

Automated Microscopy and High Content Screens (Phenotypic Screens) in Academia Labs

Dimitri Moreau* and Jean Gruenberg*

Abstract: Imaged-based screening has been developing extremely quickly in the past 10 years. Academic institutes quickly realized that the discovery capacity of this technology was huge, allowing the automatic detection and quantification of complex cell phenotypes. Associated with chemical or genetic perturbations, high content screening is the method of choice for a deep system biology analysis. The evolution of high-content screening is mainly due to the recent progress in the development of fast and high quality automated imagers and of a plethora of new very bright fluorescent markers, so that almost any cellular element can be seen and imaged. In this paper we review and summarize the major steps in the development of an image-based screening project.

Keywords: High-content screening · Image-based screening



Introduction

Today biomolecular screening is no longer the exclusive massive discovery tool of the pharmaceutical industry. Indeed, much effort is being invested in academia into setting up such research programs. The discovery of RNAi technology and the development of human genome libraries certainly have largely contributed to this new area of research.^[1] The first siRNA screens were performed in the early 2000s as were the first deep genome-wide analysis, helping to create a new discipline, which is today referred to as systems biology.^[2]

Unlike any previous research project, the screening campaigns by academia labs allowed the extremely fast discovery of new compounds, genes or pathways, because of the automation of standardized

biological or biochemical assays. In addition to drug discovery, this technology has already been applied to many areas of Life Sciences, including stem cell biology, viral infection, cellular processes, parasite entry, toxin transport, subcellular organelle maintenance, cancer biology, lipid homeostasis.^[3] Today screening projects are being developed in most academic institutions, and Systems Biology has moved into the limelight,^[4,1] in particular in Switzerland with the SystemsX initiative (<http://www.systemsx.ch/>).

Screening is all about obtaining quantitative information through the full automation of each step of the process: sample preparation, assay, reading and data analysis. Therefore, many universities and academic institutes have invested a lot in the development of screening facilities, equipped with automated tools and robots to run screening projects. Conversely, the manufacturers providing this equipment had to adapt their offers to the academic world and specific needs. Today, we have access to extremely powerful equipment, which can run any steps of a screening project.^[5]

Light microscopy, including fluorescent imaging, is central to research in cell and molecular biology to visualize the behavior of complex biological systems, and is itself undergoing major developments in recent years.^[6] Hence, the screening community had to come up with automated tools capable of running usual biological assay used in research. This led to automated microscopes that have been, invented in mid-to-late 1990s,^[7] extensively developed over the past decade, and achieve

today high quality imaging in a very fast scanning mode. Automated microscopes are now able to acquire thousands of high quality images per hour, which in turn results in very large amounts of digital image data. To cope with this new data accumulation, laboratories have developed new software for automated analysis in order to extract as much information as rapidly as possible.^[8] The aim of this paper is to highlight some of the key steps necessary to develop and run your image based screening project.

Assay Development

To design an assay in a screen format, one has to keep in mind that the rule number-one for such a project is to repeat a simple experiment under exactly the same conditions thousands of time with only one parameter that changes at a time in order to extract highly quantitative information. Hence, it is not always so straightforward to transfer a bench experiment to an automated high-throughput format (Fig. 1). The assay has to be extremely robust and consistent and should exhibit the best dynamic range that can be achieved in order to be able to select hits out of the cloud of data points. Today most high content screens (HCS) are run on cell lines or primary cells, but an increasing number of projects use yeast cells, bacteria, worms, parasites or small organism embryos.

