Research Article

Cancer Prevention Research

In Silico Systems Biology Analysis of Variants of Uncertain Significance in Lynch Syndrome Supports the Prioritization of Functional Molecular Validation

Ester Borras[1], Kyle Chang[1], Mala Pande[2,3], Amanda Cuddy[4], Jennifer L. Bosch[1], Sarah A. Bannon[3], Maureen E. Mork[3], Miguel A. Rodriguez-Bigas[3,4], Melissa W. Taggart[5], Patrick M. Lynch[2,3], Y. Nancy You[3,4], and Eduardo Vilar[1,3,6]

Abstract

Lynch syndrome (LS) is a genetic condition secondary to germline alterations in the DNA mismatch repair (MMR) genes with 30% of changes being variants of uncertain significance (VUS). Our aim was to perform an *in silico* reclassification of VUS from a large single institutional cohort that will help prioritizing functional validation. A total of 54 VUS were detected with 33 (61%) novel variants. We integrated family history, pathology, and genetic information along with supporting evidence from eight different *in silico* tools at the RNA and protein level. Our assessment allowed us to reclassify 54% (29/54) of the VUS as probably damaging, 13% (7/54) as possibly damaging, and 28% (15/54) as probably neutral. There are more than 1,000 VUS reported in MMR genes and our approach facilitates the prioritization of further functional efforts to assess the pathogenicity to those classified as probably damaging. *Cancer Prev Res; 10(10); 580–7. ©2017 AACR.*

Introduction

Lynch syndrome (LS) is the most common hereditary colorectal cancer syndrome, accounting for approximately 3% of the total of colorectal cancer (1) and affecting potentially up to 1.1 million carriers in the United States (2). LS is caused by germline alterations in one of the DNA mismatch repair (MMR) genes: *MLH1* (MIM# 120436), *MSH2* (MIM# 609309), *MSH6* (MIM# 600678), *PMS2* (MIM# 600259), or *EPCAM/TACSTD1* (MIM# 185535; refs. 3, 4). LS patients present mainly with early-onset colorectal and endometrial cancers as well as ovarian, gastric, small bowel, and urinary tract tumors, among others (5). LS-associated tumors display loss of MMR protein expression and microsatellite instability (MSI; ref. 5). The Amsterdam and revised Bethesda criteria guidelines were developed two decades ago to clinically identify suspected LS families that need genetic testing (5). More recently, the implementation of universal tumor-based testing strategies using immunochemistry (IHC) of MMR proteins or detection of MSI by PCR analysis has become the main trigger for confirmatory germline testing (6).

The majority of germline MMR mutations detected through genetic testing are short insertions, deletions, or nucleotide substitutions that create a premature stop codon or nucleotide changes at consensus splice sites affecting mRNA splicing and leading to a truncated protein (7). These mutations are directly classified as pathogenic following the American College of Medical Genetics and Genomics (ACMG) criteria, thus establishing a diagnosis of LS and triggering clear recommendations for surveillance and subsequent genetic testing of at-risk family members (5). However, 30% of the identified DNA changes in MMR genes are classified as variants of uncertain significance (VUS; ref. 7). VUS continue representing a major clinical dilemma as their functional consequences are unknown, and, therefore, there are no established risk assessment, surveillance, or chemoprevention recommendations for these patients and family members. Furthermore, patients harboring VUS are then left in a "*genetic limbo*" of uncertainty until isolated efforts from researchers resolve their genetic conundrum in one way or the other (pathogenicity vs. benign/neutral status). These mutations are typically missense, small insertions or deletions that do not encode a truncated protein or mutations close to a splice site that have unclear functional significance. Classifying VUS as pathogenic or neutral is a challenge unless subjected to an exhaustive functional analysis (8). In fact, reclassification of VUS requires an orthogonal

[1]Department of Clinical Cancer Prevention, The University of Texas MD Anderson Cancer Center, Houston, Texas. [2]Department of Gastroenterology, Hepatology, and Nutrition, The University of Texas MD Anderson Cancer Center, Houston, Texas. [3]Clinical Cancer Genetics Program, The University of Texas MD Anderson Cancer Center, Houston, Texas. [4]Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas. [5]Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, Texas. [6]Department of Gastrointestinal Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas.

Note: Supplementary data for this article are available at Cancer Prevention Research Online (http://cancerprevres.aacrjournals.org/).

This work has been presented in part during the 6th Biennial International Society for Gastrointestinal Hereditary Tumours (InSiGHT) Meeting, Sao Paulo, Brazil, June 18–20, 2015.

Corresponding Author: Eduardo Vilar, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Boulevard, Unit 1360, Houston, TX 77030. Phone: 713-745-4929; Fax: 713-794-4403; E-mail: EVilar@mdanderson.org

doi: 10.1158/1940-6207.CAPR-17-0058

approach integrating data from different sources such as family history, clinical information, tumor pathology, frequency of the VUS in the general population, and *in vitro* functional assays at the RNA and protein levels (9–13). An initial step for reclassification of VUS is to evaluate the functional impact of the VUS using *in silico* systems biology tools that predict the consequences of the change at both RNA and protein levels (14–16). Recently, a five-tiered classification system of MMR VUS has been proposed by The International Society for Gastrointestinal Hereditary Tumours (InSiGHT) to facilitate a consistent management of families suspected to have LS and to encourage an international collaboration for the curation and classification of VUS in MMR genes (7, 17).

Our aim in this study was to perform an *in silico* systems biology reclassification of VUS detected in a large institutional cohort of patients suspected to have LS that was not included in the original InSiGHT manuscript. We incorporated data from tumor pathology, family history, and *in silico* RNA and protein prediction analyses and validated the performance of our approach in a group of VUS already reclassified as pathogenic or neutral. Our approach can be used to prioritize variants for further laboratory investigation and reclassification using *in vitro* functional assays.

