

# Journal of Biomolecular Screening

<http://jbx.sagepub.com/>

---

## **Teachable, High-Content Analytics for Live-Cell, Phase Contrast Movies**

Samuel V. Alworth, Hirotada Watanabe and James S. J. Lee

*J Biomol Screen* 2010 15: 968 originally published online 16 July 2010

DOI: 10.1177/1087057110373546

The online version of this article can be found at:

<http://jbx.sagepub.com/content/15/8/968>

---

Published by:



<http://www.sagepublications.com>

**Additional services and information for *Journal of Biomolecular Screening* can be found at:**

**Email Alerts:** <http://jbx.sagepub.com/cgi/alerts>

**Subscriptions:** <http://jbx.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

# Teachable, High-Content Analytics for Live-Cell, Phase Contrast Movies

SAMUEL V. ALWORTH,<sup>1</sup> HIROTADA WATANABE,<sup>2</sup> and JAMES S. J. LEE<sup>1</sup>

CL-Quant is a new solution platform for broad, high-content, live-cell image analysis. Powered by novel machine learning technologies and teach-by-example interfaces, CL-Quant provides a platform for the rapid development and application of scalable, high-performance, and fully automated analytics for a broad range of live-cell microscopy imaging applications, including label-free phase contrast imaging. The authors used CL-Quant to teach off-the-shelf universal analytics, called standard recipes, for cell proliferation, wound healing, cell counting, and cell motility assays using phase contrast movies collected on the BioStation CT and BioStation IM platforms. Similar to application modules, standard recipes are intended to work robustly across a wide range of imaging conditions without requiring customization by the end user. The authors validated the performance of the standard recipes by comparing their performance with truth created manually, or by custom analytics optimized for each individual movie (and therefore yielding the best possible result for the image), and validated by independent review. The validation data show that the standard recipes' performance is comparable with the validated truth with low variation. The data validate that the CL-Quant standard recipes can provide robust results without customization for live-cell assays in broad cell types and laboratory settings. (*Journal of Biomolecular Screening* 2010:968-977)

**Key words:** image analysis, image recognition, machine learning, live-cell imaging, imaging and incubation systems, cell screening, cell proliferation, wound healing

## INTRODUCTION

CURRENT GENERATION HIGH-THROUGHPUT AND HIGH-CONTENT SCREENING methods have proven valuable in accelerating advances in biomedical science. A new generation of automated cell incubation and imaging systems combined with automated live-cell, high-content analysis software can provide a critical platform for quality assurance and cell screening. The ability to image and analyze the cells using phase contrast images without the introduction of fluorescent markers maintains them in their most stable and natural form. As compared with fixed-cell approaches, continuous imaging allows for more extensive and robust target identification and also the accurate identification of assay end points. It enables the analysis of how living, complex biological systems, housed

in a consistent environment and automated setting, respond to experimental treatment.

High-content, long-term, label-free assays provide a useful tool for drug discovery, toxicology, and bioproduction, as well as for stem cell and cancer research. Common applications include the quantification of cell region growth, cell counts, cell motion, wound healing, apoptosis, neurite outgrowth, angiogenesis tubule formation, and fluorescence intensity/expression levels over time. Walk-away automation is required, and high-content image analytics (analysis module) must reliably and accurately discriminate target structures (cell bodies, cell regions, etc.) without human interaction, across a range of complex imaging conditions (both biological and systematic).

Live-cell kinetic phenotypes characterizing cells undergoing dynamic interactions are highly complex. The conventional high-content analytics that rely on simple thresholding on high signal-to-noise (S/N) fluorescence staining are inadequate for measuring kinetic phenotypes. Consequently, analytics are developed by teams of highly skilled algorithm scientists for very specialized assay and imaging conditions,<sup>1-14</sup> which is a costly and time-consuming process, and the resulting analysis modules cannot be easily updated for changing assay conditions such as cell types or imaging platforms. A totally new generation of analysis technology is required.

CL-Quant (Nikon Corporation, Yokohama, Japan) is a new solution platform for broad, high-content, and live-cell image

<sup>1</sup>DRVision Technologies LLC, Bellevue, WA, USA.

<sup>2</sup>Nikon Instruments Company, Yokohama-city, Kanagawa Japan.

Received Nov 30, 2009, and in revised form Apr 26, 2010. Accepted for publication Apr 29, 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* 15(8); 2010

DOI: 10.1177/1087057110373546

analysis. Its machine learning technologies and teach-by-example interfaces make it possible for nonexperts to quickly create scalable and novel image analytics.

CL-Quant contains the following teachable core modules:

1. Image segmentation: teachable enhancement of image patterns of interest for cellular region segmentation
2. Region partition: teachable enhancement of cell boundaries for fine cellular region partition
3. Phenotyping: teachable measurement discrimination rules for object classification

The key common attributes of these machine learning-based modules are as follows:

- Intuitive learning: teach-by-example interfaces
- Fast learning: instant feedback of learning results
- Incremental learning: error correction update
- Stable learning: maintaining previous outcomes while adding new examples
- Soft outcome: output confidence values (0-255) rather than binary data for flexible use

We used CL-Quant machine learning modules to teach off-the-shelf, universal analytics called “standard recipes” for cell proliferation, cell count, wound healing, and cell motility applications, using the standard graphical user interfaces without special programming. Similar to application modules, standard recipes are intended to work robustly across a wide range of assay conditions (different cell types, magnifications, operators, etc.) without modification by the end user. However, recipes are different from traditional modules because they encode user teaching rather than image-processing programming. Thus, they are a dynamic algorithm that can be created or updated through user teaching, rather than a fixed algorithm that can be configured through parameter modification. Here we report the validation of these fully automated standard recipes.

