NanoBioPhotonixLab

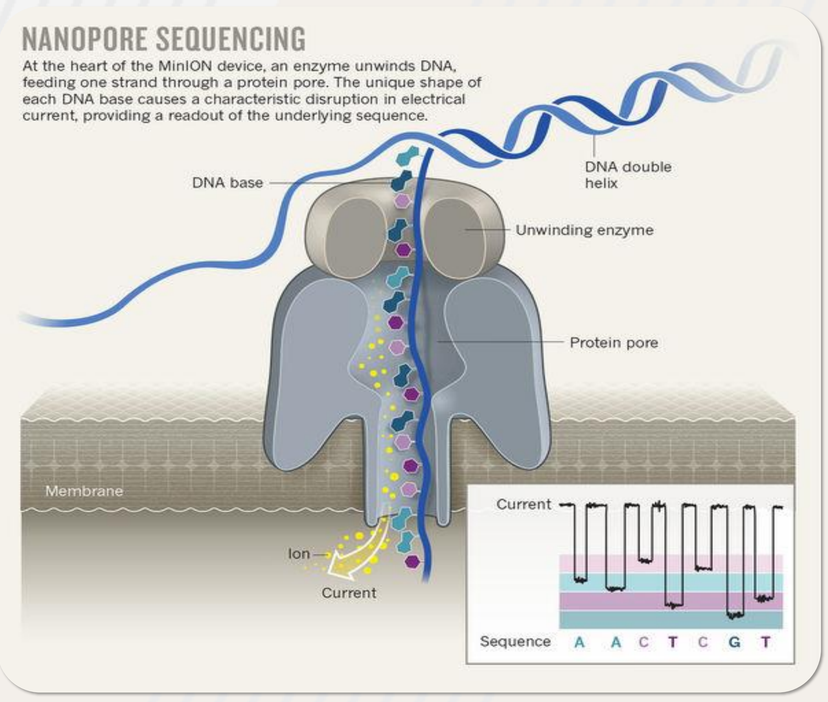
Using Nanopore Sequencing to Detect Base Modifications

Gal Goldner[1], Yael Michaeli[1], Brian Yao[2], Chloe Hsu[2], Jennifer Listgarten[2] and Yuval Ebenstein[1]

[1] School of Chemistry, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel-Aviv University

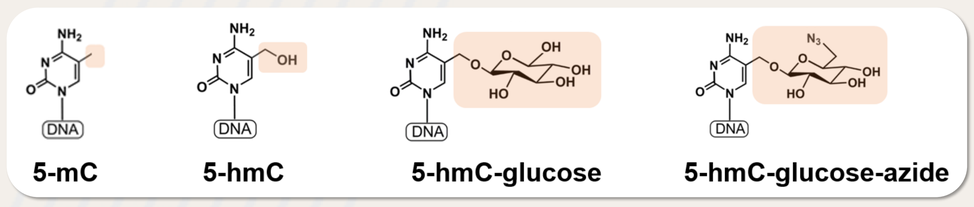
[2] Department of Electrical Engineering and Computer Science, University of California, Berkeley

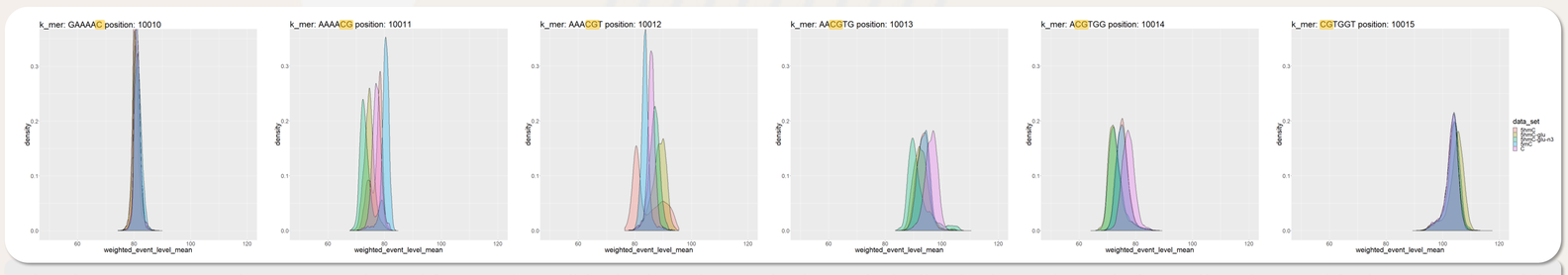
Nanopore sequencing is a third-generation approach used mainly for the sequencing of DNA or RNA. In nanopore sequencing, an ionic current passes through nanopores and the changes in current are measured, as biological molecules (such as DNA) pass through the nanopore. The information about the changes in current can be used to identify the molecules. Today, nanopore sequencing can detect, in addition to the four natural nucleotides, a very small number of modified nucleotides (mainly 5mC and 6mA). Therefore, its ability to detect epigenetics changes of modified nucleotides is very limited. In this study, we are using chemical reactions in order to modify a single nucleotide so its measured current will be very different from that of the same but unmodified nucleotide. We used four modifications of Cytosine (within CpG site), different in their chemical functional groups and sizes. We saw a unique behaviour for each of these modifications, based on an analysis of the raw data and an advanced analysis in which we used several tools that produce more data and information about the behaviour of the modified nucleotide inside the pore. Since at any given time multiple nucleotides pass through the pore, theoretically it is necessary to use the data of all possible combinations of the nucleotides with the modification (their number depends on the pore) in order to detect that modification in an unknown sequence. Because we were able to understand the modification's behaviour, this work can form the basis for reading DNA changes without the need for a specific training.



Investigating the effect of the modifications on the signal from the Nanopore sequencer

In order to examine the four Cytosine modifications and their influence on the Cytosine signal, we took 1-2kb DNA sequences (of lambda) and modified for each of them two of their Cytosines. We did it by using two PCR primers that we modified their Cytosines (only one for each). We saw that there is a difference in the average current between every modified Cytosine compared to the unmodified Cytosine in the same k-mer. This proves that the modifications affect the ionic current that passes through the pore and that this information can be used in order to detect these modifications.

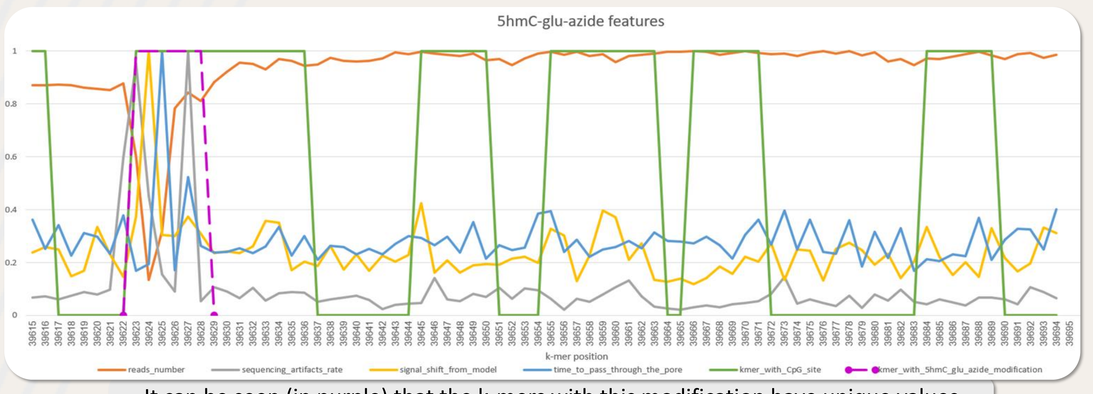




The change in current between each modification (5mC (blue), 5hmC (orange), 5hmC-glu (yellow), 5hmC-glu-azide (green)) and unmodified Cytosine (purple)

Using 5hmC-glucose-azide in order to detect 5hmC modification

After an advanced analysis we noticed that the most significant signal is obtained from the **5hmC-glucose-azide** modification. We characterized this modification's behaviour so it can be detected based on unique features regardless of the specific k-mer that it is part of. 5hmC (5-Hydroxymethylcytosine) has an important role in epigenetics and today it can't be detected by using Nanopore sequencing. Therefore, 5hmC-glucose-azide can be used to detect natural 5hmC (by using chemical reactions that attach glucose and azide to the natural modification).



It can be seen (in purple) that the k-mers with this modification have unique values compared to all the other k-mers (also the ones with CpG sites, in green)

Future work

An experiment where natural 5hmC modifications were modified to get 5hmC-glu-azide modifications (using a sample from a mouse brain - c57/bl6) was done. In the analysis of the experiment's outputs – which is still in progress – we are using a machine learning model to be able to quantify our ability to detect the modification.

The **MinION** sequencer (Oxford Nanopore Technologies)