Screens are mainly run in microtiter plates with a format varying from 96 to 1536 wells. In HCS, we are using high quality imaging plates, with a particularly

*Correspondence: Prof. D. Moreau, Prof. J. Gruenberg
ACCESS-Geneva
NCCR Chemical-Biology
Department of Biochemistry
University of Geneva
Quai Ernest Ansermet 30, 1211 Geneva 4
E-mail: Dimitri.Moreau@unige.ch,
jean.gruenberg@unige.ch

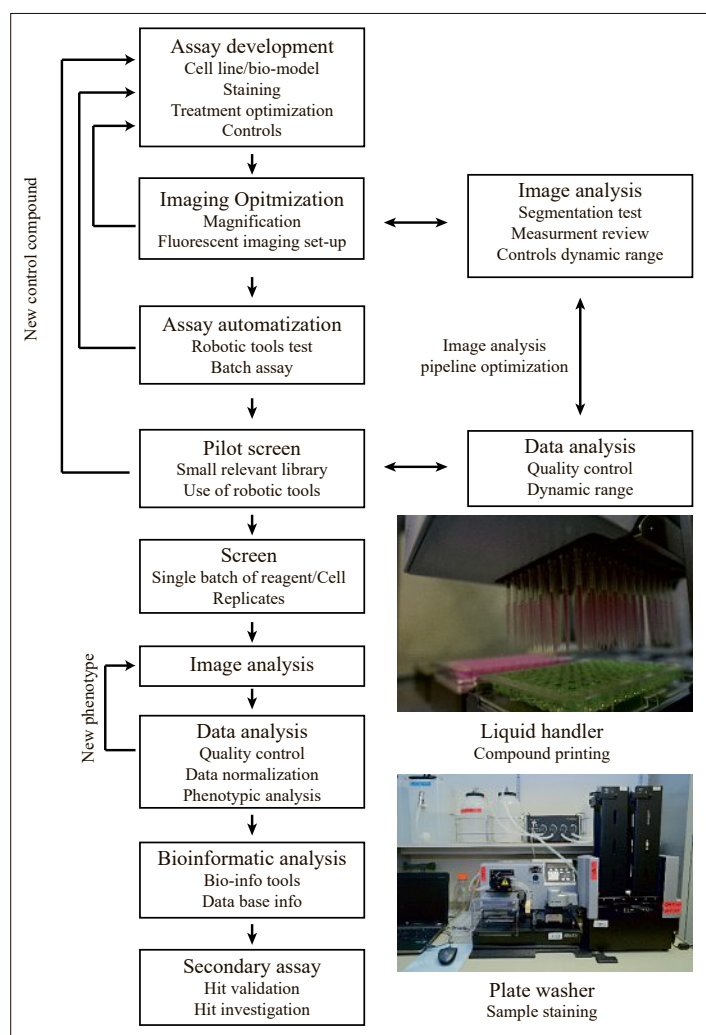


Fig. 1. Key steps for HCS screening development and run. The development part would take between 6 months and 1 year. The screen run should not take more than few weeks. The screen result analysis could take several years, depending on the complexity and the quality of the data. Pictured are robotic tools used in a screening campaign.

tively large quantity, the costs of the assay have to be considered. To optimize costs and staining reproducibility, the use of an automated plate washer (Fig. 1) is mandatory. This will make it possible to reduce the volume of reagents in each well, and will guarantee that each well from each plate from each batch is treated in a highly consistent manner.

It is highly recommended that a pilot screen be run once the optimization steps look promising with good and consistent results. The use of a small library of compounds or siRNAs will make it possible to validate the assay for the screening campaign or to make some new adjustments at all levels of the experiment from the assay to the image analysis (quality control, positive control consistency, analysis pipeline). In addition, a pilot library may also help identify new positive controls, potentially stronger or more relevant for the biology of interest or to discover new phenotypes that could require some modification of the image analysis pipeline (Fig. 1 and 3).

Automated-Microscopy Reading

Over the past 10 years, automated microscopes have drastically evolved to reach the image quality almost identical to that of any standard manual microscope. Today, three main actors (Molecular Devices™, GE healthcare™ and PerkinElmer®) of this field are leading the market with microscopes showing quasi-equivalent capacities. These microscopes are equipped with large field of view and low-light imaging sCMOS camera sensor, combined with strong LED light and fluorescent filters or colored laser. The recent fast evolution of numerical imager and source of light (much more stable) has been of major importance for the evolution of these automated microscopes. Standard objectives provide magnification possibilities from 4× to 100×, allowing most favorable imaging conditions for any object according to its size. When setting up an assay, good care must be taken to avoid any problems due to uneven illumination or image artifact. The laser-based image autofocus offers ideal image conditions for each field of view and the focus point can be adjusted for each object analyzed in the sample. These microscopes can also be used for live cell imaging, because they are equipped with temperature, CO₂, and humidity controls. However, live cell imaging is rarely suitable for large scale screening campaigns because of the time scale (*e.g.* see below).

