PMS2</td>
<td>17</td>
<td>Glu81 Lys</td>
<td>Carrier - tumor 3</td>
<td>Brain Glioblastoma 51-55</td>
</tr>
<tr>
<td>MSH2</td>
<td>35</td>
<td>Lys110Gln</td>
<td>Carrier - tumor 1</td>
<td>CRC 76-80</td>
</tr>
<tr>
<td>MSH2</td>
<td>13</td>
<td>Arg621Leu</td>
<td>Carrier - tumor 2</td>
<td>CRC 76-80</td>
</tr>
<tr>
<td>MSH6</td>
<td>24</td>
<td>Tyr313Cys</td>
<td>Carrier - tumor 3</td>
<td>CRC 41-45</td>
</tr>
</tbody>
</table>

**Table 1.** List of carriers of a germline DNA mismatch repair variant of uncertain clinical significance and their cancer-affected status that were included in this study. Abbreviations: ID, identification number; F, female; M, male; VUS, variant of uncertain clinical significance; CRC, colorectal cancer; EC, endometrial cancer; IHC, immunohistochemistry; NA, not applicable; NT, not tested; UTR, untranslated region.

+ Indicates heterogeneous / patchy loss of DNA mismatch repair protein expression by IHC.
Family members (F_161) were only tested for the germline MLH1 c.539_541del p.(Val180del) variant and not for the concomitant MSH6 c.4068_4071dup p.(Lys1358Aspfs*2) variant. The parents (ID_376-2 and ID_376-3) from carrier ID_376 were tested for the germline MSH2 exon 1-6 duplication and were found to be wildtype. The sister (ID_058-3) and relative (ID_058-7) from carrier ID_058 were tested for the germline MSH2 exon 1-6 duplication and were found to be wildtype. Participant ID_149 carries a germline MSH2 exon 14-15 duplication and a concomitant germline POLE c.1708C>A p.(Leu570Met) variant. The mother (ID_149-2) from carrier ID_149 was tested for the familial germline MSH2 exon 14-15 duplication and POLE c.1708C>A p.(Leu570Met) variants, however, was found to only carry the MSH2 germline VUS.

A biopsy or resection tumor tissue specimen was collected from each participant. Normal colonic mucosa or normal endometrium tissue were collected where possible. Pedigree information was collected during clinical work-up and segregation of the MMR VUS in family members was performed via Sanger sequencing as part of this study. A group of CRC- or EC-affected MMR pathogenic variant carriers (n=19) or non-carriers of a germline MMR pathogenic variant or VUS (n=20) who were participants of the Australasian Colorectal Cancer Family Registry (ACCFR) were included as reference groups, where n=37 underwent targeted panel sequencing or whole exome sequencing as shown in Table S1. The molecular and phenotypic characterization of these individuals and their tumors from the ANGELS [22,32,33] and the ACCFR [2,34–36] have been described previously. All studies were approved by the Human Research Ethics Committees at The University of Melbourne (HREC#1750748) and hospitals governing participating family cancer clinics. Written informed consent was obtained from all participants.

2.2 Immunohistochemical Testing for DNA Mismatch Repair Protein Expression

For the 25 MMR VUS carriers, IHC of tumor tissue was derived from either the diagnostic pathology report or from testing performed by this study. For this study, the Ventana DISCOVERY ULTRA automated stainer (Ventana Medical Systems Inc., Oro Valley, United States) was used with anti-MLH1 (M1), anti-MSH2 (G219-1129), anti-MSH6 (SP93) mouse monoclonal and anti-PMS2 (A16-4) rabbit monoclonal primary antibodies (Roche Diagnostics, Basel, Switzerland) and tested on a 4µM tissue section. All staining protocols were performed following the manufacturer’s protocol (Roche Diagnostics).

To detect dMMR colonic crypts or endometrial glands, IHC was performed on 4µM sections from a tissue block containing non-tumor-adjacent normal colonic mucosa or endometrium from the resection margins. Twenty serial 4µM sections were cut to a depth of 80µM with the 1st, 10th and 19th slide stained for the MMR protein that was concordant with the MMR gene harboring the VUS. If no dMMR crypt/gland was identified, a further twenty 4µM sections were cut and screened. This process was repeated up to three times with a maximum of
3x80µM sections of tissue screened. When a dMMR crypt/gland was identified, the subsequent section (2nd, 11th or 20th slide) was stained for the unaffected MMR gene as a control for artefactual loss of expression. For example, if an MSH2-deficient crypt/gland was identified, the next slide was stained for MLH1 protein expression. All dMMR crypts/glands identified in this study were independently confirmed by two senior pathologists (CR and RP), with 100% concordance in classifying stained slides as positive or negative for dMMR crypts/glands. Six normal colonic mucosa and six normal endometrial tissue samples were available for a total of 12 VUS carriers.

2.3 Tumor MLH1 Methylation Testing Assays

Testing for MLH1 gene promoter methylation on formalin-fixed paraffin embedded (FFPE) tissue DNA was performed as previously described [22,33]. Briefly, two independent MLH1 methylation assays, namely MethyLight [2,34] and MS-HRM (methylation-sensitive high resolution melting assay) [37], were used to test the same tumor DNA sample alongside a set of DNA standards (0% - 100% methylation) and no-template (negative) controls. Bisulfite conversion of tumor DNA was performed using the EZ DNA Methylation-Lightning™ Kit (Zymo Research, Irvine, United States). For MethyLight, MLH1 methylation was quantitatively reported based on the percentage of methylated reference (PMR) calculations [34], where tumors with a PMR \( \geq 10\% \) were considered “positive” [2,34]. For MS-HRM, the MeltDoctor™ HRM Reagent Kit (Thermo Fisher Scientific, Massachusetts, United States) was used where tumors demonstrating \( \geq 5\% \) were considered MLH1 methylation “positive”.

