In vivo cell cycle profiling in xenograft tumors by quantitative intravital microscopy

Citation

Published Version

Permanent link

Terms of Use

Share Your Story

Accessibility

In vivo cell cycle profiling in xenograft tumors by quantitative intravital microscopy

Deepak R Chittajallu[1,6], **Stefan Florian**[2,6], **Rainer H Kohler**[3], **Yoshiko Iwamoto**[3], **James D Orth**[4], **Ralph Weissleder**[2,3], **Gaudenz Danuser**[1,5], and **Timothy J Mitchison**[2]

[1]Department of Cell Biology, Harvard Medical School, Boston, Massachusetts, USA

[2]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA

[3]Center for Systems Biology, Massachusetts General Hospital, Boston, Massachusetts, USA

[4]Molecular, Cellular and Developmental Biology, University of Colorado Boulder, Boulder, Colorado, USA

Abstract

Quantification of cell-cycle state at a single-cell level is essential to understand fundamental three-dimensional biological processes such as tissue development and cancer. Analysis of 3D *in vivo* images, however, is very challenging. Today's best practice, manual annotation of select image events, generates arbitrarily sampled data distributions, unsuitable for reliable mechanistic inferences. Here, we present an integrated workflow for quantitative *in vivo* cell-cycle profiling. It combines image analysis and machine learning methods for automated 3D segmentation and cell-cycle state identification of individual cell-nuclei with widely varying morphologies embedded in complex tumor environments. We applied our workflow to quantify cell-cycle effects of three antimitotic cancer drugs over 8 days in HT-1080 fibrosarcoma xenografts in living mice using a dataset of 38,000 cells and compared the induced phenotypes. In contrast to 2D culture, observed mitotic arrest was relatively low, suggesting involvement of additional mechanisms in their antitumor effect *in vivo*.

Precise coordination of cell-cycle progression is essential to many biological phenomena, and it is of special interest in cancer research, where cell cycle deregulation is one of the disease hallmarks. Most of our knowledge in this direction stems from 2D cell culture systems. However, a large body of evidence shows that highly proliferating cells cultured on an artificial flat surface behave very differently compared to cells growing in a more

physiological 3D environment[1]. Therefore, a major challenge in cancer research is the development of accessible 3D model systems that accurately reflect *in vivo* cell-cycle behavior. Initially, this was addressed using 3D *in vitro* culture systems. More recently, intravital microscopy has allowed live imaging of tissues in living animals at single-cell resolution over time, opening the door towards 3D experiments in a real physiological environment[2–5]. Coupling this technology with recently developed fluorescent reporters of cell-cycle state[6] enables the *in vivo* study of cell cycle effects of experimental perturbations at single-cell resolution in both space and time. Today's best practice in interpreting such 3D microscopy data relies on visual inspection and manual quantification of select image events. This is tedious, prone to bias and limits us to small-scale studies resulting in arbitrarily sampled data distributions. Automated analysis of 3D microscopy data, especially in an intravital setting, is challenging because of the relatively poor image quality and the presence of cells with varying sizes, shapes and appearance in close contact with each other. Thus, while automated analysis is routine in the study of 2D monolayer cell cultures [7,8], the need for such tools for 3D image analysis is just beginning to be addressed[9].

Here, we introduce a workflow for automated *in vivo* cell cycle profiling that integrates a high-resolution intravital imaging setup for longitudinal observations of tissues with a computational framework for automated 3D segmentation and cell cycle state identification of individual cell nuclei with varying morphologies (Fig. 1). Firstly, we used a grid-based spatial reference system to noninvasively track multiple tissue locations, thereby generating a multidimensional dataset for studying tissue changes in space and time. Then, we used marker-controlled watersheds coupled with a supervised hierarchical learning-based region merging method for automatic 3D segmentation of cell nuclei and a supervised classification scheme for automatic identification of the cell cycle state of each cell based on image-derived features. In a proof of principle study, we quantified the effects of three antimitotic cancer drugs over 8 days and found that the induction of mitotic arrest was much lower than in 2D culture and each drug induced a characteristic effect on cell morphology suggesting additional, nonmitotic effects as mechanisms of action. While our workflow was developed with an eye towards our specific application of testing the effects of antimitotic drugs in xenograft tumors, it could be applied to any other problem in tissue biology or pharmacology where quantifying cell cycle progression is essential.

Results

Tumor model and imaging setup

We applied our quantitative imaging workflow to a xenograft tumor model based on the HT-1080[10] fibrosarcoma cell line implanted in a dorsal skin-fold-chamber (DSC) in nude mice[3–5], an established model in wide use for preclinical drug testing. For live detection of cell cycle state *in vivo* at the single-cell level, an HT-1080 clone with stable expression of a DNA morphology reporter (histone H2B-CFP) and the FUCCI fluorescent cell cycle reporter system[11] (G1 cells express a red fluorescent protein, S/G2/mitotic cells are green) was generated (Fig. 1). Preliminary studies with this cell line showed a mean segmentation accuracy of 83.84% in the crowded environment of xenograft tumors (Supplementary Fig. 1). To further improve accuracy, we reduced fluorescent cell density by mixing fluorescent

cells with the non-fluorescent parental cell line. To reliably identify the same tumor region during consecutive imaging sessions, a gold grid was placed on the tumor one day before drug injection, and the same 3–9 positions were imaged before and multiple times after drug injection (Fig. 1). For imaging, a two-photon/laser-scanning confocal microscope with a customized heated stage for mounting of DSC was used.

Nuclei segmentation

Segmentation of cell nuclei in 3D intravital images of complex tumor environments is very challenging due to high cell density and variability in nuclear size, shape and appearance, exacerbated further by drug action [12–16]. Our solution couples marker-controlled watersheds to a novel supervised hierarchical learning-based region merging method (Fig. 2a). Only the H2B-CFP signal was used for segmentation as it presents all nuclei irrespective of the cell cycle state and enables application of our segmentation algorithm in combination with other cell-state reporters than FUCCI.

