

# IMAGE ACQUISITION AND UNDERSTANDING IN HIGH-THROUGHPUT HIGH-RESOLUTION CELL-BASED SCREENING APPLICATIONS.

*Yuval Liron, Yael Paran, Irina Lavelin, Suha Naffar-Abu-Amara, Sabina Winograd-Katz, Benjamin Geiger and Zvi Kam*

Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel.

## ABSTRACT

Quantitative interpretation of microscope images is more challenging the higher the resolution of the images is. The reward is rich multi-parametric characterization of sub-cellular structures and detailed description of cell responses to perturbations. This information is the basis of high-throughput cell-based screening, searching to discover new drugs and understand molecular mechanisms at the cell level.

We have developed a fast screening microscope acquiring high-resolution images from cells cultured in plastic-bottom multi-well plates, [1-4] and are writing an automated pipeline for the analysis of Tera Bytes of images from high throughput screens. The platform includes database for storage and retrieval of images, visualization with easy linkage of the analyzed results to the original multi-color images, segmentation of objects in images (including cells, nuclei, cytoskeletal fibers and sub-cellular organelles), multi-parametric quantification of morphological and multicolor fluorescence intensities, and statistical comparisons to control wells displayed in color coded scores on the plate graphics. This system was successfully employed for screening of the effect of drugs, gene over-expression and siRNA of diverse cellular properties, including cell adhesion, migration, survival and cytoskeletal organization.

**Index Terms**— Fluorescence Microscopy, Biological Cells, Image Processing pipeline.

## 1. INTRODUCTION

Image understanding and quantitative analysis of microscope-acquired data is a technology of wide use in biological sciences. Automated image processing is becoming a central resource in post-genomic research. The spread in the biological questions and the information presented by the images prevented so far establishing of a generic infrastructure for quantitative analysis of experiments acquiring images. However, a unified platform for the analysis of biological images, equivalent to BLAST in DNA sequence analysis, would have high impact on post-genomic bioinformatics, creating a universal language

for referring to high-level properties of biological systems such as cells, embryos, organs and whole animals. One way to create such language is to identify and register structures and to quantify various morphological features, protein-concentrations and molecular interactions in images. It will help coherent accumulation of knowledge for biological complex systems.

In view of the number of possible molecular labels used for fluorescence microscopy and the variations between cell types and within cells of the same lineages, biological images are characterized by an extremely broad diversity in contents, and therefore by a wide range of applications developed for their analysis. Unfortunately, these applications are not easy to transfer from one laboratory to another due to different image formats, incompatibility of image processing environments, and last, but most important, the variability in the images originating from even similar biological specimens prepared in different laboratories. All these factors require expert manpower for adopting existing applications to local uses, and professionals for the development of new ones. Such manpower is not commonly available in experimental biological laboratories.

Image processing is central in many other research areas, such as applied mathematics, robotics, high-energy physics, astronomy and medicine. The establishment of a common platform in astronomy and in high-energy physics, and the agreement on a universal format for medical images demonstrated the benefits for coherent efforts by many groups all contributing collectively to the progress in these areas. However, in biological sciences, competing commercial companies offering imaging solutions, both for acquisition and analysis, dominate the scene today. Most of these solutions are optimized towards some applications, define rigid frameworks for the user, and many are closed for the researcher for modifications. Open frameworks (commercial as well as public domain solutions) are too complex to adapt or develop without specialized personnel. The diverse nature of biological images needs a platform with a rich variety of applications, open for smooth acceptance of plug-ins from many laboratories and easy to modify for evolving specific needs. Such a platform will allow biologists to pool their efforts in biological image processing, share algorithms and analysis programs, build

intuitive and user-friendly tool-boxes for image understanding, and most important, facilitate cooperative integration of high-level proteomic and cellomic information. While several years ago the development of a universal platform with the desired features was not easy, present computer and communication technologies and cross-platforms software tools are making such a task feasible.

As was emphasized above, biological image understanding platforms cannot offer a global solution to all existing or emerging problems. Instead one needs:

- I. A platform offering rich sets of solutions to many applications.
- II. A way to locate available solutions based on the problem presented in biological terms (not based on the mathematical algorithm nomenclature). The ideal platform should attempt to be an ideal problem solver: given the images, propose possible answers.
- III. Access under the same platform to more than one available solution for similar biological application with easy comparison to help biologists optimize the solution to their specific needs.
- IV. Easy tools for modification of existing solutions.
- V. Since a universal image format does not exist, import/export modules to common formats should be included, with the data describing the experimental parameters.
- VI. The same platform should allow manual processing with visualization, for development and optimization of the analysis, and support totally automated analysis of large data sets of images.
- VII. Compatibility with existing open source platforms (ImageJ, MATLAB, Cell Profiler, Open Microscopy) would make wide range of applications available.
- VIII. Rich set of examples and tutorials should be offered.
- IX. The platform should be build hierarchically, offering high level solutions for the untrained user, with capability to go into the processing steps to modify parameters, exchanged modules, and write new modules.

