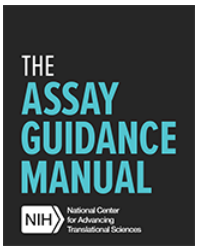
U.S. National Library of Medicine
National Center for Biotechnology Information

NLM Citation: Buchser W, Collins M, Garyantes T, et al. Assay Development Guidelines for Image-Based High Content Screening, High Content Analysis and High Content Imaging. 2012 Oct 1 [Updated 2014 Sep 22]. In: Markossian S, Grossman A, Arkin M, et al., editors. Assay Guidance Manual [Internet]. Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences; 2004-.
Bookshelf URL: https://www.ncbi.nlm.nih.gov/books/



Assay Development Guidelines for Image-Based High Content Screening, High Content Analysis and High Content Imaging

William Buchser, Ph.D.,[1] Mark Collins, Ph.D.,[2] Tina Garyantes, Ph.D.,[3] Rajarshi Guha, Ph.D.,[4] Steven Haney, Ph.D.,[5] Vance Lemmon, Ph.D.,[6,*] Zhuyin Li, Ph.D.,[7,†] and O. Joseph Trask, B.S.[8,‡]

Created: October 1, 2012; Updated: September 22, 2014.

Abstract

Automated microscope based High Content Screening (HCS, or HCA, HCI) has gained significant momentum recently due to its ability to study many features simultaneously in complex biology systems. HCS can be used all along the preclinical drug discovery pipeline, it has the power to identify and validate new drug targets or new lead compounds, to predict *in vivo* toxicity, and to suggest pathways or molecular targets of orphan compounds. HCS also has the potential to be used to support clinical trials, such as companion diagnostics. In this chapter, state of the art HCS approaches are detailed, and challenges specific to HCS are discussed. It should serve as an introduction for new HCS practitioners. More chapters will follow on specific assay examples and on high level informatics analysis.

1. Introduction

1.1. What is High Content Screening (HCS)?

High Content Screening (HCS) or automated microscope-based screening measures biological activity in single cells or whole organisms following treatment with thousands of agents, such as compounds or siRNAs, in multi-well plates. Typically, multiple features of the cell or organism are measured with one or more fluorescent dyes leading to the term High Content. At times, HCS has been called high content analysis (HCA) high content

Author Affiliations: 1 Washington University in St. Louis School of Medicine, Department of Genetics; Email: wbuchser@genetics.wustl.edu. 2 PurpleBio Consulting; Email: Collinsusa2010@gmail.com. 3 MaxSAR Biopharma; Email: garyante@optonline.net. 4 National Center for Advancing Translational Sciences, National Institutes of Health (NCATS/NIH); Email: Rajarshi.guha@nih.gov. 5 Eli Lilly and Company; Email: Shaney314@gmail.com. 6 Miami Project to Cure Paralysis, University of Miami; Email: vlemmon@med.miami.edu. 7 Bristol-Myers Squibb; Email: Zhuyin.Li@bms.com. 8 The Hamner Institutes for Health Sciences; Email: jtrask@thehamner.org.

Corresponding author.

* Editor

† Editor

‡ Editor

imaging (HCI) or image cytometry (IC). Generally, HCA, HCI and IC refer to lower throughput automated microscope based assays (<100,000 samples or data points), although HCA sometimes refers to the analysis portion of HCS. The term HCS was first used in a 1997 paper by Giuliano et al (1). It appears to be the natural successor to the automation of clinical histology (2) and an extension of HTS plate reader systems, with early references to the automation of the analysis of microscope images dating back to the advent of the desk top computing (3). Examples of early systems were the Oncor Videometric150, the BDS chromosome painting and the Meridian ACIS Ca^{2+} image tracker, among others. Our modern automated analysis solutions owe their origins to Metamorph and ImagePro as well as a 1969 publication by Rosenfeld (4).

In contrast to traditional HTS, which has a single read out of activity, HCS allows a scientist to measure many properties or features of individual cells or organisms at once. The ability to study many features and multiplex simultaneously is both what gives HCS tremendous power and challenging complexity. Like its predecessor technologies such as standalone low throughput automated image analysis systems and screening instruments including the FMAT (5), HCS can be quite effectively used simply to provide improved signal to background or signal to noise. But it can also enable both targeted and phenotypic assays that measure movement within a cell or between cells or allow analysis of specific sub-populations of cells in a heterogeneous mix that would be difficult or impossible to run with other techniques. Most powerfully, HCS can be used to help predict the efficacy of potential drugs in unique cellular niches when applied to physiologically relevant cellular systems. The predictive power of such systems can often be enhanced by working with primary cells or differentiated stem cells and in 3-D culture rather than with cell lines in a traditional 2-D culture where many unique aspects of cellular physiology have been lost (6,7). Bickmore focused on this critical problem when she noted "the exquisite cell-type specificity of regulatory elements revealed by the ENCODE studies emphasizes the importance of having appropriate biological material on which to test hypotheses" in the 2012 publication of the data from the ENCODE project (8).

1.2. Uses of HCS

The most obvious applications of HCS are primary screens of potential leads, molecules that can be further optimized into drug candidates, for cellular activities that cannot be easily measured by a single endpoint, such as spatially localized proteins or measurements of cellular morphology. In almost all instances, there are alternative assay formats that can be used for primary screening, but HCS increases the power of the experiment by measuring. For example, HCS can be used to measure the formation of gap junctions by mixing cells that have been preloaded with fluorescent dye with non-loaded cells, and measuring the spread of the fluorescent dye between cells over time in culture (9). This is a very robust way to measure the establishment of functional gap junctions; however, alternatively, a binding assay could be established that looked for an increase in the expression of the gap junction proteins on the cell surface as a primary screen. That kind of HTS approach could not, however, assess functional gap junctions, the way HCS can. Likewise, cellular differentiation can be monitored by measuring an increase in the expression of a marker of the differentiated cell type and/or a decrease in the expression of a marker of the undifferentiated cell type. However, coupling these measurements with a visual measure of differentiation can increase confidence in the outcome.

Cellular morphology changes, such as neurite outgrowth can only be measured in a microscopic image, with or without a molecular marker to confirm the relevance of observed morphology changes. Cell morphology can also be an important measure of cellular differentiation (10) such as the differentiation of epithelial cells to mesenchymal cells (11) or of precursor cells into oligodendrocytes, astrocytes and neurons (12). Here too, the concomitant use of a differentiation marker can help confirm initial analyses and clarify the lineage. Intracellular morphological changes such as protein expression, trafficking or translocation can also be studied. An example of protein translocation, NF-kB translocation will be discussed in a separate chapter. Movement between the cytoplasm and the nucleus or other organelles can be measured or trafficking between organelles and the cell surface can be studied. Likewise, one can measure internalization of proteins or vesicles. Similarly, movement of

cells across a well, such as models of wound healing, migration, or chemotaxis, can all be evaluated with HCS (13, 14).

In addition, there are many useful lead identification applications that could be done with more traditional HTS techniques but where HCS provides an advantage. Because HCS provides more than just the endpoint, off-target effects such as cytotoxicity or fluorescence from the test compounds are easily identified. Additionally, signal to background can often be improved; for example, cell surface binding can be seen in the presence of background noise by using confocal imaging and co-localizing the signal with a cellular surface marker. Alternatively, there are assays where the statistical power of a single read out is not sufficient to enable screening in singlicate but is enabled by combining multiple features to improve the predictive power of the assay.

1.3. Applications of HCS beyond lead identification

The utility of HCS goes well beyond the identification of lead compounds for the development of pharmaceuticals. HCS can be used all along the drug discovery pipeline for the identification or validation of appropriate drug targets, for predicting the pathway or molecular target of compounds identified in phenotypic screens, for lead optimization and toxicity prediction or for the analysis of clinical data. Outside of biological assays, HCS systems have also found utility in chemistry and material science, for instance for screening of crystallization conditions, corrosion resistance or ceramic formation/structure.

Target identification and validation is a very common application of HCS. Potential drug targets can be identified by assaying the effect of increasing or knocking-down RNA expression in an assay that could later be used for compound screening. The cell used in the assay must be amenable to transfection and the cell culture conditions will often need to be modified to allow for translation, transcription and expression of the desired protein to be over expressed or alternatively for the degradation of the existing mRNA of a protein to be knock-down. Alternatively, new targets and pathways can be identified by screening compounds with known mechanisms of action against a phenotype. Not surprisingly, the potential targets identified by these two approaches are often complementary rather than confirmatory since there is a bias in the timing of the effects (15, 16). Overexpression of proteins can occur very quickly, within minutes of transfections. In contrast, RNAi mediated knockdown can take 48 hours or longer to reduce protein levels. These same techniques can also be used for validating suspected targets although clinical or human genetic data is needed for full validation.

HCS can be used for target prediction or pathway profiling. Pathway profiling is an approach used to identify the pathway or target of an orphan compound coming from a phenotypic screen where assays known to be sensitive to the regulation of a pathway or target are used to suggest the mechanism of action, MOA, of an orphan compound. For example, Lonza-Odyssey Thera (17) and the Broad Institute's Metabolite profiling platform (http://www.broadinstitute.org/scientific-community/science/platforms/metabolite-profiling-platform/metabolite-profiling-platform) both have existing platforms for assigning MOA to compounds with unknown MOA.

HCS can also be used to improve our understanding of the on and off target effects of compounds identified by traditional HTS as they progress through lead optimization. For instance, a cellular enzyme activity assay can be used to confirm that a compound known to affect enzyme activity in a biochemical assay retains activity in the cellular environment. Alternatively, compounds identified by traditional screens can be clustered by cellular phenotype in an imaged based screen to suggest both selectivity and toxicity issues. More directly, *in vitro* micronuclei formation assays (18) use HCS to predict which compounds are likely to lead to DNA damage *in vivo* and developmental toxicity assays run on whole embryos can warn of potential reproductive toxicity liabilities that will need to be further evaluated (19).

Even further down the pharmaceutical pipeline, HCS is used to evaluate both *ex vivo* and clinical samples for relevant activity. For instance, both preclinical and clinical blood samples can be assays for biomarker activity of particular cell types by HCS (20).

1.4. HCS Challenges

Even with the tremendous processing capacity of modern computers, high speed inter-connections, and relatively cheap data storage available with today's computers, the application of HCS has significant technical challenges that need to be considered. Although many instruments come with powerful, easy to use image analysis software, HCS practitioners often find that third party analysis software is needed to improve the performance of an occasional assay. In these cases, fluid partnership between the assay developer and an image analysis expert can greatly reduce assay development time by balancing the effort put into optimizing each. At the same time, due to lack of image and data format standards, many solutions for data storage, data transfer, and data annotation, do not translate well between platforms. In addition, the total size of data collecting and manipulated in HCS can be daunting. Storage needs for a single academic lab are typically a few terabytes per year. The costs for on-line storage with RAID and off-line backup are not trivial (between $20K - $100K in 2012) and decisions will eventually have to be made about what data to keep and what to discard. Anticipating data and image transfer and sharing needs during the installation of your HCS solution will ultimately make your solution much more satisfying.

The other area that needs careful contemplation before embarking on an HCS campaign is the consistency and reproducibility of your cellular model. It is common for variables that would not be noticeable in a traditional HTS to become a major source of variance. Consider the effects of mechanical forces or thermal fluctuation on cellular stress responses (21, 22). Differentiation assays are acutely sensitive to changes in proliferation rates that vary with donor age (23, 24). Likewise, proliferation assays are sensitive to differentiation. Gene translation rates may be affected by what phase of the cell cycle cells begin in. The power of HCS to measure basic biological phenomenon, means that the practitioner also needs to be acutely aware of controlling the assay conditions and of interpreting results, with the latter being one of the most important aspect of HCS.

1.5. Summary

HCS has the power to identify new drug targets or new lead compounds, help predict cellular *in vivo* toxicity, suggest molecular targets of orphan compounds, and assess *in vivo* activity among other yet to be explored uses. HCS can be used to measure the effects of compounds and biological molecules such as plasmids carrying cDNAs, or RNAis on subpopulations of cells and specific cell types in mixed cultures. It can be used to measure movement, be it intracellular, cellular or intercellular. Any phenomenon that can be seen reproducibly in a microscope can ultimately be assayed with HCS. Important considerations and state of the art approaches will be detailed in rest of the chapter.

2. Image Technologies and Instruments

2.1. Introduction to Image Technologies

In theory, any instruments that produce multiparametric analysis of cellular or organismal phenotypes can produce high content data. Fluorescent Activated Cell Sorters (FACS) with multiple lasers can do this, as can MALDI-TOF mass spectrophotometers. But the term is most conventionally applied to automated microscopes using fluorescent and transmitted light to image cells, tissues, or small organisms such as Zebrafish embryos, C. Elegans or Drosophila larva followed by sophisticated image and data analysis. There are many commercial vendors of instruments in the High Content Analysis (HCA) arena and it is impossible to describe them all as the market and models are constant changing. Nonetheless, it is possible to divide the most widely used

instruments into just three categories: wide-field imagers, confocal imagers and laser scanning cytometers. Basic descriptions of features that define each category will be provided along with representative examples.

2.1.1. High Content Screening requires speed

Effective screening campaigns required that the instrumentation have as high a throughput as possible. In the high content arena many choices have to be made to enable fast data acquisition. Lower magnification allows larger image fields that can capture larger numbers of cells which is needed to obtain statistically meaningful results. Similarly, cameras with larger chips (2048×2048 pixel vs 1040×1400 pixel) can image larger fields. Use of high intensity light sources such as lasers or optimizing fluorescence staining or fluorescent reporter expression so the fluorescent signals are bright will reduce integration exposure time and therefore the image acquisition time, as does using an objective with relatively high numerical aperture. If essential phenotypic features can be recognized with two channels as opposed to three, four or more channels, large savings in screening times can be achieved. During the assay development phase selecting the appropriate features that discriminate positive and negative controls can dramatically speed the final screen design to achieve the highest throughput possible. It is important to keep in mind that the goal is not necessarily to produce attractive images of individual cells, of the sort acquired with high magnification on a confocal microscope. Rather the goal is to detect and quantify critical features from the captured image that define a large phenotypic space from hundreds of cells in a particular treatment condition in as short a time as possible.

2.1.2. High Content Imagers versus Microscopes

High content imagers (HCI) at the present time typically consist of an automated microscope or components of a microscope in a box supplied with image analysis software and commonly have image management packages to store images on a server/data storage system. They generate very large image sets and associated meta-data that can be greater than a 0.5 TB per day at full capacity in screening campaigns. Compared to research microscopes, the configuration choices are much more limited. For example, a particular instrument may have only one fluorescent light source option, one or two camera options and no choices regarding objectives. But they have rapid autofocus and very precise stages optimized for multi-well plates and most systems offer environmental control for live cell imaging.

2.1.3. High Content Analysis versus High Content Screening

HCI can be used in two contexts. In a research environment, HCA is often a medium throughput activity were a few hundred or a few thousand perturbagens (compounds, drugs, siRNAs, cDNAs) are tested and scores of parameters are recorded from each individual cell using multiple imaging channels. The readouts can be kinetic or single endpoint using live or fixed cells respectively. The images are retained and perhaps reanalyzed with the goal of getting a very complex assessment of different subpopulations of cells in each well. The purpose could be basic research or to serve as a tertiary screen to study the toxicological properties of hits from a target based screen. In contrast, HCS is like other HTS methods aimed at screening 100,000s of pertubagens. The goal is to identify hits for additional testing. The readout is typically a fixed endpoint to reduce noise and facilitate automation. A smaller number of parameters are often selected during the assay development phase to speed screening and reduce the data storage requirements.

2.1.4. Background Information on Microscopy

Fluorescence imaging is widely used in biomedical research so a detailed introduction is not appropriate here. Several commercial and non-commercial websites provide extremely detailed information about microscopy and readers are encouraged to explore them to learn details about different fluorescent techniques listed in the links provided in below:

- Florida State University

- Home: http://micro.magnet.fsu.edu/primer/index.html
- Microscopy: http://micro.magnet.fsu.edu/primer/index.html
- Fluorescence microscopy: http://micro.magnet.fsu.edu/primer/index.html
- Objectives: http://micro.magnet.fsu.edu/primer/anatomy/objectives.html
- Filters: http://micro.magnet.fsu.edu/primer/lightandcolor/filtershome.html
- Confocal: http://micro.magnet.fsu.edu/primer/techniques/confocal/index.html
- Numerical Aperture: http://micro.magnet.fsu.edu/primer/anatomy/numaperture.html
- Airy Disc: http://micro.magnet.fsu.edu/primer/java/imageformation/airydiskbasics/index.html
- Nipkow: http://micro.magnet.fsu.edu/primer/techniques/fluorescence/spinningdisk/index.html
- Laser light sources: http://micro.magnet.fsu.edu/primer/lightandcolor/lasersintro.html
- Fluorescent Proteins: http://micro.magnet.fsu.edu/primer/techniques/fluorescence/fluorescentproteins/fluorescentproteinshome.html
- FRET: http://micro.magnet.fsu.edu/primer/techniques/fluorescence/fret/fretintro.html

- Olympus
  - Microscopy: http://www.olympusmicro.com/primer/anatomy/anatomy.html
  - Fluorescence microscopy: http://www.olympusmicro.com/primer/techniques/fluorescence/fluorhome.html
  - Objectives: http://www.olympusmicro.com/primer/anatomy/objectives.html
  - Filters: http://www.olympusmicro.com/primer/lightandcolor/filter.html
  - Confocal: http://www.olympusmicro.com/primer/techniques/confocal/index.html
  - Numerical Aperture: http://www.olympusmicro.com/primer/anatomy/numaperture.html
  - Airy Disc: http://www.olympusmicro.com/primer/techniques/confocal/resolutionintro.html
  - Nipkow: http://www.olympusmicro.com/primer/techniques/confocal/confocalscanningsystems.html
  - Bleed-through: http://www.olympusmicro.com/primer/techniques/confocal/bleedthrough.html
  - Non-laser light sources: http://www.olympusmicro.com/primer/techniques/confocal/noncoherentsources.html
  - Laser light sources: http://www.olympusmicro.com/primer/techniques/confocal/confocallaserintro.html
  - Fluorescent Proteins: http://www.olympusmicro.com/primer/techniques/confocal/applications/opticalhighlighters.html
  - FRET: http://www.olympusmicro.com/primer/techniques/fluorescence/fret/fretintro.html
- Nikon:
  - Home: http://www.microscopyu.com
  - Microscopy: http://www.microscopyu.com/articles/optics/components.html
  - Fluorescence microscopy: http://www.microscopyu.com/articles/fluorescence/fluorescenceintro.html
  - Objectives: http://www.microscopyu.com/articles/optics/objectiveintro.html
  - Filters: http://www.microscopyu.com/articles/fluorescence/filtercubes/filterindex.html
  - Confocal: http://www.microscopyu.com/articles/confocal/index.html
  - Numerical Aperture: http://www.microscopyu.com/articles/formulas/formulasna.html
  - Airy Disc: http://www.microscopyu.com/articles/optics/mtfintro.html
  - Nipkow: http://www.microscopyu.com/articles/confocal/confocalintrobasics.html
  - Bleed-through: http://www.microscopyu.com/tutorials/java/cubeprofiles/triple/index.html
  - Fluorescent Proteins: http://www.microscopyu.com/articles/livecellimaging/fpintro.html
  - FRET: http://www.microscopyu.com/articles/fluorescence/fret/fretintro.html
- Zeiss:
  - Home: http://zeiss-campus.magnet.fsu.edu/index.html
  - Microscopy: http://zeiss-campus.magnet.fsu.edu/articles/basics/index.html

- Fluorescence microscopy: http://zeiss-campus.magnet.fsu.edu/articles/basics/fluorescence.html
- Objectives: http://zeiss-campus.magnet.fsu.edu/articles/basics/objectives.html
- Filters: http://zeiss-campus.magnet.fsu.edu/tutorials/matchingfiltersets/index.html
- Confocal: http://zeiss-campus.magnet.fsu.edu/tutorials/opticalsectioning/confocalwidefield/index.html
- Numerical Aperture: http://zeiss-campus.magnet.fsu.edu/print/basics/resolution-print.html
- Airy Disc: http://zeiss-campus.magnet.fsu.edu/tutorials/basics/airydiskbasics/index.html
- Nipkow: http://zeiss-campus.magnet.fsu.edu/articles/spinningdisk/index.html
- Dual-spinning disc confocal: http://zeiss-campus.magnet.fsu.edu/articles/spinningdisk/introduction.html
- Bleed-through: http://zeiss-campus.magnet.fsu.edu/articles/spectralimaging/introduction.html
- Non-laser light sources: http://zeiss-campus.magnet.fsu.edu/articles/lightsources/index.html
- Laser light sources: http://micro.magnet.fsu.edu/primer/lightandcolor/lasersintro.html
- Fluorescent Proteins: http://zeiss-campus.magnet.fsu.edu/articles/probes/index.html
- FRET: http://zeiss-campus.magnet.fsu.edu/articles/spectralimaging/spectralfret.html

2.1.5. Excitation Sources for High Content Imaging

There are three basic types of light sources for microscopy: lamps, lasers and light emitting diodes. Lamps provide a relatively broad excitation source from UV to IR. Lamps are likely to be able to excite many fluorescent dyes and proteins but the power in a given region may be low. Lamps may need to be replaced frequently and be realigned to give optimal excitation. Xenon lamps have a broad spectrum but relatively little UV and substantial IR that is beyond the region used to excite most dyes and fluorescent proteins. Halogen lamps also have a broad spectrum but, like Xenon lamps, give of large amounts of useless IR and heat. Mercury lamps have sharp and narrow emission bands that require a carefully selection of filters. They offer strong UV excitation but have relatively short useful lifetimes, compared to other light sources. Conventional lasers provide a fixed monochromatic wavelength per laser. Power is substantial but may not offer optimal excitation for certain targets. White light and tunable lasers are entering the market but have not yet been used in commercially available HC imagers. Lasers have long lifetimes but are relatively expensive to replace, although this is changing. LEDs are now bright enough to be used in microcopy and have made a rapid entry into the HC arena. LEDs offer long-life and are much more stable light output. The have much less fluctuations on the second time scale compared to lamps and also do not slowly dim in intensity on the week/month timescale the way lamps do. This can substantially reduce noise in HC assays as images are acquired across a plate.

2.1.6. Objectives

Anyone using microscopy as a primary tool needs to be familiar with microscope objectives. The imaging sites listed above provide outstanding introductions to this important knowledge area (Figure 1). Some key facts to know include the following:

1. Objectives are designed with a certain working distance and plate or coverglass thickness in mind. To get the best image there should not be a mismatch.
2. Objectives are designed to be used dry (an air gap between the objective and the specimen) or use water or oil between the objective and the plate/coverglass to enhance numerical aperture multiplier. A dry objective cannot be used with fluid and vice versa.
3. As a rule of thumb, for a given magnification, for example 20x, the objective with the higher numerical aperture (N.A.) will collect more light, reducing exposure times and increasing throughput. Similarly, water and oil objectives collect more light than dry lenses. In fluorescence applications, more light is better and in HCS, since speed is important, anything that can be done to reduce imaging time will speed screening. But using emersion objectives (water or oil) makes microscopy much more difficult, especially

in a screening environment. Therefore, most HCI uses dry objectives but with as high a N.A. as possible (Figures 2 and 3).

2.1.7. Airy Discs

When light comes from a point in a specimen and goes through an object to form a part of an image, the light from the point does not form a point in the image. Rather it forms a diffraction pattern with a central maximum surrounded by a few rings (Figure 4). The central maximum, usually called an Airy Disc, contains most (>80%) of the energy. The minimum distance between airy discs then can be resolved defines the resolution of the objective and varies with the wavelength of the light ($\mathbf{r_{Airy} = 0.61 \times (\lambda_{Ex} / NA_{Obj})}$**.** The higher the N.A. of the objective, the smaller the airy disc and the better the resolution (Figure 5).

2.1.8. Bleed Through

An important consideration for HCA is that the fluorescent dyes and proteins used in assays typically have broad excitation and emission spectra. As a result there can be significant bleed through from one fluorescent probe to another. At least four things can be done to minimize this:

1. The excitation wavelengths chosen should take into account the peak properties of the fluorescent targets to minimize cross excitation. With laser or LED light sources this is less of a burden. With other light sources, such as halogen, xenon or mercury lamps, careful selection of filters in the excitation path is required. In any case, absorption bands have tails towards shorter wavelengths. So a narrow band chosen to excite green fluorescent protein (GFP) will likely also excite a yellow fluorescent protein (YFP) (Figure 6).
2. The filters in the emission path must be optimized to minimize cross talk between the different fluorescence emitters and that is directly dependent on the dichroic filter in the optical path. Emission bands have tails towards the longer wavelengths. Fluorescence from GFP will bleed into the YFP channel. Keep in mind that imperfections in glass, optical materials and coatings used on filters in fluorescence detection commonly display multiple excitation or emission peaks; therefore it is recommended to review the specifications of the filters to understand how they perform.
3. Adopt a strategy to reduce problems with cross talk by having the brighter signal in the longer wavelength channels (Table 1).
4. Routinely assess cross talk between different channels. At a minimum, there should be control wells where the bleed through from the shorter wavelength channel into the longer wavelength channel is measured. In High Content Assays using primary and secondary antibodies, new lots of antibodies or just variations in handling on different days can lead to significant changes in fluorescent signals, with consequences for relative signal strengths in different channels. So controls need to be done on a routine basis.

