Evaluation of

MLH1 variants of unclear significance

Nicole Köger[1], Lea Paulsen[1], Francisco López-Kostner[2], Adriana Della Valle[3], Carlos Alberto Vaccaro[4], Edenir Inêz Palmero[5,6], Karin Alvarez[2], Carlos Sarroca[3], Florencia Neffa[3], Pablo German Kalfayan[4], Maria Laura Gonzalez[4], Benedito Mauro Rossi[7], Rui Manuel Reis[1,8,9], Angela Brieger[1], Stefan Zeuzem[1], Inga Hinrichsen[1], Mev Dominguez-Valentin[*10], Guido Plotz[*1]

[1]Biomedizinisches Forschungslabor, Medizinische Klinik 1, Universitätsklinik Frankfurt, Germany
[2]Laboratorio de Oncología y Genética Molecular, Clínica Los Condes, Santiago, Chile
[3]Hospital Fuerzas Armadas, Grupo Colaborativo Uruguayo. Investigación de Afecciones Oncológicas Hereditarias (GCU), Montevideo, Uruguay
[4]Hereditary Cancer Program (PROCANHE), Hospital Italiano, Buenos Aires, Argentina
[5]Molecular Oncology Research Center, Barretos Cancer Hospital, Barretos, São Paulo, Brazil
[6]Barretos School of Health Sciences - FACISB, Barretos, São Paulo, Brazil
[7]Hospital Sirio Libanes, Sao Paulo, Brazil
[8]Life and Health Sciences Q5 753 Research Institute (ICVS), Health Sciences School, University of Minho, Braga, 754 Portugal
[9]ICVS/3B's-PT Government Associate Laboratory, Braga, Guimarães, 755 Portugal.
[10]Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway

* Joint last authorship

Corresponding author:
Dr. Guido Plotz
Medizinische Klinik 1
Biomedizinisches Forschungslabor
Haus 11, EG
Universitätsklinik Frankfurt
Theodor-Stern-Kai 7
D-60590 Frankfurt
Germany
Email: plotz@med.uni-frankfurt.de
Phone: +49 69 6301 87668
Fax: +49 69 6301 84441

https://orcid.org/0000-0001-5314-9739

Abbreviated title: Evaluation of MLH1 variants

Grant sponsors: FINEP - CT-INFRA (02/2010), Radium Hospital Foundation (Oslo, Norway), Helse Sør-Øst (Norway). Deutsche Forschungsgemeinschaft

This article has been accepted for publication and 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 an 'Accepted Article', doi: 10.1002/gcc.22536

Abstract

Inactivating mutations in the *MLH1* gene cause the cancer predisposition Lynch syndrome, but for small coding genetic variants it is mostly unclear if they are inactivating or not. Nine such *MLH1* variants have been identified in South American colorectal cancer (CRC) patients (p.Tyr97Asp, p.His112Gln, p.Pro141Ala, p.Arg265Pro, p.Asn338Ser, p.Ile501del, p.Arg575Lys, p.Lys618del, p.Leu676Pro), and evidence of pathogenicity or neutrality was not available for the majority of these variants. We therefore performed biochemical laboratory testing of the variant proteins and compared the results to protein *in-silico* predictions on structure and conservation. Additionally, we collected all available clinical information of the families to come to a conclusion concerning their pathogenic potential and facilitate clinical diagnosis in the affected families. We provide evidence that four of the alterations are causative for Lynch syndrome, four are likely neutral and one shows compromised activity which can currently not be classified with respect to its pathogenic potential. The work demonstrates that biochemical testing, corroborated by congruent evolutionary and structural information, can serve to reliably classify uncertain variants when other data are insufficient.

Key words

Lynch syndrome, *MLH1*, small coding variant, variant of uncertain significance,
pathogenicity classification

Introduction

Lynch syndrome (MIM #120435) is a heritable condition associated with a greatly increased lifetime risk of colorectal cancer (CRC) (70-80%), endometrial cancer (50-60%), stomach cancer (13-19%), ovarian cancer (9-14%), cancer of the small intestine, the biliary tract, brain as well as carcinoma of the ureters and renal pelvis [1]. It is caused by an inactivating mutation of a DNA mismatch repair (MMR) gene: *MLH1* (MIM# 120436), *MSH2* (MIM #609309), *MSH6* (MIM #600678), and *PMS2* (MIM #600259) or to deletions of the 3' portion of the *EPCAM* gene [2,3].

Since Lynch syndrome does not provide a characteristic phenotype enabling clinical diagnosis, diagnosis requires identification of a causative (inactivating) mutation in a MMR gene. Besides establishing the diagnosis for the affected patient, this subsequently enables predictive testing of family members, and therefore early surveillance and cancer prevention methods provide significant benefit for affected individuals.

While the relevant genes are known and tools for detecting genetic alterations are readily available, a significant proportion of alterations identified in patients cannot straightforwardly be classified as inactivating and therefore pathogenic [3,4]. This exemplarily accounts for many small coding variants like missense alterations and small insertions or deletions, as well as for non-coding variants outside the highly conserved splicing motifs which may affect splicing in an unpredictable fashion. For these variants, pathogenicity clarification must be performed. Different lines of evidence, either directly from clinical patient or family data, or indirect, from laboratory or *in silico* predictions, may be used. Different approaches have been suggested to solve this problem [5], and for some lines of evidence, likelihood ratios (LRs) have been determined to integrate evidence into a probability score [6] of pathogenicity which allows classification in a 5-tiered system [4], ideally yielding a clear classification of a variant.

However, for other types of small coding variants, as well as for so-called private variants for which sufficient appropriate information is usually lacking, pathogenicity classification

requires alternative procedures. In these cases, it is possible to perform functional testing to assess the impact of a variant, which has frequently served to approach a classification for such variants [5,7–10].

We have previously shown that small, coding variants most frequently cause destabilization and/or inactivation of the resulting MLH1 protein, and provided reference variants that facilitate translation of biochemical test results in pathogenicity information [11].

Genetic counselling and testing has recently been introduced in Latin America and has led to the identification of several yet uncharacterized genetic variants in *MLH1* [12]. In this work, we have tested nine small coding variants identified in 9 South American individuals. We describe the experimental evaluation of their effects at protein level. We further discuss the results in context with all the available clinical information that has been gathered from the affected patients and their families, and also assess the structural role of the altered residues and their conservation in evolution. While biochemical data alone is not yet considered fully sufficient for variant classification, we show that information from different lines of evidence was consistent and strongly supported conclusions prompted by the biochemical results. Thus, we found that four alterations which disrupt protein function are causative for Lynch syndrome, while another four are likely neutral. Only one alteration evaded classification due to unclear biochemical results and therefore requires further analyses.

Materials/Subjects and Methods.

South American cancer families.

Unpublished data from hereditary cancer registries and published data from patients with suspected Lynch syndrome have been included in this work. The data include results from germline DNA testing, tumor testing (based on microsatellite instability analysis and/or immunohistochemistry) and family history [12–15].

Families that fulfilled the Amsterdam criteria and/or the Bethesda guidelines [16–18] were selected from 4 hereditary cancer registries: Hospital Italiano (Buenos Aires, Argentina), Barretos Cancer Hospital (Barretos, Brazil), Clinica Las Condes (Santiago, Chile), and Hospital de las Fuerzas Armadas (Montevideo, Uruguay).

Patients were informed about their inclusion into the registries and written informed consent was obtained from all participants during genetic counseling sessions.

Nomenclature and classification of genetic variants.

The nomenclature guidelines of the Human Genome Variation Society (HGVS) were used to describe the detected genetic variants [19]. The recurrence of the identified variants was established by interrogating four databases (in their latest releases as of September 2017): the Leiden Open Variation Database (LOVD), the Universal Mutation Database (UMD), ClinVar and the Human Gene Mutation Database (HGMD). The variants were classified according to the 5-tier classification system into the following categories: class 5 (pathogenic), class 4 (likely pathogenic), class 3 (uncertain variants or variants of unknown significance, VUS), class 2 (likely not pathogenic) and class 1 (not pathogenic) [4]. Syntax of all variants was verified using Mutalyzer [20] using the current *MLH1* reference sequence (NM_000249.3).

Cell lines.

HEK293T cells were used for this work and were kindly provided by Prof. Josef Jiricny, Zürich, Switzerland, in the year 2001. Their identity was confirmed by comparison of its genomic short tandem repeat (STR) profile from 9 loci with the source HEK293T cell line DSMZ ACC 635 and further confirmed by a variable number tandem repeat (VNTR) profile by the Leibnitz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany.

Protein expression and quantification.

pcDNA3-MLH1, pSG5-PMS2, and the HEK293T cell line have been described previously [21,22]. Missense variants were generated by site-directed mutagenesis (QuikChange 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, "Max" linear, 40 kDa, Polysciences,Warrington, PA) and extracted as described previously [11,23]. The extracts were analyzed by SDS-PAGE and immunoblotting (using anti-MLH1, G168-728, BD Biosciences, and anti-PMS2, E-19, and anti-beta-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.

Evaluation of expression defects with respect to pathogenicity.

Protein expression and quantification were performed in parallel with a stability-impaired neutral control variant (*MLH1* p.Val716Met) and a severely destabilized pathogenic control variant (*MLH1* p.Ala681Thr) as described before [11]. A clinically pathogenic defect of protein stability is present when the expression of the variant in question was similar or below that of the pathogenic control variant as extensively demonstrated before [9,11].

MMR activity.

The MMR activity of *MLH1* variants was scored *in vitro* as described previously [11,23]. 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. 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 = (intensity of bands of repaired substrate) / (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 MLH1 variants was analyzed in direct comparison to a wild-type protein that had been produced in parallel, and calculated as e(relative)=e(variant)/e(wild-type)*100.

Structural analyses.

Function-structure evaluations were performed with an updated model of human MutLα (MLH1-PMS2) [9] based on the N-terminal domains of human PMS2 NTD [24] and an homology model of MLH1-NTD [23]. The C-terminal domains (CTD) were built by homology modeling using yeast MutLα (PDB codes: 4E4W and 4FMN) and human MLH1 (PDB code: 3RBN) structures as templates. 4E4W, the highest resolution yeast MutLα structure, was used for modeling the conformation of the dimeric interface and as a template for modeling missing regions of MLH1 and the whole PMS2 subunit. 4FMN was used to model one loop missing from 4E4W structure. The MIP-box peptide (a fragment of NTG2) was taken from 4FMN structure and the zinc ions from 4E4W. The modeling templates were identified and selected using MODexplorer [25]. The target-template alignment used for modeling PMS2 was evaluated and refined using MODalign [26]. The final model was constructed after exporting the alignments from MODexplorer and MODalign and running Modeller [27] on a hybrid template

containing all necessary fragments from the selected templates, the MIP-box, and zinc ions. Figures were generated using PyMOL v.1.4.1 (Schrödinger LLC).

Results

Testing of nine small coding *MLH1* variants identified in nine South American cancer patients.

Nine small coding variants identified in nine CRC patients from South America were chosen for analysis (**Table 1**). Of these, five have not been reported before. Three of the previously described variants have been listed in InSiGHT class 3, meaning that their pathogenic potential is unclear. Only one variant (p.Lys618del), for which sufficient clinical data is available, has been classified pathogenic (class 5). We included this variant in our analysis as an additional pathogenic control.

Although family cancer disease information was available for six variant carriers (**Supplementary Figure S1**), it was not possible to perform additional genetic analyses in their relatives to assess cosegregation, which is a highly reliable method for the assessment of pathogenicity. Moreover, no sufficient information on molecular tumor traits (microsatellite instability and BRAF status) could be used for pathogenicity classification by Bayesian integrative multifactorial analysis [6] (**Table 1**). For assessing pathogenicity, we therefore had to rely on alternative ways, primarily functional testing of the genetic variants *in vitro*.

Functional analysis of the *MLH1* variants.

Loss of protein stability (resulting in absence of protein) and loss of repair activity are both directly associated with disease. Therefore, testing these parameters provides significant information for pathogenicity assessment.

The most frequent consequence of small coding alterations in MLH1 has been shown to be a decrease of the protein stability, which can be determined by expression of the variant cDNA [11]. Decreases in protein stability entail lower protein levels, which, even if the protein is functional, causes loss of repair activity below a certain threshold [11,28]. In order to translate expression defects into pathogenicity statements, we used previously established reference variants [11]. We applied reference variant MLH1 p.Ala681Thr, which allows to identify variants whose destabilization is severe enough to confer a pathogenic effect in humans due to the low cellular protein levels. Additionally, the neutral polymorphism MLH1 p.Val716Met was used as a reference for clinically neutral defects of stability [11].

