

Assignment 3: methodology and result

- **Nanopore sequencing poses a number of computational challenges, for which various methods and algorithms have been developed**
 - **WaveNano**
 - show that the indel issue can be significantly reduced via accurate labeling of nucleotide and move labels directly from the raw signal, which can then be efficiently learned by a bi-directional WaveNet model simultaneously through feature sharing
 - bi-directional WaveNet model with residual blocks and skip connections is able to capture the extremely long dependency in the raw signal.
 - Taking the predicted move as the segmentation guidance, we employ the Viterbi decoding to obtain the final base-calling results from the smoothed nucleotide probability matrix.
 - **Nanopore sequencing technology and tools for genome assembly**
 - ONT's basecalling tools, Metrichor, Nanonet and Scrappie, are the best choices for the basecalling step in terms of both accuracy and performance. Among these tools, Scrappie is the newest, fastest and most accurate basecaller. Thus, we recommend using Scrappie for the basecalling step (see analysis in section 'Basecalling tools').
 - For the polishing step, we recommend using Racon, as it is much faster than Nanopolish. Racon also produces highly-accurate assemblies (see analysis in section 'Read mapping and polishing tools').
 - For the read-to-read overlap finding step, Minimap is faster than GraphMap, and it requires low memory. Also, it has similar accuracy to GraphMap. Thus, we recommend Minimap for the read-to-read overlap finding step (see analysis in section 'Read-to-read overlap finding tools').
 - **continuous wavelet dynamic time warping algorithm (DTW)**
 - algorithm starts from low-resolution wavelet transforms of the two sequences, such that the transformed sequences are short and have similar sampling rates.
 - Then the peaks and nadirs of the transformed sequences are extracted to form feature sequences with similar lengths, which can be easily mapped by the original DTW.
 - Our algorithm then recursively projects the warping path from a lower-resolution level to a higher-resolution one by building a context-dependent boundary and enabling a constrained search for the warping path in the latter.

• **Result**

- In this work, we reported a new version of the previously published work on simulating the Nanopore sequencing, DeepSimulator1.5.
- In this updated version, we have updated all the three modules of DeepSimulator significantly with several crucial overall optimizations, resulting in a more powerful, quicker and lighter simulator.
- This major update can remarkably broaden its applications in large-scale sequencing simulations as well as studies focusing on the Nanopore signals.