

An Automated High-Content Screening Image Analysis Pipeline for the Identification of Selective Autophagic Inducers in Human Cancer Cell Lines

JANOS KRISTON-VIZI,^{1,*} CHING AENG LIM,² PETER CONDRON,²
KELVIN CHUA,² MARTIN WASSER,¹ and HORST FLOTOW²

Automated image processing is a critical and often rate-limiting step in high-content screening (HCS) workflows. The authors describe an open-source imaging-statistical framework with emphasis on segmentation to identify novel selective pharmacological inducers of autophagy. They screened a human alveolar cancer cell line and evaluated images by both local adaptive and global segmentation. At an individual cell level, region-growing segmentation was compared with histogram-derived segmentation. The histogram approach allowed segmentation of a sporadic-pattern foreground and hence the attainment of pixel-level precision. Single-cell phenotypic features were measured and reduced after assessing assay quality control. Hit compounds selected by machine learning corresponded well to the subjective threshold-based hits determined by expert analysis. Histogram-derived segmentation displayed robustness against image noise, a factor adversely affecting region growing segmentation. (*Journal of Biomolecular Screening* XXXX:xx-xx)

Key words: autophagy, image processing, cellular high-content screening, phenotypic assay, MDC, LC3, HCS

INTRODUCTION

IMAGES GENERATED DURING HIGH-CONTENT SCREENING (HCS) CAMPAIGNS contain vast amounts of data. The automated extraction of useful, decision-supporting data from these images is often a significant bottleneck in the HCS pipeline. The translation of phenotypic images into quantitative information on, for example, how compounds affect subcellular structures can benefit from automated imaging and statistical methods and allow an unbiased interpretation of screening results.

The ultimate objective of any HCS screen is the selection of compounds that may be effective against diseases such as cancer, Alzheimer's disease, and Huntington's disease.^{1,2} These diseases are linked to autophagy, a highly regulated, homeostatic, intracellular catabolic mechanism by which eukaryotes degrade superfluous or faulty organelles and long-lived

proteins.³ Autophagy activation can lead to cell death through programmed self-digestion, and thus many anticancer agents may have autophagy-inducing capability. In addition, cell growth is negatively regulated by autophagy, which may slow down tumor growth.⁴ Being an evolutionarily highly conserved process, autophagy-specific genes have been characterized in distant species ranging from yeast to human.⁵ Macroautophagy (henceforth referred to as autophagy) is one of the 3 primary autophagy forms besides microautophagy and chaperone-mediated autophagy. The process of autophagy may be upregulated during both extracellular stress conditions such as starvation, infection, hypoxia, heat, or drug treatment, as well as intracellular stress conditions, including the accumulation of protein aggregate, misfolded proteins, or defective organelles.⁶ Autophagy involves a series of steps, including the formation and expansion (vesicle elongation) of an isolation membrane (phagophore), which then fuses to form autophagosomes, also known as autophagic vacuoles (AVs), which in turn fuse with lysosomes to form an autolysosome where the cytoplasmic material is sequestered and then degraded.⁷

HCS for autophagy activators is still limited, despite an emerging research interest. A recent high-content cellular image analysis study identified 8 autophagic cell death inducer compounds,⁸ and a cell-based functional screening revealed 3 genes inducing high levels of autophagosome formation when overexpressed.⁹ Upon induction of autophagy, both studies exploited the localization of microtubule-associated protein light chain 3 (LC3) fused with green fluorescent protein (GFP)

¹Bioinformatics Institute (A*STAR), Singapore.

²Experimental Therapeutic Centre (A*STAR), Singapore.

Present address: *Translational Research Resource Centre, MRC Laboratory for Molecular Cell Biology and Cell Biology Unit, University College London, London, UK.

Received Jan 11, 2010, and in revised form Apr 9, 2010. Accepted for publication Apr 12, 2010.

Supplementary material for this article is available on the *Journal of Biomolecular Screening* Web site at <http://jbx.sagepub.com/supplemental>.

Journal of Biomolecular Screening XX(X); XXXX

DOI: 10.1177/1087057110373393

to the autophagosomal membrane.¹⁰ Increased levels of autophagy are phenotypically indicated by the increased number, size, and/or fluorescent intensity of autophagosomes that can also aggregate around the nuclear membrane. Cellular systems expressing LC3-GFP are considered as being more specific to indicate autophagy, although emission is not steady through the process because late-stage AVs emit a weaker fluorescent signal than early stage AVs. There are several alternative methods for measuring autophagosomal induction, with no clear consensus as to which is the most appropriate to use. The fluorescent dye monodansylcadaverine (MDC) accumulates in autophagic and other intracellular compartments,¹¹⁻¹³ making it a convenient indicator of autophagy activation. We have therefore combined a simple and cost-effective MDC-based primary assay with an orthogonal secondary assay using an immunofluorescent LC3B antibody marker into a high-throughput phenotypic screening strategy to identify activators of autophagy.

Using HCS also allows us to distinguish true autophagic inducer compounds from those that increase the level of autophagic degradation indirectly by cellular toxicity.⁸

Automated fluorescent confocal microscopy coupled with increased computational power has dramatically enhanced the rate of image acquisition and analysis of derived multispectral data. A typical workflow starts with image preprocessing to improve the quality of raw images, followed by image segmentation where fluorescent regions of interest are separated from the background. Segmentation is a critical component of any image-processing workflow as errors at this phase profoundly influence hit selection.¹⁴ A decline in segmentation performance will result in errors in derived numerical metrics. In contrast to the blob-like nucleus, a region-growing segmentation strategy is not optimal for the (usually) punctate AV structures and may not be able to distinguish between AV and background noise.¹⁵ Autophagosomes do not appear as continuous but as sparse regions in an image. It is coupled with spectral crosstalk (bleed-through), which leads to multiple classes in the same image such as the nuclear, the AV bleed-through, and background pixels. A segmenting algorithm can address those challenges by clustering separate pixels into multiple classes.

As the capability to accurately segment sparse objects and correctly tackle bleed-through is not present in the commercial image analysis software available to us, we developed a custom macro in ImageJ,¹⁶ an open-source image analysis application. We combined this image analysis approach with a well-established statistical and analytical tool, R¹⁷ (<http://r-project.org>). R is an extensively used open-source platform for exploratory data analysis, descriptive statistics and assay quality control,¹⁸⁻²⁰ assay performance determination, and significance tests. Interactive visualization of the data was performed using the commercial data visualization application Spotfire (Somerville, MA). A statistical method-based feature selection was used to eliminate irrelevant or unhelpful numerical

descriptors, followed by machine learning-based hit selection. Hit compounds selected from the primary screen were then confirmed in secondary screens. The initial quantitative and automated processing of the images and derived data contributes to an unbiased selection of hits from such phenotypic assays, in contrast to prevailing subjective methods based mostly on expert curation and analysis.^{21,22}

This article outlines a simple image analysis–statistical framework that can be applied readily in a high-content cellular screen of cancer cell lines to identify novel selective pharmaceutical inducers of autophagy in compound libraries.

MATERIALS AND METHODS

Compounds for autophagy phenotypic screen

We used the LOPAC¹²⁸⁰ compound library (Sigma-Aldrich, St. Louis, MO) as our test compounds, trifluoperazine dihydrochloride (Sigma-Aldrich) as our positive control, and DMSO (Sigma-Aldrich) as our negative control.

Cell culture and cell plating

Human alveolar carcinoma A549 (lung epithelial cancer) cells were seeded into 96-well transparent, flat-bottom plates (Greiner, Monroe, NC) at a density of 5000 cells per well in Dulbecco's modified Eagle's medium (DMEM) containing 4% fetal bovine serum (HyClone, Logan, UT). The cells were then incubated with library compounds at 5 concentrations (10, 2, 0.4, 0.08, and 0.016 μM) for 48 h.

Primary assay using MDC staining

The live cells were stained with MDC as follows: 100 μL staining medium containing 100 μM MDC was added to all wells and incubated for 40 min at 37°C. Excess MDC was then washed off with phosphate-buffered saline (PBS) on the Bio-Tek EL-405, followed by the addition of 100 μL of warm DMEM. The plate was then read immediately on the ImageXpress Ultra. Cell nuclei were then stained with 3 μM Hoechst 33342 (Invitrogen, Carlsbad, CA).

