A Robust High-Content Imaging Approach for Probing the Mechanism of Action and Phenotypic Outcomes of Cell-Cycle Modulators

Jeffrey J. Sutherland¹, Jonathan Low², Wayne Blosser², Michele Dowless², Thomas A. Engler³, and Louis F. Stancato²

Abstract

High-content screening is increasingly used to elucidate changes in cellular biology arising from treatment with small molecules and biological probes. We describe a cell classifier for automated analysis of multiparametric data from immunofluorescence microscopy and characterize the phenotypes of 41 cell-cycle modulators, including several protein kinase inhibitors in preclinical and clinical development. This method produces a consistent assessment of treatment-induced phenotypes across experiments done by different biologists and highlights the prevalence of nonuniform and concentration-dependent cellular response to treatment. Contrasting cell phenotypes from high-content screening to kinase selectivity profiles from cell-free assays highlights the limited utility of enzyme potency ratios in understanding the mechanism of action for cell-cycle kinase inhibitors. Our cell-level approach for assessing phenotypic outcomes is reliable, reproducible and capable of supporting medium throughput analyses of a wide range of cellular perturbations. Mol Cancer Ther; 10(2); 242–54. ©2011 AACR.

Introduction

The characterization of cell populations with immunofluorescence microscopy, or high-content screening, allows a detailed understanding of the effect of small molecule modulators on the mitotic cell cycle. By quantifying the intensity, localization, and morphology of 4 to 5 markers in individual cells, such experiments typically produce on the order of 10⁵ to 10⁶ observations per treatment. In recognition of the heterogeneity of cell populations (1), several methods have been proposed for analyzing treatment-induced perturbations by cellular imaging. Such approaches assign a phenotype to individual cells, using rules to emulate their classification by biologists (2–4), or by applying nonsupervised (5–9) or supervised (10, 11) multivariate analysis of cytological features. Methods for analysis of high-content imaging experiments have been reviewed elsewhere (12, 13).

Chemical modulation of the mitotic cell cycle has proven to be effective in treating cancer. A number of approved chemotherapeutic agents disrupt DNA replication [e.g., the topoisomerase inhibitors (14) topotecan and camptothecin] or microtubule dynamics (15; e.g., the tubulin modulators paclitaxel and vinca alkaloids). The mitotic cell cycle is regulated by many kinases, and the search for small molecule inhibitors has produced a number of agents in preclinical and clinical development.

The roles of kinases such as CDK1, AURKA, AURKB, and PLK1 in the G2-M checkpoint are well-established (16). The CDK1-cyclin B complex regulates entry into mitosis; loss of CDK1 function results in arrest at the G2-M boundary and enrichment of cell populations having large nuclei with 4N DNA present as diffuse chromatin. Because of their role at the G2-M checkpoint, inhibition of other G2-M kinases in addition to CDK1 is expected to result in a phenotype consistent with selective CDK1 inhibition (17). The kinase AURKA is involved in centrosome regulation, and its inhibition manifests itself via enrichment of cells in prometaphase. AURKA promotes bipolar spindle assembly (18), and mutations in this kinase prevent centrosome separation leading to the formation of monopolar spindles (19). In addition to its role in cytokinesis, AURKB activates the spindle-assembly checkpoint, and manifestation of AURKA inhibition requires a functional spindle-assembly checkpoint. For this reason, dual aurora A/B inhibitors are expected to yield a phenotype consistent with selective AURKB inhibition (20). Finally, PLK1 functions in spindle formation, chromosome segregation and cytokinesis, the inhibition...
of which results in failure to establish a bipolar spindle in prometaphase (21).

Protein kinases have a high degree of structural homology, and activity of small molecule inhibitors against proteins other than the intended target (i.e., off-target activity) is frequently observed in cell-free assays. To understand the relationship between selectivity and effects on cell populations, we developed a classifier of cellular phenotype in HCT-116 cells (a cell line for colorectal carcinoma), and applied it to kinase and nonkinase modulators of the cell cycle. The approach yields a highly-reproducible assessment of changes in cell populations induced by different treatments or different concentrations of the same treatment, and highlights the limited utility of kinase selectivity panels in selecting inhibitors that exhibit the desired phenotype.

Materials and Methods

Cell-cycle modulators and high-content imaging

Various inhibitors of proteins involved in cell-cycle regulation were selected from the Lilly corporate collection (Table 1). Where available, compounds were purchased from commercial vendors. Several reported kinase inhibitors not available for purchase were synthesized internally, using synthetic schemes described in the public domain. All compounds have ≥ 95% purity. Experiments were done by 3 biologists over a 6-month period, to study phenotypes induced by cell-cycle modulators. Experiments 2 and 3 were designed to specifically probe reproducibility of phenotypes, whereas experiment 1 was designed to characterize a large collection of kinase inhibitors. HCT-116 cells were plated onto 96-well dishes, treated with compounds in 10-point concentration curves, and imaged on the Arrayscan VTI platform (see Supplementary Methods). Cytological features of cells were captured with the Target Activation bio-application bundled with the imaging instrument. The selected cytological features quantify a number of important changes in cells undergoing mitosis (Supplementary Table 1). Objects are defined from nuclei identification, and mostly correspond to individual cells; clusters of nuclei from treatments causing polyploidy are captured as 1 object with ≥8N DNA content and daughter nuclei in anaphase are captured as 2 objects.

Normalization of cellular features

Numerical values of features from the instrument software are log2 transformed to increase the normality of distribution across cell populations (base 2 is convenient for counting doublings of intensity, e.g., DNA content of 2N, 4N, 8N, etc.). To account for plate-to-plate variation in cytological features (i.e., changes arising from variation in antibody staining intensity, incubator conditions, etc.), individual values for each feature are converted to Z-scores, using the mean and standard deviation obtained by pooling dimethyl sulfoxide (DMSO)-treated cells from 8 negative control wells (i.e., normalization of features on a per-plate basis). The intensity distributions for cells in 8 positive control wells containing 0.2 μmol/L of nocodazole are visually assessed for each plate and channel to verify for consistency across plates after feature scaling (Supplementary Fig. 1). This method of normalization is adequate to control for plate-to-plate variability within an experiment. All further analysis uses scaled values of cytological features.