Figure 1: Important properties of objectives are indicated on the barrel of the objective, these include magnification, numerical aperture, working distance, immersion medium and coverslip thickness. Taken from http://www.olympusmicro.com/primer/anatomy/specifications.html

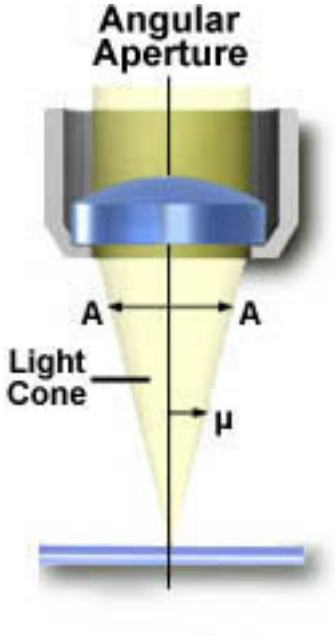


Figure 2: Numerical aperture is dependent on the half angle of the aperture (m) and the refractive index of the medium (n) between the objective and the specimen. Taken from http://www.olympusmicro.com/primer/anatomy/numaperture.html

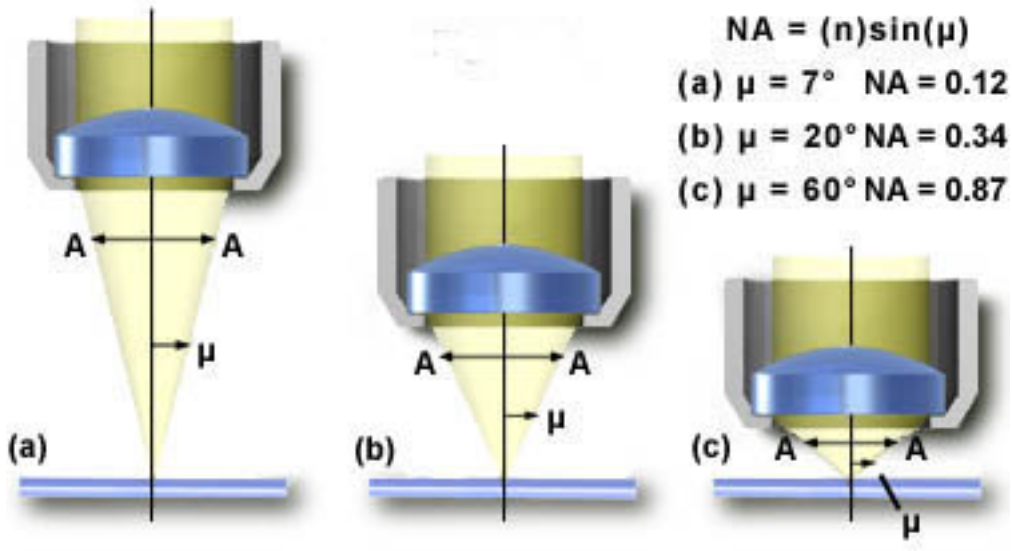


Figure 3: As the angle increases from 7o to 60o there is a 7 fold increase in N.A. Taken from http://www.olympusmicro.com/primer/anatomy/numaperture.html

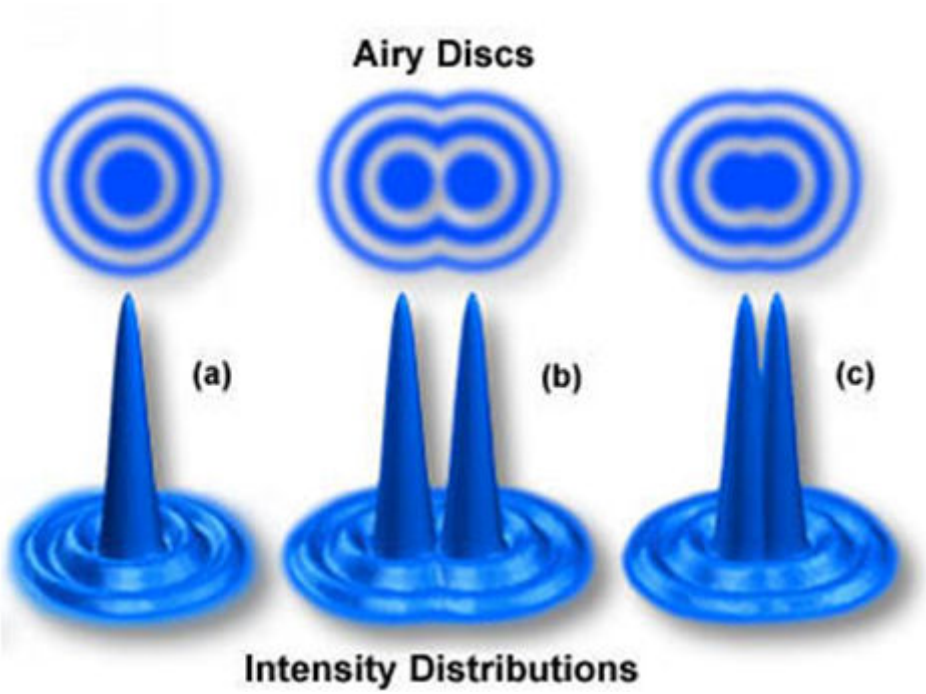


Figure 4: Schematic representation of Airy Disc intensity distributions. Adapted from http://www.olympusmicro.com/primer/anatomy/numaperture.html

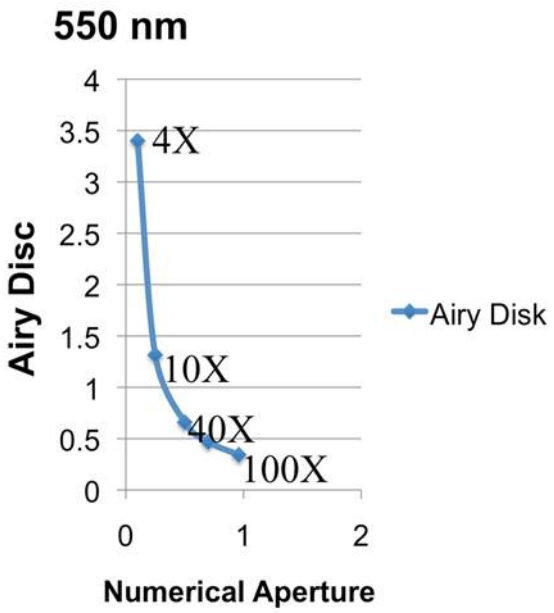


Figure 5: Airy disc versus numerical aperture (N.A.). As N.A. increases, the airy disc decreases.

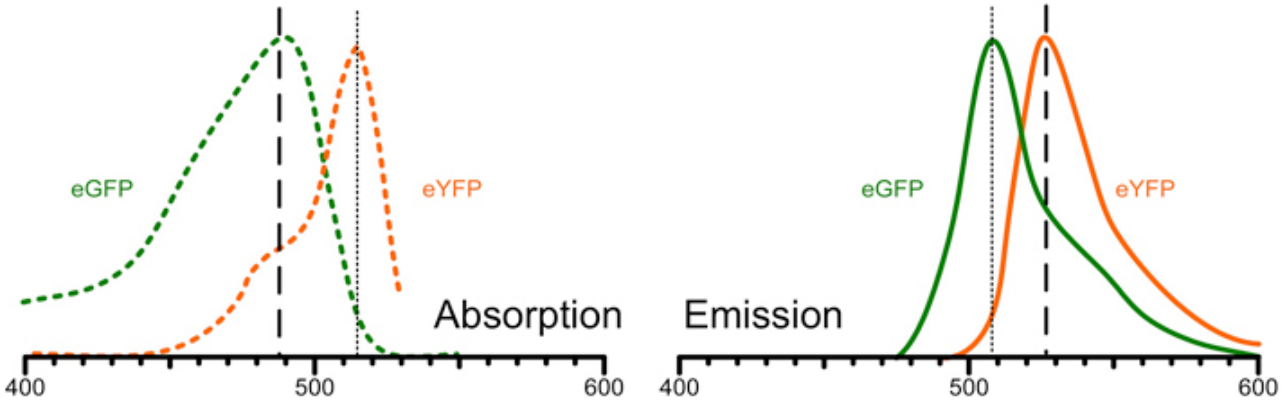


Figure 6: Fluorescent proteins often have overlapping excitation and emission spectra. eGFP and eYFP have substantial overlap, making clear separation of the signals difficult, even with the best filter choices.

Table 1: Making sure the signal from the emitter in the longer wavelength is brighter than the signal from the emitter in the short wavelength can minimize cross-talk between fluorescence emitters. If the brighter signal is from the shorter wavelength emitter, much of the signal detected in the long wavelength channel will be from the "wrong" emitter.

Fluorescence		Detection			
Green Label	Yellow Label	Green Channel		Yellow Channel	
		From Green (~ 100%)	From Yellow (~0%)	From Green (~50%)	From Yellow (~100%)
++	++	++	~	+	++
++	++++	++	~	+	++++
++++	++	++++	~	++	++
++++	++++	++++	~	++	++++

2.1.9. Detectors

HCI use two major types of detectors; digital cameras and photomultiplier tubes (PMTs). Digital cameras for HCS benefit from the large market for personal cameras that lead to decreased costs and increased chip size and defense needs for high sensitivity. HCI often use CCDs, EMCCDs and sCMOS cameras with high frame rates (100 FPS), large dynamic ranges (>20,000:1), broad spectral sensitivity (400-900 nm and higher), and high resolution (>2000 × 2000 pixels). While these cameras can provide high quality images, the files are large with consequences for image storage systems. The cameras are usually monochrome cameras. Color images are produced by acquiring images of the same field serially, using different optical filters. PMTs are based on a very mature technology, have extreme sensitivity to measure very low light intensity and fast responses with wide spectral sensitivity and are almost always used in conjugation with a laser source. To produce images, they are used in conjunction with a scanning technology that moves a light beam, typically a laser beam, across a sample. In HCS, the scanning is relatively slow but several PMTs can be used simultaneously to acquire data in different fluorescent channels.

2.1.10. Autofocus

HCI use two different approaches to focusing on the specimens. These are 1) laser-based systems that detect the bottom of the plate and 2) image analysis-based systems that step through the specimen and use algorithms to determine the optimal focus plane. Since focusing takes time it is usually not done every time an image is acquired in HCS. Testing is needed to determine the minimum number of times focusing needs to be done to provide reliable data, which is related to the plate material used. The laser-based systems are fast but can perform poorly if the plates are not extremely flat or if the specimens are thick. The image analysis-based approaches are comparatively slow and have the caveat if fluorescent debris, artifacts, material or even clumps of cells which are not within the Z focus plane of uniformed cells in the well, the focusing typically fails. For these reasons it may be prudent to focus on fluorescence outside of typical lint debris that fluorescence in blue to violet wavelength.

2.1.11. Environmental Controls

HCA can be done on live cells to study cell movement, cell proliferation, cell death, and also to use various reporters to monitor protein interactions, membrane potentials or intracellular Ca^{2+} levels. This will require that the instruments control temperature, humidity and CO_2 levels. The larger instruments from most vendors have environmental controls as standard features or optional packages. If long-term time lapse imaging is planed then testing different instruments prior to purchase is recommended. While temperature and CO_2 are relatively easy to control, humidity is not and evaporation from multiwall plates can affect cell behavior. In addition, intense illumination of cells with lasers or lamps can damage or kill the cells. Therefore it is important to verify that the imager can detect critical features over time without causing cell damage.

2.1.12. Liquid Handling

Some instruments offer liquid handling to permit the addition of compounds, drugs, etc. to cells in individual wells. This is almost always done in live cell imaging situations to measure cell responses using fluorescence reporters to monitor membrane potentials or intracellular Ca^{2+}. Typically the liquid handling is done with a pipette like device, some with disposable tips.

2.1.13. High Content Imagers

There are many HCI on the market, with vendors releasing new models regularly. Therefore, it is impossible to have a resource that is truly comprehensive and current. Individuals interested in acquiring a new instrument are encouraged to survey the current market after first developing a detailed user requirements specification. International meetings, such as the Society for Laboratory Automation and Screening, PittCon, or CHI High Content Analysis are excellent venues to view demonstrations from many HCS vendors. There are a number of

social media websites such as LinkedIn and Facebook focused on HCS/HCA that have members providing feedback about these instruments as well as dedicated user group websites. The Cold Spring Harbor Meeting on High Throughput Phenotyping is an exciting place to learn about cutting edge approaches.

2.1.14. Image and Data Analysis

Often the major factors that differentiate the High Content platforms from different vendors are in the software that acquire, analyze, and manage HC images. Most software packages from major vendors now offer advanced data analysis systems that allow tracking of entire screening campaigns. Perhaps more important, the software needs to have a comprehensive and user friendly system for developing a High Content Assay. While scientists often have a basic idea of the type of assay that will run, the optimal image analysis algorithms and features or parameters that need to be measured in a screen have to be determined using positive and negative referenced controls. The ease and speed at which this can be done is highly dependent on image analysis tools provided by the vendor. Newer software packages have improved GUIs based on real world workflows. Anyone acquiring a new HCA system should include image analysis tools in the user requirement specification.

2.2. Wide Field Imagers

These instruments are similar to and, indeed are often built around inverted research microscopes from major vendors such as Olympus, Nikon, Zeiss, etc. The hardware solutions offered by different vendors have evolved rapidly over the past 10-15 years and are now robust and provide excellent images quickly.

BD Biosciences distributes the BD Pathway 435^{tm}, a unit with metal halide and transmitted light sources. This is also equipped with a Nipkow spinning disk for confocality (Note: BD is discontinuing their HCS instruments but there are many in academic labs and core facilities)

GE Healthcare markets the IN Cell Analyzer 1000 and 2000. The InCell 2000 is the new generation instrument; it uses a metal halide lamp. Camera options are 1392 × 1040 or 2048 × 2048 pixels. Transmitted light modes include bright field, phase and differential interference contrast.

IDEA Bio-Medical has a large instrument, the WiSCAN that uses a mercury light source for fluorescence and LEDs for transmitted light. The instrument uses a 512 × 512 water-cooled EMCCD camera for fast, sensitive imaging in a HCS environment. The Hermes 100 is a small bench-top instrument for individual labs. It uses LED light sources to allow two-color and transmitted light image acquisition.

MAIA Scientific markets the MIAS-2^{tm}. This instrument can acquire 5 different transmitted light channels with a halogen light source and up to 8 fluorescent channels using a xenon light source. Imaging is done with a color camera and an intensified B&W camera.

Molecular Devices has one wide field imager, the ImageXpress Micro HCS system, which has an integrated fluidics system for delivering reagents in live cell imaging applications. It uses a xenon lamp and can use air or oil objectives.

Perkin Elmer sells the Operetta, a bench top widefield unit with a xenon lamp and an LED for transmitted light. It has a spinning disk confocal option

ThermoFisher (Cellomics) developed the first commercial HCS imager and now sells three wide field instruments. The ArrayScan VTI HCS Reader is an instrument suitable for a core facility or large laboratory. It uses a metal halide or LED light source and can be enhanced with a spinning disc confocal option. The Cellinsight is designed as a “personal” imager. It also uses an LED light source and is designed for use with four common dyes; Hoechst, FITX, TRITC and Cy5. The ToxInsight IVT Platform is designed to focus on identifying potential toxic liabilities in newly identified compounds. It is also a small footprint instrument using LEDs and a four-color approach to HCA.

Vala Sciences manufactures a Kinetic Image Cytometer (KIC) designed for kinetic analysis of calcium dynamics in an HCS system. It uses LEDs and large format cameras to acquire data.

2.3. Confocal HCA Imagers

Confocal microscopes use a light barrier with a fixed or adjustable pinhole to eliminate light that is in front or behind the focus plane of an objective. This gives much better depth resolution and improved contrast by rejecting light from out of focus sources. But it causes reductions in the light signal. It also only works for a single point in the specimen at any given moment. To overcome this problem the sample must be moved across the sampling point or the light beam and pinhole need to be scanned across the sample. One approach is the Laser Scanning Confocal Microscope (LSCM). The other is the Nipkow spinning disk that has multiple small pinholes or curved slits to increase illumination and the number of points in the specimen that can be imaged simultaneously. To obtain an optimal image with regards to light transmission and Z-axis resolution, the pinhole size must be matched precisely to the objectives Airy Disc. LSCMs have adjustable pinholes that can be varied depending on the objective and other factors, such as the wavelength of the illuminating light source. But the intense laser beam can bleach the specimen and it often takes a few seconds to scan a region of interest (ROI). Spinning discs sacrifice most of the illuminating light used to excite fluorescence in the specimen but can scan a ROI in a few hundred milliseconds and result in less bleaching and increases throughput. They are preferred for live cell imaging. Yokogawa has devised a dual spinning disc technology with lenses in the first disk that focus light on the pinholes in the second disc. This increases the illumination of the specimen and, importantly for HCS, decreases image acquisition time.

Confocal imagining is usually more expensive in terms of capital investment and screening time. It is best used for imaging small intra-cellular structures, small cells, complex 3-D structures and samples with strong background fluorescence. HCS campaigns have been run using confocal imagers to eliminate the need to wash stains from cells, a big advantage if the cells are loosely adherent. Furthermore, the sharper images obtained via confocal methods could make image analysis process easier.

BD Biosciences distributes the BD Pathway 855^{tm}, a Nipkow spinning disk system with mercury halide and transmitted light sources. It has an integrated liquid handler and integrated environmental control. It can also be used in wide field mode. This imager is often used for kinetic studies of signals relevant to physiologists, such as membrane potential or calcium (**Note**: BD is discontinuing their HCS instruments but there are many in academic labs and core facilities).

GE Healthcare markets the IN Cell Analyzer 6000. This is a line scanning LSCM with a variable aperture. It has 4 laser lines (405, 488, 561, 642) and an LED for transmitted light and a large format sCMOS camera. It has an integrated liquid handler and environmental control.

Perkin Elmer sells the Opera, a HTS system designed with water emersion lenses to give higher N.A. It uses laser based excitation combined with a Yokogawa dual spinning disc system to give confocality.

Molecular Devices has a point scanning LSCM, the ImageXpress ULTRA. This machine has four lasers and 4 PMTS that can be operated simultaneously or sequentially. It has options for air or oil objectives.

Yokogawa has two confocal imagers that exploit their dual spinning disc technologies. They have discs with different size pinholes, depending on objectives in use. The CellVoyger CV1000 is designed for long term live cell imaging with the option for oil immersion lenses. The CellVoyager CV7000 is an instrument designed for HTS, taking advantage of three large chip (2560 x 2160) cameras and a choice of lasers as well as halogen lamp and a LED for UV imaging. Live cell imaging is provided as well as liquid handling and water immersion lenses.

2.4. Laser Scanning Cytometers

These imagers are conceptually similar to a flatbed scanner with laser beams scanned across the entire surface of the plate and fluorescence detected with PMTs. They produce images equivalent to at maximum a low NA 20X objective and are good at detecting cells, including DNA content and colonies and even model organisms such as zebrafish, but not subcellular features or processes. LSCs have a very large depth of focus. They are often used to identify fluorescent intensities above a threshold. An example is nuclear translocation assays, where a diffusely localized protein in the cytoplasm gives a low signal but when concentrated in the nucleus gives a high signal. Other applications include cell proliferation, cell toxicity, protein kinase activation, and cell cycle analysis. This approach might be considered a medium content, high throughput technology.

The Acumen ^{e}X3 has 3 lasers (405, 488, 633nm) and 4 PMTS and has been used in many HTS projects.

Molecular Devices ImageXpress Velos Laser Scanning Cytometer (formerly IsoCyte) uses 2 lasers (selected from 405, 440, 488, 532, 633nm), 4 PMTs. It also uses light scattering as a method to detect non-fluorescent objects such as colonies.

The Compucyte iCyte is a hybrid instrument that uses laser scanning on an inverted microscope with objectives (10, 20, 40, 60, 100x) and up to four lasers (selected from 405, 488, 532, 561, 594, 633nm). It also uses 4 PMTs.

Slide based scanners: Some instruments offer measurement of transmitted light in different wavelengths using line scanners. When used with conventional histological stains, this can provide very useful images and information from tissues that could be of interest in disease models. A major advantage of line-based scanners is the elimination or minimization of tiling to produce very large, high magnification images.

Aperio sells three slide scanners aimed at the pathology market. The ScanScope FL uses a mercury lamp, a 20x objective and a TDI line-scan camera to acquire images in up to 4 color channels. The ScanScope CS has 2 objectives (20x, 40x). The ScanScope AT has a 20x objective but is designed for automation with a slide loader that can hold up to 400 slides.

Hamamatsu markets the NanoZoomer, which uses TDI line-scans to acquire both transmitted light images of tissues stained, for example with H&E, PAS, or NBT stains and also fluorescence. It uses a 20x objective, a mercury lamp for fluorescence.

The Leica SCN400 and SCN400F uses a linear CCD device to acquire brightfield images, The SCN400F also can acquire fluorescence channels.

2.5. FACS like instruments

As mentioned previously FACS provides multidimensional data that can be considered in a high content approach. There are some instruments that cross the border from FACS to HCA by acquiring images and not just intensity data.

The Amnis ImageStream X uses lasers and LEDs to give darkfield, side scatter (785 nm) and fluorescence images of cells using 5 lasers (405, 488. 561, 592, 658) of cells passing through a flow cuvette.

3. Assay Concept and Design

Living cells, the basic building blocks of life, are an integrated and collaborative network of genes, proteins and innumerable metabolic reactions that give rise to functions that are essential for life. Conversely, dysfunctions in these same vital networks give rise to a host of diverse diseases and disorders. Although much less complex than *in vivo* models or complete organisms, cells possess the systemic complexity needed to study the interactions between different elements of the network and the responses of the network to external stimulations. Therefore

more and more physiologically relevant cellular models are being used to validate targets or to evaluate drug efficacy and to predict potential adverse side-effects. Furthermore, advancements in cell isolation, cell line generation and cell differentiation technologies have led to more scalable and affordable cellular models, which in turn facilitate screening using more physiologically relevant cellular models. Due to its information-rich nature, high-content screening (HCS) has become the choice for many scientists to examine the complex effects of compounds or other reagents in physiologically relevant cellular models, not only against their intended targets, but also against other cellular targets and pathways (25-29).

Like standalone high-resolution microscopes, automated HCS systems can be utilized to study many cellular processes. Some of these processes, such as protein phosphorylation, cell surface ligand binding, molecular uptake, protein expression, cell cycle regulation, enzyme activation, and cell proliferation, can be analyzed by conventional methods, though image-based methods can often deliver comparably high quality results with multiple parameters. The strength of HCS is based on its ability to enable both target-based and phenotype-based assays for otherwise intractable cellular processes. These processes often play pivotal roles in cell survival and division, and can be visualized as intracellular protein translocation, organelle structure changes, overall morphology changes, cell subpopulation redistribution, and three dimensional (3-D) structure modifications. These assays not only have been used to study fundamental biological processes and disease mechanisms, but also have been applied to new drug discoveries and toxicity investigations.

3.1. Intracellular protein translocation:

Examples of HCS assays monitoring intracellular protein redistribution include translocation of a transcription factor from the cytoplasm to the nucleus to initiate or modulate gene transcription, internalization of G-protein coupled receptors (GPCR) to initiate a signaling cascade (30, 31), translocation of glucose transporter from the cytoplasm to the cell surface to facilitate glucose uptake (32), and recruitment of LC3B, an autophagy-related protein, to the autophagosome under conditions of stress (33, 34). In order to follow the translocation event, the protein must be labeled with a fluorescent probe, often by tagging/expressing the protein directly with a conjugated fluorescent protein marker (such as green or red fluorescent proteins (GFP or RFP)). This system then can be used to study the spatial and temporal effects of external stimulants in both kinetic and end-point fashions.

Different protein tagging technologies have been developed as potential substitutions of florescent proteins. For example, the SNAP-tag, which is a 20 kDa mutant of the DNA repair protein O6-alkylguanine-DNA alkyltransferase that reacts specifically and rapidly with benzylguanine (BG) derivatives (www.neb.com) The BG moiety can then be used to irreversibly label the SNAP-tag with a fluorophore. This technology allows one to label the protein in question using chemical fluorophores with different wavelengths and cell permeability, thus facilitating multiplex readout from the same cells. Halo-tag (www.promega.com) and fluorogen activating protein (FAP) (www.spectragenetics.com) are based on similar concepts, but using different proteins and probes. These proteins and their associated probes have no endogenous eukaryotic equivalent and are not toxic to cells when expressed at low levels. The covalent nature of these technologies makes them versatile for both live and fixed cell images.

The protein tagging technologies described above require overexpression of the proteins of interest. Sometimes stable cell lines are not feasible, and inducible expression systems could be used to circumvent the situation. Not surprisingly, this approach will require more intensive assay validations due to potential variations associated with the inducible expression systems. Overexpression of some proteins may disturb the delicate balance of the cellular network or the tags may disrupt the function or trafficking of the proteins in unknowable ways, and lead to results that are not physiologically relevant. Because of this, most scientists prefer antibody staining methods to track the intracellular locations of such proteins. However, antibody staining methods generally require chemical fixation of the cells, and so are limited to end-point reads (with the exception of cell surface proteins). There are many commercially available kits for specific protein translocations. These kits are validated by

vendors but their uses are narrowly defined, and one has very limited options to change the compositions of the reagents as needed. Alternatively, there are many well-characterized antibodies available via different sources. Assay developers usually need to screen multiple antibodies to find one that works in the bioimaging-based assay, and to validate the assay using know stimulators and inhibitors. If assay developers decide to use proprietary antibodies raised in house, the selectivity of the antibodies must be critically examined, and the assay must be fully validated using know stimulators and inhibitors in related and unrelated pathways to ensure the observed translocation of proteins is specific to the biological event(s) of interested.

Lastly but not least important, it is imperative to develop image analysis algorithms and phenotype clustering statistical methods (if applicable) concurrently with the development of biological assays to make sure that the assay has optimal sensitivity towards the desirable phenotypes. These algorithms and methods must be validated using know stimulators, inhibitors and/or tool compounds. For compound screening, the same compound at different concentrations could lead to different phenotypes, due to the compound's different potencies on different pathways or due to toxic effects. Therefore, it is essential to test tool compounds in a broad dose response concentration range to find all potential phenotypes associated with the assay. This information can be used to define POSITIVE calling criteria for primary screening to minimize false positives and false negatives. These principles are applicable to all HCS assay formats.