We thoroughly assessed the expression levels of all variants in direct comparison to wildtype MLH1 protein and the two reference variants (**Figure 1A**). Results of several independent experiments were analyzed (**Figure 1B**). Six of the variants were at least as strongly expressed as the wildtype protein and therefore do not display stability problems. In contrast, three variants displayed stability decreases similar or stronger than the reference variant: p.Arg265Pro, p.Lys618del and p.Leu676Pro. In two of these cases, expression was significantly below the reference variant for pathological expression defects ($p<0.05$). For these strongly destabilized variants, a pathogenic defect can be concluded since insufficient MLH1 protein is available in the cell [11].

Another major reason for pathogenicity of small coding MLH1 variants is when these confer catalytic inactivity of the variant MLH1 protein. Since the major task of MLH1 protein is supporting DNA mismatch repair, we assessed the ability of the variants to perform the repair reaction *in vitro* (**Figure 2A**). Several independent experiments were performed to validate the results (**Figure 2B**).

Four of the variant proteins were indistinguishable from wildtype in this analysis, while only a residual repair signal identical to the negative control sample demonstrated that four variants

had fully lost repair activity (p.Tyr97Asp, p.Arg265Pro, p.Lys618del and p.Leu676Pro). One variant (p.His112Gln) showed slightly reduced activity.

Conservation and structural roles of the affected residues. Analysis of the structural positions and conservation of variant residues can explain and thereby confirm the functional results of mutational studies. We therefore analyzed the conservation and considered the positions of the affected residues in a structural model of the MutLα heterodimer (**Figure 3**).

Tyr97 is located in the "ATP lid", a highly conserved structure vital for functionality of the ATPase [29–31]. The non-conservative substitution of tyrosine to aspartate is expected to disturb the ATPase, explaining the observed loss of catalytic function and corroborating a pathogenic effect.

His112 is located in an extensive β-sheet that forms the back of the ATPase pocket. This residue is involved in interaction of MLH1 with the MutSα [23]. While inversion of the side-chain charge abolished interaction and rendered MLH1 DNA repair deficient [23], the less drastic substitution to glutamine described here had a milder effect on DNA repair.

Pro141 is in a similar position, albeit on the other side of the β-sheet. A specific role in protein structure or function is not obvious. Moreover, alanine is frequently present in this position in other organisms (**Table 2** and **Supplementary Figure S1**), strongly suggesting that it is a neutral substitution, which is confirmed by the biochemical results.

Arg265 is located in a highly conserved loop. It establishes a side-chain hydrogen bond to the ATPase pocket. Therefore, it has been suggested to be a detector switch that transmits ATPase signals to conformational changes [30]. A substitution of this highly relevant residue to serine has been described before and was also found to abolish repair activity [32]. Since the residue exerts its biochemical effects through its side chain, the mutation to proline can be expected to display an even more damaging defect, explaining the biochemical loss of function of the variant and corroborating a pathogenic effect.

Asn338 is located in the unstructured linker connecting the N- and C-terminal domains. This linker is not conserved nor crystallized and therefore not present in the structural model. This alone argues against a relevant role of Asn338 in MLH1. Moreover, substitutions for serine in this position are present in many organisms (**Table 2** and **Supplementary Figure S2**), corroborating the functional finding that the substitution is neutral.

Ile501 also belongs to this linker region, but it is present in the structure of the C-terminal domains. It is located in a loop, therefore its deletion does not affect secondary structure. Since the linker region is highly variable not only in sequence but also in length, deletions or insertions seem to be largely acceptable without loss of function. These considerations are in good agreement with the experimental finding that it did not affect biochemical functionality of MLH1.

Arg575 is located in the conserved C-terminal domain. It connects a β-sheet secondary structure with a loop. It is located on the protein surface, and far from functional core regions of the CTD like the zinc-binding [33], dimerization [34,35] and endonuclease [33,36] domains, and it is poorly conserved. Moreover, this conservative substitution is also frequently present in MLH1 proteins from other organisms (**Table 2** and **Supplementary Figure S3**), corroborating that it is a neutral exchange, which is consistent with the biochemical findings.

Deletion of **lysine 618** affects the center of an α-helix. Consequently, the deletion causes a rotation of the C-terminal residues around the helix axis and distort the orientation of internal (hydrophobic) and external (hydrophilic) residues. Therefore, it can be expected to severely affect protein structure (and function), which is reflected by the low stability of the mutant protein.

Leu676 is located in the middle of another α-helix, within a conserved structural three-helix motif that we have shown before to be very sensitive to substitutions which regularly cause severe protein destabilization [11]. Therefore, the functional finding of severe instability is fully consistent with this substitution to the structure breaker amino acid proline. A different

substitution of this position (Leu676Arg) has been found to have the same defect before and was therefore considered pathogenic [11].

Comprehensive evaluation of evidence.

In the Amsterdam-positive family from Uruguay (**Uru-8249/Uru-8461**, **Supplementary Figure S4**), 11 directly related individuals in three generations were affected by CRC, with an average age of onset of 47 years. The index patient (carrier of the p.Tyr97Asp alteration) was diagnosed with CRC at 42 years. The high penetrance of Lynch-syndrome-associated cancers in the family and the low average age of onset strongly suggest a hereditary cause, and the finding that p.Tyr97Asp inactivates DNA mismatch repair strongly suggests that this alteration is the causative factor.

The index patient of family **Arg-39** carries two alterations, the substitutions p.**His112Gln** and p.**Arg575Lys**. There are two CRC cases in the family, who does not meet the Amsterdam II criteria. Since it was not possible to retrieve material from the family for determining allelic positions of the two variants, these were considered individually. While one variant is highly likely neutral, the clinical outcome of the other, biochemically compromised variant (p.His112Gln) remains unclear. The comparatively low penetrance of Lynch-syndrome-associated cancers in the family, however, rather suggests a moderate, if any, pathogenic effect. Another observation that supports neutrality of this variant is its frequency in some control populations, which approximates the general incidence of Lynch syndrome (**Table 1**): it is present at a rate of 1:8650 in East Asia, while the general incidence of Lynch syndrome, i.e. all pathogenic mutations in all relevant genes, has been estimated to 1:660-1:2200; [37].

