

COMP39/9900 Computer Science/IT Capstone Project

School of Computer Science and Engineering, UNSW

Project Number: P5

Project Title: Deep-Learning Techniques to Understand Chromosome Interactions with Enhanced Interpretability and Uncertainty Estimation

Project Clients: Hamid Alinejad-Rokny

Project Specializations: Software development; Computer Science and Algorithms; Artificial Intelligence (Machine/Deep Learning, NLP); Big data analytics and visualization; Bioinformatics/Biomedical

Number of groups: 3

Background:

Supervisory team: Dr. Hamid Rokny (GSBmE), Dr. Reza Argha, and Prof. Nigel Lovell (GSBmE).

Abstract: The 3D arrangement of our DNA critically influences gene regulation by positioning regulatory elements such as enhancers close to gene starting points, or promoters. Understanding how gene expression is regulated across various cell types has seen significant progress; however, the study of DNA's 3D structural variations among these cell types remains a pioneering area of research. Recent technological advancements have enabled the examination of DNA organization at the single-cell level, exposing cell-specific structural differences. Yet, a comprehensive understanding of how these structural variances affect gene activity requires the development of more sophisticated computational models. This project aims to develop advanced deep-learning methods that not only predict but also interpret the complex interactions within DNA's 3D structure. By integrating techniques for enhanced interpretability, our models will provide meaningful insights into the biological significance of chromosomal arrangements, making the complex interplay more accessible to biologists. Additionally, we will incorporate uncertainty estimation within our computational framework, ensuring robust predictions even when dealing with the inherent variability and noise present in single-cell genomic data. This dual approach will improve confidence in the model outputs and facilitate a deeper understanding of the dynamic 3D genome architecture's role in cell-specific gene regulation.

Resources: We have access to 3D genome organization data from more than 50 samples. We have also local high-performance GPU and CPU computing systems in UNSW Biomedical Machine Learning Lab.

The skills required for project(s): Scripting experience in Python.

What skills will you gain during your degree: Machine learning and data analytics techniques, Bioinformatics, Systems Biology.

References

<https://www.nature.com/articles/nature12593>
<https://www.nature.com/articles/s41587-021-01034-y>
<https://www.pnas.org/doi/10.1073/pnas.1901423116>

Requirements and Scope:

The project aims to advance our understanding of the 3D structure of DNA and its impact on gene regulation. Here's an outline of the project scope:

Development of Deep Learning Models: Create advanced algorithms to model DNA's 3D interactions, focusing on single-cell genomic data.

Enhanced Model Interpretability: Implement interpretability techniques to make model predictions accessible and meaningful to biological researchers.

Uncertainty Estimation: Incorporate methods to estimate prediction uncertainty, addressing the variability and noise in genomic data.

Cross-Cell Type Analysis: Conduct analyses comparing DNA structures across different cell types to understand structural influences on gene activity.

Collaborations and Data Sharing: Partner with genomic labs for data access and establish a framework for disseminating research findings.

Application and Validation: Validate the models using independent datasets to ensure their effectiveness in real-world scenarios.

This project combines computational modelling with biological insights to explore the dynamic nature of genome architecture across different cellular environments.

Resources: We have access to 3D genome organisation data from more than 50 samples. We have also local high-performance GPU and CPU computing systems at UNSW BioMedical Machine Learning Lab.

Required Knowledge and skills:

The skills required for project(s): Scripting experience in Python.
Don't worry about the biology side of the project; we'll teach you.

Expected outcomes/deliverables:

The expected outcomes and deliverables for the project are:

Expected Outcomes:

Robust Predictive Models: Development of deep-learning models for accurate predictions of DNA's 3D interactions.

Interpretability Insights: Enhanced understanding of DNA structures influencing gene regulation via interpretable models.

Reliable Predictions: Integration of uncertainty estimation to ensure confidence in predictions from noisy single-cell data.

Deliverables:

Software Tools: A comprehensive suite of software tools, complete with documentation for user guidance.

Research Publications: Research papers detailing methodologies and findings, published in peer-reviewed journals.

Data and Models: Public release of processed datasets and trained models for community use and further research.

Training Sessions: Workshops or webinars to educate researchers on the tools developed and their genomic applications.

Collaboration Reports: Documentation of collaborative efforts, showcasing practical applications in various biological studies.

Supervision:

Hamid Alinejad-Rokny

Additional resources: