Expression defect size among unclassified MLH1 variants
determines pathogenicity
in Lynch syndrome diagnosis

Inga Hinrichsen 1, Angela Brieger 1, Jörg Trojan 2, Stefan Zeuzem 2, Mef Nilbert 3,4, and Guido Plotz 1

1 Medizinische Klinik 1, Biomedical Research Laboratory, Johann Wolfgang Goethe-Universität, Frankfurt, Germany.
2 Medizinische Klinik 1, Johann Wolfgang Goethe-Universität, Frankfurt, Germany.
3 HNPCC-Register, Clinical Research Centre, Hvidovre Hospital, Copenhagen University, Denmark.
4 Department of Oncology, Lund University, Sweden.

Running title: Expression defect size determines MLH1 variant pathogenicity.
Keywords: Lynch syndrome; MLH1; mismatch repair; unclassified genetic variant; diagnostic tool.
Financial support: Research grant 2007.030.1 from the Sander-Stiftung to GP and SZ.
Corresponding author: Dr. Guido Plotz, Biomedizinisches Forschungslabor, Haus 11, Klinik der Johann Wolfgang Goethe-Universität, Theodor-Stern-Kai 7, 60590 Frankfurt, Germany. Phone: +49 69 6301 87668; Fax: +49 69 6301 87689; Email: plotz@med.uni-frankfurt.de
Conflict of interest: none declared.
Word count (excluding references): 5005; 5 Figures; 1 Table.
Abbreviations: UV, unclassified variant. MSI, microsatellite instability. MMR, mismatch repair. MMRCS, mismatch repair cancer syndrome.
ABSTRACT

Purpose: Lynch syndrome is caused by a germline mutation in a mismatch repair gene, most commonly the \textit{MLH1} gene. However, one-third of the identified alterations are missense variants with unclear clinical significance. The functionality of these variants can be tested in the laboratory, but the results cannot be used for clinical diagnosis. We therefore aimed to establish a laboratory test that can be applied clinically.

Experimental design: We assessed the expression, stability and mismatch repair activity of 38 \textit{MLH1} missense variants and determined the pathogenicity status of recurrent variants using clinical data.

Results: Four recurrent variants were classified as neutral (K618A, H718Y, E578G, V716M) and three as pathogenic (A681T, L622H, P654L). All seven variants were proficient in mismatch repair but showed defects in expression. qPCR, pulse-chase and thermal stability experiments confirmed decreases in protein stability, which were stronger in the pathogenic variants. The minimal cellular \textit{MLH1} concentration for mismatch repair was determined, which corroborated that strongly destabilized variants can cause repair deficiency. Loss of \textit{MLH1} tumor immunostaining is consistently reported in carriers of the pathogenic variants, demonstrating the impact of this protein instability on these tumors.

Conclusions: Expression defects are frequent among \textit{MLH1} missense variants, but only severe defects cause Lynch syndrome. The data obtained here enabled us to establish a threshold for distinguishing tolerable (clinically neutral) from pathogenic expression defects. This threshold allows the translation of laboratory results for uncertain \textit{MLH1} variants into pathogenicity statements for diagnosis, thereby improving the targeting of cancer prevention measures in affected families.
STATEMENT OF TRANSLATIONAL RELEVANCE

Lynch syndrome is the most common type of heritable predisposition for cancers of the colon and endometrium. To establish diagnosis of this syndrome, an inactivating mutation in a DNA mismatch repair gene has to be found. However, the most commonly affected gene (MLH1) frequently shows missense variations of unknown relevance, and no diagnosis can therefore be established for affected families. Functional tests have not yet solved this problem because no correlation between test readouts and clinical consequences has been established. Here, we report a quantitative link between impaired protein stability among MLH1 variants and cancer risk. The described testing system allows diagnosis of Lynch syndrome to be made based on a laboratory test for a variant for the first time. We also show that stability constraints are the most frequent consequence of MLH1 missense variations and are therefore the most relevant molecular cause of carcinogenesis and the most sensitive parameter for diagnosis.
INTRODUCTION

Lynch syndrome is a hereditary predisposition for cancer that accounts for 2 to 5% of all colorectal cancers (MIM #120435) (1, 2). Additionally, the tumor risk for some other organs, especially the endometrium, is increased. Lynch syndrome is a relatively common genetic disorder: approximately 1 in 660-2,000 individuals is affected (3). It is caused by heterozygous germline mutational inactivation of one of four mismatch repair (MMR) genes (MLH1, MSH2, MSH6 or PMS2). Somatic loss of the remaining wild-type allele leads to microsatellite instability (MSI), which is a hallmark of Lynch syndrome tumors (4, 5).

