GET READY FOR JUMP-CP: ANALYSIS OF A PILOT DATASET REVEALS CRITICAL INSIGHTS FOR PLATFORM DEVELOPMENT

Core Life Analytics

Victor S.C. Wong, Wienand Omta, David Egan

Core Life Analytics B.V., 's Hertogenbosch, The Netherlands

Overview

There is a growing and significant interest in adopting automated high content phenotypic profiling method for target and drug discovery pipelines. The Broad Institute has recently established a JUMP (Joint Undertaking in Morphological Profiling) Cell Painting consortium to generate a large public reference Cell Painting dataset. The aim is to present the 'ground truth' for the study of phenotypic relationships between various biological perturbations that target the same genes in cells. While this represents a huge potential, such large volume of data also presents challenges for scientists to find critical information. We took and analyzed a preliminary subset of the JUMP-CP dataset in our StratoMineR™ analytics platform. We show that our web-based data analytics platform, StratoMineR™, can detect and generate distinct plate maps, select relevant features, and perform dimensionality reduction and unbiased hit picking. We can make several comparisons to examine differences in phenotypic outcomes between cell lines, time points and conditions. We further examined datasets from compound and CRISPR experiments, and interestingly this comparison detected differential phenotypic outcomes between chemical and genetic approaches for identical gene targets in two different cell lines. Clustering analysis of our hits revealed that: 1) substantial numbers of treatments gave diverse phenotypes regardless of their common gene targets, 2) there are a few phenotypic similarities between compound and CRISPR treatments that share common gene targets, and 3) there are significant clusterings of unrelated gene targets that gave similar phenotypic outcomes. Therefore, our analysis from StratoMineR™ provides valuable insights that can break the data analytics barrier for scientists who wish to develop their own Cell Painting platforms and take advantage of the JUMP-CP dataset when it becomes available.

Introduction

The JUMP-CP is a high content screening dataset of microscopic images and image-based phenotypic profiles from two cell lines treated with chemical compounds and genetic perturbations. The cells were fixed and the standard Cell Painting assay protocol with six fluorescent dyes[1] were used to label various components of the cell (**Figure 1**). Diverse morphological features were segmented and extracted from CellProfiler.

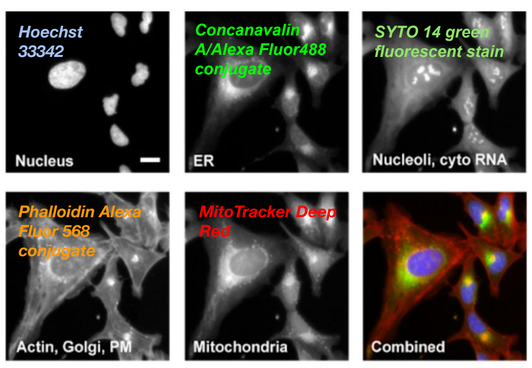

Figure 1: The Cell Painting Assay is a target agnostic phenotypic profiling method that can be used to profile bioactive molecules[1].

Experimental Parameters

Cell Line: A549 & U2OS
Plate Format: 384-well
Plates: 51
Replicates: 2-5
Features: 6400
Gene Targets: 175+
Time Points: 1, 2, 4, 14, 28 days
Treatments: Compounds (306), CRISPR sgRNAs (335), ORFs (175)

Table 1. We used the publicly available preliminary JUMP CP dataset which can be found in the Github repository[2]. Here, we provide a summary breakdown of the overall JUMP-CP experiment.

Method

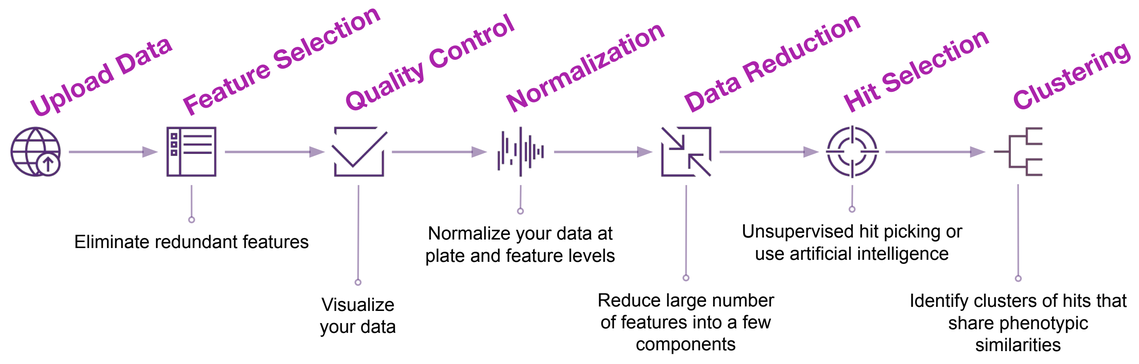

Figure 2. The StratoMineR™ workflow. StratoMineR™ is a web-based platform which guides users through a typical workflow in analysis of high content multi-parametric data[3].

