Universiteit
Leiden
The Netherlands

Development of automatic image analysis methods for high-throughput and high-content screening

Di, Z.

Citation
Di, Z. (2013, December 10). *Development of automatic image analysis methods for high-throughput and high-content screening*. Retrieved from https://hdl.handle.net/1887/22833

Version:	Corrected Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/22833

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/22833 holds various files of this Leiden University dissertation.

Author: Di, Zi
Title: Development of automatic image analysis methods for high-throughput and high-content screening
Issue Date: 2013-12-10

Chapter 1

General introduction

1.1 Introduction of the thesis

This thesis focuses on the development of image analysis methods for ultra-high content analysis of high-throughput screens where cellular phenotype responses to various genetic or chemical perturbations that are under investigation. Our primary goal is to deliver efficient and robust image analysis platforms which can 1) automatically detect cellular structures of interest from florescence microscope images and 2) quantify dynamics and organization of multi-cellular systems with phenotypic features. To recover heterogeneity of cellular behavior, we aim to develop single-cell-based image analysis methods so that cell subpopulations can be distinguished and investigated. Furthermore, we intend to develop methods to extract an ultra-high level of phenotypic details from images. This would enable system-level studies of phenotype characterization.

To promote a further understanding of this thesis, this introductory chapter firstly provides the general background and essential knowledge related to the high-throughput and high-content screening. Next, the state-of-the-art techniques and image analysis methods that have been already applied to high-throughput and high-content screenings are given. Finally, the scope and structure of this thesis is presented at the end of this chapter.

1.2 High-throughput and high-content screening

High-throughput and high-content screening is a phenotypic screening technique that utilizes automated microscope systems to identify the functional role of substances such as small molecules, peptides or RNA interference (RNAi), in the context of a pivotal, pathology relevant, cellular process, ultimately enabling the identification of drug targets and/or novel drug molecules. To prepare the screening, cells are cultured in 96- or 384-well format micro plates and treatments with substances are applied. Changes in cell phenotype can be visualized by labeling structures and molecular components of the cells with fluorescent dyes, made visible by immunofluorescence

methodologies, or by expression of green fluorescent protein (GFP)-tagged proteins. Finally, micro plates are imaged by an automated microscope system to monitor cellular response to the different perturbations.

1.2.1 "High-throughput"

In recent years, florescence microscopy technology has been developed dramatically in respect to resolution, speed, complexity and scale. The emergence of automated microscope systems with robotic handling enables the investigation of a large volume of compounds or genetic players simultaneously. Combined with genome-wide RNAi approaches [1, 2], high-throughput small-molecule-based perturbations [3] or over-expression strategies [4], high-throughput and high-content screening has become a powerful technology to thoroughly study the regulation of biological pathways that underlie the function of intact cells.

1.2.2 "High-content"

Although other high-throughput techniques, such as mass spectrometry or DNA-microarrays, have been developed and successfully applied to study diverse cellular pathway and their possible involvement in disease, these techniques, despite their great usefulness, cannot provide adequate temporal and spatial information , in the context of the structural and functional integrity of cells. Most importantly, they do not directly show whether the identified molecules have a functional role in the cellular process that is under investigation. High-throughput and high-content screening technique fills this gap by probing the function of molecules in their nature environment with exquisite and ever increasing spatial and temporal resolution [5–8]. Moreover, integrated with other high-throughput techniques, interdisciplinary information is collected, which provides us a challenging opportunity to study whole biological systems more comprehensively.

1.3 Workflow of high-throughput and high-content screenings

High-throughput and high-content screenings consist of four major stages: sample preparation, image acquisition, image analysis and data analysis. In this chapter, a general introduction for each of these stages is provided and the state-of-the-art techniques that are used in each stage are summarized.

1.3.1 Sample preparation

According to the biological question, different types of assays can be used for screening. In general, they are categorized in two groups: fixed assays and live-cell assays. For the studies where only the final status of cells needs to be investigated, fixed assays are the primary technique to be considered. The most common one is a 2D

fixed assay. In this assay, cells are grown as a cell monolayer (therefore the term "2D" is used) and only a simple and specific read out is measured. Because this type of assay is relatively easy to be processed and imaged, it is often considered for higher-throughput experimentation and larger scale automation. Many projects have already achieved a great success using this type of assay [1, 3, 7, 9, 10]. One example is the work of Marino Zerial and his co-workers [1]: a genome-wide RNAi screen was performed in HeLa cells to explore the function of human kinases in two principal types of endocytosis: clathrin- and caveolae/raft-mediated endocytosis. They showed that a high number of kinases is involved in endocytosis, and that each endocytosis route is regulated by a specific kinase subset.

