Title:

Automated workflow for the cell cycle analysis of non-adherent and adherent cells using a machine learning approach

Authors: Kourosh Hayatigolkhatmi[1*], Chiara Soriani[1*], Emanuel Soda[1*], Elena Ceccacci[1], Oualid El Menna[1], Sebastiano Peri[1], Ivan Negrelli[2], Giacomo Bertolini[2,] Gian Martino Franchi[2], Roberta Carbone[2], Saverio Minucci[1,3$], Simona Rodighiero[1$]

[1]Department of Experimental Oncology, European Institute of Oncology-IRCCS, Via Adamello 16, 20139, Milano, Italy.

[2]Tethis S.p.A., Milan, Italy.

[3]Department of Oncology and Hemato-Oncology, University of Milan, Milan, Italy.

[*]Contributed equally

[$]Corresponding authors

Summary

Understanding the details of the cell cycle at the level of individual cells is critical for both cellular biology and cancer research. While existing methods using specific fluorescent markers have advanced our ability to study the cell cycle in cells that adhere to surfaces, there is a clear gap when it comes to non-adherent cells. In this study, we combine a specialized surface to improve cell attachment, the genetically-encoded FUCCI(CA)2 sensor, an automated image processing and analysis pipeline, and a custom machine-learning algorithm. This combined approach allowed us to precisely measure the duration of different cell cycle phases in non-adherent, as well as adherent cells.

Our method provided detailed information from hundreds of cells under different experimental conditions in a fully automated manner. We validated this approach in two different acute myeloid leukemia cell lines, NB4 and Kasumi-1, which have unique and distinct cell cycle characteristics. We also measured how drugs that influence cell cycle properties affect the duration of each phase in the cell cycles of these cell lines. Importantly, our cell cycle analysis system is freely available and has also been validated for use with adherent cells.

In summary, this article introduces a comprehensive, automated method for studying the cell cycle in both non-adherent and adherent cells, offering a valuable tool for cellular biology, cancer research and drug development.

Introduction

Cell cycle dynamics coordinate cellular division and proliferation through regulating the different cell cycle phases. Dysregulation in these processes is a hallmark of malignancies such as human cancer, where aberrant activities in cyclin-dependent kinases (CDKs), cyclins, and CDK inhibitors often drive uncontrolled proliferation. Consequently, targeting cell cycle components has emerged as a pivotal therapeutic strategy, especially crucial in pre-clinical drug evaluation (Malumbres and Barbacid, 2009; Khan and Wang, 2022).

Traditional methods for assessing cell cycle dynamics have been largely dependent on the quantification of DNA content through flow or image cytometry, providing a static snapshot of cell populations in various cycle phases (Furia, Pelicci and Faretta, 2013; Ligasová, Frydrych and Koberna, 2023). Although valuable, these techniques fall short in capturing intra-population variability and require additional protein markers for precise phase determination (Ligasová, Frydrych and Koberna, 2023; Rieger, 2022).

Methods utilizing cells expressing fluorescently labelled reporters and time-lapse microscopy can discriminate cell cycle phases at the level of individual cells, thereby offering valuable insights into the variability of cell cycle and cell cycle phase durations within the overall cell population (Chao *et al.*, 2019; Hiratsuka and Komatsu, 2019).

Advanced imaging methods, such as time-lapse microscopy coupled with fluorescently tagged reporters, have shown promise in detailing cell cycle dynamics at the single-cell level. Technologies like the Fluorescent Ubiquitination-based Cell Cycle Indicator (FUCCI) have been employed for this, effectively demarcating cell cycle phases through color-coding (Sakaue-Sawano *et al.*, 2017).

FUCCI(CA)2 express the hCdt1(1/100) fused to mCherry fluorescent protein and hGem(1/110) fused to mVenus, generating a clear and distinct tricolor demarcation, separating G1 (red), S (green), and G2/M phases (yellow) (Sakaue-Sawano *et al.*, 2017).

This dynamic insight is particularly crucial in the context of acute myeloid leukemia (AML), where chromosomal translocations generate fusion genes that disrupt cellular differentiation programs and drive proliferation (Alcalay *et al.*, 2001).

The described experimental setup utilizes nanostructured titanium oxide-coated multiwell plates relying on the technology used in the commercially available Smart BioSurface (SBS) slides (Krol *et al.*, 2021). Such technology should hypothetically enable us to immobilize non-adherent cells for extended imaging durations. To overcome the limitations of manual data analysis, we introduce an automated image analysis pipeline for time-lapse movies of AML cell lines, exploiting a FUCCI-based probe for visualization. Our data analysis approach combines custom image processing, TrackMate-based cell tracking, and machine learning-based track filtering. Thereby automating the entire data analysis workflow.

In summary, we present a comprehensive, experimental protocol for cell cycle analysis in adherent and non-adherent cells (summarized in **BOX1**). The approach leverages routine imaging technologies and advanced data analyses, enhancing the precision and efficiency of drug screening protocols in oncological research.

Results

Modified conditions enable AML cells to adhere to the substrate feasible for live cell imaging

Live imaging and tracking of non-adherent cells, when multiple positions should be acquired, is challenging due to their high propensity to mechanical perturbations.

To conduct long time imaging of AML cells, we exploit the combined action of the SBS and the partial immobilization effect of methylcellulose (MC). We were able to image and track AML cells up to 72h when 20% complete medium was added to 80% MC and applied on AML cells previously adhered to the SBS (movie S1).

Two different AML cell lines, NB4 and Kasumi-1, were equipped with the FUCCI(CA)2 technology and chosen as the study models. NB4 cells have relatively faster doubling time in comparison to Kasumi-1 cells (Skopek *et al.*, 2023). Hence, NB4 cells were treated with vehicle or CDK inhibitors, being compared to the naturally slow-cycling counterparts, Kasumi-1 cells, by time-lapse imaging (Fig. 1A and Fig. S1). As a result, we were able to follow single cells and the manual assessment of cell cycle phases confirmed the impact of CDK inhibitors. Furthermore, the expected difference in cell cycle progression of NB4 vs Kasumi-1 cells was evident.

Image processing facilitates the cell tracking and profiling of cell cycle phases

The original time-lapse images were composed by Red and Green channels, detecting respectively the mCherry and mVenus markers of the FUCCI(CA)2 indicator. During the experiment each cell alternatively switches between the expression of these two markers as it goes through the different

cell cycle phases, causing the lack of a single fluorescence channel suitable for tracking. Furthermore, the automatic profiling of cell cycle phases becomes cumbersome when dealing with two distinct channels that in the end give rise to two independent fluorescence time-series. We set up an image processing pipeline utilizing the open-source software Fiji (Schindelin *et al.*, 2012) (NIH, version 2.14.0/1.54f) to transform the original images into a dataset optimized for the subsequent steps of tracking and cell cycle phase assignment. We represented the color changes that occur during the cell cycle with the Hue scale, as described in (Fujimoto *et al.*, 2020). To achieve this, as represented in Fig. 1B, stacks of Red and Green channels were overlaid and converted into an RGB stack and then transformed into an HSB stack (Hue, Saturation and Brightness channels). The Hue channel was retained for the assignment of cell cycle phases, while the Brightness channel was used as tracking reference. These two were merged to the Red and Green fluorescence channels to form the final stack used for the tracking process.

In the tracking analysis, the Fiji plugin TrackMate (Tinevez *et al.*, 2017) was employed, adapting the example script available on the dedicated TrackMate website (https://imagej.net/plugins/trackmate/scripting/scripting), to ensure the automation of the tracking step. The related parameters were selected and tuned according to each experiment, as well as the proper filters to discard uninformative tracks. The output consisted of a table with cells associated to selected tracks and their corresponding numerical features in each timeframe. Key features included the mean fluorescence intensity in the Red and Green channels, as well as the mean intensity of the Hue value.

The whole image analysis pipeline successfully recovered a unique tracking channel and effectively mapped the alternating Red and Green curves (Fig. 1C, left panel) into a single time series, as depicted in Figure 1C, right panel.

Incorrect track filtering by the machine learning model efficiently automates data cleaning and cell cycle phase assignment steps

The files generated by the Trackmate pipeline were imported into R (R Core Team, 2021) for the demultiplexing, filtering and data analysis. To address missing frames, a data imputation process was implemented, aimed at recovering instances with up to 5 consecutive missing values. This data imputation was executed using an Exponential Weighted Moving Average (EWMA) technique, employing a window size of 3, thereby encompassing 6 observations (3 preceding, 3 succeeding).

A smoothing spline approach was individually applied to the green, red, and hue channels, utilizing a regularization parameter (lambda) of 0.0001. A min-max scaling to transform value into the range [0, 1] was applied to make the fluorescence intensity and hue scale comparable among channels and samples. For the creation of the random forest model, a training pipeline was set up, as described in Fig 2A. In accordance with Fig 2B, feature extraction was carried out on the green and red channels to derive temporal descriptors of the data.

Employing these extracted features, the trained random forest model was applied to discriminate between trackable and untraceable cells. Subsequently, only the cells predicted as traceable by the model were retained for subsequent analyses. Here are reported the obtained number of cells for each condition, merging all the experiments performed:

- NB4 DMSO: 410 cells (3 experiments) out of 1862 tracked cells
- NB4 Palbociclib 50 nM: 328 cells (3 experiments) out of 2881 tracked cells
- NB4 PF-0606873600 50 nM: 206 cells (1 experiment) out of 1102 tracked cells
- NB4 Ribociclib 50nM: 119 cells (1 experiment) out of 784 tracked cells
- Kasumi-1 Untreated: 119 cells (1 experiment) out of 1604 tracked cells
- MDA-MB-231 Untreated: 1116 (1 experiment) out of 3204 tracked cells

By using the hue channel each track was partitioned into its cell cycle component using the following thresholds:

- **G1**: hue >= 0 & hue < 0.65
- **S**: hue >= 0.85
- **G2/M**: hue >= 0.65 & hue <= 0.85

Additionally, a refinement step involving cell reassignment was conducted to identify instances where the G1 to S phase transition was labelled as G2/M in frames preceding the S phase. The outcome of the analysis is visually depicted in Fig. 3A and B.

Evaluation of tracks' selection performances confirms the efficiency of the machine learning model

We carried out a validation of the performances of the whole pipeline in selecting "traceable" tracks. Specifically, we extracted from one experiment the ID of the tracks that belonged to the vehicle (DMSO) treated condition identified as "traceable" by the ML algorithm. We then validated these tracks by visual inspection of the corresponding images, checking for eventual tracker errors. We created a scoring metric defined as follows:

$Score = \frac{N_{Tracks\ (images)}}{N_{Tracks\ (ML\ algorithm)}}$	N_{Tracks} (images) = the number of tracks selected as *traceable* by the algorithm that were good also in images N_{Tracks} (ML algorithm) = the number of tracks selected as *traceable* by the algorithm

We checked up to 102 tracks that underwent to whole cell cycle profiling by the analysis pipeline and we found that, among these, 63 were also consistent with images (Score = 0.62). We then decided to select tracks identified as traceable with a probability greater than 0.70, gaining a final Score of 0.74 (57 tracks consistent with images over 77 tracks).

The method can quantify the cell cycle progression and does not impact the transcriptome over long time periods

We were able to accurately quantify the cell cycle progression of NB4 and Kasumi-1 cell lines over 72 hours at single-cell resolution (figure 3A). The empirical observation of the superimposition of red and green fluorescence signals during the transition between G1 and S makes the classification of the cell in one of the two dual states challenging. Thus, we decided to quantify the G1 to S phase transition in the cell cycle profiling, as a separate cell cycle stage.

