ABSTRACT

In this dissertation, we focus on the problem of analyzing data from high-throughput sequencing experiments, which provide information at an unprecedented resolution. However, statistical methods developed for such data rarely tackle the data at such high resolutions, and often make approximations that only hold under certain conditions. We propose a model-based approach to dealing with such data, starting from a single sample. By taking into account the inherent spatial structure present in such data, our model can accurately capture important genomic regions.

Building upon the single-sample model, we then turn to other important problems in genomics. Specifically, we develop methods aimed at detecting biologically meaningful differences between multiple samples, as well as clustering of experimental samples based on their sequencing profiles.