The Value of Long Read Amplicon Sequencing for Clinical Applications

K. Neveling[1], R. Derks[1], A. Den Ouden[1], S. van der Heuvel[1], C. Heiner[3], I. McLaughlin[3], J. Harding[3], L. Aro[3], D. Lugtenberg[1], A. Mensenkamp[1], M. Kwint[1], M. Tjon-Pon-Fong[1], M. van der Vorst[1], M. Ligtenberg[1,2], H. Yntema[1], M. Nelen[1], L. Vissers[1], L. Haer-Wigman[1], R. de Voer[1]

[1]Department of Human Genetics, Radboud university medical center, Nijmegen, the Netherlands

[2]Department of Pathology, Radboud university medical center, Nijmegen, the Netherlands

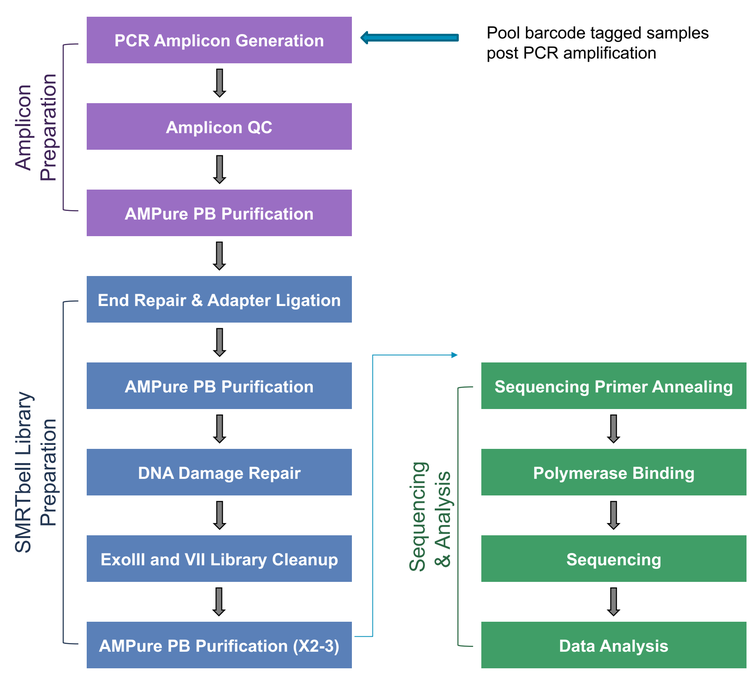
[3]Pacific Biosciences, 1305 O'Brien Drive, Menlo Park, CA 94025

Introduction

NGS is commonly used for amplicon sequencing in clinical applications to study genetic disorders and detect disease-causing mutations. This approach can be plagued by limited ability to phase sequence variants and makes interpretation of sequence data difficult when pseudogenes are present. Long-read highly accurate amplicon sequencing can provide very accurate, efficient, high throughput (through multiplexing) sequences from single molecules, with read lengths largely limited by PCR. Data is easy to interpret; phased variants and breakpoints are present within high fidelity individual reads.

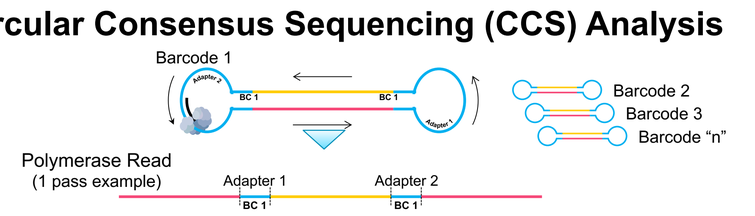
Here we show SMRT Sequencing of the ***PMS2*** and ***OPN1*** (***MW and LW***) genes using the Sequel System. Homologous regions make NGS and MLPA results very difficult to interpret.

Long-Read SMRT Sequencing Workflow



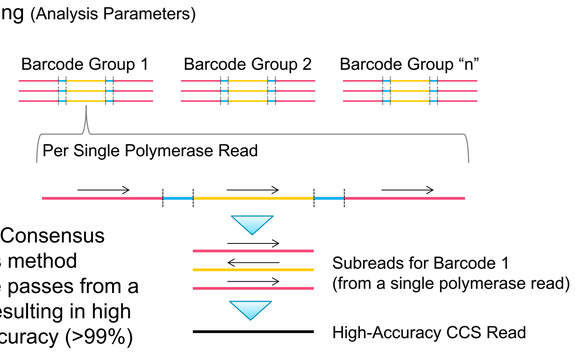
Sequencing Analysis Workflow

Circular Consensus Sequencing (CCS) Analysis



In SMRT Analysis:

1. Pre-Process Filtering (Analysis Parameters)
2. Demultiplex



3. Generate Circular Consensus
   - The CCS analysis method combines multiple passes from a single molecule resulting in high individual read accuracy (>99%)
   - CCS generate HiFi reads ready for further analysis (alignment, variant calling, etc. with standard informatic tools

Methods and Results *PMS2*

For *PMS2*, three amplicons ranging in size from 11.4 kb to 16.8 kb were designed using unique primers, covering 36 kb of sequence. SMRT Sequencing produced HiFi reads with coverage ranging from 200-fold to 1500-fold; data clearly indicated 2 deletions >1000 kb with precise breakpoint mapping.