Firstly, we used a locally-adaptive thresholding algorithm to obtain coarse foreground-background segmentation. Next, we detected seed points in cell nuclei using a 3D extension of a scale-adaptive multi-scale Laplacian-of-Gaussian (LoG) filter[17] and used them as markers for a marker-controlled watershed algorithm [18,19] to obtain an initial segmentation result. This, however, was suboptimal, especially in cases where nuclei of different sizes and/or shapes touch each other and when nuclear shape is aspherical. Both scenarios often occur in cancerous tissues: the latter leads to over-segmentation errors, while the former leads to under-segmentation errors. These problems were increasingly profound in images taken after drug treatment. To address this, we intentionally over-segmented the nuclei, by tuning parameters of the seed detection algorithm accordingly, and then corrected them using a supervised hierarchical learning-based algorithm that finds all pairs of spatially adjacent regions and presents them to a discriminative two-class classification model based on image-derived features (Supplementary Table 1). The pairs the classifier recommends for merging are iteratively merged in decreasing order of merge probability.

We trained and evaluated the segmentation algorithm using a dataset of 54 volumes (7717 nuclei) containing data from different locations of tumors grown in six mice treated with one of three antimitotic drugs. We obtained a segmentation accuracy of 91.04+/−4.87% with a range of 77.61–97.89% (Figs. 2b-d, Supplementary Table 2). The region merging algorithm improved the segmentation accuracy by 7% (Supplementary Table 3) and cell density was a key impeding factor (Supplementary Fig. 2). We also validated the algorithm on datasets from different organisms or cell lines acquired with light-sheet, spinning-disk and laser-scanning confocal microscopes (Supplementary Fig. 3). On *Drosophila*, zebrafish and mouse embryo datasets, the accuracy was consistently over 90%, similar to other solutions[20]. On undiluted tumor spheroids from 3 cancer cell lines, the accuracy was ~85%, reflecting the special challenges that high density and irregularities in nuclear shape present, similar to our *in vivo* data. Thus, our algorithm provides consistently good performance on datasets from developing embryos and diluted tumor samples. It achieves slightly lower, but still acceptable performance, in undiluted tumor samples. The performance is also consistent across a wide range of microscopy equipment.

Automated cell cycle state identification

Manual inspection and annotation of the cell cycle state at a single-cell level is very slow, tedious and often unreliable in 3D images. Hence, we developed a fully automated supervised classification scheme to identify the cell-cycle state of each detected cell based on image-derived features extracted from the histone and/or FUCCI channels (Fig. 3, Supplementary Table 4). With the FUCCI reporter, nuclei of cells in G1 appear red and change to green during the S, G2, and mitotic phases with an interim period of yellow/orange marking the transition from G1 to S, henceforth referred to as Late-G1/Early-S state. Mitotic cells can be recognized by their spherical shape after nuclear envelope breakdown and their distinct chromatin texture in the histone channel. Nuclei in Early-G1 were not clearly visible in the FUCCI channels and were only seen in the histone channel.

Specifically, we built a four-class model that associates a given cell with: (i) G1, (ii) Late-G1/Early-S, (iii) S/G2, or (iv) mitosis. To support studies interested in monitoring cell cycle progression at a coarser level, we used the same classification scheme to build two dedicated and more accurate two-class models: (i) Interphase (G1+Late-G1/Early-S+S/G2) vs mitosis, and (ii) G1 vs S+G2+mitosis.

We trained and validated the cell cycle state identification models using the above-mentioned dataset of 54 volumes. We obtained an accuracy of 90.76% and an average F-Measure=0.87 for the four-class model (Fig. 3d, Supplementary Tables 5–7). For the two-class models we obtained an accuracy of 97.42% with an average F-Measure=0.93 for the Interphase vs mitotic model (Supplementary Tables 8–10), and an accuracy of 95.63% with an average F-Measure=0.95 for the G1 vs S+G2+Mitotic model (Supplementary Tables 11–13).

Quantitative analysis of cancer drug effects in space and time

To demonstrate the value of the proposed quantitative imaging workflow, we used it to track the effects of cancer therapeutics on cell cycle progression. This is a particularly difficult problem for an automated imaging pipeline due to the large and unpredictable morphological changes these drugs elicit. We focused on testing effects of cytotoxic cancer drugs that kill cells in mitosis. Most of our current understanding of the mechanism of action of these antimitotic drugs is based on experiments conducted in 2D *in vitro* cell culture systems that are often inaccurate or incomplete in predicting drug efficacy in patients. While their cytotoxic effects in 2D cell culture are strictly mitosis dependent, increasing evidence suggests that this is likely not the case in human tumors[21,22]. To better understand clinical efficacy, we need to measure pharmacodynamics in more realistic environments while maintaining single-cell resolution, as cell cycle arrest and cell death display complex and variable kinetics at the single-cell level[23]. We first analyzed the single-cell pharmacodynamics of the established microtubule (MT)-interacting cancer drug paclitaxel, used in the treatment of breast, ovarian and lung cancer. We followed six tumor positions in a paclitaxel-treated mouse for seven days or as long as the imaging conditions allowed (Fig. 4a, b, c). With one exception, the trends in cell density over time were quite similar at all imaged positions (Fig. 4c). One position imaged outside the gold grids showed a substantial increase in tumor cell density from day 1 to day 2 after drug treatment, which may be due to

the imaging of slightly different locations on different days. This highlights the need for the grid based spatial reference system for accurate tracking of parameters in a defined tumor location. Overall, we measured a downward trend in cell density after treatment with paclitaxel, suggesting that the drug is effective. Cell cycle state quantification showed that mitotic cells accumulated after drug injection, peaking on day 2 (Fig. 4c). This was expected for a drug that greatly increases the duration of mitosis by activating the spindle-assembly checkpoint[24]. However, compared to 2D culture models (with up to 80% mitotic cells after 1 day of drug exposure), in our experiments this arrest was much more modest, in line with other *in vivo* studies[3,25–27]. To validate the results of the whole imaging and analysis pipeline, we stained tissue sections from paclitaxel treated HT-1080 tumors for the mitotic marker phospho-histone H3 and quantified mitotic arrest by histology as an orthogonal approach (Fig. 4d, Supplementary Fig. 4). The mitotic arrest observed was similar, but slightly lower because the histological approach counts also non-dividing mouse stromal cells while the fluorescent reporters used for intravital microscopy are expressed only in the human graft cells. The proposed intravital imaging pipeline was further validated with flow cytometry of cells from drug treated cancer cell spheroids, an established system for 3D cell culture (Supplementary Fig. 5). Together, these orthogonal validation assays confirmed that application of paclitaxel *in vivo* induces a relatively low mitotic arrest when compared to 2D culture.

