Integrated analysis of unclassified variants in mismatch repair genes

Chiara Pastrello, PhD[1], Elisa Pin, MSc[1], Fabio Marroni, PhD[2], Chiara Bedin, MSc[3],
Mara Fornasarig, MD[4], Maria Grazia Tibiletti, PhD[5], Cristina Oliani, MD[6],
Maurizio Ponz de Leon, MD[7], Emanuele Damiano Urso, MD[3], Lara Della Puppa, BSc[1],
Marco Agostini, PhD[3], and Alessandra Viel, PhD[1]

Purpose: Lynch syndrome is a genetic disease that predisposes to colorectal tumors, caused by mutation in mismatch repair genes. The use of genetic tests to identify mutation carriers does not always give perfectly clear results, as happens when an unclassified variant is found. This study aimed to define the pathogenic role of 35 variants present in *MSH2*, *MLH1*, *MSH6*, and *PMS2* genes identified in our 15-year case study. **Methods:** We collected clinical and molecular data of all carriers, and then we analyzed the variants pathogenic role with web tools and molecular analyses. Using a Bayesian approach, we derived a posterior probability of pathogenicity and classified each variant according to a standardized five-class system. **Results:** The *MSH2* p.Pro349Arg, p.Met688Arg, the *MLH1* p.Gly67Arg, p.Thr82Ala, p.Lys618Ala, the *MSH6* p.Ala1236Pro, and the *PMS2* p.Arg20Gln were classified as pathogenic, and the *MSH2* p.Cys697Arg and the *PMS2* p.Ser46Ile were classified as likely pathogenic. Seven variants were likely nonpathogenic, 3 were nonpathogenic, and 16 remained uncertain. **Conclusion:** Quantitative assessment of several parameters and their integration in a multifactorial likelihood model is the method of choice for classifying the variants. As such classifications can be associated with surveillance and testing recommendations, the results and the method developed in our study can be useful for helping laboratory geneticists in evaluation of genetic tests and clinicians in the management of carriers. *Genet Med* 2011:13(2):115–124.

Key Words: *unclassified variant, mismatch repair, Lynch syndrome, colorectal cancer, multifactorial likelihood.*

Lynch syndrome, also known as Hereditary Non-Polyposis Colorectal Cancer, is an autosomal dominant syndrome caused by mutations in one of the mismatch repair (MMR) genes, namely *MLH1*, *MSH2*, *MSH6*, and *PMS2*.[1] These genes are involved in the repair of DNA polymerase errors, especially those involving one or few base pairs. The loss of MMR function causes the accumulation of mutations, particularly in tandem repeat sequences leading to microsatellite instability (MSI), a typical marker of Hereditary Non-Polyposis Colorectal Cancer tumors.

Genetic tests are offered to probands that fulfill Amsterdam clinical criteria and/or that develop MSI tumors at young ages[1]; the finding of a clearly pathogenic mutation in a family allows a better surveillance of carriers. However, mutation screening with advanced DNA sequence technologies leads to the detection of an increasing number of missense, silent, and intronic variants. These changes often do not produce truncated proteins upon translation, like most disease-causing pathogenic mutations, and are so-called unclassified variants (UVs). Their relation to disease development is often not clear, but it is very important for the risk determination of the carrier to offer an adequate follow-up. This is the reason why there has been an intense debate for many years, and especially in the past 2 years, on how to assess mutation pathogenicity in cancer susceptibility genes.[2,3]

This work is in line with most recent publications and aims to provide some helpful data for the purpose of UVs classification.[4–12] It simply represents an initial attempt to evaluate UVs of MMR genes based on a relatively elementary quantitative assessment of several parameters, integrated in a multifactorial likelihood model.[13]

MATERIALS AND METHODS

Patients and controls

From 1994 to 2008, we performed genetic tests of MMR genes in 306 unrelated probands that fulfilled Amsterdam and/or Bethesda criteria (Table 1) who were recruited mainly at the C.R.O. National Cancer Institute in Aviano (recorded as A-AV), and at collaborating Centers in Padova (A-PD), Varese (A-VA), Modena (A-MD), and Montecchio (A-VR). Immunohistochemistry (IHC) and MSI were not used as selection criteria for enrollment. We selected all patients with one or more UV and, when available, tested the relatives of the proband. For every person involved in the UV analysis, we collected, under informed consent, blood samples from which we obtained DNA. Where possible, we also obtained RNA from lymphoblastoid cell lines and DNA from tumor tissues. As controls, we collected DNA from 90 unrelated subjects with negative colonoscopy (clean colon) at ages ranging from 23 to 86 years (mean 63.42 ± 13.36).

Mutation, MSI, and IHC

The DNA samples of probands were sequenced for the entire open reading frame and flanking intronic sequences of *MSH2*, *MLH1*, *MSH6*, and *PMS2* genes. Tumor DNAs were tested only

Table 1 Amsterdam I and II criteria and Bethesda revised guidelines[1]

Name	Criteria
Amsterdam I	There should be at least three relatives (one of which must be a first degree relative of the other two) with histologically verified colorectal cancer (CRC) in at least two successive generations. At least one CRC should be diagnosed before the age of 50 years and FAP should be excluded
Amsterdam II	The same as Amsterdam I, but the three relatives can have CRC, cancer of the endometrium, small bowel, ureter, or renal pelvis
Bethesda revised	Individual with CRC diagnosed at age <50 years, or with presence of synchronous, metachronous colorectal, or other HNPCC-associated tumors, regardless of age, or with CRC with the MSI-H histology (presence of tumor infiltrating lymphocytes, Crohn's-like lymphocytic reactions, mucinous/signet-ring differentiation, or medullary growth pattern) diagnosed at age <60 years, or with CRC diagnosed in one or more first-degree relatives with an HNPCC-related tumor, with one of the cancers being diagnosed at age <50 years, or CRC diagnosed in two or more first- or second-degree relatives with HNPCC-related tumors, regardless of age

FAP, familial adenomatous polyposis.

for the mutation detected in blood DNA and evaluated for loss of heterozygosity (LOH) by DNA sequencing. MSI and IHC analyses were carried out and evaluated on tumor tissues by conventional methods.[14,15]

cDNA splicing and primer extension analyses

cDNA obtained from mRNA extracted from lymphoblastoid cell lines was polymerase chain reaction (PCR)-amplified and analyzed by agarose gel electrophoresis to evaluate a potential involvement of the UV nucleotide in the correct splicing signaling. Primers used can be requested from authors.