Materials and Methods

Patients and samples

A total of 54 patients with clinical suspicion of LS that underwent genetic testing and were found to harbor a VUS in one of the MMR genes were included in this study (Supplementary Tables S1 and S2). In addition, a total of 36 patients with 21 unique missense mutations from the same data set already reclassified as InSiGHT class 1 (pathogenic) or 5 (neutral) were used for validation purposes (Supplementary Table S3). All patients were referred to genetic counseling by their primary care teams based on institutional referral criteria, which includes patients who meet Bethesda or Amsterdam criteria, or otherwise have a concerning personal or family history for hereditary cancer per clinician discretion. An additional set of patients was identified via universal MSI and/or IHC tumor testing of colorectal and endometrial tumors, which initiated in 2009 and 2012, respectively. Genetic testing for LS was recommended after genetic counseling evaluation based on the risk assessment, which included a patient's personal/family history and results of MSI and/or IHC tumor testing (see Supplementary Table S2 for patient clinical characteristics and family history). This cohort of patients was retrieved from two prospective institutional databases at The University of Texas MD Anderson Cancer Center (MDACC) from 2003 to 2012: (i) Clinical Cancer Genetics database, which includes all patients undergoing genetic counseling at UTMDACC and (2) colorectal cancer database, which collects clinical, pathologic, and molecular data of patients with colorectal cancer undergoing surgical resection at our institution. These two databases contained a total of 377 patients that underwent MMR genetic testing.

Clinical, laboratory, and pathology information was obtained from the electronic medical record. Information regarding the family history was collected by certified genetic counselors at the time of the initial assessment (S. A. Bannon and M. E. Mork). In those cases not referred for genetic counseling, information on the family history was retrieved directly from the medical record. Pedigrees made as part of the genetic counseling visit or derived from the information collected in the chart were assessed for fulfillment of Amsterdam I/II and Bethesda guidelines (18). The MDACC Institutional Review Board approved this study.

Tumor analysis

IHC was performed on 5-μm sections of formalin-fixed and paraffin-embedded (FFPE) samples according to standard procedures using a panel of four monoclonal antibodies against the MMR proteins: MLH1 (550838, clone G168-15, BD Pharmingen; BD Biosciences), MSH2 (NA27, clone FE11, Calbiochem, EMD Millipore), MSH6 (610919, clone 44, BD Pharmingen; BD Biosciences), and PMS2 (556415, clone A16-4, BD Pharmingen; BD Biosciences). Loss of expression in the tumor cells was considered solely when there was normal nuclear staining in adjacent non-neoplastic cells, which served as internal controls. These assessments were performed by an expert gastrointestinal pathologist (M.W. Taggart).

For MSI analysis, DNA was extracted from microdissected tumor and normal FFPE tissue. The MSI status was assessed using a panel of seven microsatellite markers (BAT25, BAT26, BAT40, D2S123, D5S346, D17S250, and TGFβRII) and was performed at the Molecular Diagnostic Laboratory in MDACC per standard of care. A microsatellite marker was considered positive when an allelic shift was present in the tumor compared with normal tissue. A case was considered MSI-high (MSI-H) when three of the loci ($\geq$30%) tested were positive, MSI-low (MSI-L) when one or two (<30%) were positive, and microsatellite stable (MSS) when all loci tested were negative.

DNA sequencing

To assess for the presence of germline mutations and large deletions and duplications in the MMR genes, DNA extracted from blood was tested by full genomic sequencing of the exons of the MMR genes and by gene dosage analysis [multiplex ligation-dependent probe amplification (MLPA)]. These analyses were performed in a Clinical Laboratory Improvement Amendments–certified laboratory (City of Hope Molecular Diagnostic Laboratory; Mayo Medical Laboratories; Myriad Genetics) Mutation nomenclature of MMR genes followed Human Genome Variation Society (HGVS) recommendations (version 2.0) with nucleotide 1 corresponding to the A of the ATG-translation initiation codon. Following InSiGHT Variant Interpretation Committee guidelines (7), we considered a pathogenic mutation a coding sequence variation resulting in a stop codon (i.e. nonsense or frameshift pathogenic alteration that is predicted to result in interruption of known functional protein domains). In addition, those variants where mRNA assays indicated that the variant allele will result in a splicing aberration, thus leading to premature stop codon or in-frame deletion disrupting a functional domain or protein conformation, were also considered pathogenic. VUS were checked for pathogenicity in the LOVD database (http://insight-database.org) and all those classified as class 1 (neutral) or class 5 (pathogenic) were used to evaluate the performance of our classification system but were excluded from the final analysis of VUS reported in this article. Therefore, our approach was evaluated in VUS classified as InSiGHT classes 2, 3, 4 or novel ones (not previously reported and without an InSiGHT class assigned).

Bioinformatics analysis

DNA sequences containing the identified MMR variants were analyzed with several bioinformatics tools to evaluate the impact

at the RNA and protein level. We used SAMtools (19) to extract 150 base pair flanking sequence upstream and downstream from the variant of interest. To identify potential splicing mutations, disruption or creation of splice sites were evaluated using NNSplice (20), Splicerport (21), and SoftBerry (22). These three programs are ones of the most common used for splice side variant interpretation in clinical laboratories (13). In addition, these programs have shown a high level of accuracy on their predictions (23). In splice site predictions, major alterations are those that are predicted to destroy or create splicing sites (SS) or produce a SS score modification higher than 20%. The VUS were classified in three categories at the RNA level: (i) aberrant splicing, when two or more programs predicted an aberrant splicing; (ii) noneffect, when 3 programs predicted nonaberrant splicing; and (iii) inconclusive, when only one program predicted an aberrant splicing. Furthermore, we have performed a validation of the performance of these tools integrated in our approach using a total of 24 variants already reported by others to the InSiGHT database, which have been already reclassified as class 1 or 5 (Supplementary Table S4).

