

Integrating AI, Biophysical Modelling and Experimental Validation for Enhancer Variant Interpretation

Supervisory Team:

Dr Hannah Long and Prof Joe Marsh

Project Overview:

Enhancers play a crucial role in gene regulation, and their perturbation is linked to numerous human diseases. However, predicting the functional impact of enhancer variants remains a significant challenge. This project takes an interdisciplinary approach by integrating AI-driven computational methods with biophysical modelling and high-throughput experimental validation to develop a comprehensive framework for enhancer variant interpretation.

Computational and Biophysical Modelling of Enhancer Variants:

We will employ a hybrid computational strategy that combines genome language models with biophysical simulations to predict the impact of enhancer variants. Genome language models will be benchmarked against published Massively Parallel Reporter Assay (MPRA) datasets to assess their ability to predict enhancer activity and variant effects across diverse cellular contexts. In parallel, biophysical modelling techniques, such as molecular dynamics simulations and structural modelling, will be used to investigate how enhancer variants alter transcription factor binding affinity and DNA conformational dynamics. By integrating AI-based predictions with physical principles of molecular interactions, we aim to improve the accuracy and interpretability of enhancer variant impact assessments.

Experimental Validation of Enhancer Variant Predictions:

Predictions generated from the computational and biophysical models will guide experimental validation efforts using high-throughput synthetic biology approaches. We will apply saturation mutagenesis and MPRA technologies to systematically test the regulatory activity of selected enhancer variants in disease-relevant cellular models. This experimental validation will provide ground-truth data to iteratively refine computational models, ensuring they accurately capture enhancer function in a biological context. Additionally, novel enhancer sequences predicted through modelling will be synthesized and tested to explore their regulatory potential.

Training Environment:

This interdisciplinary project offers a unique training experience at the intersection of computational biology, biophysics, and experimental genomics. The PhD candidate will gain hands-on experience with AI models, molecular simulations, and high-throughput functional assays, equipping them with a diverse skill set applicable to both academia and industry.

Impact:

By combining AI-driven predictions with biophysical modelling and experimental validation, this project will enhance our ability to interpret non-coding genetic variation, contributing to improved disease risk assessment and therapeutic target identification.