Humu-2015-0315

Review

Understanding the pathogenicity of non-coding mismatch repair gene promoter variants in Lynch syndrome

Qing Liu¹, Bryony A. Thompson^{2,3}, Robyn L. Ward^{1,4}, Luke B. Hesson¹ and Mathew A. Sloane^{1,*}

¹Adult Cancer Program, Lowy Cancer Research Centre and Prince of Wales Clinical School, UNSW Australia, Sydney NSW 2052, Australia.

²Huntsman Cancer Institute, University of Utah, 2000 Circle of Hope, Salt Lake City, UT 84112.

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1002/humu.22971](https://doi.org/10.1002/humu.22971).

This article is protected by copyright. All rights reserved.

³Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, University of Melbourne, Victoria 3010, Australia.

⁴Level 3 Brian Wilson Chancellery, The University of Queensland, Brisbane, Queensland 4072, Australia.

***Corresponding author:**

Mathew A. Sloane

m.sloane@unsw.edu.au

Funding: Robyn Ward is supported by the Cancer Institute NSW and Cancer Council NSW (RG13-07) and Bryony Thompson is supported by an NHMRC Early-Career Fellowship (1091211).

Conflict of interest: the authors declare no conflict of interest.

Abstract

Lynch syndrome is the most common familial cancer condition that mainly predisposes to tumors of the colon and endometrium. Cancer susceptibility is caused by the autosomal dominant inheritance of a loss-of-function mutation or epimutation in one of the DNA mismatch repair (MMR) genes. Cancer risk assessment is often possible with non-synonymous coding region mutations, but in many cases patients present with DNA sequence changes within non-coding regions, including the promoters, of MMR genes. The pathogenic role of promoter variants, and hence clinical significance, is unclear and this hinders the clinical management of carriers. In this review we provide an overview of the classification of MMR gene variants, outline the laboratory assays and online resources that can be used to assess the causality of promoter variants in Lynch syndrome, and highlight some of the practical challenges of demonstrating the pathogenicity of these variants. In conclusion, we propose a guide that could be integrated into the current InSiGHT classification scheme to help determine if a MMR promoter variant is pathogenic.

Key words: Lynch syndrome; cancer susceptibility; non-coding variant; promoter; 5' UTR

An introduction to Lynch syndrome

Lynch syndrome is an autosomal dominant condition associated with increased cancer susceptibility, particularly of the colon and endometrium. The clinicopathological features of the syndrome include a predominance of poorly differentiated tumors in the proximal (right-side) colon and the endometrium with an average age of onset of 45 years (Lynch and de la Chapelle, 2003; Umar, et al., 2004). An increased frequency of other cancer types has also been documented (Watson and Lynch, 1994; Watson, et al., 2008). People affected by the condition are heterozygous for a constitutional loss-of-function mutation or epimutation in one of the DNA MMR genes (*MLH1* – MIM# 120436, *MSH2* – MIM# 609309, *MSH6* – MIM# 600678, or *PMS2* – MIM# 600259). A constitutional epimutation refers to the somatic, allele-specific epigenetic inactivation of an allele that is normally active (Hesson, et al., 2010). Pathogenic variants in *MLH1* and *MSH2* (Bonadona, et al., 2011) are associated with a higher cancer risk than in *MSH6* (Baglietto, et al., 2010; Bonadona, et al., 2011) and *PMS2* (ten Broeke, et al., 2015).

The MMR pathway normally repairs single base pair mismatches and small insertion or deletion loops during DNA replication (Kunkel and Erie, 2005). The inheritance of a mutated MMR gene allele (first hit), followed by somatic loss of the remaining wild-type allele (second hit), either through point mutation (Huth, et al., 2012), deletion (Hemminki, et al.,

1994) or promoter hypermethylation (Deng, et al., 2004; Nagasaka, et al., 2010), results in loss of MMR function. This is recognizable in tumor tissue at the molecular level by measuring the length of specific microsatellite repeats that are usually monomorphic in length. Alterations in the length of these repeats, also known as microsatellite instability (MSI), as well as negative immunohistochemical (IHC) staining for the protein of the mutated MMR gene, are characteristic features of Lynch syndrome cancers (Kohlmann and Gruber, 1993). Tumors also typically have a normal *BRAF* sequence and no *MLH1* promoter methylation (with the exception of rare cases caused by a constitutional epimutation or with second hit methylation), which distinguishes them from sporadic MSI positive cancers that frequently have the *BRAF* p.V600E mutation and a methylated *MLH1* promoter (Deng, et al., 2004; Parsons, et al., 2012; Poynter, et al., 2008).

Population-based studies of Lynch syndrome have shown that most MMR gene mutations are in *MLH1* (35-40%) and *MSH2* (44-48%), and far less frequent in *MSH6* (8-10%), and *PMS2* (2-8%) (Schofield, et al., 2014; Steinke, et al., 2014), but mutation frequency can be cohort-dependent (Ward, et al., 2013b). Mutation ‘hotspots’ do not occur in the MMR genes with the spectrum of constitutional mutations including missense (Belvederesi, et al., 2006), nonsense (Yoon, et al., 2009), frameshift (Caluseriu, et al., 2001) and splicing mutations (Thompson, et al., 2013), exon deletions (van der Klift, et al., 2005), and inversions (Chen, 2008; Rhees, et al., 2014; Wagner, et al., 2002). *MSH2* epimutations

are also induced by *cis*-acting deletions of the 3' end of the upstream *EPCAM* gene (Ligtenberg, et al., 2009). A comprehensive collection of MMR gene variants can be found in the International Society for Gastrointestinal Hereditary Tumors (InSiGHT) variant database (Plazzer, et al., 2013).

Despite over twenty years of investigation into the molecular cause of Lynch syndrome, pathogenic variants have not been found in up to 30% of people with a positive family history or strong clinical suspicion of the condition (Lynch, et al., 2015). In these patients no sequence variant is identified, or a sequence variant of uncertain functional and clinical relevance is discovered (Thompson, et al., 2014). In this review we provide an overview of the classification of MMR gene variants, outline the tools that can be used to demonstrate that a MMR gene promoter variant abrogates allele-specific function, and review cases to highlight the practical challenges of demonstrating the pathogenicity of these variants. We also propose an experimental workflow that could be integrated into the current InSiGHT classification scheme to help determine if a MMR promoter variant is pathogenic.

The “state-of-the-art” classification of MMR gene variants in Lynch syndrome

The classification of MMR gene variants is important for the clinical management of families with Lynch syndrome. Intensive surveillance can be restricted to individuals with a

pathogenic variant (Syngal, et al., 1999), which is a cost-effective approach to prevent cancer and mortality in at-risk families (Jarvinen, et al., 2000; Snowsill, et al., 2014). The InSiGHT variants database uses a 5-tiered standardized classification scheme to rank the pathogenicity for constitutional variants in the MMR genes. The database serves as a resource for health care practitioners and researchers, who are also able to submit newly discovered variants, which are then graded in pathogenicity by a review committee based on prevailing evidence. Variants are classified as class 1 (not pathogenic), class 2 (likely not pathogenic), class 3 (uncertain), class 4 (likely pathogenic) or class 5 (pathogenic). Class 3 variants, also known as variants of uncertain significance (VUS), are DNA sequence alterations that have insufficient evidence to allow unequivocal determination of clinical significance (Kearney, et al., 2011). Class 4 variants have a posterior probability of pathogenicity between 95% and 99% (derived through multifactorial likelihood analysis), or fulfill criterion 1 and either criterion 2 or 3 below. Class 5 variants have a posterior probability of pathogenicity greater than 99%, or fulfill **all** of the following characteristics (readers should refer to Thompson *et al.* (2014) for a detailed description of the rules for variant classification):

Criterion 1 – allele-specific abrogated function that has been validated by functional assays, or the patient is characterized with constitutional MMR deficiency (CMMRD) (Bakry, et al., 2014), or presence of the variant is discovered in different background haplotypes.

Criterion 2 – evidence for co-segregation with disease within pedigrees.

Criterion 3 – evidence of association with a Lynch syndrome molecular phenotype (MSI and/or IHC loss) in two or more independent tumors.

Criterion 4 – an allele frequency less than 1% in healthy population controls.

Patients with class 1 or class 2 variants and their relatives are managed according to their family and personal history as if they had no pathogenic variant in a MMR gene (Plon, et al., 2008; Thompson, et al., 2014). The system recommends that individuals carrying class 4 and class 5 variants be managed in accordance with the guidelines for clinical testing and full high-risk surveillance which includes colonoscopy every one to two years commencing at age 25 (Cancer Institute NSW; Vasen, 2007).

