

# HIGH THROUGHPUT MICROSCOPY: FROM IMAGES TO DATA

Jeffrey H. Price

Burnham Institute for Medical Research, La Jolla CA 92037

Vala Sciences Inc., La Jolla CA 92037

## ABSTRACT

For academic scientists, high throughput microscopy enables a paradigm shift from labor intensive qualitative microscopy to fast image-based quantitative biology. For pharmaceutical scientists, it enables replacing whole-well averaged readouts of simple cellular biomarkers with multiplexed cellular systems biological screens that include cell subpopulation and morphological characteristics. For research pathologists, the future looks bright for automating the process of reading and scoring tissue microarrays, which consist of a few hundred to a few thousand tissue sections on each slide. For image-based (or high content) screening, various algorithms have been developed to perform the automated cytometry needed to assess compound hits and/or dose responses. For many other drug screening and academic biological assays and for tissue microarrays, substantial challenges remain to be solved by more advanced image analysis techniques. For all of these applications, raw image data can comprise tens to hundreds of gigabytes (GB) per experiment and new frameworks enabling rapid algorithm development and validation are needed.

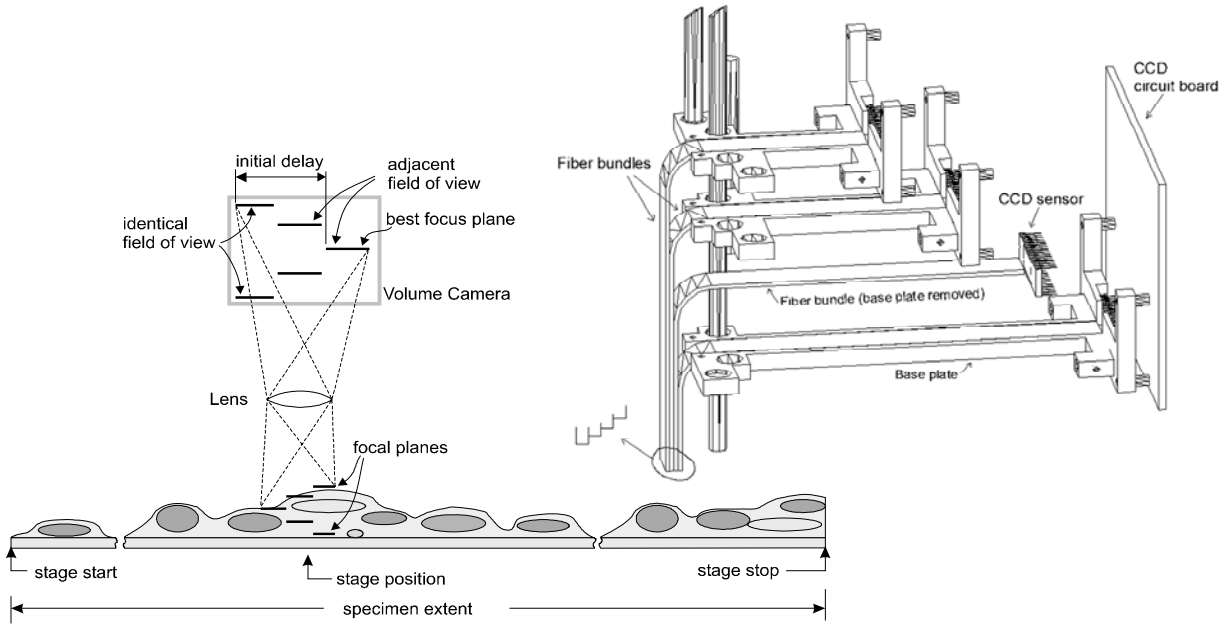
High throughput microscopes (HTM) are designed to scan slides, multi-well plates and culture chambers rapidly and automatically to image large areas. Typical slides with cells smeared or cultured on the glass surface can have 10,000 to 5 million cells [1]. Tissue microarray (TMA) slides, which are sectioned from paraffin blocks containing an array of 0.5-2.0 mm diameter by 1-2 cm long tissue biopsy cores, can have up to a few thousand tissue sections on each slide [2-4]. Image based drug screens [5,6] are carried out in multi-well plates (including 96-, 384-, 1536-, and 3456-well formats) and small screens may include tens of thousands of wells and screens can be as large as millions of wells. Instead of screening for active compounds or drugs, genetic screens are also carried out using gene activation or inhibition techniques to discover new molecular targets and pathways responsible for important cellular functions [7]. The raw image data sets for these applications can vary from a few hundred megabytes (MB) to several terabytes (TB). While with TMAs the data is still largely collected manually by pathologists, in screening applications that may comprise images of tens millions of cells it is virtually impossible for any human to view the entire image set. Thus, while some applications would benefit greatly by labor-saving automated image analysis, others require it.