One individual from Colombia carried the *MLH1* p.**Pro141Ala**, for which we did not find biochemical evidence for conferring a loss of function. Moreover, alanine in this position seems evolutionary acceptable because it occurs naturally in MLH1 proteins of 8% of all

analyzed organisms (**Table 2**), strongly suggesting neutrality of this substitution. Additionally, the variant showed co-occurrence in the carrier with a pathogenic mutation in MSH2. While this invalidates use of the medical data for interpretation of the specific effect of Pro141Ala, co-occurrence of a pathogenic variant with an unclear variant is generally considered to make pathogenicity less likely for the unclear variant [38].

In the Amsterdam-positive family **Chi-43**, 10 of 41 individuals were affected from Lynch-syndrome-associated neoplasms in 4 generations, with an average age of onset of 25 years. This family record is highly suspicious for Lynch syndrome, and the finding that the identified alteration (p.Arg265Pro) abolished the biochemical function of the protein consistently speak in favor of pathogenic effect in humans.

A comparatively young Brazilian CRC patient (38 years at diagnosis) was found to carry a variant causing the **p.Asn338Ser** substitution in MLH1. While MSI has been observed in the tumor of a German carrier of this variant [39], all other information on this variant suggest neutrality. Therefore, the singular observation of MSI in one tumor probably has been coincidence caused by somatic loss of mismatch repair, for example by MLH1 hypermethylation.

Another **Brazilian** CRC patient of 60 years of age carried a three-nucleotide deletion causing the loss of one amino acid (**isoleucine 501**) from the MLH1 protein. The consequences of amino acid deletions are, in general, expected to be worse and harder to predict than substitutions. However, this deletion had no impact on biochemical performance of MLH1, and this was in good agreement with the low conservation and structural position of the deleted residue at the end of the highly variant linker region of MLH1. We therefore consider it to be likely neutral.

In contrast, the other one-residue deletion in this study (**p.Lys618del**) was found in an Amsterdam-positive family from Argentina with multiple first-degree relatives with colorectal carcinoma (Arg-43; average age at cancer diagnosis: 44 years). Biochemical evidence

confirms that the variant is defective. The variant p.Lys618del has already been classified in class 5.

Finally, in two Amsterdam-positive Brazilian families (**BRA-1** and **SL-36D**) who presented in summary with 8 cancer cases in 26 recorded members (average age of onset 48 years) the index patients carried the alteration **p.Leu676Pro**. Molecular tumor data (MSI, IHC) are at odds (Table 1), but the variant has not been found in 280 healthy controls [15]. In our analysis, this alteration abrogated stability and repair activity of MLH1, strongly suggesting causality for the cancer predisposition in these two families.

Discussion

Pathogenicity classification of small coding *MLH1* variants is sometimes straightforwardly possible if sufficient clinical data is available. However, so-called private variants (rare variants with little clinical and family information) may fail in established classification approaches and then require alternative methods for pathogenicity assessment. Here, we evaluated a series of small coding *MLH1* variants identified in South American patients suspected to have Lynch syndrome. While extensive information on cancer cases in the families was available from some of the affected individuals, genetic testing has only been performed in the index individuals, and practically no molecular tumor data was available. Therefore, although even little clinical information can sometimes facilitate classification by integrative analysis [10], this was not possible in these cases.

We therefore had to rely on functional analysis in combination with other lines of evidence to approach a classification for these variants and facilitate diagnosis for the affected families.

We used a previously established robust method that allows to deduce pathogenicity information from functional evidence, and also used other lines of evidence (from evolutionary information, protein structure and family clinical data). Comprehensive analysis

showed that evidence for pathogenicity or neutrality was consistent: it coherently supported that two alterations confer a pathogenic effect by catalytic inactivation (the two ATPase mutations, p.Tyr97Asp and p.Arg265Pro). For two further alterations, consistent evidence suggested a pathogenic effect by destruction of protein integrity (p.Lys618del and p.Leu676Pro).

Conversely, findings for another four variants coherently supported neutrality: they showed no detectable defects in functionality. While this finding is usually hampered by the caveat that the applied biochemical assay system may not reflect all aspects of biological function, in these cases we also found that the investigated patient substitutions occur quite frequently in evolution. This observation strongly supports neutrality similarly to finding a significant frequency of a given variant in unaffected control populations. Moreover, one variant (Pro141Ala) co-occurred with a clearly pathogenic variant in its carrier (**Table 1**), which is also considered a strong indicator of neutrality.

Just one variant remained for which neither functional nor other evidence gave a clear result: for p.His112Gln, functional data showed intermediate results which would require more sensitive reference variants to determine if the effects will result in a clinical phenotype.

Taken together, the analysis of biochemical function, performed with suitable controls enabling clinically meaningful result readout, in conjunction with additional evidence is able to yield a reliable pathogenicity statement if coherent results are achieved. This approach may therefore step in to the breach if other ways of classification fail to provide sufficient evidential value.

Acknowledgements

We wish to thank the families for their participation and contribution to this study.

Fundings

Barretos Cancer Hospital was partially funded by FINEP - CT-INFRA (02/2010), Radium Hospital Foundation (Oslo, Norway), Helse Sør-Øst (Norway). This work was supported by the Deutsche Forschungsgemeinschaft (grant PL688/2-1) to G.P.

Conflict of interest

The authors declare no conflict of interest.

References

1. Kobayashi H, Ohno S, Sasaki Y, Matsuura M. Hereditary breast and ovarian cancer susceptibility genes (review). *Oncol Rep*. 2013;30(3):1019-1029.
2. Kuiper RP, Vissers LELM, Venkatachalam R, et al. Recurrence and variability of germline EPCAM deletions in Lynch syndrome. *Hum Mutat*. 2011;32(4):407-414.
3. Peltomäki P. Update on Lynch syndrome genomics. *Fam Cancer*. 2016;15(3):385-393.
4. 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.
5. Kansikas M, Kariola R, Nyström M. Verification of the three-step model in assessing the pathogenicity of mismatch repair gene variants. *Hum Mutat*. 2011;32(1):107-115.
6. Thompson BA, Goldgar DE, Paterson C, et al. 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*. 2013;34(1):200-209.
7. Borràs E, Pineda M, Brieger A, et al. Comprehensive functional assessment of MLH1 variants of unknown significance. *Hum Mutat*. 2012;33(11):1576-1588.