To enable more detailed visual inspection of drug effects over time at single-cell resolution, we arranged cell thumbnail images in a montage, grouped by timepoint and predicted cell cycle state (Fig. 4e; see Supplementary Fig. 6 for a high resolution version). The distinct color patterns for G1, Late-G1/Early-S and G2 cells allow an immediate, intuitive quantification of the cell cycle distribution: The increase in mitotic cells on days 1 and 2 was accompanied by a concomitant depletion in G1 (red) and early S (yellow cells), probably the result of delayed progression of mitotic cells into the next cell cycle. Beginning with day 3, the trend was reversed. A strong increase in G1/early S cells likely reflects the synchronous transition of the arrested mitotic cells into the next cell cycle. Faulty mitotic exit, called mitotic slippage, is a frequent outcome of prolonged mitotic arrest[28,29]. Combined with the metadata and the features calculated for each cell this data presentation approach supports qualitative validation of results and formulation of new hypotheses on the mechanism of drug action *in vivo*.

Antimitotic cancer drugs differ in pharmacodynamics and efficacy

We applied the same analysis workflow to compare the cell cycle effects of two additional antimitotic cancer drugs with different mechanisms of action and clinical efficacies. MT drugs have been in clinical use as cancer therapeutics for decades and remain among the most successful chemotherapeutic drugs available[24]. Because they perturb MT dynamics in all cells of the body, independent of cell cycle state, they also have strong neurotoxic side effects. This spurred development of non-MT-interacting molecules that induce mitotic arrest through different targets, including the mitotic spindle kinesin Kif11/Eg5/KSP. While indeed lacking neurotoxicity, most of these drugs were not effective enough to successfully pass Phase III clinical trials[22,30]. Tissue culture experiments failed to predict the striking differences in anti-tumor potency observed in patients[28,29], which we still do not understand.

We therefore treated mice carrying HT-1080 xenograft tumors with eribulin, a novel MT binding drug used in late stage breast cancer[31], and compared the effects of this drug and paclitaxel to the effects of ispinesib[30], a novel experimental KSP inhibitor that failed to achieve approval for clinical use (Fig. 5, Supplementary Figures 6–8). While the cell cycle profiles show a high amount of variability in timing and potency, in six out of seven mice (two eribulin; two paclitaxel; three ispinesib) a single dose (~60% of the published maximally tolerated dose) of drug induced a highly significant mitotic arrest of up to 20% of the total cell population. Three out of four mice treated with MT drugs showed a reduced mean cell density after six days when compared to the pre-drug time point, while in all mice treated with ispinesib the density was unchanged, suggesting lower antitumor efficacy of ispinesib in this system, similar to the clinical experience. In summary, our experiments show a similar and relatively low mitotic arrest for all three antimitotic drugs, suggesting that the reduction in tumor mass induced by the MT targeting drugs may involve other, non-mitotic mechanisms. This is in line with previous *in vivo* studies on docetaxel[25] and paclitaxel[3]. To draw definitive conclusions about drug efficacy in this xenograft model, longer post treatment measurements of tumor volume and analysis of a higher number of mice will be necessary.

We sought image features other than mitotic arrest that might explain these findings: Chromosomal arrangement in cells arrested in mitosis differed between the three antimitotic drugs (Figs. 6a-c). It has been proposed that, because it is immediate and independent of mitotic entry and exit, disturbed chromosome alignment in mitosis is a better pharmacodynamic marker for antimitotic drugs than increase in mitotic cells[32]. About 20% of mitotic cells on day 3 after injection of ispinesib showed a normal metaphase plate, and by day 6 the fraction of cells with normally aligned chromosomes in one mouse grew to 60%, approximately the same as before drug treatment, indicating that the drug was no longer active (Fig. 6b). In paclitaxel- and eribulin-treated mice, however, both on days 3 and 6, virtually all mitotic spindles had misaligned chromosomes, suggesting that a single dose of these drugs acts for much longer periods of time. This is consistent with tissue culture studies showing taxanes and eribulin both tend to persist in cells, presumably bound tightly to MTs or tubulin[33,34]. Thus, by measuring pharmacodynamic markers live in the tumor, we could estimate drug activity in the tumor over time.

G1 cell morphology after three days was also very different between the three drugs (Fig. 6d). MT drug treatment resulted in a significant increase in multinucleated nuclei, whereas ispinesib did not affect G1 cell morphology. The multinucleated phenotype suggests that MT drugs either have additional direct effects on interphase cells or induce weaker mitotic arrest with a tendency towards mitotic slippage (i.e. premature mitotic exit) and faulty division into multinucleated G1 cells. Thus, single doses of paclitaxel and eribulin seem to have longer activity in nude mouse xenografts than ispinesib, possibly a reason for the lack of clinical efficacy of the latter. The drugs also differ in their effects on G1 cells as early as day 3 after drug injection, which may further contribute to the observed differences in antitumor activity. Further studies will be necessary to build on these preliminary observations. Our automated, quantitative approach now enables such systematic studies.

Discussion

Although advances in microscopy technology have led to wide adoption of 3D imaging in tissue culture and intravital microscopy, image analysis is still tedious and a limiting factor in application of these techniques. It is difficult to score enough events to capture the full biological heterogeneity in tissues and often, unsuitable 2D culture systems are used as a substitute, limiting 3D studies to proof of principle experiments focusing on a few selected cellular events in a 3D context.