The automated microscope is driven by a dedicated software, so that image acquisition protocol can be customized, saved and used for each plate of a screen. The interface of the software is quite user friend-

thin and flat plastic bottom. These specifications are needed to ensure good and consistent finding of the autofocus plane by the microscope and optimal image resolution with high magnification objectives (most often with short working distances).

To optimize HCS assays, a certain number of parameters must be properly adjusted: the cell line (biological model) to be used, the cell density in the well, the cell growth over the time period of the assay, the amount of reagent to test, monitoring of the potential toxicity under basal conditions, the efficiency of the treatment (*e.g.* knockdown efficiency in RNAi screens), as well as the sample fixation and staining conditions. To facilitate the optimization of these critical parameters, it is essential to use an appropriate positive control. The choice of this positive control can be critical for the success of the screening project. The control has to be of the same chemical nature as the molecules to be screened in the assay (*e.g.* small molecules or siRNAs). The positive control should not directly target the reporter used in the assay, but a component of the biological pathway of interest. This is of particular importance, since one cannot expect the same dynamic range when targeting the reporter itself vs.

a component of the pathway, and since it might be difficult to clearly identify targets of the biological pathway when running the screen. Obviously, one also needs to set up the assay under the same conditions as used in the HCS itself, including by using the same automated tools. Once optimized, the assay has to be run under very stable conditions in secured incubators and any variation in temperature, humidity or CO₂ must be avoided. This will limit the danger of plate edge effects, a problem that is frequently observed and is mainly due to uneven evaporation between plate wells.

In most cases, HCS uses fluorescence imaging, in order to visualize a reporter present in the cell for example on an organelle, and therefore staining and sample preparation must be optimal. In immunofluorescence assays, the antibodies need to be specific in a very robust way, including a good signal-to-noise ratio, and no fluorescence bleed-through between fluorescence channels used in the detection. Today, commercially available antibodies can be obtained against a vast number of endogenous proteins present in many, if not all subcellular compartments. Since the primary and secondary fluorescent antibodies can be expensive and used in rela-

ly, in most cases it does not take more than half a day to become pretty familiar with the basic commands. According to one's wishes, the software can guarantee the imaging of a pre-established number of fields. Alternatively, it can adapt the number of images acquired for each condition to a pre-set minimum number of objects that are required (e.g. number of cells) before moving on to the next well. For statistical relevance of the data, a minimum number of 1000 to 2000 events (cell, object of interest) would be required for each condition. The number of images and the magnification has to be adjusted accordingly. The microscope can acquire a stack of images that can all be saved and may be used for a 3D reconstruction or compressed to a single image for maximum projection. One should bear in mind that despite fast imaging capability, the acquisition of images from a complete microtiter plate in several colors and with several fields per condition can take hours. It is important to keep this in mind when running large batches of plates, in particular to ensure that the samples to be imaged remain stable over this time course.

A microscope with high-speed imaging will automatically generate a large amount of data, e.g. terabytes every day under routine conditions. Therefore, the development of a safe and large storage system has to be planned before the installation of the microscope, and the IT side of this equipment plays a major role. It is not always easy to integrate all software and hardware. The data pipeline has to be well-defined and structured to ensure no loss of information. The acquisition software needs to be connected to a database storing metadata information (e.g. user name, experiment name, plate ID, well content), and the images should be stored on a secured server. The analysis software should be capable of retrieving the images and the metadata, and finally the numerical data generated should be accessible to the data analysis software. The integration of such a complex pipeline is not an easy task, and it is even harder to make it available to any user.