The primary enabling component of analyzing large biological microscopy data sets is the HTM instrument itself. The commercially available instruments include: the Fisher-Cellomics (Pittsburgh, PA) ArrayScan and KineticScan high content screening (HCS) instruments; the Applied Precision (Issaquah, WA) cellWoRx for HCS; the GE-Amersham (Piscataway, NJ) In Cell 1000 and In Cell 3000 for HCS; the BD Biosciences-Atto

(Rockville, MD) CARV II for HCS; the Molecular Devices/Universal Imaging/Axon (Sunnyvale, CA) ImageExpress Micro, ImageXpress 500A and Discovery1 for HCS; the Evotec Technologies (Woburn, MA) Opera™ and Insight 3D Cell for HCS; the PerkinElmer (Wellesley, MA) CellLux for HCS, the Aperio (Vista, CA) ScanScope systems for TMAs; the Clarent (San Juan Capistrano, CA) ACIS for TMAs; the Applied Imaging (San Jose, CA) CytoVision for cytogenetics and Ariol for TMAs; and Bacus Laboratories (Lombard, IL) BLISS for TMAs; My research and commercial labs first developed the Q3DM EIDAQ 100 that was purchased by Beckman Coulter and renamed the IC 100 (and was off the market at this writing). Most of these instruments are based on widefield microscope optics and a few utilize confocal optics (Evotec Opera, Atto CARV II and GE-Amersham In Cell 3000). All utilize autofocus, motorized stage scanning and digital cameras to perform automatic scanning and image capture over large areas. Most move in an incremental raster pattern to capture each image and some perform autofocus prior to capturing each time to enable high resolution (low depth of field) scanning while others autofocus on every few fields for lower resolution (higher depth of field) scanning. The mosaic of images is sometimes combined to produce a single large contiguous image. The Aperio system utilizes a pre-scan autofocus surface profiling step prior to moving the stage in a series of continuous-motion capture sequences to acquire long image strips. These long strips are then combined to produce a single large image of the scan area.

Our first instrumentation research projects that led to the IC 100 included solving the key challenges of high speed autofocus [8-10], real-time image segmentation [11] and light source stabilization for reliable image fluorometry [12]. The resulting IC 100 performs autofocus in as little as 0.25 s on slides, the fastest available for image-based autofocus, and performs sophisticated image analyses for robust high content screening [6]. The IC 100 scans in an incremental raster pattern like most of the other commercially available systems.

Noting the need for faster systems, we began building an instrument to overcome the limitations of incremental scanning in 1994. Our goal of enabling scanning with the entire range of available objective resolutions led to the design shown in Fig. 1, which utilizes parallel imaging of multiple focal planes for on-the-fly autofocus to provide high precision closed-loop-feedback control of focus with submicron depths of field [13]. This design removes the need to carry out a pre-focus step to profile the scan surface as required in the Aperio system. For high resolution objectives (i.e., NA > 0.5), the prefocus step can add substantial time to scanning and even dominate the time required. With automatic slide and multi-well plate stackers and robots, the time required to complete an entire experiment with many slides or