3.2. Organelle structure changes

Examples of organelle structure change assays include the evaluation of mitochondrial membrane potential as a marker of cell health, cytoskeletal remodeling, quantification of lipid droplet formation in metabolic disease, formation of micronuclei during genotoxicity, and quantification of endocytosis or internalization for intracellular drug delivery (35, 36). Over the years, Molecular Probes® (Life Technologies) has developed many organelle-specific chemical dyes and fluorescently labeled antibodies against specific organelle markers. Recently, they also adapted the BacMam technology to express GFP-fusion constructs of different organelle markers. These dyes, antibodies and organelle markers cover a broad spectrum of wavelengths and can be used to examine the location and structure of multiple organelles simultaneously.

Development of HCS-amenable assays for structural changes can be a challenge, due to the heterogeneous morphologies of cells in dissociated cell cultures. Cell behavior is strongly influenced by local environment. There is evidence that some cell types at the edge of a colony will behave very differently than cells in the center of the colony (37-39). Recent developments in micro-patterned plate technology could be used to address this issue. These micro-patterned plates could provide niches that mimic the extra cellular matrix (ECM) of cells in tissues, and the organized patterns facilitate more uniform cell placement and adhesion to plates, thus making assay development and image analysis straightforward (40-43).

3.3. Morphology changes

Morphology change is a hallmark assay for high-content based screening. Many assays monitor cell process extension or tube formation as markers of disease. These include the measurement of angiogenesis for anti-cancer indications, oligodendrocyte differentiation for multiple sclerosis and other neurodegenerative diseases, and neurite outgrowth for different CNS indications (44). Another important area is related to cell differentiation associated morphology changes. Epithelial-mesenchymal transition (EMT) for oncology or fibrosis indications and stem cell differentiation are two notable examples (45-47).

Morphology changes can be directly monitored using bright-field image technology, or using fluorescent images with dyes or other markers that define the boundary of cells. However, assays reliant solely on morphological changes must be tightly controlled to avoid misinterpretation of results. For example, neurons, oligodendrocytes and astrocytes have very similar branched morphology, and differentiate from neural progenitor cells, though their functions are very different. Therefore, it is important to include cell-type specific markers in assays for

phenotyping, preferably including both up regulated and down regulated markers, before making final conclusions.

3.4. Cell subpopulation redistribution

Most automated cellular imaging systems allow one to view a large population of cells at a time, often at the individual cell and organelle level. This allows one to run subpopulation analysis, including co-culture of multiple cell types to mimic tissue environment, cell-cell communication for signal transduction between cells (9 Li 2003), and determination of stem cell differentiation efficiency. Together with the multi-variant analysis ability, HCS technology empowers one to learn more about the interactions between the elements of the cellular network, such as the influence of cell cycle regulation and microenvironment on different signaling pathways, and gain in depth knowledge of the basic building blocks of our body.

The assay categories described above cover many bioassays under different biology events in different pathways and/or different cellular systems (Table 2, and Figure 7). Furthermore, a complex biological event could include multiple steps in different pathways. Frequently, there are specific imaging based assays for different steps involved in the complex biological event. Figure 8 illustrates key steps involved in apoptosis and available imaging methods (www.lifetechnologies.com). Apoptosis is a very highly regulated process leading to cell death. The biochemical and morphological changes that characterize apoptosis include the activation of caspases, the loss of mitochondrial membrane potential, the loss of plasma membrane asymmetry, the condensing and fragmentation of the cellular DNA, cytoplasmic membrane blebbing, and apoptotic body formation. Finally, apoptotic cells will be destroyed by phagocytes. Inappropriate regulation of apoptosis could lead to diseases like neurodegeneration, autoimmune, AIDS, ischemia-associated injury, and cancers. Since there is no single parameter that defines apoptosis, a combination of imaging methods is recommended for reliable detection of apoptosis when conducting HCA. However, for high throughput compound screening, one may pick one or two imaging methods due to cost and screening logistic concerns. Choosing which imaging method(s) to be used is very dependent on the goal of the screening. To detect early stage apoptosis, caspase 8 activation and/or mitochondrial membrane potential assays could be used. To detect middle stage apoptosis, phosphatidylserine exposure or membrane permeability could be considered. Finally, to detect late phase apoptosis, DNA fragmentation assays could be used. However, in follow up assays, the hits from primary screening should be examined by a combination of multiple imaging methods in order to better understand the mechanism of actions.

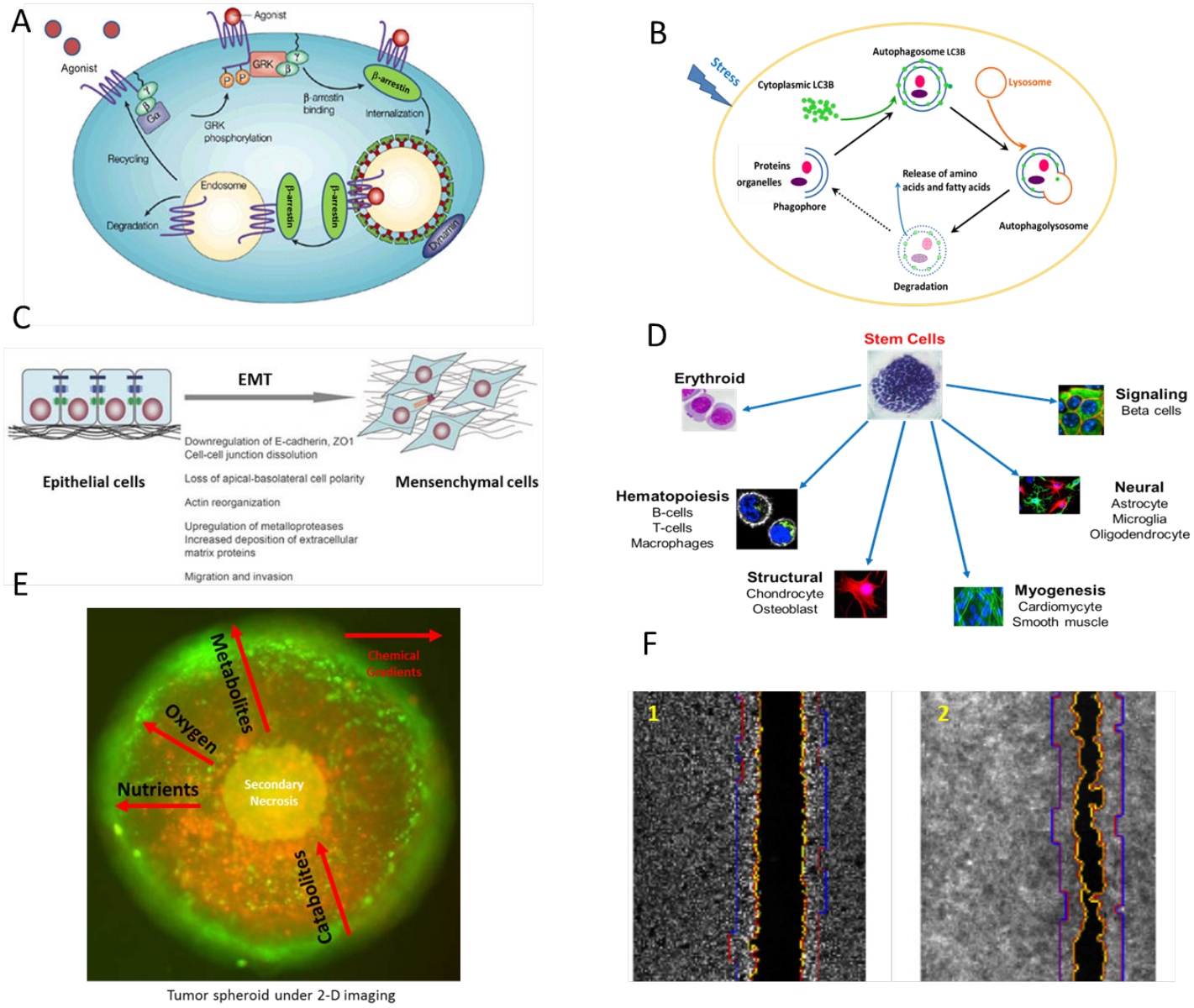


Figure 7: A) **Beta-arrestin mediated GPCR internalization** (Figure adapted from Nature.com with modifications). First, agonist-activated GPCRs are phosphorylated by GRKs (G-protein coupled receptor kinases) on their carboxyl-terminal tails. Second, arrestins translocate to and bind to the agonist-occupied, GRK-phosphorylated receptors at the plasma membrane. Third, arrestins target the desensitized receptors to clathrin-coated pits for endocytosis. Finally, receptors and arrestins are recycled or degraded. HCS can detect the internalization of GPCR by following GFP-tagged beta-arrestin (30, 31, 48). **B) Autophagy**: it is hypothesized that autophagy is up-regulated in cancer cells to promote survival in response to chemotherapy or other stresses. Autophagy includes multiple steps. The first step involves the formation and elongation of isolated membranes, or phagophores; in the second step, which involves the LC3B protein, the cytoplasmic cargo is sequestered, and the double-membrane autophagosome is formed. Fusion of a lysosome with the autophagosome to generate the autolysosome is the penultimate step. In the fourth and final phase, the cargo is degraded, and amino acids and fatty acids are released as nutrient for the cell. HCS can be used to monitor the aggregation of LC3B protein, thus following autophagy events (33,34). **C) Epithelial-mensenchymal transition** (EMT, Figure adapted from Nature.com with modifications): it is hypothesized that EMT is a key step toward cancer metastasis or toward tissue fibrosis. During EMT, cell morphology is changed from cobblestone shaped epithelial cells to elongated mensenchymal cells. Meanwhile, epithelial cell markers, such as ZO-1 and E-cadherin, are down regulated, and expression of extracellular matrix proteins, such as collagens, is increased. Using cell-mask dyes for morphology changes and specific antibodies for different cell markers, HCS can be used to detect EMT (45). **D) Stem cell differentiation:** under certain physiologic or experimental conditions, stem cells can be induced to become tissue- or organ-specific cells with special functions. These specialized cell types have distinguished shapes and biomarkers and can be picked up by HCS using cell-mask dyes and specific antibodies for different cell lineage markers (46,47). **E) 3-D multiple cell type tumor spheroids** show many differences in biological functions compared to 2-D cultures (e.g. the chemical gradients within the 3-D tumor spheroids are much

similar to in vivo while 2-D models lack such gradients) , and resemble *in vivo* tumor tissue structure (Figure adapted from Dr. Michael A. Henson Group website with modifications. Green: live cells; red: dead cells). Therefore, the spheroids have gained momentum for applications in drug discovery. HCS technology with confocality provides ways to study the 3-D structure of the spheroids (49,50). **F)** HCS technology can be used to quantify cell migration, invasion and chemotaxis in 2-D or 3-D cellular models for wound healing, cancer metastasis and inflammation studies. Mechanic scratch or micro-patterned plate technologies could be used to create a cell-free area prior to assay start. Migration of cells into the cell-free area could be measured. The key for this assay is to distinguish migrated cells from proliferated cells in the scratched area. Micro-pattern technology also is used to create micro-conduit array plates with steady chemical gradients for chemotaxis assays (51, 52).

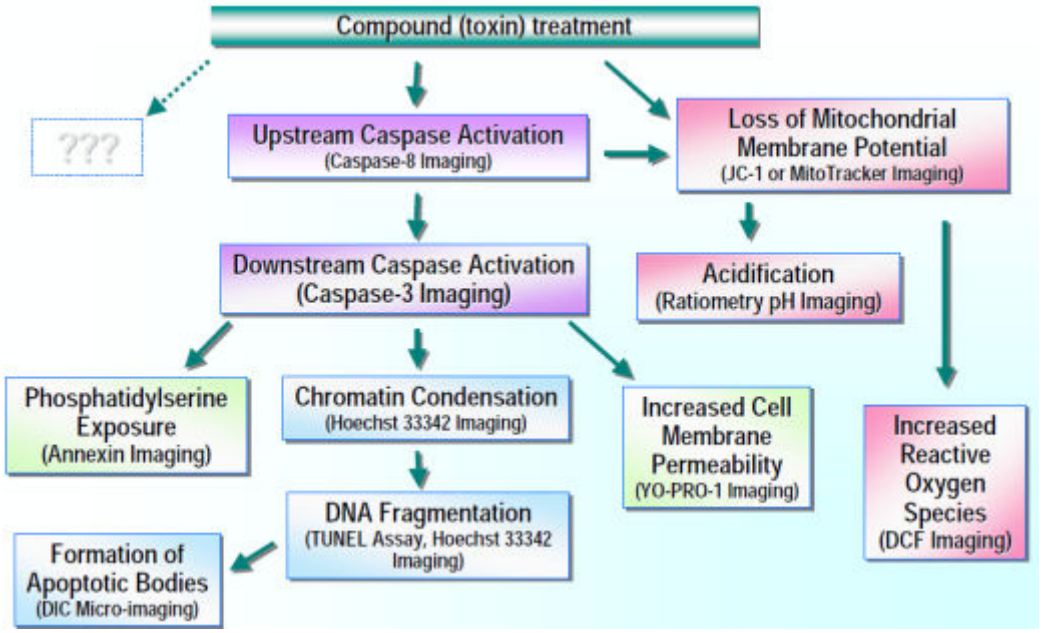


Figure 8: Key apoptotic steps and available imaging methods.

Table 2: Examples of bioassays

Assay Categories	Examples of Biology Events	Examples of Assays
Intracellular Protein Redistributions	Apoptosis	Caspases, cathepsins, calpains, Cyclins and PARP protein levels
	Autophagy	Autophagy protein LC3B aggregations
	Cytoplasm-nucleus translocation for nuclear receptors or transcriptional factors	AR, ER, GR, 5-LOX, ATF-2, ATM, beta-catenin, c-Jun, CREB ERK2, NF-κB, p53, SMAD, STATs
	Trafficking for cell surface receptors, ion channels and transporters	Beta-arrestin for GPCR internalization; ligand or receptor internalization for CB1, CB2, CRTH2, CXCR4, EGFR; Cytoplasm-cell surface membrane translocation for ion channels or transporters such as Glut1, Glut4

Table 2 continued from previous page.

Assay Categories	Examples of Biology Events	Examples of Assays
Organelle Structure and/or Function Changes	Apoptosis	Annexin V assay to detect externalization of phosphatidylserine, DNA fragmentation, mitochondria membrane potential, membrane permeability, nuclear condensation
	Autophagy	Autolysosome formation, mitochondria degradation
	Cell Division	Mitotic spindle structure by alpha-tubulin stain
	Cell polarization	Cytoskeletal re-arrangement by actin stain
	Drug delivery	Internalization of drugs via endocytosis
	Genotoxicity	Micronucleus assay to quantitate micronuclei in multinucleate cells; DNA damage indicated by phosphorylation of H2AX
	Lipid uptake and storage	Lipid droplet size and number
Morphology changes	Cell differentiation	Stem cell differentiation, epithelial-mesenchymal transition (EMT), oligodendrocyte differentiation
	Process extension	Angiogenesis, neurite outgrowth
Cell Subpopulation Redistributions	Anti-infectious	Percentage of cells infected
	Cell differentiation	Stem cell differentiation, epithelial-mesenchymal transition (EMT), oligodendrocyte differentiation
	Cell migration	Chemotaxi, wound healing, and cancer cell metastasis.

4. Cellular Models for High Content Experiments

High Content experiments depend on cell systems that serve as models for *in vivo*, typically human, biology. All models are measured by the extent to which they perform well in an assay and to the extent that they respond to stimuli in an authentic manner. Controversy exists concerning how well of in vitro cell systems accurately portray *in vivo* biology. Plating a single cell line on a two-dimensional surface in media with high levels of both oxygen and serum/growth factors may not model the in vivo situation sufficiently well for all investigations. Such systems give very robust signals in proliferation and apoptosis assays, but such responses are frequently muted *in vivo*, due to the target cells growing in an environment with multiple additional cell types. This section will explore how to insure that HCS assays best provide biologically or clinically meaningful results.

A discussion of cellular models must follow one on what needs to be modeled. In general, cell growth and the regulation of canonical signaling pathways have been modeled most frequently, particularly in the contexts of common cancers, glucose dysregulation in diabetes, neurodegeneration, pathological inflammation and toxicology. In these contexts, standard cell lines and culture conditions may be inadequate, but in other cases, such conditions may be fine. We will begin with experiments where the models are easier to establish and can be considered standard, and work towards models for more complex biological questions.

4.1. Cellular models for signal transduction pathways and other cell-autonomous responses

Much of pharmaceutical and biotechnology research is focused on finding modulators (typically inhibitors, but increasingly also to find agonists, potentiators and inverse agonists) of specific cellular target proteins. Studies on signaling pathways are also important to academic research. Such target-based research can make the search for a suitable model fairly straightforward. The easiest cellular models are immortalized and cancer cell lines. Although transformed, there are many examples of cell lines that retain the characteristics of the cell types they originated from. This includes important signaling pathways, such as estrogen receptor signaling in breast and

ovarian lines, insulin signaling in hepatic lines, and TNF-α responsiveness in immune cell-derived lines. Not all derivatives of a given cell type retain such properties, for example some breast cell lines have lost estrogen signaling. In these cases, the cell lines are better models of specific forms of cancer than of the original cell type, but then again, it is necessary to study signaling dysregulation in such diseases. Therefore, if you have a signaling pathway in mind, options for cell models can be found with a quick search of the literature. It is important to verify these cell lines are functional using known reference compounds, proteins, or other stimuli during assay development to ensure the desirable pathways are performing well in these cell lines.

The advantage of working with cell lines that are well-represented in the literature is that many additional properties of the lines that are important to consider will already be characterized and be manageable. Properties such as growth and metabolic rates can affect many assay types because some lines need to be attended more frequently than others. Other properties impact imaging assays more than other types. Colony morphology is one example. If the cells grow as clumps or clusters, then many cells will grow away from the well surface, making the imaging process more difficult than for lines that grow uniformly spread. Cell adherence can be a problem for imaging assays that require fixation and staining, as these steps add additional treatments and washing cycles to the process. Loosely adherent cells will be lost at each step unless care is taken to avoid disturbing them. This can be accomplished through automating sample preparation, where some instruments can control the rate and the placement of the reagent additions and wash steps. Common properties that can vary significantly between cell lines are summarized in Table 3.

Cancer cell lines and immortalized lines (lines that are not derived from tumors, but have inactivated senescence barriers) are easier and cheaper than primary cells, as they can be passaged in theory indefinitely; however mutations and functional response typically diminish over time. This allows both a single line to be used in experiments for many months or even years, and to expand cells prior to a screen, so that all of treatments (compounds, peptides, siRNAs, etc.) are used on cells of the same passage. Although cell lines are capable of near infinite growth, their properties do change over time and these changes can be exacerbated by inconsistent management of their growth. Understanding proper handling of cell lines by managing growth rate characteristics and cell passage number limitation is essential for consistent and biologically relevant studies. Sometimes it is necessary to sort cells to enrich the desirable cell type population. There are a few additional steps that ought to be taken when working with cell lines. Misidentification of cell lines is not rare (more than 20% error rates have been reported!). Cell lines can be mislabeled, contaminated or mishandled, making them inappropriate for the intended study. Examination of the cellular properties is essential; genotyping is inexpensive, so it is worth considering a deliberate evaluation phase for any line that is acquired through a commercial source or a collaborator.

Some cell lines are engineered for screening specific pathways. The cell lines used are chosen on the basis of their properties in cell-based assays, and the monitoring activity of the pathway through fluorescent proteins such as GFP fusion can make sample preparation much easier. In fact, they can be imaged live to better understand biological function and kinetics. Transcription factors expressed as GFP fusion proteins are common. Examples include FOXO and STAT family members, beta-catenin and TCF4/TCF7L2, NF-κB, CREB and many others. GFP fusions to other proteins are used in other robust assays, including GPCR signaling components such as PTH receptor internalization and beta-Arrestin (30, 31, 48), even protein kinases such as p38-MAPKAP2 (53), AKT kinases are activated through a transient localization to the plasma membrane and MAP/ERK family kinases can be localized to the nucleus.

In cases where the screen or assay is not specifically tied to a single pathway, but is in fact targeting a cellular response, multiple pathways may contribute to the response. The effect may be different across cell lines, even lines that are genetically and phenotypically very similar. This heterogeneity has made it difficult for many experimental results to be extended, particularly to clinically significant therapeutics. This is becoming a well-recognized issue, and some groups, both academic and industrial, have transitioned to using panels of cell lines

that are defined by both signaling characteristics and genetic background, including amplification of oncogenes and chromosomal imbalances (54) The goal is to generate data that reflects properties of cell lines grouped by common properties as reflected in the disease state in question (such as cancer subtype). The process of selecting lines is the same as outlined above, but many cell lines would need to be selected and screened. The logistical challenges are out of the scope for this chapter, but scientists looking for novel therapeutic strategies should be aware of this approach. Phenotypic questions addressed by such panels include demonstrating that blocking autophagy, ER stress response or other survival mechanisms will lead to cell death (55). Image-based approaches to phenotypic assays present unique and very valuable additions to biology and drug discovery, but the value in a discovery made in a single cell line is potentially limited unless it can be generalized or placed in a tractable signaling context.

Table 3: Summary of common properties that can vary between cell lines

Property	Impact
Growth rate	Increase in cell number; affects confluence (some properties are affected by confluence and may need to be split more frequently)
Metabolic rate	Consumption of energy and nutrients; can change health of the cell, as well as assay conditions (pH in particular)
Colony morphology	Pattern of growth as cells divide; some lines will spread evenly, others will clump
Adherence	How strongly the cells bind to the plate; influenced by the materials used to coat the plate, some lines adhere better to a collagen or fibronectin coating on the plate surface
Heterogeneity	Cell lines that appear as mixed populations morphologically
Proportion of cytoplasm	Some lines have very little cytoplasm, making many imaging assays very difficult

4.2. Primary cell models.

4.2.1. Models using differentiated primary cells

Primary cells are intrinsically more difficult to acquire and handle than most cell lines. Primary cells have limited capacities for expansion (the Hayflick Limit), or may even be post-mitotic and cannot be expanded through normal cell passaging. It can be difficult to obtain many primary cell types. Most human cell types, including pancreatic β-cells, adipose, primary tumor and tumor-associated fibroblasts, kidney and liver hepatocytes or macrophages frequently require research collaboration or material transfer agreements with hospitals or specialized procurement facilities. Commercial sources are available, but are expensive. Primary cells from animal models, particularly rodent, are much more common. On-site animal facilities can make procurement simpler to plan for, but may require a researcher to isolate the cells themselves. The two biggest logistical challenges to using differentiated primary cells as experimental models are (a) donor consistency and (b) delivery schedule. These can be more manageable for animal sources but both can be major problems for human samples, particularly irregular delivery schedules. Differentiated human cells are collected during surgical or post-mortem procedures. Although disease tissue is frequently sought as a bona-fide model of the disease itself, sample heterogeneity is typically much greater and some samples cannot be used. This is true for non-cancerous samples as well, including hepatocytes or β-cells from diabetics and synovium samples from patients with Rheumatoid Arthritis. Even when samples can be used, there is frequent variability. Some primary cell models require media changes during the first few hours in culture, and the timing can vary from sample to sample. Samples are typically collected, isolated (purified) and shipped fresh within hours of collection, so advance warning is limited and a lab that depends on these will need to be prepared. In some cases, cells can be cryopreserved, greatly simplifying the experimental process.

4.2.2. Models using primary cells produced from differentiated stem cells

An alternative to using differentiated primary cells is to differentiate stem cells in the lab for the assays. Various adult stem cell types are available, and each can be used to generate different types of cells. Mesenchymal stem cells can be used to generate hepatocytes, skeletal muscle, adipose cells, and others. The extent to which they differentiate can be variable, creating a *de facto* co-culture system with cells that did not fully differentiate. In some cases, splitting and purifying the cells is not practical. Hepatocytes form very tight junctions, making separating them difficult and adipocytes have the unexpected property of being buoyant when they have accumulated significant lipid stores. In addition to the use of partially differentiated stem cells, protocols exist for differentiating pluripotent stem cells and iPS cells. These approaches are still under development and rather specialized.

4.2.3. Establishing primary cell models

Primary cells are valuable because they retain properties beyond what cell lines can provide. All cell lines have significant genetic and regulatory alterations. The price for a steady supply of cells is that many cell type-specific properties are greatly diminished or lost entirely. Loss of CDK inhibitors and teleomerase, frequent activation of p53 and at least some chromosomal changes result in the degradation of cell type-specific functions; indeed many cell types are terminally differentiated, and this post-mitotic state is essential for physiology and morphology of the cell. Primary cells have a greater capacity to retain these properties, but they are affected by culture conditions, and therefore establishing proper culture conditions is essential to leveraging the benefits of using primary cells. For most cell models there is a strong primary literature history that describes the critical properties of the cell type in question and the culture conditions necessary to maintain them, although exceptions exist and some scholarship researching the models under consideration is important. Typically, conditions that need to be specified include media and supplements or the need for supporting feeder cells to produce native growth factors. Supplements in media for primary cells are typically titrated carefully to support growth but to avoid being higher than necessary. As such, the media may expire more rapidly than standard media preparations. In general, it is better to avoid proprietary media or supplement formulations because it is not possible to specify the experimental conditions and inter-lot variability can lead to failed assays. Therefore, some reverse engineering may be required to adapt the cell culture system to one that is appropriate for the assay being developed. As a quick example, primary human hepatocyte culture has been optimized for toxicological studies using commercial ITS (insulin-transferrin-selenium) formulations, but the level of insulin is far higher than normal, and precludes any insulin sensitivity of the hepatocytes. To adapt the hepatocyte culture system to one that can be used for the study of glucose regulation, the commercial ITS solution needs to be replaced with individual component stock solutions that can be independently varied. For proprietary formulations, manufacturers are typically reluctant to fully describe their composition, although they will often confirm whether specific materials or growth factors are present when asked as a specific question. Nevertheless, a lack of complete understanding of the culture conditions may lead to surprises later on.

