A050 · Confounder-Aware Feature Correction for Single-Cell Batch Integration

Presenter: Calvin McCarter — BigHat Biosciences

Authors: Calvin McCarter

Session: Day 1 · Thu Oct 1, 2026 · 2:15–4:15 PM

Batch integration is a central preprocessing step in single-cell genomics, where datasets collected across experiments, donors, and protocols must be combined despite pervasive technical batch effects. The leading integration methods produce a shared low-dimensional embedding, which discards the corrected gene-expression values that downstream differential-expression, marker, and reuse analyses depend on. The feature-space (expression-correcting) methods that do preserve genes typically align the marginal expression distributions of batches and thereby risk erasing genuine biological variation whenever cell-type composition differs across batches, i.e. whenever batch effect and biological signal are confounded. We recast single-cell batch integration as confounded domain adaptation and apply ConDo, a method that matches conditional expression distributions given the cell-type annotation rather than marginal distributions. To extend ConDo's pairwise source-to-target adapter to the many-batch setting, we introduce an agglomerative compatibility-graph integrator: batches are nodes connected when they share a cell type, and we greedily merge each best-scoring neighbor into a growing reference by fitting one maximum-mean-discrepancy ConDo adapter conditioned on cell type. On the Open Problems Single-Cell Integration Benchmark, ConDo is the strongest feature-space integrator, ranking first among feature methods on five of six datasets. It even outperforms deep embedding methods on the overall score, ranking first across all methods on four of six datasets while returning corrected expression rather than an opaque embedding.