

RESEARCH

Open Access



In silico splicing analysis of the *PMS2* gene: exploring alternative molecular mechanisms in *PMS2*-associated Lynch syndrome

Cătălin Vasile Munteanu^{1,3*}, Cătălin Marian^{4,5,7}, Adela Chiriță-Emandi^{2,3,6}, Maria Puiu^{2,3,6} and Adrian Pavel Trifa^{2,5,7,8}

Abstract

Lynch syndrome (LS) is one of the most common hereditary cancer syndrome in human populations, associated with germline variants in *MLH1*, *MSH2/EPCAM*, *MSH6* and *PMS2* genes. The advent of next generation sequencing has proven a significant impact in germline variant detection in the causative genes; however, a large proportion of patients with clinical criteria still receive uncertain or negative results. *PMS2* is the least frequent reported gene, associated with up to 15% of LS cases with late-onset disease and low penetrance phenotype; however, the proportion of *PMS2*-LS cases is considered to be highly underestimated. In this context, our analysis aimed to improve the current diagnostic yield by focusing on missense and intronic *PMS2* variants available in public clinical databases (ClinVar, LOVD). We performed an in silico assessment of the wild-type DNA sequence and the reported genetic variants, employing splicing bioinformatics tools known for their effectiveness in other genes. Splicing variants were predicted in silico and using GTEx short-read RNA expression data. Out of the 2384 missense variants discovered, 90% were classified with uncertain significance (VUS). 4.9% of missense variants were shown to have a potential splicing consequence ($DS > 0.2$) using SpliceAI. As described in the original publication, SpliceAI-visual was proven effective in annotation of short intronic variants (< 50 bp). Four short intronic variants were identified using SpliceAI-visual as potentially splicing disturbing, in spite of using a lower threshold ($DS > 0.1$). Exons 2, 3, 4, 5, 6, 7, 8, 11, 12 and 14 were consistently predicted in at least three out of eight software with weak canonical splice sites. Additionally, we noted that both Exonic Splicing Enhancers (ESEs) and Exonic Splicing Silencers (ESSs) contribute significantly to alternative splicing and exonic selection in *PMS2* gene. Specifically, ESE motifs were consistently more abundant in highly expressed exons 5, 11 and 14, while ESS motifs played a fundamental role in exons 6, 7 and 10. Computational analysis performed in our study serves as a valuable filtering step for guiding further RNA experiments. Additional functional data is necessary to validate our findings.

Keywords Lynch syndrome, Mismatch repair gene *PMS2*, Alternative splicing, RNA splicing, Splicing mutations, Exon skipping, Splicing regulatory elements

*Correspondence:
Cătălin Vasile Munteanu
catalin.munteanu@umft.ro

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Introduction

Lynch syndrome (LS, OMIM #120435, #609310, #614350, #614337, #613244) is one of the most prevalent cancer predisposition syndromes known in humans, with an estimated prevalence in the general population of 1 in 300. LS is characterized by pathogenic and likely pathogenic germline variants in DNA mismatch repair (MMR) genes – *MLH1*, *MSH2/EPCAM*, *MSH6* and *PMS2* – discovered by current molecular approaches [1–6]. Despite the latest advancements in high-throughput sequencing technologies, germline testing still continues to pose significant challenges in *PMS2*-associated Lynch syndrome (*PMS2*-LS) in current clinical practice. Pathogenic and likely pathogenic *PMS2* variants are reported in literature in up to 15% of all individuals diagnosed with Lynch syndrome, a recently acknowledged underestimation [1–6]. Supporting observations for this assertion include the high reported frequency of *PMS2* defects in individuals with Constitutional Mismatch Repair Deficiency (CMMRD) syndrome (OMIM #276300), a rare pediatric cancer syndrome [1, 3, 7]. In heterozygous state, *PMS2* pathogenic variants are associated with Lynch-related cancers, mainly colorectal and endometrial cancers [8]. However, the low penetrance phenotype and late-onset disease associated with heterozygous *PMS2* variants are contributing factors that could impede an accurate assessment of the true prevalence of healthy carriers in the population [5, 9, 10].

The PMS1 homolog 2, mismatch repair system component (*PMS2*, UniProt C9J167) represents an essential part in human DNA mismatch repair process. *PMS2* combines with *MLH1* in vivo to form the MutL α (*MLH1*-*PMS2*) heterodimer, which consist of N-terminal and C-terminal domains, as well as a linker region [11, 12], containing the Nuclear Localisation Signal (NLS) (Fig. 1) [13–15]. Currently, there is limited literature to describe how the MutL α complex is regulated. It is well established that *MLH1* loss of expression is associated with premature *PMS2* degradation. Additionally, several *MLH1* missense variants in the C-terminal region

have been linked to the loss of *PMS2* protein expression and mismatch repair deficiency (MMR-d) [16]. Previous research suggests that phosphorylation of *PMS2* in the C-terminal domain might play an important role, although specific amino acids involved in this process have not yet been identified [17].

Historically, the only exonic positions regarded as critical for RNA splicing were those corresponding to acceptor and donor canonical splice sites, specifically the first and last three nucleotides of the exon [18, 19]. Further research uncovered the significance of additional factors in exon recognition, such as splicing regulatory elements (SREs), chromatin structure, transcription rate, and the secondary and tertiary structure of the transcript [20–22]. With the recent development and implementation of in silico tools and deep learning-based algorithms [23, 24], many exonic variants formerly classified as missense, nonsense, or silent are susceptible to reclassification, with some revealing a predicted impact on splicing [24–28]. In this scenario, the pathogenic mechanism usually involve a mixture of transcripts with abnormal splicing patterns and transcripts carrying the causative variant [29–31].

The vast majority of *PMS2* variants documented in public databases to date are missense. Only a small proportion of these variants are, however, clinically significant, being classified as pathogenic (class 5) or likely pathogenic (class 4), according to ACMG 2015 guidelines [32–34]. Comparing to other LS-related genes, *PMS2* truncating variants, including nonsense, frameshift and splicing variants, are less frequently encountered in clinical databases [34, 35]. Nevertheless, the major mechanism of disease is loss of function (LOF), with the latest ClinGen expert group assessment supporting haploinsufficiency as the major driver of disease [36]. In other human diseases, exonic variations are linked to splicing alterations in as much as 25% of cases [37], primarily involving exon skipping, but in *PMS2*-associated LS this aspect remains largely unknown. Previous research indicates that alternative splicing is a common phenomenon during the natural processing of the *PMS2* gene [38].

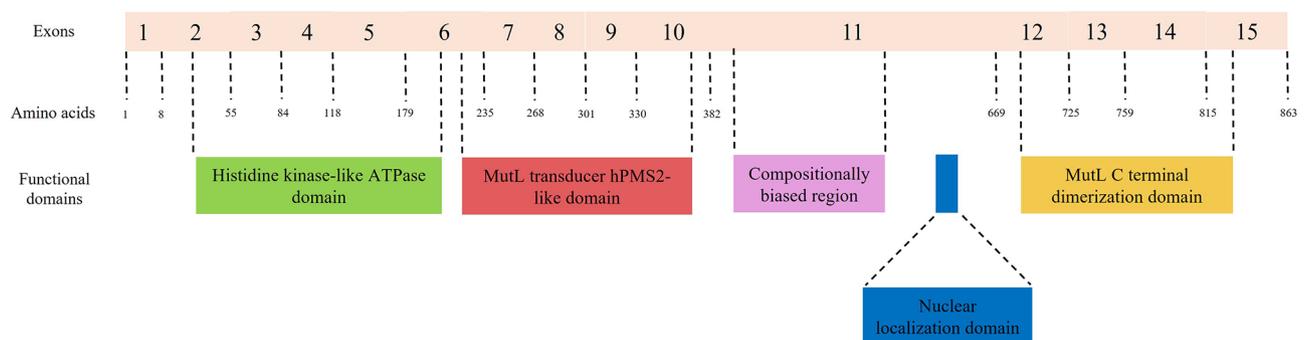


Fig. 1 The *PMS2* functional domains aligned to MANE transcript (NM_000535.7): Histidine kinase-like ATPase domain, MutL transducer hPMS2-like domain, Nuclear localization domain, MutL C terminal dimerisation domain

Limited data is available on splicing role in tissue-specific and developmental stage-dependent *PMS2* gene expression, as well as how variants altering splicing influence the protein function. Based on these observations, the main objective of our analysis was to narrow the existing diagnostic gap in *PMS2*-related LS by identifying potential significant splicing variants that could contribute to the underlying mechanisms of the disease. To assess the impact of splicing in silico, our analysis primarily focused on missense and short intronic germline variants related to Lynch syndrome and documented in public clinical databases. Bioinformatics tools, proven effective for other genes and previously employed in literature, were utilized to analyze both the wild-type DNA sequence and reported variants. To address the limitations of in silico tools and provide strength to our analysis, publicly available gene expression data was complementary employed for quantifying exon expression in a tissue-specific manner.

Results

Acceptor loss, the major mechanism for missense variants with predicted splicing impact in SpliceAI

Within the consulted databases (ClinVar and LOVD), a total of 2384 missense variants were documented (Fig. 2). Notably, the preponderant majority of these variants fell within the category of variants of uncertain significance (VUS), comprising 90.81% of the total. Variants with conflicting interpretations constituted 6.87%, while benign/likely benign variants accounted for 1.63%, and pathogenic/likely pathogenic variants were observed in 0.67%.

