A086 · MUTARA: MUTagenesis Analysis of Relative binding Affinity

Presenter: Shogan Sugumar Swamy — Boston Children's Hospital

Authors: Shogan Sugumar Swamy, Danilo Pellin

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

A systemic exploration of the protein-protein and protein-ligand interactions involves a series of tedious and customized code until any actionable insights are generated, some of which are quality control, normalization, statistical inference and visualization, often requiring extreme caution at each of those steps from naming conventions to formatting them. Especially in studies involving Deep Mutational Screening (DMS), which involves studying large numbers of Next Gen Sequencing (NGS) samples simultaneously, it becomes even more complex. We introduce an end-to-end computational platform MUTARA that integrates the complete DMS analysis and streamlines the design-test-iterate cycles. The platform provides: (1) interactive data exploration and replicate-level QC; (2) preprocessing and batch normalization; (3) per-substitution statistical analysis (GLM-based enrichment estimation with handling of sparse counts); (4) comparative testing across experimental conditions; and (5) interactive visualization of mutation landscapes and functional consequences. The platform synergizes JavaScript's interactivity, Python's large-scale data handling, and R's statistical inference. The platform has a Dual-mode architecture, an interactive web interface for exploratory analysis and filtering, and a command-line interface for high-throughput batch processing and pipeline integration, supporting seamless downstream structural or functional validation. Its applications range from directed protein evolution, particularly for antibody epitope mapping, non-genotoxic conditioning strategies, and base editor engineering. The modular architecture supports custom statistical models and integration with structural prediction tools (Boltz-2, AlphaFold). MUTARA addresses a critical computational bottleneck in functional genomics and protein engineering by providing an accessible, reproducible pipeline for DMS data analysis. Open deployment and validation across multiple protein systems will facilitate broader adoption.