However, 2D fixed assays are not representative of the cellular environment found in organisms. In fact, tissue-specific architecture, mechanical and biochemical cues, and cell–cell communication are lost under such simplified and highly biased conditions. Efforts to address this problem led to the development of 3D cell cultures. They utilize an extracellular matrix (ECM) gel to re-establish physiological cell–cell and cell–ECM interactions, thus mimicking spatial organization of real tissues in their nature environment better than 2D cultures. Recently 3D cultures have shown many advantages in a broad range of cell biology studies [11, 12], including tumorigenesis [13, 14], cell adhesion [15–17], cell migration [18] and epithelial morphogenesis [19]. Many researchers also start to establish 3D assays for high-throughput and high-content screening, for example, in the study of tumor cell migration and invasion [20]. Two 3D screening assays based on mouse and human breast cancer cells are extensively described in **chapters 3 and 5** of this thesis.

Another type of assays are live-cell assays which are used to study the dynamics of cellular processes. It requires automated microscope to monitor these processes and often fluorescence tagged protein are used that are continuously imaged over a certain period of time. This allows the collection of much more detailed phenotypic information, especially temporal information is provided that otherwise cannot be obtained. In addition, live-cell assays can reveal primary defects and secondary consequences of the phenotype and thereby allow a more precise interpretation of the function of the molecules that are under investigation. This type of assays has been widely used in the study of embryogenesis [2], cell division [21], and intracellular translocation of molecules [22, 23], which so far are the most significant achievements among all high-throughput and high-content studies. These type of studies are based on the discrimination between the cell membrane, the nucleus and the cytoplasm, and the translocation of fluorescently labeled molecules between the distinct compartments. One example is the study of NF-κB nuclear translocation. Many studies on this subject have been published in recent years, notably the work of Covert and co-workers [24, 25]. They developed a screen platform with single cell resolution that can image and determine the NF-κB activation over time in a high-throughput manner. In **chapter 2** of this thesis, a novel NF-κB screening platform is described that also measures NF-κB activation dynamics at a single cell level but is able to apply for high cell density as well. Furthermore, it is able to quantify the characteristics of all single cell dynamics for further understanding of cell-to-cell heterogeneity.

1.3.2 Image acquisition in a high-throughput manner

One of the characteristics of high-throughput and high-content screening is that a large amount of images needs to be acquired. Therefore image acquisition systems that are fully automated and time efficient are required. Most researchers choose to use commercial systems that are already available on the market [7, 26, 27]. In this chapter, the techniques that are relevant in the context of this thesis are summarized.

1.3.2.1 Microscope models

Most high-throughput and high-content image acquisition systems are equipped with wide-field fluorescence microscopy. For this type of microscopy, cellular structures of interest are tagged with fluorescent proteins and exited by light of a specific wavelength. After absorbing energy from excitation light, the fluorescent proteins emit light of longer wavelength that is captured by a detector to generate images. Confocal microscopy is another frequently used and more advanced fluorescence microscopy method. It adds a pinhole aperture in front of the light source so that light excites only one optical plane (focal plane) at a time. In addition, another pinhole is added in front of the detector to filter out the emission light generated from the plane above and beneath the focal plane. Therefore, images from confocal microscopy contain only sharp in-focus information from the focal plane. This is especially useful when a specimen is relatively thick and a series of optical sections need to be acquired through the specimen. The disadvantage is the relatively slower imaging process and longer exposure time of the specimen. Especially when a large number of sections needs to be imaged, severe photobleaching can occur and that presents an as yet unresolved problem.

The new generation of high-throughput screening systems starts to explore two-photon excitation microscopy [28], spinning disc microscopy [3] and super resolution microscopy [29]. Unfortunately, due to their relatively higher demand for both hardware and software, only few researchers have so far developed and applied them successfully in high-throughput and high-content screening.

1.3.2.2 Magnification, resolution, sampling size and camera setting

Objective lenses are the most important component of an optical system because they are the predominant factor that defines image quality. In general, objective lenses can be classified based on their magnification and resolution. High-throughput and high-content screening systems are typically equipped with$<10\times$, $20\times$ and $40\times$ magnification lenses. Different magnifications are applied according to the structure of analysis. Measurement of structures over large areas, such as cell networks or zebrafish embryos, requires lower magnification than measurement of sub-cellular structures, for example, nuclear repair foci in a DNA damage and repair assay.