As represented in figure 3A-B, cell cycle differences of NB4 and Kasumi-1 cells were quantified. Moreover, the method was able to discriminate cell cycle differences of leukemic cells in presence of various sub-optimal doses of different CDK inhibitors. These CDK inhibitors are potent inhibitors of CDKs2/4/6, which mainly regulate the G1 to S transition (Fassl, Geng and Sicinski, 2022). Hence, we expected the administration of nanomolar concentrations of these agents to lengthen mainly the G1 phase with less impact on the other phases. Our live cell imaging of the relatively fast-cycling NB4 cells in presence of these drugs at such spectrum of concentrations affirmed the significant G1-prolongation in these cells (Fig. 3B and movies S2-4). The total average duration of one full cell cycle was quantified as 21.5 h ± 6.5 h (mean ± standard deviation) for NB4 cells and 24.0 h ± 7.8 h for Kasumi-1 cells. For instance, administration of 50nM Palbociclib extended mainly the G1 phase of the NB4 cells by about 5 hours (from 9.1 h ± 5.1 h to 14.2 h ± 5.7 h in vehicle treated and Palbociclib treated NB4 cells, respectively).

It was demonstrated that the nanostructured surfaces can promote changes in the cellular protein expression profile (Schulte *et al.*, 2016). We thus questioned the possible impact of the SBS on gene expression profile of the AML cells. We made a comparison of transcriptome of the cells pre- and post-imaging by performing RNA-seq investigations. As represented in the supplementary Fig. S2, we did not observe any meaningful transcriptomic alterations throughout the process. Hence, we conclude that the protocol is feasible for biological investigations without a significant effect on the transcriptomic profile of the cells.

Discussion

In the last few years, efforts have been made to simplify the cell cycle assessment relying on FUCCI technology, resulting in the development and public availability of software tools and ImageJ plug-ins (Roccio *et al.*, 2013; Koh *et al.*, 2017; Ghannoum *et al.*, 2021; Taïeb *et al.*, 2022). However, all these methods require human intervention at different points in the analysis workflow, such as the initial selection of cells for analysis (Taïeb *et al.*, 2022) or the manual correction of inaccurate tracks during the analysis (Roccio *et al.*, 2013; Koh *et al.*, 2017; Ghannoum *et al.*, 2021; Taïeb *et al.*, 2022). In this report we describe a complete protocol for the cell cycle analysis of adherent and non-adherent cells expressing the FUCCI(CA)2 technology in a fully automated manner. The complete experimental workflow is explained in BOX1 and illustrated in Fig.4. It was applied to the analysis of the cell cycle phases of two different AML cell lines, NB4 and Kasumi-1, which have different durations. Hence, detecting this difference was a reliable initial verification strategy for the protocol's efficiency. As a second quality check step, we decided to treat the fast-cycling NB4 cells with suboptimal concentrations of three different CDK inhibitors (Palbociclib, PF-06873600 and Ribociclib). These CDK inhibitors have different inhibitory impact on various CDKs (Freeman-Cook *et*

al., 2021; Fassl, Geng and Sicinski, 2022). The effect of the three CDK inhibitors on the cell cycle duration of NB4 cells was evaluated. Up to about 400 cells in one single experimental condition were quantified, for up to 12 different conditions in a single experiment. This entire analysis process took approximately 2 hours of human involvement, for a total execution time that ranges from 12 to 48 hours, depending on dataset size, that is commissioned to a machine (Fig. 4). It should be taken into consideration that the manual correction of incorrect tracks takes approximately 2h per field of view (FOV). Hence, in the reported experiments of this study, the analysis of a single well would take about 40 to 50 hours of manual work. This amount of time would make such analysis unfeasible. Not only the suggested protocol makes these experiments doable, but also the experimental set-up can be scaled up according to the experimental needs. We successfully emphasized alterations of few hours in the duration of cell cycle phases when NB4 cells were subjected to exceedingly low concentrations of inhibitors. Moreover, the set-up pipeline makes accessible the quantification of potentially thousands of cells in an automated fashion relying on the crucial contribution of a machine learning algorithm.

Various approaches have been documented for imaging non-adherent cells. These range from creating custom supports to confine cell movement within a restricted region enabling extended imaging periods of up to 40 hours (Day *et al.*, 2009), to partially immobilizing cells using substances like gelatine (Ritter *et al.*, 2020) or low-melting-point agarose (Strong and Daniels, 2017).

The capability to track non-adherent cells for up to 72 hours (almost equivalent to three complete cell cycles in rapidly dividing NB4 cells) was achieved by employing SBS-coated glass in combination with the addition of MC to the culture medium. MC was necessary to limit the movement of single cells or cell clusters after more than 24h of observation (movies S1 and S5). We deduce that this seeding protocol has the potential to be effectively employed with various non-adherent cell types. This is based on the principle that cell adhesion on SBS is facilitated by the nanostructure's ability to engage with integrins (Schulte *et al.*, 2016), which are widely expressed in diverse cell types (Johansen *et al.*, 2018; Floren *et al.*, 2020; Kim *et al.*, 2020; Ogana *et al.*, 2023). The seeding protocol could also certainly be extended to other types of live cell imaging experiments, especially when the fluorescence intensity can be monitored for shorter durations, thereby circumventing the need for the addition of MC. Moreover, the image and data analysis pipelines can be easily applied to adherent cells, as we showed in Fig S4.

The presented image and data analysis workflows rely on 2 different software, Fiji and R, widely used by the imaging and data analysis community. The scripts were made publicly available, enabling customization to tailor the workflow according to specific requirements. This not only streamlines the cell cycle analysis of cells expressing the FUCCI probe, but also makes it accessible to researchers with basic programming knowledge.

In conclusion we confirm the efficiency of the optimized conditions in immobilizing the non-adherent cells for long time periods, without affecting their predicted response upon different environmental circumstances. The described work allows researchers in the field to analyse thousands of (non-)adherent cells per each experimental condition, by using the image and data analysis pipelines to automatically perform image processing, cell tracking, filtering of incorrect tracks and cell cycle phases identification and characterization.

BOX 1 – experimental and analysis workflow

Sample Preparation (in-suspension cells):

The cells of interest should be at optimum viability and physiological conditions prior to the mounting process on the SBS-multiwell plates. This step requires 1-2 hours of manual handling and is divided into seven steps as follows:

(I) Consider 250 thousand cells for one well of a standard 12-well SBS-multiwell. Wash the cells 3 times in sterile phosphate-buffered saline (PBS).
Note 1: PBS should be at room temperature (RT).
Note 2: The media and PBS should not be vacuumed during the process. It is crucial to make sure the cells won't get dried at any point of the process.
Note 3: The mentioned number of cells are for cells with an average size of 10 to 20 micrometers in diameter and a doubling time of 20-30 hours. The number of cells with different properties, should be optimized accordingly.

(II) Resuspend the cells in 1000 µl PBS and gently mix, achieving a homogenous cell suspension.

(III) Load the suspension slowly from one side of each well, covering all the surface.

(IV) Let the cells mount for 20-30 mins at RT, followed by 20-30 mins at 37°C.
Note 4: Avoid moving or shaking the plate or the surface underneath.

(V) Remove the PBS with pipette (we suggest not to use the vacuum and not to tilt the well).

(VI) Wash the wells 2 times with enough volume of media without fetal bovine serum (FBS) (we suggest not to use the vacuum and not to tilt the well).
Note 5: Any protein contamination, including FBS can easily perturb adhesion of the cells on the titanium-oxide particles.
Note 6: Load and aspirate slowly from the side of the wells, keeping the pipette at 45° angle.

(VII) Load enough methylcellulose gently from one side of the well and proceed to the next step.
Note 7: In case any treatments are needed during the time-lapse experiment, the desired agents to can be added to the methylcellulose compartment prior to loading on the wells.

Sample Preparation (adherent and semi-adherent cells):

In case the cells of interest are (semi-)adherent cells, plate the cells at optimum confluency (based on the planned time-lapse duration and cell properties) on glass-bottom dishes and proceed to the next step.

Live-Cell Imaging

(I) Pre-heat the microscope incubator at 37 °C and 5 % of CO_2 for standard cell culture. Otherwise tune the temperature and humidity controller according to the ideal cell viability conditions.

(II) Mount the sample on the microscope stage and choose the imaging parameters depending on the cell type.

Note 8: we suggest setting the objective and image binning to represent each cell with at least 15 pixels in diameter.

Note 9: we suggest adjusting the excitation intensity and exposure time for the detection of mCherry and mVenus signal to have a difference between maximum and minimum gray value of at least 1000 units (for 16-bit camera).

(III) Select the desired total duration of the experiment and the time-frame parameter, considering the cell type viability upon illumination and the speed of cells movement.

Note 10: for tracking purposes it is convenient to start with a minimum time interval of 30 min and evaluate the outcome of the time-lapse, adjusting eventually the described parameter in the following experiments

Image processing

(I) Open the pre-processing macro (Image_processing_HSB_v1.ijm) in your Fiji

Note 11: the execution of the macro requires the prior installation of the Basic plugin (https://github.com/marrlab/BaSiC) for flat-field correction

(II) Set the proper image filters and background subtraction processing for red and green channels at lines 58-78 and 85-94
(III) Choose the maximum values of the Brightness and Contrast for both red and green channels and set the found values at lines 116 (red) and 118 (green)
(IV) Set the proper value of the Top Hat filter on the Brightness channel (HSB stack) at line 127
(V) Set the correct position of red and green channels, depending on the experiment, at lines 173-174.
(VI) Run the macro from the Fiji script editor
(VII) Set in the dialog window the input directory (where original data are stored), the output directory (where you want to store processed images) and the original file format

Tracking Analysis

(I) Open the TrackMate script in the Fiji macro editor
(II) Adjust the parameters settings (from line 39 to 58) according to the experiment
(III) Run the script from the Fiji script editor
(IV) Set in the dialog window the input directory (where the images processed with the previous pipeline are stored) and the output directory (where you want to store the results table)

Data Analysis

Sample/condition demultiplexing

(I) Read the list of files containing tracking information for each cell. This can be done by looping through the files and reading into R using the function read.delim().
(II) Merge all the tables into a single table using the filename to create a column for the demultiplexing of the condition.

Missing frame imputation

(I) Use the na_ma() function to perform imputation. This function let the user choose among different weighting strategies, we suggest trying the “exponential” or the “simple”, as well

as trying different values for k (dimension of the moving average window) and maxgap (maximum number of successive NA values).

Data smoothing

(I) Use the function smooth.spline() to remove noise from the timeseries and enhance the trend. We suggest trying different values for lambda value starting with a very low one such as [1e-05, 1e-04, 1e-03].

Data normalization

(I) Use the function normalize_vector() to make the curves comparable to each other.

Feature extraction

(I) Use the function tk_tsfeatures () to extract features from the channel intensity, different features can be specified to be extracted such as:
    a. The number of times the time series crosses its median.
    b. Autocorrelation of the time series.
    c. Autocorrelation of the first/second-differentiated time series.
    d. Spectral entropy.
    e. Stability and lumpiness on a tilled version of the timeseries.

Traceability assessment

(I) Use the function predict() passing the table and the pre-trained model to predict whether a track is traceable or not. The output is a table containing the predicted class as well as the predicted probability of being part of a specific class.
(II) The predicted probability can be used to further filter the predicted traceable cells.

