SCOPE: a normalization and copy number estimation method for single-cell DNA sequencing

Yuchao Jiang

University of North Carolina at Chapel Hill
E-mail: yuchaoj@email.unc.edu

Abstract: Whole genome single-cell DNA sequencing (scDNA-seq) enables characterization of copy number profiles at the cellular level. This technology circumvents the averaging effects associated with bulk-tissue sequencing and increases resolution while decreasing ambiguity in tracking the evolutionary history of cancer. ScDNA-seq data is, however highly sparse and noisy due to the biases and artifacts that are introduced during the library preparation and sequencing procedure. Here, we propose SCOPE, a normalization and copy number estimation method for scDNA-seq data. The main features of SCOPE include: (i) a Poisson latent factor model for normalization, which borrows information across cells and regions to estimate bias, using negative control cells identified by cell-specific Gini coefficients; (ii) modeling of GC content bias using an expectation-maximization algorithm embedded in the normalization step, which accounts for the aberrant copy number changes that deviate from the null distributions; and (iii) a cross-sample segmentation procedure to identify breakpoints that are shared across cells from the same subclone. We evaluate SCOPE on a diverse set of scDNA-seq data in cancer genomics, using array-based calls of purified bulk samples as gold standards and whole-exome sequencing and single-cell RNA sequencing as orthogonal validations; we find that, compared to existing methods, SCOPE offers more accurate copy number estimates. Further, we demonstrate SCOPE on three recently released scDNA-seq datasets by 10X Genomics: we show that it can reliably recover 1% cancer cell spike-ins from a background of normal cells and that it successfully reconstructs cancer subclonal structure from ~10,000 breast cancer cells.