1

INTRODUCTION

Steven A. Haney

1.1 THE BEGINNING OF HIGH CONTENT SCREENING

Microscopy has historically been inherently a descriptive endeavor and in fact it is frequently described as an art as well as a science. It is also becoming increasingly recognized that image-based scoring needs to be standardized for numerous medical applications. For example, for medical diagnoses, interpretation of medical images has been used since the 1950s to distinguish disorders such as cervical dysplasias and karyotyping [1]. Cameras used in microscopes during this era were able to capture an image, reduce the image data to a grid that was printed on a dot-matrix printer and integrated regional intensities to interpret shapes and features. In essence, these principles have not changed in 50 years, but the sophistication and throughput with which it is done has increased with advances in microscope and camera design and computational power. In the early 1990s, these advances were realized as automated acquisition and analysis of biological assays became more common.

Advances in automated microscopy, namely the automated movement of slides on the stage, focusing, changing fluorophore filters, and setting proper image exposure times, were also essential to standardizing and improving biomedical imaging. Automated microscopy was necessary to reduce the amount of time required of laboratory personnel to produce these images, which was a bottleneck for these studies, especially medical diagnoses. A team of scientists from Boston and Cambridge, Massachusetts described an automated microscope in 1976 that directly anticipated its use in subcellular microscopy and image analysis [2]. The microscope, and a processed image of a promyelocyte captured using the instrument, are shown in Figure 1.1.

An Introduction to High Content Screening: Imaging Technology, Assay Development, and Data Analysis in Biology and Drug Discovery, First Edition. Edited by Steven A. Haney, Douglas Bowman, and Arijit Chakravarty.
 Published 2015 by John Wiley & Sons, Inc.

