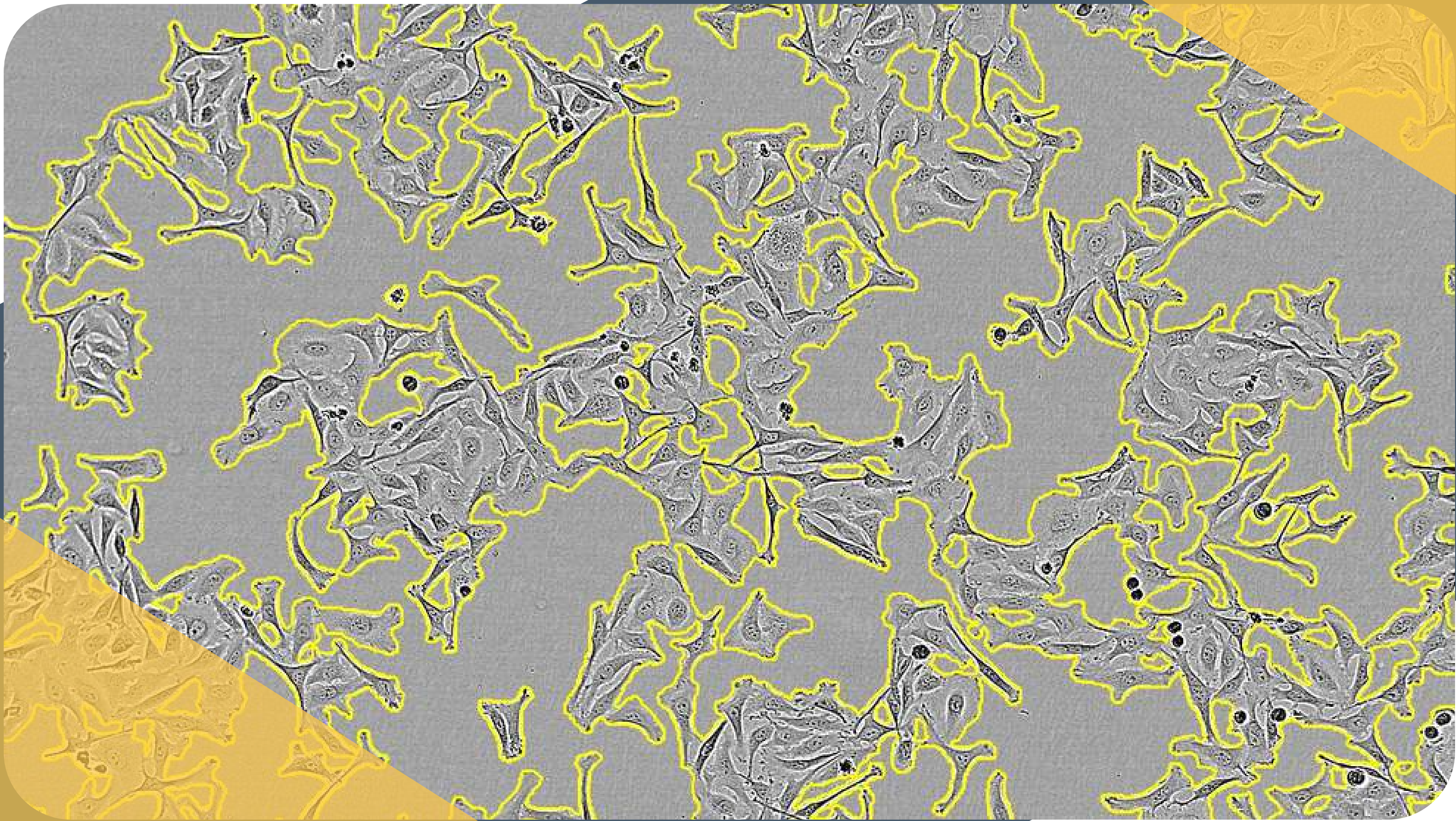
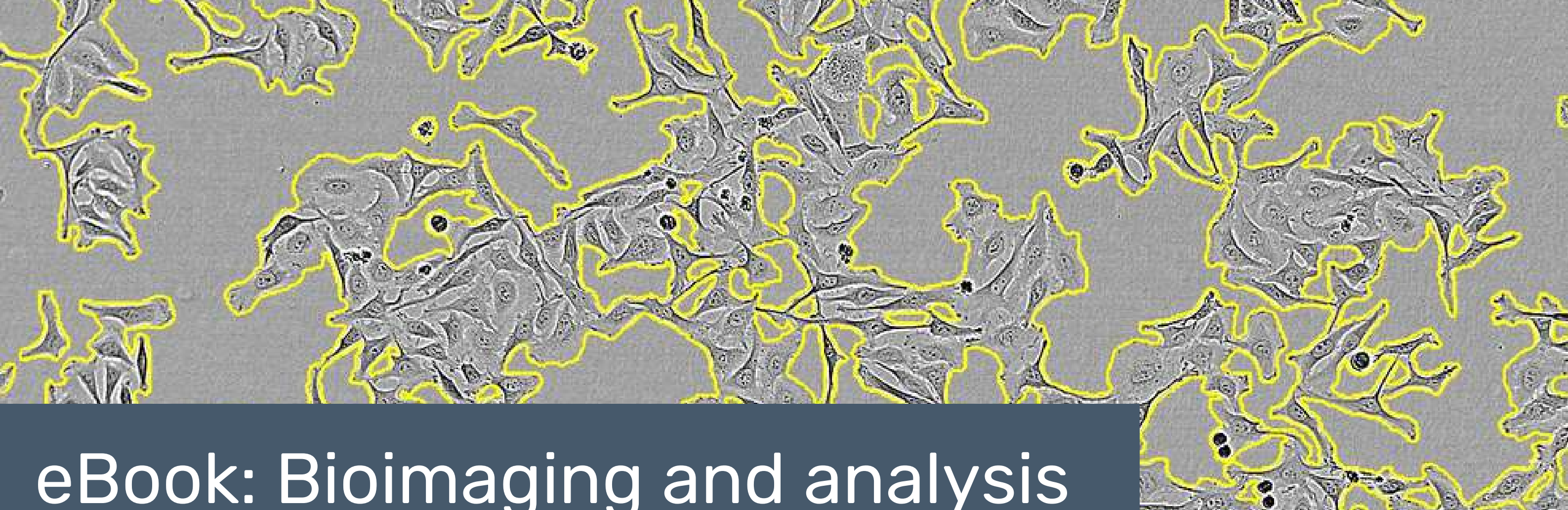


eBook: Bioimaging and analysis

SARTORIUS

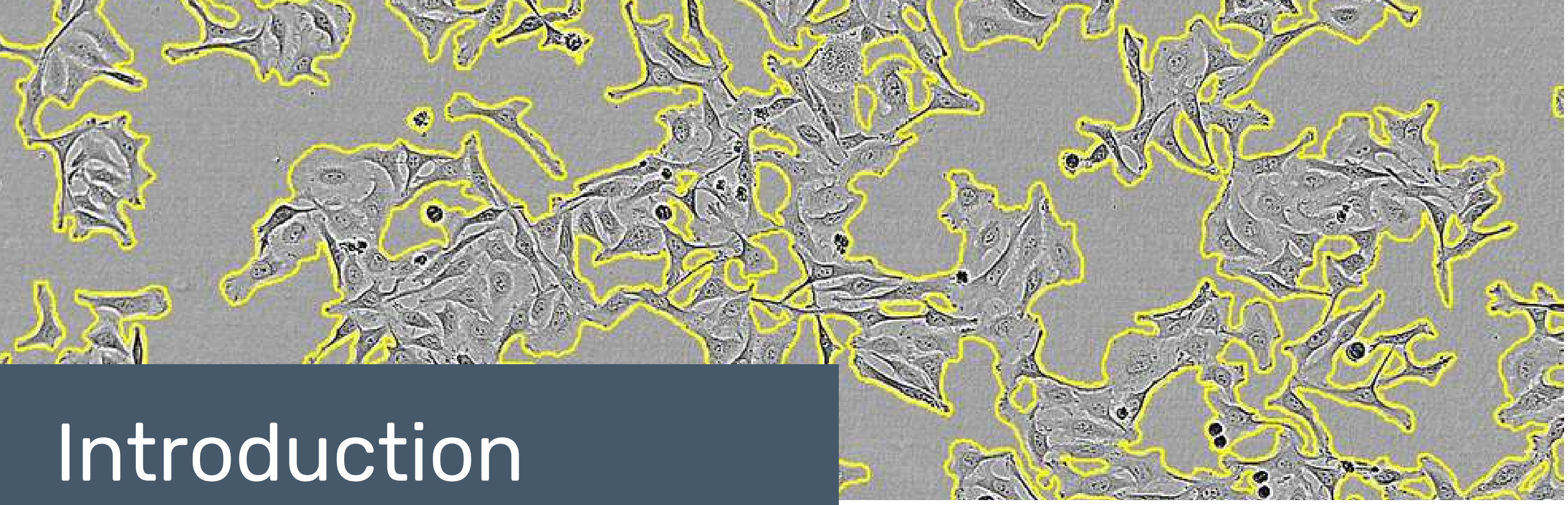


eBook: Bioimaging and analysis

CONTENTS

- Introduction
- **News:** VascuViz: a novel method to uncover the secrets of blood vessels
- **Interview:** Bioimage analysis: has deep learning changed the game?
- **Ask the Experts:** The impact of artificial intelligence in drug discovery
- **Application note:** Advanced label-free classification of cell morphology subpopulations
- **Infographic:** Key trends in image acquisition and analysis
- **White paper**: Live-cell analysis of cell subsets and heterogeneity
- **Panel discussion:** Bioimaging & analysis

SARTORIUS



Introduction

The goal of bioimaging has always been to capture the maximum amount of information about a biological process, with the minimum amount of interference. To achieve this goal there are two key aspects that need to be operating well: imaging instrumentation, to capture high-resolution images in a minimally invasive manner; and the image analysis methods, to extract as much information as possible. When working in tandem, these techniques can act as a powerful tool for therapeutic development and the advancement of personalized medicine.

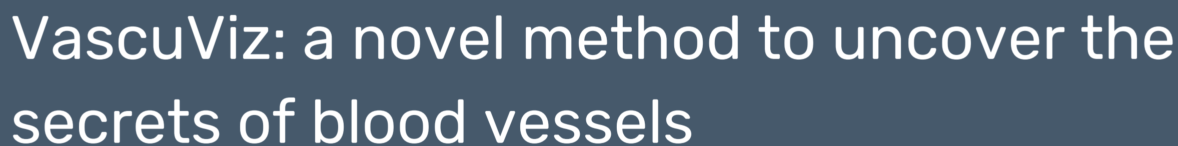
However, the pursuit of more accurate models for disease and basic biology has led to more complex, 3D subjects for image capture and the need to capture processes real-time in live cells has driven the instrumentation to become more sophisticated. These systems are capable of delivering highly complex, detailed images that require cutting-edge analysis models to mine the maximum information from each biological sample being evaluated.

This eBook rounds up some of the key features from our recent Spotlight on bioimaging and analysis, examining the impact of deep learning on bioimage analysis and how AI has changed the drug discovery space, before revealing some of the key trends in bioimage acquisition and analysis.

Discover the opportunities that cutting-edge live-cell imaging technologies have opened into the exploration of cell subsets, heterogeneity and morphology.

Tristan Free
Senior Editor,
BioTechniques
tfree@biotechniques.com

SARTORIUS



VascuViz: a novel method to uncover the secrets of blood vessels

3D images of vasculature systems can now be achieved using a combination of two contrast agents, allowing researchers to perform multiscale imaging of blood vessels.

Researchers at Johns Hopkins Medicine (MD, USA) have developed a novel imaging agent for blood vessels, referred to as 'VascuViz', which is compatible with several imaging techniques, unlike current more restrictive methods. This uses a quick-setting polymer mixture to fill blood vessels prior to imaging and allows researchers to visualize one sample at different scales.

Often researchers will use techniques such as MRI, CT scans or microscopy to capture images of blood vessels and study them. However, different imaging agents are needed to make blood vessels visible for each of these techniques and can often make them invisible to other imaging methods, presenting difficulties for observing macro- and microvasculature structures simultaneously.

"Usually, if you want to gather data on blood vessels in a given tissue and combine it with all of its surrounding context like the structure and the types of cells growing there, you have to re-label the tissue several times, acquire multiple images and piece together the complementary information," explained Arvind Pathak, who leads this research group. "This can be an expensive and time-consuming process that risks destroying the tissue's architecture, precluding our ability to use the combined information in novel ways."

The research group hopes that VascuViz will accelerate imaging-based research as it enables researchers to collect more data from a single sample by using one imaging agent that is applicable to a variety of techniques.

"Now, rather than using an approximation, we can more precisely estimate features like blood flow in actual blood vessels and combine it with complementary information, such as cell density," said Akanksha Bhargava, the lead author on this paper.

Bhargava looked at many combinations of imaging agents that are currently used and tested them with different imaging techniques. Bhargava found that combining a CT contrast agent and a fluorescently labeled MRI contrast agent (BriteVu and Galbumin-Rhodamine) would be suitable for several optical-imaging techniques and make the macro- and microvascular structures visible at the same time.

As VascuViz was successful in test tubes, the research group tested it in different mouse tissues, such as the vascular system of breast-cancer models and kidney tissues. 3D visualizations of the vasculature structure of these were created by combining the images collected using MRI and CT scans and optical microscopy. This approach can be combined with mathematical models or images of other tissue elements to understand diseases with abnormal blood flow, such as cancer and stroke.

VascuViz is especially useful for generating computerized visualizations of complex biological systems, for example the circulatory system, and is a new tool in the growing field of 'image-based' vascular systems biology. The researchers hope this will improve understanding of the structure of tissue dynamics and their response to drug treatments.

Source

1. Bhargava A, Monteagudo B, Kushwaha P, *et al*. VascuViz: a multimodality and multiscale imaging and visualization pipeline for vascular systems biology. *Nat. Methods* 19, 242–254 (2022).

Bioimage analysis: has deep learning changed the game?

We recently spoke to Beth Cimini (right), an analyst at the Broad Institute of MIT and Harvard (MA, USA), about her role at the institute and how bioimage analysis has evolved in recent years due to COVID-19 and the application of deep learning.

Bringing together researchers and labs from both Harvard and MIT, the Broad Institute is a not-for-profit research center, which was initially developed for genomic medicine and has since expanded across many disciplines of biomedical research. Nearly 20 years on from its inception, the Broad Institute focuses on doing biology at scale; for example, the institute currently performs around 1 in 30 COVID tests in the United States. Cimini is part of the imaging platform team, and develops open-access image analysis software using deep learning to streamline analysis leading to better and faster answers to biological questions.

The imaging platform contains two labs working on this mission. The first, led by Anne Carpenter and Shantanu Singh, is focused on turning the data gathered from images into answers to scientific questions using deep learning and informatics. The second lab, which Cimini leads, is focused on making image analysis easier, leading to quicker and more expansive image analysis. One function of the lab is to maintain the Cell Profiler and Cell Profiler Analyst tools, which were created by Carpenter and are freely available tools that streamline bioimage analysis.

"You shouldn't need to know how to do [coding] in order to do good microscopy and good image analysis"

When a researcher has an image to analyze, the first step is to upload it onto an image analysis program such as ImageJ or Fiji and explore the thresholding and filtering options to produce the best quality images for analysis. But what happens when a researcher has 5000 images they'd like to analyze?

Well, currently, they'd need to learn how to code. As Cimini pointed out, "you shouldn't need to know how to do [coding] in order to do good microscopy and good image analysis". Some people wouldn't know where to begin, and many of us just don't have the time to commit to mastering this skill, and that's where Cell Profiler comes into play.

Cell Profiler allows you to string together a 'pipeline' of different image analysis steps called 'modules' without getting anywhere close to a line of code. This could mean taking measurements, finding objects, smoothing an image or highlighting the edge of the nucleus – or all the above. Once you have created a pipeline using Cell Profiler all you need to do is upload the whole set of images into the program and the rest is done for you.

Cell Profiler is a powerful program with between 900 and 950 different image analysis settings; however, a powerful program like this doesn't come without a trade-off, as is often the case. Cimini recognizes that the program could initially come across as overwhelming and appreciates that this may be one of the largest challenges faced by users of Cell Profiler. "I was a Cell Profiler before I worked on the program, so I know how challenging it can be," Cimini explained. However, to combat this the team has developed a thorough guide along with video tutorials to help researchers use this tool.

"I was a Cell Profiler before I worked on the program, so I know how challenging it can be."

While Cell Profiler allows faster analysis of cells and

Bioimage analysis: has deep learning changed the game?

turning images into data, Cell Profiler Analyst enables researchers to take this data set and turn it into tangible answers to questions. For instance, if a researcher measured the cell area in 100 images using Cell Profiler, they would then be able to use Cell Profiler Analyst to visualize this data. Cimini put it nicely, "Cell Profiler gets you the important numbers and then Cell Profiler Analyst allows you to explore that data to answer your questions."

These are powerful tools that are freely available to all researchers, something the Broad Institute feels is essential for a number of reasons. The first is that sometimes there are cases when a researcher wishes to use the Cell Profiler platform to do something that is not currently available on the program. As the code is open-source, updates can be suggested to the team at the Broad, which means the tool doesn't remain stagnant and the technology continues to develop. The second is ensuring that researchers in countries with fewer resources don't have to pay for expensive software licensing, which is not only good for accessibility but will also accelerate the advancement of scientific research.

"We think it makes science move a lot faster and it is a lot fairer"

Open-source image analysis and tools such as Cell Profiler and Cell Profiler Analyst have exploded recently, which Cimini agreed could be down to COVID. She observed that when researchers were forced out of the labs and into their homes, they were finding the time to dig out the data they had been meaning to analyze, and the Broad was able to put time and energy into teaching people how to use their tools through webinars and office hours.

But, Cimini thinks the recent advancements in image analysis software itself were down to something else entirely.

When Cimini first joined the Broad Institute in 2016, the idea that deep learning could be used in bioimaging was one the team believed possible; however, the sticking point was gathering enough data to train the models.