2.4 Next-Generation Sequencing

All available FFPE tumor tissue DNA (n=28) and matched blood-derived DNA samples from 25 MMR VUS carriers (Table 2) underwent targeted multigene panel sequencing using the panel capture previously described [22,33]. This customized panel incorporated 297 genes, including the MMR and EPCAM genes, as well as other established hereditary CRC and EC genes and the BRAF p.V600E mutation (2.005 megabases). Library preparation was performed using the SureSelect™ Low Input Target Enrichment System from Agilent Technologies (Santa Clara, California, United States) using standard procedures. The median on-target coverage for the panel sequenced test tumors was 906 (interquartile range = 763 – 1099) for the tumor DNA and 154 (interquartile range = 128 – 172) for blood-derived DNA samples. Panel libraries were sequenced on an Illumina NovaSeq 6000 (San Diego, California, United States) comprising 150 base pair (bp) paired end reads performed at the Australian Genome Research Facility.
Table 2. Display of the data used to implement the ACMG/InSiGHT framework for final DNA mismatch repair variant of uncertain significance classification. Abbreviations: MSI, microsatellite instability; MSS, microsatellite stable; MMR, DNA mismatch repair; dMMR, DNA mismatch repair deficiency; pMMR, DNA mismatch repair proficiency; WT, wildtype; IHC, immunohistochemistry; MAPP, multi-variate analysis of protein polymorphisms; PP2, PolyPhen-2.1; VUS, variant of uncertain clinical significance; LB, likely benign; LP, likely pathogenic; NA, not applicable; NT, not tested.  
+ Indicates heterogeneous / patchy loss of DNA mismatch repair protein expression by IHC

1 Two unrelated families (F_058 and F_376) who carry the MSH2 exon 1-6 duplication were grouped together for the purposes of the ACMG/InSiGHT classification, however, at this stage without further investigation, it remains unknown if the breaking points of the variants identified in each family are in the same location

2 Multiple entries documented in Ambry Genetics (https://www.ambrygen.com/, last accessed date: 23rd of May 2023)
2.5 Bioinformatics Pipeline

Adapter sequences were trimmed from raw FASTQ files using trimmomatic (v.0.38) \[38\] and aligned to the GRCh37 human reference genome using Burrows-Wheeler-Aligner (v.0.7.12) to generate BAM files. Germline and somatic single nucleotide variants and somatic insertions / deletions (INDELs) were called using Strelka (v.2.9.2., Illumina, San Diego, California, United States) and Mutect2 (v.0.5) using the recommended workflows \[39,40\]. Tumor mutational signatures (TMS) were calculated using the pre-defined set of 18 small (1-50bp) insertions/deletions (ID) signatures as published on COSMIC (https://cancer.sanger.ac.uk/signatures/, last accessed date: 13th of January 2023, v.3.2) \[41\]. All variants were restricted to the panel capture region and filtered based on PASS variants called by both Strelka and Mutect2. All variants were further filtered on a minimum depth of 50bp for normal/blood and tumor samples with a minimum variant allele frequency of 10\% \[32\]. For variant calling, the following RefSeq transcripts were used (\textit{MLH1}: NM_000249.3, \textit{MSH2}: NM_000251.2, \textit{MSH6}: NM_000179.2 and \textit{PMS2}: NM_000535.5). The MMR genes were interrogated for somatic mutations, including single somatic mutations (e.g. missense, nonsense, insertion, deletion, frameshift variant type) and LOH, in the same gene as the germline MMR VUS. LOH across the MMR genes were called using LOHdeTerminator (v.0.6, https://github.com/supernifty/LOHdeTerminator). The pathogenicity of somatic MMR mutations were determined using the Varsome database \[42\] (https://varsome.com/, last accessed date: 13th of January 2023), which categorizes variants into the ACMG classification system. All likely pathogenic/pathogenic MMR mutations were manually confirmed in BAM files using the Integrative Genomics Viewer (v.2.3) \[43\].

2.6 Determination of Tumor Microsatellite Instability and Mismatch Repair Deficiency

Panel sequenced tumors were assessed for evidence of MSI-H/dMMR using: 1) four independent MSI detection tools, namely MSMuTect \[44\], MANTIS \[45\], MSIseq \[46\] and MSISensor \[47\], 2) INDEL count and 3) the combination of ID2 TMS with ID7 TMS (TMS ID2+ID7) \[32\] as described in Walker \textit{et al.} 2023 \[22\]. Overall tumor MSI-H/dMMR status was determined by combining these six features (using an additive feature combination approach), where a tumor with any 3 or more of these 6 features with positivity for dMMR was considered to be MSI-H/dMMR \[22\]. This approach has been shown to be the most robust across whole-exome sequencing and panel assays as well as across CRC and EC tumors, while presenting with the highest prediction accuracy to differentiate dMMR from pMMR (MMR-proficient) tumors \[22\].

2.7 Classifying MMR Variants Using a Combination of Existing Methodologies

We used a combination of the ACMG \[15\] and InSiGHT \[16\] criteria, hereon referred to as the ACMG/InSiGHT framework, for improved and contemporary variant classification. Briefly, the features assessed are displayed in Table 2, including:
• Rarity of MMR variant (the rarer a variant, the more likely the variant will not be present in healthy controls, with <1 in 50,000 alleles indicating MMR variant rarity in gnomAD using the non-cancer dataset);

• Incorporation of tumor characteristics generated by this study, including age of diagnosis, tumor NGS-derived MSI-H/dMMR status, tumor \textit{BRAF} V600E mutation status, tumor \textit{MLH1} methylation status and MMR IHC result;

• Prior probability scores calculated for missense variants using the \textit{in silico} prediction tools Multi-variate Analysis of Protein Polymorphisms [48] and PolyPhen-2.1 [49] (pre-computed prior probabilities with a score of >0.68 and ≤0.81 indicate variant pathogenicity as determined in https://hci-priors.hci.utah.edu/PRIORS/ (last accessed date: 18th of July 2023);

• Tumor characteristics, either generated from this study or available from external public data, for the same variant were combined to generate a tumor odds pathogenicity score [50,51];

• Evidence of functional effect on protein structure (e.g., ClinVar, https://www.ncbi.nlm.nih.gov/clinvar/, last accessed date: 1st of June 2023);

• Co-segregation of variant with disease phenotype with a combined Bayes Likelihood Ratio >18.7 in two or more families [16] (e.g., COsegregation v.2: https://fengbj-laboratory.org/cool2/manual.html, last accessed date: 1st of June 2023);

• Predicted splicing effect using SpliceAI (with a delta score of >0.2 indicating pathogenicity) (https://spliceailookup.broadinstitute.org/, last accessed date: 1st of June 2023) [52].