Image resolution (which refers here to spatial resolution) is a term used to describe how closely two objects can be resolved in an image and is directly determined by numerical aperture (NA), which is a number describing the amount of light coming

from the focus that the objective can collect. According to the Rayleigh criterion, the relation between NA and resolution is formulated as:

$$r = 0.61 \times \frac{\lambda}{\mathrm{NA}} \tag{1.1}$$

where λ is the wavelength of emitted light. For a fluorescence microscope of NA = 0.5 and λ = 500nm, this results a resolution limit of 610nm. Two distinct objects closer than this resolution will be imaged as a single object. Because in high-throughput and high-content screening image acquisition is often done (as well as in this thesis) with 4× or 10× lenses with NA equivalent to 0.13–0.3 (corresponding to resolution limits of 2346nm–1017nm at λ = 500nm), the objective lenses are the limiting factor in determining which sub-cellular structures can still be imaged.

Sampling size (also refers to sampling frequency) often refers to the distance between 2 signal-recording points. According to the Nyquist rate, image sampling size in xy-direction bigger than half of the objective lens resolution would result in a loss of information. The pixel size of the camera used often determines whether this can be achieved (the width of each pixel in the camera chip divided by the magnification gives the size of each pixel in the sample plane). Combining pixels in the camera (pixel binning) can increase the pixel size and signal intensity. The benefits of this setting are that exposure time and image file size can be reduced significantly, but the disadvantage is that spatial resolution will be decreased.

1.3.2.3 Balance between quality and quantity

Image quality is directly related to magnification, resolution, sampling size, pixel binning and other elements. Higher magnification, smaller sampling size or no pixel binning could increase the image quality and this could make complex and computational intensive image analysis methods redundant while much more detailed and accurate information can be extracted. However, better image quality sacrifices imaging time, processing time and requires more data storage. For example, reducing z sampling size from 10μm to 5μm not only doubles the size of the image file, but also doubles the time of imaging. Moreover, the limitations of hardware and software often make it impossible to deal with extremely large amount of data. For example, image files bigger than 1.5 gigabyte cannot be opened and analyzed by ImageJ on a 32-bit Windows operation system. Therefore finding the balance between image quality and quantity is very critical for high-throughput and high-content screening.

1.3.3 Image analysis

Once images are acquired, the next stage of screening is image analysis. Image analysis is a process to extract numerical information from images that are representative of cellular phenotypic responses. Those numerical parameters are then used for further data analysis, such as hit identification and compound characterization. Image analysis often comprises two general steps: segmentation and quantification.

1.3.3.1 Segmentation

Segmentation is a technique of defining regions of the images that contain signals from the molecules which are under investigation (also called region of interested or ROIs). Those molecules are often fused with fluorescence protein or antibodies, or stained with fluorescent dyes, and as a result, regions that contain these molecules present higher signal intensity in images than other regions. This difference of intensity between ROIs and background is used for most segmentation algorithms. The key of those algorithms is to automatically determine an intensity threshold to distinguish ROIs and background, or sometimes multiple-levels are selected. There are other advanced segmentation methods that are model based or using prior knowledge, for example active contour model [30, 31]. However, considering their complexity and specificity, they have not been widely applied in high-throughput and high-content screening.

Because quantification is often made based on the segmentation result, proper segmentation is the key to generate meaningful data and must be optimized for every screening. Several commercial systems already combine image acquisition and analysis [7], for example, the BD pathwayTM bioimaging system, KineticScan HCS Reader (Cellomics) and ImageXpress high-content imaging Systems (Molecular Devices). They provide many standard image analysis methods, however, it is not possible for researchers to substantially adapt these methods for more specific biological questions. Therefore many researchers choose to use independent image analysis packages or write their own macros. One of the most popular tools used in high-throughput and high-content screening is ImageJ (or Fiji, which is considered as a distribution of ImageJ). It contains various standard image analysis algorithms. More importantly, because of its free open-source feature, hundreds of state-of-the-art methods have been programmed in the form of ImageJ plugins and macros, and are provided online. Furthermore, it provides a user-friendly platform for developers to customize their own analysis. For example, image analysis algorithms developed in this thesis are programmed as ImageJ plugins.