8. Borràs E, Pineda M, Cadiñanos J, et al. Refining the role of PMS2 in Lynch syndrome: Germline mutational analysis improved by comprehensive assessment of variants. *J Med Genet*. 2013;50(8):552-563.
9. Hinrichsen I, Schäfer D, Langer D, et al. Functional testing strategy for coding genetic variants of unclear significance in MLH1 in Lynch syndrome diagnosis. *Carcinogenesis*. 2015;36(2):202-211.
10. Tricarico R, Kasela M, Mareni C, et al. Assessment of the InSiGHT Interpretation Criteria for the Clinical Classification of 24 MLH1 and MSH2 Gene Variants. *Hum Mutat*. 2017;38(1):64-77.
11. Hinrichsen I, Brieger A, Trojan J, Zeuzem S, Nilbert M, Plotz G. Expression defect size among unclassified MLH1 variants determines pathogenicity in Lynch syndrome diagnosis. *Clin. Cancer Res.* 2013;19(9):2432-2441.
12. Rossi BM, Palmero EI, López-Kostner F, et al. A survey of the clinicopathological and molecular characteristics of patients with suspected Lynch syndrome in Latin America. *BMC Cancer*. 2017;17(1):623.
13. Valentin MD, da Silva FC, dos Santos EMM, et al. Characterization of germline mutations of MLH1 and MSH2 in unrelated south American suspected Lynch syndrome individuals. *Fam Cancer*. 2011;10(4):641-647.
14. Dominguez-Valentin M, Nilbert M, Wernhoff P, et al. Mutation spectrum in South American Lynch syndrome families. *Hered Cancer Clin Pract*. 2013;11(1):18.
15. Carneiro da Silva F, Ferreira JRdO, Torrezan GT, et al. Clinical and Molecular Characterization of Brazilian Patients Suspected to Have Lynch Syndrome. *PLoS ONE*. 2015;10(10):e0139753.
16. Vasen HFA, Mecklin JP, Khan PM, Lynch HT. THE INTERNATIONAL-COLLABORATIVE-GROUP-ON-HEREDITARY-NON-POLYPOSIS-COLORECTAL -CANCER (ICG-HNPCC). *Dis. Colon Rectum*. 1991;34(5):424-425.
17. Rodriguez-Bigas MA, Boland CR, Hamilton SR, et al. A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines. *J.Natl.Cancer Inst.* 1997;89(23):1758-1762.

18. Vasen HF, Watson P, Mecklin JP, Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology*. 1999;116(6):1453-1456.
19. den Dunnen JT, Antonarakis SE. Mutation nomenclature extensions and suggestions to describe complex mutations: A discussion. *Hum. Mutat.* 2000;15(1):7-12.
20. Wildeman M, van Ophuizen E, den Dunnen JT, Taschner PE. Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. *Hum Mutat*. 2008;29(1):6-13.
21. Trojan J, Zeuzem S, Randolph A, et al. Functional analysis of hMLH1 variants and HNPCC-related mutations using a human expression system. *Gastroenterology*. 2002;122(1):211-219.
22. Plotz G, Raedle J, Brieger A, Trojan J, Zeuzem S. N-terminus of hMLH1 confers interaction of hMutLalpha and hMutLbeta with hMutSalpha. *Nucleic Acids Res.* 2003;31(12):3217-3226.
23. Plotz G, Welsch C, Giron-Monzon L, et al. Mutations in the MutSalpha interaction interface of MLH1 can abolish DNA mismatch repair. *Nucleic Acids Res*. 2006;34(22):6574-6586.
24. Guarne A, Junop MS, Yang W. Structure and function of the N-terminal 40 kDa fragment of human PMS2: a monomeric GHL ATPase. *EMBO J.* 2001;20(19):5521-5531.
25. Kosinski J, Barbato A, Tramontano A. MODexplorer: An integrated tool for exploring protein sequence, structure and function relationships. *Bioinformatics*. 2013;29(7):953-954.
26. Barbato A, Benkert P, Schwede T, Tramontano A, Kosinski J. Improving your target-template alignment with MODalign. *Bioinformatics*. 2012;28(7):1038-1039.
27. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*. 1993;234(3):779-815.
28. Kansikas M, Kasela M, Kantelinen J, Nyström M. Assessing how reduced expression levels of the mismatch repair genes MLH1, MSH2, and MSH6 affect repair efficiency. *Hum Mutat*. 2014;35(9):1123-1127.
29. Raschle M, Dufner P, Marra G, Jiricny J. Mutations within the hMLH1 and hPMS2 Subunits of the Human MutLalpha Mismatch Repair Factor Affect Its ATPase Activity, but Not Its Ability to Interact with hMutSalpha. *J.Biol.Chem.* 2002;277(24):21810-21820.

30. Ban C, Junop M, Yang W. Transformation of MutL by ATP binding and hydrolysis: a switch in DNA mismatch repair. *Cell*. 1999;97(1):85-97.
31. Ban C, Yang W. Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. *Cell*. 1998;95(4):541-552.
32. Drost M, Zonneveld JeBM, van Dijk L, et al. A cell-free assay for the functional analysis of variants of the mismatch repair protein MLH1. *Hum Mutat*. 2010;31(3):247-253.
33. Kosinski J, Plotz G, Guarne A, Bujnicki JM, Friedhoff P. The PMS2 subunit of human MutLalpha contains a metal ion binding domain of the iron-dependent repressor protein family. *J Mol Biol*. 2008;382(3):610-627.
34. Kosinski J, Hinrichsen I, Bujnicki JM, Friedhoff P, Plotz G. Identification of Lynch syndrome mutations in the MLH1-PMS2 interface that disturb dimerization and mismatch repair. *Hum Mutat*. 2010;31(8):975-982.
35. Cutalo JM, Darden TA, Kunkel TA, Tomer KB. Mapping the dimer interface in the C-terminal domains of the yeast MLH1-PMS1 heterodimer. *Biochemistry*. 2006;45(51):15458-15467.
36. Kadyrov FA, Dzantiev L, Constantin N, Modrich P. Endonucleolytic function of MutLalpha in human mismatch repair. *Cell*. 2006;126(2):297-308.
37. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285-291.
38. Karbassi I, Maston GA, Love A, et al. A Standardized DNA Variant Scoring System for Pathogenicity Assessments in Mendelian Disorders. *Hum Mutat*. 2016;37(1):127-134.
39. Hardt K, Heick SB, 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.