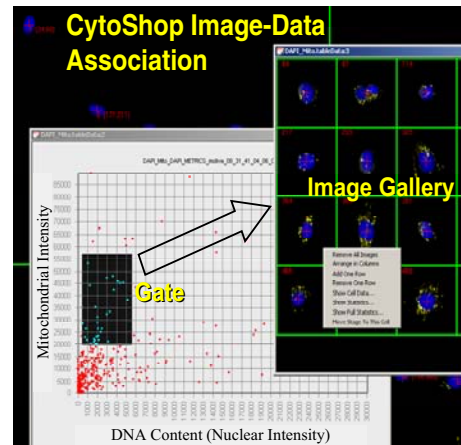


**Fig. 1.** A cartoon of parallel measurement of the degree of focus of multiple focal planes is shown (left). The specimen is magnified by the objective lens, and multiple planes within the specimen are in focus simultaneously (at different axial positions in image space). The lens magnifies the specimen into the image space, where it is captured by several sensors positioned at different axial and lateral locations. Each sensor in the array collects a unique in-focus focal plane, and each corresponding image contains different in-focus and out-of-focus information. In practice, image sensors cannot be placed on top of each other and, instead, the different focal planes in the same lateral field of view are imaged using beamsplitters (not shown). Our first prototype utilized nine fiberoptic imaging bundles to collect adjacent fields of view, and a portion of this apparatus is shown in the engineering drawing (right). Reprinted from Bravo-Zanoguera et al. [13].

plates means that the entire process must be optimized for fast image acquisition. With on-the-fly autofocus, parallelism essentially eliminates the time required for scanning. Since many cells lie adjacent to one another, continuous scanning enables highly parallel cytometry thereby overcoming the fundamental limitation of single-file cell-data collection in flow cytometry. Thus, while image-based systems have typically been slower than flow cytometers to date, continuous scanning will provide both much higher cytometry data bandwidth and cell analysis rates. Further increases in speed will be enabled by wider CCD sensors without sacrificing the signal-to-noise ratio (SNR). With TDI CCD sensors commercially available now ( $> 200$  MHz) and blood cell preparations (where small cells can be densely packed), it will be possible to reach over 100,000 cells/s with  $0.25 \times 0.25 \mu\text{m}^2$  spatial sampling, yielding about 400 pixels per cell. With 1-10  $\mu\text{s}$  dwell times in high speed flow cytometry and 1-10 ms dwell times in high speed TDI CCD sensors, the effective SNR can be over 500-fold greater ( $\sim\sqrt{400,000}$ ) without accounting for differences between PMTs and CCDs. With these instruments, it will become routine to collect over 1 TB of image data/day.

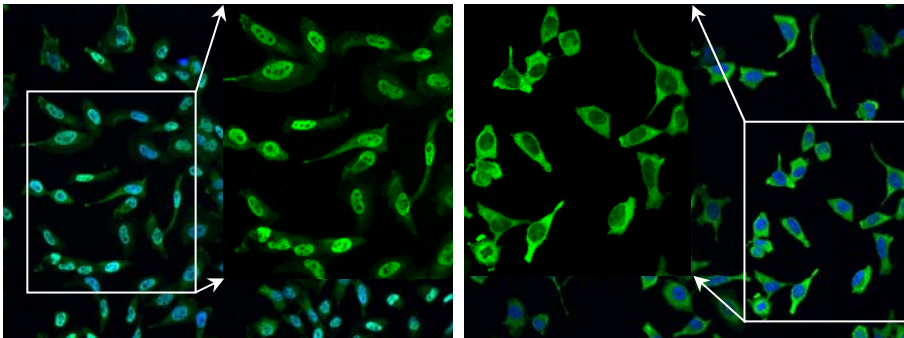
With the technology to image a slide containing 5 million blood cells in a few minutes (for a 2 Gpixel image), practical clinical and biological applications require fully automated image analysis (including image segmentation and feature extraction), cell-by-cell feature databases with links to the cell images, and sophisticated viewing tools. In the example in Fig. 2 from Q3DM-Beckman Coulter CytoShop software, fully automated image segmentation locates the nuclei [11], cells and subcellular patterns [14] and stores cell ROIs and feature sets [15]. The embedded database maintains a link between the cellular ROI and the associated feature vector to enable manual cell classification through arbitrary

gating. In Fig. 2, a scatter plot of mitochondrial intensity as a function of DNA content was gated to select for cells with relatively high mitochondrial intensities and low DNA content. As data sets grow in size, more sophisticated cell classification would aid mining the data for cells of interest and patterns in



