

[24] Compound Classification Using Image-Based Cellular Phenotypes

By CYNTHIA L. ADAMS, VADIM KUTSY, DANIEL A. COLEMAN, GE CONG, ANNE MOON CROMPTON, KATHLEEN A. ELIAS, DONALD R. OESTREICHER, JAY K. TRAUTMAN, and EUGENI VAISBERG

Abstract

Compounds with similar target specificities and modes of inhibition cause similar cellular phenotypes. Based on this observation, we hypothesized that we could quantitatively classify compounds with diverse mechanisms of action using cellular phenotypes and identify compounds with unintended cellular activities within a chemical series. We have developed Cytometrix™ technologies, a highly automated image-based system capable of quantifying, clustering, and classifying changes in cellular phenotypes for this purpose. Using this system, 45 out of 51 known compounds were accurately classified into 12 distinct mechanisms of action. We also demonstrate microtubule-binding activity in one of seven related cytochalasin actin poisons. This technology can be used for a variety of drug discovery applications, including high-throughput primary screening of chemical and siRNA libraries and as a secondary assay to detect unintended activities and toxicities.

Introduction

Recent advances in high-throughput screening, chemistry, and genomics have resulted in a proliferation of new drug candidates. However, candidate drugs often fail in the clinic due to unintended activity ([Prentis et al., 1988](#)). Crucial to successful drug discovery is a robust approach that reduces the number of compounds tested in animals by identifying compounds that have unintended activities *in vitro*. Several approaches to assess the broader biological effects of candidate drugs prior to clinical testing have been developed, such as panels of biochemical and cell assays, gene expression analyses, and proteomic profiling ([Pritchard et al., 2003](#)). These assays, however, often are limited to a few compound conditions tested late in the discovery process.

Cell biology largely remains an unexploited tool in the attrition management of early stage compounds. At the gene and protein levels, many variables can be difficult to track, including posttranslational modifications,

protein turnover, and functional redundancy, which can often be assessed by monitoring cytoskeletal, cell cycle, and organelle responses to perturbations in the pathways that affect these systems. Cytoskeletal biology provides a rich source of markers for monitoring effects of compounds of interest on cellular structures and processes long known to be indicative of cellular integrity and general health of the cell, including the cytoskeleton, signal transduction pathways, and protein trafficking machinery.

Mechanisms of action (MOA) of a compound in a cell are dependent on many parameters, including concentration, time of incubation, the genotype, and environment of the cell. A phenotypic signature of a compound can be collected using cell morphology changes measured systematically across many conditions. To do this we developed Cytometrix technologies to (1) automatically and broadly measure effects of compounds on individual cells and (2) use multivariate statistics to analyze those effects across multiple conditions and compounds. This system is applicable at the earliest stages of drug discovery, enabling identification and prioritization of compounds that are highly selective for their intended targets and that do not display other cellular effects, thus increasing the chances of success during clinical testing. We have reported on an application of the Cytometrix technologies wherein an unbiased cell morphology-based screen of 107 small molecules comprising four different kinase inhibitor scaffolds was used to identify an unexpected activity in one of the compounds deriving from its role as an inhibitor of carbonyl reductase 1 ([Tanaka *et al.*, 2005](#)).

Cytometrix technologies is a system designed to broadly characterize the effects of compounds on cells. It includes fluidic automation for handling cells and fluorescent probes, automated microscopes for imaging relevant cellular structures, and software for image and data analysis. This chapter shows that cellular morphology changes are well correlated with compound MOA and that multivariate quantification of morphological changes across many experimental variables allows for clustering and classification of different mechanistic classes as well as for the detection of secondary mechanisms within a single class. This chapter provides general guidelines for collecting and analyzing such data, as well as detailed protocols for one particular type of experiment.

Quantifying Cellular Morphology Changes

Changes in cell cycle, cell shape, cytoskeleton organization, and protein trafficking machinery in response to compounds often result in concomitant cellular morphology changes. Accordingly, we label cells after they have been treated with a compound with a panel of fluorescent markers that broadly provide surrogate measures of these events. Hoechst, a DNA

marker, reveals the morphology of the cell nucleus and cell cycle status. Antitubulin antibody DM1 α is used to assess cell shape and cytoskeletal organization (Blose *et al.*, 1984). The lectin from *Lens culinaris* labels the Golgi apparatus, whose morphology is sensitive to changes in directed intracellular traffic between the nucleus and the plasma membrane (Polishchuk and Mironov, 2004). Measurements of changes in cellular and organelle morphology are taken with segmentation and data reduction algorithms that identify cell and nuclear boundaries based on the DNA and microtubule markers.

Cell types of differing origins respond differently to compounds acting by different MOAs depending on the expression levels and posttranslational modifications of their proteomes. It follows that monitoring changes across many cell types with a few markers may provide a more complex compound signature than using a single cell type with a large number of markers, as has been described (Perlman *et al.*, 2004). The example that follows used six cell lines with three markers, but the methodology is not limited to these conditions.

Cell Culture, Compound Addition, and Image Acquisition

Cell Culture

SKOV3 (ovarian epithelial cancer), A498 (kidney epithelial cancer), A549 (lung epithelial cancer), and SF268 (central nervous system epithelial cancer) are grown and maintained in RPMI media (Mediatech, Inc., Herndon, VA) with 5% fetal calf serum (FBS, HyClone, Logan, UT). DU145 (prostate epithelial cancer) cells (ATTC, Manassas, VA) are maintained in MEM with 5% FBS. HUVEC cells (VEC Technologies, Inc., Rensselaer, NY) are maintained in MCDB131 media (VEC Technologies) with 10% FBS.

Cell Plating

1. Plate 1000–1800 trypsinized (Mediatech) cells per well into 384-well plates (Costar) using a Multidrop (Thermo Labsystems, Beverly, MA). Make 3 cell plates for each drug plate to be tested, plus three control plates with no compound.
2. Incubate for 24 h at 37°, 5% CO₂.

Preparation of Compounds Plates

Compound stocks are maintained in dimethyl sulfoxide (DMSO). Compounds are serially diluted in DMSO in separate 384-well drug plates to achieve eight concentrations in 3 \times dilution steps using a Multimek

(Beckman Coulter Inc., Fullerton, CA). Wells are reserved on every drug plate for negative and positive controls. Negative control wells contain DMSO only and positive control wells receive eight, threefold dilutions of paclitaxel (Sigma-Aldrich-Fluka, St. Louis, MO) starting at a high concentration of 1 μ M on cells.

Compound Addition

1. Add media to entire compound plate and mix using a multichannel device.
2. Add compound to three cell plates per cell line (18 cell plates total) from the diluted compound plate using a PlateTrak (CCS Packard, Torrance, CA). Compound mixed with media is added to the cells to achieve a final 0.4% DMSO concentration.
3. Incubate for another 24 h at 37°, 5% CO₂.

Immunocytochemistry

All of the following steps are done using an automated pipetting and washing system.

1. Wash cells once with TBS buffer (Teknova, Half Moon Bay, CA).
2. Fix with 4% formaldehyde (Polysciences, Inc., Warrington, PA) made up in TBS for 1 h.
3. Wash three times with TBS buffer.
4. Block in 0.01% Triton X-100 (ICN Biomedicals, Inc., Irvine, CA) and 1% bovine serum albumin (BSA) in TBS (Teknova) for 1 h.
5. Stain with 0.01% Triton-X 100, 0.1% BSA, 5 μ g/ml Hoechst 33342 (Invitrogen, Eugene, OR), 5 μ g/ml FITC-lectin *L. culinaris* (Sigma-Aldrich-Fluka), and 3 μ g/ml rhodamine red-labeled monoclonal antibody DM1 α (courtesy of Tim Stearns) for 1 h.
6. Wash three times in TBS.

Microscopy

Cells are imaged on an inverted Axiovert 100 M epifluorescence microscope (Carl Zeiss Inc., Oberkochen, Germany) with a 5 \times objective and a xenon lamp (Sutter Instruments, Novato, CA). Metamorph (Universal Imaging, Downingtown, PA) is used to control the motorized x, y, z stage (Prior Scientific, Rockland, MA) that moves the plate to each well, auto-focuses, and takes three successive fluorescent images with an Orca 100 camera (Hamamatsu, Shizuoka Pref., Japan). Exposure times are set to minimize the number of saturated pixels in the image.

Pseudocolor overlays of the images are made in Adobe Photoshop using channel layers. The three-color overlay of the entire field of view is shown in Fig. 1A. Adobe can be used to zoom in on a portion of the field for illustrative purposes (Fig. 1B).

In the presence of compounds with different MOAs, the attributes of a cell line change is characteristic of each MOA. Several examples are shown in Fig. 2. HUVEC cells are stained with markers for the nucleus, the Golgi apparatus, and tubulin as described earlier. In the presence of DMSO control, HUVEC cells show smooth yet differently sized nuclei, slightly compact Golgi, and slightly elongated and overlapping cell shape as revealed by microtubule organization. Upon addition of compound, cellular organization changes significantly. Folimycin, a vacuolar-type ATPase inhibitor, causes an increase in the overall staining and distribution of Golgi. SB 203580, a p38 MAP kinase inhibitor, causes elongation of the cells. Vincristine, a microtubule-binding compound, causes formation of microtubule paracrystals and loss of Golgi organization. Paclitaxel, at subtoxic concentrations, causes the Golgi to compact and microtubules to reorganize. Cytochalasin D (CD), an actin poison, causes cells to collapse with two nuclei and results in the appearance of many retraction fibers. All of these morphology changes are unique and mechanism dependent.

