

RawMod: A Deep Learning Approach to Modification Agnostic Classification in Nanopore Sequencing

Bhargav Srinivasan, Simon Ambrozak, Can Firtina
University of Maryland Department of Computer Science

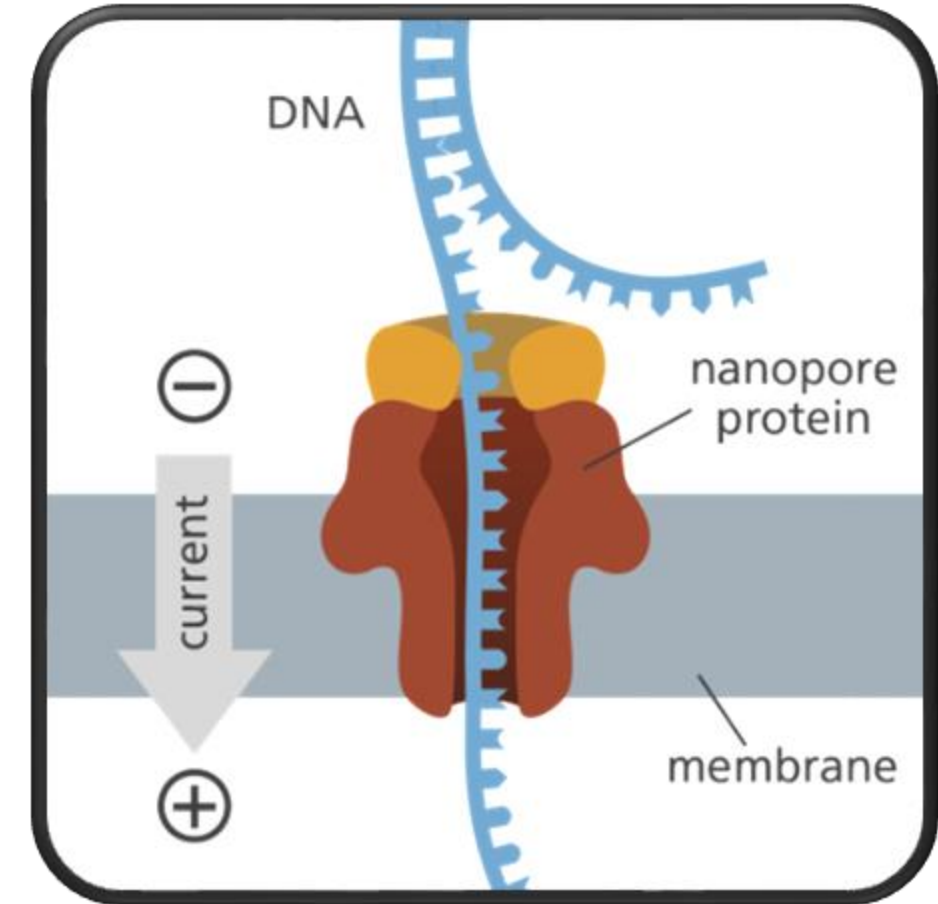
STORM



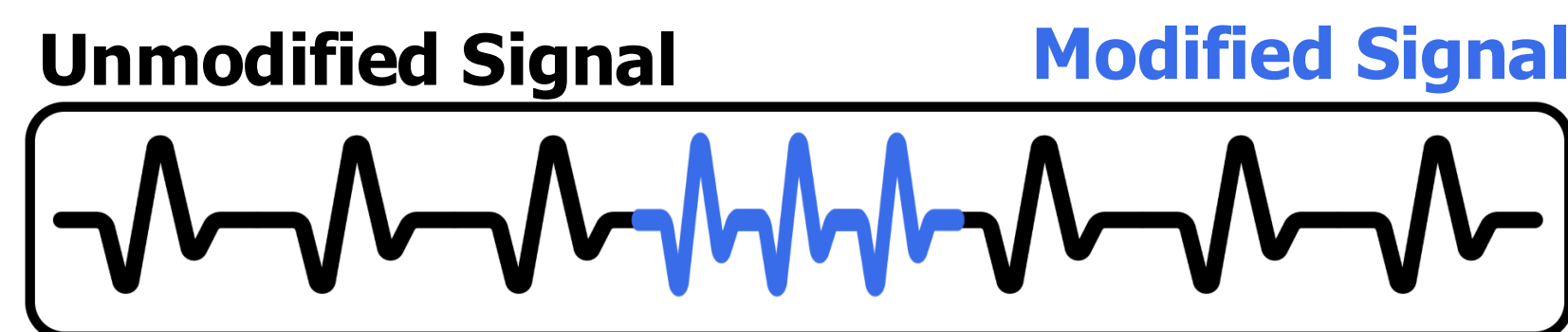
Contact: bds062@umd.edu

1. Modification Detection in Nanopore Sequencing

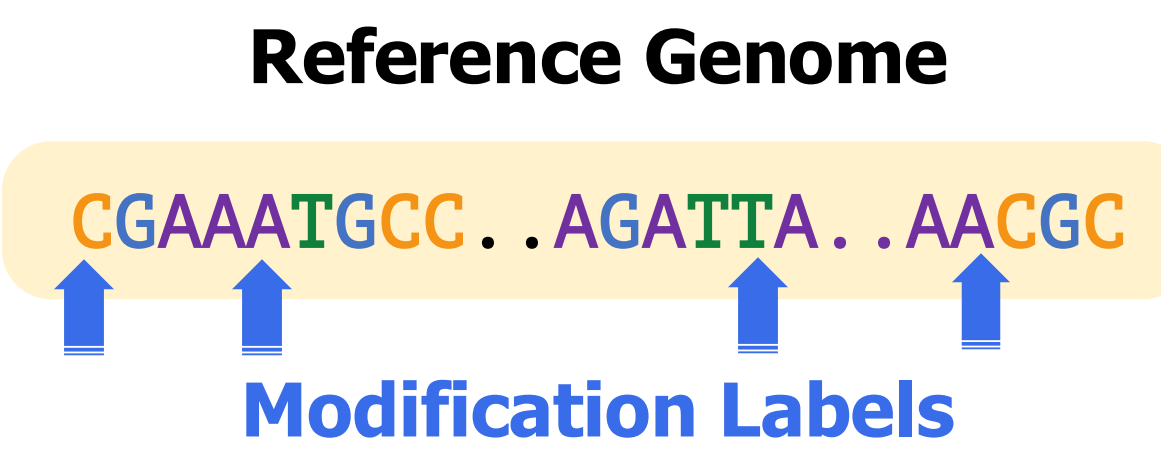
1. Nanopore Sequencing



2. Raw Nanopore Signal



