Cell Painting in Pictures and Numbers: An Integrated Image and Data Analytics Platform for Accelerating Drug Discovery

Victor S.C. Wong, Maria Roa Oyaga, Michel van Elswijk, Rik Frijman, Job Mooij, Wienand Omta, David A. Egan

Introduction

Image-based phenotypic profiling is growing in popularity for target and drug discovery processes. Such high content approaches yield rich phenotypic data that can reveal critical holistic insights into mechanisms of action, or toxicity of candidate drugs. Much of the growth has been driven by the use of Cell Painting, a standardized high content profiling method originally developed at the Broad Institute. The JUMP-CP dataset has the potential to be an invaluable resource for drug discovery but the complexity and the size of the data make it challenging for users outside of the consortium. Namely, the three main barriers are: 1) accessing the images, 2) image analysis, and 3) downstream numerical analysis. To help in surmounting these challenges, we developed an integrated cloud-based platform with no-code configuration to allow researchers to quickly translate the vast amounts of data to actionable insights on a drug candidate. Specifically, we built a cloud platform that integrates image storage and registration, image analysis and downstream numeric data analyses. We developed two different tools for image analysis: CPUltra, a distributed version of CellProfiler that leverages massively scalable parallel processing to accelerate analyses and StratoFeatures, a deep learning segmentation-free tool for feature extraction. To validate our method, we used JUMP-CP images from from thirty source 2 and 10 replicate plates (a total of sixty 384-well plates) to run image analysis with CPUltra and StratoFeatures. Numeric outputs were then compared to the original JUMP-CP numeric data using phenotypic Euclidean distance scores of compounds. Outputs from both image analysis tools demonstrated significant positive correlation with JUMP-CP output. Taken together, we show the importance and feasibility of leveraging the cloud computing infrastructure, and how they help make data-based drug discovery accessible to biologists.

Method

The JUMP-CP is a high content (HC) screening dataset of microscopic images and image-based phenotypic profiles from U2OS cell line 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. CellProfiler was used for segmentation and feature extraction. We used the publicly available JUMP-CP dataset which can be found in the Cell Painting AWS repository[2].

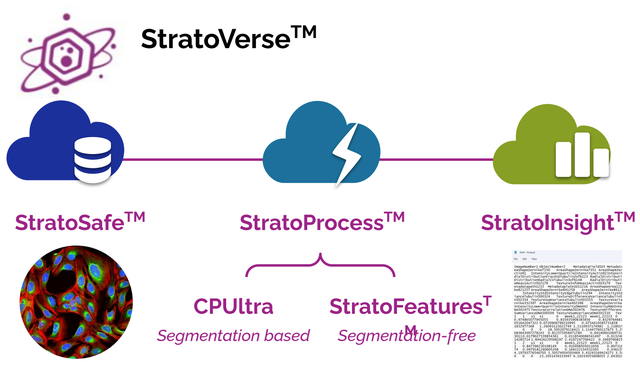
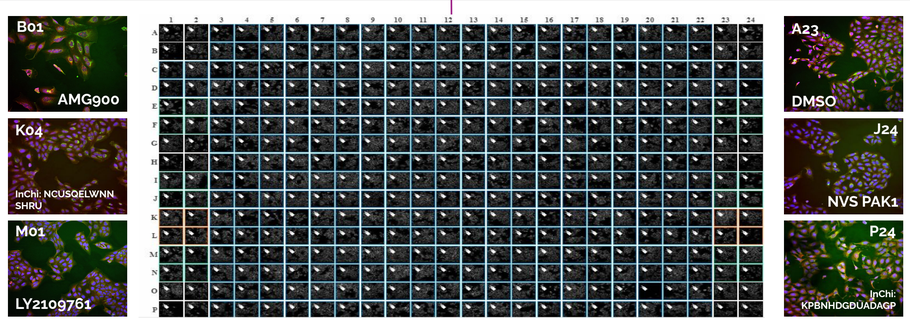


Figure 1. StratoVerse™ is a web-based platform which guides users through a typical workflow in analysis of high content image and numeric data[3]. Starting with image data upload in StratoSafe™, the user can then perform subsequent image (StratoProcess™) and numeric data (StratoInsight™) analyses steps. CPUltra and StratoFeatures™ are currently the two available apps for image analysis.