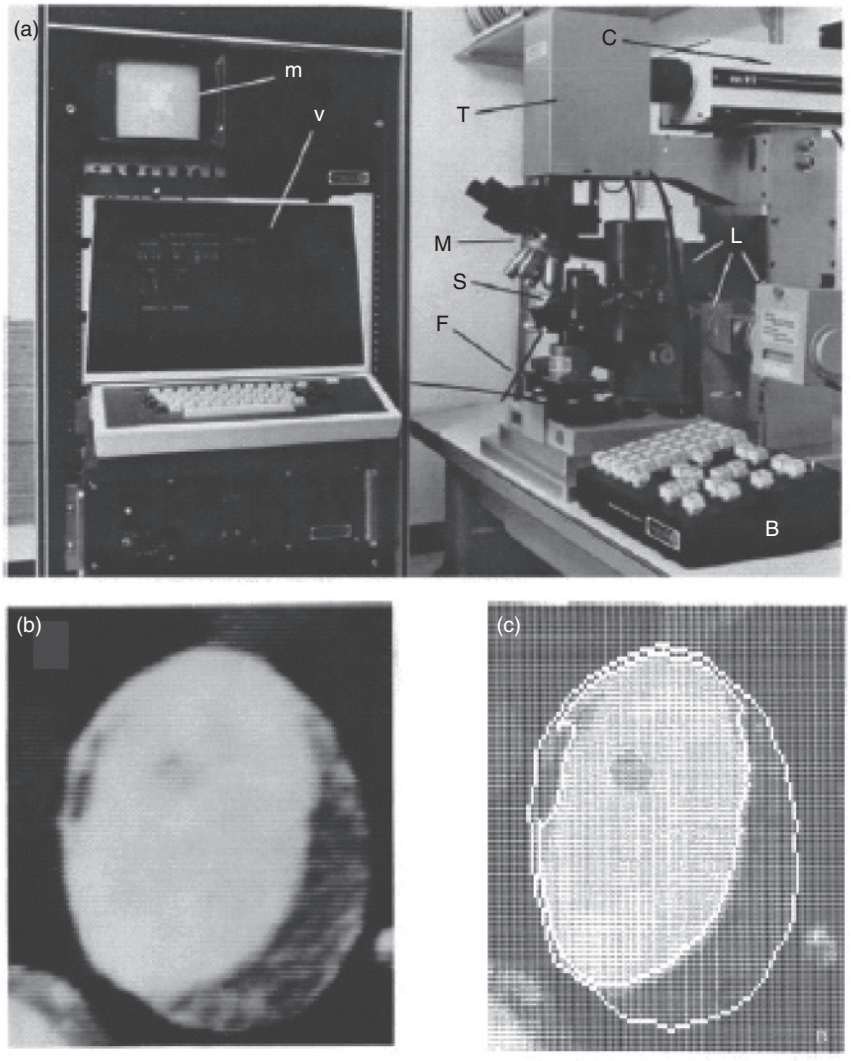

FIGURE 1.1 *An early automated microscope used in biomedical research.* (a) An example of an automated fluorescence microscope. Letters inside the figure are from the original source. The system is outfitted with controlled stage and filter movements (S and F), a push-button console for manual movements (B), a television camera and monitor (T and m) and a video terminal for digitizing video images (v). (b) A video image of a promyelocyte and (c) image analysis of (b), showing, an outline of the nucleus and cell borders, which can be used in automated cell type recognition. Reproduced with permission from [2]. Copyright 1974 John Wiley & Sons.

Until the mid-1990s, automated microscopy was applied in basic research to address areas of high technical difficulty, where rigorous measurements of subtle cellular events (such as textural changes) were needed, events that took place over long time periods or were rare (which made it challenging to acquire sufficient numbers of images of each event). In medicine, automated imaging was used to standardize the interpretation of assay results, such as for the diagnosis of disease from histological samples (where it was notoriously difficult to achieve concordance among clinical pathologists). Adapting quantitative imaging assays into a screening context was first described by Lansing Taylor and colleagues [3], who commercialized an automated microscope capable of screening samples in multiwell plates (a format that had emerged as an industry standard during this time period). The term "high content" was coined to contrast the low throughput screening in these imaging assays with the increasing scale of high throughput primary drug discovery screens. Many groups have since demonstrated the usefulness of automated microscopy in drug discovery [4, 5] and basic research [6, 7]. During this phase (the early 2000s), data acquisition, image analysis, and data management still imposed limits on image-based screening, but it did find an important place in the pharmaceutical industry, where expensive, labor-intensive assays critical for late-stage drug development were a bottleneck. One example is the micronucleus assay, an assay that measures the teratogenicity of novel therapeutics through counting the number of micronuclei (small nonnuclear chromosomal fragments that result from dysregulation of mitosis). An increase in the number of cells that contain micronuclei is indicative of genotoxicity, so this assay is frequently part of a screening program to make a go/no go decision on clinical development [8]. The assay requires finding binucleate cells and checking for a nearby micronucleus. For each compound assayed, a single technician might spend many hours in front of a microscope searching and counting nuclei. Automation of image capture and analysis not only reduced the work burden of researchers, but it also made the analysis itself more robust [9]. Similar applications were found in the field of cell biology, where automated microscopy was utilized to collect and analyze large data sets [10, 11].

Following from these early implementations, high content screening (HCS) has been widely adopted across many fields as the technology has improved and more instruments are available commercially. The speed at which images can be analyzed is limited by computer power, as more advanced computer technology has been developed, the scale at which samples can be analyzed has improved. Faster computers also mean that more measurements per cell can be made; shapes of cells and subcellular structures can be analyzed as well as probe intensities within regions of interest. This has led to the quantification of subtle morphological changes as assay endpoints. A widely used application of this approach has been receptor internalization assays, such as the Transfluor™ assay to measure the activation of GPCRs through changes in the pattern of receptor staining, from even staining over the surface of the cells to dense puncta following internalization of the activated receptors through vesicle formation [12]. Concomitant with the increase in the sophistication of the assays themselves, improvements in the mechanical process of screening samples has also fed the growth of HCS. Gross-level changes, such as integrating plate

handling robotics and fine-level changes, such as improvements in sample detection and autofocusing, have improved the scale of HCS to the point where image-based readouts are possible for true high throughput screens (screens of greater than 100,000 compounds) [5].

HCS has a strong presence in basic biological studies as well. The most widely recognized applications are similar to screening for drug candidates, including siRNA screening to identify genes that control a biological process, and chemical genetics, the identification of small molecules that perturb a specific cellular protein or process. While operationally similar to drug screening, they seek to explain and study biological questions rather than lead to therapeutics explicitly. Additional uses of HCS in basic science include the study of model organisms. Finally, the use of multiparametric single cell measurements has extended our understanding of pathway signaling in novel ways [11].