A primer extension method was used to analyze cDNA allelic expression. Single nucleotide primer extension was carried out in a final volume of 10 μl containing 0.25 pmol of purified PCR product, 0.2 μM of reverse primer located downstream of the mutation site, and 5 μl of SNaPshot Multiplex Ready Reaction Mix (Applied Biosystems, Foster City, CA). The reaction was performed as recommended by the manufacturer in a thermal cycler (10 cycles), run on the ABI PRISM 3100 Genetic Analyzer (Applied Biosystems) and evaluated with GeneScan software. The electropherogram was used to calculate the ratio between mutated and wild-type peak areas in both gDNA and cDNA.

Denaturing high-performance liquid chromatography analysis

The same PCR reactions set up for mutational screening were prepared for healthy controls, with a reference sample carrying the UV for every amplification. If a variant was present, denaturation followed by slow renaturation leads to the formation of heteroduplexes that are detected by a denaturing high-performance liquid chromatography analysis instrument (Transgenomics Inc., Omaha, NE). To confirm UV presence, every sample with a different chromatogram was sequenced.

In silico analysis

For the in silico evaluation of the selected UVs, the bioinformatics tools reported in Table 2 were used. We used the default threshold and collected all the results supplied by the tools exactly as they emerged. The A-GVGD value was obtained through the Alamut software (Interactive Biosoftware, Rouen, France) that uses International Agency for Research on Cancer (IARC)-style curated alignments. The A-GVGD Web site alignments were not used because the MSH6 and PMS2 protein alignments have not yet been released by IARC. Data on

Table 2 Web sites used for the in silico analysis

Type of resource	Web site
Analysis of exon splicing enhancer (ESE) sequence	ESEfinder (http://rulai.cshl.edu/tools/ESE/)
	RescueESE (http://genes.mit.edu/burgelab/rescue-ese/)
	PESX (http://cubweb.biology.columbia.edu/pesx/)
Analysis of canonical splice sites	NNSplice (http://www.fruitfly.org/seq_tools/splice.html)
	Splice site finder (no longer available) (http://violin.genet.sickkids.on.ca/~ali/splicesitefinder.html)
Analysis of intron/exon structure	Genescan (http://genes.mit.edu/GENSCAN.html)
Analysis of the effect of the substitution on the protein	PolyPhen (http://genetics.bwh.harvard.edu/pph/)
	SIFT (http://blocks.fhcrc.org/sift/SIFT.html)
	Pmut (http://mmb2.pcb.ub.es:8080/PMut/)
	A—GVGD (http://agvgd.iarc.fr/agvgd_input.php)

in vitro functional studies were collected from the available literature.[16–23]

Statistical evaluation

UVs were classified using likelihood of pathogenicity for both quantitative and qualitative data. For individual variants, we used a Bayesian approach to incorporate direct and indirect evidence of pathogenicity in a single model,[4,13] starting from a prior probability of 0.5[24] and assuming statistical independence of the sources of information. Posterior probability of pathogenicity was obtained from the likelihood ratio and categorized in five classes.[6] To test our method, a well-known common polymorphism (*MLH1* p.Ile219Val) was introduced as the negative control. For the positive control, there are no well-known pathogenic missense mutations, and hence, we used a truncating one (*MSH2* p.Arg406Ter).

Prior odds ratios (OR) of pathogenicity for each variant were calculated from published work applying a continuity correction.[25] For in silico analyses, OR were obtained from published results based on different tools: A-GVGD,[12] PolyPhen, Pmut, SIFT,[26] and splicing tools.[27] As in silico predictions are not independent, we calculated one average OR for the three exon splicing enhancer (ESE)-related software tools, one average OR for the two splicing site-related software, and one for the protein-related tools.

For MSI and IHC, we estimated OR from data published by Engel et al.,[28] who evaluated these molecular features as predictive of mutations in 1119 unrelated patients. Because of the correlation of absence of protein expression and presence of MSI, a single OR for these two combined data was estimated. When the UV co-occurs with a clearly pathologic mutation, it is impossible to assess whether MSI phenotype is due to the UV, to the pathologic mutation, or both, and we accordingly assigned an OR of 1. This neutral score was not applied when a typical compound heterozygous phenotype was evident.

OR for pathogenicity based on familiarity was derived from a set of patients previously screened for pathogenic mutations: the event was the presence of a clear-cut pathogenic mutation and groups were based on the fulfillment of Amsterdam criteria.

For OR related to co-occurrence, we used the formula proposed by Easton et al.[29] Co-occurrence was considered exclusively as the presence of a mutation in *trans* of the same gene, with the exception of cases with a recognizable biallelic phenotype. For families where the genetic test could be extended to at least one relative, OR of cosegregation of disease and variant were evaluated according to Thompson et al.[30] For this analysis, we assumed the age-specific risks and penetrance estimated by Marroni et al.[31] for the Italian population.

In instances in which calculating OR was not possible, we arbitrarily assigned an OR of 2 (see Table, Supplemental Digital Content 1, http://links.lww.com/GIM/A127); the value is conservative and was chosen so that data for which a quantitative assessment is difficult would not heavily bias quantifiable evidence. This estimate was tentatively applied with the aim of also incorporating "qualitative data," as LOH, allelic imbalance, absence in healthy controls, and in vitro functional effect, into the multifactorial model.

RESULTS

We analyzed 35 UVs that are listed in Table 3: 12 *MSH2*, 15 *MLH1*, 4 *MSH6*, and 4 *PMS2*. Twenty UVs were missense, 7 were silent, and 8 were intronic mutations. To test our system, we also included in the analysis the p.Arg406Ter nonsense mutation of the *MSH2* gene and the common p.Ile219Val poly-

morphism of the *MLH1* gene (Table 3). Bibliographic references for some UVs can be found online in the MMR variants database (http://www.med.mun.ca/MMRvariants/).

Globally, we found UVs in a total of 40 probands, corresponding to about 13% of the screened patients. In most cases, every patient carried only one UV, whereas five patients carried two or three UVs each.

Copresence of a clear-cut pathogenic mutation was ascertained for nine UVs (Table 3). Data concerning disease-UV segregation, presence of Amsterdam criteria in the family, MSI, and IHC are also reported in Table 3.

