Harnessing and Validating Deep Learning-based Image Analysis for Cell Painting as an Unbiased Approach to Phenotypic Drug Discovery

Maria Roa Oyaga[1], Bendeguz Zovathi[2], Thomas Ebner[2], Phillip Clarke[2] , Wienand Omta[1], David Egan[1], Philipp Kainz[2], Victor S.C. Wong[1]

1 Core Life Analytics BV, 5211 DA 's-Hertogenbosch, The Netherlands

2 KML Vision GmbH, 8020 Graz, Austria

Introduction

There is a growing interest in adopting high content (HC) image-based approaches for target and drug discovery. Cell Painting is an example of such a HC assay that is quickly gaining traction. CellProfiler is the most widely used image analysis tool for Cell Painting[1], but it requires a significant level of expertise. To address this challenge, we used a robust deep learning-based, computer vision application, IKOSA AI[2], to create an image analysis model for Cell Painting. IKOSA AI is a web-browser driven tool, developed by KML Vision, for the generation of new deep learning models that does not require any coding knowledge. To validate the IKOSA AI model and compare it to the JUMP-CP data[3], we used StratoMineR™ [4]. This software is a web-based application that guides biologist through a best-practices and intuitive data analytics workflow.

Methods

A. IKOSA AI

Image dataset
- Pilot JUMP-CP images
- A549 and U2OS
- 5 fluorescent and 3 brightfield channels

Model training and testing
- n=2780 images, 1080x1080 px
- (2208 training, 572 test)
- Feature Extraction: 1832 features were measured from each object (Cells and Nuclei)

Image Analysis
- Data: Images from 5 compound plates from Pilot JUMP-CP (different experimental conditions)
- Object level data extracted

B. StratoMineR

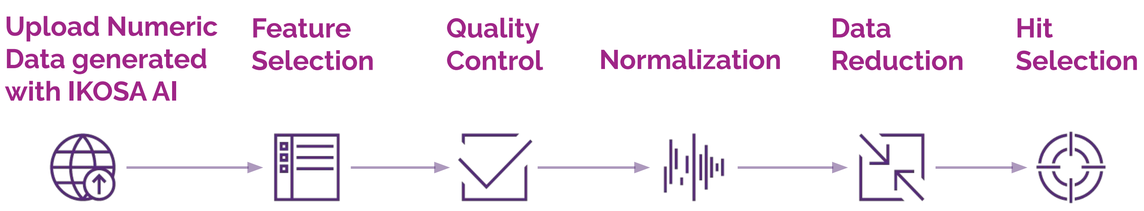

Figure 1: IKOSA AI and StratoMineR™ workflow. **A)** A deep neural network was trained in IKOSA AI with the images described above. The model was then used to analyze 5 compound plates. **B)** The numeric data extracted by IKOSA AI was uploaded to StratoMineR™ for downstream data analytics.

