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 FLOTOW[2]

[1]Bioinformatics Institute (A*STAR)
30 Biopolis Street, Matrix #07-01
Singapore 138671
Tel: +65 6478-8353
Fax:+65 6478-9045
janos@kristonvizi.hu

[2]Experimental Therapeutic Centre (A*STAR)
Experimental Therapeutics Centre
31 Biopolis Drive, The Nanos #03-01
Singapore 138669
Tel: +65 6407-0340

ABSTRACT

Automated image processing is a critical and often rate-limiting step in a high-content screening (HCS) workflow. An imaging - statistical framework is described here with emphasis on segmentation to identify novel selective pharmacological inducers of autophagy. We screen a human alveolar cancer cell line and evaluate the images both by local adaptive and global segmentation. Region growing segmentation is compared with histogram-derived segmentation, where the latter allows a sporadic-pattern foreground, achieving pixel-level precision. Single-cell phenotypic features are measured and reduced after assay quality control check. Hit compounds selected by machine-learning correspond well to the subjective threshold-based hits determined by an expert analysis. Histogram-derived segmentation shows robustness against image noise, that adversely affects region growing segmentation.

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

INTRODUCTION

High content screening derived images contain vast amounts of data. The automated extraction of useful and decision supporting data from these images often represents a significant bottleneck in the HCS pipeline. The translation of phenotypic images into quantitative information of how compounds affect subcellular structures, for example, can benefit from automated imaging and statistical methods that 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. Cancer, Alzheimer's disease and Huntington's disease are thought to affect millions of individuals worldwide[1,2]. These diseases are linked to autophagy, making it an emerging therapeutic target in drug discovery. Autophagy is a highly regulated, homeostatic, intracellular catabolic mechanism by which eukaryotes degrade superfluous or faulty organelles and long-lived

proteins[3]. Being an evolutionarily highly conserved process, autophagy-specific genes have been characterized in distant species ranging from yeast to human[4]. Macroautophagy (henceforth referred to as autophagy) is one of the three 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[5]. The process 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, AV), which in turn fuse with lysosomes to form an autolysosome where the cytoplasmic material is sequestered and then degraded[6].

Autophagosomal activation can lead to cell death through excessive self-digestion, thus many anticancer agents have autophagy inducing capability. In addition, cell growth is negatively regulated by autophagy, in which may slow down tumour growth[7].

High-content screening (HCS) for autophagy activators is still limited, despite an emerging research interest. A recent high-content cellular image analysis study identified eight autophagic cell death inducer compounds[8] and a cell-based functional screening revealed three 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) 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, though 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 compartments[11], thus it is suitable to mark the generation of autolysosomes[12]. It is considered rather unspecific due to staining not only the acidic autophagic vesicles, but also numerous other acidic organelles of the cell[13]. Notwithstanding, coupled with a confocal microscopy system rather than that of widefield, MDC staining can be a satisfactory indicator of autophagy activation. It fits well into a high-throughput phenotypic screening strategy, where the response from a simpler and cost-effective MDC-based primary assay can be confirmed by a secondary assay using immunofluorescent LC3B antibody marker.

Besides specificity, selectivity is crucial throughout hit selection process: we want 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 the multispectral data derived from them. A typical workflow starts with image preprocessing to improve the quality of raw images, followed by image segmentation, which is a step where fluorescent regions of interest are separated from the background. Segmentation is a critical component of the image processing system, errors at this phase profoundly influence hit selection. A decline in segmentation accuracy may lead to reduced segmentation performance as such an event discredits all numerical metrics derived from it. In contrast with the blob-like nucleus, 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[14]. The need for precise segmentation of a sparsely stained AV region together with the need of spectral crosstalk (bleed-through) correction, requires a multi cluster-based segmenting

algorithm as well as source code level customization. As this capability is not available in commercial image analysis software, we are incorporating it into a customizable open-source image analysis applications, ImageJ[15].

We combine this image analysis with a well-established statistical and analytical tool, R[16] (http://r-project.org), an extensively used open-source platform for exploratory data analysis, descriptive statistics, assay quality control[17,18] and performance determination as well as for significance tests, as well as a commercial data visualization application SpotFire (Somerville, Massachusetts, United States). A statistical-method based feature selection is used to eliminate irrelevant or unhelpful numerical descriptors, followed by machine learning based hit selection. Hit compounds selected from the primary screen are then confirmed in secondary screens. The initial quantitative and automated processing of the images and derived data largely contributes to an unbiased hit selection of hits from such phenotypic assays, in contrast to prevailing subjective methods based mostly on expert curation and analysis[19,20].

The objective of this paper is to outline a simple image analysis - statistical framework, that can be applied readily to a high-content cellular screen of cancer cell lines to identify novel selective pharmacological inducers of autophagy in compound libraries.

MATERIALS

Compounds for autophagy phenotypic screen

We used the LOPAC[1280] compound library (Sigma-Aldrich) as our test compounds; trifluoperazine dihydrocholoride (Sigma-Aldrich) as our positive control; and Dimethylsulphoxide (DMSO) (Sigma-Aldrich) as our negative control.

METHODS

Cell culture and cell plating

Human alveolar carcinoma A549 (lung epithelial cancer) cells were seeded into 96-well transparent, flat-bottom plates (Greiner) at a density of 5000 cells per well in DMEM containing 4% Fetal bovine serum (Hyclone). The cells were then incubated with library compounds at 5 concentrations (10uM, 2uM, 0.4uM, 0.08uM, 0.016uM) for 48 hours.

Primary assay using MDC staining

The live cells were stained with monodansylcadaverine (MDC) as follows. 100 μl staining medium containing 100 μM MDC was added to all wells and incubated for 40 minutes at 37^{o}C. Excess MDC was then washed off with 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).

Secondary assay using LC3B staining

A549 cells were fixed in 4% paraformaldehyde for 10 min at room temperature and then permeabilised with 0.1% Triton X-100 for a further 10 min. The fixed cells were then blocked with

3% BSA for 30min. To stain for LC3B, we incubated the cells with anti-LC3B rabbit polyclonal antibody at 1:2000 dilution (Novus, #NB600) and subsequently labelled it with 1:2000 dilution of anti-rabbit AlexaFluor546 secondary antibody (Invitrogen) for visualisation. 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 ImageXpress Ultra (Molecular Devices) system using 20x objective. Subsequent images were analyzed using ImageJ (http://rsb.info.nih.gov/ij/ NIH 1997-2009) application. The segmenting and feature extracting algorithm identified nuclear objects through Hoechst 33342 (HO342) dye interaction with DNA, and autophagosomes through MDC staining intensities.

Fluorescent Image Analysis

A range of computational tools, image processing and statistical softwares are available for phenotypic profiling of compounds' bioactivity at individual organelle, cell, cell subpopulation or microtiter well level. In the initial exploratory phase of the high-content cellular imaging work, the IN Cell Investigator (IN Cell Developer Toolbox, GE Healthcare) commercially available application facilitated the set-up 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 of imaging pipeline where histogram-derived techniques are commonly used. These methods choose a brightness threshold (θ) by either maximizing the variance between pixel intensities associated with the foreground and background or minimizing the intra-class variance of foreground/background objects. The Isodata algorithm[21] is a representative of the former approach, and the K-means clustering algorithm for the latter.