Image Analysis

The automated microscope will generate a very large number of images, and of course manual reviewing of these is unrealistic in most cases. An HCS project thus implies the capacity to automatize the image analysis.^[9] Several software have been developed that allow users to extract as much information as possible from each image,^[10] including freeware (e.g. CellProfiler,^[11] EBImage,^[12] Image J^[13]) and commercial software (e.g. Genedata

Screeener[®], MetaXpress[®], HCS Studio[™], Harmony[®], IN Cell Analyzer[™]). These software are dedicated to image acquisition and analysis, and they are mostly very well designed but not always fully intuitive – serious training sessions are needed in most cases.

The image analysis will be separated in two main steps: the segmentation of the image and the extraction of the measurements. The goal of the segmentation is to cut the image and generate masks for each individual object visible on the image (e.g. whole cell, nucleus, organelle, viral particle). To identify each object, the software uses several parameters, like size or intensity (Fig. 2), and thus to facilitate the segmentation process the image can be modified to reduce the background and to highlight the object of interest (e.g. deconvolution, top hat noise reduction). These modified images will be used for optimal masking but cannot be used for quantification, since the intensity information is lost. The mask is then applied onto the original image, and can be modified according to the needs (e.g. grown or shrunk), as well as associated (e.g. to obtain a cell mask from separate nuclei and cytoplasm masks) or dissociated. Specific areas (e.g. perinuclear region or the cell edge) that contain individual objects (e.g. organelles)

can be defined in order to reveal and precisely quantify a particular phenotype. In some cases, it is necessary to delete from the analysis, conditions that can affect the quantification (e.g. dividing or dying cell), by implementing a specific segmentation in the analysis pipeline to identify the problematic objects (Fig. 2). The software allows a visual verification of the segmentation quality, by applying the generated mask on the image. It is very important to double-check the masks on several images to be sure that it is applicable to all images in the screen.

Once segmentation is done, measurements are extracted from each mask and channel. One should extract as many meaningful values as possible, but still limit the number of parameters that are expected to vary in the analysis, since a phenotypic analysis will only be feasible if a precise description of each element of the image has been collected. The appropriate phenotypic description of a cell may require between 100 and 200 descriptors, which can be very diverse and reflect several parameters, including object size, intensity, shape, count, localization, granularity, texture, distribution, co-occurrence. An analysis based on a whole-cell mask as main object is often very useful, because the dataset is then generated 'per cell' and

