Localization of Large Rearrangements in *PMS2* and *PMS2CL* using PacBio

Shujuan Pan[1]; Benjamin Obadia[1]; David Lee[1]; Jack Landon[1]; Paula Esquetini[1]; Benjamin Roa[1]; Debora Mancini-DiNardo[1]

1. Myriad Genetics, Inc., Salt Lake City, UT, USA

Myriad genetics®

Background

- *PMS2,* one of the mismatch repair genes associated with Lynch syndrome, is a gene often included in hereditary pan-cancer panels.[a]
- However, testing *PMS2* by short-read next generation sequencing (NGS) is challenging due to the presence of *PMS2CL,* an unprocessed pseudogene that is highly homologous to the 3' end of *PMS2.*[b]
- In this study, we evaluated the ability of long-read sequencing technology in detecting the large rearrangements (LRs)/copy number variations that had previously been determined to be in *PMS2* or *PMS2CL* through orthogonal confirmation.

Methods

- 18 de-identified samples were tested in this study, 17 of which carried an LR that was previously determined to be either in *PMS2* or *PMS2CL*. One sample carried a terminal LR where the specific gene location was undetermined.
- These LRs were first identified via short-read NGS platform, and the gene locations of these LRs were later determined using a combination of MLPA and a long-range PCR approach.
- An aliquot of each sample was fragmented to ~10kb followed by hyb-capture of regions of interest before proceeding to long-read sequencing with PacBio Sequel II at ~50x read depth.
- BAM files from PacBio were aligned to GRCh37 using Minimap2 v2.28 and Paraphase v3.4.0. Breakpoints were examined using IGV software and a concordance check was performed against the original findings.

a) Andini KD, Nielsen M, Suerink M, et al. PMS2-associated Lynch syndrome: Past, present and future. Front Oncol. 2023 Feb 21;13:1127329. doi: 10.3389/fonc.2023.1127329. **b)** Hayward BE, De Vos M, Valleley EM, et al. Extensive gene conversion at the PMS2 DNA mismatch repair locus. Hum Mutat. 2007 May;28(5):424-30. doi: 10.1002/humu.20457.

Results

- In 16 out of 17 samples with known gene locations, PacBio ascribed the LR to the correct gene. The concordance rate with the original findings was 94.1%.
- For one sample with a terminal LR whose location was not previously determined, PacBio showed the gene location to be in *PMS2.*
- Discernment of the correct location was based on the presence of breakpoints in the sequencing reads that were mostly located within deeper intronic regions that are not typically analyzed by short-read NGS.
- The one remaining sample was previously determined to carry a full gene deletion of *PMS2CL*. No breakpoints were present in either *PMS2* or *PMS2CL* on PacBio, and a conclusive call could not be made based solely on IGV analysis. This deletion likely extends beyond the capture capability of probes used in this test.

Figure 1. Orthogonal confirmation of a potential *PMS2* LR in Sample 5. **(A)** A potential *PMS2* exon 14 deletion detected by NGS panel and **(B)** confirmed via MLPA plus **(C)** long-range PCR followed by nested PCR and **(D)** Sanger sequencing under MLPA probes.