Here, we introduce an integrated workflow to quantify the cell cycle effects of experimental perturbations on living tissues *in vivo* up to 8 days, preserving the full spatial information in the 3D reference system of a xenograft tumor. We combine intravital microscopy optimized for longitudinal observations of cell fate with a novel computational framework for automated image analysis. The computational framework uses image analysis and machine learning techniques in tandem for the robust segmentation and cell cycle state identification of individual cell nuclei, even when embedded in a highly complex tumor environment. It reliably detects cells based on DNA signal that is very heterogeneous in appearance typical in drug treated xenograft tumors, such as mitotic cells with different chromosome morphology (Figs. 2d, 6a, 6c), and other model systems (Supplementary Fig. 3). The proposed workflow has substantial advantages over existing methods (flow cytometry, immunohistochemistry) to assess the cell cycle state *in vivo*. (i) Because it is noninvasive and it does not require the sacrifice of a mouse for each measurement, our approach allows truly longitudinal observations of tumors after drug treatment at the single-cell level and also reduces the number of mice required for longitudinal observations. (ii) In addition to the quantification of cell cycle state distribution and cell density, a multitude of parameters at a single-cell level (e.g., position, volume, shape, appearance) are calculated automatically from 3D image stacks. (iii) The display of individual cells filtered based on metadata and calculated features in a 2D montage supports the conception and quick qualitative validation of new biological hypotheses and exploration of unexpected phenotypes. Overall, our workflow facilitates unbiased, consistent, and reproducible studies at a large scale. All elements required for the workflow featured here can be adapted to address different experimental questions. Therefore, we anticipate that our framework will contribute to faster, more consistent and richer quantitative analysis of 3D image datasets from living tissues.

Online Methods

Generation of cell line with stable reporter expression and cell culture

The HT-1080 cell line (ATCC) was grown in Eagle's Minimum Essential Medium (EMEM, ATCC 30-2003) supplemented with 10% FBS (Gibco) and 100 units/mL penicillin and 100 μg/mL streptomycin (Gibco) at 37°C with 5% CO_2. The FUCCI system allows discrimination between G1 and S/G2/M phases of the cell cycle in living cells using a fragment of the G1 specific protein hCdt1 labeled with a red fluorescent protein (mKO2 - hCdt1) and a fragment of geminin labeled with a green fluorescent protein (mAG - hGem), expressed during S, G2 and M phases of the cell cycle [6]. The H2B CFP signal can be used to detect chromosome condensation during mitosis[3] and thus distinguishes interphase from

mitosis. To generate a HT-1080 population stably expressing the FUCCI reporter proteins and H2B CFP, the cell line was serially transduced with lentiviral particles coding for mKO2-hCdt1 (neomycin resistance gene), mAG-hGeminin (blasticidin resistance gene) and H2B CFP (hygromycin resistance gene), respectively and kept under continuous selection using the corresponding antibiotics for at least three weeks (1 mg/ml neomycin, 10 μg/ml blasticidin, and 100 μg/ml hygromycin). Fugene6 (Roche) was used for transfection of plasmids into HEK293T packaging cells according to the manufacturer's protocol. After transfection of viral production plasmids, medium was replaced after 12 hours, and viral supernatant harvested 24 hours later. Cells were transduced and expression of the fluorescent reporter was verified by fluorescence microscopy. A stable, clonal population of cells expressing all three reporter proteins as brightly as possible and with uniform intensity was obtained by limiting dilution to isolate single cell clones on a 96 well plate under antibiotic selection.

MCF7 and T47D cell lines were obtained from ATCC and grown in RPMI 1640 (Gibco) supplemented with 10% FBS (Gibco) and 100 units/mL penicillin and 100 μg/mL streptomycin (Gibco) at 37°C with 5% CO_2. Stable H2B GFP expressing cell lines were generated by K. Krukenberg (Harvard Medical School, Boston, MA) following standard protocols. All cell lines were obtained as mycoplasma free aliquots from ATCC and not tested in house for mycoplasma contamination.

Mice

All procedures and animal protocols were approved by the subcommittee on Research Animal care at Massachusetts General Hospital. Nu/Nu mice (Cox-7; Massachusetts General Hospital) were fed 5 ml antibiotic (sulfamethoxazole/trimethoprim 200 mg/40 mg per 5 ml; Aurobindo Pharma) in 250 ml drinking water with weekly changes during the length of the experiment. For dorsal skinfold chamber (DSC) implantation, cell injection, injection of drugs, and microscopy, mice were anesthetized by isoflurane vaporization (Harvard Apparatus) with 2.0 L/min isoflurane: 2.0 l/min oxygen. After 30 min of imaging, the isoflurane flow rate was slowly reduced to less than 1.5 l/min. One hundred microliters of saline was injected i.p. every hour during imaging to maintain hydration. Surgery was conducted under sterile conditions with a zoom stereomicroscope (Olympus SZ61). To reduce discomfort mice were treated with analgesic (buprenorphine 0.1 mg/kg, Patterson Veterinary Supply) prior to surgery and every 12 hours for three days after surgery.

DSC implantation

Titanium DSCs (APJ Trading Co, Inc.) were implanted into the dorsal skinfold of Nu/Nu mice as described[3] The DSC stretches and sandwiches the 2 layers of skin on the back of the mouse. On one side, the skin is surgically removed and replaced by a 10-mm-diameter optical glass cover slip held in place with a c-clip. Spacers between the two halves of the DSC frame prevent excess compression of the tissue and vessels. The window allows free-access imaging of the remaining layers of striated skin muscle, subcutaneous tissue, deep dermis, and tumors.

Cell injection into DSC

Cells were harvested by trypsinization (0.25% trypsin-EDTA) and resuspended in growth medium. Preliminary studies with this three-color cell line grown in xenograft tumors showed a mean segmentation accuracy in the crowded tumor environment of 83.84% (Supplementary Fig. 1). We therefore reduced fluorescent cell density by mixing fluorescent cells 1:20 or 1:35 with the non-fluorescent parental cell line. These ratios were chosen as an empirical optimum between segmentability and having enough cells per image to collect reliable statistics (Supplementary Fig. 2), and resulted in an actual proportion of 20–80% of labeled tumor cells after engraftment. After DSC implantation and cell injection, tumors were allowed to vascularize and grow for 2–3 weeks with the non-fluorescent parental cell line. Mice were anesthetized and approximately 2×10^6 (50 µl) cells were injected subcutaneously into the DSC using a 0.5 ml insulin syringe (29G; BD Biosciences). The needle was bent at 90 degrees to aid injection. Injections were carried out under a stereomicroscope. After injection, sterile saline was added into the DSC and it was closed with a new cover slip. Before imaging, DSCs and tumors were evaluated and rejected if there were any gross tissue abnormalities, the possibility of tissue movement, necrosis or infection.

