Core Life Analytics

Artificial Intelligence for High Content Analysis with HC StratoMineR

Introduction

Methods such as automated microscopy that generate high content data sets are invaluable for phenotypic screening. Their utility is limited however if the user does not have access to the right data mining tools. HC StratoMineR is a powerful, yet intuitive, web-application that allows the biologist to analyze the large numeric data sets that are generated in high content cell screening projects.

HC StratoMineR, (Omta et al., 2016), walks users through a logical analytical workflow, (Fig 1), which they can use to characterize screened reagents according to the cellular phenotypic profile that they induce. The application gives the user access to a choice of well-validated methods for data normalization, data reduction, outlier detection and clustering.

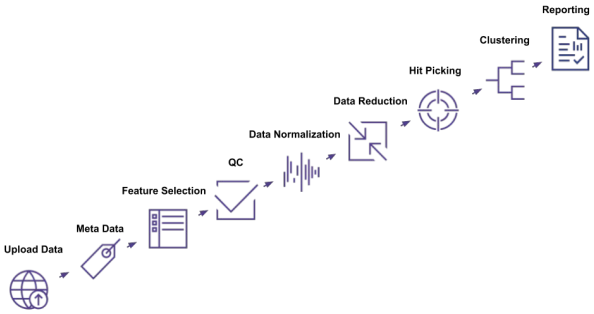

Fig 1. The HC StratoMineR workflow.

High content data sets are ideal for analysis with artificial intelligence methods. A user can build an AI model based on one or more phenotypes, and use this model to identify similar phenotypes in the same data set or in new data. The success of this approach however, is dependent on the quality of the model generated, and by extension the data used to generate the model. Fortunately in HC StratoMineR users have access to an application with which to build and validate high quality AI models.

Unsupervised Analysis

Unsupervised analytical methods are useful for evaluating known phenotypes in a high content data set, or for identifying new phenotypes. Here we demonstrate this with a subset of a genome-wide siRNA screen for regulators of the mitotic cell cycle, (van Heesbeen et al., 2017), that we have previously used to validate HC StratoMineR. In this case 24 384-well plates were used, (12 plates in duplicate), with features extracted at object level.

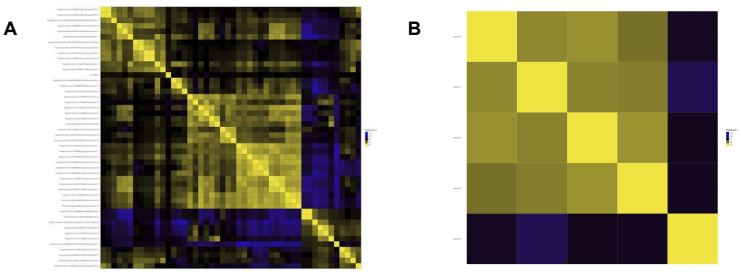

Fig 2. Correlation matrices of extracted features, (A), and common factors generated after data reduction, (B).

An siRNA against the kinesin Kif11 was used as a positive control, another control Hec1 gives a similar phenotype. 39 extracted features, (Fig2A) were subjected to data reduction using common factor analysis, generating 5 factors, (Fig 2B). These were used to calculate distance scores for each well from the median of all the samples, (Fig 3A)