Image and Data Reduction

Image Segmentation

Custom software is used to segment objects and extract attributes, but other commercially available software packages (such as MetaMorph, Axon ImageExpress, MatLab, and others) can achieve similar results. Segmentation of nuclei is based on a gradient method for edge detection (Cong and Vaisberg, 2005), the cells are segmented using a modified watershed algorithm on microtubule marker, and perinuclear area containing the Golgi complex is defined by expanding the mask from nuclei (Vaisberg *et al.*, 2002) (Fig.1C).

Image Analysis

Image masks are used to locate cell and nuclear areas for multiple shape-, texture-, and intensity-related measurements. For each object identified by segmentation algorithms, collected attributes include object location, area, perimeter, form factor, and axis ratio, as well as sum, mean, variance, moment, and kurtosis of pixel intensities.

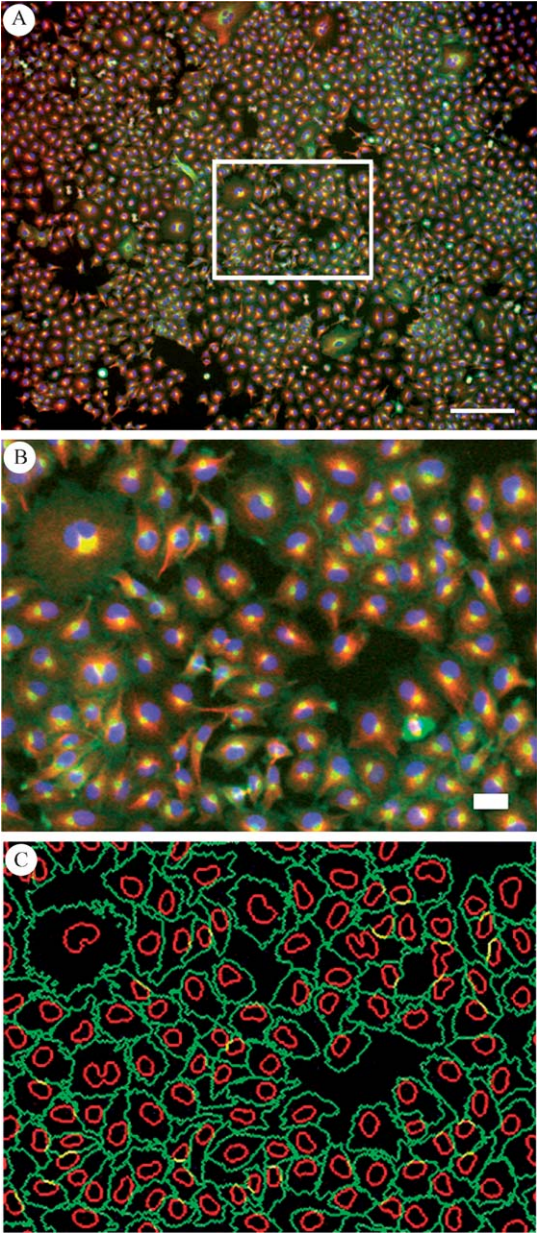


FIG. 1. Fluorescent markers and cell segmentation. (A) Image of A549 cells treated with 0.4% DMSO. The DNA marker fluorescence is colored blue, the Golgi is green, and the tubulin is red. The boxed region in A is enlarged in B. Segmentation of the same region is shown in C with DNA and tubulin segmentation outlines in red and green, respectively. Scale bars: (A) 200 μm and (B) 30 μm .

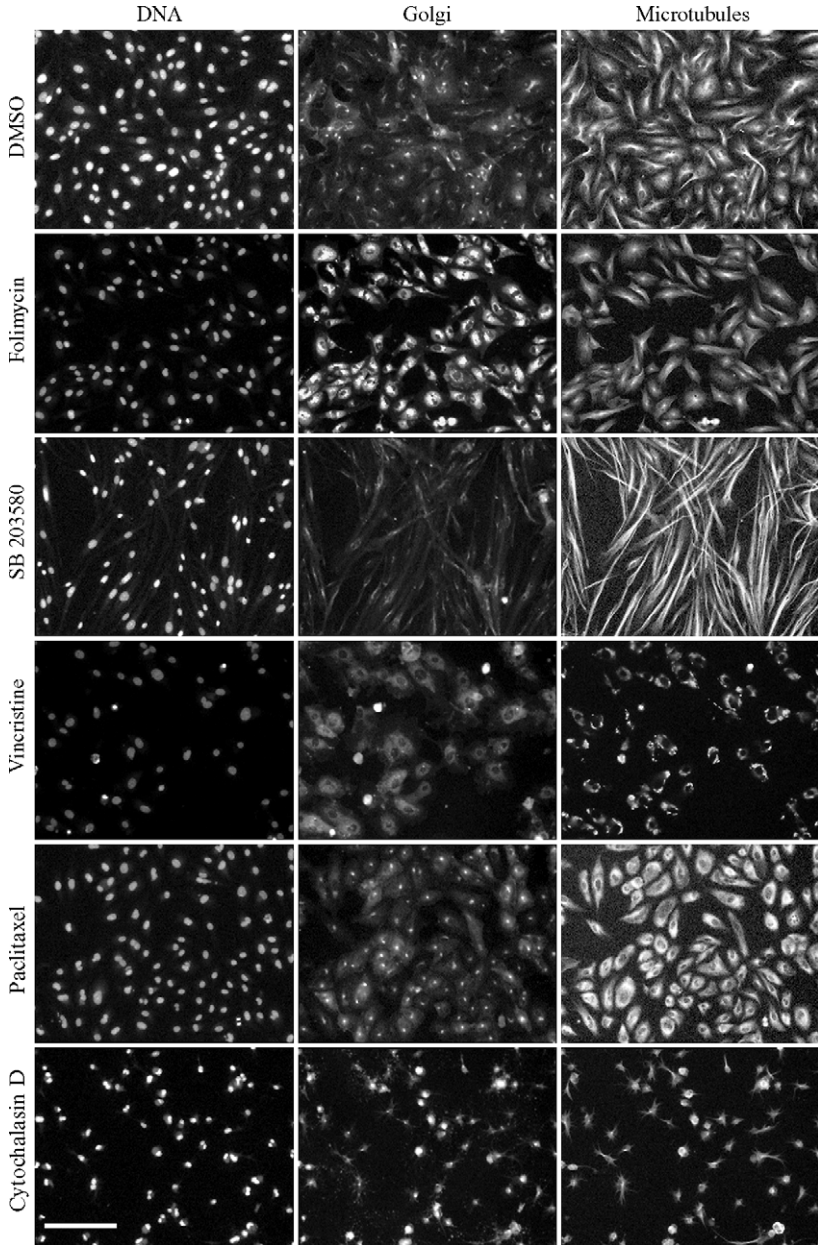


FIG. 2. Examples of compound-induced cell morphology changes. HUVEC cells in the presence of 0.4% DMSO control, folimycin (4.8 nM), SB 203580 (0.1 mM), vincristine

Object Attributes and Classification

After multiple features of individual objects are collected, each feature is summarized by a set of population statistics (called attributes) that describe distributions of a given feature across given populations of cells. In most cases we use attributes derived from subpopulations of cells in different phases of the cell cycle. Characterization of cell cycle status of each cell is performed by two algorithms that use features of cell nuclei for classification (Vaisberg and Coleman, 2002). The cell cycle algorithm classifies each cell by total intensity of its nuclei (DNA content) as gap 1, synthesis, or gap 2. The condensation algorithm classifies each cell as having condensed or not-condensed DNA using the mean intensity and area of the nuclei. We use DNA condensation as a surrogate marker for mitosis.

Attributes used for further analysis in this paper are listed in Table I. These attributes were chosen for their biological information content and relatively low cross-correlation. There are many other attributes that can be extracted from the object or image, but we have found that the set described in Table I captures most of the significant morphological reorganizations.

Data Storage

Image, feature, and attribute data, as well as experimental information, such as cell type, compound concentrations, and plate bar codes, are stored in a custom Oracle database. Actual images are stored on a file server with pointers stored in the database. Many such configurations can be designed using custom or commercial software packages, and some have been described (Swedlow *et al.*, 2003).

Analysis of Quantitative Cellular Phenotypes across Cell Lines

A broader inspection of compound-induced cellular effects can be achieved by examining attribute changes across multiple cell lines. The compounds used in Fig. 2 can produce different effects in different cell types. Figure 3 shows quantification of image attributes for the same five compounds and concentrations on six cell lines: A498, A549, DU145, HUVEC, SF268, and SKOV3. These cell lines were chosen for their broad diversity of tissue source and genetic makeup (Scherf *et al.*, 2000), but

(14 μM), paclitaxel (0.37 nM), and cytochalasin D (4.0 μM). For each condition, subregions of the entire images are shown for DNA, Golgi, and microtubule markers. Compound concentrations were chosen to highlight diverse morphology changes seen in a single cell line. Scale bar (200 μm) applies to all images.

