Discovery of genetic defects in unexplained colorectal cancer syndromes

Jansen, A.M.L.

Chapter 5

The complexity of screening *PMS2* in DNA isolated from formalin-fixed paraffin-embedded tissue

AML Jansen[1,2], CMJ Tops[3], D Ruano[1], R van Eijk[1],
JT Wijnen[3], FJ Hes[3], T van Wezel[1], H Morreau[1]

In submission

Department of [1]Pathology, [2]Human Genetics and [3]Clinical Genetics,
Leiden University Medical Center, Leiden, The Netherlands

Abstract

Background and Aims: Germline variants in the DNA mismatch repair (MMR) gene *PMS2* represent 1-14% of all MMR gene variants. Correct variant analysis of *PMS2* is complex due to the presence of multiple pseudogenes and the occurrence of gene conversion. The complexity of analysis increases in highly fragmented DNA from formalin-fixed paraffin-embedded (FFPE) tissue. We now describe and test a reliable approach to detect true *PMS2* variants in fragmented DNA.

Methods: A custom NGS panel meant for FFPE tissue was used targeting four MMR genes, *POLE* and *POLD1*. Amplicon design for *PMS2* was based on the position of paralogous sequence variants (PSV) that distinguish *PMS2* from its pseudogenes. We screened 125 MMR-deficient tumors for *PMS2* variants.

Results: We present an overview of PSVs that can be used for reliable distinction between *PMS2* and its pseudogenes. *PMS2* variants in exons 1-11 can be correctly curated on basis of this information. For exons 12-15 this is less reliable as these undergo gene conversion. Of the 125 tumors tested, six were unexplained MMR-deficient tumors with solitary PMS2 protein expression loss. In these six tumors three unclassified variants (class 3) and four (likely) pathogenic variants (class 4 and 5) were detected in *PMS2*. One microsatellite unstable tumor with positive staining for all MMR proteins was found to carry a frameshift *PMS2* variant. No pathogenic *PMS2* variants were detected in tumors with other patterns of MMR protein expression loss.

Conclusions: With a paired-end NGS approach with one or two PSVs in every amplicon, variants can reliably be detected in exons 1 to 11 of *PMS2*.

Keywords: *PMS2,* variant, next-generation sequencing, paralogous sequence variant.

Introduction

Pathogenic heterozygous germline variants in the MMR genes cause Lynch Syndrome (LS), an autosomal dominant predisposition for colorectal-, endometrial- and other cancers.[1] While the majority of the causal variants are found in *MLH1* and *MSH2*, variants in the less frequent mutated *PMS2* represent 1-14% of all MMR gene variants.[2, 3] The colorectal cancer (CRC) risk of *PMS2* variant carriers has shown to be much lower compared to *MLH1, MSH2* and *MSH6,* with risk of CRC around 11-19% by the age of 70 years.[4] Homozygous or compound heterozygous variants in the *PMS2* gene are seen more often in patients with constitutional mismatch repair deficiency (CMMRD), a recessive disorder characterized by CRC and childhood hematological- and brain malignancies.[5]

The analysis of *PMS2* is complex due to the presence of multiple pseudogenes.[3, 6, 7] Fourteen *PMS2*-pseudogenes share a high homology with the 5' end of *PMS2* (exon 1 to 5), while a fifteenth pseudogene (*PMS2CL*) shares high homology with *PMS2* exon 9 and exon 11 to 15.[2, 7-10] An additional complexity is added due to ongoing gene conversion events between *PMS2* and *PMS2CL*.[10] Germline variant screening strategies propose long-range PCR with a reverse primer in *PMS2* exon 6 or propose designing multiplex ligation-dependent amplification (MLPA) probes, and PCR primers based on paralogous sequence variants (PSVs) to distinguish *PMS2* exon 1 to 5 from the fourteen homologous pseudogenes.[2, 9, 11] These PSVs are specific nucleotides that differ between *PMS2* and the pseudogenes, and enable differentiation between two almost complete homologues sequences.[3, 8, 9] This strategy is not reliable in detecting variants in exon 12 to 15 due to gene conversion events between *PMS2* and *PMS2CL*.[10, 12] Through crossover the sequence corresponding to *PMS2* or *PMS2CL* could be present as the exon 12 to 15 sequence of *PMS2,* and subsequently expressed.[9-11] To determine which sequence is present, and expressed, long-range PCR on gDNA or cDNA is proposed using primers in the unique exon 10 and a nonspecific reverse primer in the 3' UTR.[9, 10, 12-14]