To establish a diagnosis of Lynch syndrome and offer predictive testing for family members, an inactivating germline mutation has to be identified. Mutations in the MLH1 gene (MIM #120436) account for the majority of Lynch syndrome cases. One-third of all alterations found in this gene are missense variants (6). The clinical significance of these variants is unknown a priori, and they are therefore termed unclassified variants (UVs) (7) and cannot be used to establish a diagnosis. Consequently, relatives of carriers cannot be offered predictive testing, and preventive surveillance cannot be properly targeted.

The internationally growing awareness of Lynch syndrome is leading to increases in systematic screening. In conjunction with the improved affordability of sequencing, this will likely result in identification of growing numbers of UVs in the future (8). Therefore, methods to correlate these mismatch repair gene UVs with their clinical consequences are urgently required (9). For this purpose, both clinical and functional laboratory evidence may be used.

Because few alterations occur recurrently, and sufficient cosegregation data are rare, the available clinical information is insufficient in most cases. Therefore, we and others have exerted significant efforts during the last decade to assess the effects of MLH1 UVs using functional laboratory tests (10-21). However, a correlation with a clinical phenotype has not been established for any of these assays. Moreover, many different functions have to be tested to cover all of the possible effects an alteration may have (22). The results of these heterogeneous investigations cannot currently be translated to determine disease risk, therefore failing to facilitate diagnosis (23).
We therefore attempted to identify a functional parameter that is defective for the majority of $MLH1$ variants and simultaneously provide a testing system involving a threshold to enable clinically relevant interpretations to be made. Furthermore, we aimed to develop a relatively simple assay system that is easily adoptable.

For these purposes, we functionally tested a series of $MLH1$ missense variations. We found that the majority of the $MLH1$ missense alterations compromised protein stability. Several internationally recurrent variants that are still considered to represent UVs by most clinicians were included in the analyses. We were able to define solid pathogenicity statements for these variants based on a comprehensive analysis of published and unpublished clinical data. This allowed the identification of an expression level threshold that can be used to directly recognize pathogenicity (caused by reduced protein stability) associated with novel UVs.

### MATERIALS AND METHODS.

**Selection of variants for analysis.** A total of 38 $MLH1$ missense variants were selected from public $MLH1$ variation databases; these variants included 7 repair-proficient and 3 repair-deficient recurrent alterations for which large amounts of clinical data are available. Additional variants were selected arbitrarily.

**Protein expression and expression quantification.** pcDNA3-MLH1, pSG5-PMS2, and the HEK293 and HEK293T cell lines have been described previously (13, 24). Missense variants were generated via site-directed mutagenesis (QuickChange II Kit, Stratagene) and confirmed by direct sequencing. HEK293T cells were transiently transfected with 5 µg of vector DNA and 20 µl of polyethyleneimine (1 mg/ml, linear, 25 kDa, Polysciences, Warrington, PA) and extracted as described previously (24, 25). The extracts were analyzed via SDS-PAGE and immunoblotting (using anti-MLH1, G168-728, BD Biosciences, and anti-PMS2, E-19, and anti-β-Actin, C2, from Santa Cruz Biotechnologies). Chemiluminescence signals (Immobilon, Millipore) were detected in an LAS-4000 mini camera (Fuji) and quantified using Multi Gauge v3.2.
qPCR analysis of MLH1 transcription. MLH1 transcript levels were measured using quantitative PCR (qPCR) according to the MIQE guidelines (26). Total RNA was extracted from transfected HEK293 cells using TRIzol (Invitrogen). The RNA was then dissolved in 50 µl of RNase-free water, and the RNA content was quantified via UV-spectrometry. The RNA was treated then with DNase (DNase RQ1, Promega) to remove potential residual plasmid DNA. The success of the DNase treatment was controlled via PCR analysis. cDNA was generated from 1 µg of total RNA. This RNA had either been prepared fresh or came from samples that had been stored at -80°C for less than one month and not thawed more than twice. Reverse transcription was performed for 10 min at 25°C, followed by 50 min at 50°C, with M-MLV reverse transcriptase (50U, RNase H Minus point mutant, Promega) and 250 ng of random primers (Promega) according to the manufacturers’ recommendations in a total volume of 25 µl. The cDNA samples were stored at -20°C.

Primer and probe sequences were designed using FileBuilder software and produced by Applied Biosystems. A PCR product spanning exons 12-13 was used because this area corresponds to the unconserved linker region of the MLH1 protein where few genetic alterations have been reported; therefore, mRNA from all variant cDNA constructs could be quantified using the same primers and probe, without the interference of the individual genetic alterations. The primers used in these assays were as follows (mRNA sequence NM_000249.3): AGAGAGGACCTACTTCCAGCAA (f), ATCTTCCACCATTTCACATCAGAA (r) and CCCAGAAAGACATC (hydrolysis probe). The obtained amplicon length was 71 bp. The hydrolysis probes contained FAM as a reporter dye and a non-fluorescent quencher. Calibration curves generated from dilutions of the MLH1 plasmid showed that the qPCR results were linear over a wide range (Supp. Figure S1). The only the reference gene used was GAPDH (assay #Hs99999905_m1, Applied Biosystems) because all samples were from the same source material and of the same quality.