The impact of missense variants at the protein level was analyzed using Polyphen2 (24), SIFT (25), Mutation Assessor (26), and FATHMM (27) using website default conditions and Multivariate analysis of protein polymorphisms (MAPP) as described earlier (28). We have selected these three programs are the most commonly used for missense variant interpretation in clinical laboratories (13). The VUS were classified at protein level in four categories: (i) Damaging, if four or more programs predicted the VUS as damaging, probably damaging, or high functional impact; (ii) Probably damaging, if three programs predicted the VUS as damaging, probably damaging, possibly damaging, or medium functional impact; and (iii) Neutral, if three or more programs predicted the VUS as neutral. Moreover, we analyzed the impact at the protein level by combining the results of MAPP and Polyphen2 based on prior probabilities of pathogenicity from these tools (16). The VUS with a prior probability of pathogenicity higher than 80% were classified as pathogenic, lower than 20% as neutral, and from 20% to 80% as VUS (7).

After performing these bioinformatics analyses at the RNA and protein level, we then proceed to reclassify the variants in five categories: (i) Probably damaging, if the VUS was classified as damaging by either protein or RNA *in silico* tools; (ii) Possibly damaging, if the VUS was classified as probably damaging by protein *in silico* tools; (iii) VUS, if the VUS was classified as inconclusive at RNA level and neutral by protein *in silico* tools; and (iv) Probably neutral, if the VUS was classified as neutral by protein and RNA *in silico* tools.

Sensitivity and specificity analysis

We have tallied the numbers of damaging and neutral (including those VUS classified as inconclusive) VUS predictions generated by our *in silico* approach and also by the reported InSiGHT class into a 2-by-2 table. Fisher exact test was performed to assess the level of association between our approach and the reported InSiGHT class. However, as we opted for using only VUS from our clinical data that were already classified by InSiGHT, our numbers in this validation analysis are modest, thus rendering a limited power. Sensitivity was calculated as the number of damaging VUS predictions by our algorithm divided by number of damaging VUS predictions by the reported InSiGHT classification and specificity was calculated as number of neutral VUS predictions by our algorithm divided by number of neutral VUS predictions by the reported InSiGHT classification. Ninety-five percent confidence intervals of sensitivity and specificity were calculated using GraphPad Prism 6.0 (GraphPad Software).

Results

Germline mutational analysis of the *MMR* genes in a total of 377 patients led to the identification of 54 index case carriers of 61 MMR VUS; 54 of them were unique, thus representing 14% of the mutations detected (Fig. 1; Supplementary Table S1). Four VUS were detected in more than one patient. The VUS c.306G>T (p.E102D) located in *MLH1* was detected in 5 non-related patients. The variants c.1361T>G (p.I454R) in *MSH2*, c.1295T>C (p.F432S) in *MSH6*, and c.1437C>G (p.H479Q) in *PMS2*, each were detected in 2 different non-related patients. Six patients harbored two or more VUS in MMR genes. Three patients harbored a VUS and a pathogenic mutation: the first patient had a pathogenic germline mutation in *MSH2* (c.181C>T; p.Q16*) and a VUS (c.2210+7G>T; p.?) located in the same gene; the second patient had a VUS in *PMS2* (c.1437C>G; p.H479Q) and a missense mutation in *MLH1* (p. H264Lfs*2) that caused aberrant splicing leading to a truncated protein (29). In fact, c.1437C>G was previously classified by InSiGHT as class 3 due to insufficient evidence. The third patient harbored a pathogenic mutation in *MSH2* (c.647dup; p.I216Ifs*19) with a VUS in *PMS2* (c.52A>G; p.I18V). The c.52A>G variant was previously classified by InSiGHT as class 2 (likely not pathogenic) due to its presence in more than 1% of European Americans. Moreover, it was also detected in an LS patient with a pathogenic *PMS2* mutation (c.354-1G>A, p.?; ref. 30). These three cases did not have a phenotype consistent with constitutional MMR deficiency, thus supporting their functional reclassification as neutral (Supplementary Table S2). However, we have reported them in this article based on the same approach as the rest of the VUS as we were not able to verify if they were in *trans* with respect to the pathogenic mutation.

From the total of 54 VUS, we detected 19 in *MLH1* (35%), 15 in *MSH2* (28%), 11 in *MSH6* (20%), and 9 in *PMS2* (17%; Fig. 2A; Supplementary Table S1). The vast majority of VUS were missense mutations (76%, 41/54) but we also detected indels (11%, 6/54) and alterations affecting the splicing (13%, 7/54). *MSH6* had higher numbers of indel mutations (36%, 4/11) compared with the other MMR genes (Fig. 2B; Supplementary Table S1). 61% (33/54) of these VUS were novel, 6% (3/54) were already classified by InSiGHT as class 2 (probably neutral), 24% (13/54) as class 3 (VUS), and 9% (5/54) as class 4 (probably pathogenic). *MSH2* and *PMS2* had the highest proportion of new VUS (73% and 67%, respectively; Supplementary Table S1).

Eighty-one percent (44/54) of the patients fulfilled the Amsterdam or Bethesda criteria. The vast majority of patients with VUS in *MLH1*, *MSH2*, and *PMS2* fulfilled Bethesda guidelines (68%, 53%, and 56%, respectively). A notable proportion (18%, 2/11) of patients with VUS in *MSH6* was detected by universal tumor testing without any associated family history (Fig. 2C; Supplementary Table S2). Sixty-eight percent (37/54) of patients were diagnosed with colorectal cancer with a mean age at diagnosis of 43 years, 13% (7/54) with endometrial cancer with a mean age at diagnosis of 50 years, and 19% (10/54) with other LS-associated tumors (Supplementary Table S2). Most of the tumors displayed MSI-H (57%, 31/54) and/or loss of one or more MMR proteins (61%, 33/54). We analyzed the correlation between the MSI-

