Project: Machine learning in gene regulation and epigenomics

A fundamental question of biology concerns the relationship between an organism's DNA sequence (its genotype) and its observable traits (its phenotype). Solving this genotype-phenotype problem will enable understanding and engineering organisms towards a wide variety of goals including biotechnology, agriculture and health. Towards this end, researchers have recently sequenced of thousands of genomes and collected of tens of thousands of genome-wide measurements, providing a unique opportunity to address this genotype-phenotype problem.

Most consequential genetic variation in mammals causes changes in gene regulation. Gene regulation is the process by which a cell determines which genes to express as proteins. This process allows for an organism’s single genome to express hundreds of tissues and cell types. Thus, understanding gene regulation is key to understanding genetic differences between individuals. A complete understanding of gene regulation promises to lead to monumental advances in bioengineering and health. Yet many aspects of gene regulation remain poorly understood. The recent development of sequencing-based genomic assays promises to grant insight into gene regulation. These assays use high-throughput DNA sequencing as a readout to measure genomic activities including transcription, DNA-protein binding, chromatin accessibility and others. As of this writing, international consortia such as CEEHRC, ENCODE and IHEC have performed tens of thousands of genomic assays in a variety of tissues from organisms including human, mouse and fruit fly.

The successful applicant will develop machine learning methods to solve important problems in genomics. They will work in the research group of Maxwell Libbrecht, where they will design analysis, implement software, develop visualizations and contribute to research papers for publication in computational biology conferences and journals. The applicant should be an undergraduate student in Computing Science or a related field and have experience with algorithms, probability and machine learning.

Desired qualifications:

- Undergraduate student in Computing Science or a related field.
- Coursework and experience in algorithms, probability and machine learning (CMPT 310 and 419).
- Knowledge of biology and genomics is beneficial but not required.

More information about the group’s research can be found on our website: [www.libbrechtlab.com](http://www.libbrechtlab.com)

Applicants should email a CV and transcript to Dr. Libbrecht at [maxwl@sfu.ca](mailto:maxwl@sfu.ca)