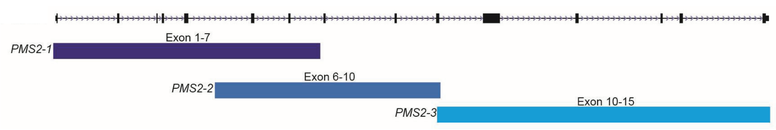


Figure 1: Design of the *PMS2* LR-PCR fragments.

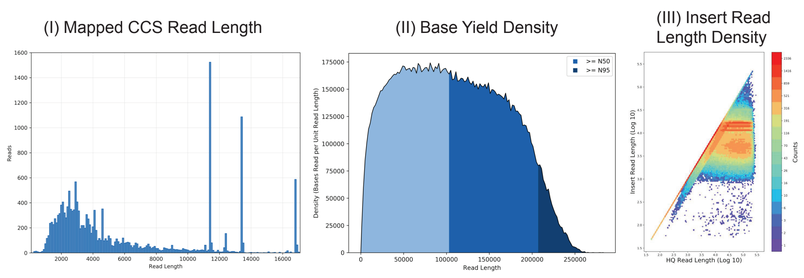


Figure 2: Run metrics of a 16 kb amplicon run: (I) the majority of mapped CCS reads (HiFi reads) represent the 11.4, 13.6 and 16.8 kb *PMS2* fragments; (II) the N50 polymerase read length is >100 kb; and (III) the insert read length density plot shows the three LR_PCR fragments.

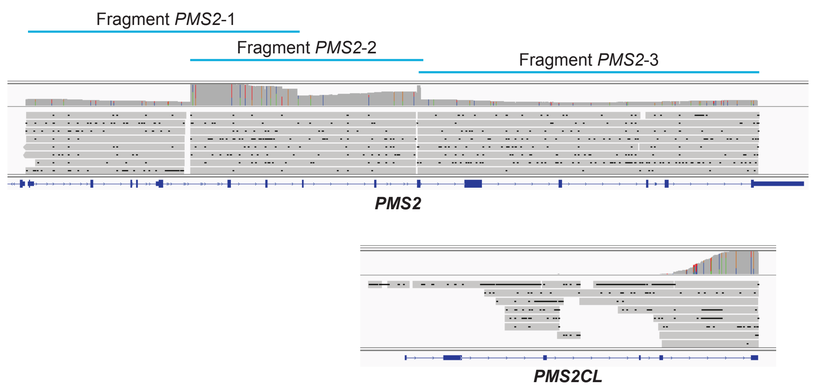


Figure 3: Complete coverage of *PMS2* by long-read sequencing (upper panel). Coverage of *PMS2* is > 6000x, whereas the coverage *of PMS2-CL* is > 30x. Due to this large difference we do not worry about sequencing the pseudogene as well (lower panel).

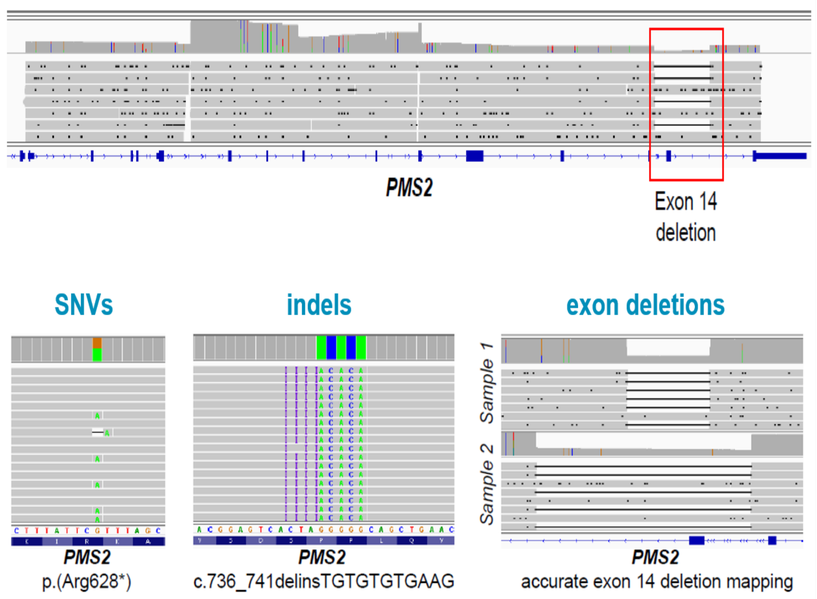


Figure 4: Long-read sequencing of *PMS2* can detect exon deletions >1 kb in size (upper panel), SNVs (lower panel; left), small indels (lower panel; middle) and accurate breakpoint mapping of most exon deletions (lower panel; right).