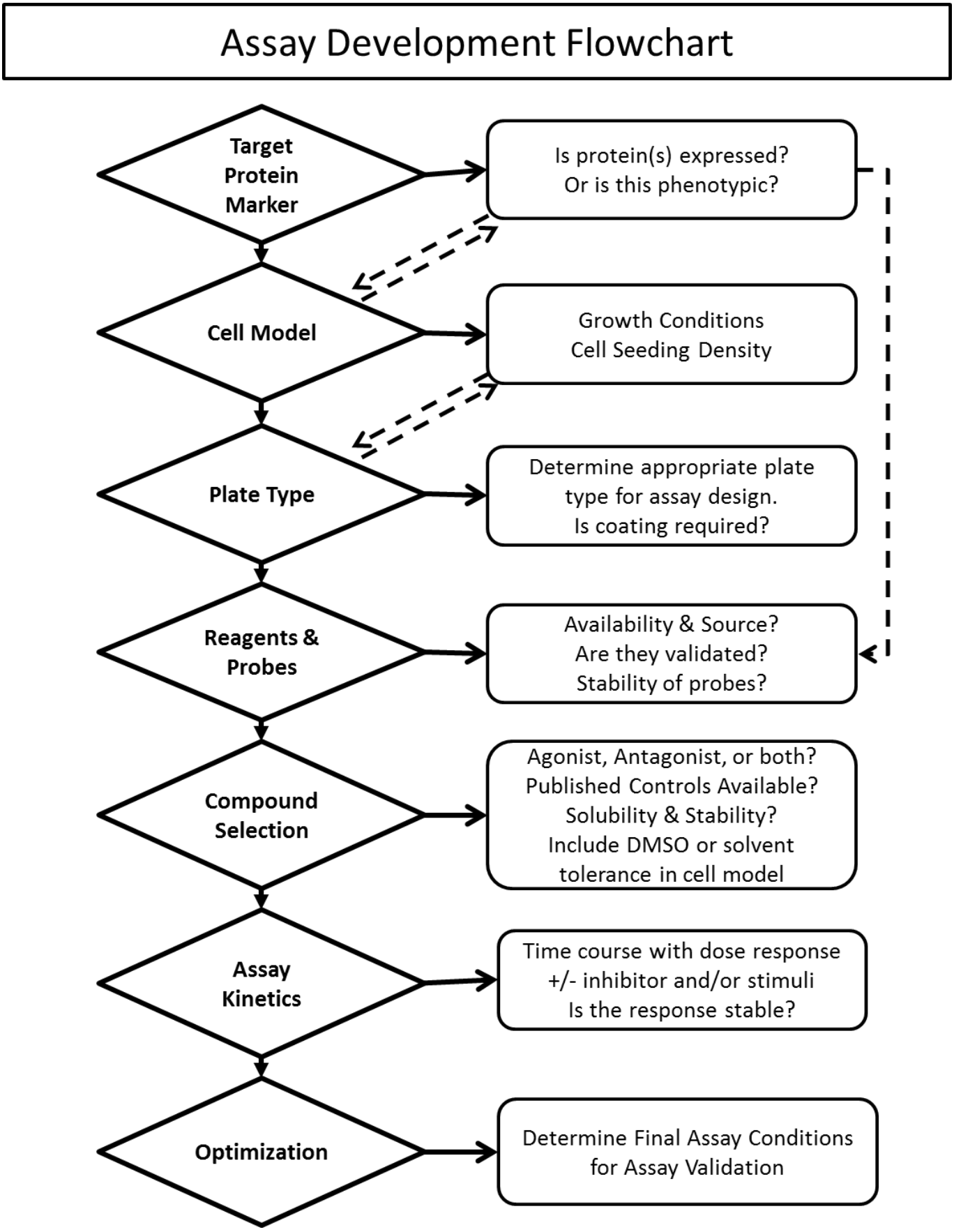
Beyond the media requirements, there are frequently additional specifications regarding seeding conditions. Cell density requirements are typically fairly rigid, particularly if they are high. Many post-mitotic cell types are plated at confluence, and deviating from this will cause the cells to dedifferentiate. This can be a very difficult step to optimize, as the fraction of surviving cells capable of adhering to the culture matrix will vary from sample to sample; for new samples, it may be necessary to plate cells across a range of seeding densities. The ability to work with a single batch of cryopreserved cells helps tremendously with this step more than any other. Addition of support extra-cellular matrices, such as collagen, laminin or fibronectin coated plates or basement membrane gels (e.g. Matrigel™) may be required. This is especially important for studies on certain cancer cells, such as breast cancer lines, primary neurons and hepatocytes. Often, it is not possible to omit these and maintain the cells for any length of time. Neurons, in particular are very sensitive to changes in substrate conditions. Different concentrations or types of poly-lysine, laminin or proteglycans produce dramatic changes in neuronal

phenotypes. Even under optimal conditions in *ex-vivo situations,* these post mitotic cells will have a defined life span.

Standardizing primary cell culture conditions is essential for robust assay performance. For experiments where cells are used from a new source (patient or animal) for each experiment, responsiveness will vary, and separate normalizations will be required for each experiment. Endothelial cells, a proliferating primary cell used in angiogenesis experiments, form tube-like channels when plated on a basement membrane matrix. The dynamics of this process is affected by modest changes in source or lot, seeding density, passage number, basement membrane matrix composition and media. The first three factors mean that the assay responsiveness will change during repeated runs of an experiment, so historical performance comparisons are difficult. There is a significant literature from the HTS field on the number of controls that are needed per plate to give reliable normalization across plates and experiments (56). This is relevant to HCS as well.

5. Assay Development Considerations and TroubleShooting

5.1. Assay Development Flowchart



5.2. Target, Protein, Marker

It is important to consider the goals of the assay and whether the desired target protein/s is/are expressed in the cell model of interest, if the protein expression is constitutively active or requires activation or stimulation. It is suggested to examine relevant reference literature or the Gene Expression Omnibus for microarray data, then measure activity with a validated assay (e.g. Western Blot, ELISA, flow cytometry, etc). Another consideration is location, location, location; verify the location of the protein or marker probe expression within the cell model to determine if it is amenable for HCS.

5.3. Cell Model

Which Cell Line Should I Use?

It is highly recommended to review the literature and references to determine if the appropriate cell model or cell line is documented for the specific assay of interest. If the cell model or cell line of interest is not referenced in the literature then it is recommended to cross validate a known cell model or cell line with a known biological endpoint before proceeding with any unknown or orphan cell models. The source of the cell line must have documentation; if using a cell line from a collaborator or colleague then obtain as much background information as possible about the history and growth characteristics. Phenotype and genotype the cell line as required. If starting with a known documented cell line, it is best to purchase from established cell bank, e.g., ATCC with history of lot details and cell growth profile specifications. If using primary cells, stem cells, *ex-vivo* tissue then establish and document as much about the growth behavior in culture before transitioning to multiwell plates for further reference, reproducibility and evaluation of likelihood of success. For transient, stable, or inducible transfection or infection of reporter proteins, such as fluorescent proteins (i.e. GFP), then steps must be taken to further validate the cell line to determine the percentage of cells expressing the reporter after cell seeding, stimulation, and other treatment.

5.3.2. Growth Conditions

Define the media, serum, and other growth factors for optimal biological response. Please note, while the optimization of health of the cells and biological conditions are needed, high levels of serum can lead to compound absorption in the assay affecting the results. When and if possible reduce the amount of serum used to a minimal level without sacrificing the overall health conditions of the cells. When miniaturizing the assay to multi-well plates, it is required to verify and/or validate if it is able to reproduce the correct biological response. It is important to know how long cells can survive and respond "normally" in culture when designing these types of assays. Determine the sensitivity of the cells outside of normal optimal environmental conditions, i.e., outside of the incubator at room temperature to mimic plate handling timing, and if the cells can tolerate changes in temperature, pH or osmolarity fluctuations. Also consider whether the addition of HEPES buffer will minimize pH changes without altering the biological model response.

5.3.3. Cell Seeding Density

Determine the cell seeding density by initially plating cells to achieve at or near confluence of the monolayer or at desired density for biological outcome. For cell types that tend to form clumps, a cell strainer could be used to de-clumping. As a general rule, cells approximately 10 microns in diameter and proliferation doubling time less than 24 hours, seed ~5,000 cells per well for 96-well plates or ~1,500 cells per well for 384-well plate. From this point forward, dilute cells by increments of 500 to 1000 cells per 96-well in replicates and incubate overnight. Label cells with an indicator to identify cells, i.e., nucleic acid dye such as Hoechst 33342, DAPI, DRAQ5, or others; this can be counterstained with a cytoplasmic indictor such as Cell Mask^{TM} or other live cell indicator such as CM-FDA (5-Chloromethylfluorescein Diacetate) or Calcein-AM. These fluorescent indicators will be used to determine if the image analysis algorithm can properly identify, segment (separate) individual cellular

objects. Perform statistical analysis of number of cellular objects per field or well to determine minimal number that can be used to provide a robust assay (see Section 6).

5.4. Plate Type

Often overlooked, the plate type chosen is critical to a successful screening campaign; keep in the mind the following when choosing a plate type.

5.4.1. Plate Material

Plates are generally made of glass, quartz, polystyrene or other composite materials. Each plate material has its advantages and disadvantages, so it is important to carefully consider what type is chosen to provide the best result in the assay.

1. **Glass and quartz** are one of the flattest and best optical materials made but are also expensive to purchase and therefore they are typically only used in specialized cases where the need for enhanced optics and flatness of the plate is required to resolve detection of subcellular structures or if capturing an entire well with a high numerical aperture, high magnification objective, i.e. 40X, 60X, 63X, 100X. Keep in mind glass and quartz material will likely require substrate coating for proper cell attachment.
2. **Polystyrene** based materials are the most common in HTS and have been adopted for HCS. The advantage of these plates is the cost is relatively low and most cell types can attach without basement substrate materials or coating.

5.4.2. Substrate requirements

The use of poly-D-lysine (PDL) coating can enhance attachment and spreading in many cells. This is commonly used in to improve attachment during compound treatment and subsequent processing and labeling steps, such as cytotoxicity assays. Extracellular matrix (ECM) proteins are used to coat plates to establish or mimic appropriate biological conditions of the assay. The most common ECM base substrates include Collagen-I, Collagen-IV, Fibronectin, Laminin, and Matrigel. If using glass bottom plates, then it is absolutely necessary to coat plates with ECM or PDL coating to promote cell adhesion and attachment. For cell spreading and migration assays it is important to test individual ECMs or a combination of these substrates as the outcome is highly dependent on the cell adhesion molecules expressed by the cells and the matrix molecules they interact with. In primary cells it is almost always necessary to use a biological substrate material to achieve appropriate conditions for an assay if feeder cells are not used. For example, appropriate substrates are required for optimal axon and dendrite outgrowth from neurons. Commercially available plates with pre-coated substrate materials are offered by many manufactures. It is recommended testing more than one lot of these plates to verify assay performance and robustness as variability in manufacturer lots are not uncommon.

5.4.3. Physical dimensions of the plate

Most plate manufactures follows SLAS standard format. Table 4 shows the approximate surface area and maximum volume for a single well, based on plate manufacturer. The flatness and bottom thickness of a plate are also important parameters. When matching a plate with numerical aperture (NA) microscope objective lenses, it is important to carefully determine the plate thickness. For example, higher NA objectives such as a 40X/0.95NA objective lens likely has a coverslip thickness of 0.17 mm. Be sure the plate bottom thickness is at or near the coverslip thickness of the objective lens and is appropriate for the working distance of the objective. In this case, do not use plate thickness near 1 mm as the microscope objective lens with high NA will fail to focus on the cells with clarity.

It is also important to determine if there is a need for evaporation wells or barriers to prevent loss of liquid in wells over time for longer term incubations, depending on your assay. Test plate types to ensure cell morphology and biological outcome are not altered. Additionally it is important that the plate does not leak over time before

scanning; if the wells dry out from leaks or from wicking, the autofocusing (image based and laser based) will likely fail. Be cautious of this as potential damage from salt based storage buffer solution including sodium azide (NaN_3) on optics and electrical components inside the imager is possible.

Table 4: Approximate surface area and maximum volume for a single well

Plate Type	Surface Area / well	Volume
96-well	0.32 cm^2	< 300 uL
384-well	0.06 cm^2	< 110uL
1536 well	0.0023 cm^2	< 10uL

5.5. Reagents, Buffers, and Probes.

When beginning a new assay, if and when possible use a validated assay kit, commercial if available, to become familiar with the steps involved in performing the assay. Then decide if the assay will continue with the "kit" or if a "home-brew" assay will be developed. A home-brew kit requires further optimization, validation, and time, but often offers cost savings in larger screening campaigns. Buffers used for imaging assays include the following:

1. **Salt based solutions:** most common solutions are Hank's Balanced Salt Solution (HBSS), Phosphate Buffered Saline (PBS), and Tris Buffered Saline (TBS).
2. **Permeabilization buffers:** salt solutions or water containing detergents such as Triton X-100, Tween-20, SDS, NP-40, or other detergents.
3. **Blocking buffers:** for antibody labeling, these are salt solution buffers containing protein such as BSA, fractionated antibody chains, or whole or fractionated serum from animal species that correspond to any secondary antibodies that will be used in the assay.
4. **Fixation buffers:** for cell preservation include formaldehyde, paraformaldehyde, glutaraldehyde, ethanol, methanol, acetone, and commercial customized propriety formulas. Typically the alcohol based fixatives serve as both a fixative and a permeabilization agent and useful for phospho-protein labeling. Combinations of multiple fixatives or even double fixation methods can improve preservation and fluorescent signal. Glutaraldehyde can provide stabilization of protein labeling but it auto-fluoresces so it is best avoided or used a low concentration, i.e. 0.01%. The presence of auto-fluorescence can be reduced with specialized treatments or quenchers but these may bring about other problems.
5. **Post staining buffer solutions**: prevent microbial growth includes salt solution (HBSS, PBS, TBS) with 0.01% NaN_3. Be cautious as NaN_3 is toxic and can be dangerous when combined with metals or acids, so precautions are needed.

5.5.1. Optimization and development of an un-validated assay

Antibody and organelle probe selection requires researching the literature and other resources to determine a starting point of antibody or probe choice. Choosing an antibody typically involves choosing one or more antibody sources for differences in epitope recognition site or phosphorylation recognition. When choosing a probe, it is important to understand the different chemistry for binding to organelles or proteins, spectral properties of the probe, how the kinetics are altered over time, and stability of probes in live or fix end point assays.

5.5.1.1. Antibody optimization

As with other antibody based staining methods, blocking with serum from the same animal species as the primary antibody is best; an alternative is to use at least 3% w/w bovine serum albumin (BSA). As a general rule, use the recommended dilution by the supplier of the antibody or if not stated start at 1:50 dilution and dilute by 2-fold. If no signal is observed, increase the antibody concentration. Use 50 µL per well for cells seeded in 96-

well plate and dilute as necessary for other plate types; confirm the cell monolayers are completely covered with antibody solution. For example, if the supplier of the antibody suggests 1:100, use a titration scheme outlined in Table 5.

Incubate primary antibody for a minimum of 60 minutes at RT. Longer incubation times may be required to improve antibody binding or to optimize work flow process such as overnight incubation at 4^{o}C. Use a secondary antibody reporter that is well established and be sure to measure secondary antibody staining alone without primary antibody to determine non-specific binding. Include no antibody staining to measure and establish background fluorescence of the cell type used, as some cell types are notorious for autofluorescence such as liver derived cells. By lowering the concentration of the fluorescent secondary reporter, the signal to noise ratio may improve. If two or more primary antibodies are used it is important to prove the secondary antibodies do not cross-react with an inappropriate primary antibody. If the signal in the antibody labeling is weak or undetectable, the use of other enhancement techniques may be required to boost the signal such as strepavidin binding complex or tyramide signal amplification.

Table 5: Example titration scheme for primary antibody

Content	Concentration of primary antibody
Negative Control	Unstained – No primary Ab and No secondary Ab
Negative Control	Non-specific Binding - Secondary Antibody Only, no primary
Experimental Conditions	1:50
	1:100
	1:200
	1:400
	1:800
	1:1600

Probe optimization

Probes include functional dyes such as calcium indictors, liposomes, lysomsomes, mitochondria indicators,cytomplasmic, and nucleic acid probes. Dilute desired probe starting at the recommend manufacture's concentration by 2-fold; additional increase concentration by 2-fold for at least one or more concentrations for a total of not less than 5 data points. Repeat concentration curve if signal is either too weak or if saturation is reached and reduce the concentration curve less than 2-fold to "dial-in" on the optimal concentration.

Not all probes are fixable and must be analyzed using live cell imaging techniques; proper design of live cell experiments with time dependent kinetics is absolutely critical to successful outcome. When planning a live cell experiment with untested bioprobes it is important to account for the time required for an HCS imaging device to acquire cells, fields or wells on a plate. For example, if a mitochondria probe requires 30 minutes to properly load in cells and fluorescence begins to decay or results in toxicity in 2 hours, then it is absolutely necessary the image acquisition is completed within this time period.

For fixable probes, determine the stability of the fluorescent signal following fixation and analyze plates appropriately. For example, measure probe fluorescent signal at time 0 post-fixation, then measure signal at day 1, 2, 3… and so on to determine the overall stability of the signal. It is also critical to establish the stability of the light source (lamp-based, laser, LED) in the high content imager to ensure it is functioning properly during the testing period. Use a known standard fluorescent dye, cells with label or other inert material to reference daily fluorescence during the study. As a general rule, most nucleic acid probes bind tightly and are very stable but organelle probes tend to be leaky and less stable over time.

5.6. Reference Compounds and Stimuli

If there are no known published reference compounds for the assay, then use untreated control or untreated control plus vehicle (i.e., DMSO) as a baseline to normalize the data. This is an acceptable approach and may be used for orphan targets or in phenotypic screenings assays when "references" are not established.

For known reference compounds, consider the commercial availability, the stability of the reference compound, and its solubility in solvent, media, or buffer. Also determine if a specific or specialized solvent is required, and its specificity for activation or inhibition.

Refer to literature to determine potential reagents to use as stimuli for signaling pathways. A few things to consider when selecting stimuli for assay are:

- What is the signal to noise ratio and window for the assay?
- Is the stimulus biologically relevant (e.g. cell types, pathways…)?
- Is it constitutively active
- Is the stimulus specific, or does it activate multiple pathways?
- Would different stimuli for the same receptor lead to activation of different signaling pathways?
    - If so, did you select the correct stimulus for the assay of interest?
- Is more than one stimulus required for the assay or to obtain an improved S.N? And is it biologically relevant?
- Does media, sera, or growth factor used to maintain the cells activate the signaling pathway in question or alter morphology or migration?
    - If so, one may need to search for suppliers for "conditional media" or find ways to remove the stimuli.
    - If the pathway is activated by serum, it may be necessary to incubate the cells in serum-free media for a few hours prior to starting the assay.

Determine the tolerance of the cells in the assay to chemical compound solvents, such as acetone, acetonitrile, chloroform, dimethyl sulfoxide (DMSO), ethanol, ethyl acetate, hexane, methanol, tetrahydrofuran, toluene, or water used to solubilize chemical compounds. Determine the maximum concentration of solvent a cell model or cells can withstand before assay performance is altered and/or results in detachment of cells or cytotoxicity.

DMSO is the most common solvent used in biological drug discovery and many compound libraries are delivered in DMSO. In most cases, a working concentration of DMSO for *in vitro* assays between 0.1 and 1% is acceptable; however, this needs to be confirmed for every assay model. Perform DMSO tolerance by beginning at either 8 or 4%, dilute 2-fold in media used in the assay to include 8, 4, 2, 1, 0.5% DMSO, then include 0.2, 0.1, 0.05, 0.01% DMSO. Also include an untreated (no DMSO) control. DMSO tolerance is used in the assay model design by mimicking compound addition for the assay. If a stimulus is applied, include DMSO concentration curve for both un-stimulated and stimulated to determine if the stimuli has an effect on DMSO tolerance. Based on these results chemical compounds stocks for screening can be made at the appropriate concentration for maximum solubility and delivery to cell plates for assay validation process. If dose response curves will be done in the assay, all wells need to have the same final concentration of vehicle (such as DMSO) to reduce variability and eliminate artifacts caused by synergies between the vehicle and compounds at some concentrations but not others, and to prevent compounds from crashing out in the solution. Other solvents mentioned above require a similar concentration curve as DMSO and may require a larger concentration range to determine both tolerance of cells to solvent and solubility of the chemical compound.

5.7. Kinetics of the Assay

5.7.1. Assay response stability

Determine if the response for the assay is stable or prone to degradation by performing a time course experiment. This is critical in fast response assays such as using an agonist to trigger calcium mobilization or in using a stimulus to activate signal transduction pathways. The time course will be dependent on the assay type chosen. For example, calcium flux assays must be performed in live cells and measured within seconds. Alternatively, for signal transduction pathways, you must determine the half-life activation time following addition of stimuli in increments not less than 5 minutes for the first 30 minutes and increments not less than 10 minutes between 30 and 60 minutes.

Determine if inhibitor compounds can be added simultaneously or if a pre-incubation of inhibitor compound is required before adding stimulus. Be sure to determine this timing with the adaptation of automation and liquid handling devices in the laboratory. If simultaneously delivery of inhibitor compound and stimuli is not possible, determine the time course required to pre-incubate with the inhibitor compound starting at 5 minutes and increasing to 30 minutes or more as necessary to optimize work flow logistics for screening and determine if it improves signal to noise ratio. It is important in fast signaling pathways that the pH and temperature are stable during compound and stimuli additions as these can affect the biological outcome. Be sure to pre-warm stimuli plates, media and compounds to room temperature if necessary. Once a method is adopted, it is critical to maintain it throughout the validation process.

For assays with long incubation time, samples in the wells at the outskirt of the plate are frequently problematic due to evaporation or inconsistent temperature controls. To minimize impacts of evaporation, for small sample numbers, one could fill the outskirt wells with buffer or media only, and use other wells for your samples; for large scale screening, one could put a tray of water with antibiotics or NaN_3 in an incubator to provide sufficient humidity; for assays that are very sensitive to evaporation, one could put plates in small boxes padded with wet paper towels. To minimize effects of temperature fluctuations across the plate, one should make sure all plates and reagents used for the assay are warmed up to room temperature prior to the addition of cells, pre-plate the cells in the tissue culture hood, allow the cells to attach before moving the plates to the incubator; for assays that are extremely sensitive to temperature changes, one should not stack plates in the incubator.

5.7.2. Cell growth characteristics

Determine if serum withdraw or serum free conditions, or addition of supplementary cell growth components or chemicals affect cell morphology, migration, or assay endpoint in cells over time. These considerations are important in several assay endpoints including

- **Cell cycle analysis**: if cells reach confluence or if they are starved of serum or growth components, cell cycle arrest can occur, which can affect the endpoint measurement.
- **Dendrite, axon or neurite extension**: typically require growth factors from supplements or from feeder cells.
- **Cell motility and migration**: typically affected by serum withdrawal and addition of serum or growth factor supplements.
- **Signal Transduction Pathway**: serum withdrawl or low serum can increase the signal in assays such as NF-κB (see NF-κB Translocation Assay Development and Validation for HCS), MAPK kinase pathways such as ERK phosphorylation or p38 (53).

5.7.3. Live cell imaging

Not all live cell imaging assays need to be performed with an environmental controlled chamber; however, in screening operations it is important to know the challenges and difficulty to control the work flow if disrupted by

automation mishaps or other failures. If the sequence of processing plates is interrupted, this typically will result in variable in the assay data. To determine if your assay is amenable to live cell imaging conditions with or without environmental control in screening operations, you should perform a time-course study on the HCS instrument following assay treatment or in environmental-controlled conditions such as in an incubator. For example, once a bioprobe indicator completes recommended incubation time for detection, acquire images on the HCS device at time zero (t=0), then in subsequent time points over a 60 minute period, capturing images at 5 or 10 minutes intervals. There are several conditions that need to be considered when performing this operation including the image acquisition time per well or per plate and the exposure time per fluorescent probe. Photobleaching and phototoxicity are possible and may affect the results. When appropriate, use more than one plate and analyze multiple wells to measure the overall variability. If the time required to capture every well or selected wells on a plate does not exceed the degradation of the fluorescent probe measured then the assay, it can be used in screening operations as long as each plate has appropriate control wells for normalization.

5.8. Optimization

Verify methodology through documentation of all previous established assay conditions along with the DMSO or other solvent used for optimal assay performance for preparation of assay validation. Standardize and create an SOP protocol that will be referenced for assay validation. To reiterate, image analysis method should be developed alongside with the biological assay development to ensure optimal sensitivity is obtained for the desirable phenotypes.

6. Image Acquisition, Analysis and Data Interpretation

6.1. Introduction

Computer assisted image analysis is the key component to most high-content screening endeavors, since a microscope generated image can contain an immense amount of information. The goals of image analysis are simple – identify objects (usually cells), accurately measure features within, about or between these objects and extract knowledge from the features. But before an image can be analyzed a few things should be quickly reviewed.

6.1.1. Capturing a Good Image

Use your imaging platform to capture *good* representative images in an unbiased manner (see Section 2 – Image Technologies and Instruments). Verify that your workflow allows for automated capture of information about the identity of the plate being scanned (user information, time, plate ID, barcode, etc). Double check the location of well A1 to ensure the plate is not loaded in reverse. Having a designated well lacking cells or containing fluorescent beads can allow unambiguous identification of plate position. Ensure that your workflow will annotate the data with assay conditions and compound treatments that will be needed for data analysis.

Every image captured regardless of the quality generates data, good or bad, so it is imperative a *good* image is captured for subsequent analysis. Some aspects of acquiring a good set of images are predetermined by the experimental design. Images should include all the fluorescent and/or brightfield channels needed for the analysis. If the analysis will require 3-D analysis or different sample or time points in a series that are required for analysis, these images must be appropriately captured. Image fields should be taken in appropriately located, predefined positions within the sample wells (cells near the edge of the well can behave differently or the images may be distorted). Pixel resolution and magnification must also be selected to balance the level of detail vs. the number of objects (eg. cells) available for analysis and will depend on the type of objects you are setting out to analyze. Sometimes multiple fields of images from the same well must be taken to obtain enough objects with appropriate resolutions.

