Decoding the epitranscriptomic code in *Plasmodium falciparum*

Steven T. Lancashire[1], Joshua Levendis[1], Lakvin Fernando[1], Emma McHugh[1], Stuart A. Ralph[1]

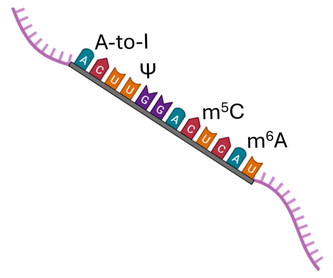
1. The Department of Biochemistry and Molecular Biology and Bio 21 Molecular Science and Biotechnology Institute, The University of Melbourne, Victoria 3010

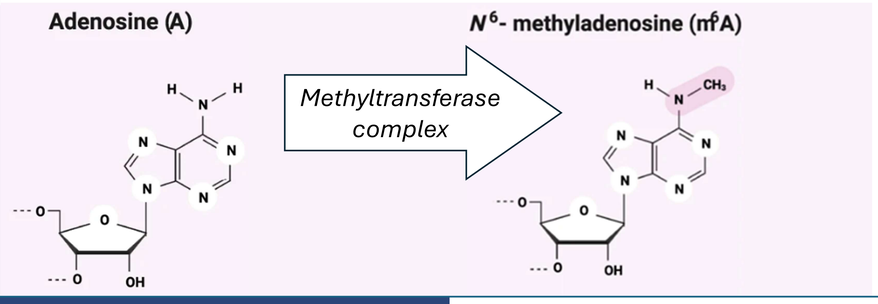
Background

Plasmodium falciparum, the infectious parasite that causes malaria in humans has an mRNA transcriptome that undergoes extensive post-transcriptional chemical modifications. These modifications are implicated in the transcriptional regulation of the parasite's complex lifecycle. The most prevalent modification is adenosine to N6-methyl-adenosine (m^6A). These modifications to mRNA occur at DRACH sequence motifs, however less than 10% of DRACH sites are modified. This bioinformatics masters project aims to identify the other primary sequence and secondary structural motifs required for m^6a modification.

What is the epitranscriptome?

The epitranscriptome is the set of chemical alterations to the nucleotides of RNA. There are 170+ known RNA modifications, with m^6A the most abundant in mRNA. Modifications to mRNA influence translational efficiency, mRNA stability and splicing.

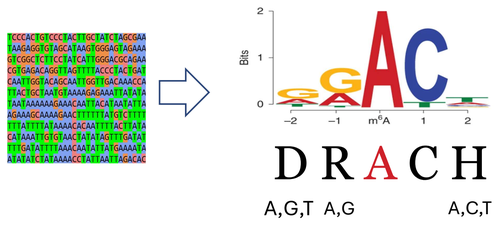




Methylation of a DRACH motif's central adenosine to m^6A is performed by a methyl-transferase heterodimer that is conserved in eukaryotes.

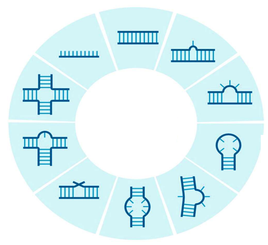
What are RNA motifs?