1.3.3.2 Quantification

Quantification aims to extract numerical features from images. According to biological questions, features can be classified as morphological based, localization based and intensity based. Morphological based features usually refer to the morphological properties of structures that are under investigation, such as the area of the nucleus (2D) or surface of neuron cell (3D). Localization based features contain relative spatial information of interested structures, for example the position of focal adhesions relative to the cell border. Both features are measured based on the segmentation result, and can be measured on the single-cell level or cell population level, or sometimes even sub-cellular level [32]. Intensity based features can be measured from the segmentation result as well. For example, in the study of NF-κB translocation [33] (**chapter 2**), dynamic changes of fluorescence intensity in the nuclear and cytoplasmic area are quantified to analyze the kinetics of NF-κB translocation. Other intensity

based features are independent of segmentation result, such as moments and wavelet. Recently those features started to draw more and more attention, and some analysis packages already embed algorithms to measure them, such as CellProfiler [34] and PSLID [32, 35].

1.3.3.3 Tracking and analysis for image stacks

For some research projects, tracking techniques are required. Tracking refers to finding the location of objects (such as nuclei or cells) in each consecutive frame of time-lapse images. It works directly on the segmentation result and it connects objects in time on the basis of criteria such as the speed of motion, the shape of the trajectory and the possibility of the objects splitting (cell division) or merging (cell fusion). Once tracking is computed, the tracks and associated object properties (morphology and intensity changes) can be combined to yield a powerful description of the phenotypic evolution of cells.

For the study of 3D structures, series of optical sections are typically acquired and compose image stacks. Those stacks can be used to reconstruct the 3D geometric models of the sample, or can be collapsed (also called projected) into a single 2D image. For the latter, further analysis such as segmentation and quantification is performed on the 2D projected images, thus saving processing time and data storage significantly. However, information in z-direction is missing. 3D analysis provides much more detailed spatial information, but requires more complicated and computational intensive reconstruction methods and quantification methods. Although it has not been widely used for high-throughput screening, especially for large-scale experiments such as genome wide siRNA screens, due to the physiological and pathological relevance of 3D culture systems, image analysis for 3D high-throughput and high-content screening will become an important area for future innovative developments in image processing. This is the subject of **chapters 4 and 5** of this thesis.

1.3.4 Data analysis

In data analysis, numerical features derived from a single treatment, for example a single RNAi gene knockdown or exposure to a small molecule, are considered as data from one sample. The purpose of data analysis varies from experiment to experiment. One basic purpose is to identify "hits", meaning to identify genes or compounds which play a functional role in the cellular process that is under investigation (genes), or may - positively or negatively - affect the process (compounds). This is determined by comparing each sample with control treatments that are carried out under the same condition but induce no change in the cellular process. Before "hits" are identified, quality control and data normalization are performed to remove systematic errors and to allow comparison and combination of samples from different plates. There are many statistical methods for normalization and quality control for single readout screens which are summarized in references [36, 37]. For the high-throughput and high-content screen where multiple readouts are measured, multi-parametric tests are

applied, for example Mahalanobis distance. Because more aspects of phenotype are taken into account, multi-parametric tests often provide more precise hits for treatments that have high variability in a single readout metric. Several multi-parametric methods will be introduced and applied in this thesis.

Recently, machine learning started to show its powerful function in high-throughput and high-content screening [38]. Clustering is one of the typical techniques. It is a form of unsupervised machine learning and can be used to group the samples based on the phenotypic similarities. For example, clustering has been used to group drugs by their effects [39] and proteins by their sub-cellular patterns [40]. Another often used technique is classification [41], which is a form of supervised machine learning. The biologists define the classes of treatments according to the biological properties in advance. For example, genes which are involved in the same signaling pathway are assigned to the same class. Those samples with known classes are used as training data to train a "classifier", which can automatically determine the class of unknown sample and which of the features are informative for distinguishing the classes. Sometimes unsupervised machine learning and supervised machine learning are combined in one experiment, to for example investigate the relationship of known biological classes with morphological clusters.

1.4 Scope and structure of this thesis

Currently, image analysis is likely the major bottleneck in high-throughput and high-content screening studies. One of the problems is the lacking of robust single-cell based analysis. Few platforms perform single cell image analysis but rely on the cell density or image resolution. When the cells are too dense or image resolution and magnification is relatively low, for example in micro-tissue study, existing methods cannot work properly. Therefore, one goal of this thesis is to develop robust and efficient image analysis methods suitable for single-cell studies that should not be largely limited to cell type, cell density and image quantity

The reason that there is a great need for the methods of single cell analysis is that cells may respond to stimuli or perturbation differently, creating distinct subpopulations. Although many studies emphasize this effect, it has almost never been quantitatively identified and the information on subpopulation behavior has not been used for later screening data analysis. Thus how important the role is of subpopulation in many high-throughput and high-content screens is still unknown. This thesis aims to develop an analysis platform that takes into account the heterogeneity of cellular behavior so that subpopulation information is collected for the further data analysis.