The denaturing high-performance liquid chromatography analysis of at least 50 individuals with clean colon revealed the presence of only 4 UVs, all within the *MSH2* gene. Three of them (p.Gly322Asp, p.Leu556Leu, and p.Lys579Lys) were present in only one person and the last one (c.1077-10T>C) was present in four individuals (Table 3). In addition, our previous studies of the *PMS2* gene revealed the presence of p.Arg20Gln in 2 of 70 control chromosomes, whereas the p.Ser46Ile variant was not detected in any control.[32,33] The LOH analysis was only possible for 21 samples, and we obtained results from 16 tumors. The *MSH2* p.Arg406Ter truncating mutation and 8 UVs (*MSH2*: p.Ala328Ala, p.Met688Arg, and p.Cys697Arg; *MLH1*: p.Gly67Arg, p.Arg265His, and p.Lys618Ala; *MSH6*: p.Ala1236Pro; and *PMS2*: p.Arg20Gln) showed loss of the wild-type allele, *MSH2* p.Lys113Lys revealed loss of the variant allele, and the remaining five UVs, as well as the p.Ile219Val polymorphism, did not show LOH (Table 3).

None of the UVs analyzed showed cDNA alteration of the physiological splicing (data not shown). The allelic expression analysis revealed a high imbalance level in the *MLH1* p.Arg265His sample, but it was probably due to the in *cis* presence of the *MLH1* c.1011delC frameshift mutation that is predicted to cause mRNA decay. The other UVs did not show a comparably high imbalance level, but *MSH2* p.Lys113Lys, p.Met688Arg, p.Cys697Arg, and *MLH1* p.Val326Ala showed imbalance ratios higher than 1.2 or lower than 0.8. Data are shown in Figure 1 and reported in Table 3.

Every web-based prediction software used for the analysis rendered a value that is indicative of the predicted effect of the nucleotidic/aminoacidic substitution on the splicing (Table 4) or on the protein (Table 5).

We categorized variants according to the IARC classification based on the posterior probability of causality (Table, Supplemental Digital Content 1, http://links.lww.com/GIM/A127). Seven UVs were classified as definitely pathogenic and two as likely pathogenic, whereas three and seven UVs resulted not pathogenic or likely not, respectively. The additional 16 UVs remained of uncertain significance (Table 6). The polymorphism used to test our method was included in the class with the lowest probability, while the truncating mutation, even if lacking the in silico protein prediction, was attributed to the highest class.

DISCUSSION

We evaluated all the UVs collected in our 15 years of genetic testing for Lynch syndrome. This integrated analysis aimed to develop a model for assigning each UV to a pathogenic or a nonpathogenic category and to elucidate their role in Lynch syndrome development. All evaluations were made under the arbitrary assumption that the analyzed variants could have the same impact on molecular and clinical phenotype as the proven pathogenic MMR gene mutations resulting in protein truncation or instability.

Table 3 Clinical and molecular data

Unclassified variant									Healthy control alleles
Protein	cDNA	ID family	Family type	Co-occurrence	MSI/IHC	LOH	Segregation[a]	Allelic imbalance	
MSH2									
p.Lys113Lys	c.339G>A	A-PD18	Like	*MLH1* c.1684insT	H/–	—	—	—	0/180
		A-PD31	Like	Del 5' *MSH2*	H/–	—	1, 0, 0, 0	—	
		A-AV114	AI	Del 5' *MSH2*[b]	H/MSH2	Mut	0, 0, 0, 2	Red	
p.Gly322Asp	c.965G>A	A-AV1	AI	—	S/–	No	0, 0, 1, 1	Bal	1/100
		A-VA13	AI	—	S/Normal	—	—	—	
p.Ala328Ala	c.984C>T	A-PD12	AII	*MSH2* c.942+2T>A	H/–	Wt	—	—	0/100
		A-PD28	Like	—	S/–	—	—	—	
p.Pro349Arg	c.1046C>G	A-PD16	AI	—	–/–	—	1, 0, 0, 3	—	0/100
		A-PD27	AI	—	H/MSH2	—	—	—	
p.Arg406Ter	c.1216C>T	A-AV68	AI	—	H/–	Wt	3, 0, 0, 0	—	0/180
		A-VA22	Like	—	H/MSH2	—	1, 1, 0, 0	—	
		A-VA36	AI	—	H/MSH2	—	1, 4, 0, 3	—	
p.Leu556Leu	c.1666T>C	A-MD17	Like	*MSH6* c.2984delA	H/–	—	—	—	1/180
p.Lys579Lys	c.1737A>G	A-PD39	AII	—	–/–	—	0, 1, 1, 1	—	1/180
p.Met688Arg	c.2063T>G	A-AV107	AII	—	H/–	Wt	—	Incr	0/180
p.Cys697Arg	c.2089T>C	A-AV87	Multiple	—	H/MSH2	Wt	0, 0, 0, 2	Incr	0/180
p.Leu814Leu	c.2442T>G	A-MD19	AI	*MSH2* c.942+3A>C[c]	H/–	—	4, 3, 1, 13	Bal	0/180
p.Ala834Thr	c.2500G>A	A-AV31	EO	—	S/–	No	0, 1, 0, 4	Bal	0/116
	c.1077-10T>C	A-VA3	Like	—	S/Normal	—	1, 0, 0, 0	—	4/180
		A-AV71	EO	—	H/–	—	—	—	
		A-AV102	Multiple	—	S/–	No	—	—	
	c.1387-8G>T	A-PD39	AII	—	–/–	—	—	—	0/130
MLH1									
p.Gly67Arg	c.199G>A	A-AV133	AI	—	H/–	Wt	0, 0, 0, 1	Bal	0/100
p.Thr82Ala	c.244A>G	A-VR3	AI	—	H/MLH1	—	1, 0, 0, 0	Bal	0/100
p.Ile219Val	c.655A>G	A-AV26	AI	—	S/–	—	—	—	34.5/100[d]
		A-AV27	AI	—	S/–	—	—	—	
		A-AV28	AI	*MSH2* p.Cys778Ter	H/–	No	1, 5, 2, 0	—	
p.Arg265His	c.794G>A	A-PD1	AI	*MLH1* c.1011delC[c]	H/–	Wt	1, 3, 1, 0	Red	0/180
		A-PD56	EO	*MLH1* c.1011delC[c]	H/–	—	—	—	
p.Val326Ala	c.977T>C	A-AV18	Like	—	S/–	—	1, 0, 2, 1	Red	0/180
		A-AV38	EO	—	S/–	—	—	—	
p.Pro355Ala	c.1063C>G	A-VA2	Like	—	H/MLH1	—	—	—	0/100
p.Ser406Asn	c.1217G>A	A-AV30	EO	—	–/–	—	1, 0, 1, 1	Bal	0/180
p.Lys488Thr	c.1463A>C	A-VA23	Like	—	H/MLH1	—	—	Bal	0/100
p.Lys618Ala	c.1852-3AA>GC	A-VR3	AI	—	H/MLH1	Wt	1, 0, 0, 0	Bal	0/180
p.Leu653Leu	c.1959G>T	A-VA18	EO	—	H/MSH2-6	—	—	Bal	0/106