While this strategy is very suitable for reliable detection of *PMS2* variants in leukocyte DNA, it is not applicable when using DNA isolated from formalin-fixed paraffin-embedded (FFPE) tissue blocks, which is highly fragmented.[15] It has been recently shown that in a large proportion of MMR-deficient tumors without pathogenic germline MMR variant and without *MLH1* promoter hypermethylation, two somatic MMR variants can explain the MMR-deficiency.[16-18] This occurrence of somatic MMR inactivation also shows the need for reliable detection of somatic *PMS2* variants in DNA isolated from FFPE tissue. Testing DNA isolated from FFPE furthermore enables detection of variants in (deceased) index patients of which only FFPE is available. Furthermore, to implement reliable *PMS2* variant screening in molecular tumor diagnostics, a high-throughput strategy should be developed.

Some studies only focus on screening for variants in *MLH1, MSH2* and *MSH6,* possibly because of the complexity of screening for true *PMS2* variants.[16] We now describe possible pitfalls in *PMS2* variant detection and a next-generation sequencing (NGS)-based approach for reliable somatic and germline *PMS2* testing in FFPE DNA.

Materials and Methods

Study Cohort

Colorectal and endometrial cancers and when available matching normal tissue of 40 patients with unexplained MMR-deficiency was screened for DNA variants with NGS to detect variants in the MMR genes in a diagnostic setting. All tumors were pre-screened with immunohistochemical (IHC) staining of the four MMR proteins and the majority (83%) showed expression loss of one or two of the MMR proteins. Many of these tumors were previously screened for microsatellite instability and showed to have high microsatellite instability (MSI-H). All MLH1/PMS2 negative tumors were tested for *MLH1* promoter hypermethylation, and somatic NGS was performed if no methylation was detected. Four tumors had solitary immunohistochemical expression loss of PMS2. Additionally, DNA isolated from FFPE tissue blocks of 85 unexplained suspected Lynch Syndrome patients (without germline MMR variants and without *MLH1* promoter hypermethylation) were screened with NGS for variants in the MMR genes in a research setting. Two of the latter tumors showed isolated PMS2 expression loss with IHC.

NGS panel

A custom paired end NGS library was designed covering *MLH1, MSH2, MSH6, PMS2, POLE* and *POLD1*. Ion AmpliSeq™ Custom Panels were designed with the Ion AmpliSeq™ Designer tool. Libraries were prepared with Ion AmpliSeq™ Library Kit 2.0 according to the manufacturer's protocol. The panel used in a diagnostic setting slightly differs from the research panel. The diagnostic panel covers the exonic regions with 99.2% coverage of *MLH1*, 99.3% coverage of *MSH2*, 100% coverage of *MSH6* and 76.5% of *PMS2* (exon 1-12) and the exonuclease domain of *POLE* (exon 7-14) and *POLD1* (exon 8-13). The research panel is comparable but covers 100% of *MLH1,* 94.9% of *MSH2,* 97.7% of *MSH6,* 79.1% of *PMS2* (exons 1-11 and exon 14) and *POLE* and *POLD1* completely. Next-generation sequencing was performed with the Ion Proton™ System (Life Technologies, Carlsbad, CA, USA).

NGS annotation

Raw data analysis, alignments, and variant calling was carried out using the default parameters in Torrent Suite. The unaligned BAM files generated by the Proton sequencer were mapped against the human reference genome (GRCh37/hg19) using the TMAP 5.0.7 software with default parameters (*https://github.com/iontorrent/TS*). A read is assigned to the genomic location with the highest mapping score. In case that a particular read gets the same alignment score at multiple locations, it will be randomly assigned to one of the loci. All (likely) pathogenic *PMS2* variants were visually inspected with the Integrative Genomics Viewer (IGV).[19, 20] The following Genbank reference sequences were used: NM_000249.3 for *MLH1*, NM_000251.2 for *MSH2*, NM_000179.2 for *MSH6*, NM_000535.5 for *PMS2*, NM_006231.2 for *POLE* and NM_001256849.1 for *POLD1*. Classification of the functional effects of the variants was done according to the five-tiered InSiGHT scheme.[21]

