New challenges and opportunities presented by advances in transcriptomics

Transcriptomics is a technology that allows us to view the activities of genes in cells of a biological specimen. Recent advances in this technology have spurred intense research activity in the bioinformatics community, which seeks to identify and solve new analytical challenges associated with the new data types and extract useful, testable biological relationships and insights from the data. In this talk, I will present two on-going projects in my group, related to transcriptomic data analysis.

First, I will present interesting new challenges and opportunities in the context of “spatial single-cell transcriptomics”. Here, a data set comprises a 2D or 3D view of each of thousands of cells in the specimen, with each cell’s view offering the locations of transcripts (physical copies) of hundreds to thousands of genes in that cell. The kinds of biological insights that data of such unprecedented resolution offer are yet to be fully understood, and data-driven approaches to identify significant patterns are expected to provide novel insights. The talk will touch upon methods of spatial analysis applied to spatial single-cell transcriptomics data.

In the second half of the talk, I will discuss the problem of deconvolving transcriptomics profiles of biological specimens into their cell type-specific components. A biological sample is often composed of cells of different types, e.g., tumor and immune cells in a cancer tissue sample. The established and inexpensive technology for transcriptomics (so-called “bulk” technology) produces a sample-level profile – a high dimensional vector of activity values of all genes – that is a weighted combination of cell type-specific profiles, with unknown weights representing the relative abundances of different cell types in the sample. If one additionally has data from single cell transcriptomics that provide reference profiles of each cell type separately, one can hope to infer the cell type abundances by solving a regression problem. This is the “deconvolution” task and various methods have been proposed for it. We are specifically interested in the common scenario where the reference profiles available to us are noisy due to experimental as well as biological reasons. We have developed a Bayesian methodology to simultaneously infer the unknown abundancies and accurate transcriptomic profiles of individual cell types from their mixtures available to us through bulk transcriptomics technology. I will present the model and its comparative assessment on standard benchmarks.