## MATERIALS AND METHODS

### *Imaging system and image acquisition*

Live-cell, phase contrast movies were acquired for this study using the BioStation CT and IM incubation and imaging systems. The BioStation CT allows imaging experiments to be conducted without removing the cells from the incubator. Consisting of a standard-sized tissue culture incubator with an inverted microscope inside, BioStation CT holds 30 vessels ranging from 96-well plates to 75-cm<sup>2</sup> flasks, which are moved between the microscope stage and the vessel rack via a robotic device while maintaining precise levels of CO<sub>2</sub>, humidity, and temperature. BioStation CT acquires images from 2× to 40× magnification with apodized phase contrast (APC) optics and

2× to 40× magnification with fluorescent images using 3-color LED illumination.

The BioStation IM is a compact cell incubation and monitoring system that facilitates a broad array of long-term time lapse experiments, including studies of cell growth, morphology, and protein expression, by providing consistent environmental control of temperature, humidity, and gas concentration in combination with phase and fluorescence imaging of exceptional quality.

### *Cell proliferation benchmark images*

Forty phase contrast movies of cells growing in culture were acquired on the BioStation CT of 7 cell types (CHO, Cos-7, HEK193, HeLa, INS1, NIH3T3, PC12), imaged every 30 min for 24 to 48 h. The movies were acquired at 2×, 4×, 10×, and 20× magnifications (10 movies per magnification). For validation, 5 image frames per movie were stratified sampled to create a benchmark data set of 200 images. The samples were selected systematically at the frame in the second time point, at the 25th percentile time point, 50th percentile (middle) time point, 75th percentile time point, and the second to last frame.

### *Cell count benchmark images*

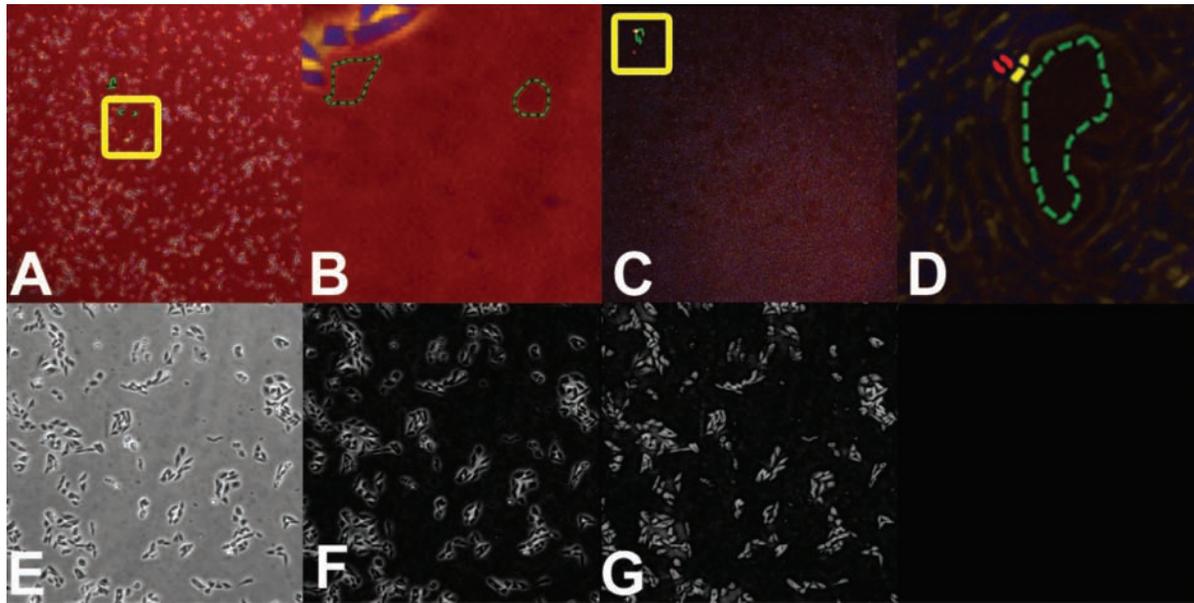
A subset of cell proliferation benchmark images where individual cells are distinguishable was used for the cell count benchmark data set. We found that individual cells could not be reliably identified by eye in the PC12 and INS1 movies even after careful image examination, so we removed these from the benchmark data set as no truth (discussed below) could be reliably created. Then, 10× images from the 25th percentile time point were chosen (40 images), and the resulting benchmark data set includes 31 images at 10× magnification.

### *Wound-healing benchmark images*

Six phase contrast movies showing wound closure (cells migrating into a scratch in the monolayer culture) of 3 different cell types (MDCK, keratinocytes, and fibroblasts) were acquired on the BioStation IM and CT at varying sampling rates and length. Five frames per movie were stratified and sampled as described above to create a benchmark data set of 30 images.

### *Cell motility benchmark images*

Thirty-nine phase contrast movies of 7 cell types (HeLa, NIH3T3, CHO-K1, Cos-7, HEK293, PC12, INS1) were acquired on the BioStation IM. Images were acquired once per minute. Sequences from the beginning of the movie (minimum 100 frames) where cells were relatively isolated were used for tracking validation.



**FIG. 1.** Segmentation teaching for 2× and 4× magnifications. (A, C) We performed segmentation teaching on derived color images combining the original image and pattern channels. The color image shows the original image in the red channel, bright pattern image in the green channel, and dark pattern image in the blue channel. (B, D) Close-up of teaching regions of interest (ROIs) shown in the yellow rectangles in (A) and (C). Green ROIs indicate image patterns for enhancement, red ROIs for suppression, and yellow ROIs for background patterns. (E) Original phase contrast, example input image (Chinese hamster ovary [CHO] cells at 2×). (F) Bright pattern image. (G) Dark pattern image.

### Truth creation strategy

Standard recipe performance is validated through comparison with truth. Typically, performance is validated through comparison with manually created truth.<sup>15,16</sup> We used this approach for cell counting and cell motility validation. For cell proliferation and wound-healing studies, we created optimal results by developing a custom recipe for each test movie and used this as the truth.

