Identification of an Alu insertion in *MSH2* by Next-Generation Sequencing in a Family with Lynch Syndrome: An 8-year Diagnostic Odyssey

Amanda Jacquart[1], Stefano Rosati[2], Donald Basel[2], Jessica Grzybowski[3] Michael Muriello[2]
The Medical College of Wisconsin[1], Genomic Sciences & Precision Medicine Center (GSPMC)[2], Ambry Genetics[3]

Genomic Sciences & Precision Medicine Center

MEDICAL COLLEGE OF WISCONSIN

Ambry Genetics A Konica Minolta Company

Background

- Lynch syndrome (LS) is a well-known cause of hereditary colon cancer.

- Pathogenic variants and likely pathogenic variants in one of the mismatch repair (MMR) genes (*MLH1*, *MSH2*, *MSH6* and *PMS2* along with deletions of *EPCAM*) are known to cause LS.

- Alu insertions are the most abundant retrotransposon in the human genome and insertions of Alu elements have been shown to cause disease by either disrupting a coding region or a splice signal.

  - Retroelement insertions have been observed in cancer predisposition genes and were recently reported to be more common (1/325; 0.3%) than previously estimated (1/600; 0.16%).[1]

- There have been previous reports of Alu insertions in MMR genes in families with LS.[2,3]

Methods

- The mobile element (ME) detection software Mobster[4] and the commercial laboratory's in-house developed software was used to detect unaligned and soft-clipped reads from the BAM file

  - The variant was confirmed by Sanger sequencing

- The mother's WGS BAM file data were again reviewed and reads covering this insertion were not identifiable (Figure 2)

  - Mobster was implemented to run on the maternal WGS

  - Split reads were detected on WGS in the same variant location

  - Standard WGS BWA alignment[5] did not map the reads which contained more than 50% Alu reads and trimmed the reads with less than 50% Alu reads

  - This splice site was not detected by standard variant calling

    - Reads are assessed for small variants and structural variation (Figure 3)[6]

- The MSH2/Alu insertion was undetected by conventional NGS variant calling methods

Case Report

- 16-year-old female diagnosed with stage 4 colon cancer (Figure 1)

  - Colon tumor specimen showed abnormal microsatellite instability (MSI) and loss of protein expression of MSH-2 and MSH-6 by immunohistochemistry (IHC)

- Clinical history was unremarkable leading up to her diagnosis and she passed away 10 months later

- Patient's mother had a history of multiple colon polyps starting in her mid-20s

  - Results from a LS screen performed on a colon tubular adenoma with focal high-grade dysplasia revealed abnormal MSI and the same absent protein expression

  - She also had a history of a sebaceous adenoma and a squamous cell carcinoma of the scalp

- Maternal family history:

  - Fulfilled Amsterdam Criteria II

- Paternal family history:

  - Significant for multiple generations of breast cancer

Key
- ⬤ = Colon Ca
- ⬧ = Colon Polyps
- ◖ = Sebaceous Neoplasms
- ⬤ = Breast Ca
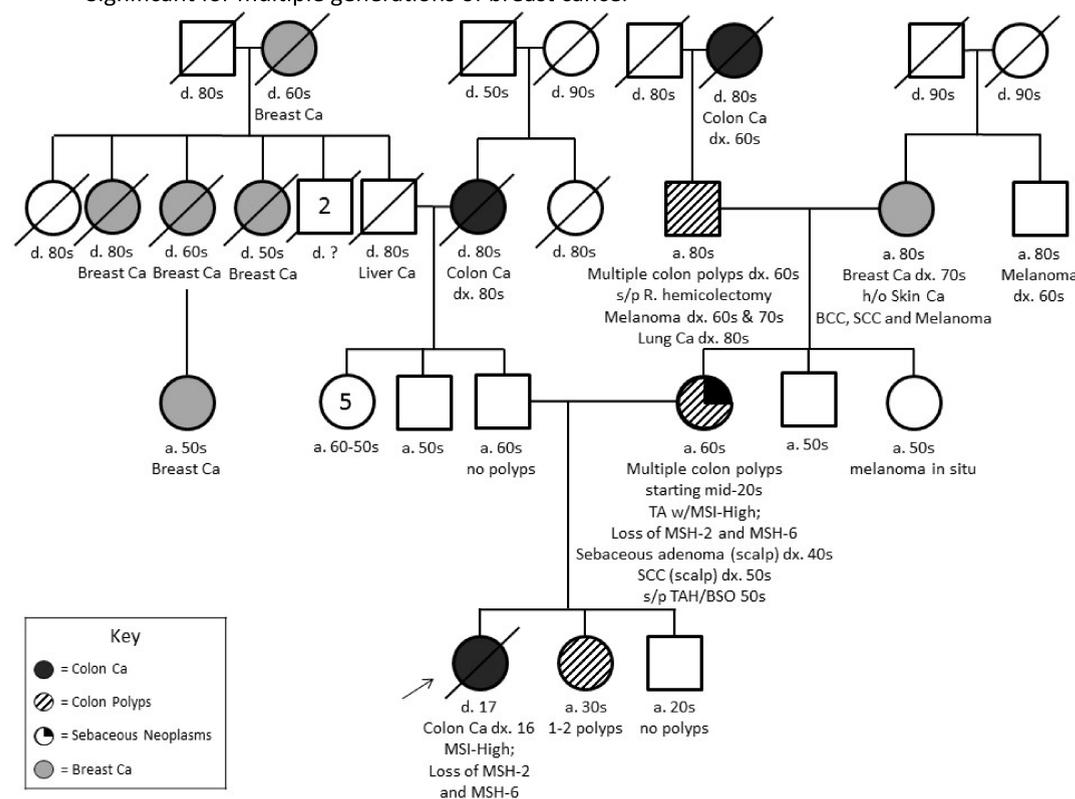
FIGURE 1. *Genetic pedigree.*

