A083 · IGVF Single-cell Perturb-Seq Pipeline, a unified framework for complex data analysis and perturbation inference

Presenter: Lucas Ferreira da Silva — Massachusetts General Hospital · Harvard Medical School

Authors: Lucas Ferreira da Silva, Logan Blaine, Sizhu Jiang, Alejandro Barrera, Olga Pushkareva, Sara Geraghty, Mphathi Nzima, Adam Klie, Alexandra Mo, Ruhi Rai, Gary Yang, Timothy Barry, Thomas Cowart, Eric Che, Ian Whaling, Andreas Gschwind, Benjamin C. Hitz, Siddharth Raghavan, Eugene Katsevich, Charles A. Gersbach, Gary C. Hon, Jesse M. Engreitz, Luca Pinello

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

Introduction: Perturb-seq links programmed genetic perturbations to single-cell molecular phenotypes, enabling direct tests of gene and regulatory-element function across heterogeneous cell states. The field now encompasses targeted and whole-transcriptome measurements, single and combinatorial perturbations, alternative guide-capture strategies, sample multiplexing, and editing modalities. This flexibility expands the biological questions that can be addressed, but it also changes how reads structure, guide metadata, cells, and transcriptional outcomes are encoded and analyzed. Consequently, cross-study comparisons are hindered and reproducibility is restricted by the use of technology-specific scripts. Methods: We developed the IGVF Single-cell CRISPR Pipeline, a Nextflow framework that runs locally, on SLURM clusters, or in the cloud and adapts processing, guide assignment, inference, and reporting to each assay while producing standardized Perturb-MuData outputs. We evaluated the pipeline in two complementary settings. First, a controlled IGVF benchmark used the same WTC11 CRISPRi system and guide library across five sequencing and guide-capture technologies, isolating the pipeline's ability to accommodate technical variation. Second, we reprocessed independent public TAP-seq, base-editing, dual-guide, and enhancer-perturbation studies to test portability across distinct biological questions and experimental designs. Results: Across the controlled IGVF benchmark, the pipeline produced comparable RNA and guide counts, also cell-retention, guide assignment, and perturbation inference while preserving assay-specific differences required for interpretation. Intended target genes were consistently repressed, with 83.2% to 94.1% recovered as negative nominally significant cis effects across tested analyses; mean intended-target effects were negative in every assay, whereas control-guide effects remained centered near zero. Reprocessing the public studies generated the same analysis-ready data structure across technologies and recovered technical summaries and principal biological patterns consistent with the original publications. These complementary evaluations show that the IGVF Single-cell CRISPR Pipeline can standardize heterogeneous experiments without erasing their design-specific

context, supporting reproducible analysis and meaningful comparison across consortia or when re-analyzing public datasets.