## **2. THE PLATFORM UNDER DEVELOPMENT:**

### **2.1. Image Analysis Modules**

The software package under development is directed mainly to cell-based screening applications, where microscope images are acquired from tissue cells cultured in dense multi-well plates. However, interactive modules with display of segments and their evaluated parameters that were included for tuning of analysis parameters were also useful for quantitative analysis of single images. The modules were compiled from our collection of mixed code languages (Fortran, C) using Portland Group compiler.

The available image processing modules can be sorted into the following categories:

1. Input/Output of images from/to the image data-base. Images can be referred through various links, typically the

images from a well in the plate, reporting a specific cell type, treatment or growth conditions.

2. Image filtration algorithms, including noise-reduction, smoothing, background subtraction, flattening, contrasting, illumination corrections, etc.

3. Segmentation algorithms, such as blobby contiguous component, watershed segmentations and recognition of fibers. These tools define cells, nuclei and other sub-cellular organelles, and identify special segments such as dead cells or debris.

4. Quantification: These include various morphological characterizations of segments (area, perimeter, shape features etc), as well as fluorescence total intensities and textural parameters.

5. Statistics: Averages, percentiles, and non-parametric statistical comparisons of the distributions of features quantified in treated wells and in control wells.

6. Visualization: including display of individual images, montages of images, for example all images taken in a given well, color superposition for multicolor images, as well as graphical display of plates, where wells can be painted with spectral colors coding for any one or combinations of the multi-parametric quantitative evaluations based on image analyses. The plate graphics interface allows clicking on a well, displaying the images taken in it and seeing the parameters obtained by the analysis.

### **2.2. Assembly of Modules into an Image Analysis Pipeline**

Using the image processing and quantification modules describe above, scripts can be executed with script lines running sequential steps in the analysis, and pass between them images, variables, and analyzed data (typically saved as lists of calculated attributes for each segment in the images, or for the whole well). The scripts execute loops (i.e. on wells in a plate, on all images acquired in a well, on colors), thus creating analysis pipeline for high-throughput experiments recording Tera Bytes of images. Since the typical acquisition time of high-resolution images, including auto-focusing and fluorescence exposures, is less than a second (0.3 Tera Bytes/week) and this volume cannot be saved (we have only 50 Tera Bytes NetApp storage that would be filled in a few months), we target the analysis time for each image to the second time frame. Presently, despite optimization, some of the algorithms processing time per image (e.g. fiber analysis) is longer by a factor of 2-8, and we use a computer cluster to compensate this factor. Images can be sent to the cluster nodes for analysis in real time, or at the end of a screen.

### **2.3. Integration of External Modules**

It would be ideal to be able to smoothly add modules from other analysis platforms. Presently, we are working to import MATLAB modules. The advantage is the ease and exchange of algorithms, but the slow analysis speed

eventually calls for implementation using faster processing codes. We are examining linking to JAVA codes.

## 2.4. Interface for the biologist

Image analysis programs cannot offer global solutions to biological images. Adaptation of existing programs to a specific application may require flexibility at different levels:

1. Designing new algorithms and adding modules.
2. Changing specific steps in the analysis process.
3. Modification of variables used for the analysis (e.g. thresholds, size parameters, fluorescence intensity ranges). The first level requires a trained researcher, and the platform is open for addition of modules. Since the typical wet-table biologist has little training in image processing, but has excellent perception for the interpretation of the images he acquires, the platform offer intuitive adaptation at the second and third levels. Examples of documented scripts that perform image-processing tasks and explained in a clear biological language can be tested with on-line visualization. Examples include nuclei counting, cell identification, quantification of fluorescence in cells and inside nuclei, scoring the microtubule or actin filament content in cells, integrating fluorescence in mitochondria and Golgi, determining distributions of morphological and fluorescence intensity parameters characterizing cell adhesions and vesicles as well as dynamic measurements such as microtubule growth dynamics.

