The Journal of Molecular Diagnostics, Vol. 26, No. 8, August 2024

ELSEVIER

the Journal of Molecular Diagnostics
jmdjournal.org

Open-Source Bioinformatic Pipeline to Improve *PMS2* Genetic Testing Using Short-Read NGS Data

Elisabet Munté,[*†] Lídia Feliubadaló,[*†‡] Jesús Del Valle,[*†‡] Sara González,[*†] Mireia Ramos-Muntada,[†‡] Judith Balmaña,[§¶] Teresa Ramon y Cajal,[‖] Noemí Tuset,[**] Gemma Llort,[††] Juan Cadiñanos,[‡‡] Joan Brunet,[*‡§§] Gabriel Capellá,[*†‡] Conxi Lázaro,[*†‡] and Marta Pineda[*†‡]

From the Hereditary Cancer Program, Catalan Institute of Oncology, L'Hospitalet de Llobregat; Hereditary Cancer Group,† Molecular Mechanisms and Experimental Therapy in Oncology Program, Institut d'Investigació Biomèdica de Bellvitge, L'Hospitalet de Llobregat; Ciber Oncología,‡ Instituto Salud Carlos III, Madrid; the High Risk and Cancer Prevention Group,§ Vall d'Hebron Institute of Oncology, Barcelona; the Medical Oncology Department,¶ University Hospital of Vall d'Hebron, Barcelona; the Familial Cancer Clinic,‖ Medical Oncology Service, Hospital Sant Pau, Barcelona; the Medical Oncology Department,** Hospital Universitari Arnau de Vilanova, Lleida; the Department of Medical Oncology Parc Taulí,†† Hospital Universitari Parc Taulí Sabadell, Barcelona; the R&D Laboratory,‡‡ Fundación Centro Médico de Asturias—IMOMA, Oviedo; and the Precision Oncology Group (OncoGir_Pro),§§ Institut d'Investigació Biomèdica de Girona, Girona, Spain*

Accepted for publication May 13, 2024.

Address correspondence to Marta Pineda, Ph.D., or Conxi Lázaro, Ph.D., Catalan Institute of Oncology, L'Hospitalet de Llobregat, Hereditary Cancer Program, Avinguda de la Granvia de l'Hospitalet, 199, L'Hospitalet de Llobregat, 08908 Barcelona, Spain. E-mail: mpineda@iconcologia.net or clazaro@iconcologia.net.

The molecular diagnosis of mismatch repair—deficient cancer syndromes is hampered by difficulties in sequencing the *PMS2* gene, mainly owing to the PMS2CL pseudogene. Next-generation sequencing short reads cannot be mapped unambiguously by standard pipelines, compromising variant calling accuracy. This study aimed to provide a refined bioinformatic pipeline for *PMS2* mutational analysis and explore *PMS2* germline pathogenic variant prevalence in an unselected hereditary cancer (HC) cohort. *PMS2* mutational analysis was optimized using two cohorts: 192 unselected HC patients for assessing the allelic ratio of paralogous sequence variants, and 13 samples enriched with *PMS2* (likely) pathogenic variants screened previously by long-range genomic DNA PCR amplification. Reads were forced to align with the *PMS2* reference sequence, except those corresponding to exon 11, where only those intersecting gene-specific invariant positions were considered. Afterward, the refined pipeline's accuracy was validated in a cohort of 40 patients and used to screen 5619 HC patients. Compared with our routine diagnostic pipeline, the PMS2_vaR pipeline showed increased technical sensitivity (0.853 to 0.956, respectively) in the validation cohort, identifying all previously *PMS2* pathogenic variants found by long-range genomic DNA PCR amplification. Fifteen HC cohort samples carried a pathogenic *PMS2* variant (15 of 5619; 0.285%), doubling the estimated prevalence in the general population. The refined open-source approach improved *PMS2* mutational analysis accuracy, allowing its inclusion in the routine next-generation sequencing pipeline streamlining *PMS2* screening. *(J Mol Diagn 2024, 26: 727—738; https://doi.org/10.1016/j.jmoldx.2024.05.005)*

Supported by Fundació La Marató de Televisió de Catalunya (TV3) grant 202028-30; Fundación Mutua Madrileña grant AP183612023; Ministerio de Ciencia e Innovación (PID2019-111254RB-100); Carlos III National Health Institute, funded by Fondo Europeo de Desarrollo Regional (FEDER) "A way to build Europe" (PI23/00017 and PI19/00553); Carlos III National Health Institute and the European Union (EU) Next Generation EU/Mecanismo para la Recuperación y la Resiliencia (PMP22/00064); Centro de Investigación Biomédica en Red Cáncer (CIBERONC) grant CB16/12/00234; and Department of Research and Universities of the Generalitat de Catalunya and Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) (2023SGR01112). Also supported by the Acció Instrumental de Formació de Científics i Tecnòlegs grant SLT017/20/000129 of the Departament de Salut de la Generalitat de Catalunya. The CERCA program/Generalitat de Catalunya and Fundación María Cristina Masaveu Peterson provided institutional support.

C.L. and M.P. contributed equally to this work.

https://doi.org/10.1016/j.jmoldx.2024.05.005

Lynch syndrome (LS) is a common, dominantly inherited, cancer-predisposing condition caused by germline pathogenic variants affecting the function of mismatch repair (MMR) genes (*MLH1*, *MSH2*, *MSH6*, and *PMS*2).[1] Despite its incomplete penetrance, individuals harboring an MMR pathogenic variant have increased chances of developing colorectal and endometrial cancers, among others.[2] Biallelic pathogenic alterations damaging the same MMR genes cause constitutional mismatch repair deficiency (CMMRD), a very rare (1 in 1,000,000) and severe condition that predisposes to multiorgan cancers—mainly brain, hematologic, and colorectal—usually with childhood onset.[3–5] Its penetrance is more than 90% at the age of 20.

The estimated population frequency of pathogenic *PMS2* variant carriers is the highest among the four MMR genes (1 in 714; 0.140%).[1] Accordingly, *PMS2* is the most frequently mutated gene in CMMRD syndrome, accounting for nearly 60% of cases.[5] In contrast, *PMS2* is the least frequently mutated gene in the LS series, probably owing to its lowest penetrance in heterozygous carriers and the former use of clinical criteria for LS tumor screening. Nevertheless, the cancer risk varies widely even among heterozygous carriers from the same family.[2,6,7]

Gene panels using targeted next-generation sequencing (NGS) of short reads are the tests most used in the field of hereditary cancer (HC) because of their optimal balance between cost and benefit. However, short-read–based NGS has significant limitations in the identification of variants in complex regions.[8] Indeed, *PMS2* gene analysis presents a major challenge mainly because of the existence of multiple pseudogenes.[9,10] Specifically, there are 14 pseudogenes located at the 5′ end, spanning exons 1 through 5, and an additional 15th pseudogene located at the 3′ end, known as PMS2CL. Remarkably, the PMS2CL pseudogene is an inverted partial duplication located on the same chromosome 7 that exhibits notable sequence homology (>98% identity) with exons 9 and 11 to 15 of the *PMS2* gene. Some bases, called paralogous sequence variants (PSVs), differ in *PMS2*-PMS2CL reference sequence.[11,12] It has been proven that sequence exchange (recombination and gene conversion) is a frequent event observed between these two loci.[13,14] This makes it difficult to discriminate reliably whether an identified variant is located in the gene or the pseudogene.[13,14]

Genomic DNA long-range PCR (LR-PCR) amplification and gene-specific cDNA amplification using primers located in less-homologous regions, and DNA/cDNA long-read sequencing, can analyze *PMS2* specifically.[7,15–20] Nevertheless, these techniques are labor-intensive, complicate routine diagnostic workflows, and present many technical challenges, which question the feasibility of implementing them in large cohorts. Bioinformatic approaches partially can palliate these difficulties. In this sense, Gould et al[11] proposed a workflow in which gene and pseudogene variants were forced to align with the *PMS2* gene reference sequence. By this means, they identified seven PSV positions in *PMS2* exon 11, where none of the 707 ethnically diverse patients from their cohort differed from the gene and pseudogene reference sequences. These positions, hereinafter referred to as invariant PSVs, were demonstrated to be useful in identifying the origin of variants identified in NGS reads overlapping them.[11] Despite these advances, to our knowledge, there is no free open-source pipeline available to analyze *PMS2* accurately. Thus, the inclusion of the *PMS2* gene in routine NGS diagnostic pipelines remains a challenge for most genetic testing laboratories.

