The handle http://hdl.handle.net/1887/62614 holds various files of this Leiden University dissertation.

**Author:** Jansen, A.M.L.
**Title:** Discovery of genetic defects in unexplained colorectal cancer syndromes
**Issue Date:** 2018-05-29
Chapter 5

The complexity of screening \textit{PMS2} in DNA isolated from formalin-fixed paraffin-embedded tissue

AML Jansen\textsuperscript{1,2}, CMJ Tops\textsuperscript{3}, D Ruano\textsuperscript{1}, R van Eijk\textsuperscript{1}, JT Wijnen\textsuperscript{3}, FJ Hes\textsuperscript{3}, T van Wezel\textsuperscript{1}, H Morreau\textsuperscript{1}

\textit{In submission}

Department of \textsuperscript{1}Pathology, \textsuperscript{2}Human Genetics and \textsuperscript{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. 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. 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. 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.

The analysis of PMS2 is complex due to the presence of multiple pseudogenes. 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. An additional complexity is added due to ongoing gene conversion events between PMS2 and PMS2CL. Germine 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. These PSVs are specific nucleotides that differ between PMS2 and the pseudogenes, and enable differentiation between two almost complete homologues sequences. This strategy is not reliable in detecting variants in exon 12 to 15 due to gene conversion events between PMS2 and PMS2CL. Through crossover the sequence corresponding to PMS2 or PMS2CL could be present as the exon 12 to 15 sequence of PMS2, and subsequently expressed. 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.

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. 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. 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. 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.
Chapter 5

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). 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.
The complexity of screening PMS2 in DNA isolated from FFPE tissue

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

<table>
<thead>
<tr>
<th>Target</th>
<th>Size (bp)</th>
<th>Amplicons needed</th>
<th># pseudogenes</th>
<th>PSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon 1</td>
<td>23</td>
<td>1</td>
<td>5</td>
<td>c.-13G&gt;C (4/5) and c.-9G&gt;A (1/5) c.1A&gt;T (4/5) and c.-4_-5delinsAG (1/5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>c.24-4C&gt;T (13/13) c.89A&gt;C (11/13)</td>
</tr>
<tr>
<td>Exon 2</td>
<td>140</td>
<td>1 or 2</td>
<td>13</td>
<td>c.117A&gt;G (13/13) c.121G&gt;A (9/13) c.125T&gt;A (9/13)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>c.164-12delT (3/14) and c.209A&gt;G (11/14)</td>
</tr>
<tr>
<td>Exon 3</td>
<td>87</td>
<td>1 or 2</td>
<td>14</td>
<td>c.187G&gt;A (10/14) c.195T&gt;C (14/14) c.240C&gt;T (3/14) and c.250+8G&gt;A (11/14)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>c.396A&gt;G, c.406A&gt;G, c.418G&gt;T, c.429T&gt;C, c.452G&gt;C, c.478C&gt;A, c.492C&gt;T (all 14/14)</td>
</tr>
<tr>
<td>Exon 4</td>
<td>103</td>
<td>2</td>
<td>14</td>
<td>c.299A&gt;G (10/14) and c.298C&gt;G (3/14) c.353+22C&gt;T (14/14)</td>
</tr>
<tr>
<td>Exon 5</td>
<td>184</td>
<td>2</td>
<td>14</td>
<td>c.396A&gt;G, c.406A&gt;G, c.418G&gt;T, c.429T&gt;C, c.452G&gt;C, c.478C&gt;A, c.492C&gt;T (all 14/14)</td>
</tr>
<tr>
<td>Exon 6</td>
<td>168</td>
<td>2</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Exon 7</td>
<td>98</td>
<td>1 or 2</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Exon 8</td>
<td>100</td>
<td>1 or 2</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>Exon 9</td>
<td>85</td>
<td>1 or 2</td>
<td>1 (PMS2CL)</td>
<td>c.924G&gt;C, c.932A&gt;G and c.934A&gt;G</td>
</tr>
<tr>
<td>Exon 10</td>
<td>156</td>
<td>2</td>
<td>0</td>
<td>-</td>
</tr>
</tbody>
</table>

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. 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

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

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