Figure legends

Figure 1: Analysis of expression of the *MLH1* variants.

A: Expression of wildtype and variant MLH1 proteins was visualized by SDS-PAGE and western blotting. The two stability reference variants Ala681Thr (pathogenic expression defect) and Val716Met (non-pathogenic expression defect, polymorphism) were transfected in parallel. The shown blots are representative for 9 independent experiments that were performed and delivered the data shown in evaluation (Figure 1B).

B: Average expression values in percent of the wildtype expression, standard deviations and results of the t-test of 9 independent experiments are shown for wildtype and variant MLH1 protein expression. The column of the reference variant for a pathogenic expression defect (p.Ala681Thr) is hatched, expression levels significantly below the dotted line established by this reference variant represent pathogenic stability defects of the variant protein (Lys618del and Leu676Pro).

Figure 2: Analysis of mismatch repair activity of the *MLH1* variants.

DNA mismatch repair activity was assessed for wildtype and variant MLH1 proteins, and a negative control (without MLH1 protein) was included as detailed in Materials and Methods.

A. The extent of repair is visible in the agarose gel electrophoresis by the generation of 2 smaller fragments of the (unrepaired) linearized plasmid. This is shown in the representative agarose gel image (composed of three experiments since always 6 samples were processed in parallel).

B. Between 3 and 9 independent experiments were performed for each variant, and repair activity was scored relative to wildtype MLH1 protein (100%). Average repair values and standard deviations are shown.

Figure 3. Structural positions of affected residues in the MutLα heterodimer. The model is shown in cartoon presentation, comprising the N-terminal (aa 1-331) and C-terminal (aa 488-756) domains of MLH1 (green) and the corresponding partner domains PMS2 (right side). The MLH1 subunits are rotated by 90° on the left side to improve presentation of the residues studied (shown as red balls). The ADP- Mg^{2+} bound by the MLH1 ATPase is shown in blue, the zinc ions of the exonuclease domain in orange. The locations of residues affected by alterations are indicated by red balls representing the alpha carbon atoms. For the two deletion variants, Ile501del and Lys618del, it is impossible to properly indicate the affected residue because both occur in the context of two (Ile500-Ile501) or three (Lys616-Lys617-Lys618) repetitions of the same amino acid; in these cases, only the residue with the lowest number is indicated.

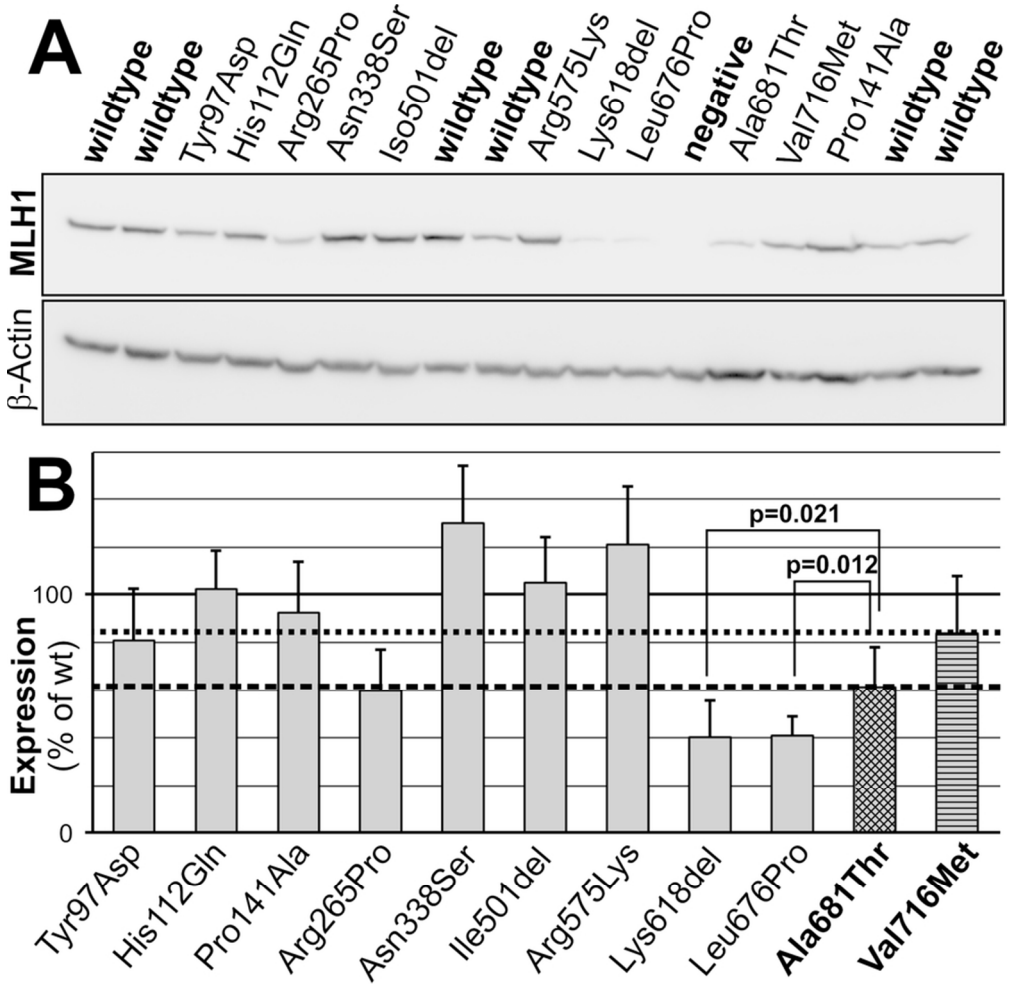

Figure 1: Analysis of expression of the MLH1 variants.
A: Expression of wildtype and variant MLH1 proteins was visualized by SDS-PAGE and western blotting. The two stability reference variants Ala681Thr (pathogenic expression defect) and Val716Met (non-pathogenic expression defect, polymorphism) were transfected in parallel. The shown blots are representative for 9 independent experiments that were performed and delivered the data shown in evaluation (Figure 1B).
B: Average expression values in percent of the wildtype expression, standard deviations and results of the t-test of 9 independent experiments are shown for wildtype and variant MLH1 protein expression. The column of the reference variant for a pathogenic expression defect (p.Ala681Thr) is hatched, expression levels significantly below the dotted line established by this reference variant represent pathogenic stability defects of the variant protein (Lys618del and Leu676Pro).