Individual cells were manually counted to provide truth for benchmark image sets. One hundred percent sampling of the cells in the image was done for cell counting. True cell counts were calculated for each data set. For cell motility, true cell tracks—linked (x, y, t) vertices for cells in the movies—were created manually. Cells that did not divide or leave the field of view were chosen (100 cells in total). The truth was reviewed and any discrepancies resolved.

### Automated quantification

All of the standard recipes perform image background detection followed by mask refinement. A segmentation recipe for image background detection is taught by drawing. To facilitate more robust analysis, multiple pattern channels enhancing image content were created, and background detection was taught using a combined image that encodes the pattern

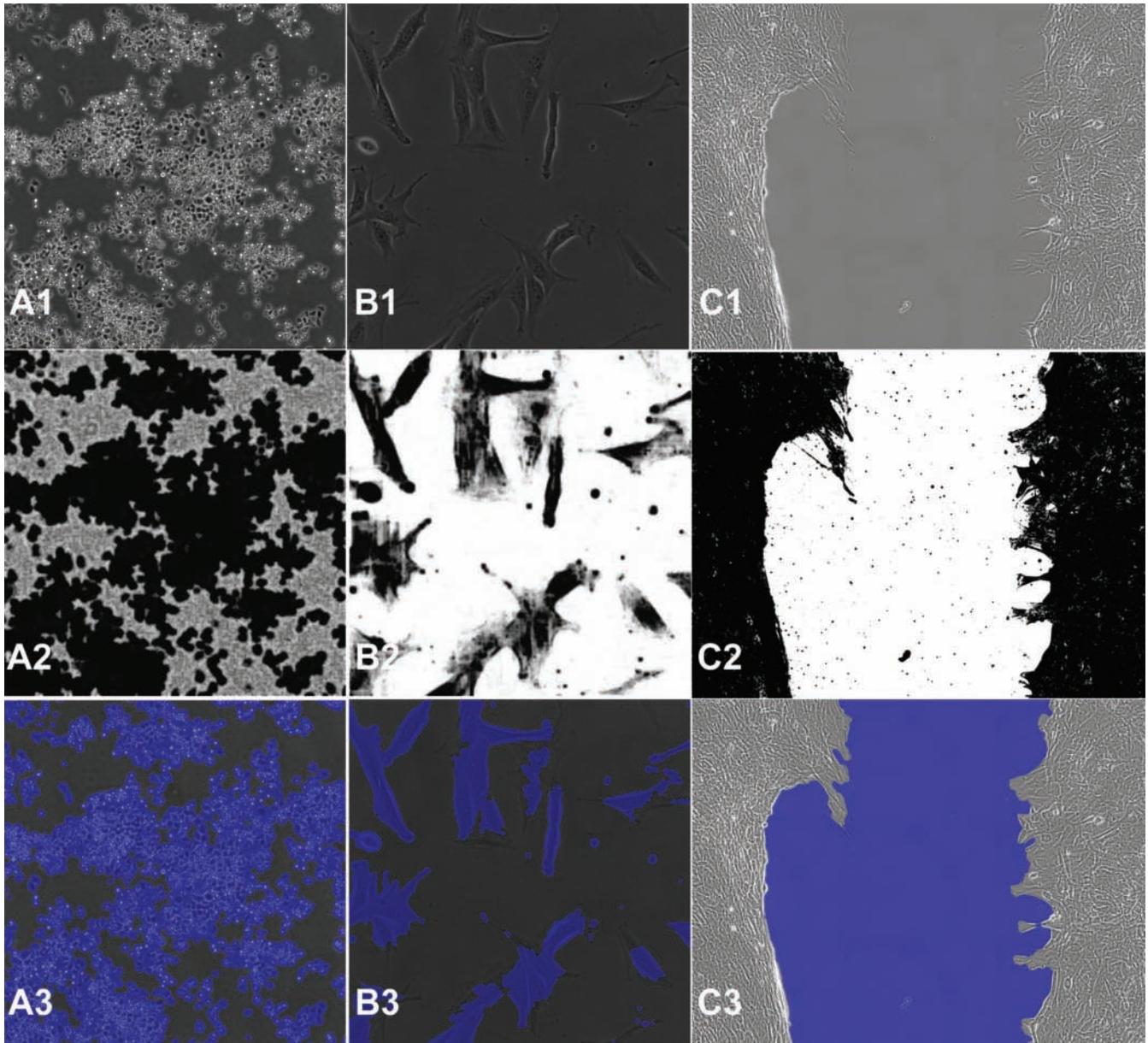
information along with the original image. A core set of pattern channel-based background segmentation recipes is used for all the applications, followed by application-specific refinements as described below. The cell-counting recipe uses a teachable machine learning technology for mask partition, also described below.

### Pattern channels

CL-Quant provides pattern channel functions that can be taught interactively to generate enhanced images isolating certain types of image patterns such as bright, dark, and texture. In CL-Quant, operators add pattern channels for use in combined channel segmentation teaching (**Fig. 1E-G**). The additional patterns improve the accuracy and robustness of segmentation by emphasizing relevant patterns for the teaching step.

