A calibrated cell-based functional assay to aide classification of *MLH1* DNA mismatch repair gene variants

Abhijit Rath¹, Alexander A. Radecki¹, Kaussar Rahman¹, Rachel B. Gilmore², Jonathan R. Hudson², Matthew Cenci¹, Sean V. Tavtigian³, James P. Grady⁴, Christopher D. Heinen^{1*}

¹Center for Molecular Oncology, UConn Health, Farmington, CT

²Department of Genetics and Genome Sciences, UConn Health, Farmington, CT

³Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, UT

⁴Connecticut Institute for Clinical and Translational Science, UConn Health, Farmington, CT

* email: cheinen@uchc.edu; tel: 860-679-8859

ABSTRACT

PURPOSE: Functional assays provide important evidence for classifying the disease significance of germline variants in the DNA mismatch repair genes. We sought to develop a cell-based approach for testing the function of variants of uncertain significance (VUS) in the *MLH1* gene.

METHODS: Using CRISPR gene editing, we knocked-in *MLH1* VUS into the endogenous *MLH1* loci in human embryonic stem cells. We examined their impact at the RNA and protein level, including their ability to maintain stability of microsatellite sequences and instigate a DNA damage response. We calibrated these assays by testing well-established pathogenic and benign control variants.

RESULTS: Five VUS resulted in functionally abnormal protein, 15 VUS resulted in functionally normal protein, and one VUS showed mixed results. Furthermore, we converted the functional outputs into a single odds in favor of pathogenicity score for each VUS.

CONCLUSION: Our CRISPR-based functional assay successfully models phenotypes observed in patients in a cellular context. Using this approach, we generated evidence for or against pathogenicity for utilization by variant classification expert panels. Ultimately, this information will assist in proper diagnosis and disease management for suspected Lynch syndrome patients.

INTRODUCTION

Lynch syndrome (LS; OMIM: 120435) is an autosomal dominant cancer predisposition condition and the predominant cause of inherited colorectal cancer (CRC)¹. LS patients constitute about 2-4% of all CRC cases². Historically, LS has been associated with increased likelihood of developing colorectal or endometrial cancer. However, recent evidence suggests far greater implications of LS diagnosis in increased predisposition to cancer across multiple tissue types³. An LS diagnosis has a profound impact on the clinical management of the patient, but also for instituting proactive cancer screening of “at-risk” relatives. LS diagnosis may be considered due to an early age of cancer presentation, family history, and certain clinical features, although a confirmatory diagnosis of LS depends upon identification of an inactivating germline pathogenic variant in one of the canonical DNA mismatch repair (MMR) genes; *MSH2* (OMIM: 609309), *MLH1* (OMIM: 120436), *MSH6* (OMIM: 600678), and *PMS2* (OMIM: 600259)¹. Therefore, in recent clinical practice, next-generation sequencing-based identification of MMR gene sequence variants is routinely employed⁴. However, accurate interpretation of the pathogenic significance of gene variants creates a bottleneck, hindering conclusive clinical diagnosis of LS.

To this end, the International Society for Gastrointestinal Hereditary Tumors (InSiGHT), proposed a tiered classification scheme for germline MMR gene variants based on clinical or functional evidence for each variant⁵. Similarly, the American College of Medical Genetics (ACMG) and Association for Molecular Pathology (AMP) generated guidelines for clinical variant interpretation in general based on the same kinds of evidence⁶. For some non-truncating variants, strong family history, variant segregation data, tumor pathology, and population frequency alone is sufficient for classification as benign or pathogenic. However, for nearly 30% of MMR variants identified in the InSiGHT variant database insufficient information exists, resulting in classification as a variant of uncertain significance (VUS)^{5,7}. The problem is most

evident in the case of a missense change, which constitutes at least 80% of the reported VUS in the four major MMR genes in the ClinVar database.

In the absence of sufficient patient and family history data, laboratory generated functional data can also provide evidence to enable successful classification⁸⁻¹⁰. Thus, the need exists for well-established and validated functional assays. Fortunately, much is known about the normal cellular function of the MMR pathway to provide the basis for such assays. MMR protects genomic integrity by removing a wrongly incorporated nucleotide from the daughter strand during DNA replication¹¹⁻¹⁴. MMR also initiates apoptotic signaling in response to persistent DNA adducts, as encountered during treatment of replicating cells with a DNA alkylating agent^{15,16}. Over the course of two decades, multiple laboratories have employed wide-ranging assays for functional interpretation of MMR-gene missense variants^{9,10}. However, often a lack of cellular context, use of artificial overexpression systems, or use of non-human model systems prevents a high confidence appraisal of the variants in question.