Results

A custom paired-end MMR panel was designed for detecting variants in DNA isolated from formalin-fixed paraffin-embedded (FFPE) tissue. On average, amplicons are 100-175 bp in size in order to be able to amplify fragmented DNA. For *PMS2*, a reliable screening panel could be made covering exons 1-11 only, with one or two paralogous sequence variants (PSV) in every amplicon. With exons 1-11 as target, a 72.9% coverage can be achieved. An overview of PSVs present in these eleven exons is shown in Table 1.

Table 1: paralogous sequence variants (PSVs) in *PMS2*

Target	Size (bp)	Amplicons needed	# pseudogenes	PSV
Exon 1	23	1	5	c.-13G>C (4/5) and c.-9G>A (1/5) c.1A>T (4/5) and c.-4_-5delinsAG (1/5)
Exon 2	140	1 or 2	13	c.24-4C>T (13/13) c.89A>C (11/13) c.117A>G (13/13) c.121G>A (9/13) c.125T>A (9/13)
Exon 3	87	1 or 2	14	c.164-12delT (3/14) and c.209A>G (11/14) c.187G>A (10/14) c.195T>C (14/14) c.240C>T (3/14) and c.250+8G>A (11/14)
Exon 4	103	2	14	c.251-11C>G (13/14) and c.251-13C>T (1/14) c.299A>G (10/14) and c.298C>G (3/14) c.353+22C>T (14/14)
Exon 5	184	2	14	c.396A>G, c.406A>G, c.418G>T, c.429T>C, c.452G>C, c.478C>A, c.492C>T (all 14/14)
Exon 6	168	2	0	-
Exon 7	98	1 or 2	0	-
Exon 8	100	1 or 2	0	-
Exon 9	85	1 or 2	1 (PMS2CL)	c.924G>C, c.932A>G and c.934A>G
Exon 10	156	2	0	-
Exon 11	862	8 or 9	1 (PMS2CL)	c.1238_1239delAAinsGG, c.1360_1361delCTinsTC, c.1379G>A, c.1556A>G, c.1559C>T, c.1567T>A, c.1688_1689delGAinsAG, c.1714G>A, c.1717A>T, c.1730dupA, c.1732C>T, c.1740A>G, c.1760G>A, c.1771T>C, c.1795G>A, c.1798A>G, c.1855G>A, c.1863_1864delTA, c.1952A>G, c.2006+26C>A and c.2006+28delC

Overview of paralogous sequence variants (PSVs) in PMS2, #/#; number of pseudogenes with the variant/total number of pseudogenes for this exon.

Using the strategy described above and in the Material and Methods, 125 MMR-deficient tumors and, when available matching normal colonic mucosa, were screened for *PMS2* variants. Six tumors showed solitary immunohistochemical PMS2 expression loss. Four (likely) pathogenic *PMS2* variants (class 4 or 5), and three variants of uncertain significance (VUS class 3) were detected (Figure 1A and Table 2). The class 3 *PMS2* c.308C>T and c.1687C>T variants were both found in tumors with a variant in the exonuclease domain of *POLE*, where the *PMS2* variant is expected to be secondary to the *POLE* variant.[22] Additionally, one tumor with positive staining for all MMR proteins and an MSI-H phenotype was found to carry a frameshift *PMS2* c.325dupG variant. In remaining cases with combined MLH1/PMS2, combined MSH2/MSH6 or solitary MSH6 expression loss no pathogenic *PMS2* variant was detected.

For all variants the IGV was used to determine the presence of PSVs. An example is shown in Figure 1B for an exon 9 *PMS2* c.955C>A variant. The NGS amplicon containing the *PMS2* c.955C>A also contains three PSVs, c.934A>G, c.932A>G and c.924G>C. None of the reads showed any of these three PSVs, indicating that this variant is truly present in *PMS2* and not in the pseudogene *PMS2CL*. This was done for all eight *PMS2* variants shown in Table 2, and all variants were found to be present in *PMS2* and not one of the pseudogenes.

