AZENTA
LIFE SCIENCES

Third-generation amplicon sequencing analysis report

Methods

Library construction and sequencing

For each sample, >2ug purified PCR fragments (1kb~10kb) was used for library preparation. Fragmented DNAs were treated with End Prep Enzyme Mix for end repairing, 5'Phosphorylation and 3'dA-tailing in one reaction, followed by a T-A ligation to add universal hairpin adapters to both ends. Then the SMRTbell libraries were validated by Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), and quantified by Qubit 3.0 Fluorometer. The final SMRTbell libraries were loaded on PacBio Sequel instrument according to manufacturer's instructions (Pacific Biosciences of California, Inc., California, USA).

Data analysis

Raw data statistics:

Perform statistical analysis on subreads to obtain information such as the number, average length and GC content of subreads sequence.

CCS analysis:

Perform Circular Consensus Calling on subreads to obtain high-quality CCS reads by using pbccs software (version 3.3.0)[1], and perform statistical analysis on CCS sequences.

Data split:

Split CCS reads according to barcode by using lima software (version 1.9.0)[2], and statistics the split results.

Statistics of target sequences abundance:

If the target area is specified, the target sequence is captured according to the 10nt of upstream and downstream of the target sequence, otherwise, the target sequence refs to the full-length sequences.

Statistics of base distribution:

The CCS reads are aligned to the reference sequence by using BWA software, and the section on the alignment is extracted and mutation abundance information is counted.

Database and software

Pbccs: https://github.com/PacificBiosciences/unanimity

Lima: https://github.com/PacificBiosciences/barcoding