To address these concerns, we developed a human embryonic stem cell (hESC)-based functional assay using clustered regularly interspaced short palindromic repeats (CRISPR) genome editing to probe MMR gene missense variants¹⁷. In the current study, we have improved and expanded the use of this approach to assess a large number of *MLH1* VUS, as *MLH1* is the most commonly affected gene in LS patients⁷. We evaluated the ability of these variants to impact MMR-associated repair and damage response functions. In the design of our assay, we followed recommendations from members of the ClinGen Sequence Variant Interpretation Working Group on the use of functional assays in clinical variant interpretation¹⁸. The strengths of our approach include an *in cellulo* environment which better recapitulates the function of the MMR proteins, the use of genome editing to introduce the variants into the endogenous *MLH1* loci - allowing for simultaneous examination of the effects on RNA and

protein, the use of technical and biological replicates to ensure rigor and reproducibility, and the use of known pathogenic and benign variants to allow for assay calibration. By testing a series of established control variants, we were able to use a quantitative framework to convert our functional results into a Bayesian estimate for odds in favor of pathogenicity (OddsPath_Functional) for each of the VUS¹⁹. This outcome will allow variant interpretation expert panels to readily integrate our data into the existing ACMG/AMP classification scheme.

MATERIALS AND METHODS

Generation of *MLH1* variant expressing cell lines

H1 hESCs were obtained from the UConn Stem Cell Core and scored to have a normal karyotype. Cells were cultured on growth factor reduced Matrigel (Corning # 356231) coated plates, fed daily with PeproGrow hESC medium (PeproTech), and maintained at 37°C at 5% CO₂ and 95% RH. They were routinely passaged either by microdissection, ReleSR (Stemcell Technologies), or StemPro Accutase Cell Dissociation Reagent (ThermoFisher Scientific) upon reaching around 80% confluency. For each *MLH1* (NCBI Gene ID: 4292, RefSeq: NM_000249.4) variant (Table S1), a suitable single guide RNA (sgRNA) to the appropriate genomic target and near a protospacer adjacent motif (PAM) was chosen using an *in silico* program (CRISPOR²⁰) (Table S2). Homozygous insertion of the given single nucleotide variant was performed as described previously¹⁷ with the following modifications: a ribonucleoprotein (RNP)-based targeting approach was used with either a Cas9 or Cas12a recombinant enzyme pre-incubated with the variant-specific Alt-R sgRNAs/crRNAs as per manufacturer's recommendations (IDT). Following transfection with the Amaxa Stem Cell Nucleofector Kit 2 (Lonza) using an Amaxa Nucleofector II machine, cells underwent selection with 1 µg/mL puromycin (Sigma) for a period of 48 h to isolate single cell clones. Surviving cells were treated with CloneR (STEMCELL Technologies) as recommended. Approximately 32-48 surviving

colonies were picked in a PCR hood for each experiment and their genotype confirmed by Sanger sequencing (Table S3 and Fig. S1).

Details of off-target analysis, immunoblotting, exon exclusion assays, cell survival assay, microsatellite instability (MSI) analysis and statistical analyses can be found in the Supplementary Materials and Methods.

RESULTS:

Generation of *MLH1* variants

We selected 11 known benign and 11 known likely pathogenic/pathogenic control variants and 21 VUS from the InSiGHT database (Fig. 1). The classification of these variants was performed by the InSiGHT Sequence Variant Interpretation committee, and reassessed without the use of prior functional data in order to avoid circularity in our analyses (Table S1).

We employed CRISPR-Cas gene editing to engineer homozygously targeted cell clones for each variant in wild-type H1 hESCs (WT) (Fig. S1). We also generated an *MLH1* knockout cell line (KO) via CRISPR-Cas targeting of exon 2. This targeting resulted in a cell clone in which the *MLH1* protein was lost (Fig. 2a). In general for the variants, we observed a wide-ranging targeting efficiency (Table S4), but the median efficiency was ~10%, in line with our previous report¹⁷. Unlike our previous work with *MSH2*, targeting of *MLH1* warranted certain important modifications. First, we opted for a ribonucleoprotein-based approach using commercially sourced sgRNAs, recombinant Cas nucleases, and single strand repair template carrying the single nucleotide variant of interest (Table S2). This approach reduced the reagent preparation time while also drastically reducing the chances of generating unwanted off-target effects²¹. Second, the final 43 cell lines contained only a single nucleotide change to avoid the potential for unintended splicing defects caused by the addition of silent mutations. As a common

Fig 1. MLH1 variants under study. Representative image of the MLH1 variants in this study mapped to the N-terminal structure of human MLH1 (Protein Data Bank ID code 4P7A) and C-terminal structure of *Saccharomyces cerevisiae* MLH1 (Protein Data Bank ID code 4E4W), connected by a theoretical linker from the human protein. **(a)** The known benign (green) and pathogenic (red) variants selected for functional evaluation. **(b)** The missense variants of uncertain significance (VUS) examined in this study.

strategy, introduction of a silent mutation in the PAM site can prevent re-cutting of the sgRNA target sequence by Cas nuclease, and thus increase the chances of recovery of the correctly edited clone. However, in the context of *MLH1*, it may confuse variant interpretation owing to the multiple relatively short exons in *MLH1* and the proclivity for single nucleotide variants of *MLH1* to impart a splicing defect²². Instead, we were able to retain a similar editing efficiency by strategically selecting PAM sites, such that the target nucleotide was positioned within the seed sequence of the Cas nuclease target, thereby preventing recurrent cleavage without introduction of an undesired change. Third, use of both Cas9 and Cas12a nucleases offered a greater flexibility to engineer a greater number of variants based on availability of the different PAM sites.