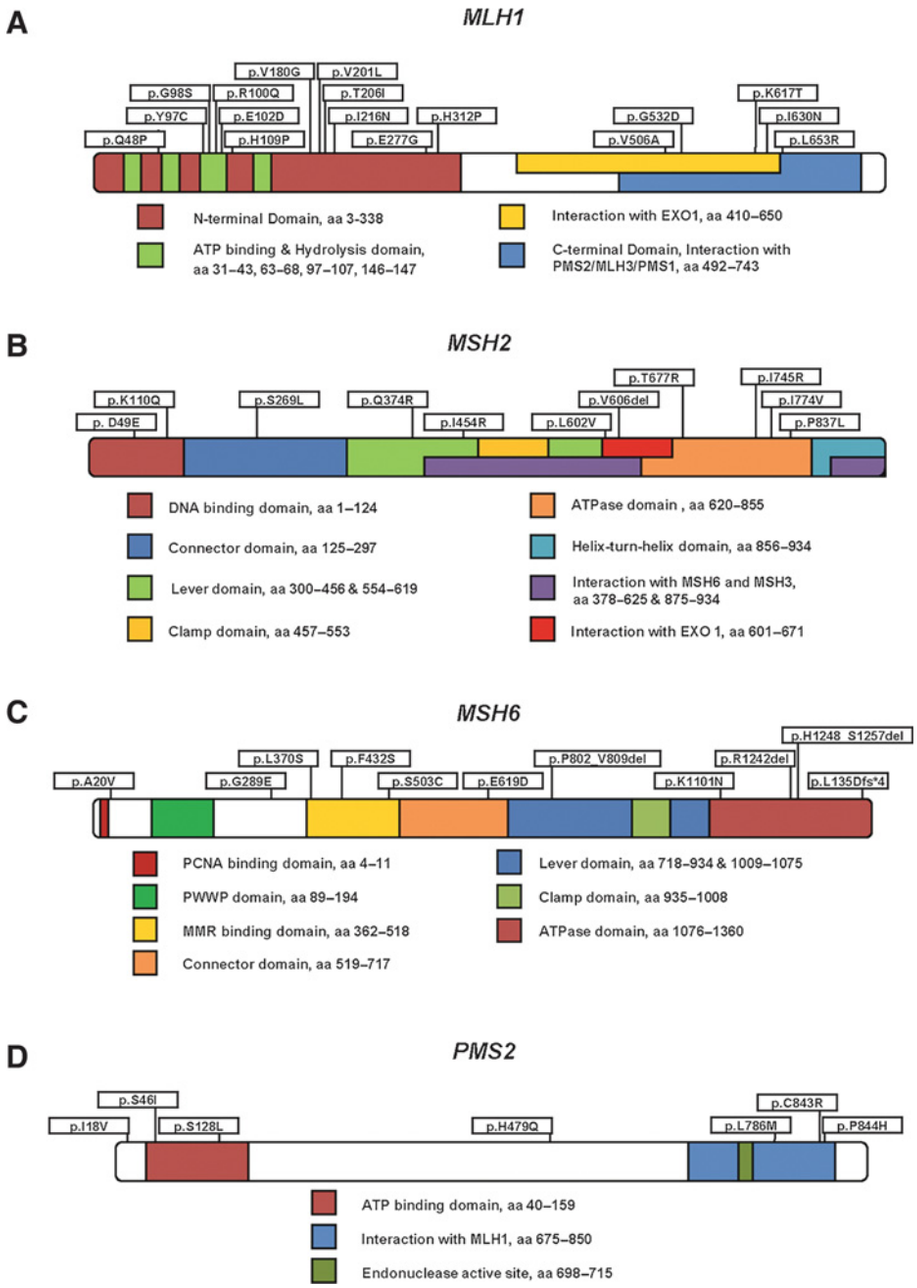

Figure 1.
Schematic representation of (**A**) *MLH1*, (**B**) *MSH2*, (**C**) *MSH6*, and (**D**) *PMS2* showing the functional domains and VUS localization.

status and the gene harboring the VUS and observed the lowest percentage of MSI-H tumors among the *MSH6* VUS carriers (67%, 6/9; Supplementary Table S2; Fig. 2D). Then, we analyzed the correlation between IHC and genotype of the VUS detected and observed that 57% (26/45) of them showed correlation between the gene where the mutation was detected and the loss of protein by staining. VUS detected in *PMS2* (33%, 2/6) had the lowest correlation with IHC compared with *MLH1* (59%, 10/17), *MSH2* (64%, 9/14), and *MSH6* (56%, 5/9; Supplementary Table S2; Fig. 2E).

Then, we used three different programs to predict changes in the acceptor and donor splice sites. Six VUS showed alterations in splice sites by two or more *in silico* programs representing 11% of all VUS analyzed (Fig. 3A and B; Supplementary Tables S5 and S6). Three of these variants were located in *MLH1*: c.208–3C>G (p.?), c.292G>A (p.G98S), and c.306G>T (p.E102D). In *MSH2*, two variants were predicted to be pathogenic at the RNA level, c.2320A>G (p.I774V) and c.2635-3C>G. Finally, one *PMS2* variant was also predicted to be pathogenic at RNA level, c.2276-17T>C (Supplementary Tables S5 and S6). Moreover, five variants showed an aberrant alteration by only one program and, therefore, they were classified as inconclusive. Three of them were located in *MLH1* (c.140-10C>T, c.1958T>G, and c.299G>A) and two in *PMS2* (c.251-18C>G and c.383C>T). We validated our