Control qPCR assays with RNA but without reverse transcription were performed, and no amplification occurred. Additionally, untransfected HEK293T cells were always analyzed in parallel and yielded much lower levels of amplification (Cq 10 cycles higher than the transfected samples on average). The qPCR assays were performed in a total volume of 15 µl, which included TaqMan universal master mix, an assay mixture containing the primers and hydrolysis probe, and 1.5 µl of a
Hinrichsen et al.: Expression defect size determines MLH1 variant pathogenicity

sample. The cycling conditions were as follows: 2 min at 50°C, 10 min at 95°C, followed by 60 cycles of 15 sec at 95°C and 1 min at 60°C. qPCR was performed in a StepOnePlus Realtime cycler (Applied Biosystems). The StepOne 2.0 software was employed to generate qPCR curves and Cq values. To calculate MLH1 transcript expression, the samples were normalized based on the results for GAPDH (ΔCq=Cq(MLH1)-Cq(GAPDH)). Subsequently, the variants were compared to the calibrator (wild-type MLH1) by calculating the ΔΔCq value (ΔΔCq(variant)=ΔCq(variant)-ΔCq(wild-type)). Relative expression was calculated using the standard formula f= 2^(-ΔΔCq).

**Determination of MMR activity.** The MMR activity of MLH1 variants was scored in vitro as described previously (25). Briefly, protein extracts were mixed with 35 ng of DNA substrate containing a G-T mismatch and a 3’ single-strand nick at a distance of 83 bp with reaction buffer in a total volume of 15 µl. After incubation at 37 °C, the DNA substrate was purified and digested with EcoRV and AseI. The restriction fragments were separated in agarose gels and analyzed using GelDoc XR plus detection and QuantityOne software (Bio-Rad). The repair efficiency (e) was calculated as: 

\[ e = \frac{\text{intensity of bands of repaired substrate}}{\text{intensity of all bands of substrate}} \]

This result is independent of the amount of DNA recovered through plasmid purification. The typical total repair efficiencies ranged from 50-90%. The repair efficiency of MutLα variants was analyzed in direct comparison to a wild-type protein that had been produced in parallel, and calculated as 

\[ e(\text{relative})=\frac{e(\text{variant})}{e(\text{wild-type})}\times100. \]

**Collection of clinical data and in silico analyses.** Publications addressing variants were identified using the Leiden Open Variation Database (LOVD; www.lov.d.nl/MLH1), the Mismatch Repair Gene Variants Database (www.med.mun.ca/ mmrvariants) and the MMR Gene Unclassified Variants Database (http://www.mmrmissense.net). Moreover, searches were performed for all alternative variant descriptions (e.g., MLH1 “V716M”, “Val716Met”, “2146G>A”) in PubMed and Google. Comprehensive clinical information was obtained from a recent publication by Hardt et al. from the German HNPCC consortium (15). Each reported patient carrying an alteration and exhibiting a Lynch-syndrome-associated tumor was listed in Suppl. Table S1 with all available information. Great care was taken to detect multiple reports of identical patients based on their patient/family identifiers, the reporting authors, the country of origin or other strikingly identical features; when such cases were
found, all information was summarized in a single entry. In some cases, the authors were contacted to resolve contradictions.

**Structural and bioinformatic analyses.** Function-structure evaluations were performed with a model of human MLH1-PMS2 based on the structure of human PMS2 NTD (27) and homology models of MLH1-NTD and MLH1-PMS2-CTD (24, 25, 28). Figures were generated using PyMOL v.1.4.1 (Schrödinger LLC).

**RESULTS**

**Screening of the expression and MMR efficacy of MLH1 missense variants.**

A total of 38 MLH1 missense variants were selected from public databases of MLH1 variations. These included recurrent alterations (shown in gray in Figure 1), for which large amounts of clinical data are available, and additional arbitrarily selected variants. We assessed two major functional protein parameters (expression and MMR activity), because missense alterations rarely affect transcriptional integrity (29, 30).

The obtained expression values covered the entire range from 0%-100% of wildtype (Figure 1A), while the variants formed two groups in terms of their repair capacity (Figure 1B): 21 variants showed MMR efficiencies similar to wildtype (>70% of wildtype activity), whereas MMR activity was largely absent in 15 variants (<30% of wildtype activity). These two groups comprised 55% and 40% of all variants, respectively (Figure 1C). Overall, more variants showed low expression than low MMR activity (71% versus 45%; "low" here means <70% of wildtype).