To detect evidence of off-target cleavage in the clonally generated cell lines, we bioinformatically identified the most likely off-target sites for each sgRNA. The top five off-target sites were PCR amplified and sequence verified using Sanger sequencing. We did not detect any off-target cleavage in agreement with previous observations from us and others (Table S5)^{17,23,24}.

Impact of missense variants on MLH1-PMS2 steady state level

Loss of stability of MLH1-PMS2 protein product is one way that MMR function is impaired by a pathogenic missense change^{9,10}. Thus, as an initial evaluation, we examined steady state levels of MLH1-PMS2 heterodimer in the variant cell lines. As shown in Fig. 2a., we observed a high degree of variability in MLH1 and PMS2 protein levels both between and within the control variants and VUS compared to WT cells. As expected for known benign controls, MLH1 protein levels remained greater than 80% for most of the 11 variants with the exception of V213M, K618A, S692P, and V716M, which show a greater decrease relative to the WT levels. For the known pathogenic controls, six variants (S44F, G67R, I107R, T117M, R265S, and R687W)

Fig. 2. Steady state levels of MLH1 and PMS2 proteins and aberrant splicing variants. (a)

Representative immunoblots show the steady state levels of MLH1 and PMS2 in benign, pathogenic, and VUS expressing cell lines. Quantification of the protein levels with respect to the WT levels are shown after normalizing for loading. WT, wild-type hESCs; KO, MLH1 knockout hESCs **(b)** Reverse transcribed total RNA from R265S, S265R and A42V cells was PCR amplified to detect the presence of variant containing exons. Representative agarose gel images showing aberrant splicing in indicated cell lines. CHX= Cycloheximide.

showed drastically reduced MLH1-PMS2 levels, while the remaining five (I19F, D41H, R100P, S247P, and I276R) retained relatively stable levels. For the VUS, we observed markedly reduced MLH1 levels (<20% of WT) for C39R, D63N, L73P, and A111P, while the remaining 16 variants showed a varying degree of expression. Since MLH1-PMS2 exists as an obligate heterodimer²⁵, an MLH1 variant impacting this interaction can also ultimately affect PMS2 stability. Not surprisingly then, we found that the pattern of PMS2 levels among the variants mostly followed that observed for MLH1.

Variant induced impact on MLH1 mRNA splicing:

One manner in which MLH1 protein levels may be affected by variants is through disruption of mRNA splicing. An advantage of our CRISPR editing-based approach is that we can assess the impact of any variant at both the protein and RNA level. Thus, for each VUS that showed a dramatically reduced protein level (C39R, A42V, D63N, N64S, L73P, A111P, G244V), we examined the impact of the variant on its endogenous mRNA exon. We also examined a set of pathogenic controls that displayed reduced protein levels (S44F, R100P, I107R, T117M, S247P, and R265S). We observed evidence of aberrant splicing for two variants, the pathogenic control R265S located in exon 10 and the VUS A42V located in exon 2. As shown in Fig. **2b**, we failed to detect the expected full-length PCR product spanning exons 7-11 in cDNA from R265S cells, but rather detected two lower migrating bands. Upon gel extraction and Sanger sequencing of the aberrant bands, we confirmed a predominant excision of exon 10 leading to shortening of the amplicon (Fig. S2a). We also detected a fraction of mRNA where both exons 9 and 10 were excluded. Of note, the c.793C>A (R265S) variant has been reported to confer a splicing defect in patient-derived lymphocytes²⁶, thus suggesting our CRISPR-based approach can detect variants that impact splicing. To validate the specificity of the variant for this aberrant splicing phenotype, we re-edited the R265S cell line using CRISPR-Cas targeting to re-establish the WT sequence (S265R). We observed a nearly complete rescue of the splicing defect and MLH1

steady-state protein level in the S265R cell line (Fig. **2b** and S3a). The A42V cells, on the other hand, still display a majority of the full-length PCR product spanning exons 1-3 as in WT cells, however, a weaker lower migrating band was observed suggesting exon exclusion (Fig. **2b**). Sanger sequencing of the aberrant splice product confirmed exclusion of the exon 2 (Fig. S2b). The remaining variants tested all displayed normal exon splicing, suggesting that the changes in protein levels are likely due to impacts of the variant on protein stability (Fig. S4 and Table S6).