Grid implantation

To allow reliable identification of the same tumor region during consecutive imaging sessions we tested different techniques such as the use of fluorescent fiducial beads, vascular landmarks, relative position to the external metal frame of the DSC and the use of various metallic grids. Gold grids with mesh size of 75G routinely used in electron microscopy proved most reliable for imaging the same regions repeatedly. One day before the first imaging session the cover slip was removed, and a 75 mesh gold grid (G-75-G; Energy Beam Science, CT, USA) was carefully positioned onto the tumor, covered with saline, and the DSC closed with a fresh cover slip. The grid attaches to the tumor and tumor cells grow between and over the gold grid. As a result the tumor moves with the grid if it shifts in small increments. The grid size was selected so that the imaging field (25x objective with 2 zoom) is just inside of the grid openings and does not include the gold frame.

Drug treatment

Random animals were assigned to each treatment group. To allow neovascularization, HT-1080 xenografts were grown for 2 to 3 weeks before drug treatment. For our experiments, we selected a single dose between 50–70% of published maximally tolerated doses. Mice treated with paclitaxel were injected in the tail vein with a single bolus of 30 mg/kg paclitaxel (injectable formulation, Novaplus; Bedford Laboratories). A total of 30 mg/kg paclitaxel was used based on published data[26,35]. Eribulin (Halaven) was injected in the injectable formulation (Eisai) into the tail vein at a dose of 1.2 mg/kg[31]. The KSP inhibitor SB-715992 (ispinesib, Seleckchem) was dissolved in 10% Ethanol, 10% Cremophor EL (Sigma Aldrich) and 80% D5W (5% w/v aqueous dextrose solution) according to published protocols[30] and injected intraperitoneally. The dose of 20 mg/kg body weight was selected based on the data in AACR 2002, Poster 1335, available from

http://www.cytokinetics.com/pdf/AACR_2002_Poster_1335.pdf, where the MTD in female BDF mice was determined as 36 mg/kg.

Microscopy

To reduce motion artifacts and permit high-resolution microscopy over extended periods, a custom made DSC holder was used[3] The setup consists of an aluminum holder attached to an aluminum platform. The platform and DSC holder were kept at 38°C to reduce thermal drift. The screws of the DSC fit into machined holes in the holder and the DSC frame was further immobilized using small plates and screws. The microscope was mounted on a floating air table to eliminate vibration external to the instrumentation. A customized Olympus FV1000 confocal/multi-photon microsocope was used. Olympus objectives were as follows: 25× XPlan N [numerical aperture (NA) = 1.05, water], 2×/340 XLFluor (NA = 0.14, air).

Because of the significant advantages of 2 photon excitation in depth penetration and tissue we tried to excite all 3 channels using 2 photon excitation. However, with the lasers currently available in our laboratories, it was not possible to efficiently excite mKO2 and achieve clean channel separation with 2 photon imaging. Thus, to achieve the best image quality possible for nuclear segmentation, the H2B channel was acquired using two-photon microscopy while the red and green FUCCI channels were re-acquired in a second run using single-photon laser-scanning confocal microscopy.

Monomeric Azami Green (mAG) and monomeric Kusabira Orange (mKO2) were excited using a 473 nm or 559 nm pumped diode laser, respectively, in combination with a DM405/473/559-nm dichroic beam splitter. Emitted light was separated and collected with beam splitters SDM560 and SDM640 and band-pass filters BA490-540 and BA575-620. In addition, each z-stack was acquired using two-photon excitation at 830 nM with a FV10-MRCYR/XR filter cube (Olympus) to image H2B CFP. Using a motorized xy stage, three to ten z-stacks of 20 to 50 optical sections (2 μm per section), were collected. With these acquisition conditions, in some cases, the faster attenuation of green fluorescence intensity with depth vs. red may cause some yellow (Late-G1/early S) cells to appear red. If this issue is relevant to the question addressed, our computational framework offers cell cycle models where red and yellow cells are pooled to avoid this possible issue. The grid based tracking method used here is only suitable for treated tumors or tissues with low proliferation rates. If untreated, proliferating tumor cells tend to overgrow the grid within a few days. Image quality during intravital imaging can be affected by many factors, like edema formation, growth of fibrous tissue over the tumor, tissue drift (i.e. drift of the tumor tissue away from the cover slip), all of which can also be induced by drug treatment. When assessing the growth or recession of tumorous tissue, it is important to ensure that changes in cell count are not due to such phenomena. This can be reliably judged by the appearance of the images: very weak contrast or, in general, uniform reduction in the signal to noise ratio reflects one of the problems mentioned above. When individual positions showed such problems at late timepoints, tracking was stopped. In general, rarely encountered problems with image quality up to day 7 post-treatment, but later timepoints showed major variation, and we

would not recommend experiments longer than 8 days. In other xenograft systems or when unperturbed tissues are imaged, longer imaging may be possible.

Histopathological analysis

Unlabeled HT-1080 tumors were either left untreated or treated with a single dose of 40 mg/kg paclitaxel i.v. for 1, 4 or 7 days. Three tumors for each condition were embedded in O.C.T. compound (Sakura Finetek) and serial 6 μm-thick frozen sections were prepared for histopathological analysis. For immunohistochemistry, the tissue sections were treated with 0.3% H_2O_2 in dH_2O to suppress endogenous peroxidase activity, and then blocked using 4% goat normal serum in PBS for 30 minutes at room temperature. The sections were incubated with a primary antibody, phospho-histone H3 (Ser10) (1:200, #9710, Cell Signaling), overnight at 4°C. The following day, the sections were washed in PBS three times for five minutes each and incubated with biotinylated anti-rabbit IgG (1:100, BA1000, Vector Laboratories) for 30 minutes at room temperature. After washing the sections in PBS three times for 5 minutes each, VECTASTAIN ABC kit (Vector Laboratories) was applied according to the manufacturer's protocol, and a 3-amino-9-ethylcarbazole (AEC) substrate (Dako) was used for color development. All the sections were counterstained with Harris hematoxylin (Sigma-Aldrich) and images were captured using NanoZoomer 2.0RS (Hamamatsu).