Neural networks are trained by data sets that contain labels. For example, if you are training a neural network to recognize a bus or a plane in an image, you would train it with a data set of pictures labeled either as 'bus' or 'plane'. Each picture would then run through the neural network, which would decide if the picture was more likely to contain a bus or a plane. At this point, the label on the picture would tell the model if it was right or wrong. The model gets smarter by prioritizing the route through the neural network that results in the best predictions of what an image contains. So, if the goal is to train a neural network to isolate more specific features such as nuclei in squamous endothelial cells, a data set with these features and corresponding labels are required.

Industry leaders in deep learning and artificial intelligence, such as Facebook, had enormous data sets created by millions of people tagging their friend's faces for years that could be used to train neural networks. So, for deep learning to be applied to bioimaging, relevant data sets need to be created to train the computer models.

Currently, Cell Profiler doesn't have its own component of deep learning but can incorporate deep learning networks and data sets such as Cellpose and StarDist. As data sets improve, Cimini and the team at Broad Institute hope their tool can come with a neural network that works straight out of the box rather than with the complicated settings and extra plug-ins required now. Currently, Cimini spends time helping researchers understand how to pick the best thresholding algorithm for their work, which she thinks will eventually become redundant, as the tool will be able to do this itself. This will allow more time

Bioimage analysis: has deep learning changed the game?

for the group to work with biologists to help them make Cell Profiler pipelines or create more open-source image analysis tools and workflows.

While deep learning is, as Cimini describes, “undoubtedly valuable” it does not come without limitations. There is no understanding of how a neural network makes certain decisions, which Cimini explained using the example of a deep learning algorithm trained to classify handwritten numbers. Once sorted and classified, Cimini filtered the results to show only the elements of the dataset that the algorithm had assigned as a three. While most of the results were a three, there were a few twos and an eight. “When mistakes are made it’s hard to know why and it is hard to know how to fix them,” says Cimini, and often, the only option is to continue training the network. For this reason, Cimini doesn’t believe that deep learning tools will become the be-all and end-all of bioimage analysis but there is no doubt it will continue to develop further.

When asked what makes working on the Cell Profiler and Cell Profiler Analyst technology so rewarding, Cimini revealed that Cell Profiler is cited in over a thousand papers a year, meaning that there are “at least a thousand people a year who might not have been able to get an answer before.” Cimini adds that, “there is a little piece of science that couldn’t be done and now can”. For example, Cell Profiler has been used to pick which therapeutic drugs would be best suited to patients with leukemia or lymphoma. The interface was able to predict which drugs would work best for a patient and made better judgments about treatment plans than a doctor alone, increasing the life expectancy of patients with cancer. “When you have a huge impact on somebody’s life, that is great, but having thousands of tiny impacts on people’s lives is also really rewarding.”

As deep learning becomes more incorporated into bioimage analysis programs, like Cell Profiler and Cell Profiler Analyst, the speed at which data can be sorted and analyzed will lead to faster science dissemination. While these tools may look daunting at first, those developing and maintaining them have created a range of informative materials and enjoy assisting researchers to find answers to a scientific questions within their images.

Ask the Experts: The impact of artificial intelligence in drug discovery

Artificial intelligence (AI) has become more common, both in our research labs and in our homes, but what are the limitations of AI?

We turn to Anne Carpenter (Broad Institute; MA, USA), Wengong Jin (Eric and Wendy Schmidt Center; MA, USA), and Jürgen Bajorath (University of Bonn; Germany) to answer our questions about developing computational techniques for drug discovery, the challenges of doing so, and how this technology might evolve in the future.

Contents

- What is the difference between AI, machine learning and deep learning? 2
- In what ways can AI be used to accelerate drug discovery?... 3
- What have been some significant advancements or successes of AI in drug discovery?.. 4
- How is AI being used within your own research?.. 5
- What do you think is a common misconception about using AI in drug discovery?... 6
- What are the challenges of developing AI for drug discovery?...................................... 6
- What are the current limitations of using AI in drug discovery?..................................... 7
- How do you think AI will evolve in the next decade to accelerate drug discovery?... 8
- Meet the Experts.. 10

What is the difference between AI, machine learning and deep learning?

Anne Carpenter

These terms can be confusing because some envelop the others, and some have both a technical meaning and an everyday meaning. Simply speaking, in machine learning (ML) you aim to teach a computer to answer questions correctly by providing it with examples (either examples with the correct answers, in supervised learning, or just examples of the data in unsupervised learning). The computer aims to discover general rules rather than just memorizing answers. ML can be trained to answer questions like, “Where are the nuclei in this image?” or “Where are the transcription factor binding sites in this genome sequence?” or “What groups of similar samples exist in this dataset?”

To understand deep learning (DL), it helps to know that in most ML applications to date, specific features were extracted intentionally from the data in the hopes that those features would make it easier for the computer to learn correct answers. For example, we design features in images relating to the texture, shape, and size of cytoplasmic staining to try to detect whether a cell is metastatic or not.

However, DL is a type of ML where instead, you feed the raw data to the computer, usually in huge quantities, and let it sort out how to best extract features from the data in order to make its decisions. The decision-making system has many internal layers, which sparked the name 'deep' learning. In my metastatic cell example, we would just give the system the raw image pixel data and let the system figure out how to distinguish metastatic cells by giving it many examples.

Now, AI: some use the term artificial intelligence to refer to any computer system that can make good decisions, whereas others use it to refer just to a branch of ML where the computer is forced to understand how it can learn generally, rather than being trained just for a specific task.

Jürgen Bajorath

ML is a sub-discipline of AI and DL uses deep neural network (DNN) architectures and is a sub-discipline of ML.

In what ways can AI be used to accelerate drug discovery?

Anne Carpenter

This is such an exciting time to be working at the interface of computer science and drug discovery because there are so many applications! Many individual steps of drug discovery can be accelerated using AI.

For example, you can train systems to predict a given compound's activity in an assay based on its chemical structure or other pre-existing information about the compound. You can train a system to sift through millions of images of cells treated with compounds to identify a favorable phenotype. You can predict the structure of a protein involved in a disease so that compounds can be designed to fit into them better. You can even test millions of chemical structures virtually to assess how they bind with the protein. Systems for these and many more tasks are not perfect, but they can assist experts in making better and faster decisions.

Wengong Jin

Molecular screening is a crucial step in drug discovery, where a chemist puts a library of existing compounds into a biological assay to measure their biological properties, such as potency, toxicity and solubility. The number of chemicals that could be potential drug candidates is estimated to be at least 10^{60}– these are all the molecules that obey Lipinski's rule-of-five for oral bioavailability – creating a major bottleneck in screening for drug candidates.

Standard high-throughput screening facilities in the pharmaceutical industry can only test around 10^{5} compounds per day. It is, therefore, crucial to restrict the size of compound libraries to make the screening time and associated costs feasible. We seek to accelerate and automate drug discovery using AI.

Previous screening efforts in the pharmaceutical industry have generated many datasets of molecules with labeled properties. This allows us to build molecular property models that can predict the properties of a compound without testing it in a wet lab. We can then use these models to virtually screen a much larger collection of molecules at a much faster speed (10^{8} compounds/day) than is possible with current high-throughput screening facilities in a wet lab.

Jürgen Bajorath

The hope is that AI approaches will further expand the currently charted chemical/target space and accelerate discovery paths from targets and novel chemical entities to drug candidates.

AI enterprises engaged in drug discovery already claim such accomplishments on a case-by-case basis. However, a word of caution is advisable since there is typically a gap between claims, promotional efforts (for example, for fundraising), and the scientific reality when it comes to pushing 'new' technologies in drug discovery. In this context, it should be noted that ML has a long history in pharmaceutical research and that DL represents an extension of this framework, rather than a truly novel approach.

What have been some significant advancements or successes of AI in drug discovery?

Anne Carpenter

I serve on the scientific advisory board of a company called Recursion, which uses ML to identify changes in cell morphology that occur when genes associated with disorders are perturbed. The team then screen compounds to identify those that could reverse disease-associated changes. Using computers to analyze images makes these decisions fast and objective. They now have four candidate therapeutics entering clinical trials!

Wengong Jin

In 2020, whilst I was at MIT (MA, USA), we successfully used AI to discover a new antibiotic called Halicin. We did this by training a DNN to become capable of predicting molecules with antibacterial activity. We performed predictions on multiple chemical libraries and discovered Halicin, a compound that is structurally divergent from conventional antibiotics and displays bactericidal activity against a wide spectrum of pathogens including *Mycobacterium tuberculosis* and carbapenem-resistant *Enterobacteriaceae*.

Halicin also effectively treated *Clostridioides difficile* and pan-resistant *Acinetobacter baumannii* infections in mice. This work highlights the utility of DL approaches to expand our antibiotic arsenal through the discovery of structurally distinct antibacterial molecules. This discovery was published in *Cell* and received significant attention because there is an urgent need to discover new antibiotics due to the rapid emergence of antibiotic-resistant bacteria.

Jürgen Bajorath
Currently, AI in drug discovery mostly refers to DL and robotics, while the adaptation of other AI sub-disciplines is still at very early stages. One of the areas where DL has recently made a substantial impact is in computer-aided synthesis planning and prediction, at least at the methodological level. However, many medicinal chemists attest to the fact that these DL-driven advances are yet to be made practically applicable in their day-to-day efforts, aside from raising awareness of what this technology can do. Time and substantial efforts will be required until AI/DL tools measurably impact the practice of drug discovery on a larger scale.

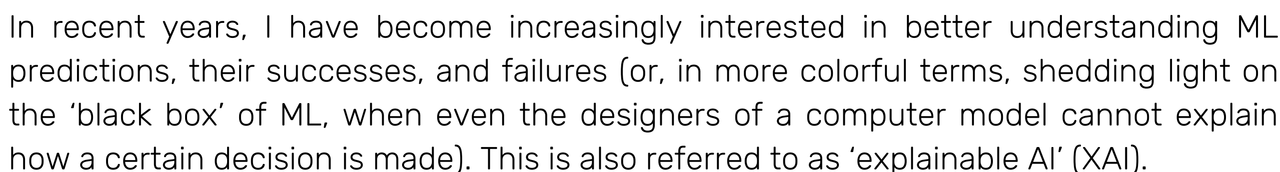
How is AI being used within your own research?

Anne Carpenter
We are teaching the computer to see things that humans cannot see in images. For example, by eye, humans cannot distinguish cells with a certain type of leukemia from those without, so biomarkers were developed that could be detected by fluorescence flow cytometry. We recently used DL to teach the computer to identify those leukemic cells based on just unstained microscopy images, without any biomarker labels, and it succeeded!

Wengong Jin
I am currently using AI to search for synergistic drug combinations to treat COVID-19. Drug combinations make promising therapeutic candidates for COVID-19, but the lack of high-quality training data makes it difficult for DL to predict drug synergy accurately.

To address this challenge, I proposed a novel DL model called ComboNet, which jointly models drug-target interaction and drug synergy. Together with the National Center for Advancing Translational Sciences (MA, USA), we discovered two novel drug combinations (remdesivir and reserpine; remdesivir and IQ-1S) with strong synergy. This work was published in *PNAS* in 2021 and we are currently applying this model to find effective drug combinations for pancreatic cancer.

Jürgen Bajorath
Our research largely focuses on computer-aided medicinal chemistry and chemoinformatics. Like other groups in this area, we have been using ML for molecular property predictions and other applications for many years. Furthermore, we have also developed ML approaches for a number of specific tasks such as predicting activity cliffs (structurally similar compounds that are active against the same target but with large differences in potency) or compound-target screening matrices.



In recent years, I have become increasingly interested in better understanding ML predictions, their successes, and failures (or, in more colorful terms, shedding light on the 'black box' of ML, when even the designers of a computer model cannot explain how a certain decision is made). This is also referred to as 'explainable AI' (XAI).

XAI refers to methods that allows humans to comprehend the results outputted by ML algorithms. Notably, one of the attractions of DL is that DNNs enable us to tackle problems that are difficult, if not impossible, to address using standard ML approaches such as molecular image-based predictions or chemical representation learning. This is another major driver for increasingly investigating DL in our research environment.

What do you think is a common misconception about using AI in drug discovery?

Jürgen Bajorath

Firstly, it is often not sufficiently understood what AI is – and what it is not. We are still far away from a situation where computers make autonomous decisions beyond human reasoning, at least in pharmaceutical research. DL is data-driven, statistical in nature, and far from being some form of 'magic' for unsolved problems in drug discovery, such as high attrition rates.

Secondly, high expectations that AI might 'revolutionize' the drug discovery process are on rather fragile grounds. No single scientific approach or technology has ever come close to revolutionizing drug discovery and there are good reasons to anticipate that this will also apply to AI. Hence, in light of the drug discovery history, arriving at a better general understanding of current AI approaches, their opportunities and limitations, would be beneficial for pharma environments and help to avoid unrealistic expectations.

What are the challenges of developing AI for drug discovery?

Anne Carpenter

It's fairly easy to achieve successful results for a supervised ML problem if you try enough parameters or architectures and test it on only a small sample that is very similar to what you've it trained on. The challenge is to create something that works reliably in the real world, and that takes a serious investment in creating the training and testing data to be sure that you are not fooling yourself with a system that has just memorized the correct answers for a small dataset.

Wengong Jin
The major challenge of developing AI for drug discovery is data scarcity and bias, as training data is usually limited in molecular property prediction, or is otherwise biased. Additionally, molecular assays used for learning property predictors involve many sources of spurious correlations, as a result of the choice of chemical libraries, batch effects, or measurement biases, for example. Therefore, effective molecular property prediction requires that models generalize beyond the chemical space of training examples and avoid learning spurious correlations introduced by these biases.

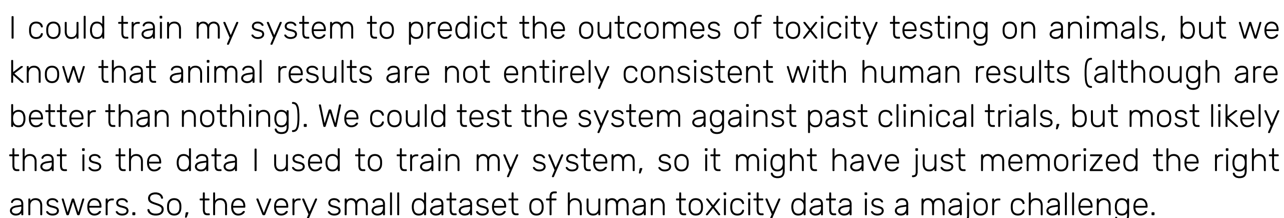
It is also challenging to design proper evaluation protocols to measure the generalization power of a method when applied to a new chemical space, as is common in drug discovery.

Jürgen Bajorath
Unlike other fields where AI/DL has made a strong impact, drug discovery is overall not a data-rich discipline. The use of limited amounts of mostly structured data does not play into the strengths of 'data-hungry' DL approaches. Consequently, consistent improvements of DL predictions over other ML approaches are not expected across typical applications such as compound activity or property predictions and are currently not observed.

In drug discovery settings, it will be important to identify applications where DL is most likely to outperform standard ML approaches (for example, image-based analysis of high-content assays) and concentrate on novel applications that are essentially enabled through DL (such as advanced synthesis design). In addition to data constraints, it should also be taken into consideration that drug discovery is a highly interdisciplinary process with intrinsic scientific heterogeneity, making it rather unlikely that 'one-size-fits-all' AI systems will be easy to conceptualize and implement.

What are some of the current limitations of using AI in drug discovery?

Anne Carpenter
One of the biggest challenges I see is in predicting the toxicity of compounds. Solving this problem would have a HUGE impact on the pharmaceutical industry but it's very challenging to design AI solutions for it. For example, if I invented a new AI-based tool that could tell you whether a given chemical structure would be toxic to humans, how could I prove it works? We can't give lots of different compounds to humans outside of clinical trials, and there are very few trials of new compounds each year to validate my system.



I could train my system to predict the outcomes of toxicity testing on animals, but we know that animal results are not entirely consistent with human results (although are better than nothing). We could test the system against past clinical trials, but most likely that is the data I used to train my system, so it might have just memorized the right answers. So, the very small dataset of human toxicity data is a major challenge.

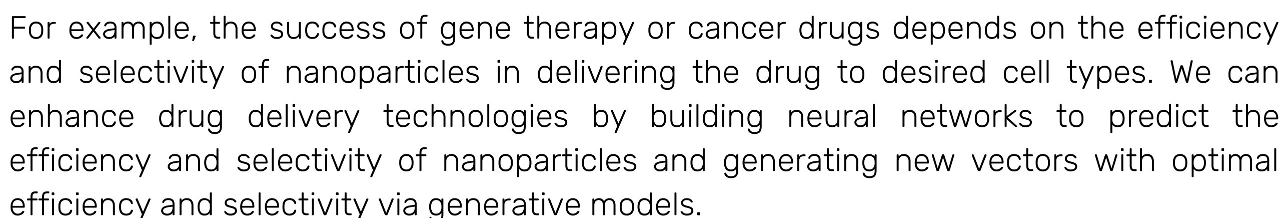
Jürgen Bajorath
In addition to general limitations resulting from data sparseness, the black box character of AI/DL is another important issue. Drug discovery practitioners are typically reluctant to rely on predictions that cannot be understood in chemical or biological terms, which works against the acceptance of black box approaches for practical applications. This emphasizes the need for XAI methods to rationalize predictions and communicate them in an intuitive manner.

Since operating in discovery project teams typically requires multi-tasking and working under time pressure, ease-of-use and robustness of new computational methods and tools are essential for using them in practical applications and for making progress. While developing consistently accurate predictive models is a formidable challenge, transforming expert domain models into widely accessible tools presents another challenge of similar magnitude.

How do you think AI will evolve in the next decade to accelerate drug discovery?

Anne Carpenter
ML will be incorporated more seriously at each step in the pipeline, providing assistance to experts and making their work more efficient. On top of this, I imagine we will see improvements in generative ML systems. So, instead of telling you whether a proposed compound is likely to be effective, this can instead generate a structure from scratch that is predicted to have properties of interest, and even generate a proposed 'recipe' for how to synthesize the compound. The real-world testing of compounds in biological systems will always be a bottleneck and an important step in the process, but it’s exciting to see how much acceleration we can get from computational predictions.