Two aspects that may require fine adjustment on a day-by-day basis for a given screening campaign are the focus and exposure. An out-of-focus image, even slightly, will impact the apparent size and intensity of the objects and sub-objects within the image, and can quickly increase the noise within the analysis. Most platforms provide multiple ways to auto-focus fields before taking the image, and they should be tested for their accuracy, speed and robustness. A good exposure is also very important, and should aim to optimize the *dynamic range* of the detector such as a monochrome CCD camera or PMT (photo multiplier tube). The result is a gradient scale with dark intensity in an image as black pixels and bright intensity as white pixels representing low and high numbers respectively.

Finally, do the settings used to capture an image of a neutral control field also allow other images in the experiment to be acquired accurately and with high quality? Depending on the dynamic range of the image acquisition system and the intensities in different fluorescent or brightfield channels, it may be necessary to adjust the image acquisition settings using positive or negative controls. Compromises may need to be made because it is usually preferable to have all the images within one experiment taken with exactly the same parameters rather than have to spend time normalizing after the acquisition.

In the process of taking a good image you may want try to limit the time it takes to actually acquire the images; for example, ensure that exposure times are not needlessly long or that resolution is unnecessarily high. Long exposures generally mean more time needs to be spent in image acquisition, High Throughput and High Content Screening often need to balance the amount of time taken to achieve the desired assay quality with the throughput needed for the screening.

6.1.2. Overview of Image Analysis

You have the images, so now what?

Analyzing an image with the goal of measuring features within the image or objects requires several steps (Figure 9). Once the image is acquired, it often needs to be adjusted to get the best quality by use of flat-field or post acquisition background correction. Next a threshold is applied to identify objects from background followed by segmentation to separate individual objects. These objects are often further selected based on a variety of criteria and finally the features are extracted. While the primary object identification is usually done using a nuclear or cytoplasm specific stain in one channel to identify the cell, additional objects are often identified with other stains and acquisition channels to generate additional feature data from the image. All these steps will be discussed in detail in the following sections.

6.2. Segmentation (Image Processing)

Computer based image processing has been an important part of most industries since the late sixties when the first graphics programs were developed (57). Now, image analysis and processing is a normal end point for many biological assays. Most images processing in biology has a simple goal—to separate the *signal* from the *background*. This is accomplished in a few steps that involve optimizing the image, reducing background artifacts, and then applying a threshold.

6.2.1. Notes about Images

Images come in many forms, including different file types, resolutions, color depth, pages, stacks, montages, and usually have associated metadata. Image formats vary, and different platforms provide options for how to acquire and store images. These same considerations will also be important when processing the image.

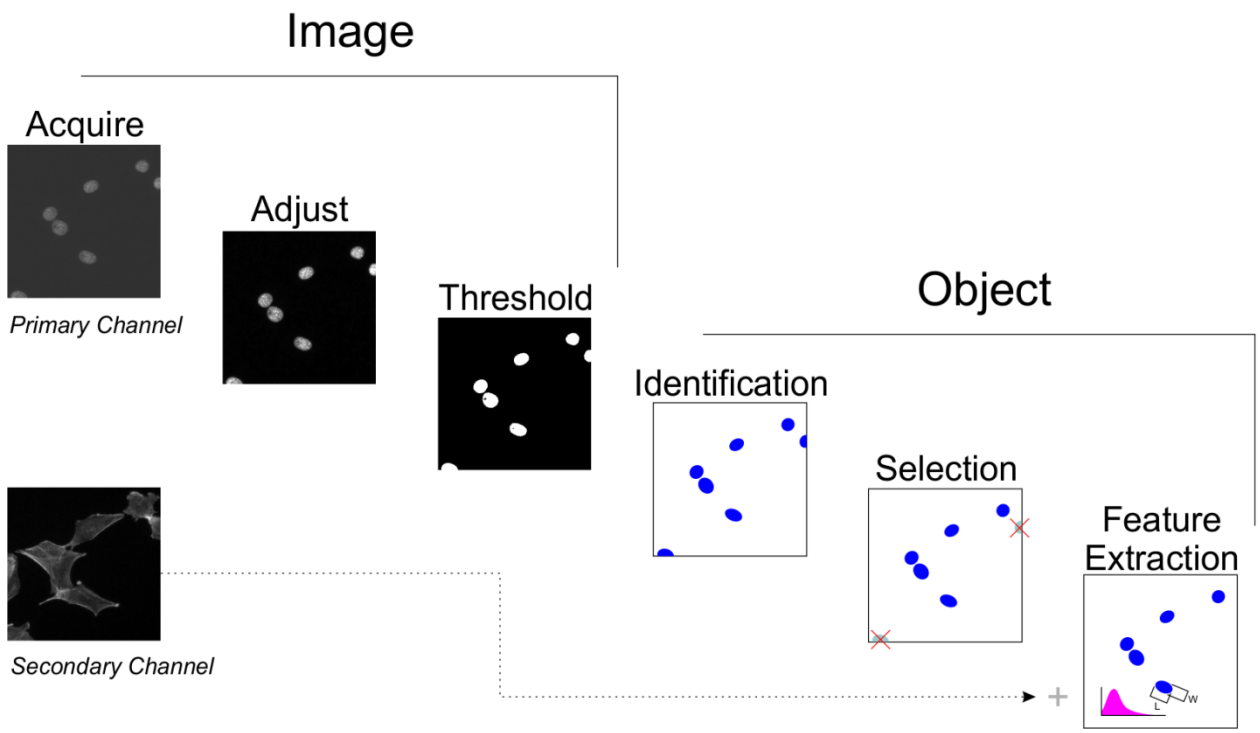


Figure 9: Typical Steps of Image Analysis. After acquisition, an image is adjusted to reduce noise and optimize the signal, then a threshold is applied and the information is converted into an object-based format. The objects are selected based on criteria and features of each object are extracted.

6.2.1.1. File Type

There currently is no universal image standard for HCS, although OME offers one such solution. Use a lossless format to work with images, such as TIF or BMP raster formats. Copies can be made in JPEG or PNG, but these formats will lose information so they should not be used as primary storage.

6.2.1.2. Resolution

Generally the image should be analyzed at the same resolution it was captured. In some situations, where the signal to noise is low, an image can be "binned" so that a group of pixels (2×2 for example) will be averaged into one pixel. This process decreases the resolution, noise and the image size for storage. It is generally more beneficial to do this during image acquisition, because it can reduce scan time, but could also be beneficial during analysis in some cases.

6.2.1.3. Color depth

Color or bit depth is a very important parameter of the image. It is the number of bits (1 and 0s) used to represent the staining intensity. The more bits associated with an image, the more shades of gray or color that can be represented. Larger bit depths expand the intensity range, allowing for the inclusion of pixels with fewer photons (previously black) or many more photons (previously white). Most professional cameras (including those used in HCS) operate in the 12 bit range, giving 4096 shades of intensity to work with, and this is generally preferred for image analysis due to practicality considerations. Larger dynamic range cameras may be used including 14-bit (2^{14}=16,384 shades) or 16-bit (2^{16}=65,536 shades). Unfortunately, the majority of computer displays (PC, Mac, Linux) are only capable of displaying 8 bits (256 shades) of information per color channel, causing some information in the acquired image to be lost upon viewing. 12-bit images will generally show up as black if you use standard software to display them. Programs like ImageJ are useful as they provide a solution to

automatically stretch the color depth to fit into the 8-bit range so it can be properly displayed, but remember that the actual image is actually more nuanced.

Users will commonly see images that are labeled 16 bit and 24 bit as well. 16 bit TIF images are a common "container" that can hold 12, 14, and 16 bit images, so if your camera is 12 bit, then these are actually 12 bit images in a 16 bit container. Similarly, 24 bit images are often combinations of the three color channels (R, G, B) each of which are 8 bit. Most image based software can read and translate these image variations.

Proprietary formats often group images together within one file. For example, TIF images support groupings of images, but this can make the TIF images complicated to work with. Often, multiple color channels will be saved in one TIF image. It is also possible to bundle multiple pages, frames of a time lapse, Z-stacks of images, or even montages of images within one image format. These types of groupings may not be natively read by image processing software and will therefore need conversion to separate and store individual images. The diversity of storage solutions adapted by instrument manufacturers makes reanalysis of images across software platforms challenging.

6.2.2. Image Optimization and Background Correction

Before an image can be processed for object identification, it often needs to be "adjusted" or optimized to achieve optimal contrast and reduce errors that would otherwise be confused as objects or alter the object tracing. There are two common "types of error" that are seen with microscopic images. These can be thought of as bias and imprecision or systematic and random. Bias or systematic errors are reproducible or predictable. Bias is an overall or local deviation in image intensity, which can be caused by variation in the output from the light source, uneven field illumination, optical aberrations, an artifact floating in the well, focus failures, or other reasons. On the other hand, all images contain some *imprecision* or random error, which we call noise. Random error arises from variations in the number of observed events (photons) stimulating a dye molecule, number of dye molecules, electrons emitted per stimulation, etc. These two classes of errors are present in any kind of measurement and imaging is no exception.

Bias correction is usually called "background correction" imaging platforms. Since background imperfections are generally low frequency (not in sharp focus), they are easily dealt with by two methods. The first group of methods takes the form of a moving average (a smoothing function) and can both reduce noise and reduce background. These functions are also called "rolling ball" methods. Background correction of this type is usually defined by a pixel radius to sample from and depends on both the size of the objects you are trying to identify and the size of the objects that tend to be causing artifacts. Slight changes in the sample size of the background function can have major effects on the image (Figure 10). For instance, if the background correction averages across 4 pixels and you are looking for features that are on average 4 pixels across, you will lose most of your sensitivity to that feature.

Filtering the image based on spatial frequency is also an effective way to eliminate out of focus background artifacts (a high pass filter will remove variations that change slowly, across many pixels, such as something that is out of focus, while retaining abrupt changes seen in objects in focus). In addition, microscope systems often provide a way of eliminating common artifacts produced in the optic pathway of the instrument. These are often termed "peripheral illumination" or "illumination" errors, and are produced as the light is channeled through the various lenses and apertures. The peripheral illumination artifacts can be easily corrected by sampling an empty plate, and using that image to compute a "flat field". This operation may not eliminate the need for additional computational adjustments.

6.2.3. Object Identification

The primary goal of image processing is to distinguish the *signal* from the *background*. Once the background has been corrected, a threshold is typically set which cuts off the pixels which are too dim (in a fluorescent image) or

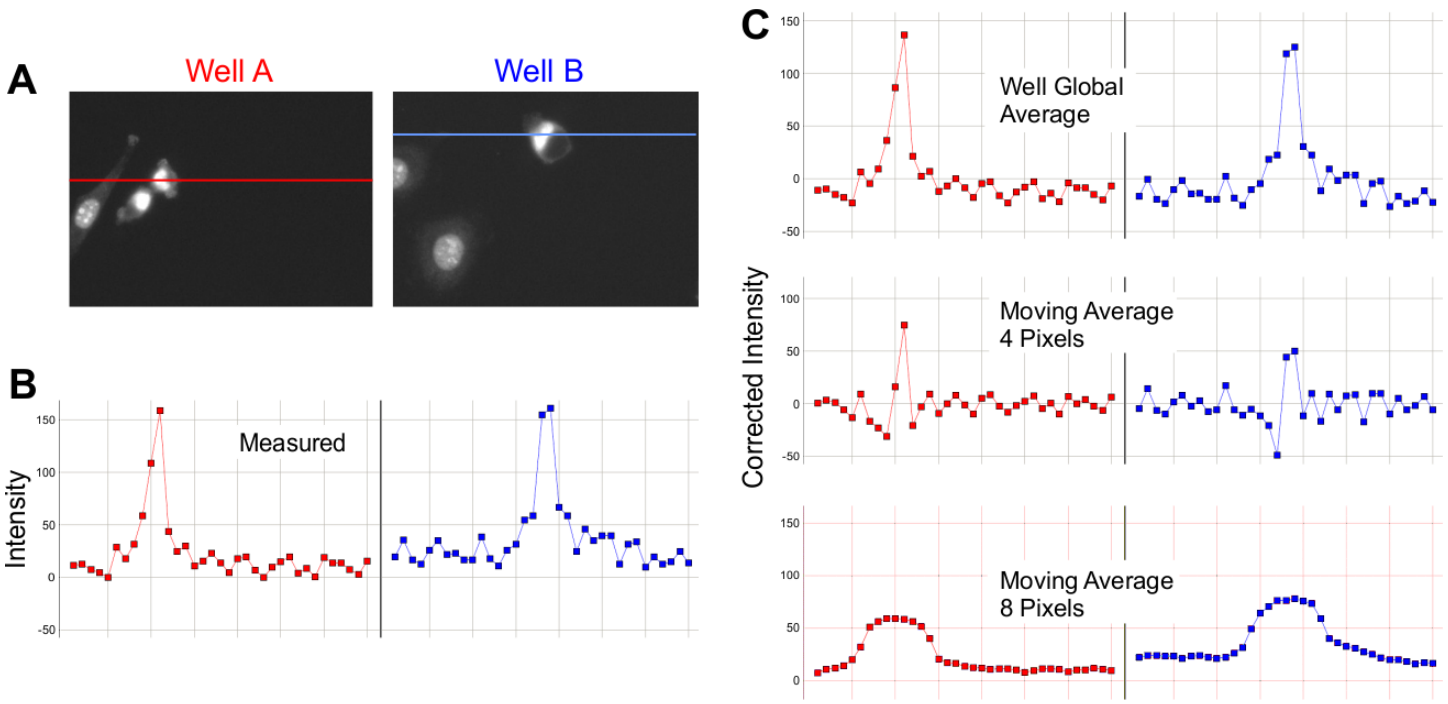


Figure 10: Background correction can have significant effects on image analysis. A) Two images from different "wells" are shown in raw grayscale, pre-corrected. A "line scan" is done across the image producing two scatter plots. B) Measuring the raw intensity for each well. C) Application of basic background correction on the two line scans. First, each point is simply subtracted from the global average of the scan. This preserves the full detail but shifts the baselines so they fall at zero. If a moving average (rolling ball) is applied with a radius of 4 pixels, the result is drastic, actually decreasing the signal-to-noise ratio and burying the peaks. An increase of the moving average radius to 8 pixels reduces the background noise and smoothes the scan.

too bright (in a brightfield image) and are thus ignored as background. Setting the threshold can be simple if images are taken with consistent exposures and with a very stable dye (Hoechst or DAPI for example). In these cases it is often possible to have a manual or fixed threshold which works across an entire plate. In other cases, more sophisticated approaches must be taken that account for changes in signal intensity with time or position.

Automated thresholding algorithms analyze the pixel intensities to determine which pixels are associated with background and which belong to the objects. If certain assumptions can be made, then these automated methods work well. The most common methods assume that the majority of the image pixels will be background, and uses an offset from the mode of the image histogram (Figure 11) to set the threshold.

The result of thresholding is a "binary" image or mask, which has only negative and positive pixels. After thresholding, image "segmentation" can divide positive pixels into separate entities or "objects". This process can be a simple algorithm which scans through the image until a positive pixel is found, then scans all connected positive pixels which are added into the first object. This process is repeated until all the positive pixels are accounted for by objects.

Another type of segmentation is often preformed either before or after object identification with the goal of splitting apart two objects that are associated with one another. This is achieved, for example, by applying a watershed algorithm on the binary image ("fills" the image with water until boundaries are established) or searching for intensity peaks or computing shape features. Other segmentation algorithms may divide the entire image into a grid for subsequent object or pixel based segmentation. Improving and developing segmentation algorithms is an active area of research.

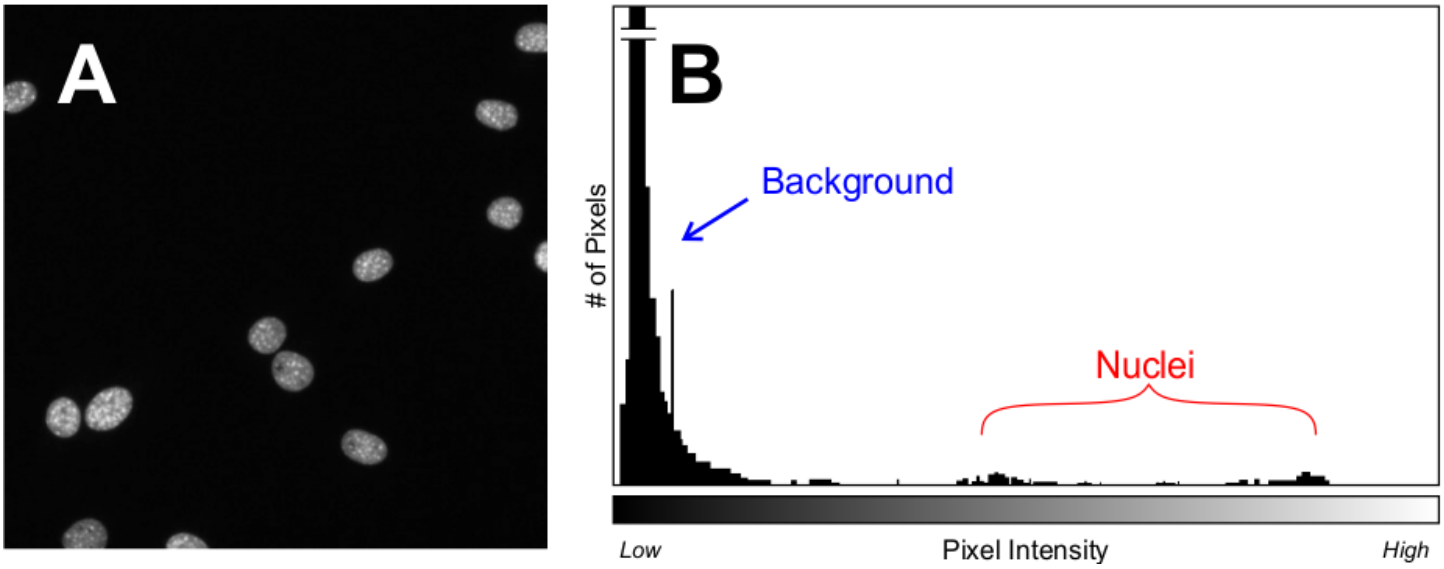


Figure 11: Image and Corresponding Histogram. A histogram plotting the number of pixels (Y-axis) which have a particular intensity (X-axis). Here, almost all the pixels are close to black, and only a small number occupy the lighter bins of the histogram.

6.2.3.1. Border or Edge Objects

Most image analysis software has options for inclusion or exclusion of objects which intersect the border of the image. These objects should only be included if complete sampling is most important (total counts for instance) and it is known that a small gap exists between image fields. If there is no gap between fields then these objects would be counted twice, which would overestimate the total count. Border objects should be excluded if information about particular object shape or structure is most important (i.e. size of cell or length of neurites).

The processing of objects provides a new set of feature data from the analytical software. Each object has many properties, i.e., shape, size, texture, and intensity that can be used in analysis. But usually objects are first "selected" to determine whether they are of interest in the analysis or just part of the noise. Object selection requires a training set or specific parameters to refine objects by their properties. For example, if objects are identified to represent cell nuclei in a homogenous culture of cells, then it is likely that the object area and object intensity criteria will be in a relatively small range, and all other objects are considered noise, debris or something else in the well. Processing can be done to identify objects similar to a known object or to identify objects that are different from a known object (i.e. training).

6.2.3.2. Secondary and Tertiary objects

To extract other features or to gain information about entities near the primary object, secondary and tertiary "sub-objects" are often identified. The simplest algorithm uses a mask or halo around the primary object at a predefined pixel width. Other algorithms use additional channels (from actin or cytoplasmic staining for example) to define the border of the secondary objects.

6.3. Feature Extraction (Object Processing)

The actual measurements generated from an image are called "features". Usually these are cell or object-based measurements like number, size, shape, intensity, texture, or kinetic measurements. Most software analysis packages provide many more features per cell than is useful to ultimately report from a high content screen and therefore in it is important to understand what the features represent and how they are derived. This understanding will allow you to choose appropriate set to analyze.

There are no standards for output data features; therefore, there remains a wide range of interpretation of the generated output features for each manufacture's image analysis algorithms. The specifics for each feature vary from platform to platform, but a few stand out for their broad use such as object counts, object size / shape and object intensity. It is important to read the manufacture's description of each feature and to remember that features can be used in combinations.

For each of these measurements, there are two basic ways to determine their inclusion in downstream analysis. First, hypothesis driven: is a particular cellular feature which is being directly measured by a feature relevant to the dataset and worthy of inclusion. The advantage is that it will always be easy to interpret this data. The disadvantage is that these hypothesized features are often *not* the most robust features for measuring differences between samples and controls. The second method looks to best compare the sample phenotype to that of the controls. Both positive and negative controls could be used for comparison.

Before starting on analysis, it is very important to think about the organization of the various forms of data. This is primarily the images, the results of image analysis, and the metadata (experimental setup, etc). Other sections of the book will discuss this in detail (see Section 8 – Data Management for High Content Screening).

6.3.1. Intensity measures

Measuring intensity should be simple, it is after all the most basic measurement that comes from the image sensor and is related to the number of photons captured on the sensor during the exposure time. The numbers attained are usually just called intensity *units*, since the camera isn't scaled or calibrated. Raw intensities are processed over some unit of area to give a meaningful value. The unit of area is a single pixel, set of pixels, or an object area (a cell or nucleus). Usually, at least two measurements are given for a particular object. The "Sum" or "Total" measurement, and an "Average" or "Mean" measurement. The Sum or Total intensity represents an aggregation of all the pixel intensities combined to make up the unit area, so these are directly affected by object size. This is also sometimes called integrated intensity. The Average intensity feature is an average accumulation of pixel intensities across an area so that the number of pixels doesn't affect the measurement as compared to Sum intensity, and therefore is a feature that is orthogonal to area measurements. There can also be intensity features related to the variation of the intensity compared to surrounding pixel intensities.

6.3.2. Nuclear features

With the appropriate fluorescent probes, nuclear stain is one of the best markers for cell identification because of its distinct edge detection and relatively uniform staining. Common features include the "Nuclear Area" and the "Nuclear Intensity", which are simple and useful calculations. Most screening paradigms should include these measures at least in quality control assessment to identify abnormal nuclei. Cell death and proliferation can dramatically impact other features, and should be considered assessed for removal from primary analysis or used in gating strategies for measuring subpopulations. In cell death, apoptotic nuclei are often smaller and more intense, while necrotic nuclei can be larger. In addition, dividing cells may have smaller or brighter nuclei. For cell cycle assays, the Total or Sum nuclear intensity is the most useful measurement for distinguishing G0/G1 and G2/M phases but not S-phase due to variability of intensity measurements, since it most closely reflects DNA content.

6.3.3. Position measurement

Each object can be tracked for its relative XYZ position within the coordinate system of the platform or relative to other features of the cell or neighboring cells. As before, careful reading of the manufacturer's description of these measures is necessary to avoid confusion. Position measurements are very important for any analysis that involves populations of cells (clusters, colonies, stem cells, population analysis). Proximity to neighboring cells has also been shown to be an important factor in predicting viral infection, neurite outgrowth, autophagy and a host of other phenomenon (40). The user should be concerned with two things – from which point is the

reference for the position measurement (the field or image, well, entire plate), and what position within the cell *itself* (upper left point, centroid, bounding box).

6.3.4. Regional analysis

Measuring specific regions or compartments within or around a cell is important for many assays. As with NF-κB translocation, for example, the cytoplasm must be distinguished from the nucleus to assess which compartment the protein is occupying. The fastest algorithms to process this information use a mask or halo, which is dilated out or contracted in from the primary object. If a nuclear stain was used a Nuclear Mask is constructed either by directly copying the primary object, or often constricting by 1 or 2 pixels. The Cytosolic Mask is then constructed by dilating a few pixels away from the nucleus and then is active over an additional width of pixels (creating a ring overlaying the cytoplasm, Figure 12). There are two important considerations when defining the width and distances of these masks. First, should the perinuclear area be part of the nuclear mask, the cytoplasmic ring, or excluded? Many biological processes (autophagy, nuclear import/export, protein synthesis, etc) take place in this perinuclear area, so its placement is often relevant. Second, how far out should the cytoplasmic ring or halo extend? Most cells have projections which extend some distances, but if this mask is being used to identify cytosol then extending the mask too far will include too much background, decreasing the value of the measurement. On the other hand, if a spot or strand (see below) is expected out in these processes, then the mask should be extended. Sometimes this outer ring is used to detect nearby objects that aren't even part of the cell, so it could be dilated quite far depending on the assay.

Cell area

Although the cytoplasmic ring discussed above produces a boundary whose area can be measured, the shape of this simple object doesn't reflect the true cellular boundary (also known as the cellular *extent*). Secondary object algorithms that use an additional channel which demark the cytoplasm, cytoskeleton, or membrane can return an accurate measurement of the cellular area. These other channels can build up objects from the pixels, where groups of similar pixels are combined into groups that are eventually defined as a particular kind of object (Figure 13).

6.3.6. Puncta identification

Spots, strands and sub-regions are important measures in many cell biological assays. Most image analysis software includes feature extraction to look for "sub-objects" or "spots" within a primary object. These are powerful measurements, but rely on consistent primary object identification and cell extents (at least approximated) to be useful. Sub-object identification usually works by setting another threshold and operates within a boundary defined by a mask (Cytoplasmic Mask or Cell Area for example). Puncta (spots or small regions) can be identified in these regions (Figure 12). Typical examples of small punctate regions that can be identified and measured are lipid rafts, ribosomes, micronuclei, mitochondria and autophagesome to name a few. Strands like actin or tubulin filaments can also be identified in a similar manner.

6.3.7. Edges

Tracing and analyzing neurites often uses different types of algorithms due to the semi one dimensional nature of these extensions. Algorithms rely on a basic method sometimes called "skeletonize". This involves converting pixels into line segments, which branch out and behave just like axons or dendrites. These lines can then pruned to remove small extensions (noise) and are analyzed for length and branching. Gap junction or tight junction formation between neighboring cells could be analyzed by the similar way.