Results

Prior to patient's passing — Multiple genes analyzed, no causative variants identified

- Clinical sequencing and deletion/duplication analysis of *APC*, *MLH1*, *MSH2*, *MSH6*, and *PMS2*
- Research testing of *EPCAM* (*TACSTD1*) though Dr. Ligtenberg's laboratory in Nijmegen
- DNA was isolated and banked

One year after — Patient's mother underwent whole genome sequencing (WGS)

- No causative variants identified to explain family history of cancer
- A pathogenic variant was found in the *FBN1* gene, leading to a diagnosis of Marfan syndrome in the mother

Eight years later — Testing performed on banked DNA utilizing a commercial laboratory's custom cancer panel on a Next-Generation sequencing (NGS) platform

- A total of 81 cancer susceptibility genes analyzed
- A likely pathogenic variant was observed at c.1442_1443insAlu in the *MSH2* gene
- Patient's parents underwent confirmatory genetic testing via the same laboratory's NGS panel
  - Mother: positive for the likely pathogenic variant in the *MSH2* gene at c.1442_1443insAlu
  - Father: negative/normal genetic findings
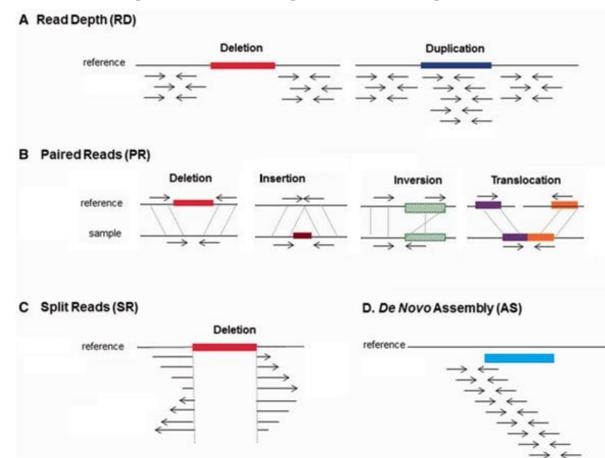
FIGURE 3. *Methods of structural variation detection.*

Conclusion

- There is a subset of patients with a phenotype strongly suggestive of LS and no identifiable germline pathogenic variant.

- This case demonstrates the importance of critically assessing the testing methodologies previously performed in this patient cohort.

  - Specifically taking into account if previous testing was capable of identifying retroelement insertions.

- This case demonstrates the value of reanalyzing short-read sequencing data for structural variants and retroelement events for cases that have not been previously diagnosed.

FIGURE 2. *IGV screenshot of the BAM files from GATK BWA (bottom) and Mobster (top) from the same raw fastq input.*

While GATK shows a more specific alignment to reads in MSH2 exon 9, a clear region of read trimming can be seen at the ALU insertion point. Mobster's alignment (top) shows only the split reads that have mapping to other portions of the genome.

While BWA's trimming is indicative of the insertion, this type of trimming common in alignments- especially in regions of low complexity. This makes it difficult to detect and review regions with this type of insertion. This image shows the value of an event specific aligner such as Mobster.

References

1. Qian Y et al. *Cancer Genet.*, 2017, *216-217*, 159–169.
2. Kloor et al. *Hum. Genet.*, 2004, *115*(5), 432–438.
3. Solassol et al. *Hum. Mutat.*, 2019, *40*(6), 716–720.
4. Thung DT et al. *Genome Biol.*, 2014, *15*(10).
5. Li H et al. *Bioinformatics*, 2009, *25*(14), 1754–1760.
6. *Structural Variation Detection Analysis* (n.d.). Retrieved from https://www.1010genome.com/sv-detection/