**Fig. 2.** These printscreen overlays show the image-data association in the Q3DM-Beckman Coulter CytoShop software. Arbitrarily large montaged fields of view are cached on disk and displayed as a single image comprised of reconstructed regions of interest (ROIs) each containing one cell (partially visible in background) [15]. Each cell has up to 75 features that can be plotted as shown in the large inset. Manually selected gates set (in this case for low DNA and high mitochondrial intensities) automatically generate galleries of cell images as shown in the right inset.

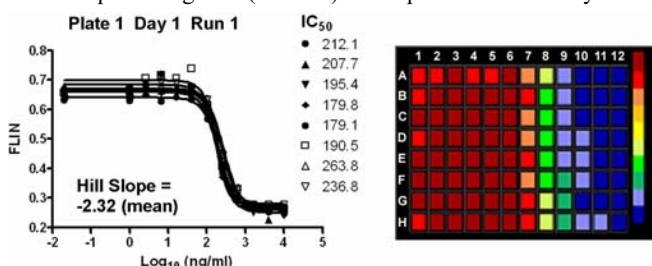
**Fig. 3.** Inhibition of translocation of the cytoplasmic NF $\kappa$ B p65 regulatory subunit to the cell nucleus of HeLa cells. Left: maximum IL-1 $\alpha$  stimulation with translocation to the nucleus. Right: maximum inhibition by IL1-RA in the presence of 10 ng/ml IL-1. Stimulation with IL-1 causes phosphorylation and separation release of the inhibitory subunit I $\kappa$ B, allowing the regulatory subunit p65 to translocate to the nucleus. Inhibition of the pathway results in a higher fractional distribution of the Alexa488-immunolabeled p65 in the cytoplasm. The original images are 662x495  $\mu\text{m}^2$  and the magnified insets are 248x298  $\mu\text{m}^2$  images of only the Alexa 488 channel. Reprinted from Morelock et al. [6].



heterogeneous cell populations. The image processing and analysis subroutines in

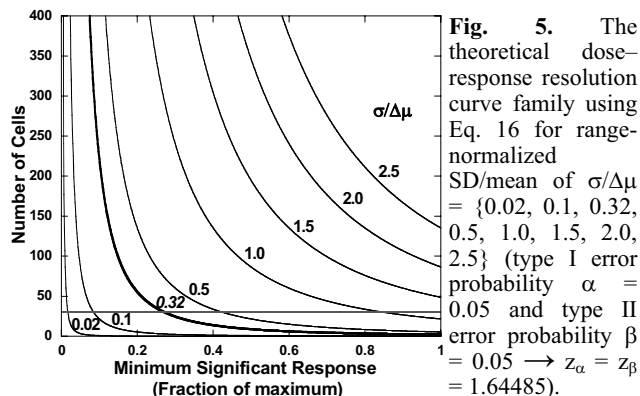
CytoShop are plug-ins [16] and the plug-in architecture is available to 3<sup>rd</sup> parties and users to create additional features. Without tools like these, it will be impossible to realize the full potential of experiments comprising hundreds of millions of cells. Even more challenging will be performing data mining across years of large experiments. With ever evolving image analysis tools, it would be very useful to continue to update feature sets and perform increasingly more sophisticated data mining.

As an example of these tools applied to HCS, Fig. 3 shows two control images for the translocation of nuclear factor  $\kappa$ B (NF $\kappa$ B) from the cytoplasm to the nucleus [6]. NF $\kappa$ B is involved in immune and cell stress response signaling pathways [5, 17-22]. Upon stimulation, for example by proinflammatory cytokines, the NF $\kappa$ B inhibitory subunit (I $\kappa$ B) is phosphorylated and subsequently proteolysed. Loss of the inhibitory subunit frees the modified protein's p65 regulatory subunit to translocate from the cytoplasm into the nucleus where it can activate defense genes in response to the external stimuli, e.g., from viruses and bacteria. For control studies for statistical validation of HCS methods, inhibition of translocation was accomplished by pretreatment of cells with an IL-1 receptor antagonist (IL-1RA) in the presence of IL-1 cytokine