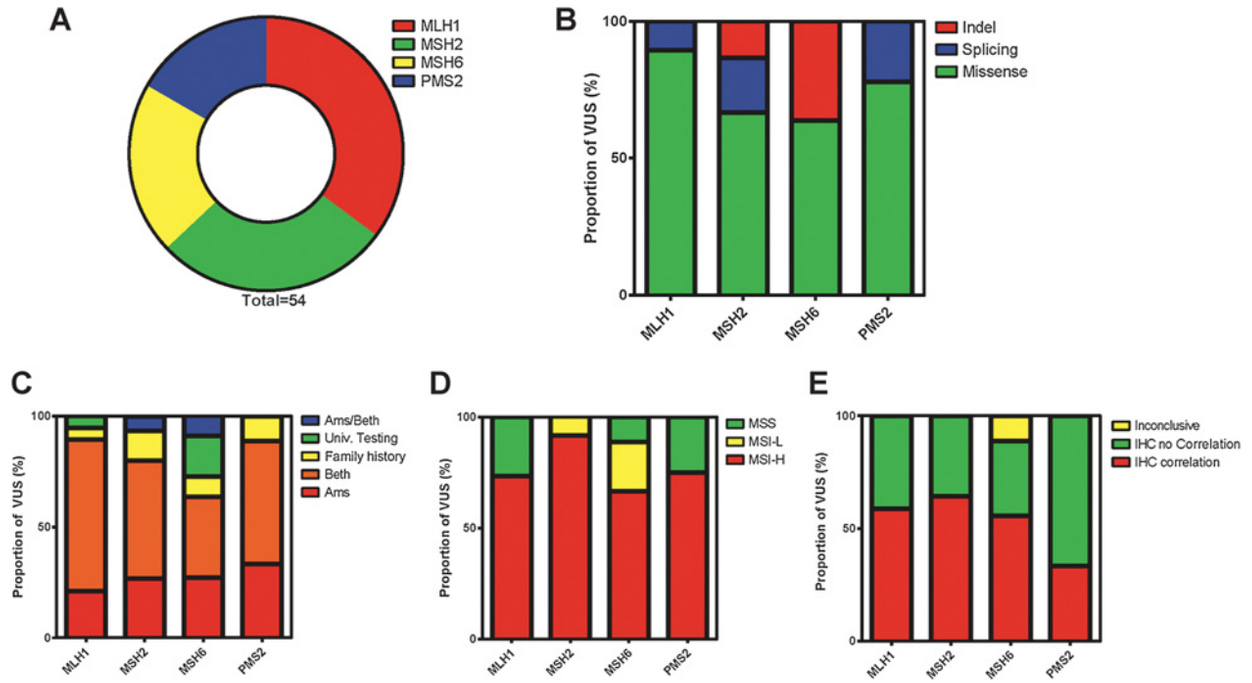

Figure 2.

A, Proportion of variants detected in the MMR genes. **B,** Proportion of different types of variants in MMR genes. **C,** Proportion patients with positive family history criteria per MMR gene mutated. **D,** Proportion of tumors displaying MSS, MSI-L and MSI-H per MMR gene. **E,** Correlation between loss of IHC expression and gene variant location. Abbreviations: Ams, Amsterdam criteria; Beth, Bethesda criteria; Ams/Beth, more than one family with the same variant with a different criteria; MSS, microsatellite stable, MSI-L, microsatellite instability-low; MSI-H, microsatellite instability-high.

approach combining these three RNA splicing *in silico* tools by collecting a total of 24 mutations that have been demonstrated to affect the splicing of MMR genes and as such they have been already reported to InSiGHT as classes 1 and 5 (Supplementary Tables S7 and S8). All class 1 variants were predicted to be neutral by these bioinformatic tools and only 2 class 5 variants resulted in an inconclusive prediction. Therefore, we decided to integrate them along with the neutral variants in order to be very conservative in our assessment (Supplementary Fig. S1). This validation concluded that our approach had a sensitivity of 83.3% (95% CI, 0.51–0.97) and specificity of 100% (95% CI, 0.73–1.0), thus being highly reliable for *in silico* prediction of RNA splicing changes ($P < 0.0001$).

At the protein level, we have reclassified missense VUS using five *in silico* tools as damaging (25 of 41, 61%), probably damaging (9 of 41, 22%), and neutral (7 of 41, 17%; Fig. 3; Supplementary Tables S5 and S6). *MLH1* and *PMS2* variants were classified mostly as damaging (82% and 57%, respectively; Fig. 3C and D; Supplementary Tables S9 and S10). Interestingly, *MSH6* had the highest proportion of neutral variants (43%) and *MSH2* the lowest (0%). Finally, it has been reported that the combination of the "prior probability of pathogenicity" from MAPP and Polyphen2 *in silico* tools has the highest sensitivity to classify VUS at the protein level (16). Using this approach, we classified 21 of our variants as damaging (51%), 9 as true VUS (22%), and 11 as neutral (27%; Supplementary Fig. S2 and Tables S9 and S10). When we compare our classification approach using five different *in silico* tools and the "prior probability" method with MAPP-Polyphen2 both agreed in their classification in 25 of 41 variants with the highest agreement among the ones classified as damaging (72% called as damaging based on our method were also damaging by MAPP-Polyphen2; Supplementary Table S9). Seven variants that were classified as neutral by our approach were assigned the same category by the "prior probability" method. Nine variants classified as probably damaging by our agreement of five *in silico* tools were classified differently by the "prior probability" approach being 3 damaging, 2 VUS, and 4 neutral (Supplementary Tables S9 and S10).

Overall, our approach combining a total of 8 *in silico* RNA and protein tools assigned a status of either probably or possibly damaging to 67% of the variants, which will be recommended to proceed for functional molecular analysis with top priority, and probably neutral to 28%. Only 5% were still being considered VUS (Fig. 3E and F; Supplementary Table S11).

Finally, in order to validate the sensitivity and specificity of our reclassification approach, we took 21 variants that have been detected in patients evaluated at MD Anderson during the same period of time but that have been already reclassified as InSiGHT class 1 (neutral) and 5 (pathogenic; 8 and 13 variants, respectively; Supplementary Tables S3 and S12–S15). The vast majority of these variants were in *MLH1* and *MSH2* (13 and 6, respectively). Our *in silico* approach assigned VUS status to 2 out of 21 of these variants. Among pathogenic variants, 77% were correctly classified as probably damaging. For neutral variants, the degree of misclassification was higher with 50% of the variants predicted to be damaging. Therefore, our approach had a sensitivity of 91.6% (95% CI, 0.61–0.99) and specificity of 42.8% (95% CI, 0.09–0.81, $P = 0.11$; Supplementary Fig. S3; Supplementary Tables S3 and S12–S15).