82x81mm (300 x 300 DPI)

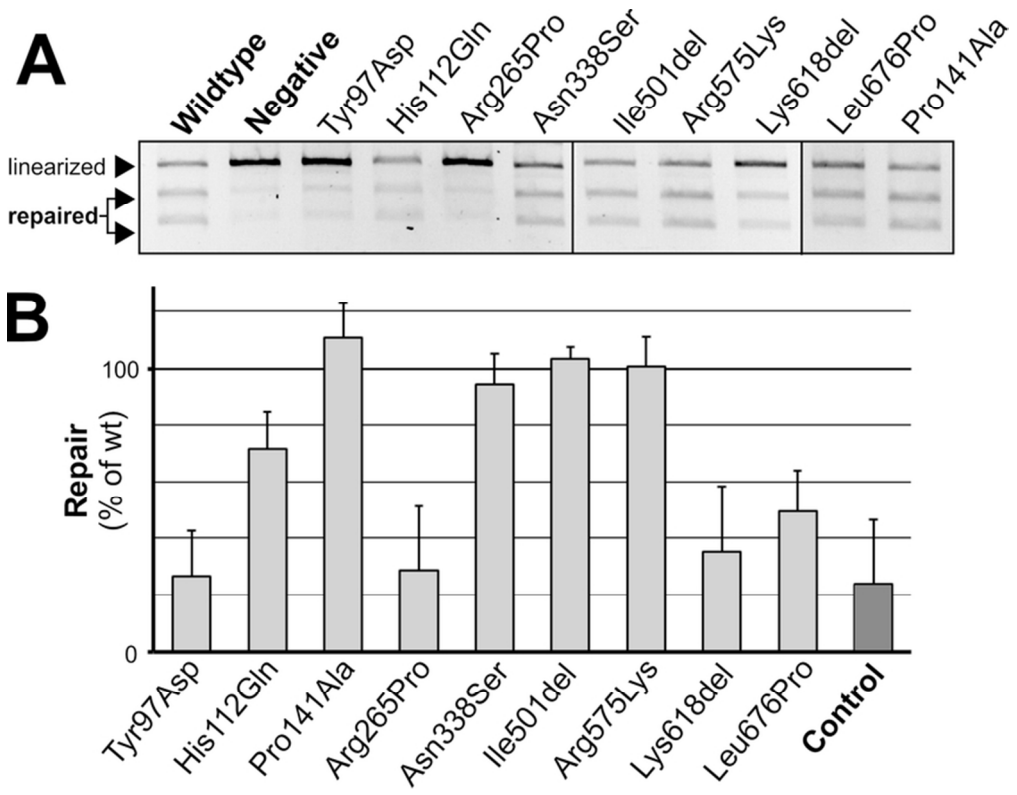

Figure 2: Analysis of mismatch repair activity of the MLH1 variants.
DNA mismatch repair activity was assessed for wildtype and variant MLH1 proteins, and a negative control (without MLH1 protein) was included as detailed in Materials and Methods. A. The extent of repair is visible in the agarose gel electrophoresis by the generation of 2 smaller fragments of the (unrepaired) linearized plasmid. This is shown in the representative agarose gel image (composed of three experiments since always 6 samples were processed in parallel).
B. Between 3 and 9 independent experiments were performed for each variant, and repair activity was scored relative to wildtype MLH1 protein (100%). Average repair values and standard deviations are shown.

71x55mm (300 x 300 DPI)

