3

Linking Microscopy and High Content Screening in Large-Scale Biomedical Research

James G. Evans and Paul Matsudaira

Summary

Microscopy has been a cornerstone of discovery in the academic life sciences for more than 100 yr. This comes from a unique ability to provide extremely rich information of biological structure and dynamics. The advent of digital imaging and machine vision has brought within itself the ability to collect images more easily and critically, the ability to measure objects and intensities within images. Although many continue to use microscopy in a qualitative manner, the analytical capabilities afforded by machine vision are increasingly being applied to basic cell biology and biomedical research. Scalable quantitative imaging technology might enable scientists and engineers to determine structure, dynamics, and function of entire biological systems rather than individual molecules or pathways. This chapter will provide an overview of early efforts in the academic community to apply high content screening to the study of biological systems.

Key Words: Adhesion; high content screening (HCS); macrophage; morphology; motility; podosome; quantitative microscopy; signal transduction; wound healing.

1. Introduction

One approach to understand better the biological processes is to model them as complex dynamic systems. However, to accurately model large systems requires significant amounts of experimentally derived quantitative data. Because biological systems are regulated spatially and temporally, it is critical that the tools used to manipulate and measure these systems are adapt in the spatial and temporal domains. Microscopy has been a cornerstone of discovery in the academic life sciences for more than 100 yr. This comes from a unique ability to provide extremely rich information of biological structure and dynamics.

2. Yeast Biology

The benchmark of large-scale biology in academia remains the human genome sequencing project *(1)*. However, the minimal information content of an image-based time-resolved map of the human proteome is estimated to be nearly six log orders of magnitude greater than that contained within the DNA sequence requiring petabytes of data storage.* As previously occurred with genome sequencing, pioneering proteome-level imaging efforts are being undertaken in *Saccharomyces cerevisiae*. Initial genome-wide analysis of protein localization for about 4200 genes in *S. cerevisiae*

*Based on a 3×10^9 base sequence stored as a 30 MB ASCII file and collection of 4D image data requiring 900 TB storage (two 16-bit channels, 1024 × 1024 pixels and 150 Z slices with 50 stacks per 25 h time-series, 24,000 genes).

From: *Methods in Molecular Biology, vol. 356:*
High Content Screening: A Powerful Approach to Systems Cell Biology and Drug Discovery
Edited by: D. L. Taylor, J. R. Haskins, and K. Giuliano © Humana Press, Inc., Totowa, NJ

has been reported and although this used standard fluorescence microscopy and qualitative assessment of protein distribution, it provides a framework on which dynamic models of protein interaction may be based *(2)*. More recently a genome-wide deletion mutant screen was used to identify nine missing components of the phosphate-responsive signal transduction PHO pathway in yeast *(3)*. Similarly in the mammalian U2OS osteosarcoma cell line, large-scale automated screening has been successfully employed through individual overexpression of 7364 microarrayed cDNAs to identify genes involved in cell proliferation under a variety of growth conditions *(4)*.

It is hoped that by quantitatively determining the relative concentration and localization of each protein to a particular subcellular compartment, one might be able to constrain interactions within existing highly convoluted protein–protein interaction maps. In addition, it will likely be important to determine the degree of localization heterogeneity for each labeled protein among the population of cells.

As initial global protein mapping efforts lacked both quantitative assessment of protein localization and the variance within the cell population, several academic groups are using commercially available green fluorescent protein (GFP)-fusion libraries (Invitrogen, Carlsbad, CA) to quantify the degree of localization for proteins between subcellular compartments. Although currently available high content screening (HCS) platforms offer modest resolution (×40 magnification, 0.75 NA), initial automated screens provide relative localization to either two or three sub-compartments such as nucleus, cytoplasm, and plasma membrane. However, as it is carried out quantitatively and on thousands of cells, the statistical significance of the measurements is anticipated to be high. Higher resolution and time-resolved analyses undertaken as follow-up experiments use confocal deconvolution microscopy and specialized image analysis software, which provides additional information regarding the three dimensional localization and dynamic properties of each protein in relation to one or more marker proteins for structures of interest. To provide mechanistic information, these experiments must be combined with manipulations of the biological system. Indeed, screens underway involve use of gene knock-outs, function-blocking antibodies, metabolic constraints, and chemical libraries to quantify alterations vs the baseline behavior of key proteins.