### Segmentation teaching

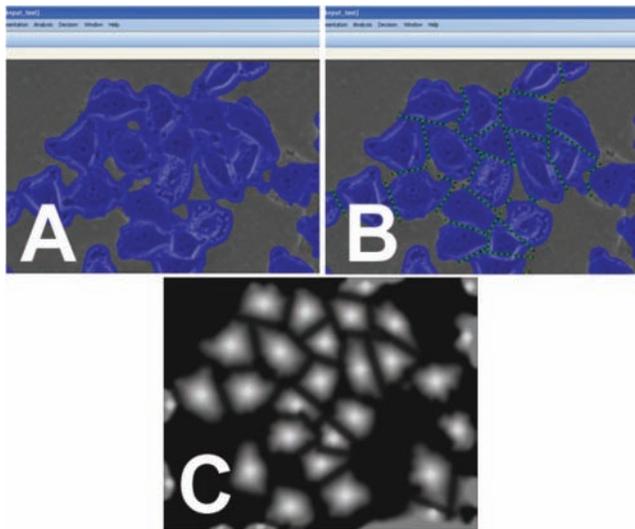
Users can flexibly teach segmentation recipes using a teach-by-example interface and CL-Quant's machine learning technology for image segmentation.<sup>17</sup> **Figure 1A,C** shows 2 of the teaching images used to create the segmentation recipe for 2× and 4× images. The zoomed-in teaching regions (yellow boxes) are shown in **Figure 1B,D**. The user provides the software examples by drawing multiple regions of interest



**FIG. 2.** Original and confidence images, as well as segmentation masks for the cell proliferation and wound-healing standard recipes. By column: (A) example 4× cell proliferation image (HeLa), (B) example 20× cell proliferation image (NIH3T3), and (C) example wound-healing image. By row: (1) original images, (2) confidence images, and (3) final detection masks shown overlain on the original image. Standard recipes must robustly transform a broad range of input images without requiring user modification.

(ROIs). Green ROIs specify patterns for enhancement, red ROIs specify patterns for suppression, and yellow ROIs specify background patterns. The user input creates a soft or fuzzy matching model<sup>17</sup> where input image pixels are compared with the user teaching, and their likelihood for enhancement set membership is encoded into an output 8-bit

image called a confidence image (**Fig. 2**, row 2). **Figure 2** illustrates applying the recipe to new images in an execution mode. Input images (row 1) are transformed to create confidence images (row 2), which are thresholded to create the segmentation masks (blue overlays in row 3), which are subsequently refined and quantified. Of course, in execution



**FIG. 3.** Illustration of cell partition teaching. (A) Input detection mask created by the segmentation recipe. (B) Green lines drawn by user to teach the desired partitioning of the input mask. The user-defined cut regions are used to generate teaching data for the machine learning engine. (C) Membership image generated by the machine learning engine given the user training. It is used to partition the input detection mask into discrete single-cell components.

mode, no user teaching is necessary, and the analysis is fully automated.

CL-Quant makes it possible to transform an input image to an enhanced image that can be easily segmented using a drawing interface, replacing traditional methods of image processing (i.e., programming or applying image-processing functions). This is the most challenging step in image analysis. It can be taught using multiple images and is updateable at a later time with additional images. It creates a model that is encoded into a segmentation recipe that can be automatically applied to transform and segment new images.

### Partition teaching

Cell region partitioning is the process of identifying individual cells in segmented regions. Cell region partition is an extremely difficult problem in computing because there are too many variables and uncertain factors for a simple algorithm. In CL-Quant, users can flexibly teach a partition recipe using a drawing interface and a teachable region boundary enhancement technology.

Figure 3A shows an input mask covering the cell region. In Figure 3B, the user draws the cutting lines to partition the regions into individual cells for cell count and various cell measurements. On the basis of the input mask and user cutting, CL-Quant obtains individual cell regions that it uses to generate

teaching data. The machine learning engine creates a membership image (Fig. 3C) based on the individual cell regions. This membership image has high values in the center of cells and low values in cell boundaries.

The machine learning engine finds the optimal relationship between the membership image and the feature images that are derived from multiple filtering of the original image in a multi-resolution manner, resulting in the recipe that is the set of the best-fitting coefficients. Then the recipe can be applied to another (similar) image for cell partitioning. If the cutting result is not satisfactory, the user can perform additional teaching in an error correction mode. The created recipe can be applied to new images where features are combined into the membership images using the fitting coefficients (recipe), and cell partitioning is performed based on multiple thresholding levels applied to the membership image.

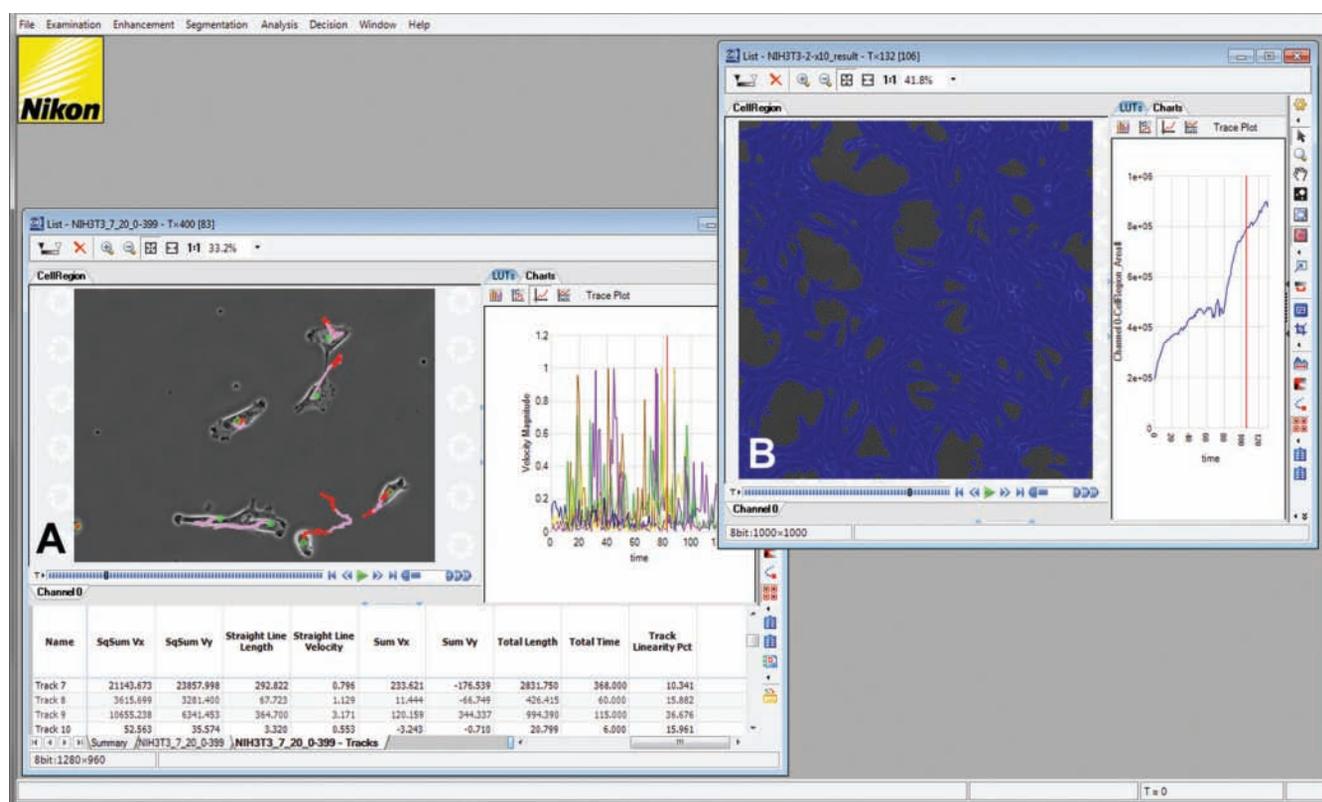
### Cell proliferation recipe

Two standard recipes were created: one for 2 $\times$  and 4 $\times$  magnifications and one for 10 $\times$  and 20 $\times$ . Both recipes generate pattern channels and were taught on the combined pseudo-color images where the original and pattern channels are mapped to R,G,B (see Fig. 1). Two images from the Chinese hamster ovary (CHO) 2 $\times$  image set were used to teach the 2 $\times$ /4 $\times$  segmentation recipe, and 3 images from the CHO 20 $\times$  image set were used to teach the 10 $\times$ /20 $\times$  segmentation recipe. In both cases, the image background is taught to generate a transformed confidence image where cell regions are dark and the background is bright (Fig. 2, row 2). The confidence image is thresholded for dark intensities to mask, or segment, the cellular regions. The cell region detection mask is then refined by filling small holes and removing (gating) mask regions based on their morphological and intensity characteristics. The area and confluency (mask area divided by image area) of the final detection mask is measured.

The cell proliferation recipe can be applied to any phase contrast movie to quantitatively measure the rate of cell growth in terms of cell area and confluency (ratio of the cell region area to the image area). Figure 4B shows a representative movie and associated growth curve in the CL-Quant software. The movie can be advanced using the controls just below the displayed image, and the trace plot shows the cell region area measurement over time.

### Cell counting recipe (10 $\times$ )

The cell-counting recipe uses the same 10 $\times$ /20 $\times$  segmentation recipe as the cell proliferation recipe above. Cell partition teaching was done initially using an image from the Cos-7 cell line and updated with an image of the HeLa cell type. The cell partition recipe divides a cell region mask into a number of



**FIG. 4.** CL-Quant screenshot showing processing results for the (A) cell motility and (B) cell proliferation  $10\times$  and  $20\times$  standard recipes. The track overlays (pink and red) show the location of the center of the detection mask in previous time points. Red points in the overlay indicate periods where the cell was moving in isolation, and pink points indicate periods where the cell was in contact with other cells. The trace plot shows velocity measures over time for the cells shown here. The spreadsheet displays motility metrics (i.e., straight-line distance, total distance) at the current timeframe. The timeframe can be advanced using the movie controls just below the image. The cell proliferation processing results (C) include cell region detection mask and associated measurements (i.e., cell region area, confluency). Here the detection mask is shown overlain, and the cell region area measurement over time is displayed in the trace plot.

individual regions corresponding to individual cells (Suppl. Fig. S1); cell counts and individual cell morphology and intensity measures are calculated.

#### Wound-healing recipe

The wound-healing standard recipe uses the same  $2\times/4\times$  segmentation recipe as the cell proliferation recipe above, but its threshold settings are adjusted to segment the “wound” or substrate area rather than the cell regions. The detection mask is refined by filling holes and removing small detection regions. Because the application is intended to measure the shrinkage of the wound, the wound is restricted to be less than its original size by conditioning later frames with earlier frames, thereby ensuring background regions that may appear outside the wound area are ignored. The wound area is measured.

#### Cell motility recipe

The cell motility standard recipe uses the same  $10\times/20\times$  segmentation recipe as the cell proliferation standard recipe. After mask refinement, CL-Quant’s motility tracking recipe is applied. The tracking module can be simply configured using common parameters (max velocity, size, etc.) and performs mask-based tracking. It will maintain cell separation over time when cells come together. One advantage of using masks for tracking is that detailed morphological characterization can be done over time, which can be a sensitive metric for comparing different experimental conditions.<sup>18</sup>

The motility recipe can be used to extract a wide array of single-cell tracking metrics, including single-cell descriptors of movement (i.e., displacement, velocity acceleration, etc.), morphology (i.e., compactness, length-width ratio, etc.), and intensity

(i.e., mean, total, standard deviation of intensity in the cell). Moving averages of these measures can be calculated. In addition, measurements and model parameters that describe the entire track (i.e., total length, diffusion constant, persistent, etc.) are provided. **Figure 4A** shows a screenshot of a movie after processing in CL-Quant. The trace plot data show individual cell velocities for the 7 cells (2 are touching) shown in the image over time. The spreadsheet below shows motility measurements for the cells at the current time point.