Secondary assay using LC3B staining

A549 cells were fixed in 4% paraformaldehyde for 10 min at room temperature and then permeabilized with 0.1% Triton X-100 for a further 10 min. The fixed cells were then blocked with 3% bovine serum albumin (BSA) for 30 min. To stain for LC3B, we incubated the cells with anti-LC3B rabbit polyclonal antibody at 1:2000 dilution (#NB600; Novus, Littleton, CO) and subsequently labeled it with 1:2000 dilution of anti-rabbit AlexaFluor546 secondary antibody (Invitrogen) for visualization. Cell nuclei were then stained with 3 μM Hoechst 33342 (Invitrogen).

Fluorescent image acquisition

Images of the microwells (4 fields per well) were acquired by an ImageXpress Ultra (Molecular Devices, Sunnyvale, CA) system using 20 \times objective. Subsequent images were analyzed using the ImageJ (<http://rsb.info.nih.gov/ij/> [NIH 1997-2009]) application. The macro can be freely downloaded from the Web site of the institute at http://imaging.bii.a-star.edu.sg/projects/autophagy/Kriston_HCS_ImageJ_macro.zip. The segmenting and feature extracting algorithm identified nuclear objects through Hoechst 33342 (H0342) dye interaction with DNA and autophagosomes through MDC staining intensities.

Fluorescent image analysis

A range of computational, image processing, and statistical software is available for the phenotypic profiling of compound bioactivity at organelle, cell, cell subpopulation, or microtiter well levels. In the initial exploratory phase of this high-content cellular imaging work, the IN Cell Investigator (IN Cell Developer Toolbox, GE Healthcare, Piscataway, NJ), a commercially available application, facilitated the setup of a region growing-based imaging pipeline in a user-friendly manner and served as a benchmark imaging tool.

Image segmentation is a critical point in the imaging pipeline. Histogram-derived techniques are commonly used to perform this segmentation. These methods choose a brightness threshold (θ) by either maximizing the variance between pixel intensities associated with the foreground and background or minimizing the intraclass variance of foreground/background objects. The Isodata algorithm²³ is a representative of the former approach and the K-means clustering algorithm of the latter.

We used the fast, built-in ImageJ implementation of the Isodata algorithm to segment an image region into 1 foreground region and 1 background region. This technique was selected based on the observation that each image contained only 1 foreground and 1 background class. The Isodata algorithm initializes segmentation by dividing the maximal dynamic range of the image into 2 parts, representing foreground and background. The sample mean brightness of the foreground ($m_{f,0}$) and background ($m_{b,0}$) pixels is then calculated, and a new θ_1 is calculated by averaging the 2 mean values. Based on this new θ_1 value, the process is iterated until convergence to result in θ_k .²⁴

In those cases when images showed the existence of multiple foreground or background classes, we segmented those images with the K-means clustering^{25,26} algorithm. Given the fixed number of clusters based on a priori assumptions, K-means is a simple unsupervised learning algorithm to place pixels into clusters whose centroids have maximal separation.²⁷ The objective of the K-means segmentation algorithm is to minimize the total intracluster pixel brightness variance. Although the algorithm is theoretically considered to be

sensitive to randomly chosen pixels at initialization, it resulted in satisfactory segmentation in our practical application. We performed K-means clustering using Jarek Sacha's "K-means clustering" plug-in²⁶ (<http://ij-plugins.sourceforge.net/plugins/clustering/>).

Cell proliferation measurement requires the precise identification of the nucleus number in each well. Spatial staining variability in H0342 staining can, however, result in a mixed population of weakly and strongly fluorescent nuclei. This variation complicates the accurate identification of these nuclei across the image. Segmentation methods can compensate for these variations in staining quality. These methods can be applied on an image either globally or locally. However, global whole-image segmentation with a single threshold (θ) cannot compensate for variable weak and strong staining resulting from variations in protocol or tissue type. In general, a global segmentation method can be effective when working with uniformly fluorescent foreground pixels over uniformly dark background pixels with no spatial changes.

In our case, segmentation with a global threshold can correctly identify strongly fluorescent nuclei but leads to false negatives by missing weakly stained nuclei. In addition, histogram-derived segmentation techniques using a single global threshold also cause false-negative pixels and/or false positives, if the global range of θ is spanned toward darker values.

Spectral bleed-through is the other factor that hinders the global application of a single-threshold segmentation. Both H0342 and MDC dyes are excited on the same wavelength (360 nm), and given some overlap between the broad emission spectra of both dyes, it is inevitable that some fluorescence emission from the MDC staining is detected in the nuclear fluorescence (H0342-optimized) channel (Ch1) as shown in **Figure 1a** and vice versa in the autophagosomal channel (Ch2). This may result in MDC-origin staining contaminating the nuclear channel (AV bleed-through) to such an extent that it appears brighter than weakly stained nuclei. This can especially pose difficulties in positive controls and wells containing higher compound concentrations.

For assay quality control, we used 3 well-established quality control (QC) metrics: coefficient of variation (CV), Z' factor, and signal-to-background (S/B) ratio.²⁰ Coefficient of variation was calculated to measure the precision relative to the mean values calculated for c^- as minimum and c^+ as maximum signals.

In the subsequent plate, uniformity analysis Z' values^{19,20} were calculated for total and mean AV intensities of the control wells. Z' takes into account the c^- and c^+ variability and the dynamic range of the assay too, and it is therefore an invaluable and widely used QC measure. The S/B ratio was expressed as $S/B = \mu_{c^+}/\mu_{c^-}$, denoting the ratio of the mean values of positive and negative controls, respectively. The R script that we developed for the calculation is freely downloadable from the

