Functional testing strategy for coding genetic variants of unclear significance in MLH1 in Lynch syndrome diagnosis

Inga Hinrichsen, Dieter Schäfer1, Deborah Langer, Nicole Köger, Margarethe Wittmann, Stefan Aretz2, Verena Steinke2, Stefanie Holzapfel2, Jörg Trojan3, Rainer König1, Stefan Zeuzem3, Angela Brieger and Guido Plotz*

Biomedical Research Laboratory, Department of Internal Medicine 1 and 1Department of Human Genetics, Universitätsklinikum Frankfurt, Frankfurt D-60590, Germany, 2Institute of Human Genetics, University of Bonn, Bonn D-53127, Germany and 3Department of Internal Medicine 1, Universitätsklinikum Frankfurt D-60590, Frankfurt, Germany

Abstract

Lynch syndrome is caused by inactivating mutations in the MLH1 gene, but genetic variants of unclear significance frequently preclude diagnosis. Functional testing can reveal variant-conferred defects in gene or protein function. Based on functional defect frequencies and clinical applicability of test systems, we developed a functional testing strategy aimed at efficiently detecting pathogenic defects in coding MLH1 variants. In this strategy, tests of repair activity and expression are prioritized over analyses of subcellular protein localization and messenger RNA (mRNA) formation. This strategy was used for four unclear coding MLH1 variants (p.Asp41His, p.Leu507Phe, p.Gln689Arg, p.Glu605del + p.Val716Met). Expression was analyzed using a transfection system, mismatch repair (MMR) activity by complementation in vitro, mRNA formation by reverse transcriptase–PCR in carrier lymphocyte mRNA, and subcellular localization with dye-labeled fusion constructs. All tests included clinically meaningful controls. The strategy enabled efficient identification of defects in two unclear variants: the p.Asp41His variant showed loss of MMR activity, whereas the compound variant p.Glu605del + p.Val716Met had a defect of expression. This expression defect was significantly stronger than the pathogenic expression reference variant analyzed in parallel, therefore the defect of the compound variant is also pathogenic. Interestingly, the expression defect was caused additively by both of the compound variants, at least one of which is non-pathogenic when occurring by itself. Tests were neutral for p.Leu507Phe and p.Gln689Arg, and the results were consistent with available clinical data. We finally discuss the improved sensitivity and efficiency of the applied strategy and its limitations in analyzing unclear coding MLH1 variants.

Introduction

Lynch syndrome (MIM #120435), previously called hereditary non-polyposis colorectal cancer, is a relatively frequent hereditary predisposition for cancer of the colorectum, the endometrium and some other organs (1,2): ~1 in 660–2000 individuals is affected (3). For establishing diagnosis, an inactivating germine mutation in one of four mismatch repair (MMR) genes (MLH1, MSH2, MSH6 or PMS2) has to be found. This also permits predictive genetic testing in relatives, enabling targeted cancer surveillance and prevention measures in affected families (4).

In Lynch syndrome, MLH1 (MIM #120436) is the gene most frequently affected by mutations. One-third of the observed alterations are non-synonymous, non-truncating variants in the coding region: mostly missense variants, but also small in-frame insertion–deletion mutations and read-through alterations of the translation termination codon (5,6). These variants a priori have an unknown impact on health and therefore are called variants of uncertain (clinical) significance. They need to be classified concerning their pathogenic potential based on

Received: September 9, 2014; Revised: November 13, 2014; Accepted: November 26, 2014

© The Author 2014. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com.

© The Author 2014. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com.
**Abbreviations:**
cDNA complementary DNA
MMR mismatch repair
mRNA messenger RNA.

suitable evidence (7). Without classification, the variant cannot be used to establish a diagnosis in a patient, and relatives therefore cannot be offered predictive testing and targeted cancer prevention.

Two lines of evidence can serve to classify a variant of uncertain (clinical) significance. Suitable information on a variants’ effect is available directly from observations in humans (e.g. cosegregation of the variant with disease, frequency in healthy controls, age of cancer onset and occurrence of microsatellite instability in the tumor). Alternatively, information can be acquired by functional laboratory analyses. These investigations are a powerful alternative tool for classification and are the only way of performing classification in all cases with a paucity of information from affected individuals.