These parameters were cumulatively considered for final MMR variant classification and variants were categorized into the recommended five-tier ACMG classifications (class 5 – pathogenic (P), class 4 – likely pathogenic (LP), class 3 – variant of uncertain significance (VUS), class 2 – likely benign (LB), class 1 – benign (B)) [15]. The final ACMG/InSiGHT classification for each of the 25 MMR VUS was then assessed for concordance with the tumor NGS-derived MSI-H/dMMR status, somatic second hit and dMMR crypt/gland testing results.

3. Results

3.1 Characteristics of Patients with MMR VUS

An overview of the study design is shown in Figure 1. A total of 24 carrier families with 25 unique germline MMR VUS were included in the study comprising VUS in \textit{MLH1} (n=6), \textit{MSH2} (n=9), \textit{MSH6} (n=6) and \textit{PMS2} (n=4) with 14/25 (56%) VUS resulting in missense changes (Table 1, Table 2). Testing for segregation of the VUS in relatives identified an additional 18 carriers, where the cancer-affected status of each of the 42 germline VUS carriers (probands and relatives) included in the study are shown in Table 1. The tumor type and age at diagnosis (ranging from 15-20 and 81-85 years) for each of the carriers are listed in Table 1, where 16/42 (38.1%) carriers developed multiple tumors. The pedigree for each of the VUS carrying families...
can be requested from the corresponding author upon reasonable request. For 25/42 (59.5%) VUS carriers in this study, we tested one or more tumors (Table 1). The pattern of loss of MMR protein expression by IHC was concordant with the MMR gene harboring the VUS in 13/25 (52%) of the cases, discordant in 8/25 (32%) of the cases, while for a further four carriers (16%), no loss of MMR protein expression by IHC was reported (pMMR) (Table 1, Table 2).

Information collected for current variant classification approaches (ACMG/InSiGHT criteria)
- Variant frequency
- \(BRAF\) V600E mutation status
- \(MLH1\) methylation status
- MMR IHC pattern
- Prior probabilities
- Functional data
- Segregation odds ratio
- \textit{In-silico} splicing predictions

Assessment of two screening assays for their complimentary potential to existing variant classification methodologies

Figure 1. Overview of study design. Schema presenting the study inclusion criteria, the breakdown of the germline MMR VUS distribution and the testing assays applied.

Abbreviations: CRC, colorectal cancer; EC, endometrial cancer; MMR, DNA mismatch repair; VUS, variant of uncertain clinical significance; IHC, immunohistochemistry; MSI, microsatellite instability; MSI-H, high levels of microsatellite instability; MSS, microsatellite stable; dMMR, DNA mismatch repair deficiency; pMMR, DNA mismatch repair proficiency; ACMG, American College of Medical Genetics and Genomics; InSiGHT, International Society for Gastrointestinal Hereditary Tumours.

3.2 Variant Classification Using the ACMG/InSiGHT Framework
The re-classification of the 25 MMR VUS based on the ACMG/InSiGHT framework (see Methods above) are shown in Table 2. A total of 10/25 (40%) VUS were reclassified, with seven
VUS (28%) reclassified as likely pathogenic (class 4), all of which showed a concordant pattern of MMR protein loss of expression by IHC. Two out of 25 (8%) were reclassified as likely benign (class 2) and one VUS was reclassified as benign (class 1), where each of the three LB/B variants had a pattern of MMR protein loss that was discordant with the MMR gene harboring the variant (Table 2). None of the VUS were categorized as pathogenic (class 5). For the remaining 15 VUS, the additional information provided by this study did not change their classification as a class 3 variant. Of these, six VUS were concordant with the observed IHC pattern of loss, five VUS were discordant to the observed IHC pattern of loss, with four VUS displaying no MMR protein loss by IHC (pMMR) (Table 2).

3.3 Determining Microsatellite Instability / DNA Mismatch Repair Deficiency using Tumor Sequencing Data

To assess whether tumor features and somatic profiles generated using tumor panel sequencing could inform MMR VUS classification, tumor features associated with Lynch syndrome were assessed on 28 tumors collected from the 25 VUS carriers in this study (Table 1), and compared with a reference group of dMMR tumors from known germline MMR pathogenic variant carriers (n=16) and pMMR tumors from non-MMR carriers (n=18) that previously underwent tumor panel sequencing (Table S1). The aim was to determine tumor MSI/MMR status from NGS by applying the previously described additive feature combination approach [22] and compare this with the observed MMR IHC status. For the reference group of tumors, the MSI/MMR status from tumor sequencing was 100% concordant with the MMR IHC dMMR or pMMR result. As expected for the Lynch syndrome tumors, the pattern of MMR protein loss was concordant with the MMR gene harboring the germline pathogenic variant and all were MSI-H/dMMR by tumor sequencing (Table S1, Figure S1A). In the test group of 28 tumors from 25 VUS carriers, 25/28 were classified as MSI-H/dMMR from tumor sequencing, of which 23/25 (92%) were dMMR by MMR IHC (Table 3, Figure S1B). Three tumors were classified as MSS/pMMR from tumor sequencing with two (66%) of these also confirmed to be pMMR by MMR IHC (Table 3, Figure S1B). All 7 of the VUS reclassified by ACMG/InSiGHT framework to LP had at least one tumor that was MSI-H/dMMR (Table 3). There were an additional 8 tumors from 7 VUS carriers that were not reclassified by ACMG/InSiGHT framework that were MSI-H/dMMR from tumor sequencing and demonstrated loss of the MMR protein/s concordant with the MMR gene harboring the VUS (Table 3).
Table 3. Overview of targeted tumor sequencing results from the test group including sequenced family members. Abbreviations: ID, identification number; CRC, colorectal cancer; EC, endometrial cancer; LP, likely pathogenic; VUS, variant of uncertain clinical significance; LB, likely benign; B, benign; IHC, immunohistochemistry; MMR, DNA mismatch repair; MSI, microsatellite instability; MSI-H, high levels of microsatellite stability; MSS, microsatellite stable; dMMR, DNA mismatch repair deficient; pMMR, DNA mismatch repair proficient; LOH, loss of heterozygosity; mut., single somatic mutation; NT, not tested.

+ Indicates heterogeneous / patchy loss of DNA mismatch repair protein expression by IHC

1 Participant (ID_161) developed a single endometrial cancer showing loss of MLH1/PMS2 but carried two VUS; one in MLH1 and one in MSH6

2 Participant (ID_352) carried a PMS2 VUS but developed two different CRCs; one with loss of MSH2/MSH6 protein expression and one with no loss of MMR protein expression (pMMR)

3.4 Determining Somatic MMR Gene Second Hit using Tumor Sequencing Data

Following the “two-hit” model for Lynch syndrome, a somatic second hit would be expected in the same gene as the gene carrying the germline variant. We aimed to identify the presence of a somatic MMR mutation as the second hit in the gene carrying the germline MMR VUS. For the reference group of dMMR Lynch syndrome tumors, 13 of the 16 (81.3%) harbored a detectable somatic second hit in the same MMR gene harboring the germline pathogenic variant, which...
included single somatic mutations (6/13, 46%) and LOH (7/13, 53.8%) (Table S1, Figure S2A). Of the 18 reference pMMR non-MMR carrier tumors, only 2/18 (11.1%) harbored a somatic mutation in one of the four MMR genes, both of which were LOH events (Table S1, Figure S2A). The findings from these reference tumors highlights the enrichment of somatic MMR mutations as second hits in Lynch syndrome related CRCs and ECs. Similarly, in the test group, for 86.7% of the cases (13/15), a second hit could be identified where the MMR VUS was concordant to the IHC pattern of loss. The second hit was more commonly a single somatic mutation (Figure S2B). Equivalent to the reference group, for cases with no MMR loss of protein expression by IHC, the second hit was exclusively of LOH mutation type (Figure S2B).

For 7/9 (77.8%) tumors from the 7 VUS reclassified by ACMG/InSiGHT framework to LP, a second hit was identified (Table 3). There were two exceptions. The first was an EC diagnosed at 55-60 years from person ID_315-2 who carried the MSH2 c.1862G>T p.(Arg621Leu) variant but the tumor showed loss of MLH1/PMS2+ expression related to tumor MLH1 promoter methylation. The sister (ID_315) was also a carrier and developed an EC at 61-65 years, which demonstrated loss of MSH2/MSH6 expression and a somatic second hit in MSH2. The second exception was an EC diagnosed at 51-55 years from ID_156 who carried the MSH6 c.3556+5_3556+9delins variant that was MSI-H/dMMR from tumor sequencing, showed solitary loss of MSH6 expression but no somatic mutation in MSH6, however, a CRC diagnosed at 61-65 years in ID_156 also showed solitary loss of MSH6 expression with a second hit in MSH6.

There were an additional 8 tumors from 7 VUS carriers that were not reclassified by ACMG/InSiGHT framework, but did demonstrate a second hit (Table 3), including ID_176 who carried the MLH1 c.1594G>C p.(Gly532Arg) variant and developed MLH1/PMS2 deficient CRC and EC, in the absence of MLH1 methylation, where a second hit was observed in both tumors. Two VUS carriers, ID_170 (MLH1 c.-117G>T p.? ) and ID_111 (MSH6 c.1153_1155del p.(Arg385del)), demonstrated MSI-H/dMMR by tumor sequencing and a second hit in their respective CRCs, however, both CRCs were pMMR by IHC suggesting a false negative MMR IHC result. An additional VUS carrier, ID_376 (MSH2 exon 1-6 duplication) whose CRC showed loss of MSH2/MSH6 by IHC, MSI-H/dMMR by tumor sequencing, did not harbor a second hit in the MSH2 gene (Table 3).

There were a further 8 VUS carriers whose tumors were all MSI-H/dMMR by tumor sequencing but their pattern of MMR protein loss by IHC indicated a different MMR gene was defective to the one harboring the VUS (Table 3). Tumor sequencing revealed two somatic MMR mutations (also known as “double somatics”), which were likely responsible for the pattern of MMR protein loss by IHC in these 8 carriers. For two of these VUS (ID_263: MSH6 c.*85T>A and ID_202: MSH2 c.138C>G p.(His46Gln)) a second hit was observed, however, the ACMG/InSiGHT framework reclassified these variants as benign and likely benign, respectively (Table 3).
The specificity of the somatic MMR mutations to the MMR gene harboring the VUS was assessed. For the reference group of dMMR Lynch syndrome tumors, a somatic second hit was identified in 100% of \textit{MLH1} (n=4), in 70% of \textit{MSH2} (n=10), in 100% of \textit{MSH6} (n=1) and 100% of \textit{PMS2} (n=1) germline pathogenic variant carriers but somatic MMR mutations in the other MMR genes were rarely observed (Table S1). For the reference pMMR non-Lynch syndrome tumors, the presence of any MMR somatic mutation was found in only 11.1% (2/18) of the cases screened (Table S1). In the test group, for the MMR VUS categorized as LP by the ACMG/InSiGHT framework, only 3/9 (33.3%) tumors presented with $\geq$1 somatic event in the gene that did not harbor the germline VUS (Table 3, Figure 2). For the three cases where the germline VUS was classified as LB/B, a different molecular mechanism, e.g., double somatic MMR mutations (n=2) or a concomitant germline variant plus somatic second hit in the same gene (n=1, ID_161), is the likely cause for the observed tumor dMMR (Table 3, Figure 2).
**Figure 2.** Overview of the number and type of somatic events by the 25 germline DNA mismatch repair variants of uncertain significance screened in this study. Red highlighted text indicates duplicate entries per row. **Abbreviations:** VUS, variant of uncertain significance; germ, germline variant; som, somatic mutation; ACMG, American College of Medical Genetics and Genomics; InSiGHT, International Society for Gastrointestinal Hereditary Tumours; MMR, DNA mismatch repair; dMMR, DNA mismatch repair deficient; pMMR, DNA mismatch repair proficient; IHC, immunohistochemistry; MSI, microsatellite instability; MSI-H, high levels of microsatellite instability; MSS, microsatellite stable; LOH, loss of heterozygosity; LP, likely pathogenic; LB, likely benign; B, benign.