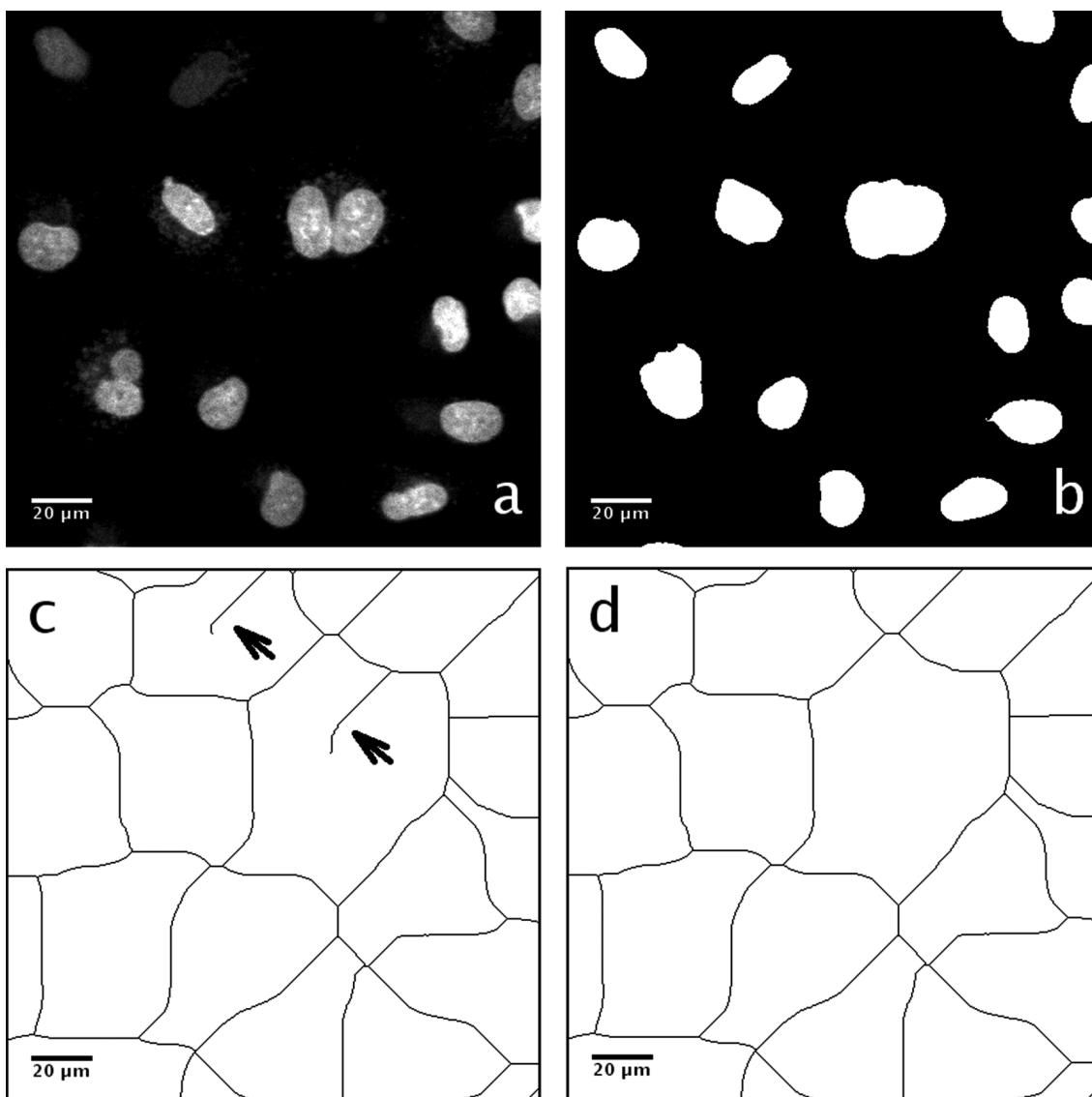


FIG. 1. Generation of the influence zone binary image. **(a)** Weakly and strongly stained H0342 fluorescent A549 nuclei in the nuclear channel, enclosed with monodansylcadaverine (MDC)-originated fluorescence emission. **(b)** Binary mask of the manual segmentation using θ_{stack} value. Note the unconstrained contour being adequate to generate the influence zone. **(c)** Skeletonized background of image **b**, with branch artifacts indicated by the arrows. **(d)** The influence zone grids after the pruning operation.

Bioinformatics Institute's Web site: http://imaging.bii.a-star.edu.sg/projects/autophagy/Kriston_HCS_R_script.R.

Bioimage processing and statistical analysis was embedded into the HCS workflow as **Figure 2** shows. Subsequent to the HCS assay preparation and image acquisition, subimages were defined through image preprocessing. Local segmentation was performed, followed by validation. Quantitative data were measured in the frame of feature extraction followed by assessing the quality of the assay. Statistical evaluation required the removal of irrelevant features, followed by data visualization.

Machine learning-based hit selection was the final stage of the workflow.

RESULTS

Preprocessing

Our proposed imaging workflow addresses both problems caused by the spectral bleed-through and punctate AV foreground by decomposing the image into subimages (influence

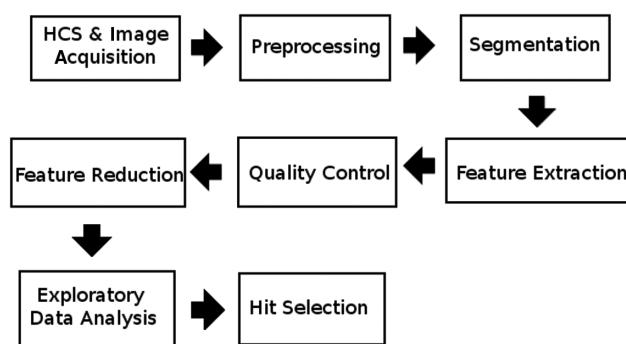


FIG. 2. Overall workflow diagram.

zones). An influence zone represents an equally divided image partition approximating cell boundaries around one or a few nuclear “seeds” that serve as predefined markers.²⁸ An influence zone image was generated based on the influence zone image generation pipeline presented in Supplemental Figure S1 using a Ch1 image (**Fig. 1a**, representative selection). A strong smoothening with a 15-pixel window size median filter was applied to remove noise. This was followed by a histogram-based global thresholding, with a manually selected common θ_{stack} value applied for all images (a stack) acquired from a microplate (**Fig. 1b**). An optional size filter removed any small segmentation artifacts, and the binary hole-filling algorithm reduced skeletonization artifacts. The background of each binary image was then thinned using the skeletonizing algorithm²⁹ (**Fig. 1c**) implemented in ImageJ, resulting in a grid image containing influence zones. Branch artifacts (**Fig. 1c**) still occurred as a by-product of the skeletonization. We customized Gabriel Landini’s “PruneAll” macro³⁰ and removed all branches of the binary skeleton, leaving only the closed loops (**Fig. 1d**).

Segmentation: nuclear channel

Local segmentation was designed to analyze each influence zone individually. Assumptions were chosen to be as permissive as possible. An influence zone on Ch1 was supposed to contain either 1 or more nuclei with optional autophagosomal bleed-through pixels forming weaker stained punctate regions than the nuclei. Because H0342-derived fluorescent pixels’ appearance was assumed on Ch2, the nuclear pixels were blanked to zero to eliminate nuclear bleed-through.

As shown in **Figure 3**, the next module of the imaging pipeline is composed of image preprocessing, local segmentation, and feature extraction of nuclei. The local region was extracted—that is, a Ch1 region enclosed by the contour of the underlying influence zone was duplicated and the copy processed as a global image. Image preprocessing included the application of a median filter to remove the noise of the

detector and at the same time keep structures and contours mostly intact.

Segmentation was performed on the median filtered result image. The K-means algorithm was used due to our observation that multiple foreground and/or background clusters exist in each Ch1 influence zone. Autophagosomal bleed-through led to multiple background clusters, where the centroid brightness value of the AV bleed-through cluster was typically lower than that of a strongly stained nucleus and higher than that of the local background. Based on this observation, 3 fixed clusters were selected indicating the local background, AV bleed-through, and nuclear foreground clusters.

A current limitation of the pipeline is the assumption that if multiple nuclei exist in a single influence zone, then they have similar brightness and thus fall into the same cluster. When the source of multiple foreground clusters is that of the H0342 staining variance, only those detected nuclei are clustered into the brightest centroid cluster. This limitation does not cause a systematic error, and the number of such influence zones is low. There is room for improvement of the speed of the imaging macro; currently, images of 1 to 3 plates can be processed overnight by a PC equipped with an Intel Core 2 Extreme X9650 quad core 3-GHz CPU and 16 GB memory. Influence zone generation takes an additional hour per plate, but the process can be run in parallel with other applications in a current multicore system.

During the binarization step, the 3 clusters with centroids μ_1 , μ_2 , and μ_3 were classified as foreground or background. The darkest cluster with the lowest centroid brightness value (μ_3) was always classified as local background and the brightest (μ_1) as foreground (i.e., nucleus). The middle cluster (μ_2) between those two acted as either nucleus or AV bleed-through depending on an empirically determined cluster centroid value. Because bitdepth of the images was 16, pixel intensity values ranged between zero and $2^{16} - 1 = 65,535$. We used θ_{nuc} as a threshold between strongly and weakly stained nuclei with a constant value such as $\theta_{nuc} = 10,000$. In practice, μ_2 was considered a nuclear cluster if $\mu_1 > \theta_{nuc}$ and $\mu_2 > \theta_{nuc}$. Furthermore, μ_2 was considered an AV bleed-through cluster if $\mu_1 > \theta_{nuc}$ and $\mu_2 < \theta_{nuc}$.

Ambiguity occurred when an influence zone contained a weakly stained nucleus ($\mu_1 < \theta_{nuc}$ and $\mu_2 < \theta_{nuc}$). In such cases, the centroids of nuclear and AV bleed-through clusters (if applicable) had similar brightness, and therefore it was not possible to separate them based on intensity. In the presence of an AV bleed-through, there was a chance for false-positive pixels if both μ_1 and μ_2 clusters were selected to be foregrounds. In the absence of an AV bleed-through, there was a chance for falsely segmented pixels, if the cluster μ_2 alone was selected to be foreground. Choosing a segmentation threshold as $\theta = (\mu_1 + \mu_2)/2$ resulted in a satisfying compromise when segmenting those influence zones.