Among the recurrent MLH1 variants, 7 were proficient in mismatch repair while displaying expression defects. These variants, hereafter referred to as *validation variants*, were (sorted by expression level): K618A>H718Y>E578G>V716M>A681T>L622H>P654L. Because they are catalytically active, the potential pathogenic effects of these variants are likely attributable to their expression defects. Therefore, they are suitable for assessing the correlation between expression levels and clinical outcomes.

**Decreased protein stability underlies reduced MLH1 expression levels.**
To verify that the expression defects reflected reduced stability of the variant proteins, we quantified the levels of MLH1 transcripts in the expression system. Under the applied conditions, MLH1 transcript levels were quite robust to variations in the mass of transfected plasmid DNA. They decreased only when the amount of DNA transfected missed the standard amount by a factor ≥2 (Suppl. Figure S2A). However, even this situation did not affect protein levels (Suppl. Figure S2B), suggesting that MLH1 transcript abundance was not limiting for protein production. This was confirmed by the finding that the transcript levels in transfected HEK293T cells far exceeded those of endogenous MLH1 in HEK293 cells, while no corresponding increase in protein levels occurred (Suppl. Figure S2C). Therefore, translation was the rate-limiting step of the expression system.

Consequently, small differences in the transfection efficiency or plasmid transcription did not influence the resulting protein levels, and expression level differences necessarily reflect properties inherent to the protein variant. Transcript quantification for the MLH1 validation variants additionally confirmed that their expression level defects were not caused by poor transfection or transcription (Suppl. Figure S2D).

Because these results suggested that the substitutions caused decreased stability in the MLH1 proteins, we analyzed the protein degradation rates of the validation variants using an in vitro pulse-chase method developed for this purpose (31). Four of the validation variants (K618A, V716M, H718Y, and E578G) showed similar reductions of the half-life time to 64% on average. The A681T variant was clearly less stable (half-life time decreased to 43%, Suppl. Figure S3A). The expression of the L622H and P654L variants was too low to assess the degradation rate in this system, but this finding also corroborates the notion that low stabilities underlies the expression defects observed in the transfection system.

Proteolytic stability of missense variants usually correlates with thermal stability (32). Therefore, we tested defolding temperatures via differential scanning fluorimetry (33). Thermal stability was slightly compromised in the K618A variant but was decreased more strongly in V716M, H718Y and E578G (Suppl. Figure 3B). The low-stability variants were again poorly expressed (in a different expression system) which prevented their purification for this analysis.
In conclusion, all of the obtained data confirmed that the decreases in expression levels observed in the transfection system reflected protein stability constraints due to missense substitutions. Although the precise sorting of the stability of the variants differed slightly between the experimental systems, the following consensus was obtained: wt > K618A ≥ E578G ≈ V716M ≈ H718Y > A681T > L622H > P654L.

Clinical phenotypes of recurrent MLHI variants.

To investigate the correlation between MLHI variant expression level and clinical phenotypes, we performed a comprehensive analysis of clinical data for the recurrent variants (the 7 MLHI validation variants and three additional, repair-deficient pathogenic control variants: T117M, R659P and L749P). Altogether, reports from 350 patients carrying these 10 variants in the germline were evaluated (Table 1 summarizes the complete data presented in Suppl. Table S1). For the purpose of comparison, a Lynch-syndrome control group comprised of carriers of pathogenic mutations was formed. In parallel, all published information on the frequency of these variants in unaffected control individuals was collected.

The average age at diagnosis, frequency of tumor MSI and fulfillment of the Amsterdam criteria displayed a clear gradient in carriers of the analyzed variants (Figure 2). In carriers of K618A, E578G, V716M and H718Y, the average age at diagnosis was greater than in the Lynch syndrome control group (47.4 years versus 38.8 years, p=0.00006). Additionally, few carriers of these four alterations were positive for the Amsterdam criteria (27% versus 74%) or showed tumor MSI (35% versus 97%).

We evaluated four additional clinical parameters that can provide evidence of pathogenicity or neutrality: i) the frequency of the alteration in control populations (which is expected to be low for pathogenic alterations); ii) co-segregation of the alteration with disease; iii) the identification of homozygous variant carriers (who would, in case of pathogenic variants, be affected not by Lynch syndrome but by mismatch repair cancer syndrome, MMRCS, MIM #276300, a severe condition that presents with malignancies in early childhood) (34); and vi) co-occurrence with other pathogenic mutations (frequent co-occurrence means that pathogenicity of the variant is unlikely).
The alterations K618A, E578G, V716M and H718Y were found in control populations. They mostly did not cosegregate with disease, and homozygosity was observed in cancer patients as well as in healthy controls, none of whom displayed MMRCs. Finally, co-occurrence with pathogenic mutations was frequent (10-20%). Furthermore, four case-control analyses showed that there was no association of K618A with increased cancer risk (35-38) (Table 1).

In summary, all of these findings provide strong evidence that K618A, E578G, V716M and H718Y are not causative for Lynch syndrome.

