ChIP-seq is a powerful technology for detecting the protein binding sites in the whole genome scale. Although the method for single ChIP-seq data analyses (e.g., “peak detection”) has been well developed, rigorous statistical method for comparing multiple ChIP-seq datasets with the considerations of control data, biological variation, and multiple-factor experimental designs are not yet available.

In this work we develop a statistical method to perform quantitative comparison of multiple ChIP-seq datasets and detect genomic regions showing differential binding (DB). We first take an union of the peaks called from all datasets to form candidate DB regions. For each dataset, the count from IP sample at a candidate region is assumed to follow Poisson distribution. The underlying Poisson rate is assumed to be the multiplication of artifact and true biological signal, and the artifact is modeled as a function of the control data. For each individual dataset, we first estimate and remove the artifacts to obtain estimated biological signals. These quantities are then passed into hypothesis testing procedures for differential binding analysis. Simulations and real data analyses demonstrate that the proposed method performs favorably in detecting differential binding compared with existing methods.