1.2 SIX SKILL SETS ESSENTIAL FOR RUNNING HCS EXPERIMENTS

At this point we want to touch on the fundamental skill sets required to successfully set up and use an HCS system to address a biological problem, and how responsibilities might be divided up in different settings. The six major skill sets required to develop and run an HCS project are shown in Figure 1.2. Each area is distinct enough as to be a full-fledged area of expertise (hence introducing these areas as "skill sets"), but

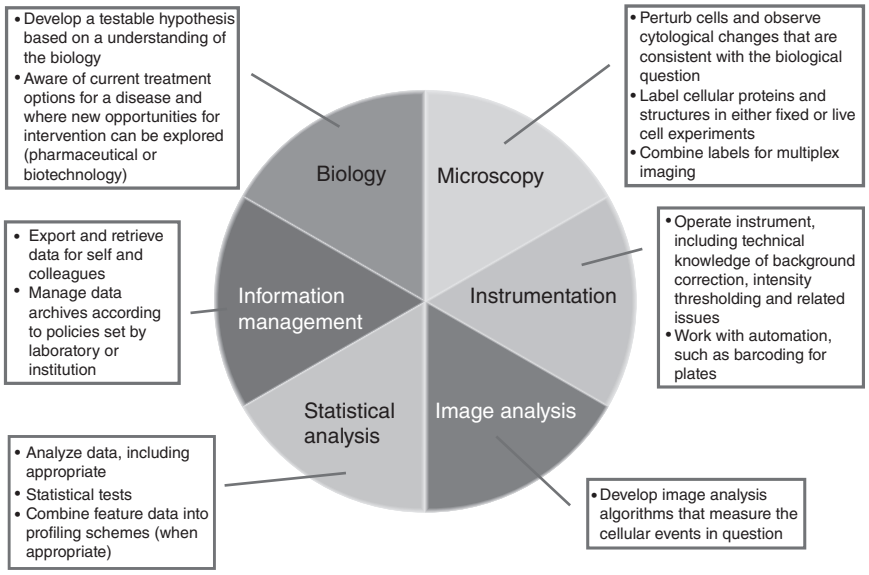

FIGURE 1.2 *The basic skill sets essential for establishing and running HCS experiments*. Skills noted in the figure are discussed in detail in the text.

typically a person is competent in more than one area. It is rare that all roles can been successfully filled by one person. Therefore, the ability to develop a collaborative team is essential to HCS. It is also very important to understand that these roles vary between groups, and this can cause problems when people move between groups or as groups change in size. The skill sets are the following.

1.2.1 Biology

The biologist develops the question that needs to be answered experimentally. In academia, the biologist is typically a cell biologist and oftentimes is also capable of collecting images by HCS as well. In industrial circles (pharma and biotech), a therapeutic team may be led by a biochemist or *in vivo* pharmacologist, who may have little training in fluorescence microscopy. The key area of expertise here is an appreciation of these problems and an ability to formulate strategies (experimental systems and assays) to address them. There is also a significant understanding of how cellular models in the laboratory relate to the biology *in vivo*. In addition to understanding the fundamental biological question, understanding how to establish a cellular model that incorporates relevant aspects of the biological environment is important.

1.2.2 Microscopy

Although many of the HCS systems are sold as turnkey "black-boxes," it is important to have a good understanding of fundamental microscopy components (staining techniques, reagents, and optics) as each has a significant impact on the quality of data generated by the instruments. For example, the choice of illumination system and filter sets determine which fluorescence wavelengths (fluorophores) you can use to stain specific cellular compartments. Other microscope objective characteristics (numerical aperture, magnification, and working distance) also impact both the types of samples one can image as well as the spatial resolution of the resulting images. More information on these topics are covered in Chapters 3 and 7. If the biological question is posed by someone who is not a trained microscopist, then it is important to discuss technical aspects with someone who has such training, which is why these skills are frequently part of the Platform Manager responsibilities (see below), particularly when the HCS instrument is used in a core facility.