Spheroid culture experiments and imaging

Cancer cell line spheroids were grown as described[36]. Briefly, 10000 cells per well were seeded in 96 well plates (Greiner Bio-One, 655090) coated with 1.5% agarose dissolved in complete growth media and polymerized at room temperature. Half the media was exchanged every 2–3 days. Three to six days after plating, spheroids were moved for imaging to 96 well plates coated with Poly-HEMA as described[37]. At this point, the spheroids had usually reached a diameter >300μM. Imaging was performed on a Nikon Ti motorized inverted microscope equipped with a Yokogawa CSU-X1 spinning disk confocal head with Spectral Applied Research Aurora Borealis modification and Spectral Applied Research LMM-5 laser merge module with AOTF controlled solid state lasers. A 445nm (80mW) laser was used to excite CFP, a 488nm (100mW) laser for mAG and GFP and a 561nm (100mW) laser to for mKO2. Images were acquired with a Hamamatsu ORCA-AG cooled CCD camera controlled by MetaMorph image acquisition software in whole media at room temperature. The following lasers and filter sets were used: CFP: Excitation: 447nm, Dichroic: Triple 447/515/642, Emission: 480/40; mAG/GFP: Excitation: 491nm, Dichroic: QUAD 405/491/561/642, Emission: 525/50; mKO2: Excitation: 561nm, Dichroic: QUAD 405/491/561/642, Emission: 620/60. The typical depth of z-stacks was 40–50μM, binning was set to 1 and the z spacing was 2μm.

Flow cytometry of spheroids

In the flow cytometry validation experiments (Supplementary Fig. 5), 96 HT-1080 spheroids were grown per treatment condition on 96 well Nunclon Sphera U-Bottom plates with a nonadherent surface (Thermo Scientific, 174925) for 4 days. Then, three spheroids per condition were treated with either 2 μM paclitaxel or DMSO as a control for 18 hr and imaged on poly-HEMA coated plates as described above. Immediately after imaging, all 96

spheroids were pooled for each condition and spun down by centrifugation (400g for 5 min), followed by dispersion of the pellet into a single-cell suspension in 100 µl 0.25% Trypsin solution. After 1 min of trypsin treatment, cells were resuspended by pipetting, 1ml of complete culture media was added and 500µl of the suspension were transferred into FACS tubes. Living cells were stained with DyeCycle Violet dye (Molecular Probes, V35003) at a concentration of 10 µM for 20 min at 37°C and analzyed on a BD LSRII flow cytometer. The dye was excited with a 405 nm laser and emitted fluorescence was detected with a 450/50 bandpass filter. For data analysis, FlowJo v.X.0.7 was used. First, living single-cells were gated through a FSC/SSC live cell gate combined with a single-cell gate on the Alexa Fluor 405-A vs. Alexa Fluor 405-W plot. This population was then analyzed using the built in cell cycle quantification platform using the univariate model without any adjustments.

Nuclei segmentation

Firstly, the images were preprocessed by applying a $3\times3\times3$ median filter to reduce noise. Then, a rough foreground-background segmentation was obtained using a Poisson-based minimum error thresholding method[38], wherein the image histogram is modeled as a mixture of two poisson models, one for the foreground and one for the background, and the threshold is computed by minimizing the relative entropy between the image histogram and the poisson mixture model. Because of the high variability in foreground-background contrast, caused by attenuation of intensity values with depth and by changes in the brightness of cell nuclei depending on their cell cycle state, applying a global threshold to the entire 3D volume did not provide an accurate segmentation. To address this issue, the thresholding method was applied in a locally adaptive fashion within each slice of the 3D volume. The result obtained was then cleansed with a set of refinement steps that included hole-filling, morphological opening to remove thin structures such as vessels and the removal of small regions such as fragments of dead cells whose size is below a sanity threshold. Typically, the choice of a good thresholding algorithm varies based on the nature of the data being analyzed and the aforementioned approach may not yield an optimal performance on other biological systems and/or data acquired using other microscopic setups. Hence, our software provides a set of foreground-background segmentation algorithms, each of which was found to work better in specific scenarios, that the user can quickly try on sample data and pick a method that works well.

Next, seed points were detected in the cell nuclei by analyzing the response of a scale-adaptive multiscale Laplacian-of-Gaussian (LoG) filter. Specifically, the filter was applied at a series of scales within a user-specified scale-range and the local maxima in the 4D scale-space response were defined as seed points. This multiscale approach provides scale-invariance enabling the detection of seed points in nuclei with varying size. While searching for local maxima, the use of a constant scale-range for all voxels led to inaccurate results, especially for clusters of nuclei of different sizes touching each other with weak edge information in between. Al-Kofafi et. al. proposed an elegant solution[17] to this problem in the context of nuclei segmentation in 2D histopathological images. The authors incorporated a finer per-pixel control over the scale-range using the Euclidean distance map of the binary foreground-background mask. Specifically, they constrained the upper-bound of the scale-

range to that of a blob with radius equal to the distance map value at each voxel. Here, we adapt this method to the context of nuclei seed point detection in 3D intravital images.