3. Modification Detection

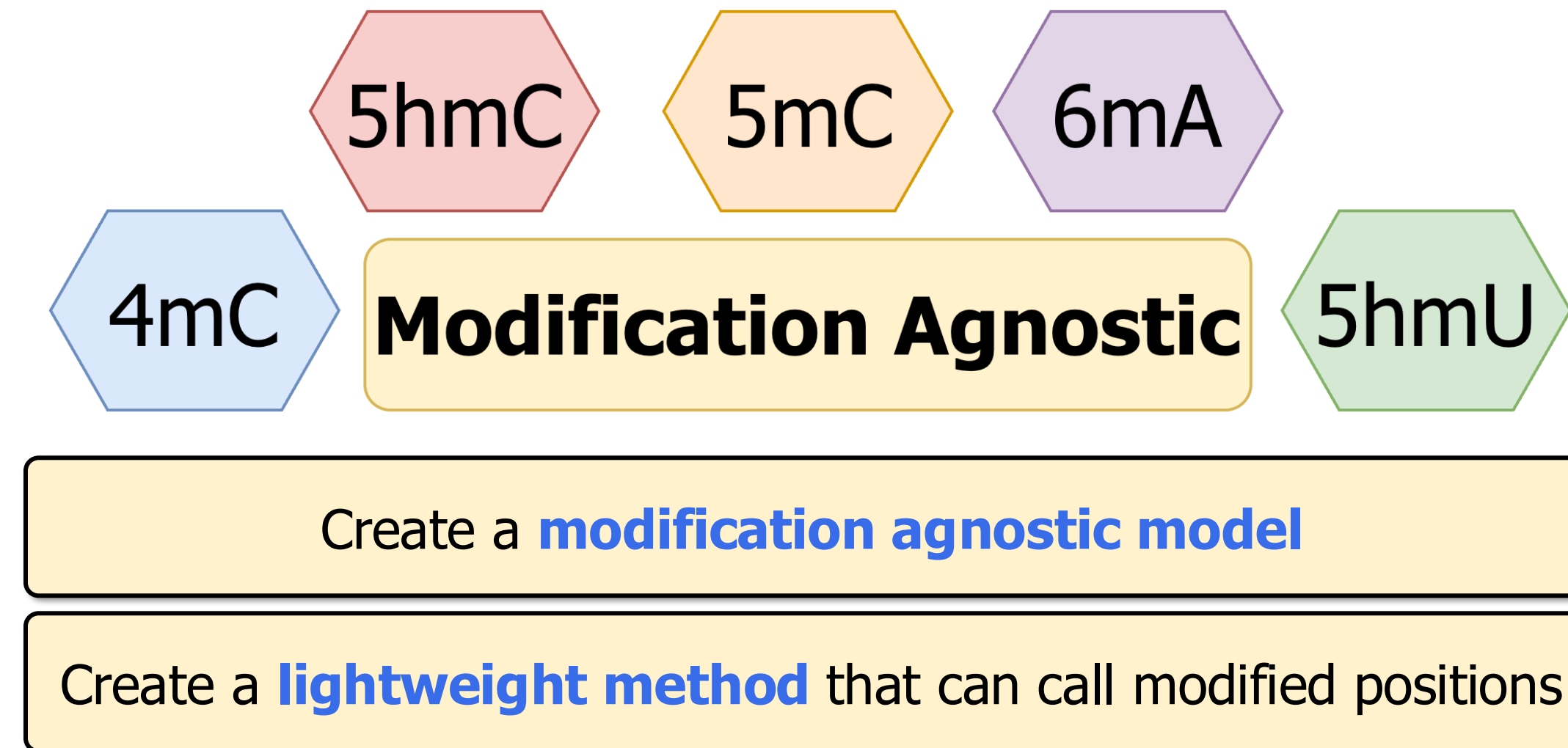


Epigenetic modifications act as chemical mechanisms appended to canonical nucleotides for genetic expression, signaling, and inheritance

2. Current Limitations

- Modification specific tools prevent epigenetic generalization
- Rare modifications are inaccurately identified
- Modification methods are **time and computationally costly**
- General modification methods **require a control sample**

