

Developing new bioinformatic methods to identify modifications from Nanopore sequencing



Søren Heidebach¹, Lucas N.L. van der Maas³, Mantas Sereika¹, André Lamúrias², Sheila I. Jensen³, Katja Hose², Thomas D. Nielsen² and Mads Albertsen¹

¹Albertsen Lab, Center for Microbial Communities, Aalborg University. ²Department of Computer Science, Aalborg University.

³The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark.

Background

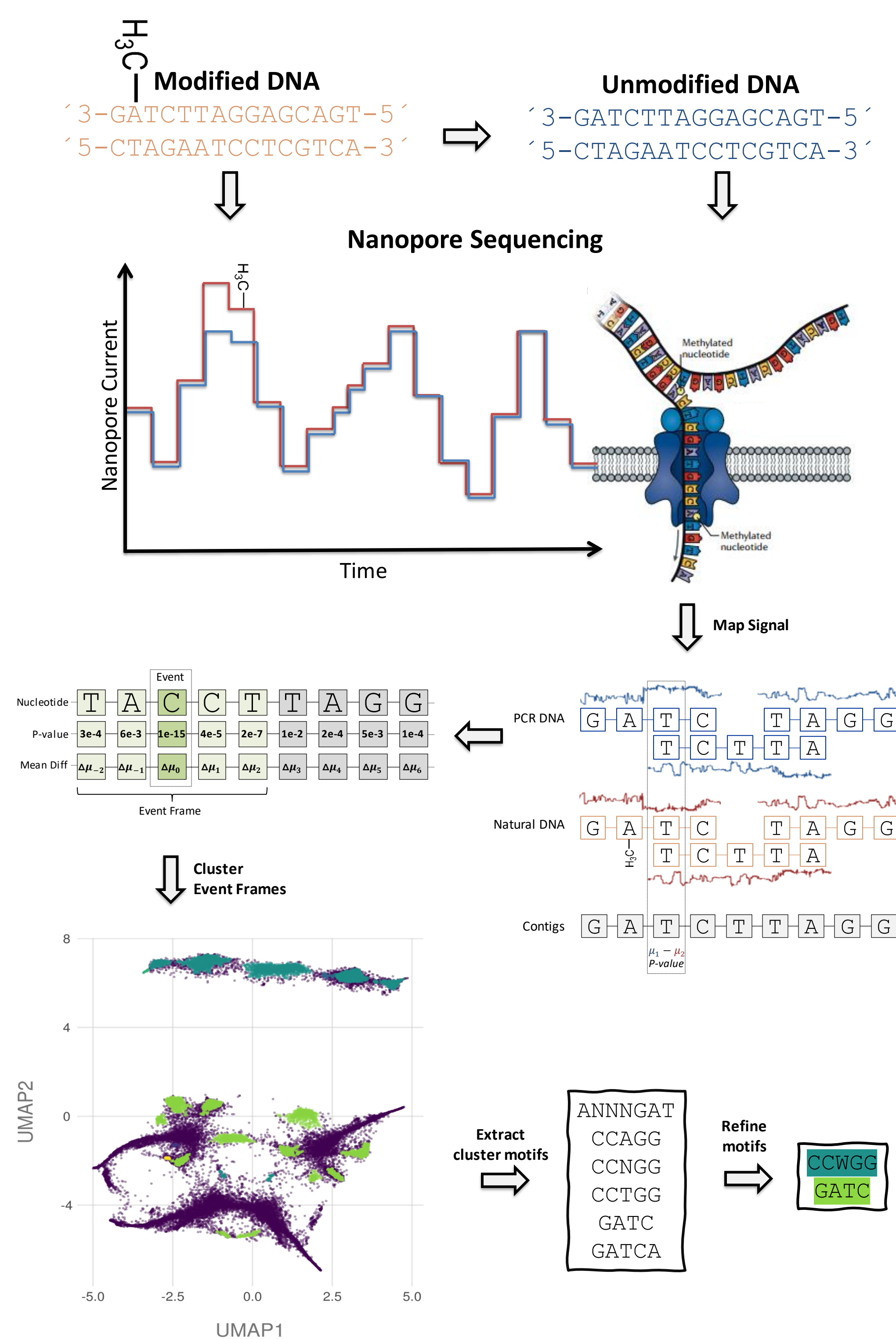
Extraction of microbial genomes in complex samples requires *in silico* methods, as monoculture cultivation is often extremely difficult or even impossible. The methylation of DNA represents a useful feature, as different microbial taxonomies have different methylomes, a difference driven primarily by DNA restriction mechanisms. To ensure all motifs are included a *de novo* detection approach is preferable, such novel modifications and motifs are also included in modification features. Additionally, detection must be possible on shorter metagenomic contigs to be useful in binning.

