Distinguishing cell types by phenotypic profiling of the nucleus.

Key features

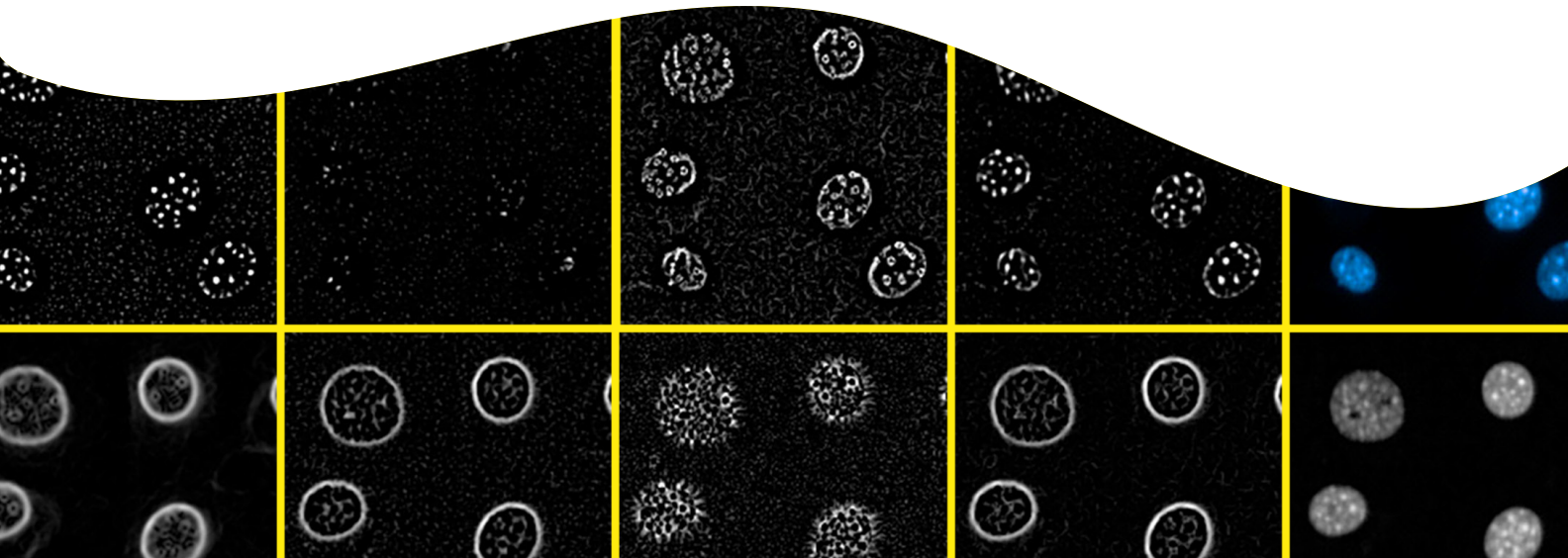
- Distinguish cell types based exclusively on nuclear staining
- Generate detailed phenotypic profiles using texture and advanced morphology parameters
- Leverage easy to use supervised machine learning tools to accurately classify cell types in co-cultures
- See how texture and advanced morphology parameters feed into Principal Component Analyses for separating more complex data sets

Introduction

The promise of high-content screening is the acceleration of discovery by extracting as much relevant information as possible from cells. Nevertheless, a large percentage of high-content screens analyze only a small number of image-based properties.[1] As a result, valuable information from precious cells and disease models is not utilized. As nearly all screening approaches require a nuclear counterstain such as Hoechst to facilitate segmentation, phenotypic profiling of the nuclei can offer new and additional perspectives on assays at no extra cost. Hoechst "total sum intensity" distribution is sometimes used to analyze cell cycle distribution, in particular G0/G1, S and G2/M populations. However, besides cell cycle analysis, more information can be retrieved from the nuclear "counterstain". Using Harmony® high-content imaging and analysis software, this study shows how a single nuclear stain enables phenotypic profiling and how phenotypic profiles can be used to distinguish cell types within co-cultures or even within seven different cell types without any further staining or additional phenotypic markers.