Wengong Jin
I think AI will be applied to a much broader range of biological applications like structural biology, immunology, gene therapy and drug delivery. Therapeutic development in these areas has been hindered by the enormous time and cost associated with experimental processes. AI-based therapeutic design may become the next-generation technology in these fields.



For example, the success of gene therapy or cancer drugs depends on the efficiency and selectivity of nanoparticles in delivering the drug to desired cell types. We can enhance drug delivery technologies by building neural networks to predict the efficiency and selectivity of nanoparticles and generating new vectors with optimal efficiency and selectivity via generative models.

Jürgen Bajorath
For the reasons discussed above, I do not anticipate 'revolutionary' AI-driven developments in drug discovery and design over the next years. Provocatively put, making better drugs through AI probably is an elusive goal for the next decade, given that the discovery process is multi-factorial and much too complex and time-consuming for a single technology to be a game-changer.

Instead, incremental advances in early-phase discovery such as in synthesis prediction, targeted compound design, or in vivo drug property predictions are expected and will certainly be helpful. However, for AI/DL to mature in discovery settings, there is an urgent need for more prospective applications (that is, demonstrating what has been accomplished, rather than what could be done). This will primarily depend on the confidence of drug discovery investigators to translate predictions into experiments.

Practical applications in high-profile discovery projects will be essential for establishing AI within the drug discovery spectrum and increasing its acceptance among experimentalists. It is also anticipated that further progress will be made in integrating predictive modeling with robotics in lab automation. Although this might not always require rocket science, the potential impact of such efforts should not be underestimated, especially if they lead to substantial reductions in the workload required for standard procedures in chemical labs, biological screening, or in the scale-up of experiments.

Last but not least, going beyond DL, it will also be very interesting to see a more extensive deployment of other AI methods and tools such as recommender systems that have the potential to impact the practice of drug discovery.

Meet the Experts

Anne Carpenter is the Senior Director of the imaging platform at the Broad Institute (MA, USA) and co-leads a research group developing computational techniques for use across multiple disease areas. Carpenter's research focuses on developing AI methods for biological image analysis. Carpenter acknowledges funding for her laboratory from the National Institutes of Health (R35 GM122547).

Wengong Jin is a Postdoctoral Associate at the Eric and Wendy Schmidt Center at the Broad Institute (MA, USA). Jin's research looks at developing novel machine learning algorithms for a variety of biological applications, most recently using deep learning to identify synergistic drug combinations for treating COVID-19.

Jürgen Bajorath is a Professor and Chair of Life Science Informatics at the University of Bonn (Germany), where his research focuses on developing computational methods for medicinal chemistry, chemoinformatics and chemical biology. Bajorath also studies structure-activity relationships in drug design, Big Data analytics and the use of AI in the life sciences.

SARTORIUS

Application Note

January 11, 2022

Keywords or phrases:
Label-Free Classification, Morphology, Cytotoxicity, Differentiation, Cell Cycle, Label-Free Analysis

Advanced Label-Free Classification of Cell Morphology Subpopulations

Gillian Lovell[1], Jasmine Trigg[1], Daniel Porto[2], Timothy Jackson[1], Nicola Bevan[1], Timothy Dale[1]
[1] Sartorius UK Ltd., Units 2 & 3 The Quadrant, Newark Close, Royston Hertfordshire SG8 5HL UK
[2] Sartorius Corporation, 300 West Morgan Road, Ann Arbor, MI 48108 USA

Correspondence
Email: AskAScientist@Sartorius.com

Introduction

The morphology of a cell contains vast amounts of information on cell health and differentiation state, and yields insight into cell phenotype. Biologists use this information daily to drive decisions around cell culture conditions and responses during assay development. Traditionally this important information has been described qualitatively or via the use of single metrics as a surrogate for total cell shape. However, these methods are subjective evaluations and can lead to loss of data and a lack of robustness and reproducibility within cell-based assays.

Objective quantification of morphology enables researchers to make data-driven decisions for successful cell culture propagation and experimentation. Morphological data can be used as a kinetic readout to determine compound effects—for example, counting the number of cells with viable versus apoptotic morphology yields a direct measure of cytotoxicity. Furthermore, using label-free image analysis to derive these measurements has the advantage of being completely non-perturbing to cell cultures, ensuring that the data generated is not an artifact of the detection method. This can be vital when using highly sensitive or rare cell types.

The Incucyte® Advanced Label-Free Classification Software Module enables automated quantification of cell morphology by employing multivariate analysis to identify multiple morphological features such as cell area, texture, brightness, and symmetry. These parameters are then used to create an unbiased, meaningful score value that enables cell subpopulations to be classified into two user-identified groups.

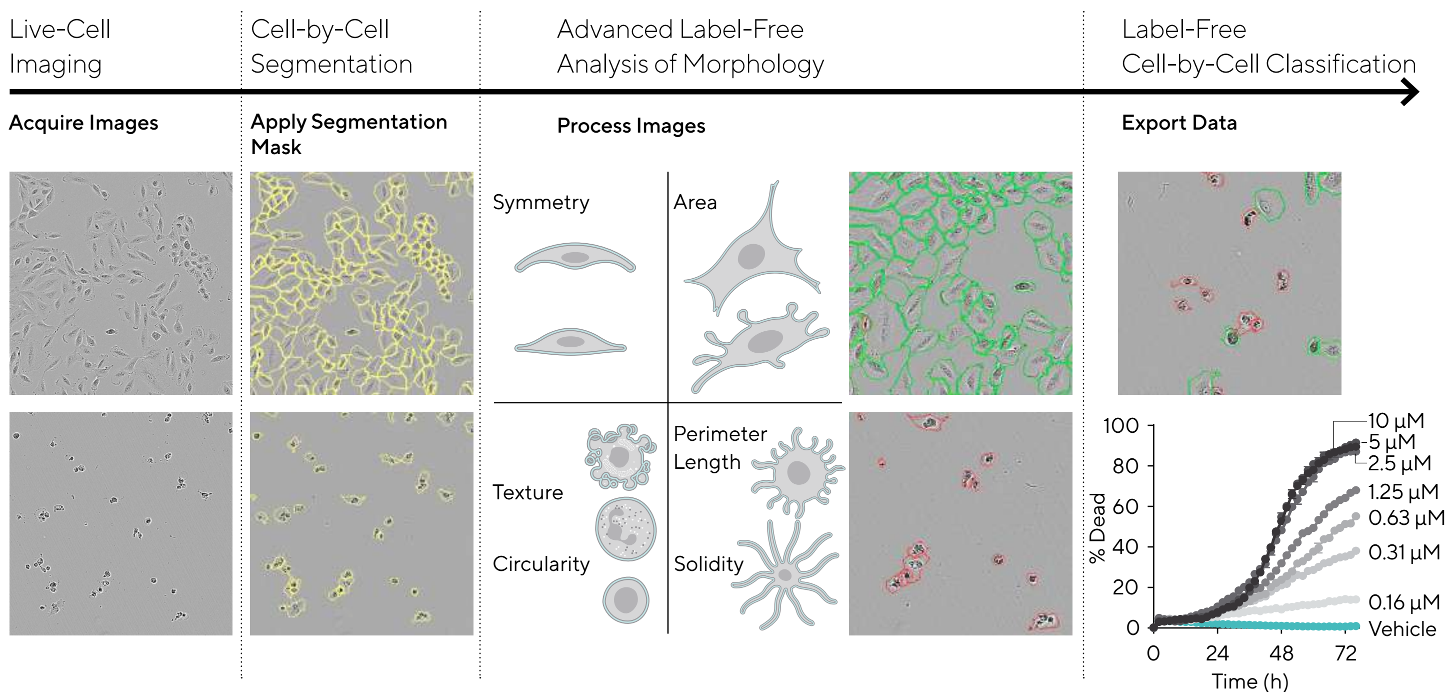
Overview of Incucyte® Advanced Label-Free Classification Analysis

The Incucyte® Advanced Label-Free Classification Analysis Software Module is an add-on to the Incucyte® Cell-by-Cell Analysis Software Module. It enables two classes of cells to be identified by their morphology and quantified over time in kinetic assays.

This workflow is summarized in Figure 1. Images of cells are acquired using the Incucyte® Adherent Cell-by-Cell scan setting, and individual cells are segmented using the integrated software. Advanced classification can then be applied, where the classifier is trained using control images of the two classes of interest. For example, to perform a label-free live | dead assay, the live class is represented using healthy, growing cells at a range of confluence values and the dead class is represented by images of dead cells when a cytotoxic compound has taken effect. Once the classifier has been trained to detect these two morphological classes, it can be applied to any other images containing the same biological model. Integrated software automatically classifies individual cells and the percentage of cells in each class over time can be visualized.

Figure 1

The Incucyte® Advanced Label-Free Classification Analysis Workflow



Note. This workflow can be applied to any use case in which two subpopulations of cells have distinct morphology. For example, mitotic cells can be identified within a culture; undifferentiated monocytes can be distinguished from macrophages.

Method

Mammalian cells have a wide range of different morphologies, which can be characterized in several ways—they vary in size (area, outer perimeter length), shape (aspect ratio, solidity) and texture. Incorporating all these features within the Incucyte® Advanced Label-Free Classification Analysis Software Module, we have employed multivariate analysis that uses over 20 metrics describing different cell attributes. For every cell these metrics are distilled onto a single axis, resulting in a score value between 0 and 1. Dead cells will have a score close to 0 whereas live cells will have a score nearer to 1. A threshold is then applied to group the cells into one of two classes. Where the threshold is set at 0.5, all cells with scores < 0.5 will be classed as 'dead' and those > 0.5 will be classed as 'live.'

Figure 2 demonstrates this classification process. A549 cells were treated with a concentration range of camptothecin to induce cell death in the presence of Incucyte® Annexin V reagent for the purpose of comparing the label-free multivariate response to that of a known apoptosis detection reagent.

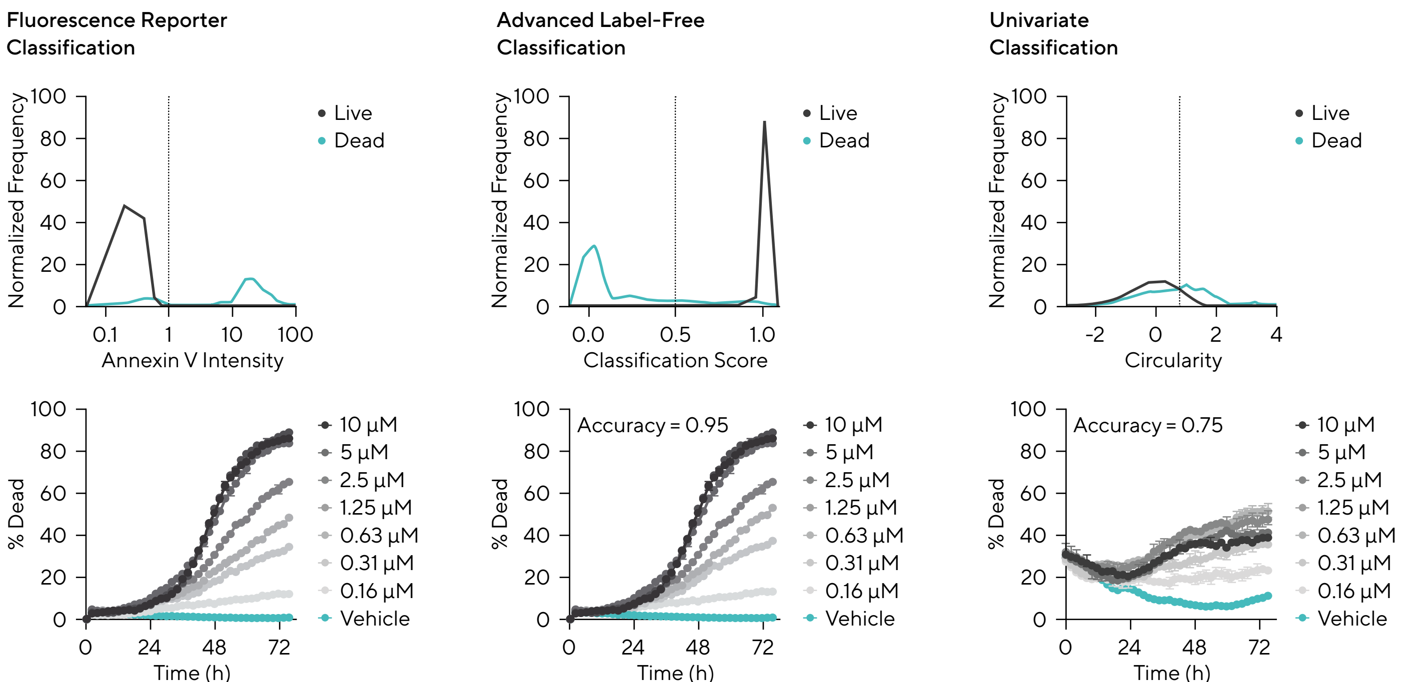
Histograms show the fluorescence intensity, advanced label-free classification score value, or a univariate circularity value for control images of live and dead cells (Figure 2, top row). Both fluorescence and Incucyte® Advanced Label-Free Classification Analysis methods show clear separation between the classes; however the use of the label-free circularity metric on its own results in overlapping populations. Thresholds were used in each case to identify live versus dead cells (indicated by the dashed line on the histogram plots), and the time courses below show the percentage of dead cells per image through time (Figure 2, bottom row).

Fluorescence and Incucyte® Advanced Label-Free Classification Analysis show similar time- and concentration-dependent increases in the percentage of dead cells. While classification based only on the circularity of the cell yields concentration-dependent effects, the time course displays a high percentage of cell death in the untreated (vehicle) cells—an observation that is not reflected upon examining the images of cells.

Incucyte® Advanced Label-Free Classification Analysis and univariate (circularity) classification methods were compared to the standard fluorescence classification method using a confusion matrix. These results confirmed that Incucyte® Advanced Label-Free Classification Analysis is more accurate (accuracy = 0.95) than the label-free univariate method (accuracy = 0.75).

Figure 2

Fluorescent, Incucyte® Advanced Label-Free, and Univariate Classification Analyses



Note. Incucyte® Advanced Label-Free Classification Analysis of live and dead cells yields similar results to the use of fluorescent Incucyte® Annexin V reagent. Univariate analysis is less accurate (relative to fluorescence classification) than multivariate analysis via Incucyte® Advanced Label-Free Classification Analysis.

Applications

This workflow can be adapted to a wide range of applications, including label-free detection of dead cells. In this application note we will demonstrate the use of Incucyte® Advanced Label-Free Classification Analysis Software Module within three biological models: 1) a label-free live | dead assay; 2) detection of mitotic cells within a cell cycle assay; and 3) label-free differentiation and morphological analysis of macrophage subpopulations.

Label-Free Live | Dead Assay

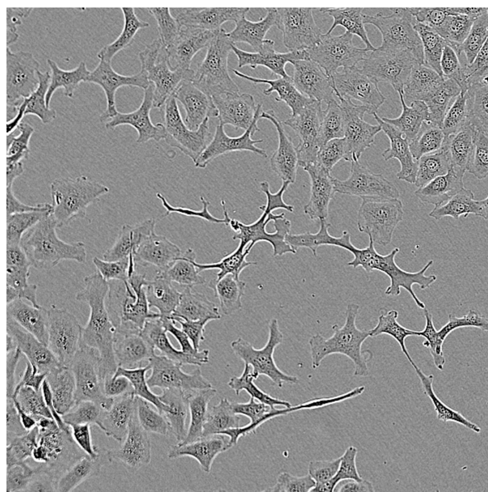
Advanced label-free classification can identify dead cells without the requirement for a fluorescent reagent. Therefore, it is an ideal solution for measurements of cytotoxicity where highly sensitive cell types are used and in cases where fluorescent channels are being dedicated to monitor other biologically relevant events. The label-free live | dead assay was validated against a panel of cancer cell lines with a wide variety of morphologies. Each cell type was treated with a concentration range of camptothecin (CMP), cisplatin (CIS), staurosporine (STP) and nocodazole (NOC) in the presence of Incucyte® Annexin V reagent for comparison. Figure 3 displays the morphology of SKOV3 ovarian cancer cells upon treatment, demonstrating that each compound results in a different morphological change.

Cell death was determined using Incucyte® Advanced Label-Free Classification Analysis Software Module, as well as fluorescence classification to identify annexin V-positive (apoptotic) cells. Phase HD images show that dead cells are visible in CMP and CIS treated conditions, while NOC treatment alters the form of the cell without cytotoxicity. STP induces rapid cell death and the apoptotic bodies are accompanied by a large amount of dead cell debris. These four compounds have different mechanisms of action and yield varied cytotoxic responses. The plate view shows the time course of percentage of dead cells calculated using the Incucyte® Advanced Label-Free Classification Analysis tool. STP induces rapid cell death even at low concentrations, while CMP induces cell death more slowly. In comparison, CIS induces only partial cytotoxicity at the highest concentration tested and NOC, which targets the cytoskeleton, appears to lack any concentration-dependent cytotoxicity.