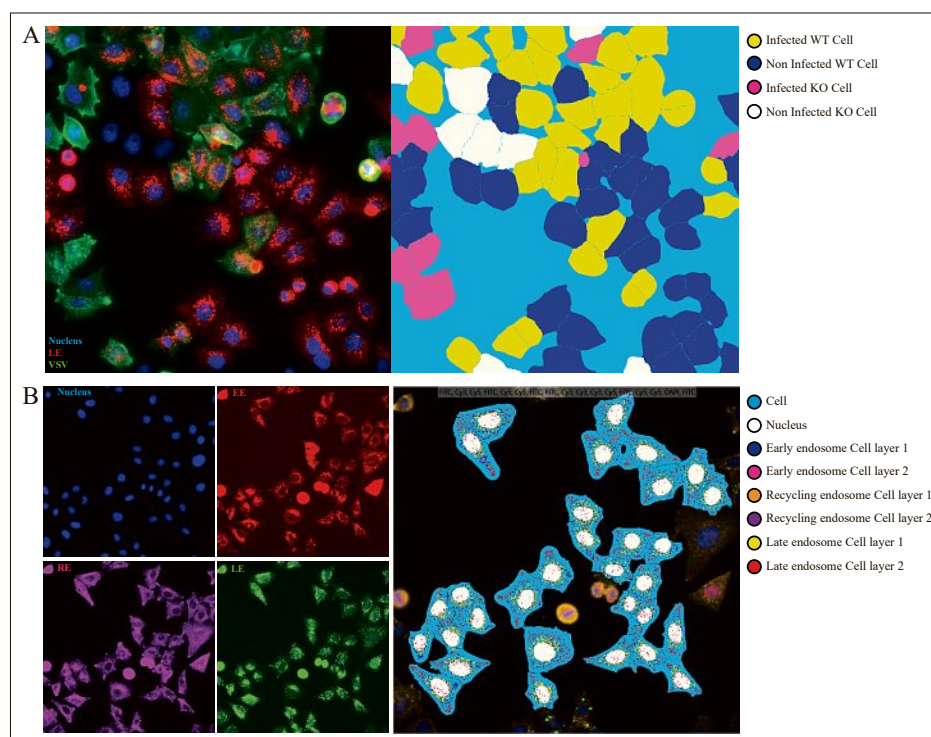


Fig. 2. Examples of image segmentation performed with MetaXpress[®]Custom Module Editor. A. Identification of different cell population based on the presence/absence of fluorescent staining. Masking of KO and WT cells infected or not by fluorescent VSV virus. B. Left: Immunofluorescence images of fixed HeLa cell, acquired with an ImageXpress[®] Micro XLS from Molecular Devices[™], Nucleus in blue, early endosome (EE) in red, recycling endosome (RE) in pink, late endosome (LE) in green. Right: Advance segmentation of early and late endosome in HeLa cells. Generation of a main cell mask subdivided in cell layer the pixel distance from the nucleus. The endosome masks are attributed to a specific region in the cell. Border, dividing and dying cells have been excluded from the final mask, by detecting them or through filtering different channels.

each individual cell is analyzed independently. This may allow the identification of a rare event or phenotype lost in averaged image data, and may help revealing population context effects.^[14]

This image analysis of a full experiment is computer-intensive and may take hours to days of calculation. As mentioned above, it is advisable to secure appropriate IT support and to run the analysis on a core cluster.

Data Analysis

The output of the image analysis results in a large amount of numerical data, which need to be processed to yield the analysis of the phenotype. The true phenotypic analysis begins at the data analysis level, when numbers are describing cells and cellular objects in the images. Not only is it impossible to manually review all images in a screen, but also the human eye may fail to associate some parameters (*e.g.* position, intensity, size) or to separate mildly different phenotypes. The descriptors and numbers generated with the image analysis render each object unique and precisely defined: a proper analysis will allow appropriate data mining and hit selection.

The first step in the data analysis is to establish the quality control of the screen. It is indeed essential to ensure that no experimental artefact perturbs the screening experiment. One needs to check the normal distribution of the data, the consistency between plates and replicates, the good quality of positive and negative controls, possible unexpected toxicity during the experiment, and the danger of intra-plate (*e.g.* edge-effect) or batch artefacts. This step is absolutely critical, because artefacts can mask real hits or generate false positive hits.

As discussed above, HCS cannot focus on a single parameter, and the goal is to provide deep phenotypic analysis and thus to integrate all data in the analysis. Since data are usually in very diverse scales, some normalization is needed. The most commonly used one is the Z-value or control-based normalization, but new methods are constantly being developed.^[8a,15] With the normalized data it becomes possible to carry out the phenotypic analysis in different ways. The unsupervised clustering makes it possible to group in an automated fashion similar phenotypes without prior knowledge (*e.g.* organelle distribution in the cell),^[16] using statistical methods that use all data available to cluster together similar events (*e.g.* principal component analysis^[17]). Alternatively, cellular profiling using machine learning allows the automated detection of specific phenotypes^[18] that are already known (*e.g.* cell-

cycle arrest). In this case, the software is trained to recognize a specific phenotype by the analysis of specifically affected descriptors – a ‘per cell’ analysis of the data is mandatory to ensure proper teaching and detection of the desired phenotype.^[3a,19] The goal of those two methods is different and can depend on the screen. Unsupervised clustering may lead to the discovery of unpredicted phenotypes, different from wild type, and caused by the treatment (Fig. 3), while machine learning leads to the identification of profiles or patterns previously observed with a positive control. Although the two methods can be complementary, deep phenotypic analysis usually implies the discovery of unexpected effects of the treatment and for this the unsupervised clustering would be more relevant.