Non-synonymous genetic variants located in the coding sequence (‘coding variants’) can affect gene function in different ways, from transcription and splicing to protein stability, catalytic activity, binding to its functional protein partners and subcellular distribution. For clarification of pathogenicity by functional analyses, a testing strategy should optimally meet two challenges: for best performance, it should focus first on functional parameters that most frequently cause pathogenic effects in patients. Furthermore, the experimental result should be translatable into a disease risk statement by inclusion of meaningful controls.

Based on this premise, we have considered current information on functional test systems for rationally developing a functional testing strategy for coding MLH1 variants. We have exemplarily applied it to MLH1 and PMS2 variants identified in six colorectal cancer patients in order to investigate their pathogenic potential and test the performance of the strategy.

**Materials and methods**

**Cell line authentication**

HEK293T cells were used for this work and were kindly provided by Prof. J. Jiricny, Zürich, Switzerland, in the year 2001. Their identity was confirmed by comparison of its genomic short tandem repeat profile from nine loci with the source HEK293T cell line DSMZ ACC 635 and further confirmed by a variable number tandem repeat profile by the Leibniz Institute DSMZ—German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany, in 2009, just before the experiments described in this paper were initiated.

**Variants for analysis**

Genetic variants that have been found in six cancer patients who were referred to the Human Genetics Department for counseling and fulfilled clinical diagnostic criteria for Lynch syndrome (Bethesda or Amsterdam criteria) were used for the analysis (see Table I for comprehensive data). The study was approved by the local ethics committee. Syntax of all variants was verified using MutaTyper (8) with the current reference sequences (MLH1: NM_000249.3; PMS2: NM_000535.5).

**Protein expression and quantification**

pcDNA3-MLH1, pSG5-PMS2, and the HEK293 and HEK293T cell lines have been described previously (9,10). Missense variants were generated by site-directed mutagenesis (QuickChange II Kit; Stratagene) and confirmed by direct sequencing. Transfected 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 (9,11,12). The extracts were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis 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 (Millipore) were detected in an LAS-4000 mini camera (Fuji) and quantified using Multi Gauge v3.2.

**Clinical evaluation of expression defects**

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 (12). A clinically pathogenic defect of protein stability was assumed when the expression of the variant in question was similar or below that of the pathogenic control variant.

**MMR activity**

The MMR activity of MLH1 variants was scored in vitro as described previously (11,12). 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 Asel. 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 to 90%. The repair efficiency of MutLα variants was analyzed in direct comparison with a wild-type protein that had been produced in parallel and calculated as e (relative) = e (variant/e (wild-type) × 100.

**Allelic expression**

RNA analysis was performed for the MLH1 variants p.Leu507Phe, p.Glu605del and p.Asn414His. RNA was extracted from total blood using PAXgene Blood RNA Tubes (PreAnalytix; Qiagen, Hilden, Germany) by use of the PAXgene Blood RNA Kit (Qiagen) according to the manufacturer’s protocol. Complementary DNA (cDNA) was synthesized with the Superscript first strand system for reverse transcription–PCR (Invitrogen) according to the manufacturer’s protocol. Variant containing MLH1 exons were amplified by primers in the adjacent exons. PCR products were analyzed by Sanger sequencing.

**RNA extraction and RT–PCR**

Peripheral blood mononuclear cells were isolated from 10 ml blood using the CPT-Vacutainer system (BD Life Sciences). Total RNA was extracted using Trizol (Invitrogen) according to the manufacturer’s recommendations, and RNA was dissolved in 30 μl RNase-free water after ethanol precipitation. cDNA was created from 1 μg total RNA; 50 ng random primers (Promega) according to the manufacturer’s recommendations in a volume of 25 μl. Transcription was performed for 10 min at 25°C, followed by 50 min at 50°C. cDNA samples were stored at −20°C.

cDNA was amplified in four PCR reactions containing combinations of the following specific primers: MLH1-wt–1993F: 5′-GAGGAAGATGGTCCCAAAGGAC-3′; MLH1-605del–1993F: 5′-GAGGAAGATGGTCCCAAAGGAC-3′; MLH1-wt–2368R: 5′-AGGGCTTTATAGACAATGTGTTCCAC-3′; MLH1-V716M–2368R: 5′-AGGGCTTTATAGACAATGTGTTCCAC-3′. PCR was performed with AmpliTaq Gold (Applied Biosystems) and conditions were adjusted so that only correct template–primer combinations yielded a product (95°C for 2 min, 45 cycles with: 95°C for 30 s, 65°C for 30 s, 72°C for 1:30 min).

