NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

School of Biological Sciences
College of Science

Reg. No. 200604393R

Research Theme: **Molecular Biology or Computational Biology**

PhD Research Project Title:

Building Predictive Models of Human Development Using Single-Cell Perturbation Data

Scholarship category (Please indicate the source of funding for this project):

SBS Research Student Scholarship

Principal Investigator/Supervisor: **Richard She**

Co-supervisor/ Collaborator(s) (if any):

Project Description

a) Background:

A central goal in biology is to understand how the genome encodes the regulatory instructions that shape human development. While the protein-coding functions of genes are relatively well understood, we still lack a systematic, predictive model for how gene expression is regulated across different human cell types. Understanding how these regulatory networks are wired—and how they respond to genetic perturbation—is one of the most fundamental questions in biology. It is also essential for interpreting how genetic risk factors for complex diseases affect different organ systems during development.

To address this, our lab is building a platform for mapping gene regulatory networks at scale in human stem cell–derived tissues. We begin with induced pluripotent stem cells (iPSCs), which can be differentiated into nearly any human cell type. By directing these cells into three well-characterized lineages—**neurons**, **cardiomyocytes**, and **hepatocytes**—we capture the developmental trajectories of the three primary germ layers (ectoderm, mesoderm, and endoderm). These cell types are not only central to human physiology but also representative of distinct transcriptional programs, making them an ideal testbed for exploring both conserved and tissue-specific regulatory logic.

Using CRISPR-based perturbation and single-cell RNA sequencing (Perturb-seq), we can measure how hundreds of gene perturbations affect transcriptional output in each cell type. These data allow us to reconstruct functional relationships between genes and begin to train machine learning models that predict how regulatory interactions vary across developmental contexts. While many efforts have focused on predicting gene expression from DNA sequence or motif content, our approach instead learns directly from empirical perturbation data—grounding predictive models in the observed behavior of real human cells during differentiation.

b) Proposed work:

Incoming Ph.D. students will have the opportunity to lead a large-scale experimental and computational effort to map and model human gene regulatory networks. This project is designed to produce meaningful results within a four-year timeline, with well-scaffolded modules spanning cell line engineering, data generation, and model development

Project 1: Cross-lineage functional genomics of human development using Perturb-seq and predictive modeling

This project begins with the differentiation of iPSCs into neurons, cardiomyocytes, and hepatocytes using robust, well-validated protocols. These three cell types model early development of the brain, heart, and liver, and are representative of major developmental lineages. Once differentiated, each cell type will be transduced with a pooled CRISPRi/a library targeting approximately 500 regulatory genes—including transcription factors, chromatin modifiers, and pathway components. Single-cell RNA sequencing will then be used to quantify the effect of each perturbation on gene expression in thousands of individual cells.

The resulting datasets will allow us to map gene regulatory relationships at high resolution and identify both shared and lineage-specific modules of gene expression control. These data will also support the development of AI models that predict the transcriptional consequences of gene perturbations—learning not just whether a gene is essential, but how its influence depends on the cellular context in which it operates. For example, does perturbing a given transcription factor alter neurogenesis but leave hepatic development unchanged? Can we predict these effects in advance based on co-expression patterns or prior perturbations?

Students will have the opportunity to take ownership of both experimental and computational components of the project. These include optimizing lineage-specific differentiation protocols, designing and validating CRISPR libraries, building high-quality single-cell datasets, and participating in the development and benchmarking of predictive models. The project is well-positioned for early progress: cell lines and protocols are already in place, and the first round of pooled screens is underway.

Students in this project will receive training in stem cell differentiation, pooled CRISPR screening, single-cell transcriptomics, and data integration using statistical and machine learning tools. These skills are broadly applicable in both academic and industry settings, and will prepare students to work at the interface of functional genomics, systems biology, and disease genetics.

c) Preferred skills:

Willingness to learn and passion for science are most important. Prior wet lab experience is helpful but can be taught. Computational skills will be essential for large scale data analysis.

Supervisor contact:

If you have questions regarding this project, please email the Principal Investigator:

richardsheshe@gmail.com

SBS contact and how to apply:

Associate Chair-Biological Sciences (Graduate Studies): AC-SBS-GS@ntu.edu.sg

Please apply at the following:

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

School of Biological Sciences
College of Science

Reg. No. 200604393R

Application portal:
https://venus.wis.ntu.edu.sg/GOAL/OnlineApplicationModule/frmOnlineApplication.ASPX

60 Nanyang Drive, Singapore 637551