Class 3 variants are problematic for cancer risk assessment because it is challenging to perform laboratory experiments to show allele-specific abrogation of function. Investigations of class 3 variants involve multiple, complex *in vitro* assays and possibly repeated contact and consent from patients to request additional samples for analysis (Syngal, et al., 1999). The current standard-of-care for persons with class 3 variants is based on their family and personal history (Plon, et al., 2008; Thompson, et al., 2014).

In 2013, 765 class 3 MMR gene variants were listed in InSiGHT. These variants were categorized as non-truncating non-synonymous variants (524; 68 %), variants in introns outside the canonical splicing dinucleotides (127; 17 %), synonymous (63; 8 %), initiation

codon (8; 1 %), 5' untranslated region (UTR) (15; 2 %) and 3' UTR (6; 0.4 %) variants, large genomic duplications (17; 2 %) and premature terminating codons in the last exon (5; 0.7 %) (Thompson, et al., 2014). Variants within MMR gene promoters are likely to have been underestimated because sequencing of these regions is rarely performed. A list of all MMR gene promoter variants listed in InSiGHT and population-based databases (October, 2015) is shown in Supp. Table S1. The current version of the InSiGHT classification scheme does not include specific rules for the interpretation of promoter variants, yet being able to demonstrate that a variant causes allele-specific abrogated function is required to assign a variant to class 4 and 5 (Thompson, et al., 2014).

Mechanisms by which promoter variants effect gene expression

Variants within the promoter region of an MMR gene can be pathogenic due to a loss of promoter function that results in the reduction or loss of gene expression. Gene promoters are regions of DNA that bind and position the transcription preinitiation complex (PIC) to regulate the transcriptional output of the linked gene (Fig. 1A). The PIC consists of a DNA-dependent RNA polymerase, such as RNA Polymerase II (RNAPII), and transcription factors (TFs) that recognize sequence elements in the core promoter that overlaps the transcription

start site (TSS) and 5'UTR. The binding of RNAPII can also be coordinated by TFs that bind to more distant enhancer regions (Lenhard, et al., 2012; Riethoven, 2010).

The major types of vertebrate promoters are categorized based on their DNA sequence, expression pattern and chromatin structure. Type I or 'sharp' promoters are often found at genes with tissue-specific expression patterns, a highly defined TSS, a TATA-box and high levels of nucleosome occupancy. Type II or 'broad' promoters have a short CpG island with multiple TSSs, and a nucleosome depleted region across the TSS. Type III promoters have large CpG islands with bivalent chromatin domains that extend into the gene body and are typically associated with developmentally regulated genes (Lenhard, et al., 2012). The DNA MMR genes have the typical features of type II promoters – widespread expression (Fink, et al., 1997; Leach, et al., 1996), a CpG island (Clark, et al., 1994), multiple TSSs (Hesson, et al., 2015) and highly positioned nucleosomes flanking a nucleosome depleted region across the TSS (Hesson, et al., 2013; Lin, et al., 2007).

Promoter variants affect gene expression by altering TF binding motifs (Fig. 1B, C) as illustrated by the following examples. Telomerase reverse transcriptase (*TERT*, MIM# 187270) is a catalytic subunit of telomerase that is repressed in normal somatic cells, but is reactivated in cancer (Kim, et al., 1994; Meyerson, 2000). *TERT* promoter variants were recently implicated as a cause of sporadic and hereditary melanoma (Horn, et al., 2013; Huang, et al., 2013). Increased *TERT* expression in melanoma is thought to be mediated by

promoter variants that generate novel binding motifs for transcriptional activators in the ETS TF family (Horn, et al., 2013; Huang, et al., 2013).

Death-associated protein kinase 1 (*DAPK1*, MIM# 600831) is a positive mediator of apoptosis that is often silenced by DNA hypermethylation in sporadic chronic lymphocytic leukemia (CLL). In a case of familial CLL an SNV located approximately 6.5 kb (c.-6531A>G) upstream of the *DAPK1* TSS generated a binding site for the transcriptional repressor, HOXB7, that was associated with a constitutional epimutation and allele-specific loss of expression (Raval, et al., 2007).

X-chromosome inactivation is the process by which one X-chromosome is randomly transcriptionally silenced in the somatic cells of female placental mammals. In approximately 1% of females there is preferential inactivation (95:5) of one X-chromosome (Sapienza, et al., 1997). Two variants (c.-43C>G and c.-43C>A) in the *XIST* (MIM# 314670) promoter have been associated with skewed inactivation (Sapienza, et al., 1997; Tomkins, et al., 2002). These two SNVs occur in a conserved CTCF binding site which has enhanced affinity for c.-43C>G, but binding is abolished by c.-43C>A. The site probably regulates higher-order chromatin structure at the *XIST* promoter (Pugacheva, et al., 2005).

Guidelines for investigating causality of sequence variants in human disease have been proposed, but they do not provide details of the experimental techniques to guide researchers

(MacArthur, et al., 2014). *In vitro* assays to help assess the pathogenicity of missense variants have been reviewed, but techniques interrogating coding variants cannot be applied to promoter variants (Rasmussen, et al., 2012), and although a general experimental approach has been suggested for the study of promoter variants (de Vooght, et al., 2009), this was not in the context of Lynch syndrome.

In the following section, the laboratory assays and online resources that can be used to evaluate the causality of a MMR gene promoter variant in Lynch syndrome are outlined. The advantages and disadvantages of each tool are also summarized in Supp. Table S2. It is appreciated that each research team will perform experiments in different orders depending on their expertise and availability of resources.

Laboratory assays and online resources to assess the pathogenicity of a MMR gene promoter variant in Lynch syndrome

I. Laboratory assays

(a) Prerequisites for screening for MMR gene promoter variants

Only patients that fulfill the following criteria should be screened for promoter variants: (i) tumor IHC-negative for MLH1/PMS2, MSH2/MSH6, PMS2 only or MSH6 only, (ii) a tumor with MSI (Umar, et al., 2004), (iii) no sequence changes in the exons or splice sites of the candidate MMR gene/s, (iv) no exon copy number changes in the candidate MMR gene/s detected by multiplex ligation-dependent probe amplification (MLPA) and (v) no 3' *EPCAM* deletion. Lynch syndrome tumors will also usually have a normal *BRAF* sequence, however, the somatic p.V600E mutation has been reported in some epimutation carriers (Crepin, et al., 2012; Goel, et al., 2011; van Roon, et al., 2010), so such patients should not necessarily be excluded from promoter variant analysis if the other criteria are fulfilled.

(b) Sample collection

The collection of appropriate tissue samples is critical to assess the physiological relevance of a promoter variant. It can be difficult to collect samples because some family members may be deceased, others may be unwilling or unable to consent, or there may be logistical challenges associated with collecting tissues from separated families. This will vary on a case-by-case basis so communication between clinicians and researchers is important to ensure such challenges are clear at the outset, and to maximize the chance of obtaining the most appropriate samples to perform informative experiments.

Currently, only DNA is routinely isolated from peripheral blood for genetic testing. A lack of RNA is not problematic if an obvious pathogenic coding variant is detected, however, RNA from tissues is critical to demonstrate that a promoter variant is associated with an *in vivo* allele-specific reduction in expression. A concurrent extraction of DNA and RNA at the time of initial sampling avoids a *post hoc* blood collection for RNA isolation, which can be inconvenient or stressful for the patient and sometimes not possible if they have passed away in the interim. RNA from normal colon mucosa is more physiologically relevant, but this is usually not critical for establishing whether a change in *in vivo* expression is physiologically relevant.

An immortalized lymphoblastoid cell line (LCL), generated by Epstein-Barr virus (EBV) transformation of peripheral B lymphocytes, can also serve as a renewable source of nucleic acids to optimize experiments when patient samples are limited, and for performing more sophisticated experiments (Frisan, et al., 2001). However, it is important to appreciate that during the process of immortalization clonal selection may occur, possibly leading to alterations in the levels of gene expression. Consequently, LCLs may not be representative of the *in vivo* state so experiments on patient samples should be the priority.

(c) Promoter sequencing and DNA methylation analysis

Variants in DNA MMR gene promoters are not normally reported in routine genetic test reports because only the exons and splice sites are sequenced. The MMR genes have type II promoters so standard Sanger sequencing of the 1 kb region upstream of codon one of the relevant MMR gene/s (as indicated by IHC analysis of tumors) should identify any potential pathogenic variants in the promoter.