The detected seed points were then used as markers for the marker-controlled watershed algorithm[18,19] on the negative response of the multi-scale LoG filter to obtain an initial segmentation result. (Background regions were pruned by computing an intersection with the foreground mask obtained using the thresholding method). This result, however, was unsatisfactory, especially in cases where nuclei of different sizes and/or shapes touch each other and when nuclear shape is aspherical. Both scenarios often occur in our data and became increasingly profound drug treatment. This leads to the detection of multiple seed points within a single nucleus resulting in over-segmentation errors or to the detection of a single seed point for a cluster of touching nuclei resulting in under-segmentation errors. The relative proportion of these two types of errors can be manipulated by tuning the scale-range parameter of the multi-scale LoG filter. Decreasing the lower-bound of the scale-range results in relatively more over-segmentation errors and increasing it results in relatively more under-segmentation errors. Since it is hard to design a principled approach to correct under-segmentation errors, we chose to intentionally over-segment the nuclei by pushing down the lower-bound of the scale-range to a value that empirically results in little or no under-segmentation errors and then correct the over-segmentation errors using a region merging approach. There have been a few interesting efforts reported in the literature that adopt a similar but varied strategy to correct over- and under-segmentation errors resulting from the use of the watershed algorithm for 3D cell nuclei segmentation[12–16] and the method proposed here is based on constructive insights drawn from each of them. The most promising of these is a series of three incremental and increasingly effective methods proposed by Lin et.al.[13,15,16] wherein adjacent regions in the result obtained from the watershed algorithm are iteratively merged in a greedy fashion based on confidence scores derived from a generative classification model. The key ingredient of success in their method is the underlying generative model, which constitutes a probabilistic model of a well-segmented cell nucleus based on a set of features quantifying the appearance, boundary edgeness, size, and shape of the region. However, given the high amount of heterogeneity in the size, shape and appearance of the cell nuclei in our data exacerbated further by the action of cancer drugs, it becomes difficult to design an accurate probabilistic model that encompasses all of the nuclei irrespective of whether a parametric or a non-parametric distribution is employed. On the contrary, for the present task of region merging, one has to only decide whether or not to merge a given pair of mutually adjacent regions. This is relatively less complicated than answering the more general question of whether or not a given region is a well-segmented cell nucleus. These insights suggest that the use of a purely discriminative classification approach may be more amicable to adopt than a generative one which is also in line with the widely believed design philosophy articulated succinctly by Vapnik[39] that "one should solve a problem directly and never solve a more general problem as an intermediate step". For a very similar line of reasoning discriminative classifiers in general are known to perform better than their generative counterparts[40]. Based on these inferences, we chose to guide the region merging process using a discriminative classification model with a richer set of image-derived features carefully designed to just solve the classification problem at hand.

Given an initial segmentation of the cell nuclei obtained from the marker-controlled watershed algorithm, a supervised hierarchical learning-based region merging approach was used to detect and correct any over-segmentation errors. Firstly, a non-negative edge-weighted region adjacency graph (RAG) was constructed wherein each vertex represents a distinct region in the watershed segmentation result and the edges connect vertices corresponding to the regions that share a common boundary in the image space. The weight of an edge connecting two vertices represents the confidence in merging the associated regions to correct a potential over-segmentation error computed using a two-class classification model that is trained to detect over-segmentation errors. Specifically, if the classifier senses a potential over-segmentation error and recommends merging of two adjacent regions, then the corresponding edge-weight in the RAG is set equal to the classifier's prediction probability for the merge decision. If the classifier recommends against merging a given pair of adjacent regions, then the corresponding edge is removed from the RAG. After graph construction, adjacent regions were iteratively merged in decreasing order of merge confidence given by the weight of the corresponding edge in the RAG. In each iteration of the region merging process, the edge with maximum weight in the RAG was picked, the two associated regions were merged and the RAG was updated by recomputing the weights of the edges to all the neighboring regions using the classification model thereby allowing the regions to be merged in a hierarchical fashion. To efficiently carry out the operations of retrieving the edge with maximum weight in each iteration and subsequently updating the RAG, all the edges of the RAG were added to a heap-based max-priority queue data structure with edge-weights as priorities. Additionally, before adding an edge to the priority queue, a check was performed to make sure it has a dominant weight among its neighboring edges to instill an extra amount of confidence into the merge-decision[41]. The key component of success in the proposed region merging method is the underlying discriminative two-class classification model that decides whether or not to merge two adjacent regions using a carefully designed set of image-derived features. The classification model was built in a supervised fashion from a semi-automatically generated training set containing examples of both merge and don't-merge classes: (i) the merge class contains pairs of adjacent regions that belong to a single-cell nucleus that was over-segmented by the watershed algorithm, and (ii) the don't-merge class contains pairs of adjacent regions that belong to two distinct cell nuclei. To efficiently obtain the necessary training data, deliberately over-segmented results were generated using the watershed algorithm and presented to a human expert in the form of a user-friendly software tool. The tool allowed manual classification of each segmented region/fragment as over-segmented, under-segmented, or well-segmented. Isolated pairs of adjacent over-segmented regions with no other regions beside them were then extracted and used as examples of the merge-class. Pairs of adjacent regions in which either both or one was correctly segmented were extracted and used as examples of the don't-merge class. The generated training dataset was then used to train a two-class random forest classifier[42] to decide whether a given pair of adjacent regions should be merged or not based on a set of 20 image-derived features categorized in three groups, namely: (i) appearance similarity group containing features quantifying the similarity in appearance of the two adjacent regions in terms of overall brightness, intensity distribution, and texture, (ii) boundary saliency group containing features quantifying the amount of edge information at the common boundary shared by the two regions which we

refer to as the watershed region, (iii) boundary continuity group containing features quantifying the continuity in the bounding surfaces of the two regions in terms of curvature at the neck, the amount of shared boundary, the gap between the bounding surfaces of the two regions, and the change in convexity introduced by a potential region merging decision (Supplementary Table 1, Fig. 2b). Additionally, to prevent the heavy imbalance between the proportions of examples in the merge and don't-merge classes in our training dataset from biasing the classifier towards the majority class, an ensemble of five random forest classifiers was created wherein each component of the ensemble is trained on a randomly under-sampled balanced subset of the training data. In the test phase, the final classification decision was made by fusing the decisions of all the components in the ensemble using the average of probabilities rule. Lastly, after the region merging process, segmented regions touching the image border in the X and Y dimensions and whose size and volume are below a sanity threshold were pruned. Note, that only the nuclear marker channel was used to segment cell nuclei for two primary reasons: (i) all the nuclei are visible in the nuclear marker channel at all time points, unlike the FUCCI cell cycle marker, and (ii) future studies may rely on other biosensors to study different aspects of chemotherapeutic drug action.

