Characterization of complex hereditary cancer associated germline variants with long read sequencing

Sami Belhadj, PhD[1]; Felicia Hernandez, PhD[1]; Terra Brannan, PhD[1]; Colin Young, PhD[1]; Jesus Ramirez Castaño, BS[1]; Cassidy Carraway, BS[1]; Jessica Gryzbowski, MS-CGC[1]; Carrie Horton, MS-CGC[1]; Marcy Richardson, PhD[1]; Rachid Karam, MD, PhD[1]

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

Introduction: Long read (LR) sequencing is emerging as a key technology to elucidate complex variants that are missed/not fully characterized by conventional short read sequencing approaches. Among these, structural variants and mobile elements are particularly challenging. We sought to apply LR sequencing to resolve several complex germline variants and to determine its clinical utility.

Methods: DNA samples submitted for multigene panel clinical diagnostic testing underwent multiple LR-based strategies including genome sequencing, targeted DNA amplicon sequencing and hybrid capture on the PacBio Sequel IIe. Short read DNA and RNA sequencing data was available for all studied cases as part of clinical testing.

Results: By using LR DNA sequencing, we fully characterized a 3Kb SVA (SINE-VNTR-Alu) insertion in *MSH2* (c.212-15_212-14insSVA) that caused skipping of exon 2 in a family fulfilling Amsterdam II criteria. In another case where RNA data showed skipping of *APC* exon 14, we identified a previously reported Alu-mediated complex recombination event in intron 14 near the splice donor site. Using targeted LR hybrid capture, we fully elucidated tandem duplication breakpoints in cases carrying *PALB2* EX13dup or *APC* 5'UTR-Ex8dup. Additionally, we partially resolved the breakpoints of a case with *ATM* EX34-3'UTRdup, which received a reclassification from variant of uncertain significance to likely-pathogenic after combining and assessing evidence from LR, RNA and MLPA data. Finally, preliminary hybrid capture data showed LR utility in identifying variants affecting *PMS2* and separating them from those affecting its pseudogene, *PMS2CL*.

Conclusions: We successfully employed LR DNA sequencing to characterize germline variants that were challenging to resolve with conventional approaches. Targeted LR applications are cost effective, allow testing of multiple cases simultaneously and provide increased read depth compared to LR genome sequencing. The combination of LR DNA data and previously available RNA data was key for the correct interpretation of splice altering variants. Further studies are warranted to uncover previously uncharacterized germline DNA variation in cancer predisposing genes that could represent the underlying genetic cause for missing heritability cases.