Related Work

1-Base calling : is the process of assigning nucleobases to chromatogram peaks or electrical current changes resulting from nucleotides passing through a nanopore.

Tracy also supports re-estimating basecalling qualities. The output of Tracy can be in JSON, FASTA, FASTQ or TSV format. The JSON and TSV formats output the trace at every sampling position. These formats also list the basecalling positions,

2-Sequence alignment or sequence comparison lies at heart of the bioinformatics, which describes the way of arrangement of DNA/RNA or protein sequences, in order to identify the regions of similarity among them.

In illumina : During the alignment step, the banded Smith-Waterman algorithm aligns clusters from each sample against amplicon sequences specified in the manifest file.

The banded Smith-Waterman algorithm performs local sequence alignments to determine similar regions between 2 sequences. Instead of comparing the total sequence, the Smith-Waterman algorithm compares segments of all possible lengths. Local alignments are useful for dissimilar sequences that are suspected to contain regions of similarity within the larger sequence. This process allows alignment across small amplicon targets, often less than 10 bp

In Nanopore sequencing: has emerged as a major sequencing technology and many long-read aligners have been designed for aligning nanopore reads. However, the high error rate makes accurate and efficient alignment difficult. Utilizing the noise and error characteristics inherent in the sequencing process properly can play a vital role in constructing a robust aligner.

3-Assembly: Sequence assembly is the initial step towards downstream data analysis of the sequencing data.

Comparative Assembly : reference based assembly or mapping to genome of a closely related species .

De Novo Assembly : assembly in the strict sense . No or little information about the genome , transcriptome or proteins.

-Falcon assembler

The Falcon assembler developed by Jason Chin from PacBio is another pipeline adopting the strategy of HGAP. It shares many features with PBcR, such as raw reads overlapping for base error correction using DALIGNER and overlap filtering. The major difference lies in its contig consensus generation

-Miniasm assembler

Read error correction is the most CPU-intensive stage of the whole assembly process, and assemblies on gigabase-sized genomes are still out of reach for many projects due to high sequencing costs and large computational requirements. The Miniasm assembler developed by Heng Li takes a different approach to deal with noisy long reads by skipping the step of read error correction completely.

-Hybrid assemblers

The NGS platforms such as Illumina's HiSeq and MiSeq have played a dominant role in genomic research and applications. It is foreseeable that short read data will continue to be a very important part of data sources for years to come. Different algorithms have been explored for genome assembly and many pipelines have been developed for various applications.