Results

We began our approach to analyze the JUMP-CP image dataset by accessing the publicly available repository: **https://registry.opendata.aws/cellpainting-gallery/**. A total of 60 384-well replicate plate images from source 2 and 10 were uploaded into our image data management platform, StratoSafe™, and metadata was extracted from image file names (e.g., Plate Name, Well Location, Field, Channel. These images can then be viewed at the plate and well level for quality control and inspection in StratoVieweR™.

	File ID	File Name	File Size	File Extension	File Separator	Plate Name	Plate ID	Well Location	Field	Field ID	Channel ID	Channel	Row	Col
1	1	1_Dest210531-152149...	2002944	tif	_	Dest210531-152149	1	A01	F001	1	25	Z01C01	A	01
2	9	1_Dest210531-152149...	2002944	tif	_	Dest210531-152149	1	A01	F002	2	25	Z01C01	A	01
3	17	1_Dest210531-152149...	2002944	tif	_	Dest210531-152149	1	A01	F003	3	25	Z01C01	A	01
4	25	1_Dest210531-152149...	2002944	tif	_	Dest210531-152149	1	A01	F004	4	25	Z01C01	A	01
5	33	1_Dest210531-152149...	2002944	tif	_	Dest210531-152149	1	A01	F005	5	25	Z01C01	A	01
6	41	1_Dest210531-152149...	2002944	tif	_	Dest210531-152149	1	A01	F006	6	25	Z01C01	A	01



Since image analysis using local-installed version of CellProfiler[3] represents a huge barrier for many biologists due to computational power and time, we developed a cloud-deployable version of CellProfiler, **CPUltra**, in our scalable cloud computing module (StratoProcess™). We obtained the publicly available JUMP-CP CellProfiler pipeline from the github repository and applied this pipeline across all the plates to extract features. Additionally, we analyzed the same images in **StratoFeatures™**, a deep learning approach that extracts features from images using an intermediate layer of a repurposed convolutional neural network with 50 layers, trained on the ImageNet dataset[4]. The neural network input layer is 224 x 224, requiring JUMP-CP images to be down-sampled. The deep learning features represent patterns and details present in images. They identify and emphasize information such as cellular structures, morphological changes, and variations in fluorescence intensity.

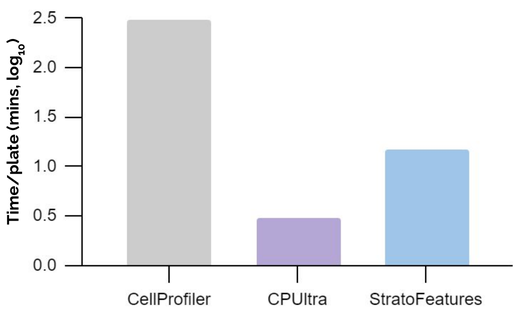
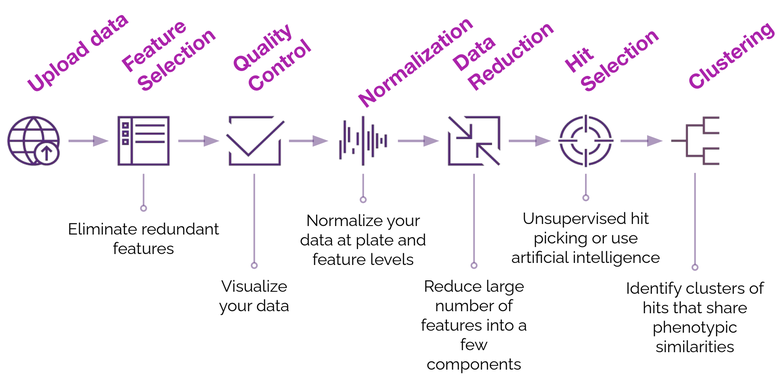


Figure 2. Compute time comparison. CPUltra uses scalable cloud computing, which allows every job/task (input) to be spread across multiple computing nodes in parallel to accelerate image analysis. StratoFeatures™ presented here does not use parallel computing, although this is currently under development. Both CPUltra and StratoFeatures™ show shorter times per plates compared to compared to local-installation of CellProfiler (~ 5 hours/plate[1]).