Quantifying antiproliferative effects of compound treatments

For each well, the cell density is calculated by counting the number of objects (cells) per field of view, and averaging across all fields for a given well. For a treatment compound, cell density is converted to a percentage relative to the plate-averaged cell density from DMSO treatment (i.e., 100% corresponds to the average cell density for DMSO treatment). Logistic regression curve fits were done using TIBCO Spotfire (Version 2.1; TIBCO Software, Inc.), and the concentration at which the curve crosses 50% is reported as the EC50 of the compound.

Quantifying cell phenotypes induced by compound treatments

A set of 8 reference compounds of reported mechanism were selected for the purpose of classifying cells by phenotype (designated in bold type in Table 1). These were the CDK inhibitors AG-024322 and R-547, the aurora kinase inhibitors AZD-1152 and tozasertib, the PLK1 inhibitor BI-2536, and the microtubule modulators ON-01910, nocodazole and paclitaxel. ON-01910 was originally reported as a non-ATP competitive PLK1 inhibitor; in subsequent reports it has been found to not inhibit PLK1 biochemically and generates a cell phenotype consistent with microtubule modulation (22, 23). The compounds were selected to represent a diversity of phenotypes observed for G2-M modulators. Visual analysis of images reveals cell populations consistent with the mechanism of action of the compounds at all concentrations above the antiproliferation EC50, allowing those wells to be pooled for the purpose of training a classifier. Cells in these wells are described by the cytological features in Supplementary Table 1, and assigned the class label of the treatment compound. To our knowledge, there are no cell-cycle modulators that arrest cells in metaphase or anaphase. To train the classifier in the identification of these states, images for DMSO-treated cells in experiment 3 were reviewed and ~30 cells of each type identified and labeled accordingly. The total number of cells across pooled wells used for training the classifier was 63,575, 58,298, and 78,552 for experiments 1 to 3.

The reference compounds were used to develop a classifier of cells for each experiment (i.e., 3 experiments, 3 classifiers). The classifier was developed with the recursive partitioning algorithm in Jmp (Version 7.0.2, SAS Institute, Inc), using the compound class label as response variable and cytological features as factors. The
Table 1. Compounds characterized by high-content imaging in HCT-116

<table>
<thead>
<tr>
<th>Name*</th>
<th>EC_{50}, \mu mol/L</th>
<th>Dominant phenotype^c</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experiment^b 1</td>
<td>Experiment 2</td>
</tr>
<tr>
<td><strong>Aurora inhibitors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AZD1152</td>
<td>0.026</td>
<td>0.023</td>
</tr>
<tr>
<td>AZD1152 metabolite</td>
<td>0.034</td>
<td>0.025</td>
</tr>
<tr>
<td>CYC116</td>
<td>0.649</td>
<td>0.519</td>
</tr>
<tr>
<td>ENMD-2076</td>
<td>0.247</td>
<td>&gt;2</td>
</tr>
<tr>
<td>MLN-8054</td>
<td>0.036</td>
<td>0.075</td>
</tr>
<tr>
<td>PHA-739358</td>
<td>0.045</td>
<td>0.050</td>
</tr>
<tr>
<td>Tazosertib (VX-680; MK0457)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CDK inhibitors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AG-024322</td>
<td>0.173</td>
<td>0.116</td>
</tr>
<tr>
<td>AG-12286 (WO 9921845 A2 223784-75-6)</td>
<td>0.249</td>
<td>G2 + proM</td>
</tr>
<tr>
<td>Alvocidib (flavopiridol)</td>
<td>0.153</td>
<td>0.256</td>
</tr>
<tr>
<td>BMI-1026 (39)</td>
<td>0.030</td>
<td>G2</td>
</tr>
<tr>
<td>BMS-265246 (40)</td>
<td>0.293</td>
<td>0.492</td>
</tr>
<tr>
<td>JNJ-7706621 (41, 42)</td>
<td>0.524</td>
<td>G2</td>
</tr>
<tr>
<td>PD-171851 (43)</td>
<td>0.570</td>
<td>G2 + proM</td>
</tr>
<tr>
<td>PD-0332991</td>
<td>&gt;2</td>
<td>G1-S</td>
</tr>
<tr>
<td>Aminopurvalanol</td>
<td>2.500</td>
<td>G2</td>
</tr>
<tr>
<td>R-547</td>
<td>0.077</td>
<td>0.144</td>
</tr>
<tr>
<td>SCH-727965</td>
<td>0.018</td>
<td>0.008</td>
</tr>
<tr>
<td>SNS-032</td>
<td>0.074</td>
<td>0.119</td>
</tr>
<tr>
<td>WO 2001064655 A1 358788-29-1</td>
<td>0.292</td>
<td>G2 + proM</td>
</tr>
<tr>
<td>WO 2001064656 A1 358789-50-1</td>
<td>0.250</td>
<td>G2 + endo</td>
</tr>
<tr>
<td><strong>PLK1 inhibitors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BI-2536</td>
<td>0.009</td>
<td>0.008</td>
</tr>
<tr>
<td>GSK-461364</td>
<td>0.012</td>
<td>0.014</td>
</tr>
<tr>
<td>HMN-176 (28)</td>
<td>0.164</td>
<td>0.260</td>
</tr>
<tr>
<td>WO 2006049339 A1 886856-66-2</td>
<td>0.101</td>
<td>0.072</td>
</tr>
<tr>
<td>WO 2006066172 A1 893440-87-4</td>
<td>1.861</td>
<td>proM + M-apopt</td>
</tr>
<tr>
<td><strong>Wee1 inhibitor</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>US 2007254892 A1 955365-24-9</td>
<td>2.340</td>
<td>proM + M-apopt</td>
</tr>
<tr>
<td><strong>DNA intercalators; Topoisomerase inhibitors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aclarubicin</td>
<td>0.238</td>
<td>0.255</td>
</tr>
<tr>
<td>Camptothecin</td>
<td>0.011</td>
<td>0.008</td>
</tr>
<tr>
<td>Doxorubicin</td>
<td>0.254</td>
<td>0.247</td>
</tr>
<tr>
<td>Topotecan</td>
<td>0.020</td>
<td>0.033</td>
</tr>
<tr>
<td><strong>DNA synthesis inhibitors</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-Fluoro-2’ deoxyuridine (5-fluoridine)</td>
<td>&gt;2</td>
<td>&gt;2</td>
</tr>
<tr>
<td>Herboxidene</td>
<td>0.010</td>
<td>0.011</td>
</tr>
<tr>
<td>Illudin S</td>
<td>0.015</td>
<td>0.019</td>
</tr>
<tr>
<td>Mitomycin</td>
<td>0.187</td>
<td>0.253</td>
</tr>
<tr>
<td><strong>Microtubule modulators</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ON-01910</td>
<td>0.100</td>
<td>0.052</td>
</tr>
<tr>
<td>Albendazole</td>
<td>0.448</td>
<td>0.305</td>
</tr>
<tr>
<td>Ciclobendazole</td>
<td>0.547</td>
<td>0.691</td>
</tr>
<tr>
<td>Mebendazole</td>
<td>0.318</td>
<td>0.483</td>
</tr>
<tr>
<td>Nocodazole</td>
<td>0.055</td>
<td>0.072</td>
</tr>
<tr>
<td>Paclitaxel</td>
<td>0.003</td>
<td>0.007</td>
</tr>
</tbody>
</table>