Variant effect on DNA damage signaling

Mammalian cells mount a cell cycle arrest and/or apoptotic response upon treatment with a DNA alkylating agent in an MMR-dependent fashion^{15,16}. To assess the ability of *MLH1* VUS expressing cells to successfully instigate such a response, we measured cell survival 48 h post treatment with the S_N1 alkylating agent MNNG. We also subjected WT and *MLH1* KO cells as well as the known benign and pathogenic controls to MNNG. As expected, greater than 80% of *MLH1* KO cells survived MNNG treatment, whereas only 15% of WT cells did (Fig **3**). The eleven benign controls responded similarly to the WT cells, though two cell lines, E578G and K618A, displayed slightly increased resistance to MNNG. The known pathogenic variant expressing cells demonstrated resistance to treatment similar to KO cells. To evaluate the VUS, we used statistical clustering to identify those variants whose survival level was similar to either benign or pathogenic controls. Two, non-overlapping clusters were generated demonstrating a clean separation of the results for the benign and pathogenic controls. Of the 21 *MLH1* VUS examined, C39R, D63N, L73P, R100Q, A111P, and G244V showed resistance to MNNG, clustering with the pathogenic controls. The remaining 16 VUS clustered with the benign controls. Two variant lines, A42V and S627T, demonstrated a slightly increased resistance to MNNG compared to WT cells, though still segregated with the benign cluster. As an additional control to confirm that the change in damage response in a pathogenic variant line was indeed due to the sequence variant, we tested the response to MNNG in the S265R revertant cell line.

Fig. 3. The mismatch repair-dependent damage response in variant cell lines. Cell survival after 48 hr treatment with 2 μM of the DNA alkylating agent *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine. The values are represented as the mean ± SEM; n=3-5. A statistical clustering analysis was performed which grouped the variants into two populations based on their percent survival (blue and orange).

We found that re-editing this variant back to the WT sequence restored the normal response to DNA damage (Fig. S3b).

Examining microsatellite instability (MSI) in *MLH1* VUS

Deficient MMR function leads to instability in microsatellite repeats (MSI) and the presence of MSI is the best-characterized molecular evidence for LS associated tumors²⁷. We examined MSI in the *MLH1* VUS expressing cells as another read-out for MMR function. The benign and pathogenic variants were also examined to serve as controls. We passaged each cell line for 15-20 generations and then isolated 32 single cell clones to evaluate for MSI at five mononucleotide repeats (Table 1 and Table S7). We observed few or no unstable clones in the WT and benign control cells, with higher numbers of unstable clones in the KO and pathogenic controls. To more quantitatively determine if MSI can be used to distinguish benign and pathogenic variants, we again employed a statistical clustering approach. Using the results for each cell line from all five MSI markers, we asked whether the benign and pathogenic controls would segregate into two separate clusters. We observed this distinct clustering of controls for three markers, BAT-25, BAT-26 and NR-27 with only two discordant results (G67R for BAT-26 and NR-27) (Fig. S5). However, the controls failed to segregate into two separate clusters for the NR-21 and NR-22 markers (Fig. S5), suggesting that these MSI markers may not be informative for distinguishing variants.

Odds of Pathogenicity Scores

In order to convert the results from our functional assays into a value that can be utilized in the ACMG/AMP classification scheme, we utilized a Bayesian framework that allows for the expression of evidence into a quantitative OddsPath score¹⁹. To generate this score, we obtained