Co-cultured cells can be distinguished based on Hoechst nucleus staining only

To maintain primary cells *in vitro*, they are often co-cultured with other cells which provide pro-survival signals in the form of trophic factors and cell-cell interactions. Typical examples are primary hepatocytes co-cultured with fibroblasts or primary neurons co-cultured with astrocytes. Besides this, co-cultures are also used to study the interaction between cell types, e.g. cancer cells with tumor-derived fibroblasts, or epithelial



cells with lymphocytes. In the typical direct co-culture setup, cell types are mixed within the same well, posing challenges to analyze them separately. To show how nuclear counterstaining, which is normally only used for cell segmentation, can also be used for cell classification, human hepatocytes (HepG2) and mouse fibroblasts (NIH/3T3) were co-cultured and analyzed.

Application

HepG2 liver and NIH/3T3 fibroblast cells were seeded into a PhenoPlate 384-well microplate (Revvity, # 6057300) either alone or as co-culture at different ratios (2:1, 1:1, 1:2). Prior to mixing, HepG2 cells were stained with CellTracker Green CMFDA (Thermofisher, # C2925) and NIH/3T3 cells with CellTracker Red CMTPX (Thermofisher, # C34552) to enable validation of the accuracy of the phenotypic classification. For each individual cell type, 96 wells, and for each co-culture condition, 64 wells, were used. The next day, the cells were fixed, stained with Hoechst 33342 (Thermofisher, # H3570) and single plane images were acquired on an Opera Phenix® high-content screening system using a 20x water immersion objective in confocal mode. A total of 9 fields per well were acquired corresponding to approximately 1100 cells. Figure 1 shows example images of individual cultures and co-cultures.

Image analysis

To classify individual cells based on Hoechst nuclear staining as either HepG2 or NIH/3T3, images were analyzed using Harmony software. As a first step, nuclei were segmented using the *Find Nuclei* building block and basic morphology (i.e. area, roundness, width, length) and intensity (i.e. mean, max, sum) properties were calculated using the *Calculate Morphology Properties* and *Calculate Intensity Properties* building blocks. Border objects were removed using the *Select Population* building block which was followed by another *Select Population* building block to remove mitotic cells based on the previous calculated morphology and intensity properties. Mitotic cells were eliminated from further analysis based on the assumption that these nuclei should have less distinctive texture features than G0/G1, S and G2 nuclei. To calculate detailed phenotypic profiles, SER texture (Spots, Edges and Ridges) and advanced STAR morphology (Symmetry, Threshold compactness, Axial or Radial) parameters were calculated using *Calculate Texture Properties and Calculate Morphology Properties* building blocks. SER texture quantifies the occurrence of eight characteristic intensity patterns such as spots, edges and ridges within the image (see Figure 2 for visualizations). To capture texture structures with different sizes, three independent *Calculate Texture Properties* building blocks were used with different settings for the scale parameter (0, 1 and 2px). STAR morphology parameters are a set of properties that quantify the distribution of either texture features or fluorescence intensities inside a region of interest.

Glossary of terms

Phenotype: The collection of observable traits of an organism, e.g. at the minimal level of a cell, properties such as size, shape or molecular content. Due to interaction and alteration with, or of the environment, these characteristics can change, e.g. cells passing through cell cycle.

Phenotypic marker: A marker that allows the identification of a specific phenotype, e.g. phospho-histone H3 is a marker for mitotic cells.

Phenotypic profiling: Extraction of a large number of quantitative features from microscopy images of cells to identify biologically relevant similarities or differences among samples based on these profiles.[2]