DNA methylation screening can be done in conjunction with sequencing and should be performed on patients fulfilling the prerequisite criteria for promoter sequencing, as described above. Constitutional epimutations can occur with (secondary) (Fig. 1D) (Hitchins, et al., 2011) or without (primary) (Goel, et al., 2011; Hitchins, et al., 2007; Suter, et al., 2004) a co-segregating genetic variant. It has been hypothesized that secondary epimutations are a consequence of altered TF binding (Hitchins, et al., 2011). Combined bisulfite restriction analysis (COBRA) (Xiong and Laird, 1997), methylation-specific MLPA (MS-MLPA) (Nygren, et al., 2005), methylation-specific PCR (MSP) (Herman, et al., 1996) and pyrosequencing (Tost and Gut, 2007) are commonly used assays to initially measure the average level of methylation across both alleles of a MMR gene promoter (Supp. Table S2). It is important to examine the correct regions for hypermethylation and the locations within the promoters of the *MLH1* (Deng, et al., 1999; Hitchins, et al., 2007) and *MSH2* (Ligtenberg, et al., 2009) genes have been previously described. The spatial distribution and allele-specificity of DNA methylation can then be assessed at nucleotide resolution using sodium

bisulfite sequencing if the promoter region contains a heterozygous variant (Hitchins, et al., 2007; Sloane, et al., 2015).

(d) Assays to measure *in vivo* allele-specific loss of MMR gene expression

RNA from peripheral blood is a **priority** in establishing if a MMR gene promoter variant is associated with reduced expression. Commonly used techniques to measure allele-specific gene expression include cDNA-restriction fragment length polymorphism (RFLP) analysis (Shin, et al., 2002), cDNA pyrosequencing (Ronaghi, et al., 1998) and allele-specific quantitative reverse transcriptase PCR (qRT-PCR) (Gaudet, et al., 2009) (Supp. Table S2).

All allele-specific gene expression assays are reliant on an expressed (exonic) heterozygous single nucleotide polymorphism (SNP) or the promoter variant itself. In cases where a heterozygous exonic SNP is identified it must be determined which polymorphic SNP variant is *in cis* with the promoter variant (i.e. variant haplotype must be established) using cDNA sequencing.

(e) *In vitro* assays to measure the effect of a promoter variant on expression

In vitro assays can demonstrate that a promoter variant can reduce promoter activity.

Luciferase reporter assays (LRAs) (Smale, 2010) have been the standard method for investigating the effects of variants on promoter activity but are likely to be superseded by CRISPR (clustered regularly interspaced short palindromic repeat)–Cas9 technology (Sander and Joung, 2014).

LRAs quantify promoter activity by linking a promoter to the luciferase gene in a plasmid construct and measuring the light generated by the luciferase protein as it oxidizes the chemical substrate luciferin (Smale, 2010). Reduced activity from a variant construct relative to a wild-type construct supports a causal role for a promoter variant in the *in vivo* loss of expression. Firefly and *Renilla* are used to normalize for transfection efficiency, however, luciferase activity may be affected by the genetic background of the cell types used and there is often high variation in activity between plasmid constructs. For *MLH1* c.-27C>A, LRAs measured more than an 8-fold reduction in activity relative to a wild-type sequence in HCT116 cells, whereas only a 2-fold reduction was observed with the same construct in HEK293 cells (Hitchins, et al., 2011). Ward et al., 2013a found differences in normalized luciferase output of up to 30% for the same variant in different cell lines. For this reason, LRAs should be performed in more than one cell line at a minimum. LRAs cannot reproduce an *in vivo* chromatin environment, which may also affect TF binding. For example, the A allele at *MLH1* c.-27C>A is associated with a constitutional epimutation in patient samples

but not on plasmid constructs (Hitchins, et al., 2011). Furthermore, LRAs cannot be performed in the context of the colonic epithelium, which could be important in determining the pathogenic effects of the variant.

There is currently no accepted cutoff in promoter activity to define a pathogenic variant. Hinrichsen et al., 2013, however, used an *in vitro* MMR assay to show that MMR efficiency progressively dropped when MLH1 protein concentration fell below 50% of control cells proficient in MMR (Hinrichsen, et al., 2013). Similarly, Kansikas et al., 2014, showed that there was a complete loss of MMR when *MLH1* or *MSH2* mRNA were reduced to 25% of control cell levels (Kansikas, et al., 2014). This would be the *in vivo* equivalent of a 50% loss of expression from a variant allele combined with LOH in a tumor, suggesting that a 50% reduction or greater in promoter activity may be an ideal cutoff to define abrogated allele-specific expression, as proposed previously (Green, et al., 2003; Hesson, et al., 2015).

It is possible that the introduction of a variant directly into the genome of a cell line or model organism using CRISPR-Cas9 technology might overcome the limitations of LRAs (Sander and Joung, 2014). CRISPR-Cas9 is a component of the adaptive immune system of prokaryotes that uses RNA-guided nucleases to cleave foreign DNA (Makarova, et al., 2011). It has been adapted as a molecular tool to precisely edit eukaryotic genomes. In the CRISPR-Cas9 system, the Cas9 nuclease is directed by a guide RNA that introduces a double stranded break in the complementary DNA sequence. DNA repair by non-homologous end joining

introduces genetic alterations such as indels that impair gene function (Sander and Joung, 2014). Modifications to the CRISPR-Cas9 system have enabled the introduction or substitution of single nucleotide changes via the high-fidelity homology directed repair pathway (Chen, et al., 2011; Chu, et al., 2015; Yin, et al., 2014). Although CRISPR-Cas9 can have off-target effects it is likely to become a valuable tool to investigate the causal role of MMR gene promoter variants in Lynch syndrome.

(f) Assays to investigate the mechanism-of-action of a promoter variant

Assays that investigate the mechanism-of-action of a MMR gene promoter variant principally aim to identify the loss of a transcriptional activator, or the gain of a transcriptional repressor. Historically, electrophoretic mobility shift assays (EMSAs) have been the method of choice for investigating whether a promoter variant creates or destroys a TF binding site (Hellman and Fried, 2007). EMSAs, however, have several limitations and in only one instance, when combined with mass spectrometry, has the technique identified a specific TF with altered binding at a MMR gene promoter variant (E1A-F binding loss caused by *MSH2* c.-78_-77del) (Yan, et al., 2007).

Proteome-wide analysis of SNPs (PWAS) is an alternative unbiased method that could be used in LCLs to identify differential binding of TFs. Two cell populations are

differentially labelled using stable isotope labeling by amino acids in cell culture (SILAC). Nuclear lysates are then incubated with a wild-type and variant oligonucleotide, and differential binding of proteins is quantified with mass spectrometry (MS)-based proteomics. PWAS has been used to identify differential binding of TFs at variants highly associated with type 1-diabetes at the *IL2RA* locus (Butter, et al., 2012). A potential limitation of PWAS is that TFs with chromatin-dependent binding may not bind to the naked DNA probes.

If a candidate TF is identified with techniques such as EMSA or PWAS, chromatin immunoprecipitation (ChIP) (O'Neill and Turner, 2003; Orlando and Paro, 1993) performed in LCLs or patient tissues is the best approach to validate that the TF binds a MMR promoter *in vivo*. NOME-seq, a technique that can map the position of nucleosomes and endogenous CpG methylation on the same molecule, can show if altered TF binding is associated with alterations in nucleosome positioning or DNA methylation (Fatemi, et al., 2005).

(g) Tools to identify co-segregating variants *in cis*

If a variant has no effect on promoter activity in *in vitro* assays then it may be a 'bystander' variant that is co-segregating on the same haplotype as another *cis*-acting genetic variant that may be the cause of *in vivo* allele-specific loss of expression (see *MLH1* c.-269C>G in Supp. Table S3). There are no strict criteria in place for the size of the genomic region that should

be investigated to exclude other co-segregating pathogenic alterations. In previous studies, we sequenced a 10 kb region from the 3' end of *EPM2AIP1* to *MLH1* exon 2 to exclude a genetic cause for an epimutation (Sloane, et al., 2015) or performed targeted enrichment and sequencing of a 1.7 Mb region encompassing 15 genes across the *MLH1* locus (Hesson, et al., 2015). Moreover, Kwok *et al.* (2014) excluded copy number variation using SNP arrays on a 30 Mb haplotype. Ideally high-throughput sequencing of the candidate MMR gene and flanking regions should be performed on germline DNA from cancer-affected and unaffected family members. Bioinformatic software such as the Genome Analysis Toolkit (GATK) (McKenna, et al., 2010), Pindel (Ye, et al., 2009) and CREST (Wang, et al., 2011) are able to identify small variants and complex rearrangements including inversions, tandem duplications and large deletions in these datasets.