Although spatial resolution is relatively limited in the majority of HCS instruments, compared with high-end research light microscopes, simple but highly effective primary screens may be performed in *S. cerevisiae*, a yeast that is only 3–5 microns in diameter. An example of early success is a screen that uses a rhodamine-conjugated β-glucan antibody to detect mutants defective in cell wall synthesis by virtue of increased epitope accessibility (Wheeler and Fink, in preparation). Using AlexaFluor 488-conjugated concanavalin A as a cell marker, this ×20 magnification based screen segmented cells in the field of view analyzing each cell mask for coincident localization of the β-glucan rhodamine signal indicative of defective cell wall. Populations of cells with either higher threshold average intensity or an higher average standard deviation in the rhodamine channel were scored as hits. This screen was able to find genes previously reported to be involved in cell wall synthesis such as GAS1, KRE6, and OCH1 as well as a conserved, interconnected network of genes regulating polar cell wall remodeling in both *S. cerevisiae* and the pathogenic fungus *Candida albicans* (Wheeler and Fink, in preparation).

3. Multiparametric Population Profiling

Hypothesis-free molecular cytology is an a HCS-based methodology that can be used in drug discovery, target validation, and systems biology. Essentially this quantitative imaging based approach uses a suite of multiplexed assays to screen for effects of drugs or other perturbing agents such as RNA interference (RNAi). Because a range of diagnostic antibodies and stains are used, the number of readouts is large providing a profile for each compound at a particular dose. Clustering compounds based on these response profiles, rather than similarity of chemical structures, provides a means of target identification for novel compounds by clustering with

well-characterized drugs *(5,6)*. Perlman et al. *(5)* used this approach to screen 100 compounds that included 96 of known mechanism (six were duplicated at an alternate stock concentration), three with unknown targets, and didemnin B that has several molecular targets. Compounds with similar targets clustered based on similar cellular effects. Austocystin, one compound with an unknown target, showed association with the cluster of transcription and translation inhibitors demonstrating the validity of this unsupervised approach *(5,6)*.

4. Subpopulation Dynamics in Morphology Profiling

The mammalian cytoskeleton represents several exquisitely regulated mechano-chemical subsystems that coordinately assemble to provide dynamic cellular structures in response to external cues. As cell morphology represents the sum of these subcellular structures it provides a quantifiable readout for cell function and thus a powerful tool for dissection of signaling pathways.

Previously, we have shown that multidimensional image-based analyses for cytoskeletal dynamics in macrophages can provide additional insight into the molecular regulation of adhesion complex assembly *(7)*. Microtubule perturbation affects the rate of assembly for specific adhesions called podosomes that are formed in macrophages *(7,8)*, which in turn affects cell–substrate adhesivity and cell motility. The morphological profiles for cells undergoing rapid translocation to those that are stationary, can be established using an HCS approach through correlation with similar time-resolved analyses. Indeed, using such an approach we aim to model cell shape and cytoskeletal assembly in response to a wide concentration range for the microtubule-modulating drugs paclitaxel and demecolcine. In addition, potential correlations of cell shape or cytoskeletal arrangement with nuclear morphology can be explored. The assay is being performed on a variety of cell lines including IC-21 macrophages and Src-transformed 3T3 cells *(9)*. This fixed end-point assay is performed in 96-well format and includes three fluorescent channels demarking the nucleus, cell shape, and F-actin. In early experiments we have determined that the high degree of morphological variation from cell to cell renders mean values at the population level refractive. Instead, population analyses must be carried out with single cell resolution so that subpopulation distributions may be assessed. To efficiently handle millions of cell records each with approx 100 measured features, we have adopted a *k*-means clustering approach whereby each cell is classified based on a set of compartment-based morphological descriptors for cell shape, nuclear morphology, and cytoskeletal arrangement. Using this approach we have been able to establish morphologically distinct subpopulations of cells and aim to monitor their flux in response to microtubule perturbations. Complementary time-resolved morphological analyses are likely to reveal temporal relationships between morphology profiles. And, although large-scale time-resolved HCS experiments present considerable challenges to analysis, visualization, and data management, the correctly interpreted results might simplify analysis of larger-scale fixed end-point screens.

5. Time-Resolved HCS

The majority of kinetic HCS studies completed to date have involved measurement of calcium fluxes, cytotoxicity, and receptor internalization in pharma and biotech labs. And, analyses for cell-level dynamics such as proliferation, differentiation, and motility have predominately been performed as fixed end-point assays. Indeed, the advantages in throughput afforded by fixed end-point assays and the often simpler sample preparation has led to their domination of the HCS field. The scarcity of kinetic assays might be partially to blame for the relatively poor tools currently available to analyze, mine, and visualize large amounts of dynamic data. Ultimately, it might remain unclear how significant missing data in fixed end-point assays is until equivalent time-resolved assays are performed.