Results

We began our approach to analyze the JUMP-CP data set by performing several critical steps in pre-processing of the data. We first eliminated redundant measurements with binary data types, low or no variation standard deviation, or high Spearman's correlation values (data not shown). Moreover, measurements lacking variation in data distribution were also considered redundant. Since critical annotations (e.g., cell line, reagent classes, compound names, gene targets, time) were not included in the raw JUMP-CP data files, they were added the experiment using the Merge Metadata module in StratoMineR™, which were used to generate plate maps. All data could be labeled with relevant annotations, and separated by cell line using the Visual Data Mining (VDM) interactive data visualization tool in Quality Control (**Figure 3**). Further downstream analysis was performed such as plate normalization, data transformation to handle any skewness in the data, and feature scaling to normalize the range of independent measurements (data not shown).

Data visualization and quality control

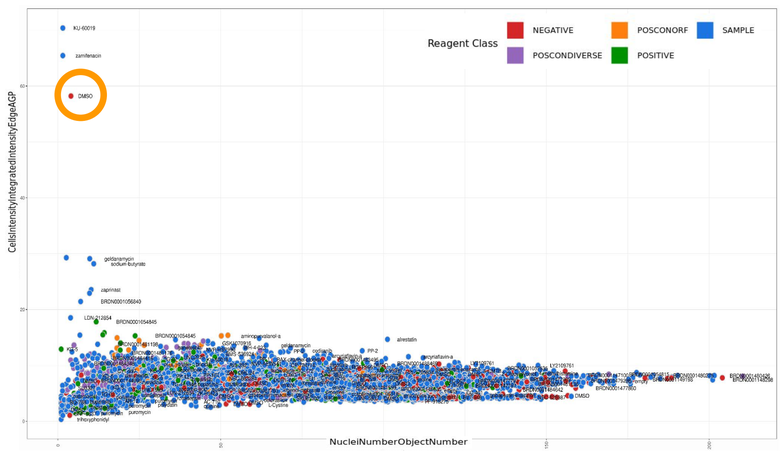

Figure 3: Using the VDM QC interactive data visualization module, we can quickly get an overview of the entire preliminary JUMP-CP dataset. Indeed, we identified a potential outlier for one of the negative control wells (circled). We used the merged metadata module to combine annotation file with the raw data file to label samples, such as compound name or gene target.

Data visualization and quality control

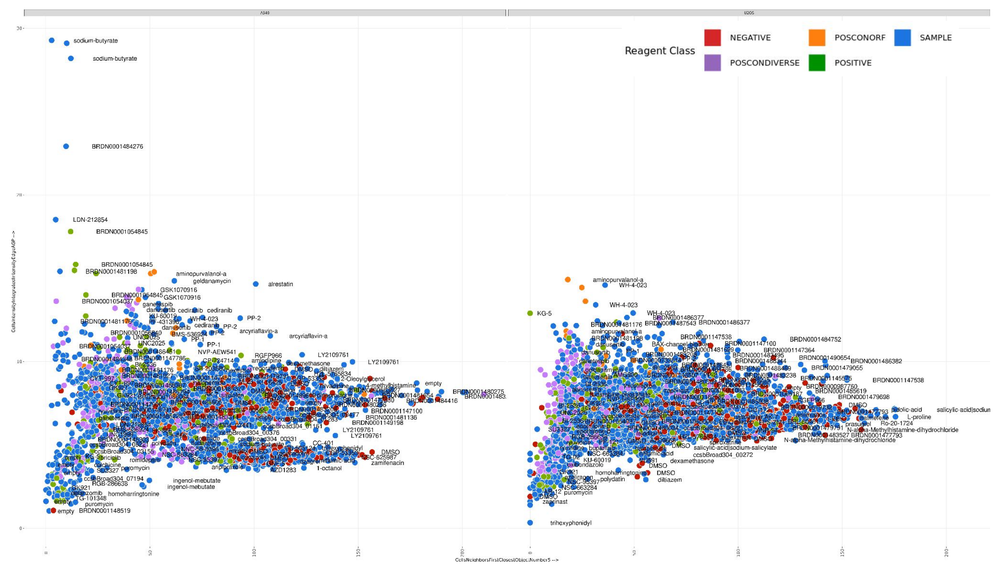

Figure 4. In continuation of the previous figure (Figure 3), as part of this exercise to demonstrate the importance of quality control, there is also a functionality that allows for the tiling of data based on a metadata feature. As an example, we can tile the data based on cell lines, A549 (left), and U2OS (right). Other examples includes time, concentration and molecular target(s).