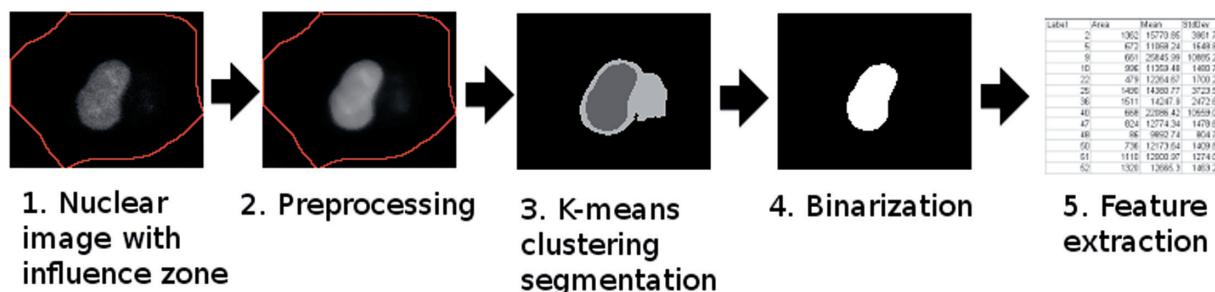


FIG. 3. Imaging pipeline of the proposed local segmentation and feature extraction of nuclei.

Nuclear segmentation can result in merged nuclei artifacts when nuclei stained with a similar fluorescence intensity were located near each other in the same influence zone. A watershed algorithm was applied to separate such clumped nuclei if its circularity exceeded 0.63, an empirically determined value.

Feature extraction: nuclear channel

Because the number of influence zones in an image did not always correlate with the number of nuclei, the nuclei were counted based on a separate nuclear segmentation and binarization. Median area size of $n = 4986$ negative control (DMSO-treated) nuclei were measured as $\bar{a} = 150 \mu\text{m}^2$, $\sigma = 71 \mu\text{m}^2$. A highly permissive empirical size threshold of $32 \mu\text{m}^2$ ($= 200 \text{ pixel}^2$) was applied, and smaller objects were not considered nuclei and were filtered out.

Morphology, intensity, and texture-based nuclear features were extracted after nuclear segmentation to describe nuclei numerically. Basic shape descriptors such as area, circularity, and Feret's diameter (also termed maximum caliper length, the longest distance between any 2 points along the nuclear contour) were computed using built-in ImageJ measures. Several additional nuclear morphological features were also computed using the Particles8 plug-in³⁰ (version 2.10) by Gabriel Landini: breadth (the largest axis perpendicular to Feret's diameter), area of the convex hull polygon, and radius of the minimal bounding circle, to name a few. The complete list is available at the author's Web site.³⁰ Particle coordinates were also extracted such as the binary blob's centroid and the center of mass based on the brightness-weighted centroid. Intensity-based nuclear features included total, mean, median, and standard deviation of intensity.

For textural features extraction, the Gray Level Correlation Matrix Texture Analyzer plug-in was used (GLCM_Texture version 0.4, authored by Julio E. Cabrera and available at <http://rsb.info.nih.gov/ij/plugins/texture.html> [accessed March 30, 2010]), which calculated 5 standard texture features (angular second moment, contrast, correlation, inverse difference moment, entropy) from the co-occurrence matrices.

Segmentation: AV channel

The result of nuclear segmentation was used in the local segmentation of autophagosomes (Fig. 4). The same influence zone used to analyze Ch1 was also used to analyze the corresponding Ch2. It was assumed at preprocessing that any pixel intensities located in the projected nuclear area on Ch2 (other than background) was the consequence of nuclear bleed-through. Therefore, Ch2 pixels superimposed by the nuclear mask were blanked to zero intensity.

During AV segmentation, Ch2 influence zone pixels (apart from the nuclear projection) were classified into 2 groups: local background and AV. This binary presumption allowed us to use the computationally fast Isodata algorithm.

In contrast to the region-growing segmentation approach, which allocated (4 or 8) connected pixels to a central seed (e.g., here the nucleus) to form a continuous region, possibly including background pixels, the histogram-based Isodata algorithm developed a sporadic-pattern foreground presuming no connection between pixels, hence identifying more strictly the punctate AV structures. The thresholding was automatic, and θ was calculated from the image histogram.

If more than one nucleus was found in the same influence zone, the matching AV region was calculated for each nucleus separately to ensure one-to-one correspondence between an AV region and a nucleus. A region-growing method was applied from the nuclear centroids to partition the given influence zone into subzones corresponding to each nucleus. Histogram-based segmentation was then applied to those AV subzones.

Feature extraction: AV channel

The number and total area of puncta per influence zone was calculated. Intensity-based features were also calculated: minimum, maximum, standard deviation, mean, median, and total intensity of the whole AV mask.

Precision of the segmentation was validated visually by superimposing the nuclear and AV mask contours on the respective original image. Both the proposed and the region-growing

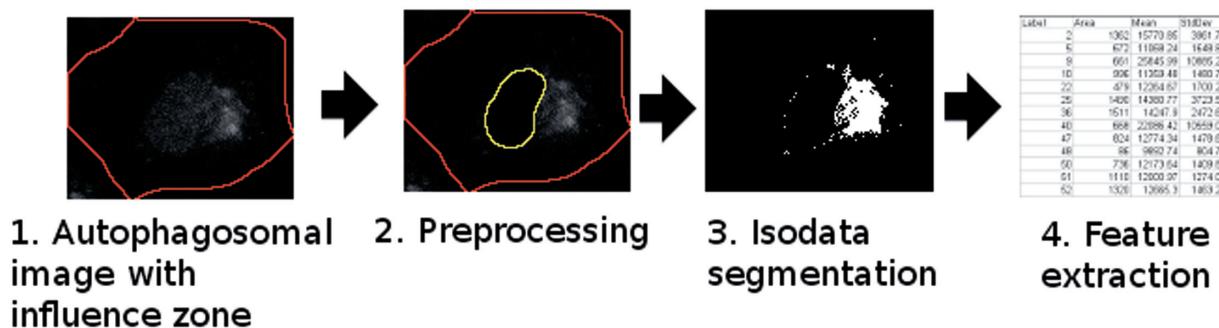


FIG. 4. The proposed autophagic vacuole (AV) segmentation and feature extraction pipeline.

segmentation method distinguished a hit from a nonhit, as depicted in **Figure 5c**.

Precise measurement of the number of nuclei is important to determine the toxicity of a compound. A ground truth image was created by manually segmenting a representative sample image in Ch1 based on pixel intensities evaluated by a human expert. The Ch1 ground truth image contained 365 nuclei (100%), the proposed segmentation resulted in the identification of 401 nuclei (110%), and the region-growing segmentation resulted in the identification of 427 nuclei (117%). Even though the proposed method had better performance, both segmentation methods measured the number of nuclei within the 20% error range that is used as a threshold to identify toxicity.

The quantification of pixel-level accuracy gives a numerical insight into the precision of the Ch2 AV segmentation. Common segmentation performance metrics³¹ were used for quantification. The measurement of true-positive, false-positive, true-negative, and false-negative pixel numbers enabled the calculation of precision, sensitivity (also called true-positive rate or recall), specificity, false-positive rate, and F-measure. Perfect segmentation is indicated by 100% values for precision, sensitivity, specificity, and F-measure. A near-zero false-positive rate value indicates good segmentation as well. One hundred percent sensitivity of pixel-level segmentation means that all real AV pixels are segmented as foreground pixels. The sensitivity was 71.5% for the proposed segmentation method and 71.0% for the region-growing segmentation. The specificity measure was 99.8% and 54.9% for the proposed and region-growing methods, respectively. The lower value of the latter was due to the high number of false-positive pixels (**Fig. 6**). For similar reason, the precision was 98% for the proposed segmentation method and 16% for the region-growing segmentation. The F-measure values were calculated as 82% and 26% for the proposed and region-growing segmentation, respectively. The false-positive rate was 0.2% and 45% for the proposed and region-growing segmentation, respectively.