We use the fast, built-in ImageJ implementation of Isodata algorithm to segment an image region into one foreground region and one background region by observing on the images that only one foreground and one background class exists there. The Isodata algorithm initializes the segmentation by dividing the maximal dynamic range of the image into two parts, representing foreground and background. Then the sample mean brightness of the foreground ($m_{f,0}$) and background ($m_{b,0}$) pixels are calculated, and a new θ_1 is calculated by averaging the two mean values. Based on this new θ_1 value, the process is iterated until idempotence to result θ_k [22].

In those cases when images showed the existence of multiple foreground or background classes, we segmented those images with K-means clustering[23,24] 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 centroid brightnesses from each other locate as far as possible[25]. The objective of the K-means segmentation algorithm is to minimize the total intra-cluster pixel brightness variance. Although the algorithm is theoretically considered to be sensitive to the randomly chosen pixels at initialization, it results satisfactory segmentation in our practical application, using Jarek Sacha's K-means clustering plugin implementation under ImageJ.

Cell proliferation measurement requires the precise identification of nucleus number in each well, though the spatial staining variability in HO342 staining leads a mixed population of weakly and strongly fluorescent nuclei. Imaging techniques can compensate variations in staining quality. Segmentation methods can be applied on an image either globally or locally. However the whole

image level (global) use of a segmentation algorithm with a global θ can not compensate the defiency of weak and strong staining accounted for the 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 θ can correctly segment strongly fluorescent nuclei, but leads to false negatives by missing the randomly located weakly stained nuclei. Histogram-derived segmentation techniques using any single global θ causes false negative pixels and/or false positives, if the global range of θ is spanned towards darker values. Spectral bleed-through is the other factor that hinders the global application of a single-threshold segmentation. Both HO342 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 (HO342-optimized) channel (Ch1) as shown in Fig. 2a and vice versa in the autophagosomal channel (Ch2). This causes MDC-origin staining to appear on nuclear channel (AV bleed-through) that are brighter than weakly stained nuclei especially in positive controls and wells containing higher compound concentrations.

For assay quality control we use three well established QC metrics: coefficient of variation (CV), Z' factor and signal to background (SB) ratio.
Coefficient of variation is calculated as in Eq. (1) to measure the precision relative to the mean values calculated for c^- as minimum and c^+ as maximum signals.

$$CV = \frac{\sigma}{\mu} \qquad [1]$$

where μ is the mean, σ is the standard deviation of the assay signal in terms of AV fluorescence or cell number.

In the consequent plate uniformity analysis Z' (Eq. 2), values are 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 is therefore an invaluable and widely used QC measure.

$$Z' = 1 - \frac{(3\sigma_{c^+} + 3\sigma_{c^-})}{|\mu_{c^+} - \mu_{c^-}|} \qquad [2]$$

The signal to background (SB) ratio is expressed as SB = μ_{c+} / μ_{c-} denoting the mean values of positive and negative controls respectively.

Bioimage processing and statistical analysis is embedded into the HCS workflow as Figure 1 shows. Subsequent to the HCS assay preparation and image acquisition, sub-images are defined through image preprocessing. Local segmentation is performed followed by validation. Quantitative data is measured in the frame of feature extraction followed by assessing the quality of the assay. Statistical evaluation requires the removal of irrelevant features, followed by the data visualization. Machine learning-based hit selection is the final stage of the workflow.

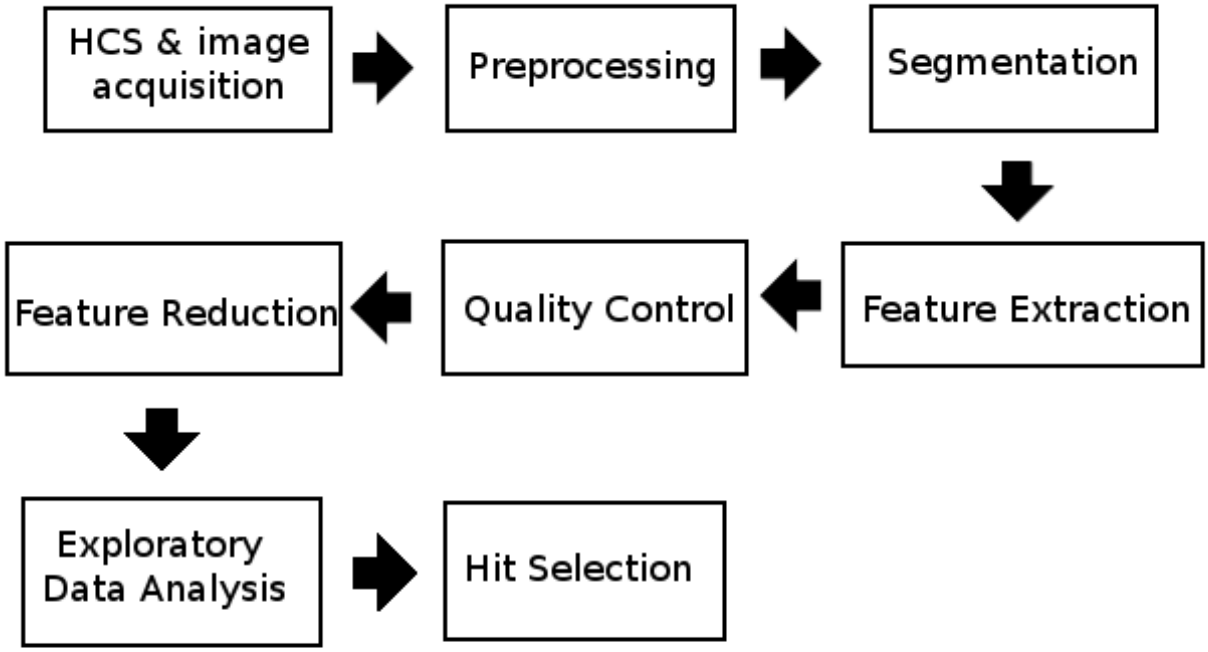


FIG. 1. Overall workflow diagram.

RESULTS

Preprocessing

Our proposed imaging workflow addresses both problems caused by the spectral bleed-through and punctate AV foreground by decomposing the image into sub-images (influence zones). An influence zone represents an equally divided image partition approximating cell boundaries around one or a few nuclear "seeds" that serves as predefined markers[26]. An influence zone image is generated based on the imaging pipeline presented on Figure 3 using a Ch1 image (Fig. 2a., representative selection). A strong smoothening with a 15 pixel window size median filter is applied to remove noise. This is followed by a histogram-based global thresholding, with a manually selected common θ_{stack} value applied for all images (a stack) acquired from a microtiter plate (Fig. 2b.). An optional size filter removes the small segmentation artifacts, and the binary hole filling algorithm reduces skeletonization artifacts. The background of each binary image is then thinned using the skeletonizing algorithm[27] (Fig. 2c.) implemented in ImageJ, resulting a grid image containing influence zones. A customized pruning algorithm based on Gabriel Landini' PruneAll macro is applied until idempotence to remove all branches of the binary skeleton, leaving only the closed loops (Fig. 2d.).

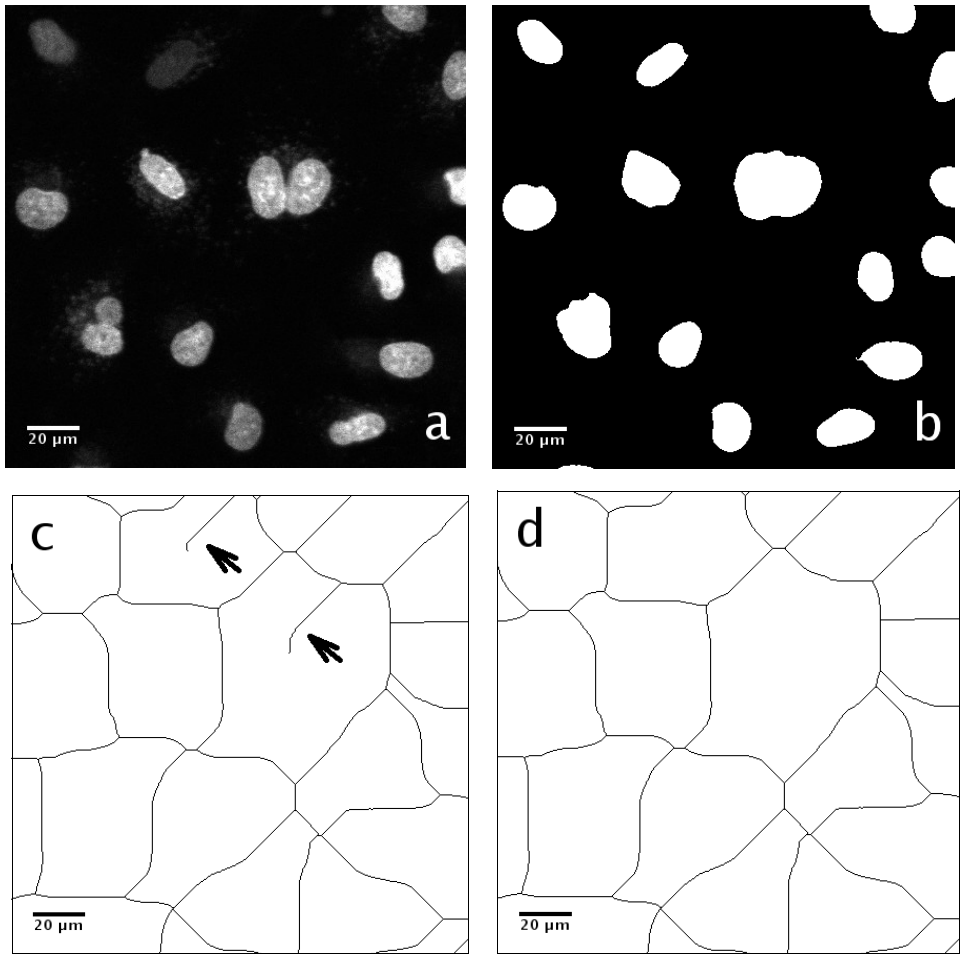


FIG. 2. Generation of the influence zone binary image. (a) Weakly and strongly stained HO342 fluorescent A549 nuclei in the nuclear channel, enclosed with MDC originated fluorescence emission. (b) Binary mask of the manual segmentation using θ_{stack} value. Note the unconstrained contour, being adequate to generate influence zone. (c) Skeletonized background of image b, with branch artifacts indicated by the arrows. (d) The influence zone grids after pruning operation.

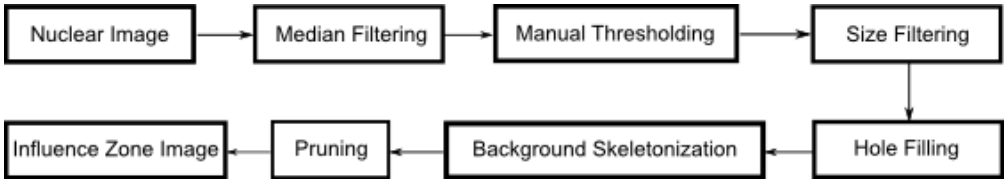


FIG. 3. Imaging pipeline of the proposed influence zone image generation.

Segmentation: Nuclear channel

Local segmentation is designed to analyze each influence zone individually. Assumptions are chosen to be as permissive as possible. An influence zone on Ch1 is supposed to contain either one or more nuclei with optional autophagosomal bleed-through pixels forming weaker stained punctate regions than that of nuclei. Since HO342 derived fluorescent pixels appearence is assumed on Ch2,

therefore the nuclear pixels are blanked to zero in order to eliminate nuclear bleed-through.

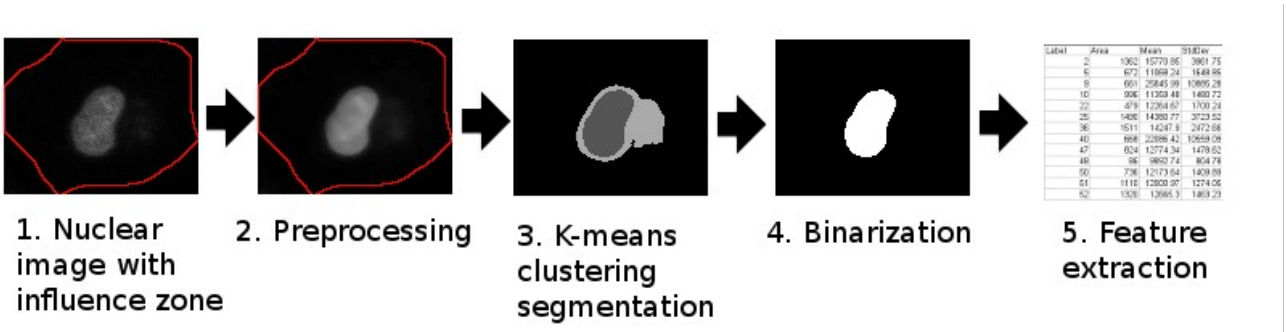


FIG. 4. Imaging pipeline of the proposed local segmentation and feature extraction of nuclei.

As shown in Figure 4, the next module of the imaging pipepline is composed of image preprocessing, local segmentation and feature extraction of nuclei. The local region is extracted, that is a Ch1 region enclosed by the contour of the underlying influence zone is duplicated, and the copy is processed as a global image. Image preprocessing includes the application of a median filter that removes the noise of the detector and the same time keeps structures and contours mostly intact.

Segmentation is performed on the median filtered result image. K-means algorithm is used due to our observation that multiple foreground and/or background clusters exist in each Ch1 influence zone. Autophagosomal bleed-through leads to multiple background clusters, where the centroid brightness value of AV bleed-through cluster is typically lower then that of a strongly stained nuclei's and higher then that of the local background's. That observation results to set the fixed cluster number to three, indicating the local background, AV bleed-through and nuclear foreground clusters.

Current limitation of the pipeline is derived from the assumption that an influence zone contains multiple nuclei, with similar brightness, thus falling into the same cluster. When the source of multiple foreground clusters is that of the HO342 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 was low. There is room for improvement at the speed of the imaging macro, currently images of 1-3 plates can be processed overnight by a PC equipped with an Intel Core 2 Extreme X9650 quad core 3GHz CPU, 16 GB memory. Influence zone generation takes an additional hour per plate, however the process can be run parallel with other applications in a current multicore system.