The wound-healing assay is a standard method used to measure epithelial cell motility and when performed as a fixed end-point assay can be scaled to 384-well format providing a relatively high throughput means to screen for cell migration defects *(10,11)*. However, time-resolved analyses permit measurement of rate changes during closure of an individual wound as

well as the relative morphology and motility of cells in the leading edge and trailing cell sheet. Using a Cellomics Kineticscan HCS instrument, capable of maintaining normal cell culture conditions, wound-healing assays were performed for 20 h with images being collected at 15 min intervals and MatLab scripts used to calculate the rate of wound closure. Using this approach Kumar et al. compared the closure rates of two human mammary epithelial cell lines (HMEC) with differing surface expression of HER2 receptors. The 184-A1 HMEC line has approx 20,000 HER2 receptors expressed on the cell surface whereas the 24H clone has approx 600,000. Under full serum conditions both cell lines exhibit full wound closure. Under EGF, HRG, or serum free conditions; however, the HMEC cell line exhibits slow and incomplete wound closure as compared with the 24H cell line. The kinetic assay enabled a relatively simple comparison of widely differing migration rates. Although a fixed end-point approach might be less data intensive, several replicates would be required to cover a range of time-points and analyses would likely exhibit greater variation. However, the major advantage of time-resolved analyses is the potential incorporation of additional dynamic measurements such as calcium flux, receptor internalization, morphology, and motility at the individual cell level (Kumar et al., in preparation).

6. Whole Tissue HCS

An exciting recent development is the ability to survey intact tissues at subcellular resolution using custom multiphoton instruments *(12,13)*. Analysis of tissue might be performed *in situ* for tissues such as epithelia or on whole mount biopsies using a multiphoton scanner with an integrated microtome. The promise of this technology is to be able to perform HCS on cells in a physiological 3D context. This extra spatial dimension permits each cells morphological or intensity profiles to be taken in context of its relative location within the tissue architecture. One major challenge for high-resolution tissue HCS is the ability to store and manage datasets that are individually in the multiterabyte scale. Another is that despite advances in the medical fields for imaging applications such as magnetic resonance imaging, it is beyond current commodity compute technology to interactively visualize multiterabyte datasets.

7. Continuing Technology Development

HCS requires a large breadth of technologies ranging from sample preparation to data management. The academic community continues to develop critical tools in several technology areas discussed briefly below.

7.1. Sample Manipulation

Micropatterned cell arrays for delivery of nucleic acids or viruses for protein expression or specific protein knockdown through RNAi *(14)* provided a cost-effective means to screen the effects of a wide variety of conditions. The advantage of this miniaturization approach is that it not only reduces reagent costs but might also reduce data collection time because each microwell can be patterned with several overexpression or RNAi constructs. Used in conjunction with micropatterned reagent delivery arrays, electrically controlled cell sorting arrays might enable positioning of sorted cells to specific locations *(15)*.

7.2. Optics

Although increasing optical resolution is an ongoing goal for many groups *(16,17)*, the optical demands of HCS are relatively low. In contrast, the pursuit of increased sample throughput is a major emphasis of HCS. Although optical instrumentation is only one component of a complete HCS system, with considerable efforts to accelerate performance of other key areas such as image processing, analysis and data management, it is inevitable that optical throughput will become a major limiting factor to sample throughput within 5 yr. In terms of ultimate optical performance, all current imaging systems are limited to a single active objective. As compute

power can easily be scaled and applied in a distributed fashion to cope with increases in data acquisition, multiobjective parallel imaging platforms are currently being developed at the Whitehead MIT BioImaging Center to address the need for log orders of magnitude increases in sample throughput.

7.3. Data Processing, Analysis, and Management

Although hardware development remains firmly in the commercial sector, development of freely distributed image processing and analysis applications has long been available through the academic community. For small scale processing and analysis, the java-based ImageJ (formerly NIH Image) provides novice users the ability to segment and quantify cell features although this does not yet provide the necessary scalability to be effective in the HCS field. In contrast, the MatLab-based CellProfiler package (www.cellprofiler.org) was developed with large-scale analyses in mind and offers developers complete flexibility as an open source project.