Segmentation of AV areas using either the region-growing (**Fig. 6b**) or the histogram-derived (**Fig. 6c**) method yielded different results. By superimposing the contours of the 2 segmenting pipelines (**Fig. 6d**) on the original image (**Fig. 6a**), it becomes obvious that the proposed pipeline segmented the AV pixels only, whereas the region-growing pipeline included numerous background pixels as well. The proposed imaging pipeline demanded an extended early software development period, but it yielded significant advantages in segmentation precision. In the current assay, nuclear bleed-through is the most significant source of noise in the AV channel, which can be eliminated by blanking the nuclear region. Because the cytoplasmic region (marked in orange in **Fig. 6b**) contains minimal noise, both region-growing and histogram-derived methods yielded similar total AV intensities. This feature enables us to use the 2 methods for validation purposes.

Quality control

Following phenotypic feature extraction, statistical analyses were carried out to serve 2 main purposes: (1) to calculate plate quality metrics and (2) to provide decision support for hit selection. We developed a custom script in the R language (R Project for Statistical Computing) and used TIBCO Spotfire (Somerville, MA) for data visualization.

QC plays a crucial role in assay performance evaluation to determine if we can identify a hit with confidence. AV fluorescence intensities of negative control (c^-) and positive control (c^+) wells were used to compute an overall plate quality. Three established QC metrics were used: CV, Z' factor, and S/B ratio.

CV values were calculated on cell number as well as on the total and mean AV intensity of negative and positive control wells. For the majority of the plates, CV values fell under 0.15 using any segmentation method. Z' values of both total and mean AV intensities were above our range of acceptance³² ($Z' > 0.40$) in a representative plate, with $Z'_{AV\ mean} = 0.59$ and

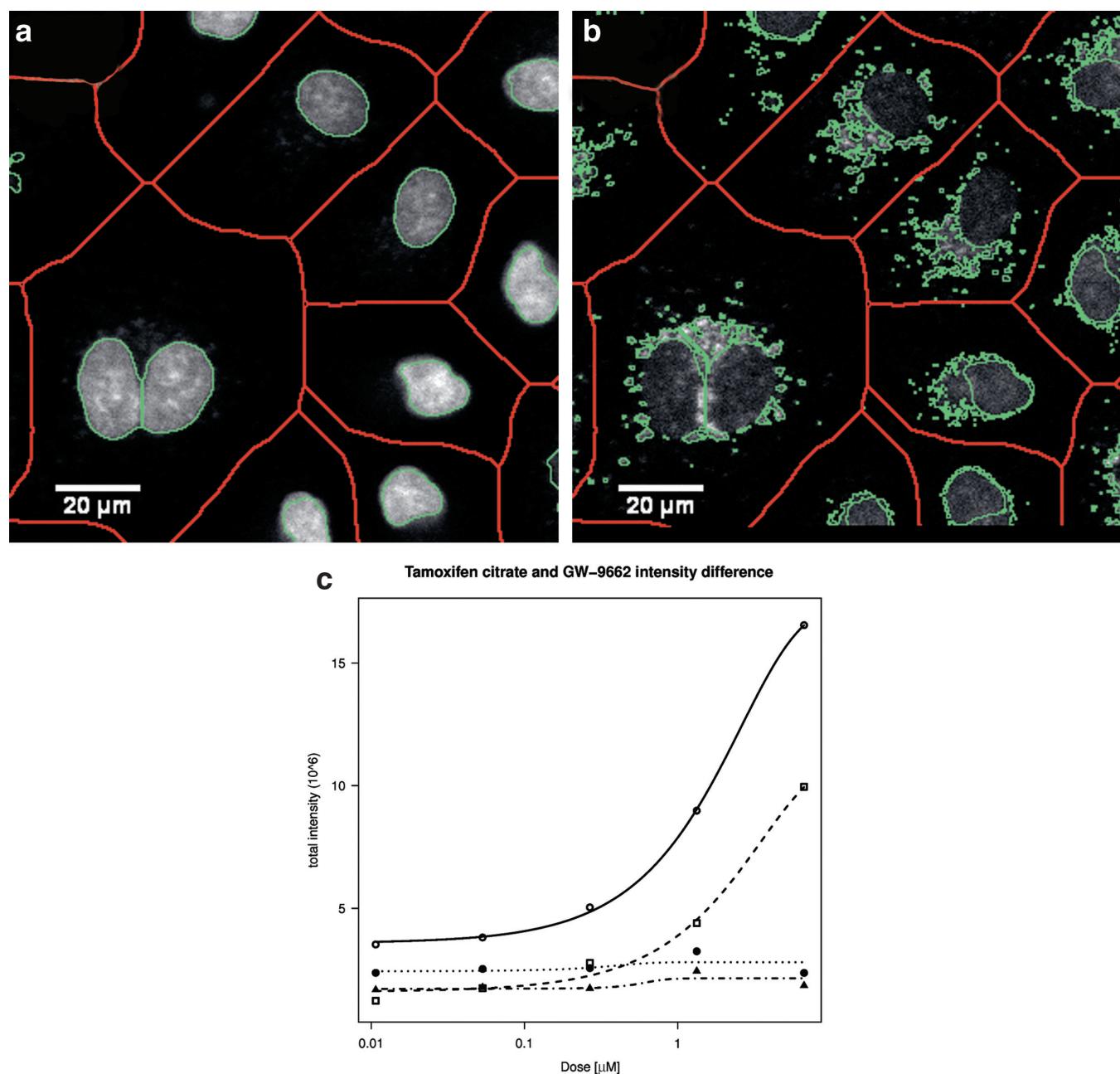


FIG. 5. Validation of the proposed segmentation. Green contours of the (a) nuclear and (b) autophagosomal binary mask are superimposed on the original images. Borders of influence zones are marked in red. Separation of hit compounds (c) is visualized by dose-response curves of hit compound tamoxifen citrate (open circles, solid) and nonhit compound GW-9662 (filled circles, dotted) processed by the proposed segmentation and the region-growing segmentation (open boxes, dashed and filled triangles, dash dotted, respectively).

$Z'_{AV\ total} = 0.66$. The medians of $Z'_{AV\ total}$ were 0.41 and 0.39 in plates of the primary and secondary screen, respectively. A high S/B ratio ($S/B > 2$) is required to evaluate the assay as screenable.²⁰ S/B distributions of our assays ranged above that limit, as shown in Supplemental Figure S2. In the supplemental figure, quality control boxplots of S/B ratio represent primary

(MDC) and secondary (LC3) assays, segmented by histogram-based (IJ) and region-growing (GE) methods. All assays by both methods spread above the acceptable $S/B > 2$ range.²⁰ The IJ and GE data sets showed significant difference ($p\text{-value}^{MDC} = 0.0016$ using the 2-sample t -test and $p\text{-value}^{LC3} = 0.0015$ using the Welch 2-sample t -test).