The subsequent bioinformatics analysis is performed using the output of the phenotypic analysis. Several bioinformatics databases are available and can help understand the biology behind the phenotypic analysis. The compounds or genes clusters need to be investigated using structure activity analysis or biological pathways database (*e.g.* Gene ontology – <http://geneontology.org/>, Kegg pathway – <http://www.genome.jp/kegg/pathway.html>, DAVID – <https://david.ncifcrf.gov/>). The phenotypic analysis can be extremely powerful to drive the identification of compound target, *e.g.* the small molecule U18666A^[20] is able to phenocopy the cholesterol storage disorder Niemann-Pick C.^[21] In a phenotypic analysis, a screen carried out with small molecules can be combined with genetic screens after RNAi, leading to a faster identification of the biological targets,

pathways, or modes of action. Indeed, the phenotype numerical description can be done with such high precision that similar activity of compound or RNAi could be detected with the statistical analysis of numerical data.

Conclusion

Any new user of automated microscopy will realize very rapidly the high discovery potential of the technology. Unlike any precedent microscopy method, HCS makes it possible to obtain independent deep cell phenotypic analysis on a statistically relevant (very) large number of events. It is worth emphasizing that this major advantage of automated microscopy applies not only to screening campaigns using compounds or siRNAs, but also to routine work in research labs. Unbiased and highly statistically relevant data can be obtained for essentially any fluorescence microscopy experiment through the quantitative analysis of a very large number of events by automated microscopy.

While the technology necessary for HCS is available, both at the hardware and software level, it is not straightforward to become an expert and to run a screen. However, screening facilities are being developed by academic institutions and provide the necessary expertise and equipment.

Some may be tempted to believe that it will always be possible to find ‘something’ of interest by testing so many different perturbations. However, if the assay is not developed properly, *e.g.* with a good dynamic range between positive and negative

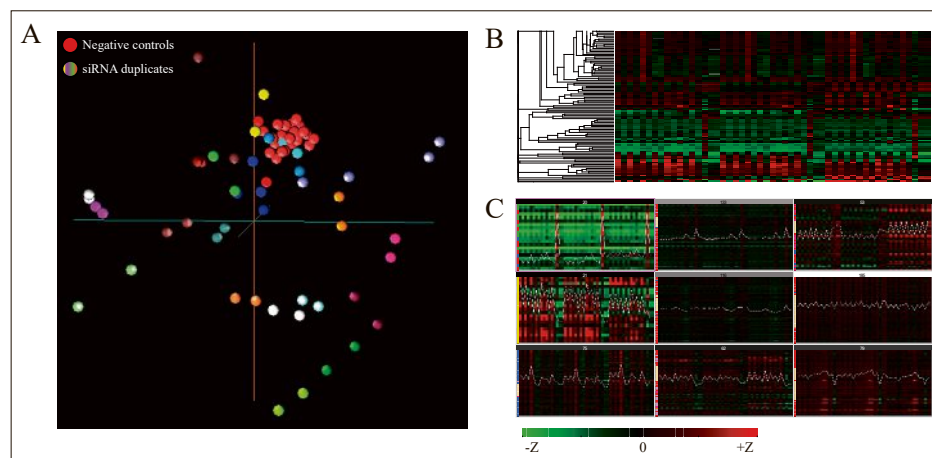


Fig. 3. Examples of phenotypic data analysis performed with AcuityXpress® on image analysis Z normalized data set. Each image is quantified with more than 50 descriptors. A. PCA analysis performed on a targeted siRNA screen. The proximity of the negative control dots and the duplicate siRNA dots confirm the quality of the screen assay and of the image analysis performed. B. Data hierarchical clustering on small molecule screen data. The analysis connects similar phenotypes and gives an estimated distance between them. C. Phenotype clustering by self-organizing map method performed on a small molecule screen. The clustering highlights compound families inducing similar phenotype on tested cells. Each column is a descriptor, each row is a compound test, and the coloring is according to Z-score value.

conditions, the output data may look like a complex cloud of points that cannot be interpreted. Also, not every experiment can be run on a screen format – proper advice before starting can save a lot of time and effort. One should remember that the screen output is always limited by the detection system – an inadequate assay cannot provide adequate data. The quality of the output depends on the quality of the assay, as well as its development and optimization. Trying to save time by skipping necessary validation steps can be very counter-productive and rarely saves time in the end. It is important to keep in mind that a screen campaign is a one shot experiment, and is very costly not only in reagents but mostly in the downstream efforts invested in the subsequent analysis of possible targets.