Table 1 – Microsatellite instability in MLH1 variant-expressing cell lines

Variant/MSI Marker	NR-21^a	NR-22^a	NR-27^a	BAT-25^a	BAT-26^a
WT	0/30	0/31	0/29	0/30	0/31
KO	28/32	30/30	31/31	15/31	31/31
<i>Benign</i>					
D132H	0/30	0/31	0/32	0/32	0/32
V213M	1/32	0/32	0/30	1/31	0/32
I219V	5/32	0/32	0/30	0/32	1/32
E268G	1/32	0/32	0/32	0/32	0/31
S406N	6/32	0/32	0/28	0/29	0/32
E578G	2/32	0/32	1/27	1/28	0/32
K618A	0/32	0/32	0/32	1/32	0/32
Q689R	5/30	0/30	0/29	0/29	0/31
S692P	7/32	0/31	6/30	0/29	2/31
V716M	0/31	0/31	1/31	0/31	0/30
H718Y	4/32	0/32	0/32	0/32	4/32
<i>VUS</i>					
R9W	6/32	0/32	11/32	0/31	2/32
C39R	25/32	14/32	32/32	17/32	21/32
A42V	2/32	1/32	0/32	0/32	0/32
D63N	13/30	27/29	13/30	23/30	28/31
N64S	0/32	0/32	0/32	0/32	0/32
L73P	24/31	5/31	22/31	21/29	16/32
R100Q	13/32	2/32	9/31	9/32	6/32
A111P	15/32	14/32	15/32	16/32	20/32
H112D	0/32	5/32	0/31	0/32	0/31
L135V	2/32	0/32	0/32	1/32	0/32
E234G	3/32	0/32	3/31	0/32	0/32
G244V	25/28	6/28	15/26	15/26	18/27
K286Q	8/31	0/31	4/28	0/31	0/31
N338S	0/32	1/32	0/32	0/32	0/32
G454R	0/32	0/32	0/32	0/32	0/32
R474P	0/32	0/32	0/31	0/32	0/32
R474Q	6/31	0/30	0/29	0/30	0/31
V506A	0/31	0/29	0/25	0/29	0/30
Q542L	0/27	0/31	0/31	0/31	0/32
S627T	10/29	0/29	0/29	0/29	0/32
D748A	1/32	0/32	0/30	0/32	0/32
<i>Pathogenic</i>					
I19F	21/32	11/32	23/32	13/32	30/32
D41H	13/32	21/31	15/30	16/31	23/32
S44F	15/31	10/31	29/31	30/31	28/31
G67R	27/32	7/32	13/31	16/30	10/31
R100P	15/31	5/31	22/32	24/32	18/32
I107R	15/32	18/29	17/28	12/25	25/31
T117M	14/29	11/29	22/28	17/25	22/29
S247P	31/32	10/32	17/29	11/32	23/32
R265S	15/32	7/31	28/32	23/28	29/30
I276R	17/32	13/31	25/25	17/30	19/31
R687W	14/30	6/30	23/28	11/27	25/31

^a Ratio of clones with altered alleles compared with WT sequence to total clones tested

previously established OddsPath values for the 11 pathogenic and 11 benign variants that were derived based on existing personal and clinical data (Table S1). The contribution of any prior functional assay was removed from the calculations in order to avoid circularity¹⁸. To calibrate our functional assays, linear regression analyses were performed with the OddsPath score as the dependent variable and the assay readout as the independent variable. We initially obtained a single OddsPath score from the MSI assay for each VUS by averaging the results from the three informative MSI markers (BAT25, BAT26 and NR27). A separate OddsPath score was derived from the MNNG survival data and then both functional scores were combined to create a single OddsPath_Functional score for each VUS (Table 2). Using the draft ACMG/AMP classification criteria for MMR genes (<https://www.insight-group.org/criteria/>), we assigned the appropriate evidence code corresponding to these OddsPath_Functional scores. Five variants had OddsPath_Functional scores > 18.7 consistent with strong evidence of Pathogenicity (PS3), whereas 15 VUS had OddsPath_Functional scores < 0.05, consistent with strong evidence of a benign classification (BS3). We then combined these new evidence codes with existing clinical, genetic and *in silico* evidence available for each variant to determine a predicted classification. From these results, we expect that 13 VUS will be re-classified (Table 2). One variant that remains a VUS is R100Q, which had an OddsPath_Functional score of 0.778, which is indeterminate for pathogenicity.

DISCUSSION:

The ACMG/AMP guidelines for clinical variant interpretation include the use of functional studies as one means to predict the impact of a genetic variant⁶. A point of emphasis in these guidelines is that researchers should consider the degree to which a functional assay reflects the biological environment and disease mechanism. The CRISPR-based approach to study MMR gene variants described here examines each variant in the context of the human cell, which allowed us to determine the impact of variant proteins in their normal nuclear environment. The use of a

Table 2 Summary of Functional Study Outcomes and Predicted Classifications for VUS

Variant	Protein Stability ^a	Damage Response Signaling ^a	DNA Repair Function ^a	OddsPath_Functional Score ^b	ACMG/AMP Evidence Code ^c	Predicted Classification ^d
R9W	+	+	+	8.70E-5	BS3	VUS
C39R	-	-	-	508	PS3	LP
A42V	+/-	+	+	2.54E-4	BS3	VUS
D63N	-	-	-	1,350	PS3	LP
N64S	+/-	+	+	5.65E-6	BS3	B
L73P	-	-	-	108	PS3	LP
R100Q	+	-	+/-	0.778	Ind	VUS
A111P	-	-	-	379	PS3	LP
H112D	+	+	+	4.58E-6	BS3	VUS
L135V	+	+	+	1.31E-5	BS3	LB
E234G	+	+	+	2.19E-5	BS3	LB
G244V	+/-	-	-	57.5	PS3	LP
K286Q	+	+	+	9.15E-6	BS3	VUS
N338S	+	+	+	7.25E-6	BS3	LB
G454R	+	+	+	1.95E-6	BS3	LB
R474P	+	+	+	8.63E-6	BS3	LB
R474Q	+	+	+	3.11E-5	BS3	LB
V506A	+/-	+	+	2.60E-5	BS3	VUS
Q542L	+/-	+	+	2.55E-6	BS3	VUS
S627T	+	+	+	6.41E-4	BS3	LB
D748A	+	+	+	2.97E-5	BS3	VUS