**Subcellular localization**

HEK293T cells were grown on glass cover slides coated with gelatin and transfected with the following vectors: pEGFP-C1 (Clontech), pEGFP-C1-MLH1 (wild-type or variants, p.Leu507Phe or p.Gln689Arg) and pEGFP-C1-MLH1-NLSdead (the MLH1 mutant mNL5bip for which an abrogation of the bipartite nuclear localization has been experimentally shown, ref. 13). The vectors were transfected either alone or cotransfected with pSG5-PMS2. Coverslips were retrieved 48 h after transfection, washed,
Table I. Clinical data and molecular genetics of the patients

<table>
<thead>
<tr>
<th>Patient number</th>
<th>Sex</th>
<th>Disease (age at diagnosis)</th>
<th>Tumor molecular markers</th>
<th>Molecular genetic analysis</th>
<th>Previous published data on patients (with their age at diagnosis) carrying the genetic variant (see text for references)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F001</td>
<td>F</td>
<td>Cancer col. asc. (52)</td>
<td>MSI-H (4/5)</td>
<td>PMS2-loss</td>
<td>Negative c.943C&gt;T, p.Arg315X None ID 204: right-sided CRC, 32 years, Amsterdam-negative family, IHC normal</td>
</tr>
<tr>
<td>F003</td>
<td>F</td>
<td>Cancer cecum (44)</td>
<td>MSI-H</td>
<td>MLH1-loss</td>
<td>Negative c.583A&gt;T p. Lys195X ID 204: right-sided CRC, 32 years, Amsterdam-negative family, IHC normal</td>
</tr>
<tr>
<td>F008</td>
<td>F</td>
<td>Rectosigmoidal cancer (40)</td>
<td>MSI-H</td>
<td>MLH1 and PMS2: loss</td>
<td>Negative c.121G&gt;C p. Asp41His Sequencing: negative ID 204: right-sided CRC, 32 years, Amsterdam-negative family, IHC normal</td>
</tr>
<tr>
<td>F010</td>
<td>F</td>
<td>Cancer sigmoidal colon (50)</td>
<td>MSI-H /2/5)</td>
<td>MLH1-loss Suspcion of PMS2-loss</td>
<td>Negative c.1521G&gt;C p. Leu507Phe MLPA: negative Sequencing: negative None</td>
</tr>
<tr>
<td>F011</td>
<td>M</td>
<td>Cancer col. asc. (50)</td>
<td>MSI-H</td>
<td>MLH1-loss</td>
<td>Negative 1) c.1814_1816delAAG p. Glu605del 2) c.2146G&gt;A p. Val716Met Negative ID 158-1: Cancer rectum, 42 years, MSS, IHC normal</td>
</tr>
<tr>
<td>F012</td>
<td>M</td>
<td>Cancer rectum (46)</td>
<td>MSS in metastatic tissue</td>
<td>MLH1/MSH2/ PMS2/MSH6: Normal</td>
<td>Negative c.2066A&gt;G p. Gln689Arg Negative ID 541: Endometrial cancer, 59 years, MSI, Amsterdam/Bethesda negative, not in 140 controls. ID 7: Small bowel cancer, 51 years, MSS. 2/932 CRC cases, 5/1066 controls, MSS, Amsterdam negative, IHC normal</td>
</tr>
</tbody>
</table>
incubated with Hoechst 33342 (1:5000 in phosphate-buffered saline) for nuclear staining, washed again and mounted. Subcellular localization was assessed on a Leica TCS SP5 confocal laser scanning microscope.

**Clinical data and in silico analyses**

Publications addressing the variants were identified using the Leiden Open Variation Database (LOVD; www LOVD nl MLH1). Moreover, entries in dbSNP were investigated and searches were performed for all alternative variant descriptions (e.g. MLH1 ’E605del,’ ‘Glu605del,’ ‘c.1814_1816delAAAG’) in PubMed and Google. Variant classifications were looked up at http://www.insight group org/variants/classifications/. Potential effects of the coding variants were assessed using SplicePort (14) on (splicing) and MAPP-MMR (15), Sift (16) and PolyPhen-2 (17) (on protein effects).