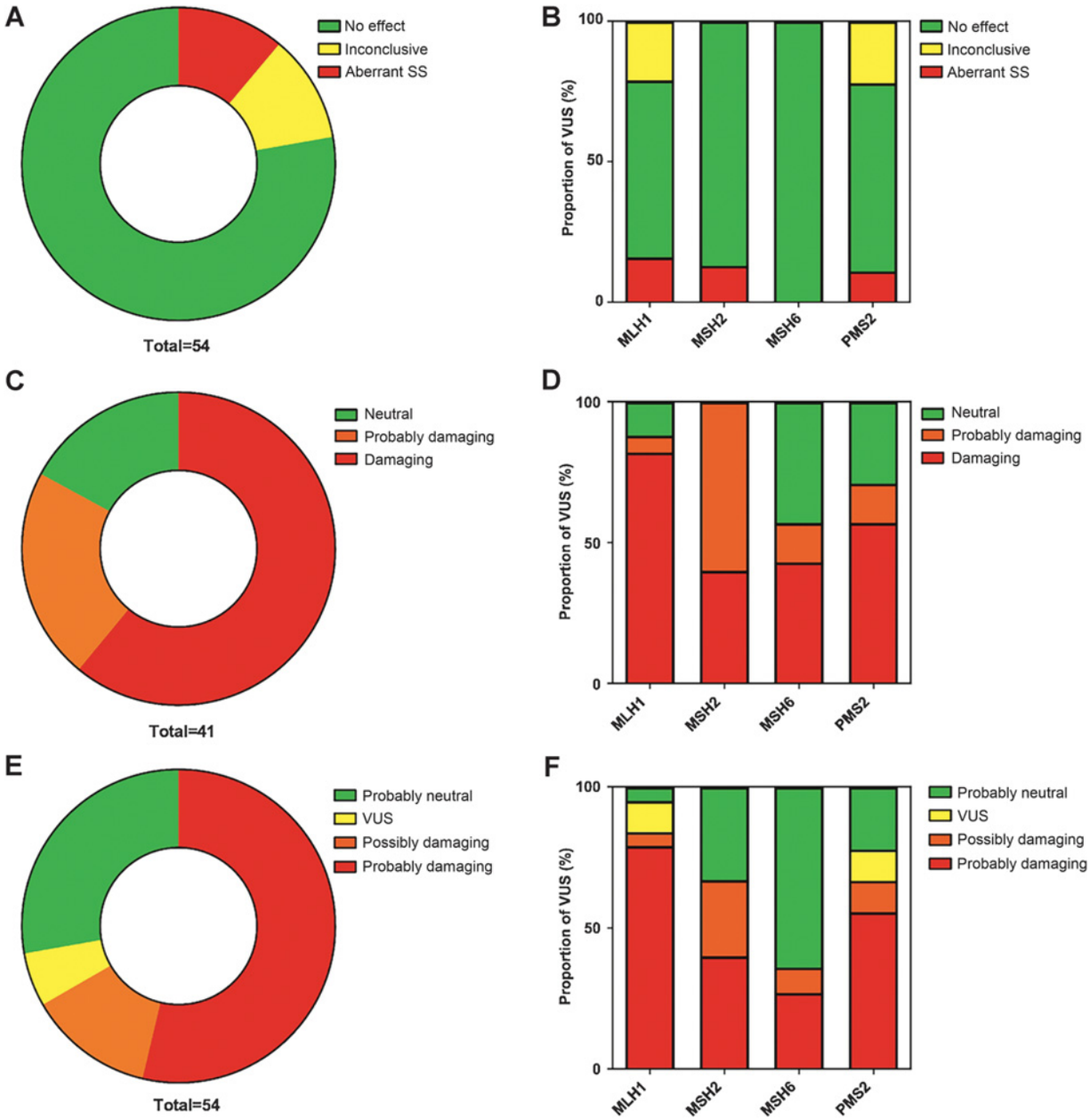

Figure 3.

A, Proportion for different *in silico* predictions at the RNA level. **B,** Classification of variants at the RNA level per gene. **C,** Representation of proportion for different classifications at the protein level. **D,** Classification of variants at the protein level per gene. **E,** Final prior-classification using RNA and protein *in silico* assays for all four MMR genes. **F,** Final prior-classification using RNA and protein *in silico* assays per gene.

Discussion

In this study, we have reported a large set of novel VUS detected in LS patients from a single institution registry along with a detailed analysis of the clinical data and *in silico* predictions at the RNA and protein levels. We have also validated our approach using VUS detected among our patients that have been already reclassified as pathogenic (class 5) or neutral (class 1) by InSiGHT and observed a high sensitivity with modest specificity. These analyses have allowed us classifying 67% (36/54) of the variants as probably or possibly damaging. Our goal was to offer an approach that will help others to prioritize those variants for further functional analysis, thus permitting a definitive reclassification of the variants as suggested by the InSiGHT working group on MMR VUS (7).

VUS variants in MMR genes are common genetic test results that continue presenting a considerable clinical challenge. A number of *in silico* programs have been devised to assist with the prediction

of RNA and protein consequences. As reported previously, we used several RNA splicing software packages to improve our predictions (31–34). We decided not to use enhancer elements as their results were not concordant with each other and also, as reported previously, their predictions did not always correlate with experimental evidence (23, 33, 35). At protein level, the *in silico* tools are generally based on amino acid conservation, severity of substitution, and structural changes of the wild-type and variant proteins. The *in silico* tools used by us have been reported to have strong correlations with experimental functional validation (16, 33, 36).