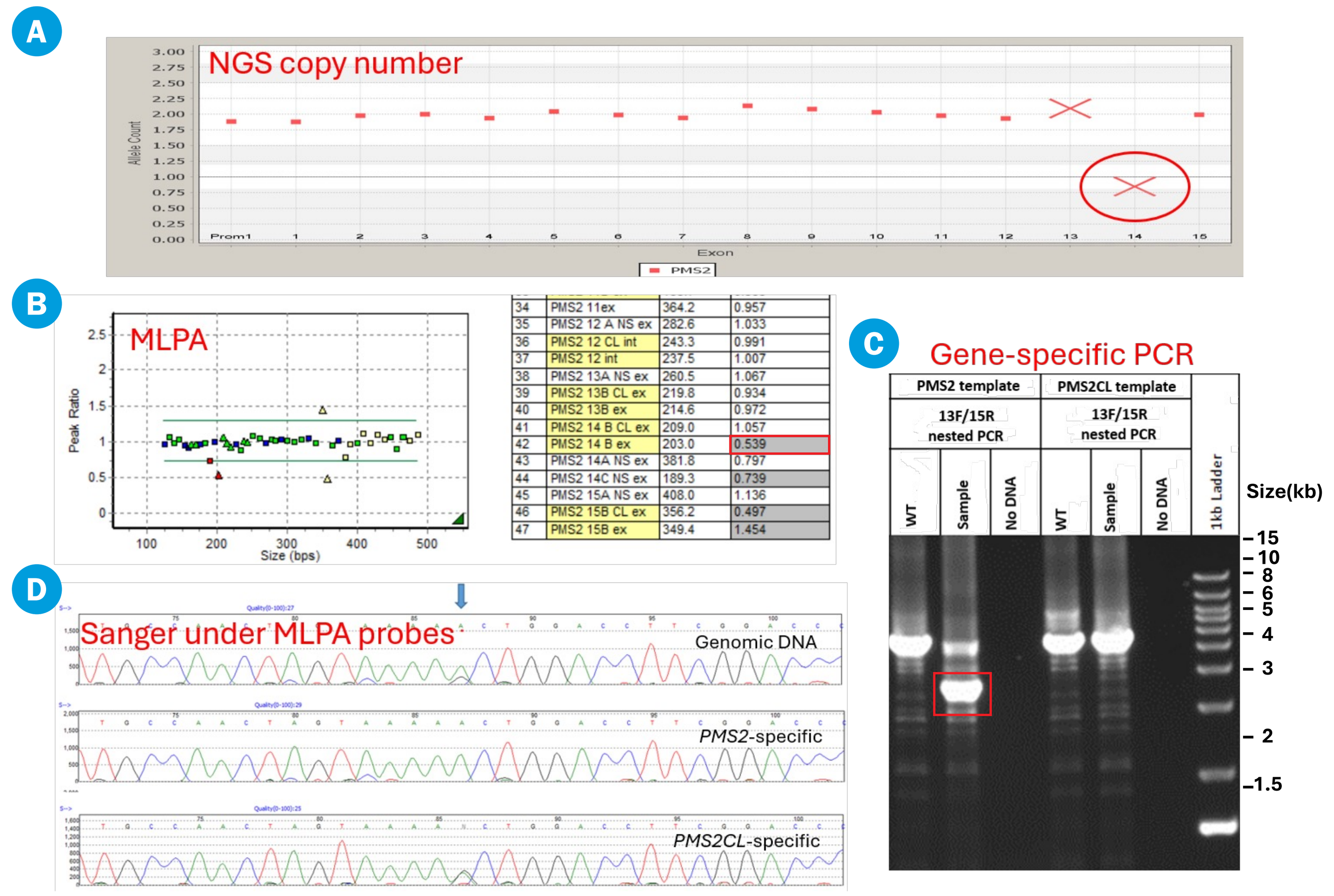

A
NGS copy number
B
MLPA
Peak Ratio
Size (bps)
34 PMS2 11ex 364.2 0.957
35 PMS2 12 A NS ex 282.6 1.033
36 PMS2 12 CL int 243.3 0.991
37 PMS2 12 int 237.5 1.007
38 PMS2 13A NS ex 260.5 1.067
39 PMS2 13B CL ex 219.8 0.934
40 PMS2 13B ex 214.6 0.972
41 PMS2 14 B CL ex 209.0 1.057
42 PMS2 14 B ex 203.0 0.539
43 PMS2 14A NS ex 381.8 0.797
44 PMS2 14C NS ex 189.3 0.739
45 PMS2 15A NS ex 408.0 1.136
46 PMS2 15B CL ex 356.2 0.497
47 PMS2 15B ex 349.4 1.454
C
Gene-specific PCR
PMS2 template
PMS2CL template
13F/15R nested PCR
WT
Sample
No DNA
1kb Ladder
Size(kb)
D
Sanger under MLPA probes
Genomic DNA
PMS2-specific
PMS2CL-specific

Figure 2. PacBio reads show a 1879bp deletion encompassing *PMS2* exon 14, consistent with NGS finding in Sample 5. No deletion was detected in *PMS2CL.*