TABLE I
NAMES AND DESCRIPTIONS OF ATTRIBUTES MEASURED

Attribute	Description
Adhesion coefficient	Relative change in cell counts upon extensive washing of wells
Average area	Average area of interphase cells in a population
Average area, mitotic	Average area of mitotic cells in a population
Average cell axes ratio	Average ratio of short axis of an ellipse that has the same second moment as a cell mask to its long axis in interphase cells
Average cell axes ratio, mitotic	Average ratio of short axis of an ellipse that has the same second moment as a cell mask to its long axis in all mitotic cells
Average cell axes ratio, preanaphase	Average ratio of short axis of an ellipse that has the same second moment as a cell mask to its long axis in preanaphase mitotic cells
Average cell form factor	Average form factor (ratio of area to square of perimeter) for all interphase cells in a population
Average cell form factor, mitotic	Average form factor (ratio of area to square of perimeter) for all mitotic cells in a population
Average cell form factor, preanaphase mitotic	Average form factor (ratio of area to square of perimeter) for preanaphase mitotic cells in a population
Average cell moment	Average moment of masks of all interphase cells in a population
Average cell moment, mitotic	Average moment of masks of all mitotic cells in a population
Average cell moment, preanaphase mitotic	Average moment of masks of all preanaphase mitotic cells in a population
Average Golgi intensity	Average of mean lectin pixel intensities for all interphase cells in a population
Average Golgi intensity, mitotic	Average of mean lectin pixel intensities for all mitotic cells in a population
Average Golgi kurtosis	Average kurtosis, or shape of the probability distribution, of lectin in interphase cells.
Average Golgi kurtosis, mitotic	Average kurtosis, or shape of the probability distribution, of lectin in mitotic cells.
Average nuclei area	Average area of nuclei of interphase cells
Average nuclei axes ratio	Average ratio of short axis of an ellipse that has the same second moment as a nuclear mask to its long axis in interphase cells
Average nuclei intensity	Average of mean Hoechst pixel intensities within nuclear masks for all interphase cells in a population
Average total tubulin intensity over DNA mask, mitotic	Average of the sum of Dm1 α pixel intensities for all mitotic cells in a population. Only pixels contained within nuclear masks are included in computation
Average total tubulin intensity, interphase	Average of the sum of Dm1 α pixel intensities for all interphase cells in a population

(continued)

TABLE I (continued)

Attribute	Description
Average total tubulin intensity, mitotic	Average of the sum of Dm1 α pixel intensities for all mitotic cells in a population
Average total tubulin intensity, preanaphase mitotic	Average of the sum of Dm1 α pixel intensities for preanaphase mitotic cells in a population
Average tubulin intensity	Average of mean Dm1 α pixel intensities for all interphase cells in a population
Average tubulin kurtosis	Average kurtosis, or shape of the probability distribution, of Dm1 α intensity in interphase cells
Average tubulin kurtosis, mitotic	Average kurtosis, or shape of the probability distribution, of Dm1 α intensity in mitotic cells
Average tubulin kurtosis, preanaphase mitotic	Average kurtosis, or shape of the probability distribution, of Dm1 α intensity in preanaphase mitotic cells
G1 proportion	Proportion of cells in G1 phase of cell cycle in a population
G2 proportion	Proportion of cells in G2 phase of cell cycle in a population
Mitotic index	Proportion of mitotic cells in a population
Postanaphase proportion	Proportion of postanaphase mitotic cells in a population
Preanaphase proportion	Proportion of preanaphase mitotic cells in a population
Proportion of branched cells, interphase	Proportion of interphase cells with three or more well-formed projections in a population of all cells
Proportion of branched cells, mitotic	Proportion of mitotic cells with three or more well-formed projections in a population of all cells
Relative number of cells	Relative number of cells with respect to DMSO control cells
S proportion	Proportion of cells in S phase of cell cycle in a population
Standard deviation nuclei intensity	Average standard deviation of Hoechst pixel intensities within nuclear masks for all interphase cells in a population
Standard deviation tubulin intensity	Average standard deviation of Dm1 α pixel intensities for all interphase cells in a population

many other adherent cell types may be used. Average Golgi intensity changes in HUVEC cells seen in the presence of folimycin (yellow line and arrow, Fig. 3) are seen in all other cell lines except DU145 cells. Golgi intensity changes are significant in all CD-treated cells (red line, Fig. 3) except HUVEC cells (red arrow, Fig. 3). In SB 203580-treated cells, the average cell area of A498, HUVEC, and SKOV3 cells increases, whereas A549, DU145, and SF268 remain unchanged compared to DMSO (black arrow, Fig. 3). Subtle changes in morphology are also resolved. For example, differences in Golgi morphology between DMSO and paclitaxel seen in Fig. 2 are evident in the quantification shown in Fig. 3 (green arrow). Overall, each compound has a unique signature, or combination of attribute changes, across multiple cell lines.

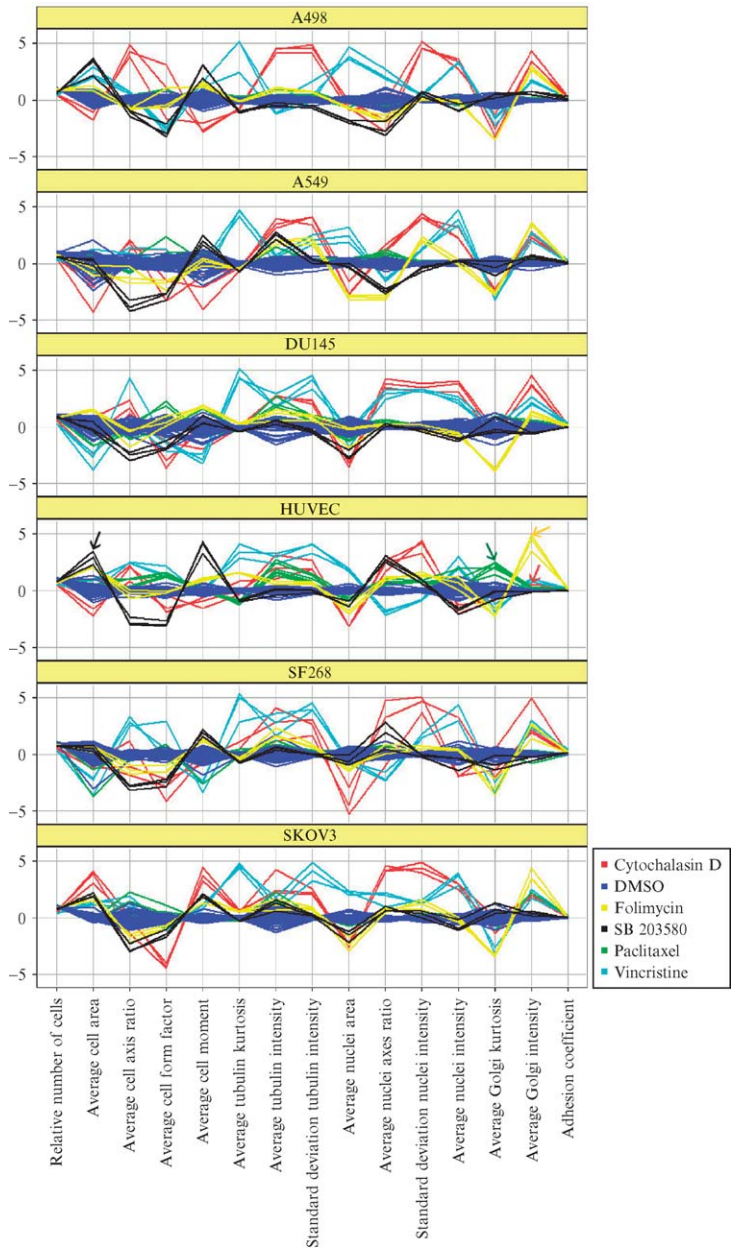


FIG. 3. Attribute profiles. A profile plot showing the relative changes in many attributes in six cell lines for six different compounds at a single concentration of each, where HUVEC

Comparing Attribute Changes Across Cell Types

1. Normalize attributes to the negative control by expressing their value as standard deviations from the mean of the DMSO control on each plate. Standard deviation of DMSO is usually stable from plate to plate; as a result pooled standard deviation from the experiment is used for the normalization.

2. Choose representative conditions for comparison, for example:
 - a. One compound at the same level of growth inhibition in different cell types
 - b. Multiple compounds at the same level of growth inhibition in a single cell type
 - c. One compound at the same concentration in multiple cell types
3. Use SpotFireTM or other graphical program to visualize changes (Fig. 3).

Additional information comes from analyzing dose–response curves (Fig. 4). Dose–response curves are best viewed using normalized features. Attributes change in both drug and cell line-dependent ways with respect to increasing doses of compound. For example, the cell axis ratio becomes larger (or the cells become rounder) in CD-treated cells in a dose-dependent manner when compared to DMSO-treated cells. A498 and HUVEC cells exhibit changes in the cell axis ratio at lowest concentrations of compound, whereas SF268 exhibit changes only after 1 μM , even though the relative effect on cell number is largely the same for all cell lines. If the cell axis ratio had been compared on different cell lines at a concentration lower than 1 μM , the morphology change of SF268 would not have been apparent.

Cell morphology changes are often detected at concentrations much lower than those needed to induce reduction in cell number. For example, SB 203580 shows subtle effects on cell number, but does change the cell axis ratio in a cell line-specific way (Fig. 4). SB 203580 causes the cell axis ratio to decrease as cells become elongated in HUVEC and A498 cells well before there are any significant changes in cell number. DU145 cell counts were most sensitive to paclitaxel treatment, but HUVEC cells were more

data are from the images seen in Fig. 2. The x axis lists representative attributes calculated using automated algorithms. Lines connect the values between attributes, but are themselves irrelevant. The relative difference of each attribute compared to DMSO is plotted on the y axis as the number of standard deviations from the mean of DMSO for each cell line. Individual values of DMSO deviate from the mean values for each cell line and are indicative of the noise in that feature. Colored arrows highlight feature changes in HUVEC cells for different compounds compared to other cell lines. Red, cytochalasin D (4.0 μM); dark blue, DMSO control; yellow, folimycin (4.8 nM); black, SB 203580 (0.1 mM); green, paclitaxel (0.37 nM); and light blue, vincristine (14 μM).

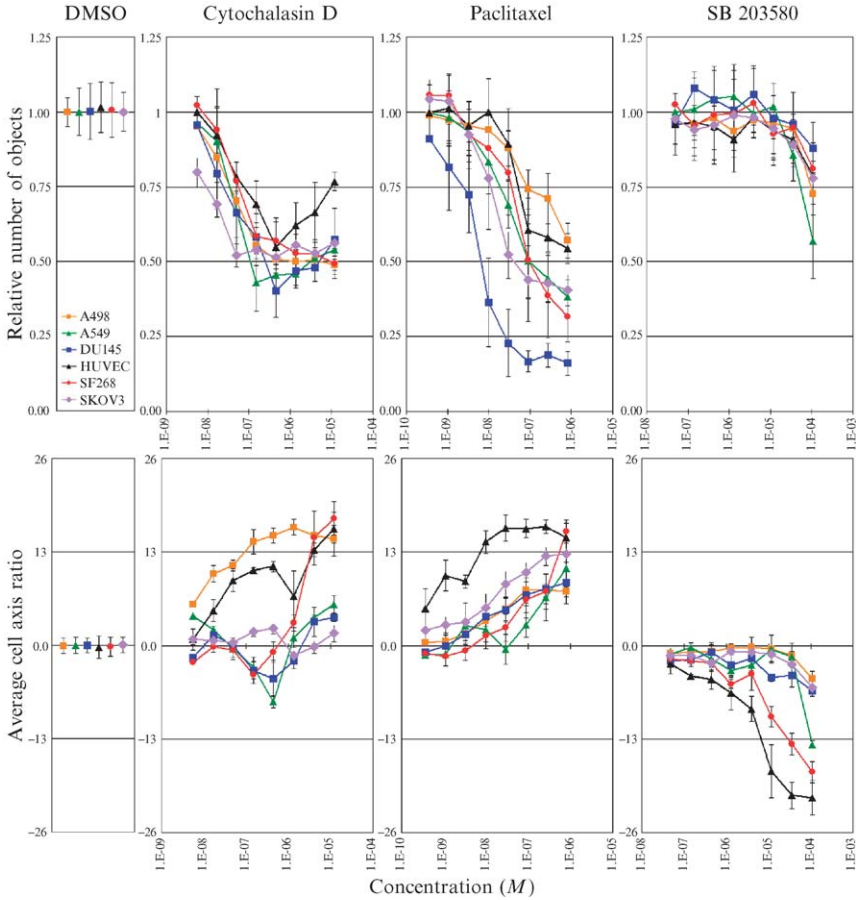


FIG. 4. Differential cell line response. The top row shows plots of the relative number of objects, compared to DMSO, against multiple concentrations of CD, paclitaxel, and SB 203580. The bottom row shows dose–response curves of the average cell axis ratio for the same concentrations of drug, where the results from each well are normalized by dividing by the median axis ratio obtained from eight DMSO control wells on the same plate. The DMSO controls are shown in the far left column. Each graph shows dose–response curves for six cell lines. Each data point is the average value of three replicate wells. Error bars show the standard deviation. Yellow square, A498; green triangle, A549; blue square, DU145; black triangle, HUVEC; red circle, SF268; purple diamond, SKOV3.

sensitive to changes in the cell axis ratio compared to the other cell lines. These data show that, depending on the mechanism of action of the compound, different cells types respond differentially. They also show that analysis of morphological attributes allows detection of effects of

compounds at concentrations lower than those that cause noticeable changes in growth. The underlying causes of the cell line-specific changes may not be clear, but we hypothesize that differential responses across multiple attributes and cell types are related to mechanisms of action of compounds that generate these responses.

To test whether a combination of attributes can be used to describe MOA-related changes in cells, we applied a set of 51 commercially available compounds (Table II) that broadly represent 12 MOA classes (Table III) to six cell lines using the protocols outlined earlier. We combined normalized values of attributes collected at a single concentration of compound into a vector, or signature. In order to visually investigate patterns in the multivariate signature data set, we used principal component analysis (PCA) on signature data (Fig. 5). This linear transformation converts a number of correlated variables into an equal number of uncorrelated variables, or principal components, in such a way that first components account for as much of the variability in the data set as possible (Johnson and Wichern, 1998).

Signature Construction and Visualization

1. Collect well level attributes for all conditions (all attributes in Table I from all cell lines were chosen for the signature used to construct the PCA plots in this example).
2. Normalize attributes to negative controls and build a table of attribute data for all cell lines of interest (columns) for each compound at each concentration (rows). One can report replicate data (multiple rows at each compound concentration) or use median results (single row for a compound concentration) at this step.
3. Compute principal components.
4. Plot first two or three principal components using a graphical rendering program.

In this data set the first three principal components account for 37% of the variation. Each point represents a signature, or a compound at a single concentration. A solid line connects the signatures of each compound according to concentration and forms a multivariate dose–response curve. The multivariate dose response of a compound is the piecewise linear interpolation of its signatures ordered by concentration. In Fig. 5 the compounds are color coded by the biochemical mechanism of action listed in Table II. The dose–response curves emanate from the origin of the untreated cells' signatures. Closely adjacent points in PCA space have similar values for most of the attributes across all of the cell lines. Although adjacent points for different compounds may represent substantially different absolute

TABLE II
TEST COMPOUNDS^a

Compound (abbreviation)	Class	Highest concentration used (μM)
2,5-Di- <i>t</i> -butylhydroquinone (DBHG)	ER-ATPase	20.6
Amsacrine	Topo 2	10.7
Butyrolactone I	Cdc2/cyclin B	10.8
Calmidazolium chloride (calmidazolium)	Calmodulin	6.6
CCCP	Mitochondria	18.0
Colcemid	Tubulin	0.6
Cyclopiazonic acid	ER-ATPase	13.6
Cytochalasin A (CA)	Actin	9.6
Cytochalasin B	Actin	9.5
Cytochalasin C	Actin	9.0
Cytochalasin D (CD)	Actin	1.4
Cytochalasin E	Actin	0.1
Cytochalasin H	Actin	2.1
Cytochalasin J	Actin	10.1
DMSO	Control	0
Dolostatin	Tubulin	5.4
Doxorubicin	Topo 2	4.7
Ellipticine	Topo 2	18.6
Epirubicin	Topo 2	2.4
Etoposide	Topo 2	7.8
Fluphenazine- <i>N</i> -2-chloroethane (FCE)	Calmodulin	8.7
Latrunculin B	Actin	11.6
Mastoparan, <i>Polistes jadwagae</i> (Mas PJ)	Gi and Go	2.8
Mastoparan, <i>Vespa xanthoptera</i> (Mas VX)	Gi and Go	2.9
Mastoparan, <i>Vespula lewisii</i> (Mas VL)	Gi and Go	3.1
Mitoxantrone	Topo 2	0.003
Nocodazole	Tubulin	0.03
Olomoucine	Cdc2/cyclin B	15.3
Phomopsin A	Tubulin	5.8
Phorbol 12,13-dibutyrate	PKC	0.7
Phorbol 12,13-didecanoate	PKC	0.3
Phorbol 12-myristate 13-acetate (PMA)	PKC	7.4
Purvalanol A	Cdc2/cyclin B	11.8
Roscovitine	Cdc2/cyclin B	12.9
Paclitaxel	Paclitaxel	2.9
Thapsigargin	ER-ATPase	7.0
Tubulazole-C	Tubulin	9.0
Vinblastine	Tubulin	0.4
Vincristine	Tubulin	1.6
GGTI-2147	GGTase1	3.4
GGTI-298	GGTase1	3.3
GGTI-286	GGTase1	3.7

(continued)

TABLE II (continued)

Compound (abbreviation)	Class	Highest concentration used (μM)
W-13	Calmodulin	13.1
W-12	Calmodulin	14.5
W-7	Calmodulin	12.2
W-5	Calmodulin	13.4
Mas 7	Gi and Go	3.1
Tyrphostin 9	Mitochondria	16.2
SKF 86002	p38 MAPK	15.4
PD 169316	p38 MAPK	12.7
SB 203580	p38 MAPK	12.1

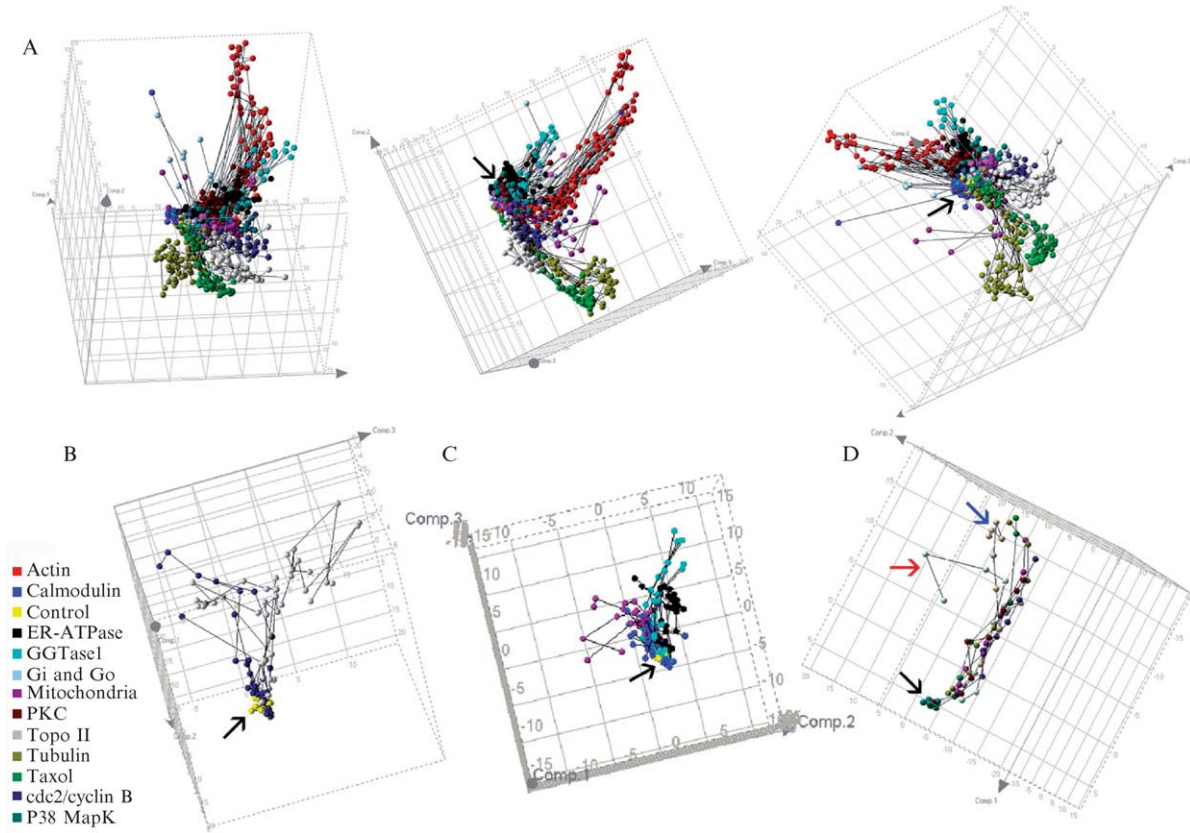
^a Compounds used, the abbreviations used in the text and figures, assigned MOA class, and highest concentrations used in the assay.

TABLE III
MECHANISM OF ACTION CLASSES OF TEST COMPOUNDS^a

MOA classes	Abbreviation	Number of representatives
Actin poisons	Actin	8
Calmodulin antagonists	Calmodulin	6
Endoplasmic reticulum Ca^{2+} -ATPase inhibitors	ER-ATPase	3
G protein-coupled receptor: $G_{\alpha i}$ and $G_{\alpha o}$ activators	Gi and Go	4
Geranylgeranyltransferase 1 inhibitors	GGTase1	3
Oxidative phosphorylation uncoupler	Mitochondria	2
p34 cdc2/cyclin B inhibitors	Cdc2/cyclin B	4
p38 MAP kinase inhibitors	p38 MAPK	3
PKC activator	PKC	3
Topoisomerase II inhibitors	Topo II	6
Tubulin destabilizers	Tubulin	7
Tubulin stabilizer	Paclitaxel	1 (two replications)

^a MOA classes, the abbreviations used in the text and figures, and the number of representative compounds in each class.

concentrations, the effects of compounds at those concentrations on all attributes are similar and thus likely have similar overall effects on all cell types and measurements used. Conversely, distant points represent distinct signatures, and thus attribute values, which reflect differences in effects of compounds. On the whole, signature data in PCA space show patterns of similarity and differences between mechanisms.



On the PCA plot, differences among compound classes increase with concentration. Clusters of compounds corresponding to MOA are apparent. For example, compounds known to inhibit tubulin, actin, mitochondria, GGTase, and Topo II are shown to form discrete clusters in principal component space as the concentrations of the compounds increase (Fig. 5A). The classes Topo II, cdc2/cyclinB, ER-ATPase, GGTase, and mitochondria inhibitors are difficult to visualize in Fig. 5A despite the three different angles, as their signatures are masked by signatures of other compounds. When shown separately, they have unique trajectories as well (Fig. 5B and C). Overall, the distinct clusters show that cellular phenotyping can discriminate different MOAs.

Compounds can affect a common, primary target by different molecular mechanisms. This is exemplified by application of the microtubule stabilizer paclitaxel and microtubule destabilizers (Jordan and Wilson, 2004) (Fig. 5A). Changes in overall cellular morphologies at increasing concentrations of these compounds cause the classes to diverge in the PCA view. The related attributes responsible for the divergence include differences in tubulin attributes, as expected (Peterson and Mitchison, 2002) (data not shown). This example shows that cellular phenotyping can precisely distinguish compounds that affect the same target, but in different ways.

Not all compounds assigned to a particular class may actually be specific for the target of interest. For example, cytochalasin A (CA) is similar to the other actin inhibitors at low concentrations, but deviates at higher concentrations (red arrow, Fig. 5D). CA is known also to inhibit microtubules (Himes and Houston, 1976), unlike cytochalasins B, D, E, H, or J. Consistent with this broader specificity, attributes relating to tubulin, cell shape, and adhesion are differentially affected by CA at high concentrations. Latrunculin inhibits actin by binding only to monomeric actin, whereas cytochalasins (Spector *et al.*, 1989) bind to both monomeric actin and

FIG. 5. Principal component plots of multivariate signatures. Color legend applies to A, B, and C. Black arrows point to location of DMSO control data points in all panels. (A) Three different views of the three-dimensional plots of the first three principal components of a multivariate signature made up of 32 attributes from each cell line (198 attributes total) for replicate data sets for the compounds listed in Table III. Data points for which the relative number of cells for any cell line was less than 50% of control are not shown. Lines connect increasing concentrations of a single compound in a single well. Dose-response curves are colored by MOA as listed in the color legend. (B) All compounds from the Topo II (white), cdc2/cyclin B (blue), and control (yellow) classes. (C) All compounds from the calmodulin (dark blue), ER-ATPase (black), GGTase1 (light blue), mitochondria (purple), and control (yellow) classes. (D) Actin MOA cluster. Each color is a different actin inhibitor, with DMSO at the bottom (black arrow). Red arrow points to cytochalasin A. Blue arrow points to latrunculin.

barbed ends of actin filaments. Latrunculin clusters with actin inhibitors, but it follows a slightly deviant trajectory (blue arrow, Fig. 5D). If the mechanisms of CA or latrunculin had been unknown, cellular phenotyping would have alerted the researcher to the probability that these compounds have unintended activities or deviant mechanisms compared to the rest of the compounds in that class.

While the markers used in this experiment label only tubulin, DNA, and Golgi, the test set of compounds broadly represents MOAs of interest in drug discovery, many of which affect pathways not related to cellular components visualized by the markers. For example, the actin inhibitor class is a well-defined cluster in the PCA graph even though the organization of the actin cytoskeleton was not measured directly. In fact, most of the molecular targets of the compounds tested were not measured directly, yet inhibition of those targets causes distinct morphological changes in the cell that were detected using surrogate markers.

Clustering and Classification of Compounds

Concentration-dependent differences can also be used to form a single phenotypic descriptor. Differences between the effects of a single compound at two different concentrations can be greater than differences between the effects caused by compounds from different classes. Moreover, different compounds might cause similar changes in cells at substantially different concentrations. We thus developed a quantitative measure that describes the biological effects of a compound regardless of the efficacy (i.e., an absolute concentration that induces a particular phenotype) and that is concentration independent using an average angle dissimilarity measure. This measure aligns multivariate dose responses by their potencies, or distances from control. For every pair of aligned points on two curves, the angle between points with the origin as the vertex is computed. The dissimilarity, d , is the average of these angles, mathematically:

$$d(D_1, D_2) = \frac{1}{b-a} \int_a^b \text{angle}(D_1(p), D_2(p)) dp$$

where $D_i(p)$ is the i th dose response at potency p and a and b are the minimum and maximum common potency obtained by the two drugs. The dissimilarity is scaled between 0 and p . A dissimilarity of 0 is obtained for dose responses that are identical over their common potency range. While there are multiple signatures, there is only one dissimilarity value per compound pair per dose response. Compared to PCA, this measure is more sensitive to the subtle changes that occur at lower concentrations of

compounds, which are hard to visualize by PCA. Further, the angle dissimilarity measure allows reduction of the effects of a compound on multiple cell types, attributes, and compound concentrations to a single value.

Calculation and Presentation of Angle Dissimilarity Measure Data

1. Calculate dissimilarity using normalized signature data.
2. Visualize using heat map ordered by hierarchical clustering derived from the angle dissimilarities for all pairs of compounds with the Ward clustering technique (Ward, 1963) (Fig. 6).

High correlation is seen within compound classes, as well as some similarities between compound classes. The microtubule, actin, G_i and G_o , PKC, and p38 mitogen-activated protein kinase classes all fall within their own branches of the tree (Fig. 6a–c, f, and g). There are two other branches that show mixed compound classes. In one branch (Fig. 6d), p34cdc2/cyclin B and Topo 2 classes intermix. In another branch (Fig. 6e), mitochondria, GGTase1, calmodulin, and ER Ca^{2+} -ATPase inhibitors intermix. Cytochalasin A does not appear very different from the actin inhibitors because the CA class deviates from the actin class only at high concentrations. However, the two classes are very similar at low concentrations, a finding that correlates with the PCA view. This type of plot would have been nearly impossible to interpret using the signatures at each concentration, as there would have been eight different branches for each compound and they would not have clustered together. Clustering compounds as shown in Fig. 6 is only possible by reducing multivariate dose–response data to a single variable.

While the heat map can be used to visualize correlations among compounds, the angle dissimilarity measure can be used to assign a single compound to a MOA class using a classifier (Tables IV and V). If one is interested in which class out of many predetermined classes a compound is most similar to (such as toxic vs nontoxic), this technique would be used. Overall, for 12 different MOAs we achieved 88% classification accuracy (45 out of 51 compounds were classified correctly). The confusion matrix for classification shows that all compounds in 9 out of 12 MOAs were classified correctly (Table V).

Classification

1. Find the optimal subset of attributes for a model to achieve <15% error using leave-one-out cross-validation (Hastie *et al.*, 2001). For the classification example presented earlier, we used mitotic index and G2 interphase proportion in SKOV3, average nuclei area in A549 and SF268,

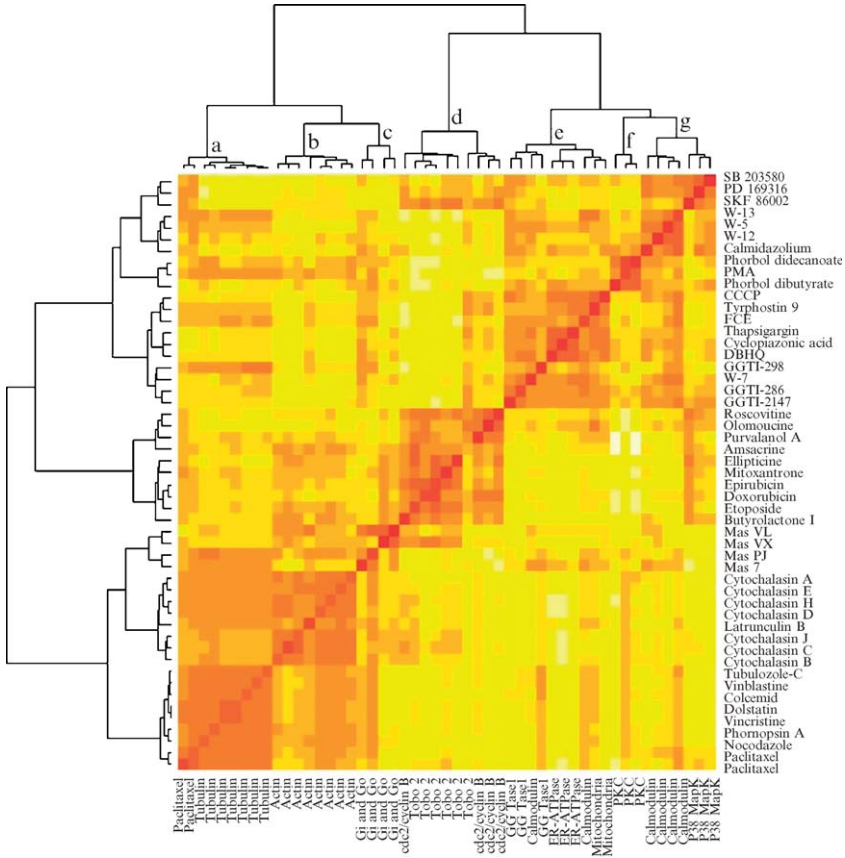


FIG. 6. Clustering using angle dissimilarity measure. Heat plot showing the correlation between the angle dissimilarity measure of each compound. The Ward clustering tree is shown on both axes. Compound names are seen along the y axis and their corresponding MOA classes are listed along the x axis. The color scale from red to yellow indicates compounds that are more or less correlated. Colors are scaled by row.

G1 mitotic proportion in DU145, average Golgi kurtosis in SF268, and average cell axis ratio in interphase cells in A498 cells.

2. Run leave-one-out cross-validation using the model to classify each compound using all other compounds as a training set (Hastie *et al.*, 2001).
 - a. Calculate the square distance of each compound to the mean of each MOA class (Table IV).
 - b. Assign classification of compound to class with minimum square distance.
3. Visualize using confusion matrix (Table V).

Many compounds are known to affect more than one target, and our analysis shows that at least some of the compounds tested have multiple activities. The compounds that were misclassified in Table V are calmidazolium, FCE, W-7, butyrolactone 1, amsacrine, and doxorubicin. FCE and W-7 are not tightly clustered with the other calmodulin members in the dendrogram, and their distances are larger as well. Review of the PCA plot (Fig. 5D) shows that the calmodulin class of compounds is not tightly clustered and has only one or two data points clearly distinguishable from control. There may not have been a sufficiently pronounced cellular response using this assay to allow robust classification of this MOA because we observed only weak responses for all attributes and cell types tested. However, if these compounds represented a chemical series that was predicted to be specific for calmodulin, these data would warn the researcher that perhaps the compounds have additional targets. Calmidazolium, amsacrine, butyrolactone, and doxorubicin are very narrowly misclassified, as judged by the fact that the difference in the distances to the misclassified class to their assigned class is less than 0.2 (compared to a median separation distance of 2.5 ± 1.8 and a max of p^2 for all compounds to all classes) and they are tightly clustered with the other members in their class in the dendrogram. Misclassification of the cdc2/cyclin B and Topo II inhibitors, butyrolactone 1, amsacrine, and doxorubicin, is expected, as compounds from both mechanisms intermix in both the clustering (Fig. 6) and the PCA plot (Fig. 5C) and may indicate these compounds inhibit the same pathway, albeit via different targets or mechanisms. Indeed, cdc2 and topoisomerase II have been shown to interact (Escargueil *et al.*, 2001).

Conclusions

We have shown that it is possible to successfully classify inhibitors of a diverse array of targets using Cytometrix technologies using relatively generic morphological changes in a number of different cell types by successfully classifying 45 out of 51 known compounds into 12 distinct MOA classes. We show that compounds that inhibit the same target via different MOAs can be distinguished, for example, the microtubule-binding taxanes and vinca alkaloids. Moreover, Cytometrix technologies can identify unexpected activities in members of a closely related compound series, for example, microtubule binding in the actin poison cytochalasin A. In a previous publication, Cytometrix technologies were used to discover a novel inhibitor of carbonyl reductase 1 (Tanaka *et al.*, 2005). While the techniques outlined are specific for this one experiment, they are applicable to many different biological assay designs where cell lines, markers, time points, and reagents are varied.

TABLE IV
CLASSIFICATION RESULTS^a

Treatment name	Assigned MOA	Classification	Correctness	Actin	Calmodulin	Cdc2/cyclin B	ER-ATPase	GGTase1	Gi and Go	Mitochondria	p38 MAPK	PKC	Paclitaxel	Topo 2	Tubulin
Cytochalasin A	Actin	Actin	Correct	0.2	3.1	3.3	5.3	2.8	1.6	2.3	3.6	1.7	0.7	4.5	1.1
Cytochalasin B	Actin	Actin	Correct	0.2	3.7	2.5	6.2	3.9	1.2	2.5	3.2	2.5	1.1	3.2	1.9
Cytochalasin C	Actin	Actin	Correct	0.3	4.6	2.1	6.8	4.2	1.4	3.4	3.0	3.3	1.5	2.5	3.1
Cytochalasin D	Actin	Actin	Correct	0.1	3.0	2.0	6.4	3.1	1.4	2.3	2.7	2.9	0.5	2.9	0.8
Cytochalasin E	Actin	Actin	Correct	0.2	3.2	3.1	4.8	2.9	1.3	2.5	3.3	2.4	0.6	3.7	0.9
Cytochalasin H	Actin	Actin	Correct	0.1	3.3	1.8	6.5	3.1	1.3	2.4	3.0	3.2	0.5	2.7	0.8
Cytochalasin J	Actin	Actin	Correct	0.2	4.4	2.9	6.7	4.0	1.4	3.1	3.0	2.2	1.2	3.1	2.3
Latrunculin B	Actin	Actin	Correct	0.7	3.5	3.1	3.0	2.9	1.0	1.6	3.4	3.2	0.9	3.5	1.4
W-12	Calmodulin	Calmodulin	Correct	4.1	0.8	3.4	2.5	1.4	2.0	1.5	1.6	2.1	1.5	5.3	4.6
W-13	Calmodulin	Calmodulin	Correct	3.1	0.1	2.8	1.5	0.5	1.7	0.4	1.1	1.8	0.6	4.5	1.3
W-5	Calmodulin	Calmodulin	Correct	6.2	0.3	3.5	1.8	0.6	2.6	1.0	1.4	2.6	1.3	5.2	3.2
Olomoucine	Cdc2/cyclin B	Cdc2/cyclin B	Correct	7.0	2.8	0.7	2.5	2.5	3.8	1.1	1.4	5.6	2.5	1.6	7.8
Purvalanol A	Cdc2/cyclin B	Cdc2/cyclin B	Correct	3.8	3.9	0.6	3.7	2.6	2.8	1.5	2.1	6.9	1.8	1.3	4.5
Roscovitine	Cdc2/cyclin B	Cdc2/cyclin B	Correct	5.7	3.9	0.3	3.8	3.3	4.1	2.1	1.2	6.7	2.6	1.0	7.8
Cyclopiazonic acid	ER-ATPase	ER-ATPase	Correct	10.0	1.5	3.6	0.2	1.2	2.6	0.6	2.8	2.9	3.8	6.3	7.3
DBHQ	ER-ATPase	ER-ATPase	Correct	8.8	1.4	4.3	0.2	0.9	2.3	0.7	2.9	2.6	3.2	6.3	5.9
Thapsigargin	ER-ATPase	ER-ATPase	Correct	8.3	1.5	3.2	0.2	0.9	2.6	0.4	3.0	3.2	3.4	5.4	6.5
GGTI-2147	GGTase1	GGTase1	Correct	6.3	0.8	3.4	1.2	0.4	2.8	0.7	1.2	2.6	2.6	5.7	5.3
GGTI-286	GGTase1	GGTase1	Correct	8.9	0.9	4.1	0.9	0.4	3.3	0.8	1.8	2.3	2.5	6.4	5.2
GGTI-298	GGTase1	GGTase1	Correct	4.1	1.3	3.5	2.0	0.9	1.9	1.3	2.7	3.0	1.0	4.4	1.2
Mas 7	Gi and Go	Gi and Go	Correct	3.5	1.1	3.3	1.6	1.3	0.7	1.2	3.2	2.6	1.1	4.1	1.8
Mas PJ	Gi and Go	Gi and Go	Correct	1.1	1.8	4.2	2.8	1.7	0.6	1.4	3.8	1.7	0.6	4.5	0.7
Mas VL	Gi and Go	Gi and Go	Correct	2.5	3.0	3.2	2.8	3.0	0.3	2.1	4.0	3.3	2.8	3.2	4.6

Mas VX	Gi and Go	Gi and Go	Correct	3.1	3.8	2.0	4.1	3.9	1.4	2.6	3.7	3.6	2.1	1.8	5.2
CCCP	Mitochondria	Mitochondria	Correct	6.6	1.2	2.8	0.8	0.9	2.2	0.2	2.5	3.2	2.8	4.8	6.1
Tyrphostin 9	Mitochondria	Mitochondria	Correct	4.2	1.0	2.1	0.7	0.9	2.1	0.2	2.6	2.9	1.2	4.2	2.5
PD 169316	P38 MapK	P38 MapK	Correct	5.7	1.6	1.6	3.5	1.6	4.2	2.2	0.4	3.6	1.2	2.6	7.5
SB 203580	P38 MapK	P38 MapK	Correct	6.5	1.4	2.4	2.5	1.7	3.4	2.0	0.5	2.4	1.6	3.7	8.6
SKF 86002	P38 MapK	P38 MapK	Correct	5.2	3.1	0.9	4.2	2.1	4.6	2.5	0.6	5.1	1.5	1.3	7.4
Phorbol dibutyrate	PKC	PKC	Correct	7.8	2.9	6.2	2.3	2.5	3.0	2.5	2.9	0.6	3.8	8.7	8.5
Phorbol didecanoate	PKC	PKC	Correct	4.2	2.7	6.4	3.4	2.5	2.6	2.6	3.3	0.2	2.4	9.6	4.6
PMA	PKC	PKC	Correct	2.0	2.4	6.3	3.5	1.8	2.3	1.9	3.5	0.1	1.3	7.7	1.8
Paclitaxel	Paclitaxel	Paclitaxel	Correct	1.5	1.7	1.9	4.4	1.8	1.6	1.7	1.2	3.3	0.0	2.3	0.3
Paclitaxel	Paclitaxel	Paclitaxel	Correct	1.2	1.3	2.9	4.2	1.8	1.3	1.6	1.4	2.1	0.0	3.2	0.2
Ellipticine	Topo 2	Topo 2	Correct	3.3	5.7	1.1	5.6	5.0	2.5	3.5	1.5	5.7	2.5	0.9	7.5
Epirubicin	Topo 2	Topo 2	Correct	4.4	6.0	0.8	6.0	5.6	2.2	3.5	2.1	6.7	2.0	0.3	5.6
Etoposide	Topo 2	Topo 2	Correct	2.8	5.4	0.6	5.3	4.2	2.4	2.8	2.0	8.1	1.7	0.3	5.4
Mitoxantrone	Topo 2	Topo 2	Correct	3.3	3.9	1.2	5.0	3.7	1.8	3.2	1.4	5.2	1.3	0.9	4.6
Colcemid	Tubulin	Tubulin	Correct	1.1	1.7	3.5	3.8	1.3	1.3	1.7	3.5	2.0	0.1	4.3	0.0
Dolostatin	Tubulin	Tubulin	Correct	1.1	1.9	3.0	3.6	1.6	1.1	1.7	3.7	2.9	0.1	3.6	0.0
Nocodazole	Tubulin	Tubulin	Correct	0.8	2.1	3.8	3.5	1.6	0.8	1.5	5.0	1.7	0.2	4.3	0.1
Phomopsin A	Tubulin	Tubulin	Correct	0.7	2.1	3.6	4.5	1.9	1.0	1.9	3.6	2.0	0.2	4.3	0.1
Tubulozole-C	Tubulin	Tubulin	Correct	1.2	1.7	3.2	3.4	1.3	1.2	1.5	3.8	2.5	0.1	4.1	0.0
Vinblastine	Tubulin	Tubulin	Correct	1.2	1.8	3.3	3.3	1.4	1.2	1.6	3.9	2.6	0.1	4.1	0.0
Vincristine	Tubulin	Tubulin	Correct	1.1	1.9	3.1	3.5	1.6	1.2	1.6	3.7	2.8	0.1	3.7	0.0
Calmidazolium	Calmodulin	p38 MAPK	Incorrect	6.0	1.1	3.4	1.9	1.5	1.8	1.5	1.0	1.6	1.7	5.5	6.8
FCE	Calmodulin	Mitochondria	Incorrect	3.8	1.0	2.8	0.8	0.5	1.5	0.4	2.2	2.8	1.0	4.2	1.8
W-7	Calmodulin	GGTase1	Incorrect	8.0	1.0	2.8	0.8	0.7	1.9	1.1	2.2	3.7	2.5	4.3	4.8
Butyrolactone I	Cdc2/cyclin B	Topo 2	Incorrect	2.5	4.9	1.5	6.1	5.3	2.2	4.0	1.5	5.5	2.2	1.2	6.2
Amsacrine	Topo 2	Cdc2/cyclin B	Incorrect	2.3	3.3	1.0	3.6	3.1	1.8	2.0	1.6	6.0	1.7	1.1	4.6
Doxorubicin	Topo 2	Cdc2/cyclin B	Incorrect	5.4	4.5	0.4	4.9	4.0	2.6	2.7	1.6	8.0	2.0	0.6	5.8

^aThe classification assignment of each compound and assessment of correctness and the square distance of each compound to the mean of each MOA class are shown. Lower numbers denote more similarity.

TABLE V
CONFUSION MATRIX OF CLASSIFICATION^a

	Actin	Calmodulin	Cdc2/cyclin B	ER-ATPase	GGTase1	Gi and Go	Mitochondria	p38 MAPK	PKC	Paclitaxel	Topo 2	Tubulin	Grand total
Actin	8												8
Calmodulin		3											3
Cdc2/cyclin B			3								2		5
ER-ATPase				3									3
GGTase1		1			3								4
Gi and Go						4							4
Mitochondria		1					2						3
p38 MAPK		1						3					4
PKC									3				3
Paclitaxel										2			2
Topo 2			1								4		5
Tubulin												7	7
Grand total	8	6	4	3	3	4	2	3	3	2	6	7	51

^a Summary of Table IV such that compounds in the MOA classes listed in the first row are scored as having been classified as any of the other MOAs listed in the first column. A correct classification is one where the compound being classified is classified as its assigned MOA, or along the diagonal. Misclassifications are seen off the diagonal.

The multivariate approach to analyzing image data has many additional applications beyond the data set discussed in this chapter, including but not limited to siRNA and compound screening, toxicity profiling, and clinical pharmacodynamics. There have been many studies in the literature to which this approach may be applied, for example, to identify new compounds similar to desired control compounds or with novel phenotypes (Haggarty *et al.*, 2000; Mayer *et al.*, 1999; Yarrow *et al.*, 2003) or the approach could have been applied to characterize morphological changes observed in siRNA screens (Goshima and Vale, 2003; Kiger *et al.*, 2003) or switch-of-function mutation screens (Heo and Meyer, 2003). The approach can be applied using siRNA phenotypes as control morphologies to suggest primary screening targets, as has been done manually (Eggert *et al.*, 2004). When coupled with toxicity data and profiles, appropriate assay combinations may enable safety profiling of compounds.

Other high-content imaging assays have been described and most share similar approaches to data collection and image quantification as those described here (Abraham *et al.*, 2004). However, most of these data are collected using one cell line with many markers, and it is analyzed primarily

using dose responses of single features, which tend to be simple averages of image data that do not distinguish subpopulations of cells. [Perlman *et al.* \(2004\)](#) tested 100 compounds on HeLa cells using 11 markers and presented clustering data using a different method than that described here. While they used a different set of compounds, it is noteworthy to state that in this publication only 61 of the compounds showed any morphological changes, and we suspect that additional cell lines would have increased the number of responding compounds. Also, while they used different analytical techniques, some of the mechanistic classes did not cluster together, including the actin inhibitors, which clustered robustly in our study. It remains to be seen how the results from both studies would change using the different analyses.

Most cell-based assays that measure cell death or gene or protein expression quantify averages for cell populations or the concentration at which a singular effect is achieved. A cell-based screening approach that uses multiple GI50 measurements was developed at the National Cancer Institute ([Weinstein *et al.*, 1997](#)). This approach is based on constructing profiles of compounds based on their growth inhibition effects in a panel of 60 cell lines and then using that multivariate profile to compare mechanisms. The inherent limitations of this method are the lack of activity seen for compounds that do not cause significant growth inhibition and the logistics of handling many cell lines. The strength of our approach lies in the discovery that morphological changes in cells can precede measurable growth inhibition by over two orders of magnitude and that these changes are cell line dependent. Thus, quantification of the morphology of cellular structures may be more predictive of an *in vivo* response.

Microarray technology has been used for discovering molecular targets and for classifying compounds ([Gunther *et al.*, 2003](#); [Hughes *et al.*, 2000](#); [Marton *et al.*, 1998](#)). One of the differences between microarrays and Cytometrix technologies is that the former approach generally relies on a measure of the population average. In contrast, Cytometrix algorithms measure the response of individual cells and thus are sensitive to changes in as low as 10% of the population (data not shown). A 10% change may be in the noise level for microarray analysis. Ultimately, we expect Cytometrix technologies and microarray analysis to be complementary. Integration of data sets from these two forms of analysis therefore likely will have additional advantages.

In conclusion, we have introduced a new approach for generically quantifying the effects of compounds on cells. Cytometrix technologies are highly flexible and can be configured to characterize compounds on multiple cell lines, time points, markers, and concentrations, depending on the application. Cellular compound activity can be measured quantitatively

using a multivariate experimental and analytical approach. We showed that the differential effects of compounds on cells of different genetic backgrounds facilitate discrimination among MOAs and that measuring a full range of biological activity provides a powerful approach to novel compound characterization and identification.