During the binarization step, the 3 clusters such as μ_1, μ_2, μ_3 are classified as foreground or background. The darkest cluster with the lowest centroid brightness value (μ_3) is always classified as local background, and the brightest (μ_1) as foreground i.e. nucleus. The middle cluster (μ_2) between those two can act as either nucleus or AV bleed-through depending on an empirically determined cluster centroid value. Since bitdepth of the images is 16, pixel intensity values range between zero and 2^{16} -1 = 65535. In the following, θ_{nuc} is used as a threshold between strongly and weakly stained nuclei with a constant value such as θ_{nuc} = 10000. Practically μ_2 is considered as a nuclear cluster and $\mu_1 > \theta_{nuc}$ and $\mu_2 > \theta_{nuc}$. Furthermore μ_2 is considered as an AV bleed-through cluster if $\mu_1 > \theta_{nuc}$ and $\mu_2 < \theta_{nuc}$.

Ambiguity occurs when an influence zone contains 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) are similarly bright, and therefore it is not possible to separate them based on intensity. In the presence of an AV bleed-through there is a chance for false positive pixels if both μ_1 and μ_2 clusters are selected to be foregrounds. In the absence of an AV bleed-through for false negative pixels, if cluster μ_2 alone is selected to be foreground. Choosing a segmentation threshold as $\theta = (\mu_1 + \mu_2)/2$ results in a satisfying compromise when segmenting those influence zones.

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

Feature extraction: Nuclear channel

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

Morphology, intensity and texture-based nuclear features are extracted after nuclear segmentation in order to describe nuclei numerically. Basic shape descriptors such as area, circularity and Feret's diameter (syn. caliper length, the longest distance between any two points along the nuclear contour) are computed using built-in ImageJ measures. Several additional nuclear morphological features are also computed using Particles8 plugin (ver. 2.10) by Gabriel Landini[28]: breadth (the largest axis perpendicular to the Feret's diameter), area of the convex hull polygon, radius of the minimal bounding circle, to name a few. The complete list is available at the author's website. Particle coordinates are also extracted such as the binary blob's centroid and the center of mass based on the brightness-weighted centroid. Intensity-based nuclear features include total, mean, median, standard deviation of intensity.

For textural features extraction, the plugin – Gray Level Correlation Matrix Texture Analyzer is used (GLCM_Texture ver. 0.4 authored by Julio E. Cabrera), which calculates four 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 is used in the local segmentation of autophagosomes (Figure 5). The same influence zone is used to analyze Ch1 that was used to analyze the corresponding Ch2. It is assumed at preprocessing, that any pixel intesities located in the projected nuclear area on Ch2 (other than background) is the consequence of nuclear bleed-through. Therefore, Ch2 pixels superimposed by the nuclear mask are blanked to zero intensity.

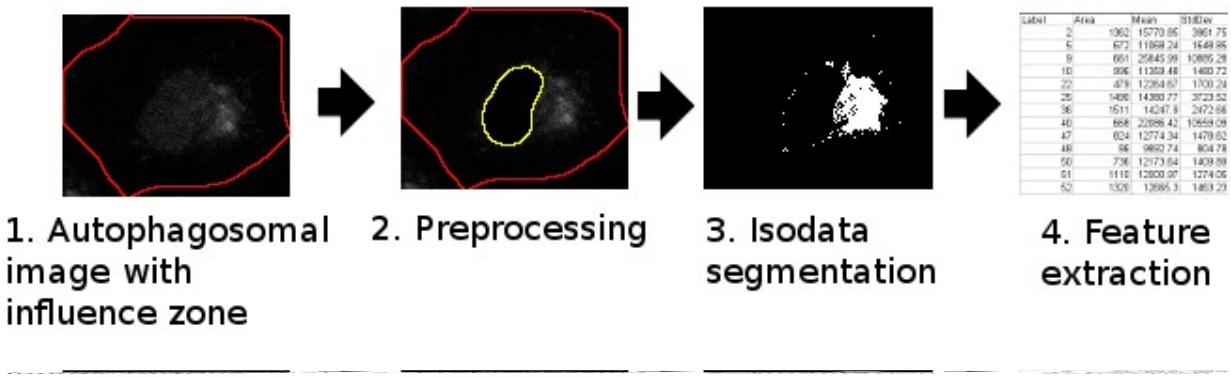


FIG. 5. The proposed AV segmentation and feature extraction pipeline.

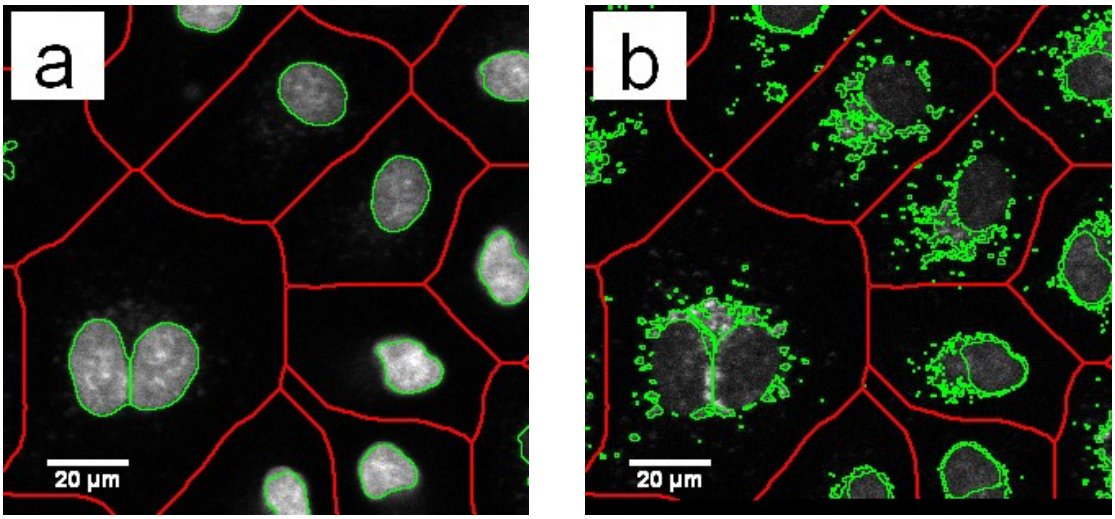
During AV segmentation, Ch2 influence zone pixels apart from the nuclear projection are classified into two groups: local background and AV. This binary presumption allows us to use the computationally fast Isodata algorithm. Distinct to region growing segmentation, that allocates (4- or 8-) connected pixels to a central seed (e.g. here the nucleus) to form a continuous region by possibly including background pixels, the histogram-based Isodata algorithm develops a sporadic-pattern foreground presuming no connection between pixels, extracting more strictly the punctate AV structure. The thresholding is automatic and θ is calculated from the image histogram.

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

Feature extraction: AV channel

The number and total area of pucta per influence zone are calculated. Intensity-based features are 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 segmentation method distinguish a hit from nonhit as depicted in Figure 6c.