Cell phase assignment

(I) The cell cycle phase quantification is performed on the cells predicted as traceable; this is done by setting two thresholds on the Hue intensities. We suggest trying different values according to the original Hue intensity reached by the fluorescence observed.
(II) The first round of phase assignment divides the tracks into:
    a. G1
    b. S
    c. G2/M
(III) By looping through each track, a second round of phase assignment is performed to re-assign frames classified as G2/M that occur before S. Three different strategies can be followed:
    a. Re-assign the G2/M frames to G1
    b. Re-assign the G2/M frames to S
    c. Re-assign the G2/M frames to a new G1/S phase

Cell phase quantification

(I) Group your data using the group_by() function using the Track_ID and Condition columns as grouping variables then use the function add_count() to count the number of frames in each condition for each track.
(II) Divide the number obtained by the number of frames acquired in an hour to obtain the time spent in each phase (in hours units).
(III) Sum the time spent in each phase to obtain the total cell cycle duration.

Materials and Methods

Key resource table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Cell line (*Homo sapiens*)	NB4	Cellosaurus	CVCL_0005	
Cell line (*Homo sapiens*)	Kasumi-1	ATCC	CRL-2724™	
Cell line (*Homo sapiens*)	MDA-MB-231	ATCC	CRM-HTB-26™	
Chemical compound and drug	Palbociclib	Selleckchem	S4482	
Chemical compound and drug	Ribociclib	Selleckchem	S7440	
Chemical compound and drug	PF-06873600	Selleckchem	S8816	
Plasmid	tFUCCI(CA)2/pCSII-EF	RIKEN BRC through the National BioResource Project of the MEXT, Japan (cat. RDB15446); Sakaue-Sawano, et al., Mol. Cell 68 (3): 626-640.e5. 2017.	RDB15446	
Software	LAS X	Leica Microsystems Inc	https://www.leica-microsystems.com/products/mic	

			roscope-software/	
Software	FIJI	Schindelin et al., 2012	https://fiji.sc/	
Software	TrackMate	https://doi.org/10.1016/j.ymeth.2016.09.016	https://imagej.net/plugins/trackmate/	
Software	R	R Core Team,2023	https://www.r-project.org/about.html	
Software	Tidyverse	Wickham et al, 2019		
Software	Tidymodels	Khun et al, 2020		
Software	tsfeatures	Hyndamn et al, 2023	https://pkg.robjhyndman.com/tsfeatures/	
Software	timetk	Dancho et al, 2023	https://business-science.github.io/timetk/	
Widefield Microscope	Leica Thunder Imager	Leica	https://www.leica-microsystems.com/products/thunder-imaging-systems/	
Camera	DFC9000 GTC	Leica	https://www.leica-microsystems.com/products/microscope-cameras/p/leica-dfc9000/	

Multi-color illumination	Lumencor Spectra X	Lumencor	https://lumencor.com/products/spectra-x-light-engine	
Other	Glass bottom multiwell plate	MatTek Corporation	P12G-1.5-14-F	

Cell lines and growing conditions

NB4 and Kasumi-1 cells were grown in 3ml of a mix of 80ml MethoCult™ H4230 RPMI-1640 plus 20ml Roswell Park Memorial Institute (RPMI) 1640 medium with final concentrations of 10% fetal bovine serum (FBS), 2mM glutamine and 1% Penicillin/Streptomycin. MDA-MB-231 cells were grown in Dulbecco's Modified Eagle Medium (DMEM) with 10% FBS+2mM L-Glutamine and 1% Penicillin/Streptomycin. All cells were grown according to ATCC recommendations in a humidified tissue culture incubator at 37 °C with 5% CO_2 environment.

Generation of FUCCI(CA)2-equipped cell lines

Cells were put in 24 well-plates and plated at a density of 500,000 cells in 500 µl of medium per well. The Lentiviral vector carrying the tFUCCI(CA)2/pCSII-EF (Sakaue-Sawano *et al.*, 2017) plasmid was diluted in RPMI 10% serum pen/strep in order to add 500 µl to the cells. Three rounds of infection were carried out in the presence of 5ug/mL of polybrene (Sigma). Cells were infected simply by adding the 100X concentrated virus supernatant (20-35µL/well) onto the cells; the plate was centrifuged for 1 hour at RT at 2500rpm. The medium is added 2 hours post-infection up to 1ml final. Then incubated at 37˚C overnight. Infected cells expressed FUCCI(CA)2 probes, emitting mCherry and mVenus. Therefore, the selection of the cells expressing the FUCCI(CA)2 was done by sorting the fluorescence markers, mCherry and mVenus on fluorescence-activated cell sorting (FACS) instrument (FACSAria cell sorter,BD Biosciences, Oxford, UK).

SBS coated bottom-glass multiwell plates production

Glass bottom multiwell plates are coated by Tethis using the Supersonic Cluster Beam Deposition (SCBD) technology (a detailed description of SCBD and its principle of operation can be found in Refs.(Piseri, Tafreshi and Milani, 2004; Wegner *et al.*, 2006). A supersonic seeded beam of titania nano-clusters is produced, under high vacuum, by a pulsed microplasma cluster source (PMCS) (Piseri, Tafreshi and Milani, 2004; Wegner *et al.*, 2006) and deposited on glass bottom multiwell plates. The process is tuned to produce nanostructured TiO2 films (thickness from 50 to 200 nm) with a controlled nanoscale morphology. The titania nanostructured coating of the plates is transparent and biocompatible and has a surface topography that promotes the spontaneous adhesion and immobilization of living cells (Carbone *et al.*, 2006).

Sample preparation

Approximately 250,000 of Kasumi-1 and NB4 cells and 100,000 of MDA-MB-231 cells were plated either in each well of the Tethis SBS 12-well multi-wells or 8-well ibidi plates, in triplicates in the presence of the aforementioned compounds at day 0 (T0). Cells were treated with final 2nM, 10nM, 50nM, 250nM, and 1250nM of Palbociclib, 50nM, 75nM, 100nM, 125nM and 150nM of Ribociclib, and 25nM, 50 nM, 75nM, 100nM and 125nM of PF-06873600, diluted in advance and directly in the relevant culture medium for each cell line. Only the 50nM treatments are shown for Ribociclib and PF-06873600.

Three cell suspension samples were prepared for triplicate independent counting and the average of three readings used as the cell count. Cell suspension was diluted with Trypan Blue dye (Sigma) at 1:1 ratio to identify cell viability using Biorad TC20™ automated cell counter.

RNA extraction and RNA-seq protocol

Total RNA was extracted from dry pallets of cells collected prior- and post-acquisitions and purified using the Zymo Research Quick-RNA Miniprep (W/O directzol). Reverse transcription was performed with the SuperScript II Kit (Invitrogen), according to the manufacturer's protocol. RNA-seq was performed according to the True-seq Low sample protocol selecting only polyadenylated transcripts. In brief, before starting mRNA isolation and library preparations the integrity of the total RNA was evaluated by running samples on a Bioanalyzer instrument by picoRNA Chip (Agilent), then converted into libraries of double stranded cDNA appropriate for next generation sequencing on the Illumina platform. The Illumina TruSeq v.2 RNA Sample Preparation Kit was used following manufacturer's recommendations. Briefly, 0.1-1 µg of total RNA were subjected to two rounds of mRNA purification by denaturing and letting the RNA bind to Poly-T oligo-attached magnetic beads. Then fragmentation was performed exploiting divalent cations contained in the Illumina fragmentation buffer and high temperature. First and second strand cDNA is reverse transcribed from fragmented RNA using random hexamers. First strand cDNA was synthesized by SuperScript II (Invitrogen) reverse transcriptase transcriptase and random primers and second strand cDNA synthesized by DNA polymerase I and Rnase H. The subsequent isolation of the cDNA was achieved by using AMPure XP beads (depending on the concentration used, these beads can efficiently recover PCR products of different sizes). The product recovered contained overhanging strands of various lengths due to the fragmentation procedure. The 5' and 3' ends of cDNA are repaired by the 3'-5' exonuclease activity and the polymerase activity and adenylated at 3' extremities before ligating specific Illumina oligonucleotides adapters followed by 15 cycles of PCR reaction using proprietary Illumina primers mix to enrich the DNA fragments. Prepared libraries were quality checked and quantified using Agilent high sensitivity DNA assay on a Bioanalizer 2100 instrument (Agilent Technologies).

RNA sequencing data analysis

Raw reads 51bp PE for NB4 and Kasumi cells were quality-filtered and aligned to the hg18 reference genome using nf-core/rnaseq v3.9 pipeline using STAR as aligner and Salmon for quantification with default parameters. Gene counts for each sample were log1p transformed, mean value among the

two replicates was taken to compute Pearson correlation among gene expression pre- and post-time-lapse acquisition.

Image Acquisition

Images were acquired with a Leica Thunder Imager (Leica Microsystems, Wetzlar, Germany), equipped with a Lumencor Spectra X Light Engine (Lumencor, Beaverton, USA) for fluorescence excitation, a motorized stage and a Leica DFC9000 GTC camera. For non-adherent cells, images were acquired with LAS X software (Leica Microsystems, Wetzlar, Germany, version 3.7.5.24914) using a 20X/0.75NA air objective and a binning 2x2 was applied to increase the SNR. The mCherry and mVenus signal were detected respectively with 540-580 nm and 460-500 nm excitation filters, 585 and 505 nm dichroic mirrors and 592-668 nm and 512-542 nm emission filters. The brightfield channel was also acquired for representation purposes. We imaged 20 to 25 fields of views per well and focal points were manually set in each position before starting the acquisition and kept constant during the whole time-lapse thanks to the Adaptive Focus Control (AFC, Leica Microsystems). The total duration of the time-lapse on non-adherent cells was 72 hours, and the time interval was set to 1 hour to prevent cell phototoxicity.

Regarding the experiment on adherent cells, a 10X/0.32 NA PH1 dry objective was used and the time-lapse duration was set to 120 hours, with 30 min as timestep. In this case 10 fields of view per well were acquired.

Image Analysis

The image pre-processing step was performed using a custom-made Fiji macro. Briefly, the pipeline executed a flat-field correction on the fluorescence channels with the plugin Basic (Peng *et al.*, 2017), then applied a gaussian blur (sigma=1 px), a top hat filter (radius=20 px) and a background subtraction to enhance the cells' signal; duplicated images of Red and Green channels were merged in a multichannel stack and brightness and contrast values were adjusted, according to each experiment, within the range of 0 to the maximum gray value of the stack's histogram; the multichannel images were then converted into RGBs for each frame, and finally into a HSB stack from which the Brightness channel and the Hue channel were kept respectively for tracking and cell-phase profiling purposes; the final processed image was saved for further analysis and was composed of 4 channels: the Red and Green channels, the Brightness channel and the Hue channel.