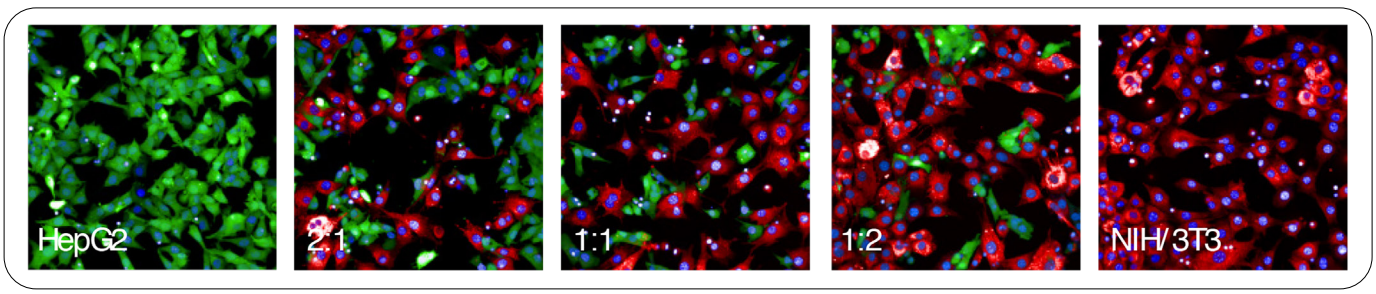


Figure 1. Representative images from wells containing either HepG2 or NIH/3T3 alone or co-cultures of HepG2 and NIH/3T3 cells mixed at ratios 2:1, 1:1 and 1:2 respectively. HepG2 cells are stained with Hoechst 33342 and CellTracker Green CMFDA. NIH/3T3 cells are stained with Hoechst 33342 and CellTracker Red CMTPX. Images were acquired on the Opera Phenix system in confocal mode using a 20x water immersion objective.

They also include profiles in which the distribution of texture features or fluorescence intensity distributions are weighted depending on their localization inside the region of interest. In the case of the nuclear analysis used here, two profiles exist. Profile 1/2 starts at the nuclear membrane and weighting factors decrease towards the nuclear center. This parameter is very sensitive to phenotypic changes within this outer region of the nucleus. Profile 2/2 has the highest weight factor in the nuclear center and decreases towards the nuclear membrane. Therefore, this parameter sensitively captures changes within the inner region of the nucleus. STAR morphology properties also include a sliding parabola filter that can be used to remove smooth and continuous background from the image. Example images showing profiles, SER texture and sliding parabola filtered images of NIH/3T3 nuclei are shown in Figure 2.

Using PhenoLOGIC machine-learning to select the best parameters for distinguishing cell types

A total of 230 parameters were calculated for every nucleus. The PhenoLOGIC machine-learning option in Harmony was then used to select the parameters best suited to discriminate between the two cell types. PhenoLOGIC requires the user to supervise training by simply clicking on about 100 representative objects per class to train the software to distinguish different phenotypes (Figure 3). After training, the software performs a linear discriminant analysis (LDA)[3] to create a linear combination of the most relevant parameters that is then applied to untrained sample wells to classify cells either as HepG2 or NIH/3T3.

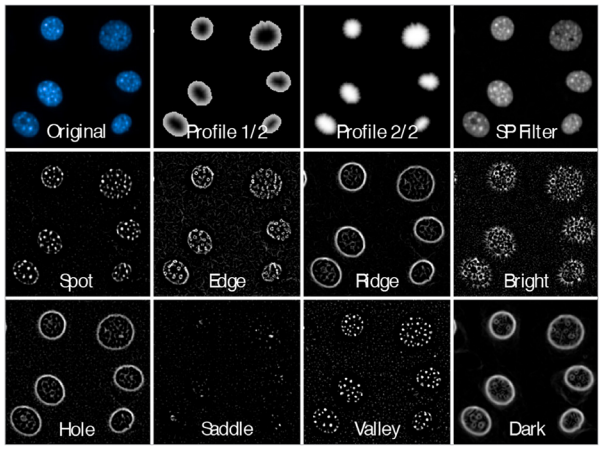
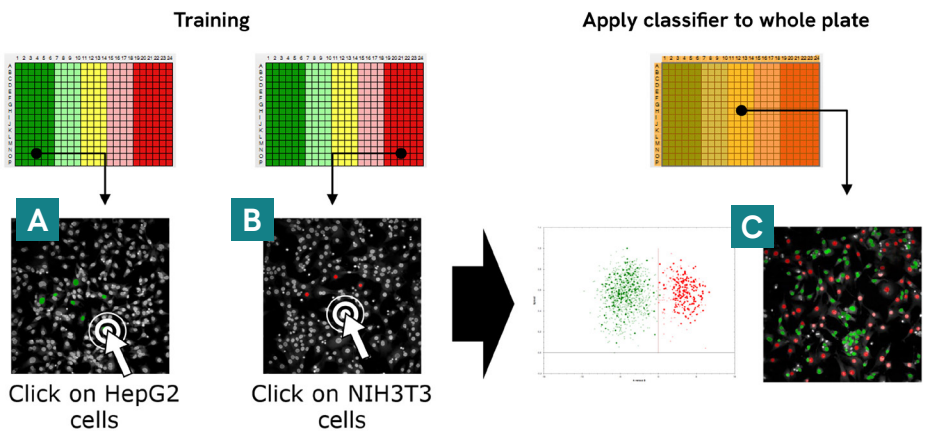