(*Continued*)

Table 3 *Continued*

Unclassified variant									
Protein	cDNA	ID family	Family type	Co-occurrence	MSI/IHC	LOH	Segregation[a]	Allelic imbalance	Healthy control alleles
p.Val716Met	c.2146G>A	A-AV23	AII	—	S/–	—	0, 1, 0, 1	Bal	0/160
		A-AV44	EO	—	S/–	No	0, 1, 0, 1	—	
	c.116+8G>A	A-PD45	Like	—	–/–	—	—	—	0/180
	c.307–29C>A	A-MD11	Like	—	–/–	—	0, 1, 0, 0	—	0/100
	c.790+10A>G	A-AV41	EO	—	S/–	No	—	—	0/100
	c.884+4A>G	A-PD52	Like	—	H/–	—	—	—	0/180
	c.1039-8T>A	A-PD17	EO	—	S/–	—	—	—	0/100
		A-AV94	EO	—	–/–	—	—	—	
MSH6									
	c.628-56C>T	A-AV13	EO	—	H/–	—	—	—	0/100
p.Glu983Gln	c.2949G>C	A-PD31	Like	Del 5' *MSH2*	H/–	—	1, 0, 0, 0	—	0/180
p.Pro1082Pro	c.3246G>T	A-PD31	Like	Del 5' *MSH2*	H/–	—	1, 0, 0, 0	—	0/170
p.Ala1236Pro	c.3706G>C	A-VA30	AII	—	H/MSH6	Wt	0, 0, 1, 2	—	0/180
PMS2									
p.Arg20Gln	c.59G>A	A-AV4	AI	*MSH6* c.1960-1ins4	H/–	Wt	—	—	2/70
		A-PD2	AI	—	–/–	—	0, 0, 2, 0	—	
		A-AV13	EO	—	H/–	—	—	—	
p.Ser46Ile	c.137G>T	A-VA17	EO	*PMS2* p.Gln643Ter[e]	H/PMS2	—	—	—	0/118
p.Thr511Ala	c.1531A>G	A-AV13	EO	—	H/–	—	—	—	—
p.Thr597Ser	c.1789A>T	A-PD2	AI	—	S/–	—	1, 0, 1, 0	—	—

AI or AII, Amsterdam criteria I or II; EO, early onset; Like, Amsterdam missing one criteria; Multiple, multiple tumors; S, microsatellite stability; H, high microsatellite instability; Wt, loss of wild-type allele; mut, loss of mutated allele; incr, increased; red, reduced; bal, balanced; —, missing data because not tested or not evaluable.
[a]Number of affected carrier, healthy carrier, affected noncarrier, healthy noncarrier among relatives.
[b]In *trans* with pathogenic mutation.
[c]In *cis* with pathogenic mutation.
[d]Hapmap data.
[e]Biallelic MMR mutation phenotype.

Among the UVs, the *MSH2* p.Pro349Arg, p.Met688Arg, the *MLH1* p.Gly67Arg, p.Thr82Ala, p.Lys618Ala, the *MSH6* p.Ala1236Pro, and the *PMS2* p.Arg20Gln were the most likely to be pathogenic. All these UVs and two others in the likely pathogenic class (*MSH2* p.Cys697Arg and *PMS2* p.Ser46Ile) are missense variants, whereas deep intronic and silent variants never reached the highest probability of pathogenicity. In some cases, this lack is associated with a real absence of mRNA alteration, but in other situations, it could be partially due to data absence, because in silico protein prediction is not possible for nonmissense variants.

Ten UVs resulted nonpathogenic or likely nonpathogenic, whereas 16 remained uncertain. However, these three categories could be biased because of paucity of data or conflicting results. Therefore, we cannot exclude that some of the uncertain and nonpathogenic variants could represent low penetrance alleles, unable to fully manifest their (low) pathogenic role in our limited-size families.

Importantly, the positive control p.Arg406Ter presented the highest probability of pathogenicity, whereas the negative control, p.Ile219Val, had one of the lowest. Moreover, three of the class 1 or 2 UVs (*MSH2* p.Gly322Asp, *MLH1* p.Ser406Asn and p.Val716Met) were classified as benign in a structured assessment study on MMR ambiguous mutations by Barnetson et al.[34] *MSH2* p.Gly322Asp and *MLH1* c.1039-8T>A were also considered by Arnold et al.,[24] but not studied further, because they were found at polymorphic frequency in unaffected controls. This points to the potential value of our classification scheme.