Results: IKOSA AI Model

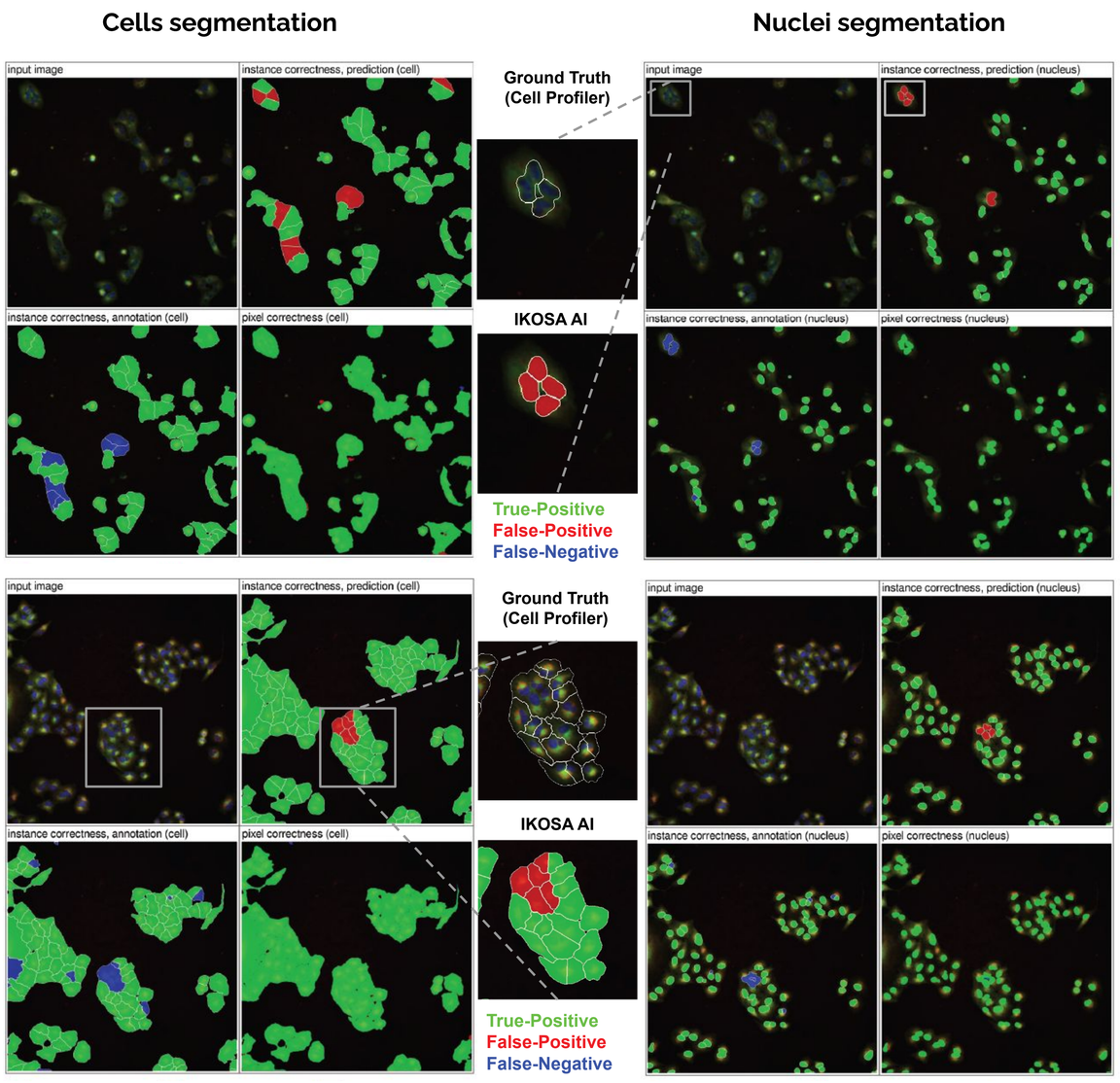

Figure 2: IKOSA AI Segmentation Performance. The segmentation performance was evaluated using the test images shown in **Figure 1A**. This resulted in precision/recall/average precision scores of 0.97/0.89/0.86 and 0.98/0.94/0.92 for cell and nucleus instances respectively. The ground truth labels from CellProfiler were generated with the masks available in the Cell Painting Gallery.

Conclusions

Our results demonstrate that the IKOSA AI model is reproducible and yields consistent outcomes with the publicly available JUMP-CP numeric data. Further investigation is currently underway on how cytoplasm features contribute to the phenotypic diversity of the Cell Painting assay. The combination of IKOSA AI and StratoMineR represent an accessible, unbiased, fully-automated and cloud-based solution that does not require computer programming, delivering comparable results to the current State of the Art. Alternative to CellProfiler manual process, our approach to Cell Painting, from image analysis to downstream data science, greatly accelerates novel biological insights from high content data in a holistic and reproducible fashion.

References

1. Bray MA et al. Nat Protoc. 2016 Sep; 11(9): 1757-1774.
2. KML Vision GmbH, IKOSA (software), Graz, Austria, software available at https://app.ikosa.ai
3. We used the dataset cpg0000-jump-pilot, available from the Cell Painting Gallery on the Registry of Open Data on AWS (https://registry.opendata.aws/cellpainting-gallery/)
4. Omta W et al. Assay Drug Dev Technol. 2016; 14(8): 439-452.