Figure 2. SER texture and STAR morphology properties are key parameters for phenotypic profiling within Harmony software. Original input image, profile images (Profile 1/2 and Profile 2/2), sliding parabola filtered (SP Filter) image and SER texture filtered (Spot, Edge, Ridge, Bright, Hole, Saddle, Valley and Dark) images of the same NIH/3T3 nuclei.



Properties (ordered by relevance)	Linear coefficient
Nucleus Hoechst 33342 SER saddle 0 px	102,042
Nucleus Hoechst 33342 SER bright 2 px	208,404
Nuckus profile1/2 SER-Ridge	-133,82
Nucleus threshold compacthness 30% SER-Ridge	-14,4307
Nucleus axia length ratio SER-Spot	3,74013
Nucleus threshold compactness 50%	-1,41603
Nucleus profile 1/2	-9,9629

Figure 3. Identifying cellular phenotypes using PhenoLOGIC machine learning. In "Training" mode, about 100 single cells within different wells were selected to teach the software to identify the different cell types in "mono cultures" (A and B). Once cells for each class were marked, the resulting classifier was applied to the whole data set. PhenoLOGIC combines the most meaningful parameters, to achieve accurate classification of the two cell types (panel C). In this case, seven properties were chosen to distinguish HepG2 from NIH/3T3 (properties shown in table below the scatter plot). Note how advanced SER (first 2) and STAR (position 3 to 7) properties dominate the selection.

To check the accuracy of the classification, the CellTracker intensity in a perinuclear region was calculated. If a cell was classified as one cell type but the respective CellTracker intensity was below a defined threshold, the cell was counted as a "falsely classified" cell. For this purpose, it was important to calculate CellTracker intensities after the PhenoLOGIC *Select Population* building block to avoid including this information in the classifier.

In mono cultures, 97.5 – 97.8% of the cells are classified correctly. Incorrectly classified cells, i.e. NIH/3T3 in pure HepG2 cultures are identified as false positive based on the CellTracker staining (Figure 4C, first red and light red column on the left). In addition a negligible number of HepG2 cells (0.01%, Figure 4B, first light green column on the left) is identified as false positive in pure HepG2 cultures. These represent cells that did not take up sufficient CellTracker during staining. The same is true for HepG2 cells identified in pure NIH/3T3 cells. All falsely classified cells are identified by CellTracker staining (Figure 4B, last green and light green columns on the right) and a very low and negligible number of NIH/3T3 cells (0.01%, Figure 4C, last light red column on the right) are identified as false positives based on the insufficient CellTracker staining. The percentage of falsely classified cells decreases in the co-cultures. In a 1:1 co-culture, the percentage of false positives drops to 0.34 – 0.65% of all cells (Figure 4). This verification clearly shows that the advanced texture and STAR morphology properties, together with the PhenoLOGIC™ machine learning option, all built-in to Harmony high-content imaging and analysis software, allow the phenotypic differentiation of cell types in co-cultures based on Hoechst nuclear staining alone.

Phenotypic profiling of the nucleus allows distinguishing of seven different cell types

The PhenoLOGIC-based classification of HepG2 and NIH/3T3 cells in co-cultures showed that the features used to distinguish the two cell types were all SER and STAR morphology properties. This prompted an assessment of whether these properties alone would be sufficient to separate even more cell types from one another. Therefore, seven different cell lines, mouse fibroblasts (NIH/3T3), canine kidney epithelial cells (MDCK), human breast adenocarcinoma (MCF7), human lung carcinoma (A549), human hepatocellular carcinoma (HepG2) and human fibrosarcoma (HT1080) were seeded into a PhenoPlate 384-well microplate (three columns = 48 wells per cell type). The following day the cells were fixed and stained with Hoechst 33342 only. This time, single plane images were acquired on the Operetta CLS high-content analysis system using a 20x water immersion objective in confocal mode. A total of nine fields per well were acquired. Example images of the different cell types are shown in Figure 5.