To address this need, PMS2_vaR is presented, the first free open-source pipeline written in R, which integrates and upgrades the previously reported strategy. The aim of this study is to increase the accuracy of *PMS2* variant detection using routine NGS panel data, in addition to reducing the number of samples that need to undergo LR-PCR. This study also aimed to assess the prevalence of *PMS2* pathogenic variants in an HC cohort upon implementation.

Materials and Methods

Study Cohorts

Table 1 and Supplemental Figure S1 provide an overview of the cohorts used.

The PMS2_vaR pipeline was optimized using samples from two cohorts: optimization cohort A, comprising 192 HC patients used to assess the allelic ratio of PSVs in unselected samples; and optimization cohort B, enriched in samples harboring PMS2 (likely) pathogenic variants, composed of 13 cancer patients in whom blood DNA was analyzed previously by *PMS2* LR-PCR, enabling the identification of *PMS2* variants. For validation purposes, a LS suspicion cohort of 40 patients analyzed previously by PMS2 LR-PCR was used to determine the pipeline's accuracy.

Finally, a large HC cohort of 5619 patients was studied. According to clinical phenotypes, the cohort comprised 13 LS-suspected patients showing exclusive PMS2 loss of expression in tumors (blood samples were not analyzed previously by LR-PCR), 36 patients diagnosed with early onset colorectal cancer (age, <50 years at diagnosis) showing MMR-conserved protein expression or with no available immunohistochemistry information, 798 patients fulfilling other LS suspicion criteria (Amsterdam criteria or MMR expression loss but not exclusively of PMS2), and 4772 patients tested for suspicion of other HC syndromes.

Sample Collection and Preprocessing

DNA samples were obtained from peripheral blood leukocytes of individuals with HC suspicion referred to the Molecular Diagnostics Service at the Institut Català d'Oncologia from its network of genetic counseling units. Informed written consent for both diagnostic and research purposes was obtained from this cohort of patients. The study protocol was approved by the Ethics Committee of the Catalan Institute of Oncology–Bellvitge University Hospital (PR278/19).

Table 1 Summary of Cohorts Analyzed

Group	Subgroup	Clinical and molecular criteria	*n*	Recommendation for *PMS2* LR-PCR analysis	Purpose
Optimization cohort	Cohort A	HC suspicion, unselected	192	No	Determine allele ratio
	Cohort B	LS suspicion; IHC: PMS2^{-}; *PMS2* LR-PCR analysis previously performed Enriched in PMS2 (L)PAT variants	13	Previously performed	Determine pipeline accuracy
Validation cohort		LS suspicion; *PMS2* LR-PCR analysis previously performed	40	Previously analyzed	Determine pipeline accuracy
Hereditary cancer cohort		LS suspicion; IHC: PMS2^{-}; *PMS2* LR-PCR analysis not performed	13	Yes, in samples with an identified *PMS2* (likely) pathogenic variant*	Determine prevalence in this subgroup
		Early onset CRC suspicion; IHC: conserved expression	36	Yes, in samples with an identified *PMS2* (likely) pathogenic variant*	Determine prevalence in this subgroup
		LS suspicion (Amsterdam criteria or IHC MMR expression loss excluding *PMS2*)	798	Yes, in samples with an identified *PMS2* (likely) pathogenic variant*	Determine prevalence in this subgroup
		Other HC suspicions	4772	Yes, in samples with an identified *PMS2* (likely) pathogenic variant*,†,‡	Determine prevalence in this subgroup

*Pathogenic paralogous sequence variants (c.1864_1865del and c.1730dup) will be considered for LR-PCR testing only if the variant allele frequency is >60%.

†Pseudogenic exon 13 c.2186_2187del and c.2243_2246del *PMS2* variants will be considered for LR-PCR testing only if the tumor molecular characteristics are indicative of a *PMS2* alteration (microsatellite instability or exclusive IHC PMS2 loss) or when IHC analysis is not possible.

‡(Likely) pathogenic variants called by the general approach but filtered out after the refined E11 approach will not be tested unless the tumor molecular characteristics are indicative of a *PMS2* alteration (microsatellite instability or exclusive IHC PMS2 loss) or when IHC analysis is not possible.

CRC, colorectal cancer; HC, hereditary cancer; IHC, immunohistochemistry; (L)PAT, (likely) pathogenic; LR-PCR, long-range PCR; LS, Lynch syndrome; MMR, mismatch repair.

Routine Diagnostics Pipeline

Genetic testing was conducted on peripheral blood DNA using NGS custom panel ICO-IMPPC Hereditary Cancer Panel (I2HCP).[21] This panel encompasses a comprehensive selection of 122 to 168 genes (depending on the version used) associated with HC susceptibility. The bioinformatics approach used for the routine diagnostics pipeline was described previously.[21,22] The selection of genes for analysis was based on the phenotype of each patient[23] and their family, following the Catalan Health Service guidelines.

PSV Determination

A list of 31 exonic base differences was obtained by comparing the *PMS2* (NM_000535.7; *https://www.ncbi.nlm.nih.gov/nuccore/1519311653*, last accessed May 15, 2024) and PMS2CL (NR_002217.1; *https://www.ncbi.nlm.nih.gov/nuccore/NR_002217.1*, last accessed May 15, 2024) sequences using the BlastN tool from the National Center for Biotechnology Information. Because some variants were consecutive, they were considered compound variants, thus the final list was 28 PSVs, 23 of which were located within exon 11. Supplemental Table S1 contains the complete list of all 28 PSVs.[12]

Bioinformatic Pipeline Development

The PMS2_vaR pipeline was conceived for the R statistical computing environment (v4.2.1; *https://github.com/emunte/PMS2_vaR*, last accessed March 22, 2024). It requires the installation of the following software: SAMtools (v1.10; *https://www.htslib.org*), Picard (v.2.26.4.jar; *https://github.com/broadinstitute/picard*), BWA (0.7.17; *https://github.com/lh3/bwa*), and VarDictJava (v1.8.3; *https://github.com/AstraZeneca-NGS/VarDictJava*). It also uses functions from both R/Bioconductor and CRAN packages (*https://cran.r-project.org*; see the required libraries in the GitHub space).

The pipeline consists of two scripts: modify_reference and run_PMS2_vaR (Figure 1).

modify_reference Script

Given a human reference genome sequence FASTA file and its PMS2CL FASTA sequence, the workflow generates a PMS2CL-masked reference genome in which the PMS2CL genomic sequence is replaced by Ns (any base). This file is needed as an input file for the Run_PMS2_vaR algorithm. This step only needs to be executed once (per human reference genome version).

run_PMS2_vaR Script

To feed the pipeline, the user is required to provide several input files, including a text file containing paths to the BAM