^a From this study

^b OddsPath_Functional scores calculated as described in the Methods using the known benign and pathogenic variants for calibration

^c ACMG/AMP evidence code derived from OddsPath scores in accordance with the draft ACMG/AMP classification criteria for MMR genes (<https://www.insight-group.org/criteria/>). PS3, strong pathogenic; BS3, strong benign; Ind, indeterminate

^d Predicted ACMG/AMP classification derived from combining OddsPath_Functional score with prior *in silico*, clinical or genetic data (see Table S1) using draft ACMG/AMP classification criteria for MMR genes. LP, likely pathogenic; LB, likely benign; B, benign

cell-based assay also allowed us to examine both repair of endogenous microsatellites and the initiation of a global cell response to DNA damage, both mechanisms that are likely involved in disease etiology^{1,15,28}. By introducing the variant into the endogenous *MLH1* gene loci, we ensured that the *MLH1* endogenous promoter and regulatory sequences control its expression. This also allowed us to assess the effects of the variant on the *MLH1* RNA and protein simultaneously. By doing so, we found that two variants altered mRNA splicing, R265S and A42V. Whereas the R265S variant led to exclusion of exon 10, or 9 and 10, in all transcripts, resulting in an abnormally functioning protein, the A42V variant had only a partial defect, producing enough full-length protein to retain normal function. Combining mRNA analyses with MMR pathway functional testing in the same cells allows us to resolve challenges that arise in interpreting bioinformatic predictions or *ex vivo* assays of mRNA splicing²⁹.

To align the output of functional assays such as ours with the ACMG/AMP guidelines, one of us previously derived a quantitative approach to convert a functional result into an OddsPath score that could then be converted to ACMG/AMP evidence codes for or against pathogenicity¹⁹. Adopting this overall approach, members of InSiGHT created draft ACMG/AMP guidelines specific to *MSH2*, *MSH6*, *MLH1* and *PMS2* (<https://www.insight-group.org/criteria/>). In these guidelines, OddsPath scores > 18.7 indicate strong evidence for pathogenicity (PS3). Based on our data, five VUS can be assigned this evidence strength (C39R, D63N, L73P, A111P and G244V; see Table 2). The remainder of the VUS tested, with one exception, all had OddsPath scores < 0.05, indicative of strong evidence for a benign classification (BS3). However, functional results are not considered to be stand-alone evidence for a pathogenic or benign classification¹⁸. Thus, functional evidence must be combined with at least one other piece of data including *in silico*, clinical or genetic data for the given variant. Using prior data available on the InSiGHT database for the 21 VUS tested here, we generated classification predictions (Table 2). Five variants are expected to reach the level of likely pathogenic, while 7 variants are

expected to reach the level of likely benign. One variant (N64S) reached the level of benign due to another existing piece of strong benign evidence. This variant was reported to co-occur with a known pathogenic *MLH1* variant in a patient who did not demonstrate a Constitutive MMR-Deficiency Disease phenotype (evidence code BS4)³⁰. The 8 remaining variants should still be considered VUS at this time, most of them due to discordance between our functional results and prior *in silico* evidence from MAPP and PolyPhen-2. We have observed some discordance between *in silico* data and functional data in our previous study of *MSH2* variants as well¹⁷, as the *in silico* approaches tend to over-predict pathogenicity. This result further highlights why neither approach should be used alone for direct classification.

The one OddsPath_Functional score exception was for R100Q, which had a score of 0.778, indicating indeterminate evidence for classification. Interestingly the R100Q cells displayed a lack of response to DNA damage similar to a functionally abnormal variant, but only limited MSI. This contrasts with all other functionally abnormal variants, which showed both the loss of damage response and high levels of MSI. Of note, an analogous change in the corresponding amino acid for *Escherichia coli* MutL (R95F) has undergone interrogation in single molecule studies³¹. Because of its location in the hinge region of the ATP lid of the MutL/MLH1 protein, the R95F variant inhibits ATP binding. While this variant does not prevent ternary complex formation between MutL, MutS, and DNA, it does impact the formation of MutL sliding clamps on DNA. Future studies will be needed to determine whether the R100Q variant similarly reduces the levels of MLH1-PMS2 sliding clamps on DNA. Why this variant appears so disruptive for the DNA damage signaling function of MLH1, but not quite as disruptive for repair of microsatellite sequences is unclear. One hypothesis is that the damage signaling function requires more efficient MLH1-PMS2 sliding clamp formation compared to standard microsatellite repair. This would be consistent with previous observations that the damage signaling function is more sensitive to significantly reduced levels of MLH1 expression than standard repair³².