1.2.3 HCS Instrumentation (Platform Manager)

The platform manager focuses on the hardware and software needed to keep an HCS facility running smoothly. Much of the information needed to run a particular HCS instrument is obtained from the instrument vendor, including instrument operation and, in some cases, strategies for data analysis; the Platform Manager will be the one who interacts with the vendors directly, particularly for handling challenging problems with the instrumentation or for scheduling updates to the instrument. Although the image acquisition configuration is often simplified by the user interface software,

a solid understanding of imaging hardware (laser autofocus, CCD camera exposure time, PMT amplifier gain) is needed for optimal use of the instrument. In addition, it is important to know how to integrate the HCS instrument with other laboratory automation instruments to increase overall throughput. Relevant automation instrumentation includes robotic plate handling devices (robot arms with plate "stackers" or "hotels") that serve plates to the HCS instrument one at a time and store plates after they are read, automated tissue culture incubators to store plate used for live cell imaging, and plate barcoding equipment. We go into more detail on these in Chapters 5 and 12.

1.2.4 Image Analysis

The identity of the person that contribute to the image analysis can be very fluid. In many cases, this position functions as an extension of the microscopy skill set, but it is also becoming a more specialized position, particularly as HCS experiments grow in complexity or subtlety, such as spheroids and primary cells (which may be plated at confluence) require more work to develop an algorithm that is suitable. Many of the instruments include "canned" algorithms that can be applied to a range of assays (cell counting, mitotic index, cell health, neurite outgrowth, etc.). There are also third-party applications, such as the open-source CellProfiler™ and the commercial analytical package Definiens™ that are compatible with all of the common platform image set files and require more effort to understand how to use them than the "shrink-wrapped" algorithms. The skills are covered in more detail in Chapters 4, 5, and 15.

1.2.5 Statistical Analysis

The data analyst needs to understand the design and objectives of the experiment and apply the appropriate statistical tests to characterize any conclusions. The scope of the data analysis needs can vary greatly depending on the user needs: whether you are using the instrument for single experiments or you are running a screen with hundreds of compounds. Depending on the experiment, analysis can be straightforward, even routine, or it can be complex with a potential for making false conclusions if the proper statistical tests are not used.

HCS in a screening environment typically means that an assay is being run to identify hits, and the methods for determining how robust a screen is can be evaluated by someone with good screening experience, but not necessarily having a rigorous background in statistics. Assays used for HTS are typically well validated and use a limited set of well-characterized cell lines. Positive and negative controls produce visually distinct results, and are therefore easy to measure. As such, few images need to be obtained (as few as one per well) and compound (or RNAi reagent) effects are evaluated relative to these controls. Analysis of the data typically involves measures of automation and cell culture performance, in the form of heatmaps of individual plates to locate patterns that suggest systematic or spurious problems, followed by normalizations of the data across plates and an evaluation of each treatment according to a single measure, typically a *Z*-score (see Chapter 9). Such an analytical stream is

fairly straightforward. Phenotypic patterns, such as composite measures of multiple features, require testing of the metrics themselves and a scheme for integrating features. Features can be measured according to *Z*-scores, much like in a single endpoint HTS assay. However, using multiple morphological features to evaluate the effect treatment effects can lead to false conclusions, because testing a wide number of potential features as potential assay endpoints leads to spurious associations. The latter is a subtle statistical problem, and because of pitfalls such as this, analyzing such data requires stronger training or experience than is typical for a bench scientist [13]. In addition to Chapter 9, data analysis is covered in Chapters 8–13, which cover the concept of assay metrics in HCS and progress through statistical analysis of screening data and multivariate methods.

1.2.6 Information Technology Support

Information technology (IT) expertise is needed to implement a data management solution according to mandates from the research team and their institution. HCS generates unprecedented volumes of data. Storing and retrieving data in a stable, validated IT environment that conforms to institutional guidelines requires both a thorough understanding of how to manage data, but also an understanding of the needs of different scientists who use the HCS instrument. HCS vendors are well aware of the data management requirements, but rarely provide complete solutions. There may also be need to integrate with user databases, such as linking the plate barcodes with a compound library as described above. Further details related to informatics requirements are described in Chapter 6.