In many high-throughput and high-content screens specific biological effects are expected, and in the analysis by conventional methods only representative features are measured from the images. Those methods largely depend on the expertise of biologists, but often even biologists are not certain of all possible effects. To solve this problem, many methods measure morphological parameters thoroughly, however,

they miss the information which cannot be revealed by morphological properties. This thesis aims to develop an ultra-high content analysis which collects maximum information of phenotype from images, which can be applied for various types of screening assays.

The structure of this thesis is as follows: **chapter 2** introduces an automated image analysis for the study of NF-κB nuclear translocation kinetics in high-throughput screening. **Chapter 3** moves away from the live-cell screening and presents another image analysis platform suitable for the study of 3D cultured micro-tissues. This platform firstly projects each image stack into a single in-focus 2D image, and then ultra-high content measurement is carried out on the 2D projected image. The work in **chapter 4** extends this 2D projection based analysis to real 3D analysis, and demonstrates that it is suited for screenings which use wide-field microscopy. In **chapter 5**, this 3D analysis platform is applied to analyze the effects of various anti-cancer drugs on 3D micro-issues of mouse breast cancer. The performance of 2D projection analysis and 3D analysis is compared in **chapter 5**. Finally, **chapter 6** provides a general discussion on the results obtained in our studies and on the implications for future research.

References

[1] Pelkmans, L. et al. Genome-wide analysis of human kinases in clathrin- and caveolae/raft-mediated endocytosis. Nature 436, 78-86 (2005).

[2] Sonnichsen, B. et al. Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. Nature 434, 462-469 (2005).

[3] Korn, K. & Krausz, E. Cell-based high-content screening of small-molecule libraries. Curr Opin Chem Biol 11, 503-510 (2007).

[4] Starkuviene, V. et al. High-content screening microscopy identifies novel proteins with a putative role in secretory membrane traffic. Genome Res 14, 1948-1956 (2004).

[5] Chudakov, D. M., Lukyanov, S. & Lukyanov, K. A. Fluorescent proteins as a toolkit for in vivo imaging. Trends Biotechnol 23, 605-613 (2005).

[6] Lippincott-Schwartz, J. & Patterson, G. H. Development and use of fluorescent protein markers in living cells. Science 300, 87-91 (2003).

[7] Pepperkok, R. & Ellenberg, J. High-throughput fluorescence microscopy for systems biology. Nat Rev Mol Cell Biol 7, 690-696 (2006).

[8] Tsien, R. Y. Building and breeding molecules to spy on cells and tumors. FEBS Lett 579, 927-932 (2005).

[9] Puigvert, J. C. et al. Systems Biology Approach Identifies the Kinase Csnk1a1 as a Regulator of the DNA Damage Response in Embryonic Stem Cells. Sci Signal 6 (2013).

[10] Zanella, F., Lorens, J. B. & Link, W. High content screening: seeing is believing. Trends Biotechnol 28, 237-245 (2010).

[11] Huh, D., Hamilton, G. A. & Ingber, D. E. From 3D cell culture to organs-on-chips. Trends Cell Biol 21, 745-754 (2011).

[12] Pampaloni, F., Reynaud, E. G. & Stelzer, E. H. K. The third dimension bridges the gap between cell culture and live tissue. Nat Rev Mol Cell Bio 8, 839-845 (2007).

[13] Petersen, O. W., Ronnov-Jessen, L., Howlett, A. R. & Bissell, M. J. Interaction with basement membrane serves to rapidly distinguish growth and differentiation pattern of normal and malignant human breast epithelial cells. Proc Natl Acad Sci U S A 89, 9064-9068 (1992).

[14] Wang, F. et al. Reciprocal interactions between beta1-integrin and epidermal growth factor receptor in three-dimensional basement membrane breast cultures: a different perspective in epithelial biology. Proc Natl Acad Sci U S A 95, 14821-14826 (1998).