Abbreviations: G1-S, G1 or S-phase; proM, prometaphase; M-apopt, M-phase apoptotic; endo, endoreduplication.

*Structures are given in Supplementary scheme 1; compounds in bold are used as references in calibrating the cell classifier; kinase inhibitors with no reference given are described at http://www.clinicaltrials.gov; inhibitors described in patents, but not the journal literature, are identified by the CAS number retrieved in SciFinder (Version 2007.3, American Chemical Society).

bExperiments in which the compound was characterized, each of which was done by a different biologist on a different date.

cPhenotype from cell-classifier for dominant cell population(s), that is, polyploidy; and # indicate phenotypes at lower and higher concentrations, respectively.
maximum significance rule was used for selecting splits. Recursive partitioning is a greedy approach, selecting the best feature for each split in an incremental manner regardless of its appeal from a biological perspective. To emulate the manner in which biologists analyze cells, the first splits are obtained by visual binning of DNA intensity into 2N, 4N, 8N, and >8N categories. This is done manually via histogram plots in TIBCO Spotfire, allowing the boundaries between 2N, 4N, 8N, etc. to change from experiment to experiment (Supplementary Fig. 2). For experiment 3, the subsequent splits used features and rules selected by the recursive partitioning algorithm. The terminal nodes are assigned names consistent with the values of cytological features and review of images. For the other experiments, the features selected in experiment 3 were retained for each split, but the split value was allowed to change. The rules for classifying metaphase and anaphase cells are not updated in experiments 1 to 2 because of their rarity in the study of cell-cycle modulators, and the need to identify representative cells by review of images. The complete decision tree as depicted in Jmp is shown in Supplementary Figure 3.

Comparing treatment wells

For the purpose of contrasting the cell classifier to other approaches, it is necessary to quantitatively assess the similarity of phenotypes observed under different treatment conditions (i.e., comparing cell phenotypes between wells). The phenotype profile for a well is represented as a vector of length 9, indicating the proportion of cells belonging to each phenotype. For comparison, a vector of length 5 represents the well-averaged values for 5 DNA-related features (excluding ObjectVarIntenDNA due to its high correlation with ObjectAvgIntenDNA; Pearson r > 0.9 for the 3 experiments). We employed the Euclidian and cosine distance metrics for comparing pairs of vectors:

\[ \mathbf{p} = (p_1, p_2, \ldots, p_n); \quad \mathbf{q} = (q_1, q_2, \ldots, q_n) \]

\[ \text{d}_{\text{Euclidian}} = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \]

\[ \text{d}_{\text{cosine}} = 1 - \frac{\sum_{i=1}^{n} p_i q_i}{\sqrt{\sum_{i=1}^{n} p_i^2} \sqrt{\sum_{i=1}^{n} q_i^2}} \]

Results

Antiproliferation effects of cell-cycle inhibitors

A total of 41 cell-cycle modulators were characterized via high-content imaging of HCT-116 cells. Cells were imaged on the Arrayscan VTI platform, quantifying intensity, localization and/or morphology of Hoechst stain (DNA), terminal deoxynucleotidyl transferase dUTP nick end labeling (TUNEL; a marker for apoptosis), and antibodies for cyclin B1, phospho-histone H3 (pHH3), and α-tubulin (see Supplementary Methods). To assess the reproducibility of imaging experiments, 29 compounds were tested in 2 or 3 separate experiments, done by different biologists over a 6-month period. Most compounds inhibit cell proliferation at a concentration of 1 μmol/L or less (Table 1). Using the 58 pairs of defined EC50s for the same compound (i.e., EC50s not prefixed with the symbol ‘>’), the fold difference between experiments ranges from 1.0 to 3.9, with an average of 1.5. The assessment of antiproliferative properties of treatments, measured by quantifying changes in cell density (i.e., cell count per field of view), is highly reproducible across experiments.

Classification of cell phenotypes from immunofluorescence microscopy

A classifier of cell phenotypes was developed using 8 reference compounds, selected to represent a diversity of mechanisms among G2-M modulators (see the Methods section). A visual assessment of images from treatment concentrations above the antiproliferation EC50 reveals cell populations consistent with published reports on standard tubulin modulators (15) and RNAi approaches for kinase targets (17, 20, 23). Wells containing a reference compound at concentrations above the antiproliferation EC50 were pooled, and cells described using cytological features (e.g., nuclear area, cyclin B1 intensity, etc.) listed in Supplementary Table 1. The classifier begins with assessment of DNA content (2N, 4N, 8N, >8N) using boundaries defined by visual analysis of Hoechst DNA staining intensity (Supplementary Fig. 2), and creates a decision tree by applying recursive partitioning using cytological features as factors and the compound mechanism of action as response variable. The rules that constitute the decision tree are selected incrementally in a manner that distinguishes cells treated with compounds having different mechanisms. A cell’s phenotype is defined by the terminal node into which it falls (Fig. 1). As such, each cell is assigned 1 of 9 possible phenotypes, including the phenotype ‘other’ that corresponds mostly to cell debris. The model is recalibrated in each experiment by analysis of cells treated with reference compounds, without the need to examine and identify cells for training as required with other approaches.