The tracking analysis was realized through TrackMate (version 7.10.2). We automatized the execution of TrackMate over all the fields of view acquired in the experiment by adapting a Jython script, freely available on the website (https://imagej.net/plugins/trackmate/scripting/scripting). As the shape of our cells was approximately round, the Laplacian of Gaussian (LoG) detector was selected as the cell identifier on the Brightness channel, and the Linear Assignment Problem (LAP) Tracker algorithm was utilized for linking phase, adjusting the Max Linking distance parameter for each experiment. Gap closing was enabled for up to 3 frames, considering the significant decrease in fluorescence intensity during mitosis and the transition between G1 and S phase, that can eventually affect the detection of a cell, due to the rapid decrease of the Red signal and the relatively slow increase of the Green one (Sakaue-Sawano *et al.*, 2017). Identified tracks were filtered for total

duration, track displacement, and starting frame of the track. Specifically, the duration filter maximized the chances to follow at least one cell cycle (Track Duration > 25 hours), while the cutoff on total displacement assures the deletion of dead cells that are usually motionless (Track displacement > 10 px). Tracks that start after the first 10 hours were discarded to avoid tracker errors that may give rise to unreliable cell cycle quantification as the cells tend to form clusters above 24-30 hours of experiment (movie S1). The specific values of parameters of the LoG cells' detector (i.e., spot radius, quality threshold) were preliminarily checked on images via the TrackMate GUI, as well as the linking and gap closing max distances and the final filters on tracks' duration, displacement and starting frame. The parameters selected manually were then inserted in the script for the batch execution. As an output, we automatically saved a table of spots for each field of view, corresponding to cells in identified tracks, with the selected TrackMate features (See Suppl. Table I).

Model Creation

To assess the traceability of each track in a fast, efficient and scalable manner, we employed a machine-learning approach. A dataset consisting of 2319 manually annotated tracks, with 1,939 tracks classified as untraceable and 380 tracks marked as traceable (∼1/5) was used. The dataset was divided into training and test sets, with 1855 tracks (~ 80%) allocated for training and 464 tracks (~20%) reserved for testing while also stratifying for the outcome variable. To address the problem of the unbalanced outcome a downsampling procedure is employed in the training set.

Time-series-like features associated with each track are extracted (Wang, Smith and Hyndman, 2006; Hyndman *et al.*, 2023), as illustrated in Figure S3A, displaying the distribution of the 70 features utilized for distinguishing between traceable and untraceable cells.

To optimize model performance and address potential biases of the random forest model employed, we applied stratified 10-fold cross-validation on the training set, creating resampling folds. Leveraging these folds, we conducted hyperparameter tuning. Specifically, we explored different values for Randomly Selected Predictors (mtry), number of trees (trees), and Minimal Node Size (min_n) to identify the combination that maximized the Area under the receiver operator curve (AUC) (Sacks *et al.*, 1989). Subsequently, an additional round of training was conducted using the selected hyperparameters, this time employing the entire training set. The trained model was then evaluated on an independent test set to assess its performance. The results, as depicted in Figure S3B, revealed an AUC value of 0.971 as well as a sensitivity of 0.897, a specificity of 0.974 and an accuracy of 0.905 confirming the validity of the training procedure. These values were comparable to those obtained from the resampling folds, indicating the absence of overfitting.

Given the relatively small size of the manually annotated training set, we performed an additional and final round of training using the pseudo-labelling framework (Lee, 2013). The original model was used to predict unlabelled data, any predicted probability exceeding 0.5 was considered indicative of confidence in the prediction. A new semi-supervised random forest model was trained on this augmented dataset, employing the same hyperparameters as determined previously. This final model achieved an improved AUC of 0.975 on the test set, indicating enhanced track trackability assessment. The schematics of the pipeline, implemented in R (v. 4.3.0) (R Core Team,

2021) using Tidymodels (v. 1.1.0) (Kuhn and Wickham, 2020), and Tidyverse (v.2.0.0) (Wickham *et al.*, 2019) can be seen in Figure 3 (panels A).

Time series analysis

To generate time-series-like information, we utilized the above-mentioned Fiji pipeline, which provided a table containing intensity values for the red, green, and hue channels. To address missing frames resulting from tracking gaps, an Exponential Weighted Moving Average (EWMA) was fitted to impute these frames.

To further refine the track curves, we applied a two-step smoothing process. Initially, a Simple Moving Average (SMA) was applied, followed by fitting a fixed lambda Smoothing Spline with $\lambda = 0.0001$. Finally, min-max scaling is used to normalize the tracks within the range of [0, 1] to make the red and green intensity comparable.

To tackle tracking errors introduced by the Fiji pipeline, we applied the aforementioned random forest model to assess the traceability of each cell (Fig. 2B).

On the tracks classified as trackable, a manual threshold on the hue intensity is employed to determine the cell cycle phases (G1, S, and G2/M). Furthermore, to quantify the cell cycle at the single-cell level, track splitting to isolate individual cell components and perform phase reassignment into G1, G1/S, S, and G2/M is performed using a custom R function.

The resulting single-cell tracks enabled the quantification of individual phase durations for each cell. As an example, the cell phase quantification for five different conditions can be observed in Figure 3.

Model Deployment

The resulting model was saved as a .rds file to facilitate practical implementation. Moreover, we encapsulated the model as an API in a Docker container using Vetiver (v. 0.2.1) (Silge, 2023), enabling easy deployment and usage.

Code availability

The source code and user manual for the Fiji pipeline is available at https://github.com/ieoresearch/cellcycle-image-analysis.

The source code and user manual for the R pipeline is available at GitHub repository.

Funding and Acknowledgements

This work was partially supported by the Italian Ministry of Health with Ricerca Corrente and 5x1000 funds and by the Grant AIRC IG20 to SM. KH received a fellowship from AIRC.

CS is grateful to Giulia Tini for the helpful discussions.

References

Alcalay, M. *et al.* (2001) 'Common themes in the pathogenesis of acute myeloid leukemia', *Oncogene*, 20(40), pp. 5680–5694. Available at: https://doi.org/10.1038/sj.onc.1204642.

Carbone, R. *et al.* (2006) 'Biocompatibility of cluster-assembled nanostructured TiO2 with primary and cancer cells', *Biomaterials*, 27(17), pp. 3221–3229. Available at: https://doi.org/10.1016/j.biomaterials.2006.01.056.

Chao, H.X. *et al.* (2019) 'Evidence that the human cell cycle is a series of uncoupled, memoryless phases', *Molecular Systems Biology*, 15(3), p. e8604. Available at: https://doi.org/10.15252/msb.20188604.

Day, D. *et al.* (2009) 'A method for prolonged imaging of motile lymphocytes', *Immunology & Cell Biology*, 87(2), pp. 154–158. Available at: https://doi.org/10.1038/icb.2008.79.

Fassl, A., Geng, Y. and Sicinski, P. (2022) 'CDK4 and CDK6 kinases: From basic science to cancer therapy', *Science*, 375(6577), p. eabc1495. Available at: https://doi.org/10.1126/science.abc1495.

Floren, M. *et al.* (2020) 'Tetraspanin CD82 drives acute myeloid leukemia chemoresistance by modulating protein kinase C alpha and β1 integrin activation', *Oncogene*, 39(19), pp. 3910–3925. Available at: https://doi.org/10.1038/s41388-020-1261-0.

Freeman-Cook, K.D. *et al.* (2021) 'Discovery of PF-06873600, a CDK2/4/6 Inhibitor for the Treatment of Cancer', *Journal of Medicinal Chemistry*, 64(13), pp. 9056–9077. Available at: https://doi.org/10.1021/acs.jmedchem.1c00159.

Fujimoto, K. *et al.* (2020) 'Tracking and Analysis of FUCCI-Labeled Cells Based on Particle Filters and Time-to-Event Analysis', *International Journal of Bioscience, Biochemistry and Bioinformatics*, 10(2), pp. 94–109. Available at: https://doi.org/10.17706/ijbbb.2020.10.2.94-109.

Furia, L., Pelicci, P.G. and Faretta, M. (2013) 'A computational platform for robotized fluorescence microscopy (I): High-content image-based cell-cycle analysis', *Cytometry Part A*, 83A(4), pp. 333–343. Available at: https://doi.org/10.1002/cyto.a.22266.

Ghannoum, S. *et al.* (2021) 'CellMAPtracer: A User-Friendly Tracking Tool for Long-Term Migratory and Proliferating Cells Associated with FUCCI Systems', *Cells*, 10(2), p. 469. Available at: https://doi.org/10.3390/cells10020469.

Hiratsuka, T. and Komatsu, N. (2019) 'Single-Cell Live Imaging', in V. Proserpio (ed.) *Single Cell Methods*. New York, NY: Springer New York (Methods in Molecular Biology), pp. 409–421. Available at: https://doi.org/10.1007/978-1-4939-9240-9_24.

Hyndman, R. *et al.* (2023) *tsfeatures: Time Series Feature Extraction*. Available at: https://pkg.robjhyndman.com/tsfeatures/.

Johansen, S. *et al.* (2018) 'The Possible Importance of β3 Integrins for Leukemogenesis and Chemoresistance in Acute Myeloid Leukemia', *International Journal of Molecular Sciences*, 19(1), p. 251. Available at: https://doi.org/10.3390/ijms19010251.

Khan, M.G.M. and Wang, Y. (2022) 'Cell Cycle-Related Clinical Applications', in Z. Wang (ed.) *Cell-Cycle Synchronization: Methods and Protocols*. New York, NY: Springer US, pp. 35–46. Available at: https://doi.org/10.1007/978-1-0716-2736-5_3.

Kim, H.N. *et al.* (2020) 'Cadherins, Selectins, and Integrins in CAM-DR in Leukemia', *Frontiers in Oncology*, 10. Available at: https://doi.org/10.3389/fonc.2020.592733.

Koh, S.-B. *et al.* (2017) 'A quantitative FastFUCCI assay defines cell cycle dynamics at a single-cell level', *Journal of Cell Science*, 130(2), pp. 512–520. Available at: https://doi.org/10.1242/jcs.195164.

Krol, I. *et al.* (2021) 'Detection of clustered circulating tumour cells in early breast cancer', *British Journal of Cancer*, 125(1), pp. 23–27. Available at: https://doi.org/10.1038/s41416-021-01327-8.

Kuhn, M. and Wickham, H. (2020) *Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles.* Available at: https://www.tidymodels.org.

Lee, D.-H. (2013) 'Pseudo-Label : The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks', in. Available at: https://api.semanticscholar.org/CorpusID:18507866.

Ligasová, A., Frydrych, I. and Koberna, K. (2023) 'Basic Methods of Cell Cycle Analysis', *International Journal of Molecular Sciences*, 24(4). Available at: https://doi.org/10.3390/ijms24043674.

Malumbres, M. and Barbacid, M. (2009) 'Cell cycle, CDKs and cancer: a changing paradigm', *Nature Reviews Cancer*, 9(3), pp. 153–166. Available at: https://doi.org/10.1038/nrc2602.

Ogana, H.A. *et al.* (2023) 'Targeting integrins in drug-resistant acute myeloid leukaemia', *British Journal of Pharmacology*, n/a(n/a). Available at: https://doi.org/10.1111/bph.16149.

Peng, T. *et al.* (2017) 'A BaSiC tool for background and shading correction of optical microscopy images', *Nature Communications*, 8(1), p. 14836. Available at: https://doi.org/10.1038/ncomms14836.

Piseri, P., Tafreshi, H.V. and Milani, P. (2004) 'Manipulation of nanoparticles in supersonic beams for the production of nanostructured materials', *Current Opinion in Solid State and Materials Science*, 8(3–4), pp. 195–202. Available at: https://doi.org/10.1016/j.cossms.2004.08.002.

R Core Team (2021) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Rieger, A.M. (2022) 'Flow Cytometry and Cell Cycle Analysis: An Overview', in Z. Wang (ed.) *Cell-Cycle Synchronization*. New York, NY: Springer US (Methods in Molecular Biology), pp. 47–57. Available at: https://doi.org/10.1007/978-1-0716-2736-5_4.