In contrast, carriers of the validation variants presenting severe expression defects (A681T, L622H and P654L) showed a younger average age at diagnosis, frequent tumor MSI and fulfillment of the Amsterdam criteria (Figure 2). Notably, these clinical features corresponded to the size of the expression defect (A681T < L622H < P654L). Additionally, neither homozygosity nor co-occurrence was reported for these variants. Moreover, comprehensive positive co-segregation information was available for these three alterations (Table 1): A681T has been found to co-segregate with disease in 11 individuals from an extended Scottish kindred (35) and in five Polish cancer families (39); L622H is a Spanish founder mutation that co-segregates with disease in families (40); P654L was shown to co-segregate with disease in 9 relatives from 4 German families (15).

In summary, the clinical evidence demonstrates that K618A, E578G, V716M and H718Y, which all resulted in MLH1 expression levels above 60%, represent neutral variants. In contrast, the variants A681T, L622H and P654L, which were associated with lower expression levels (52%, 42% and 25%, respectively) are causative for Lynch syndrome.

**Effect of MLH1 protein concentrations on MMR activity.**

The analysis suggested that the four MMR-proficient variants (K618A, E578G, V716M and H718Y) moderately destabilize the MLH1 protein but are not causative for Lynch syndrome, whereas the strongly destabilizing variants (A681T, L622H and P654L) are pathogenic. In the tumor cells of carriers of these variants, the low intracellular MLH1 protein concentration (after loss of the wildtype allele) likely caused an MMR defect. We therefore investigated the minimal MLH1 protein level required for efficient mismatch repair. We used the cell line HEK293 and its clone HEK293T, which
are MLH1- and repair-proficient and -deficient, respectively. We confirmed the mismatch repair defect by testing microsatellite instability in these cells: HEK293T, but not HEK293, displayed strong MSI (Suppl. Figure S4).

Dilution experiments showed that reduction of the MLH1 concentration to 50% does not affect MMR activity, though it is reduced at lower concentrations (Figure 3). Notably, this finding is in good agreement with observations regarding MSI in healthy tissues of Lynch syndrome individuals: normally, both MLH1 alleles contribute equally to MLH1 expression (41). Consequently, individuals heterozygous for an inactivating MLH1 mutation constitutively present cellular MLH1 levels that are 50% of normal levels. In the healthy tissues of these individuals, MSI is not prominent, but is detectable in in-depth analyses (42, 43). This confirms that the intracellular MLH1 concentration in mutation carriers is just sufficient to retain (almost normal) MMR activity.

Pathogenicity threshold for expression defects in MLH1 variants.

Taken together, the data suggest that decreases in MLH1 protein stability are compatible with normal health to some degree, but below a certain threshold, MMR function is insufficient, and carriers display typical traits of Lynch syndrome (Figure 4). This threshold, established based on the neutral V716M and pathogenic A681T variants, is corroborated by clinical and functional data for six additional missense alterations (Figure 4, see also Suppl. Table S2). Furthermore, it is consistent with the immunohistochemistry of MLH1 in tumors of affected patients: the tumors of carriers of variants showing stability above the threshold level were mostly positive for MLH1 immunostaining, while MLH1 was consistently undetectable in carriers of variants with a stability below the threshold (Figure 4 and Table 1).

Analysis of stability defects in silico and the relationships with protein structure.

Several in silico algorithms have been developed to predict the effect of missense variants. Of these algorithms, MAPP-MMR (44) was most consistent with the experimental data. It correctly predicted a high probability of a damaging effect for most repair-deficient variants (Suppl. Figure S5), whereas repair-proficient variants were mostly scored as “neutral” or “borderline”, irrespective of their stability.
Hinrichsen et al.: Expression defect size determines MLH1 variant pathogenicity

defect. Other algorithms specifically designed for predicting the effect of substitutions on protein
stability (Cupsat, i-Mutant, PBSA) were not correlated with the observed experimental effects.
Consequently, in silico stability determinations lack accuracy (45) and can currently not replace
experimental analysis of stability.
We also analyzed potential associations of the effects of the substitutions with their structural
groups. Strongly destabilizing substitutions frequently affected residues in the C-terminal “In”
subdomain of MLH1, especially in its core three-helix motif (Suppl. Figure S6). This suggests that
the structural integrity of this subdomain is important for MLH1 protein stability. Repair-deficient
substitutions specifically affected sites showing catalytic activity (Suppl. Figure S7).