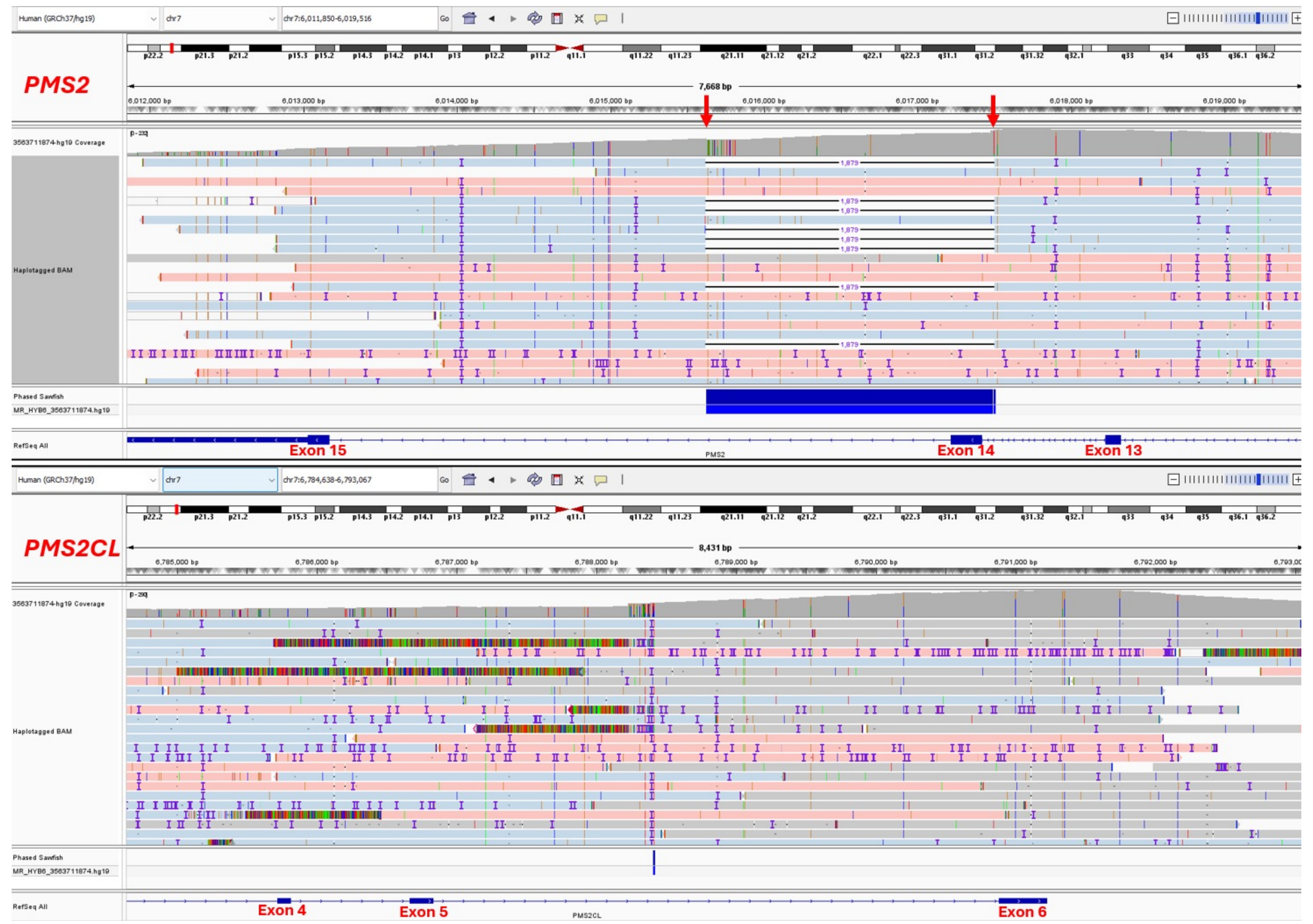

PMS2
Exon 15
Exon 14
Exon 13
PMS2CL
Exon 4
Exon 5
Exon 6

Figure 3. PacBio reads show a breakpoint consistent with dup exons 11-15 in *PMS2* in Sample 18, which was undetermined by orthogonal methods due to the LR being terminal. No breakpoints detected in *PMS2CL.*