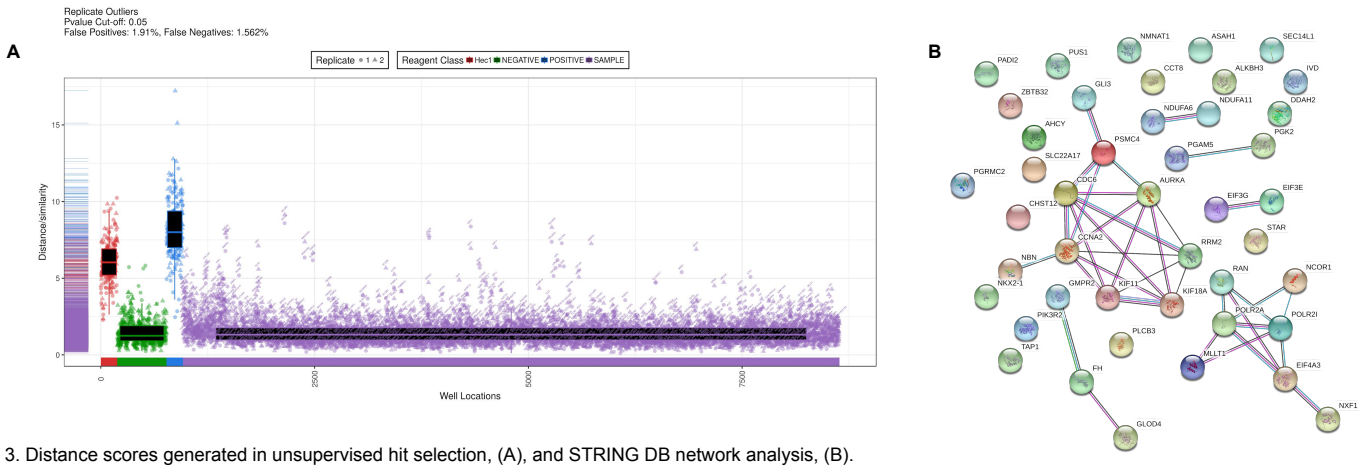

Fig 3. Distance scores generated in unsupervised hit selection, (A), and STRING DB network analysis, (B).

This analysis generated 46 significant hits including Kif11 and another kinesin Kif18a. Network analysis in STRING DB however, (Fig 3B), did not show an enrichment for biological processes associated with cell cycle regulation. This is not surprising as our analysis would be expected to identify hits with diverse phenotypes.

Supervised Analysis with Artificial Intelligence

We also analyzed the same data set using the recently implemented Artificial Intelligence module in HC StratoMineR. Data reduction was not performed but 39 extracted features were used to generate a two classifier Random Forest model based on the negative controls and the Kif11 positive control. Using this model wells that were significantly similar to the Kif11 positive control were identified (Fig 4A).

This generated 40 hits including Kif11, and Kif18a but also Kif20a. Network analysis in this case generated 128 significantly enriched biological processes in STRING DB (Fig 4B).These were rich in those involved in the cell cycle and more specifically the mitotic cell. Of the top 10, 5 were related to the cell cycle.

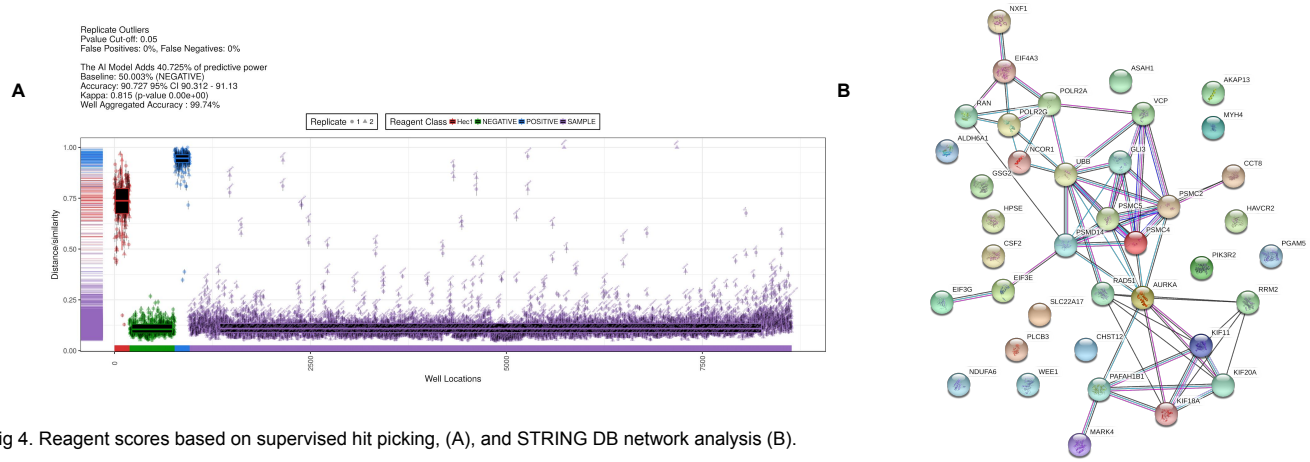

Fig 4. Reagent scores based on supervised hit picking, (A), and STRING DB network analysis (B).