From all class 1–3 and conflicting missense variants (2384 variants) included in the study, 117 variants (4.90%) were anticipated to exhibit at least a mild effect on RNA splicing ($DS > 0.2$) and 34 (1.42%) a moderate or high impact ($DS > 0.5$) (Figure S1). Sixteen variants of uncertain significance were reported in 3' canonical splice sites (first exonic position), with 7 (43.75%) having a potential impact (5 of acceptor loss type and 2 of acceptor gain type). Thirty-eight variants of uncertain significance and one conflicting variant were reported in 5' canonical splice sites (last 3 exonic positions). Among these, 14 variants (35.89%) were expected to have an impact

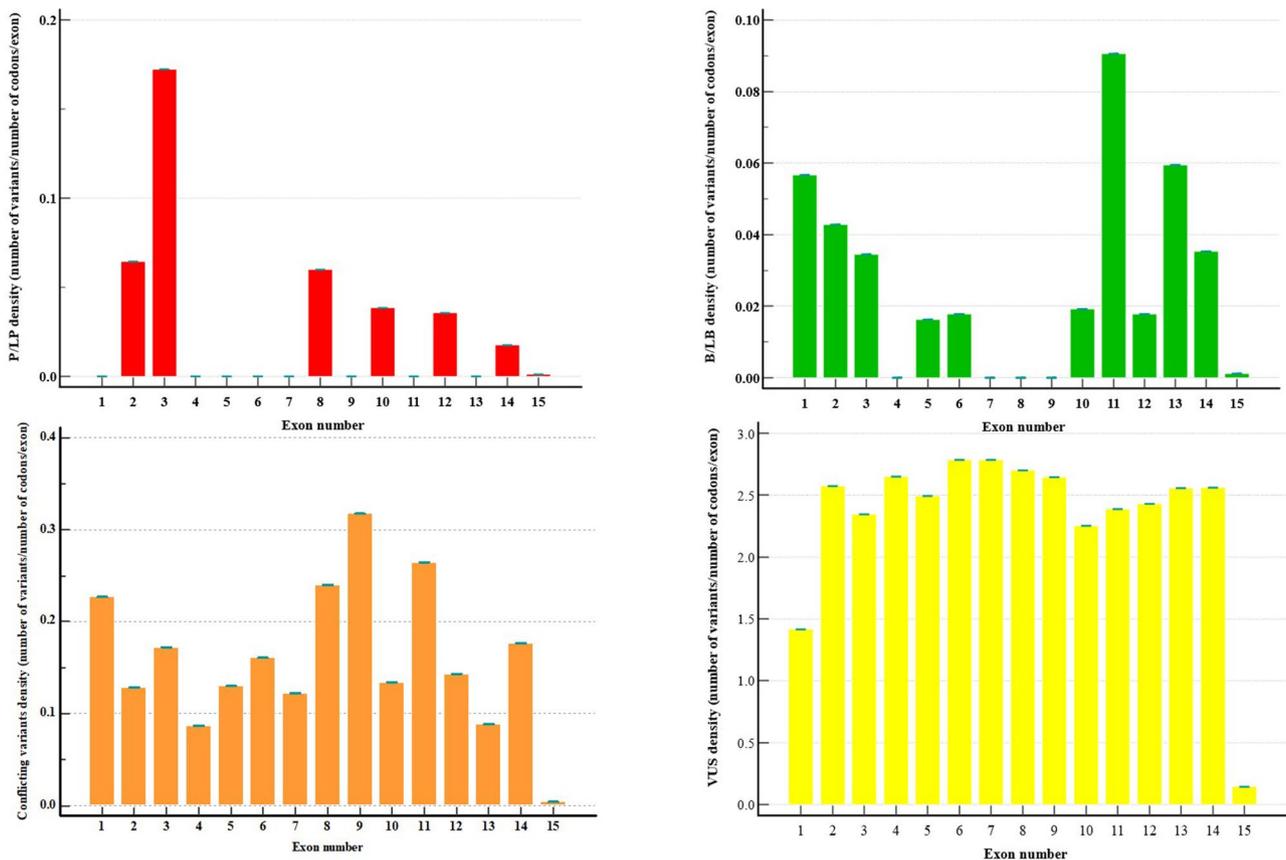


Fig. 2 Distribution of reported *PMS2* missense variants across exons, stratified depending on ACMG 2015 classification. Red (P/LP – pathogenic/likely pathogenic variants), green (B/LB – benign/likely benign variants), orange (Conflicting variants), yellow (VUS – variants of uncertain significance). Each bar signifies the calculated density of variants, depending on the exon length and total number of reported variants

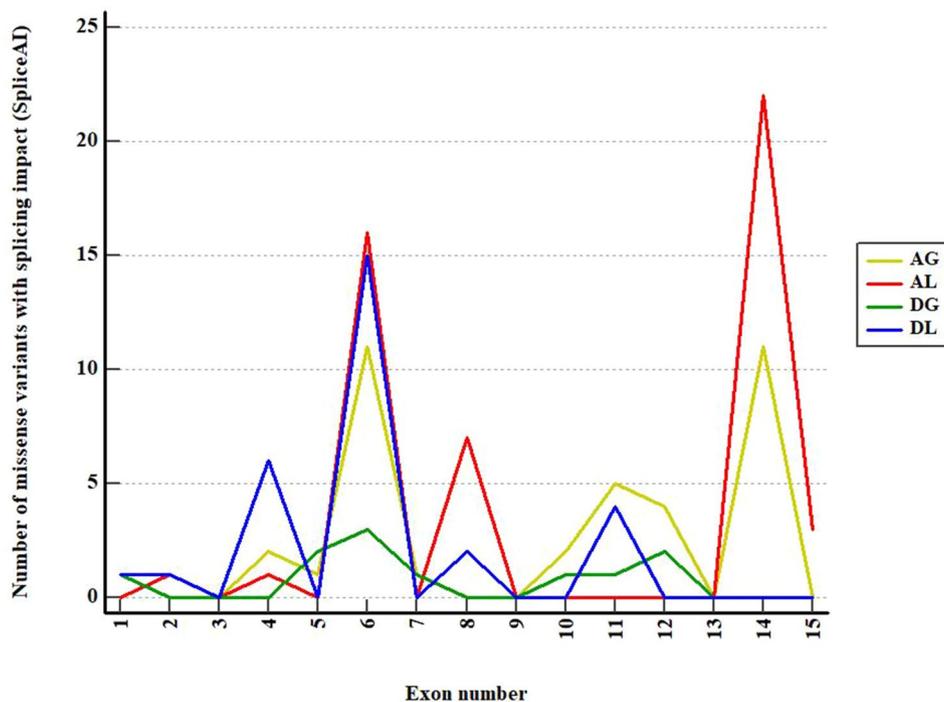


Fig. 3 Class 1–3 and conflicting *PMS2* missense variants with predicted splicing impact (DS > 0.2) in SpliceAI. Yellow (AG—Acceptor Gain), red (AL—Acceptor Loss), green (DG—Donor Gain), blue (DL—Donor Loss)

Table 1 *PMS2* exons harboring missense variants with predicted splicing impact (SpliceAI)

Predicted mechanism	Total number of splicing variants	Overrepresented exons	Mean (CI 95%)
Acceptor gain	38 (32.47%)	6, 11, 14	2.53 (0.45–4.61)
Acceptor loss	50 (42.73%)	6, 8, 14	3.33 (0–7.05)
Donor gain	11 (9.40%)	1, 5, 6, 11	0.40 (0–0.80)
Donor loss	29 (24.78%)	4, 6, 11	1.93 (0–4.16)

Overrepresented exons denote exons with a number of splicing variants above the 95% confidence interval. Percentages are relative to the total number of predicted splicing variants (117 variants with DS > 0.2)

of donor loss type. When categorized by predicted consequences, the exons most commonly linked to specific splicing mechanisms were as presented in Fig. 3; Table 1. Additionally, wild-type ESRseq scores (HEXoSplice in silico score calculated for enhancer splicing elements predicted in the wild type sequence) and Δ ESRseq scores (HEXoSplice in silico score calculated for enhancer splicing elements induced by genetic variants) corresponding to splicing variants (DS > 0.2) in exons enriched in acceptor loss (AL) and donor loss (DL) variants (Table 1) were compared to those from variants without such a prediction (Table 2, Figure S3).