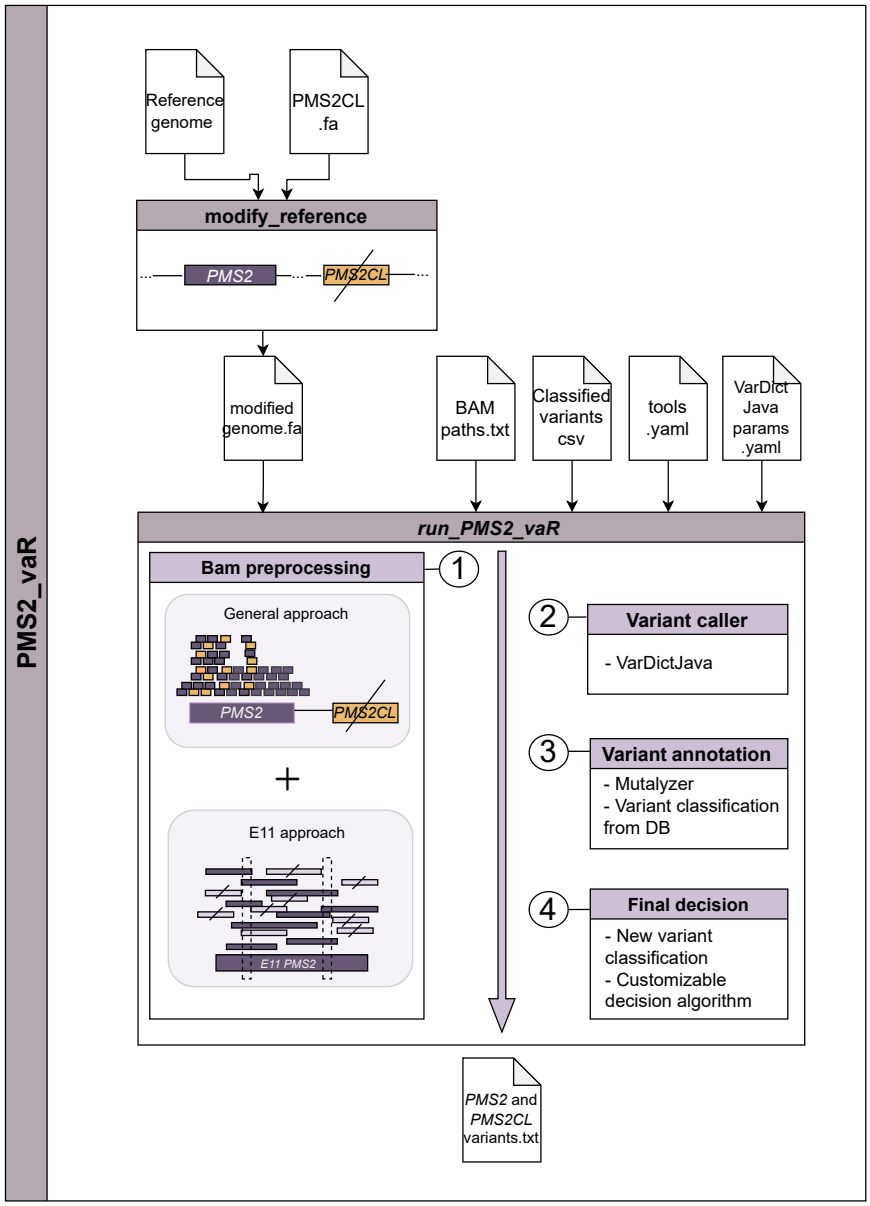

Figure 1 Schematic representation of the PMS2_vaR pipeline. The modified reference genome sequence file is obtained using the "modify_reference" script. Given a BAM file, a classified variants csv file and the yaml files specifying the tool's paths and parameters for VardictJava, the run_PMS2_vaR script produces a text file containing all *PMS2* and PMS2CL variants. The steps to construct the final *PMS2* candidate variants list are detailed in *run_PMS2_vaR Script*. .fa, FASTA file.

files, a yaml file detailing the paths to the necessary tools (SAMtools, Picard, BWA, and VarDictJava), another yaml file specifying the parameters to be used with VarDictJava, and a comma delimited file listing classified *PMS2* variants. The template for these files is available at *https://github.com/emunte/PMS2_vaR*.

In the general approach, to obtain the list of candidate variants that need to be validated further by LR-PCR, gene and pseudogene reads in the highly homologous regions were aligned with the PMS2CL-masked human genome reference sequence. To this end, first, reads aligning with *PMS2* or PMS2CL in the standard pipeline BAM were selected using SAMtools. The resulting BAM file was transformed into paired-end FASTQ files using Picard software. Afterward, the FASTQ files were realigned with the modified reference genome using BWA-MEM. The SAM file was converted to a BAM file, sorted, and indexed using SAMtools.

Subsequently, exon (E)11 was analyzed based on the approach of Gould et al[11] (hereinafter called the E11 approach). Following their recommendations, only reads that intersected with any of the seven invariant PSVs were included (Supplemental Table S1). Read names overlapping the corresponding positions were obtained using SAMtools and the reads then were filtered by name using Piccard. These were aligned to the standard (nonmasked) reference genome. The resulting E11 BAM was merged with the BAM obtained for the other exons with the general approach.

Variant calling was performed for both approaches using VarDictJava. The following parameters were modified: i)

Table 2 Primer Sequences for *PMS2* Amplification

Target	Template	Forward primer	Reverse primer
LR1 (exons 1—5)	gDNA	5′-ACGTCGAAAGCAGCCAATGGGAGTT-3′*,†	5′-CTTCCACCTGTGCATACCACAGGCT-3′*,†
LR2 (exons 7—9)	gDNA	5′-GGTCCAGGTCTTACATGCATACTGT-3′*,†	5′-CTGACTGACATTTAGCTTGTTGACA-3′*,†
LR3 (exons 11—15)	gDNA	5′-GCGTTGATATCAATGTTACTCCAGA-3′*,†	5′-CCTTCCATCTCCAAAACCAGCAAGA-3′*,†
Exon 1	LR1	5′-M13F-ACGTCGAAAGCAGCCAATGGGAGTT-3′*,†	5′-M13R-CAGGTAGAAAGGAAATGCATTCAGT-3′*,†
Exon 2	LR1	5′-ACAGTGTTGAGTCATTTCCCACAGT-3′*,†	5′-TTCTTAGCATAACACCTGCCTGGCA-3′*,†
Exon 3—4	LR1	5′-M13F-CTGGGCTAGTAAATAGCCAGAAAG-3′†	5′-M13R-TATGACTTAGATTGGCAGCGAGACA-3′*,†
Exon 5	LR1	5′-M13F-CTTGATTATCTCAGAGGGATCGTCA-3′*,†	5′-M13R-TCTCACTGTGTTGCCCAGTCCTAAT-3′*,†
Exon 6	gDNA	5′-M13F-TGCTTCCCTTGATTTGTGCGATGAT-3′*,†	5′-M13R-CATTCTACTGGAAGGGACAATGGA-3′
Exon 7	gDNA	5′-M13F-ACCCACGAGTTTGACATTGCAGTGA-3′*	5′-M13R-AAAAGACACGAAACTATTAGCCTTAGA-3′
Exon 8	gDNA	5′-M13F-AGATTTGGAGCACAGATACCCGTGA-3′*,†	5′-M13R-TGCGGTAGACTTCTGTAAATGCACA-3′*,†
Exon 9	LR2	5′-M13F-CCTTCTAAGAACATGCTGGTTGGTT-3′*,†	5′-M13R-ATCTCATTCCAGTCATAGCAGAGCT-3′*,†
Exon 10	gDNA	5′-M13F-AATTAGCCAGTGTGGTGGCACTTG-3′†	5′-M13R-AGCTTTAGAAGCTGTTTGTACAC-3′†
Exon 11a	LR3	5′-M13F-TCACATAAGCACGTCCTCTCACCAT-3′*,†	5′-M13R-GAATGGCAGTCCACATCTGAAAAAG-3′
Exon 11b	LR3	5′-M13F-CAGAGCGGAGGTGGAGAAGGAC-3′	5′-M13R-GTGAAACCCTGTTTCCACCAAAAAT-3′
Exon 12	LR3	5′-M13F-GCCAAGATTGTGCCATTGCACTGTA-3′*	5′-M13R-AGTAGATACAAGGTCTTGCTGTGTT-3′*,†
Exon 13	LR3	5′-M13F-TTGTTTTCATTTCATTTCTGCTG-3′	5′-M13R-ATGTTAGCCAGGCTGGTCTCAAACT-3′*,†
Exon 14	LR3	5′-M13F-GCTTTCAAGTGAAACGTGTTTGTCA-3′	5′-M13R-GCACGTAGCTCTCTGTGTAAAATGA-3′*,†
Exon 15	LR3	5′-M13F-GCTGAGATCTAGAACCTAGGCTTCT-3′*,†	5′-M13R-ACACACGAGCGCATGCAAACATAGA-3′*,†