Acknowledgments

We gratefully acknowledge G. Alexander, J. Armstrong, R. de la Rosa, C. Elliot, J. Finer, A. Fritsch, E. Kiro, S. Kriz, P. Julius, C. Lei, P. Ling, P. McMahon, R. Moody, B. Muller, K. Penhall-Wilson, D. Platz, L. Reddy, A. Rao, C. Shumate, G. Singh, D. Snyder, K. Suekawa, B. Tang, R. Venkat, W. Zink, S. Wu for their technical contributions to the project; C. Beraud, L. Goldstein, B. Jack, F. Malik, J. Sabry, S. Smith, J. Spudich, R. Vale, and K. Wood for their support and D. Drubin, C. Hazuka, D. Lenzi, M. Maxon, S. Ramchandani, and R. Sakowicz for their support and critical reading of the manuscript.

References

- Abraham, V. C., Taylor, D. L., and Haskins, J. R. (2004). High content screening applied to large-scale cell biology. *Trends Biotechnol.* **22**, 15–22.
- Blose, S. H., Meltzer, D. I., and Feramisco, J. R. (1984). 10-nm filaments are induced to collapse in living cells microinjected with monoclonal and polyclonal antibodies against tubulin. *J. Cell Biol.* **98**, 847–858.
- Cong, G., and Vaisberg, E. (2005). Extracting shape information contained in cell images. U.S. Patent 6,956,961.
- Eggert, U. S., Kiger, A. A., Richter, C., Perlman, Z. E., Perrimon, N., Mitchison, T. J., and Field, C. M. (2004). Parallel chemical genetic and genome-wide RNAi screens identify cytokinesis inhibitors and targets. *PLoS Biol.* **2**, 0001–0009.
- Escargueil, A. E., Plisov, S. Y., Skladanowski, A., Borgne, A., Jeijer, L., Gorbysky, G. J., and Larsen, A. K. (2001). Recruitment of cdc2 kinase by DNA topoisomerase II is coupled to chromatin remodeling. *FASEB J.* **15**, 2288–2290.
- Goshima, G., and Vale, R. D. (2003). The roles of microtubule-based motor proteins in mitosis: Comprehensive RNAi analysis in the *Drosophila* S2 cell line. *J. Cell Biol.* **162**, 1003–1016.
- Gunther, E. C., Stone, D. J., Gerwien, R. W., Bento, P., and Heyes, M. P. (2003). Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles *in vitro*. *Proc. Natl. Acad. Sci. USA* **100**, 9608–9613.
- Haggarty, S. J., Mayer, T. U., Miyamoto, D. T., Fathi, R., King, R. W., Mitchison, T. J., and Schreiber, S. L. (2000). Dissecting cellular processes using small molecules: Identification of colchicine-like, paclitaxel-like and other small molecules that perturb mitosis. *Chem. Biol.* **7**, 275–286.
- Hastie, T., Tibshirani, R., and Friedman, J. H. (2001). Model assessment and selection. In “The Elements of Statistical Learning: Data Mining, Inference, and Prediction,” pp. 214–217. Springer-Verlag, New York.
- Heo, W. D., and Meyer, T. (2003). Switch-of-function mutants based on morphology classification of Ras superfamily small GTPases. *Cell* **113**, 315–328.
- Himes, R. H., and Houston, L. L. (1976). The action of cytochalasin A on the *in vitro* polymerization of brain tubulin and muscle G-actin. *J. Supramol. Struct.* **5**, 81–90.

- Hughes, T. R., Marton, M. J., Jones, A. R., Roberts, C. J., Stoughton, R., Armour, C. D., Bennett, H. A., Coffey, E., Dai, H., He, Y. D., Kidd, M. J., King, A. M., Meyer, M. R., Slade, D., Lum, P. Y., Stepaniants, S. B., Shoemaker, D. D., Gachotte, D., Chakraburty, K., Simon, J., Bard, M., and Fried, S. H. (2000). Functional discovery via a compendium of expression profiles. *Cell* **102**, 109–126.
- Johnson, R. A., and Wichern, D. W. (1998). Principal components. In “Applied Multivariate Statistical Analysis,” 3rd Ed. pp. 356–380. Prentice-Hall, Upper Saddle River, NJ.
- Jordan, M. A., and Wilson, L. (2004). Microtubules as a target for anticancer drugs. *Nature Rev. Cancer* **4**, 253–265.
- Kiger, A. A., Baum, B., Jones, S., Jones, M. R., Coulson, A., Echeverri, C., and Perrimon, N. (2003). A functional genomic analysis of cell morphology using RNA interference. *J. Biol.* **2**, 27.1–27.15.
- Marton, M. J., DeRisi, J. L., Bennett, H. A., Iyer, V. R., Meyer, M. R., Roberts, C. J., Stoughton, R., Burchard, J., Slade, D., Dai, H., Bassett, D. E., Jr., Hartwell, L. H., Brown, P. O., and Friend, S. H. (1998). Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nature Med.* **4**, 1293–1301.
- Mayer, T. U., Kapoor, T. M., Haggarty, S. J., King, R. W., Schreiber, S. L., and Mitchison, T. J. (1999). Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. *Science* **286**, 971–974.
- 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.
- Peterson, J. R., and Mitchison, T. J. (2002). Small molecules, big impact: A history of chemical inhibitors and the cytoskeleton. *Chem. Biol.* **12**, 1275–1285.
- Polishchuk, R. S., and Mironov, A. A. (2004). Structural aspects of Golgi function. *Cell. Mol. Life Sci.* **61**, 146–158.
- Prentis, R. A., Lis, Y., and Walker, S. R. (1988). Pharmaceutical innovation by the seven UK-owned pharmaceutical companies. *Br. J. Clin. Pharmacol.* **3**, 387–396.
- Pritchard, J. F., Jurima-Romet, M., Reimer, M. L., Mortimer, E., Rolfe, B., and Cayen, M. N. (2003). Making better drugs: Decision gates in non-clinical drug development. *Nature Rev. Drug Discov.* **7**, 542–553.
- Scherf, U., Ross, D. T., Waltham, M., Smith, L. H., Lee, J. K., Tanabe, L., Kohn, K. W., Reinhold, W. C., Myers, T. G., Andrews, D. T., Scudiero, D. A., Eisen, M. B., Sausville, E. A., Pommier, Y., Bostein, D., Brown, P. O., and Weinstein, J. N. (2000). A gene expression database for the molecular pharmacology of cancer. *Nature Genet.* **24**, 236–244.
- Spector, I., Shochet, N. R., Blasberger, D., and Kashman, Y. (1989). Latrunculins: Novel marine macrolides that disrupt microfilament organization and affect cell growth. I. Comparison with cytochalasin D. *Cell Motil. Cytoskel.* **13**, 127–144.
- Swedlow, J. R., Goldberg, I., Brauner, E., and Sorger, P. K. (2003). Informatics and quantitative analysis in biological imaging. *Science* **300**, 100–102.
- Tanaka, M., Bateman, R., Rauh, D., Vaisberg, E., Ramchandani, S., Zhang, C., Hansen, K. C., Burlingame, A. L., Trautman, J. K., Shokat, K. M., and Adams, C. L. (2005). An unbiased cell morphology-based screen for new, biologically active small molecules. *PLoS Biol.* **3**, 0764–0776.
- Vaisberg, E., and Coleman, D. A. (2002). Classifying cells based on information contained in cell images U.S. Patent 6,876,760..
- Vaisberg, E. A., Cong, G., and Hsien-Hsun, W. (2002). Image analysis of the Golgi complex. World International Property Organization 02/067195.
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **58**, 236–244.

- Weinstein, J. N., Myers, T. G., O'Connor, P. M., Friend, S. H., Fornace, A. J., Jr., Kohn, K. W., Fojo, T., Bates, S. E., Rubinstein, L. V., Anderson, N. L., Buolamwini, J. K., van Osdol, W. W., Monks, A. P., Scudiero, D. A., Sausville, E. A., Zaharevitz, D. W., Bunow, B., Viswanadhan, V. N., Johnson, G. S., Wittes, R. E., and Paull, K. D. (1997). An information-intensive approach to the molecular pharmacology of cancer. *Science* **275**, 343–349.
- Yarrow, J. C., Feng, Y., Perlman, Z. E., Kirchhausen, T., and Mitchison, T. J. (2003). Phenotypic screening of small molecule libraries by high throughput cell imaging. *Comb. Chem. High Throughput Screen.* **6**, 279–286.

[25] High-Content Screening: Emerging Hardware and Software Technologies

By SEUNGTAEK LEE and BONNIE J. HOWELL

Abstract

The field of high-content screening has flourished since 2000 with advancements in automated fluorescence microscopy technologies, fluorescent labeling techniques, and sophisticated image analysis software. Through the use of these technologies, researchers can now monitor cellular and molecular events in individual cells *in vitro* following drug treatment or RNAi and rapidly screen compound and siRNA libraries. This chapter discusses current and next-generation hardware and software features and capabilities.

Introduction

High-content screening (HCS) has gained tremendous popularity in the past few years in the drug discovery industry from advances in fluorescence microscopy and automated screening technologies. HCS provides an opportunity to rapidly screen chemical or siRNA libraries by imaging subcellular and molecular events of individual cells with an automated fluorescent microscope. HCS gets its name from the rich and diverse set of information provided by analyzing the phenotype of whole cells. The ability to multiplex with multiple markers or fluorescent probes and even multiple cell lines simultaneously allows for more efficient screens for target identification, target validation, lead identification, and lead optimization. Although fluorescence microscopy and digital image analysis have been around for decades, the complete integration and automation