Table 2 Wild-type ESRseq and Δ ESRseq scores comparison between predicted splicing variants (DS > 0.2) and variants with no predicted splicing impact (DS < 0.2) in exons enriched in acceptor loss (AL) and donor loss (DL) variants

Exon	ESRseq score wild type <i>p</i> -value	Δ ESRseq score <i>p</i> -value
Exon 4	0.7868	0.0598
Exon 6	0.5272	0.0013*
Exon 8	0.1429	0.5495
Exon 11	0.005	0.2413
Exon 14	0.0076	0.8120

Statistically significant results ($p < 0.05$, Mann-Whitney U test) are highlighted in bold. *Student's *t*-test

Exons 11 and 14 are enriched in missense variants with predicted splicing impact that are located in enhancer splicing elements. Exon 6 is enriched in missense variants with predicted splicing impact that modify the strength of enhancer splicing elements (increase or decrease the wild type ESRseq score)

SpliceAI-visual, a valuable prediction tool for *PMS2* complex short intronic variants

Out of 838 intronic variants reported in ClinVar, we discovered 71 non-point short genetic variants (< 50 bp). Upon filtering by clinical significance, 61 (85.91%) variants had conflicting interpretation or were interpreted as class 1–3 (benign, likely benign and VUS) according to ACMG 2015 criteria. Among these, 11 out of 61 variants affected the intronic canonical splice sites (the first 6 and last 3 intronic nucleotides). Notably, 14 variants were predicted to have a potential splicing impact when we used a lower threshold (DS < 0.2) for DS in conjunction

with SpliceAI-visual. To the best of our knowledge, at the time when this manuscript was written, with one exception, no clinical or functional data were available for the mentioned variants. Additional details regarding the selected variants are available in the Table S1.

Canonical splice site interpreted as novel splice site by SpliceAI

The variant of uncertain significance NM_000535.7:c.1970_2006+9dup is a 46 bp duplication that spans over the 3' end of exon 11 and exon-intron junction (Fig. 4A). In this particular case, relying solely on the DS provided by SpliceAI might suggest a mild increase in the strength of the canonical donor splice site.



Fig. 4 SpliceAI-visual predictions (IGV interface, MobiDetails) for *PMS2* short intronic variants: **(A)** NM_000535.7:c.1970_2006+9dup, **(B)** NM_000535.7:c.538-5_538-4del, **(C)** NM_000535.7:c.354-18_354-15dup. REF—reference (wild-type) score, ALT—alternative (variant) score. Vertical blue bars signify donor sites and orange bars signify acceptor sites

When the sequence was inspected using SpliceAI-visual, it was visible that SpliceAI located the original donor site in the duplicated region, showing only a slight decrease in strength (0.07). The canonical donor site was consequently interpreted as a novel splice site with DS score of 0.27. Upon further analysis, the duplicated region was anticipated to induce a frameshift and introduce a premature termination codon in the sequence. This event is expected to trigger transcript degradation via nonsense-mediated decay (NMD).

Variants increasing the strength of a weak canonical splice site

NM_000535.7:c.538-5_538-4del is a likely benign variant located in the proximity of acceptor site of exon 6 (Fig. 4B). As previously described, the 5' end of exon 6 corresponds to a relatively weak splice site (REF score: 0.67). Within the exonic sequence, there is at least one alternative cryptic acceptor splice site that is stronger than the canonical site (REF score: 0.81). While the variant DS (0.28) suggests a mild increase in strength of the acceptor splice site, the cumulative effect renders the canonical splice site stronger than cryptic sites (ALT score: 0.94). This, in turn, could consequently alter the natural proportion of $\Delta 6$ and $\Delta 6p$ transcripts with potential clinical impact. Similarly, variant NM_000535.7:c.538-12dup with conflicting interpretation of pathogenicity increases the strength of the same acceptor splice site (ALT score: 0.78), despite having only a mild DS (0.12).

Intronic inclusion and premature termination predicted by SpliceAI-visual

The variant NM_000535.7:c.354-18_354-15dup previously classified as likely benign, constitutes a 4 bp intronic duplication near the 3' splice site of exon 5 (Fig. 4C). SpliceAI indicates a mild acceptor gain effect (DS: 0.17) that could be easily filtered out using a standard DS cutoff value of 0.2. When using SpliceAI-visual, we observed that the variant enhances the strength of an intronic cryptic acceptor site (ALT score: 0.57). This site could compete with the canonical site and lead to intronic inclusion. Using the alternative splice site is important in this context since the variant induces a frameshift and includes a premature termination codon naturally present in the intronic sequence. This, consequently, is predicted to induce NMD and eventual transcript degradation.

Bioinformatics assessment of donor and acceptor splice sites strength

Canonical exon-intron junction motifs were evaluated using eight alternative software tools (Table 3, Figure S4). As was previously indicated [39], splice sites with scores falling below the lower boundary of the 90% confidence interval (90% CI) were deemed weak. Eleven exons displayed at least one putative weak splicing signal, with exons 2, 3, 4, 5, 6, 7, 8, 11, 12 and 14 consistently reported in at least three out of eight predictions. Exons 9, 10, 13, and 15, were lacking identifiable weak canonical splice sites.

Table 3 Summary of *PMS2* exons predicted to harbor strong and weak canonical splice sites

Software	Weak donor	Strong donor	Mean (CI 90%)	Weak acceptor	Strong acceptor	Mean (CI 90%)
ESEfinder 3.0	4, 5, 7, 8, 12, 14	1, 2, 9, 10, 11, 13	5.09 (2.83–7.36)	3, 5, 8, 11	4, 7, 9, 10, 13, 14	6.24 (4.18–8.30)
FSplice	4, 6, 8, 12	2, 9, 10, 13, 14	11.38 (9.38–13.38)	2, 3, 8, 14	7, 10, 12, 13	7.77 (6.20–9.33)
MaxEntScan	4, 6, 8	2, 9, 10, 13, 14	8.67 (7.60–9.74)	2, 3, 8, 14	6, 7, 10, 13	8.17 (7.06–9.27)
NetGene2	4, 7, 8	3, 9, 10, 12, 13, 14	0.68 (0.50–0.86)	3, 5, 8, 11, 14	4, 6, 9, 12, 13, 15	0.51 (0.35–0.67)
NNSplice	4, 6	1, 2, 3, 9, 10, 13, 14	0.88 (0.79–0.96)	2, 8, 15	7, 9, 10, 13, 14	0.66 (0.50–0.83)
SpliceAI	4, 6, 8	1, 3, 5, 9, 10, 12, 13, 14	0.94 (0.91–0.98)	4, 6, 8	2, 3, 5, 9, 10, 11, 12, 13, 14	0.91 (0.84–0.97)
Spliceator	4, 7, 12	1, 2, 3, 5, 6, 8, 9, 10, 11, 13, 14	0.78 (0.58–0.98)	5, 8, 11	2, 3, 4, 6, 7, 9, 10, 12, 13, 14	0.77 (0.57–0.97)
SpliceRover	1, 4	3, 5, 8, 10, 11, 13, 14	0.82 (0.69–0.95)	8	-	0.96 (0.93–1)

Weak splice sites were defined as those with scores below the lower bound of the 90% confidence interval (90% CI), while strong splice sites had scores above the upper bound

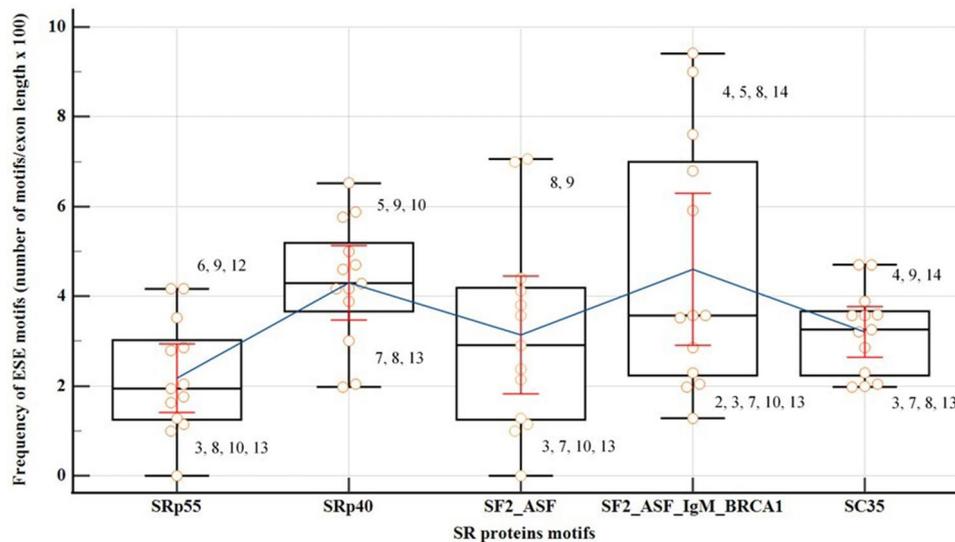


Fig. 5 Exonic density of SR proteins motifs identified using ESEfinder 3.0 software. Red lines represent the 95% CI for means. Exons outside the interval are depleted (below the lower bound of 95% CI) or enriched (above the upper bound of 95% CI) in SR protein motifs and were noted in the graph

Table 4 *PMS2* exons predicted to be prone to exon skipping, based on ESEs and ESSs densities and ESS/ESE ratio

Software	Low ESEs density	Mean (CI 90%)	High ESSs density	Mean (CI 90%)	High ESS/ESE ratio	Mean (CI 90%)
ESEfinder 3.0	3, 7, 10, 13	17.23 (14.26–20.19)	-	-	-	-
HExoSplice	2, 3, 9, 10	21.96 (17.91–26.01)	2, 6, 7, 10	26.65 (21.63–31.67)	2, 6, 9, 10	119.86 (89.60–150.11)
HOT-SKIP	8, 9	80.26 (75.13–85.39)	6, 7, 10, 13	57.72 (52.26–63.18)	6, 7, 9, 10, 13	74.03 (63.67–84.39)

Exonic splicing regulatory elements (SREs) predictions: ESEs, ESSs and ESS/ESE ratio

The coding regions of all *PMS2* exons, except from the first and last exons, were assessed utilizing motif matrices in ESEfinder to detect potential binding sites for SR proteins: SF2/ASF, SF2/ASF (IgM-BRCA1), SC35, SRp40 and SRp55 (Fig. 5, Table S3). Significance in the analysis was attributed solely to motifs with high scores surpassing the standard thresholds recommended by developers. Exons 3, 7, 10, and 13 exhibited the lowest overall density of ESE motifs, with an isolated overrepresentation of binding sites for SRp40 in exon 10. The findings from the parallel analysis, employing both HOT-SKIP and HExoSplice, are comprehensively outlined in Table 4 and Table S4.