\footnote{\textsuperscript{1} Determination of the MSI/MMR status using the additive feature combination approach as previously described in Walker et al., 2023}
3.5 Detection of DNA Mismatch Repair Deficient Crypts / Glands in Normal Tissue

To establish the protocol, screening of normal colonic mucosa for dMMR crypts was performed for three pathogenic variant carriers from the reference group with available tissue. Two crypts from two different carriers demonstrated a loss of expression of the MLH1 protein which was concordant with the germline pathogenic variant in MLH1 (Figure 3A, Table S1, Table 4). The screening did not identify a dMMR crypt in the third reference case (Ref_411) from 2x80µM tissue screening, after which the tissue was depleted. In addition to finding a dMMR crypt in the normal colonic mucosa of MLH1 pathogenic variant carrier Ref_029, a dMMR crypt was identified in the normal colonic mucosa of their relative (Ref_029-2) who was also a carrier of the family MLH1 pathogenic variant (Table 4). No dMMR crypts were identified from 3x80µM tissue screening of the normal colonic mucosa from two CRC-affected people who did not carry a germline MMR pathogenic variant (Table 4).
**Figure 3.** Detection of DNA mismatch repair deficient crypts or glands in (A) the reference and (B) the test group with three cases identified each. **Abbreviations:** MMR, DNA mismatch repair; dMMR, DNA mismatch repair deficient; IHC, immunohistochemistry.
<table>
<thead>
<tr>
<th>#</th>
<th>Carrier Type</th>
<th>Carrier ID</th>
<th>Tissue</th>
<th>MMR Mutation</th>
<th>Gene</th>
<th>MSH2/MSH6 Exon</th>
<th>MSHH/MMR by Pathogenic Exon</th>
<th>Presence of a Second Hit</th>
<th>Presence of a dMMR Crypt/Gland</th>
<th>Presence of a Second Hit</th>
<th>Final Tumor Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Reference</td>
<td>Ref_229</td>
<td>CRC</td>
<td>MSH2 +5 3556+5_3556+8delins</td>
<td>MLH1</td>
<td>3:1862G&gt;T p.(Arg621Leu)/2</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>2-80µM</td>
<td>dMMR - 1st 2nd hit (LS)</td>
</tr>
<tr>
<td>2</td>
<td>Reference</td>
<td>Ref_2025</td>
<td>CRC</td>
<td>MSH2 +5 3556+5_3556+8delins</td>
<td>MLH1</td>
<td>3:1862G&gt;T p.(Arg621Leu)/2</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>2-80µM</td>
<td>dMMR - 1st 2nd hit (LS)</td>
</tr>
<tr>
<td>3</td>
<td>Reference</td>
<td>Ref_315</td>
<td>CRC</td>
<td>MSH2 +5 3556+5_3556+8delins</td>
<td>MLH1</td>
<td>3:1862G&gt;T p.(Arg621Leu)/2</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>2-80µM</td>
<td>dMMR - 1st 2nd hit (LS)</td>
</tr>
<tr>
<td>4</td>
<td>Reference</td>
<td>Ref_376</td>
<td>CRC</td>
<td>MSH2 +5 3556+5_3556+8delins</td>
<td>MLH1</td>
<td>3:1862G&gt;T p.(Arg621Leu)/2</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>2-80µM</td>
<td>dMMR - 1st 2nd hit (LS)</td>
</tr>
<tr>
<td>5</td>
<td>Reference</td>
<td>Ref_051</td>
<td>CRC</td>
<td>MSH2 +5 3556+5_3556+8delins</td>
<td>MLH1</td>
<td>3:1862G&gt;T p.(Arg621Leu)/2</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>2-80µM</td>
<td>dMMR - 1st 2nd hit (LS)</td>
</tr>
<tr>
<td>6</td>
<td>Reference</td>
<td>Ref_156</td>
<td>CRC</td>
<td>MSH2 +5 3556+5_3556+8delins</td>
<td>MLH1</td>
<td>3:1862G&gt;T p.(Arg621Leu)/2</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>2-80µM</td>
<td>dMMR - 1st 2nd hit (LS)</td>
</tr>
<tr>
<td>7</td>
<td>Reference</td>
<td>Ref_153</td>
<td>CRC</td>
<td>MSH2 +5 3556+5_3556+8delins</td>
<td>MLH1</td>
<td>3:1862G&gt;T p.(Arg621Leu)/2</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>2-80µM</td>
<td>dMMR - 1st 2nd hit (LS)</td>
</tr>
<tr>
<td>8</td>
<td>Reference</td>
<td>Ref_154</td>
<td>CRC</td>
<td>MSH2 +5 3556+5_3556+8delins</td>
<td>MLH1</td>
<td>3:1862G&gt;T p.(Arg621Leu)/2</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>2-80µM</td>
<td>dMMR - 1st 2nd hit (LS)</td>
</tr>
<tr>
<td>9</td>
<td>Reference</td>
<td>Ref_155</td>
<td>CRC</td>
<td>MSH2 +5 3556+5_3556+8delins</td>
<td>MLH1</td>
<td>3:1862G&gt;T p.(Arg621Leu)/2</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>2-80µM</td>
<td>dMMR - 1st 2nd hit (LS)</td>
</tr>
<tr>
<td>10</td>
<td>Reference</td>
<td>Ref_156</td>
<td>CRC</td>
<td>MSH2 +5 3556+5_3556+8delins</td>
<td>MLH1</td>
<td>3:1862G&gt;T p.(Arg621Leu)/2</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>2-80µM</td>
<td>dMMR - 1st 2nd hit (LS)</td>
</tr>
</tbody>
</table>

Table 4. Overview of normal tissue screening for DNA mismatch repair deficient crypts and glands in the reference and test groups. Abbreviations: ID, identification number; CRC, colorectal cancer; IHC, immunohistochemistry; P, pathogenic; LP, likely pathogenic; VUS, variant of uncertain significance; LB, likely benign; MMR, DNA mismatch repair; dMMR, DNA mismatch repair deficiency; pMMR, DNA mismatch repair proficiency; MSI-H, high levels of microsatellite stability; MSS, microsatellite stable; NA, not applicable; NT, not tested; LS, Lynch syndrome; MLH1me, MLH1 gene promoter hypermethylation.