[15] Cukierman, E., Pankov, R., Stevens, D. R. & Yamada, K. M. Taking cell-matrix adhesions to the third dimension. Science 294, 1708-1712 (2001).

[16] Meshel, A. S., Wei, Q., Adelstein, R. S. & Sheetz, M. P. Basic mechanism of three-dimensional collagen fibre transport by fibroblasts. Nat Cell Biol 7, 157-164 (2005).

[17] Walpita, D. & Hay, E. Studying actin-dependent processes in tissue culture. Nat Rev Mol Cell Biol 3, 137-141 (2002).

[18] Zaman, M. H. et al. Migration of tumor cells in 3D matrices is governed by matrix stiffness along with cell-matrix adhesion and proteolysis. Proc Natl Acad Sci U S A 103, 13897-13897 (2006).

[19] O'Brien, L. E., Zegers, M. M. & Mostov, K. E. Opinion: Building epithelial architecture: insights from three-dimensional culture models. Nat Rev Mol Cell Biol 3, 531-537 (2002).

[20] Echeverria, V. et al. An Automated High-Content Assay for Tumor Cell Migration through 3-Dimensional Matrices. J Biomol Screen 15, 1144-1151 (2010).

[21] Neumann, B. et al. Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. Nature 464, 721-727 (2010).

[22] Bartfeld, S. et al. High-throughput and single-cell imaging of NF-kappaB oscillations using monoclonal cell lines. BMC Cell Biol 11, 21 (2010).

[23] Azorsa, D. O. et al. High-content siRNA screening of the kinome identifies kinases involved in Alzheimer's disease-related tau hyperphosphorylation. BMC Genomics 11, 25 (2010).

[24] Lee, T. K. & Covert, M. W. High-throughput, single-cell NF-kappaB dynamics. Curr Opin Genet Dev 20, 677-683 (2010).

[25] Tay, S. et al. Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing. Nature 466, 267-271 (2010).

[26] Abraham, V. C., Taylor, D. L. & Haskins, J. R. High content screening applied to large-scale cell biology. Trends Biotechnol 22, 15-22 (2004).

[27] Starkuviene, V. & Pepperkok, R. The potential of high-content high-throughput microscopy in drug discovery. Br J Pharmacol 152, 62-71 (2007).

[28] Lakowicz, J. R., Gryczynski, I. I. & Gryczynski, Z. High Throughput Screening with Multiphoton Excitation. J Biomol Screen 4, 355-362 (1999).

[29] Ries, J., Kaplan, C., Platonova, E., Eghlidi, H. & Ewers, H. A simple, versatile method for GFP-based super-resolution microscopy via nanobodies. Nat Methods 9, 582-584 (2012).

[30] Maksimovic, R., Stankovic, S. & Milovanovic, D. Computed tomography image analyzer: 3D reconstruction and segmentation applying active contour models–'snakes'. Int J Med Inform 58-59, 29-37 (2000).

[31] Maksimovic, R. et al. Computed tomography image analyzer: segmentation applying active contour models–"snakes". Stud Health Technol Inform 68, 395-399 (1999).

[32] Glory, E. & Murphy, R. F. Automated subcellular location determination and high-throughput microscopy. Dev Cell 12, 7-16 (2007).

[33] Di, Z. et al. Automated analysis of NF-κB nuclear translocation kinetics in high-throughput screening. PloS one 7, e52337 (2012).

[34] Carpenter, A. E. et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol 7 (2006).

[35] Huang, K. & Murphy, R. F. Boosting accuracy of automated classification of fluorescence microscope images for location proteomics. Bmc Bioinformatics 5 (2004).

[36] Birmingham, A. et al. Statistical methods for analysis of high-throughput RNA interference screens. Nat Methods 6, 569-575 (2009).

[37] Malo, N., Hanley, J. A., Cerquozzi, S., Pelletier, J. & Nadon, R. Statistical practice in high-throughput screening data analysis. Nat Biotechnol 24, 167-175 (2006).

[38] Murphy, R. F. An active role for machine learning in drug development. Nat Chem Biol 7, 327-330 (2011).

[39] Perlman, Z. E. et al. Multidimensional drug profiling by automated microscopy. Science 306, 1194-1198 (2004).

[40] Chen, X. & Murphy, R. F. Objective clustering of proteins based on subcellular location patterns. J Biomed Biotechnol 2005, 87-95 (2005).

[41] Boland, M. V. & Murphy, R. F. A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of HeLa cells. Bioinformatics 17, 1213-1223 (2001).