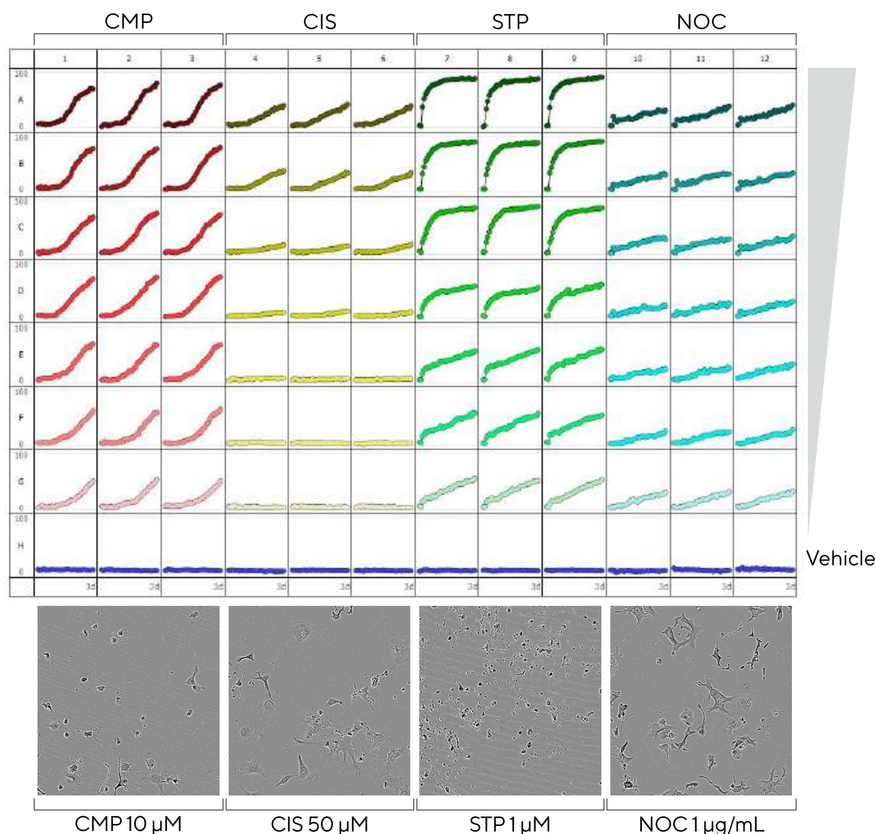
Figure 3

Label-Free Live | Dead Analysis of SKOV3 Cells

Vehicle SKOV3 Cell Morphology, 72 h



Advanced Label-Free Classification % Dead SKOV3 Cells



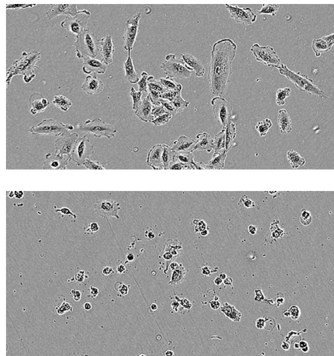
Note. Plate view shows the percentage of dead cells over time calculated using the Incucyte® Advanced Label-Free Classification Analysis tool. Phase HD images show cell morphology at 72 h post-treatment.

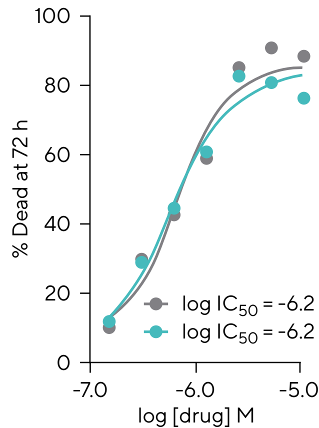
Validation studies revealed that Incucyte® Advanced Label-Free Classification Analysis yielded comparable results to fluorescence classification across a wide range of adherent cell types and compound treatments. Figure 4 shows that EC_{50} values for three cytotoxic compounds (CMP, STP, and CIS) calculated using Incucyte® Advanced Label-Free Classification Analysis were similar to those calculated using fluorescent cell health reagents across multiple cell types.

Figure 4

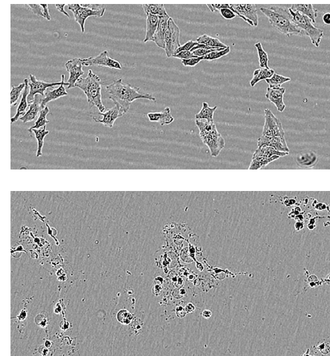
Response Curves for Incucyte Advanced Label-Free Classification Analysis and Fluorescent Reagents

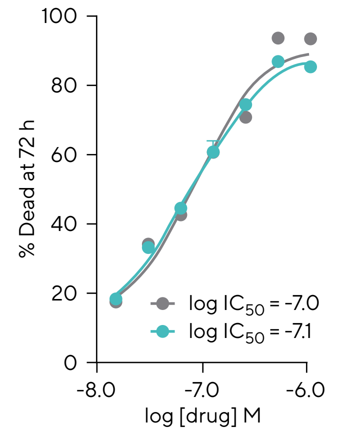
A549 and Camptothecin



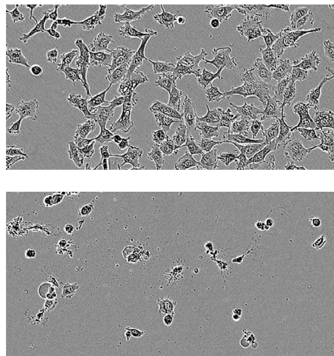


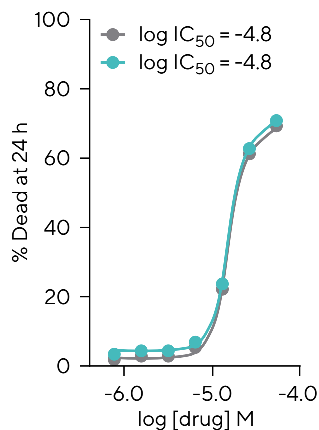
HeLa and Staurosporine



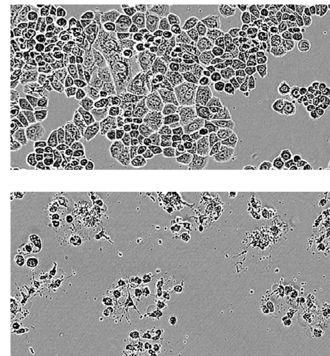


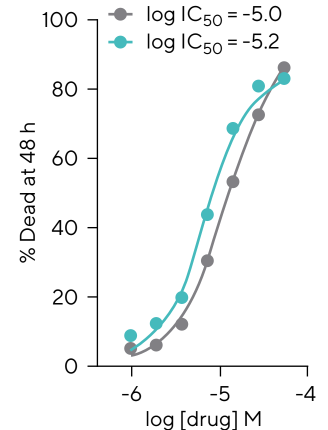
HT1080 and Cisplatin





AU565 and Cisplatin





Note. Diverse cell types display comparable concentration response curves using fluorescent cell health reagents (gray) and Incucyte® Advanced Label-Free Classification Analysis (teal). Pairs of images display untreated (vehicle, top) and treated cells (bottom image, highest concentration).

Label-Free Live | Dead Assay With Cell Cycle Multiplex

Label-free analysis is beneficial in circumstances where the cell type under investigation is highly sensitive, and the use of a cell health reagent is not desirable. It can also add valuable information in situations where the fluorescence channels are reporting other data such as cell cycle phase.

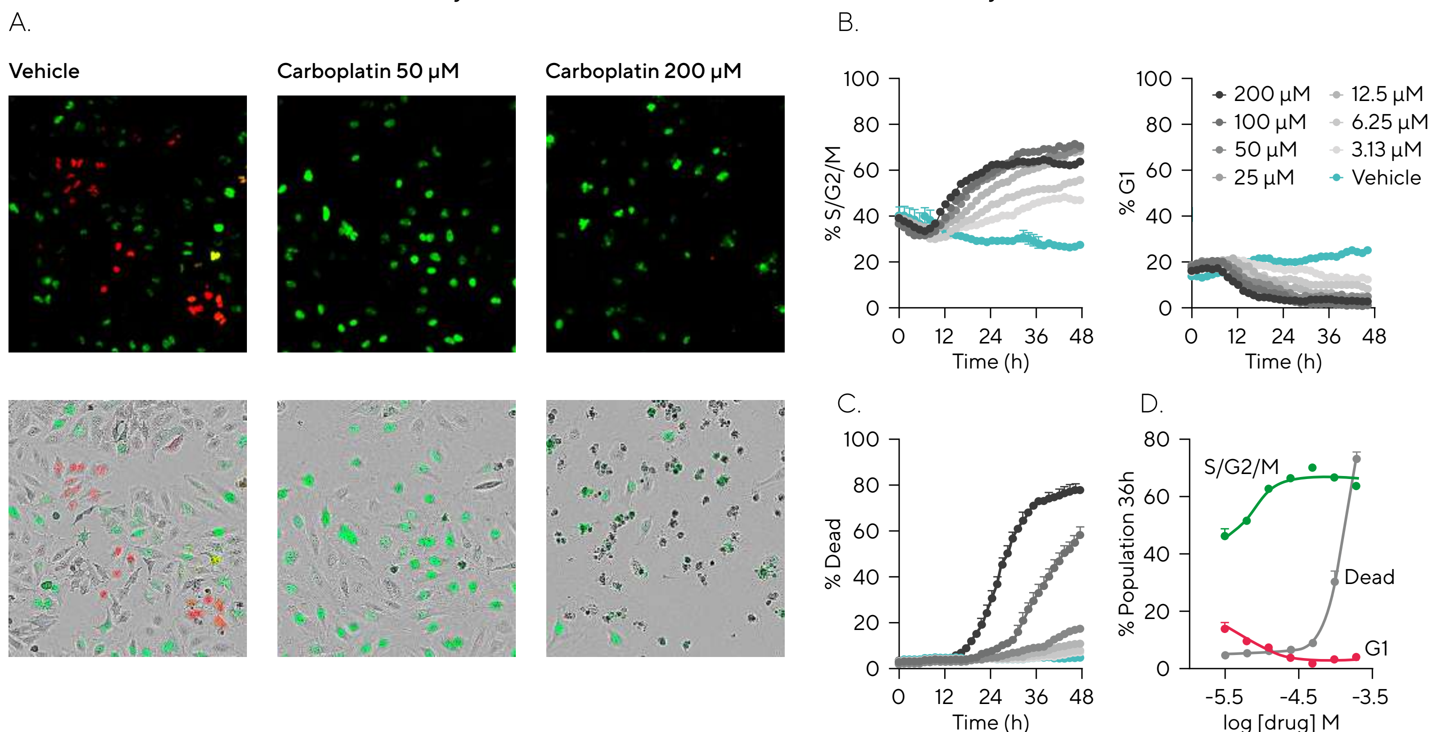
Figure 5 demonstrates the use of Incucyte® Advanced Label-Free Classification Analysis with cells expressing Incucyte® Cell Cycle Lentivirus reagent. These cells express green fluorescence in the S | G2 | M phases of the cell cycle, non-fluorescence in the transition phase M→G1, red or orange fluorescence in G1, and yellow fluorescence (red or orange + green) in the transition phase G1→S. A healthy, growing culture will display a mixture of all four populations as displayed in Figure 5A (vehicle).

HeLa cells stably expressing Incucyte® Cell Cycle Lentivirus reagent were treated with increasing concentrations of carboplatin, a DNA-binding chemotherapeutic that induces cell cycle arrest and apoptosis. Fluorescence images indicate cell cycle arrest at 50 µM and 200 µM, where a high percentage of cells are in S | G2 | M and display green fluorescence (Figure 5A, carboplatin). This observation is reflected in the time course of percent cells in S | G2 | M and G1 phases (Figure 5B).

Phase HD images reveal that in the presence of 50 µM carboplatin, cells have a normal morphology resembling that of the vehicle cells while those treated with 200 µM carboplatin have an apoptotic morphology. Incucyte® Advanced Label-Free Classification Analysis was used to identify live and dead cells, and the time course of percentage of dead cells indicates a cytotoxic effect at the two highest concentrations (Figure 5C). Overlay of the concentration response curves (Figure 5D) indicates the window between maximal cell cycle arrest and induction of apoptosis.

Figure 5

Cell Death Measurements With Incucyte® Advanced Label-Free Classification Analysis



Note. Incucyte® Advanced Label-Free Classification Analysis enables cell death measurements in cells expressing Incucyte® Cell Cycle Lentivirus reagent. A healthy, growing culture displays a mixture of all four cell populations in various phases of the cell cycle (A). After the addition of carboplatin, fluorescent images indicate cell cycle arrest, where a high percentage of cells are in S | G2 | M and display green fluorescence (B). Incucyte® Advanced Label-Free Classification Analysis indicates a cytotoxic effect at the two highest concentrations of carboplatin (C). Overlay of the concentration response curves indicates the window between maximal cell cycle arrest and apoptosis (D).

Label-Free Mitotic Cell Detection With Cell Cycle Multiplex

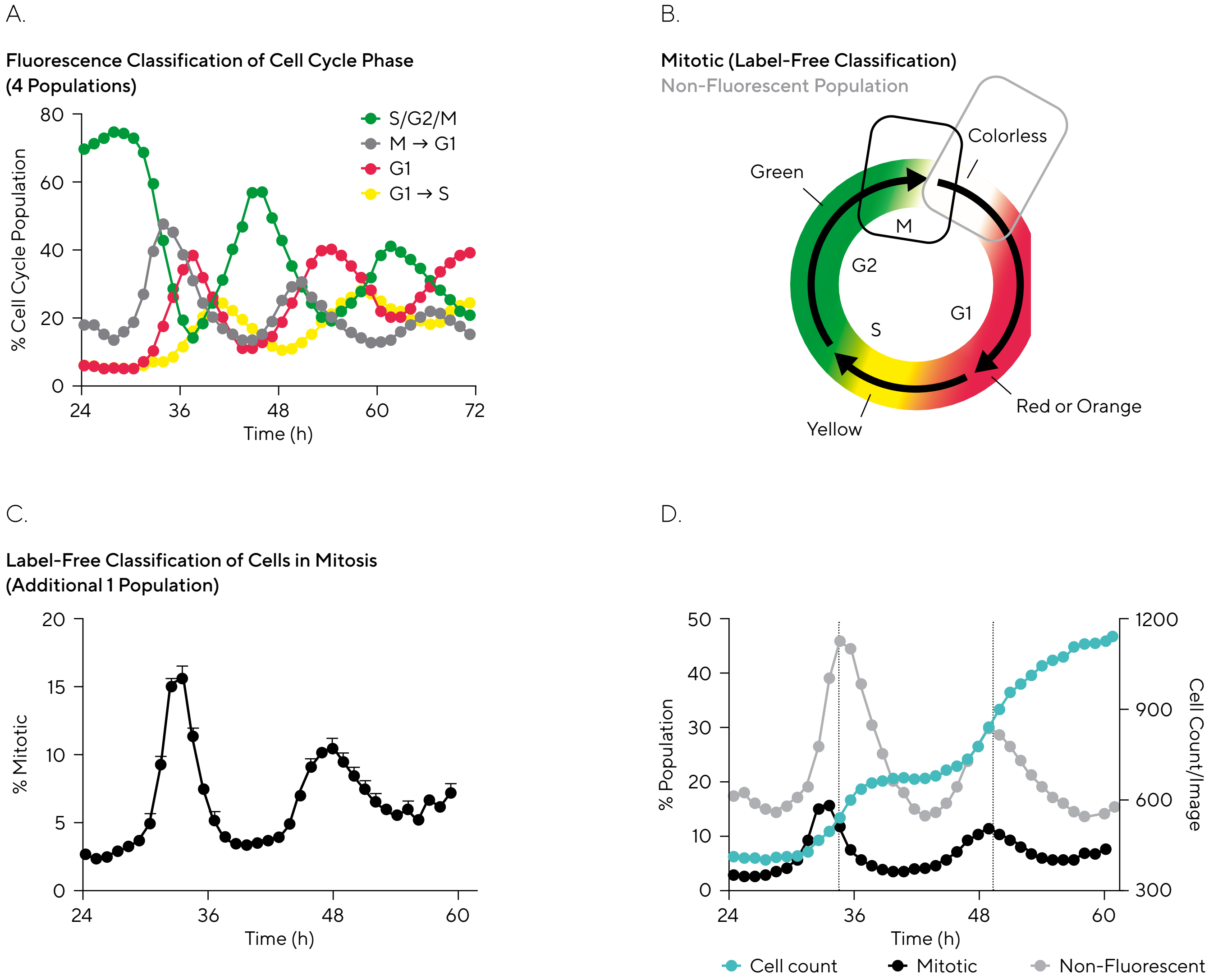
In addition to detection of dead cells, Incucyte® Advanced Label-Free Classification Analysis can be used to identify other morphologies of interest, such as mitotic cells. Using fluorescence classification, the Incucyte® Cell Cycle Lentivirus reagent enables users to detect four distinct populations of cells based on their stage of the cell cycle: S | G2 | M (green), M→G1 (non-fluorescent) transition, G1 (red or orange), G1→S (red or orange + green) transition. With Advanced Label-Free Classification, cells in mitosis can be identified by their unique morphology, providing quantification of a fifth population.

To exemplify this, HeLa cell cycle cells were synchronized using a thymidine block. Cells were treated with thymidine (2.5 mM) for 24 hours until 80% accumulated in S | G2 | M. At 24 hours, the block was removed and the cells progressed synchronously through the cell cycle and began to divide once again. The time course of four fluorescent populations in Figure 6A demonstrates that each population peaks in sequence as the cells move through the cycle.

As the schematic shows (Figure 6B), after S | G2 | M, the cells move through the non-fluorescent M→G1 transition into expressing red or orange fluorescence in G1, and red or orange + green in the G1→S transition. Approximately 9 hours after the removal of the thymidine block, images were observed that contained a high percentage of mitotic cells as indicated by their small, circular and dense morphology. Advanced training set selection enabled this mitotic cell subpopulation to be used to train a classifier to detect mitotic versus non-mitotic cells.

Figure 6

Identification of Five Subpopulations Within a Synchronized HeLa Cell Cycle Culture



Note. Time course of fluorescent populations (A) indicate the changing cell cycle phase (B); time course of mitotic cells (C) are indicated in black. Cell count (teal circles, D) overlays the mitotic (black circles) and non-fluorescent (gray circles) populations.

The time course of mitotic cells in Figure 6C displays the same peaking profile as the fluorescent populations. Overlay of time courses (Figure 6D) confirms that the mitotic population (black) peaks immediately prior to that of the non-fluorescent M -> G1 transition (gray), and that during this time the cell count (teal) experiences a step-wise increase.

Overall, these data demonstrate that the use of fluorescent markers with label-free morphological information enables researchers to extend the biological insight of subpopulations within live cells simultaneously.

Label-Free Differentiation Assay

The activation and differentiation of immune cells is often accompanied by morphological changes. For example, monocytes are a key component of the innate immune system and can differentiate into a number of functional immune cells such as macrophages. Under the influence of pro- or anti-inflammatory cytokines at the site of recruitment, these macrophages can be further activated to M1 or M2 phenotypes.