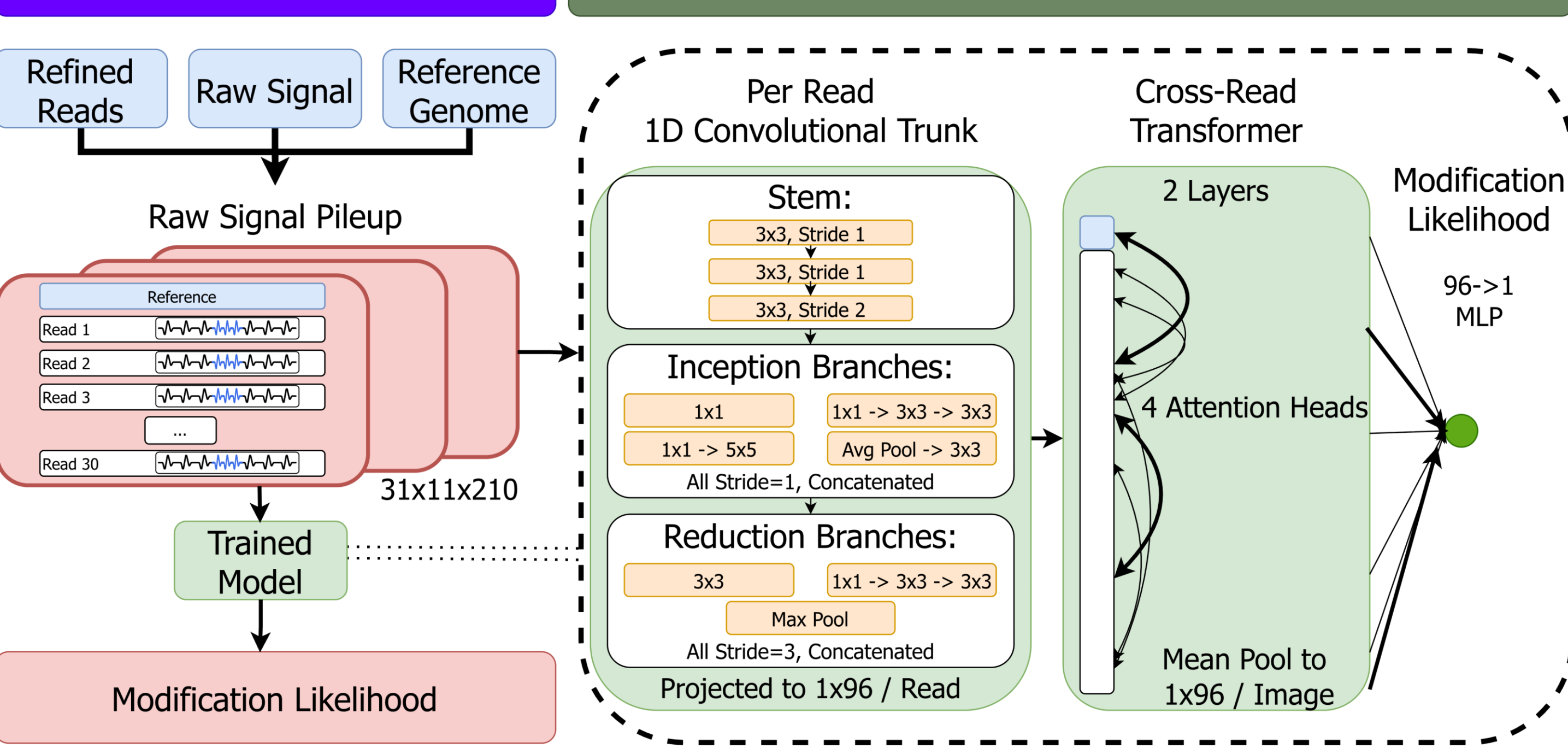
3. Our Goal



4. Our Method

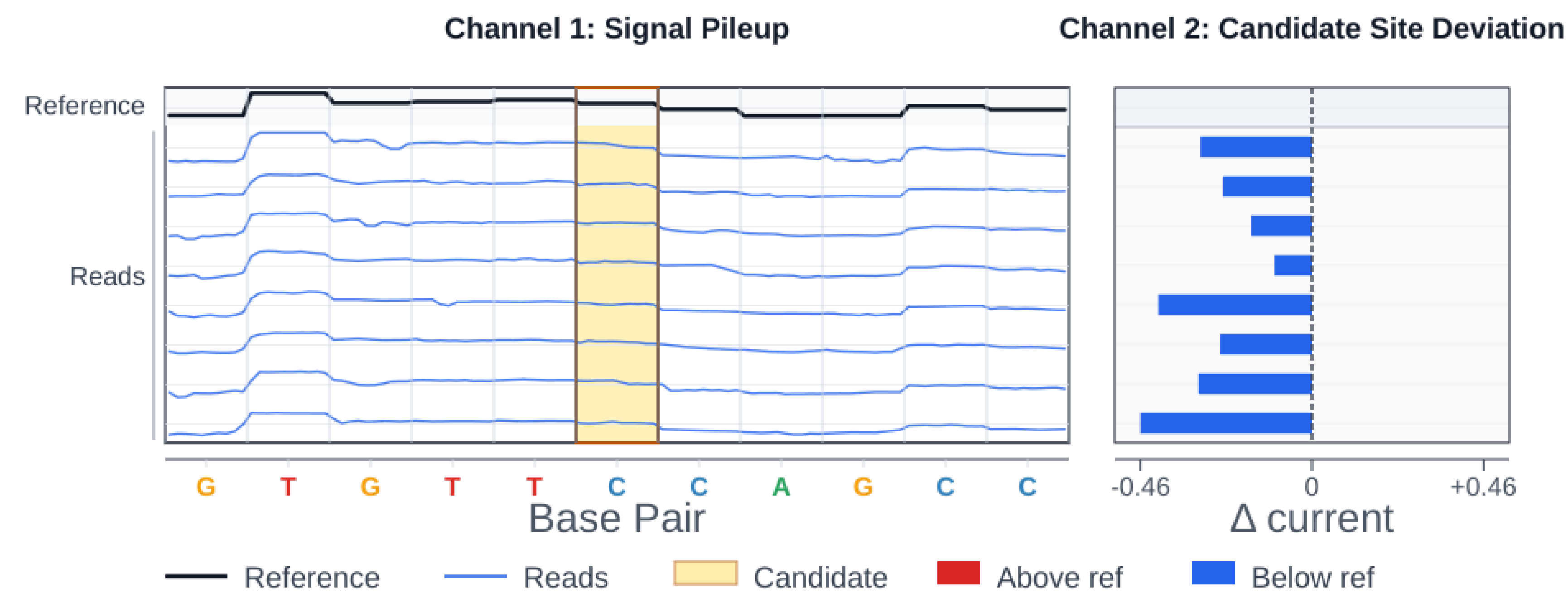
A: RawMod

B: Model Architecture



- A) RawMod consists of deriving modification likelihoods from a raw signal pileup
- B) We employ a Transformer block appended to a Convolutional Trunk to collect refined information from each read, then come to a consensus regarding the queried site

5. Candidate Site Raw Signal Pileup