**Structural analyses**

Function–structure evaluations were performed with an updated model of human MutLx (MLH1–PMS2). The N-terminal domains were based on the structure of human PMS2 N-terminal domains (58) and an homology model of MLH1-N-terminal domains (11). The C-terminal domains were built by homology modeling using yeast MutLx (PDB codes: 4E4W and 4FMN) and human MLH1 (PDB code: 3RBN) structures as templates. 4E4W, the highest resolution yeast MutLx 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 MIF-box peptide (a fragment of NYT2) was taken from 4FMN structure and the zinc ions from 4E4W. The modeling templates were identified and selected using MODexplorer (19). The target–template alignment used for modeling PMS2 was evaluated and refined using MODalign (20). The final model was constructed after exporting the alignments from MODexplorer and MODalign and running Modeller (21) on a hybrid template containing all necessary fragments from the selected templates, the MIF-box and zinc ions. Figures were generated using PyMOL v1.4.1 (Schrödinger LLC).

**Results**

**Strategy for functional analysis of coding MLH1 variants**

Considering the plethora of functional analyses that can in principle be performed to assess potential defects of unclear coding MLH1 variants, we first addressed the question of the most rational and efficient analytical strategy for detecting pathogenic defects. For prioritizing functional parameters to be tested, we applied the following criteria: (i) functional parameters which most frequently cause pathogenicity should be prioritized in testing; (ii) functional parameters that provide a defined discrimination between pathogenic and neutral (clear association of the parameter with disease, low rate of ambiguous results) are preferable.

We therefore assessed functional defect frequencies that we and others have observed in coding MLH1 variants (Supplementary Table 1, available at Carcinogenesis Online). MMR activity defects (34–62%) (12,22–24) and strong expression defects (6–78%) (12,22,23,25,26) were frequent, with roughly three quarters (68–87%) of the investigated variants showing defects in either of these two parameters (12,22,23). In contrast, messenger RNA (mRNA) aberrations and subcellular localization defects have been reported more rarely (0–20% and 32%, respectively; Supplementary Table 1, available at Carcinogenesis Online) (23,27–30).

Loss of MMR activity is considered causative for Lynch syndrome (6). Concerning the association of expression defects with disease, we have recently provided discriminators to distinguish neutral from pathogenic expression defects (12). Therefore, both parameters provide suitable controls for achieving a clinically meaningful conclusion.

For these reasons, functional assessment of coding MLH1 variants reasonably starts with investigating protein expression and MMR activity (Figure 1), which both provide established association with disease and represent the most frequent functional defects in coding MLH1 variants. Defects of mRNA integrity or subcellular localization, which are observed much more rarely and which have poorly defined correlation to pathogenicity, are suggested to be tested only in variants that are neutral in MMR activity and expression analysis (Figure 1).

**Protein expression and MMR activity of variants identified in six cancer patients**

The analytical strategy (Figure 1) was applied for MMR gene variants identified in six colorectal cancer patients. Three patients carried missense variants in MLH1 (p.Asp41His, p.Leu507Phe, p.Gln689Arg). One patient carried two variants in the MLH1 gene: p.Val716Met + p.Glu605del (Table I). Two patients carried truncated non-sense variants (MLH1 p.Lys195X and PMS2 p.Arg315X). The PMS2 protein forms a functional heterodimer with MLH1, therefore PMS2 variants can compromise MMR activity as MLH1 variants do (31). The truncating variants were included as controls.

Expression was determined by transient transfection into HEK293T cells, which do not express the endogenous MLH1 gene due to a promoter hypermethylation (10). Expression analysis showed signals for all non-truncating MLH1 variants, and a protein fragment was detectable for one truncating variant (PMS2 p.Arg315X) (Figure 2A). PMS2 was properly coexpressed in all MLH1 variants except for the truncating MLH1 variant, since PMS2 is destabilized on the protein level in the absence of proper dimerization with the MLH1 C-terminus, which is absent in this variant (9).

In order to detect and confirm potential reductions of expression levels, we performed multiple independent transfection experiments and quantifications of all variants. This revealed that average expression levels were decreased in two MLH1 variants: p.Asp41His and p.Glu605del (Figure 2B).