*Primers from Clendenning et al.[15]
†Primers from Vaughn et al.[16]
LR, long range; gDNA, genomic DNA; M13F sequence, 5′-TGTAAAACGACGGCCAGT-3′; M13R sequence, 5′-CAGGAAACAGCTATGACC-3′.

the minimum allele frequency was set to 0.1 to accommodate the factual tetra-allelic situation because the new alignments (without the PMS2CL sequence in the reference) combine four alleles in *PMS2* E9, 11 to 15, and nearby positions (Supplemental Figure S2); ii) the region of interest (-R) was set to chr7:6012350-6049257 for hg19 and chr7:5972719-6009626 for hg38; iii) the minimum phred score (-q) was set to 15; and iv) the number of mismatches allowed in a read (-m) was set to 10 for greater permissiveness. In addition to variants with all filters passed, variants tagged for mean mismatches in reads ≥5.25 (NM5.25) or for being adjacent to an insertion variant (InIns) were kept. In a diploid variant calling situation, these filters would point to likely false-positive variants, however, this study tried to be conservative.

The two variant calling files were converted into txt files using the vcfR R package and were merged into the same document. This allows the user to verify whether the variant was found by one or both approaches. The pipeline followed the decision algorithm described in *Results* to suggest whether LR-PCR should be performed or not for each variant.

PMS2 Variant Validation by LR-PCR

Candidate variants in *PMS2* (NM_000535.7, NG_008466.1, *https://www.ncbi.nlm.nih.gov/nuccore/NG_008466.1*, last accessed May 15, 2024) identified by vaR_PMS2 were analyzed using previously described LR-PCR procedures.[15,16] A schematic representation detailing the annealing positions of all the primers used can be found in Supplemental Figure S3. In brief, amplicons encompassing entire exons 1 to 5 (long-range amplicon LR1), 9 (LR2), and 11 to 15 (LR3) were generated using LaTaq polymerase (TaKaRa Bio, Inc, Otsu, Shiga, Japan) and the corresponding primers are listed in Table 2. Amplification of LR-PCR products was confirmed by agarose gel electrophoresis. The LR3 product was purified by gel extraction to avoid pseudogene amplification from genomic DNA instead of from the LR-PCR product in the following exon-specific PCR. LR-PCR products (or purified products) were diluted 1:10 and 1 μL of this dilution was used as the template for nested exon-specific amplifications. Exon-specific PCRs were performed using DreamTaq DNA polymerase (Thermo Fisher Scientific, Waltham, MA) and the corresponding primers (Table 2). For exons 6, 7, 8, and 10, genomic DNA was used as the PCR template. Amplification was confirmed by agarose gel electrophoresis and PCR products were sequenced using the Big Dye Terminator v.3.1 Cycle Sequencing kit (Applied Biosystems, Waltham, MA) and an Applied Biosystems 3130XL Genetic Analyzer.[15,16]

Variant Classification

A list of 129 *PMS2* classified variants was provided to feed the pipeline (Supplemental Table S2). Variants initially were classified using the vaRHC R package[24] and subsequently curated by the Catalan Institute of Oncology

Hereditary Cancer Molecular Diagnostics Service. The draft version of the InSiGHT-ClinGen–specific MMR variant classification guidelines were followed (*https://www.insight-group.org/content/uploads/2021/11/DRAFT_Nov_2021_TEMPLATE_SVI.ACMG_Specifications_InSiGHT_MMR_V1.pdf*, last accessed March 22, 2024). Users have the flexibility to incorporate additional classified variants or modify the classification of existing ones, tailoring the system to their specific requirements (see GitHub for further details).

Assessment of Routine and PMS2_vaR Pipelines Performance

The performance of the routine and the PMS2_vaR pipelines was analyzed against the results obtained from LR-PCR in both the optimization and validation cohorts. A comprehensive set of performance metrics was computed, including accuracy, sensitivity, specificity, positive predictive value, and negative predictive value. The McNemar test was used to determine significance, with a predefined P value of 0.05. Statistical analyses were conducted using R v.4.2.2, leveraging the CRAN package DTComPair v.1.2.2 (*https://CRAN.R-project.org/package=DTComPair*). To assess the reduction in long and short PCRs [(LR)-PCR] workload achieved by using PMS2_vaR, the location of each candidate variant identified in the HC *PMS2* cohort was examined to determine the precise PCR reaction required for validation. This result was compared with the total number of PCR reactions needed in the *PMS2* analysis by (LR)-PCR.

Results

Bioinformatic Pipeline Development: Data Processing and Variant Calling

A bioinformatic pipeline was developed to identify *PMS2* variants from multigene panel NGS data with high accuracy. First, reads aligning to *PMS2* or PMS2CL were converted into FASTQ files and realigned with a human reference genome with the PMS2CL sequence masked. This forced both *PMS2* and PMS2CL reads to map to the *PMS2* reference. Subsequently, the PSV variant allele frequency (VAF) was assessed from a cohort of unselected samples (optimization cohort A). Given that PSVs are positions where gene and pseudogene reference sequences diverge, it would be expected that when one of these variants is called, the reads that support it come from the pseudogene. Consequently, PSVs would have an expected VAF of approximately 50% (present in two of four alleles). However, in the analysis of samples of the optimization cohort A, the observed VAF ranged from 35% to 45% for most PSVs (Supplemental Figure S4). This reduction suggested that the probes have a slightly weaker affinity for PMS2CL regions harboring PSVs according to the reference genome. Moreover, some PSVs deviated strongly from the expected proportions: three exhibited a VAF below 25% across multiple samples, indicating that these variants likely are pseudogene polymorphisms, and another three displayed VAFs exceeding 60%, suggesting that they likely are gene polymorphisms (Supplemental Figure S4).

In addition to this general approach, a refined method for analyzing exon 11 subsequently was introduced, including only reads overlapping invariant PSVs for this exon. Results from the two approaches were integrated into the same data frame.

Decision-Making Algorithm

An algorithm was designed to assess the presence of variants in MMR genes and to recommend if a called *PMS2* variant would need confirmation by (LR-)PCR analysis (Figure 2). For each *PMS2* candidate variant, the algorithm first assessed if the variant passed the quality filters and if it was in a region of interest (this study's setting was a coding region ± 20 bp). Next, it checked if it was a PSV. Among the 28 exonic PSVs listed (Supplemental Table S1), 2 were classified as (likely) pathogenic variants if located in the gene (c.1730dup and c.1864_1865del). These two variants were called in all samples from the optimization cohort A and, in most cases, they corresponded to the pseudogene reference sequence rather than being a gene variant.[11] To avoid the need for LR-PCR analysis of each sample, paralogous variants were regarded exclusively as potential gene candidates if their VAF exceeded 60% (Table 1 and Figure 2). The selected threshold presumes that candidate variants at PSV positions should be present in at least three of the four alleles of *PMS2* and PMS2CL (Supplemental Figure S2).

For non-PSV variants or those PSVs with VAFs ≥60%, the algorithm examined whether the variant was classified or not. If the variant was considered pathogenic or likely pathogenic and also was called with the E11 approach, LR-PCR was recommended (Figure 2). However, in samples with any called *PMS2* pathogenic variant discarded by the E11 approach, LR-PCR analysis would be recommended only when the family phenotype strongly indicated Lynch or CMMRD syndromes (ie, loss of PMS2 protein expression or microsatellite instability). Variants of unknown significance did not undergo LR-PCR because they are not currently clinically actionable. They were reported only if the variant was detected with both approaches, with a disclaimer clarifying that they were not validated by LR-PCR. Finally, benign or likely benign variants were not reported.