II. Online resources

(a) *In silico* tools

In silico tools are useful for identifying putative TFs that may be affected by a MMR gene promoter variant. JASPAR is an open-access database of non-redundant, TF binding sites in vertebrates that is based on experimental data (Sandelin, et al., 2004a). ConSite integrates TF binding site prediction, generated with high-quality transcription factor models and cross-

species comparison, to predict the location of *cis*-regulatory elements in genomic sequences (Sandelin, et al., 2004b). The ENCODE project has also mapped 12 histone modifications and the binding sites for 187 TFs in 118 cell lines across 1% of the human genome using ChIP and the profiles can be visualized on the UCSC epigenome browser (Birney, et al., 2007) (Supp. Table S2). These open-access, web-based resources are useful starting points for designing assays such as EMSA and ChIP to experimentally validate that a MMR gene promoter variant alters TF binding.

(b) Variant databases

Prior to publically available variant databases, large cohorts of unrelated healthy controls had to be sequenced to demonstrate a promoter variant was not a common polymorphism.

Researchers can now consult population-based databases such as the 1000 Genomes project (Genomes Project, et al., 2015), the NCBI dbSNP Short Genetic Variations database (Sherry, et al., 2001), the NHLBI grand opportunity exome sequencing project (ESP) (NHLBI Exome Sequencing Project (ESP) Exome Variant Server, 2015) and the Exome Aggregation Consortium (ExAC) (Exome Aggregation Consortium (ExAC), 2015) to determine the population frequency of a newly identified variant. A pathogenic variant is predicted to have

a minor population allele frequency of less than 1%. A summary of databases is provided in Supp. Table S2.

The InSiGHT variants database is not population-based and contains variants in the MMR genes, and other colorectal cancer risk genes (Thompson, et al., 2014). It should be consulted for pre-existing entries of the identified promoter variant. Data accumulated for the same variant from independent studies provides more certainty about pathogenicity. To assist in the classification of variants, researchers should submit all relevant data to the Leiden Open Variation Database (LOVD) or via the InSiGHT template (<http://insight-group.org/variants/classifications/>).

Databases have limitations with respect to depth of coverage, regions sequenced, or the number of individuals investigated. For example, the full 1000 Genomes project has sequenced about 2,500 samples at 7× mean coverage, but this level of coverage is insufficient to provide a complete genotype of each sample (Genomes Project, et al., 2015). A limitation of the ESP database is that gene promoter regions are often sequenced to low coverage, or not at all. Researchers should bear these limitations in mind when assessing variant allele frequencies in control populations.

Case studies of MMR gene promoter variants in Lynch syndrome

The InSiGHT variants database currently contains 10 *MLH1*, 6 *MSH2*, 7 *MSH6* and 7 *PMS2* promoter variants (Fig. 2). With the exception of one class 4 promoter variant (*MSH2* c.-78_-77del), all promoter variants are classified as class 1 (11), class 2 (3) or class 3 (15). Supp. Table S3 provides a summary of the experimental and clinical data that have been used to classify all promoter variants with accompanying publications in InSiGHT (excluding the class 1 variants with an allele frequency >1%, see Supp. Table S1). The criteria that still need to be fulfilled to reclassify these variants as likely pathogenic or pathogenic are also provided. The following examples outline in more detail the assays that have been used for four of the most intensively interrogated MMR gene promoter variants with an explanation for their current classification.

(i) *MLH1* c.-27C>A

MLH1 c.-27C>A is a class 3 promoter variant that co-segregates with DNA hypermethylation and repressive histone modifications (H3K9me3 and H3K27me3) on a transcriptionally inactive allele (constitutional epimutation). It occurs on a haplotype with c.85G>T (p.A29S), but no other detectable variants (frequency <1%) or copy number changes. LRAs showed that c.-27C>A alone significantly reduced the transcriptional activity of the *MLH1* promoter (criterion 1) (Hitchins, et al., 2011). This variant co-segregated with early-onset CRC and

endometrial cancer in five families (criterion 2) (Hitchins, et al., 2011; Kwok, et al., 2014) and six *MLH1*-negative tumors with MSI have been identified in patients carrying the variant (criterion 3). In addition, it has not been reported in any population-based variant databases (criterion 4). Functional assays have demonstrated that c.85G>T has no effect on transcription (Hitchins, et al., 2011), and the p.A29S amino acid substitution has normal MMR activity, localization and interaction with PMS2 in *in vitro* assays (Raevaara, et al., 2005). However, a role for c.85G>T in attracting methylation that can extend into intron 1 has not been excluded. It is for this reason, and because InSiGHT does not have promoter specific functional criteria for classification, that c.-27C>A is listed as a class 3 variant even though it has all four characteristics of a pathogenic variant.

(ii) *MLH1* c.-42C>T

c.-42C>T is a class 3 promoter variant that was first reported in a Newfoundland family that met the Amsterdam criteria for Lynch syndrome (Green, et al., 2003). It is not associated with a constitutional epimutation (Ward, et al., 2013a). Although c.-42C>T had only 37% of the activity of the wild-type promoter in LRAs, possibly caused by altered c-Myb binding, criterion 1 has not been fulfilled because the LRA was only performed in one cell line, *in vivo* allele-specific loss of *MLH1* expression was not investigated in any patient samples, nor is it

known if it shares a haplotype with other genetic alterations (Green, et al., 2003). Criterion 2 has also not been met, because there was insufficient co-segregation data from affected family members. c.-42C>T has been reported in two unrelated individuals with *MLH1*-negative/MSI colorectal cancer (criterion 3) and is very rare in population databases (criteria 4). The class 3 status is based on insufficient data supporting criteria 1 and 2.

(iii) *MLH1* c.-7C>T/c.-28A>G

A recent study (Hesson, et al., 2015) investigated a rare *MLH1* haplotype containing the c.-7C>T and c.-28A>G variants in two individuals from suspected Lynch syndrome families. Expression from the variant allele was reduced to 50% of the wild-type allele in peripheral blood from both patients in the absence of an epimutation or altered nucleosome structure. A strength of the study was the high-throughput sequencing of a 1.7 Mb region encompassing the *MLH1* gene, as well as detailed microsatellite and LOH analysis, which confirmed these two variants were the only potentially pathogenic sequence alterations within or near the *MLH1* gene. Despite this work, the data in support of criterion 1 are inconclusive because the *in vitro* assays have not been performed to show the variants *per se* cause a reduction in expression. Additional data to support criteria 2 and 3 are also needed before either of these variants is changed from class 3. The presence of a homozygote carrier without features of

CMMRD in the ExAC database, suggests that it is unlikely to be high-risk, but it could be a moderate risk haplotype due to the 50% reduction in expression.

(iv) *MSH2* c.-78_-77del

The c.-78_-77del variant in the 5' UTR of *MSH2* is the only class 4 variant in InSiGHT, based on a posterior probability of pathogenicity of 0.989 using multifactorial likelihood analysis (Thompson, et al., 2014). LRAs showed promoter activity from the variant was 16.8 % of the wild-type sequence. *In silico* analysis predicted that the reduced promoter activity was caused by loss of an E1A-F TF binding site, and this was confirmed with a combined EMSA/mass-spectrometry approach. LOH also occurred in two tumors. Unfortunately, criterion 1 was not completely met because RNA was not isolated from patient tissues to show the variant allele was associated with a loss of expression. The variant fulfills criterion 2 for classification as a class 4 variant (odds of segregation with disease >5:1, but not >10:1) but not as a class 5 variant (Thompson, et al., 2014). It was found in a family (Yan, et al., 2007) where it co-segregated over two generations with MSI-high/IHC-negative *MSH2* tumors (criterion 3) in three family members, but was absent in eight cancer-free family members as well as 220 control subjects (criterion 4). Classification of this variant could be reassessed in light of *in vivo* data demonstrating allele-specific loss of expression.