Ritter, P. *et al.* (2020) 'A method for high-content functional imaging of intracellular calcium responses in gelatin-immobilized non-adherent cells', *Experimental Cell Research*, 395(2), p. 112210. Available at: https://doi.org/10.1016/j.yexcr.2020.112210.

Roccio, M. *et al.* (2013) 'Predicting stem cell fate changes by differential cell cycle progression patterns', *Development*, 140(2), pp. 459–470. Available at: https://doi.org/10.1242/dev.086215.

Sacks, J. *et al.* (1989) 'Design and Analysis of Computer Experiments', *Statistical Science*, 4(4), pp. 409–423. Available at: https://doi.org/10.1214/ss/1177012413.

Sakaue-Sawano, A. *et al.* (2017) 'Genetically Encoded Tools for Optical Dissection of the Mammalian Cell Cycle', *Molecular Cell*, 68(3), pp. 626-640.e5. Available at: https://doi.org/10.1016/j.molcel.2017.10.001.

Schindelin, J. *et al.* (2012) 'Fiji: an open-source platform for biological-image analysis', *Nature Methods*, 9(7), pp. 676–682. Available at: https://doi.org/10.1038/nmeth.2019.

Schulte, C. *et al.* (2016) ‘Conversion of nanoscale topographical information of cluster-assembled zirconia surfaces into mechanotransductive events promotes neuronal differentiation’, *Journal of Nanobiotechnology*, 14(1), p. 18. Available at: https://doi.org/10.1186/s12951-016-0171-3.

Silge, J. (2023) *vetiver: Version, Share, Deploy, and Monitor Models*.

Skopek, R. *et al.* (2023) ‘Choosing the Right Cell Line for Acute Myeloid Leukemia (AML) Research’, *International Journal of Molecular Sciences*, 24(6), p. 5377. Available at: https://doi.org/10.3390/ijms24065377.

Strong, A.D. and Daniels, R.L. (2017) ‘Live-cell calcium imaging of adherent and non-adherent GL261 cells reveals phenotype-dependent differences in drug responses’, *BMC Cancer*, 17(1), p. 516. Available at: https://doi.org/10.1186/s12885-017-3507-y.

Taïeb, H.M. *et al.* (2022) ‘FUCCItrack: An all-in-one software for single cell tracking and cell cycle analysis’, *PLOS ONE*, 17(7), pp. 1–14. Available at: https://doi.org/10.1371/journal.pone.0268297.

Tinevez, J.-Y. *et al.* (2017) ‘TrackMate: An open and extensible platform for single-particle tracking’, *Methods*, 115, pp. 80–90. Available at: https://doi.org/10.1016/j.ymeth.2016.09.016.

Wang, X., Smith, K. and Hyndman, R. (2006) ‘Characteristic-Based Clustering for Time Series Data’, *Data Mining and Knowledge Discovery*, 13(3), pp. 335–364. Available at: https://doi.org/10.1007/s10618-005-0039-x.

Wegner, K. *et al.* (2006) ‘Cluster beam deposition: a tool for nanoscale science and technology’, *Journal of Physics D: Applied Physics*, 39(22), pp. R439–R459. Available at: https://doi.org/10.1088/0022-3727/39/22/R02.

Wickham, H. *et al.* (2019) ‘Welcome to the tidyverse’, *Journal of Open Source Software*, 4(43), p. 1686. Available at: https://doi.org/10.21105/joss.01686.

Figure legends

Figure 1: Cell Tracking. A) Examples from Kasumi-1 cell and NB4 cell lines (upon the different conditions: DMSO, Palbociclib 50 nM, PF-0606873600 50 nM, Ribociclib 50 nM, Untreated), showing a tracked cell that explores the different cell-cycle phases (Scale bar is 5 µm). B) Fluorescence images with R and G channels were processed in Fiji to create a HSB stack, from which the Brightness and Hue channel were extracted to be used respectively as tracking channel and cell phase identification channel (Scale bar is 50 µm). C) Plot curves generated using the data extracted from the TrackMate script execution. The variations over time of mCherry (red curve) and mVenus (green curve) fluorescence intensities (left plot) and of the Hue scale (right plot) of a single NB4 cell in DMSO condition are shown.

Figure 2: Training and inference pipelines. A) The Machine Learning pipeline followed to create the quality model. Using timetk time-series associated features are extracted from the list of manually annotated tracks. A random forest model is then trained to predict whether a track is cycling or not. B) An unannotated track can be fed to the model to predict whether it is cycling or not.

Figure 3: Cell Cycle Phase Assignment. A) Waterfall plot of the ”sorted” and “splitted” cells. Each row corresponds to a cell. B) Boxplot of the cell phase duration of the first cell cycle. The asterisks are the adjusted significance level given by Wilcox test of the sample for each over the NB4 DMSO for each phase according to the color (ns: $p > 0.05$, *: $p <= 0.05$, **: $p <= 0.01$, ***: $p <= 0.001$, ****: $p <= 0.0001$).

Figure 4: Experimental and analysis workflow. Summary of the entire experimental workflow, from cell seeding to the calculation of cell cycle phases. Asterisks in the Live Cell Imaging section mean optional settings. Image created on Biorender.com.

Supplementary Figure legends

Figure S1: Cell Cycle Phase Assignment. **A)** Examples from NB4 cell line upon different Palbociclib treatment concentrations showing a cell that explores the cell cycle phases (Scale bar is 5 µm). For Palbociclib 250 nM and Palbociclib 1250 nM conditions, the cell was manually tracked for presentation purposes.

Movie S1: Representative movie of NB4 cells seeded on SBS + MC.

Movies S2-S4: Representative movies of 3 cells tracked with TrackMate expressing the FUCCI(CA)2 probe, showing the sequence of color changes (1 frame/hour).

Movie S5: Representative movie of NB4 cells seeded on SBS without MC.

Figure S2: A) Pearson correlation among pre and post time-lapse acquisition of Kasumi-1 (left) and NB4 (right) cell lines. In the x axis is reported the log1p counts mean between the two replicates before the acquisition, while on the y axis the log1p counts mean between the two replicates after the acquisition.

Figure S3: A) Boxplot showing the distribution of the feature extracted from the training set using timetk. B) ROC curve of the random forest model computed in the test set.

Figure S4: A) Waterfall plot of the ”sorted” and “splitted” cells. Each row corresponds to a cell. B) Boxplot of the cell phase duration of the first cell cycle. A total of 1116 cells were analysed, obtaining a mean (± SD) cell cycle duration of 24.5 ± 8.5 h. **C)** Example of an adherent cell in control condition tracked with the described pipeline (Scale bar is 5 µm).

FIG 1

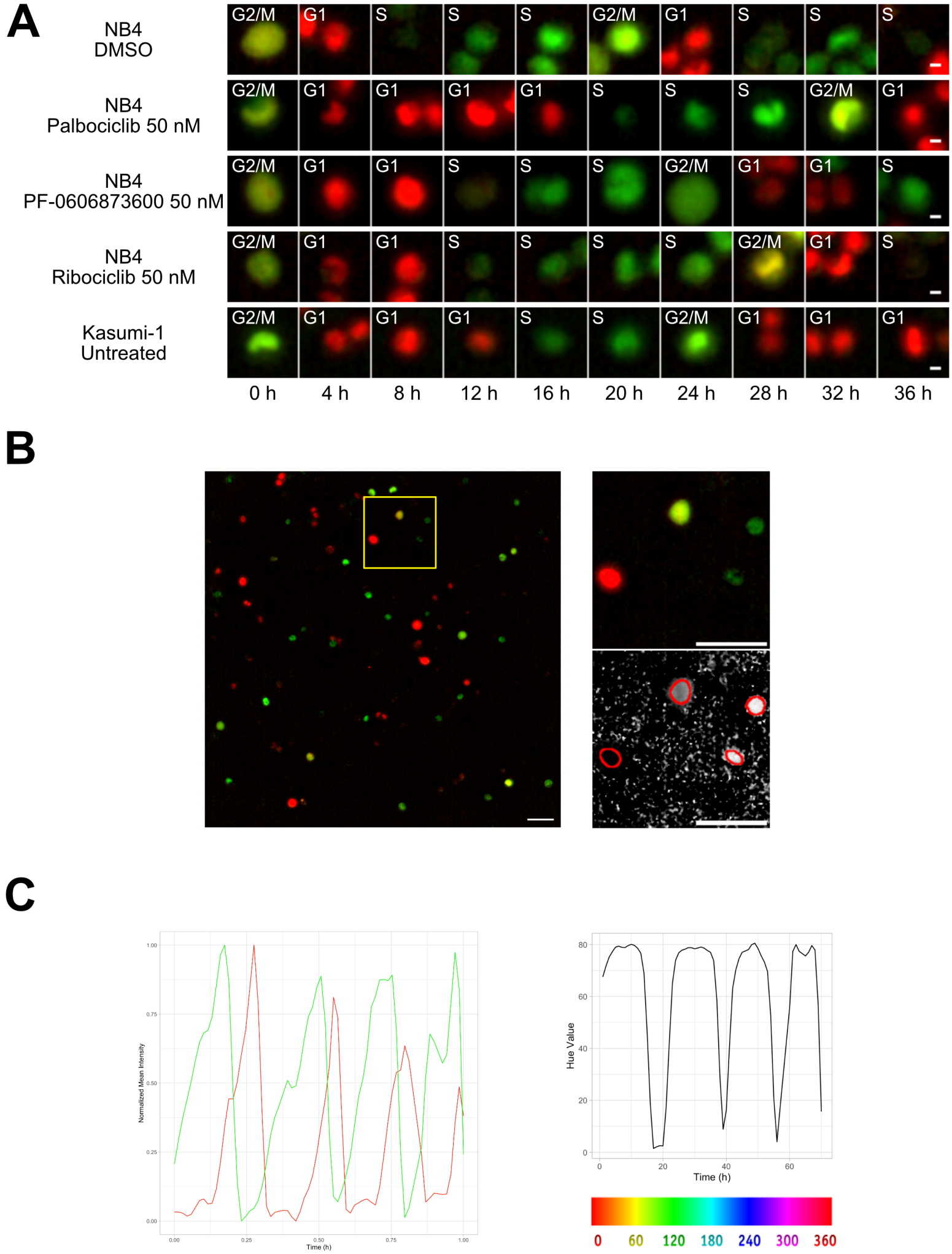
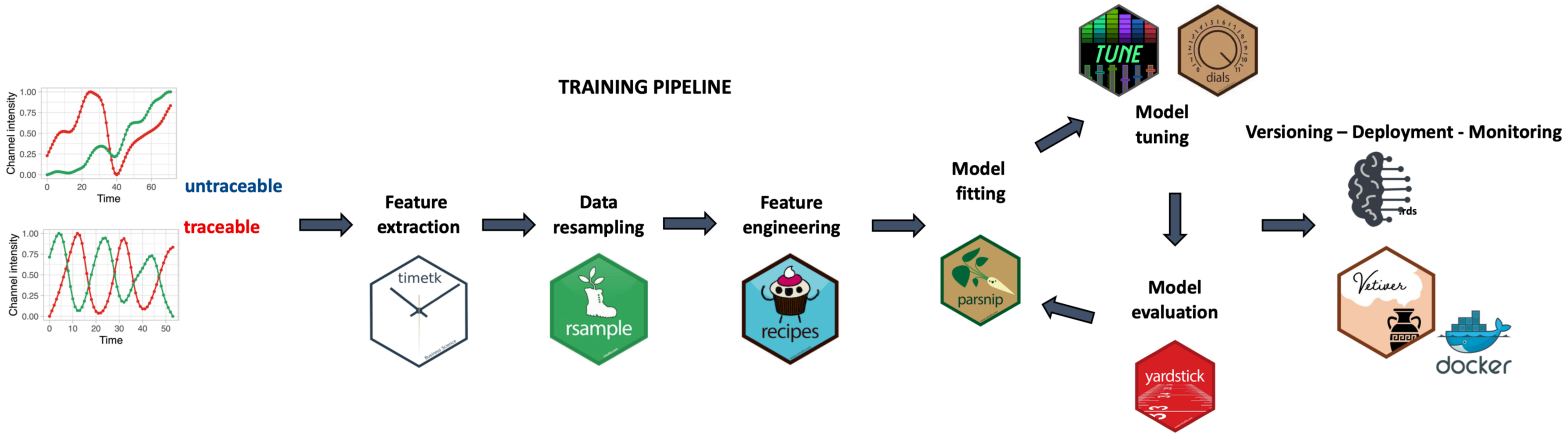