Figure 7 demonstrates the morphological changes observed upon differentiation of primary human monocytes to M1 and M2 macrophages. While monocytes are small, dense cells the M1 macrophage population is comprised of large, flat amoeboid cells. Incucyte® Advanced Label-free Classification Analysis was employed to distinguish monocytes (Figure 7A, pink segmentation mask) from M1 macrophages (Figure 7A, teal segmentation mask). The time course revealed that differentiation to M1 macrophages occurred over a seven-day period; the differentiation process was non-linear.

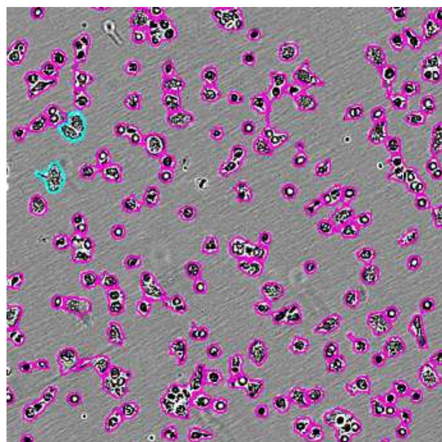
Figure 7

Quantification of Morphological Populations Within Differentiation Assays With Incucyte® Advanced Label-Free Classification Analysis

A.

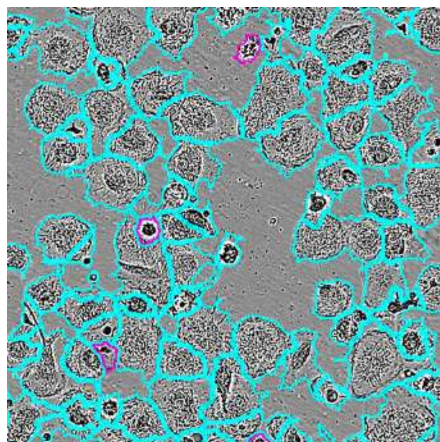
Monocytes, 0 d

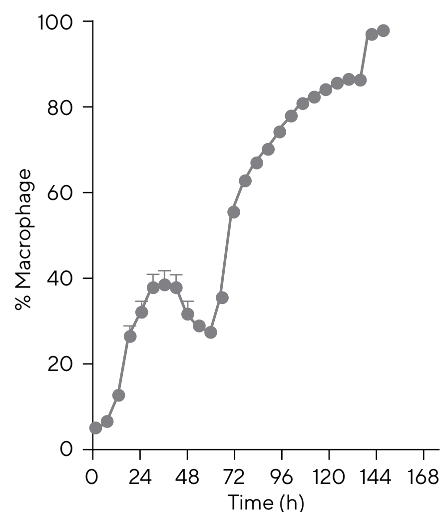
M1 Macrophages Monocytes



M1 Macrophages

M1 Macrophages Monocytes

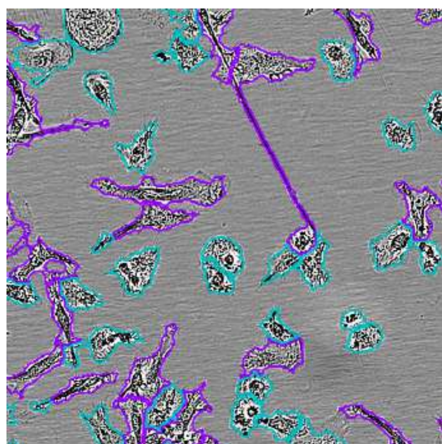


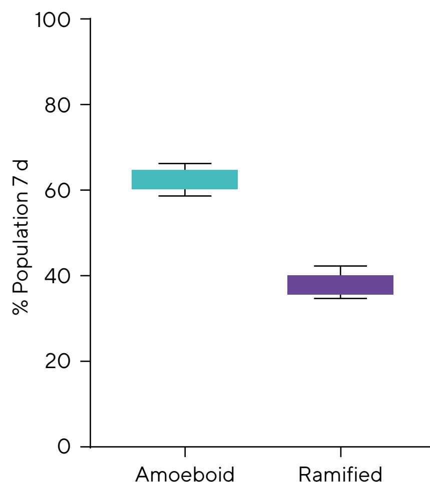


B.

M2 Macrophages

Ramified Amoeboid





Note. Incucyte® Advanced Label-free Classification Analysis was employed to distinguish monocytes (A, pink segmentation mask) from M1 macrophages (A, teal segmentation mask). The M2 macrophage population displayed a mixture of ramified (long, thin) and amoeboid (round, flat) morphologies. Using Advanced Training set selection, these two subpopulations were manually identified and used to train the software. Segmentation color indicates classification results (B). The data showed that within this biological model, 38% of cells were ramified (B, purple segmentation) and 62% cells were amoeboid (B, teal segmentation).

Summary and Conclusion

Using Phase HD images of control wells, the Incucyte® Advanced Label-Free Classification Analysis can be trained following a simple workflow to yield robust and reproducible data without the need for fluorescent reporters. Advanced training set selection allows users to identify cells of interest within an image, enabling enhanced control over training set selection. This software module can be applied to a variety of biological models including a live | dead assay, which can be employed using label-free images or multiplexed with additional fluorescent readouts; identification of mitotic cells adds another population of interest to cell cycle assays. Differentiation of monocytes to macrophages can be quantified without the requirement for fluorescent reporter reagents, and mixed morphologies such as ramified versus amoeboid can be further investigated.

Incucyte® Advanced Label-free Classification is a versatile software module enabling quantification of user-defined morphological subpopulations through time.

North America
Sartorius Corporation
300 West Morgan Road
Ann Arbor, Michigan 48108
USA
Phone +1 734 769 1600

Europe
Sartorius UK Ltd.
Longmead Business Centre
Blenheim Road
Epsom
Surrey, KT19 9QQ
United Kingdom
Phone +44 1763 227400

Asia Pacific
Sartorius Japan K.K.
4th Floor, Daiwa Shinagawa North Bldg.
1-8-11, Kita-Shinagawa 1-chome
Shinagawa-Ku
Tokyo 140-0001
Japan
Phone +81 3 6478 5202

Specifications subject to change without notice. © 2022. All rights reserved. Incucyte and all names of Sartorius products are registered trademarks and the property of Sartorius AG and/or one of its affiliated companies. Advanced-Label-Free-Classification-Analysis-Application-Note-en-L-8000-0794-A00-Sartorius Status: 01 | 2022

BioTechniques Spotlight SARTORIUS

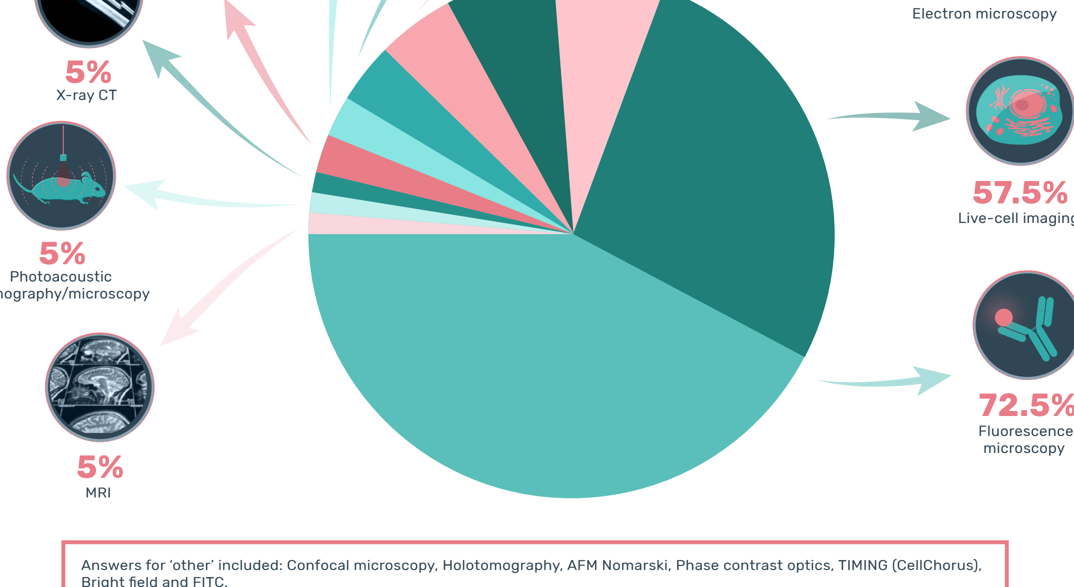
Key trends in image acquisition and analysis

For our spotlight on bioimaging and analysis, we surveyed our audience to find some key trends in image acquisition and analysis

IMAGE ACQUISITION

What imaging technique do you most commonly use? (%)

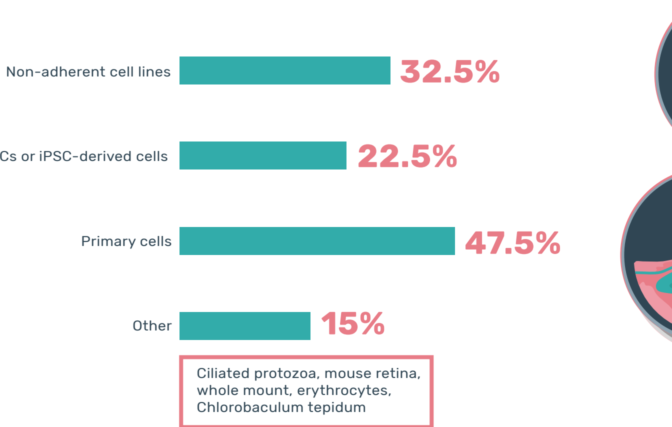
Fluorescence Microscopy was the most popular imaging technique used, followed by **live-cell imaging**



Answers for 'other' included: Confocal microscopy, Holotomography, AFM Nomarski, Phase contrast optics, TIMING (CellChorus), Bright field and FITC.

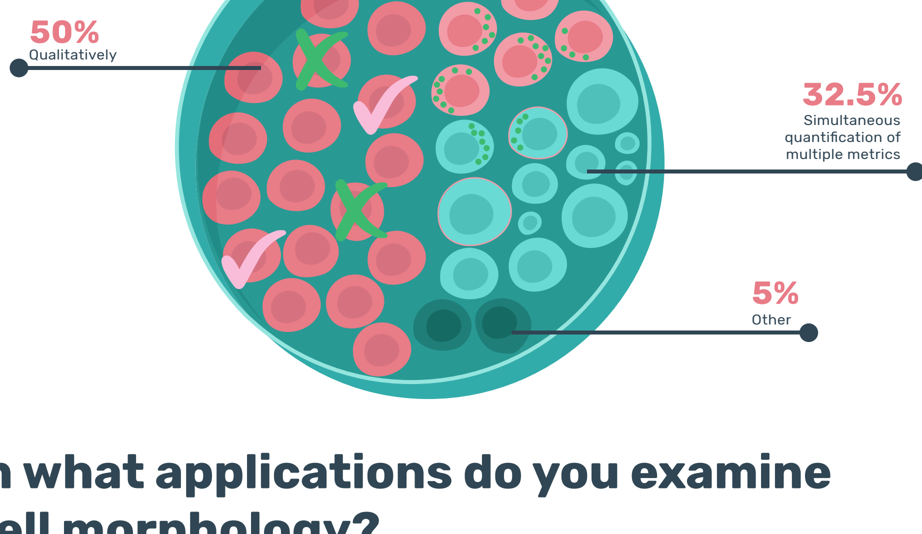
What types of cells do you image?

Adherent cell lines proved the most popular cell type studied



How do you describe cell morphology?

Most researchers analyzed cell morphology **qualitatively**

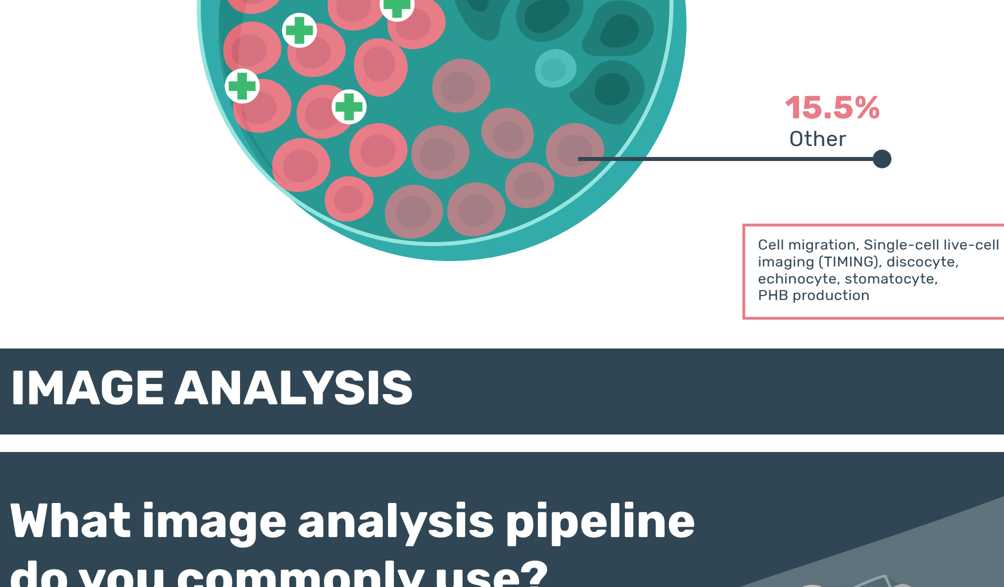


In what applications do you examine cell morphology?

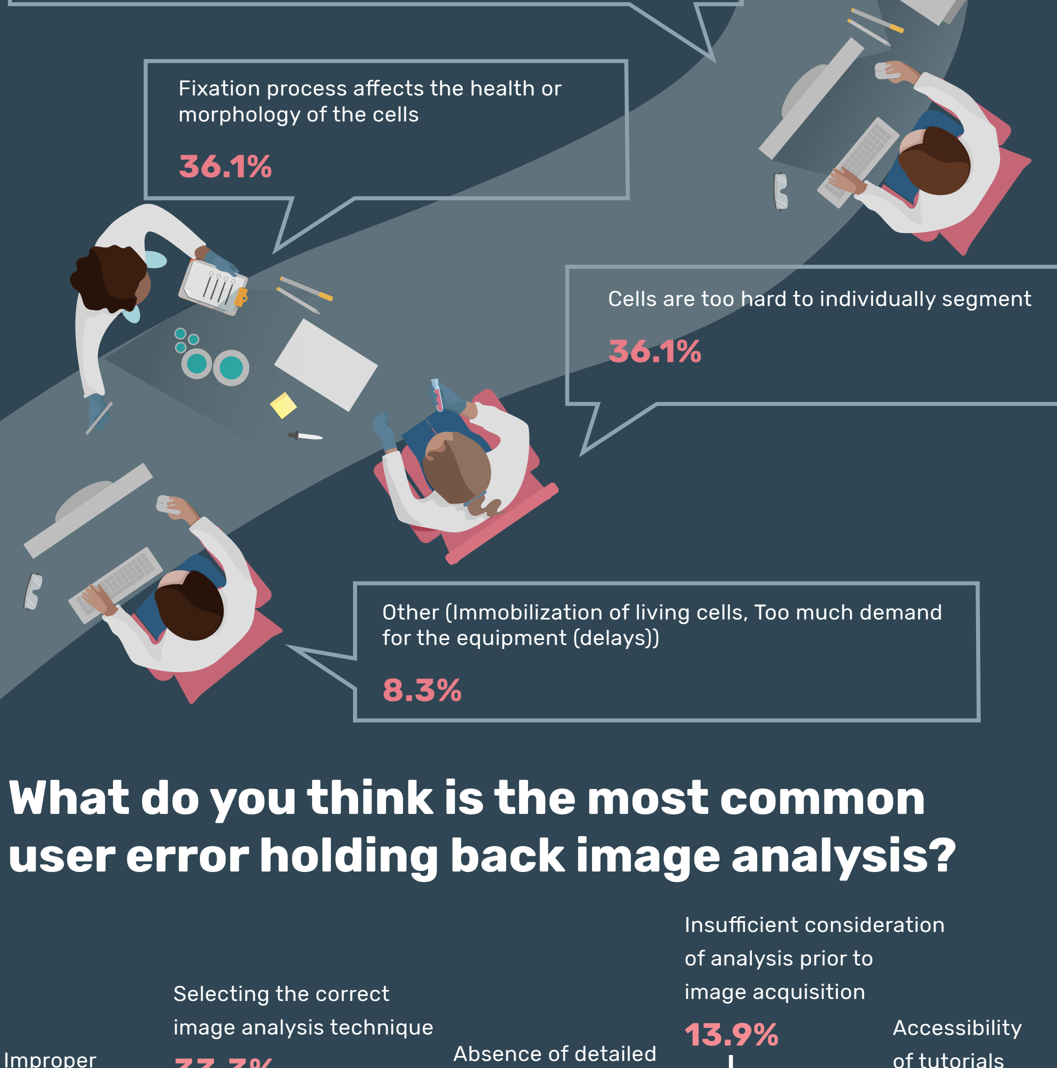
And it was a close split between examining **cell health** and **differentiation**

IMAGE ANALYSIS

What image analysis pipeline do you commonly use?



What is the biggest challenge you face in image analysis?



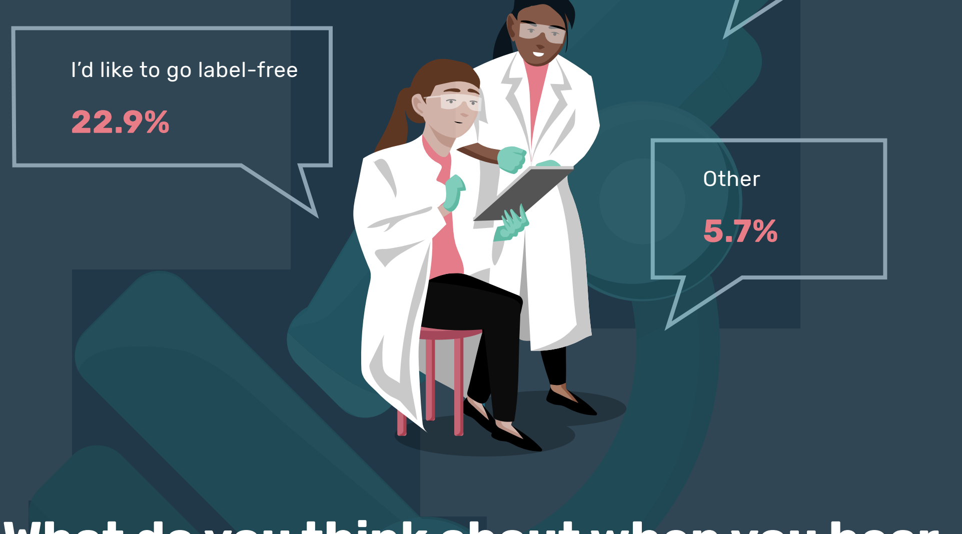
What do you think is the most common user error holding back image analysis?