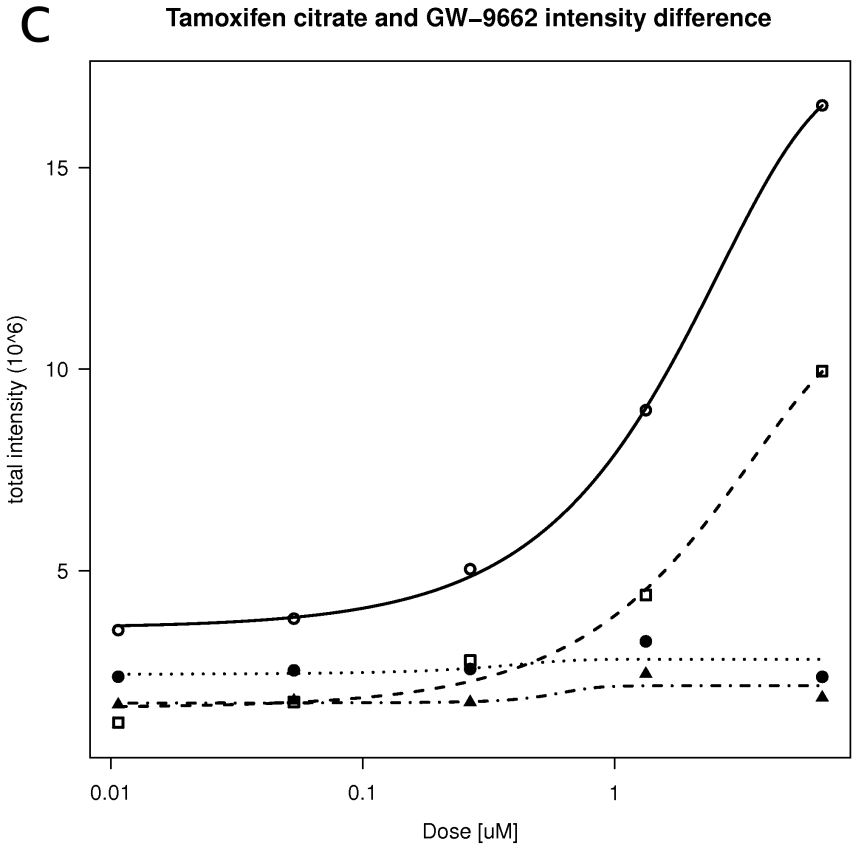


FIG. 6 Validation of the proposed segmentation. Green contours of nuclear (a) and autophagosomal (b) binary mask are superimposed on the original images. Borders of influence zones are marked with red color. Separation of hit compounds (c) visualized by dose response curves of hit compound tamoxifen 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 trianges, dashdotted respectively).

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 contains 365 nuclei (=100%), the proposed segmentation results 401 nuclei (110%) and the region growing segmentation results 427 nuclei (117%). Even though the proposed method shows better performance, both segmentation methods measures 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 metrics[29] are used for quantification. The measurement of true positive, false positive, true negative and false negative pixel numbers enables calculating precision, sensitivity, (also called as true positive rate or recall), specificity, false positive rate, and F-measure. Accurate segmentation is represented by 100% in case of precision, sensitivity, specificity and F-measure, whereas the closer false positive rate value is to zero the better. 100% sensitivity of pixel level segmentation means that all AV real 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 found 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. 7). For similar reason, the precision was found 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.

DISCUSSION

Segmentation of AV areas using either the region growing (Fig. 7b) or the histogram-derived (Fig. 7c) method yields different results. By superimposing the contours of the two segmenting pipelines (Fig. 7d) on the original image (Fig. 7a), it becomes obvious that the proposed pipeline segments the AV pixels only, while the region growing pipeline includes 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. Since the cytoplasmic region (marked as orange in Fig. 7b) contains minimal noise, both region growing and histogram-derived methods yielded similar total AV intensities. This feature enables us to use the two methods for validation purposes.

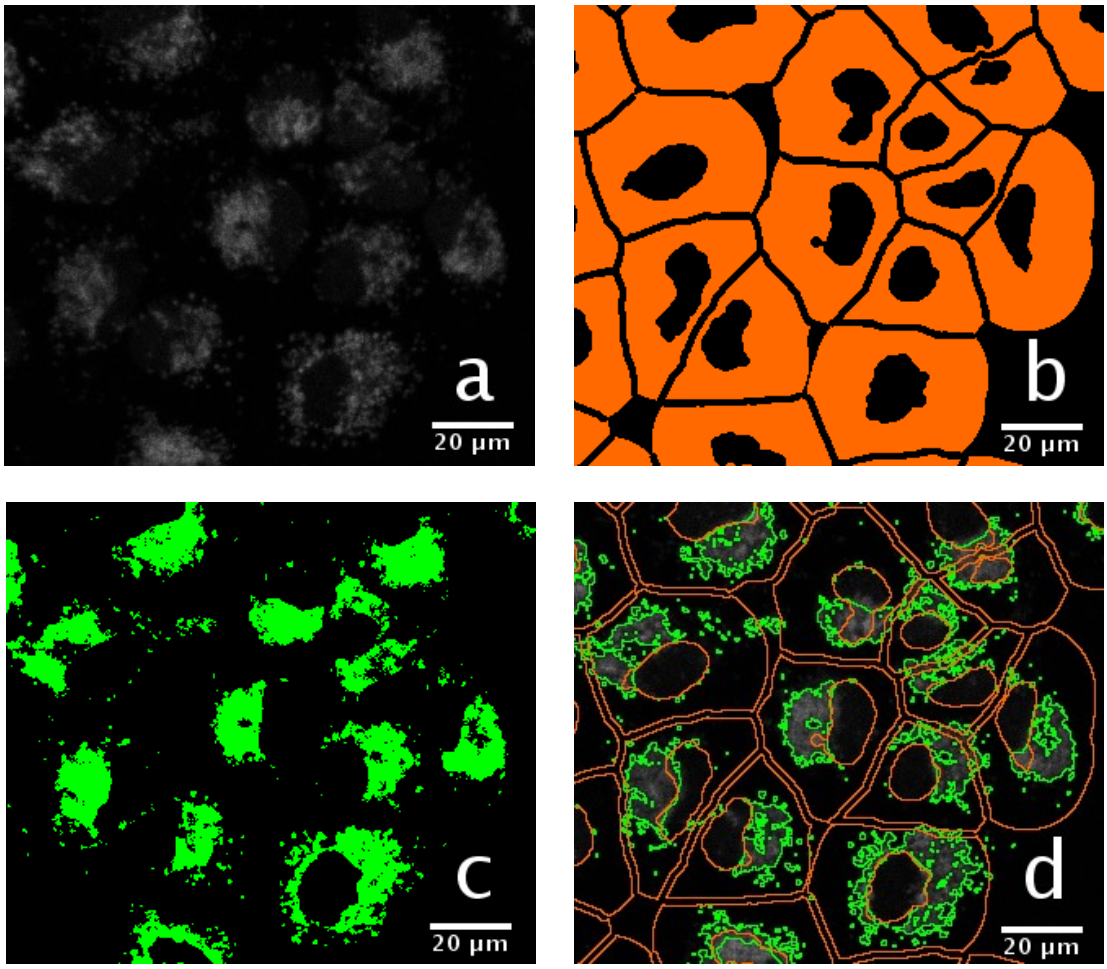


FIG. 7 Comparison of the proposed and the region growing pipelines. (a) Original Ch2 AV fluorescence image of compound pimozide with 6.67 µM concentration. (b,c) Binary mask of image

a) segmented by the region growing (b) and the proposed (c) pipeline. (d) Binary mask contours of the region growing (orange) and the proposed pipeline (green) superimposed on the original image.