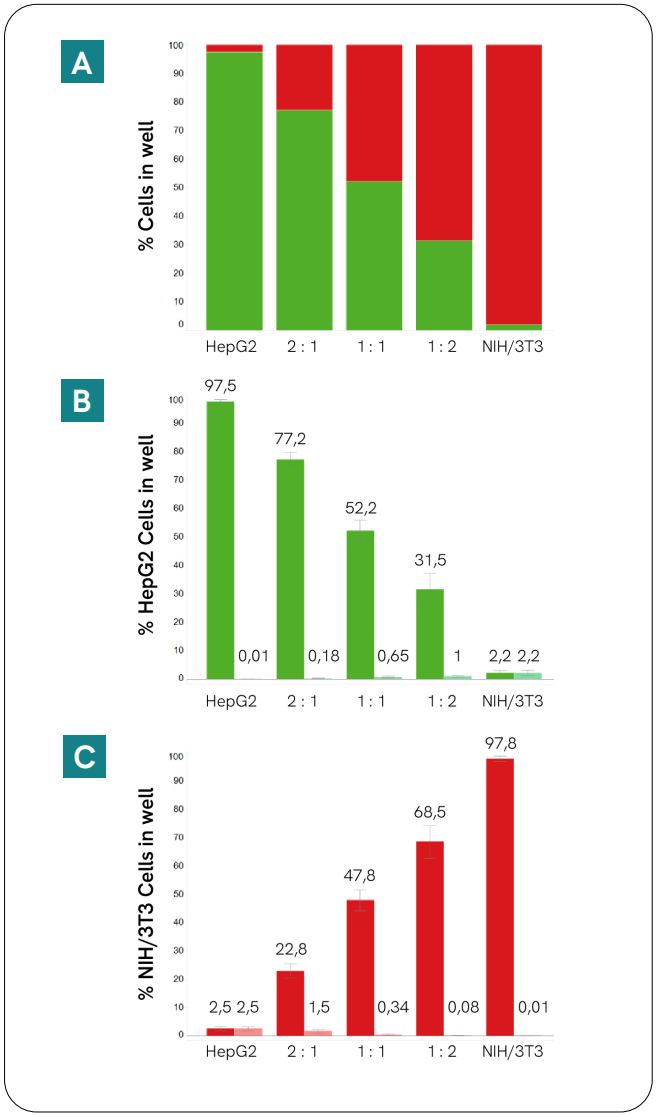


Figure 4. Results of linear classification for HepG2 and NIH/3T3 cells. HepG2 (green bars) and NIH/3T3 (red bars) cells were either cultured individually or as co-cultures at different ratios. (A) The graph shows the percentage of cells that were classified as either HepG2 or NIH/3T3 by PhenoLOGIC. (B) The percentage of cells classified as HepG2 (green) are plotted next to the percentage of cells falsely classified as HepG2 (light green) identified based on the CellTracker staining. The percentage of false positive cells ranges between 0.01 – 2.2%. (C) The percentage of cells classified as NIH/3T3 (red) are plotted next to the percentage of cells falsely classified as NIH/3T3 (light red) identified based on the CellTracker staining. The percentage of false positive cells ranges between 0.01 – 2.5%. HepG2 = only HepG2, 2:1 = 2x HepG2 in co-culture with 1x NIH/3T3 , 1:1 = 1x HepG2 in co-culture with 1x NIH/3T3, 1:2 = 1x HepG2 in co-culture with 2x NIH/3T3, NIH/3T3 = only NIH/3T3, n=96 wells for mono cultures, n=64 wells for co-cultures, error bars represent ± one standard deviation.

Image analysis was performed as for the co-culture experiment. However, this time only advanced SER texture and STAR morphology properties were calculated, not basic features such as intensity and classic morphology. As PhenoLOGIC can only distinguish up to six different classes, the SER and STAR morphology parameters were subjected to unsupervised principle component analysis (PCA) using High Content Profiler™ secondary data analysis software. Principal component analysis is a visualization method especially suited for multiparametric datasets like phenotypic profiles. It reduces the dimensionality of data sets allowing visualization of the similarities or differences among samples. As can be seen in Figure 6A, the seven cell lines form seven different well separated clusters. Each spot in the PCA corresponds to one well. When the plate layout information is used for annotation it becomes visible that each cluster is formed by wells from one cell line only (Figure 6B).

