2024 Multidisciplinary Research Program in Medicine Project*: Machine learning enabled high-throughput single cell isolation for cancer genomics using inkjet printing*

Hypothesis or Research Question(s): A new system that incorporates optimized inkjet dispensing and machine learning can enable rapid and accurate isolation of single cells for genomic and transcriptomic high-throughput sequencing applications.

PROJECT BACKGROUND & SUMMARY

The cell is the basic unit of life, and new methods for profiling the genetic characteristics of individual cells are now transforming our understanding of early development, tissue regeneration, immunology and disease. Cancer arises when a single cell acquires genetic alterations leading to uncontrolled replication. As tumour cells divide they continue to acquire mutations which they pass on to their descendants, forming sub-populations with different characteristics. This genetic diversity can allow tumours to evolve resistance to treatment and eventually spread. As such, profiling the genetic alterations and patterns of gene expression within tumours and their surrounding environment can reveal mechanisms underlying cancer relapse and metastatic progression.

Existing single cell omics platforms present deficiencies in throughput, accuracy and customizability. We are developing an inkjet-based instrument that integrates high-resolution imaging and neural network-based object recognition to rapidly isolate cells for genetic profiling. Building on a working prototype recently installed in the Genome Sciences Centre at the BC Cancer Research Institute, we will engineer new features to ensure accurate cell classification, temperature control, and ease of operation. We will validate the platform by implementing protocols for single cell DNA and RNA sequencing, benchmarking efficiency and data quality relative to an existing commercially available instrument.

The project requires multidisciplinary expertise spanning biomedical and electrical engineering, molecular biology, and bioinformatics. Key steps in a single cell genomics workflow include: cell culture or tissue dissociation, single cell isolation, preparation of nucleic acid libraries, high-throughput sequencing, and analysis of large-scale genomic datasets. Feedback from each step in the process will be used to optimize earlier steps. As such, collaboration and cross-disciplinary communication will be critical for the success of the project.

Single cell genomic technologies are being widely adopted in academic research, as well as in the biotechnology, pharmaceutical and diagnostic sectors. The device under development will provide an open and customizable platform for a range of single cell omic workflows, permitting in-depth characterization of cellular composition in healthy tissue and disease.

BENEFIT TO THE STUDENTS

The project involves a collaboration between Principal Investigators from the Departments of Medical Genetics/BC Cancer and Electrical Engineering.

An undergraduate student with a background in engineering will assist with implementing new platform features. The prototype device is an inkjet spotter capable of dispensing single cells onto a substrate in preparation for sequencing. TS1 will have the opportunity to work alongside an engineering PhD student to support implementation of computational elements of the device. TS1 will assist with advancing machine learning capabilities, such as in the capture and labeling of training data and implementation of different neural network architectures. The goal will be to improve upon the current capabilities of the system as well as expanding classification categories (e.g. differentiation of cells from extracellular debris within a sample). TS1 will have the opportunity to gain experience using different machine

learning frameworks such as TensorFlow and PyTorch in addition to deploying trained networks on the cell isolation devices for performance validation with the goal of use in a production environment.

 An undergraduate student with experience in Python programming will contribute to analysis of high-throughput single cell sequencing datasets derived from the instrument. TS2 will have the opportunity to work alongside a graduate student, implementing and deploying bioinformatic workflows to assess data quality and identify genetic mutations and altered RNA expression in cancer cells. The student will gain experience with the Linux command line and large-scale parallelization using the Genome Sciences Centre's powerful high performance computing (HPC) cluster. TS2 will expand their data science and machine learning capabilities, using packages such as Pandas, Matplotlib and Scikit-learn. The analysis will inform further optimization of the instrument, as well as enable detailed characterization of the genomic and transcriptomic variation that exists between cancer cells and underlies tumour evolution. A medical student will contribute to processing and dissociation of tumour material. The device requires a cell suspension as input, meaning solid tissue must be disaggregated prior to loading. This is a delicate process that can yield clumps of cellular debris or the remnants of dead cells depending on the quality of the sample. In order to ensure that the instrument's imaging and neural network classifier can accurately discriminate single intact cells, TS3 will be trained in established protocols for tissue digestion and nuclear isolation. They will have the opportunity to optimize the process and apply these techniques to a range of tissue samples, generating critical training data that will be used to expand the instrument's machine learning capabilities. TS3 will interact closely with TS1 and TS2, evaluating the impact of the input cell suspension on neural network classification performance and sequencing data quality.

All students will have the opportunity to learn about the research process in a dynamic and collaborative interdisciplinary environment. By the end of the summer they will prepare a presentation on their contribution and progress, and will have an opportunity to contribute to a research manuscript. This experience will help the students to develop interdisciplinary communication skills, while fostering critical thinking through planning and troubleshooting their experiments. These practical skills (including device engineering, computational cancer genomics, and tumour tissue processing) will be valuable to their future careers, in industry, medicine, or in preparation for graduate studies.