Quality Control

Following phenotypic feature extraction, the statistical analysis is carried out by serving two main purposes: (i) compute quality metrics of a plate and (ii) provide decision support for hit selection of the screen. We developed a custom script in R language for statistical analysis (R Project for Statistical Computing) and used TIBCO SpotFire (Somerville, Massachusetts, United States) for data visualization.

Quality control (QC) plays a crucial role in an assay performance evaluation in order to determine if we can identify a hit of an assay with confidence. AV fluorescence intensities of negative control (c^-) and positive control (c^+) wells are used to compute an overall plate quality. We use three well established QC metrics: coefficient of variation (CV), Z' factor and signal to background (SB) ratio.

CV values are calculated on cell number as well as total and mean AV intensity of negative and positive control wells. Great majority of the plate CV values fell under CV < 0.15 using any segmentation method. Z' values of both total and mean AV intensities are found to be above our range of acceptance (Z' > 0.50) in our sample plate, with $Z'_{AV\ mean} = 0.59$ and $Z'_{AV\ total} = 0.66$. A high SB ratio (SB > 2) is required to evaluate the assay as screenable. SB distributions of our assays ranged above that limit as it is shown in Figure 8.

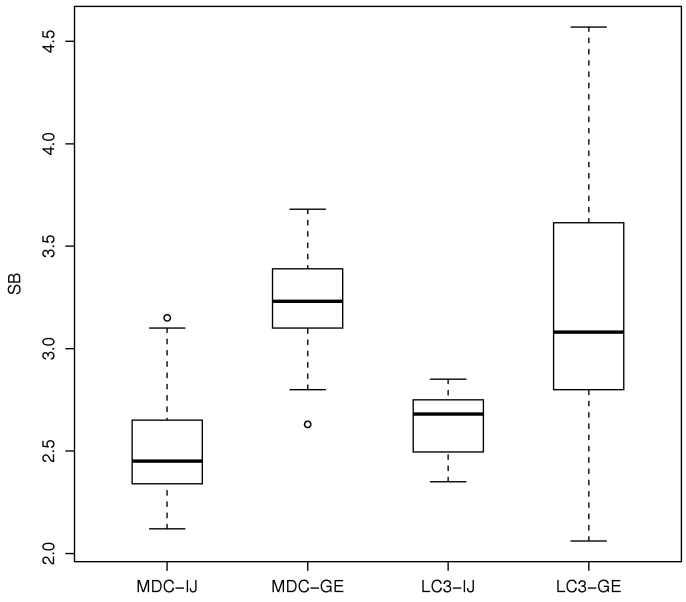


FIG. 8 Quality control boxplots of signal to background ratio (SB) representing primary (MDC) and secondary (LC3) assays, segmented by histogram-based (IJ) and region growing (GE) methods. All assays by both methods spread above the acceptable SB > 2 range. The IJ and GE datasets shows significant difference ($p\text{-value}^{MDC} = 0.0016$ and $p\text{-value}^{LC3} = 0.0015$).

Feature reduction

Throughout the feature collection phase of image processing, a set of 44 nuclear and AV features are

collected (intensity, morphology and texture-based) for each cell. The statistical correlation between these features is investigated by correlation analysis and hierarchical clustering. The results are shown on a heatmap with a dendrogram in Figure 9. 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 two 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 two additional groups of correlated features. Nuclear textural features form two additional clusters. The minor and major axes of best fitted ellipse on a nucleus are correlated with the nuclear area and Feret's diameter.

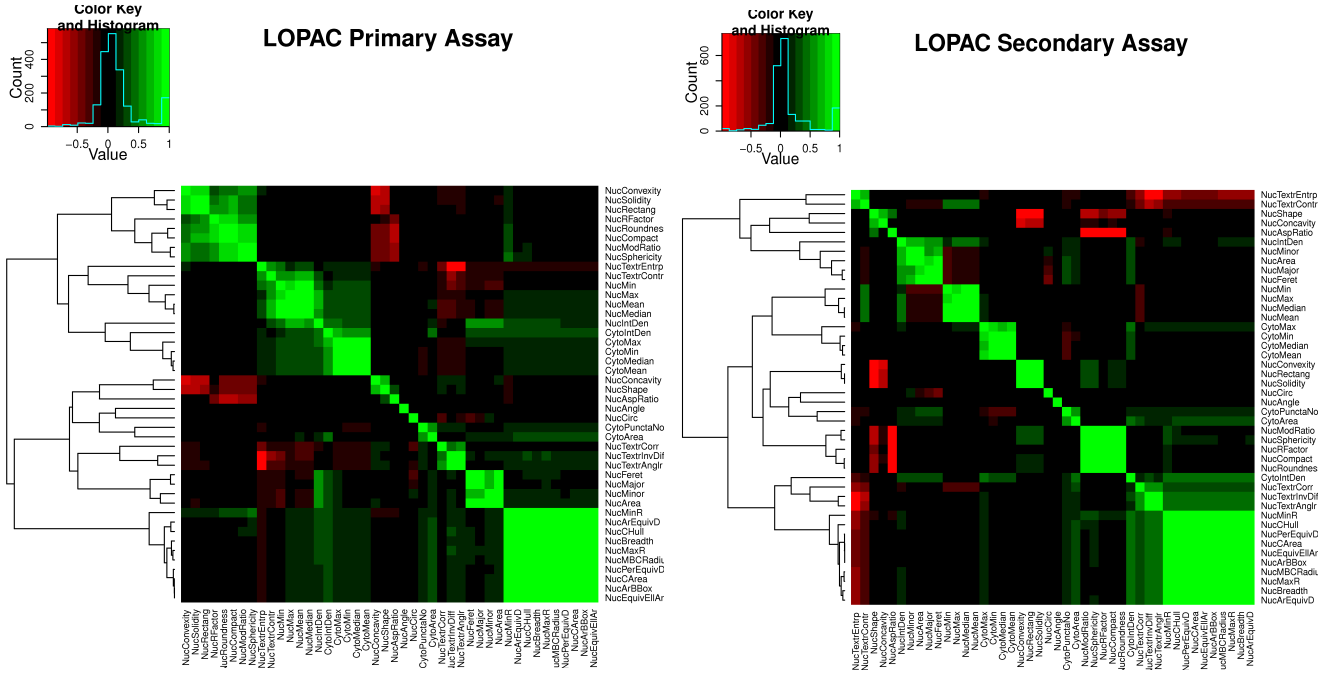


FIG. 9 Heatmap and dendrogram representation of primary screen (left) and secondary screen (right) correlation coefficients between the 44 collected features using R statistical software. Sample sizes: $N_{primary} = 1473461$ cells, $N_{secondary} = 395017$ cells.

Feature reduction is an important practical step in data mining, since irrelevant features "confuses" machine learning systems[34].

