

## **Project Abstract:**

Intratumor genetic and transcriptional heterogeneity is a common feature across diverse cancer types, including. CLL is a particular cancer that exhibits genetic and transcriptional heterogeneity along with a highly variable disease course among patients that remains poorly understood. Previous research has established that the presence of particular subclonal mutations in CLL can be linked with adverse clinical outcomes and that these subclonal mutations change over time in response to therapy. Therefore, genetic and transcriptional characterization of these subclonal populations will be paramount to enabling precision medicine and synergistic treatment combinations that target subclonal drivers and eliminate aggressive subpopulations thereby improving clinical outcome.

While bulk measurements and analysis has provided key insights into cancer biology, etiology, and prognosis in the past, this approach does not provide the resolution that is critical for understanding the interactions between different genetic events within the same environmental and genetic backgrounds to drive metastatic disease, drug resistance and disease progression. Single cell measurements are uniquely able to definitively unravel and connect these relationships. However, simultaneous extraction of DNA and RNA from the same single cells is currently not reliable. Therefore, new statistical methods and computational approaches are needed to identify and resolve genetic subpopulations using single cell transcriptional data alone.

In this proposed research, I will develop statistical methods and computational software to analyze single cell RNA-seq data derived from CLL patient samples. Specifically, I will develop methods to identify aspects of genetic heterogeneity, such as the presence of small single nucleotide mutations and regions of copy number variation, in single cells. I will then reconstruct the genetic subclonal architecture and characterize the gene expression profiles of identified subclonal populations. The proposed work will yield innovative statistical methods to enable the identification and characterization of subclonal populations in cancer and yield opensource software that can be tailored and applied to diverse cancer types. Ultimately, application of these developed methods to CLL will provide a better understanding of CLL development and progression.