Signal pileup is done by resampling signal corresponding to each base with Remora Refinement

Reference signal is generated with the k-mer lookup table for each site in the image

Additional channels include **dwell time**, base pair encoding, mapping quality, strand direction, matches reference, and window difference from reference

6. Evaluation Methodology

Datasets:

- ONT DNA Modification Benchmarking Collection
 - 5mC, 5hmC, and 6mA synthetic datasets
- UMCES SPO1 Collection
 - 5mC, 5hmC, 6mA, and **5hmU** WGS datasets

Model Training: Train on a random sample of sites from a majority of ONT and SPO1 reference genomes

Evaluation:

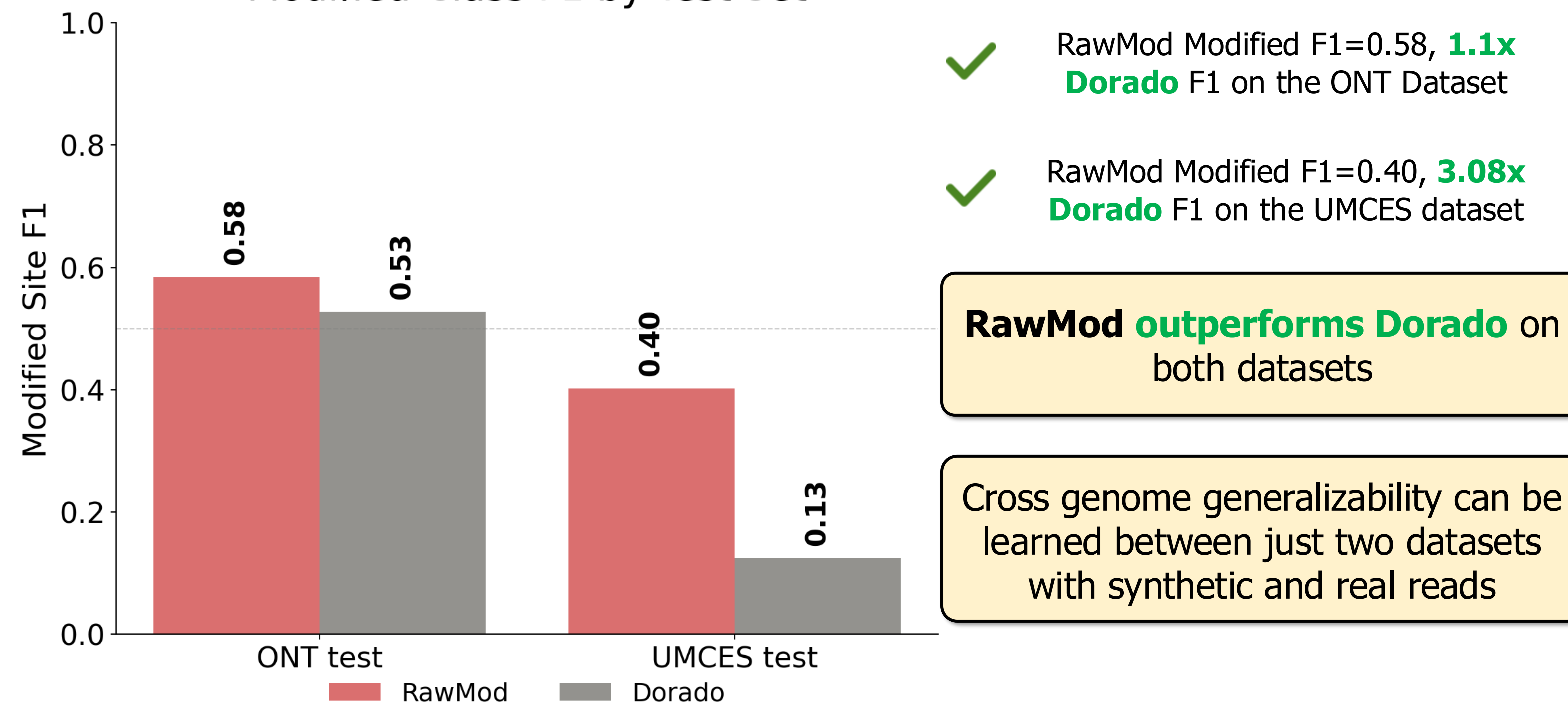
- Modification Agnostic Accuracy
 - Tested on **unseen regions** of both genomes
- Modification Generalization
 - Leave one modification out** experiment during training