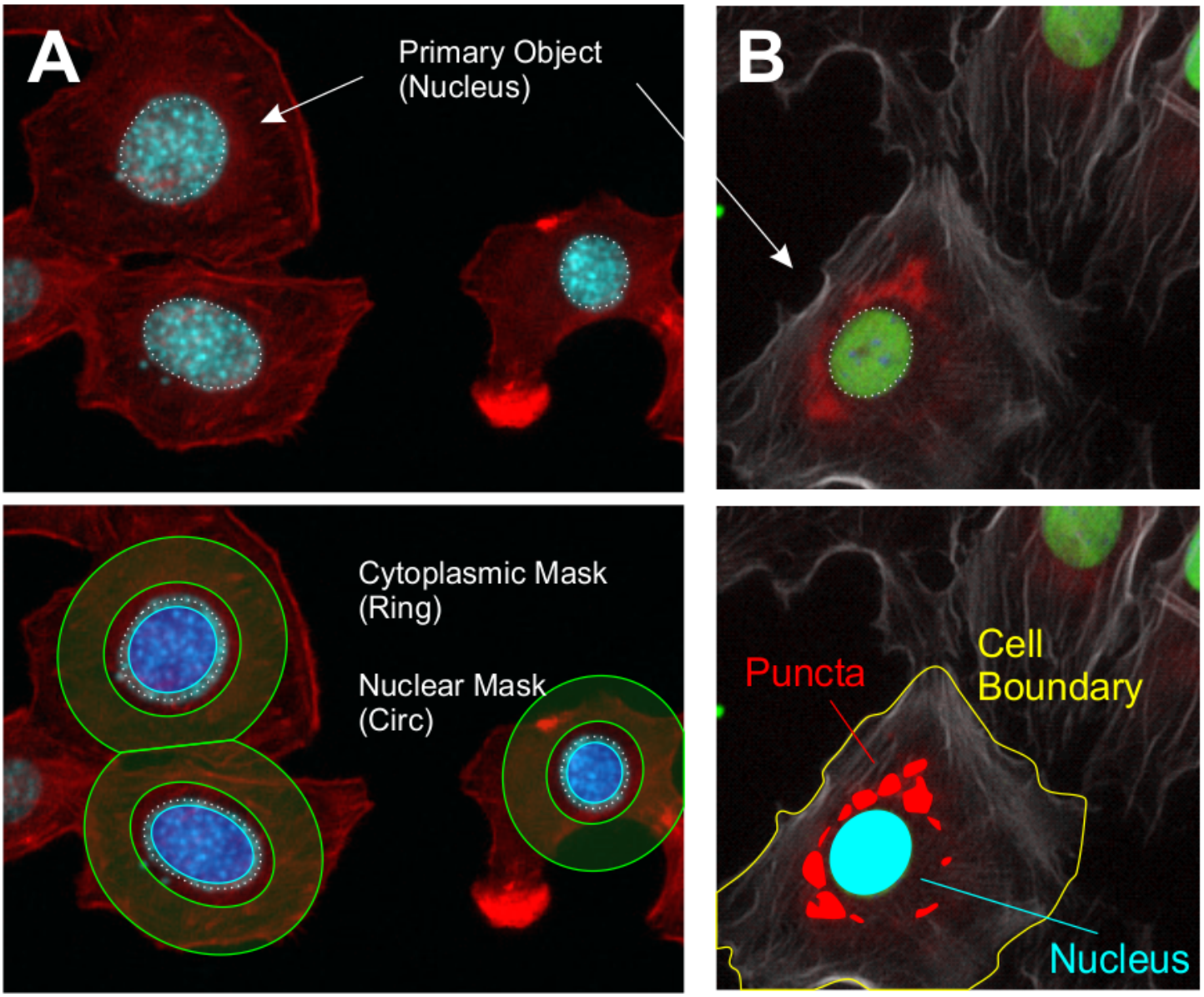


Figure 12: Regional Analysis, Masks, and Common Features. Images of cells in culture demonstrate primary object identification (dotted outline) and secondary analyses in the lower panels. A) Renal carcinoma cells stained with phalloidin (red) and Hoechst (blue) above. Below a 1 unit contracted nuclear mask (light blue outline) and a 2 unit dilated cytoplasmic ring (green). The relevance of the width of the ring is evident, since if the actin adhesion in the cell on right was an important feature, it would have been excluded from the mask. On the other hand, the masks on the two cells at the left do a good job of including cytoplasm and excluding background. B) mouse fibroblasts stained with phalloidin (gray), a nuclear protein (HMGB1, green), and an autophagic protein (LC3, red). Analysis of the cell boundary as well as cytoplasmic puncta in the red channel are displayed.

6.4. Informatics Analysis

High content screening seems simple: take good images, get good data. But once it is time to look at the data, it gets complicated. Once the images are analyzed, the results are a number of different features that quantitatively describe each and every cell present in the imaged samples. This means not only a lot of data, but a lot of multi-dimensional (multiple parameters) and hierarchical (embedded groups, cells are all related inside a well, wells are related by experimental treatments) data. This section will act as a quick description of points to think about when starting informatics analysis, and will be followed by advanced chapter(s) in areas such as machine-learning, image analysis for whole-organism, phenotypic clustering etc..

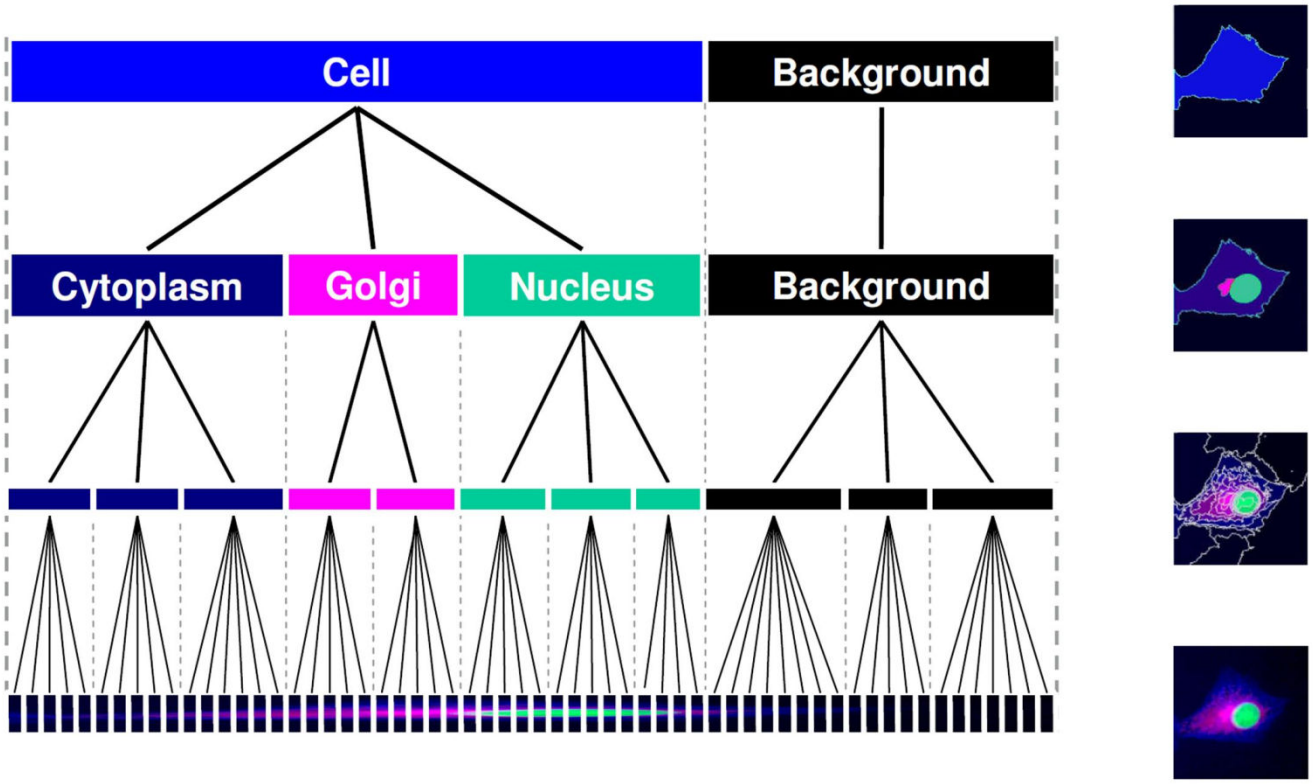


Figure 13: Pixels to objects. Using microscope channels that are derived from different cell strains, groups of pixels can be combined into a variety of objects, as in this example from Definiens: the cytosol, the golgi, and the nucleus.

6.4.1. A Brief Informatics Pipeline

An analytical pipeline (Figure 14) starts out with the raw cell-based image data. Immediately after, the metadata that describes the experimental conditions should be connected to raw data. A good next step is for a few standardized reports to be automatically generated. These reports should show well layout and plate overviews. The purpose of these reports is to get an overview of the current experiment and briefly check for large errors that can be easily spotted (such as edge effects).

A quality control / quality assurance step should be placed early in the pipeline. Many kinds of errors can occur during image acquisition or downstream analysis and common ones for the platform should be checked for in an automated or semi-automated way. Focus imperfections, incorrect exposures, background problems, artifacts, and tracing errors need to be identified so records which are affected by them can be excluded. At some point in a HCS campaign (at least in the beginning and the end, if not more frequently) images with mask overlays visible should be directly reviewed by a human observer to vet the images and ensure that tracing is correct. This process of manual image vetting can be assisted by software which let the user directly annotate the image or an attached data table with their findings (www.fastpictureviewer.com for example). Vetting the image analysis early in the screen can help to hone the algorithms and produce better data.

The data may be normalized at this point if normalization controls are built in. Some form of normalization is likely to be necessary to compare screening runs from separate days, etc.

The final steps of the analysis involve aggregating cell information together until the data can be analyzed at the treatment level (i.e. each record or row of the data table represents a particular compound or gene or condition tested in the assay). The simplest form of aggregation is to take the average (mean) or median, but one additional measure should often be included. While analyzing the cell-based data, thresholds should be set for

measurements of interest. Due to the fact that most measurements are not normally distributed, averages may produce inaccurate results. Therefore setting a threshold and aggregating the % of records above or below the threshold is sometimes preferred. Flow cytometry assays commonly call for this type of analysis called "gating". Some form of a histogram (bar, cumulative, or 2 dimensional) is often utilized to aid in setting the threshold or gate.

Cell-level data may be aggregated up to treatment level directly, or taken to an intermediate "replicate-level" first. If replicate wells were used in the experiment, then the average nuclear intensity (for example) may be average for all cells within a well to produce replicate-level data. These replicates can then be further aggregated to treatment level, allowing for a different set of statistics to be used. The reasons to aggregate to different levels are usually statistical or based on weighting.

Finally, at the treatment level, the data should be in the simplest form, but still retain deviation information so statistics can be performed. All the data, but especially the treatment level data should be stored in a database and/or exported as charts. In addition, it should be possible to easily export the primary treatment data to software (such as Excel or Spotfire DecisionSite) so that personnel have easy access to it (also see Section 8, Data Management for High Content Screening)

6.4.2. Software for Data Analysis

To choose the best software to analyze the high content data, remember that ultimately, the purpose of the analysis is to make a decision or a figure. Decisions will need to be made about whether to proceed with a particular gene or compound, or whether an assay is working. Usually a figure with statistics will be needed to convince someone else of that decision. The ideal software would allow all the different forms of data to be present (completeness), and be able to operate on them quickly (speed). Being able to change the representation or the form of the data non-destructively (dynamically) is also an ideal characteristic. Below: a list of a few solutions.

Tibco Spotfire http://spotfire.tibco.com/

Databases (Microsoft Access for example) http://office.microsoft.com/

MATLAB http://www.mathworks.com/products/matlab/

R http://cran.r-project.org/

CellProfiler Analyst http://ww.cellprofiler.org/

6.5. Image Analysis Solutions

HCS platforms are discussed elsewhere in this book (see Section 2 – Image Technologies and Instruments). Most of these vendors also produced Image analysis software that runs in real-time or just after the acquisition. But several good image analysis solutions exist that are free and open source. Two of these are listed below with examples demonstrating their basic use.

6.5.1. Free Open Source Image Analysis Software

6.5.1.1. ImageJ

ImageJ is a freely available open source, multi-platform project from the NIH. A closely related package "FIJI" (FIJI Is Just ImageJ) is usually preferred since it includes many useful modules and keeps itself up-to-date. Fiji and ImageJ are toolbox based and work much more like classic graphics software where an image is loaded, and then commands are run on it in real time and are destructive (i.e. they will change the image that has been loaded, such that if you save by accident, it would destroy the original image).

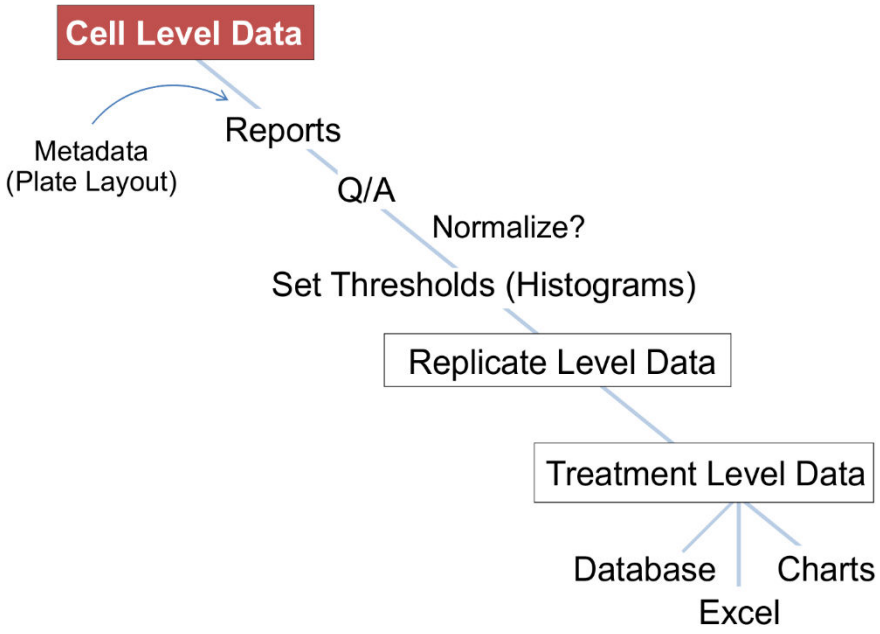


Figure 14: Informatics analysis pipeline

ImageJ can take advantage of multicore processors on most modern desktop and laptop machines. This means that programs can be written directly with multithread capabilities or that multiple scripts can be run simultaneously to greatly decrease the processing time (https://www.ncbi.nlm.nih.gov/pubmed/17936939).

Example:

Example image used are human renal cancer cells stained with Hoechst to mark the nucleus. The following case shows how to identify objects based on the nuclei.

1. File > Open Next Image (Figure 15A)
2. Background Subtraction (rolling ball)
   a Or FFT and get rid of all the really low frequency stuff (but slower)
3. Image > Adjust > Window/Level > "Auto"
4. Image > Adjust > Threshold (Figure 15B)
5. Process > Binary > Make Binary (Figure 15C)
   a Fill holes / Close (Figure 15D)
6. Process > Binary > Watershed (of touching)
7. Analyze > Analyze Particles (Figure 15E)
8. Apply Mask to original image, and other channels
9. Measure
10. Export results

6.5.1.2. CellProfiler

CellProfiler (http://www.cellprofiler.org/) is a free, open-source image analysis package that comes out of MIT's Broad Institute from David Sabatini and Polina Golland's lab by Anne Carpenter and Thouis Jones. It is a "pipeline" based tool which lets you add simple modules that work on a sequence of images {http://genomebiology.com/2006/7/10/R100}. Unlike more classic software, the modules don't run until scheduled (by clicking analyze), and they are completely non-destructive. These tools allow for quick assay design since they are already tuned for the processing of cell biological images (for the most part). Below we load an example image (Figure 16).

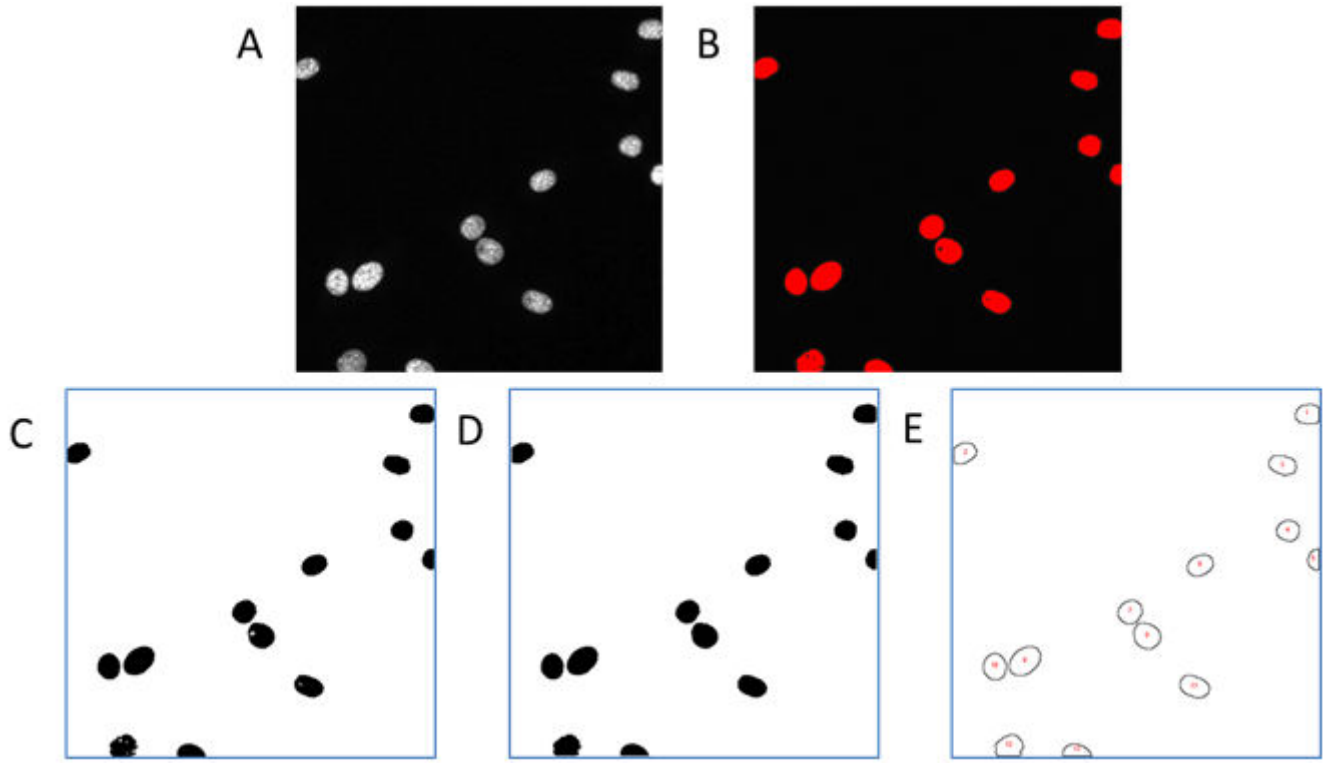


Figure 15: Example images from ImageJ analysis software. A) Representative image from Step 1. B) Representative image from Step 4 C and D) Representative images from step5 and 5a. E) Representative image from step 7.

Step 1: LoadImages – use "elsewhere" and specify a directory, enter in a part of the filename or TIF or BMP etc

Step 2: RescaleIntensity and background correction (Figure 17).

Step 3: Check "Allow Overwrite".

See Figure 17 for representative images using CellProfiler for analysis.

Step 4: ApplyThreshold

Threshold (as above) can be done in an automated way or manually. Although CellProfiler has thresholding built into its PrimaryObjectIdentification module, it is nice to do it separately so that the results of the threshold are clear. Aside from manual, Otsu, MoG (Mixture of Gaussian) and Background methods are provided (global is usually the best sub-option). Otsu is the most automated, while MoG and Background assume that the amount of background vs. foreground is known. If it is constant among the images (for example because cultures had a very constant confluence) then these will give slightly better results. The background method is similar to many classic methods which assume the background predominates in the image and uses the mode of the histogram to set the threshold. In this example (Figure 18a), Otsu global is used (with all the other defaults – Figure 18b).

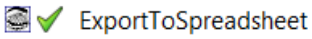
Then just have to take the measurements that you are interested in and export to spreadsheet!

MeasureObjectIntensity
MeasureObjectSizeShape

For Object Intensities – there is more than just the object mask information to consider, there is also which image should the masks be overlaid to make the intensity calculations. Here we selected the raw nuclear image,

but one could easily select a background-corrected image or even another channel. In CellProfiler, the MEAN Intensity measure is called “Intensity_MeanIntensity_” and the TOTAL Intensity measure is called “Intensity_IntegratedIntensity_”.

To measure the object size, area and shape, only the object mask is needed, so no image input is necessary. You will probably also want to uncheck the Zernicke features box, since these polynomials take a longer time.



After export, two files will appear in the default output director: “DefaultOUT_Image.csv” and “DefaultOUT_Nuclei.csv”. The first, “DefaultOUT_Images.csv” is important because it gives the list of images that were analyzed and the corresponding ID (an index) which can be used to match up additional information from the other spreadsheets.

CellProfiler has many powerful functions that are completely focused on Life Sciences research, including time lapse, worms, neuronal tracing, texture/granularity and neighbors. It is also helpful to free up some space by letting CellProfiler dispose its internal images using “Other > Conserve Memory”. CellProfiler can run very effectively on an enterprise multi-processor architecture (cluster computing), but is not currently configured to be able to run with parallel processes on a standard consumer machine.

6.5.2. Proprietary Image Analysis Software

In addition to the open source software described above, there are many proprietary image analysis software programs available. Table 6 provides a listing of some of these software programs.

Unlike the top four products, neither the Adobe nor Corel products are designed for image analysis. Since they are pervasive, many add-ons and custom scripts have been written for these to allow fairly sophisticated image analysis processes. But they are likely to require substantial more development.

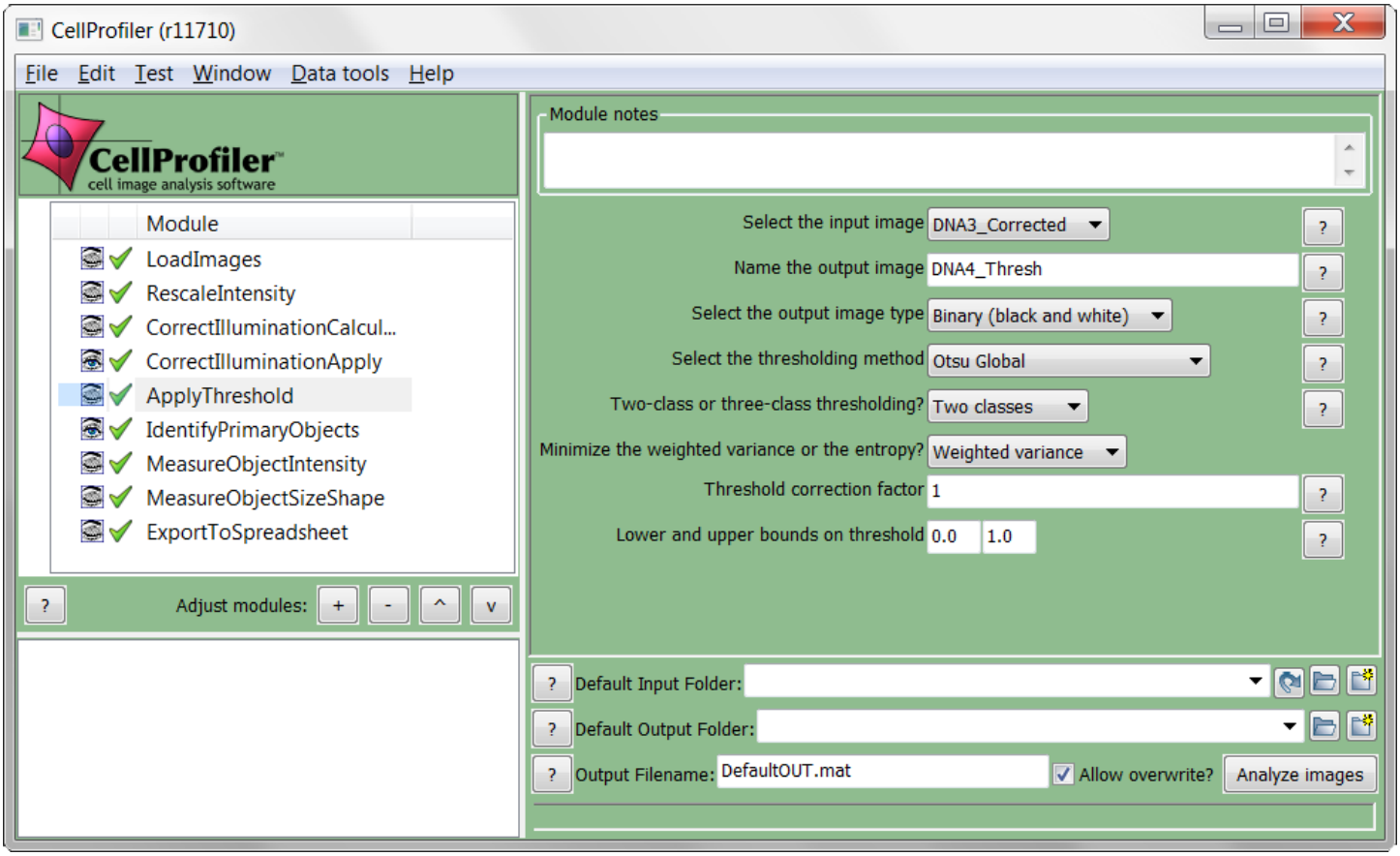


Figure 16: Example CellProfile image.