To identify redundant, irrelevant features and reduce dimensionality, we use a correlation based feature selector to determine the predictive power of each feature. Features are restricted by a correlation based feature selecting subset evaluator[30] algorithm (CfsSubsetEval) using the implementation of Weka[31] machine learning software. It 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. Best first approach is used to search the subset of features that performs greedy hill climbing with backtracing. The feature selection method above shows how many times each feature is selected during a 10-fold cross validation.

Two features were selected by the above-mentioned algorithms: 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}) 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 eight DMSO wells' CDM (= 100%). Cytotox refers to compound toxicity where the cell number of a compound is divided by the mean cell numbers of eight DMSO wells (= 100%).

The 2 selected features AV_{fold} and Cytotox are the dominant features with trivial biological interpretation, both in the primary and secondary assay. However, other features can 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 nuclear area and AV mean intensity shows a 10 (100%) entry in the primary assay only as well as the mean, median and standard deviation of AV total intensity values per well has a 100% entry value in the secondary assay. The significance of these features are not consequent in the two assays even though those are intensity and area features, biologically related to the dominant features.

Exploratory Data Analysis

The scatterplot of the two selected features (Figure 10) displays two 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 is found with smaller Cytotox value.

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

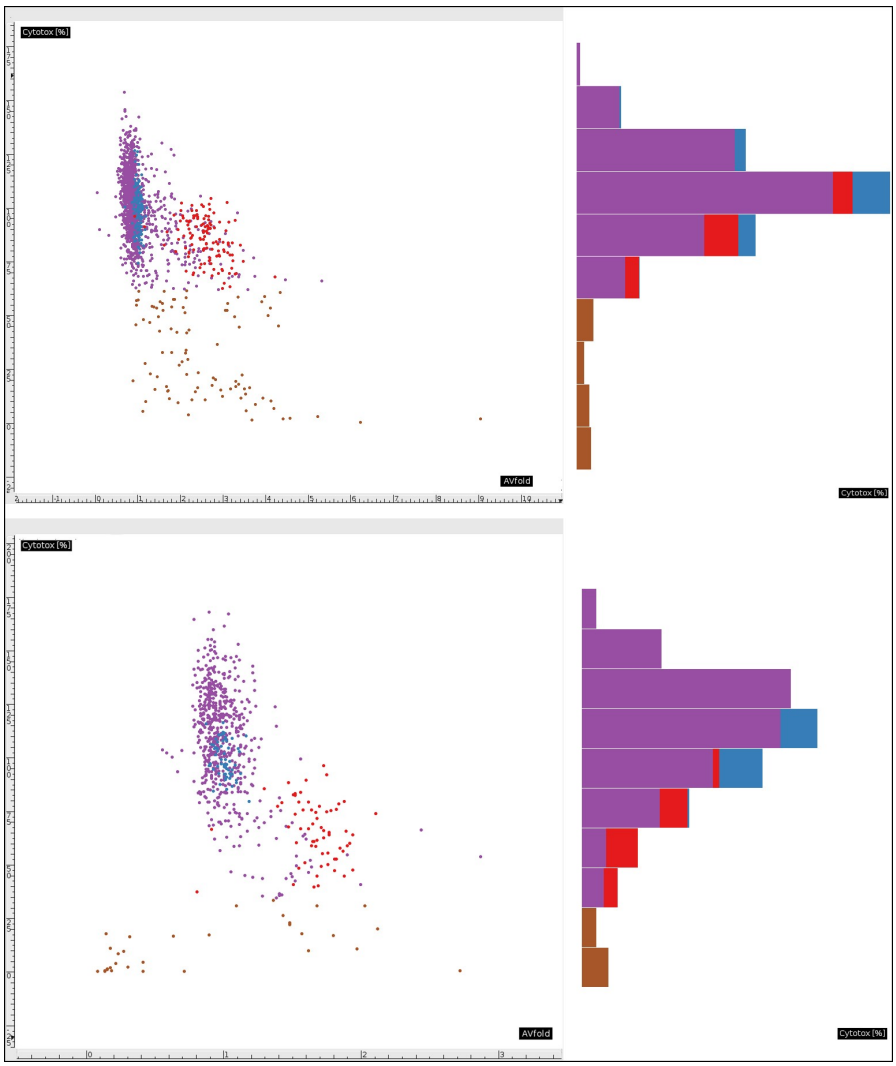


FIG. 10. Scatterplot of primary (upper left) and secondary (lower left) screen data forming two clusters of c^- (blue) and c^+ (red) by AVfold values on x axis and Cytotox [%] values on the y axis. Compound data is shown as purple dots. Compound values with Cytotox < 61.79% (primary) and 33.33% (secondary) (brown dots) are considered as outliers by the histograms (right).

Hit selection

Supervised classification method Support Vector Machine (SVM) was used for automatic hit selection because it was found to be superior in high-content cell identification[32]. We used the SVM implementation of the R-project (package e1071). The classifier was trained using the negative and positive controls. Outliers were removed from the training set in order 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 linear kernel with default settings ($C = 1$, $\gamma = 0.5$) resulted in a model (Figure 11) with total accuracy of 99.6%. The verification rate increased to 100% after removing c^+ outliers. No further SVM parameter optimization seemed necessary due to the high correct classification rate.