### Test metrics: cell proliferation and wound healing

Standard recipe result masks and truth masks are compared using the following test metrics for sample frame  $t$ :

- CPoint ( $t$ ): total truth mask area normalized by the image size averaged over all benchmark images
- $\epsilon_p(t)$ : oversegmentation area averaged over all benchmark images evaluated relative to truth mask
- $\epsilon_m(t)$ : undersegmentation area averaged over all benchmark images evaluated relative to truth mask
- cP+ ( $t$ ): expresses oversegmentation as the positive error bar extending from the truth mask mean = CPoint ( $t$ ) +  $\epsilon_p(t)$
- cP- ( $t$ ): expresses undersegmentation as the negative error bar extending from the truth mask mean = CPoint ( $t$ ) -  $\epsilon_m(t)$

Because the exact cell border is difficult to recognize and  $\pm 1$  pixel detection discrepancy around the lengthy cell region border can accumulate to large error, we therefore excluded a 4-pixel region around the truth mask when calculating under- and oversegmentation metrics.

The test metrics are defined as follows:

$$\text{CPoint}(t) = \frac{\sum_{i \in D} \text{area}_{\text{truth}}(i, t)}{\text{ImageSize} * D} \quad (1)$$

$$\epsilon_p(t) = \sqrt{\frac{\sum_{i \in D} (\epsilon^+(i, t) / \text{ImageSize})^2}{D}} \quad (2)$$

$$\epsilon_m(t) = \sqrt{\frac{\sum_{i \in D} (\epsilon^-(i, t) / \text{ImageSize})^2}{D}} \quad (3)$$

where

- $D$  = number of data sets
- ImageSize = image area (must all be the same)
- $i$  = data set index
- $t$  = sample or time index
- $\text{area}_{\text{truth}}(i, t)$  = area of truth mask at time point  $t$  and data set  $i$
- $\epsilon^+(i, t)$  = area of oversegmentation of detection mask at time sample  $t$ , data set  $i$ . Physical meaning = area of ( $\text{Detection}_{\text{mask}} - \text{Truth}_{\text{mask}}$ )

$\epsilon^-(i, t)$  = area of undersegmentation of detection mask at time sample  $t$ , data set  $i$ . Physical meaning = area of ( $\text{Truth}_{\text{mask}} - \text{Detection}_{\text{mask}}$ )

### Test metrics: cell count

Standard recipe cell counts are compared with manual cell counts for each image. The absolute value of the difference between the two is divided by the manual counts to obtain the percent error of the standard recipe on that image. The mean error over all images is reported.

### Test metrics: cell motility

Standard recipe tracks and manual tracks are compared using the following metrics:

- Average object tracking error: number of incorrect tracking time points over the entire track divided by the total number of time points in the track
- Average matching tracking sensitivity: for each truth track, the number of objects in the detected track that has  $\geq 10\%$  overlay with the truth track, divided by all objects in the truth track

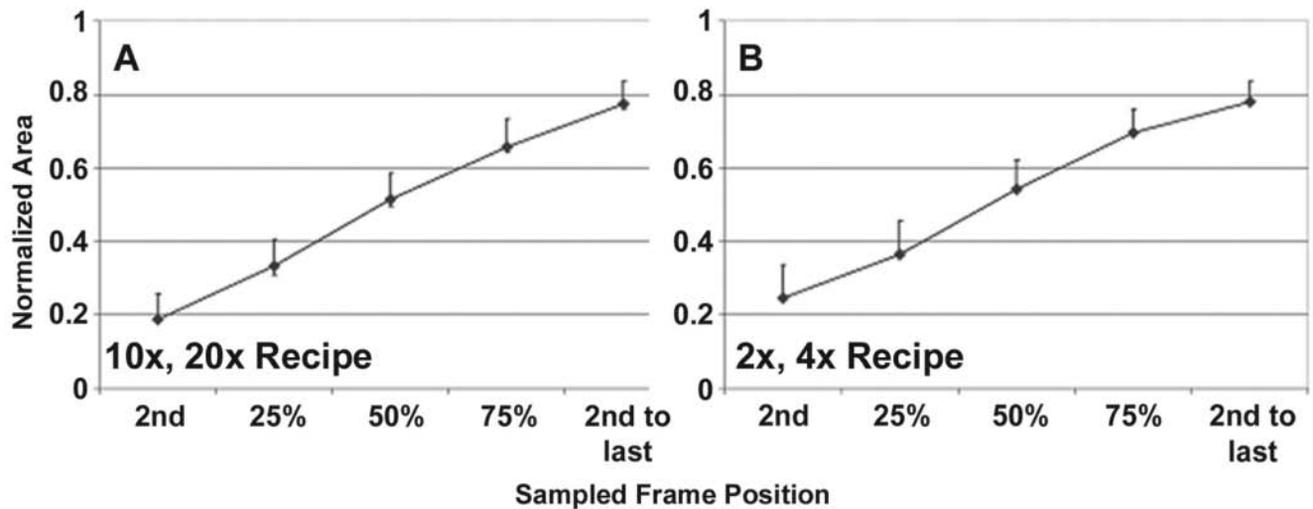
## RESULTS

### Cell proliferation and wound-healing results

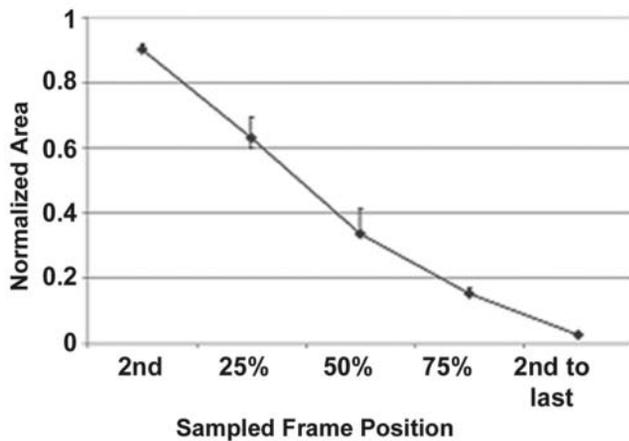
The 3 standard recipes for cell proliferation (2 recipes) and wound healing (1 recipe) prove to be robust and accurate across a large range of cell lines and movies acquired by different operators at different times (**Figs. 5-6**). The standard recipes are closely aligned with custom recipes, proving that a single recipe can be created to work well across a broad range of cell types. The standard recipes have a tendency to systematically oversegment, but the degree of oversegmentation is stable at different levels of confluency. Thus, the standard recipe will be able to detect the response of the biological system to different experimental conditions equally well as the custom recipes.