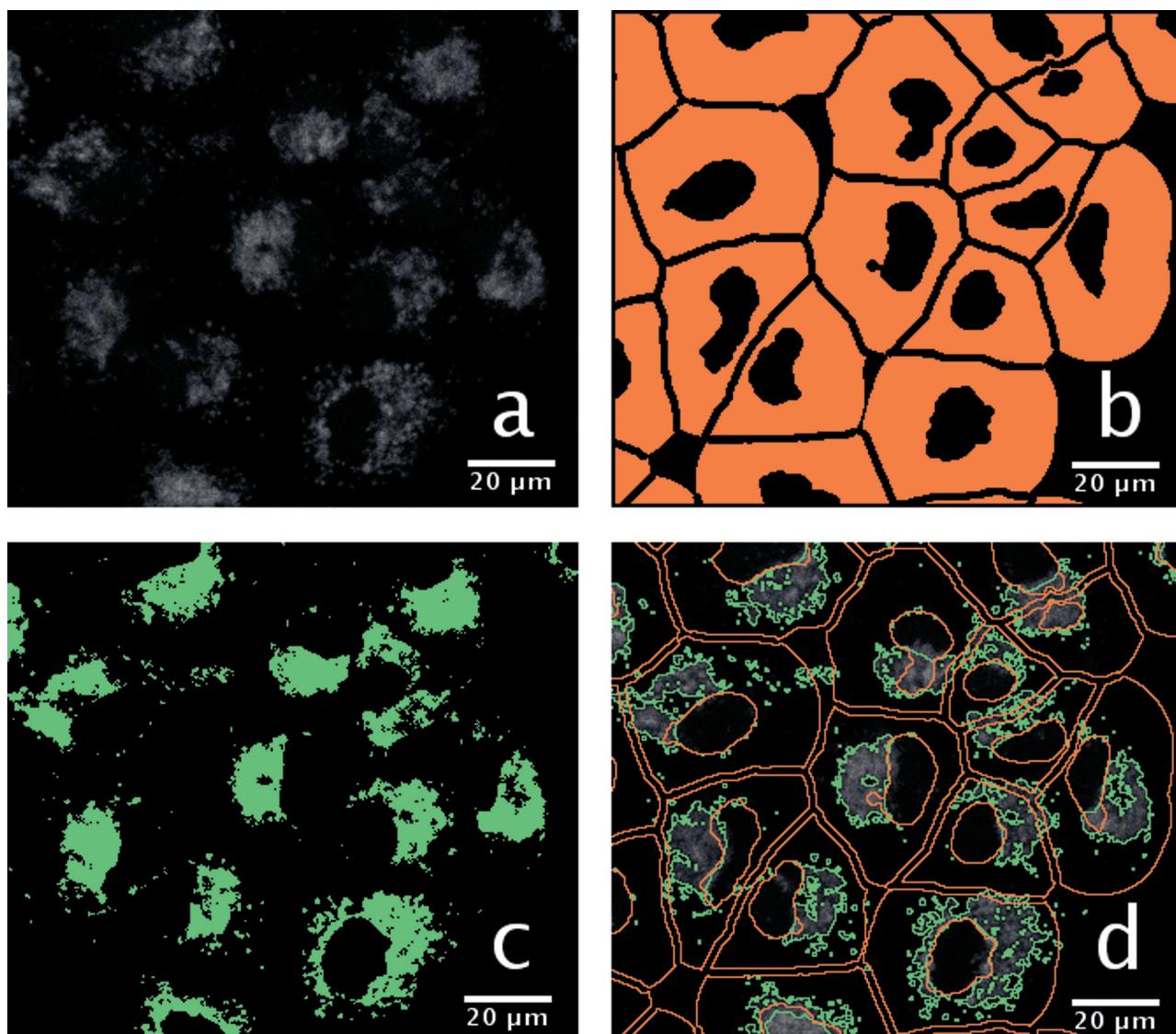


FIG. 6. Comparison of the proposed and the region-growing pipelines. **(a)** Original Ch2 autophagic vacuole (AV) fluorescence image of compound pimozone with 6.67 μ M concentration. **(b, c)** Binary mask of image **(a)** segmented by the **(b)** region-growing and the **(c)** proposed pipeline. **(d)** Binary mask contours of the region-growing (orange) and the proposed pipeline (green) superimposed on the original image.

Feature reduction

Throughout the feature collection phase of the workflow, a set of 44 nuclear and AV features was collected (intensity, morphology, and texture based) for each cell. The statistical correlation between these features was investigated by correlation analysis and hierarchical clustering. The results are shown on a heatmap with a dendrogram in **Figure 7**. The analysis revealed several groups of features with strong correlations. The alignments of these groups are highly similar in the primary and secondary screens. Correlated nuclear morphology features form 2 large groups shown on the upper left and lower right

parts of the primary assay correlation coefficient matrix. Nuclear intensity features (NucMin, NucMax, NucMean, NucMedian) and AV intensity features (CytoMin, CytoMax, CytoMean, CytoMedian) form 2 additional groups of correlated features. Nuclear textural features form 2 additional clusters. The minor and major axes of best-fitted ellipse on a nucleus are correlated with the nuclear area and Feret's diameter.

Feature reduction is an important practical step in data mining because irrelevant features “confuse” machine learning systems.³³

To identify redundant, irrelevant features and reduce dimensionality, a correlation-based feature selector was used to

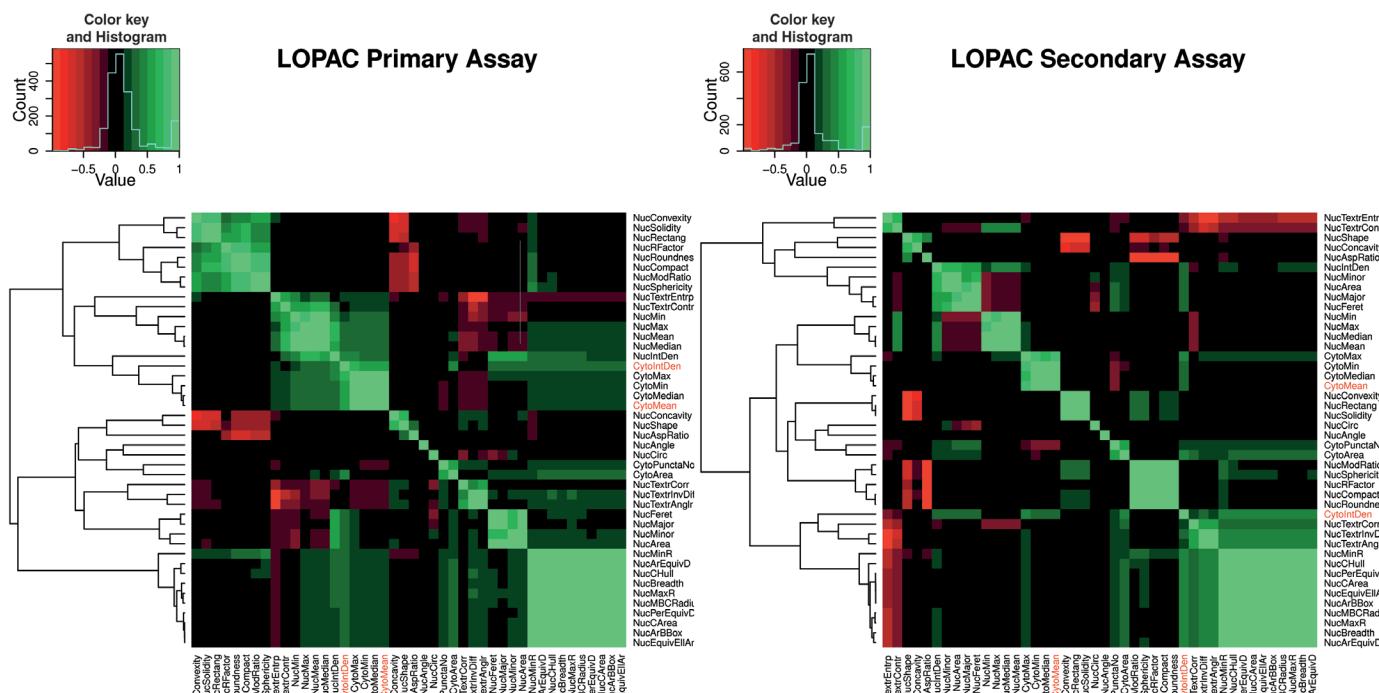


FIG. 7. Heatmap and dendrogram representation of primary screen (left) and secondary screen (right) correlation coefficients between the 44 collected features using R statistical software. Sample sizes: $N_{\text{primary}} = 1,473,461$ cells; $N_{\text{secondary}} = 395,017$ cells.

determine the predictive power of each feature. Features were restricted by a correlation-based feature selecting subset evaluator³⁴ (CfsSubsetEval). That algorithm was implemented under the Weka³³ machine learning software. CfsSubsetEval provides the advantage to assess the predictive ability of each feature individually, selecting those being highly correlated with the hit/nonhit class, but have low intercorrelation with other features. Weka's "BestFirst"³³ search method was performed on the feature space to find the subset that predicts the class best.

The feature selection method above determined how many times each feature was selected during a 10-fold cross-validation. Two features were selected: AV_{fold} and Cytotox; both are present in 9 out of 10 folds of our cross-validation, and hence the 9 (90%) entry in the primary assay and 80% (AV_{fold}) and 100% (Cytotox) in the secondary assay. AV_{fold} is a feature derived from MDC staining intensity. The AV median (CDM) of a well is calculated by the mean AV pixel intensities of each cell. AV_{fold} is calculated by dividing a compound's CDM by the median of the 8 DMSO wells' CDM (100%). Cytotox refers to compound toxicity where the cell number of a compound is divided by the mean cell numbers of 8 DMSO wells (100%).