What are the biggest barriers to accessing machine learning (ML) analysis techniques?

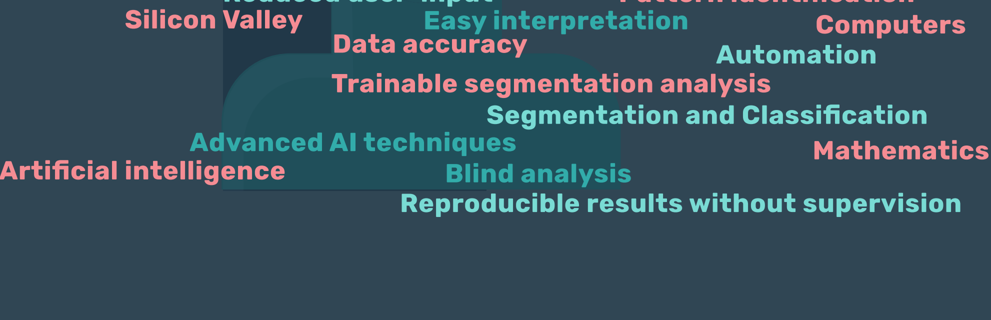
Gaining information and technical skills were the **biggest barrier** to completing successful machine learning (ML) analysis of images.

Recent surveys have highlighted that the most commonly analyzed images are still 2D. Why do you think this may be?

Many lab users are still hesitant about the use of label-free analysis and require more proof before utilizing it in their work



What do you think about when you hear the term "machine learning"?



BioTechniques Spotlight
Learn more: www.sartorius.com/incucyte
SARTORIUS

SARTORIUS

White Paper

July 2018

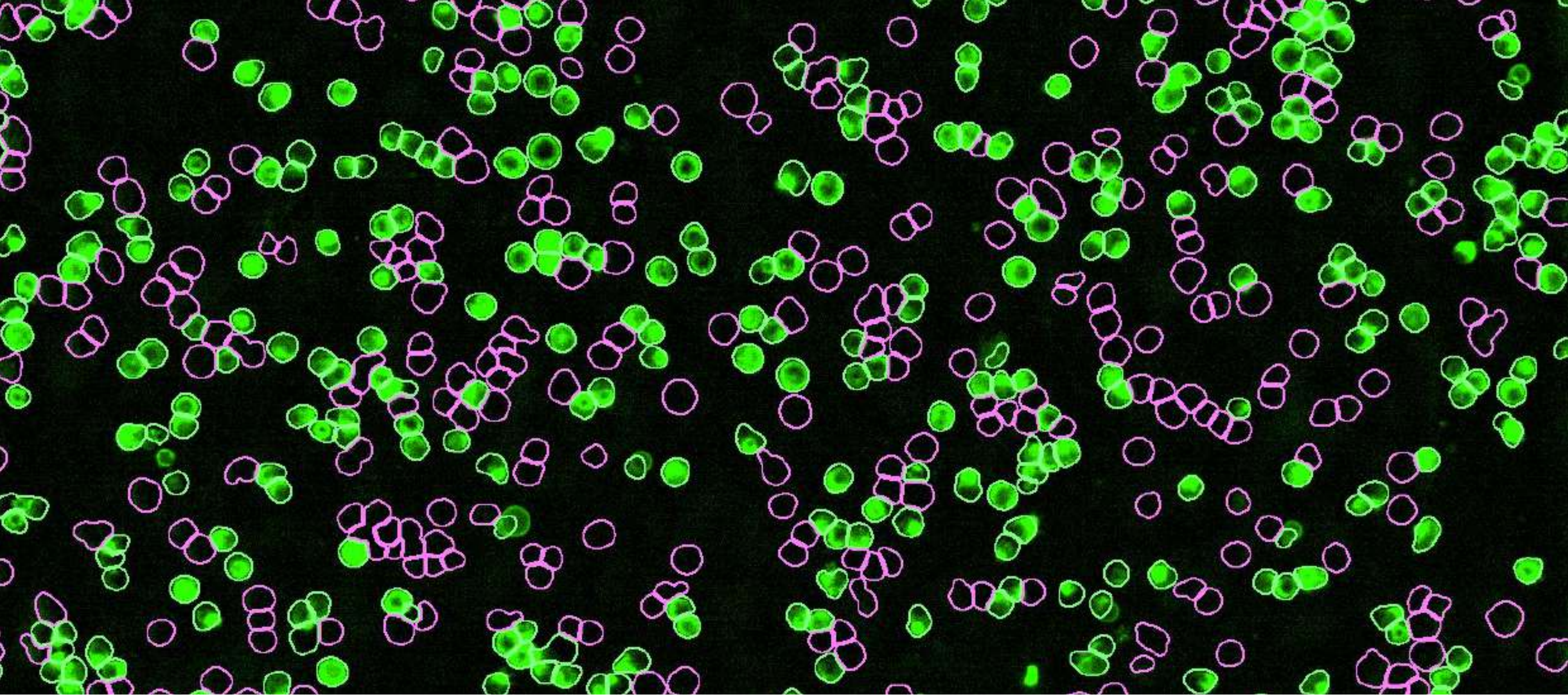
Keywords or phrases:
Subpopulation Analysis, Live-Cell Analysis, Immunocytochemistry, Phenotyping, Immune Cell Activation, Immunology, Immuno-oncology, Cell Surface Markers

Live-Cell Analysis of Cell Subsets and Heterogeneity

Clare Szybut, Nicola Bevan, Hinnah Campwala, Lauren Kelsey, Vincent Blancheteau, Nicholas Dana, Tim Jackson, Nevine Holtz, Eric Endsley, Tim Dale and Del Trezise

Essen BioScience Inc., a Sartorius Company, Ann Arbor, MI USA
Essen BioScience Ltd., a Sartorius Company, Welwyn Garden City UK

Considerable heterogeneity exists in the properties and activity of individual cells, even in the simplest cell system. This arises from fundamental differences in the basic cell types present (e.g., tumor cells and fibroblasts), genetic or epigenetic variations, the stage of cell cycle or differentiation, and the impact of each cells' unique and local dynamic microenvironment. Cellular plasticity and age-related changes add a further, long-term temporal complexity. Such heterogeneity is mirrored by the diversity of pharmacological response at the cellular level, where even seemingly identical cells may respond differently and at different times to drug treatments and perturbagens. Accordingly, analysis at the cell-by-cell level promises valuable and additional biological insight beyond which whole population measures may deliver. In this article, we consider this opportunity from a perspective of analyzing living cells over time, and describe new, enabling and industrial-scale, live-cell analysis solutions for quantifying the phenotypic biology of cell subsets in heterogeneous cultures.



Standard Techniques for Analyzing Cellular Subsets: Strengths and Weaknesses

Flow cytometry is the best characterized and established method for true cell-by-cell analysis. Cells are analyzed one by one in a flow stream for light scattering and fluorescence. When combined with labeled antibodies to cell surface markers and fluorophores, a wide range of multivariate analyses can be applied. Each cell can be classified (binned) into a discrete subset to quantify how many of each cell type are present, and what the associated properties (e.g., cell health, size etc.) of each subset are. Using multiple fluorophore labels and high-end instrumentation, it is possible to create tens of distinct subsets simultaneously in a cell population (e.g., PBMCs). High-content imaging (HCI) methods can also be applied to resolve individual cells for heterogeneity and subset analyses. Typically, nuclear labels (e.g., DAPI) and advanced image analysis algorithms are combined to identify single cells and their boundaries and to extract features on a cell-by-cell basis. Again, subsets can be analyzed separately based on gating criteria. Mostly, this work is done using fix-and-stain protocols to maximize fluorescent signals to background and to permit labeling of intracellular structures (e.g., organelles). The third main approach is time-lapse microscopy. Here, researchers observe living cells under closely controlled environmental conditions, typically with single cell tracking and high spatial and temporal resolution. Unlike the other approaches, this gives critical insight into cell morphology, dynamics, and cell | cell interactions. Together, these three approaches provide an impressive toolkit for analysis at the cell-by-cell level.

Nevertheless, there remains a significant unmet need for additional technical solutions in this area. For both flow cytometry and HCI, the cell preparation and labeling methods cause significant disruption to the cellular environment potentially introducing experimental artefacts. HCI is suitable for analyzing adherent cells, but not non-adherent cells, while for flow cytometry the converse is true. In both cases, the analytical approaches are only amenable to a single time-point measure and, thus, do not easily report population shifts and changes in cell subsets over time. While microscope-based live-cell imaging methods can address these limitations, the identification of subsets (cell labeling) is typically more problematic in living cells, and only small numbers of cells (tens) can be routinely quantified. Thus, conclusions about subsets in the population may be statistically limited. Moreover, environmental control is hard to maintain on microscope stages for extended periods of time (e.g., > 24 h), which prohibits long term studies. Ideally, large numbers of cells (cf. flow cytometry, HCI) could be observed with high spatial resolution (cf. HCI, live-cell imaging) over long periods of time (hours–weeks), using non-perturbing, cell labeling strategies.

Incucyte® Cell-by-Cell Analysis: An Overview

Incucyte® Live-Cell Imaging and Analysis is now a well-established method for quantifying cell behaviors over time at an industrial scale. In brief, the imaging platform resides within a standard cell incubator for full environmental control, automatically capturing and analyzing time-lapse images from up to 2304 assay wells in parallel (6 x 384-well plates). A wide variety of integrated application solutions (software, reagents, protocols) are available including assays for apoptosis, immune cell killing, neurite outgrowth, phagocytosis and 3D tumor growth and viability.

To date, all Incucyte® analysis is based on 'whole-image' measures whereby the information from all objects (cells) in the field of view is consolidated into an average, or aggregate, metric. While this allows for powerful data comparisons from well to well (as would be applied in drug screening assays, for example), it does not reliably inform of cell heterogeneity or readily allow cell subsets to be characterized. Indeed, it is not possible to differentiate a large effect of a treatment on a subset of cells from a small effect on every cell in the population. Small signal changes in subsets of cells are potentially masked or overlooked completely by dilution of the measured signal from non-responding cells. Worse still, an unknowing 'net result zero' may be returned if two subsets respond in differing directions (e.g., Manshian *et al.*, 2015). To address this, we introduce new Incucyte® Cell-by-Cell Analysis Software Module and labeling methods that combine elements of time-lapse microscopy, HCI and flow cytometry approaches into an integrated solution.

A central element of the solution is the identification of individual cells using new image-processing algorithms that segment HD phase-contrast images. This permits: (1) label-free true cell counting, (2) extraction of basic morphological features of individual cells (e.g., size, shape), and (3) measurement of fluorescence intensity that originates from within each cell ('fluorescence within a phase boundary'). Incucyte® live-cell analysis fluorescence reagents can be used including apoptosis probes (Incucyte® Caspase 3/7 Dye or Incucyte® Annexin V Dye), live | dead markers (e.g., Incucyte® Cytotox Green Dye) and green | orange | red | Near-IR fluorescent labels. To enable identification of cell subsets in heterogeneous cultures based on protein surface markers (e.g., CDs), we introduce a new, simple, no-wash Ab-labeling approach for fluorescent tagging of antibodies (Incucyte® Fabfluor-488 Dye), as well as live-cell immunofluorescence protocols that allow non-perturbing, long-term monitoring of protein expression. Finally, using purpose-built, flow cytometry-like Incucyte® software tools, cell populations can then be visualized by either density plots or histograms, and classified into subsets using simple 'gating' thresholds. Importantly, changes in the population and different subsets over time can be explored via the interface and linked back to the raw images with simple color-coding of the classified objects. Together, the tool set enables researchers to easily observe and analyze subsets of living cells over time based on morphological, surface marker, cell health and/or functional properties (Figure 1).

1. Cell segmentation

2. Fluorescence intensity

Features: Phase and Fluorescence

3. View populations over time and classify cells

4. Plot subset time-courses

5. Verify with backgated images

Figure 1: Incucyte® Cell-by-Cell Analysis: concept and workflow

Segmentation Validation and Subset-Labeling

For validation of the cell segmentation method, we directly compared the changes in cell number for proliferating Incucyte® Nuclight Red-labeled Jurkat T lymphocytes using phase-image segmentation and fluorescent nuclei counting (Incucyte® Nuclight Red Lentivirus: Figure 2). Jurkat cells were plated at different densities (5K–40K cells per well) on poly-L-ornithine coated 96-well plates and scanned every 2 h for 96 h with an Incucyte® Live-Cell Analysis System (20X magnification). Throughout the duration of the experiment, the label free and fluorescent nuclei cell count values tracked very closely, indicative of a robust segmentation algorithm and cell counting method. Similar observations were made in a range of other non-adherent cell types including Raji, Ramos, and primary PBMCs.

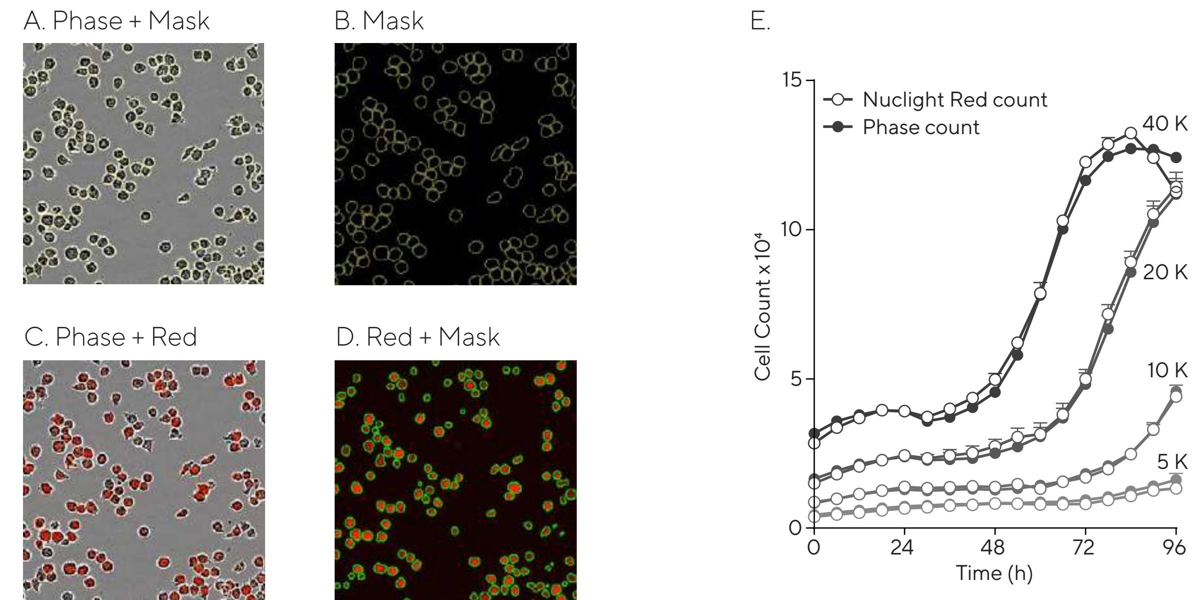


Figure 2: Label-free cell counting: validation of cell segmentation method. Incucyte® Live-Cell Analysis System (20X) images of Incucyte® Nuclight Red-labeled Jurkat cells: (A) HD Phase + Phase Mask. (B) Phase Mask only. (C) Phase + Red Fluorescence. (D) Phase Mask + Red Fluorescence. Scale bar = 50 µm. (E) Time-course analysis of label free cell count (open symbols) and nuclear count (closed symbols) at different initial cell plating densities (5K-40K, 96-well plate). Note the close similarity of values obtained by the phase object and nuclear label counting methods.

To illustrate the live-cell subset identification approach, specific antibodies to the leucocyte common antigen, CD45 and the B-lymphocyte specific antigen CD20, were labeled with Incucyte® Fabfluor-488 Dye (Figures 3 and 4). The Fabfluor | antibody conjugates were then directly added to mixed cultures of Jurkat and Ramos B cells in full cell culture media. Incucyte® Opti-Green, a background fluorescence suppressing reagent, was included to minimize non-specific fluorescence from unbound Fab | antibody complexes. Images were analyzed using Incucyte® Cell-by-Cell Analysis, then gated for CD45 positive and CD20 positive fluorescence (separate wells). Importantly, the optimal gate positions at different time points of the experiment were sufficiently close as to be the same, such that time-dependent gating was not required. In line with expectations, > 95% of cells were labeled positive for CD45, irrespective of the proportion of Jurkat or Ramos cells added to the mix. CD20 positive cells were only observed in Ramos containing cultures at the proportions expected (Figure 3).

In the continued presence of the Fab | antibody, an increase in the number of fluorescently labeled cells was observed over 48 h as the cells proliferated. Interestingly, the relative proportion of CD20 positive cells increased over this period, indicating a faster growth rate of the Ramos cells as compared to the Jurkats. This simple proof of concept experiment demonstrates the ability to specifically label and quantify subsets of cells in mixed cultures and to subsequently track long-term changes in these subsets over time.

As a wider illustration of the applicability of the method for different CD markers, primary human PBMCs were labeled with seven different anti-CD marker antibodies, each coupled to Fabfluor-488 (Figure 5). For each of three individual donors, the proportion of cells identified as marker positive by Incucyte® Cell-by-Cell Analysis correlated extremely well (R2 > 0.95) to that obtained by conventional flow cytometry analysis.