The second somewhat complex aspect of an HCS screen is the analysis of the data. Thanks to the recent development of several analysis softwares, it is becoming relatively straightforward to obtain a good image analysis followed by a detailed phenotypic analysis. The image analysis tools allow deep image processing and extraction of a large number of features so that the description of the phenotype becomes extremely precise. With good logic and knowledge of the biology behind the assay, it is possible to develop an analysis pipeline in few hours. The analysis of a screen, however, can take months or even years, and may extend with findings coming from the screen itself or with other discoveries. Indeed, the amount of information content in each image is such that it is clearly impossible to extract every relevant information with the first analysis. Therefore, step by step, the analysis will focus on specific new phenotypes, perturbations or compound clusters and a constant update of the image analysis pipeline will be needed. Hence, the knowledge extracted from a ‘good’ image-based screen is clearly incomparable with any other screening or research technique.

At the University of Geneva, with the support of the NCCR Chemical Biology (<http://www.nccr-chembio.ch/>), we have developed the ACCESS-Geneva screen-

ing facility. We have established a fully equipped wet lab able to run any type of screening, with a clear focus on high content screening. We have also developed a complete dedicated and customized IT structure to store, on a secured server, and analyze the data, with dedicated software, which we make available from any computer connected to internet. Today, we have an expertise to drive almost any image based screen project, thanks to a strong experience in assay development and strong knowledge in image and data analysis.