A
NB4 DMSO
G2/M G1 S S S G2/M G1 S S S
NB4 Palbociclib 50 nM
G2/M G1 G1 G1 G1 S S S G2/M G1
NB4 PF-0606873600 50 nM
G2/M G1 G1 S S S G2/M G1 G1 S
NB4 Ribociclib 50 nM
G2/M G1 G1 S S S S G2/M G1 S
Kasumi-1 Untreated
G2/M G1 G1 G1 S S G2/M G1 G1 G1
0 h 4 h 8 h 12 h 16 h 20 h 24 h 28 h 32 h 36 h
B
C
Normalized Mean Intensity
Time (h)
Hue Value
Time (h)
0 60 120 180 240 300 360

FIG 2

A


TRAINING PIPELINE
Channel intensity
Time
untraceable
traceable
Feature extraction
timetk
Data resampling
rsample
Feature engineering
recipes
Model fitting
parsnip
Model tuning
TUNE
dials
Model evaluation
yardstick
Versioning – Deployment - Monitoring
.rds
Vetiver
docker

B

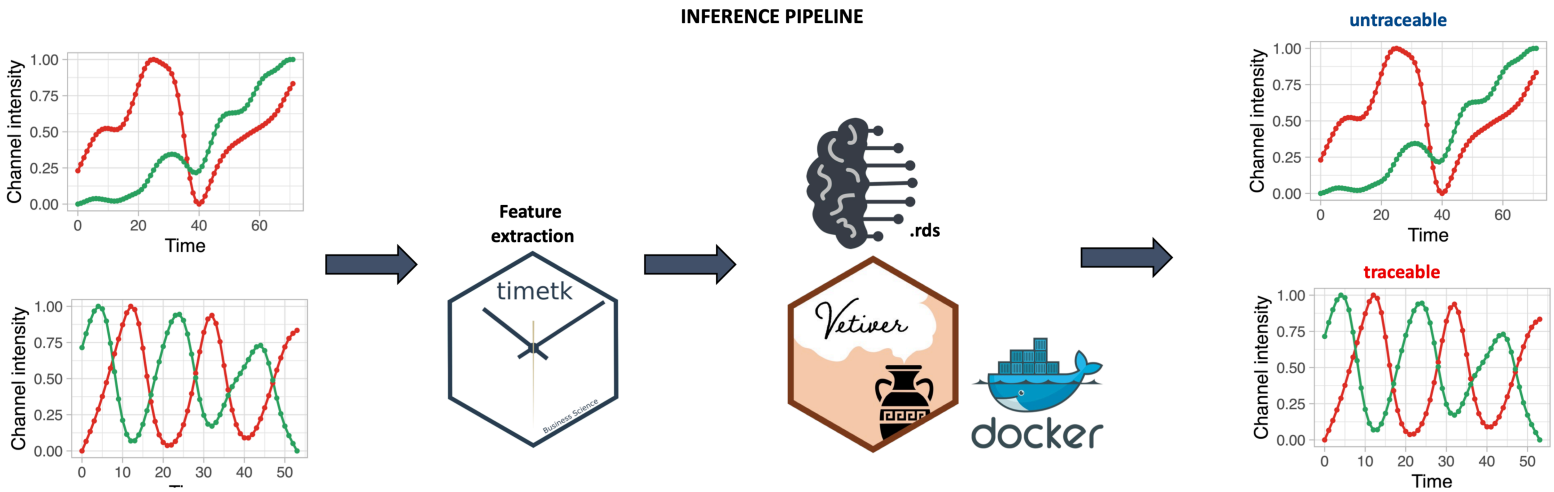
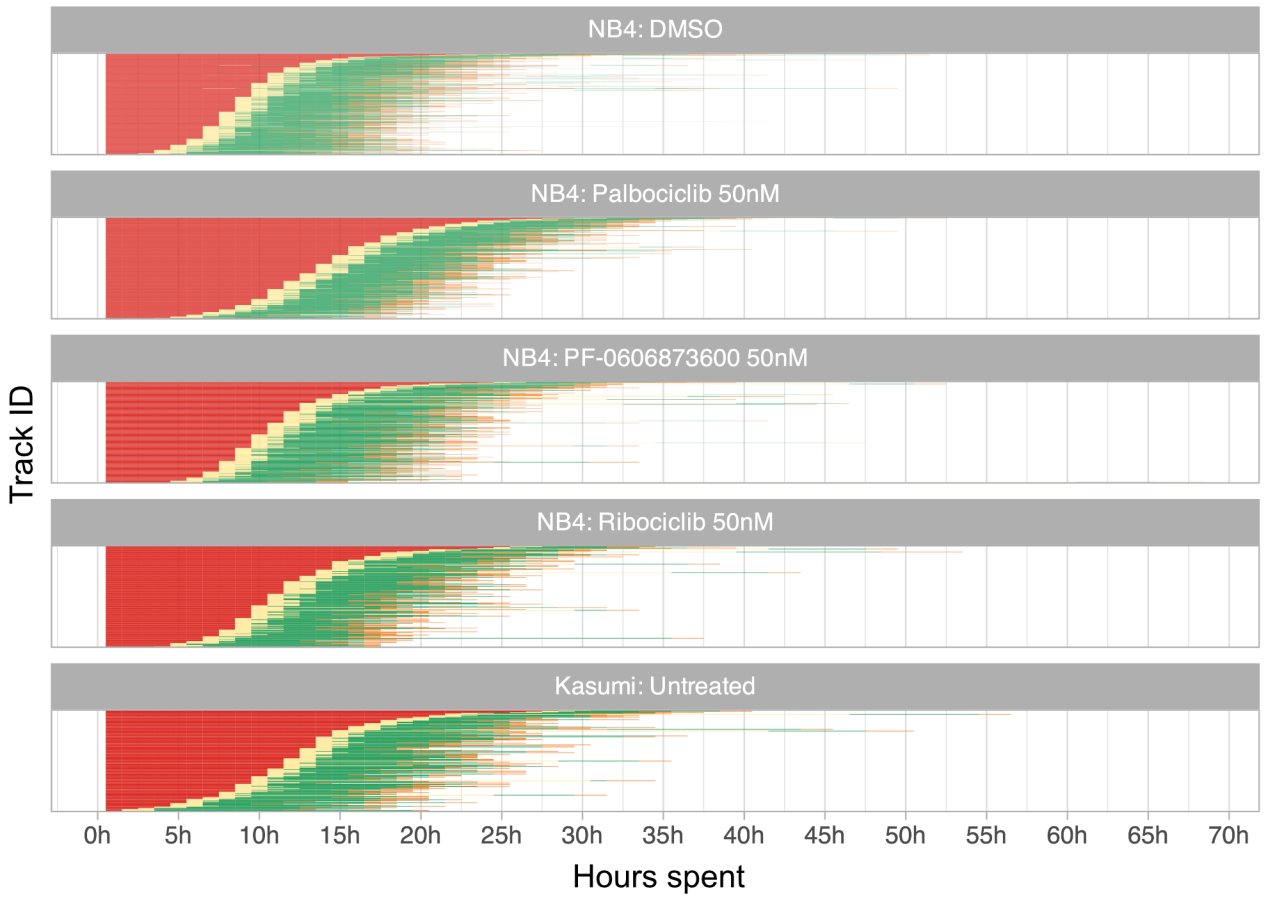

INFERENCE PIPELINE
Channel intensity
Time
Feature extraction
timetk
.rds
Vetiver
docker
untraceable
traceable

FIG 3

A


NB4: DMSO
NB4: Palbociclib 50nM
NB4: PF-0606873600 50nM
NB4: Ribociclib 50nM
Kasumi: Untreated
Track ID
0h
5h
10h
15h
20h
25h
30h
35h
40h
45h
50h
55h
60h
65h
70h
Hours spent