Recommendations to assist in the classification of MMR gene promoter variants

The three key criteria that need to be addressed to conclusively demonstrate a promoter variant is the cause of allele-specific abrogated function are i) an association with allele-specific loss of expression *in vivo*, (ii) an *in vitro* reduction in promoter activity or expression relative to the wild-type sequence, and (iii) the exclusion of other rare *cis* variants on the same haplotype. Assays investigating the mechanism-of-action will corroborate these findings but we do not deem them necessary to demonstrate pathogenicity. The assays outlined in this review have been integrated into a flowchart that we propose could be integrated with the existing InSiGHT classification scheme (Fig. 3) to help assign pathogenicity to MMR gene promoter variants.

Our ability to prove that non-coding sequence variants increase cancer risk has lagged behind the sequencing technology that has identified them. This review shows that the availability of tissue samples from patients and their relatives, combined with insufficient experimental data showing allele-specific abrogated function, have been the most significant impediments to investigating promoter variants. It is hoped that the guidelines proposed in this review will be used by researchers investigating promoter variants in MMR genes, and ultimately enable better cancer risk assessment in people that carry them.

References

- Baglietto L, Lindor NM, Dowty JG, White DM, Wagner A, Gomez Garcia EB, Vriends AH, Dutch Lynch Syndrome Study G, Cartwright NR, Barnetson RA, Farrington SM, Tenesa A, et al. 2010. Risks of Lynch syndrome cancers for MSH6 mutation carriers. *J Natl Cancer Inst* 102(3):193-201.
- Bakry D, Aronson M, Durno C, Rimawi H, Farah R, Alharbi QK, Alharbi M, Shamvil A, Ben-Shachar S, Mistry M, Constantini S, Dvir R, et al. 2014. Genetic and clinical determinants of constitutional mismatch repair deficiency syndrome: report from the constitutional mismatch repair deficiency consortium. *Eur J Cancer* 50(5):987-996.
- Belvederesi L, Bianchi F, Loretelli C, Gagliardini D, Galizia E, Bracci R, Rosati S, Bearzi I, Viel A, Cellerino R, Porfiri E. 2006. Assessing the pathogenicity of MLH1 missense mutations in patients with suspected hereditary nonpolyposis colorectal cancer: correlation with clinical, genetic and functional features. *Eur J Hum Genet* 14(7):853-859.

Birney E, Stamatoyannopoulos JA, Dutta A, Guigó R, Gingeras TR, Margulies EH, Weng Z,

Snyder M, Dermitzakis ET, Stamatoyannopoulos JA. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447(7146):799-816.

Bonadona V, Bonaiti B, Olschwang S, Grandjouan S, Huiart L, Longy M, Guimbaud R,

Buecher B, Bignon YJ, Caron O, Colas C, Nogues C, et al. 2011. Cancer risks associated with germline mutations in MLH1, MSH2, and MSH6 genes in Lynch syndrome. *JAMA* 305(22):2304-2310.

Butter F, Davison L, Viturawong T, Scheibe M, Vermeulen M, Todd JA, Mann M. 2012.

Proteome-wide analysis of disease-associated SNPs that show allele-specific transcription factor binding. *PLoS Genet* 8(9):e1002982.

Caluseriu O, Cordisco EL, Viel A, Majore S, Nascimbeni R, Pucciarelli S, Genuardi M.

2001. Four novel MSH2 and MLH1 frameshift mutations and occurrence of a breast cancer phenocopy in hereditary nonpolyposis colorectal cancer. *Hum Mutat* 17(6):521.

Cancer Institute NSW. eviQ Cancer Treatments Online v. 1.4.0 ed. <https://www.eviq.org.au/>

-
- Chen F, Pruett-Miller SM, Huang Y, Gjoka M, Duda K, Taunton J, Collingwood TN, Frodin M, Davis GD. 2011. High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat Methods* 8(9):753-755.
- Chen JM. 2008. The 10-Mb paracentric inversion of chromosome arm 2p in activating MSH2 and causing hereditary nonpolyposis colorectal cancer: re-annotation and mutational mechanisms. *Genes Chromosomes Cancer* 47(6):543-545.
- Chu VT, Weber T, Wefers B, Wurst W, Sander S, Rajewsky K, Kuhn R. 2015. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol* 33(5):543-548.
- Clark SJ, Harrison J, Paul CL, Frommer M. 1994. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res* 22(15):2990-2997.
- Crépin M, Bieu MC, Lejeune S, Escande F, Boidin D, Porchet N, Morin G, Manouvrier S, Mathieu M, Buisine MP. 2012. Evidence of constitutional MLH1 epimutation associated to transgenerational inheritance of cancer susceptibility. *Hum Mutat* 33(1):180-181.
- de Vooght KM, van Wijk R, van Solinge WW. 2009. Management of gene promoter mutations in molecular diagnostics. *Clin Chem* 55(4):698-708.

Deng G, Bell I, Crawley S, Gum J, Terdiman JP, Allen BA, Truta B, Sleisenger MH, Kim

YS. 2004. BRAF mutation is frequently present in sporadic colorectal cancer with methylated hMLH1, but not in hereditary nonpolyposis colorectal cancer. Clin Cancer Res 10(1 Pt 1):191-195.

Deng G, Chen A, Hong J, Chae HS, Kim YS. 1999. Methylation of CpG in a small region of the hMLH1 promoter invariably correlates with the absence of gene expression. Cancer Res 59(9):2029-2033.

Exome Aggregation Consortium (ExAC) C, MA (URL: <http://exac.broadinstitute.org>) [accessed October, 2015].

Fatemi M, Pao MM, Jeong S, Gal-Yam EN, Egger G, Weisenberger DJ, Jones PA. 2005. Footprinting of mammalian promoters: use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level. Nucleic Acids Res 33(20):e176.

Fink D, Nebel S, Aebi S, Zheng H, Kim HK, Christen RD, Howell SB. 1997. Expression of the DNA mismatch repair proteins hMLH1 and hPMS2 in normal human tissues. Br J Cancer 76(7):890-893.

-
- Fredriksson H, Ikonen T, Autio V, Matikainen MP, Helin HJ, Tammela TL, Koivisto PA, Schleutker J. 2006. Identification of germline MLH1 alterations in familial prostate cancer. *Eur J Cancer* 42(16):2802-2806.
- Frisan T, Levitsky V, Masucci M. 2001. Generation of lymphoblastoid cell lines (LCLs). *Methods Mol Biol* 174:125-127.
- Gaudet M, Fara AG, Beritognolo I, Sabatti M. 2009. Allele-specific PCR in SNP genotyping. *Methods Mol Biol* 578:415-424.
- Gazzoli I, Kolodner RD. 2003. Regulation of the human MSH6 gene by the Sp1 transcription factor and alteration of promoter activity and expression by polymorphisms. *Mol Cell Biol* 23(22):7992-8007.
- Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. 2015. A global reference for human genetic variation. *Nature* 526(7571):68-74. [database accessed October, 2015].
- Goel A, Nguyen TP, Leung HC, Nagasaka T, Rhee J, Hotchkiss E, Arnold M, Banerji P, Koi M, Kwok CT, Packham D, Lipton L, et al. 2011. De novo constitutional MLH1 epimutations confer early-onset colorectal cancer in two new sporadic Lynch syndrome cases, with derivation of the epimutation on the paternal allele in one. *Int J Cancer* 128(4):869-878.

-
- Green RC, Green AG, Simms M, Pater A, Robb JD, Green JS. 2003. Germline hMLH1 promoter mutation in a Newfoundland HNPCC kindred. *Clin Genet* 64(3):220-227.
- Hellman LM, Fried MG. 2007. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat Protoc* 2(8):1849-1861.
- Hemminki A, Peltomaki P, Mecklin JP, Jarvinen H, Salovaara R, Nystrom-Lahti M, de la Chapelle A, Aaltonen LA. 1994. Loss of the wild type MLH1 gene is a feature of hereditary nonpolyposis colorectal cancer. *Nat Genet* 8(4):405-410.
- Hendriks YM, Jagmohan-Changur S, van der Klift HM, Morreau H, van Puijenbroek M, Tops C, van Os T, Wagner A, Ausems MG, Gomez E, Breuning MH, Brocker-Vriends AH, et al. 2006. Heterozygous mutations in PMS2 cause hereditary nonpolyposis colorectal carcinoma (Lynch syndrome). *Gastroenterology* 130(2):312-322.
- Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB. 1996. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proc Natl Acad Sci U S A* 93(18):9821-9826.
- Hesson LB, Hitchins MP, Ward RL. 2010. Epimutations and cancer predisposition: importance and mechanisms. *Curr Opin Genet Dev* 20(3):290-298.