For the CDK1 inhibitors AG-024322 and R-547, the dominant cell populations are G2-arrested cells having 4N DNA content, large round nuclei, and low DNA intensity (i.e., diffuse chromatin). By contrast, treatment with the PLK1 inhibitor BI-2536 results mostly in cells characteristic of prometaphase arrest with 4N DNA content, large round nuclei, and low DNA intensity (i.e., diffuse chromatin). The aurora kinase inhibitors AZD-1152 and tozasertib induce polyploidy (via endoreduplication), consistent with the dominance of the AURKB phenotype over AURKA. With a 48-hour incubation (allowing 2 cell number doublings), the dominant population should consist of cells with 8N DNA content; smaller populations with 4N and >8N DNA content arise from missegmentation of nuclei clusters and mostly represent...
artifacts of image analysis (some >8N cells are present). Representative images from each mechanistic class are shown in Figure 2 and Supplementary Figure 4.

The classifier makes significant use of features derived from DNA staining (Fig. 1). In characterizing G2-M modulators, other markers such as cyclin B1 and pHH3 are frequently examined in conjunction with DNA content. While the proportion of cells for a given phenotype varies widely across mechanistic classes, cyclin B1 and pHH3 staining intensities of cells for a given phenotype are similar across mechanistic classes (Supplementary Fig. 5), except for DMSO-treated cells having lower cyclin B1 intensity for all phenotypes, and higher pHH3 intensity for prometaphase and metaphase cells. TUNEL staining, a measure of apoptosis through DNA end-labeling following DNA fragmentation, shows greater differentiation across mechanistic classes for cells of a given phenotype: cells treated with CDK1 and PLK1 inhibitors have high induction of apoptosis compared to DMSO or aurora inhibitor treatments. Most aurora inhibitors induce cytokinesis defects, but cells continue cycling beyond 48 hours. The intensity of α-tubulin immunostaining from treatment with paclitaxel, a microtubule stabilizer, is increased compared to destabilizers such

Figure 1. Decision tree used for classifying cells according to cytological features (described in Supplementary Table 1). DNA content is defined by manual binning of ObjectTotalIntenDNA (left rectangles; see Supplementary Fig. 2). Other splits are obtained from recursive partitioning in Jmp, with those in gray defined by partitioning cells treated with 8 calibration compounds, using the compound mechanism of action as class label. The splits in black are defined by partitioning examples of cells in G1, metaphase, and anaphase; approximately 30 metaphase and 30 anaphase cells were identified by manual review of images from DMSO treatment in experiment 3. A cell is assigned the phenotype for the terminal node (rectangles) into which it falls. Cells in the nodes endo 4N, endo 8N, and endo >8N denote multinucleated cells; all are combined into 1 phenotype. Likewise, cells assigned to M-apoptotic and M-apoptotic brightest are combined, since they arise from staining variation with Hoechst 33258 in apoptotic cells (see text). With the exception of the nodes in black, numerical values used in the decision tree for each cytological feature are adjusted by recalibrating the model in each experiment; the notation used to identify nodes [e.g., S1: 1.2, 1.0, 1.1 (0.95)] indicates the split number (S1), the numerical values of a cytological feature used for separating cells in 3 experiments (1.2, 1.0, and 1.1 for experiments 1–3, respectively), and the range observed in 17 subsequent experiments done with compounds from lead optimization programs (0.95).
as nocodazole, allowing some degree of differentiation; all mechanistic classes result in higher tubulin intensity than DMSO treatment.

In addition to phenotypes commonly observed in non-treated cells (e.g., G2, prometaphase, etc.), the classifier identifies “apoptotic” phenotypes for G2 and M states (Fig. 1). These cells have apparent 8N or >8N DNA content, and occur with higher prevalence in CDK1 (G2-apoptotic cells) and PLK1 or microtubule modulators (M-phase apoptotic cells), suggesting the presence of polyploidy for these classes. However, this observation is inconsistent with G2-M arrest. Staining with TUNEL indicates a high induction of apoptosis, and suggests that Hoechst dye has higher affinity for unwinding chromatin in apoptotic cells. DNA staining with propidium iodide, a DNA intercalator that binds with stoichiometry of 1 dye per 4–5 base pairs, is not affected by DNA coiling and reveals a single 4N peak for treatments such as nocodazole, vs. the 2 peak population distribution for Hoechst (Supplementary Fig. 1). As such, apparent DNA intensity of 8N or >8N, in the absence of other cytological features such as a high DNA perimeter-to-area ratio (due to multilobed nuclei) or low pHH3 intensity (consistent with AURKB inhibition), cannot be used to infer polyploidy.

In spite of this, Hoechst is preferred due to signal quenching from other fluorescent channels that occurs with propidium iodide.

**Summarizing concentration-dependent phenotypes**

The classifier assigns a phenotype to every imaged cell. A population of cells within a well can be summarized as the percentage of cells exhibiting each of the 9 phenotypes reported by the classifier. Changes in populations along a concentration-response curve can be summarized via stacked bar graphs: for the aurora inhibitor PHA-739358, the classifier reveals a mixed aurora A/B cell population at the antiproliferation EC50, which becomes consistent with AURKB inhibition at intermediate concentrations, and exhibits a dominant AURKA profile at higher concentrations (Fig. 3). This representation is useful for elucidating the structure-phenotype relationship in lead optimization programs.

**Comparing phenotypes from the cell classifier to averaged cytological features**

A simple approach for quantifying treatment-induced phenotypes consists of averaging cytological features within a well (e.g., the average DNA content of cells).
To understand how this approach differs from our phenotype classifier, we describe every treatment well using 3 approaches: (a) the average value for each of 12 cytological features used alone, (b) a vector of length 5 consisting of the average values of DNA-related cytological features (DNA well-average features), and (c) a vector of length 9 indicating the percentage of cells for each phenotype reported by the classifier (cell phenotype profiles).