We also generated a three class model which included the Hec1 control. Even though this has a similar phenotype to the Kif11 positive control we were able generate an AI model that could differentiate between Kif11 and Hec1, (Fig 5A).

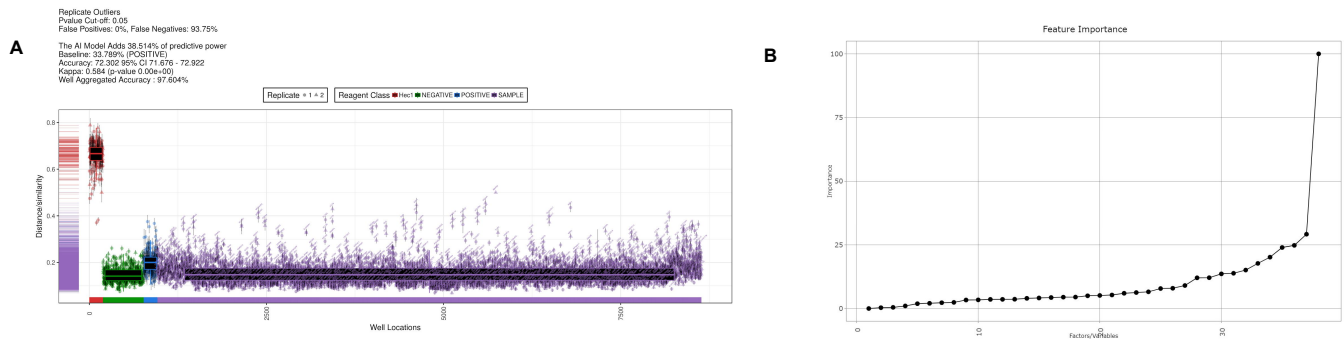

Fig 5, Scores based on 3-class AI model, (A), and a feature importance plot, (B).

Towards a Robust and Efficient Phenotypic Screening Strategy

HC StratoMineR empowers biologists to rapidly develop and validate a phenotypic screening platform. HC StratoMineR users can identify novel phenotypes using unsupervised methods and then if desirable these can be used to build high quality AI models for subsequent use. As shown above the AI methods can be used to rapidly identify phenotypes similar to the known one.

One output from the generation of the AI model is a visualization of the relative importance of the original features in the AI model, (Fig 5B). Image analysis platforms such as Cell Profiler can be used to extracted thousands of features from HC images. This however is computationally intensive and can put a strain on local image analysis infrastructure. With HC StratoMineR however, users can rapidly identify the key features that are sufficient for generation of a robust AI model.

Learn More

You can learn more about HC StratoMineR at www.corelifeanalytics.com where you can register for a free 30-day subscription to the StratoMineR platform. You can also contact us at info@corelifeanalytics.com for an online demonstration and a discussion about your specific needs.

References

Omta, WA, van Heesbeen, RG, Pagliero, RJ, van der Velden, LM, Lelieveld, D, Nellen, M, ... & Spruit, M (2016). HC StratoMineR: A web-based tool for the rapid analysis of high-content datasets. *Assay and drug development technologies*, *14*(8), 439-452.

van Heesbeen RGHP, Raaijmakers JA, Tanenbaum ME, Halim VA, Lelieveld D, Lieftink C, Heck AJR, Egan DA, Medema RH. Aurora A, MCAK, and Kif18b promote Eg5-independent spindle formation. Chromosoma. 2017 Aug;126(4):473-486.