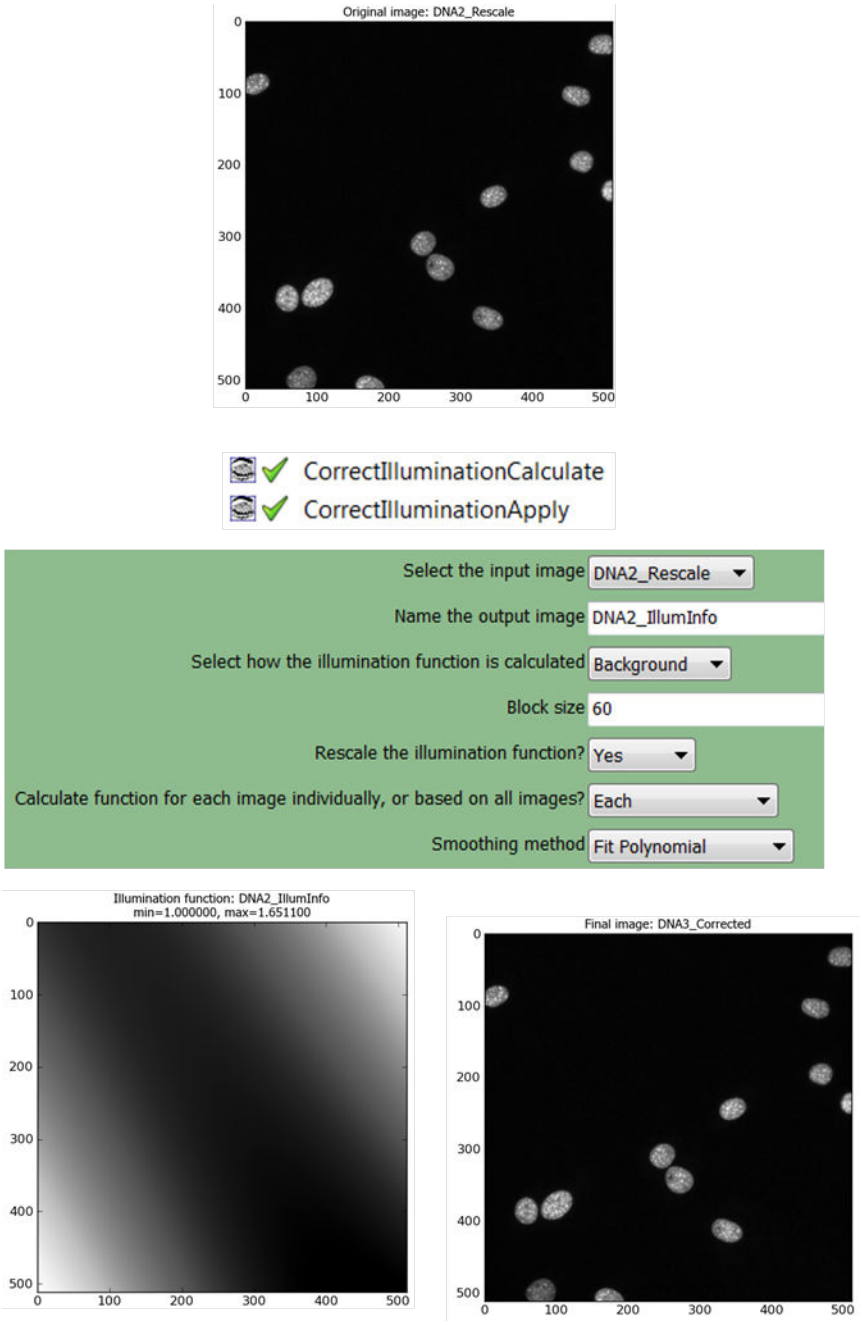


Figure 17: Representative screen captures using CellProfiler for image analysis.

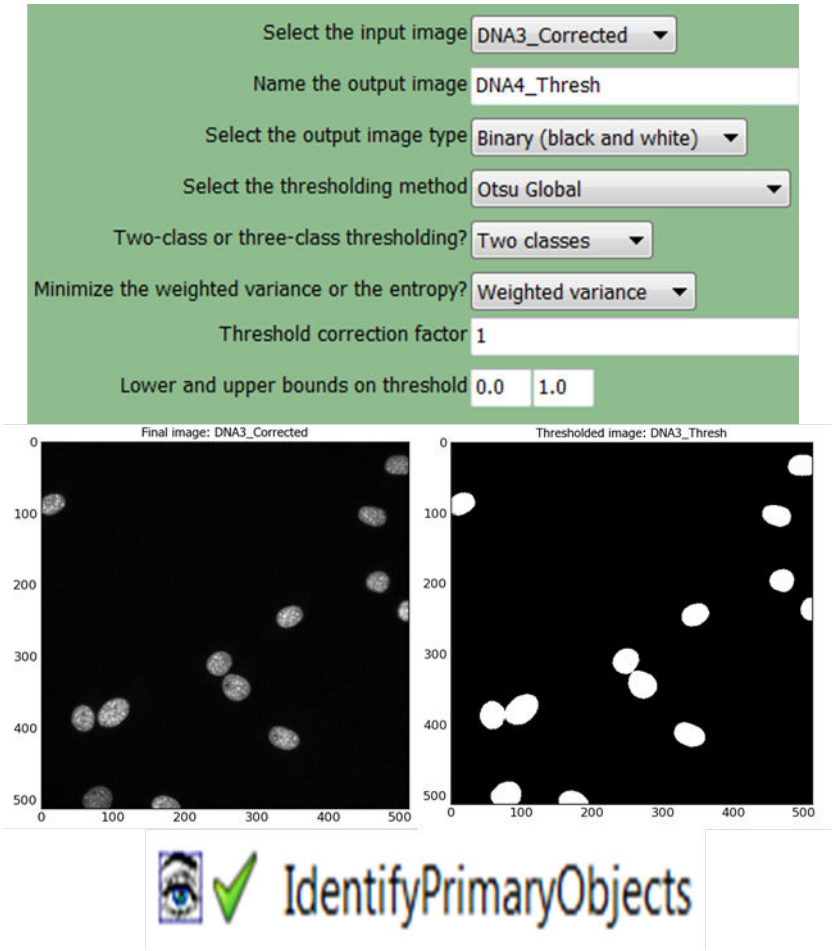


Figure 18a: Example using Otsu global

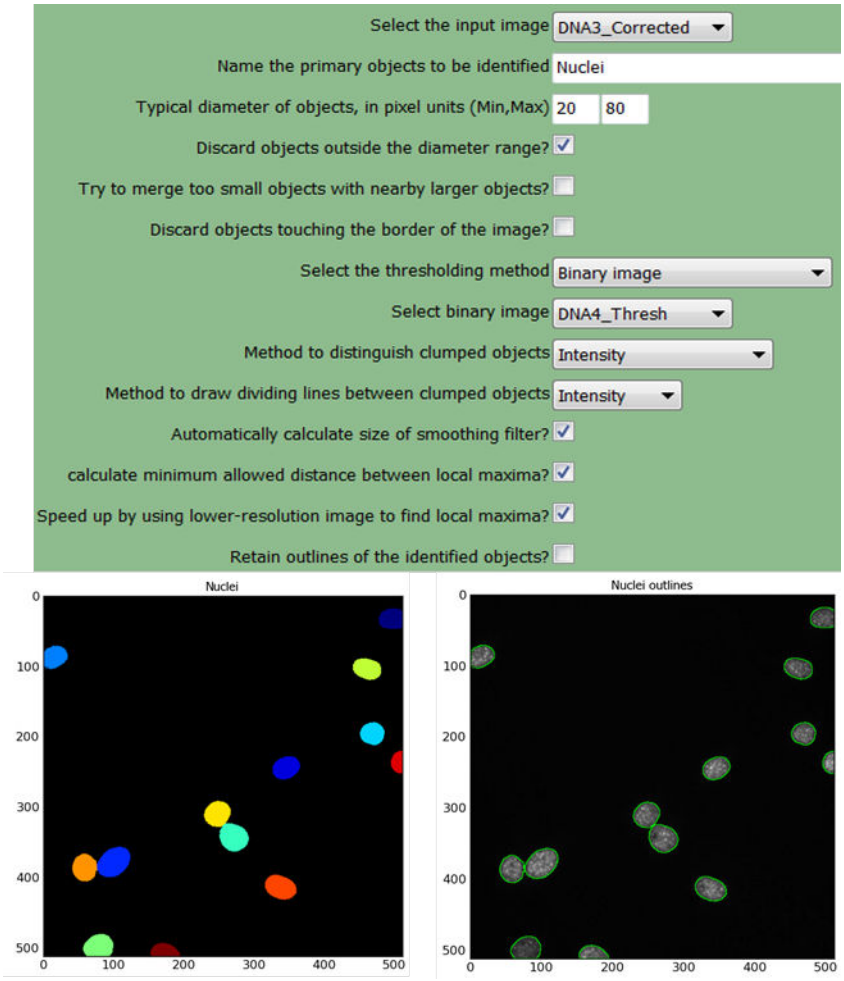


Figure 18b: Default Settings except for using the already-thresholded binary image and setting the typical diameter to match with the nuclear size of these cells.

Table 6: Examples of proprietary image analysis software

Company	Product	Website
Definiens	TissueStudio	http://www.definiens.com/
Media Cybernetics	ImagePro+	http://www.mediacy.com/
Mathworks	MATLAB	http://ww.mathworks.com/products/matlab/
Molecular Devices	Metamorph/MetaExpress	http://www.moleculardevices.com/
Adobe	Photoshop / Lightroom	http://www.adobe.com/products/photoshop-lightroom.html
Corel	Photopaint	http://www.corel.com/

7. Assay Validation for HCA Screens

This section on High Content Screen Assay Validation serves as an introduction to the topic but the reader is referred to the AGM HTS Assay Validation chapter.

HCA screens can be target defined but more often are phenotypic in nature and include measurements of dozens of features. Measured features include size, shape, intensity, texture, and dynamics. Simple examples include nuclear area and intensity, or derived measurements, such as nuclear or cytoplasmic translocation. The data is

usually based on analysis of the phenotype of single cells or objects within cells. The data can be reported for individual cells or the data from cells may be aggregated to produce data at the well or treatment level. HCA screens, unlike HTS cell based screens, often have large variability because of heterogeneous cellular populations in a given well. The data can also have large variability because the distributions of particular features are non-Gaussian when measured at the single cell level. There are many reasons for this, but the local cellular environment can have a very major effect on a given cell's response to a perturbagen (38-40, 58). There can be especially large inter-plate variability in HCA screens and this can be approached by each plate having appropriate positive and negative control wells that can be used to normalize data across plates and days (56, 59).

In HTS research plate variability studies commonly use a minimal of two signals: "Positive" and" Negative" signals. This makes the assumption that the study involves perturbagens, typically compounds that are known to be active in the assay. Many HCA phenotypic screens involve siRNAs or shRNAs to knockdown mRNA levels and subsequently protein levels or cDNA overexpression to express various proteins. In these cases it is not possible to construct dose response curves or even predict perturbagens that give "Negative" signals. Therefore standards used in HTS variability studies may not be useful in HCS assays using RNAi or over-expression approaches. During the assay development phase, it may be necessary to determine the mean response to a large number of treatments (100-1000) and then identify two or three treatments that reproducibly give a response at the level of the mean of the total set of treatments. This prescreen phase may also uncover treatments that can serve as robust positive controls. These can then be used in each plate to allow normalization across plates. The number of wells per plate in the negative reference group has a large impact on reliability of genome wide screens (56) and varies with the number of wells per plate. In 384 well plates, 16 negative control wells are acceptable and 20 or more is preferable to have acceptable false non-discovery rates (Figure 19).

There are different measurements of assay performance. Classic ones include signal to noise (S/N) and signal to background (S/B). The most widely used measurement to assess an assay is the so-called Z'-factor (60). This measurement gives insight into how much negative and positive controls are separated. The formula of Z' factor depends on the means of the positive and negative controls (μ_+, μ_-) and their standard deviations (σ_+, σ_-).

$$Z' \text{ factor} = 1 - \frac{(3\sigma_+ + 3\sigma_-)}{|\mu_+ - \mu_-|}$$

The constant factor "3*" assumes the data has a Gaussian distribution and that 3 standard deviations would encompass 99% of the values. Assays with a Z'-factor between 0.5 and 1 are considered excellent. However, often HCA data is not Gaussian and can have long tails in one or both directions. Neurite length distributions are well known for having distributions with a very long tail in one direction. An experimental approach to dealing with non-Gaussian data is to use the one-tailed Z' factor which only uses samples between the positive and negative population means (CellProfiler Statistics Module). A recently introduced measure, which is an alternative to Z-factor, is the strictly standardized mean difference (SSMD) (61).

Once an assessment measurement, such as the Z-factor, is selected, then a series of validation experiments should be performed. These include:

1. A full plate with minimum and maximum signal conditions to determine Z'-factor
2. Full plates (5-10 plates at a minimum) with minimum and maximum signal conditions to assess edge effects and other patterns of variability (pipetting, etc.) between plates
3. Full plates with a range of DMSO concentrations to assess solvent tolerance
4. Full plates with minimum and maximum signal conditions in a compound dilution scheme to ensure expected EC_{50}/IC_{50} determinations are accurate
5. Three days of EC_{50}/IC_{50} runs with reference compounds to demonstrate reproducible Z'-factor and EC_{50}/IC_{50} values. Without reproducible EC_{50}/IC_{50} values, it is not possible to do reliable SAR studies.

Edge effects are readily observed in HCA screens, so much so that one should assume they exist and explicitly test for them at the beginning of assay development. There are several strategies to deal with edge effects, the most common being media-only wells around the outside of the plate and specialized plates which have water "motes" built in (from Thermo Fisher, Nunc, and Aurora Bio, for example), or simply use a water tray in the incubator or surround assay plates with wet paper towels. Another common practice is to pre-plate the cells in the tissue culture hood, allowing the cells to attach before moving the plates to the incubator, which normalizes seeding densities on the plate (62). Position effects are important, so having controls scattered through the plate and avoiding having hits always in the same part of the plate is advantageous (also see Section 5.7.1 Assay Response Stability).

Involving statisticians during the planning process of HTS and HCA campaigns is wise. Power analyses are generally better suited for non-discovery studies; therefore, HTS approaches which seek to control the false discovery rate (FDR) (63) and balance it with the false non-discovery rate (64) are generally used (65). Finding appropriate and realistic negative and positive controls is important in this effort, and negative controls often end up being especially difficult with HCA since the treatment procedures (transfection, for example) often manipulates some of the many parameters being measured. Because screening campaigns now can involve very large numbers of perturbagens, especially compounds, special statistical methods may be called for to eliminate systemic biases introduced into an entire screen or into some plates of a screen (66)

It is also very important to understand the signal derived from a particular analytical algorithm. A common issue with HCA is to misinterpret a particular parameter due to very strong results comparing the positive and negative control. For example, many small molecule inhibitors that produce a strong effect on the parameter of interest can also diminish or enhance cell viability. Changes in overall cell health tend to have a direct impact on other measured parameters. Cell morphology changes can be due to reduced viability or increased proliferation, with many cells "rounding" up, appearing as a decreased in size. With everything else being the same, the average intensity measurements of a cytosolic marker will increase since the cell occupies less horizontal space and more vertical space. The problem can be made worse by inappropriate use of background correction or other normalization schemes. The net result is that a small effect in the parameter of interest is made to look large and significant when compounded with other variables that the investigators may not actually be interested in studying. It is a multi-parameter analysis after all, so make sure to take into account all the parameters measured.

A review of recent HCA assay development papers shows that most published HCA assays, such as nuclear translocation assays and beta-arrestin internalization, involving compound screens have Z' factors greater than 0.7. However, most HCA screens of complex biological processes, such as neurite growth, angiogenesis and tube formation, do not report Z' factors. Those studies that do provide Z' factors report values of around 0.5 or less. This is considered within the range of screenable HTS biochemical assays. However, even lower Z' factor screens can contain considerable information and a Z' factor for a single parameter around 0 may still allow hits to be identified reliably (Figure 20). Importantly, the Z' factor and most other calculations only inform the user of the strength of the positive and negative controls, and may not necessarily inform about the assay if these controls are not realistic or appropriate.

For many large scale, truly multiple-parameter HCS based compound screens, it may be difficult to validate an assay using Z' factor calculated based on only one parameter, even on a derived parameter. An advantage of HCA screens is that by combining data from multiple output parameters including ratiometric scoring it is possible to improve Z' factors from 0.3 to 0.7 (67). Figure 21 illustrates one such alternative. In this assay, each derived measurement (or classifier) alone is unable to distinguish phenotype 1 and 2 (if one projects both red and blue does onto a single axis, there will be significant overlap between the two populations). But when all 3 derived measurements are used in a three-dimensional plot, one can make a clear decision surface to separate these two populations. These advanced statistical analysis and computation need support from experienced

statisticians or informatists, and must have comparable high-speed computational infrastructure in the research facility. Future HCS chapters will discuss alternatives to the Z' factor in greater detail.

8. Data Management for High Content Screening

High throughput screening technologies by definition generate large amounts of data. Within this field, however, high content screening methods standout as they are capable of generating massive amounts of data – even when run in a non-high throughput mode. This is largely driven by the fact that a given well is characterized not by a single experimental readout (say, fluorescence) but first by a set of images with associated metadata and second, multiple numerical readouts derived from the well images. Given (uncompressed) image sizes on the order of 3 MB, a single 384 well plate imaged using a single field of view leads to 1.1 GB of image data alone. Assuming a small pilot screen of 10 plates run in single dose format, this generates 10 GB of data for that single run. While this is not particularly large, given today's storage systems, a medium sized lab can easily generate tens of such screens a month, and if dose response is considered, the image storage requirements increase by an order of magnitude. This implies corresponding increases in storage requirements including investment in software (for data management and analysis). Software costs are likely one-time investments (or will only increase slowly). It is useful to note that Open Source solutions can be employed on the software side, reducing initial costs, though of course, such solutions invariably require customization and maintenance and dedicated manpower.

8.1. Not Just Images

However, image capture and storage is just the first step in a high content screening experiment. Following imaging, the images must be registered in a repository, suitable for long-term storage and supporting efficient access by screening campaign, plate barcode and well location. Images will be processed to generate descriptors, numerical features that characterize aspects such as the number of cells, their shape, size, intensity, and texture and more complex features such as translocation of proteins, number of neurites and so on. Such analysis protocols can easily generate tens to hundreds of such features for *each* cell in a well – leading to millions of data points for a single plate. This numerical dataset must be stored and linked to the images (via plate barcode, well location and cell identifier). Finally, imaging and analysis metadata must also be recorded. This includes information such as focus settings, wavelength details, object masks, operator information and so on. These pieces of information are associated with different "levels" of the screening analysis– some are associated with the screen itself, others are relevant to individual plates or images and so on. Importantly, users may generate some metadata after the screen. Examples of this type of metadata are annotations, where a user might highlight a set of wells or even a selection of cells within a well for follow-up and include some free-text comment indicating their interest in the selection. Thus, this metadata must also be stored and linked back to images and numerical results. Any useful data management solution must be capable of supporting all these data types as well provide the flexibility to query this information in a variety of ways (68).

8.2. Image repositories

The key component of a HCS data management solution is a centralized location to collect images. In absence of a formalized management solution, the simplest approach is to simply acquire images in a file folder and inform users of folder location. This is clearly a crude and brittle approach that does not scale beyond one or two users. Invariably, folder locations may be forgotten, there is no explicit link between images and downstream data and images may not be accessible over a network easily. In addition, one must always work with the raw images, even though for many purposes (e.g., thumb nailing) they may not be required.

Most modern image repositories will employ a file system-based approach – where images are organized using a hierarchy of folders, usually located on a network-accessible storage device. But more importantly, the repository will also usually include a relational database system that records the file system path to the individual images

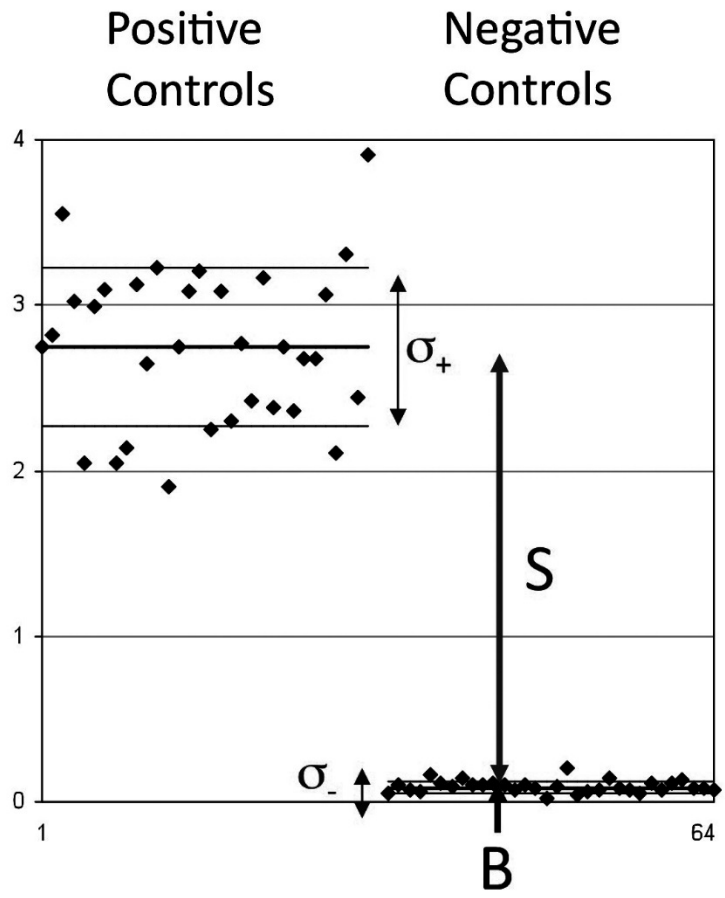


Figure 19: Standard measures of HTS assay performance:
Signal = mean of (C+) – mean (C-). Background = median (C-).
Signal to Background = S/B. σ+ = Std Dev (C+), σn = Std Dev (C-).
N = SqRt (σ2+ + σ2-)_ Signal to Noise = S/N

along with metadata such as plate barcodes, well location, imaging details (focus setting, wavelength, etc.). Import of images into a repository will usually convert them to some standardized image format specified by the system, generate thumbnail views, record meta-data and so on.

A user versed in SQL, MySQL, or Oracle can query the database to locate individual images via the database. But obviously such an approach does not lend itself to daily usage by bench scientists! To address this most vendors of image repositories will provide a graphical user interface allowing users to easily browse the image collection, searching for individual plates or wells, retrieve or archive, record notes and so on. In addition, some vendors will also include an application programming interface (API) that allows users to develop their own applications that interact with the image repository, without having to directly work with the internals (which may be subject to arbitrary changes).

Most image repositories are designed to work in an integrated fashion with a vendors imaging platform. All such repositories allow one to export and import images, though this task may not be easy. Usually, when loading images into a repository, they are converted into a common image format, such as TIFF. Some repositories may use specialized versions (OME employs the OME-TIFF format, which is a superset of the standard TIFF format).

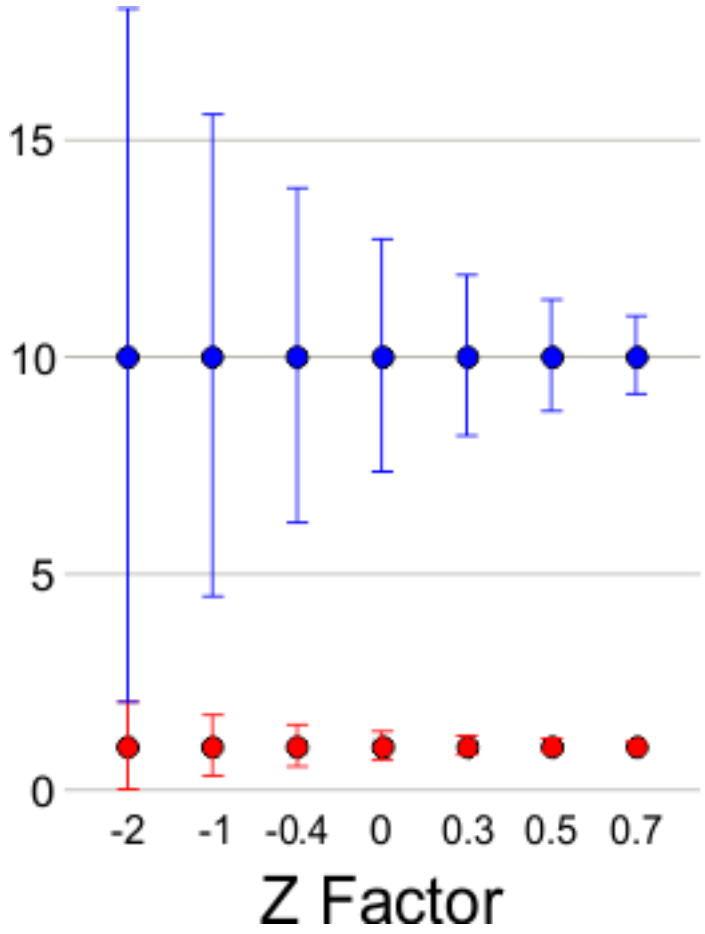


Figure 20: Z Factors comparing negative (red) and positive (blue) controls. The averages are fixed at 1 and 10, but the standard deviation is varied from 1 to 0.1 for the negative and 8 to 1 for the positive. A Z factor of -1 might still give significant results in a HCS assay.

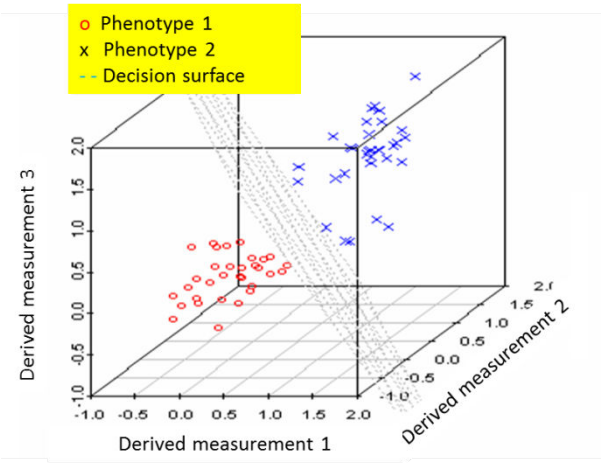


Figure 21: Validation of bioimaging based assay for primary screening using multiple parameter analysis.