Quick Guide

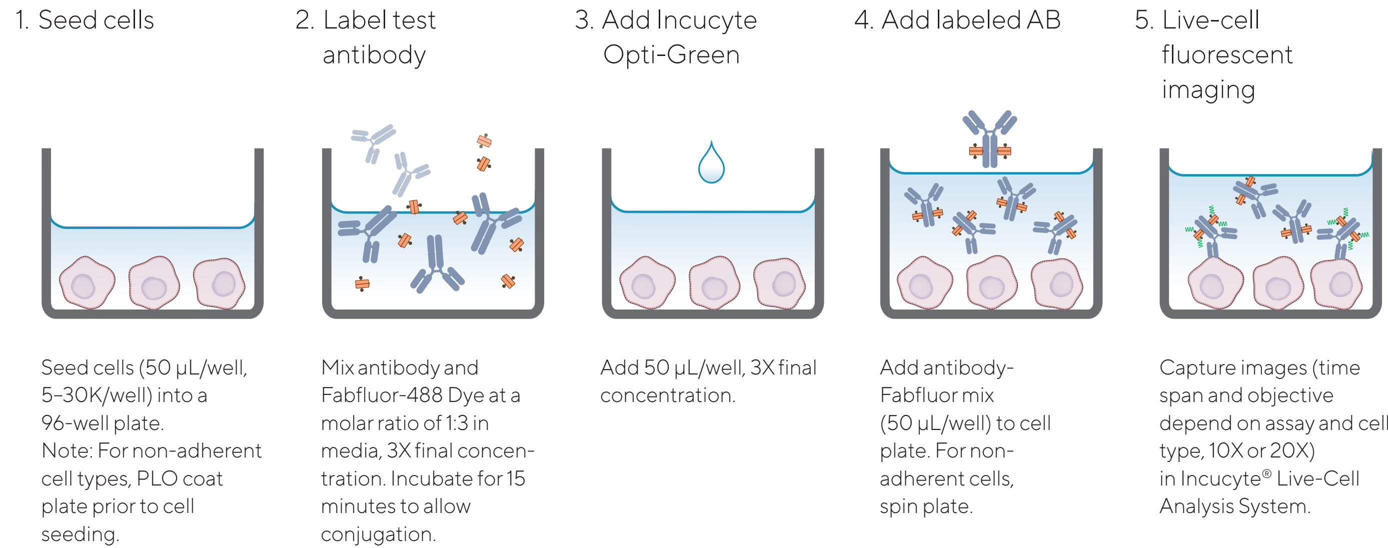


Figure 3: Single step antibody labeling (Incucyte® Fabfluor-488) and live-cell analysis protocol for quantifying cell subsets based on surface markers: concept and workflow.

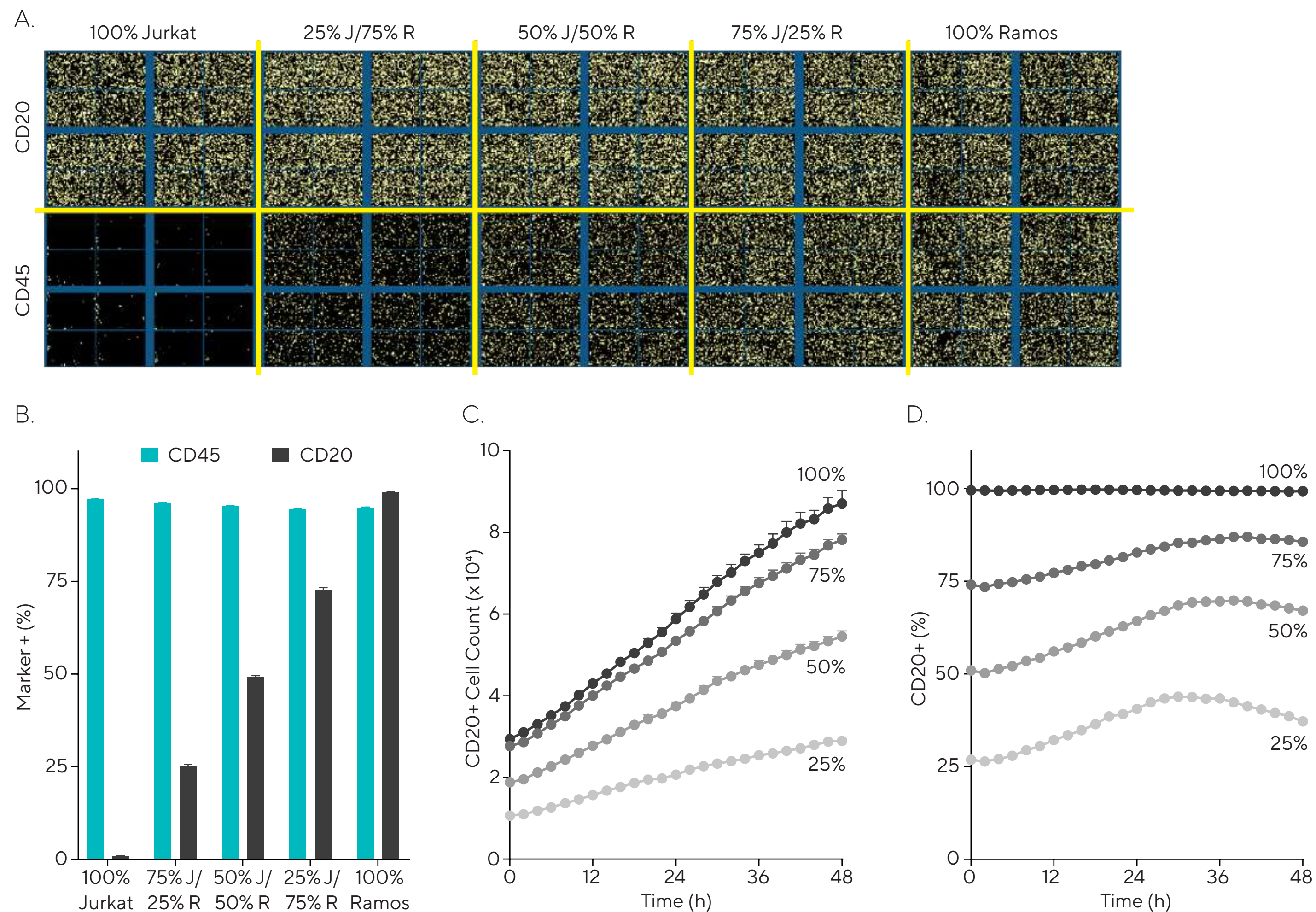


Figure 4: Identification of cells in heterogeneous (mixed) cultures using cell-by-cell analysis: Incucyte® Fabfluor-488 coupled anti-CD20 and anti-CD45 labeling of Jurkat and Ramos cells. Fabfluor-labeled antibodies were added to the cultures post-plating (A) Incucyte® Vessel View images (Yellow = antibody-labeled cell) of mixed cell populations in the ratios shown (J=Jurkat, R=Ramos). Note the greater proportion of CD20-labeled (Ramos) cells as the ratio of R:J increases. CD45 labels both cell types. (B) Incucyte® Cell-by-Cell Analysis was used for quantification of % expression in the mixed culture. (C, D) Time-courses of CD20+ cell count and % of population. Note the proliferation (increase in cell count) of CD20 cells, and time-dependent increase in proportion of CD20+ cells within the mixed culture. Values shown are the mean ± S.E.M. of 4 wells.

Subset and Heterogeneity Application Examples

1. Subset Classification Based on Morphology: Activation and Shape Change in Human PBMCs

T cell activation is an antigen-dependent process leading to proliferation and differentiation of naïve T cells into effector cells. Upregulation of specific cell surface markers, including CD71 (transferrin receptor), is associated with activation as is an increase in cell size and shape change. To observe these effects at the cellular level and correlate surface marker protein dynamics with morphology, human PBMCs were stimulated with anti-CD3 | IL-2 (10 ng/mL) or vehicle and monitored over time (120 h) in the presence of Fabfluor-488-labeled anti-CD71. Phase and fluorescence images were analyzed cell-by-cell to extract information regarding the size distribution, eccentricity (an index of shape) and expression of CD71.

In control, vehicle treated wells, the average cell area and eccentricity of the entire population at t = 0 h was 81 ± 0.5 μm^2 and 0.57 ± 0.002, respectively and remained relatively constant over 120 h in culture. In contrast, following treatment with anti-CD3 | IL-2, the average cell area and eccentricity increased markedly to 117 ± 4 μm^2 and 0.69 ± 0.004 at 120 h, respectively. The shape change was rapid and preceded the size increase by > 24 h. There was greater heterogeneity in the individual cell morphological parameters of activated cells at the later compared to the earlier time points, indicating that not all cells responded identically (Figure 6). Following activation, there was a marked, rapid time-dependent increase in the fraction (%) of cells that were CD71 positive (Figure 7). As a control, the CD4 positive fraction remained relatively constant (15–25%) throughout the duration of the experiment. From inspection of the cell images and time-lapse movies, it was clear that the large, less rounded cells were preferentially labeled with CD71 compared to the smaller rounded cells. This was borne out by the cell-by-cell analysis: when cells were classified into distinct subsets based on size (> or < 110 μm^2) at 48 h, 75 ± 1% of large cells were CD71 positive compared to 12 ± 1% of smaller cells. By the end of the experiment, > 90% of the larger cells were CD71 positive. Together, these data nicely illustrate the value of independently analyzing subsets of cells, and demonstrate how cell surface marker expression can be dynamically linked to morphological change in a cell subset that responds to an exogenous stimulus.

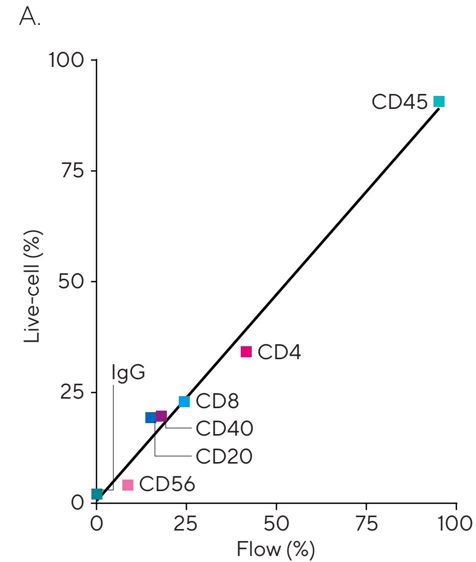
2. Subset Classification Based on CD Markers: CD8 Positive T Lymphocytes

Cytotoxic T lymphocytes are a subset of white blood cells that directly kill target cells that are either damaged, infected with bacteria or viruses, or recognized as cancerous. These T cells express the cell surface glycoprotein CD8, which is involved in the recognition of target cells via the T cell receptor | class 1 MHC antigen complex. Typically, between 15–35% of human PBMCs are CD8 positive.

As a proof of concept, we applied our cell identification method (Fabfluor-488-labeled CD8 antibody) and Cell-by-Cell Analysis techniques to study CD8 positive cells in mixed and co-culture models. First, we analyzed the sensitivity of CD8 positive cells to vincristine, a naturally occurring cytotoxic alkaloid used to treat a range of white blood cell cancers and thrombotic thrombocytopenic purpura. Human PBMCs were labeled with a Fabfluor-488 conjugated CD8 antibody in the presence of Incucyte® Opti-Green. Incucyte® Annexin V Red Dye was used to detect cellular apoptosis. Cells were treated at t = 0, with either vehicle or a range of concentrations of vincristine (1–300 nM), and images were gathered every 2 h for 120 h (Incucyte® Live-Cell Analysis System, 20X). Cells were segmented via their phase boundary and classified as positive and/or negative for the CD8 surface marker (green) and Annexin V (red) fluorescence intensity. As anticipated in these longer-term cultures, there was a high proportion (30–40%) of non-viable cells in the control wells. Nevertheless, it was possible to construct a full temporal and concentration-response drug analysis for the CD8 positive subset. Vincristine-induced toxicity was first detected 12 h post drug treatment; the threshold concentration was 3 nM, IC_{50} value 8 nM and maximal killing was observed at 100 nM (Figure 8). A true dynamic apoptotic index was subsequently calculated from the cell-by-cell analysis. Inspection of the Incucyte® images verified that the appearance of the Annexin V signal was coincident with morphological cell changes and a drop in total cell number.

In a separate study, we quantified the enrichment of CD8 positive T lymphocytes in a co-culture model of PBMCs and tumor cells (Incucyte® Cytolight Red labeled A549 adenocarcinoma). The intent here was to characterize the immune cell subset in the context of the tumor microenvironment. The phase segmentation algorithm was able to selectively identify immune cells (vs. tumor cells based on size and texture) within the co-culture. The absolute number and proportion of CD8 positive cells was determined using Fabfluor-488-labeled anti-CD8, and the Incucyte® Cell-by-Cell Analysis Software Module. PBMCs were activated with anti-CD3 | IL2 for 24 h prior to addition to plated tumor cells. Over time, there was a substantial rise in absolute cell number, and increase in the percentage of CD8 positive cells (from 35 ± 2% at the

formation of the culture to 60 ± 1% at 72 h), indicating an enrichment of the CD8 positive population. If PBMCs were not pre-activated, proliferation and enrichment were not observed. By establishing image masks around the tumor cells, it was possible to quantify 'proximity' analyses for the immune-cells and CD8 subset. Overall, CD8 cells were more closely spatially associated with tumor cells than the non-CD8 subset. Together, these two studies illustrate how a subset of immune cells can be identified and quantified in the context of other cells and monitored non-invasively over time.



B.

	Donor 1		Donor 2		Donor 3	
	Flow	Live-cell	Flow	Live-cell	Flow	Live-cell
IgG	0	0	0	3	0	3
CD45	96	91	98	87	92	94
CD20	9	11	18	20	19	27
CD4	49	42	40	30	36	31
CD8	22	19	33	31	18	19
CD40	10	10	32	22	12	27
CD56	11	2	15	9	1	2
R^2	0.98		0.99		0.96	
Slope	0.88		0.94		0.96	

Figure 5: Immunophenotyping: comparison of Incucyte® live-cell analysis and flow cytometry. Peripheral Blood Mononuclear Cells (PBMCs) from three donors were characterized for % expression for 6 CD markers and IgG control using live-cell analysis and flow cytometry. CD markers were identified using Fabfluor-488-labeled specific antibodies. A strong correlation was observed between the two methods when considering the mean values from the three donors (A), or each donor alone (B).

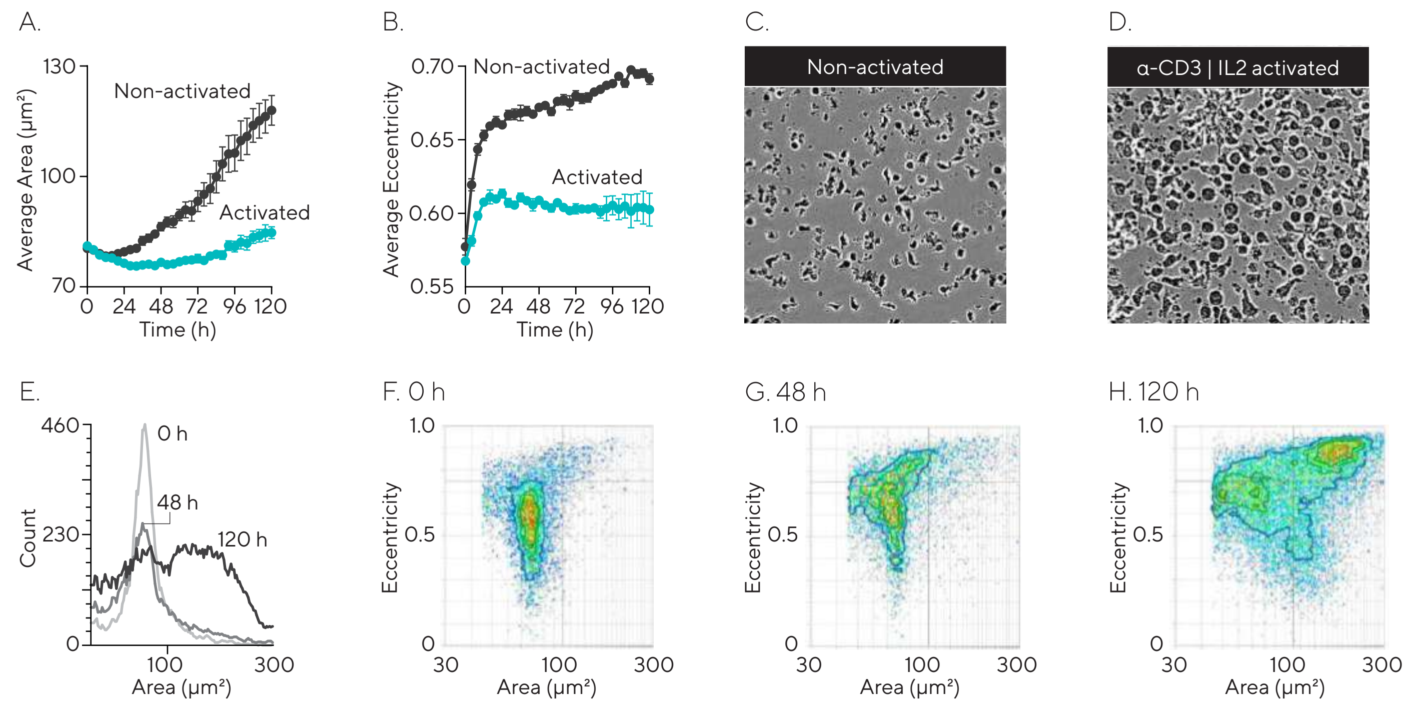


Figure 6: Cell-by-Cell Analysis: enlargement and morphological change in activated T cells. PBMCs were treated with anti-CD3 and IL-2, or vehicle control, and monitored over time with Incucyte® Live-Cell Analysis System. (A-D) Activation induced a time-dependent increase in average area and eccentricity (all cells). Note the change in eccentricity preceded the increase in area. The cell-by-cell area distribution (E) and density plots (F-H) highlight the increased heterogeneity over time following activation, and the appearance of a population of large cells with high eccentricity. Values shown are the mean ± S.E.M. of 4 wells.