+ Indicates heterogeneous / patchy loss of DNA mismatch repair protein expression by IHC

1 Relative of Ref_029 who is also carrier of the family MLH1 pathogenic variant was identified to have a dMMR crypt in normal colonic mucosa, but their CRC did not undergo tumor sequencing

2 Block was depleted after screening of 2x80µM of normal tissue

3 This sample did not undergo next-generation sequencing

4 Only received slides

For 12/25 (48%) of MMR VUS carriers, normal tissue specimens (6x normal colonic and 6x normal endometrial tissue) were available for dMMR crypt/gland screening. A single case (ID_143) was a technical failure for PMS2 IHC staining due to poor tissue fixation of the normal tissue. In 7/11 (63.6%) of the remaining cases, normal tissue blocks were available for screening while for 4/11 (36.4%) of the cases, only 4µM normal tissue sections on slides were available for screening. A total of three carriers (27.3%) had a dMMR crypt/gland identified out of 11 carriers tested (Table 4, Figure 3B). Two of these were in normal endometrium tissue with the remaining dMMR crypt identified in normal colonic mucosa. Out of four cases that were reclassified as likely pathogenic based on the ACMG/InSiGHT criteria and where normal tissue was available for testing, 2/4 (ID_315: MSH2 c.1862G>T p.(Arg621Leu) and ID_156: MSH6 c.3556+5_3556+8delins) had a dMMR gland and dMMR crypt identified, respectively (Table 4). A dMMR endometrial gland was identified in the carrier of MSH2 exon 1-6 duplication (ID_058), where the tumor also demonstrated MSI-H/dMMR by tumor sequencing, a somatic second hit in MSH2 and showed loss of MSH2/MSH6 expression by IHC, however, the
ACMG/InSiGHT framework did not result in a reclassification of the VUS (Table 4). The pathogenic criterion (PVS1) from the ACMG guidelines could not be applied for a predicted loss of function as the location of the partial gene duplication was unknown, making it uncertain if nonsense mediated mRNA decay would take place [53].

4. Discussion

In this study, tumor and non-malignant tissue features associated with germline pathogenic MMR variant carriers were investigated to determine their utility to aid MMR variant classification. Our findings from the investigation of 28 tumors from 25 VUS carriers showed that tumor MSI-H/dMMR status, determined by tumor sequencing and an additive feature combination approach [22], agreed with variant LP/P classification (Figure 4). We found MSI-H/dMMR status by tumor sequencing was 100% concordant with dMMR status by IHC in both our reference group of dMMR Lynch syndrome tumors (Figure S3) and in the tumors from 7 VUS carriers that were reclassified to LP by the ACMG/InSiGHT framework (Table 3), while the reference group of pMMR non-MMR carrier tumors were MSS/pMMR by tumor sequencing (Figure S3). Furthermore, the identification of a somatic second hit was also consistent with variant LP/P classification. A second hit was observed in 81.3% of the reference group of dMMR Lynch syndrome tumors (Figure S3) and 77.8% of the tumors from VUS reclassified to LP (Table 3) in contrast to only 11.1% of tumors from the reference group of pMMR non-MMR carriers having a somatic MMR mutation (Figure S3). In light of these findings, a further 7 VUS, that could not be reclassified by the ACMG/InSiGHT framework demonstrated tumors with MSI-H/dMMR and a second hit, suggesting that these 7 VUS could be upgraded to an LP classification (Table 3). Screening for the presence of a dMMR crypt/gland also showed potential for clinical utility for LP/P variant classification. In addition to the three known pathogenic variant carriers from the reference group, three additional VUS carriers were found to have a dMMR crypt/gland, where in two of these the ACMG/InSiGHT framework reclassified the VUS to LP (Table 4). The remaining VUS case with a dMMR endometrial gland was identified in the carrier of MSH2 exon 1-6 duplication (ID_058), and together with the tumor also demonstrating MSI-H/dMMR by tumor sequencing and a somatic second hit, is supportive of an LP classification for this variant (Table 4). Therefore, the application of tumor sequencing for MSI/dMMR status and presence of a second hit together with testing for dMMR crypts/glands is likely to improve MMR variant classification.

The use of tumor characteristics to support MMR variant classification continues to evolve. MMR IHC and testing for somatic BRAF p.V600E mutation and/or MLH1 promoter methylation are commonly used to identify or exclude CRCs and ECs for germline MMR gene testing and thus these features have been incorporated into existing MMR VUS classification approaches [15,16]. The investigation of somatic features derived from tumor next generation sequencing to support MMR variant classification has important potential. The determination of MSI/dMMR by NGS has resulted in the development of multiple accurate MSI calling tools [44–47], where our previous study compared these tools and other dMMR-related tumor features to
ultimately determine the combination of features (an additive feature combination approach) which demonstrated the highest accuracy across tumor types [22]. In this study, we showed the determination of MSI-H/dMMR by our additive feature combination approach was highly concordant with MMR IHC results, potentially identifying two VUS with false negative pMMR IHC results (ID_170 and ID_111) (Table 3). Despite this promising finding, MMR IHC currently has a diagnostic advantage over MSI-H/dMMR determination by tumor sequencing of indicating which MMR gene is likely to be the defective gene. This was evident for 8 VUS in this study where the pattern of MMR protein loss by IHC was discordant with the MMR gene harboring the VUS suggesting the VUS was not likely to be pathogenic. In these 8 cases, tumor sequencing revealed that two somatic MMR mutations were the cause for the pattern of MMR protein loss rather than the VUS. Future incorporation of tumor next generation sequencing into MMR variant classification framework should derive MSI-H/dMMR status together with identification of somatic mutations across the four MMR genes to elucidate either germline or biallelic somatic MMR gene inactivation as the likely cause for MSI-H/dMMR status.
**Figure 4.** Sunburst diagram displaying the prevalence of the three Lynch syndrome associated features in the test group. The diagram incorporates the tumor sequencing and dMMR crypt/gland screening for determining pathogenicity against the ACMG/InSiGHT framework for MMR VUS included in the test group. Abbreviations: MMR, DNA mismatch repair; dMMR, DNA mismatch repair deficient; pMMR, DNA mismatch repair proficient; MSI, microsatellite instability; MSI-H, high levels of microsatellite instability; MSS, microsatellite stable; VUS, variant of uncertain clinical significance; LP, likely pathogenic; LB, likely benign; B, benign.