In addition, we have compared two different reclassification approaches: (i) establishing a consensus among 8 different *in silico* tools to reclassify variants as damaging if all of them predicted pathogenicity and (ii) assessing "prior probability" by combining the results from MAPP-MMR and Polyphen2 (7). The comparison of both approaches has shown that 61% of the variants were classified within the same category with 72% of agreement in variants classified as damaging. In our hands, the prior probability approach classified more variants as VUS compared with ours based on 8 *in silico* tools. Therefore, we conclude that the prior probability approach is more conservative. Our classification using RNA and protein *in silico* prediction tools is allowing us to prioritize these variants for the further studies to be performed. For example, variants that were classified as pathogenic at RNA level such as c.208-3C>G (p.?), c.292G>A (p.G98S) and c.306G>T (p.E102D) in *MLH1*, which also present lost or gain in the acceptor or donor splicing sites, are good candidates to be functionally validated starting to use first *in vitro* expression analysis. Determination of MMR deficiency by IHC and MSI analysis, and other tools to assess protein expression, subcellular localization, and protein–protein interactions are required to start investigating those variants that appear to have a pathogenic effect from the *in silico* protein predictions. Finally, those variants that are classified as probably neutral or VUS should also be studied with both methods but with much lower priority.

Besides using *in silico* tools, we have also analyzed family history, Amsterdam and Bethesda criteria, MSI and MMR expression in the tumor, and also the co-occurrence with other pathogenic mutations in the same patient. Three variants were found in co-occurrence with other pathogenic mutation in the same gene or in other MMR gene without a constitutional MMR deficient (CMMR-D) phenotype. The co-occurrence of a VUS along with a pathogenic mutation in the same gene without a CMMR-D phenotype is one way to reclassify a variant as neutral, thus making further *in vitro* characterization unnecessary. However, it is always necessary to know if the VUS and the pathogenic mutation are in *trans* in order to finally classify the variant as neutral, which has not been performed in these three cases due to lack of germline from the parents of the carriers.

Interestingly, *MLH1* had the lowest proportion of neutral variants (5%) and *MSH6* had the highest proportion (64%). These results can be in line with the fact that Amsterdam and Bethesda criteria works well to detect pathogenic mutations in *MSH2* and *MLH1*, but are insufficiently sensitive to detect pathogenic mutations in *MSH6* or *PMS2* (37). On the other hand, a high proportion of variants that were classified as probably/possibly damaging showed a lack of correlation with IHC (36%, 13/36). This can be explained by the fact that the protein can still be expressed but not functional as reported before (33).

Recently, InSiGHT has developed a set of criteria for the interpretation of MMR gene variants using a multifactorial Bayesian quantitative approach and/or on stringent combinations of qualitative clinical and functional data (7). In fact, to finally classify the VUS it is necessary to have different levels of evidence. In the majority of the cases, it is very difficult to obtain all these data. The final evidence is derived from *in vitro* analysis such as MMR assay, protein interaction, protein expression, and subcellular localization, being act of them accessible only to a minority of groups. Therefore, it is important to report all the new VUS with their clinical information and integrate them in databases to disseminate the information to the research community.

In order to reclassify novel VUS, a joint effort that allows integration of the clinical and research information is necessary. We are in a critical moment for research endeavors in the field of LS, as the US Congress has passed the 21st Century Cures Act providing funding to the Cancer Moonshot over 7 years and one of the demonstration projects is focusing specifically on cancer prevention among LS patients (38). In addition, chemo- and immune-prevention continues advancing, thus placing the development of novel tools to clarify the status of VUS in LS at the center of this important initiative (2). Clinicians will need to continue contributing with their knowledge regarding phenotype, family history, and tumor pathology and basic scientists contributing with the evidence obtained from the different RNA and protein analyses. In the interphase of these two groups, there are several research initiatives looking at systems biology tools to develop probabilistic models that improve our pathogenicity estimations because the reality is that *in vitro* functional validation of all VUS is a goal that is unlikely to be achieved. Only following with this collaborative pathway, we will be able to finally reclassify these variants and help LS patients with an accurate prediction of their cancer risk so that they and affected members of their family can receive appropriate cancer surveillance and chemopreventive interventions.

Disclosure of Potential Conflicts of Interest

M. Taggart has received speakers bureau honoraria from PeerView Institute for Education. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: E. Borras, M. Pande, P.M. Lynch, E. Vilar
Development of methodology: E. Borras, K. Chang, M. Pande, E. Vilar
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): E. Borras, M. Pande, A. Cuddy, S. Bannon, M.E. Mork, M. Taggart, P.M. Lynch, Y.N. You
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): E. Borras, K. Chang, M. Pande, Y.N. You, E. Vilar
Writing, review, and/or revision of the manuscript: E. Borras, K. Chang, M. Pande, A. Cuddy, J.L. Bosch, S. Bannon, M.E. Mork, M. Rodriguez-Bigas, M. Taggart, P.M. Lynch, Y.N. You, E. Vilar
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): M. Pande, A. Cuddy, M.E. Mork, P.M. Lynch, E. Vilar
Study supervision: E. Vilar

Acknowledgments

The authors thank the patients and their families for their participation.

Grant Support

This work was supported by a gift from the Feinberg Family to E. Vilar, the Janice Davis Gordon Memorial Postdoctoral Fellowship in Colorectal Cancer Prevention (Division of Cancer Prevention/MD Anderson Cancer Center) to E. Borras, Cancer Prevention Educational Award (R25T CA057730, NIH/NCI) to K. Chang, and grant K07CA160753 (NIH/NCI) to M. Pande.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 28, 2017; revised June 9, 2017; accepted July 26, 2017; published OnlineFirst August 1, 2017.