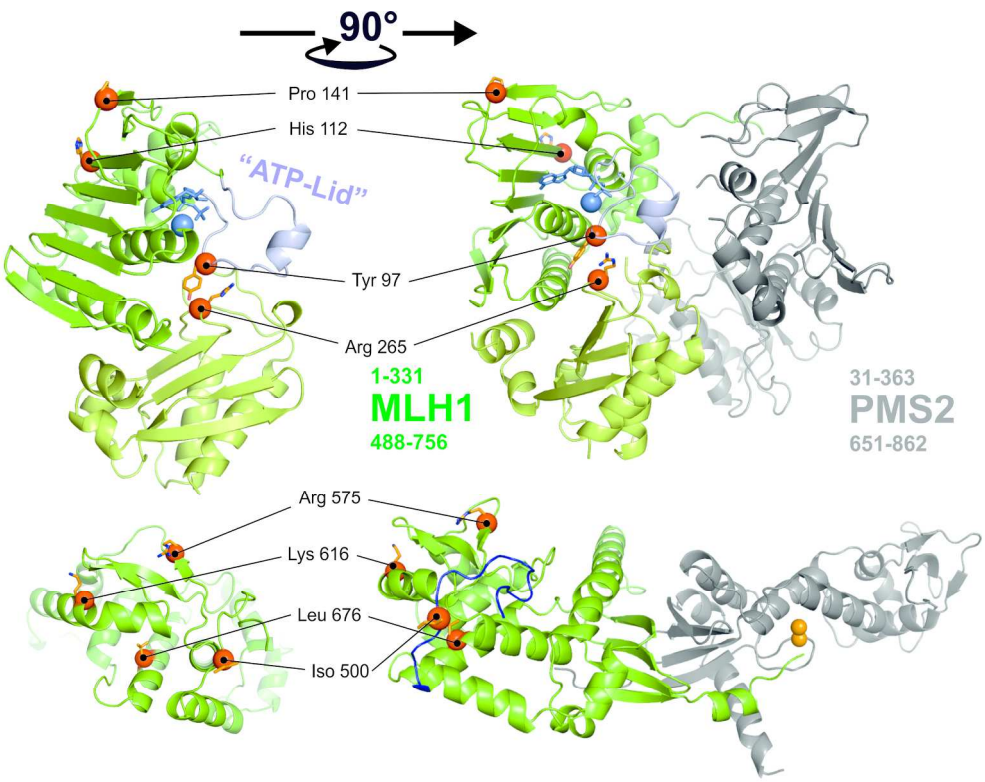

Figure 3. Structural positions of affected residues in the MutLα heterodimer. The model is shown in cartoon presentation, comprising the N-terminal (aa 1-331) and C-terminal (aa 488-756) domains of MLH1 (green) and the corresponding partner domains PMS2 (right side). The MLH1 subunits are rotated by 90° on the left side to improve presentation of the residues studied (shown as red balls). The ADP- Mg2+ bound by the MLH1 ATPase is shown in blue, the zinc ions of the exonuclease domain in orange. The locations of residues affected by alterations are indicated by red balls representing the alpha carbon atoms. For the two deletion variants, Ile501del and Lys618del, it is impossible to properly indicate the affected residue because both occur in the context of two (Ile500-Ile501) or three (Lys616-Lys617-Lys618) repetitions of the same amino acid; in these cases, only the residue with the lowest number is indicated.

200x155mm (300 x 300 DPI)

Table 1:
Clinical and genetic information of the investigated South American patients.

Familiy ID	Origin	Clinical criteria	Index patient (tumor) information	Alteration (gDNA)	Alteration (Protein)	Population frequency (Exac database)	Previous reports of the variant
Uru-8249 Uru-8461	Uruguay	ACII ACII	CRC 42 y, MSI-L CRC 52 y, MSI-L	c.289T>G	p.Tyr97Asp	n.r.	Dominguez-Valentin et al. 2013, Rossi et al. 2017
Arg-39	Argentina	ACI	CRC 54 y, MSI-H, IHC: normal MMR	c.336T>A	p.His112Gln	0.0001156 (East Asia) 0.0000864 (Latino)	Dominguez-Valentin et al. 2013, Rossi et al. 2017
				c.1724G>A	p.Arg575Lys	n.r.	
UN-23	Colombia	ACII	CRC 45 y; additional pathogenic alteration: *MLH1* c.1856delG	c.421C>G	p.Pro141Ala	n.r.	Giraldo 2005; Chao 2002; Rossi et al. 2017
Chi-43	Chile	ACII	CRC 25 y	c.794G>C	p.Arg265Pro	n.r.	Rossi et al. 2017
1	Brazil	ACII	38 y	c.1013A>G	p.Asn338Ser	0.0000902 (European) 0.0000607 (South Asian)	Rossi 2002; Tournier 2008; Hardt 2011;
2	Brazil	ACII	60 y	c.1500_1502delCAT rs587778920	p.Ile501del	0.0000864 (Latin America)	Rossi 2002 Valentin 2011 Rossi et al. 2017
Arg-43	Argentina	ACI	CRC 44 y IHC: Loss MLH1, MSI-H	c.1852_1854delAAG	p.Lys618del	n.r.	InSIGHT
BRA-1	Brazil	ACI	CRC 53 y; BRAF wildtype, MSI-H, IHC: Loss MLH1/PMS2	c.2027T>C	p.Leu676Pro	n.r. 0/280	Hardt 2011; Dominguez-Valentin et al. 2013, Carneiro 2015; Rossi et al. 2017
SL-36D		ACI	CRC 43 y, BC 50 y, IHC normal, MSS.				

Abbreviations: CRC, colorectal carcinoma. BC: breast cancer. IHC, immunohistochemistry results. MSI-H, microsatellite instability-high in tumor (when ≥ 2 markers where unstable. MSI-L: microsatellite instability low (when one of the markers was unstable). MSS, microsatellite stable (when none of the markers were unstable). GC, gastric cancer. ACI/ACII: family meets Amsterdam criteria I or II; n.r., not reported; n.a., not applicable; n.d., no data.

Table 2.
Summary of the analyses on the variants

Alteration (gDNA)	Alteration (Protein)	Evolutionary information	Structural information relevant for pathogenic potential	Biochemical effect(s) (stability/functionality)	Database classifications			Proposed classification	Classification based on
					InSiGHT	UMD	ClinVar		
c.289T>G	p.Tyr97Asp	Highly conserved (8/9)	ATPase involved, functional loss plausible	Stable Non-functional	-	-	US	Likely pathogenic	Biochemistry, conservation, structure.
C.336T>A	p.His112Gln	Highly conserved (8/9); Gln rarely in this position (1%)		Stable Compromised	-	-	-	Unclear	
c.1724G>A	p.Arg575Lys	Intermediate conservation (5/9); Lys occurs in this position (13%)	No effect on protein function plausible	Stable Functional	-	-	-	Likely neutral	Biochemistry, evolutionary occurrence, structure.
c.421C>G	p.Pro141Ala	Intermediate conservation (6/9); Ala occurs in this position (8%)	No effect on protein function plausible	Stable Functional	3	-	US	Likely neutral	Biochemistry, evolutionary occurrence, structure.
c.794G>C	p.Arg265Pro	Highly conserved (8/9)	ATPase involved, functional loss plausible	Unstable (=A681T) Non-functional	-	-	LP	Likely pathogenic	Biochemistry, conservation, structure.
c.1013A>G	p.Asn338Ser	Intermediate conservation (7/9); Ser occurs in this position (7%)	No effect on protein function plausible	Stable Functional	3	UV	US	Likely neutral	Biochemistry, evolutionary occurrence, structure.
c.1500_1502delCAT rs587778920	p.Ile501del	Intermediate conservation (4/7); Indels occur in linker	No effect on protein function plausible	Stable Functional	-	-	US	Likely neutral	Biochemistry, conservation, structure.
c.1852_1854delAAG	p.Lys618del	Intermediate conservation (4/7); No Indels: Lys615-Met621	Protein distortion highly likely	Unstable (<A681T)** Non-functional	5	Causal	Patho-genic	Pathogenic	Biochemistry, length conservation, structure.
c.2027T>C	p.Leu676Pro	Intermediate conservation (6/9);	Loss of stability plausible Arg variant has same effect	Unstable (<A681T)** Non-functional	3	-	US	Likely pathogenic	Biochemistry, structure; observation of another pathogenic substitution in this position.

Abbreviations: InSiGHT: International Society for Gastrointestinal Hereditary Tumors; UMD: Universal mutation database; UV: Unclassified variant; US: Uncertain significance; LP: likely pathogenic.