A recent study has investigated the benefit of identifying the somatic second hit for variant classification. Scott *et al.* (2022) showed that somatic second hit mutations in *MSH2* were significantly more common in tumors from *MSH2* missense variant carriers that had multiplexed analysis of variant effect (MAVE) data, indicating, the germline variant was functionally disruptive (i.e., pathogenic variant) when compared with tumors from *MSH2* missense variant carriers with MAVE scores indicating the germline variant was functionally normal (i.e., benign variant) [53]. This supports the observations from this study where a somatic second hit was more prevalent in both known pathogenic variant carriers as well as VUS that were reclassified to LP but rare in pMMR tumors.

A study performed by Shirts *et al.* (2018) demonstrated that tumor mutations in the MMR genes can support both pathogenic and benign variant classification by identifying somatic driver mutations compared with passenger mutations in patients with unexplained dMMR (i.e., suspected Lynch syndrome or Lynch-like syndrome) [54]. Furthermore, the authors propose that the cumulative evidence from independent mutations identified from sequencing unexplained dMMR tumors will ultimately classify more germline MMR gene variants. Given the rarity of some individual constitutional MMR gene variants, the observation of these same variants as somatic mutations in multiple dMMR tumors may expedite their classification. The detection of a somatic second hit, as we have shown in this study, as well as the work described by Shirts and colleagues, demonstrates that the detection of somatic MMR mutations in tumors, with confirmed MSI-H/dMMR status, can support MMR variant classification and warrants modifications of the ACMG/InSiGHT MMR variant classification guidelines to incorporate the characteristics of somatic mutations from tumor sequencing data.

An important finding from this study was the identification of double somatic MMR mutations in an MMR gene that was not the gene harboring the VUS. Double somatic MMR mutations are a recognized cause of somatic biallelic MMR gene inactivation that can lead to tumor MSI-H/dMMR phenotype [23,33,55,56]. The additional information provided by the pattern of MMR protein loss by IHC was supportive that the MSI-H/dMMR tumor phenotype was caused by two somatic MMR mutations and not related to the VUS. Two of these VUS were reclassified as LB/B, supporting the somatic mutation data but in MSI-H-dMMR tumors (Table 3). A caveat to these findings was the presence of two somatic *MSH2* mutations in the CRC from person ID_138 carrying the *MSH2* c.328A>C p.(Lys110Gln) VUS (Table 3). One of these two somatic *MSH2* mutations may represent the second hit to the germline VUS, however, the two
somatic MSH2 mutations may represent somatic biallelic inactivation (Table 3). Of interest, the MAVE data for this MSH2 missense VUS suggests it is likely benign [57] supporting a “double somatic” rather than a germline cause of MSI-H/dMMR for this tumor. Consideration of the number of somatic MMR mutations identified together with MMR IHC findings will help to interpret tumor sequencing data for MMR variant classification.

There were 4 tumors from 4 carriers where no loss of MMR protein expression was observed by IHC (Table 3). Two tumors (CRC_240 and CRC1_352) were MSS/pMMR by tumor sequencing supporting IHC result. The other two tumors (CRC_170 and CRC_111) were MSI-H/dMMR by NGS and showed a somatic second hit in the gene with the VUS (both LOH events), which may suggest false negative MMR IHC result.

The presence of a dMMR crypt or gland is a strong predictor for a variant being pathogenic given its specificity for Lynch syndrome [24–29]. In this study, a single endometrial gland showed loss of MSH2 expression in the patient harboring the MSH2 exon 1-6 duplication (ID_058) which would support this variant being pathogenic (Table 4). The absence of detectable dMMR crypts/glands does not necessarily indicate a LB/B classification and could simply reflect insufficient tissue was screened. The exact prevalence of dMMR crypts/glands across normal tissues still needs to be assessed in ancillary studies, however, Kloor et al. (2012) have indicated the detection of dMMR crypts in 1cm² of colonic mucosa in Lynch syndrome patients [24]. The feasibility in terms of the amount of biopsy needed to get at least 1cm² and cost-effectiveness of screening for dMMR crypts/glands in clinical setting needs to be determined but may offer an alternate approach to reclassify an MMR variant particularly when evidence from the existing ACMG/InSiGHT framework is insufficient.

A strength of the study was the comparison of data from the existing gold standard MMR variant classification framework to the application of novel features, particularly those derived from NGS which is increasing in clinical diagnostics. The detection of MSI-H/dMMR and a second hit from tumor sequencing is unlikely to be influenced by the type of variant. Further studies are needed to determine if the detection of dMMR crypts/glands is likely to be influenced by variant type. Furthermore, the implementation of screening for dMMR crypts or glands would be based on established MMR protein antibodies and immunohistochemical protocols and, therefore, potentially more applicable to a broader spectrum of laboratories once the tissue is available. A further demonstrated strength by this study was the ability to detect dMMR crypts/glands on FFPE archival tissue that was up to 20 years old (e.g., reference group tissue) with only a single case (ID_143) failing testing.

