Long Read Sequencing Elucidates Complex Germline Variants in Individuals Undergoing Hereditary Gastrointestinal Cancer Testing

Sami Belhadj[1], Felicia Hernandez[1], Ashley PL Marsh[1], Colin Young[1], Jessica Grzybowski[1], Jesus Ramirez Castaño[1], Cassidy Carraway[1], Carrie Horton[1], Marcy Richardson[1], Rachid Karam[1]

[1] Ambry Genetics. 1 Enterprise, Aliso Viejo, CA 92656, USA.

Background and Aim: A considerable fraction of patients who undergo clinical genetic testing present with a personal and family history suggestive of hereditary gastrointestinal (GI) cancer syndromes, and often receive negative germline results. Conventional testing, based on short read (SR) sequencing, may fail to detect structural variants (SV) and mobile element insertions (MEI), which could be resolved with long read sequencing (LRS). We sought to evaluate the clinical utility of LRS by exploring cases suspected to have Lynch syndrome (LS) missing heritability, by elucidating DNA variation in cases with abnormal RNA findings and by resolving a conflicting *PMS2* duplication.

Methods: Multiple LRS approaches were leveraged to study cases with negative or dubious germline findings. These included genome sequencing and a multigene hybrid capture panel covering all exons and introns of bona fide hereditary GI cancer genes. Further details about selected cases are included in Table 1. Patient samples underwent SR capture DNA and RNA sequencing, as well as MLPA as part of their clinical diagnostic testing.

Results: We identified the genetic cause of LS in 1/17 (5.9%) patients in the LS missing heritability cohort. The proband was a carrier of a novel 90kb *PMS2* exon 1 Inversion and was diagnosed with CRC at age 60, with PMS2 loss in his tumor. The proband has four siblings with LS-related cancers (ages at diagnosis: 51-67). Review of SR RNA data uncovered complete monoallelic expression, supporting the pathogenic nature of the inversion. We also evaluated the use of LRS to resolve cases with abnormal RNA findings. In two cases with >50 colonic polyps where RNA data supported skipping of *APC* exon 14, we uncovered an Alu-mediated complex recombination event in intron 14, in proximity to the splice donor site. Genome sequencing in a proband from a family fulfilling Amsterdam II criteria identified a 3kb SVA (SINE-VNTR-Alu) insertion that caused skipping of *MSH2* exon 2. In a cancer-free case with a suspected *PMS2* exon 11 duplication, we relied on PacBio's Paraphase tool to fully phase *PMS2*/*PMS2CL* haplotypes, which led to the identification of *PMS2CL* variants that drove the false positive result reported by MLPA.

Conclusions: By using DNA LRS, we identified and fully characterized SVs and MEIs, previously missed by conventional approaches. This allowed us to uncover the genetic basis of LS in one patient in the missing heritability cohort, providing evidence to further expand these efforts to larger discovery series. Leveraging available RNA data was crucial to elucidate the effect of the identified DNA variants. Similarly, cases with abnormal RNA findings could be resolved with LRS. Finally, LRS constitutes a new approach to address the uncertainty related to *PMS2* variants located within the region of high sequence similarity to *PMS2CL*.

Table 1. Selected cases for DNA long read sequencing.

Selected cases	Personal cancer history	Ages of diagnosis	Family history[1]	DNA LRS	Findings
LS Missing heritability (n=17)	CRC/Endometrial (n=15) Biliary tract (n=1) Urothelial (n=1)	39-77 (Median 50)	≥ 1 FDR (n=13) ≥ 2 SDRs (n=4)	Hybrid capture	*PMS2* Ex1 Inversion (1/17 cases)
APC Ex14 skipping (n=2)	Proband 1: >50 polyps[2] Proband 2: >50 polyps[3] + Breast cancer	57 71	None 2 FDRs, 1 SDR	Hybrid capture, Amplicon sequencing	*APC* Intron 14 complex rearrangement. Alu-mediated.
MSH2 Ex2 skipping (n=1)	Endometrial	55	2 FDRs	Genome sequencing	*MSH2* c.212-15_212-14insSVA
PMS2 Ex11 duplication[4] (n=1)	No cancer	NA	1 FDR, 1 SDR, 1 TDR	Hybrid capture	No *PMS2* Ex11 duplication identified

Abbreviations: LS, Lynch Syndrome; CRC, colorectal cancer; FDR, First-degree relative; SDR, Second-degree relative; TDR, Third-degree relative; LRS, long-read sequencing; Ex, exon.

1. LS related cancer types identified: colorectal, endometrial, ovarian, gastric, small bowel, urinary tract, biliary tract, prostate, pancreatic.

2. Adenomatous/hyperplastic.

3. Unspecified type.

4. Detected by multiplex ligation-dependent probe amplification (MLPA).