### Cell counting and motility

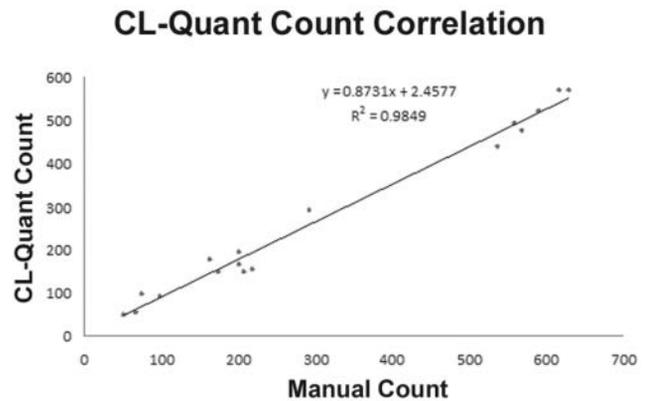
The cell-counting and motility standard recipes are likewise robust and accurate across several cell types and a sizable set of validation images and movies (**Fig. 7** and **Suppl. Fig. S1**). Cell-counting scoring correlates highly with manual analysis ( $r = 0.985$ ) and has a high degree of accuracy; the mean error over all images is only 11.9%. Note that human counters make errors as well, and a difference of 16% between human counters has been reported,<sup>15</sup> so this amount of error is very reasonable. The cell motility performance is very good with 100% average matching tracking sensitivity and 1.63% average object tracking error.



**FIG. 5.** Cell proliferation validation data are shown for the (A) 10 $\times$ , 20 $\times$  recipe and the (B) 2 $\times$ , 4 $\times$  recipe. The curves show the segmentation accuracy at the sampled time points ( $n = 20$  images at each time point covering 7 cell lines). The upper error bar, the  $cP^+$  ( $t$ ) value, indicates the amount of oversegmentation as compared with the truth, and the lower error bar, the  $cP^-$  ( $t$ ) value, indicates the amount of undersegmentation relative to the truth. The actual data are shown in Supplemental Tables S1 and S2.



**FIG. 6.** Wound-healing recipe validation data. The curves show the segmentation accuracy at the sampled time points ( $n = 6$  images at each time point covering 3 cell lines). The upper error bar, the  $cP^+$  ( $t$ ) value, indicates the amount of oversegmentation as compared with the truth, and the lower error bar, the  $cP^-$  ( $t$ ) value, indicates the amount of undersegmentation relative to the truth. The actual data are shown in Supplemental Table 3.



**FIG. 7.** Cell count results. Manual and machine scoring of cell counts show close correlation ( $r^2 = 0.9849$ ). The mean error over all images is 11.9%, which is within the range you would expect to see between human counters. Forty images covering 5 cell lines were used in the study.

## DISCUSSION

Screening systems require walk-away automation, and therefore validation of the automated image analysis is necessary. Standard recipes are meant to be used off-the-shelf

without modification by end users, and therefore they must be robust for a given application over a variety of cell lines and operators.

Here we have compared the performance of 2 cell proliferation standard recipes, one for 2× and 4× magnifications and one for 10× and 20× magnifications, with that of multiple recipes developed for each cell type and magnification combination. Similarly, a single wound-healing standard recipe is compared with multiple custom wound-healing recipes. The results show that the standard recipe performance is very close in accuracy to the custom recipes at all time points as the cells undergo proliferation and migration. Generally, the standard recipes tend to slightly oversegment relative to the custom recipe, and this is a necessary accuracy trade-off to achieve robustness across the many cell lines and conditions. However, the degree of oversegmentation is stable at the different confluency levels, and therefore the standard recipes are equally sensitive to detected changes in migration or proliferation as the custom recipes.

In this study, we have also compared cell-counting and cell motility standard recipe performance with manual truth. The goal is to show good correlation with manual truth, even though manual analysis is also not perfect. The results show good correlation and accuracy as compared with truth for the cell-counting and cell motility recipes.

### Teachable analytics

Teachable analytics address a critical issue in the high-content analysis (HCA) industry—how to develop novel analyses that are robust and scalable with less cost and risk. For end users, teachability allows the configuration of accurate analyses for their specific research. Indeed, in some cases, it could provide a software tool that can be tailored to address novel applications that were not even conceived of at the time the software was sold. For industry, it provides a way to create a lineup of application analyses without needing a team of costly image analysis and pattern recognition specialists.

CL-Quant's learning technologies provide nearly instant results in the teaching mode; recipes can literally be developed in minutes. However, it can take more time to review the recipe performance on multiple images and then update the recipe with additional teaching in an error correction mode. The recipes validated in this study were taught by our junior technical staff, and we estimate that the process of recipe creation, performance review, and teaching update took less than 40 working hours each.