This study has several limitations. An important caveat for interpreting the presence of a dMMR crypt/gland for VUS classification is the concept that another undetected pathogenic variant underlies the dMMR crypt/gland rather than the VUS. Therefore, interpretation of the presence of a dMMR crypt/gland should be considered alone but together with additional information used to classify MMR variants. Another limitation of the study was access to normal tissue as we were only able to acquire normal tissue specimen for half of the cases (48%, 12/25). Broader recognition that screening for dMMR crypts/glands has utility for variant classification
may encourage better collection and access to normal tissue. A single Lynch syndrome tumor phenocopy was identified in the case of ID_315-2, where the tumor was positive for MLH1 methylation despite the person carrying the MSH2 LP variant. Although phenocopies in Lynch syndrome are rare, the interpretation of tumor data for MMR variant classification needs detailed examination. Lastly, it is possible that somatic second hits were missed in some of the tumors. This was evident for the Lynch syndrome reference tumors where a second hit was identified in only 81.3% of sequenced tumors. Challenges in identifying more complex/cryptic variants from capture-based sequencing data or the possibility the second hit is an intronic variant not targeted by the capture may explain the missing second hits. These challenges may underlie second hit detection in the VUS cases tested in this study where, for example, the EC from person ID_156 who carried the MSH6 c.3556+5_3556+9delins variant and CRC from person ID_376 who carried the MSH2 exon 1-6 duplication did not identify a second hit despite the other cumulative evidence suggesting these variants are likely pathogenic. Lastly, complementary data could be gained from functional assays such as RT-PCR or minigene constructs to provide further functional evidence to support variant classification.

5. Conclusions

This study evaluated novel approaches to classify MMR variants, providing support for their potential incorporation into current variant classification guidelines as additional independent lines of evidence to aid MMR variant classification. Currently, somatic MMR mutation data is not used in MMR gene variant classification frameworks, but this study and other studies provide support for information gained from sequencing of dMMR tumors. Although the presence of a somatic second hit was concordant with LP/P variant classification, the knowledge that the presence of two somatic MMR gene mutations (double somatics) can also result in MSI-H/dMMR tumor phenotype needs to be acknowledged when interpreting tumor sequencing findings for variant classification. Furthermore, somatic MMR mutation data from tumor sequencing needs to be considered in conjunction with confirmation the tumor is MSI-H/dMMR. The presence of a dMMR crypt/gland in normal colonic or endometrial tissue represents a novel approach to guide LP/P MMR variant classification. The identification of germline MMR VUS prior to surgery may facilitate preservation of more normal tissue for testing but the application of dMMR crypt/gland detection using normal colonic biopsies from colonoscopy in unaffected VUS carriers needs further investigation. Our findings have shown the potential utility of tumor sequencing to determine both MSI/MMR status and presence of single point mutation/LOH as a somatic second hit, and with assessment of normal tissue for the presence of dMMR crypts/glands for improving MMR variant classification and warrants consideration for inclusion in the ACMG/InSiGHT framework.

Supplementary Materials: Figure S1: Bar plot presenting the feature count for targeted panel sequenced (A) reference and (B) test groups to determine the DNA mismatch repair status in next-generation sequencing screened tumors after applying the additive feature combination
approach; Figure S2: Pie charts presenting the proportions of the presence of a somatic second hit and by which mutation type for (A) the reference and (B) the test groups; Figure S3: Flowchart displaying the prevalence of the three Lynch syndrome associated features in the reference group; Table S1: Overview of tumors included in the reference group.

Author Contributions: RW, IMW and DDB contributed to the conception and design of the study. IMW, DDB, MAJ, FAM, SJ, RA, JB, MB, AC, EE, MG, AG, MTH, EI, JK, JM, HM, CN, NP, AR, AS and RS contributed to the recruitment of study participants. RW, MC, JEJ, JC and SGP contributed to sample preparation and quality control. RW, MC, JEJ, KM, PG and DDB contributed to the data analysis. RW, MC, JEJ, KM, PG and DDB contributed to the interpretation of findings. RW and JMC generated the pedigrees using Python. CR, RKP and MC reviewed IHC findings. RW and DDB drafted and substantively revised the manuscript. RW, IMW and DDB contributed to the overall supervision of the project. All authors have read and agreed to the published version of the manuscript.

Funding: This project was funded by a National Health and Medical Research Council of Australia (NHMRC) project grant GNT1125269 (PI- Daniel Buchanan). DDB is supported by an NHMRC Investigator grant (GNT1194896) and University of Melbourne Dame Kate Campbell Fellowship. RW is supported by the Margaret and Irene Stewardson Fund Scholarship and by the Melbourne Research Scholarship. PG is supported by the University of Melbourne Research Scholarship. MAJ is supported by an NHMRC Investigator grant (GNT1195099). The Colon Cancer Family Registry (CCFR, www.coloncfr.org) is supported in part by funding from the National Cancer Institute (NCI), National Institutes of Health (NIH) (award U01 CA167551). Support for case ascertainment was provided in part from the Surveillance, Epidemiology, and End Results (SEER) Program and the following U.S. state cancer registries: AZ, CO, MN, NC, NH; and by the Victoria Cancer Registry (Australia) and Ontario Cancer Registry (Canada).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of The University of Melbourne (protocol code HREC#1750748, approval date: 30th of July 2018).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original contributions presented in the study are included in the article/supplementary files, further inquiries can be directed to the corresponding author.

Acknowledgments: The authors thank members of the Colorectal Oncogenomics Group and members from the Genomic Medicine and Family Cancer Clinic for their support of this
manuscript. We thank the participants and staff from the ANGELS study and from the
Australasian site of the Colon Cancer Family Registry (ACCFR), in particular Maggie
Angelakos, Samantha Fox, and Allyson Templeton for their support of this manuscript. The
authors thank the Australian Genome Research Facility for their collaboration on this project.
The CCFR graciously thanks the generous contributions of their 42,505 study participants,
dedication of study staff, and the financial support from the U.S. National Cancer Institute,
without which this important registry would not exist. The content of this manuscript does not
necessarily reflect the views or policies of the NIH or any of the collaborating centers in the
CCFR, nor does mention of trade names, commercial products, or organizations imply
endorsement by the US Government, any cancer registry, or the CCFR.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the
design of the study; in the collection, analyses, or interpretation of data; in the writing of the
manuscript; or in the decision to publish the results.
References