Conclusion

- *De novo* identification of modified bacterial sequence motifs is possible with Nanopore sequencing
- Detection of motifs is possible in short and low coverage references, often present in metagenomic assemblies
- 6mA modifications are more difficult to detect than 5mC

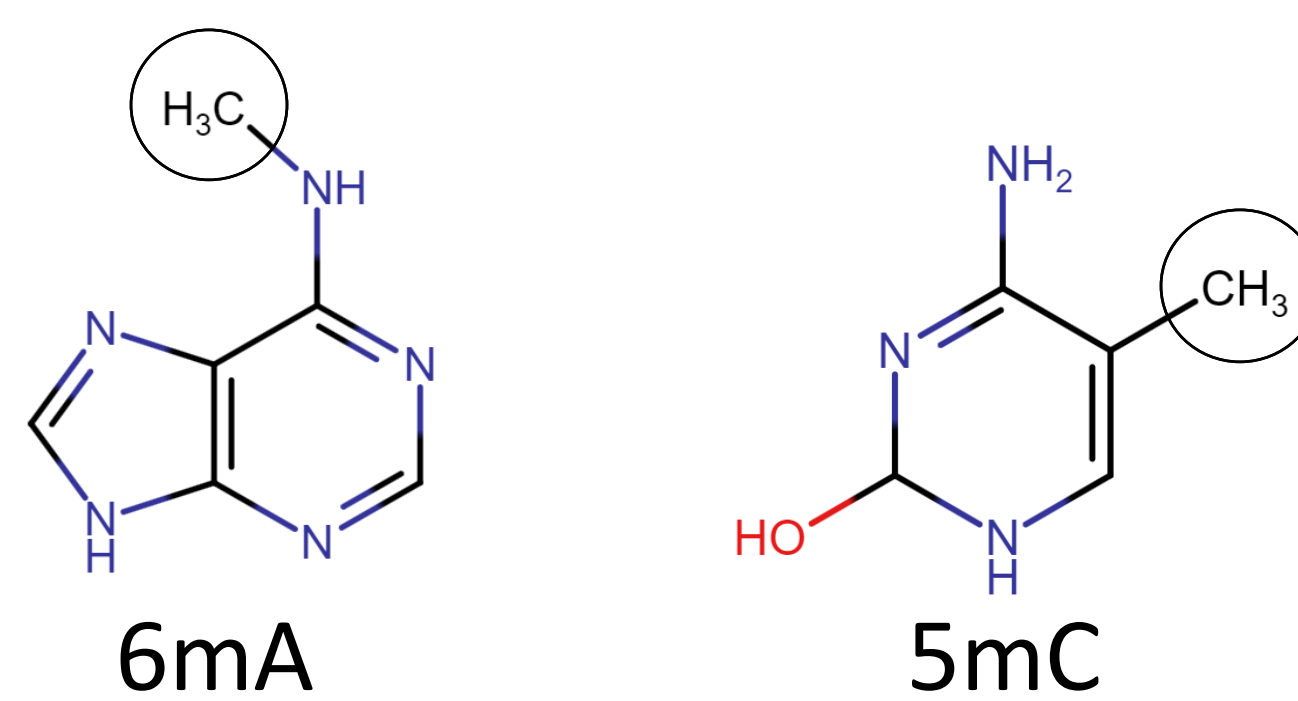
Methods

- ▶ The modification of DNA disturbs the measured current during Nanopore sequencing
- ▶ Differences between modified and unmodified DNA currents allow identification of abnormalities, termed events
- ▶ Embedding and clustering of events allow aggregation of similar modification types and detection of consensus sequences