Hesson LB, Packham D, Kwok CT, Nunez AC, Ng B, Schmidt C, Fields M, Wong JW,

Sloane MA, Ward RL. 2015. Lynch Syndrome Associated with Two MLH1 Promoter Variants and Allelic Imbalance of MLH1 Expression. *Hum Mutat* 36(6):622-630.

Hesson LB, Patil V, Sloane MA, Nunez AC, Liu J, Pimanda JE, Ward RL. 2013. Reassembly of nucleosomes at the MLH1 promoter initiates resiliencing following decitabine exposure. *PLoS Genet* 9(7):e1003636.

Hinrichsen I, Brieger A, Trojan J, Zeuzem S, Nilbert M, Plotz G. 2013. Expression defect size among unclassified MLH1 variants determines pathogenicity in Lynch syndrome diagnosis. *Clin Cancer Res* 19(9):2432-2441.

Hitchins MP, Rapkins RW, Kwok CT, Srivastava S, Wong JJ, Khachigian LM, Polly P, Goldblatt J, Ward RL. 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):200-213.

Hitchins MP, Wong JJ, Suthers G, Suter CM, Martin DI, Hawkins NJ, Ward RL. 2007. Inheritance of a cancer-associated MLH1 germ-line epimutation. *N Engl J Med* 356(7):697-705.

-
- Horn S, Figl A, Rachakonda PS, Fischer C, Sucker A, Gast A, Kadel S, Moll I, Nagore E, Hemminki K. 2013. TERT promoter mutations in familial and sporadic melanoma. *Science* 339(6122):959-961.
- Huang FW, Hodis E, Xu MJ, Kryukov GV, Chin L, Garraway LA. 2013. Highly recurrent TERT promoter mutations in human melanoma. *Science* 339(6122):957-959.
- Huth C, Kloor M, Voigt AY, Bozukova G, Evers C, Gaspar H, Tariverdian M, Schirmacher P, von Knebel Doeberitz M, Blaker H. 2012. The molecular basis of EPCAM expression loss in Lynch syndrome-associated tumors. *Mod Pathol* 25(6):911-916.
- Isidro G, Matos S, Goncalves V, Cavaleiro C, Antunes O, Marinho C, Soares J, Boavida MG. 2003. Novel MLH1 mutations and a novel MSH2 polymorphism identified by SSCP and DHPLC in Portuguese HNPCC families. *Hum Mutat* 22(5):419-420.
- Ito E, Yanagisawa Y, Iwahashi Y, Suzuki Y, Nagasaki H, Akiyama Y, Sugano S, Yuasa Y, Maruyama K. 1999. A core promoter and a frequent single-nucleotide polymorphism of the mismatch repair gene hMLH1. *Biochem Biophys Res Commun* 256(3):488-494.
- Jarvinen HJ, Aarnio M, Mustonen H, Aktan-Collan K, Aaltonen LA, Peltomaki P, De La Chapelle A, Mecklin JP. 2000. Controlled 15-year trial on screening for colorectal

-
- cancer in families with hereditary nonpolyposis colorectal cancer. *Gastroenterology* 118(5):829-834.
- Jeske YW, So A, Kelemen L, Sukor N, Willys C, Bulmer B, Gordon RD, Duffy D, Stowasser M. 2008. Examination of chromosome 7p22 candidate genes RBAK, PMS2 and GNA12 in familial hyperaldosteronism type II. *Clin Exp Pharmacol Physiol* 35(4):380-385.
- Kansikas M, Kasela M, Kantelinen J, Nystrom M. 2014. Assessing how reduced expression levels of the mismatch repair genes MLH1, MSH2, and MSH6 affect repair efficiency. *Hum Mutat* 35(9):1123-1127.
- Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST, Working Group of the American College of Medical Genetics Laboratory Quality Assurance C. 2011. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet Med* 13(7):680-685.
- Kim NW, Piatyszek MA, Prowse KR, Harley CB, West MD, Ho PL, Coviello GM, Wright WE, Weinrich SL, Shay JW. 1994. Specific association of human telomerase activity with immortal cells and cancer. *Science* 266(5193):2011-2015.

-
- Kohlmann W, Gruber SB. 1993. Lynch Syndrome. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJH, Bird TD, Fong CT, Mefford HC, Smith RJH et al, editors. GeneReviews(R). Seattle (WA).
- Kunkel TA, Erie DA. 2005. DNA mismatch repair. *Annu Rev Biochem* 74:681-710.
- Kwok CT, Vogelaar IP, van Zelst-Stams WA, Mensenkamp AR, Ligtenberg MJ, Rapkins RW, Ward RL, Chun N, Ford JM, Ladabaum U, McKinnon WC, Greenblatt MS, et al. 2014. The MLH1 c.-27C>A and c.85G>T variants are linked to dominantly inherited MLH1 epimutation and are borne on a European ancestral haplotype. *Eur J Hum Genet* 22(5):617-624.
- Lamberti C, Mangold E, Pagenstecher C, Jungck M, Schwering D, Bollmann M, Vogel J, Kindermann D, Nikorowitsch R, Friedrichs N, Schneider B, Houshdaran F, et al. 2006. Frequency of hereditary non-polyposis colorectal cancer among unselected patients with colorectal cancer in Germany. *Digestion* 74(1):58-67.
- Leach FS, Polyak K, Burrell M, Johnson KA, Hill D, Dunlop MG, Wyllie AH, Peltomaki P, de la Chapelle A, Hamilton SR, Kinzler KW, Vogelstein B. 1996. Expression of the human mismatch repair gene hMSH2 in normal and neoplastic tissues. *Cancer Res* 56(2):235-240.

-
- Lee SC, Guo JY, Lim R, Soo R, Koay E, Salto-Tellez M, Leong A, Goh BC. 2005. Clinical and molecular characteristics of hereditary non-polyposis colorectal cancer families in Southeast Asia. *Clin Genet* 68(2):137-145.
- Lenhard B, Sandelin A, Carninci P. 2012. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet* 13(4):233-245.
- Ligtenberg MJ, Kuiper RP, Chan TL, Goossens M, Hebeda KM, Voorendt M, Lee TY, Bodmer D, Hoenselaar E, Hendriks-Cornelissen SJ, Tsui WY, Kong CK, et al. 2009. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. *Nat Genet* 41(1):112-117.
- Lin JC, Jeong S, Liang G, Takai D, Fatemi M, Tsai YC, Egger G, Gal-Yam EN, Jones PA. 2007. Role of nucleosomal occupancy in the epigenetic silencing of the MLH1 CpG island. *Cancer Cell* 12(5):432-444.
- Lynch HT, de la Chapelle A. 2003. Hereditary colorectal cancer. *N Engl J Med* 348(10):919-932.
- Lynch HT, Snyder CL, Shaw TG, Heinen CD, Hitchins MP. 2015. Milestones of Lynch syndrome: 1895-2015. *Nature Reviews Cancer* 15(3):181-194.

MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, Adams

DR, Altman RB, Antonarakis SE, Ashley EA, Barrett JC, Biesecker LG, et al. 2014.

Guidelines for investigating causality of sequence variants in human disease. *Nature*

508(7497):469-476.

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S,

Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV. 2011. Evolution and

classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 9(6):467-477.

Mangold E, Pagenstecher C, Friedl W, Mathiak M, Buettner R, Engel C, Loeffler M,

Holinski-Feder E, Muller-Koch Y, Keller G, Schackert HK, Kruger S, et al. 2005.

Spectrum and frequencies of mutations in MSH2 and MLH1 identified in 1,721

German families suspected of hereditary nonpolyposis colorectal cancer. *Int J Cancer*

116(5):692-702.

Meyerson M. 2000. Role of telomerase in normal and cancer cells. *J Clin Oncol* 18(13):2626-

2634.