High density/ratio values are considered those above the upper bound of 90% CI, whereas low density/ratio values are below the same interval.

Exons displaying consistently high predictions for ESE (exons 5, 11, 14) or ESS (exons 6, 7, 10) densities in all three SREs-specific software (Table S4), were further examined based on their wild-type ESRseq scores returned by HExoSplice, in order to explore motifs distribution across exons (Fig. 6, Table S5). Correlation between exonic position and ESRseq score in exon 5

($R = -0.34$, $p = 0.001$) and exon 14 ($R = 0.30$, $p = 0.006$) was weak-moderate and statistically significant. In exon 5, ESEs were located preferentially in the first 50 bp, while ESSs were found in the latter half of the exon. This pattern was reversed in exon 14.

PMS2 expression data—GTEx database and RefSeq coding transcripts

The available data regarding *PMS2* gene expression revealed that exons 11, 13, and 14 had a significantly higher median count per base in all tissues included in the study (Table 5, Table S6). On the other hand, exons 1, 2, and 15 consistently exhibited low expression in almost all tissues. 60 protein coding transcripts were included in the study (Table S7, Figure S5). The most frequent splicing variants in studied transcripts were presented in Fig. 7.

Discussion

Low level of exonic splicing variants in *PMS2* predicted by SpliceAI

In our study, only 4.9% of *PMS2* missense variants were predicted to have a significant splicing consequence ($DS > 0.2$) using SpliceAI. This value is lower than

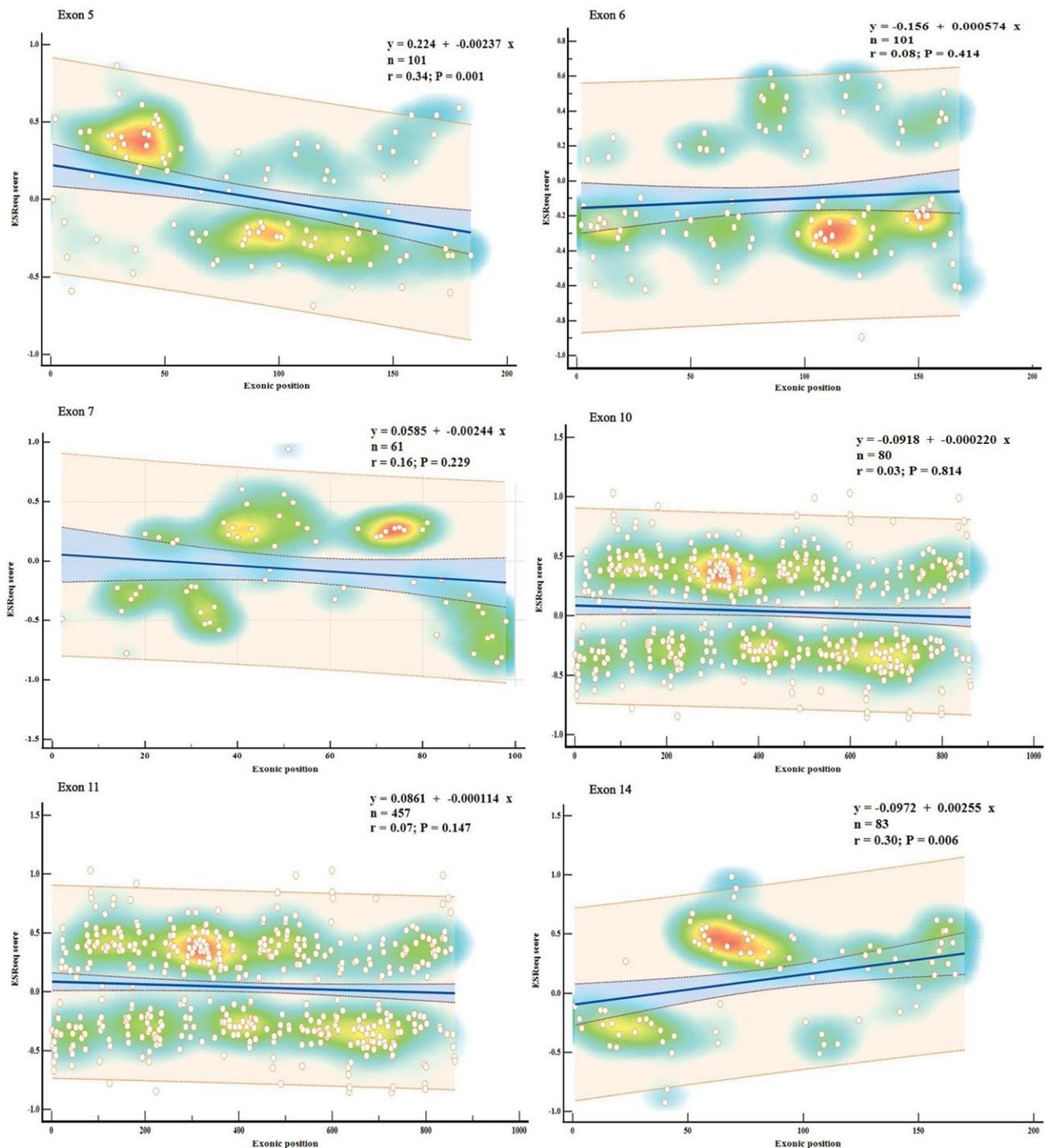


Fig. 6 Exonic distribution (HEXoSplice) of SRE motifs in exons with high density of ESEs or ESSs (exons 5, 6, 7, 10, 11 and 14). ESRseq scores > 0 correspond to ESE motifs and ESRseq scores < 0 represent ESS motifs. The red color signifies the high density of SRE motifs in a genomic region

expected given the data available in literature which states that up to 25% of exonic disease-causing variants may disturb exon definition in mature transcripts [37]. There are several potential explanations for this result: (1) A lower performance of bioinformatics predictions in exonic positions outside the canonical splice sites

[19, 40]; (2) Variants situated deep in the exonic regions are less likely to influence exon inclusion [41]; (3) The DS threshold > 0.2 could impact variant classification in *PMS2* gene, clinically significant variants being reportedly overseen in other genes when a standard cutoff value was used [24]; (4) Alternative splicing have a secondary

Table 5 Mean RNA expression (median count/base) of *PMS2* exons in Lynch syndrome-related tissues

Tissue type	Overex-pressed exons	Underex-pressed exons	Mean (CI 95%)
Bladder	11, 13, 14	1, 2, 15	0.1937 (0.1502–0.2372)
Brain—cortex	11, 13, 14	1, 2, 15	0.2263 (0.1767–0.2760)
Breast—mammary tissue	11, 13, 14	1, 2, 15	0.1883 (0.1536–0.2230)
Colon—transverse	11, 13, 14	1, 2, 15	0.1392 (0.1114–0.1671)
Colon—sigmoid	11, 13, 14	1, 2, 15	0.1516 (0.1195–0.1837)
Kidney—medulla	10, 11, 13, 14	1, 2, 3, 5	0.1818 (0.1309–0.2327)
Ovary	10, 11, 13, 14	1, 2, 15	0.1728 (0.1371–0.2085)
Pancreas	11, 13, 14	1, 2, 15	0.0618 (0.0501–0.0735)
Prostate	11, 13, 14	1, 2, 4	0.1585 (0.1163–0.2007)
Skin—sun exposed	10, 11, 13, 14	1, 2, 15	0.2941 (0.2360–0.3522)
Stomach	11, 13, 14	1, 2, 15	0.10568 (0.0842–0.1270)
Uterus	11, 13, 14	1, 2, 15	0.1714 (0.1363–0.2065)
Whole blood	1, 11, 13, 14	2, 3, 4, 5, 7	0.0525 (0.0404–0.0646)

Overexpressed exons denote exons with mean expression above the 95% CI. Underexpressed exons denote exons with mean expression below the 95% CI

role in the biology of *PMS2* gene, according to what was previously reported in literature [42].