Apart from PSVs, some common pseudogenic variants also can be found in the gene, albeit at very low population frequencies. Specifically, exon 13 recurrent variants c.2186_2187del (p.Leu729Glnfs*6) and c.2243_2246del (p.Lys748Metfs*19) are of particular interest because they

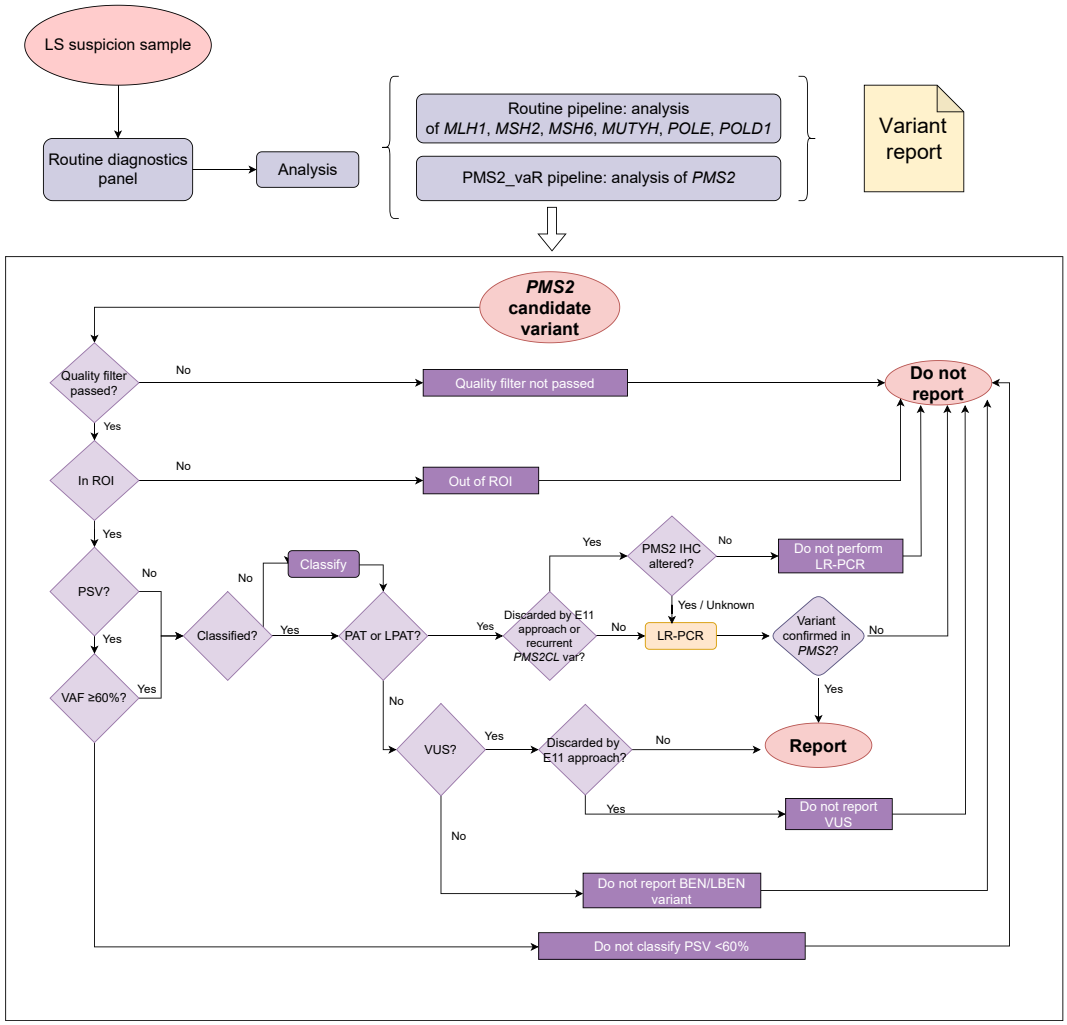

Figure 2 Algorithm used to analyze samples with suspected Lynch syndrome (LS). It assesses the presence of variants in MMR genes using a panel approach. For *PMS2* gene analysis, the PMS2_vaR pipeline indicates when long-range (LR)-PCR analysis should be recommended to confirm a *PMS2* called variant. Recurrent PMS2CL pathogenic variants c.2186_2187del and c.2243_2246del will be examined only if PMS2 expression is lost in tumors. BEN, benign; IHC, immunohistochemistry; LBEN, likely benign; LPAT, likely pathogenic; PAT, pathogenic; PSV, paralogous sequence variant; ROI, region of interest; VAF, variant allele frequency; VUS, variant of unknown significance.

attain a (likely) pathogenic classification within the gene context. The presence of these two variants within the *PMS2* gene was identified in 0.01% (1 of 7593) and 0.07% (2 of 2739) of HC-suspected patients in whom the variants had been called by NGS, respectively.[25] Taking this into account, LR-PCR was performed only on samples that harbored these two recurrent pseudogenic variants when the clinical criteria and tumor molecular characteristics of the carriers indicated a potential *PMS2* alteration or when the VAF was $\geq$60%.

Assessment of PMS2_vaR Pipeline Performance in the Optimization and Validation Cohorts

To assess the accuracy of the newly developed PMS2_vaR pipeline, the optimization cohort B, including

Table 3 Accuracy, Sensitivity, Specificity, Positive Predicted Value, and Negative Predicted Value Obtained by the Previous Routine Diagnostic Pipeline and the Refined Pipeline (General + E11 Approach) in ROIs (±20 bp)

	Accuracy		Sensitivity		Specificity		PPV		NPV	
Pipeline	Op B	Val	Op B	Val	Op B	Val	Op B	Val	Op B	Val
Diagnostics pipeline	0.9993	0.9993	0.9067	0.8528	0.9994	0.9996	0.7391	0.7818	0.9998	0.9997
PMS2_vaR pipeline	0.9994	0.9996	0.9733	0.9561	0.9994	0.9997	0.7449	0.8435	0.9999	0.9999

NPV, negative predictive value; Op B, optimization cohort B; PPV, positive predictive value; ROI, region of interest; Val, validation cohort.

Table 4 Detailed Information of (Likely)-Pathogenic *PMS2* Variants Identified by PMS2_vaR in HC Clinical Phenotype Cohorts

Group	ID	Variant	Protein	Location	Personal phenotype
LS suspicion with exclusive PMS2 IHC^{-}	1	c.312del	p.(Phe104Leufs*8)	E04	ENDO (54 y), CRC (67 y)
	2	c.584C>A	p.(Ser195*)	E06	Sebaceoma
	3	c.706-1G>T	p.?	I06	CRC (44 y)
	4	c.1144+2T>G	p.?	I10	CRC (73 y)
	5	c.1687C>T	p.(Arg563*)	E11	ENDO (48 y)
Early onset CRC with IHC conserved					
Other LS suspicion criteria (Amsterdam criteria or MMR expression loss but not exclusive of PMS2)	6	c.717_723dup	p.(Phe242Hisfs*9)	E07	CRC (49 y), BBC (51 y)
	7	c.904-1G>A	p.?	I08	CRC (43 y)
	8	c.988+1G>A	p.?	I09	ENDO (62 y)
	9	c.1145-1_1145del	p.?	I10—E11	CRC (68 y)
	10	c.1882C>T	p.(Arg628*)	E11	CRC(44 y)
	11	c.1239dup	p.(Asp414Argfs*44)	E11	CRC (37 y)
Other HC syndromes	12	c.137G>T	p.Ser46Ile	E02	OV (66 y)
	13	c.137G>T	p.Ser46Ile	E02	BR (33 y)
	14[†]	c.989-2A>G	p.?	I09*	PAN (68 y)
	15[†]	c.989-2A>G	p.?	I09	LG (18 y) BR (43 y)
	16	c.2341C>T	p.(Gln781*)	E14	BR (49 y), PAN (49 y)

(*table continues*)

Tumors that were not confirmed by medical reports have the suffix _nc (not confirmed). Cancer family history is broken down by first-degree relatives and second- and third-degree relatives. Each bullet point refers to an individual.