The 2 selected features, AV_{fold} and Cytotox, are the dominant features with obvious biological interpretation, both in the primary and secondary assays. However, other features could also be considered to separate novel compounds based on either only the primary or the secondary assay 10-fold cross-validation.

The standard deviation of the nuclear area and the AV mean intensity shows a 10 (100%) entry only in the primary assay but not in the secondary assay. The mean, median, and standard deviation of AV total intensity features have a 100% entry value in the secondary assay but not in the primary assay. The above-mentioned features are similar to the dominant intensity and area features but were filtered out because the significance of these features was not consistent in the 2 assays.

Exploratory data analysis

The scatterplot of the 2 selected features (**Fig. 8**) displays 2 visually distinct, separated clusters of c^- and c^+ data. Cytotox thresholds of 61.79% and 33.33% are suggested for primary and secondary screens, respectively, by inspecting the Cytotox histogram. No c^- or c^+ data are found with a smaller Cytotox value.

Because outliers strongly affected the Support Vector Machine hyperplane calculation, we removed those c^+ values.

Hit selection

The supervised classification method Support Vector Machine (SVM) was used for automatic hit selection because it has been found to be superior in high-content cell identification.³⁵ We used the SVM implementation of the R-project (package e1071). The classifier was trained using the negative

HCS Image Analysis Pipeline for Identifying Selective Autophagic Inducers

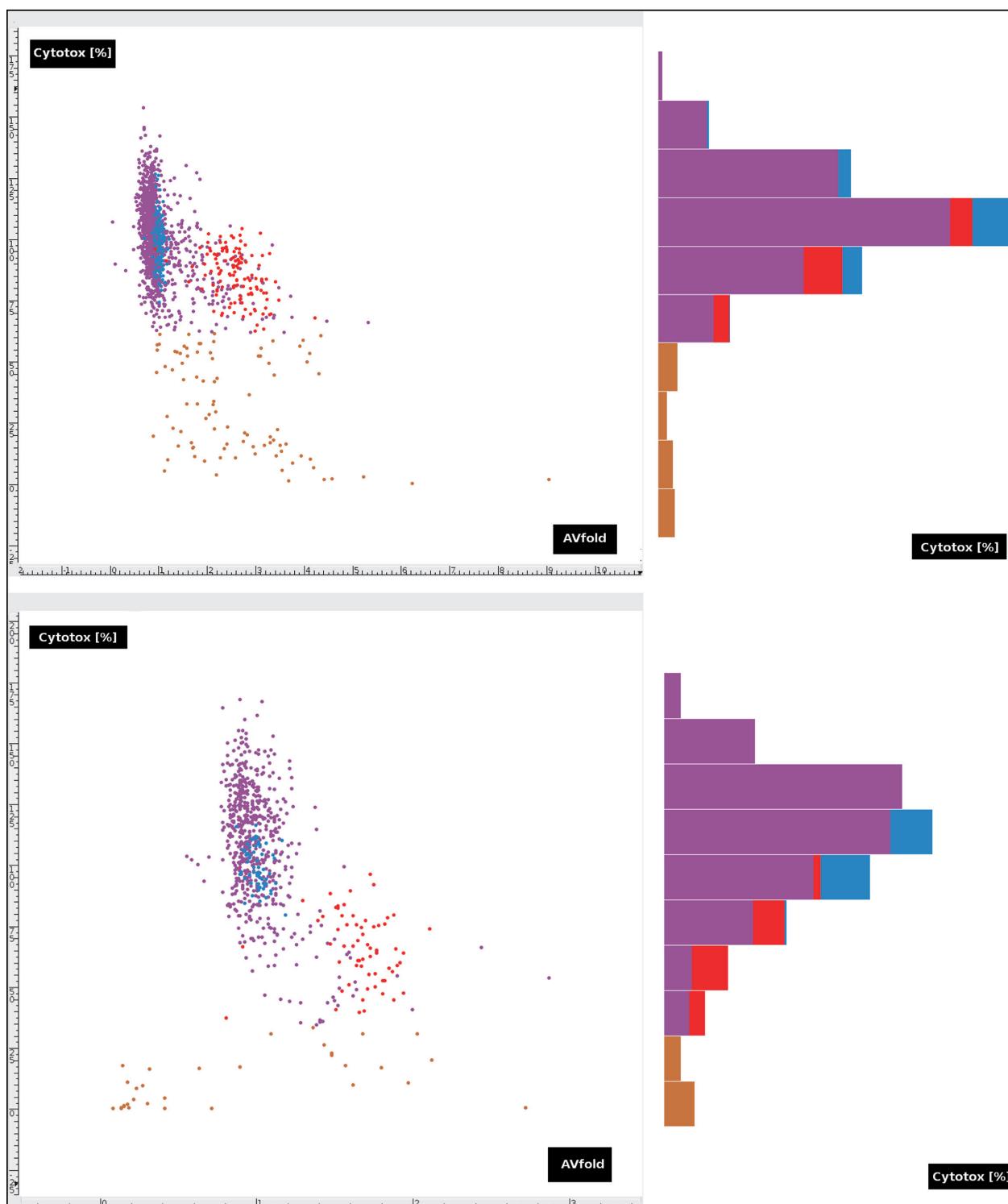


FIG. 8. Scatterplot of primary (upper left) and secondary (lower left) screen data forming 2 clusters of c⁻ (blue) and c⁺ (red) by AV_{fold} values on the x-axis and Cytotox (%) values on the y-axis. Compound data are shown as purple dots. Compound values with Cytotox <61.79% (primary) and 33.33% (secondary) (brown dots) are considered outliers by the histograms (right).

and positive controls. Outliers were removed from the training set to increase the reliability of the classification. To test our classifier, we applied a 100-fold cross-validation with a test set fraction of c^- and c^+ . The SVM using the linear kernel with default settings ($C = 1$, $\gamma = 0.5$) resulted in a model (see Suppl. Fig. S3) with total accuracy of 99.6%. The supplemental scatterplot shows the primary screen (left) and secondary screen (right) models calculated by the SVM classifier (linear kernel, $C = 1$, $\gamma = 0.5$) with c^- (red dots in region a) and c^+ (green dots in region b) data as the training set. Crosses indicate support vectors. The verification rate increased to 100% after removing c^+ outliers. No further SVM parameter optimization seemed necessary due to the high correct classification rate.

The c^+ component in the training data set of the secondary screen shows a slight cytotoxicity by shifting the c^+ cluster toward the lower Cytotox domain. In addition to the highly accurate classification, the large margins between the 2 training clusters resulted in a higher number of automatically selected hits than that of the threshold-based selected hits.

Using SVM as a machine learning approach and the nuclear number ratio cutoff for hit selection, 163 compounds were selected as hits out of 1280 in the primary assay. Following the prevailing practice, a subjective, expert-determined (threshold-based) hit selection procedure was also carried out using the results of the region-growing imaging pipeline, with thresholds $AV_{fold} > 2$ and Cytotox $> 80\%$. The cutoff value $AV_{fold} > 2$ was chosen because the human expert could explicitly detect such an intensity difference by eye. The threshold-based method suggested 159 compounds with an overlap of 96 compounds of the 162 proposed by the machine learning method (see the Venn diagram in Suppl. Fig. S4 where TH_PR and ML_PR are mean hits of the primary screen by threshold-based and machine learning methods, respectively).

In the secondary assay, 30 hits (100%) were suggested by the threshold-based method, of which 26 were also selected by the machine learning-based method, resulting in an 87% overlap (see the supplementary Venn diagram in Suppl. Fig. S5 where TH_SC and ML_SC are mean hits of the secondary screen by threshold-based and machine learning methods, respectively). The mutual overlap of the 4 distinct sets contains 19 compounds (shown in Suppl. Fig. S6). Four threshold-based selected compounds were not listed among the secondary hits selected by machine learning. The machine learning-based method resulted in 19 additional secondary hits due to the more permissive Cytotox $> 33.33\%$ threshold.