All the class 4 and 5 UVs present MSI and/or IHC absence of the UV-carrying protein while almost all showed LOH of the wild-type allele. All are present in at least one family fulfilling the Amsterdam criteria and only one was present in healthy controls. Some of them fall in fundamental functional domains, where the aminoacidic sequence is highly conserved and the perfect folding of the domain is very important. This is the case for *MSH2* p.Pro349Arg and p.Cys697Arg, the latter located in the highly conserved adenosine triphosphate domain where a similar variant, p.Cys697Phe, is reported to completely inactivate MSH2.[35] It is also the case for *MLH1* p.Gly67Arg that is located in the adenosine triphosphate binding region and showed a reduced MMR activity of the mutated protein in functional in vitro assays.[16–18] p.Cys697Arg was
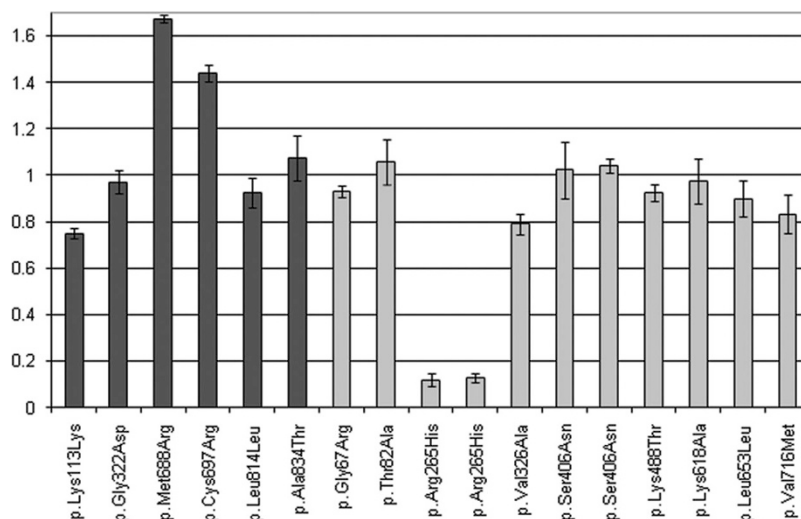
Fig. 1. cDNA UV expression. Primer extension analysis was used to calculate the ratios between mutant and wild-type peak area. Allelic imbalance was estimated dividing cDNA value by gDNA value. Dark column, *MSH2* UV; light column, *MLH1* UV.

overexpressed in cDNA. The increased expression of the allele with p.Cys697Arg could be due to an increased transcription or mRNA stability associated with a reduced protein functionality. Some other UVs fall in domains involved in repair function, like *MSH2* p.Met688Arg that is located in a site that, when changed, leads to a moderate reduction of repair capacity, as reported for p.Met688Ile.[36] *MLH1* p.Thr82Ala has been reported to reduce MMR activity in vitro[19] and seems to cosegregate in this family with p.Lys618Ala, an UV that modifies protein stability leading to increased degradation, without perturbing PMS2 binding and functionality.[20] However, in the literature, the outcome of the functional tests for the latter variant are contradictory and it is possible that both UVs have some pathogenic role and they could act in synergy. According to an alternative classification,[34] p.Lys618Ala should be a benign variant and consequently only p.Thr82Ala could be truly responsible for the observed phenotype. For *MSH6* and *PMS2*, only p.Ala1236Pro and p.Arg20Gln, respectively, resulted as class 5 pathogenic, but this conclusion should be considered with caution. There are no functional information on these UVs, *MSH6* and *PMS2* variants are less studied, and the literature data are truly scant. Both p.Ala1236Pro and p.Arg20Gln are present in families fulfilling the Amsterdam criteria and in probands with highly unstable tumors that are features associated with the highest probability of carrying a pathogenic mutation. This could be the reason that p.Arg20Gln was class 5 even if it is present in healthy controls and, in a family, occurs with a pathogenic mutation. Inclusion of p.Ala1236Pro class 5 seems to be further justified by a high probability of causality derived also from multiple OR values, sometimes low, but always concordant and in favor of pathogenicity. p.Ser46Ile was placed in class 4 by our evaluation. We have already described it to be present in a peculiar Turcot Syndrome family,[33] where the carrier showed the typical clinical signs of a compound heterozygote. This feature, together with clinical and molecular data, strongly suggests the pathogenic role of the UV. In addition, it has been described as present in a patient with early onset sporadic colorectal cancer.[37]

There are several critical aspects in the evaluation of UVs that have been well-debated several times. Every positive datum

is useful in recognizing pathogenic variants, whereas the negative data increase the UV's probability of being classified as nonpathogenic, even if a negative datum is not always an index of nonpathogenicity. MSI is a useful marker for MMR deficiency; however, MSI absence is not only related to nonpathogenic UVs but also to mutations that induce low or null degrees of instability, like some *MSH6* ones.[38] Conversely, MSI presence may not be related to the UV, as a coexisting pathogenic mutation may not have been identified or *MLH1* promoter hypermethylation is present. IHC is also useful, but a lack of expression is not always related to functional alteration of the protein: a missense variant does not truncate the protein but it could induce a conformational change that alters the binding of the antibody, although not necessarily inactivating the protein. Conversely, a MMR protein with a single amino acid change could be functionally inactivated but still recognized by the antibody. Future studies would warrant a likelihood model accounting for these caveats.

UV-disease segregation is considered one of the most important parameters to be evaluated for the assessment of UVs pathogenicity.[4] However, this is not always easy to analyze as it may be difficult to obtain DNA samples from several members of the same family: if the genetic test reveals only an UV in the proband, lack of data for pathogenicity does not allow a complete and correct risk evaluation and genetic counseling in the family. In fact, we found many difficulties in recruiting relatives of the UV carriers, mainly due to the unclear pathogenic role of the UV itself and the consequent uncertainty of the genetic test result. The small number of relatives is the cause of poor informative value of segregation data in some families. Moreover, it is also possible that an UV segregates with the disease because it is in linkage with a pathogenic mutation: this occurred in two UVs (*MSH2* p.Leu814Leu and *MLH1* pArg265His). Alternatively, the variant could cosegregate with disease by chance, as is probable for those UVs that were shared by two members of the same sibship (*MSH2* p.Lys113Lys, *MLH1* p.Thr82Ala and p.Lys618Ala, and *MSH6* p.Glu983Gln and p.Pro1082Pro).