Benchmarking Comparison:

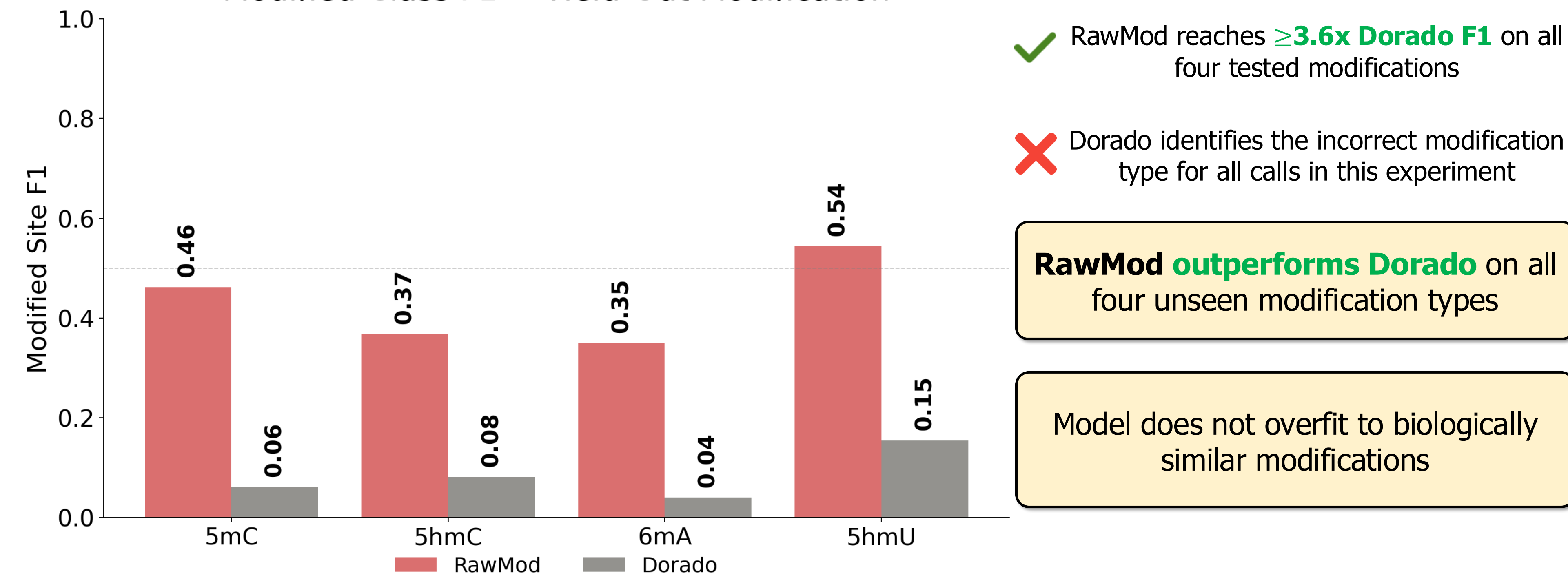
Predicted modified if Dorado **5mC, 5hmC, or 6mA models** predict the site to be modified – modification agnostic simulation

7. Results: Detection Accuracy | Modification Generalization

Modified-Class F1 by Test Set



Modified-Class F1 — Held-Out Modification



8. Key Contributions

The first modification agnostic model to achieve **high accuracy** on uncommon modification types

Lightweight method with low parameter count and **low runtime**

9. Future Work

Expand to **additional modification types and genomes**

Benchmark against **additional methods** such as RockFish, MethyNano, and DeepMod