## 3. APPLICATIONS

The platform has been applied to images acquired by a screening microscope developed in our laboratory for drugs and iRNA screening projects [1-4]. Here we describe a search for the effects of the National Cancer Institute COMBO drugs library on microtubules in several cell lines. In Figure 1 we show the titration curve for 14 drugs from the library. The wells in the plate graphics are painted with spectral colors indicating the analysis scored in each well compared to the values in control wells. To generate such plate-wide analysis we identified individual cells in all images acquired in the plate (Figure 2), and quantified for each cell a number of attributes, including cell area, total microtubule-associated fluorescence and diffused background fluorescence within the cell area. The score for each well is defined here as the averaged cell-by-cell background-subtracted microtubule-associated fluorescence normalized to cell area [5]. The concentration dependent microtubule-disrupting effect of Nocodazole and the microtubule stabilizing and bundling effect of the Taxol-like drug Paclitaxel (rows marked) is displayed by blue and red decaying to the control-green scores. Such data is being used for evaluating cell-line-dependence drug potency, for example in cancer-derived cells (Figure 3), and for quantifying combinatorial drug effects. The plate displays of the scores are mouse-click-linked to all the digital data (statistics and full data of all attributed calculated for all cells) as well as to montages of all images acquired in each

well and the individual images. This allows very convenient survey of the whole data set at all stages of the analysis.

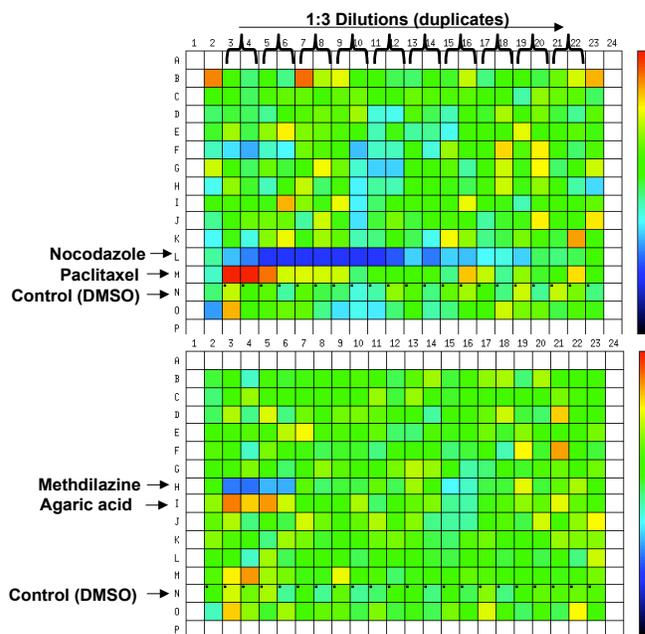


Fig. 1 Titration curves showing the effects of drugs on microtubules. The color indicated the integrated microtubule-associated fluorescence after subtraction of intracellular background, divided by cell area and normalized by the average on control wells. Control wells are marked with small black squares. The strong disrupting effect of Nocodazole and the stabilizing effect of Paclitaxel is clearly displayed. Drugs were applied in duplicates (columns). Random distribution of red and blue wells is due to artifacts, such as dead cell debris. We are working on procedures for deleting such outliers from the analysis.

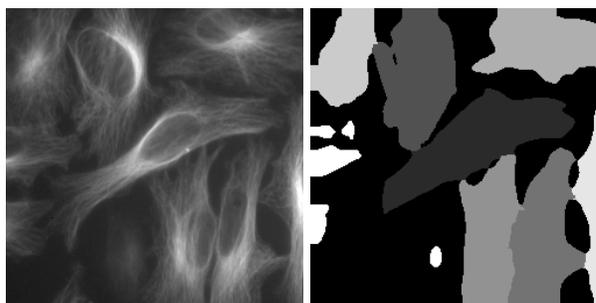


Fig. 2 The segmentation of individual cell objects in images taken from tissue cultures grown in multiwell plates. Segmentation is based on watershed performed on DAPI and YFP-tagged microtubules followed by segment aggregation guided by the nuclei.