However, the effect of MLH1-PMS2 protein levels is not quite as clear from our data. A number of variants appear to impact protein levels that then result in reduced MMR function, however, some variants show significantly reduced protein levels, yet maintain normal MMR function (Table 2). This includes a Class 1 benign variant (K618A) along with several VUS that have OddsPath_Functional scores consistent with benign interpretation. Interestingly, of those VUS with reduced protein levels, all but one show slightly poorer DNA damage responses than the wild-type protein including two variants that are borderline intermediate in their response (A42V and S627T; Fig. 3). Yet, all of these variants have very stable microsatellites. Overall, these results suggest that the DNA damage response may be more sensitive to the levels of MLH1-PMS2 sliding clamp formation that is different from regular repair. The potential for a mechanistic separation between repair and the damage response, particularly as seen with the R100Q variant, will be interesting to examine in future studies to help clarify the molecular mechanism of the MMR-dependent damage signaling response. In contrast to R100Q, another variant tested here, G67R, had previously been reported as a separation of function allele when engineered in mice³³. *Mlh1*^{G67R/G67R} mouse embryonic fibroblasts retained an apoptotic response to the DNA crosslinking agent cisplatin, even though the *Mlh1*^{G67R/G67R} animals developed tumors marked by MSI similar to *Mlh1*^{-/-} mice. However, in our studies, we found that the human G67R variant led to both increased MSI and a loss of DNA damage sensitivity. While species-specific differences are one possible explanation, it may also speak to differences in the mechanism of MMR-dependent signaling generated by crosslinking damage compared to DNA alkylation.

In summary, our CRISPR based functional assays provide a means to assess the effects of MMR gene VUS in a human cellular context. By calibrating the assays with multiple known pathogenic and benign controls, an OddsPath_Functional score can be derived for direct use in variant classification schemes. While we have provided predictions for how these results may

impact classification, ultimately, this task will require review by an expert panel such as the InSiGHT variant interpretation committee⁵. The challenge that remains is to perform these functional assays in a more high-throughput manner in order to more rapidly address the many remaining MMR gene VUS. A recent high-throughput approach examining *MSH2* variants was reported, providing encouragement that this large-scale problem may be manageable in the near future³⁴. However, in this study, the researchers used lentivirus to express variant-containing *MSH2* cDNAs. This approach has a level of artificiality with regards to regulation of gene expression, while also missing a potential role for variants in splicing anomalies. Their assay also did not examine direct repair function such as maintaining stability of microsatellites. Thus, confidence in using these results for classification may be limited. The ability to scale-up the use of CRISPR gene editing for testing VUS may help overcome these challenges.

Data Availability

Information on variants tested in this study can be found at https://www.insight-database.org/classifications/mmr_integrative_eval.html. We will supply all other data and materials upon request.

Acknowledgments

The authors wish to thank John-Paul Plazzer for his assistance in obtaining relevant variant data from the InSiGHT database and advising on the use of the MMR gene draft classification criteria. This work was supported by the National Institutes of Health grants CA222477 and CA237920 (C.D.H.), and CA164944 (S.V.T.).

Author Information

Conceptualization: Ab.R., S.V.T., J.P.G., C.D.H.; Data curation: Ab.R., A.A.R., J.P.G., C.D.H.; Formal Analysis: Ab.R., J.P.G., C.D.H.; Funding Acquisition: S.V.T., J.P.G., C.D.H.;

Investigation: Ab.R., A.A.R., K.R., R.B.G., J.R.H. M.C., J.P.G.; Methodology: Ab.R., S.V.T., J.P.G., C.D.H.; Project Administration: C.D.H.; Supervision: C.D.H.; Validation: Ab.R., A.A.R., K.R., C.D.H.; Writing – original draft: Ab.R., C.D.H.; Writing – reviewing & editing: Ab.R., A.A.R., K.R., R.B.G., J.R.H., M.C., S.V.T., J.P.G., C.D.H.

REFERENCES

1. Lynch HT, Snyder CL, Shaw TG, Heinen CD, Hitchins MP. Milestones of Lynch syndrome: 1895-2015. *Nat Rev Cancer*. 2015;15(3):181-194 (Perspectives).
2. Moreira L, Balaguer F, Lindor N, et al. Identification of Lynch syndrome among patients with colorectal cancer. *J Amer Med Assoc*. 2012;308(15):1555-1565.
3. Latham A, Srinivasan P, Kemel Y, et al. Microsatellite Instability Is Associated With the Presence of Lynch Syndrome Pan-Cancer. *J Clin Oncol*. 2019;37(4):286-295.
4. Leclerc J, Vermaut C, Buisine MP. Diagnosis of Lynch Syndrome and Strategies to Distinguish Lynch-Related Tumors from Sporadic MSI/dMMR Tumors. *Cancers (Basel)*. 2021;13(3).
5. Thompson BA, Spurdle AB, Plazzer J-P, et al. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nat Genet*. 2014;46(2):107-115 (Analysis).
6. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17:405 (ACMG Standards and Guidelines).
7. Peltomäki P. Update on Lynch syndrome genomics. *Fam Cancer*. 2016;15(3):385-393.
8. Brnich SE, Rivera-Munoz EA, Berg JS. Quantifying the potential of functional evidence to reclassify variants of uncertain significance in the categorical and Bayesian interpretation frameworks. *Hum Mutat*. 2018;39(11):1531-1541.
9. Heinen CD, Rasmussen LJ. Determining the functional significance of mismatch repair gene missense variants using biochemical and cellular assays. *Hered Cancer Clin Practice*. 2012;10(1):9.