Methods and Results *OPN1 (MW and LW)*

Full-length amplicons for *OPN1LW* and *OPN1MW*, 14 kb and 16 kb, respectively, were generated from samples with different known gene conversions / hybrid genes and subjected to SMRT Sequencing. For all cases, **PacBio sequencing was 100% concordant,** finding all gene conversions and hybrid genes originally identified by orthogonal technologies. Plus, in some cases SMRT sequencing generated additional relevant data.

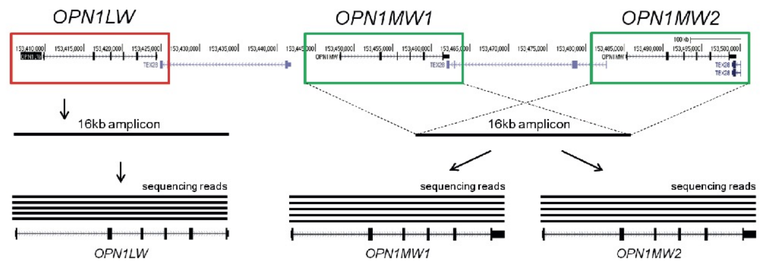


Figure 5: Representation of 16 kb LR-PCRs for *OPN1 LW* and *MW.*

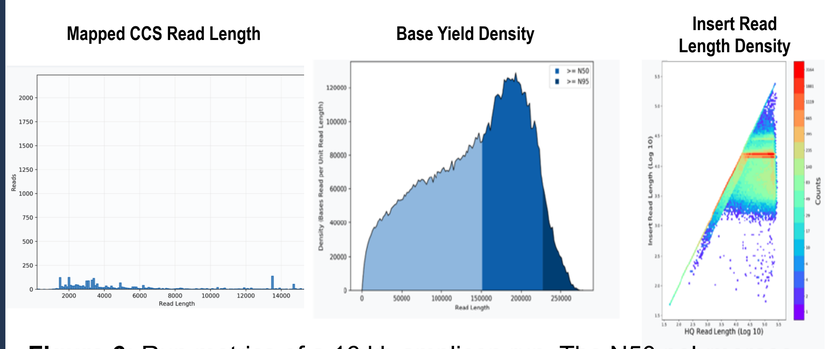


Figure 6: Run metrics of a 16 kb amplicon run. The N50 polymerase read length is >150 kb. The insert read length density plot shows the ~16 kb amplicons. The run output was 18.5 Gb.

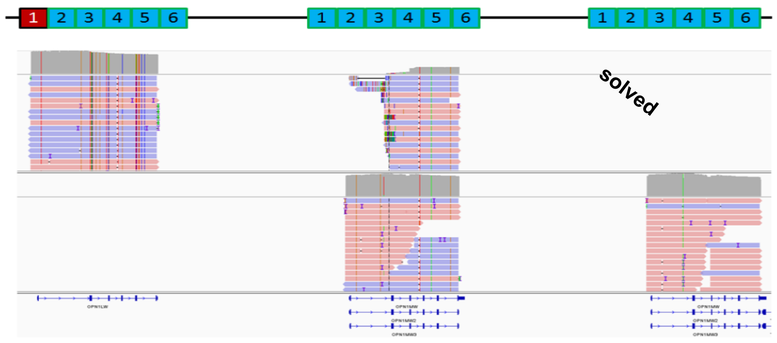


Figure 7: Protanopia patient, with two PCR products (1xLW and 1x MW) . Following sequencing, three MW copies were detected, one has an exon 1 that belongs to LW. All three copies map to different locations in the genome. The data confirm the patient's phenotype.

Conclusion

Targeted long-read sequencing with PacBio is highly accurate (>99.99%) and detects all types of variants, sequencing through various contexts. These results demonstrate the added value of long-read amplicon sequencing:

Efficiency

- Less PCR, no nesting
- Fewer added tests (i.e. MLPA)
- Multiplexing for high throughput

Improved results, easier data interpretation and analysis

- Distinguish between genes and pseudogenes
- Variant phasing within long reads
- Precise breakpoint detection