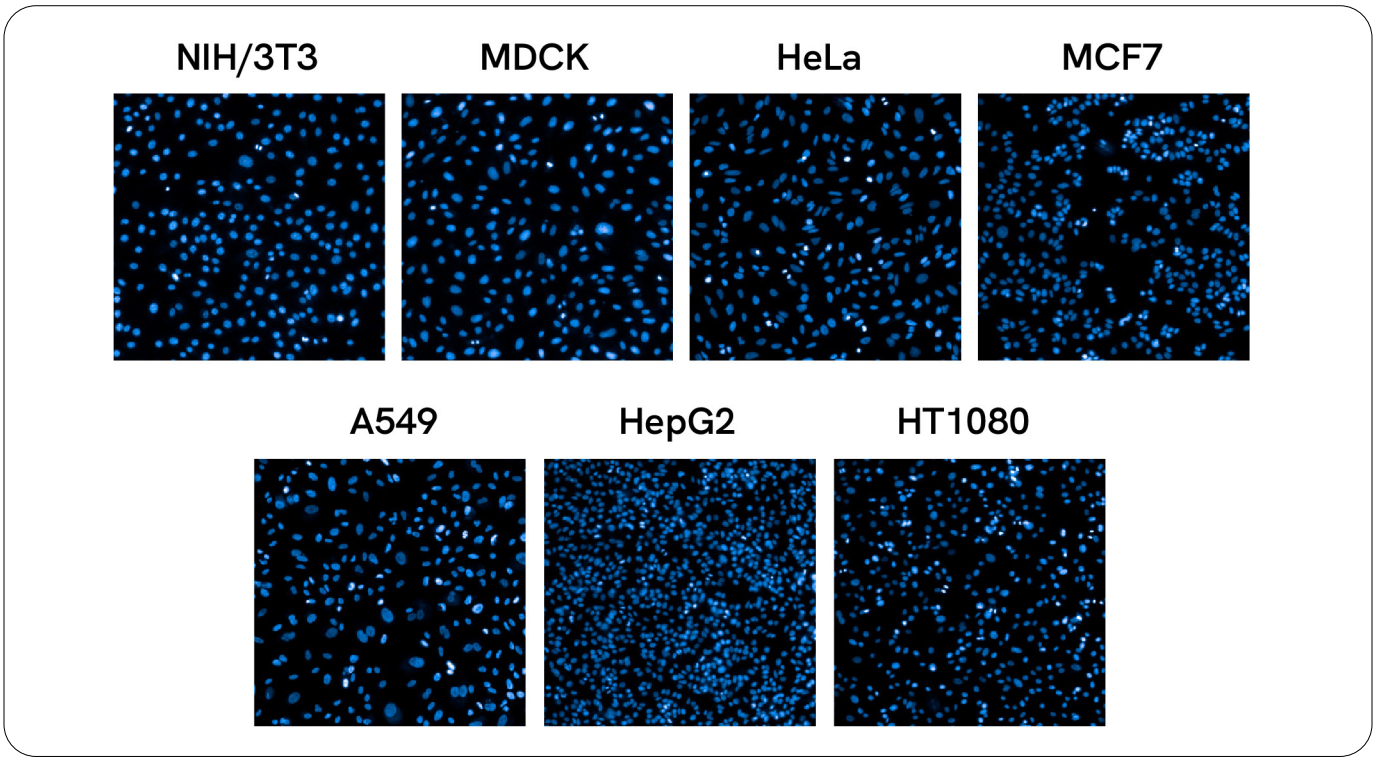


Figure 5. Nuclear phenotypes of seven different cell lines. Cells were stained with Hoechst and imaged on an Operetta CLS high-content analysis system using a 20x water immersion objective.