Core Life Analytics

Results: Comparison to Original JUMP-CP Data

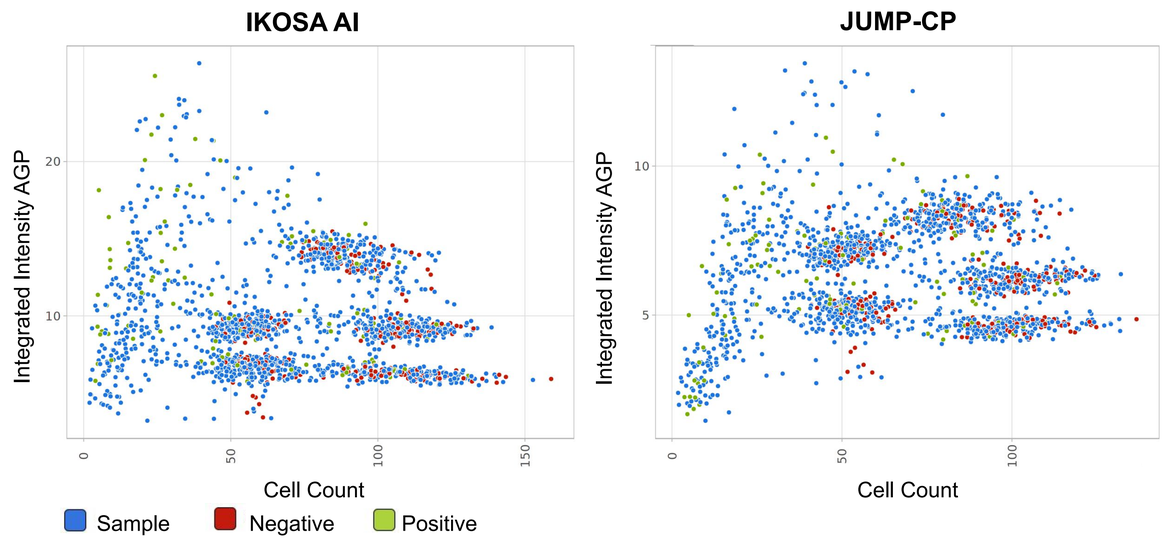

Figure 3: Raw data from IKOSA AI and JUMP-CP. We used the StratoMineR™ Quality Control interactive data visualization module to get an overview of the data (n=5 plates per source). We used the Merge Metadata module to combine an annotation file with the raw data. This supports inclusion of details about the experiment (compound names, reagent classes, etc) which results in more plotting options. Plotted here are scatter plots of two features out of the 2000 features present after Feature Selection.

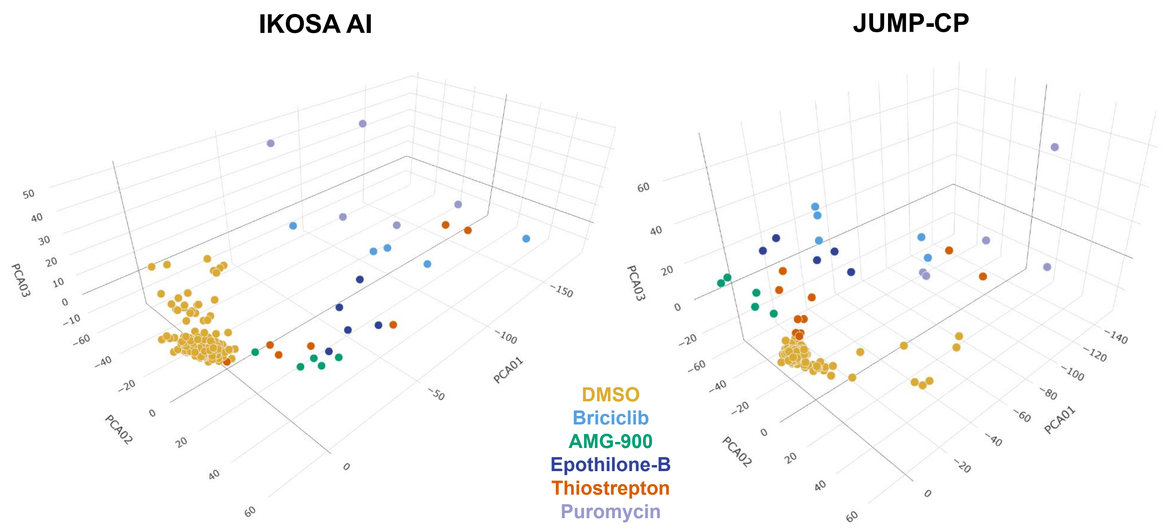

Figure 4: Dimensionality Reduction, Principal Component Analysis (PCA). Due to the large number of features, we performed PCA to reduce the computational load and redundancy, and reveal the biology behind the data. Shown in the graph are the first 3 components for each data source, data points are colored by compound name (n=5 plates per source).