Most UVs are rare and, consequently, their frequency in healthy controls should be determined in prohibitively large samples, not

Table 4 Predictions of in silico software for nucleotide sequence

Unclassified variant	ESEfinder				RescueESE	PESX	NNSplice		Splice site finder		Genescan
	SF2/ASF (br)	SC35	SRP40	SRP55			Donor	Acceptor	Donor	Acceptor	
MSH2											
p.Lys113Lys	+	=	=	=	+	+	=	=	=	=	+
p.Gly322Asp	+	+	+	=	=	+	=	=	=	=	+
p.Ala328Ala	+	+	=	=	=	=	=	=	=	+	+
p.Pro349Arg	=	+	=	+	=	=	=	=	=	+	+
p.Arg406Ter	=	=	+	=	=	=	=	=	–	–	+
p.Leu556Leu	+	=	=	=	=	+	=	=	=	=	+
p.Lys579Lys	=	=	=	=	=	=	=	=	=	=	+
p.Met688Arg	+	=	+	=	=	=	=	=	=	=	+
p.Cys697Arg	+	+	+	+	=	=	+	=	+	+	+
p.Leu814Leu	=	=	=	+	=	=	+	=	=	+	+
p.Ala834Thr	=	+	+	=	=	=	=	=	=	=	=
c.1077-10T>C	–	–	–	–	–	–	=	+	=	+	=
c.1387-8G>T	–	–	–	–	–	–	=	+	=	+	=
MLH1											
p.Gly67Arg	+	+	=	=	=	=	=	=	=	+	+
p.Thr82Ala	=	+	+	=	=	=	=	=	=	=	+
p.Ile219Val	=	=	=	=	+	=	=	=	–	–	=
p.Arg265His	=	+	+	=	=	+	=	=	=	=	=
p.Val326Ala	+	=	=	+	=	=	=	=	=	=	+
p.Pro355Ala	=	+	=	+	=	=	=	=	=	+	+
p.Ser406Asn	=	=	=	=	+	+	=	=	=	=	=
p.Lys488Thr	+	=	+	=	+	=	=	=	=	+	+
p.Lys618Ala	=	+	=	=	+	+	=	=	=	+	+
p.Leu653Leu	+	=	=	+	=	=	=	=	=	=	+
p.Val716Met	+	+	=	+	+	+	=	=	=	=	+
c.116+8G>A	–	–	–	–	–	–	+	=	=	=	=
c.307–29C>A	–	–	–	–	–	–	=	=	=	+	+
c.790+10A>G	–	–	–	–	–	–	=	=	=	=	+
c.884+4A>G	–	–	–	–	–	–	+	+	+	=	=
c.1039-8T>A	–	–	–	–	–	–	=	+	=	+	+
MSH6											
c.628–56C>T	–	–	–	–	–	–	=	=	=	=	=
p.Glu983Gln	+	=	=	=	=	+	=	=	=	=	+
p.Pro1082Pro	+	+	=	=	+	+	=	=	=	+	+
p.Ala1236Pro	=	=	=	=	+	+	=	=	=	=	+
PMS2											
p.Arg20Gln	+	+	+	=	+	+	=	=	=	+	+
p.Ser46Ile	=	=	+	=	=	=	=	=	=	=	+
p.Thr511Ala	+	=	+	+	=	=	=	=	=	=	+
p.Thr597Ser	=	=	+	=	=	+	=	=	=	+	+

+, Alteration predicted; =, alteration not predicted; –, missing data because not tested or not evaluable.

Table 5 Predictions of in silico software for protein and functional data

Unclassified variant	In vitro studies	Polyphen	SIFT	PMut	A-GVGD
MSH2					
p.Lys113Lys	—	—	—	—	—
p.Gly322Asp	=	Possibly	Tolerated	Pathological	C0
p.Ala328Ala	—	—	—	—	—
p.Pro349Arg	—	Probably	Not tolerated	Pathological	C65
p.Arg406Ter	—	—	—	—	—
p.Leu556Leu	—	—	—	—	—
p.Lys579Lys	—	—	—	—	—
p.Met688Arg	—	Probably	Not tolerated	Pathological	C45
p.Cys697Arg	—	Probably	Not tolerated	Pathological	C65
p.Leu814Leu	—	—	—	—	—
p.Ala834Thr	=	Possibly	Tolerated	Pathological	C0
c.1077-10T>C	—	—	—	—	—
c.1387-8G>T	—	—	—	—	—
MLH1					
p.Gly67Arg	Red	Probably	Not tolerated	Pathological	C65
p.Thr82Ala	Red	Probably	Not tolerated	Pathological	C35
p.Ile219Val	=	Benign	Tolerated	Neutral	C0
p.Arg265His	=	Probably	Not tolerated	Pathological	C25
p.Val326Ala	—	Possibly	Not tolerated	Pathological	C25
p.Pro355Ala	=	Possibly	Tolerated	Pathological	C0
p.Ser406Asn	=	Benign	Tolerated	Pathological	C0
p.Lys488Thr	—	Benign	Tolerated	Pathological	C0
p.Lys618Ala	=	Probably	Not tolerated	Pathological	C0
p.Leu653Leu	—	—	—	—	—
p.Val716Met	=	Benign	Tolerated	Pathological	C0
c.116+8G>A	—	—	—	—	—
c.307-29C>A	—	—	—	—	—
c.790+10A>G	—	—	—	—	—
c.884+4A>G	—	—	—	—	—
c.1039-8T>A	—	—	—	—	—
MSH6					
c.628-56C>T	—	—	—	—	—
p.Glu983Gln	—	Possibly	Tolerated	Neutral	C0
p.Pro1082Pro	—	—	—	—	—
p.Ala1236Pro	—	Possibly	Tolerated	Pathological	C0
PMS2					
p.Arg20Gln	—	Benign	Tolerated	Pathological	C0
p.Ser46Ile	—	Probably	Not tolerated	Pathological	C15
p.Thr511Ala	—	Benign	Tolerated	Pathological	C0
p.Thr597Ser	—	Benign	Tolerated	Pathological	C0

Red, reduced functionality; =, normal functionality; —, missing data because not tested or not evaluable.

Table 6 The classification of unclassified variants[a]