The fact that the image repository is usually tightly integrated with a HC instrument's platform usually means that one is constrained to using the repository that is provided by the vendor. In other words, mixing and matching components of a HCS platform is not easy and in many cases impossible. Thus one cannot (usually) employ an image repository from vendor X and expect that the analysis or viewing applications from vendor Y

will work seamlessly. More often than not, such a mix will not work without significant investments in time and effort from the vendor (or custom development on the part of the user, requiring manual export of images from one repository and possibly manual import into another repository). For smaller laboratories, such restrictions on interoperability may not be a problem, given that they may only work with a single platform. For larger facilities, however, the lack of interoperability can become a major hindrance to the effective use of multiple imaging platforms. Recent software upgrades from several of the vendors are addressing this issue but it is an on-going concern of the HCA community.

There are certain software platforms that have been designed to work with multiple imaging vendor platforms. An example would be GeneData (http://www.genedata.com/), a commercial software package, which can access images and data stored by PerkinElmer, Thermo Fisher, Molecular Devices and OpenBIS.

8.3. Data retention policies

It is important to realize that while image repositories can be very large (greater than 50-100 TB), they are finite in size. As a result, it is infeasible to continuously add images to a repository, without some policy in place to delete or archive images. Such data retention policies are obviously local to an organization. In some cases, with low enough throughput, one can retain raw images for many years. But invariably this is not practical. A more realistic goal would be to provide sufficient space to store raw images for say two years on high-speed disk (exactly how much space would be required would depend on estimates of screening throughput) after which images would be converted to a lossy compressed form (such as JPEG). These images could be retained on the high-speed disk or else moved to a slower device such as tape. This set of data would be retained for longer periods – say 5 to 10 years. This of course will depend on the study design and format. In cases of GLP, longer-term data retention policies are mandated.

A primary role of retaining the raw images is to go back to them, say for reanalysis or visual examination. In many scenarios one can get away by conversion to a format that supports different levels of compression – allowing one to quickly access a high-resolution version or a low-resolution version from the same image. A primary example is the JPEG2000 format. One could argue that a reasonable level of image compression might not affect image analysis (though we are not aware of any benchmarks that have quantitatively measured this), and thus one could directly store images in a compressed format such as JPEG instead of the raw data (even compressed TIFFs).

8.4. Linking images and data

Handling images is obviously the core responsibility of a HCS data management system. However, images are just the first step and a standard task is to process the images to evaluate numerical features (cell counts, shape, size, intensity, etc.). As noted above, a 384-well plate can easily lead to millions of data points being generated. All this data must be stored and efficiently retrieved. In addition to numerical features, other forms of image related data such as overlays and masks must also be stored for rapid access. Most HCS management systems will make use of a backend relational database, and these are usually suitable to support the large storage requirements of high content image analyses. The use of such databases allows users to easily write back new numerical results or updated pre-existing data, say based on a new or updated calculation procedure. Obviously, it is vital that this numerical data be linked to the actual images (and even cells within an image) that they are associated with. For management systems provided by vendors, this link is always present. However, when one analyzes an image using software different from the vendor provided software, the link between numerical data, overlays, etc. and images is not present – unless somehow explicitly made. This is a bottleneck for many larger organizations that operate multiple imaging platforms, and solutions to this involve using an external data management system such as GeneData, Pipeline Pilot, or OpenBIS, or developing custom software to capture links between images and numerical analyses, overlays, compound IDs and other related information.

8.5. Commercial and Open Source solutions

All imaging hardware vendors provide a HCS data management system. In some cases, the default system may contain a small amount of functionality, sufficient for handling and viewing images of a single instrument used by a few operators. But in most cases, a more comprehensive management system will also be available. As noted above, such commercial management systems are invariably proprietary.

Depending on the scale of the HCS operation, one solution may be favored over another. Thus, a small lab, using a single instrument, is probably well served by employing the vendor provided data management system. For many such groups, the tools provided by the vendor to browse, analyze and annotate are sufficient. While we do not comment on a preferred system, most of the commercial vendors provide a capable image management solution. All of these platforms are able to export images in a variety of formats and also export numerical data obtained from image processing in standard formats (tab delimited, comma separated, etc.). In addition, many of these can be integrated with advanced visualization and reporting tools such as GeneData, SAS JMP, and Spotfire (http://spotfire.tibco.com/). However, given the high costs associated with these tools, they may be out of reach for smaller groups and in such cases it is paramount that image and feature data be exportable to tools such as Matlab, Excel, R and so on.

For larger operations that have multiple imaging platforms, the lack of interoperability between vendors is a significant bottleneck in the development of a unified interface to all the imaging data generated across platforms. While some progress has been made by commercial tools external to the imaging platform (such as Spotfire and GeneData), an integrated solution invariably requires custom development by the organization (69).

Such custom development can be impossible, when vendors employ completely closed systems and are unwilling to provide access to the internals via a public API. While such cases are becoming fewer, lack of a public API to all aspects of a vendors data management system should be considered a significant shortcoming and hindrance to integration with an organizations pre-existing informatics infrastructure.

On the Open Source side, there are relatively few comprehensive HCS data management systems. The two primary systems that are currently undergoing active development are the Open Microscopy Environment (OME http://www.openmicroscopy.org/site) and OpenBIS (http://www.cisd.ethz.ch/software/openBIS). The former system provides a comprehensive file system-based image repository, coupled with the use of PostgreSQL as the relational database that stores image locations, metadata and so on. The infrastructure provides a browser-based interface to the system and allows users to access images and perform some simple operations on them. Importantly, the OME infrastructure supports varying levels of security, allowing one to restrict images and their data to certain individuals or groups, with varying degrees of accessibility (read only, read write, etc.). A key feature of this system is that it provides a completely open, well defined API, allowing users to develop applications that interact with all aspects of the repository. This makes it much easier to integrate an OME repository with an organizations pre-existing infrastructure. It is important to note that the OME platform focuses only on image data management and not other aspects of a HCS workflow such as image analysis. However, a number of open source software packages such as ImageJ and CellAnalyst (associated with CellProfiler) can interact with an OME installation, thus enabling a fully Open Source HCS data workflow.

The OpenBIS platform is a more general biological data management platform that supports a variety of technologies including high content screening, sequencing and proteomics. In terms of functionality the system supports image export and import, metadata and annotations and also links to the KNIME workflow tool to allow integrated analyses. In addition, the system comes with a number of analysis modules built in. As with the OME platform, it exposes an API allowing users to develop novel applications on top of the OpenBIS platform.

8.6. Visualization and reporting

Efficient and robust management of imaging data is a key to ensuring reproducible and rigorous scientific studies. But equally important is the ability to interact with the data to enable mining and visualization of phenotypic data. There is an abundance of tools and techniques for the visualization of data, though visualizing datasets characteristic of high content screens requires certain capabilities. Primary among them is the ability to handle millions of data points on a plot, yet retain interactivity. Importantly, plots with millions of data points are not usually informative, so visualization platforms should have the ability to generate on the fly summaries of the entire datasets (e.g., density and contour plots, binned plots, etc.). A common visualization platform is Spotfire, which has extensive capabilities and is able to connect to a number of HCS data management systems. This allows it to integrate images with various plots – click on a data point in a scatter plot displays the image(s) associated with that data point. Such integration is vital to the analysis and exploration of high content data; in the absence of image viewing capabilities, one is faced with a mass of (usually uninterpretable) numbers. Another commercial platform for such visualization tasks is GeneData. However, Spotfire and GeneData are commercial tools and can be very expensive for smaller groups. Alternatives include Miner3D (http://www.miner3d.com/) and Tableau (http://www.tableausoftware.com/). While there are a number of cheaper or free visualization tools, they are not always intuitive to use and some (such as R) while providing very sophisticated visualization capabilities, lack easy interactivity.

It is important to note that most visualization platforms will also be tightly coupled to data mining capabilities, as one usually wishes to perform some analytical operation (clustering, predictive modeling, etc.) on the phenotypic data. While data mining of phenotypic data is out of the scope of this section we note that the platforms such as SAS JMP, Spotfire and GeneData provide extensive modeling capabilities.

Another class of application that is commonly used to interact with HCS data are workflow tools such as Pipeline Pilot and KNIME. Both tools allow non-experts to easily construct analysis pipelines, in which individual components perform specific tasks such as retrieve images from repository, perform thresholding and then calculate summary statistics over a plate. Note that such tools do not play a direct role in terms of managing HCS data, but serve to hide the data management system from the users, enabling them to interact with the data in a sophisticated manner. Some significant progressions have been made in this area (69).

8.7. Towards a Unified HCS Management System

Given the variety of vendors in the HCS field, it is not surprising that there are a many choices of HCS data management system solutions. However, given that the fundamental goal of such a system is to keep track of images, their meta-data and downstream analytical results, it is not unreasonable to desire a unified management system that allows interoperability between different vendor solutions. Unfortunately, this is currently not the case.

Given the current state of HCS management systems, any unified approach must recognize that some imaging systems will be black boxes and cannot be replaced with a common repository and associated components. One unification strategy is to implement a software interface that hides the details of individual HCS data management systems, shown schematically in Figure 22. The interface would provide access to raw images, thumbnails (if present) and associated meta-data. Depending on the scope of the repository, it may also provide read/write access to numerical data calculated from images. However, the latter is very specific to individual installations and in general the interface proposed here focuses on image repositories.

Then, new applications that are developed communicate with each system via the intermediate software interface. While this certainly allows one to have a uniform interface to multiple vendor platforms this is not an ideal situation. It is completely dependent on individual vendor platforms to provide a public, documented API, which is not always the case. Furthermore while such an interface allows an organization to develop custom

applications across all their imaging platforms, it does not necessarily allow specific vendor platforms to interact with other platforms (say, vendor platform X imports image from platform Y, performs an analysis and stores results in platform Z). In the end, this approach does not solve the fundamental interoperability problem.

What are the requirements for a unified HCS data management infrastructure? In fact very little! It is perfectly fine for each individual vendor to have a proprietary database with their own schema and formats. However, the key to supporting interoperability is the provision of a uniform API to all the data. If such an API were available, one vendor would be able to develop their applications independent of where the images and associated data are stored. One could argue that certain platforms provide certain advantages that are not available on other platforms. While this is certainly true on the hardware side, it is not clear how much vendors can (and do) differentiate themselves in terms of the actual data types that are managed by their systems. From this point of view, a set of industry standards for data management is not unthinkable.

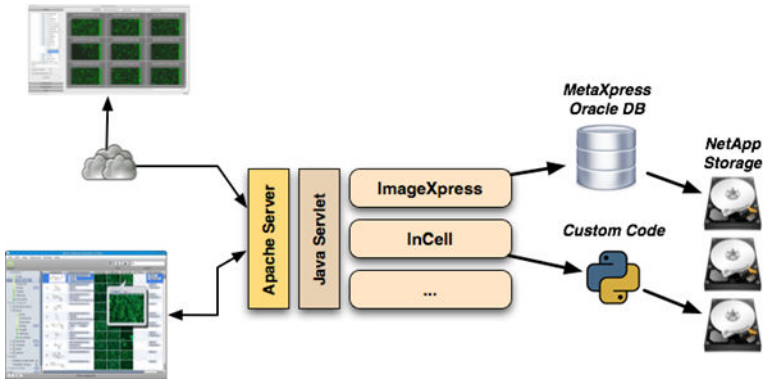


Figure 22: An overview of a software interface approach to providing a uniform interface to multiple imaging vendor platforms

9. References

1. Giuliano KA, DeBiasio RL, Dunlay RT, et al. High-Content Screening: A new approach to easing key bottlenecks in the drug discovery process. J Biomol Screen. 1997;2:249–259.
2. Pernick B, Kopp RE, Lisa J, et al. Screening of cervical cytological samples using coherent optical processing: Part 1. Appl Opt. 1978;17:21–34. PubMed PMID: 20174348.
3. Goldrosen MH, Miller GA, Warshaw WB. Comput A program for the control of an image analysis system used in in vitro assays of cell-mediated immunity. Programs Biomed. 1982;14:121–5. PubMed PMID: 7044672.
4. Rosenfeld A. Pictorial pattern recognition. Curr Mod Biol. 1969;3:211–20. PubMed PMID: 5359039.
5. Swartzman EE, Miraglia SJ, Mellentin-Michelotti J, et al. A homogeneous and multiplexed immunoassay for high-throughput screening using fluorometric microvolume assay technology. Anal Biochem. 1999;271:143–51. PubMed PMID: 10419629.
6. Liu H, Lin J, Roy K, et al. Effect of 3D scaffold and dynamic culture condition on the global gene expression profile of mouse embryonic stem cells. Biomaterials. 2006;27:5978–89. Epub 2006 Jul 7. PubMed PMID: 16824594.
7. Levin VA, Panchabhai S, Shen L, Baggerly KA. Protein and phosphoprotein levels in glioma and adenocarcinoma cell lines grown in normoxia and hypoxia in monolayer and three-dimensional cultures. Proteome Sci. 2012;10:5. PubMed PMID: 22276931.
8. Ecker JR, Bickmore WA. Genomics: ENCODE explained. Nature. 2012;489:52–66. Barroso Ines, et al. PubMed PMID: 22955614.
9. Li Z, Yan Y, Powers EA, Ying X, et al. Identification of gap junction blockers using automated fluorescence microscopy imaging. J Biomol Screen. 2003;8:489–499. PubMed PMID: 14567776.
10. Chang KH, Zandstra PW. Quantitative screening of embryonic stem cell differentiation: endoderm formation as a model. Biotechnol Bioeng. 2004;88:287–298. PubMed PMID: 15486933.
11. Lim J, Thiery JP. Epithelial-mesenchymal transitions: insights from development. Development. 2012;139:3471–86. PubMed PMID: 22949611.
12. Lagarde WH, Benjamin R, Heerens AT, et al. A non-transformed oligodendrocyte precursor cell line, OL-1, facilitates studies of insulin-like growth factor-I signaling during oligodendrocyte development. Int J Dev Neurosci. 2007;25:95–105. PubMed PMID: 17306496.
13. Mastyugin V, McWhinnie E, Labow M, Buxton F. A quantitative high-throughput endothelial cell migration assay. J Biomol Screen. 2004;9:712–718. PubMed PMID: 15634798.
14. Vidali L, Chen F, Cicchetti G, et al. Rac1-null mouse embryonic fibroblasts are motile and respond to platelet-derived growth factor. Mol Biol Cell. 2006;17:2377–2390. PubMed PMID: 16525021.
15. Moreau D, Scott C, Gruenberg J. A novel strategy to identify drugs that interfere with endosomal lipids. Chimia (Aarau). 2011;65:846–8. PubMed PMID: 22289369.

16. Schulte J, Sepp KJ, Wu C, et al. 2011, High-content chemical and RNAi screens for suppressors of neurotoxicity in a Huntington's disease model. PLoS One. 2011;6(8):e23841. PubMed PMID: 21909362.
17. Westwick JK, Lamerdin JE. Improving drug discovery with contextual assays and cellular systems analysis. Methods Mol Biol. 2011;756:61–73. PubMed PMID: 21870220.
18. Shi J, Bezabhie R, Szkudlinska A. Further evaluation of a flow cytometric in vitro micronucleus assay in CHO-K1 cells: a reliable platform that detects micronuclei and discriminates apoptotic bodies. Mutagenesis. 2010;25:33–40. PubMed PMID: 19843589.
19. Hill A, Mesens N, Steemans M, et al. Comparisons between in vitro whole cell imaging and in vivo zebrafish-based approaches for identifying potential human hepatotoxicants earlier in pharmaceutical development. Drug Metab Rev. 2012;44:127–40. PubMed PMID: 22242931.
20. Wickström M, Danielsson K, Rickardson L, et al. 2007, Pharmacological profiling of disulfiram using human tumor cell lines and human tumor cells from patients. Biochem Pharmacol. 2007;73:25–33. PubMed PMID: 17026967.
21. Xu Q, Schett G, Li C, et al. Mechanical stress-induced heat shock protein 70 expression in vascular smooth muscle cells is regulated by Rac and Ras small G-proteins but not mitogen-activated protein kinases. Circ Res. 2000;86:1122–8. PubMed PMID: 10850962.
22. Mackanos MA, Helms M, Kalish F, Contag CH. Image-guided genomic analysis of tissue response to laser-induced thermal stress. J Biomed Opt. 2011;16:058001. PubMed PMID: 21639585.
23. Lohmann M, Walenda G, Hemeda H, et al. Donor age of human platelet lysate affects proliferation and differentiation of mesenchymal stem cells. PLoS One. 2012;7(5):e37839. PubMed PMID: 22662236.
24. Mantovani C, Raimondo S, Haneef MS, et al. Morphological, molecular and functional differences of adult bone marrow- and adipose-derived stem cells isolated from rats of different ages. Exp Cell Res. 2012;318:2034–48. PubMed PMID: 22659169.
25. Giuliano KA, Johnston PA, Gough A, Taylor DL. Systems cell biology based on high-content screening. Methods in Enzymology. 2006;414:601–619. PubMed PMID: 17110213.
26. Dragunow M. High-content analysis in neuroscience. Nature Reviews Neuroscience. 2008;9:779–788. PubMed PMID: 18784656.
27. Haney SA. Increasing the robustness and validity of RNAi screens. Pharmacogenomics. 2007;8:1037–1049. PubMed PMID: 17716236.
28. Heynen-Genel S, Pache L, Chanda SK, Rosen J. Functional genomic and high-content screening for target discovery and deconvolution. Expert Opin Drug Discov. 2012;(Aug):4. PubMed PMID: 22860749.
29. Taylor DL. A personal perspective on high-content screening (HCS): from the beginning. J. Biomol Screen. 2010;15:720–725. PubMed PMID: 20639498.
30. Pierce KL, Premont RT, Lefkowitz RJ. Seven-transmembrane receptors. Nature Reviews Molecular Cell Biology. 2002;3:639–650. PubMed PMID: 12209124.
31. Garippa RJ, Hoffman AF, Gradl G, Kirsch A. High-throughput confocal microscopy for beta-arrestin-green fluorescent protein translocation G-protein-coupled receptor assays using the Evotec Opera. Methods in Enzymology. 2006;414:99–120. PubMed PMID: 17110189.
32. Baus D, Yan Y, Li Z, Garyantes T., et al. A robust assay measuring Glut4 translocation in rat myoblasts overexpressing Glut4-myc and AS160_v2. Anal. Bio. 2010;397:233–240. PubMed PMID: 19854150.
33. Zhang L, Yu J, Pan H, et al. Small molecule regulators of autophagy identified by an image-based high-throughput screen. PNAS. 2007;104:19023–19028. PubMed PMID: 18024584.
34. Yang Z, Klionsky DJ. Eaten alive: a history of macroautophagy. Nature Cell Biology. 2010;12:814–822. PubMed PMID: 20811353.
35. Delehanty JB, Boeneman K, Bradburne CE, et al. Peptides for specific intracellular delivery and targeting of nanoparticles: implications for developing nanoparticle-mediated drug delivery. Therapeutic Delivery. 2010;1:411–433. PubMed PMID: 22816144.
36. Rajendran L, Knolker HJ, Simons K. Subcellular targeting strategies for drug design and delivery. Nature Reviews Drug Discovery. 2010;9:29–42. PubMed PMID: 20043027.

37. Altschuler SJ, Wu LF. Cellular heterogeneity: do differences make a difference? Cell. 2010;141:559–563. PubMed PMID: 20478246.
38. Snijder B, Pelkmans L. Origins of regulated cell-to-cell variability. Nature Reviews Molecular Cell Biology. 2011;12:119–125. PubMed PMID: 21224886.
39. Snijder B, Sacher R, Ramo P, et al. Population context determines cell-to-cell variability in endocytosis and virus infection. Nature. 2009;461:521–523. PubMed PMID: 19710653.
40. Schauer K, Duong T, Bleakley K., et al. Probabilistic density maps to study global endomembrane organization. Nature Methods. 2010;(7):560–566. PubMed PMID: 20512144.
41. Chevrollier A, Cassereau J., Ferre M, et al. Standardized mitochondrial analysis gives new insights into mitochondrial dynamics and OPA1 function. Int J Biochem Cell Biol. 2012;44:980–988. PubMed PMID: 22433900.
42. Gautrot J, Wang C, Liu X, et al. Mimicking normal tissue architecture and perturbation in cancer with engineered micro-epidermis. Biomaterials. 2012;33:5221–5229. PubMed PMID: 22541538.
43. Samora CP, Mogessie B., Conway L, et al. MAP4 and CLASP1 operate as a safety mechanism to maintain a stable spindle position in mitosis. Nature Cell Biology. 2011;13:1040–1050. PubMed PMID: 21822276.
44. Blackmore MG, Moore DL, Smith RP, et al. High content screening of cortical neurons identifies novel regulators of axon growth. Mol Cell Neurosci. 2011;44:43–54. PubMed PMID: 20159039.
45. Xu J, Lamouille S., Derynck R. TGF-beta-induced epithelial to mesenchymal transition. Cell Research. 2009;19:156–172. PubMed PMID: 19153598.
46. Erdmann G, Volz HC, Boutros M. Systematic approaches to dissect biological process in stem cells by image-based screening. Biotechnology Journal. 2012;7:768–778. PubMed PMID: 22653826.
47. Desbordes SC, Placantonakis DG, Ciro A., et al. High-throughput screening assay for the identification of compounds regulating self-renewal and differentiation in human embryonic stem cells. Cell Stem Cell. 2008;2:602–612. PubMed PMID: 18522853.
48. Zhang J, Ferguson SS, Barak LS, et al. Molecular mechanisms of G-protein-coupled receptor signaling: role of G-protein-receptor kinases and arrestins in receptor desensitization and resensitization. Receptors Channels. 1997;5:193–199. PubMed PMID: 9606723.
49. Iwanicki MP, Davidowitz RA, Ng MR., et al. Ovarian cancer spheroids using myosin-generated force to clear the mesothelium. Cancer Discovery. 2011;1:1–14. PubMed PMID: 22586302.
50. Labarbera DV, Reid BG, Yoo BH. The multicellular tumor spheroid model for high-throughput cancer drug discovery. Expert Opinion. 2012;7:819–30. PubMed PMID: 22788761.
51. Quintavalle M, Elia L, Price J, et al. 2011. A cell-based, high content screening assay reveals activators and inhibitors of cancer cell invasion. Science Signaling. 4:183 ra49
52. Gough W, Hulkower KI, Lynch R, et al. A quantitative, facile and high-throughput image-based cell migration method is a robust alternative to the scratch assay. J. of Biomol Screen. 2011;15:155–163. PubMed PMID: 21297103.
53. Trask OJ, Baker A, Williams RG, Nickischer D, et al. Assay development and case history of a 32K-biased library high-content MK2-EGFP translocation screening to identify p38 mitogen-activated protein kinase inhibitors on the ArrayScan 3.1 imaging platform. Methods Enzymol. 2006;414:419–439. PubMed PMID: 17110205.
54. Barretina J., et al. The Cancer cell line Encyclopedia enables predictive modeling of anticancer drug sensitivity. Nature. 2012;483:603–607. PubMed PMID: 22460905.
55. Edinger AL, Thompson CB. Death by design: apoptosis, necrosis and autophagy. Curr Opin Cell Biol. 2004;16:663–669. PubMed PMID: 15530778.
56. Zhang XD, Heyse JF. Determination of sample size in genome-scale RNAi screens. Bioinformatics. 2009;25:841–844. PubMed PMID: 19223447.
57. Hiltzik, MA. Dealers of lightning: Xerox PARC and the dawn of the computer age. 1999. (1st ed.). New York: HarperBusiness. ISBN 0-88730-891-0.

58. Singh DK, Ku CJ, Wichaidit C, Steininger RJ 3rd, Wu LF, Altschuler SJ. Patterns of basal signaling heterogeneity can distinguish cellular populations with different drug sensitivities. Mol Syst Biol. 2010;6:369. PubMed PMID: 20461076.
59. Buchser W.J., Slepak T.I., Gutierrez-Arenas O., Bixby J.L., Lemmon V.P. Kinase/phosphatase overexpression reveals pathways regulating hippocampal neuron morphology. Mol Syst Biol. 2010;6:391. PubMed PMID: 20664637.
60. Zhang JH, Chung TD, Oldenburg KR. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. J Biomol Screen. 1999;4:67–73. PubMed PMID: 10838414.
61. Zhang X.D. Illustration of SSMD, z score, SSMD*, z* score, and t statistic for hit selection in RNAi high-throughput screens. J Biomol Screen. 2011;16:775–785. PubMed PMID: 21515799.
62. Lundholt BK, Scudder KM, Pagliaro L. A simple technique for reducing edge effect in cell-based assays. J Biomol Screen. 2003;8:566–70. PubMed PMID: 14567784.
63. Benjamini, Y; Hochberg, Y. 1995. "Controlling the false discovery rate: a practical and powerful approach to multiple testing". Journal of the Royal Statistical Society, Series B (Methodological) 57: 289–300. MR 1325392.
64. Genovese C., Wasserman Y. Journal of the Royal Statistical Society: Series B. Statistical Methodology. 2002;64:499–517.
65. Zhang XD, Lacson R, Yang R, et al. The use of SSMD-based false discovery and false nondiscovery rates in genome-scale RNAi screens. J Biomol Screen. 2010;15:1123–31. PubMed PMID: 20852024.
66. Dragiev P, Nadon R, Makarenkov V. Two effective methods for correcting experimental high-throughput screening data. Bioinformatics. 2012;28:1775–1782. PubMed PMID: 22563067.
67. Kummel AH, Gubler P. Integration of multiple readouts into the z' factor for assay quality assessment. J Biomol Screen. 2010;15:95–101. Gehin, et al. PubMed PMID: 19940084.
68. Collins, MA. 2009. Generating “omic knowledge”: the role of informatics in high content screening. Combi Chem & high throughput screening. 12:917-925.
69. Cornelissen F, Cik M, Gustin E. Phaedra, a protocol-driven system for analysis and validation of high-content imaging and flow cytometry. J. Biomol Screen. 2012;17:496–508. PubMed PMID: 22233649.

License