Wells for treatment concentrations below the antiproliferation EC50 are discarded to focus on those wells in which cells are responding to treatment. For approach 1, 2 wells are compared by taking the absolute value of the difference between wells. For approaches 2 and 3 that describe each well as a vector, the difference between wells is calculated using the Euclidean and cosine distances (see Materials and Methods). Small differences or distances indicate consistent assessment of phenotype.

For the purpose of assessing reproducibility between experiments, we identified 16 compounds tested in all 3 experiments, and set aside the 6 compounds used for model calibration. This yielded 159 pairs of wells from different experiments that contain the same compound at the same (or similar) concentration. Because the measures above have different natural scales, the 3 approaches for comparing wells are calculated using the Euclidian and cosine distances (see Materials and Methods). Small differences or distances indicate consistent assessment of phenotype.

For the purpose of assessing reproducibility between experiments, we identified 16 compounds tested in all 3 experiments, and set aside the 6 compounds used for model calibration. This yielded 159 pairs of wells from different experiments that contain the same compound at the same (or similar) concentration. Because the measures above have different natural scales, the 3 approaches for comparing wells are calculated using the Euclidian and cosine distances (see Materials and Methods). Small differences or distances indicate consistent assessment of phenotype.

For the purpose of assessing reproducibility between experiments, we identified 16 compounds tested in all 3 experiments, and set aside the 6 compounds used for model calibration. This yielded 159 pairs of wells from different experiments that contain the same compound at the same (or similar) concentration. Because the measures above have different natural scales, the 3 approaches for comparing wells are calculated using the Euclidian and cosine distances (see Materials and Methods). Small differences or distances indicate consistent assessment of phenotype.

For the purpose of assessing reproducibility between experiments, we identified 16 compounds tested in all 3 experiments, and set aside the 6 compounds used for model calibration. This yielded 159 pairs of wells from different experiments that contain the same compound at the same (or similar) concentration. Because the measures above have different natural scales, the 3 approaches for comparing wells are calculated using the Euclidian and cosine distances (see Materials and Methods). Small differences or distances indicate consistent assessment of phenotype.

For the purpose of assessing reproducibility between experiments, we identified 16 compounds tested in all 3 experiments, and set aside the 6 compounds used for model calibration. This yielded 159 pairs of wells from different experiments that contain the same compound at the same (or similar) concentration. Because the measures above have different natural scales, the 3 approaches for comparing wells are calculated using the Euclidian and cosine distances (see Materials and Methods). Small differences or distances indicate consistent assessment of phenotype.

For the purpose of assessing reproducibility between experiments, we identified 16 compounds tested in all 3 experiments, and set aside the 6 compounds used for model calibration. This yielded 159 pairs of wells from different experiments that contain the same compound at the same (or similar) concentration. Because the measures above have different natural scales, the 3 approaches for comparing wells are calculated using the Euclidian and cosine distances (see Materials and Methods). Small differences or distances indicate consistent assessment of phenotype.

For the purpose of assessing reproducibility between experiments, we identified 16 compounds tested in all 3 experiments, and set aside the 6 compounds used for model calibration. This yielded 159 pairs of wells from different experiments that contain the same compound at the same (or similar) concentration. Because the measures above have different natural scales, the 3 approaches for comparing wells are calculated using the Euclidian and cosine distances (see Materials and Methods). Small differences or distances indicate consistent assessment of phenotype.

For the purpose of assessing reproducibility between experiments, we identified 16 compounds tested in all 3 experiments, and set aside the 6 compounds used for model calibration. This yielded 159 pairs of wells from different experiments that contain the same compound at the same (or similar) concentration. Because the measures above have different natural scales, the 3 approaches for comparing wells are calculated using the Euclidian and cosine distances (see Materials and Methods). Small differences or distances indicate consistent assessment of phenotype.

For the purpose of assessing reproducibility between experiments, we identified 16 compounds tested in all 3 experiments, and set aside the 6 compounds used for model calibration. This yielded 159 pairs of wells from different experiments that contain the same compound at the same (or similar) concentration. Because the measures above have different natural scales, the 3 approaches for comparing wells are calculated using the Euclidian and cosine distances (see Materials and Methods). Small differences or distances indicate consistent assessment of phenotype.

For the purpose of assessing reproducibility between experiments, we identified 16 compounds tested in all 3 experiments, and set aside the 6 compounds used for model calibration. This yielded 159 pairs of wells from different experiments that contain the same compound at the same (or similar) concentration. Because the measures above have different natural scales, the 3 approaches for comparing wells are calculated using the Euclidian and cosine distances (see Materials and Methods). Small differences or distances indicate consistent assessment of phenotype.

For the purpose of assessing reproducibility between experiments, we identified 16 compounds tested in all 3 experiments, and set aside the 6 compounds used for model calibration. This yielded 159 pairs of wells from different experiments that contain the same compound at the same (or similar) concentration. Because the measures above have different natural scales, the 3 approaches for comparing wells are calculated using the Euclidian and cosine distances (see Materials and Methods). Small differences or distances indicate consistent assessment of phenotype.

For the purpose of assessing reproducibility between experiments, we identified 16 compounds tested in all 3 experiments, and set aside the 6 compounds used for model calibration. This yielded 159 pairs of wells from different experiments that contain the same compound at the same (or similar) concentration. Because the measures above have different natural scales, the 3 approaches for comparing wells are calculated using the Euclidian and cosine distances (see Materials and Methods). Small differences or distances indicate consistent assessment of phenotype.

For the purpose of assessing reproducibility between experiments, we identified 16 compounds tested in all 3 experiments, and set aside the 6 compounds used for model calibration. This yielded 159 pairs of wells from different experiments that contain the same compound at the same (or similar) concentration. Because the measures above have different natural scales, the 3 approaches for comparing wells are calculated using the Euclidian and cosine distances (see Materials and Methods). Small differences or distances indicate consistent assessment of phenotype.