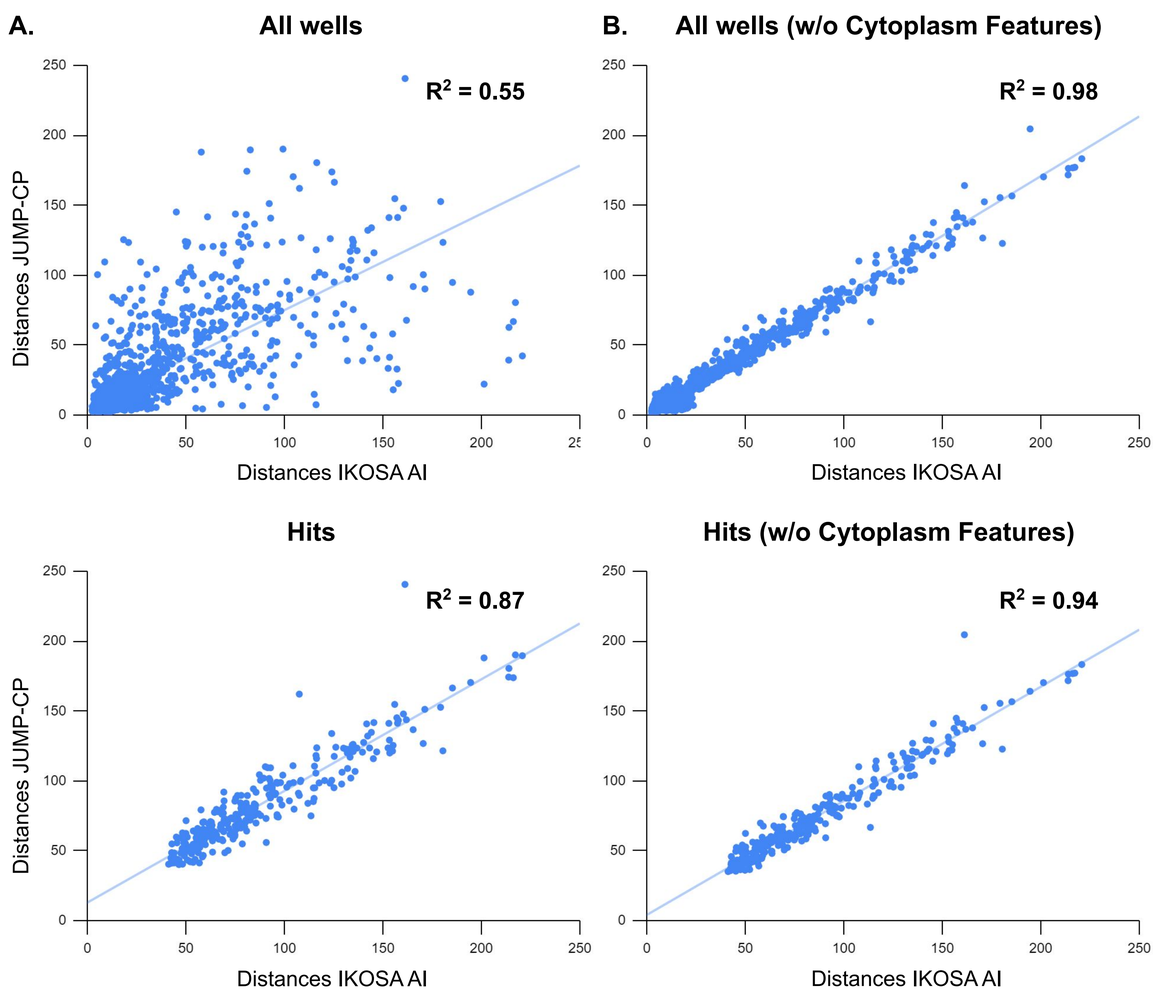

Figure 5: Comparison between IKOSA AI and JUMP-CP. Euclidean distance scores from the negative control median to all the wells (top graphs) and distance scores to hits only, $p<0.05$ (bottom graphs). IKOSA AI extracts measurements from cells and nuclei, while the JUMP-CP data contains features from cells, nuclei and cytoplasm. **A)** JUMP-CP data with all features (including Cytoplasm) and IKOSA AI data. **B)** JUMP-CP data excluding Cytoplasm features and IKOSA AI data.

KMLVISION