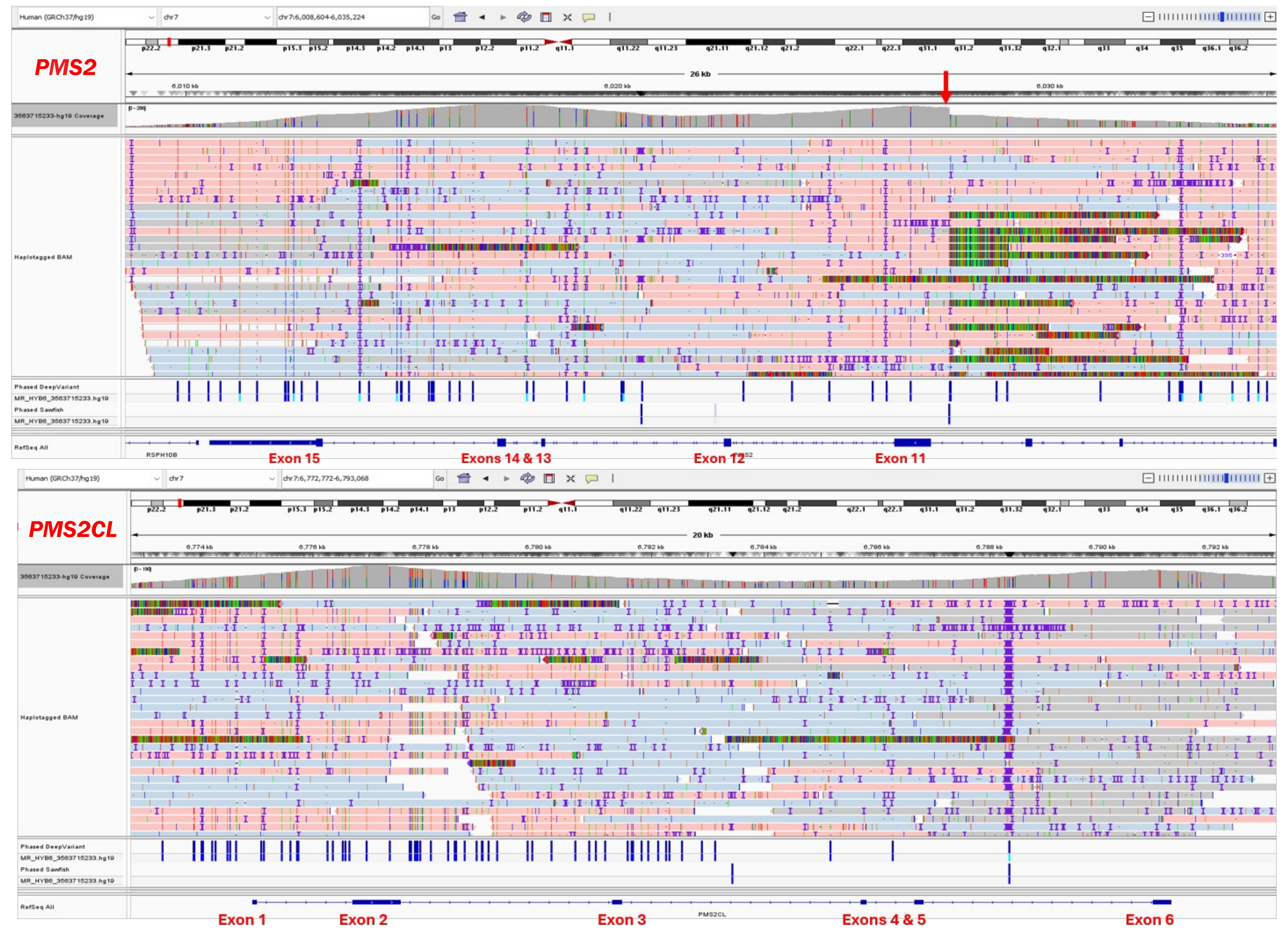

PMS2
Exon 15
Exons 14 & 13
Exon 12
Exon 11
PMS2CL
Exon 1
Exon 2
Exon 3
Exons 4 & 5
Exon 6

Table 1. Summary of PacBio results in confirming the locations of LRs previously identified in *PMS2* or *PMS2CL*

Sample ID	NGS	Gene location based on orthogonal data	PacBio confirmed gene location?
1	Del exon 12	*PMS2*	Yes
2	Del exon 12	*PMS2*	Yes
3	Del exon 12	*PMS2*	Yes
4	Del exon 12	*PMS2CL*	Yes
5	Del exon 14	*PMS2*	Yes
6	Del exon 14	*PMS2*	Yes
7	Del exon 14	*PMS2CL*	Yes
8	Del exon 14	*PMS2CL*	Yes
9	Del exon 15	*PMS2CL*	Yes
10	Del exon 15	*PMS2CL*	Yes
11	Del exons 12-14	*PMS2CL*	Yes
12	Del exons 12-15	*PMS2CL*	Yes
13	Del exons 12-15	*PMS2CL*	No. Full *PMS2CL* deletion, no breakpoints on PacBio
14	Del exons 13-15	*PMS2*	Yes
15	Del exons 14-15	*PMS2*	Yes
16	Dup exon 12	*PMS2*	Yes
17	Dup exons 11-12	*PMS2*	Yes
18	Dup exons 11-15	Terminal duplication; Location undetermined	Yes. Determined to be in *PMS2* by PacBio

Conclusions

- Our study indicates that at ~50x read depth, long-read sequencing analysis is capable of correctly identifying, in most cases, whether an LR is present in *PMS2* or *PMS2CL.*
- PacBio was able to determine the location of a terminal duplication, which was unable to be correctly located with orthogonal methods
- The correct localization of the LR by PacBio mostly relies on the breakpoints revealed in the intronic regions and the long reads that allow accurate alignment.
- In cases without breakpoint sequences, additional investigations are warranted to ascribe the correct location of the LR to *PMS2* or *PMS2CL.*

Disclosures: All authors were employees of Myriad Genetics, Inc. at the time of this study and received salaries and stock as compensation.

Shujuan Pan (span@myriad.com)