Numeric output from CPUltra and StratoFeatures™ was then fed into StratoInsight™ (without download/re-upload cycle), wherein StratoMineR™ was used to perform downstream analysis[5]. StratoMineR™ takes the multiparametric numeric dataset through a workflow (pictured below), allowing rapid data mining, visualization and analysis. Moreover, we compared CPUltra and StratoFeatures™ results with the original JUMP-CP dataset from the public repository.



Results (Cont'd)

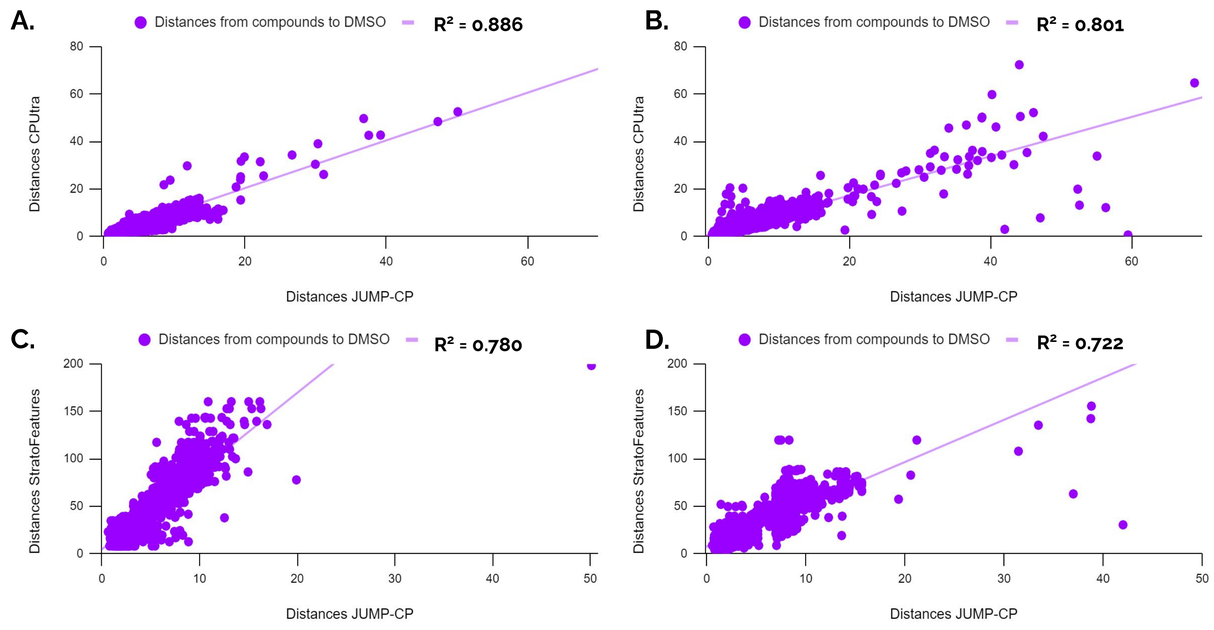


Figure 4. StratoMineR™ for numeric data analysis. CPUltra, StratoFeatures™ and the original JUMP-CP numeric datasets were visualized and compared. We compared Euclidean distance scores for all the compounds between CPUltra vs. JUMP-CP (source 10, **A**; source 2, **B**), and StratoFeatures vs. JUMP-CP (source 10, **C**; source 2, **D**) datasets and saw strong R^2 correlations.