Unclassified variant	Likelihood ratio	Probability of pathogenicity	Class
MSH2			
p.Lys113Lys	0.004	0.003982	2
p.Gly322Asp	0.011	0.010921	2
p.Ala328Ala	0.112	0.100846	3
p.Pro349Arg	192.521	0.994833	5
p.Arg406Ter[b]	942167.365	0.999999	5
p.Leu556Leu	0.221	0.180913	3
p.Lys579Lys	1.260	0.557569	3
p.Met688Arg	4726.470	0.999788	5
p.Cys697Arg	63.304	0.984449	4
p.Leu814Leu	0.815	0.449177	3
p.Ala834Thr	0.002	0.002285	2
c.1077-10T>C	0.00002	0.000021	1
c.1387-8G>T	2.332	0.699918	3
MLH1			
p.Gly67Arg	1580.878	0.999368	5
p.Thr82Ala	1698.382	0.999412	5
p.Ile219Val[c]	0.001	0.000595	1
p.Arg265His	17.744	0.946649	3
p.Val326Ala	0.001	0.000618	1
p.Pro355Ala	6.629	0.868927	3
p.Ser406Asn	0.046	0.044017	2
p.Lys488Thr	3.409	0.773184	3
p.Lys618Ala	677.637	0.998526	5
p.Leu653Leu	0.009	0.009055	2
p.Val716Met	0.001	0.000866	1
c.116+8G>A	0.192	0.160982	3
c.307-29C>A	0.263	0.208338	3
c.790+10A>G	0.009	0.008654	2
c.884+4A>G	2.163	0.683865	3
c.1039-8T>A	0.001	0.001072	2
MSH6			
c.628-56C>T	6.675	0.869702	3
p.Glu983Gln	1.147	0.534338	3
p.Pro1082Pro	0.648	0.393185	3
p.Ala1236Pro	1793.804	0.999443	5
PMS2			
p.Arg20Gln	186.603	0.994670	5
p.Ser46Ile	36.795	0.973542	4
p.Thr511Ala	6.931	0.873909	3
p.Thr597Ser	0.286	0.222641	3

[a]Final results reported in this table are not intended for clinical use.
[b]Positive control.
[c]Negative control.

available for this study. However, if a variant showed a frequency of 1% or more in our limited-size control sample, we could infer that it most likely does not have a pathogenic role.[4]

The experimental analysis of splicing alteration was complicated by the difficulty in obtaining viable lymphocytes, necessary for the establishment of lymphoblastoid cell lines. Moreover, the only extra bands seen in our RNA assays seemed to be physiologic alternative splicings,[39] but we cannot exclude that the reverse transcriptase-PCR we used for the analysis was unable to highlight the existing alterations.

A number of databases and software to evaluate UVs have been developed. None of them, individually, is able to predict the real molecular and functional effect in all cases. However, a combination of analyses seems to be more reliable than a single-software test, even if an analysis performed using different web tools is complicated by prediction discordance, particularly for ESE/exon splicing silencer creation or elimination, due to different algorithms used to evaluate nucleotide variations. A study conducted on ESEfinder and then confirmed on RescueESE8,[40,41] demonstrated that the prediction does not always correlate with an in vitro effect. Protein-prediction software tools are more reliable and there is greater agreement in their output data.

Our choice to consider all these clinical, molecular, and bioinformatics features is an attempt to remedy every single method defect and lack of an acknowledged evaluation. To this aim, an assessment of all these features in very large cohorts is important for establishing the true power of each one. It is especially fundamental to have access to large and appropriate reference sample sets for derivation of the OR necessary to predict the probability of pathogenicity of mutations that do not clearly affect mRNA or protein.

To be as objective as possible, we calculated the probability of pathogenicity of every UV with the method proposed by Goldgar et al.,[4,13] calculating OR for every feature that was possible. The probabilities calculated were then classified in five classes, following the guidelines of the Special Issue of Human Mutation.[3–12] An increasing number of groups are trying to classify UVs using these recommendations, but most of them focus on the *BRCA1* and *BRCA2* genes, for which a multifactorial likelihood classification has already been developed and refined. Instead, for MMR genes, there are not well-established models or well-characterized features[42] so that, at the time of writing, a very few groups have attempted to classify MMR UVs with the Bayesian likelihood method. The study by Arnold et al.[24] investigated several variants, three of which are in common with our data (*MLH1* c.307-29C>A, c.1039-8 T>A and *MSH2* p.Gly322Asp): only c.307-29C>A was analyzed by a similar comprehensive approach and assigned to class 3, as it is in our study.

As each pathogenicity class is associated with surveillance and testing recommendations, these integrated analyses could help laboratory geneticists better understand the pathogenic role of UVs and obtain useful data for genetic counselors. However, our work has to be considered among the pioneers in this framework and most of our evaluations need to be further refined. At the moment, this evaluation has to be considered only for research purposes and not for clinical use.