Exons harboring weak canonical splice sites may require regulatory elements for proper definition

Critical for exon definition, canonical splice sites may demand additional signals, such as ESEs or ESSs, when their sequence is less conserved [18, 23]. Not surprisingly, in the past, other analysis of variants that disrupt SREs supported the same idea of SREs exhibiting activity predominantly in exons having splice sites of weak or moderate consensus [40]. In accordance with these observations, in our study, exons 2, 4, 5, 6, 8, 12, 14 predicted to have weak splice sites exhibited a high frequency of at least one ESE motif in ESEfinder. Exon 3, however, although predicted with a weak 3' splice site, had a lower than average level of ESEs in ESEfinder and HExoSplice, but a significantly higher one in HOT-SKIP, suggesting a potential for alternative splicing events [43]. Notably, the expression of exon 3 was below average in all tissues, with whole blood and medullary kidney showing significantly low expression levels. In contrast, the median read count per base in bladder tissue was above the average. When taken globally, in all exons with a weak 5'ss or 3'ss, except for exon 4, ESE or ESS motifs were overrepresented in at least one prediction. Nevertheless, in exon 4, ESEfinder detected a high density of SC35 and SF2/ASF (IgM-BRCA1) specific domains, but the overall ESEs density was above the average, reaching the upper bound of the confidence interval.

PMS2 exons with weak splice sites may be prone to exon skipping

As indicated by the low ESEs frequency, high ESSs frequency, or high ESS/ESR ratio [44, 45], exons 2, 3, 6, 7 and 8 were predicted by at least one SRE software as

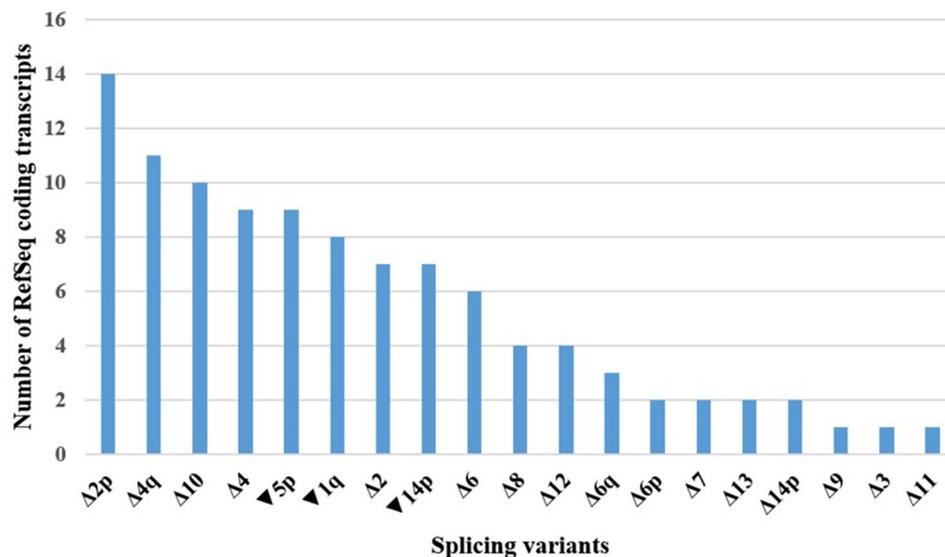


Fig. 7 Splicing variants encountered in *PMS2* RefSeq coding transcripts. Δ—complete or partial (adjacent to canonical splice sites) exonic deletion, ▼—intronic sequence inclusion (adjacent to canonical splice sites), p—acceptor site shift, q—donor site shift

being potentially prone to exon skipping. This association is plausible, given similar in silico observations reported in the literature for other genes [41], which were subsequently confirmed in functional studies [39]. Indeed, when consulting transcripts reported to date, high number of wild type and mutant isoforms aligned with this prediction [42, 46]. Bulk RNA-seq data was concordant over exons 2, 3 and 7, with the mention of variations in exon 3 and 7 that were tissue-dependent. On the other hand, exons 6 and 8 displayed an average median read count per base across all tissues, as indicated by GTEx database. Moreover, exon 8 was shown to harbor higher than average levels of SF2/ASF sites, but the overall density of SR proteins-binding domains remained low. SpliceAI was concordant with canonical splice site software regarding exon 8, indicating a high density of acceptor loss variants, which may serve as an indirect proof of the same biological event. Indeed, $\Delta 8_{11}$ transcript, naturally occurring by a splice acceptor shift mechanism, was reported [42]. Other similar ‘multi-cassette’ RefSeq transcripts were curated, suggesting once again the role of alternative splicing in processing of this exon. Intriguingly, exon 4, also featuring a weak donor site, undergoes alternative splicing in certain transcripts, while exons 5 and 14 utilizes distinct acceptor sites [46], an aspect also foreseen by SpliceAI-visual (Figure S2), which identifies strong cryptic acceptor sites. Recently, several exonic and intronic variants associated with exon 4 skipping were reported in literature, highlighting even more the relevance of alternative splicing in *PMS2*-related LS [40, 46, 47].

Other exons potentially sensitive to skipping with no weak consensus were 9, 10 and 13. Exon 10 was mentioned by all 3 tools, with a low ESEs incidence in ESEfinder and high ESSs levels in HOT-SKIP and HExoSplice. On the other hand, exon 10 is one of the most highly expressed exons across all tissues. Of note, several protein-coding RefSeq transcripts revealed, however, an isolated exclusion of this exon. Minor $\Delta 10$ transcripts were also evident in vitro assays [46]. Exon 9 is sometimes skipped in a ‘multi-cassette’ event with other exons. However, from our knowledge, there are no natural transcripts where exon 9 is skipped individually. In this case, ESRs predictions were conflictual, with ESEfinder indicating a high density of ESE motifs, while the other tools supported a low level. Similarly, exon 13 is a constitutive exon, being included and highly expressed in the majority of transcripts. This comes as no surprise, given that it encodes the MutL C terminal dimerization domain, a region known for its high conservation in the protein [48, 49].

High ESE levels concordantly predicted in exons 5, 11 and 14 – critical for *PMS2* function

The vast majority of ESE domains in exon 5 are located next to the acceptor site in HExoSplice. Interestingly, an alternative 3'ss located proximally in the intronic sequence (and predicted by SpliceAI-visual—Figure S2) was observed in at least one naturally occurring transcript. This distribution prompts questions regarding the role of ESEs in the region, particularly considering that 3 out of 8 predictions indicated a 3' weak splice site [50]. Of note, exons 2–5 codify N-terminal ATPase domain of *PMS2* (*HATPase_c_3*, InterPro, PF08676), which is important for maintaining mismatch repair proficiency. However, it is worth highlighting that this domain may be working in an asymmetric manner with the similar domain in MLH1, with the later appearing to play a more decisive role in this biological process [51–53].

By comparison, in exon 11, ESEs and ESSs hexamers are evenly distributed throughout the exon, with a region with high ESSs density and a relative ESEs depletion near the 3'ss. Exon 11, being a long exon, suggests that ESEs may play a fundamental role in facilitating the accurate selection of exonic boundaries by the splicing machinery [43]. Additionally, certain isoforms include a shorter variant of this exon due to a shift in the 5' splice site, in contrast with the presence of a relatively strong 5' donor consensus. Similarly, SpliceAI predicted a significant number of variants in exon 11 that could lead to donor site gain or loss. At least one transcript, $\Delta 11q_{14p}$, has been reported to occur via splice donor shift process [42]. An analogous event was noticed in atypical CMMRD related to a benign missense founder variant that created a novel donor splice site in intron 11, generating a 5 bp deletion frameshift at the exon 11–12 junction [54, 55]. Remarkably, exon 11 is rarely spliced out in reported transcripts, which indicates a fundamental biological significance of the regional coding sequence. Upon delving into the GTEx exon expression data, we observed a consistent trend of high expression for exon 11 across all investigated tissues. Upon reviewing the literature, we discovered that *PMS2* interacts with MLH1 in a critical area spanning amino acid residues 675 to 850 [11]. This sequence defines the C-terminus of the protein, a region sensitive to phosphorylation and involved in regulating the degradation of *PMS2*, that overlaps with exons 12–15 [17]. Additionally, missense variants within exon 11, located in close proximity to this hotspot region, may impact MutL α heterodimer formation by altering the binding affinity of the protomers and nuclear localization domain [56].

Within exon 14, ESSs are primarily located at the 5' end (3' acceptor site), whereas ESEs are complementarily positioned in the middle and 3' end (5' donor site). Since exon 14 was consistently predicted to have a weak

acceptor site, and confirmed by wild type and mutant $\Delta 14p$ transcripts [46], the high ESEs incidence may be relevant for its inclusion in mature transcripts, as it was reported in other exons with similar properties [43]. As mentioned earlier, exon 14 encodes the MutL C terminal dimerisation domain (*MutL_C*, InterPro, IPR014790) and, thus, critically contributes to MutL α heterodimerization [48, 49]. As anticipated, the RNA expression analysis revealed, once again, a consistent pattern of elevated expression across all considered tissues.