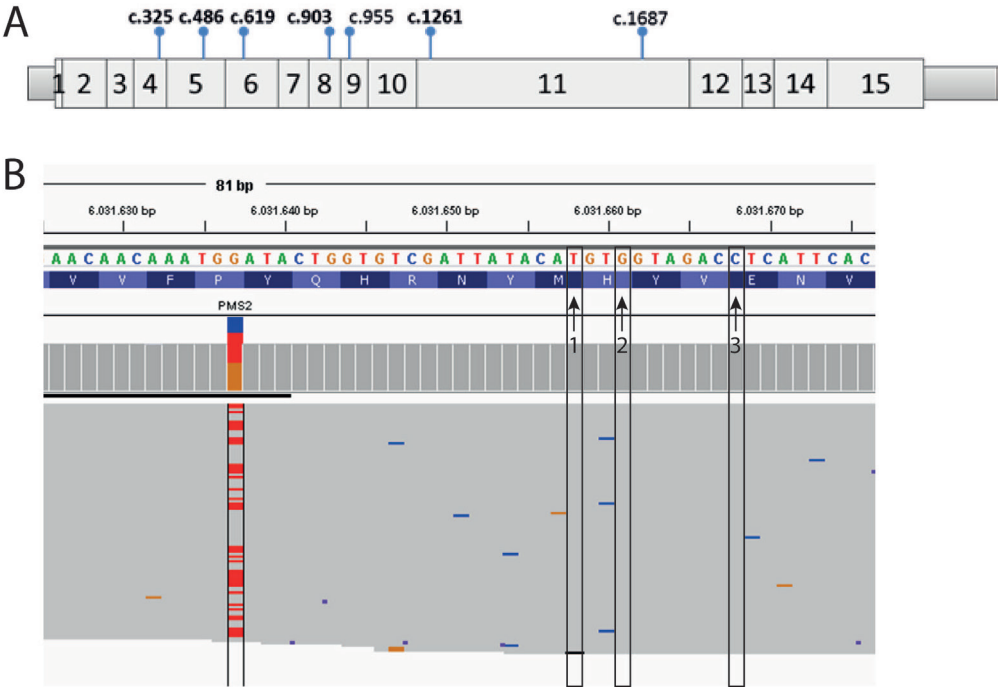

Figure 1: *PMS2* variants detected with NGS

[A] PMS2 variants found with NGS, Class 4 and class 5 variants shown in bold. [B] IGV printout of the PMS2 c.955C>A shown (in red). Arrows show the location of three PSVs present in the amplicon (1. c.934A>G, 2. c.932A>G and 3. c.924G>C). All three are absent in the reads, indicating that this variant is present in PMS2 and not one of the pseudogenes. PMS2 is shown in reverse complement, because PMS2 is present on the reverse strand.

Table 2: Overview PMS2 variants

Tumor characteristics	Variant	Exon	%	Class	PSVs in amplicon
CRC45, PMS2-, MSI-H	*PMS2* c.308C>T†	4	11%	3	c.299, c.298
CRC38, MMR+, MSI-H	*PMS2* c.325dupG	4	77%	4	c.299, c.298
CRC31, PMS2-, MSI-H	*PMS2* c.486delA	5	34%	4	c.406, c.418, c.429, c.452, c.478, c.492
CRC48, PMS2-, MSI unknown	*PMS2* c.619G>T	6	48%	4	NA
CRC67, PMS2-, MSI-H	*PMS2* c.903G>T	8	52%	4	NA
	PMS2 c.1261C>T	11	26%	5	c.1238_1239, c.1360_1361*
EC58, PMS2-, MSI unknown	*PMS2* c.955C>A	9	41%	3	c.924, c.932, c.934
EC55, PMS2-, MSI-H	*PMS2* c.1687C>T†	11	30%	3	c.1556, c.1559, c.1567, c.1688_1689

*Tumor is shown as type of tumor, followed by age of onset; CRC, colorectal cancer; EC, endometrial cancer; MSI, microsatellite instability, high (H) or unknown; PMS2-, PMS2 negative staining; MMR+, positive IHC MMR staining; % is variant allele frequency; NA, not applicable, this exon is unique; *, PSV is present in primer sequence; †variant present in POLE mutated tumor.*

