Specific Aims:

Intratumor heterogeneity is a common feature across diverse cancer types[1,2 3]. Chronic lymphocytic leukemia (CLL) exhibits particularly diverse combinations of clonal and subclonal somatic mutations and copy number variations (CNVs) along with a highly variable disease course among patients that remains poorly understood[3,4]. Dynamic changes can be observed among intra-tumoral subclonal populations over time and following therapy, suggesting an active evolutionary process[5]. Such tumoral evolution can lead to therapeutic resistance and relapse thereby presenting challenges to current standards of cancer treatment[6,7,8].

Transcriptional characterization of these subclonal populations is integral to understanding this evolutionary process. We hypothesize that different combinations of subclonal mutations in CLL will present different transcriptional changes that affect key pathways involved in RNA splicing, apoptosis, cell proliferation, cellular senescence, DNA damage repair, inflammation, Wnt and Notch signaling to ultimately provide particular subclones with enhanced tumorigenic efficiency. 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.

I will take advantage of single-cell RNA-seq datasets generated by the Wu lab at the Dana-Farber Cancer Institute as a part of separate research efforts (1R01HL11645201, 1R01CA15501002) to study subclonal evolution in CLL. I will develop innovative statistical methods to resolve genetic subclonal populations and link transcriptional profiles at the single cell level.

Aim 1: Inferring somatic mutations from single-cell RNA-seq data. Somatic copy number variants (CNVs) and single nucleotide variants (SNVs) play an important role in cancer pathogenesis and progression[9,10,11]. Inferring CNVs from transcriptomic data remains difficult due to uneven coverage across deletion and amplification sites[12]. Similarly, inferring somatic mutations from transcriptomic data is liable to false negatives due to high rates of mono-allelic gene expression[13,14]. Aim 1 will develop a Bayesian hierarchical approach to leverage information from across multiple genetic loci to make probabilistic inferences on presence or absence of CNVs and SNVs. My approach will take into consideration biases introduced by mono-allelic expression and other technical artifacts such as sequencing errors.

Aim 2: Reconstructing subclonal architecture and dissecting subclonal evolution on the single cell level. Intratumoral heterogeneity is a key factor in determining cancer progression, drug resistance, and clinical outcomes[1,2,3,4,5]. Traditional bulk measurements are unable to resolve whether mutations are mutually exclusive or co-occurring[15] and are subject to averaging artifacts[16]. Aim 2 will develop statistical methods to reconstruct subclonal architectures, impute the order of genetic alterations incurred, and identify genetic subclones based on somatic mutations inferred from Aim 1 from within a probabilistic framework. These methods will be applied to primary, metastatic, pre- and post-treatment CLL samples to assess proportion, frequency, and evolution of subclonal populations and their impact on clinical outcome.

Aim 3: Transcriptomic characterization of genetic subclonal populations. While traditional bulk characterization of genetic clonal populations have revealed immense transcriptional differences putatively linked to particular somatic mutations, single-cell analysis of subclonal populations will allow us to characterize genetically distinct subpopulations within the same environment and genetic background to identify and tease out potentially more subtle or environment-dependent effects. Building on my lab's previous work in characterizing transcriptional heterogeneity in single cells[17,18], Aim 3 will analyze the transcriptional state(s) of distinguishable genetic subclones to identify features associated with clonal growth rate, metastatic transition, and drug resistance. I will work closely with collaborators in the Wu lab to validate findings using in-vitro and in-vivo techniques.

Successful completion of this proposal will yield new insight into subclonal evolution in CLL and provide powerful new open-source computational software for identifying and characterizing subclonal populations that can be tailored and applied to diverse cancer types.