Primary sequence



An overrepresented pattern of nucleotides in the RNA's sequence

Secondary structure



Conserved RNA structures arising from heterogenous sequences

Less than 10% of DRACH motifs carry m^6A

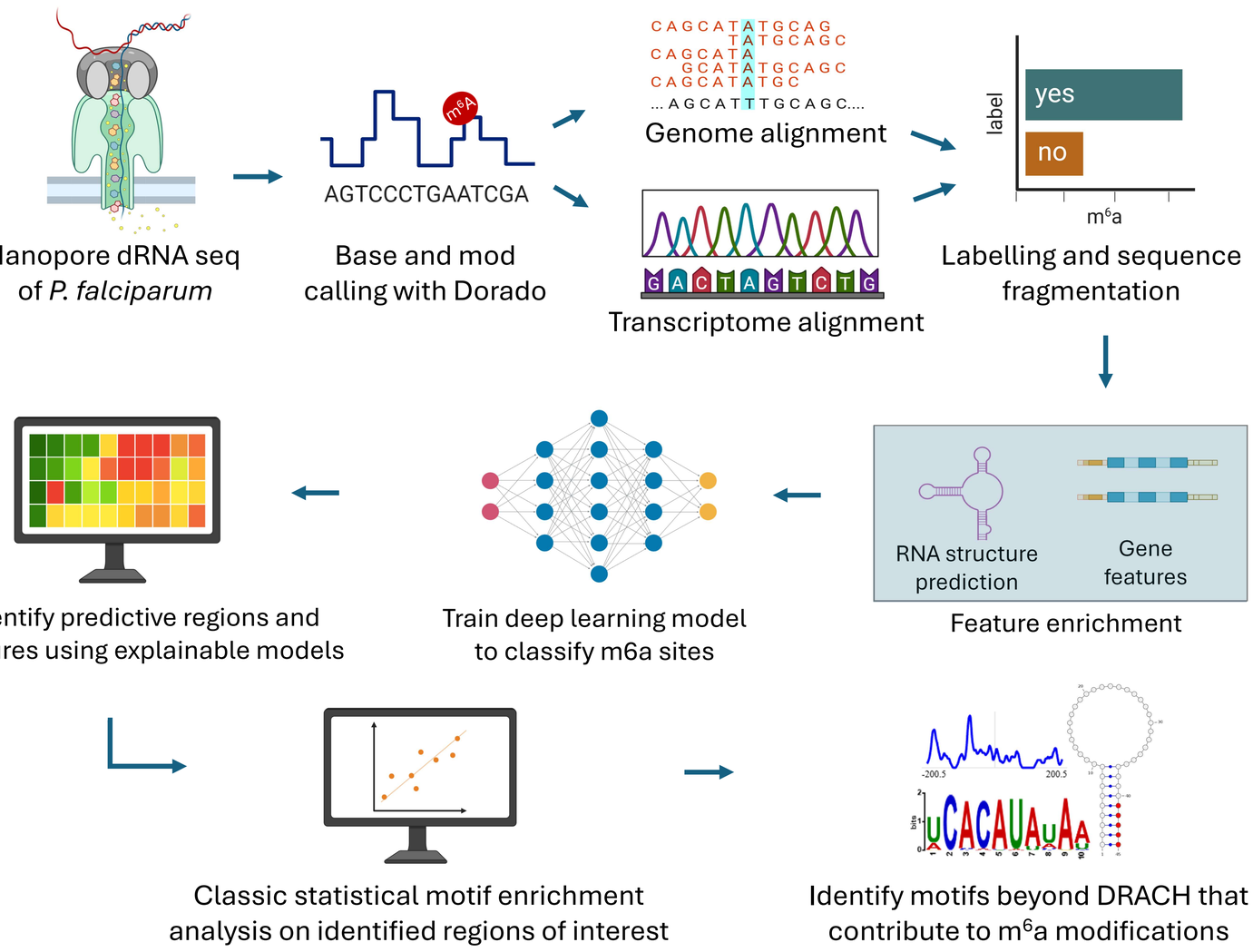
m^6A impacts translational plasticity across *P. falciparum's* lifecycle

- Modification reduces and delays translation, resulting in parasite lifecycle supporting just-in-time translation
- The m^6A reader, PfYTH2, represses protein translation and is essential for parasite survival
- m^6A in 3' UTRs facilitates 3' cleavage of the mRNA transcript

What RNA sequence and structural motifs do m^6A modifications require?

Motif discovery

Plasmodium falciparum long read dRNA-seq data are combined with RNA structure and gene features. Deep learning models are used to identify RNA regions predictive of m^6A. Statistical motif enrichment analysis is then applied to these regions to discover motifs.