B

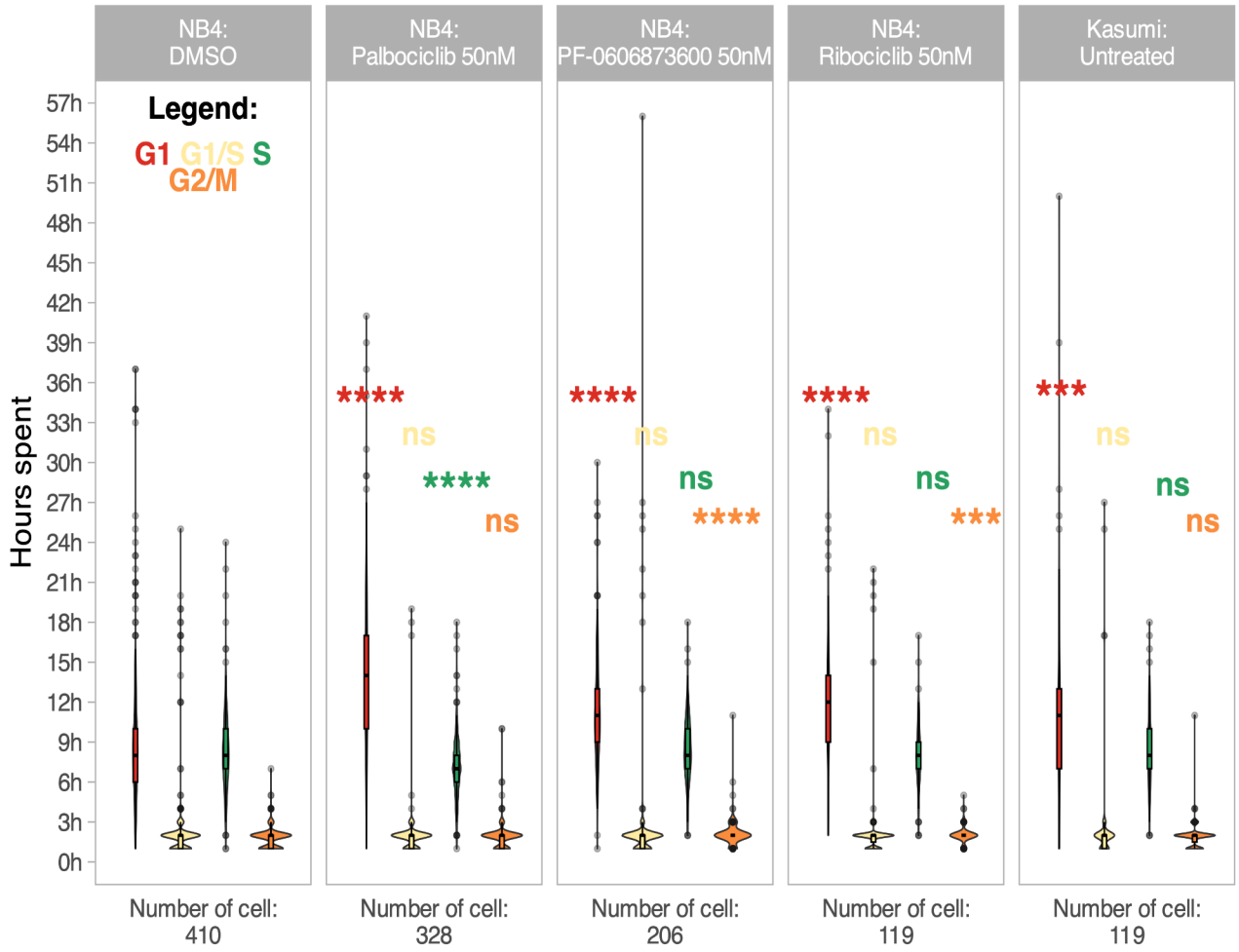

NB4: DMSO
NB4: Palbociclib 50nM
NB4: PF-0606873600 50nM
NB4: Ribociclib 50nM
Kasumi: Untreated
Legend:
G1 G1/S S
G2/M
Hours spent
57h
54h
51h
48h
45h
42h
39h
36h
33h
30h
27h
24h
21h
18h
15h
12h
9h
6h
3h
0h
****
ns
****
ns
****
ns
****
****
ns
ns
***
***
ns
ns
ns
Number of cell: 410
Number of cell: 328
Number of cell: 206
Number of cell: 119
Number of cell: 119

FIG 4

Workflow

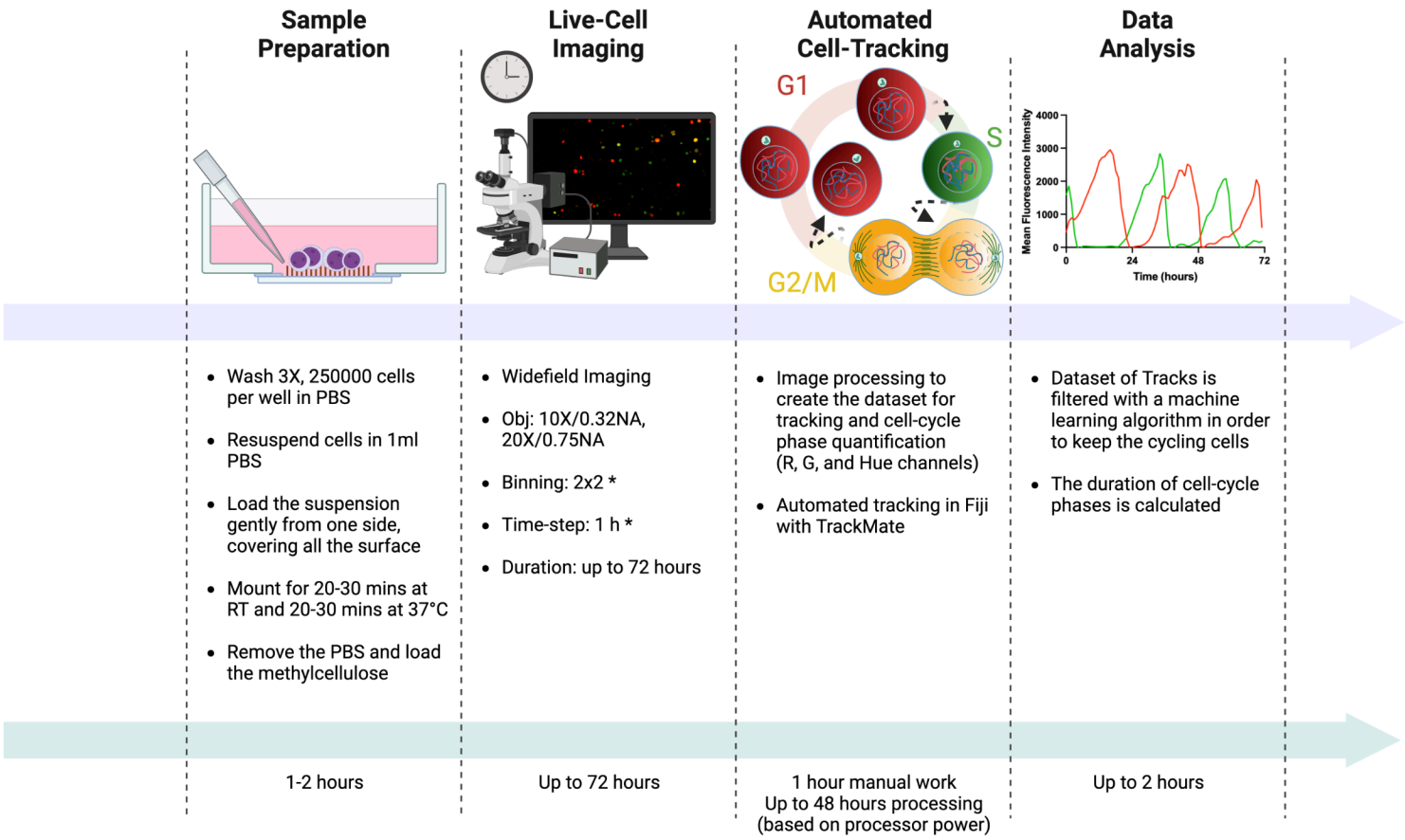


FIG S1

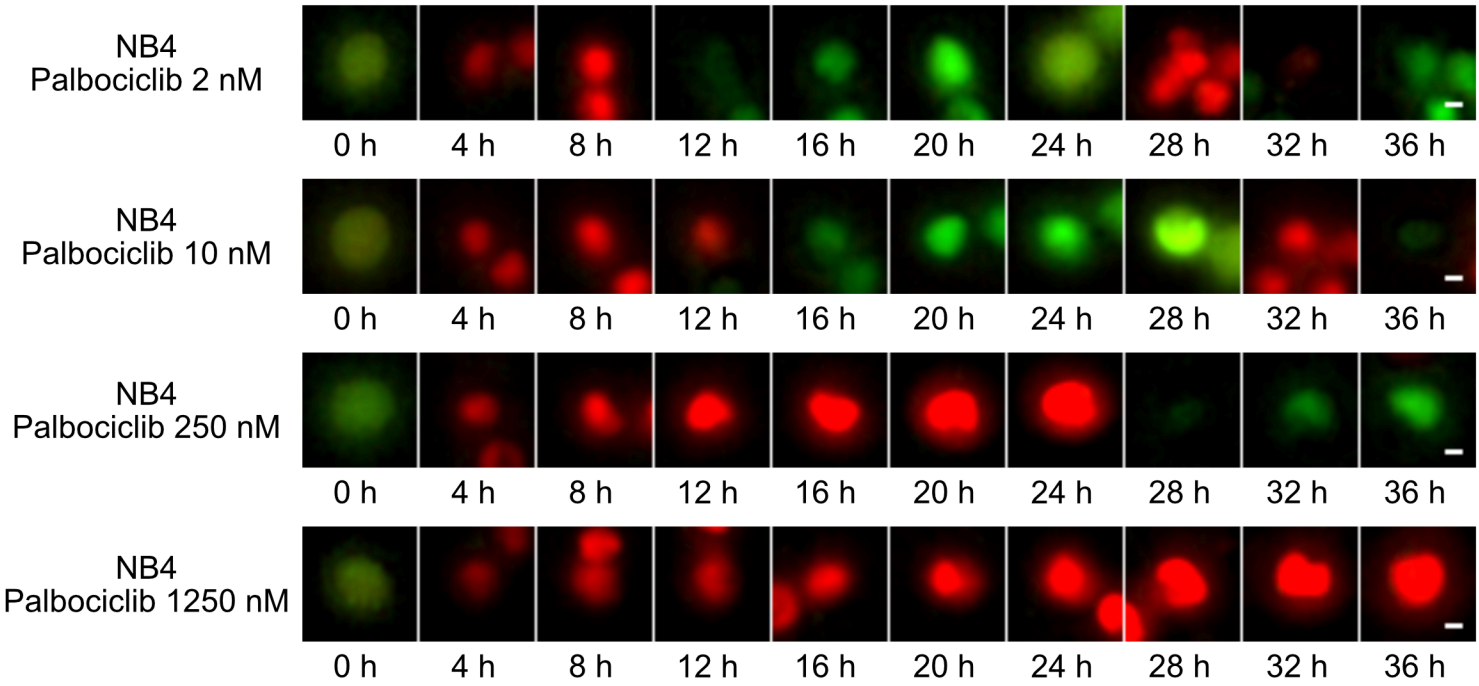

NB4
Palbociclib 2 nM
0 h
4 h
8 h
12 h
16 h
20 h
24 h
28 h
32 h
36 h
NB4
Palbociclib 10 nM
0 h
4 h
8 h
12 h
16 h
20 h
24 h
28 h
32 h
36 h
NB4
Palbociclib 250 nM
0 h
4 h
8 h
12 h
16 h
20 h
24 h
28 h
32 h
36 h
NB4
Palbociclib 1250 nM
0 h
4 h
8 h
12 h
16 h
20 h
24 h
28 h
32 h
36 h