Muller-Koch Y, Kopp R, Lohse P, Baretton G, Stoetzer A, Aust D, Daum J, Kerker B, Gross

M, Dietmeier W, Holinski-Feder E. 2001. Sixteen rare sequence variants of the

hMLH1 and hMSH2 genes found in a cohort of 254 suspected HNPCC (hereditary

-
- non-polyposis colorectal cancer) patients: mutations or polymorphisms? *Eur J Med Res* 6(11):473-482.
- Nagasaka T, Rhees J, Kloor M, Gebert J, Naomoto Y, Boland CR, Goel A. 2010. Somatic hypermethylation of MSH2 is a frequent event in Lynch Syndrome colorectal cancers. *Cancer Res* 70(8):3098-3108.
- NHLBI Exome Sequencing Project (ESP) Exome Variant Server B, MD (URL: <http://evs.gs.washington.edu/EVS/>) [accessed October, 2015].
- Nilbert M, Wikman FP, Hansen TV, Krarup HB, Orntoft TF, Nielsen FC, Sunde L, Gerdes AM, Cruger D, Timshel S, Bisgaard ML, Bernstein I, et al. 2009. Major contribution from recurrent alterations and MSH6 mutations in the Danish Lynch syndrome population. *Fam Cancer* 8(1):75-83.
- Nygren AO, Ameziane N, Duarte HM, Vijzelaar RN, Waisfisz Q, Hess CJ, Schouten JP, Errami A. 2005. Methylation-specific MLPA (MS-MLPA): simultaneous detection of CpG methylation and copy number changes of up to 40 sequences. *Nucleic Acids Res* 33(14):e128.
- O'Neill LP, Turner BM. 2003. Immunoprecipitation of native chromatin: NChIP. *Methods* 31(1):76-82.

Orlando V, Paro R. 1993. Mapping Polycomb-repressed domains in the bithorax complex using in vivo formaldehyde cross-linked chromatin. *Cell* 75(6):1187-1198.

Parsons MT, Buchanan DD, Thompson B, Young JP, Spurdle AB. 2012. Correlation of tumour BRAF mutations and MLH1 methylation with germline mismatch repair (MMR) gene mutation status: a literature review assessing utility of tumour features for MMR variant classification. *J Med Genet* 49(3):151-157.

Plazzer JP, Sijmons RH, Woods MO, Peltomaki P, Thompson B, Den Dunnen JT, Macrae F. 2013. The InSiGHT database: utilizing 100 years of insights into Lynch syndrome. *Fam Cancer* 12(2):175-180.

Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, Hogervorst FB, Hoogerbrugge N, Spurdle AB, Tavtigian SV, Iarc Unclassified Genetic Variants Working Group. 2008. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat* 29(11):1282-1291.

Poynter JN, Siegmund KD, Weisenberger DJ, Long TI, Thibodeau SN, Lindor N, Young J, Jenkins MA, Hopper JL, Baron JA, Buchanan D, Casey G. 2008. Molecular characterization of MSI-H colorectal cancer by MLH1 promoter methylation,

-
- immunohistochemistry, and mismatch repair germline mutation screening. *Cancer Epidemiol Biomarkers Prev* 17(11):3208-3215.
- Pugacheva EM, Tiwari VK, Abdullaev Z, Vostrov AA, Flanagan PT, Quitschke WW, Loukinov DI, Ohlsson R, Lobanenko VV. 2005. Familial cases of point mutations in the XIST promoter reveal a correlation between CTCF binding and pre-emptive choices of X chromosome inactivation. *Human molecular genetics* 14(7):953-965.
- Raevaara TE, Korhonen MK, Lohi H, Hampel H, Lynch E, Lönnqvist KE, Holinski-Feder E, Sutter C, McKinnon W, Duraisamy S. 2005. Functional significance and clinical phenotype of nontruncating mismatch repair variants of MLH1. *Gastroenterology* 129(2):537-549.
- Rasmussen LJ, Heinen CD, Royer-Pokora B, Drost M, Tavtigian S, Hofstra RM, de Wind N. 2012. Pathological assessment of mismatch repair gene variants in Lynch syndrome: past, present, and future. *Hum Mutat* 33(12):1617-1625.
- Raval A, Tanner SM, Byrd JC, Angerman EB, Perko JD, Chen S-S, Hackanson B, Grever MR, Lucas DM, Matkovic JJ. 2007. Downregulation of Death-Associated Protein Kinase (DAK1) in Chronic Lymphocytic Leukemia. *Cell* 129(5):879-890.

-
- Rhees J, Arnold M, Boland CR. 2014. Inversion of exons 1-7 of the MSH2 gene is a frequent cause of unexplained Lynch syndrome in one local population. *Fam Cancer* 13(2):219-225.
- Riethoven JJ. 2010. Regulatory regions in DNA: promoters, enhancers, silencers, and insulators. *Methods Mol Biol* 674:33-42.
- Ronaghi M, Uhlen M, Nyren P. 1998. A sequencing method based on real-time pyrophosphate. *Science* 281(5375):363-365.
- Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B. 2004a. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* 32(Database issue):D91-4.
- Sandelin A, Wasserman WW, Lenhard B. 2004b. ConSite: web-based prediction of regulatory elements using cross-species comparison. *Nucleic Acids Res* 32(Web Server issue):W249-52.
- Sander JD, Joung JK. 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* 32(4):347-355.
- Sapienza C, Winter RM, Willard HF. 1997. A promoter mutation in the XIST gene in two unrelated families with skewed X-chromosome inactivation. *Nature genetics* 17:353.

-
- Schafmayer C, Buch S, Egberts JH, Franke A, Brosch M, El Sharawy A, Conring M, Koschnick M, Schwiedernoch S, Katalinic A, Kremer B, Folsch UR, et al. 2007. Genetic investigation of DNA-repair pathway genes PMS2, MLH1, MSH2, MSH6, MUTYH, OGG1 and MTH1 in sporadic colon cancer. *Int J Cancer* 121(3):555-558.
- Schofield L, Grieu F, Amanuel B, Carrello A, Spagnolo D, Kiraly C, Pachter N, Goldblatt J, Platell C, Levitt M, Stewart C, Salama P, et al. 2014. Population-based screening for Lynch syndrome in Western Australia. *Int J Cancer* 135(5):1085-1091.
- Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 29(1):308-311.
- Shin KH, Shin JH, Kim JH, Park JG. 2002. Mutational analysis of promoters of mismatch repair genes hMSH2 and hMLH1 in hereditary nonpolyposis colorectal cancer and early onset colorectal cancer patients: identification of three novel germ-line mutations in promoter of the hMSH2 gene. *Cancer Res* 62(1):38-42.
- Sloane MA, Nunez AC, Packham D, Kwok CT, Suthers G, Hesson LB, Ward RL. 2015. Mosaic Epigenetic Inheritance as a Cause of Early-Onset Colorectal Cancer. *JAMA Oncol* 1(7):953-957
- Smale ST. 2010. Luciferase assay. *Cold Spring Harb Protoc* 2010(5):pdb prot5421.

Snowsill T, Huxley N, Hoyle M, Jones-Hughes T, Coelho H, Cooper C, Frayling I, Hyde C.

2014. A systematic review and economic evaluation of diagnostic strategies for Lynch syndrome. *Health Technol Assess* 18(58):1-406.

Steinke V, Holzapfel S, Loeffler M, Holinski-Feder E, Morak M, Schackert HK, Gorgens H,

Pox C, Royer-Pokora B, von Knebel-Doeberitz M, Buttner R, Propping P, et al. 2014.

Evaluating the performance of clinical criteria for predicting mismatch repair gene mutations in Lynch syndrome: a comprehensive analysis of 3,671 families. *Int J*

Cancer 135(1):69-77.

Suter CM, Martin DI, Ward RL. 2004. Germline epimutation of MLH1 in individuals with

multiple cancers. *Nat Genet* 36(5):497-501.

Syngal S, Fox EA, Li C, Dovidio M, Eng C, Kolodner RD, Garber JE. 1999. Interpretation of

genetic test results for hereditary nonpolyposis colorectal cancer: implications for

clinical predisposition testing. *JAMA* 282(3):247-253.

ten Broeke SW, Brohet RM, Tops CM, van der Klift HM, Velthuisen ME, Bernstein I,

Capella Munar G, Gomez Garcia E, Hoogerbrugge N, Letteboer TG, Menko FH,

Lindblom A, et al. 2015. Lynch syndrome caused by germline PMS2 mutations:

delineating the cancer risk. *J Clin Oncol* 33(4):319-325.

Thompson BA, Goldgar DE, Paterson C, Clendenning M, Walters R, Arnold S, Parsons MT, Michael DW, Gallinger S, Haile RW, Hopper JL, Jenkins MA. 2013. A multifactorial likelihood model for MMR gene variant classification incorporating probabilities based on sequence bioinformatics and tumor characteristics: a report from the Colon Cancer Family Registry. *Hum Mutat* 34(1):200-209.