Nanopore dRNA seq of *P. falciparum*

Base and mod calling with Dorado

Genome alignment

Transcriptome alignment

Labelling and sequence fragmentation

Feature enrichment

Train deep learning model to classify m6a sites

Identify predictive regions and features using explainable models

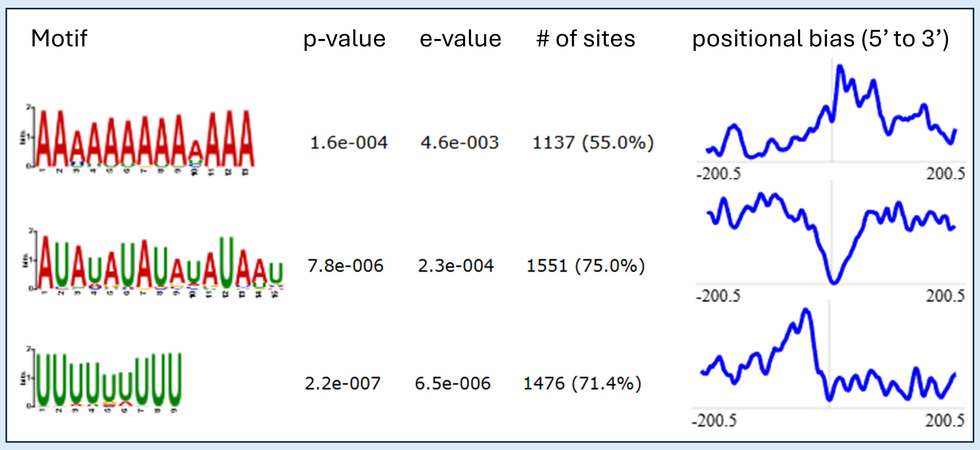
Classic statistical motif enrichment analysis on identified regions of interest

Identify motifs beyond DRACH that contribute to m^6a modifications

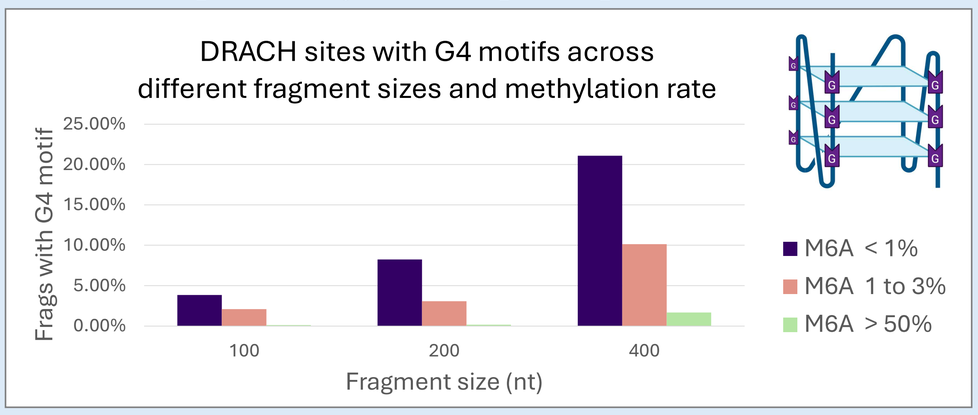
Preliminary results

m^6A is selectively enriched and depleted in different motif contexts

Result 1: Enriched homopolymer motifs show directional bias around m^6a modified DRACH sites.



Result 2: DRACH sites are less likely to be modified near motifs that potentially form G-quadruplex (G4) structures.



References

Sinha, A. *et al.* Functional Characterization of the m6A-Dependent Translational Modulator PfYTH.2 in the Human Malaria Parasite. *mBio* **12**, 10.1128/mbio.00661-21 (2021).

Baumgarten, S. *et al.* Transcriptome-wide dynamics of extensive m6A mRNA methylation during *Plasmodium falciparum* blood-stage development. *Nat Microbiol* **4**, 2246–2259 (2019).

Bailey, T. L. STREME: accurate and versatile sequence motif discovery. *Bioinformatics* **37**, 2834–2840 (2021).

Acknowledgements

Stuart Ralph, Joshua Levendis, Lakvin Fernando, Emma McHugh, Sophie Collier, Long Huynh, Haowen Deng, Zoe Tregloan-Dunn

stevelan.github.io/m6a-motifs