SubOrder_ Nanopore Sequencing Result Report

Experimental Methods

The target plasmid DNA has been extracted and a sequencing library was generated. Library preparation and library quality control for the Nanopore platform: Fixed quantities of plasmid DNA were extracted, followed by fragmentation using transposase enzymes with integrated barcodes. Subsequently, the samples were combined and subjected to purification. The purified products were then ligated with a standardized adapter, facilitating the construction of a library compatible with the ONT sequencing platform. Once the library was appropriately diluted, it was loaded onto a sequencing chip and subjected to sequencing using the ONT platform.

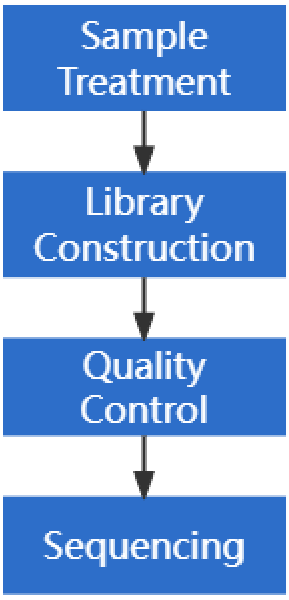

Figure 1. Experimental Procedure

Bioinformatics Analysis

（1） Data quality control: Quality control of raw data based on quality values, the quality-controlled data was used for subsequent comparison and validation analysis.

（2） Reads Mapping: Mapping of the quality-controlled data to the reference sequences was done with minimap2 (version 2.15-r905) [1] software.

（3） Validation analysis: After quality control by Samtools (Version: 1.9) [2], the mapping rate, Coverage Analysis, and Variant Calling information were calculated based on the mapping results and visualized for display thereafter.

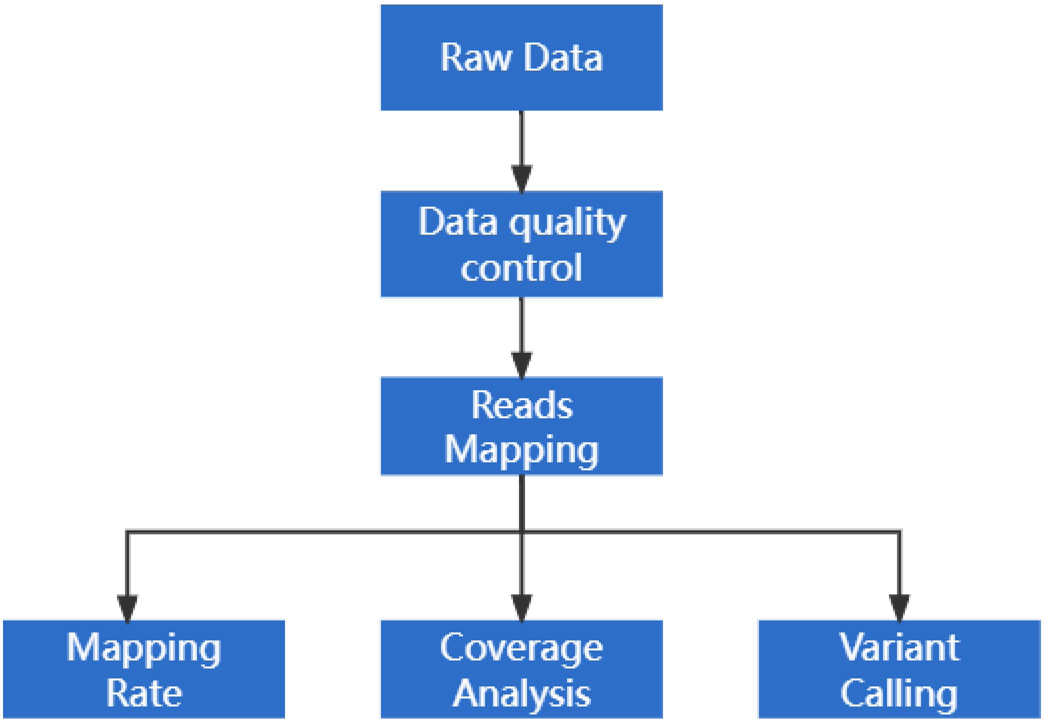

Figure 2. Bioinformatics Analysis Process

Results

Based on the comparison results, the coverage and variation information of the sequencing data was determined for the reference sequence, these statistical results are reported in the file " SubOrder_snp_indel.xls".

Abbreviations used in the file:

（1） Id: Reference sequence name;
（2） Pos: The reference position, with the 1st base having position 1. Positions are sorted numerically, in increasing order, within reference;
（3） Ref: Reference base;
（4） Depth: Read depth at this position for this sample;
（5） A: Number of reads that support adenine at this position;
（6） C: Number of reads that support cytosine at this position;
（7） G: Number of reads that support guanine at this position;
（8） T: Number of reads that support thymine at this position;
（9） MutRate(%): Percentage of reads that support non-reference base at this position;
（10） Mut: Mutation genotype;
（11） InsCov: Counts of reads that support insertion at this position;
（12） InsRate(%): Percentage of reads that support insertion at this position;
（13） Ins: Insertion genotype;
（14） DelCov: Counts of reads that support deletion at this position;
（15） DelRate(%): Percentage of reads that support deletion at this position;
（16） Del: Deletion genotype;

The coverage and variation information of the reference sequence by sequencing data were visualized based on the statistical results. The results are shown in Figure 3, where the abscissa is the position information of the reference sequence (unit bp), and the ordinate is the corresponding statistical value. Please note, nanopore may have high errors in the methylation modification site or homopolymer region (poly region), and the methylation modification site region is controlled by determining the depth of DNA double stranded sequencing for quality control; The poly region, on the other hand, uses the genotype corresponding to the maximum allele frequency as the candidate result. When the candidate result is inconsistent with the reference sequence, it is recommended to use Sanger or PacBio sequencing for validation and integrate the results. The positive mutation information detected is indicated on the upper left of the picture, it includes the location, type, and proportion of the variation. As a an example, not linked to the analyzed sample, 10362 G->A 99.77%, indicates that at base pair position 10,362 in the reference sequence there is a guanine nucleobase, while 99.77% of the sequencing data support an adenine nucleobase at this position in the sample. The insertion-deletion type variation only indicates the base type(s) supported by the sequencing data at the corresponding position in the sample. The result (Figure 3) below demonstrates that the synthesized sequence is accurate.

Figure 3. Overlay of Reference Sequence by Sequencing Results

Regarding the sample sequence, the consensus sequence of the genotype with maximal allelic frequency at each locus can be found in SEQ format in the file " SubOrder_cons_seq.seq ".

References

[1] Heng. Minimap2: pairwise alignment for nucleotide sequences[J]. Bioinformatics, 2018, 34(18):3094-3100.

[2] Danecek P , Bonfield J K , Liddle J , et al. Twelve years of SAMtools and BCFtools[J]. GigaScience, 2021, 10(2):1-4.