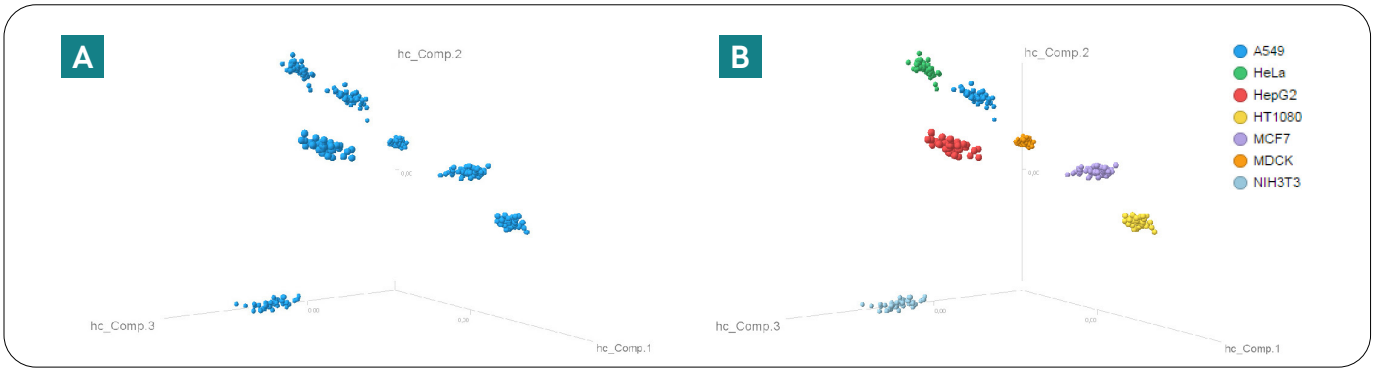


Figure 6. Three-dimensional Principle Component Analysis (PCA) of phenotypic profiles from seven different cell lines. The set of SER texture and STAR morphology parameters calculated in Harmony was subjected to PCA within High Content Profiler software. As can be seen in (A), seven distinct clusters are formed. If the plate layout information is used for annotation, it becomes visible that these clusters represent the seven different cell lines (B). Each spot represents one well. This clearly shows that SER texture and STAR morphology properties of the Hoechst nuclear staining are sufficient to distinguish seven different cell lines from one another.

Conclusions

The Hoechst staining of cell nuclei contains a plethora of information that can be used for much more than just aiding in segmentation during image analysis. As we have shown here, phenotypic profiling of the nucleus enables distinguishing of cells in co-cultures. Even up to seven individual cell lines can be separated by leveraging the Hoechst nucleus staining. This type of phenotypic analysis can be directly applied to other cell types such as primary cells co-cultured with feeder cells and furthermore, phenotypic profiling is not limited to the nucleus. Applying it to other fluorescent labels or even cells labeled by the more broad cell painting approach[2], opens up new horizons for unbiased drug discovery and disease research. The prerequisites for this type of phenotypic analysis are high-quality images, software for image segmentation and generation of phenotypic profiles and a solution to help with processing complex multiparametric datasets (reduction of dimensionality, hit selection). Revvity offers a complete solution for phenotypic profiling applications. Imaging on either the Opera Phenix or Operetta CLS High Content Screening systems allows users to generate the high quality images required. Harmony software enables primary image analysis with accurate image segmentation and advanced morphology and texture quantification methods to generate highly descriptive phenotypic profiles. With PhenoLOGIC, Harmony software also provides an easy to use machine learning-based classifier that helps with dimensionality reduction. Further secondary analysis tools for data exploration and analysis are available in High Content Profiler. Revvity's suite of products for phenotypic profiling enables you to leverage the real "content" of your high-content screening applications.

References

1. Shantanu Singh, Anne E. Carpenter and Auguste Genovesio Increasing the Content of High-Content Screening J Biomol Screen. 2014 Jun; 19(5): 640–650.

2. Bray, M.-A., Singh, S., Han, H., Davis, C. T., Borgeson, B., Hartland, C., Kost-Alimova, M., Gustafsdottir S. M., Gibson, C. C, Carpenter, A. E. (2016). Cell Painting, a high-content imagebased assay for morphological profiling using multiplexed fluorescent dyes. Nature Protocols, 11(9), 1757–1774. doi:10.1038/nprot.2016.105.

3. Fisher, R. A. (1936) The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7: 179-188 DOI:10.1111/j.1469-1809.1936.tb02137.x.

Authors

Alexander Schreiner
Mandy Malle
Karin Böttcher

Revvity, Inc.
Hamburg, Germany