[†]These two individuals are family members.

BBC, basocellular carcinoma; BileDuc, bile duct cancer; BL, bladder cancer; BR, breast cancer; CNS, central nervous system; CRC, colorectal cancer; E, exon; ENDO, endometrial cancer; FDR, first-degree relative; HC, hereditary cancer; I, intron; ID, identification; IHC, immunohistochemistry; KID, kidney cancer; LG, lung cancer; LK, leukemia; LS, Lynch syndrome; MMR, mismatch repair; OV, ovarian cancer; PAN, pancreatic cancer; PV, pathogenic variant; SDR, second-degree relative; STO, stomach cancer; TDR, third-degree relative; U, unknown.

samples previously analyzed by LR-PCR, was analyzed. An increased sensitivity in variant identification compared with the routine diagnostic pipeline was found, increasing it from 0.907 (95% CI, 0.841 to 0.972) to 0.973 (95% CI, 0.937 to 1), while maintaining specificity (0.999) (Table 3). This improvement allowed us to identify all pathogenic variants (Supplemental Table S3), but not two benign polymorphic PSVs with a VAF below 60% (Supplemental Figure S5). These variants were ignored intentionally according to the decision-making algorithm, and in agreement with their benign classification, to reduce the number of LR-PCR confirmations needed.

In the analysis of 40 samples from the validation cohort, the PMS2_vaR pipeline improved sensitivity significantly from 0.853 (95% CI, 0.807 to 0.899) to 0.956 (95% CI, 0.930 to 0.983), in comparison with the routine diagnostic pipeline (McNemar test; score = 24; $P = 9.634 \times 10^{-7}$). Again, all pathogenic variants were identified (Table 3 and Supplemental Table S3). As in the optimization cohort B, there were variants (10 in this case) that were not called by the PMS2_vaR pipeline (Supplemental Figure S5), and all

Table 4 (*continued*)

Family history of FDR	Family history of SDR or TDR	MMR expression in proband's tumors	Comments	True variant?	Prevalence
• BR (50 y), ENDO (55 y)	No	PMS2^{-}	*BRCA2* PV carrier	Yes	38.462% (5/13)
• OV (58 y), ENDO (58 y) • PAN (56 y)	No	PMS2^{-}	IHC conserved of the FDR OV cancer	Yes	
No	No	PMS2^{-}		Yes	
• ENDO (46 y) • PR (79 y)	• STO_nc (55 y)	PMS2^{-}		Yes	
No	• LK_nc (3 y)	PMS2^{-}		Yes	
					0% (0/16)
• ENDO (55 y)	• CRC (U y) • ENDO (75 y)	MSH6 and PMS2^{-}		Yes	0.627% (5/798)
• BR (61 y) • BL (77 y)	• STO (54 y) • CRC (58 y) • CRC (83 y) • CRC_nc (55 y) • STO_nc (37 y)	U	MMR conserved expression in the SDR STO and CRC (58 y) cancers	Pseudogenic	
• BileDuc_nc (58 y)	• ENDO_nc (40 y) • CNS_nc (30 y)	MLH1 and PMS2^{-}		Yes	
• ENDO (58 y) • CRC (61 y)	No	MSH6 and PMS2^{-}		Yes	
No	No	MSH6 and PMS2^{-}		Yes	
No	• BR_nc (55 y) • PAN_nc (55 y)	MSH6 heterogenous expression and PMS2^{-}		Yes	
No	• KID (44 y) • BR (70 y) • BR_nc (55 y) • PAN_nc (63 y) • CRC_nc (80 y)	PMS2 heterogenous expression, MLH1 conserved		Yes	0.105% (5/4772)
No	• CRC_nc (44 y)	Conserved		Yes	
• BR (43 y)	• CRC_nc (60 y)	Conserved		Yes	
• BR (61 y), CRC (64 y) • PAN (68 y)	• CRC _nc (66 y)	Conserved (LG and BR)	MMR conserved expression in FDR CRC and PAN tumors	Yes	
• PR (77 y)	• ENDO_nc (74 y)	Conserved (PAN)	*BRCA2* germline PV carrier (proband)	Yes	

of them corresponded to polymorphic PSVs with a VAF below 60%, classified as benign following Insight-ClinGen MMR-specific guidelines.

Prevalence of *PMS2* Pathogenic Variants in a Hereditary Cancer Cohort

The implementation of the refined PMS2_vaR pipeline in samples from a HC cohort of 5619 patients identified 16 samples harboring a putative (likely)-pathogenic *PMS2* variant (0.285%) (Table 4). Subsequent (LR-)PCR analysis confirmed a *bona fide PMS2* variant in 15 of these 16 cases: five patients harbored tumors showing exclusive PMS2 loss with immunohistochemistry, five patients met other LS suspicion criteria [four displayed tumor DNA mismatch repair protein Msh6 (MSH6)/mismatch repair endonuclease PMS2 (PMS2) loss and one exhibited DNA mismatch repair protein Mlh1 (MLH1)/PMS2 loss], and five individuals were tested for other HC suspicions (PMS2 expression was later reported as heterogeneous in one ovarian tumor and MMR expression was conserved in the remaining four tumors) (Table 4).

Only variant c.904-1G>A in intron 8, found in a patient with an unavailable tumor, was found to be pseudogenic (case

7) (Table 4). Colorectal and stomach cancers of their relatives showed preserved MMR protein expression. The alignment of the region, assessed with the Integrative Genomics Viewer, showed that the variant was in *cis* with PSVs, supporting its pseudogenic origin (Supplemental Figure S6).

Recurrent pseudogenic exon 13 variants, c.2186_2187del and c.2243_2246del, were detected in 39 and 30 samples of the HC cohort, respectively, at a VAF ranging from 12.28% to 34.67% (Supplemental Table S4). None of them had clinical criteria or tumor molecular characteristics suggesting a *PMS2* alteration, thus LR-PCR was not performed according to the proposed algorithm.

Before implementing PMS2_vaR, 3 LR-PCRs and 15 short PCRs of *PMS2* (Supplemental Figure S3) were conducted on each sample exhibiting exclusive loss of PMS2 in immunohistochemistry. Therefore, the analysis of the 13 PMS2-suspected samples of the HC cohort (Table 1) resulted in 39 LR-PCR and 195 short PCRs. With the implementation of PMS2_vaR, only five samples were recommended for PCR analysis (two LR-PCRs and five short PCRs), resulting in a reduction of 95% of LR-PCRs and 99% of short PCRs.

Discussion

Gene panels are used widely for genetic testing purposes in HC. However, they face challenges when detecting variants in genes that share high homology with pseudogenes.[8] LR-PCR using primers outside the highly homologous regions is the gold standard for discriminating these cases.[7,15,16] Nonetheless, because of its complexity and high costs, it becomes unfeasible as a screening tool in most clinical contexts. In this work, PMS2_vaR was developed, a pipeline designed to address this clinical need in the mutational analysis of the *PMS2* gene. This open-source code uses data already available as the output of a standard panel testing analysis and requires the installation of a few commonly used bioinformatic tools, making it easy to implement in diagnostic pipelines. The pipeline identifies candidate *PMS2* variants and classifies them according to the variant classification list provided. Only samples carrying putative (likely) pathogenic *PMS2* variants are recommended for subsequent LR-PCR analysis.

Our results demonstrated substantial clinical improvements, significantly increasing sensitivity for variant identification from 0.853 to 0.956 in the validation cohort while preserving specificity. Notably, all pathogenic variants were identified. PSVs were regarded as potential gene variants only if their VAF was over 60%, reducing the number of samples requiring confirmation by LR-PCR or cDNA analysis. As an illustration, in the HC cohort, consisting of 5619 samples, the pipeline only recommended (LR-)PCR analysis in 16 cases (0.28%), a number that can be handled in a routine clinical setting. The implementation of PMS2_vaR significantly reduced the number of required PCR analyses, highlighting its efficiency within the diagnostic workflow. By selectively targeting candidate variants identified by PMS2_vaR, this study was able to streamline the analysis process, minimizing unnecessary PCR reactions and conserving valuable resources. Moreover, this also accelerates the diagnostic process, ultimately reducing the turnaround times of the reports.