**Fig. 4.** An example family of inhibition curves is shown for one scan on one plate that was prepared and inhibited with IL-1RA (in the presence of 10 ng/ml of IL-1) in a dose-dependent manner, incubated for 20 min at 37°C, and then imaged five times on each of days 1 and 2. Titrations from low to high concentrations of IL-1RA are also shown in each row of the 96-well plate. At low IL-1RA concentrations, translocation is maximal and color-coded red. As [IL-1RA] is increased, the color changes to blue, indicating maximum inhibition, i.e., minimal p65 nuclear translocation. The far-left plot points are 0.0 ng/ml concentrations arbitrarily represented as 0.02 ng/ml. The color scale represents minimum (dark blue) and maximum (dark red) fractional localized intensity of the nucleus (FLIN). IC<sub>50</sub> is the 50% inhibitory concentration. Reprinted from Morelock et al. [6].

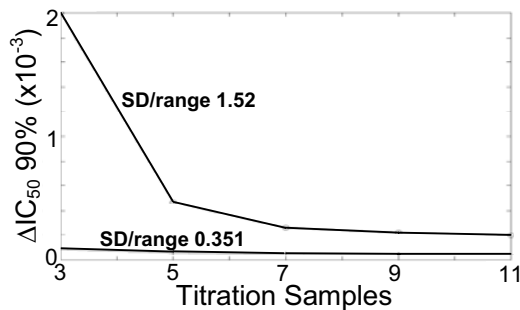
(Fig. 3) [6]. IL-1RA is an endogenous cellular protein that, e.g., is produced by human in vitro-derived macrophages [23] and has been implicated in the pathogenesis/inhibition of atherosclerosis [24]. HeLa cells were cultured overnight in four 96-well plates and IL-1RA was added to each 12-well row at concentrations of 10,000, 4,000, 1,600, 640, 256, 102, 41, 16, 6.6, 2.6, 1.0, and 0.0 ng/ml. An example family of dose-response curves is shown in Fig. 4 with the cell measurement of fractional localized intensity of the nucleus (FLIN), which is the ratio of the NF $\kappa$ B immunofluorescence integrated nuclear intensity to the whole cell (nuclear + cytoplasmic) integrated intensity. The FLIN data was measured cell-by-cell and mean FLIN was reported for each well. The relationship between the needed minimum significant response between doses and the number of cells is plotted in a family of cell-by-cell FLIN SD/range ( $\sigma/\Delta\mu$ ) curves in Fig. 5. These results show that the dynamic range can have a much greater impact on statistical significance than increasing the number of cells measured per well, which requires increasing the number of images scanned per well. The improved  $\sigma/\Delta\mu$  in these data translated to 1/10<sup>th</sup> the number of cells required for the same statistical significance as in a previously reported NF $\kappa$ B HCS screen [5]. The relationship between cell-by-cell FLIN  $\sigma/\Delta\mu$  and



**Fig. 5.** The theoretical dose-response resolution curve family using Eq. 16 for range-normalized SD/mean of  $\sigma/\Delta\mu = \{0.02, 0.1, 0.32, 0.5, 1.0, 1.5, 2.0, 2.5\}$  (type I error probability  $\alpha = 0.05$  and type II error probability  $\beta = 0.05 \rightarrow z_\alpha = z_\beta = 1.64485$ ).

The number of cells required to reach statistical significance between individual doses (y-axis) is plotted against the FLIN of the NF $\kappa$ B translocation measurement range that is resolved (x-axis). These relationships allow for easy visualization of the effect of improved or degraded measurement fidelity and resolution on cell number requirements. For both runs on all four plates, the FLIN SD range was 0.049–0.154, the combined SD was 0.095, and the mean FLIN range was 0.2975, giving a combined  $\sigma/\Delta\mu = 0.32$  (bold curve) over all experiments. For comparison, solving for  $\sigma$  using the data in Table 1 of Ding et al. [5] using Eq. 16 of Morelock et al. [6] yielded a SD of 83.5 with a reported range of 55, for  $\sigma/\Delta\mu = 1.52$ . The line at 30 cells represents a practical limit for the Gaussian assumption, below which Poisson statistics are required. Reprinted from Morelock et al. [6].