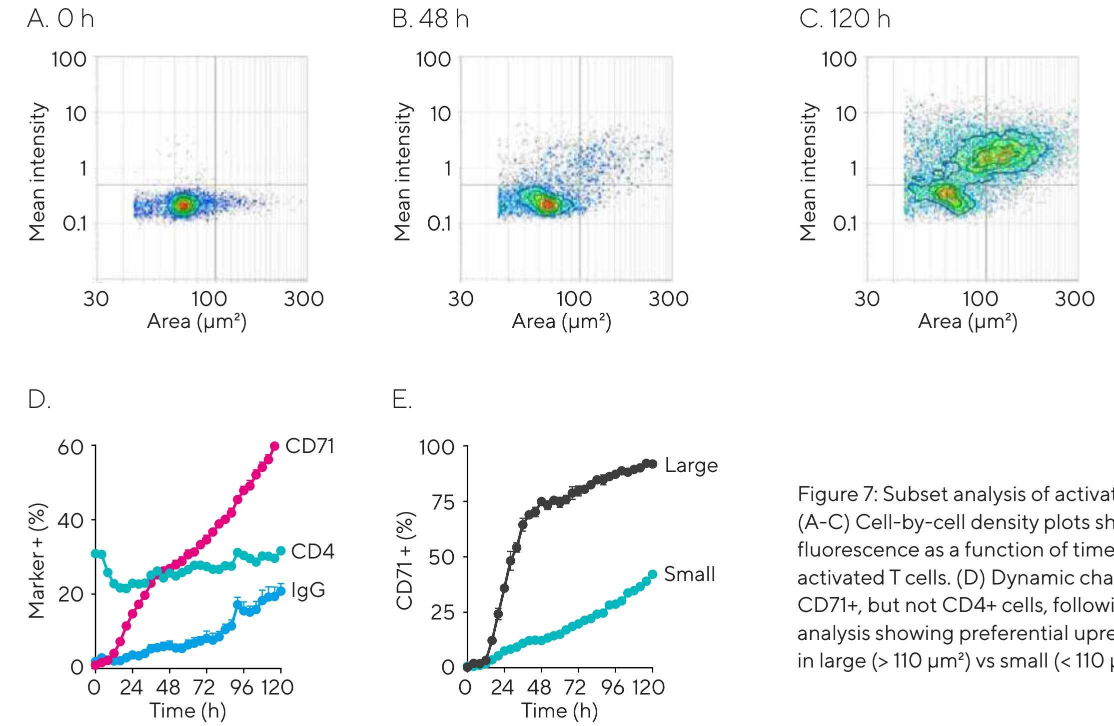


Figure 7: Subset analysis of activated T cells: CD71 upregulation. (A-C) Cell-by-cell density plots showing increase in CD71-fluorescence as a function of time and cell area (phase) in activated T cells. (D) Dynamic changes in the proportion of CD71+, but not CD4+ cells, following activation. (E) Subset analysis showing preferential upregulation of CD71 expression in large (> 110 µm²) vs small (< 110 µm²) cells.

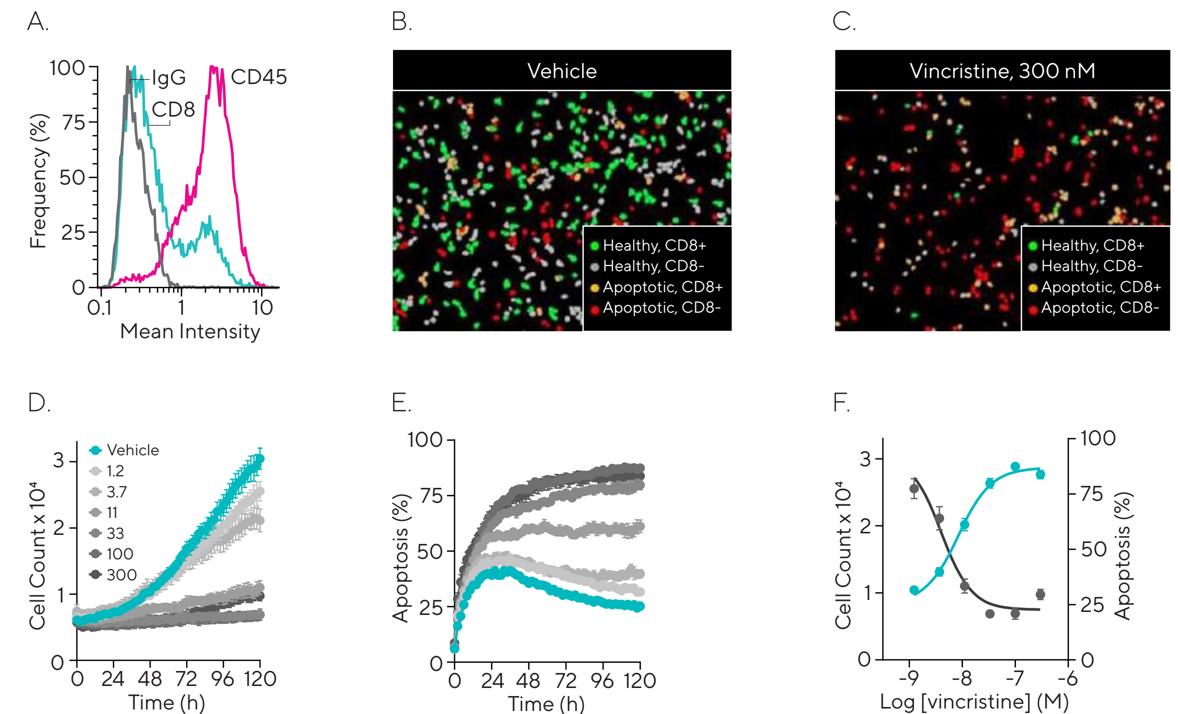


Figure 8: Susceptibility of human CD8+ T lymphocytes to vincristine-induced apoptotic cell death. (A) Frequency histograms of CD8, CD45 and IgG-labeled hPBMCs. (B and C) Segmented Incucyte® images showing color-coded subsets of healthy and apoptotic (Annexin V+) CD8+ or CD8-cells (4 groups) following treatment with vincristine (300 nM) or vehicle (48 h). Vincristine induced a concentration- and time-dependent reduction in the proliferation of CD8+ cells (D) and a concomitant increase in apoptosis (E). Concentration-response curves yielded IC_{50} or EC_{50} values of 4 nM for anti-proliferation and 8 nM for induction of apoptosis (F). Values shown are the mean ± S.E.M. of 4 wells.

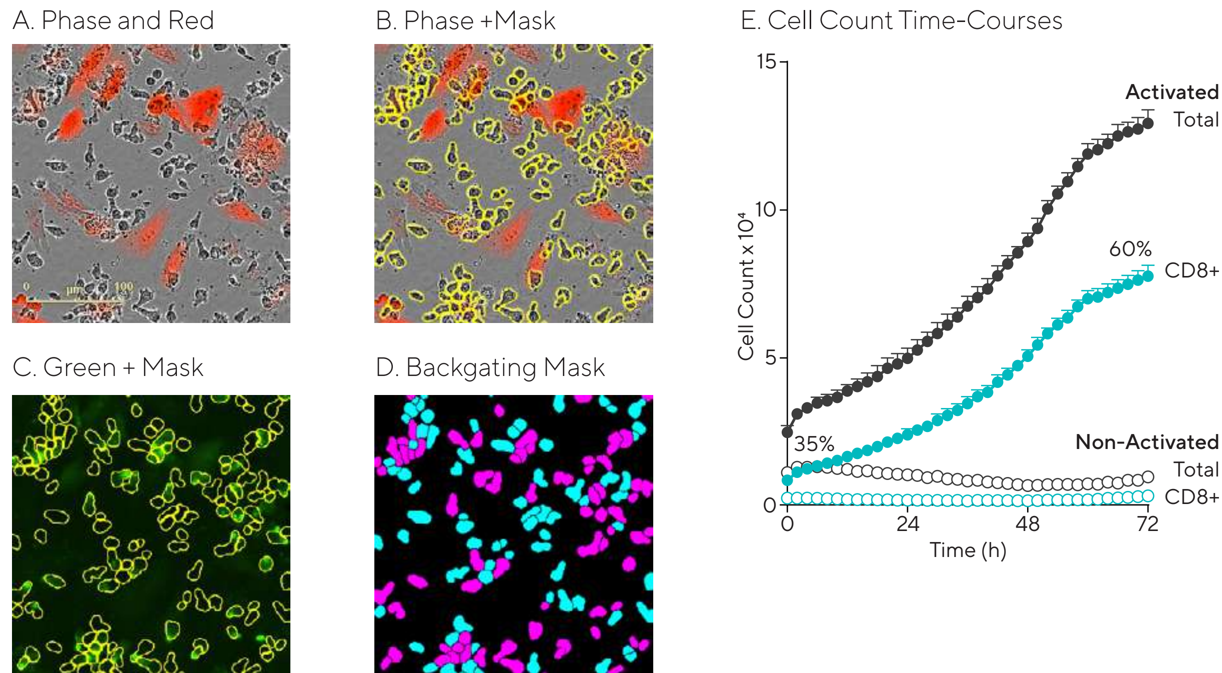


Figure 9: CD8 positive T cell enrichment in immune | tumor cell co-cultures. (A-C) Identification of hPBMCs (phase, yellow mask) and CD8+ (green, Fabfluor-488 | anti-CD8) subset in co-cultures with Incucyte® Cytolight Red A549 tumor cells. (D) Classified segmentation mask: purple = CD8+, cyan= CD8-. (E) Activation with anti-CD3 | IL2 induced a marked proliferative effect in the total population and enrichment of the CD8 subset (35% at t = 0 vs. 60% @72 h)

Summary and Future Perspectives

In this white paper, we have described an integrated solution for identifying and quantifying subsets of living cells in ...heterogeneous cultures over extended time periods and at industrial scale. Specifically, the ability to temporally analyze subsets of cells based on morphology (e.g., size, shape), biomarkers (e.g., CD antigens, nuclear labels), health | function (e.g., live | dead) and even spatial parameters (e.g., proximity to other cells) is tremendously powerful. While our method does not provide for true single-cell tracking, the solution affords both temporal and spatial perspective, which is otherwise missing from flow cytometry or high-content imaging approaches. The method obviates the need for cell lifting or fix | wash protocols and the experimental and analysis workflows are amenable to the everyday lab user as opposed to an imaging specialist.

In the examples shown, we illustrate the time-courses of apoptotic cell death of a lymphocyte population in mixed PBMC cultures, the differentiation of PBMCs and associated cell shape changes in cell subsets, and the enrichment of CD8 positive T lymphocytes in an immune | tumor cell co-culture system. In each case, the binning and classification of cells into subsets is an absolute requirement for meaningful quantitation, and the temporal analysis provides valuable, functional kinetic insight into the timing of changes in the cell populations. As with any repeated measure over time, there is a greater confidence and statistical power associated with this data compared to arbitrary single time point measures, and outcomes can be validated with images and movies.

Beyond these examples, there are numerous other biological applications that would benefit from cell-by-cell and live-cell subset analysis over time. Cancer stem cells show resistance to chemotherapeutics and are drivers of the disease but are typically a minority component of a tumor sample. Pathogens may only infect a subset of cells to propagate. Typically, only small numbers of stem cells and precursors will differentiate to downstream phenotypes depending on the conditions. More generally, the quantification of heterogeneity within a cell system enables researchers to probe novel questions as to whether cells with a particular stratified phenotype respond differently than their neighbors. In principle, this can be applied to any subset providing that there is a suitable reporter, biomarker or surrogate metric to enable the classification.

Going forward, there is opportunity for far deeper multi-parametric analysis of these living cell datasets. Similar to high-content image data processing, machine learning and deep learning algorithms could be applied to improve classification and subset resolution by better feature extraction, combining multiple variables and applying more rigorous statistical criteria (Caicedo *et al.*, 2017; Kraus *et al.*, 2017). Specific heterogeneity index tools could be introduced to report on variability in the cell populations (e.g., Gough *et al.*, 2014). A major challenge is how best to incorporate the temporal dimension for this large-scale analysis at the individual cell level. Coupling information extracted from hundreds of individual cells from one image to the next in a time sequence, when the images may be taken many minutes or hours apart, is not straightforward. There are a number of recent publications describing methods for individual cell dynamics using rapid (minutes) image acquisition paradigms (e.g., Heldt *et al.*, 2018), but extending this to low frequency sampling makes tracking each cell far more challenging. For the time being, we consider the Incucyte® approach of conducting Cell-by-Cell Analysis over time, without individual cell tracking, to be a valuable step forward.

Overall, the new approach outlined here provides a further dimension to live-cell analysis, where researchers can now probe questions at the cell-by-cell level and drill down into the characteristics and behaviors of subsets of cells. Increasingly, as live-cell analysis is applied to more complex and advanced cell systems with intrinsic heterogeneity, this analysis solution-set will yield the additional biological insight promised by these models.

References

1. Manshian BB, Munck S, Agostinis P, Himmelreich U, Soenen SJ. **High content analysis at single cell level identifies different cellular responses dependent on nanomaterial concentrations**. *Sci Rep*, Sept 8;5;13890 (2015)
2. Caicedo JC, Cooper S, Heigwer F, Warchal S, Qiu P, Molnar C, Vasilevich AS, Barry JD, Bansal HS, Kraus O, Wawer M, Paavolainen L, Herrmann MD, Rohban M, Hung J, Hennig H, Concannon J, Smith I, Clemons PA, Singh S, Rees P, Horvath P, Linington RG, Carpenter AE. **Data-analysis strategies for image-based cell profiling.** *Nat Methods,* 14(9);849–863 (2017)
3. Kraus OZ, Grys BT, Ba J, Chong Y, Frey BJ, Boone C, Andrews BJ. **Automated analysis of high-content microscopy data with deep learning.** *Mol Syst Biol,* Apr;13(4);924 (2017). Published online 2017 Apr 18. https://doi.org/10.15252/msb.20177551
4. Gough AH, Chen N, Shun TY, Lezon TR, Boltz RC, Reese CE, Wagner J, Vernetti LA, Grandis JR, Lee AV, Stern AM, Schurdak ME, Taylor DL. **Identifying and Quantifying Heterogeneity in High Content Analysis: Application of Heterogeneity Indices to Drug Discovery.** *Plos One,* Published: July 18, (2014) https://doi.org/10.1371/journal.pone.0102678
5. Schorpp K, Rothenaigner I, Maier J, Traenkle B, Rothbauer U, Hadian K. **Multiplexed High-Content Screening Approach Using the Chromobody Technology to Identify Cell Cycle Modulators in Living Cells.** *Journal of Biomolecular Screening,* 21(9);965–977 (2016)
6. Heldt FS, Barr AR, Cooper S, Bakal C, and Novák B. **A comprehensive model for the proliferation—quiescence decision in response to endogenous DNA damage in human cells.** *PNAS,* Feb 20, (2018) https://doi.org/10.1073/pnas.1715345115

North America
Essen BioScience Inc.
300 West Morgan Road
Ann Arbor, Michigan, 48108
USA
Phone: +1 734 769 1600
Email: orders.US07@sartorius.com

Europe
Essen BioScience Ltd.
Units 2 & 3 The Quadrant
Newark Close
Royston Hertfordshire
SG8 5HL
United Kingdom
Phone: +44 1763 227400
Email: euorders.UK03@sartorius.com

Asia Pacific
Sartorius Japan K.K.
4th Floor Daiwa Shinagawa North Bldg.
1-8-11 Kita-Shinagawa
Shinagawa-ku, Tokyo
140-0001
Japan
Phone: +81 3 6478 5202
Email: orders.US07@sartorius.com

Specifications subject to change without notice. © 2020. All rights reserved. Incucyte, Essen BioScience, and all names of Essen BioScience products are registered trademarks and the property of Essen BioScience unless otherwise specified. Essen BioScience is a Sartorius Company. Publication No.: 8000-0625-B00 Version: 12|2020

Panel Discussion

Bioimaging and analysis

Evolving requirements for drug discovery and basic research, such as the use of 3D cell cultures and live-cell studies, have led to more sophisticated image capture technologies. These technologies can provide highly detailed and complex images, which in turn require new analysis methods.

In this Panel Discussion for our Spotlight on bioimaging and analysis, we explore how these technological developments have changed the landscape of bioimaging and detail key tools that have facilitated improved image analysis with experts in the field.

Panelists

Irene Martinez Carrasco (left) was born and raised in Madrid (Spain) where she studied Molecular Biology. She completed a PhD at the Children's Hospital Boston (MA, USA), after which she was appointed with a technician position as a microscopist at the Severo Ochoa Molecular Biology Center (Madrid, Spain). After this period, she was recruited as the manager of the microscopy service at Umeå University (Sweden). During this time, she has been in contact with many research groups and a broad range of research topics. She mainly supports users with the use of the microscopes but also with image analysis.

Beth Cimini (left) is a Senior Group Leader and CZI Imaging Scientist in the Imaging Platform at the Broad Institute in Cambridge (MA, USA). The Cimini lab focuses on bioimage analysis tool creation (Piximi) and maintenance (CellProfiler), as well as on applying open-source tools to novel biological problems. She created and directs the Platform's Postdoctoral Training Program in Bioimage Analysis, and also leads the Broad's efforts toward community engagement and driving biological projects for the Center for Open Bioimage Analysis (COBA).

Gillian Lovell (left) is a Senior Scientist at Sartorius. Within the BioAnalytics group, she has worked across multiple research areas developing functional live-cell imaging assays and building these into customer-facing applications. Gillian previously obtained a PhD in Chemical Biology from Imperial College using multidisciplinary techniques to synthesize probes for biological targets and evaluate their effects with biochemical and cell-based assays. After a period of postdoctoral research, she joined Sartorius in 2015 working on the Incucyte® Live-Cell Analysis Systems.

Sheraz Ahmed (left) is a Senior Researcher at DFKI GmbH in Kaiserslautern (Germany), where he is leading the area of Time Series Analysis and Life Science. He received his MS and PhD degrees in Computer Science from the Technical University of Kaiserslautern under the supervision of Andreas Dengel and Marcus Liwicki. His research interests include pattern recognition, anomaly detection, gene analysis, medical image analysis, and natural language processing. He's published more than 100 papers on these, and related, topics.

BioTechniques Spotlight

Follow the QR code to access the Panel Discussion:

SARTORIUS

Contact us

Editorial Department

Senior Editor
Tristan Free
tfree@biotechniques.com

Business Development and Support

Commercial Director
Sarah Mayes
s.mayes@future-science-group.com

This supplement is brought to you by *BioTechniques* in association with:

SARTORIUS