The selection of invariant positions to filter candidate gene variants in exon 11 was based on the analysis of 707 patient samples.[11] In rare cases, this method may lead to erroneous variant assignments owing to gene conversion−related sequence exchange. Therefore, *PMS2* gene variants potentially might be lost when following the PMS2_vaR algorithm. To reduce this possibility, *PMS2* (likely) pathogenic variants filtered out after the E11 approach should be confirmed by LR-PCR if tumor molecular characteristics are indicative of a *PMS2* alteration. A similar strategy is recommended for the recurrent pseudogenic c.2186_2187del and c.2243_2246del variants in exon 13.

The integration of the PMS2_vaR pipeline into the daily diagnostics routine may produce a notable clinical impact by improving the identification of CMMRD and LS. Because of the complexity of clinically diagnosing CMMRD,[4,5] an accurate and prompt diagnosis is essential for genetic counseling, surveillance recommendations, as well as therapeutic decisions.[3] In contrast, the identification of germline *PMS2* monoallelic variant carriers is more controversial because of its relatively lower penetrance compared with the other MMR genes,[2,26,27] although significant phenotypic variability has been observed among monoallelic carriers, even between individuals from the same family.[7,26,27] The use of effective screening tools for accurate *PMS2* variant detection will help in refining the LS phenotype associated with germline alterations in this gene.

MMR genes, including *PMS2,* are considered clinically actionable because pathogenic variant identification has high benefits for the patient and family in clinical practice.[28] The American College of Medical Genetics proposed reporting secondary incidental findings in these genes in clinical exome and genome sequencing analyses.[29] For panel testing under HC suspicion, analysis of the *MLH1*, *MSH2*, *MSH6*, *BRCA1*, and *BRCA2* genes has been recommended as opportunistic testing in adult cancer patients, regardless of the main clinical phenotype.[23] Expanding this framework to encompass *PMS2* requires the availability of optimized pipelines such as PMS2_vaR.

Nine of the 10 *PMS2* pathogenic variant carriers identified in the LS suspicion cohorts harbored tumors displaying isolated PMS2 or MSH6/PMS2 loss patterns, highly indicative of PMS2 deficiency, as the main driver of carcinogenesis. In contrast, in the HC cohort, four of the five *PMS2* carriers identified presented tumors showing conserved MMR expression. Although an immunohistochemistry test can yield false-negative results, especially

for missense MMR variants,[30] this also could agree with recent findings describing that some individuals carrying *PMS2* pathogenic variants may develop MMR-proficient tumors.[31] Nevertheless, the prevalence of *PMS2* pathogenic variants was enriched in the HC cohort (0.285%) compared with the estimated prevalence in the general population (0.140%).[1]

As a limitation, the PMS2_vaR pipeline has not been optimized to detect copy number variants in the *PMS2* gene. However, the assessment of copy number variant detection tools tailored for panel data using the modified BAM files obtained by PMS2_vaR represents a promising strategy for the future. Of note, one of the major strengths of the PMS2_vaR tool is that it can be adapted for *PMS2* variant calling in the analysis of NGS panels, exomes, and genomes. Moreover, there is potential for extension to other genes in the same situation through necessary code adjustments (eg, the *PRSS1* gene in the context of HC gene panels).

Conclusions

We have developed a pipeline to improve the accuracy of *PMS2* genetic testing by using standard NGS diagnostic workflows. The results show that its use reduces the number of samples that need to undergo LR-PCR and clearly improves the identification of *PMS2* variant carriers.

Acknowledgments

We thank the patients who participated in this study and the members of the Genetic Counseling Units and the Genetic Diagnostics Laboratory of the Hereditary Cancer Program of the Catalan Institute of Oncology.

Disclosure Statement

None declared.

Supplemental Data

Supplemental material for this article can be found at *http://doi.org/10.1016/j.jmoldx.2024.05.005*.

References

1. Win AK, Jenkins MA, Dowty JG, Antoniou AC, Lee A, Giles GG, Buchanan DD, Clendenning M, Rosty C, Ahnen DJ, Thibodeau SN, Casey G, Gallinger S, Le Marchand L, Haile RW, Potter JD, Zheng Y, Lindor NM, Newcomb PA, Hopper JL, MacInnis RJ: Prevalence and penetrance of major genes and polygenes for colorectal cancer. Cancer Epidemiol Biomarkers Prev 2017, 26:404
2. 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
3. Tabori U, Hansford JR, Achatz MI, Kratz CP, Plon SE, Frebourg T, Brugieres L: Clinical management and tumor surveillance recommendations of inherited mismatch repair deficiency in childhood. Clin Cancer Res 2017, 23:e32−e37
4. Bakry D, Aronson M, Durno C, Rimawi H, Farah R, Alharbi QK, Alharbi M, Shamvil A, Ben-Shachar S, Mistry M, Constantini S, Dvir R, Qaddoumi I, Gallinger S, Lerner-Ellis J, Pollett A, Stephens D, Kelies S, Chao E, Malkin D, Bouffet E, Hawkins C, Tabori U: Genetic and clinical determinants of constitutional mismatch repair deficiency syndrome: report from the constitutional mismatch repair deficiency consortium. Eur J Cancer 2014, 50:987−996
5. Wimmer K, Kratz CP, Vasen HFA, Caron O, Ruiz-Ponte C, Slavc I, Burkhardt B, Brugieres L: Diagnostic criteria for constitutional mismatch repair deficiency syndrome: suggestions of the European consortium "Care for CMMRD" (C4CMMRD) on behalf of the EU-Consortium Care for CMMRD (C4CMMRD). J Med Genet 2014, 51:283−293
6. Andini KD, Nielsen M, Suerink M, Helderman NC, Koornstra JJ, Ahadova A, Kloor M, Mourits MJE, Kok K, Sijmons RH, Bajwa−ten Broeke SW: PMS2-associated Lynch syndrome: past, present and future. Front Oncol 2023, 13:547
7. Senter L, Clendenning M, Sotamaa K, Hampel H, Green J, Potter JD, Lindblom A, Lagerstedt K, Thibodeau SN, Lindor NM, Young J, Winship I, Dowty JG, White DM, Hopper JL, Baglietto L, Jenkins MA, de la Chapelle A: The clinical phenotype of Lynch syndrome due to germline PMS2 mutations. Gastroenterology 2008, 135:419
8. Lincoln SE, Hambuch T, Zook JM, Bristow SL, Hatchell K, Truty R, Kennemer M, Shirts BH, Fellowes A, Chowdhury S, Klee EW, Mahamdallie S, Cleveland MH, Vallone PM, Ding Y, Seal S, DeSilva W, Tomson FL, Huang C, Garlick RK, Rahman N, Salit M, Kingsmore SF, Ferber MJ, Aradhya S, Nussbaum RL: One in seven pathogenic variants can be challenging to detect by NGS: an analysis of 450,000 patients with implications for clinical sensitivity and genetic test implementation. Genet Med 2021, 239:1673−1680
9. Nicolaides NC, Carter KC, Shell BK, Papadopoulos N, Vogelstein B, Kinzler KW: Genomic organization of the human PMS2 gene family. Genomics 1995, 30:195−206
10. Nakagawa H, Lockman JC, Frankel WL, Hampel H, Steenblock K, Burgart LJ, Thibodeau SN, De La Chapelle A: Mismatch repair gene PMS2: disease-causing germline mutations are frequent in patients whose tumors stain negative for PMS2 protein, but paralogous genes obscure mutation detection and interpretation. Cancer Res 2004, 64: 4721−4727
11. Gould GM, Grauman PV, Theilmann MR, Spurka L, Wang IE, Melroy LM, Chin RG, Hite DH, Chu CS, Maguire JR, Hogan GJ, Muzzey D: Detecting clinically actionable variants in the 3′ exons of PMS2 via a reflex workflow based on equivalent hybrid capture of the gene and its pseudogene. BMC Med Genet 2018, 19:176
12. Jansen AML, Tops CMJ, Ruano D, van Eijk R, Wijnen JT, ten Broeke S, Nielsen M, Hes FJ, van Wezel T, Morreau H: The complexity of screening PMS2 in DNA isolated from formalin-fixed paraffin-embedded material. Eur J Hum Genet 2020, 28:333
13. Hayward BE, De Vos M, Valleley EMA, Charlton RS, Taylor GR, Sheridan E, Bonthron DT: Extensive gene conversion at the PMS2 DNA mismatch repair locus. Hum Mutat 2007, 28:424−430
14. Ganster C, Wernstedt A, Kehrer-Sawatzki H, Messiaen L, Schmidt K, Rahner N, Heinimann K, Fonatsch C, Zschocke J, Wimmer K: Functional PMS2 hybrid alleles containing a pseudogene-specific missense variant trace back to a single ancient intrachromosomal recombination event. Hum Mutat 2010, 31:552−560
15. Clendenning M, Hampel H, LaJeunesse J, Lindblom A, Lockman J, Nilbert M, Senter L, Sotamaa K, De La Chapelle A: Long-range PCR facilitates the identification of PMS2-specific mutations. Hum Mutat 2006, 27:490−495
16. Vaughn CP, Robles J, Swensen JJ, Miller CE, Lyon E, Mao R, Bayrak-Toydemir P, Samowitz WS: Clinical analysis of PMS2:

mutation detection and avoidance of pseudogenes. Hum Mutat 2010, 31:588−593
17. van der Klift HM, Mensenkamp AR, Drost M, Bik EC, Vos YJ, Gille HJJP, Redeker BEJW, Tiersma Y, Zonneveld JBM, García EG, Letteboer TGW, Olderode-Berends MJW, van Hest LP, van Os TA, Verhoef S, Wagner A, van Asperen CJ, ten Broeke SW, Hes FJ, de Wind N, Nielsen M, Devilee P, Ligtenberg MJL, Wijnen JT, Tops CMJ: Comprehensive mutation analysis of PMS2 in a large cohort of probands suspected of Lynch syndrome or constitutional mismatch repair deficiency syndrome. Hum Mutat 2016, 37:1162−1179
18. Etzler J, Peyrl A, Zatkova A, Schildhaus HU, Ficek A, Merkelbach-Bruse S, Kratz CP, Attarbaschi A, Hainfellner JA, Yao S, Messiaen L, Slave I, Wimmer K: RNA-based mutation analysis identifies an unusual MSH6 splicing defect and circumvents PMS2 pseudogene interference. Hum Mutat 2008, 29:299−305
19. Wagner J, Olson ND, Harris L, Khan Z, Farek J, Mahmoud M, et al: Benchmarking challenging small variants with linked and long reads. Cell Genom 2022, 2:100128
20. Schwenk V, Leal Silva RM, Scharf F, Knaust K, Wendlandt M, Häusser T, Pickl JMA, Steinke-Lange V, Laner A, Morak M, Holinski-Feder E, Wolf DA: Transcript capture and ultradeep long-read RNA sequencing (CAPLRseq) to diagnose HNPCC/Lynch syndrome. J Med Genet 2023, 60:747−759
21. Castellanos E, Gel B, Rosas I, Tornero E, Santín S, Pluvinet R, Velasco J, Sumoy L, Del Valle J, Perucho M, Blanco I, Navarro M, Brunet J, Pineda M, Feliubadaló L, Capellá G, Lázaro C, Serra E: A comprehensive custom panel design for routine hereditary cancer testing: preserving control, improving diagnostics and revealing a complex variation landscape. Sci Rep 2017, 71:1−12
22. Moreno-Cabrera JM, del Valle J, Feliubadaló L, Pineda M, González S, Campos O, Cuesta R, Brunet J, Serra E, Capellà G, Gel B, Lázaro C: Screening of CNVs using NGS data improves mutation detection yield and decreases costs in genetic testing for hereditary cancer. J Med Genet 2020, 59:75−78
23. Feliubadaló L, López-Fernández A, Pineda M, Díez O, del Valle J, Gutiérrez-Enríquez S, Teulé A, González S, Stjepanovic N, Salinas M, Capellá G, Brunet J, Lázaro C, Balmaña J; Catalan Hereditary Cancer Group: Opportunistic testing of BRCA1, BRCA2 and mismatch repair genes improves the yield of phenotype driven hereditary cancer gene panels. Int J Cancer 2019, 145:2682−2691
24. Munté E, Feliubadaló L, Pineda M, Tornero E, Gonzalez M, Moreno-Cabrera JM, Roca C, Bales Rubio J, Arnaldo L, Capellá G, Mosquera JL, Lázaro C: vaRHC: an R package for semi-automation of variant classification in hereditary cancer genes according to ACMG/AMP and gene-specific ClinGen guidelines. Bioinformatics 2023, 39:btad128
25. Pan S, Brown A, Leclair B, Elias M, Chen D, Kidd J, Bowles K, Coffee B, Roa B, Mancini-Dinardo D: Common variants in PMS2CL that can present in PMS2 as pathogenic variants with extremely low frequencies. Houston TX, American Socierty of Human Genetics, 2019, [abstract 2424], October 15-19, 2019
26. Ten Broeke SW, Klift HMV, Tops CMJ, Aretz S, Bernstein I, Buchanan DD, et al: Cancer risks for PMS2-associated Lynch syndrome. J Clin Oncol 2018, 36:2961
27. Ten Broeke SW, Brohet RM, Tops CM, Van Der Klift HM, Velthuizen ME, Bernstein I, Munar GC, Garcia EG, Hoogerbrugge N, Letteboer TGW, Menko FH, Lindblom A, Mensenkamp AR, Moller P, Van Os TA, Rahner N, Redeker BJW, Sijmons RH, Spruijt L, Suerink M, Vos YJ, Wagner A, Hes FJ, Vasen HF, Nielsen M, Wijnen JT: Lynch syndrome caused by germline PMS2 mutations: delineating the cancer risk. J Clin Oncol 2015, 33:319−325
28. Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, Herman GE, Hufnagel SB, Klein TE, Korf BR, McKelvey KD, Ormond KE, Sue Richards C, Vlangos CN, Watson M, Martin CL, Miller DT; on behalf of the ACMG Secondary Findings Maintenance Working Group: Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genet Med 2017, 19:249−255
29. Miller DT, Lee K, Abul-Husn NS, Amendola LM, Brothers K, Chung WK, Gollob MH, Gordon AS, Harrison SM, Hershberger RE, Klein TE, Richards CS, Stewart DR, Martin CL: ACMG SF v3.2 list for reporting of secondary findings in clinical exome and genome sequencing: a policy statement of the American College of Medical Genetics and Genomics (ACMG). Genet Med 2023, 25:100866
30. Hechtman JF, Rana S, Middha S, Stadler ZK, Latham A, Benayed R, Soslow R, Ladanyi M, Yaeger R, Zehir A, Shia J: Retained mismatch repair protein expression occurs in approximately 6% of microsatellite instability-high cancers and is associated with missense mutations in mismatch repair genes. Mod Pathol 2020, 33:871−879
31. Ranganathan M, Sacca RE, Trottier M, Maio A, Kemel Y, Salo-Mullen E, Catchings A, Kane S, Wang C, Ravichandran V, Ptashkin R, Mehta N, Garcia-Aguilar J, Weiser MR, Donoghue MTA, Berger MF, Mandelker D, Walsh MF, Carlo M, Liu YL, Cercek A, Yaeger R, Saltz L, Segal NH, Mendelsohn RB, Markowitz AJ, Offit K, Shia J, Stadler ZK, Latham A: Prevalence and clinical implications of mismatch repair-proficient colorectal cancer in patients with Lynch syndrome. JCO Precis Oncol 2023, 7:e2200675