IC<sub>50</sub> precision was studied using Monte Carlo simulations and as shown in Fig. 6, improved precision in the cell-by-cell measurement enables substantially fewer doses to be used for obtaining high IC<sub>50</sub> precision.



**Fig. 6.** Difference in 50% inhibitory concentration ( $\Delta IC_{50}$ ) 90% confidence intervals for  $\sigma/\Delta\mu$  values of 1.52 and 0.351 from 500 Monte Carlo simulations for each number of sample concentrations (three, five, seven, nine, and 11) in the titration is shown. Measurement precision becomes more important as sampling and coverage of the relevant concentration ranges decrease. The elbow in the top curve corresponds to a regression breakdown. Reprinted from Morelock et al. [6].

In these examples, automated image analysis was used to create high precision data. There are many remaining challenges for creating similar data for other cellular dynamics; and additional challenges for making both the images and data available in a public database for improved analyses and data mining.

#### ACKNOWLEDGEMENTS

Support for the research reported here includes: the Whitaker Foundation Biomedical Engineering Research Grants program; the WPC Research and Education Fund; NSF Major Research Instrumentation grant BES-9871365; NIH grant R01 HD37782; UC MEXUS; NIH SBIR Fast-Track 1R44 RR015170; grant C01-0094, A-1 from the California Technology, Trade and Commerce Agency, California Technology Investment Partnership; and NIH grant R01 EB006200.

#### REFERENCES

- [1] S. Bajaj, J. B. Welsh, R. C. Leif, J. H. Price, "Ultra-Rare-Event Detection Performance of a Custom Scanning Cytometer on a Model Preparation of Fetal nRBCs," *Cytometry*, 39:285-294, 2000.
- [2] J. Kononen, I. Bubendorf, A. Kallioniemi: Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* 4:844-7, 1998.
- [3] M. S. Skacel, B. Skilton, J. D. Pettay, R. R. Tubbs, "Tissue microarrays: a powerful tool for high-throughput analysis of clinical specimens," *Applied Immunohistochemistry & Molecular Morphology* 10(1):1-6, 2002.
- [4] A. Rabinovich, S. Krajewski, M. Krajewska, A. Shabaik, S. M. Hewitt, S. Belongie, J. C. Reed, J. H. Price, "Framework for Parsing, Visualizing and Scoring Tissue Microarray Images," *IEEE Transactions on Information Technology in BioMedicine*, In Press.
- [5] G. Ding, P. Fischer, R. Boltz, J. Schmidt, J. Colaianne, A. Gough, R. Rubin, D. Uiller, "Characterization and quantitation of NF $\kappa$ B nuclear translocation induced by interleukin-1 and tumor necrosis factor- $\alpha$ ," *J Biol Chem* 273:28897-28905, 1998.
- [6] M. M. Morelock, E. A. Hunter, T. J. Moran, S. Heynen, C. Laris, M. Thieleking, M. Akong, I. Mikic, S. Callaway, A. Goodacre, R. P. DeLeon, D. A. Zacharias, J. H. Price, "Statistics of Assay Validation in High Throughput Cell