Machine learning-based primary hits contained 21 compounds (70%) of the 30 secondary assay hits suggested by the threshold-based method.

DISCUSSION

We have presented an image-processing high-content analytical pipeline that is readily applicable to high-content

cellular screens to identify novel selective pharmacological inducers of autophagy in compound libraries. Two different image segmentation approaches were evaluated. Histogram-derived segmentation algorithms such as K-means clustering and Isodata allowed pixel-level precision at segmenting the sporadic-pattern autophagosomes. Applying those algorithms locally enabled us to demonstrate a superior segmentation precision.

In cases where high levels of image noise are present, the region-growing segmentation approach may be adversely affected. However, the proposed segmentation approaches outlined here are less sensitive to image noise and should provide superior hit identification.

The speed of the proposed imaging macro currently limits the processing to 1 to 3 plates by overnight by a desktop PC, leaving potential room for future improvement. Implementing the current macro in the form of an ImageJ plug-in will provide an opportunity to exploit more multicore functions. A dedicated high-performance image-processing server would also improve the processing speed.

Several 2D single-cell phenotypic, morphology-based intensity and texture features were collected. QC measures remained above our range of acceptance.

To study the correlations between features, we applied correlation analysis and hierarchical clustering data analysis techniques. The features chosen for the actual hit selection were the same when chosen by the CFS subset evaluator and the “expert user.”

Finally, we applied the SVM machine learning technique that confirmed the histogram-derived image segmentation and algorithm methods run on the commercial software.

The hit selections obtained by the 2 methods were in very good agreement. The proposed pipeline can be done only with a robust HCS assay, where an expert user can decide on meaningful features to be extracted.

ACKNOWLEDGMENTS

The authors thank Wee Choo Puah, Chinta Rambabu, and Tiehua Du for fruitful discussions.

REFERENCES

1. Armstrong L: Fighting cancer is everyone's obligation. *J Clin Oncol* 2008;21:3473-3474.
2. Donnelly A, Galasko D, Golde T, Mulvany F, Wilcock G: Welcome to Alzheimer's research & therapy. *Alzheimers Res Ther* 2009;1:1.
3. Levine B, Kroemer G: Autophagy in the pathogenesis of disease. *Cell* 2008;132:27-42.
4. Levine B: Cell biology: autophagy and cancer. *Nature* 2007;446:745-747.
5. Klionsky DJ: The molecular machinery of autophagy: unanswered questions. *J Cell Sci* 2005;118:7-18.
6. Levine B, Klionsky DJ: Development by self-digestion: molecular mechanisms and biological functions of autophagy. *Dev Cell* 2004;4:463-477.

HCS Image Analysis Pipeline for Identifying Selective Autophagic Inducers

7. Kundu M, Thompson CB: Autophagy: basic principles and relevance to disease. *Annu Rev Pathol Mech Dis* 2008;3:427-455.
8. Zhang L, Yu J, Pan H, Hu P, Hao Y, Cai W, et al: Small molecule regulators of autophagy identified by an image-based high-throughput screen. *Proc Natl Acad Sci U S A* 2007;104:19023-19028.
9. He P, Peng Z, Luo Y, Wang L, Yu P, Deng W, et al: High-throughput functional screening for autophagy-related genes and identification of TM9SF1 as an autophagosome-inducing gene. *Autophagy* 2009;1:52-60.
10. Kabeya Y, Mizushima N, Ueno T, Yamamoto A, Kirisako T, Noda T, et al: LC3, a mammalian homologue of yeast Apg8p, is localized in autophagosome membranes after processing. *EMBO J* 2000;19:5720-5728.
11. Biederbick A, Kern HF, Elsasser HP: Monodansylcadaverine (MDC) is a specific in vivo marker for autophagic vacuoles. *Eur J Cell Biol* 1995;66:3-14.
12. Longo L, Platini F, Scardino A, Alabiso O, Vasapollo G, Tessitore L: Autophagy inhibition enhances anthocyanin-induced apoptosis in hepatocellular carcinoma. *Mol Cancer Ther* 2008;7:2476-2485.
13. Tasdemir E, Galluzzi L, Maiuri MC, Criollo A, Vitale I, Hangen E, et al: Methods for assessing autophagy and autophagic cell death. *Methods Mol Biol* 2008;445:29-76.
14. Hill A, LaPan P, Li Y, Haney S: Impact of image segmentation on high-content screening data quality for SK-BR-3 cells. *BMC Bioinform* 2007;8:340.
15. Forero MG, Pennack JA, Learte AR, Hidalgo A: DeadEasy caspase: automatic counting of apoptotic cells in Drosophila. *PLoS One* 2009;4:e5441.
16. Abramoff MD, Magelhaes PJ, Ram SJ: Image processing with ImageJ. *Biophotonics Int* 2004;7:36-42.
17. R Development Core Team R: *A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2009. <http://www.r-project.org>
18. Iversen PW, Eastwood BJ, Sittampalam GS, Cox KL: A comparison of assay performance measures in screening assays: signal window, Z' factor, and assay variability ratio. *J Biomol Screen* 2006;11:247-252.
19. Zhang J, Chung T, Oldenburg KR: A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen* 1999;2:67-73.
20. Inglese J, Johnson RL, Simeonov A, Xia M, Zheng W, Austin CP, et al: High-throughput screening assays for the identification of chemical probes. *Nat Chem Biol* 2007;3:466-479.
21. Jorgensen P, Nishikawa JL, Breitkreutz BJ, Tyers M: Systematic identification of pathways that couple cell growth and division in yeast. *Science* 2002;297:395-400.
22. Carpenter AE, Sabatini DM: Systematic genome-wide screens of gene function. *Nat Rev Genet* 2004;5:11-22.
23. Ridler TW, Calvard S: Picture thresholding using an iterative selection method. *IEEE Trans Syst Man Cybern* 1978;8:630-632.
24. Young IT, Gerbrands JJ, Vliet LJ: *Fundamentals of Image Processing*. Delft, The Netherlands: TU Delft, 1995.
25. MacQueen JB: Some methods for classification and analysis of multivariate observations. *Proc 5th Berkeley Symp Math Statist Prob* 1967; 1:281-297.
26. Jain AK, Dubes RC: *Algorithms for Clustering Data*. Englewood Cliffs, NJ: Prentice Hall, 1988.
27. Matteucci M: A tutorial on clustering algorithms [Online]. Retrieved from http://home.dei.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html. Accessed April 27, 2009.
28. Lantuejoul C, Maisonneuve F: Geodesic methods in quantitative image analysis. *Pattern Recognit* 1984;2:177-187.
29. Zhang TY, Suen CY: A fast parallel algorithm for thinning digital patterns. *Comm ACM* 1984;3:236-239.
30. Landini G: Advanced shape analysis with ImageJ. Paper presented at the ImageJ User and Developer Conference, Luxembourg, November 6-7, 2008. <http://www.dentistry.bham.ac.uk/landinig/software/software.html>. Accessed March 30, 2010.
31. Fawcett T: An introduction to ROC analysis. *Pattern Recognit Lett* 2006;27:861-874.
32. Lee JA, Cox K, Kriauciunas A, Chu S: The roles of high content cellular imaging in lead optimization. In Haney S (ed): *High Content Screening*. New York: John Wiley, 2008:249-268.
33. Witten IH, Frank E: *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco: Morgan Kaufmann, 2005.
34. Hall MA: *Correlation-based Feature Subset Selection for Machine Learning*. Hamilton, New Zealand: University of Waikato: Hamilton, 1998.
35. Dürr O, Duval F, Nichols A, Lang P, Brodte A, Heyse S, et al: Robust hit identification by quality assurance and multivariate data analysis of a high-content, cell-based assay. *J Biomol Screen* 2007;12:1042-1049.

Address correspondence to:

Janos Kriston-Vizi
Translational Research Resource Centre
MRC Laboratory for Molecular Cell Biology
University College London
Gower Street, WC1E 6BT
London, UK

E-mail: j.kriston@ucl.ac.uk