REFERENCES

1. Lynch HT, Lynch PM, Lanspa SJ, Snyder CL, Lynch JF, Boland CR. Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clin Genet* 2009;76:1–18.
2. Genuardi M, Carrara S, Anti M, Ponz de Leòn M, Viel A. Assessment of pathogenicity criteria for constitutional missense mutations of the hereditary nonpolyposis colorectal cancer genes MLH1 and MSH2. *Eur J Hum Genet* 1999;7:778–782.
3. Tavtigian SV, Greenblatt MS, Goldgar DE, Boffetta P. Assessing pathogenicity: overview of results from the IARC Unclassified Genetic Variants Working Group. *Hum Mutat* 2008;29:1261–1264.
4. Goldgar DE, Easton DF, Byrnes GB, Spurdle AB, Iversen ES, Greenblatt MS. Genetic evidence and integration of various data sources for classifying uncertain variants into a single model. *Hum Mutat* 2008;29:1265–1272.
5. Greenblatt MS, Brody LC, Foulkes WD, et al. Locus-specific databases and recommendations to strengthen their contribution to the classification of variants in cancer susceptibility genes. *Hum Mutat* 2008;29:1273–1281.
6. Plon SE, Eccles DM, Easton D, et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat* 2008;29:1282–1291.
7. Hofstra RM, Spurdle AB, Eccles D, et al. Tumor characteristics as an analytic tool for classifying genetic variants of uncertain clinical significance. *Hum Mutat* 2008;29:1292–1303.
8. Spurdle AB, Couch FJ, Hogervorst FB, Radice P, Sinilnikova OM. Prediction and assessment of splicing alterations: implications for clinical testing. *Hum Mutat* 2008;29:1304–1313.
9. Couch FJ, Rasmussen LJ, Hofstra R, Monteiro AN, Greenblatt MS, de Wind N. Assessment of functional effects of unclassified genetic variants. *Hum Mutat* 2008;29:1314–1326.
10. Tavtigian SV, Greenblatt MS, Lesueur F, Byrnes GB. In silico analysis of missense substitutions using sequence-alignment based methods. *Hum Mutat* 2008;29:1327–1336.
11. Ou J, Niessen RC, Vonk J, Westers H, Hofstra RM, Sijmons RH. A database to support the interpretation of human mismatch repair gene variants. *Hum Mutat* 2008;29:1337–1341.
12. Tavtigian SV, Byrnes GB, Goldgar DE, Thomas A. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. *Hum Mutat* 2008;29:1342–1354.
13. Goldgar DE, Easton DF, Deffenbaugh AM, et al. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *Am J Hum Genet* 2004;75:535–544.
14. Baudhuin LM, Burgart LJ, Leontovich O, Thibodeau SN. Use of microsatellite instability and immunohistochemistry testing for the identification of individuals at risk for Lynch syndrome. *Fam Cancer* 2005;4:255–265.
15. Laghi L, Bianchi P, Malesci A. Differences and evolution of the methods for the assessment of microsatellite instability. *Oncogene* 2008;27:6313–6321.
16. Shimodaira H, Filosi N, Shibata H, et al. Functional analysis of human MLH1 mutations in *Saccharomyces cerevisiae*. *Nat Genet* 1998;19:384–389.
17. Kondo E, Suzuki H, Horii A, Fukushige S. A yeast two-hybrid assay provides a simple way to evaluate the vast majority of hMLH1 germ-line mutations. *Cancer Res* 2003;63:3302–3308.
18. Raevaara TE, Korhonen MK, Lohi H, et al. Functional significance and clinical phenotype of nontruncating mismatch repair variants of MLH1. *Gastroenterology* 2005;129:537–549.
19. Takahashi M, Shimodaira H, Andreutti-Zaugg C, Iggo R, Kolodner RD, Ishioka C. Functional analysis of human MLH1 variants using yeast and in vitro mismatch repair assays. *Cancer Res* 2007;67:4595–4604.
20. Perera S, Bapat B. The MLH1 variants p.Arg265Cys and p.Lys618Ala affect protein stability while p.Leu749Gln affects heterodimer formation. *Hum Mutat* 2008;29:332.
21. Drost M, Zonneveld JB, van Dijk L, et al. A cell-free assay for the functional analysis of variants of the mismatch repair protein MLH1. *Hum Mutat* 2010;31:247–253.
22. Ollila S, Dermadi Bebek D, Jiricny J, Nyström M. Mechanisms of patho-
genicity in human MSH2 missense mutants. *Hum Mutat* 2008;29:1355–1363.
23. Trojan J, Zeuzem S, Randolph A, et al. Functional analysis of hMLH1 variants and HNPCC-related mutations using a human expression system. *Gastroenterology* 2002;122:211–219.
24. Arnold S, Buchanan DD, Barker M, et al. Classifying MLH1 and MSH2 variants using bioinformatic prediction, splicing assays, segregation, and tumor characteristics. *Hum Mutat* 2009;30:757–770.
25. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Stat Med* 2004;23:1351–1375.
26. Karchin R, Agarwal M, Sali A, Couch F, Beattie MS. Classifying variants of undetermined significance in BRCA2 with protein likelihood ratios. *Cancer Inform* 2008;6:203–216.
27. Lastella P, Surdo NC, Resta N, Guanti G, Stella A. In silico and in vivo splicing analysis of MLH1 and MSH2 missense mutations shows exon- and tissue-specific effects. *BMC Genomics* 2006;7:243.
28. Engel C, Forberg J, Holinski-Feder E, et al. Novel strategy for optimal sequential application of clinical criteria, immunohistochemistry and microsatellite analysis in the diagnosis of hereditary nonpolyposis colorectal cancer. *Int J Cancer* 2006;118:115–122.
29. Easton DF, Deffenbaugh AM, Pruss D, et al. A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. *Am J Hum Genet* 2007;81:873–883.
30. Thompson D, Easton DF, Goldgar DE. A full-likelihood method for the evaluation of causality of sequence variants from family data. *Am J Hum Genet* 2003;73:652–655.
31. Marroni F, Pastrello C, Benatti P et al. A genetic model for determining MSH2 and MLH1 carrier probabilities based on family history and tumor microsatellite instability. *Clin Genet* 2006;69:254–262.
32. Viel A, Novella E, Genuardi M, et al. Lack of PMS2 gene-truncating mutations in patients with hereditary colorectal cancer. *Int J Oncol* 1998;13:565–569.
33. Agostini M, Tibiletti MG, Lucci-Cordisco E, et al. Two PMS2 mutations in a Turcot syndrome family with small bowel cancers. *Am J Gastroenterol* 2005;100:1886–1891.
34. Barnetson RA, Cartwright N, van Vliet A, et al. Classification of ambiguous mutations in DNA mismatch repair genes identified in a population-based study of colorectal cancer. *Hum Mutat* 2008;29:367–374.
35. Ollila S, Sarantaus L, Kariola R, et al. Pathogenicity of MSH2 missense mutations is typically associated with impaired repair capability of the mutated protein. *Gastroenterology* 2006;131:1408–1417.
36. Gammie AE, Erdeniz N, Beaver J, Devlin B, Nanji A, Rose MD. Functional characterization of pathogenic human MSH2 missense mutations in *Saccharomyces cerevisiae*. *Genetics* 2007;177:707–721.
37. Nakagawa H, Lockman JC, Frankel WL, et al. Mismatch repair gene PMS2: disease causing germline mutations are frequent in patients whose tumors stain negative for PMS2 protein, but paralogous genes obscure mutation detection and interpretation. *Cancer Res* 2004;64:4721–4727.
38. Lagerstedt Robinson K, Liu T, Vandrovcova J, et al. Lynch syndrome (hereditary nonpolyposis colorectal cancer) diagnostics. *J Natl Cancer Inst* 2007;99:291–299.
39. Genuardi M, Viel A, Bonora D, et al. Characterization of MLH1 and MSH2 alternative splicing and its relevance to molecular testing of colorectal cancer susceptibility. *Hum Genet* 1998;102:15–20.
40. Lastella P, Resta N, Miccolis I, Quagliarella A, Guanti G, Stella A. Site directed mutagenesis of hMLH1 exonic splicing enhancers does not correlate with splicing disruption. *J Med Genet* 2004;41:e72.
41. Auclair J, Busine MP, Navarro C, et al. Systematic mRNA analysis for the effect of MLH1 and MSH2 missense and silent mutations on aberrant splicing. *Hum Mutat* 2006;27:145–154.
42. Spurdle AB. Clinical relevance of rare germline sequence variants in cancer genes: evolution and application of classification models. *Curr Opin Genet Dev* 2010;20:315–323.