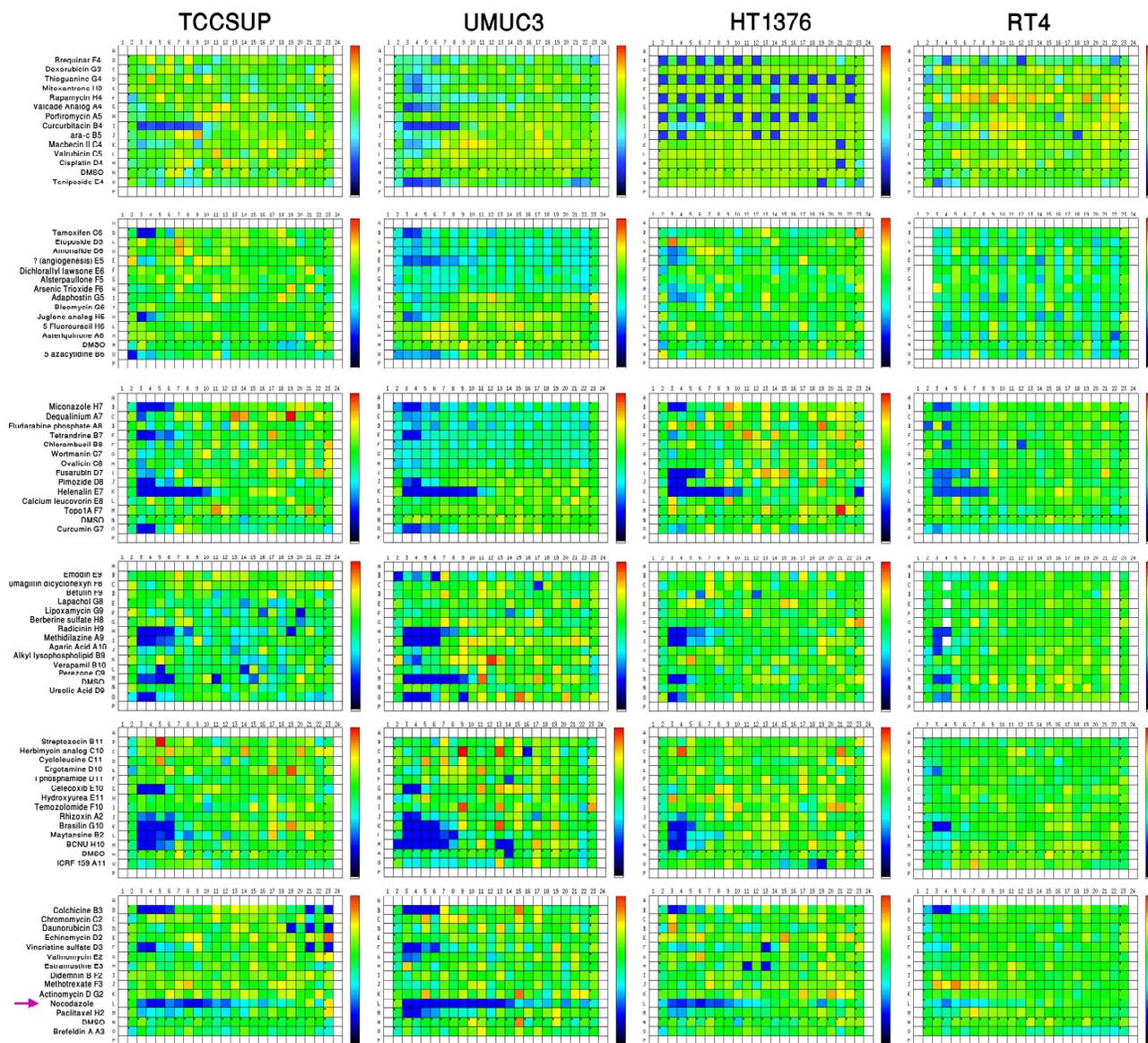


Fig. 3 Screen summary displaying the effects of the NCI COMBO drug library at decreasing concentrations on four bladder-derived cell lines. Despite the noise (due to robot errors and cell debris) the different sensitivities of the cell lines is clearly seen.

### ACKNOWLEDGEMENTS

The research was supported by the Kahn fund for Systems Biology at the Weizmann Institute.

### REFERENCES

[1] Liron, Y., Paran, Y., Zatorsky, N. G., Geiger, B. and Kam, Z. (2006) Laser Autofocusing System for High Resolution Cell Biological Imaging. *J. Microsc.* 221: 145-151.

[2] Paran, Y., Lavelin, I., Naffar-Abu-Amara, S., Winograd-Katz,

S., Liron, Y., Geiger, B., and Kam, Z. (2006) Development and application of automatic high resolution light microscopy for cell-based screens. *Methods Enzymology* 414:228,247.

[3] Paran, Y., Y., Ilan, M., Kashman, Y., Goldstein, S., Liron, Y., Geiger, B. and Kam, Z. (2007). High-throughput screening of cellular features using high resolution light microscopy; Application for profiling drug effects on cell adhesion. *J. Struc. Biol.*, 158: 233-243.

[4] Naffar-Abu-Amara, S., Shay, T., Galun, M., Cohen, N., Isakoff, S.J., Kam, Z. and Geiger, B. (2008) Identification of Novel Pro-Migratory, Cancer-Associated Genes Using Quantitative, Microscopy-Based Screening. *PLoS ONE* 1,e1457.

[5] Lichtenstein, N., Geiger, B. and Kam, Z. (2003). Quantitative analysis of cytoskeletal organization by digital fluorescence microscopy. *Cytometry* 54A:8-18.