High ESS levels concordantly predicted in exons 6, 7 and 10

Both HOTS-SKIP and HExoSplice indicated potential elevated levels of exonic splicing silencers in exons 6, 7 and 10. When we visually inspected exon 6 in HExoSplice, ESSs were strikingly located in the 5' half of the exon. In combination with ESE depletion in the same region, it could explain, at least partially, the acceptor splice shift, $\Delta 6p$ and $\Delta 6$ transcripts described in literature [42, 46] and present in RefSeq data. SpliceAI-visual identified at least 2 exonic cryptic acceptor sites in the region (Figure S2) that may play a decisive role in the process. In line with this, SpliceAI predicted a significant number of acceptor gain and acceptor loss variants in exon 6 based on the variants collected in this study. Moreover, the median Δ ESRseq score of predicted splicing variants was negative and significantly lower from other reported variants in this exon. In exon 7, ESSs are located preferentially at both 5' and 3' ends of the exon, middle of the exon being enriched in ESEs instead. Besides the predicted weak donor site, we suspect that this particular distribution of SREs may provide a possible explanation for the observed exon 7 skipping and intron 7 inclusion in some transcripts [42]. In exon 10, ESSs are relatively evenly distributed, interspersed with ESEs, showing no specific pattern.

Limitations

The present study conducted a bioinformatics evaluation of the significance of alternative splicing in the expression of the *PMS2* gene. This analysis is relevant given the numerous *PMS2* variants of uncertain significance reported, that complicate the translation of DNA sequencing data into clinical practice. However, we acknowledge that the present study has several limitations, primarily related to the fact that the analysis was conducted exclusively *in silico*. Despite including bioinformatics tools commonly used and verified by other authors, future experimental data are still required to validate the results. Additionally, the included expression data may exhibit certain known biases [57, 58]. The analyzed transcripts are solely protein-coding, excluding non-coding transcripts, which may not fully capture the diversity and complexity of alternative splicing.

Furthermore, while the primary focus of the current study was represented by the *PMS2* coding variants, the potential importance attributed to noncoding variants and antisense transcripts in exon selection should not be disregarded, warranting further investigation.

Materials and methods

Reference sequence and variant nomenclature

PMS2 variants were described according to Human Genetic Variation Society guidelines (<https://hgvs-nomenclature.org/stable/>). MANE Select transcript (NM_000535.7, ENST00000265849.12) was considered the reference sequence, position c.1 corresponding to the first coding nucleotide. The variants analyzed were retrieved from the public databases ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar>) and LOVD (<https://databases.lod.nl/shared/genes/PMS2>) on 4 December 2023. These are the most comprehensive clinical genomic databases currently available for free. Variants with no reported classification based on ACMG criteria [32] have been excluded from the analysis. Variant annotation was performed using Ensembl Variant Effect Predictor (VEP) (https://ensembl.org/Homo_sapiens/Tools/VEP/). For exonic regions, we focused on nonsynonymous missense variants. Truncating variants, known to be deleterious, and synonymous variants, very rarely reported in consulted databases, were excluded. Among the selected intronic variants, we specifically included short genetic variants (<50 base pairs). Complex and large variants were beyond the scope of our study.

Tissue-specific bulk RNA expression data and public available transcripts

Protein-coding *PMS2* transcripts under analysis were retrieved from NCBI RNA reference sequences collection (RefSeq), a public database that provides an extensive and carefully curated collection of sequences [59, 60]. RefSeq transcripts available on UCSC genome browser (<https://genome.ucsc.edu>) were included in the analysis. To further enhance our understanding, the GTEx database was employed for quantifying exon expression using median read per base provided in a tissue-specific manner [58]. The analysis focused on tissues more commonly affected in Lynch syndrome.

Bioinformatics analysis of splicing impact and statistical analysis

Several bioinformatics approaches were employed to assess the impact of *PMS2* variants on RNA splicing. The strength analysis of canonical acceptor and donor splice sites was conducted using freely available online tools, including ESEfinder 3.0 for splice sites (<https://esefinder.auc.hc.umn.edu/>), FSplice (<http://www.softberry.com/>), MaxEntScan (<http://hollywood.mit.edu/>), NetGene2 (

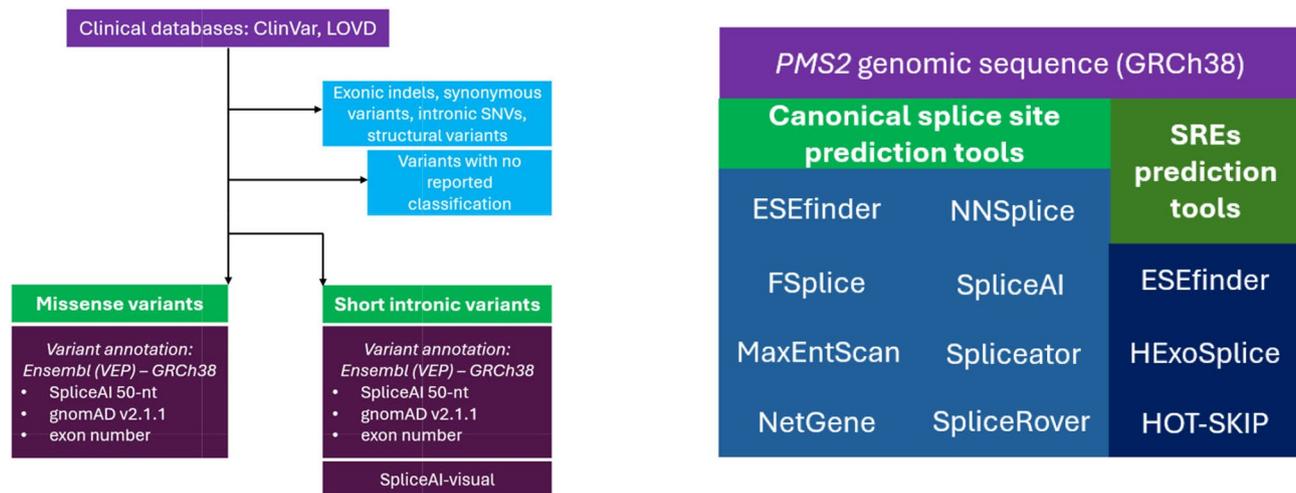


Fig. 8 An overview for the methods used in this study: **a**) data extraction from clinical databases – ClinVar and LOVD, and variant annotation using Ensembl (VEP – variant effect predictor) **b**) splicing predictions tools employed for canonical splice sites and SREs predictions

services.healthtech.dtu.dk/) and NNSplice (https://www.fruitfly.org/seq_tools/splice.html), SpliceAI [23, 24, 61], Spliceator (<https://www.lbgi.fr/spliceator/>), SpliceRover (<http://bioit2.irc.ugent.be/rover/splicerover>). The potential splicing impact of reported variants was estimated using SpliceAI, one of the most proficient deep learning-based tool reported to date [23, 24, 61]. Delta scores (DS) greater than 0.2 were used to screen for splice-altering variants, as outlined in the original publication [23]. To enhance the characterization of intronic variants, we utilized the SpliceAI-visual IGV interface [62] accessible on MobiDetails (<https://mobidetails.iurc.montp.inserm.fr>). Moreover, SpliceAI-visual was used to predict the presence of cryptic splice sites within the wild type sequence. Splicing regulatory elements (SREs) were screened across the coding regions using three alternative strategies: ESEfinder 3.0 for SR proteins (<https://esefinder.ahc.umn.edu/>), HOT-SKIP (<https://hot-skip.img.cas.cz/>) and HExoSplice (<http://bioinfo.univ-rouen.fr/HExoSplice/>). Exon skipping was considered in either case based on the resulting low putative ESEs and high ESSs densities, as well as increased ESS/ESE ratios [44, 45]. All splicing predictions were done using default specifications for each in silico tool (Table S8). A brief flowchart for the methods used in this study is shown in Fig. 8. Statistical analysis was performed using MedCalc® Statistical Software version 22.019 (MedCalc Software Ltd, Ostend, Belgium; <https://www.medcalc.org>; 2024). Statistical significance was assigned for p -value < 0.05. Except where otherwise stated, a 95% confidence interval (CI) was employed. The Student's t-test was used to statistically evaluate numerical and normally distributed values. The Mann-Whitney U test was used as a non-parametric test in non-normally distributed data.

Conclusions

This study highlights the importance of in silico tools to further guide the *PMS2* mRNA analysis and potentially increase the diagnostic yield in Lynch syndrome. Bioinformatics software used underlines the frequency of splicing alterations associated to *PMS2* gene, providing a possible explanation for the current underdiagnosis in *PMS2*-associated LS. In this regard, we identified several missense and short intronic variants that were predicted to have a splicing impact, making them good candidates for future functional analysis. Exons 2, 3, 4, 5, 6, 7, 8, 11, 12 and 14 were shown to have weak canonical splice sites in at least three prediction tools. Moreover, we observed the importance of ESEs and ESSs in the *PMS2* gene, their density and localization across the sequence, providing valuable input about splice site selection and overall exon expression. ESE motifs were more prevalent in highly expressed exons 5, 11 and 14, while ESS motifs were overrepresented in exons 6, 7 and 10. Some limitations of in silico tools emerged, however; therefore, additional functional data are necessary to evaluate the biological significance of our computational observations.

Abbreviations

AG	Acceptor gain
AL	Acceptor loss
CI	confidence interval
CMRD	Constitutional Mismatch Repair Deficiency
DG	Donor gain
DL	Donor loss
DS	Delta score
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
LS	Lynch syndrome
MMR-d	Mismatch repair deficiency
NLS	Nuclear Localisation Signal
PMS2	PMS1 homolog 2, mismatch repair system component
SRE	Splicing regulatory element

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12863-024-01281-3>.