References

1. Yurgelun MB, Kulke MH, Fuchs CS, Allen BA, Uno H, Hornick JL, et al. Cancer susceptibility gene mutations in individuals with colorectal cancer. J Clin Oncol 2017;35:1086–95.
2. Spira A, Disis ML, Schiller JT, Vilar E, Rebbeck TR, Bejar R, et al. Leveraging premalignant biology for immune-based cancer prevention. Proc Natl Acad Sci U S A 2016;113:10750–8.
3. Ligtenberg MJ, Kuiper RP, Chan TL, Goossens M, Hebeda KM, Voorendt M, et al. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3' exons of TACSTD1. Nat Genet 2009;41:112–7.
4. Lynch HT, Snyder CL, Shaw TG, Heinen CD, Hitchins MP. Milestones of lynch syndrome: 1895–2015. Nat Rev Cancer 2015;15:181–94.
5. Giardiello FM, Allen JI, Axilbund JE, Boland CR, Burke CA, Burt RW, et al. Guidelines on genetic evaluation and management of Lynch syndrome: a consensus statement by the US multi-society task force on colorectal cancer. Am J Gastroenterol 2014;109:1159–79.
6. Pearlman R, Frankel WL, Swanson B, Zhao W, Yilmaz A, Miller K, et al. Prevalence and spectrum of germline cancer susceptibility gene mutations among patients with early-onset colorectal cancer. JAMA Oncol 2017; 3:464–71.
7. Thompson BA, Spurdle AB, Plazzer JP, Greenblatt MS, Akagi K, Al-Mulla F, 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;46:107–15.
8. Peltomaki P, Vasen H. Mutations associated with HNPCC predisposition – Update of ICG-HNPCC/INSiGHT mutation database. Disease Markers 2004;20:269–76.
9. Couch FJ, Rasmussen LJ, Hofstra R, Monteiro AN, Greenblatt MS, de Wind N, et al. Assessment of functional effects of unclassified genetic variants. Hum Mutat 2008;29:1314–26.
10. Hofstra RM, Spurdle AB, Eccles D, Foulkes WD, de Wind N, Hoogerbrugge N, et al. Tumor characteristics as an analytic tool for classifying genetic variants of uncertain clinical significance. Hum Mutat 2008;29:1292–303.
11. Spurdle AB, Couch FJ, Hogervorst FB, Radice P, Sinilnikova OM, Group IUGVW. Prediction and assessment of splicing alterations: implications for clinical testing. Hum Mutat 2008;29:1304–13.
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;29:1342–54.
13. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American college of medical genetics and genomics and the association for molecular pathology. Genet Med 2015;17:405–24.
14. Tavtigian SV, Greenblatt MS, Goldgar DE, Boffetta P, Group IUGVW. Assessing pathogenicity: overview of results from the IARC unclassified genetic variants working group. Hum Mutat 2008;29:1261–4.
15. Thompson BA, Goldgar DE, Paterson C, Clendenning M, Walters R, Arnold S, 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;34:200–9.
16. Thompson BA, Greenblatt MS, Vallee MP, Herkert JC, Tessereau C, Young EL, et al. Calibration of multiple in silico tools for predicting pathogenicity of mismatch repair gene missense substitutions. Hum Mutat 2013;34:255–65.
17. Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat 2008;29:1282–91.
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;116:1453–6.
19. Li H.A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 2011;27:2987–93.
20. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. J Computat Biol 1997;4:311–23.
21. Dogan RI, Getoor L, Wilbur WJ, Mount SM. SplicePort–an interactive splice-site analysis tool. Nucleic Acids Res 2007;35:W285–91.
22. Burset M, Seledtsov IA, Solovyev VV. SpliceDB: database of canonical and non-canonical mammalian splice sites. Nucleic Acids Res 2001;29:255–9.
23. Jian X, Boerwinkle E, Liu X. In silico tools for splicing defect prediction: a survey from the viewpoint of end users. Genet Med 2014;16:497–503.
24. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248–9.
25. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res 2003;31:3812–4.
26. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res 2011;39:e118.
27. Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GL, Edwards KJ, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum Mutat 2013; 34:57–65.
28. Stone EA, Sidow A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. Genome Res 2005;15:978–86.
29. Yuen ST, Chan TL, Ho JW, Chan AS, Chung LP, Lam PW, et al. Germline, somatic and epigenetic events underlying mismatch repair deficiency in colorectal and HNPCC-related cancers. Oncogene 2002;21:7585–92.
30. van der Klift HM, Mensenkamp AR, Drost M, Bik EC, Vos YJ, Gille HJ, et al. Comprehensive mutation analysis of PMS2 in a large cohort of probands suspected of lynch syndrome or constitutional mismatch repair deficiency syndrome. Hum Mutat 2016;37:1162–79.
31. Arnold S, Buchanan DD, Barker M, Jaskowski L, Walsh MD, Birney G, et al. Classifying MLH1 and MSH2 variants using bioinformatic prediction, splicing assays, segregation, and tumor characteristics. Hum Mutat 2009; 30:757–70.
32. Betz B, Theiss S, Aktas M, Konermann C, Goecke TO, Moslein G, et al. Comparative in silico analyses and experimental validation of novel splice site and missense mutations in the genes MLH1 and MSH2. J Cancer Res Clin Oncol 2010;136:123–34.
33. Borras E, Pineda M, Brieger A, Hinrichsen I, Gomez C, Navarro M, et al. Comprehensive functional assessment of MLH1 variants of unknown significance. Hum Mutat 2012;33:1576–88.
34. Jian X, Boerwinkle E, Liu X. In silico prediction of splice-altering single nucleotide variants in the human genome. Nucleic Acids Res 2014;42: 13534–44.
35. Tournier I, Vezain M, Martins A, Charbonnier F, Baert-Desurmont S, Olschwang S, et al. A large fraction of unclassified variants of the mismatch repair genes MLH1 and MSH2 is associated with splicing defects. Hum Mutat 2008;29:1412–24.
36. Kansikas M, Kariola R, Nystrom M. Verification of the three-step model in assessing the pathogenicity of mismatch repair gene variants. Hum Mutat 2011;32:107–15.
37. Sjursen W, Haukanes BI, Grindedal EM, Aarset H, Stormorken A, Engebretsen LF, et al. Current clinical criteria for Lynch syndrome are not sensitive enough to identify MSH6 mutation carriers. J Med Genet 2010;47:579–85.
38. Spira A, Yurgelun MB, Alexandrov L, Rao A, Bejar R, Polyak K, et al. Precancer atlas to drive precision prevention trials. Cancer Res 2017;77:1510–41.