Abstract: Rapidly emerging single-cell DNA sequencing technologies offer promising datasets to further our understanding of diverse facets of cancer biology and genetics. Specifically, it will have a profound impact in resolving the tumor heterogeneity that complicates the diagnosis and treatment of cancer patients and causes relapse and drug resistance. However, novel computational methods are required to perform this task, which is challenged by uncertainties in the underlying evolutionary processes. Moreover, technical artifacts introduced during the sequencing process further complicate this task. Novel computational methods are required for the analysis and interpretation of large-scale single-cell genomic datasets for elucidating tumor heterogeneity and evolution.

In this talk, I will introduce probabilistic models and statistical inference algorithms for elucidating tumor heterogeneity and evolution from single-cell DNA sequencing data. These algorithms probabilistically model the possible mutational histories as well as the sources of uncertainties due to technical artifacts in the data. The mutation discovery algorithm employs a probabilistic model of the technical artifacts and dynamic programming for detecting point mutations from raw single-cell sequencing data. The second method introduces a continuous-time Markov chain to model the underlying mutational events in cancer and infers a tumor phylogeny, a binary leaf-labeled tree that represents the mutational history of a tumor helping guide patient-specific treatment. The final method introduces a tree-structured non-parametric Bayesian clustering framework to reconstruct the cell subpopulations in a tumor and their mutation content. Using these methods, I analyze several cancer datasets and uncover tumor phylogenies, driver mutations and cell subpopulations that are more biologically plausible than previously reported analyses. To close, I will give a brief outlook on a wider range of future directions towards providing novel computational and data-driven approaches aimed at improvement of patient well-being through advancements in the understanding of biological processes and phenomena.

Bio: Hamim Zafar is a graduate researcher in the department of Computer Science at Rice University in Houston, USA. He also has a graduate trainee appointment at the University of Texas MD Anderson Cancer Center. Hamim's research interests lie at the intersection of novel computational innovations and applications in biology and include probabilistic graphical models, machine learning, statistical inference, computational biology and bioinformatics. His dissertation research focuses on developing probabilistic models and statistical inference algorithms for elucidating heterogeneity and evolution in tumors from single-cell DNA sequencing data. Prior to joining Rice in fall 2012, Hamim completed his undergraduate studies in Electronics and Tele-communication Engineering from Jadavpur University, India.