We furthermore tested MMR activity by assessing the capability of the variant proteins to complement a MMR-deficient cell extract from 293T cells in vitro (Figure 2C, top panel). The truncating variants as well as the p.Asp41His missense variant were defective in MMR activity (Figure 2C). The other MLH1 variants showed repair activity like the wild-type protein.

The two variants with partial expression defects (MLH1 p.Asp41His and p.Glu605del) were subject to a more detailed analysis to classify whether these defects are pathogenic or not. Since the patient with the p.Glu605del variation also carried a secondary alteration (the clinically neutral, but protein-destabilizing polymorphism p.Val716Met, ref. 12), their allelic location was tested. Transcript analysis demonstrated that both are in cis (Figure 3A), therefore the compound variant (p.Glu605del + p.Val716Met) was used for functional testing.

The expression level of MLH1 p.Asp41His was similar to the non-pathogenic expression reference variant (Figure 3B). In contrast, expression of the compound variant (p.Glu605del + p.Val716Met) was significantly (p < 0.05) lower than the reference variant for pathogenic instability p.Ala681Thr (Figure 3B), demonstrating that the protein stability of this variant is too low for functional MMR in the cell (12). This low stability also explains the loss of expression in the tumor tissue (Table I) as we have observed before to occur in such destabilized variants (12). Interestingly, the destabilization was caused additively by both variants, which had a weaker destabilizing effect.
individually (Figure 3B). However, the compound variant still retained complete MMR activity (Supplementary Figure S1, available at Carcinogenesis Online), similar as the individual variants p.Glu605del (Figure 2C and Supplementary Figure S1, available at Carcinogenesis Online) and p.Val716Met (12).

Taken together, the functional analysis demonstrated that p.Asp41His caused a complete loss of function. The expression level of the compound p.Glu605del + p.Val716Met variant was below the reference variant and therefore can be expected to be pathogenic due to protein destabilization (12). Although general functional result interpretations would not have been able to detect this pathogenic destabilization, the applied analytical strategy allowed this. In contrast, expression level and repair activity of p.Leu507Phe and p.Gln689Arg were identical to the wild-type.

**Analysis of subcellular localization and mRNA formation**

To test if the p.Leu507Phe and p.Gln689Arg variants confer other defects which might interfere with protein function and potentially human health, we assessed their subcellular distribution and investigated their effect on mRNA formation.

Subcellular distribution was tested by transfecting optimized constructs (32) of the variants in HEK293T cells in parallel with suitable positive and negative controls (13). Subcellular distribution of the variants was identical to the wild-type, while the negative control, which contains a mutation in the nuclear localization sequence, was retained in the cytoplasm (Supplementary Figures S2 and S3, available at Carcinogenesis Online).

Neither of the DNA changes coding for the p.Leu507Phe and p.Gln689Arg variants is in immediate proximity to splice sites, and none has a significant effect on splice site qualities as predicted by Spliceport (14), an algorithm which was most consistent with experimental results in a previous analysis (33) (Supplementary Table 2, available at Carcinogenesis Online).

Transcript quantity and splicing at splice sites neighboring the p.Leu507Phe variant was identical to wild-type in the lymphocyte mRNA of the carrier (data not shown). p.Gln689Arg has previously been shown not to affect splicing (30). p.Asp41His and the compound variant p.Glu605del + p.Val716Met had no effect on transcript quantity or splicing either (Table II).

Taken together, there is no evidence for a functional impairment of the p.Leu507Phe and p.Gln689Arg variants in comparison with wild-type concerning mRNA production, protein stability, DNA repair activity and subcellular localization, while p.Asp41His was defective in MMR activity and the combined variant was defective in stability.

**Structure, in silico analyses and clinical data**

We used a revised structural model of the human MutLα (MLH1-PMS2) heterodimer to assess molecular roles of the affected residues (Supplementary Figure S4, available at Carcinogenesis Online). Additionally, we assessed the MAP-MMR and Polyphen scores, which aim at predicting protein effects of substitutions and have been found to correlate best with clinical effects (15–17,29) (Table II).

p.Asp41 is a highly conserved residue which tethers Gly98 of the ‘lid’ of the MLH1 adenosine triphosphatase with a side-chain hydrogen bond (Supplementary Figure S4, Box 1, available at Carcinogenesis Online). Gly98 and the ‘lid’ are important for the adenosine triphosphatase cycle (34), which is essential for MMR activity (35). Substitution of the negatively charged Asp41 to the positive histidine destroys this interaction, explaining the loss of repair activity. In silico analyses are consistent with a damaging effect (Table II). The clinical picture of the carrier patient is consistent with Lynch syndrome (Table I), similar as other, previous clinical observations (36). Taken together, these observations strongly support that this variant is pathogenic.