**DISCUSSION**

Missense variants in *MLH1* in potential Lynch syndrome patients represent a long-standing problem,
and much work has been invested in providing laboratory evidence supporting the classification of
these variants as either pathogenic or neutral (10-21). However, this laboratory evidence is not yet
being used for diagnosis because no testing system comprises clinically established thresholds to
distinguish normal function and tolerable functional impairment from pathogenic dysfunction.
In our analysis, seven MMR-proficient *MLH1* variants displayed defects of stability whose degree was
correlated with the clinical phenotype. This allowed us to establish a threshold defining what degree of
stability reduction is associated with Lynch syndrome (≤52%, A681T) versus not associated (≥65%,
V716M) (Figure 4). The intermediate zone is quite narrow, suggesting that there is a discrete (rather
than continuous) increase in cancer risk, similar to what is observed in *BRCA1* assays (46). For the
purpose of classifying uncharacterized variants, the small size of this ambiguous zone is beneficial.
For establishing the threshold, four and three recurrent *MLH1* variants were defined as neutral or
pathogenic, respectively. These definitions were based on an analysis of clinical data. Some of the
analyzed parameters are routinely applied for selecting patients for genetic analysis (age at diagnosis,
Amsterdam criteria, tumor MSI). This causes a strong sampling bias, explaining why carriers of
neutral variants deviated in terms of these parameters from sporadic cancer patients. Nevertheless,
even these biased parameters allowed us to distinguish pathogenic from neutral phenotypes (Figure
Further evidence that is free of bias (co-segregation, homozygosity, co-occurrence, frequency in controls, and case-control studies) confirmed the pathogenicity classifications. Consequently, the classifications used for threshold determination can be considered highly reliable.

Compromised expression levels were found for 84% of MLH1 variants (p<0.05, Suppl. Table S3). Similarly high rates have been reported previously: 59% of variants were classified as showing “decreased” expression by Raevaara et al. (10), and reductions of expression below 75% of normal level were found in 48% and 87% of variants in two other studies (14, 15). Impaired protein stability is therefore a major consequence of amino acid substitutions in MLH1, which is most likely also the major reason for MLH1 inactivation associated with missense variants. Why is protein destabilization observed at such a high frequency? Practically all residues of a protein contribute in many ways to its stability, which is therefore quite likely to be disturbed by a substitution (47). In contrast, interfering with catalysis requires that the affected residue plays a (more or less direct) catalytic role (Suppl. Figure S7), which is, thus, a much rarer event. In investigating MLH1 missense UVs, it hence appears to be reasonable to first test their expression (Figure 5A) and to then analyze additional functional parameters (Figure 5B) only when the level of a variants' expression is above the threshold provided by the A681T variant.

In this analysis, the average expression of half of the variants (53%) was below this threshold, and statistical significance was achieved for 24% of the variants (p<0.05; Figure 4 and Suppl. Table S3). These variants can, thus, be considered pathogenic with high certainty.

In conclusion, our analysis showed that stability constraints of the MLH1 protein are a major consequence of missense alterations in MLH1 at the cellular level (Figure 5B). Precise determination of the expression defects associated with catalytically active variants allowed us to establish a threshold below which an expression defect is associated with Lynch syndrome. This threshold is compatible with the minimal cellular MLH1 protein concentration required for unimpaired MMR activity. It is also compatible with data obtained from MLH1 immunostaining in tumors of variant carriers. For future efforts to classify uncertain MLH1 missense variants, we therefore propose first determining expression in direct comparison with clinically validated standard variants (A681T and V716M).
ACKNOWLEDGMENTS

We are grateful to Dr. Mev Dominguez Valentin, who performed the MSI analyses and collected clinical data, and to Prof. Dr. Brigitte Royer-Pokora for helpful discussions and critical review of the manuscript. The clinical evaluations incorporated in this study are based mostly on published data, demonstrating the value of accurate reporting of phenotypic expression in germline variant carriers; we are therefore grateful to all of the medical staff who gathered this information as well as the authors who made it available. We are also grateful to the national and international consortia who enforce the documentation of clinical and genotypic data and to the curators of the public databases that assemble this information.

GRANT SUPPORT

This work has been supported by research grant 2007.030.1 from Wilhelm Sander-Stiftung to GP and SZ.

REFERENCES

Hinrichsen et al.: Expression defect size determines MLH1 variant pathogenicity


FIGURE LEGENDS

Figure 1: Expression and MMR activity of MLH1 missense variants.
The expression levels (A) of MLH1 (variants) transfected into HEK293T cells and their mismatch repair activity (B) were determined as detailed in the Materials and Methods. The average values and standard deviations (bars) from several independent experiments are shown. C. Two-dimensional representation of relative MLH1 variant fitness in terms of expression and repair in comparison with wild-type MLH1. Recurrent variants are shown in gray.

Figure 2: Prime clinical markers of Lynch syndrome in recurrent MLH1 variants.
Average age at cancer diagnosis, frequency of patients fulfilling the Amsterdam criteria and frequency of tumor MSI from Table 1 for carriers of the indicated MLH1 germline variants. The asterisks mark data based on less than 5 independent reports (see Table 1).

1 In sporadic CRC, the age at diagnosis is approximately 69 years, and tumor MSI (mostly because of somatic hypermethylation of MLH1) is found in 15% of patients (4).