Dimensionality reduction & Hit Selection

Due to the large number of features in the JUMP-CP dataset, we performed dimensionality reduction to reduce the complexity of the data (**Figure 5A**). This is useful for three critical reasons: 1) reduces computational load, 2) reduces redundancy, and 3) reveals the biology behind the data by highlighting important features.

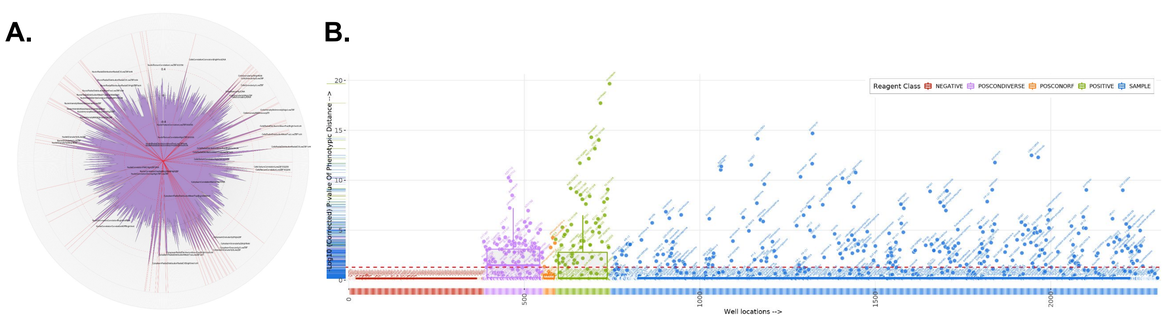

Figure 5. **A.** A spider plot showing one of the principal components (PCAs), features with significant loading in each PCA are labelled. **B.** Unsupervised hit selection using Euclidean distance scoring was used for all wells calculated from the median of the negative controls with $p < 0.05$ (red dotted line) and based on 10 PCAs. This identified 214 hits that were phenotypically far from the negative controls. Shown here is a hit selection graph for A549 cell line (U2OS is not shown).

Clustering analysis

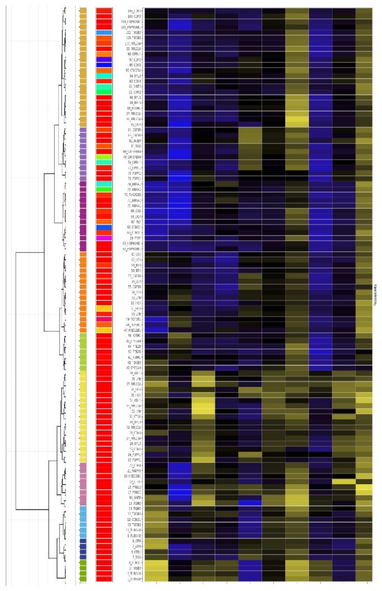

Figure 6. Hierarchical clustering of hits from compound and CRISPR treated A549 and U2OS cell lines revealed groups of gene targets based on 10 PCAs that contributed to the distance score from the negative controls. Substantial numbers of treatments gave diverse phenotypes and this is irrespective of the gene targets. Substantial numbers of treatments gave diverse phenotypes and this is irrespective of the gene targets. Surprisingly, there are few phenotypic similarities between compound and CRISPR treatments that share common annotated gene targets. For example, cycloheximide and TG-02 compounds cluster with the CRISPR guide that targets the *Rpl3* and *Cdk9* genes in A549 cells. Moreover, SB-505124 and CRISPR guide that targets Tgfbr1 gene are clustered in U2OS cells (data not shown). Lastly, there is significant clustering of compounds and CRISPR guides that have unrelated gene targets.

Conclusion

We demonstrate that StratoMineR™ can be used to generate critical knowledge from JUMP-CP data set. This online data analytics tool gives the user the ability to dynamically label and tile data points using external annotations via Merge Metadata. Our analysis shows that substantial numbers of treatments gave diverse phenotypic outcomes. We found that most of the treatments that have unrelated gene targets gave phenotypically similar outcomes. This is not surprising as these treatments of different gene targets can have overlapping effects related to their off-target mechanism of action. We would further suggest that the Artificial Intelligence functionality of StratoMineR™ will be a good tool for building specific phenotypic classes but this requires object-level data, which for the JUMP-CP data set is not yet publicly available.

More JUMP-CP analyses are available at **booth F4**!

Read more!

References

1. *Bray MA et al. Nat Protoc. 2016 Sep; 11(9): 1757-1774.*
2. *https://github.com/jump-cellpainting/2021_Chandrasekaran_submitted*
3. *Omta W et al. Assay Drug Dev Technol. 2016; 14(8): 439-452.*

Your Experiment. Your Data. Your Analysis.

Sint Janssingel 88
5211 DA, 's-Hertogenbosch
The Netherlands