Both p.Glu605del and the p.Val716Met variants are located in a domain of MLH1 which we have previously found to be
Fig. 2. Protein expression level and repair activity of the variants. (A) Wild-type MLH1-PMS2 and the indicated variants were transfected into HEK293T cells. Cells were harvested after 48h and expression was analyzed by SDS–PAGE and immunoblotting using specific antibodies against MLH1 and PMS2. β-Actin was detected as loading control. n.s., non-specific signals. The signal of the putative PMS2 fragment (PMS2 p.Arg315X, calculated weight 35 kDa) is indicated. (B) Expression signals of MLH1 were quantified from diverse (>10) independent experiments in which MLH1 wild-type protein and variants had been analyzed in parallel. Expression was scored in percentage of the wild-type. Bars indicate the standard error of the mean. (C) MMR activity of the variants was scored in comparison with MLH1-PMS2 wild-type. Top panel: representative agarose gel of a single analysis of repair activity of wild-type MLH1 and the variants. Signals comprise the linearized plasmid (lin.) which is digested into two fragments (frag.) when MLH1 is proficient in DNA MMR. Bottom panel: average repair activities of MLH1 wild-type and variants from at least six independent experiments are shown in percentage of wild-type activity. Bars indicate standard deviations. SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis.
p.Leu507Phe is a conservative substitution that affects a weakly conserved residue at the end of an α-helix (Supplementary Figure S4, Box 2, available at Carcinogenesis Online). In silico results do not suggest a damaging effect on protein function (Table II). The clinical picture was in principle compatible with Lynch syndrome (Table I), but no other Lynch syndrome cancers have been diagnosed in the family 8 years after diagnosis of the patients’ tumor (Supplementary Figure S5, available at Carcinogenesis Online). The variant did not show defects in stability, DNA repair activity, mRNA formation or subcellular localization, therefore the p.Leu507Phe variant was most likely not causally related to the cancer disease. The incomplete MSI phenotype and reduction of MLH1 and FMS2 in the tumor probably occurred somatically by another mechanism, most likely by promoter methylation, which is a frequent somatic cause of MMR deficiency in tumors.

The weakly conserved p.Glu689 residue is located in a loop at the protein surface. The conservative substitution is unlikely to disturb the protein structure (Supplementary Figure S4, Box 2, available at Carcinogenesis Online). The clinical data were not suggestive of Lynch syndrome, and microsatellite stable tumors have been reported in carriers of this variant before (22,37,38) (Table I). The protein was as stable as the wild-type in our analysis, which is consistent with the presence of the protein in the tumor in immunohistochemistry (Table I). In a case–control analysis, the variant was found more frequently in controls than in cases (39). Taken together, the functional analysis and the additional data strongly suggest that this variant is not pathogenic.

Discussion

Functional testing is often the only way to determine pathogenicity of uncertain genetic variants, especially in so-called ‘private’ variants for which clinical information is scarce, but also for combinations of variants like the combined variant investigated in this study. For these, efficient strategies for functional analyses that allow clinically relevant conclusions are required. Schemes for performing functional analyses or for interpretation of functional investigations have recently been suggested (6,40,41), but they generally apply to various types of genetic alterations in any MMR gene. Moreover, these recommendations focus on assessment of MMR function although pathogenic variants with proficient function are not infrequent, especially in variants C-terminally located in the protein (12,23,42). Therefore, we suggest an analytical strategy which enables efficient identification of functional defects with the best currently available association with pathogenicity for coding variants in MLH1.

DNA MMR capacity and protein expression were tested first, since both represent the most frequent causes of functional defects and both have well-defined correlations with disease risk. For expression, we used specific reference variants for establishing the clinical outcome of expression defects (12). General evaluation limits for expression have been suggested (<25% for pathogenic and >75% for non-pathogenic, ref. 6). However, the expression reference variants provided in the current strategy provide a much smaller diagnostic gap (15%, ref. 12 versus 50%, ref. 6) and have the additional advantage that they can serve for internal, direct comparisons in the applied analytical test system. The smaller diagnostic gap enables improved classification: for example, the pathogenic potential of the compound MLH1 variant p.Glu605del + p.Val716Met is not detectable using general recommendations (6), whereas it was obvious in our analysis using the reference variants (Figure 3B).