FIG S2

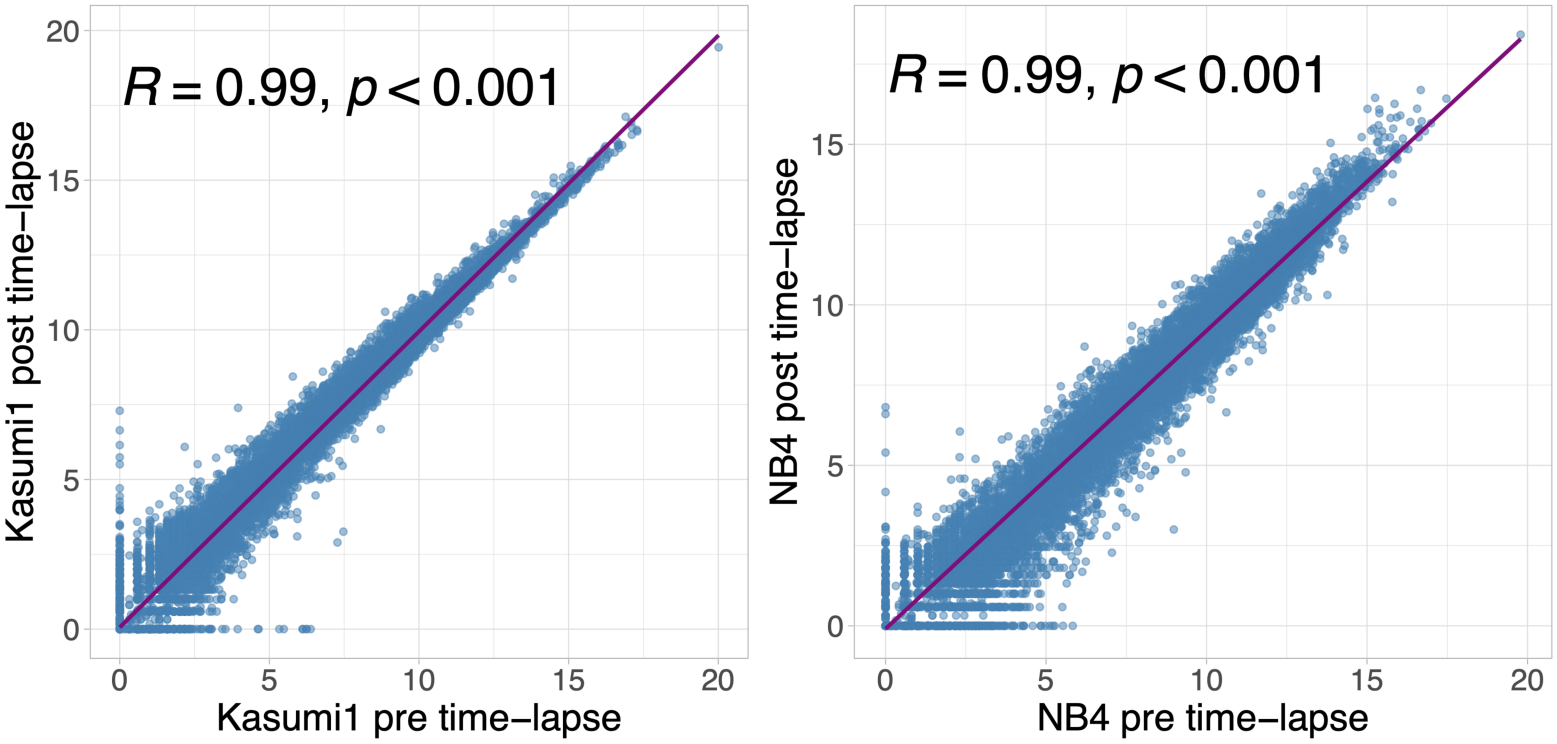

R = 0.99, p < 0.001
Kasumi1 post time−lapse
Kasumi1 pre time−lapse
0
5
10
15
20
R = 0.99, p < 0.001
NB4 post time−lapse
NB4 pre time−lapse
0
5
10
15
20

FIG S3

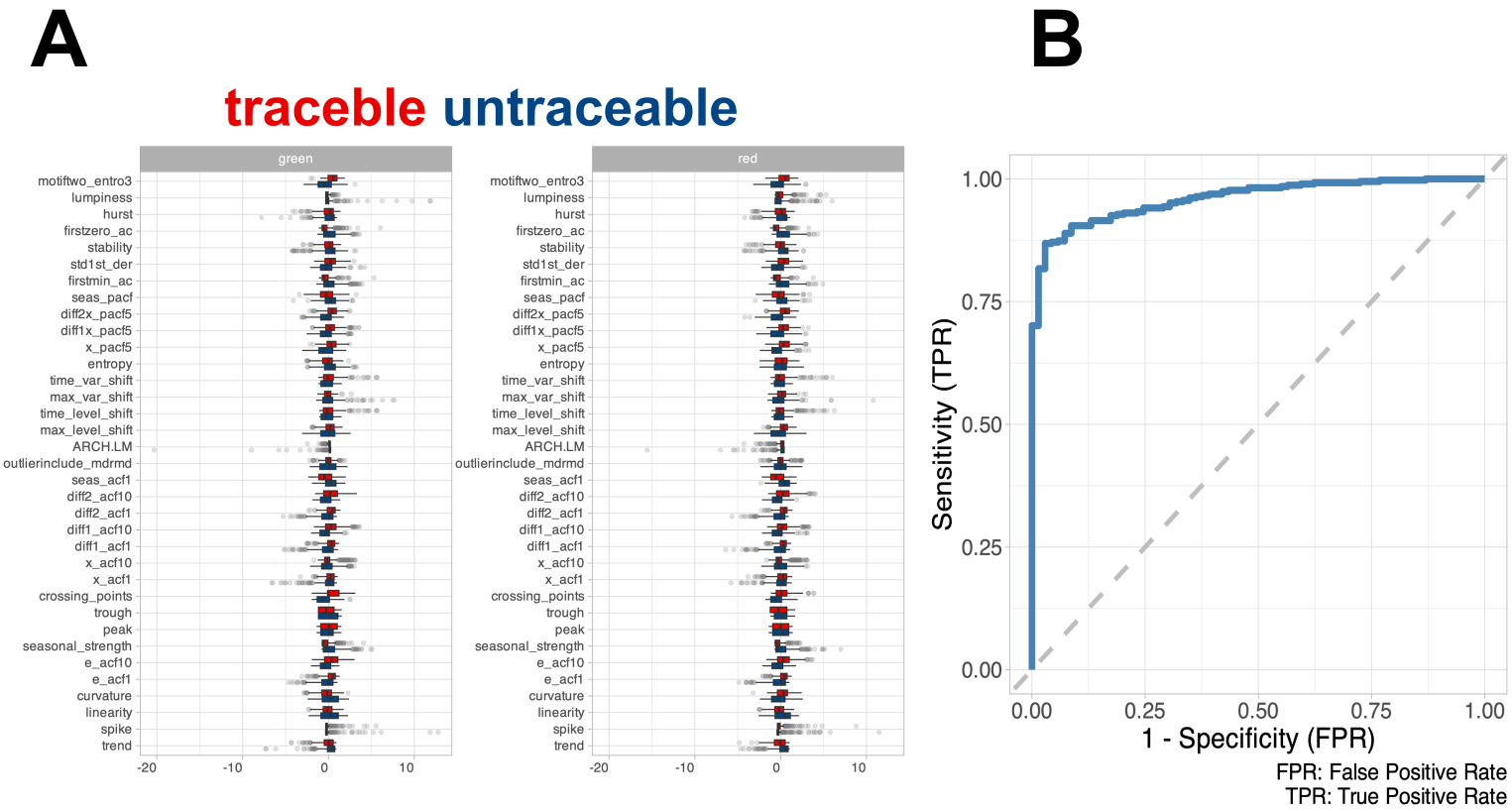
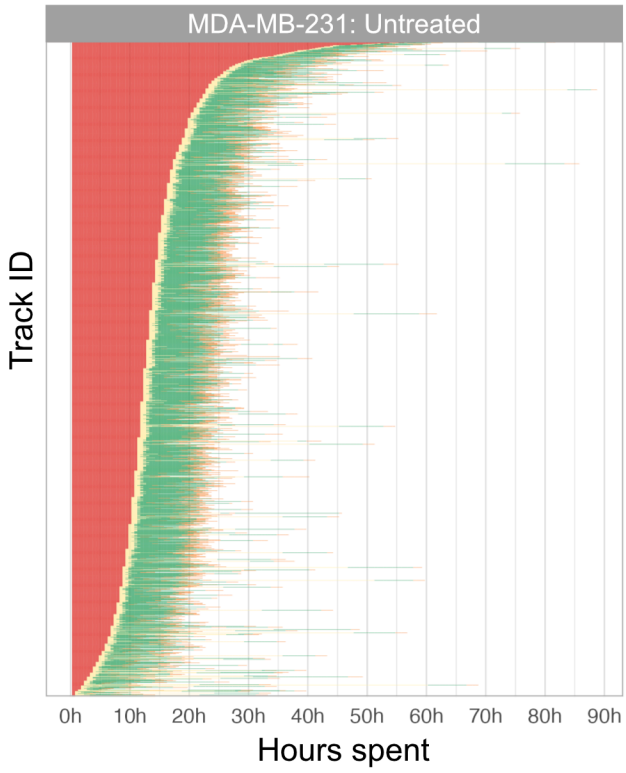
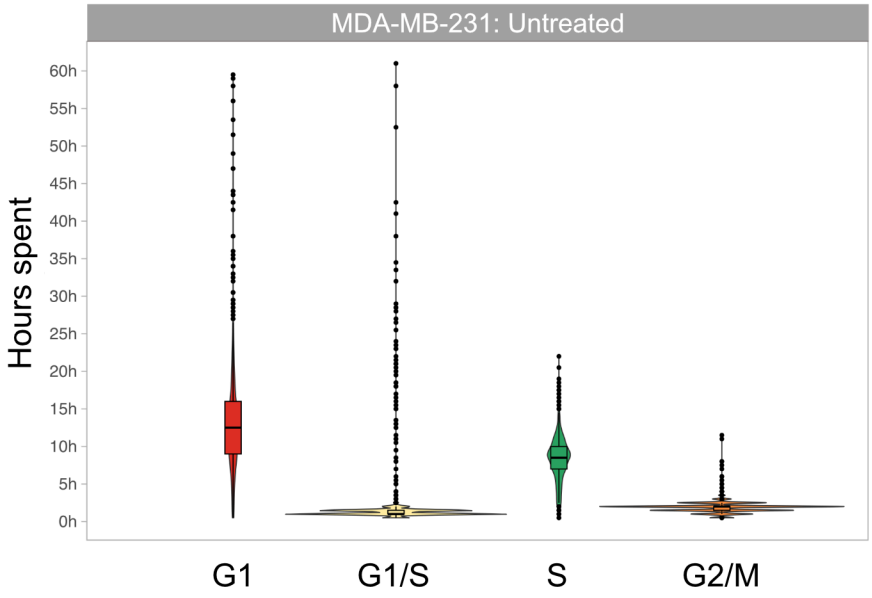

A
traceble untraceable
green
red
motiftwo_entro3
lumpiness
hurst
firstzero_ac
stability
std1st_der
firstmin_ac
seas_pacf
diff2x_pacf5
diff1x_pacf5
x_pacf5
entropy
time_var_shift
max_var_shift
time_level_shift
max_level_shift
ARCH.LM
outlierinclude_mdrmd
seas_acf1
diff2_acf10
diff2_acf1
diff1_acf10
diff1_acf1
x_acf10
x_acf1
crossing_points
trough
peak
seasonal_strength
e_acf10
e_acf1
curvature
linearity
spike
trend
-20
-10
0
10
B
Sensitivity (TPR)
1 - Specificity (FPR)
0.00
0.25
0.50
0.75
1.00
FPR: False Positive Rate
TPR: True Positive Rate

FIG S4

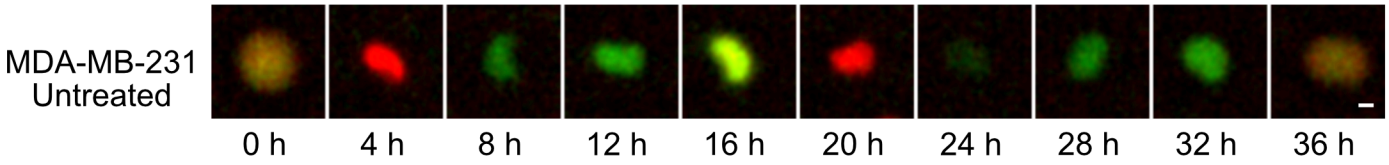
A


MDA-MB-231: Untreated
Track ID
0h
10h
20h
30h
40h
50h
60h
70h
80h
90h
Hours spent

B


MDA-MB-231: Untreated
Hours spent
60h
55h
50h
45h
40h
35h
30h
25h
20h
15h
10h
5h
0h
G1
G1/S
S
G2/M

C


MDA-MB-231
Untreated
0 h
4 h
8 h
12 h
16 h
20 h
24 h
28 h
32 h
36 h