

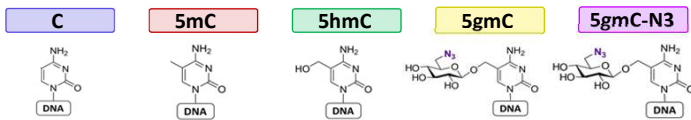
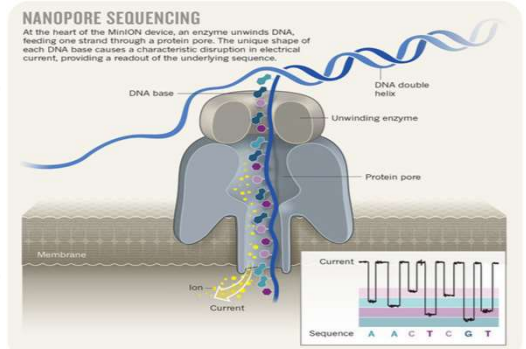


Exploring the Potential of Chemo-Enzymatic Technology for Improved Detection of Epigenetic Modifications with Nanopore Sequencing

Assaf Grunwald¹, Eliran Eitan¹, Gal Goldner¹, Brian Yao², Jennifer Listgarten² and Yuval Ebenstein¹

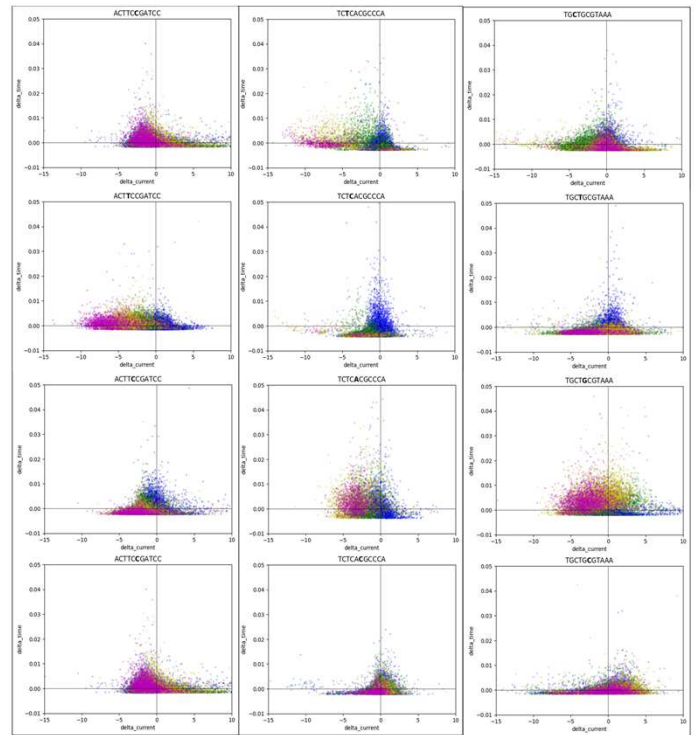
¹ School of Chemistry, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel-Aviv University ² Department of Electrical Engineering and Computer Science, University of California, Berkeley

Detecting DNA epigenetic modifications is crucial for understanding their role in gene expression, development, and disease providing insights into underlying mechanisms and new therapeutic strategies. While in mammal genomes DNA modifications are limited to several chemical groups and occur predominantly in CG context, in other organisms modifications display larger variety complexity and occur on wide and not fully known sequence motifs. Oxford Nanopore's technology offers a promising approach for the detection of epigenetic modifications in DNA. However, expanding the basecalling model of nanopore sequencing to new modifications requires a large, diverse training dataset, and the dynamic and context-dependent nature of modifications makes it difficult to generalize the model. This poses a challenge for detection and interpretation, despite the promising approach of using electric nanopore sensors. To address this challenge, we developed enzymatic technology that adds bulky chemical moieties to modifications, enhancing detection sensitivity during nanopore sequencing and allowing genome-wide detection with a small train-set. Specifically, we have shown that attachment of UDP-Glucose-Azide to 5hmC moieties significantly improves nanopore sensing ability, providing a promising avenue for characterizing other epigenetic modifications.



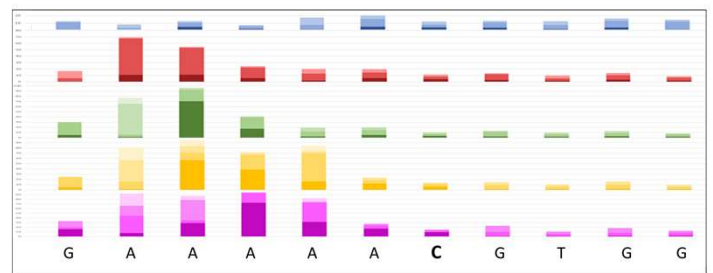
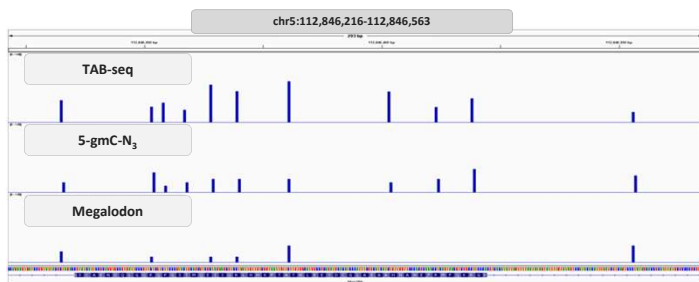
Establishment of a K-mer Train Set

To study the influence of Cytosine modifications and their chemo-enzymatic alterations, a training set was established. This training set consisted of 25 different 1-2 kb DNA sequences with modified cytosines at known locations in different k-mer contexts. The modifications included 5mC, 5hmC, 5gmC, and 5gmC-N3. The aim was to evaluate how these modifications influenced the measured parameters (namely dwell time and current), and how the k-mer positions around them were affected with respect to their distance from the modified cytosine (illustrated on the right). We have also checked how the Guppy basecaller interprets the bases around the modification and in which cases it failed to detect the modified base (characterized as either N, skipped or shifted by Guppy, illustrated below the scatter plots). Based on the shift in all of these parameters from the native cytosine, we have trained a model for the global detection of chemo-enzymatically modified 5hmC.



Scatter plots are presented to describe the shift from the average current and dwell time of the native k-mer time for each individual read spanning that k-mer. The plots depict three representative k-mers (out of the 25 in the training set), with each k-mer represented by a column. For each k-mer, the plots show three bases before the cytosine and the cytosine itself.

Testing the 5hmC Detection model on Mouse Brain DNA



Comparison of Guppy basecaller interpretation of bases around the modification (NNN, shifted, skipped) between all modifications, in a specific k-mer. Bases along the k-mer are indicated below the bars

To test our model, we used DNA extracted from mouse brain due to its known high levels of 5hmC. We have either sequenced the native DNA and called 5hmC with Megalodon or modified the DNA with glu-N3 and called 5hmC with the model we trained. The results were compared to TAB-seq, the current gold standard for 5hmC detection. Results (below) show better correlation to TAB-seq results in 5hmC detection when our model used compared to Megalodon detection.