- Imaging of Nuclear Factor  $\kappa$ B Nuclear Translocation," *Assay and Drug Development Technologies*, 3(5): 483-500, 2005.
- [7] J. N. Harada, K. E. Bower, A. P. Orth, S. Callaway, C. G. Nelson, C. Laris, J. B. Hogenesch, P. K. Vogt, S. K. Chanda, "Identification of novel mammalian growth regulatory factors by genome-scale quantitative image analysis," *Genome Research* 15:1136-1144, 2005.
- [8] J. H. Price, D. A. Gough, "Comparison of Digital Autofocus Functions for Phase-Contrast and Fluorescence Scanning Microscopy," *Cytometry*, 16(4): 283 - 297, 1994.
- [9] M. Bravo-Zanoguera, B. von Massenbach, A. L. Kellner and J. H. Price, "High Performance Autofocus Circuit for Biological Microscopy," *Review of Scientific Instruments*, 69(11):3966-3977, 1998.
- [10] M. Oliva, M. E. Bravo-Zanoguera, J. H. Price, "Filtering out contrast reversals for microscopy autofocus," *Applied Optics, Optical Tech. & Biomed. Optics*, 38(4):638-646, 1999.
- [11] J. H. Price, E. A. Hunter, and D. A. Gough, "Accuracy of Least Squares Designed Spatial FIR Filters for Segmentation of Images of Fluorescence Stained Cell Nuclei," *Cytometry*, 25(4): 303 - 316, 1996. 1998.
- [12] S. Heynen, D. A. Gough, J. H. Price, "Optically Stabilized Mercury Vapor Short Arc Lamp as UV-Light Source for Microscopy," *SPIE Photonics West '97*, vol. 2982: 430-434, Feb. 8 - 14, 1997.
- [13] M. Bravo-Zanoguera, B. von Massenbach, J. H. Price, "Automatic on-the-fly focusing for continuous image acquisition in high-resolution microscopy," *SPIE Proc. Optical Diagnostics of Biological Fluids and Advanced Techniques Analytical Cytology*, San Jose, 3604:243-252, 1999.
- [14] E. A. Hunter, R. S. Ingermanson, W. S. Callaway, "System and Method for Automatic Color Segmentation and Minimum Significant Response for Measurement of Fractional Localized Intensity of Cellular Compartments," *International Patent Cooperation Treaty Publication Number WO 03/078965 A2*, 2003.
- [15] W. S. Callaway, E. A. Hunter, "Managing images of microscopic specimens," *United States Patent Application Number 20030033090*, 2002.
- [16] W. S. Callaway, E. A. Hunter, "Method and System for Extensible Data Processing," *United States Patent 6,886,186*, 2005.
- [17] I. M. Verma, J. K. Stevenson, E. M. Schwarz, D. Van Antwerp, S. Miyamoto. Rel/NF- $\kappa$ B/I $\kappa$ B family: intimate tales of association and dissociation. *Genes Dev.*, 9:2723-2735, 1995.
- [18] N. Auphan, J. A. DiDonato, C. Rosette, A. Helmberg, M. Karin, "Immunosuppression by glucocorticoids: inhibition of NF- $\kappa$ B activity through induction of I $\kappa$ B synthesis," *Science* 270:286-290, 1995.
- [19] Zabel U, Henkel T, dos Santos Silva M, Baeuerle PA, "The NF- $\kappa$ B precursor p105 and the proto-oncogene product Bcl-3 are I $\kappa$ B molecules and control nuclear translocation of NF- $\kappa$ B," *EMBO J* 812:201-211, 1993.
- [20] N. S. C. van Oers, Z. J. Chen, "Kinasing and clipping down the NF- $\kappa$ B trail," *Science* 308:65-66, 2005.
- [21] K.-Y. Lee, F. D'Acquisto, M. S. Hayden, J.-H. Shim, S. Ghosh, "PDK1 nucleates T cell receptor-induced signaling complex for NF- $\kappa$ B activation. *Science* 308:114-118, 2005.
- [22] H. Su, N. Bidère, L. Zheng, A. Cubre, K. Sakai, J. Dale, L. Salmena, V. Hakem, S. Straus, M. Lenardo, "Requirement for caspase-8 in NF- $\kappa$ B activation by antigen receptor," *Science* 307:1465-1468, 2005.
- [23] R. W. Janson, K. R. Hance, W. P. Arend, "Production of IL-1 receptor antagonist by human in vitro-derived macrophages: effects of lipopolysaccharide and granulocyte-macrophage colony-stimulating factor," *J Immunol* 147:4218-4223, 1991.
- [24] R. Dewberry, H. Holden, D. Crossman, S. Francis, "Interleukin-1 receptor antagonist expression in human endothelial cells and atherosclerosis," *Arterioscler Thromb Vasc Biol* 20:2394-2400, 2000.