Much of the discussion of supervised machine learning technologies for use in HCA has centered on its use for phenotyping<sup>19-21</sup> after the images have been segmented using traditional image analysis methods. In addition to phenotyping, CL-Quant can be taught for image segmentation, which

is arguably the most fundamental step in HCA. CL-Quant provides teachable interfaces to supervised machine learning technologies providing segmentation, object partition, and object phenotyping (classification) capabilities. As compared with unsupervised learning methodologies, teaching in CL-Quant is much faster, providing instant feedback of the teaching results, and only the desired data are generated. For example, if you want to segment punctate cells, only cells that have a punctate appearance will be detected. In contrast, unsupervised learning approaches will detect and cluster all foreground patterns (i.e., punctate and flat cells) in the image. These would then need to be filtered using a gating or phenotype classification step, which increases processing and teaching time.

### CONCLUSION

CL-Quant's machine learning technologies are unique in the field. CL-Quant's flexible segmentation learning technologies enable fundamental image analytics for broad applications to be quickly developed with less time and cost. In this study, we prove that these teachable analytics can be used to provide robust and accurate off-the-shelf performance for target applications. We are leveraging this strength of CL-Quant to quickly develop and validate a suite of new application recipes.

### ACKNOWLEDGMENT

This research was supported in part by grant number 6R44MH075498 from the National Institute of Mental Health.

### REFERENCES

1. Pescini Gobert R, Joubert L, Curchod ML, Salvat C, Foucault I, Jorand-Lebrun C, et al: Convergent functional genomics of oligodendrocyte differentiation identifies multiple autoinhibitory signaling circuits. *Mol Cell Biol* 2009;29:1538-1553.
2. Mutka SC, Yang WQ, Dong SD, Ward SL, Craig DA, Timmermans PB, et al: Identification of nuclear export inhibitors with potent anticancer activity in vivo. *Cancer Res* 2009;69:510-517.
3. Lee S, Lee HG, Kang SH: Real-time observations of intracellular Mg<sup>2+</sup> signaling and waves in a single living ventricular myocyte cell. *Anal Chem* 2009;81:538-542.
4. Chan EY, Longatti A, McKnight NC, Tooze SA: Kinase-inactivated ULK proteins inhibit autophagy via their conserved C-terminal domains using an Atg13-independent mechanism. *Mol Cell Biol* 2009;29:157-171.
5. Wu HL, Li YH, Lin YH, Wang R, Li YB, Tie L, et al: Salvianolic acid B protects human endothelial cells from oxidative stress damage: a possible protective role of glucose-regulated protein 78 induction. *Mol Cell Biol* 2009;29:157-171.
6. Okawa Y, Hideshima T, Steed P, Vallet S, Hall S, Huang K, et al: SNX-2112, a selective Hsp90 inhibitor, potently inhibits tumor cell growth, angiogenesis,

- and osteoclastogenesis in multiple myeloma and other hematologic tumors by abrogating signaling via Akt and ERK. *Blood* 2009;113:846-855.
7. Tierno MB, Kitchens CA, Petrik B, Graham TH, Wipf P, Xu F, et al: Microtubule binding and disruption and induction of premature senescence by disorazole C1. *J Pharmacol Exp Ther* 2008;328:715-722.
  8. Muniz-Medina V, Jones S, Maglich J, Galardi C, Hollingsworth R, Kazmierski W, et al: The relative activity of "function sparing" HIV-1 entry inhibitors on viral entry and ccr5 internalization: is allosteric functional selectivity a valuable therapeutic property? *Mol Pharmacol* 2008;75:490-501.
  9. Moffat J, Grueneberg DA, Yang X, Kim SA, Kloepper AM, Hinkle G, et al: A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* 2006;124:1283-1298.
  10. Perlman ZE, Slack MD, Feng Y, Mitchison TJ, Wu LF, Altschuler SJ: Multi-dimensional drug profiling by automated microscopy. *Science* 2004;306:1194-1198.
  11. Zhou X, Cao X, Perlman Z, Wong ST: A computerized cellular imaging system for high content analysis in Monastrol suppressor screens. *J Biomed Inform* 2006;39:115-125.
  12. Lindblad J, Wahlby C, Bengtsson E, Zaltsman A: Image analysis for automatic segmentation of cytoplasm and classification of Rac1 activation. *Cytometry A* 2004;57:22-23.
  13. Garippa RJ: A multi-faceted approach to the advancement of cell-based drug discovery. *Drug Discov World* 2004;6:43-55.
  14. Harada JN, Bower KE, Orth AP, Callaway S, Nelson CG, Laris C, et al: Identification of novel mammalian growth regulatory factors by genome-scale quantitative image analysis. *Genome Res* 2005;15:1136-1144.
  15. Ramm P, Alexandrov Y, Cholewinski A, Cybuch Y, Nadon R, Soltys BJ: Automated screening of neurite outgrowth. *J Biomol Screen* 2002;8:7-18.
  16. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, et al: CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* 2006;7:R100.
  17. Lee JSJ, Oh S: Learnable object segmentation. US Patent Number 7,203,360, April 10, 2007.
  18. Oh S, Alworth SV, Kandere-Grzybowska K, Borisy GG, Grzybowski BA, Lee JSJ: Automated kinetic analysis in individual cell motility assays [poster]. Paper presented at the 46th annual meeting of the American Society for Cell Biology, December 2006, San Diego, CA.
  19. Huh S, Lee D, Murphy RF: Efficient framework for automated classification of subcellular patterns in budding yeast. *Cytometry A* 2009;75:934-940.
  20. Solyts B, Alexandrov Y, Remezov D, Swiatek M, Dagenais L, Murphy S, et al: Learning algorithms applied to cell subpopulation analysis in high content screening [poster]. Paper presented at the 10th annual conference of the Society for Biomolecular Sciences, September 2004, Orlando, FL.
  21. Jones TR, Carpenter AE, Lamprecht MR, Moffat J, Silver SJ, Grenier JK, et al: Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. *Proc Natl Acad Sci USA* 2009;106:1826-1831.

Address correspondence to:

Samuel V. Alworth

DRVision Technologies LLC

15921 NE 8th St., Suite 200, Bellevue, WA 98008

E-mail: sama@drvtechnologies.com