For the purpose of assessing reproducibility between experiments, we identified 16 compounds tested in all 3 experiments, and set aside the 6 compounds used for model calibration. This yielded 159 pairs of wells from different experiments that contain the same compound at the same (or similar) concentration. Because the measures above have different natural scales, the 3 approaches for comparing wells are calculated using the Euclidian and cosine distances (see Materials and Methods). Small differences or distances indicate consistent assessment of phenotype.

For the purpose of assessing reproducibility between experiments, we identified 16 compounds tested in all 3 experiments, and set aside the 6 compounds used for model calibration. This yielded 159 pairs of wells from different experiments that contain the same compound at the same (or similar) concentration. Because the measures above have different natural scales, the 3 approaches for comparing wells are calculated using the Euclidian and cosine distances (see Materials and Methods). Small differences or distances indicate consistent assessment of phenotype.

For the purpose of assessing reproducibility between experiments, we identified 16 compounds tested in all 3 experiments, and set aside the 6 compounds used for model calibration. This yielded 159 pairs of wells from different experiments that contain the same compound at the same (or similar) concentration. Because the measures above have different natural scales, the 3 approaches for comparing wells are calculated using the Euclidian and cosine distances (see Materials and Methods). Small differences or distances indicate consistent assessment of phenotype.
average vs. 84th percentile for cell phenotype profiles). This recalls the aphorism "the average cell does not exist" (I), where the tozasertib average arises from a bimodal distribution of cells in prometaphase (small bright nuclei) and polyploidy (very large diffuse nuclei), and appears similar to the CDK1 inhibitor with predominantly G2-arrested cells having nuclei of intermediate size and intensity (Supplementary Fig. 6). Other examples include the CDK1 inhibitors R-547 vs. BMS-265246, where the former has a higher proportion of G2-apoptotic cells, and...
the PLK1 inhibitors HMN-176 vs. BI-2536, where the former has a higher fraction of M-phase apoptotic cells.

**Cell population responses to chemotherapeutic agents**

A large number of chemotherapeutic agents are known to affect the G2-M transition of the mitotic cell cycle. We interrogated the connection between the mechanism of action of these modulators and the effect on HCT-116 cell populations. The phenotypic profiles were assessed in 2 experiments, and yielded highly consistent results (Supplementary Fig. 7).

The quinoline alkaloids camptothecin and topotecan stabilize the topoisomerase I-DNA complex, resulting in single-strand DNA breaks. Both agents produce cell populations arrested in G2 at concentrations near the antiproliferation EC50, consistent with their reported cell-cycle effects (24, 25). However, an increasing population of cells in prometaphase is apparent at higher concentrations, suggesting cellular effects unrelated to topoisomerase I inhibition. The alkylating agent mitomycin also results in a large enrichment of cells in G2. While we observed consistent phenotypes for structurally related camptothecin and topotecan, the anthracycline antibiotics doxorubicin and aclarubicin result in distinct phenotypes, even though both stabilize the topoisomerase II-DNA complex. Doxorubicin, at concentrations higher than what is necessary for topoisomerase I inhibition, is capable of inhibiting topoisomerase I and will also compete for binding sites for the various DNA stains and thus, is capable of inhibiting topoisomerase I-DNA complex. Doxorubicin, at concentrations lower than that which is necessary for topoisomerase II inhibition, is capable of inhibiting topoisomerase I and will also compete for binding sites for the various DNA stains and will therefore elicit multiple concentration-dependent phenotypic outcomes (unpublished results). Most cells from aclarubicin treatment have G1-S properties, possibly through more potent inhibition of topoisomerase I. As observed for aclarubicin, the antitubulin 5-fluoro-2'-deoxyuridine (5-FUDR or flouxuridine) significantly inhibits proliferation without resulting in a large enrichment of cells in G2-M. The classifier is not optimized for characterizing treatments that modulate the G1/S mitotic cell cycle; cells in G2 are difficult to distinguish from those in S phase using Hoechst staining and light microscopy, necessitating additional markers such as 5-ethyl-2'-deoxyuridine (EDU) or bromodeoxyuridine (BrdU) or G1-specific markers such as phosphor retinoblastoma protein 1 (pRBl). The tubulin depolymerizers mebendazole, albendazole, ciclohexadazole, and nodocazole, and the tubulin stabilizer paclitaxel all produce populations enriched in prometaphase and apoptotic cells (15). The proportion of prometaphase cells increases with concentration, presumably due to a diminishing proportion of nonviable cells in the population.

**Cell population responses to inhibitors of cell-cycle kinases**

Inhibitors of kinases tend to exhibit varying degrees of off-target activity (26), making it difficult to anticipate the level of selectivity from cell-free assays expected to yield a phenotype consistent with modulation of the intended target. The cell classifier was used to characterize changes in cell populations induced by treatment with several G2-M kinase inhibitors (Fig. 5 for selected inhibitors; Supplementary Fig. 8). Most CDK1 and pan-CDK inhibitors cause enrichment of cells in G2, consistent with the role of CDK1 in cell-cycle regulation. Many of these molecules inhibit interphase CDKS (CDK2/4/5/6). In particular, the inhibitors AG-024322, JNJ-7706621, and aminopurvalanol have potencies vs. interphase CDKS that are similar or greater than that for CDK1, yet show a phenotype consistent with CDK1 inhibition. The residual G1-S cells may be nonresponders or G1-S-arrested cells arising from inhibition of interphase CDKS. For example, the exclusively selective CDK4 inhibitor PD-0332991 inhibits proliferation, and induces mostly G1-S cells according to the classifier (Supplementary Fig. 8). Additional markers such as pRBl staining and/or EDU labeling are required for effective characterization of compounds having dominant G1/S mechanisms.