Standard file formats in cellular imaging remain to be established, although many point to the success of the DICOM standard in medical imaging as a possible platform on which to build. Initatives such as the XML-based open microscopy environment (OME) and ExperiBase, although both still nascent, might become standards for general quantitative microscopy that also enable sharing of published HCS data ***(18,19)***.

The increased sample throughput of HCS, compared with standard fluorescent microscopy, enables greater exploration of parameter space. Coupled with increasing resolution afforded by multiplexed and time-resolved assays, HCS is positioned as the predominant tool to used address the complexity of cell and tissue biology in the postgenomic era.

References

1. Lander, E. S., Linton, L. M., Birren, B., et al. (2001) Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921.
2. Huh, W. K., Falvo, J. V., Gerke, L. C., et al. (2003) Global analysis of protein localization in budding yeast. *Nature* **425,** 686–691.
3. Huang, S. and O'Shea, E. K. (2005) A systematic high-throughput screen of a yeast deletion collection for mutants defective in PHO5 regulation. *Genetics* **169,** 1859–1871.
4. Harada, J. N., Bower, K. E., Orth, A. P., et al. (2005) Identification of novel mammalian growth regulatory factors by genome-scale quantitative image analysis. *Genome Res.* **15,** 1136–1144.
5. Perlman, Z. E., Slack, M. D., Feng, Y., Mitchison, T. J., Wu, L. F., and Altschuler, S. J. (2004) Multidimensional drug profiling by automated microscopy. *Science* **306,** 1194–1198.
6. Giuliano, K. A., Chen, Y. T., and Taylor, D. L. (2004) High content screening with siRNA optimizes a cell biological approach to drug discovery: Defining the role of p53 activation in the cellular response to anticancer drugs. *J. Biomol. Screen.* **9,** 557–568.
7. Evans, J. G., Correia, I., Krasavina, O., Watson, N., and Matsudaira, P. (2003) Macrophage podosomes assemble at the leading lamella by growth and fragmentation. *J. Cell Biol.* **161,** 697–705.
8. Linder, S., Hufner, K., Wintergerst, U., and Aepfelbacher, M. (2000) Microtubule-dependent formation of podosomal adhesion structures in primary human macrophages. *J. Cell Sci.* **113(Pt 23),** 4165–4176.
9. Blake, R. A., Broome, M. A., Liu, X., et al. (2000) SU6656, a selective src family kinase inhibitor, used to probe growth factor signaling. *Mol. Cell Biol.* **20,** 9018–9027.
10. Yarrow, J. C., Perlman, Z. E., Westwood, N. J., and Mitchison, T. J. (2004) A high-throughput cell migration assay using scratch wound healing, a comparison of image-based readout methods. *BMC Biotechnol.* **4,** 21.
11. Yarrow, J. C., Totsukawa, G., Charras, G. T., and Mitchison, T. J. (2005) Screening for cell migration inhibitors via automated microscopy reveals a Rho-kinase inhibitor. *Chem. Biol.* **12,** 385–395.
12. Ragan, T., Kim, K. H., Bahlmann, K., and So, P. T. (2004) Two-photon tissue cytometry. *Methods Cell Biol.* **75,** 23–39.
13. Ragan, T. M., Huang, H., and So, P. T. (2003) In vivo and ex vivo tissue applications of two-photon microscopy. *Methods Enzymol.* **361,** 481–505.

14. Ziauddin, J. and Sabatini, D. M. (2001) Microarrays of cells expressing defined cDNAs. *Nature* **411,** 107–110.
15. Voldman, J., Gray, M. L., Toner, M., and Schmidt, M. A. (2002) A microfabrication-based dynamic array cytometer. *Anal. Chem.* **74,** 3984–3990.
16. Gustafsson, M. G. (2000) Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *J. Microsc.* **198(Pt 2),** 82–87.
17. Egner, A., and Hell, S. W. (2005) Fluorescence microscopy with super-resolved optical sections. *Trends Cell Biol.* **15,** 207–215.
18. Swedlow, J. R., Goldberg, I., Brauner, E., and Sorger, P. K. (2003) Informatics and quantitative analysis in biological imaging. *Science* **300,** 100–102.
19. Dewey, C. F., Jr., Downes, A., Chou, H., and Zhang, S. (2005) ExperiBase—an object model implementation for biology. *Proceedings of the Eighth Annual Bio-Ontologies Meeting*, June 24, 2005.