Supplementary Material 1

Acknowledgements

Not applicable.

Author contributions

Data extraction and bioinformatics analysis, C.M. and A.T.; writing—original draft preparation, C.M. and A.T.; writing—review and editing, A.C.-E., C.M., M.P.; All authors have read and agreed to the published version of the manuscript.

Funding

We would like to acknowledge *Victor Babeş University Of Medicine And Pharmacy Timisoara* for their support in covering the costs of publication for this research paper.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Doctoral School, Victor Babeş University of Medicine and Pharmacy, 2 Eftimie Murgu Square Street, Timișoara 300041, Romania

²Department of Microscopic Morphology, Genetics Discipline, Victor Babeş University of Medicine and Pharmacy, 2 Eftimie Murgu Square Street, Timișoara 300041, Romania

³Regional Center of Medical Genetics Timiș, Louis Țurcanu Clinical Emergency Hospital for Children, 2 Iosif Nemoianu Street, Timișoara 300011, Romania

⁴Department of Biochemistry and Pharmacology, Biochemistry Discipline, Victor Babeş University of Medicine and Pharmacy, 2 Eftimie Murgu Square Street, Timișoara 300041, Romania

⁵Center for Research and Innovation in Personalized Medicine of Respiratory Diseases, Victor Babeş University of Medicine and Pharmacy, 2 Eftimie Murgu Square Street, Timișoara 300041, Romania

⁶Center for Genomic Medicine, Victor Babeş University of Medicine and Pharmacy, 2 Eftimie Murgu Square Street, Timișoara 300041, Romania

⁷Center of Expertise on Rare Pulmonary Diseases, Victor Babeş Clinical Hospital of Infectious Diseases and Pneumophysiology, 13 Gheorghe Adam Street, Timișoara 300310, Romania

⁸Breast Cancer Center, The Oncology Institute "Prof. Dr. Ion Chiricuta", 34-36 Republicii Street, Cluj-Napoca 400015, Romania

Received: 1 July 2024 / Accepted: 14 November 2024

Published online: 27 November 2024

References

- Wimmer K, Etzler J. Constitutional mismatch repair-deficiency syndrome: have we so far seen only the tip of an iceberg? *Hum Genet.* 2008;124:105–22. <https://doi.org/10.1007/s00439-008-0542-4>.
- Palomaki GE, McClain MR, Melillo S, Hampel HL, Thibodeau SN. EGAPP supplementary evidence review: DNA testing strategies aimed at reducing morbidity and mortality from Lynch syndrome. *Genet Med.* 2009;11:42. <https://doi.org/10.1097/GIM.0b013e31818fa2db>.
- Goodenberger ML, Thomas BC, Riegert-Johnson D, Boland CR, Plon SE, Clendenning M, et al. PMS2 monoallelic mutation carriers: the known unknown. *Genet Med.* 2016;18:13–9. <https://doi.org/10.1038/gim.2015.27>.
- Vaughn CP, Baker CL, Samowitz WS, Swensen JJ. The frequency of previously undetectable deletions involving 3' exons of the PMS2 gene. *Genes Chromosomes Cancer.* 2013;52:107–12. <https://doi.org/10.1002/gcc.22011>.
- Senter L, Clendenning M, Sotamaa K, Hampel H, Green J, Potter JD, et al. The clinical phenotype of Lynch syndrome due to germ-line PMS2 mutations. *Gastroenterology.* 2008;135. <https://doi.org/10.1053/j.gastro.2008.04.026>.
- Yuan L, Chi Y, Chen W, Chen X, Wei P, Sheng W et al. Immunohistochemistry and microsatellite instability analysis in molecular subtyping of colorectal carcinoma based on mismatch repair competency. *Int J Clin Exp Med.* 2015, 8, 20988. /pmc/articles/PMC4723875/.
- Wimmer K, Kratz CP, Vasen HFA, Caron O, Colas C, Entz-Werle N, et al. Diagnostic criteria for constitutional mismatch repair deficiency syndrome: suggestions of the European consortium 'Care for CMMRD' (C4CMMRD). *J Med Genet.* 2014;51:355–65. <https://doi.org/10.1136/jmedgenet-2014-102284>.
- Andini KD, Nielsen M, Suerink M, Helderman NC, Koornstra JJ, Ahadova A, et al. PMS2-associated Lynch syndrome: past, present and future. *Front Oncol.* 2023;13:1–12. <https://doi.org/10.3389/fonc.2023.1127329>.
- Dominguez-Valentin M, Sampson JR, Seppälä TT, ten Broeke SW, Plazzer JP, Nakken S, et al. Cancer risks by gene, age, and gender in 6350 carriers of pathogenic mismatch repair variants: findings from the prospective Lynch Syndrome Database. *Genet Med.* 2020;22:15–25. <https://doi.org/10.1038/s41436-019-0596-9>.
- ten Broeke SW, Suerink M, Nielsen M, Response to, Roberts et al. 2018: is breast cancer truly caused by MSH6 and PMS2 variants or is it simply due to a high prevalence of these variants in the population? *Genet Med.* 2019, 21, 256–7. <https://doi.org/10.1038/s41436-018-0029-1>
- Guerrette S, Acharya S, Fishel R. The interaction of the human MutL homologues in hereditary nonpolyposis colon cancer. *J Biol Chem.* 1999;274:6336–41. <https://doi.org/10.1074/jbc.274.10.6336>.
- Kondo E, Horii A, Fukushige S. The interacting domains of three MutL heterodimers in man: hMLH1 interacts with 36 homologous amino acid residues within hMLH3, hPMS1 and hPMS2. *Nucleic Acids Res.* 2001;29:1695. <https://doi.org/10.1093/nar/29.8.1695>.
- de Barros AC, Takeda AAS, Dreyer TR, Velazquez-Campoy A, Kobe B, Fontes MRM. DNA mismatch repair proteins MLH1 and PMS2 can be imported to the nucleus by a classical nuclear import pathway. *Biochimie.* 2018;146:87–96. <https://doi.org/10.1016/j.biochi.2017.11.013>.
- Brieger A, Plotz G, Raedle J, Weber N, Baum W, Caspary WF, et al. Characterization of the nuclear import of human MutLa. *Mol Carcinog.* 2005;43:51–8. <https://doi.org/10.1002/mc.20081>.
- Wu X, Platt JL, Cascalho M. Dimerization of MLH1 and PMS2 limits nuclear localization of MutLa. *Mol Cell Biol.* 2003;23:3320–8. <https://doi.org/10.1128/MCB.23.9.3320-3328.2003>.
- Kosinski J, Hinrichsen I, Bujnicki JM, Friedhoff P, Plotz G. Identification of Lynch syndrome mutations in the MLH1-PMS2 interface that disturb dimerization and mismatch repair. *Hum Mutat.* 2010;31:975–82. <https://doi.org/10.1002/humu.21301>.
- Hinrichsen I, Weißbecher IM, Huhn M, Passmann S, Zeuzem S, Plotz G, et al. Phosphorylation-dependent signaling controls degradation of DNA mismatch repair protein PMS2. *Mol Carcinog.* 2017;56:2663–8. <https://doi.org/10.1002/mc.22709>.
- Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet.* 2002;3:285–98. <https://doi.org/10.1038/nrg775>.
- Walker LC, Hoya M, Wiggins GAR, Lindy A, Vincent LM, Parsons MT, et al. Using the ACMG/AMP framework to capture evidence related to predicted and observed impact on splicing: recommendations from the ClinGen SVI Splicing Subgroup. *Am J Hum Genet.* 2023;110:1046–67. <https://doi.org/10.1016/j.ajhg.2023.06.002>.
- Wang Z, Burge CB. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA.* 2008;14:802–13. <https://doi.org/10.1261/rna.876308>.
- Georgakopoulos-Soares I, Parada GE, Hemberg M. Secondary structures in RNA synthesis, splicing and translation. *Comput Struct Biotechnol J.* 2022;20:2871–84. <https://doi.org/10.1016/j.csbj.2022.05.041>.
- De Conti L, Baralle M, Buratti E. Exon and intron definition in pre-mRNA splicing. *Wiley Interdiscip Rev RNA.* 2013;4:49–60. <https://doi.org/10.1002/wrna.1140>.

23. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. Predicting Splicing from primary sequence with deep learning. *Cell*. 2019;176:535–48. <https://doi.org/10.1016/j.cell.2018.12.015>.
24. de Sainte Agathe JM, Filser M, Isidor B, Besnard T, Gueguen P, Perrin A, et al. SpliceAI-visual: a free online tool to improve SpliceAI splicing variant interpretation. *Hum Genomics*. 2023;17. <https://doi.org/10.1186/s40246-023-00451-1>.
25. Pagani F, Baralle FE. Genomic variants in exons and introns: identifying the splicing spoilers. *Nat Rev Genet*. 2004;5:389–96. <https://doi.org/10.1038/nrg1327>.
26. Valentine CR. The association of nonsense codons with exon skipping. *Mutat Res*. 1998;411:87–117. [https://doi.org/10.1016/s1383-5742\(98\)00010-6](https://doi.org/10.1016/s1383-5742(98)00010-6).
27. Yamaguchi T, Wakatsuki T, Kikuchi M, Horiguchi SI, Akagi K. The silent mutation MLH1 c.543C>T resulting in aberrant splicing can cause Lynch syndrome: a case report. *Jpn J Clin Oncol*. 2017;47:576–80. <https://doi.org/10.1093/jjco/hyx023>.
28. Horton C, Hoang L, Zimmermann H, Young C, Grzybowski J, Durda K, et al. Diagnostic outcomes of concurrent DNA and RNA sequencing in individuals undergoing Hereditary Cancer Testing. *JAMA Oncol*. 2023;9:2656:212–9. <https://doi.org/10.1001/jamaoncol.2023.5586>.
29. Kim E, Goren A, Ast G. Alternative splicing: current perspectives. *BioEssays*. 2008;30:38–47. <https://doi.org/10.1002/bies.20692>.
30. Majewski J, Ott J. Distribution and characterization of regulatory elements in the human genome. *Genome Res*. 2002;12:1827–36. <https://doi.org/10.1101/gr.606402>.
31. Cooper TA, Wan L, Dreyfuss G, RNA, Disease. *Cell*. 2009;136:777–93. <https://doi.org/10.1016/j.cell.2009.02.011>.
32. 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. <https://doi.org/10.1038/gim.2015.30>.
33. Landrum MJ, Kattman BL. ClinVar at five years: delivering on the promise. *Hum Mutat*. 2018;39:1623–30. <https://doi.org/10.1002/humu.23641>.
34. 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. <https://doi.org/10.1038/ng.2854>.
35. Lagerstedt-Robinson K, Rohlin A, Aravidis C, Melin B, Nordling M, Stenmark-Askalm M, et al. Mismatch repair gene mutation spectrum in the Swedish Lynch syndrome population. *Oncol Rep*. 2016;36:2823–35. <https://doi.org/10.3892/or.2016.5060>.
36. Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, et al. ClinGen—the Clinical Genome Resource. *N Engl J Med*. 2015;372:2235–42. <https://doi.org/10.1056/NEJMs1406261>.
37. Lim KH, Ferraris L, Filloux ME, Raphael BJ, Fairbrother WG. Using positional distribution to identify splicing elements and predict pre-mRNA processing defects in human genes. *Proc Natl Acad Sci U S A*. 2011;108:11093–8. <https://doi.org/10.1073/pnas.110113510>.
38. Landrith T, Li B, Cass AA, Conner BR, LaDuca H, McKenna DB, et al. Splicing profile by capture RNA-seq identifies pathogenic germline variants in tumor suppressor genes. *NPJ Precis Oncol*. 2020;4. <https://doi.org/10.1038/s41698-020-0109-y>.
39. Aissat A, de Becdelièvre A, Golmard L, Vasseur C, Costa C, Chaoui A, et al. Combined computational-experimental analyses of CFTR exon strength uncover predictability of exon-skipping level. *Hum Mutat*. 2013;34:873–81. <https://doi.org/10.1002/humu.22300>.
40. Canson D, Glubb D, Spurdle AB. Variant effect on splicing regulatory elements, branchpoint usage, and pseudoexonization: strategies to enhance bioinformatic prediction using hereditary cancer genes as exemplars. *Hum Mutat*. 2020;41:1705–21. <https://doi.org/10.1002/humu.24074>.
41. Tubeuf H, Charbonnier C, Soukariéh O, Blavier A, Lefebvre A, Dauchel H, et al. Large-scale comparative evaluation of user-friendly tools for predicting variant-induced alterations of splicing regulatory elements. *Hum Mutat*. 2020;41:1811–29. <https://doi.org/10.1002/humu.24091>.
42. Thompson BA, Martins A, Spurdle AB. A review of mismatch repair gene transcripts: issues for interpretation of mRNA splicing assays. *Clin Genet*. 2015;87:100–8. <https://doi.org/10.1111/cge.12450>.
43. Wu Y, Zhang Y, Zhang J. Distribution of exonic splicing enhancer elements in human genes. *Genomics*. 2005;86:329–36.
44. Sterne-Weiler T, Howard J, Mort M, Cooper DN, Sanford JR. Loss of exon identity is a common mechanism of human inherited disease. *Genome Res*. 2011;21:1563–71. <https://doi.org/10.1101/gr.118638.110>.
45. Raponi M, Kralovicova J, Copson E, Divina P, Eccles D, Johnson P, et al. Prediction of single-nucleotide substitutions that result in exon skipping: identification of a splicing silencer in BRCA1 exon 6. *Hum Mutat*. 2011;32:436–44. <https://doi.org/10.1002/humu.21458>.
46. van der Klift HM, Jansen AML, Steenstraten N, Bik EC, Tops CMJ, Devilee P, et al. Splicing analysis for exonic and intronic mismatch repair gene variants associated with Lynch syndrome confirms high concordance between minigene assays and patient RNA analyses. *Mol Genet Genomic Med*. 2015;3:327. <https://doi.org/10.1002/mgg3.145>.
47. Bouras A, Naibo P, Legrand C, Marc'hadour F, Ruano E, Grand-Masson C, et al. A PMS2 non-canonical splicing site variant leads to aberrant splicing in a patient suspected for lynch syndrome. *Fam Cancer*. 2023;22:303–6. <https://doi.org/10.1007/s10689-022-00323-y>.
48. Guarné A, Ramon-Maiques S, Wolff EM, Ghirlando R, Hu X, Miller JH, et al. Structure of the MutL C-terminal domain: a model of intact MutL and its roles in mismatch repair. *EMBO J*. 2004;23:4134–45. <https://doi.org/10.1038/sj.emb.0j.7600412>.
49. Mohd AB, Palama B, Nelson SE, Tomer G, Nguyen M, Huo X, et al. Truncation of the C-terminus of human MLH1 blocks intracellular stabilization of PMS2 and disrupts DNA mismatch repair. *DNA Repair (Amst)*. 2006;5:347–61. <https://doi.org/10.1016/j.dnarep.2005.11.001>.
50. Ke S, Shang S, Kalachikov SM, Morozova I, Yu L, Russo JJ, et al. Quantitative evaluation of all hexamers as exonic splicing elements. *Genome Res*. 2011;21:1360–74. <https://doi.org/10.1101/gr.119628.110>.
51. Johnson JR, Erdeniz N, Nguyen M, Dudley S, Liskay RM. Conservation of functional asymmetry in the mammalian MutLa ATPase. *DNA Repair (Amst)*. 2010;9:1209. <https://doi.org/10.1016/j.dnarep.2010.08.006>.
52. Tomer G, Buermeyer AB, Nguyen MM, Michael Liskay R. Contribution of human Mlh1 and Pms2 ATPase activities to DNA mismatch repair. *J Biol Chem*. 2002;277:21801–9. <https://doi.org/10.1074/jbc.M111342200>.
53. D'Arcy BM, Arrington J, Weisman J, McClellan SB, Vandana J, Yang Z, et al. PMS2 variant results in loss of ATPase activity without compromising mismatch repair. *Mol Genet Genomic Med*. 2022;10:1908. <https://doi.org/10.1002/mgg3.1908>.
54. Li L, Hamel N, Baker K, McGuffin MJ, Couillard M, Gologan A, et al. A homozygous PMS2 founder mutation with an attenuated constitutional mismatch repair deficiency phenotype. *J Med Genet*. 2015;52:348–52. <https://doi.org/10.1136/jmedgenet-2014-102934>.
55. Biswas K, Couillard M, Cavallone L, Burkett S, Stauffer S, Martin BK, et al. A novel mouse model of PMS2 founder mutation that causes mismatch repair defect due to aberrant splicing. *Cell Death Dis*. 2021;12. <https://doi.org/10.1038/s41419-021-04130-8>.
56. Yuan ZQ, Gottlieb B, Beitel LK, Wong N, Gordon PH, Wang Q, et al. Polymorphisms and HNPCC: PMS2-MLH1 protein interactions diminished by single nucleotide polymorphisms. *Hum Mutat*. 2002;19:108–13. <https://doi.org/10.1002/humu.10040>.
57. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, et al. A Novel Approach to High-Quality Postmortem tissue procurement: the GTEx Project. *Biopreserv Biobank*. 2015;13:311–7. <https://doi.org/10.1089/bio.2015.0032>.
58. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The genotype-tissue expression (GTEx) project. *Nat Genet*. 2013;45:580–5. <https://doi.org/10.1038/ng.2653>.
59. Frankish A, Uszczyńska B, Ritchie GRS, Gonzalez JM, Pervouchine D, Petryszak R, et al. Comparison of GENCODE and RefSeq gene annotation and the impact of reference geneset on variant effect prediction. *BMC Genomics*. 2015;16:1–11. <https://doi.org/10.1186/1471-2164-16-S8-S2>.
60. O'Leary NA, Wright MW, Brister JR, Ciufos S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016;44:D733–45. <https://doi.org/10.1093/nar/gkv1189>.
61. Jang W, Park J, Chae H, Kim M. Comparison of in Silico Tools for splice-altering variant prediction using established spliceogenic variants: an end-user's point of view. *Int J Genomics*. 2022. <https://doi.org/10.1155/2022/5265686>.
62. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative Genomics Viewer. *Nat Biotechnol*. 2011;29:24. <https://doi.org/10.1038/nbt.1754>.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.