Discussion

We now describe how to interpret *PMS2* variants present in DNA isolated from formalin-fixed paraffin-embedded (FFPE) tissue using paired-end targeted NGS with paralogous sequence variants (PSVs) in every amplicon. Six of the eight *PMS2* variants detected were located in exons that have high homology with one or more of the *PMS2* pseudogenes. Of all variants it could be concluded that they were truly present in *PMS2* and not in the pseudogenes by analysing the presence of PSVs in the amplicon (Table 2). Through automatically assigning the read to the genomic location with the highest mapping score, a reliable distinction could be made between *PMS2* and the pseudogenes, for *PMS2* exon 1 to 11. For exons 12 to 15 of *PMS2* it is not possible to reliably detect variants in FFPE derived DNA due to the existence of continuous gene conversion targeting these exons. The only solution to this challenge is long range PCR of fragments covering *PMS2* exons 12-15, but this is not applicable on the fragmented FFPE derived tissue DNA.[9, 10]

Studies that aim to detect *PMS2* variants in DNA from FFPE tissues are very limited. Only five studies describe somatic analysis of *PMS2*.[17, 18, 22-24] We and others achieve a coverage of 75-80% and do not sequence *PMS2* exon 12 to 15 completely, because variants cannot be curated in this region due to sequence exchange events. One previous study suggested full sequencing (100%) coverage of *PMS2* in tumor tissue, but did not fully explain how was coped with gene conversion of exons 12 to 15 (*http://tests.labmed.washington.edu/COLOSEQ#Introducing_ColoSeq.2BISI_Tumor*).[24] One *PMS2* splice site variant in intron 12 was shown without confirmation of its presence in *PMS2* and not in *PMS2CL* through gene conversion (previously shown to occur in 69% of tested individuals).[10] This example typically highlights the existing problem with sequencing of *PMS2* exons 12 to 15. Consensus should be reached whether it is preferable to test patients for variants in this region, without being able to confirm that the variant is expressed, or to not sequence this region and possibly missing somatic *PMS2* variants. In the currently described research panel exons 12 and exon 14 are included, but caution is needed when analysing the variants, since it cannot be established whether this variant is expressed. It could be considered that when a (likely) pathogenic *PMS2* exon 12 to 15 variant is detected in a tumor with solitary PMS2 loss of expression with no other *PMS2* variants, this variant is likely expressed and the cause of the immunohistochemical loss of PMS2 expression. Additionally, since expressed genes have elevated mutation rates, if a somatic variant is detected in *PMS2* exon 12-15 it is likely that *PMS2* is expressed.[25] However, it cannot be confirmed whether *PMS2* is truly expressed.

In conclusion, with a custom NGS panel with one or two PSVs we were able to reliably detect eight variants in *PMS2* exon 1 to 11 in six tumors with solitary PMS2 loss, and one tumor with positive MMR staining and microsatellite instability. Previous studies describe comprehensive strategies for accurate mutation detecting in *PMS2,* but mainly focus on testing genomic DNA extracted from blood.[9, 26] However, since recent studies have shown biallelic somatic inactivation of the MMR genes, there is a growing need for reliable detection of somatic variants in *PMS2*.[16-18, 22] With this guide we show a reliable method to detect *PMS2* variants in DNA from FFPE tissue.