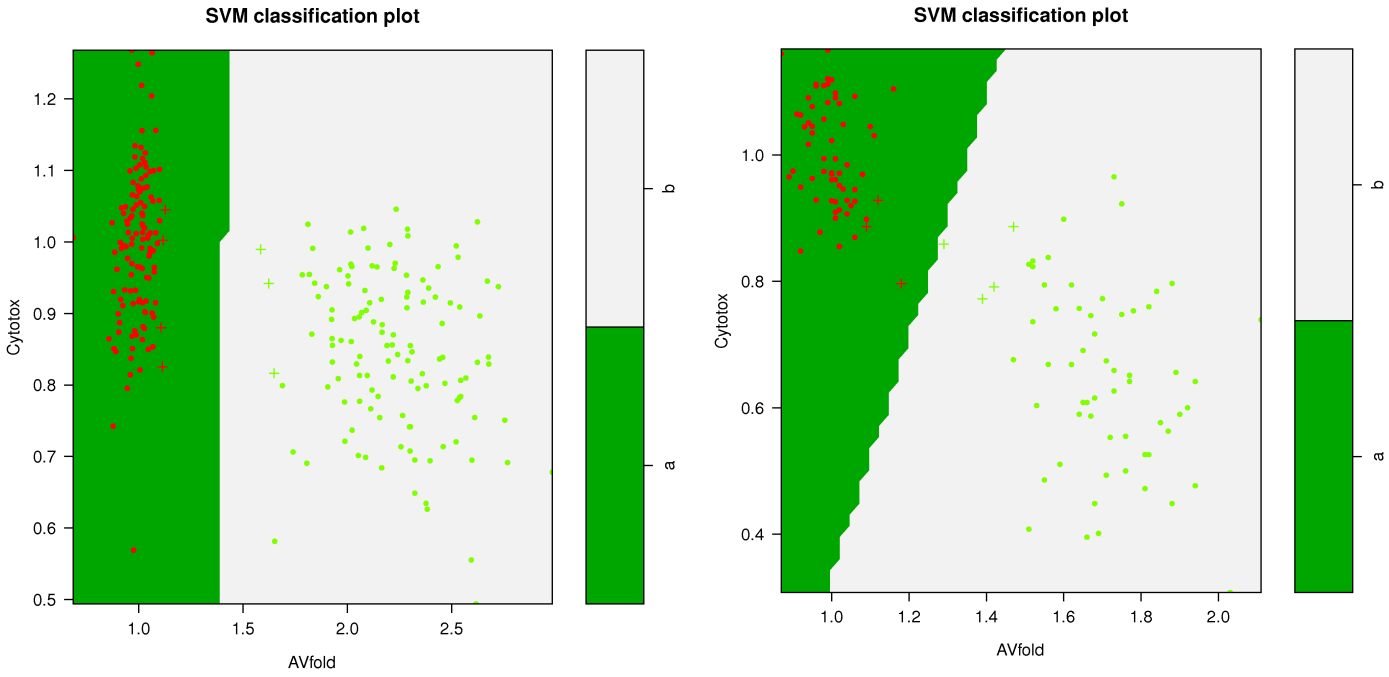


FIG. 11. Scatterplot of the primary screen (left) and secondary screen (right) models calculated by Support Vector Machine classifier (linear kernel, C = 1, γ = 0.5) with c^- (red dots in region "a") and c^+ (green dots in region "b") data as training set. Crosses indicate support vectors.

The c^+ component in the training dataset of the secondary screen shows a slight cytotoxicity by shifting the c^+ cluster towards the lower Cytotox domain. Besides the highly accurate classification, the large margins between the two 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, with thresholds $AV_{fold} > 2$ and Cytotox > 80%. The threshold-based method suggested 97 compounds with an overlap of 94 compounds (96%) of the 163 proposed by the machine learning method.

In the secondary assay 30 hits (100%) were suggested by the threshold-based method overlapping 26 hits (87%) selected by the machine-learning-based method. 4 threshold-based selected compounds (13%) were not listed among the machine-learning selected hits. The machine-learning-based method resulted 19 additional hits due to the more permissive Cytotox > 33.33% threshold. Machine-learning based primary hits contained 21 compounds (70%) out of the 30 secondary assay hits suggested by the threshold-based method.

In conclusion, we have presented an image processing - high-content analytical pipeline, that is readily applicable for a high-content cellular screen 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, however slower

speed was a limitation. The results of a region growing segmentation could be adversely affected by higher level of image noise.
The speed of the proposed imaging macro currently limits the processing to 1-3 plates by overnight, leaving a potential room of future improvement. Implementing the current macro in the form of an ImageJ plugin will give the opportunity to exploit more multicore functions.
Several 2 dimensional single-cell phenotypic features were collected, morphology-based, intensity and texture features as well. QC measures remained above our range of acceptance.
To study the correlations between features, we applied data analysis techniques such as correlation analysis and hierarchical clustering. The features chosen for the actual hit selection were the same when chosen by CFS subset evaluator and the "expert user".

Finally we applied SVM machine learning technique that recapitulated the histogram-derived image segmentation and algorithm methods run on the commercial software.
The hit selections obtained by the two methods were in very good agreement.The proposed pipeline can only be done with a robust HCS assay, where an expert user can decide on meaningful features to be extracted.

ACKNOWLEDGEMENTS
The authors thank Wee Choo Puah, Chinta Rambabu and Tiehua Du for the 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 Klionsky D J: The molecular machinery of autophagy: unanswered questions. J Cell Sci 2005;118:7-18.

5 Levine B, Klionsky D: Development by self-digestion: molecular mechanisms and biological functions of autophagy. Dev. Cell 2004;4:463-477.

6 Kundu M, Thompson CB: Autophagy: basic principles and relevance to disease. Annu Rev Pathol Mech Dis 2008;3:427-455

7 Levine B: Cell biology: Autophagy and cancer. Nature 2007;446:745-747.

8 Zhang L, Yu J, Pan H, Hu P, Hao Y, Cai W, Zhu H, Yu AD, Xie X, Ma D, Yuan J: Small molecule regulators of autophagy identified by an image-based high-throughput screen. PNAS 2007;48:19023-19028.

9 He P, Peng Z, Luo Y, Wang L, Yu P, Deng W, An Y, Shi T, Ma D: 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, Kominami E, Ohsumi Y, Yoshimori T: LC3, a mammalian homologue of yeast Apg8p, is localized in autophagosome membranes after processing. EMBO J 2000;19:5720-5728.

11 Biederbick A, Kern H F, Elsasser H P: 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(8):2476-2485.

13 Tasdemir E, Galluzzi L, Maiuri MC, Criollo A, Vitale I, Hangen E, Modjtahedi N, Kroemer G: Methods for Assessing Autophagy and Autophagic Cell Death. Methods Mol Biol 2008;445:29-76.

14 Forero MG, Pennack JA, Learte AR, Hidalgo A: DeadEasy caspase: automatic counting of Apoptotic cells in Drosophila. PLoS ONE 2009;4:e5441.

15 Abramoff MD, Magelhaes PJ, Ram SJ: Image Processing with ImageJ. Biophotonics Int 2004;7:36-42.

16 R Development Core Team: R: a language and Environment for Statistical Computing. 2009; R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http: //www.r-project.org .

17 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.

18 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.

19 Jorgensen P, Nishikawa JL, Breitkreutz B J, Tyers M: Systematic identification of pathways that couple cell growth and division in yeast. Science 2002;297:395-400.

20 Carpenter AE, Sabatini DM: Systematic genome-wide screens of gene function. Nat Rev Genet 2004;5:11-22.

21 Ridler TW, Calvard S: Picture Thresholding Using an Iterative Selection Method. IEEE Trans. Syst. Man Cybern. 1978;8:630-632.

22 Young IT, Gerbrands JJ, Vliet LJ: Fundamentals of Image Processing. Delft: TU Delft, 1995.

23 MacQueen JB: Some Methods for classification and Analysis of Multivariate Observations. Proc 5th Berkeley Symp Math Stat Probab, Univ. Calif. 1965/66, 1967;1:281-297.

24 Jain AK, Dubes RC: Algorithms for clustering data Englewood Cliffs: Prentice Hall, 1988.

25 Online article: Matteucci M: A tutorial on clustering algorithms. [Online]. Retrieved from http:// home.dei.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html, accessed on 27/04/2009.

26 Lantuejoul C, Maisonneuve F: Geodesic methods in quantitative image analysis. Pattern Recognit 1984;2:177-187.

27 Zhang TY, Suen CY: A fast parallel algorithm for thinning digital patterns. Comm ACM 1984;3:236-239.

28 Landini G: Advanced shape analysis with ImageJ. Proceedings ImageJ User and Developer Conference 2008;116-121.

29 Fawcett T: An introduction to ROC analysis. Pattern Recognit Lett 2006;27:861-874.

30 Hall MA: Correlation-based Feature Subset Selection for Machine Learning. University of Waikato: Hamilton, New Zealand, 1998.

31 Witten IH, Frank E: Data Mining : Practical Machine Learning Tools and Techniques. Morgan Kaufmann: San Francisco, USA, 2005.

32 Dürr O, Duval F, Nichols A, Lang P, Brodte A, Heyse S, Besson D: Robust Hit Identification by Quality Assurance and Multivariate Data Analysis of a High-Content, Cell-Based Assay. J Biomol Screen 2007;12:1042-1049.