Fig. 3. (A) Determination of allelic location of the compound alterations in patient F011. In order to test the allelic location of the two genetic MLH1 variants present in patient F011, RNA was extracted from peripheral blood mononuclear cells of the patient and reverse-transcribed as detailed in Materials and methods. Four analytical PCRs were performed using primers which carried specific sequences for the MLH1 wild-type or the variant sequences (forward: either wt-wild-type or del-p.Glu605del; reverse: either wt-wild-type or Met-p.Val716Met). cDNA from patient F011 and from one control subject with wild-type MLH1 sequence were analyzed. A PCR signal is detectable when the indicated combination of sequences is present on the template cDNA. The two signals in F011 indicate that the patient is carrying a wild-type allele of MLH1 and one that carries both alterations (coding for p.Glu605del and for p.Val716Met). cDNA from patient F011 and from one control subject with wild-type MLH1 sequence were analyzed. A PCR signal is detectable when the indicated combination of sequences is present on the template cDNA. The two signals in F011 indicate that the patient is carrying a wild-type allele of MLH1 and one that carries both alterations (coding for p.Glu605del and for p.Val716Met). (B) Comparison of stability-impaired MLH1 variants with clinical reference variants. Expression levels of all stability-impaired variants, including the mono-allelic compound variant p.Val716Met + p.Glu605del, were assessed in direct comparisons to the neutral and pathogenic reference variants (p.Val716Met and p.Ala681Thr, respectively) in HEK293T cells as described in Figure 2B. Average expression values were calculated from at least 10 independent experiments. Bars indicate the standard error of the mean.

Susceptible to protein destabilization without affecting catalytic activity (Supplementary Figure S4, Box 2, available at Carcinogenesis Online) (12). The patient carrying this compound variant showed microsatellite instability (MSI) and immunohistochemical loss of MLH1 in the tumor tissue as well as the wild-type sequence at the BRAF V600 codon (Table I), which is consistent with Lynch syndrome. In summary, the data suggest that the compound variant is pathogenic due to a reduction of protein stability (12). Interestingly, this pathogenic destabilization resulted from additive effects conferred by both variants, one of which (p.Val716Met) is non-pathogenic (6,12). The low stability is in agreement with the observed loss of MLH1 expression in the tumor tissue (Table I). Indeed, practically all variants with pathogenic stability defects also showed loss of expression in the tumor tissue in our previous analysis (12).
<table>
<thead>
<tr>
<th>Patient number</th>
<th>Variant</th>
<th>Functional data</th>
<th>Repair</th>
<th>mRNA</th>
<th>Structural analysis (see Discussion for details)</th>
<th>Conservation</th>
<th>MAPPP2 prior probability of pathogenicity</th>
<th>InSiGHT classification</th>
<th>Functional classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>F001</td>
<td>p.Arg315X</td>
<td>No full-length protein detectable</td>
<td>Loss of function</td>
<td>n.t.</td>
<td>Early truncation</td>
<td>n.a.</td>
<td>n.a.</td>
<td>Pathogenic (class 5)</td>
<td>Deficient (truncation – loss of function)</td>
</tr>
<tr>
<td>F003</td>
<td>p.Lys195X</td>
<td>No full-length protein detectable</td>
<td>Loss of function</td>
<td>n.t.</td>
<td>Early truncation</td>
<td>n.a.</td>
<td>n.a.</td>
<td>Not available</td>
<td>Deficient (truncation – loss of function)</td>
</tr>
<tr>
<td>F008</td>
<td>p.Asp41His</td>
<td>Reduced, but above reference limit</td>
<td>Loss of function</td>
<td>Expression and splicing like wild-type</td>
<td>Molecular interaction in ATP catalytic pocket affected</td>
<td>9</td>
<td>0.8695</td>
<td>Likely pathogenic (class 4)</td>
<td>Deficient (loss of function)</td>
</tr>
<tr>
<td>F010</td>
<td>p.Leu507Phe</td>
<td>Like wild-type</td>
<td>Like wild-type</td>
<td>Expression and splicing like wild-type</td>
<td>No reason for damaging effect obvious</td>
<td>4</td>
<td>0.1203</td>
<td>Not available</td>
<td>Neutral</td>
</tr>
<tr>
<td>F011</td>
<td>p.Glu605del + p.Val716Met</td>
<td>Below reference limit</td>
<td>Like wild-type</td>
<td>Expression and splicing like wild-type</td>
<td>Located in region relevant for protein stability</td>
<td>1 + 6</td>
<td>Not available +</td>
<td>Not available + not pathogenic (class I)</td>
<td>Deficient (stability below reference limit)</td>
</tr>
<tr>
<td>F012</td>
<td>p.Gln689Arg</td>
<td>Like wild-type</td>
<td>Like wild-type</td>
<td>No effect on splicing (B0)</td>
<td>No reason for damaging effect obvious</td>
<td>3</td>
<td>0.0051</td>
<td>Not pathogenic (class 1)</td>
<td>Neutral</td>
</tr>
</tbody>
</table>