Thompson E, Meldrum CJ, Crooks R, McPhillips M, Thomas L, Spigelman AD, Scott RJ. 2004. Hereditary non-polyposis colorectal cancer and the role of hPMS2 and hEXO1 mutations. *Clin Genet* 65(3):215-225.

Thompson BA, Spurdle AB, Plazzer JP, Greenblatt MS, Akagi K, Al-Mulla F, Bapat B, Bernstein I, Capella G, den Dunnen JT, du Sart D, Fabre A, et al. 2014. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nat Genet* 46(2):107-115.

Tomkins DJ, McDonald HL, Farrell SA, Brown CJ. 2002. Lack of expression of XIST from a small ring X chromosome containing the XIST locus in a girl with short stature, facial dysmorphism and developmental delay. *European journal of human genetics: EJHG* 10(1):44-51.

Tost J, Gut IG. 2007. Analysis of gene-specific DNA methylation patterns by pyrosequencing technology. *Methods Mol Biol* 373:89-102.

Tulupova E, Kumar R, Hanova M, Slysikova J, Pardini B, Polakova V, Naccarati A,

Vodickova L, Novotny J, Halamkova J, Hemminki K, Vodicka P. 2008. Do polymorphisms and haplotypes of mismatch repair genes modulate risk of sporadic colorectal cancer? *Mutat Res* 648(1-2):40-45.

Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Ruschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R, Hamilton SR, Hiatt RA, et al. 2004. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst* 96(4):261-268.

van der Klift H, Wijnen J, Wagner A, Verkuilen P, Tops C, Otway R, Kohonen-Corish M, Vasen H, Oliani C, Barana D, Moller P, Delozier-Blanchet C, et al. 2005. Molecular characterization of the spectrum of genomic deletions in the mismatch repair genes MSH2, MLH1, MSH6, and PMS2 responsible for hereditary nonpolyposis colorectal cancer (HNPCC). *Genes Chromosomes Cancer* 44(2):123-138.

Van Roon EH, van Puijenbroek M, Middeldorp A, van Eijk R, de Meijer EJ, Erasmus D, Wouters KA, van Engeland M, Oosting J, Hes FJ, Tops CM, van Wezel T, Boer JM, Morreau H. 2010. Early onset MSI-H colon cancer with MLH1 promoter methylation, is there a genetic predisposition? *BMC Cancer* 10: 180.

Vasen HF. 2007. Review article: The Lynch syndrome (hereditary nonpolyposis colorectal cancer). *Aliment Pharmacol Ther* 26 Suppl 2:113-126.

Wagner A, van der Klift H, Franken P, Wijnen J, Breukel C, Bezrookove V, Smits R, Kinarsky Y, Barrows A, Franklin B, Lynch J, Lynch H, et al. 2002. A 10-Mb paracentric inversion of chromosome arm 2p inactivates MSH2 and is responsible for hereditary nonpolyposis colorectal cancer in a North-American kindred. *Genes Chromosomes Cancer* 35(1):49-57.

Ward RL, Dobbins T, Lindor NM, Rapkins RW, Hitchins MP. 2013a. Identification of constitutional MLH1 epimutations and promoter variants in colorectal cancer patients from the Colon Cancer Family Registry. *Genet Med* 15(1):25-35.

Ward RL, Hicks S, Hawkins NJ. 2013b. Population-based molecular screening for Lynch syndrome: implications for personalized medicine. *J Clin Oncol* 31(20):2554-2562.

Watson P, Lynch HT. 1994. The tumor spectrum in HNPCC. *Anticancer Res* 14(4B):1635-1639.

Watson P, Vasen HF, Mecklin JP, Bernstein I, Aarnio M, Jarvinen HJ, Myrhoj T, Sunde L, Wijnen JT, Lynch HT. 2008. The risk of extra-colonic, extra-endometrial cancer in the Lynch syndrome. *Int J Cancer* 123(2):444-449.

Xiong Z, Laird PW. 1997. COBRA: a sensitive and quantitative DNA methylation assay.

Nucleic Acids Res 25(12):2532-2534.

Yan H, Jin H, Xue G, Mei Q, Ding F, Hao L, Sun SH. 2007. Germline hMSH2 promoter mutation in a Chinese HNPCC kindred: evidence for dual role of LOH. Clin Genet 72(6):556-561.

Yin H, Xue W, Chen S, Bogorad RL, Benedetti E, Grompe M, Koteliensky V, Sharp PA, Jacks T, Anderson DG. 2014. Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. Nat Biotechnol 32(6):551-553.

Yoon S, Park TS, Kim NK, Lee KA, Kim J, Song J, Kim BY, Choi JR. 2009. Three new nonsense mutations of MLH1 and MSH2 genes in Korean families with hereditary nonpolyposis colorectal cancer. Cancer Genet Cytogenet 188(2):61-64.

Zavodna K, Bujalkova M, Krivulcik T, Alemayehu A, Skorvaga M, Marra G, Fridrichova I, Jiricny J, Bartosova Z. 2006. Novel and recurrent germline alterations in the MLH1 and MSH2 genes identified in hereditary nonpolyposis colorectal cancer patients in Slovakia. Neoplasia 53(4):269-276.

Zhong X, Arita M, Yamada K, Sugiyama H, Tan K, Kanazawa S, Koike J, Teramoto T, Hemmi H. 2007. A single nucleotide substitution (-107C-->G) in the hMLH1

promoter found in colorectal cancer population reduces transcriptional activity.

Biochem Genet 45(9-10):671-681.

Figure Legends

Fig. 1. Abrogated allele-specific expression caused by altered transcription factor

binding. (a) The structure of a gene showing the location of the core (black) and proximal (gray) promoter. RNAP II is recruited to the core promoter by sequence specific TFs. Arrows indicate possible looping interactions between proximal and distal regulatory enhancers that can also mediate recruitment of RNAP II. Insulators bound by chromatin regulators help to maintain the chromatin structure of a locus. The loss of expression from a variant allele may be caused by the loss of a transcriptional activator (b) or the gain of a transcriptional repressor (c) and this may be associated with a constitutional secondary epimutation (d) in a MMR gene.

Fig. 2. Class 2, 3 and class 4 promoter variants in the MMR genes. The four MMR genes (*MLH1* – NM_000249.3, *MSH2* – NM_000251.2, *MSH6* – NM_000179.2, and *PMS2* – NM_000535.5) showing the location of class 2, 3 and 4 promoter variants from the InSiGHT variants database. For cDNA numbering the nucleotide numbering uses +1 as the A of the

ATG translation initiation codon in the reference sequence, with the initiation codon as codon

1. The region depicted by the black bar below each CpG island is shown in detail below each gene. ENCODE TF binding sites in humans with a Z-score cutoff of 1.64 ($p < 0.05$) are shown for the forward (white box) and reverse strand (light gray box).

Fig. 3. A proposed guide to classify MMR gene promoter variants using the InSiGHT 5-tiered standardized classification scheme. The flow diagram illustrates the integration of current InSiGHT classification rules with proposed guidelines to investigate the functional effect of a promoter variant. The InSiGHT criteria (1-4) and variant classes (class 1-5) are described in the main text. The thick vertical line in the upper part of the figure is the pathway of a promoter variant. Experiments do not need to be performed in a set order and it is acknowledged that research groups will gather data in a different order to that specified in the workflow. *, a promoter variant associated with a secondary epimutation should also fulfil criteria 1-4. †, the odds for segregation with disease must be >5:1 for class 4 and >10:1 for class 5. ‡, multifactorial likelihood analysis only involves data from criteria 2 and/or 3. Posterior probabilities of pathogenicity derived from this analysis use the cut-offs reported in the InSiGHT classification rules: Class 1, <0.001; Class 2, 0.001-0.049; Class 3, 0.05-0.949; Class 4, 0.95-0.99; Class 5, >0.99.

Author Manuscript

This article is protected by copyright. All rights reserved.

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Liu, Q;Thompson, BA;Ward, RL;Hesson, LB;Sloane, MA

Title:

Understanding the Pathogenicity of Noncoding Mismatch Repair Gene Promoter Variants in Lynch Syndrome

Date:

2016-05-01

Citation:

Liu, Q., Thompson, B. A., Ward, R. L., Hesson, L. B. & Sloane, M. A. (2016). Understanding the Pathogenicity of Noncoding Mismatch Repair Gene Promoter Variants in Lynch Syndrome. HUMAN MUTATION, 37 (5), pp.417-426. <https://doi.org/10.1002/humu.22971>.

Persistent Link:

<http://hdl.handle.net/11343/291076>