A075 · scBrieflow: a single-cell analysis platform for understanding morphological readout of optical pooled screens

Presenter: Ege Topkoc — Whitehead Institute

Authors: Ege Topkoc, [Please note

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

Optical pooled screens (OPS) link CRISPR perturbations to imaging phenotypes at single-cell resolution, but outputs are conventionally collapsed to per-gene averages early in analysis, discarding single-cell heterogeneity. We present scBrieflow, a standardized, config-driven pipeline repurposing scverse tools alongside Harmony batch correction for imaging-derived morphology, treating cells-by-features as the morphology analog of AnnData's traditional cells-by-genes structure. Built on brieflow's feature-extraction output, scBrieflow creates an AnnData object per screen and applies flag-don't-delete QC, feature filtering, NTC-referenced normalization, and six embedding methods: PCA, Harmony, and four batch-conditional deep generative models (VAE, contrastiveVI, CPA, SAMS-VAE), outputting a final AnnData object and HTML report. As a validation case study, nuclear DAPI intensity resolves G1/S/G2M cell-cycle phases via non-parametric peak-reflection deconvolution, chosen for generalizing across screens where two-component methods collapse the S-phase plateau. Where available, an independent proliferation marker, like Cyclin B1 or Ki67, correlates with DNA content. Phase structure is preserved after Harmony batch correction across all four screens (55,000–270,000 control cells each). To compare embeddings without circularity, we score every candidate against gene sets defined independently of any embedding, like CORUM complexes, essential-gene panels, and guide-to-gene concordance. By complex recovery across perturbed populations (2.7–11.3M cells per screen, 5,000–20,000+ perturbations), we find no universal best embedding: Harmony-corrected PCA is matched by uncorrected PCA on every screen (within 0.02 AUROC), while deep embeddings outperform Harmony on the largest screen and underperform on the second-largest. scBrieflow provides a reproducible foundation for embedding selection and genome-scale perturbation-phenotype quantification.

scBrieflow: a single-cell analysis platform for understanding morphological readout of optical pooled screens

Ege Topkoc[1, 2, *], Ana Karla Cepeda Diaz[1], Matteo Di Bernardo[1,3], Iain Cheeseman[1,4]

[1]Whitehead Institute for Biomedical Research, Cambridge, MA.; [2]Boston University, Boston, MA.; [3]Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, MA.; [4]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA.

Optical pooled screens (OPS) link CRISPR perturbations to imaging phenotypes at single-cell resolution, but outputs are conventionally collapsed to per-gene averages early in analysis, discarding single-cell heterogeneity. We present scBrieflow, a standardized, config-driven pipeline repurposing scverse tools alongside Harmony batch correction for imaging-derived morphology, treating cells-by-features as the morphology analog of AnnData's traditional cells-by-genes structure. Built on brieflow's feature-extraction output, scBrieflow creates an AnnData object per screen and applies flag-don't-delete QC, feature filtering, NTC-referenced normalization, and six embedding methods: PCA, Harmony, and four batch-conditional deep generative models (VAE, contrastiveVI, CPA, SAMS-VAE), outputting a final AnnData object and HTML report. As a validation case study, nuclear DAPI intensity resolves G1/S/G2M cell-cycle phases via non-parametric peak-reflection deconvolution, chosen for generalizing across screens where two-component methods collapse the S-phase plateau. Where available, an independent proliferation marker, like Cyclin B1 or Ki67, correlates with DNA content. Phase structure is preserved after Harmony batch correction across all four screens (55,000–270,000 control cells each). To compare embeddings without circularity, we score every candidate against gene sets defined independently of any embedding, like CORUM complexes, essential-gene panels, and guide-to-gene concordance. By complex recovery across perturbed populations (2.7–11.3M cells per screen, 5,000–20,000+ perturbations), we find no universal best embedding: Harmony-corrected PCA is matched by uncorrected PCA on every screen (within 0.02 AUROC), while deep embeddings outperform Harmony on the largest screen and underperform on the second-largest. scBrieflow provides a reproducible foundation for embedding selection and genome-scale perturbation-phenotype quantification.