Statistical Analysis of Spatial Expression Pattern for Spatially Resolved Transcriptomic Studies

Xiang Zhou

University of Michigan
E-mail: xzhousph@umich.edu

Abstract: Recent development of various spatially resolved transcriptomic techniques has enabled gene expression profiling on complex tissues with spatial localization information. Identifying genes that display spatial expression pattern in these studies is an important first step towards characterizing the spatial transcriptomic landscape. Detecting spatially expressed genes requires the development of statistical methods that can properly model spatial count data, provide effective type I error control, have sufficient statistical power, and are computationally efficient. Here, we developed such a method, SPARK. SPARK directly models count data generated from various spatial resolved transcriptomic techniques through generalized linear spatial models. With a new efficient penalized quasi-likelihood based algorithm, SPARK is scalable to data sets with tens of thousands of genes measured on tens of thousands of samples. Importantly, SPARK relies on newly developed statistical formulas for hypothesis testing, producing well-calibrated p-values and yielding high statistical power. We illustrate the benefits of SPARK through extensive simulations and in-depth analysis of four published spatially resolved transcriptomic data sets. In the real data applications, SPARK is up to ten times more powerful than existing approaches. The high power of SPARK allows us to identify new genes and pathways in these data that otherwise cannot be revealed by existing approaches.