

A DRUG DESIGN PLATFORM LEVERAGING FLANKING SEQUENCES TO ENHANCE ANTI-MIRNA BINDING



LU, TZU-TSANG
NATIONAL CHENG KUNG
UNIVERSITY, DEPT. OF
BIOMEDICAL ENGINEERING

Learning how flanking sequences influence miRNA-target binding from real CLASH chimeric reads



ABSTRACT

This study builds a deep learning model to resolve how flanking sequences influence miRNA-target binding. Using CLASH chimeric reads, we construct a high-quality dataset and train a dual-tower cross-attention model, aiming to guide anti-miRNA linker design.



INTRODUCTION

- Anti-miRNA oligonucleotides that inhibit target miRNAs are promising molecular drugs.
- Existing platforms design linkers by intramolecular complementarity, ignoring flanking sequences.
- We model how flanking sequences affect binding, using real interaction data.



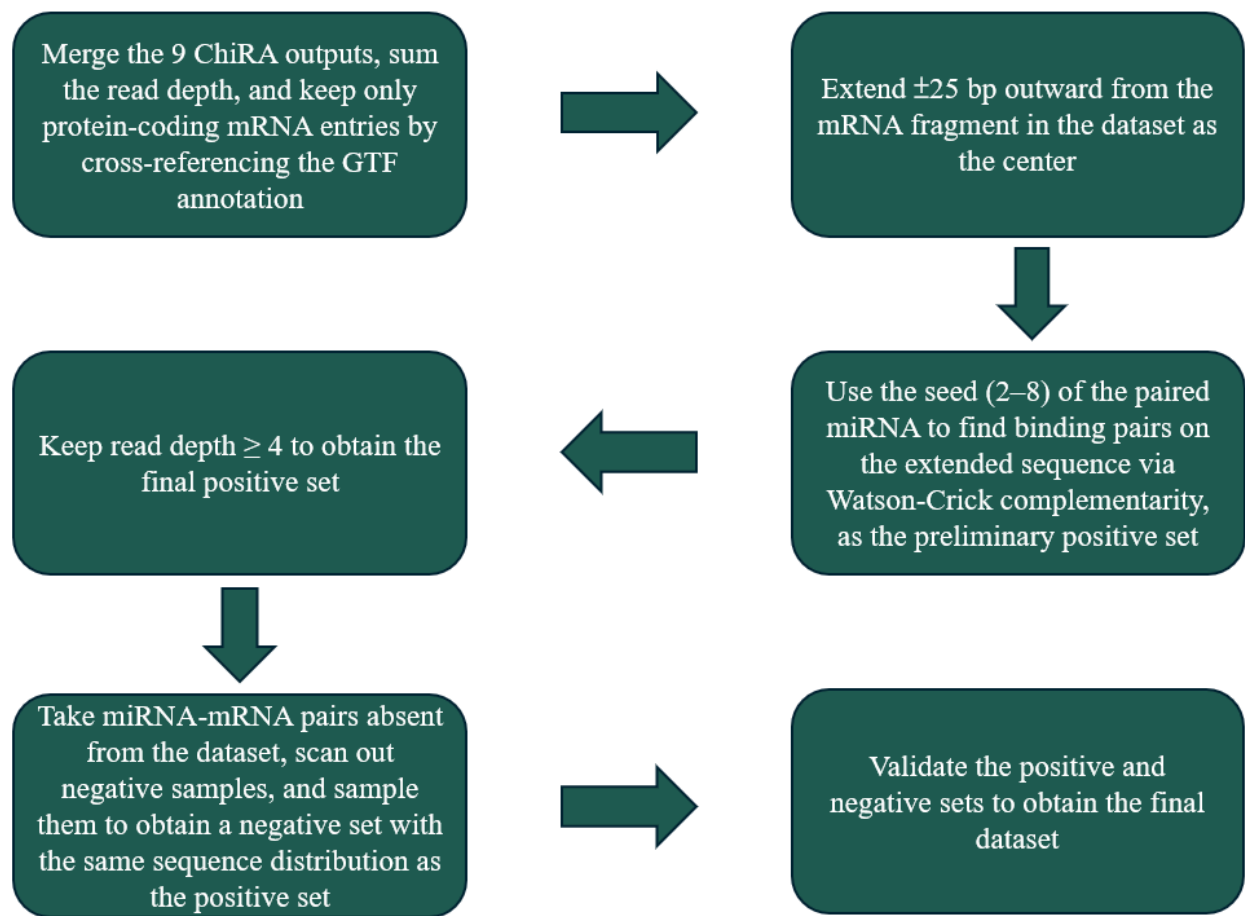
METHODOLOGY

Data:
CLASH data processed via ChiRA pipeline to obtain real chimeric reads

Dataset:
Positive set filtered by seed match, flanking extension, and read depth; leakage-free hard negatives.