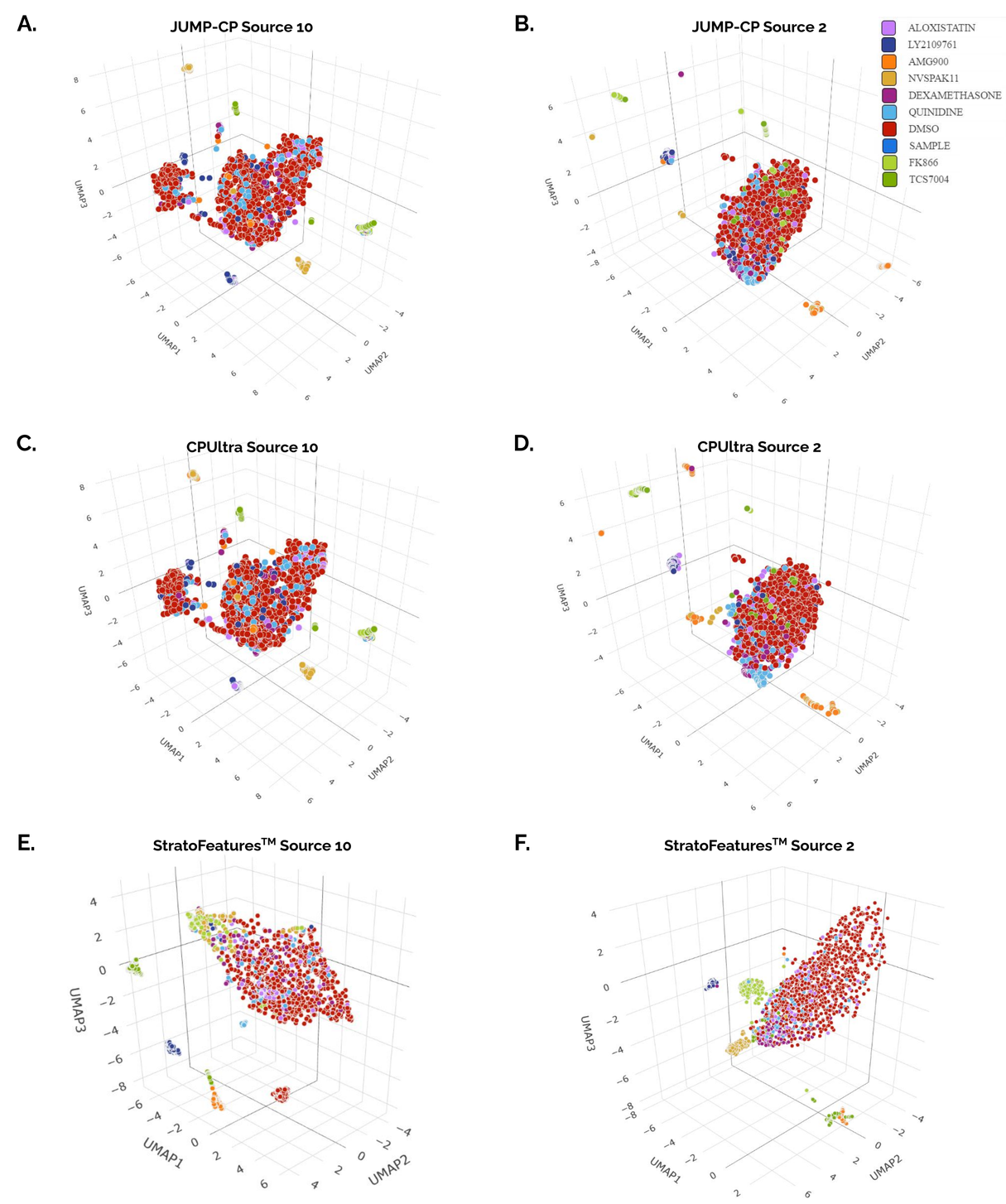


Figure 5. StratoMineR™ for phenotypic hit selection and UMAP clustering. UMAP clustering can be applied to various JUMP-CP outputs from raw data (**A, B**), CPUltra (**C, D**), StratoFeatures™ (**E, F**). Regardless of the approach, all outputs show different cluster groups in UMAP that give distinct phenotypes .

Conclusion

The StratoVerse™ platform demonstrates reproducible results that is comparable to the original JUMP-CP dataset. Moreover, our integrated environment allows rapid analysis of both image and numeric data without the need to program or code, thereby giving scientists the tools to maximize the potential and utility of complex and large phenotypic data.

References

1. *Bray MA et al. Nat Protoc. 2016; 11(9): 1757-1774.*
2. *Chandrasekaran SN et al. bioRxiv https://doi.org/10.1101/2022.01.05.475090.*
3. *Carpenter AE et al. Genome Biol 2006; 7(10): R100.*
4. *Krizhevsky A et al. Advance in Neural Information Processing Systems 2012; 1097-1105.*
5. *Omta W et al. Assay Drug Dev Technol. 2016; 14(8): 439-452.*

Core Life Analytics

Your Experiment. Your Data. Your Analysis.