References

1. Lynch HT, de la Chapelle A. Hereditary colorectal cancer. N Engl J Med 2003;348:919-32.
2. Borras E, Pineda M, Cadinanos J, et al. Refining the role of PMS2 in Lynch syndrome: germline mutational analysis improved by comprehensive assessment of variants. J Med Genet 2013;50:552-63.
3. Hendriks YM, Jagmohan-Changur S, van der Klift HM, et al. Heterozygous mutations in PMS2 cause hereditary nonpolyposis colorectal carcinoma (Lynch syndrome). Gastroenterology 2006;130:312-22.
4. ten Broeke SW, Brohet RM, Tops CM, et al. Lynch syndrome caused by germline PMS2 mutations: delineating the cancer risk. J Clin Oncol 2015;33:319-25.
5. Vasen HF, Ghorbanoghli Z, Bourdeaut F, et al. Guidelines for surveillance of individuals with constitutional mismatch repair-deficiency proposed by the European Consortium "Care for CMMR-D" (C4CMMR-D). J Med Genet 2014;51:283-93.
6. Kondo E, Horii A, Fukushige S. The human PMS2L proteins do not interact with hMLH1, a major DNA mismatch repair protein. J Biochem 1999;125:818-25.
7. Nicolaides NC, Carter KC, Shell BK, et al. Genomic organization of the human PMS2 gene family. Genomics 1995;30:195-206.
8. De Vos M, Hayward BE, Picton S, et al. Novel PMS2 pseudogenes can conceal recessive mutations causing a distinctive childhood cancer syndrome. Am J Hum Genet 2004;74:954-64.
9. van der Klift HM, Mensenkamp AR, Drost M, et al. Comprehensive Mutation Analysis of PMS2 in a Large Cohort of Probands Suspected of Lynch Syndrome or Constitutional Mismatch Repair Deficiency Syndrome. Hum Mutat 2016;37:1162-1179.
10. van der Klift HM, Tops CM, Bik EC, et al. Quantification of sequence exchange events between PMS2 and PMS2CL provides a basis for improved mutation scanning of Lynch syndrome patients. Hum Mutat 2010;31:578-87.
11. Wernstedt A, Valtorta E, Armelao F, et al. Improved multiplex ligation-dependent probe amplification analysis identifies a deleterious PMS2 allele generated by recombination with crossover between PMS2 and PMS2CL. Genes Chromosomes Cancer 2012;51:819-31.
12. Hayward BE, De Vos M, Valleley EM, et al. Extensive gene conversion at the PMS2 DNA mismatch repair locus. Hum Mutat 2007;28:424-30.
13. Clendenning M, Walsh MD, Gelpi JB, et al. Detection of large scale 3' deletions in the PMS2 gene amongst Colon-CFR participants: have we been missing anything? Fam Cancer 2013;12:563-6.
14. Etzler J, Peyrl A, Zatkova A, et al. RNA-based mutation analysis identifies an unusual MSH6 splicing defect and circumvents PMS2 pseudogene interference. Hum Mutat 2008;29:299-305.
15. Daugaard I, Kjeldsen TE, Hager H, et al. The influence of DNA degradation in formalin-fixed, paraffin-embedded (FFPE) tissue on locus-specific methylation assessment by MS-HRM. Experimental and Molecular Pathology 2015;99:632-640.
16. Mensenkamp AR, Vogelaar IP, van Zelst-Stams WA, et al. Somatic mutations in MLH1 and MSH2 are a frequent cause of mismatch-repair deficiency in Lynch syndrome-like tumors. Gastroenterology 2014;146:643-646 e8.
17. Geurts-Giele WR, Leenen CH, Dubbink HJ, et al. Somatic aberrations of mismatch repair genes as a cause of microsatellite-unstable cancers. J Pathol 2014;234:548-59.
18. Jansen AM, Geilenkirchen MA, van Wezel T, et al. Whole Gene Capture Analysis of 15 CRC Susceptibility Genes in Suspected Lynch Syndrome Patients. PLoS One 2016;11:e0157381.
19. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. Nat Biotech 2011;29:24-26.
20. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in Bioinformatics 2013;14:178-192.
21. Thompson BA, Spurdle AB, Plazzer JP, 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.
22. Jansen AM, van Wezel T, van den Akker BE, et al. Combined mismatch repair and POLE/POLD1 defects explain unresolved suspected Lynch syndrome cancers. Eur J Hum Genet 2015.
23. Stelloo E, Jansen AM, Osse EM, et al. Practical guidance for mismatch repair-deficiency testing in endometrial cancer. Ann Oncol 2016.
24. Haraldsdottir S, Hampel H, Tomsic J, et al. Colon and endometrial cancers with mismatch repair deficiency can arise from somatic, rather than germline, mutations. Gastroenterology 2014;147:1308-1316 e1.
25. Zhang J, Yang J-R. Determinants of the rate of protein sequence evolution. Nat Rev Genet 2015;16:409-420.
26. Li J, Dai H, Feng Y, et al. A Comprehensive Strategy for Accurate Mutation Detection of the Highly Homologous PMS2. J Mol Diagn 2015;17:545-53.