Model:
AIDO.RNA embeddings with a dual-tower cross-attention architecture.

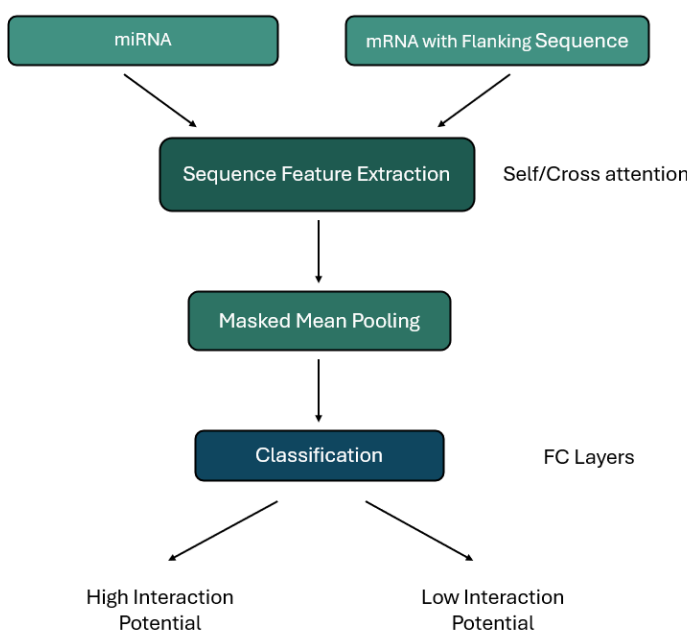
DATASET CONSTRUCTION PIPELINE



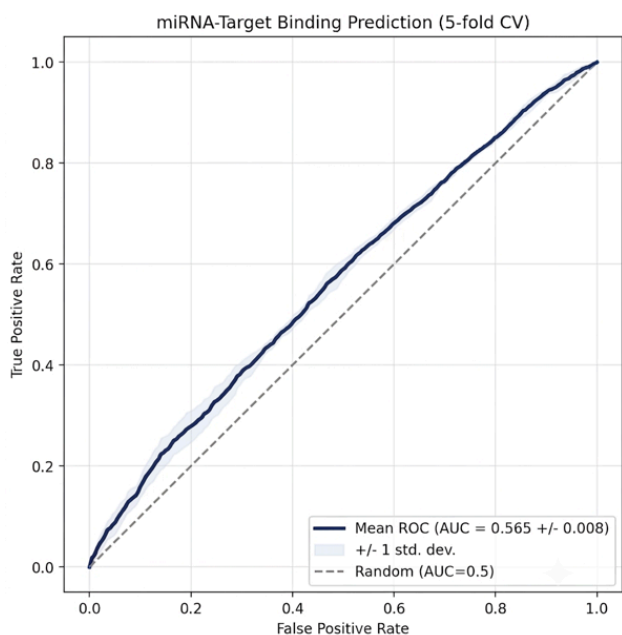
RESULT

- 5-fold cross-validation split by miRNA, with zero leakage (3,799 samples each).
- Mean validation AUC ≈ 0.57 , above random baseline.
- High training but lower validation AUC indicates overfitting.

MODEL ARCHITECTURE



MODEL PERFORMANCE



DISCUSSION

A validation AUC of ~ 0.57 indicates that flanking sequences contain potential binding signals. However, the prominent train-validation gap reveals that this signal is subtle, requiring stronger regularization or expanded datasets to mitigate overfitting.



CONCLUSION

This study provides a rigorous benchmark for flanking sequence impact in miRNA binding. The preliminary validation AUC of ~ 0.57 confirms learnable signals, though model performance requires further refinement. Future work will focus on regularization and feature encoding to support automated linker design.