In contrast to the dominance of the CDK1 phenotype noted above, AG-12286, alvocidib, PD-171851, SCH-727965, and SNS-032 potently inhibit CDK9 in addition to other CDKs, and produce mixed populations of cells in G2, prometaphase, and advanced stages of apoptosis. The role of CDK9 in transcriptional regulation (27), coupled with this observation, suggests that manifestation of the CDK1 phenotype is distorted by CDK9 activity. However, other inhibitors (e.g., AG-024322, BMI-1026, BMS-265246, and R-547) significantly inhibit CDK9 in vitro yet induce a CDK1 phenotype. The inhibitors JNJ-7706621, 358788-29-1, and 358789-50-1 (reported by Astrazeneca) exhibit AURKB activity in enzyme assays, which is evident in the cell population profile for the latter despite the role of that kinase beyond the G2-M checkpoint. The relationship between enzyme inhibition in cells and phenotype is complex and not fully understood from cell-free assays.

We investigated the relationship between aurora kinase activity and cell phenotype using a number of inhibitors at our disposal. The prodrug AZD-1152 and its metabolite differ by a phosphate group used to improve solubility of the prodrug; both are selective AURKB inhibitors and induce polyploidy in treated cells, consistent with their enzyme activity. As they are the only selective AURKB inhibitors we have characterized, it is interesting to note that both retain larger populations of G1-S cells than the other aurora kinase inhibitors. As cells responding to AURKB inhibition continue cycling beyond 48 hours, the absence of other arrest mechanisms may explain this observation. The inhibitors tozasertib and CYC116 induce polyploidy as expected for dual A/B inhibitors, unlike PHA-739358 that exhibits a concentration-dependent change from dominant polyploidy (consistent with AURKB) to dominant prometaphase arrest (consistent with AURKA). In our hands, MLN-8054 and ENMD-2076 both inhibit AURKB more potently in biochemical assays, yet induce cell populations consistent with AURKA at lower concentrations and AURKB at higher concentrations. For
Figure 5. Kinase profiling data from biochemical enzyme assays and cell phenotypes in HCT-116 cells. Right, proportion of cells classified into each of 9 phenotypes versus treatment concentration; the curve indicates inhibition of proliferation measured by counting cells per field of view. Left and middle, results from kinase enzyme assay profiling, with insets showing detail for CDK, aurora, and PLK kinases. Changes in markers from green to red (and small to large) indicate increasing binding affinity, on a log10 scale. Labeled markers indicate the IC50 in μmol/L (no units) or % inhibition at 20 μmol/L (values followed by %); the absence of labels denote inactive results (i.e., IC50 > 10 μmol/L or % inhibition < 80 for single point results). See Supplementary Methods for details on enzyme assays. Additional inhibitors are shown in Supplementary Fig. 8. Human kinome provided courtesy of Cell Signaling Technology, Inc. www.cellsignal.com.
tozasertib, its significant binding affinity for CDK1 is not apparent in the induced cell population.

In contrast to the variation in phenotypes observed for CDK and aurora kinase inhibitors, small molecules targeting PLK1 generally exhibit similar phenotypes consisting of mixed populations of cells in prometaphase and advanced apoptosis. A notable exception to the class is the Banyu inhibitor 886856-66-2 that exhibits a small but increasing population of G2 cells with increasing concentration. HMN-176 is not an ATP-competitive inhibitor, but interferes with the cellular localization of PLK1 (28). ON-01910 was originally described as a PLK1 inhibitor, but is now thought to be a tubulin modulator (22, 23; Supplementary Fig. 7).

Relationship between selectivity in cell-free assays and phenotype in HCT-116 cells

The variation in phenotypic response among the kinase inhibitors in Table 1 prompted us to examine the concordance between enzyme inhibition profiles and phenotypic response for a larger number of internal compounds exhibiting activity against G2-M kinases. For this purpose, we selected for analysis compounds having antiproliferation EC₅₀ < 1 μmol/L in HCT-116 cells and with fully determined G2-M kinase enzyme profiles (IC₅₀ vs. CDK1, AURKA, AURKB and PLK1, or ≤80 percent inhibition in single concentration testing at 20 μmol/L). For wells having concentrations above the antiproliferation EC₅₀, we summarized the proportion of cells belonging to non-G₁-S phenotypes from the cell classifier (Supplementary Fig. 9). While selective AURKA inhibitors induce populations dominated by prometaphase arrest and advanced apoptosis, the dual A/B inhibitors induce small populations of multinucleated cells, but a larger proportion of multinucleated cells vs. PLK1-selective inhibitors. For CDK inhibitors, a decreasing proportion of cells in G2 is apparent as selectivity vs. the aurora kinases and PLK1 increases. However, the corresponding increase in prometaphase and M-phase apoptotic cells may be attributed to CDK9 inhibition as noted above, and most compounds were not tested vs. other CDKs. For PLK1 inhibitors, there is no discernable change in the dominant prometaphase and M-phase apoptotic phenotype with increasing selectivity vs. the CDKs and aurora kinases. This suggests a dominant role for PLK1 enzyme inhibition over other cell-cycle targets, despite a role for PLK1 in mitosis. A conventional view holds that simultaneous inhibition of targets involved earlier in the cell cycle would manifest over M-phase arrest. The lack of clear relationships between enzyme inhibition profiles and cell phenotype supports the importance of phenotype determination via high-content imaging to verify that cell death arises from modulation of the targeted kinase (or kinases).

Discussion

The cell classifier described in this work enables the investigation of mechanism of action for cell-cycle modulators by summarizing cell-level results from high-content imaging experiments. While an increasing body of literature describes approaches that distill millions of cellular measurements for interpretation of phenotype, only a few approaches have explored the reproducibility of phenotype across experiments (7, 10). The approach described in this work has been applied to experiments done over 6 months by 3 different biologists, and significantly reduces variability in results that often afflict industrial application of immunofluorescence-based microscopy. The methodology is applied within lead optimization programs at Lilly to understand changes in phenotype that arise from structural modification of lead series, and the extent to which activity at other kinases translates to deviation from the desired effect (i.e., obtaining a phenotype consistent with inhibition of a given kinase target).

The application of our classifier highlights the prevalence of concentration-dependent phenotypes among inhibitors purported to inhibit the same kinase. While this work does not evaluate the effects of RNAi treatments, the cell morphologies that we identify as consistent with the intended target are informed from published reports using RNAi (17, 20, 22, 23, 29, 30). We postulate that departure from the expected phenotype arises from off-target effects, rather than variable inhibition of the intended target. Although the effect can sometimes be rationalized from enzyme activity profiles (e.g., PHA-739358), there is generally no discernable quantitative relationship between enzyme selectivity and phenotype. It is noteworthy that exquisitely selective kinase inhibitors studied in this work (PD-0332991 and the AZD-1152 metabolite) induce the expected phenotypes.