10. Rasmussen LJ, Heinen CD, Royer-Pokora B, et al. Pathological assessment of mismatch repair gene variants in Lynch syndrome: Past, present, and future. *Hum Mutat.* 2012;33(12):1617-1625.
11. Fishel R. Mismatch Repair. *Journal of Biological Chemistry.* 2015;290(44):26395-26403.
12. Jiricny J. Postreplicative Mismatch Repair. *CSH Perspec Biol.* 2013;5(4).
13. Kolodner RD. A personal historical view of DNA mismatch repair with an emphasis on eukaryotic DNA mismatch repair. *DNA Repair.* 2016;38:3-13.
14. Kunkel TA, Erie DA. DNA MISMATCH REPAIR. *Ann Rev Biochem.* 2005;74(1):681-710.
15. Gupta D, Heinen CD. The mismatch repair-dependent DNA damage response: Mechanisms and implications. *DNA Repair.* 2019;78:60-69.
16. Li Z, Pearlman AH, Hsieh P. DNA mismatch repair and the DNA damage response. *DNA Repair.* 2016;38:94-101.
17. Rath A, Mishra A, Ferreira VD, et al. Functional Interrogation of Lynch Syndrome Associated MSH2 Missense Variants via CRISPR-Cas9 Gene Editing in Human Embryonic Stem Cells. *Hum Mutat.* 2019;40(11):2044-2056.
18. Brnich SE, Abou Tayoun AN, Couch FJ, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med.* 2020;12(1):3-13.
19. Tavtigian SV, Greenblatt MS, Harrison SM, et al. Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genet Med.* 2018;20(9):1054-1060.
20. Concordet J-P, Haeussler M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nuc Acids Res.* 2018;46(W1):W242-W245.
21. Vakulskas CA, Behlke MA. Evaluation and Reduction of CRISPR Off-Target Cleavage Events. *Nucleic Acid Ther.* 2019;29(4):167-174.

22. Tournier I, Vezain M, Martins A, et al. A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. *Hum Mutat.* 2008;29(12):1412-1424.
23. Smith C, Gore A, Yan W, et al. Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell.* 2014;15(1):12-13.
24. Veres A, Gosis BS, Ding Q, et al. Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem Cell.* 2014;15(1):27-30.
25. Yao X, Buermeier AB, Narayanan L, et al. Different mutator phenotypes in Mlh1- versus Pms2-deficient mice. *Proc Nat Acad Sci.* 1999;96(12):6850-6855.
26. Soukarieh O, Gaildrat P, Hamieh M, et al. Exonic Splicing Mutations Are More Prevalent than Currently Estimated and Can Be Predicted by Using In Silico Tools. *PLoS Genet.* 2016;12(1):e1005756.
27. de la Chapelle A, Hampel H. Clinical relevance of microsatellite instability in colorectal cancer. *J Clin Oncol.* 2010;28(20):3380-3387.
28. Duval A, Hamelin R. Mutations at Coding Repeat Sequences in Mismatch Repair-deficient Human Cancers: Toward a New Concept of Target Genes for Instability. *Cancer Res.* 2002;62(9):2447-2454.
29. Thompson BA, Martins A, Spurdle AB. A review of mismatch repair gene transcripts: issues for interpretation of mRNA splicing assays. *Clin Genet.* 2015;87(2):100-108.
30. Hardt K, Heick S, Betz B, et al. Missense variants in hMLH1 identified in patients from the German HNPCC consortium and functional studies. *Fam Cancer.* 2011;10(2):273-284.
31. Liu J, Hanne J, Britton BM, et al. Cascading MutS and MutL sliding clamps control DNA diffusion to activate mismatch repair. *Nature.* 2016;539(7630):583-587.

32. Cejka P, Stojic L, Mojas N, et al. Methylation-induced G2/M arrest requires a full complement of the mismatch repair protein hMLH1. *EMBO J.* 2003;22(9):2245-2254.
33. Avdievich E, Reiss C, Scherer SJ, et al. Distinct effects of the recurrent Mlh1G67R mutation on MMR functions, cancer, and meiosis. *Proc Nat Acad Sci.* 2008;105(11):4247-4252.
34. Jia X, Burugula BB, Chen V, et al. Massively parallel functional testing of MSH2 missense variants conferring Lynch syndrome risk. *Am J Hum Genet.* 2021;108(1):163-175.