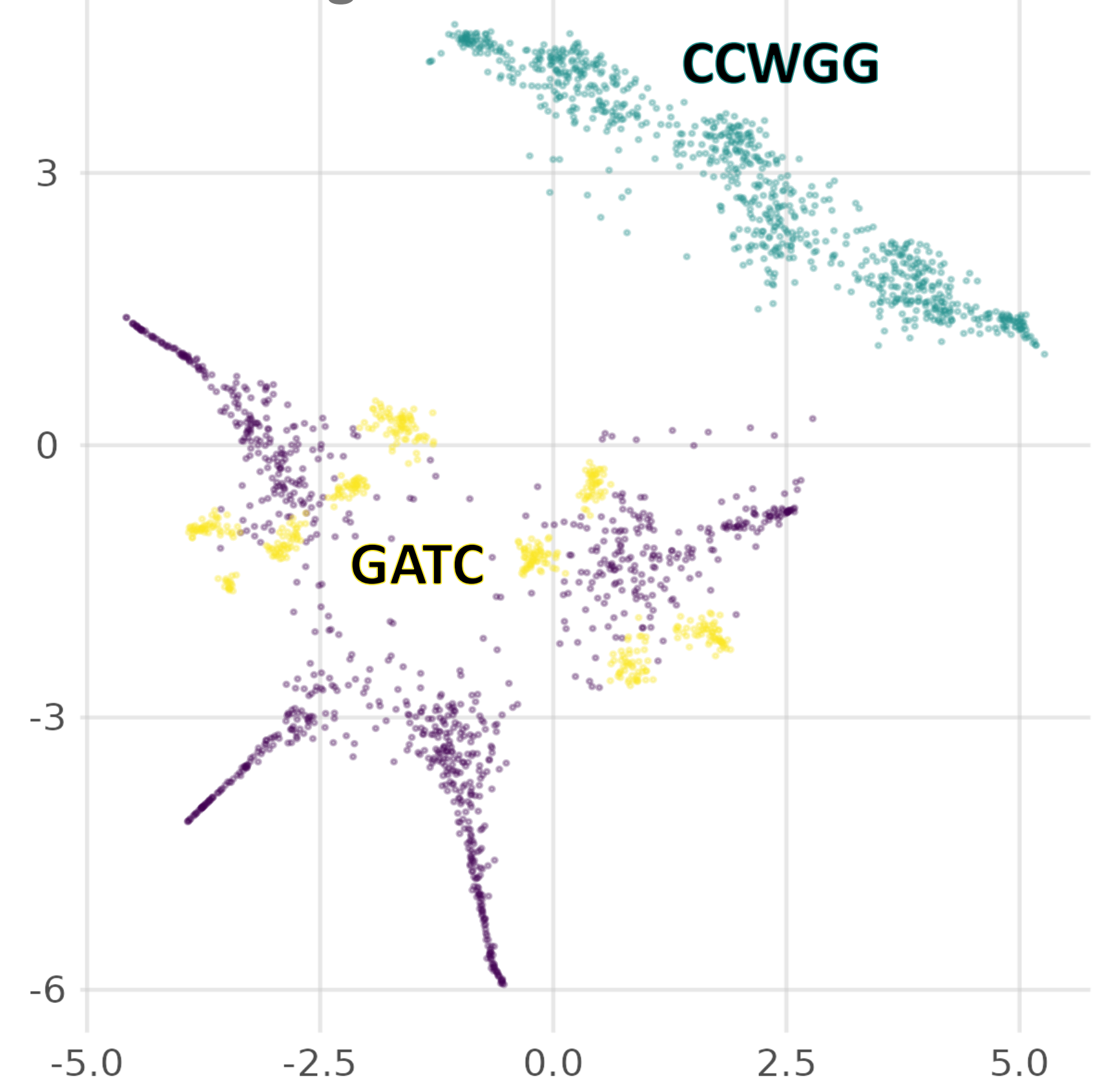
Detection of Methylated Motifs

- ▶ Evaluation using *E. coli* K12 with G6mATC and C5mCWGG motifs

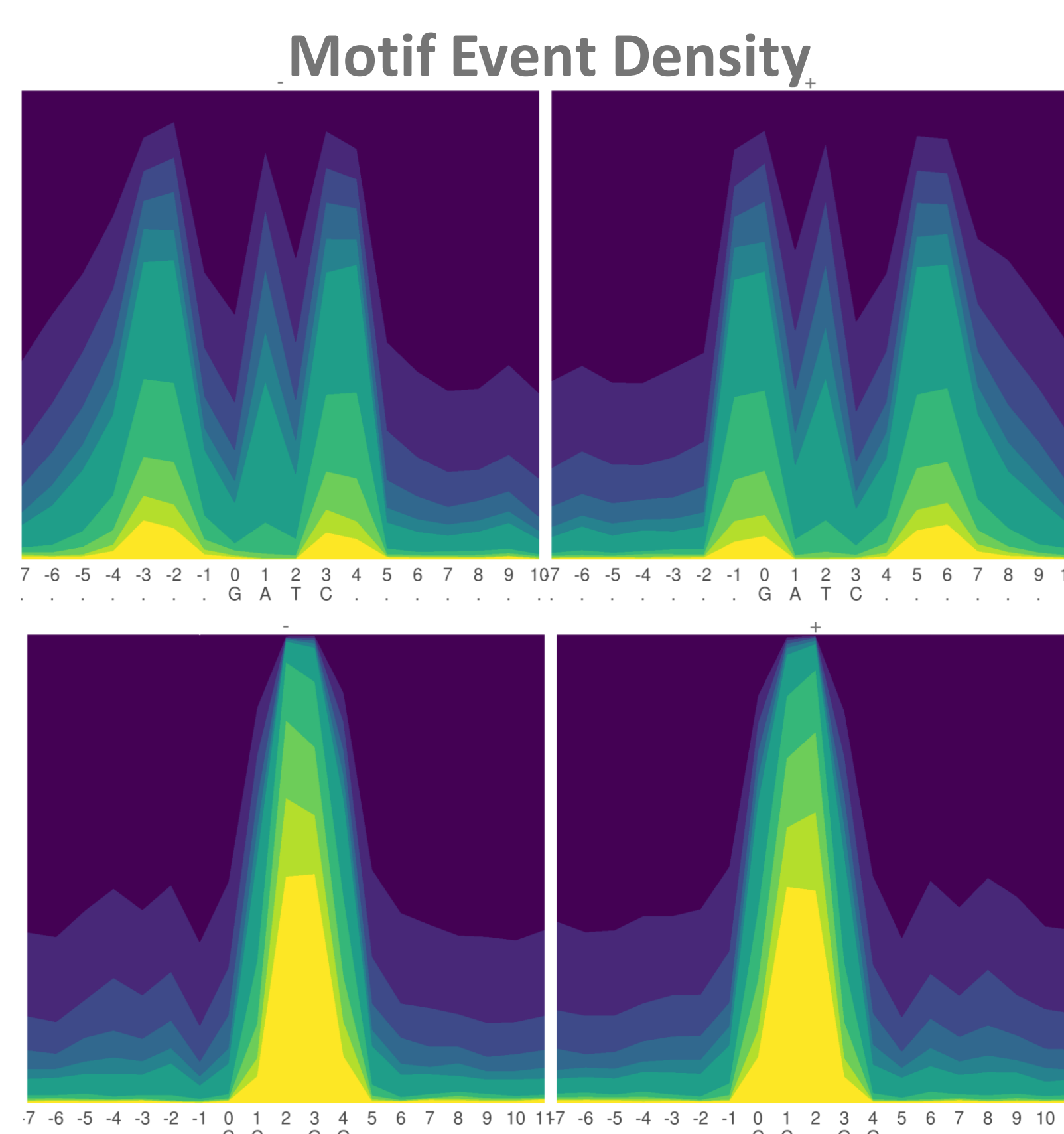


- ▶ *E. coli* -dcm/-dam used as unmodified DNA comparison
- ▶ Only 100 kb and 60x coverage required for stable detection of motifs

Motif detection using 60x coverage & 100 kb reference subset



Modification Type Affects Detectability



Test of Detection Sensitivity

| Motif | Ref Subset [kb] | Coverage | Ref Count | % Clustered |
|----------|-----------------|----------|-----------|-------------|
| CC[AT]GG | 100 | 20 | 276 | 100 |
| CC[AT]GG | 50 | 20 | 125 | 100 |
| GATC | 100 | 20 | 454 | 21 |
| GATC | 50 | 20 | 241 | 15 |

- ▶ C5mCWGG and G6mATC have very different current disturbance patterns
- ▶ Disturbances from 5mC are greater and therefore more readily detected



SørenHeidebach
shei@bio.aau.dk
SørenHeidebach/nanomotif