Conceptually, our approach is similar to methods that extract classification rules from cells representing distinct phenotypes identified by review of images (2–4). Such methods can overcome variation in staining intensity, etc. by repeating the identification of representative cells and rule training in every experiment. By recalibrating the cell classifier from reference molecules selected to represent the relevant cell phenotypes, the need for manual review of images is substantially reduced. On the other hand, the use of reference inhibitors renders the approach less effective in the identification of novel or unexpected cell morphologies (6). The simplicity of a decision tree for cell classification is appealing, but uses artificial rectangular boundaries in cytological feature space. It can be difficult to ascertain whether rarer cell populations are artifacts of classification; most PLK1 inhibitors appear to induce small populations of multinucleated cells, but these are M-phase apoptotic cells with irregular shapes and lower DNA staining intensity. In our hands, the use of mixture models (9) does not resolve the phenotype of cells falling between major phenotypic classes, as expected given the probabilistic assignment of cells to phenotype classes (unpublished results). Higher resolution imaging, perhaps with additional markers, may be
required to fully resolve distinct cell populations (10). As cellular imaging technology continues to evolve, the ability to monitor additional fluorescence channels will allow for more detailed analyses of complex cellular processes.

The imaging assay and phenotype classifier presented in this work are configured for medium-throughput characterization of G2-M modulators, and can be used to identify the probable mechanism of action in concert with biochemical profiling results. The classifier does not differentiate G1 cells from those in early S phase. For example, the CDK inhibitor R-547 was shown to inhibit Rb1 phosphorylation in phase I studies and induces populations of G1- and G2-arrested HCT-116 cells when studied via flow cytometry (31). The cell classifier indicates a dominant population of G2 cells, with some residual G1-S cells that could in fact be G1/S-arrested (i.e., blocked cell cycle). Other markers, such as antibodies for pRb1 and/or EDU labeling are probably necessary to allow full characterization of cells in G1-S. Likewise, cells responding to CDK1 and topoisomerase inhibitors (e.g., camptothecin and doxorubicin) are not differentiated with the markers employed in this study, but are readily distinguished via immunostaining for γH2AX, due to chain breaks induced by topoisomerase inhibitors (32).

In a similar vein, the classifier does not differentiate cells responding to AURKA inhibitors, PLK1 inhibitors, tubulin stabilizers (i.e., paclitaxel) or tubulin destabilizers, all of which directly modulate cytoskeleton function resulting in prometaphase arrest. High-resolution microscopy of paclitaxel reveals a metaphase-like arrangement of most chromosomes, with some lagging at the spindle pole (33). Such features are not discernable at the resolution used in this assay. At best, simple intensity-derived cytological features from α-tubulin staining provide slight differentiation of mechanistic classes, with paclitaxel and PLK1 inhibitors having higher staining intensity compared with tubulin destabilizers (whether considering all cells or only those in prometaphase/M-apoptotic states; Supplementary Fig. 5). This observation is consistent with the absence of changes in microtubule mass at lower compound concentration, despite their impact on microtubule dynamics as determined via high-resolution microscopy (34, 35). In cases where cross-reactivity is suspected (e.g., potent PLK and AURKA enzyme activity), high-resolution microscopy of centrosomes and microtubule topology to detect monopole spindle assemblies having reduced amounts of centrosomal γ-tubulin (PLK1; refs. 22, 23, 36), circular monopole spindles (AURKA; refs 20, 29), or the presence of multipolar spindles (tubulin destabilizers; refs 15, 33, 37) may be required to further clarify a compound’s mechanism of action. Higher-resolution microscopy can provide greater cellular detail for understanding mechanism, but reduces throughput due to longer scan times while also posing challenges with image storage. High-resolution microscopy is also not without its ambiguities: unlike the monopolar spindles induced by treatment with PLK1 RNAi or BI-2536, the inhibitor GSK-461364 potently inhibits the PLK1 enzyme in biochemical and substrate phosphorylation assays, yet induces multipolar spindles characteristic of tubulin modulators in H460 human lung cancer cells (38).

Although the cell classifier system described here focuses on characterization of cell-cycle modulators, the broad concept of summarizing cell-level results can be applied across high-content screenings and phenotypic drug discovery. The automated classification of phenotypic screening and structure-activity relationship data represents a solution to what is now the principal difficulty in taking raw data from a high content assay and using it to make a statistically robust and reproducible decision of phenotypic outcome.

Disclosure of Potential Conflicts of Interest

All the authors have Eli Lilly shares received via 401(k) and bonus plans.

Acknowledgments

The authors thank the Phenotypic Drug Discovery Working Group and PD² Management for guidance on screen implementation, follow-up, and general scientific discussion; Mark Marshall and Jake Starling for helpful discussions and project guidance.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received August 2, 2010; revised November 12, 2010; accepted November 24, 2010; published OnlineFirst January 7, 2011.

References


Molecular Cancer Therapeutics

A Robust High-Content Imaging Approach for Probing the Mechanism of Action and Phenotypic Outcomes of Cell-Cycle Modulators

Jeffrey J. Sutherland, Jonathan Low, Wayne Blosser, et al.


**Updated version**
Access the most recent version of this article at:
doi:10.1158/1535-7163.MCT-10-0720

**Supplementary Material**
Access the most recent supplemental material at:
http://mct.aacrjournals.org/content/suppl/2011/01/07/1535-7163.MCT-10-0720.DC1

**Cited articles**
This article cites 43 articles, 20 of which you can access for free at:
http://mct.aacrjournals.org/content/10/2/242.full#ref-list-1

**E-mail alerts**
Sign up to receive free email-alerts related to this article or journal.

**Reprints and Subscriptions**
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

**Permissions**
To request permission to re-use all or part of this article, use this link
http://mct.aacrjournals.org/content/10/2/242.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.