

Supplementary data files description: [Supplementary Data 1 - 6](#)

A statistical framework for high-content phenotypic profiling using cellular feature distributions

Yanthe E. Pearson¹, Stephan Kremb¹, Glenn L. Butterfoss¹, Xin Xie¹, Hala Fahs¹, Kristin C. Gunsalus^{1,2*}

¹Center for Genomics and Systems Biology, New York University Abu Dhabi, P. O. Box 129188, Abu Dhabi, United Arab Emirates

²Department of Biology and Center for Genomics and Systems Biology, New York University, New York, NY 10003, USA

* Corresponding author: Kristin C. Gunsalus
Email: kcg1@nyu.edu

File name: Supplementary Data 1

Description: Summary of two-way ANOVA analysis including F-statistic and corresponding p-value for each plate and feature (see Fig. 3b).

File name: Supplementary Data 2

Description: Output summary and metadata of replicate reproducibility analysis of treatment replicates and control-DMSO replicates includes well id, sample size, KS distance (corresponding p-value), EMD distance, Z-score (see Fig. 4c).

File name: Supplementary Data 3

Description: List of raw cell counts (n) corresponding to each measured EMD score calculated for 330 control cell samples, and 455 treatment samples relative to the global control population (see Fig. 5b, c).

File name: Supplementary Data 4

Description: Per-compound metadata from multiple sources, including compound dilutions, plate location, biological targets, MOA and chemical descriptors.

File name: Supplementary Data 5

Description: Summary of feature analysis includes full feature set, irreproducible features, circ features, redundant (by correlation) features, low relative variance features, and the final set of active features.

File name: Supplementary Data 6

Description: Three sets of treatment profiles and accompanying metadata; Raw EMD profiles include the full 174 feature set, profiles of 52 treatments causing strong cell loss (> 70%) and scaled EMD profiles (log transformed and min-max scaled) with the reduced 69 feature set.