1.3 INTEGRATING SKILL SETS INTO A TEAM

While skill sets were delineated above, it does not necessarily take six people to run an HCS experiment. Two or three functions might be covered by a single person, but rarely more. Therefore, HCS is a collaborative endeavor, and instead of challenging oneself with learning several complex roles, it is most productive to consider what roles each person can carry well and what skill sets need to be filled. The ability to play more than one role is influenced by the scientific and business environment. A large pharmaceutical company typically has institutional policies that mandate an IS/IT group implement data management policies that insure the security and preservation of data. This group will typically identify existing server space or install local servers for image and data storage, and will establish procedures for backing-up and archiving data. An academic lab or a small biotech company using HCS may have lesser needs for dedicated IT support since many image or data analysts (particularly those that have experience in high volume data-driven projects such as transcriptional profiling or proteomics) will be able to set up a fileserver and help to organize data. In such a case, the roles of image/data analyst and IT manager might be combined.

Most commonly, the roles of biologist and microscopist will be combined. Sometimes, a biologist who is not trained in cell biology might articulate an important

question that can be addressed by HCS, but that person may not have sufficient microscopy experience to establish a robust HCS assay. In such a case, a scientist should not be dissuaded from proposing a high content experiment, but needs to collaborate with a microscopist. The roles of HCS instrumentation specialist, image and data analyst can be combined in drug screening groups. A high throughput screening (HTS) group in a pharmaceutical company typically manages many instruments, and works with biologists to adapt assays for HTS, but this can come at the expense of flexibility, as rapid progression through many screens can limit the time that can be spent on cell lines or imaging challenges for a particular project.

Who, then, develops the image analysis algorithms? This is probably the most collaborative piece of running an HCS experiment. Certainly, the person running the HCS instrument should have experience using the standard image analysis tools packaged with most HCS instruments, but some assays might require help from an expert image analyst to write custom image analysis algorithms, using software such as MATLAB™, FIJI™, or CellProfiler™, that are not packaged with the HCS instrument. As noted above, this is becoming common in larger facilities. The biologist/microscopist functions are also intrinsically involved in developing image analysis algorithms as the process is iterative: an algorithm is developed and tested on control samples, the results are evaluated by the microscopist/biologist to confirm that they capture physiologically and experimentally relevant parameters, the algorithm is then improved further and re-evaluated, until an optimal and robust image algorithm is produced.

1.4 A FEW WORDS ON EXPERIMENTAL DESIGN

Finally, it is worth a few minutes to discuss HCS experiments from a broader perspective. Researchers beginning to appreciate the power of HCS can become overwhelmed. One common assumption for users embarking on HCS is that obtaining a meaningful result requires imaging many, many cells at high magnification using the most sensitive camera and analyzed using the most complex imaging algorithms. In truth, even the most basic HCS experiments are substantially richer in data and more statistically significant than traditional cell biological and biochemical assays (Western blotting and ELISA technology) that measure responses of cell populations only and do not provide information about individual cells. So much so, in fact, that it is worth taking some time to consider how much (or rather, how little) sophistication is necessary to answer the scientifically relevant question. As an example, a dose response curve typically requires at least three replicates per dose, and 5 to 12 compound concentrations to determine the potency of a small molecule. In many cases, more than triplicate values are used per dose. The additional replicates are necessary because dose–response curves are typically quite noisy near the IC50. In contrast, an HCS translocation assay at a moderate magnification will determine the extent of activity of the compound on 30–150 cells per field. As such a single field will capture enough that truly spurious noise is not a problem. Such assays still require replicates, due to systematic or experimenter error, but the historical problem of noise and scatter are handled much better by imaging technologies, a detailed

treatise on this is presented in Chapter 9. Lastly, the sensitivity of an HCS imager is very high, and it can measure very subtle changes. As such, it is common that a low magnification objective is usually sufficient to observe the change as a function of compound dose, and using a lower magnification objective means faster acquisition times and fewer images that need to be collected.

HCS has found a place in the highly rigorous and standardized discipline of HTS. HTS in modern drug discovery relies on the ability to robustly make singular measurements over very many samples (upward of 1 million in large-pharma HTS campaigns), and HCS accomplishes this by capturing a single image per well. At the other end of the continuum, there are approaches to drug development and basic biology that leverage the sensitivity of HCS to integrate many (largely) independent effects of a perturbation to determine the extent and similarity of it to other perturbations. These approaches do in fact benefit from better imaging and large numbers of cells, but they are far less common that the simpler HCS assays.