2 “Pathogenic” indicates the data from all carriers of variants classified as pathogenic in Table 1.

Figure 3: MMR activity depending on MLH1 protein level.
A nuclear extract from HEK293 cells (proficient in MLH1 and MMR) was gradually diluted with a nuclear extract from cells of its clone HEK293T (deficient in MLH1 and MMR), and MMR efficiency was scored. The average repair activity relative to the undiluted HEK293 extract and representative gels showing repair activity (top) as well as western blots of MLH1 (middle) and β-Actin (bottom) are presented. In the agarose gel images, the bands at the height of the 2.0 kbp marker represent the unrepaired, mismatched substrate, whereas this substrate was broken down into the lower-running bands at 1.2 kbp and 0.8 kbp when MMR was successful.

Figure 4. Pathogenicity threshold for stability defects in MLH1 variants.
The average expression levels of wild-type MLH1 and its variants are shown; bars indicate the standard errors of the mean. The expression levels of clinically neutral (left, white) and pathogenic (middle, black squares) variants were used to define the pathogenicity threshold (hatched). 1 For other
variants (gray squares), the clinical information is compatible with a pathogenic effect (see Suppl. Table S2). The fraction of tumors in variant carriers in which MLH1 expression was lost. For details, see Table 1 and Suppl. Table S2. Yes/no are indicated when only a few reports addressing IHC status were available (number of reports in brackets).

Variants whose expression was significantly lower than that of the A681T variant (P<0.05 after correction for multiple testing, see Suppl. Table S3) are marked by an asterisk.

Figure 5. MLH1 missense variant protein stability and pathogenicity classification.

A. The pathogenicity classification of an MLH1 variant can be based on clinical data and/or functional evaluation. The current work provides thresholds for determining pathogenicity based on reduced protein stability.

B. Overview of parameters for functional evaluations of the potential effects of a missense variant in the context of a cell.
Hinrichsen et al.: Expression defect size determines MLH1 variant pathogenicity.

<table>
<thead>
<tr>
<th>MLH1 variant</th>
<th>Conservation score ¹ (residues in this position)</th>
<th>MAPP-MMR score</th>
<th>Total case #</th>
<th>Average age at diagnosis /cases</th>
<th>Amsterdam-positive % /cases</th>
<th>MSI % /cases</th>
<th>Loss of MLH1 in IHC % /cases</th>
<th>Frequency in controls %/number of controls</th>
<th>Cosegregation ³</th>
<th>Homozygosity ⁴</th>
<th>Co-occurrence (%/cases) ⁵</th>
<th>Case-control analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Repair-proficient validation variants, classified neutral based on the shown data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K618A (HKMNRTWY)</td>
<td>5.06</td>
<td>123</td>
<td>49.3</td>
<td>35</td>
<td>22</td>
<td>49</td>
<td>30</td>
<td>54</td>
<td>26</td>
<td>35</td>
<td>0.67</td>
<td>4652</td>
</tr>
<tr>
<td>E578G (DENPQSTV)</td>
<td>4.85</td>
<td>11</td>
<td>51.6</td>
<td>14</td>
<td>29</td>
<td>40</td>
<td>27</td>
<td>30</td>
<td>21</td>
<td>37</td>
<td>0.47</td>
<td>30</td>
</tr>
<tr>
<td>V716M (EFILMNV)</td>
<td>2.78</td>
<td>55</td>
<td>45.2</td>
<td>15</td>
<td>25</td>
<td>45</td>
<td>24</td>
<td>44</td>
<td>18</td>
<td>38</td>
<td>0.32</td>
<td>1747</td>
</tr>
<tr>
<td>H718Y (DEFHKNQSV)</td>
<td>3.45</td>
<td>22</td>
<td>44.0</td>
<td>8</td>
<td>25</td>
<td>54</td>
<td>5</td>
<td>80</td>
<td>5</td>
<td>(0)</td>
<td>1</td>
<td>0.84</td>
</tr>
<tr>
<td><strong>Repair-proficient validation variants, classified pathogenic based on the shown data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A681T (AGSV)</td>
<td>4.41</td>
<td>43</td>
<td>42.6</td>
<td>15</td>
<td>22</td>
<td>48</td>
<td>22</td>
<td>88</td>
<td>17</td>
<td>100</td>
<td>13</td>
<td>0.00</td>
</tr>
<tr>
<td>L622H (FKLMY)</td>
<td>12.9</td>
<td>23</td>
<td>44.5</td>
<td>23</td>
<td>69</td>
<td>13</td>
<td>100</td>
<td>14</td>
<td>100</td>
<td>12</td>
<td>0.00</td>
<td>1185</td>
</tr>
<tr>
<td>P654L (GKPY)</td>
<td>19.20</td>
<td>20</td>
<td>39.2</td>
<td>20</td>
<td>75</td>
<td>20</td>
<td>100</td>
<td>11</td>
<td>100</td>
<td>9</td>
<td>0.00</td>
<td>1493</td>
</tr>
<tr>
<td><strong>Repair-deficient pathogenic control variants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T117M (ST)</td>
<td>21.05</td>
<td>38</td>
<td>40.5</td>
<td>20</td>
<td>85</td>
<td>27</td>
<td>91</td>
<td>11</td>
<td>100</td>
<td>9</td>
<td>0.00</td>
<td>1493</td>
</tr>
<tr>
<td>L749P (L)</td>
<td>37.37</td>
<td>8</td>
<td>38.8</td>
<td>6</td>
<td>100</td>
<td>6</td>
<td>100</td>
<td>2</td>
<td>50</td>
<td>4</td>
<td>0.00</td>
<td>1493</td>
</tr>
<tr>
<td>R659P (ACEHKLQRST)</td>
<td>11.70</td>
<td>5</td>
<td>38.5</td>
<td>2</td>
<td>100</td>
<td>3</td>
<td>100</td>
<td>5</td>
<td>(100)</td>
<td>1</td>
<td>0.00</td>
<td>1493</td>
</tr>
<tr>
<td>Carriers of variants classified neutral ³</td>
<td>166</td>
<td></td>
<td>47.4</td>
<td></td>
<td>77</td>
<td>27</td>
<td>86</td>
<td>35</td>
<td>87</td>
<td>27</td>
<td>35</td>
<td>0.00</td>
</tr>
<tr>
<td>Carriers of variants classified pathogenic ⁷</td>
<td>79</td>
<td></td>
<td>38.8</td>
<td></td>
<td>55</td>
<td>74</td>
<td>62</td>
<td>97</td>
<td>32</td>
<td>na</td>
<td>na</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 1. Evaluation of clinical parameters of MLH1 variant carriers.