Received: October 11, 2016

- [1] a) S. E. Mohr, J. A. Smith, C. E. Shamu, R. A. Neumuller, N. Perrimon, *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 591; b) T. M. Rana, *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 23.
- [2] a) V. C. Abraham, D. L. Taylor, J. R. Haskins, *Trends Biotechnol.* **2004**, *22*, 15; b) P. Liberali, B. Snijder, L. Pelkmans, *Nat. Rev. Genet.* **2015**, *16*, 18; c) M. Boutros, J. Ahringer, *Nat. Rev. Genet.* **2008**, *9*, 554.
- [3] a) J. Chia, G. Goh, V. Racine, S. Ng, P. Kumar, F. Bard, *Mol. Syst. Biol.* **2012**, *8*, 629; b) N. Y. Chia, Y. S. Chan, B. Feng, X. Lu, Y. L. Orlov, D. Moreau, P. Kumar, L. Yang, J. Jiang, M. S. Lau, M. Huss, B. S. Soh, P. Kraus, P. Li, T. Lufkin, B. Lim, N. D. Clarke, F. Bard, H. H. Ng, *Nature* **2010**, *468*, 316; c) D. Moreau, P. Kumar, S. C. Wang, A. Chaumet, S. Y. Chew, H. Chevalley, F. Bard, *Dev. Cell* **2011**, *21*, 231; d) H. H. Wong, P. Kumar, F. P. Tay, D. Moreau, D. X. Liu, F. Bard, *J. Virol.* **2015**, *89*, 11116; e) L. Pelkmans, E. Fava, H. Grabner, M. Hannus, B. Habermann, E. Krausz, M. Zerial, *Nature* **2005**, *436*, 78; f) A. Karlas, S. Berre, T. Couderc, M. Varjak, P. Braun, M. Meyer, N. Gangneux, L. Karo-Astover, F. Weege, M. Raftery, G. Schonrich, U. Klemm, A. Wurzlbauer, F. Bracher, A. Merits, T. F. Meyer, M. Lecuit, *Nat. Commun.* **2016**, *7*, 11320; g) C. Marugan, R. Torres, M. J. Lallena, *Front. Oncol.* **2015**, *5*, 299; h) P. Liberali, B. Snijder, L. Pelkmans, *Cell* **2014**, *157*, 1473; i) C. Collinet, M. Stoter, C. R. Bradshaw, N. Samusik, J. C. Rink, D. Kenski, B. Habermann, F. Buchholz, R. Henschel, M. S. Mueller, W. E. Nagel, E. Fava, Y. Kalaidzidis, M. Zerial, *Nature* **2010**, *464*, 243.
- [4] G. Pegoraro, T. Misteli, *Methods* **2016**, *96*, 1.
- [5] G. Turcatti, *Comb. Chem. High Through. Screen.* **2014**, *17*, 369.
- [6] D. Clery, *Science* **2014**, *346*, 290.
- [7] K. A. Giuliano, J. R. Haskins, D. L. Taylor, *Assay Drug Dev. Technol.* **2003**, *1*, 565.
- [8] a) B. Antal, A. Chessel, R. E. Carazo Salas, *Genome Biol.* **2015**, *16*, 283; b) F. Zanella, J. B. Lorens, W. Link, *Trends Biotechnol.* **2010**, *28*, 237.
- [9] V. Fetz, H. Prochnow, M. Bronstrup, F. Sasse, *Nat. Prod. Rep.* **2016**, *33*, 655.
- [10] K. W. Eliceiri, M. R. Berthold, I. G. Goldberg, L. Ibanez, B. S. Manjunath, M. E. Martone, R. F. Murphy, H. Peng, A. L. Plant, B. Roysam, N. Stuurman, J. R. Swedlow, P. Tomancak, A. E. Carpenter, *Nat. Methods* **2012**, *9*, 697.
- [11] A. E. Carpenter, *Genome Biol.* **2006**, *7*, R100.
- [12] G. Pau, F. Fuchs, O. Sklyar, M. Boutros, W. Huber, *Bioinformatics* **2010**, *26*, 979.
- [13] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, A. Cardona, *Nat. Methods* **2012**, *9*, 676.
- [14] a) B. Snijder, R. Sacher, P. Ramo, P. Liberali, K. Mench, N. Wolfm, L. Burleigh, C. C. Scott, M. H. Verheije, J. Mercer, S. Moese, T. Heger, K. Theusner, A. Jurgeit, D. Lamparter, G. Balistreri, M. Schelhaas, C. A. De Haan, V. Marjomaki, T. Hyypia, P. J. Rottier, K. Sodeik, M. Marsh, J. Gruenberg, A. Amara, U. Greber, A. Helenius, L. Pelkmans, *Mol. Syst. Biol.* **2012**, *8*, 579; b) B. Snijder, R. Sacher, P. Ramo, E. M. Damm, P. Liberali, L. Pelkmans, *Nature* **2009**, *461*, 520.
- [15] a) J. Rameseder, K. Krismer, Y. Dayma, T. Ehrenberger, M. K. Hwang, E. M. Airolidi, S. R. Floyd, M. B. Yaffe, *J. Biomol. Screen.* **2015**, *20*, 985; b) L. Guyon, C. Lajaunie, F. Fer, R. Bhajun, E. Sulpice, G. Pinna, A. Campalans, J. P. Radicella, P. Rouillier, M. Mary, S. Combe, P. Obeid, J. P. Vert, X. Gidrol, *Sci. Rep.* **2015**, *5*, 14221; c) P. Kumar, G. Goh, S. Wongphayak, D. Moreau, F. Bard, *BMC Bioinformatics* **2013**, *14*, 290.
- [16] W. C. Tjhi, K. K. Lee, T. Hung, I. W. Tsang, Y. S. Ong, F. Bard, V. Racine, *Int. J. Comp. Biol. Drug Des.* **2011**, *4*, 194.
- [17] I. T. Jolliffe, J. Cadima, *Phil. Trans A* **2016**, *374*, 20150202.
- [18] A. L. Tarca, V. J. Carey, X. W. Chen, R. Romero, S. Draghici, *PLoS Comp. Biol.* **2007**, *3*, e116.
- [19] C. Bakal, J. Aach, G. Church, N. Perrimon, *Science* **2007**, *316*, 1753
- [20] L. Liscum, J. R. Faust, *J. Biol. Chem.* **1989**, *264*, 11796.
- [21] T. Kobayashi, M.-H. Beuchat, M. Lindsay, S. Frias, R. D. Palmiter, H. Sakuraba, R. G. Parton, J. Gruenberg, *Nat. Cell Biol.* **1999**, *1*, 113.