1.5 CONCLUSIONS

HCS is not a single technological innovation, but the aggregation of a handful of independent technologies that give a highly flexible approach to quantitative cell biology. No single aspect of HCS is truly difficult to learn, but pulling together an understanding of each of the core technologies takes time. Most vendors of HCS instruments commit a lot of effort to training their users, and these efforts are essential to becoming fluent with their instruments.

There is greater variability of educational opportunities for learning the complete set of skills that contribute to HCS, in part because there are many places where HCS is used. Screening laboratories will place a premium on reliability and minimizing day-to-day variability. Drug discovery laboratories (and many academic laboratories that study cellular signaling pathways) will focus on the flexibility of the platform, and the capability of measuring the activity of a large number of signaling pathways and functional outcomes. Systems biology, tissue biology, pharmacology, and other disciplines also make use of the unique capabilities of HCS. All of these will be discussed in this book. In each case, the core process of HCS will be described but linking it to the needs of the laboratory will depend on the HCS team.

KEY POINTS

1. HCS represents the integration of diverse skills. In general, scientists working in HCS will have a high level of expertise in a few areas, but will rely on others with complementary expertise to form a robust team.
2. An appreciation of the power of HCS is invaluable, ironically because there are many occasions where a simple and efficient assay is optimal. Such cases are common and will not call on all of the experimental and analytical power of HCS that is available, just a clear vision of the problem and how it can be solved.

FURTHER READING

There are many review articles available that discuss the role of HCS in biology and drug discovery. In addition, the following books are multi-author efforts that present many perspectives and case studies in the practice of HCS.

Haney, S. (ed.). *High Content Screening: Science, Techniques and Applications*. John Wiley and Sons, Hoboken, NJ, 2008.

Inglese, J. Measuring biological responses with automated microscopy. *Methods in Enzymology*, 2006, **414**: 348–363. Academic Press, New York, NY.

Taylor, D.L. et al. *High Content Screening: A Powerful Approach to Systems Cell Biology and Drug Discovery*. Humana Press, New York, NY, 2006.

REFERENCES

1. Eaves, G.N. Image processing in the biomedical sciences. *Computers and Biomedical Research*, 1967, **1**(2): 112–123.
2. Brenner, J.F. et al. An automated microscope for cytologic research: a preliminary evaluation. *Journal of Histochemistry and Cytochemistry*, 1976, **24**: 100–111.
3. Giuliano, K. et al. High-content screening: a new approach to easing key bottlenecks in the drug discovery process. *Journal of Biomolecular Screening*, 1997, **2**: 249–259.
4. Haney, S.A. et al. High content screening moves to the front of the line. *Drug Discovery Today*, 2006, **11**: 889–894.
5. Hoffman, A.F. and Garippa, R.J. A pharmaceutical company user's perspective on the potential of high content screening in drug discovery. *Methods in Molecular Biology*, 2006, **356**: 19–31.
6. Abraham, V.C., Taylor, D.L., and Haskins, J.R. High content screening applied to large-scale cell biology. *Trends Biotechnology*, 2004, **22**(1): 15–22.
7. Evans, J.G. and Matsudaira, P. Linking microscopy and high content screening in large-scale biomedical research. *Methods in Molecular Biology*, 2007, **356**: 33–38.
8. Ramos-Remus, C. et al. Genotoxicity assessment using micronuclei assay in rheumatoid arthritis patients. *Clinical and Experimental Rheumatology*, 2002, **20**(2): 208–212.
9. Smolewski, P. et al. Micronuclei assay by laser scanning cytometry. *Cytometry*, 2001, **45**(1): 19–26.
10. Feng, Y. et al. Exo1: a new chemical inhibitor of the exocytic pathway. *Proceedings of the National Academy of Sciences*, 2003, **100**(11): 6469.
11. Perlman, Z.E. et al. Multidimensional drug profiling by automated microscopy. *Science*, 2004, **306**(5699): 1194–1198.
12. Oakley, R.H. et al. The cellular distribution of fluorescently labeled arrestins provides a robust, sensitive, and universal assay for screening G protein-coupled receptors. *Assay and Drug Development Technologies*, 2002, **1**(1 Pt 1): 21–30.
13. Malo, N. et al. Statistical practice in high-throughput screening data analysis. *Nature Biotechnology*, 2006, **24**(2): 167–175.