Clinical data summary from Supplementary Table S1, see there for detailed information and references [in square brackets].

¹ Scored from 1 (no conservation) to 9 (highly conserved) using ConSeq.
² see Supplementary Table S1, columns AF-AI.
³ Yes: number of related alteration carriers affected by a Lynch syndrome tumor.
No: number of individuals related to index patient who are affected by a tumor from the Lynch syndrome spectrum but do not carry the alteration, plus number of further related carrier individuals without Lynch syndrome tumor.
⁴ Reports on individuals carrying the alteration homozygously in the germline („case“ refers to cancer patient, „control“ refers to an unaffected individual).
⁵ Frequency of carrier individuals who carry a second, damaging germline mutation.
⁶ Co-occurrence in compound heterozygosity with the damaging mutation K618del, but this patient had Lynch syndrome and not MMRCS.
⁷ A Student-T-test (two-sided) was applied to assess if the age at diagnosis significantly differs from the group of carriers of pathogenic alterations.
⁸ Summary of the information on carriers of K618A, E578G, V716M and H718Y; see columns J, AP-AU in Supplementary Table S1.
Summary of the information on carriers of repair-deficient control variants (T117M, R659P, L749P) and all other individuals who carried (additional) pathogenic mutations; see columns K and AK-AO of Supplementary Table S1.
Figure 2

![Bar chart showing age at diagnosis, Amsterdam criteria positivity, and MSI in tumors.](image-url)
Figure 3
Figure 4

[Diagram showing data with expression (percent) on the y-axis and pathogenic expression on the x-axis, with validation variants and other repair proficient and deficient variants highlighted.]
Figure 5

A Unclassified $MLH1$ missense variant

Clinical evaluation
- Cosegregation
- Case-control analysis
- Clinical features:
  - Homozygosity
  - Amsterdam criteria
  - Age at diagnosis
  - MSI
  - IHC of MMR proteins
  - Tumor location
  - Co-occurrence
  - Frequency in controls

Insufficient clinical data

Functional evaluation
- Expression level:
  - $> V716M$
  - $< A681T$
- Stability sufficient
- Pathogenic (Stability insufficient)
- Further functional tests (currently without thresholds)

B

Cell

$MLH1$ missense alteration

Transcriptional integrity
- Transcription
- Splicing
- Transcript stability
- etc.

Function
- Catalytic function(s)
- Protein interaction(s)
- Subcellular distribution
- etc.

Stability

mRNA
Expression defect size among unclassified MLH1 variants determines pathogenicity in Lynch syndrome diagnosis

Inga Hinrichsen, Angela Brieger, Joerg Trojan, et al.

Clin Cancer Res  Published OnlineFirst February 12, 2013.

Updated version  Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-12-3299

Supplementary Material  Access the most recent supplemental material at: http://clincancerres.aacrjournals.org/content/suppl/2013/02/25/1078-0432.CCR-12-3299.DC1

Author Manuscript  Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.