ATP, adenosine triphosphate; n.a., not applicable; n.t., not tested.

*Extracted from the in silico prior probabilities [http://hci-lovd.hci.utah.edu/variants.php?action=search_unique].
Interestingly, the pathogenic effect of this compound variant was additively caused by both variants, one of which (the polymorphism p.Val716Met) is non-pathogenic when occurring alone despite its destabilizing effect (6,12). Additive effects resulting in repair deficiency have previously been suggested for an MSH2 variant pair (43). To our knowledge, this is the first time that an MLH1 variant pair has been shown to confer pathogenicity by additive effects.

Interpretation of MMR activity measurements relies on direct comparisons with repair-proficient or repair-deficient controls (6). We and others have shown before that MLH1 alleles with intermediate repair activity are rare (12,23), at least when measured with in vitro systems which generally have a surplus of the variant MLH1-PMS2 dimer. Although these systems will reliably reflect a full loss of catalytic activity, they may fail to indicate weakly compromised activity. As a consequence, false negatives (undetected moderate defects of activity) are more likely than false positives (falsely proficient-tested defective variants). However, considering the comparatively sensitive dependence of MMR activity from protein concentration (12) it is not only methodologically demanding to attempt detecting weakly compromised variants, but also associated with an increased risk of getting significantly more false-positive results. Studies are required to test the actual relevance of intermediate alleles in repair activity and potentially establish suitable test systems with reference limits.

Two variants (p.Leu507Phe and p.Gln689Arg) were neutral in expression, catalytic activity, mRNA formation and subcellular protein localization. They therefore did not show any functional defects (with the caveat that weakly compromised MMR activity may have been missed, see above). However, since clinical and in silico data independently corroborated neutrality of these variants, and neutrality has been suggested for p.Gln689Arg based on clinical information before, such a missed functional defect is unlikely.

Taken together, we have outlined an analytical testing strategy for coding MLH1 variants with the purpose of efficiently identifying functional defects that cause Lynch syndrome. Using this analytical strategy exemplarily with four unclear MLH1 variants, two variants with functional defects that are expected to cause pathogenicity in humans were identified: one of these because of a MMR defect (p.Asp41His), the other because of decreased stability (p.Glu605del + p.Val716Met). Interestingly, the pathogenic effect of this compound variant was additively caused by both variants and would not have been detected using the general functional assay recommendations. The suggested analytical strategy therefore likely currently represents the most sensitive and economic strategy to identify clinically relevant effects in coding MLH1 variants.

**Supplementary material**

Supplementary Tables 1 and 2 and Figures S1–S5 can be found at [http://carcin.oxfordjournals.org](http://carcin.oxfordjournals.org)

**Funding**

Deutsche Krebshilfe (110799 to G.P.).

**Acknowledgements**

We are grateful to J.Kosinski (Structural and Computational Biology, EMBL, Heidelberg, Germany) for revising the structural model of the MLH1-PMS2 dimer, and to S.Genssler, Georg-Speyer-Haus, Institute for Tumor Biology and Experimental Therapy, Frankfurt, Germany, for aid using the confocal laser microscope.

**Conflict of Interest Statement:** No conflicts of interests are declared.

**References**