Assignment 2: Related works summary

- The nanopore sequencing algorithm is still developing and has a growing number of applications. In this section, I am going to discuss some of them.

- In first, we have an interesting application studying DNA methylation, that is involved in many biological processes. The nanopore technology helps to make it directly, as the signal of methylated nucleotides differ than that for unmethylated ones. Mainly it have two different software, Nanopolish and SignalAlign, are using HMMs to identify C5-methylcytosine (5mC). The newest addition to the methylation toolbox is DeepMod, it is using raw electric signals and a bidirectional recurrent neural network to detect 5mC and N6-methyldeoxyadenosine (6 mA).

- Another application depends on metagenomic studies. As sometimes we cannot distinguish between closely related species or microbial strains. Several reports said that using long read might be a solution for that, and there is some software solving this problem, including MetaG and the ONT's own EMPI2ME.

- Another application helps direct sequencing of RNA which difines complexity of alternative transcripts, unbiased quantification of transcriptome and detection of methylated nucleotides. most of them was made to deal with noisy long reads, like SQANTI which is good with PacBio and any long reads as it takes FASTA files as an input. also, there is FLAIR which enables the correction, isoform definition, and alternative splicing analysis of noisy reads. ONT developed a set of tools called pinfish for long transcriptomics data analyses, which was inspired by Mandalorion pipeline. In the end, we have LoReAn for eukaryotic utilizing short- and long-reads cDNA sequencing, protein evidence, and ab initio gene prediction.

- then, we have NanoSim-H that simulate nanopore reads for test performance. Also, we have NanoDJ which integrates many tools together enabling tasks such as base calling, ...etc in a single environment. Tandem-genotypes is an interesting software that finds changes in length of tandem repeats, from "long" DNA reads aligned to a genome.

- The final piece of software that is good to be mentioned is RUBRIC. It implements ONT's molecule-by-molecule real-time selective sequencing (Read Until) for real-time sequencing.