Automatic cell cycle state identification

Firstly, the two FUCCI channels were aligned to the histone channel using a masked image registration algorithm. This was necessary because the FUCCI and histone channels were acquired in single- and two-photon fluorescence imaging modalities, respectively. A purely translational transform was sufficient in our case, however, the following procedures would also be applicable to more complex misalignments such as rotation. Since only a subset of cell nuclei present is visible in the FUCCI channels, it would be inaccurate to use all the voxels in the volume to compute the similarity metric in the registration algorithm. To address this problem, a rough estimate of the foreground in each of the two FUCCI channels was obtained using the Otsu thresholding algorithm[43] and then only the voxels inside the estimated foreground mask were used to compute the similarity metric in the registration algorithm. The registration was done in a multi-scale fashion using the Elastix registration toolkit[44] with an adaptive stochastic gradient descent optimizer and the mutual information metric.

After aligning the channels, each detected cell was represented by 73 quantitative features computed from the histone and/or FUCCI channels (Supplementary Table 4) encoding two different but complementary characteristics: (i) histone feature group containing features characterizing the appearance of the cell nucleus in the histone channel using a set of intensity statistics, Haralick texture features[45] computed at multiple scales, and a few geometric features quantifying the shape of the cell nucleus, and (ii) the FUCCI feature group containing features characterizing the appearance of the cell nucleus in the FUCCI channels using a set of intensity statistics computed from each of the two FUCCI channels and a set of measures quantifying the relative dominance and colocalization between channels using the Pearson correlation coefficient and the ratios of a selected set of intensity statistics between channels. The features in the FUCCI group are effective in distinguishing cells in G1, Late-G1/Early S, and S/G2/mitosis, and the features in the histone group are effective in distinguishing a cell in mitosis from others (Fig. 3c). A random forest

classifier[42] was trained with a database of manually labeled cell nuclei (Fig. 3b) to automatically identify the cell cycle state of each cell given its representation in the high-dimensional feature space. To prevent the heavy class imbalance in the training dataset from biasing the classification, an ensemble of five random forest classifiers was created wherein each component of the ensemble is trained on a randomly under-sampled balanced subset of the training data. In the test phase, the final classification decision was made by fusing the decisions of all the components in the ensemble using the average of probabilities rule. The aforementioned classification scheme was then applied to build a four-class model that associates a given cell to: (i) G1, (ii) Late-G1/Early S, (iii) S/G2, or (iv) mitosis. Additionally, to facilitate studies that want to monitor the cell cycle progression at a coarser level, the same classification scheme was used to build two dedicated and more accurate two-class models: (i) Interphase vs mitotic model to distinguish cells in Interphase (G1 + Late-G1/Early-S + S/G2) against mitotic cells, and (ii) G1 vs S/G2/mitotic model to distinguish cells in G1 phase from the cells in S, G2, and mitotic phases of the cell cycle.

Software

The proposed computation framework was mostly implemented in MATLAB (Mathworks) taking advantage of its parallel computing toolbox to reduce computational time when multiple processing cores are available. We used a machine learning library called WEKA[46] to explore, design, conceptualize, train, and validate all the machine learning models used in our framework and we interface our software with Imaris (Bitplane) for 3D visualizations. Manual annotation of all training datasets was performed blinded. The plots in the paper were generated using MATLAB, Python and Mathematica 9 (Wolfram Research) and plotly (www.plot.ly). Total computational time, including both nuclei segmentation and cell cycle state identification, for a typical volume of size $512 \times 512 \times 40$ voxels with 200–600 cells is 5–10 minutes on a desktop workstation with 8 processing cores (2.93 GHz per core) with 12 GB of RAM and 10–20 minutes when run on a single processing core.

The software and the underlying source code will be made publicly available for use by the scientific community as a supplementary file and under http://lccb.hms.harvard.edu/doc/InvivoCytometer_v2.0.zip. It includes a user-friendly interface for biologists and is implemented in a modular way so the nuclear segmentation algorithm can be used to address biological questions requiring other reporter systems.

Statistical analysis

In Fig. 5, we used Dunnett's t-test to compare the mean values (normalized cell density, percentage of cells in each cell cycle state) computed accross different positions within the tumor on each day after drug treatment with the pre-drug time point, and a one-sided regression slope t-test to check if there is a significant downward trend in the cell density over time. In Fig. 6, we used the one-sided t-test to see if there is a significant increase in multinucleation 3-days after drug treatment when compared to the pre-drug time point. For both figures, since the variation seen at each timepoint is across different locations in the same tumor the test statistic was assumed to be normally distributed. All statistical tests were performed using JMP Pro 11.0.0 (SAS).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Baker BM, Chen CS. Deconstructing the third dimension: how 3D culture microenvironments alter cellular cues. J Cell Sci. 2012; 125:3015–3024. [PubMed: 22797912]

2. Condeelis J, Weissleder R. In vivo imaging in cancer. Cold Spring Harb Perspect Biol. 2010:2.10.1101/cshperspect.a003848

3. Orth JD, et al. Analysis of Mitosis and Antimitotic Drug Responses in Tumors by In Vivo Microscopy and Single-Cell Pharmacodynamics. Cancer Research. 2011; 71:4608–4616. [PubMed: 21712408]

4. Pittet M, Weissleder R. Intravital Imaging. Cell. 2011; 147:983–991. [PubMed: 22118457]

5. Thurber GM, et al. Single-cell and subcellular pharmacokinetic imaging allows insight into drug action in vivo. Nat Commun. 2013; 410.1038/ncomms2506

6. Sakaue-Sawano A, et al. Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. Cell. 2008; 132:487–498. [PubMed: 18267078]

7. Carpenter AE, et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol. 2006; 710.1186/gb-2006-7-10-r100

8. Held M, et al. CellCognition: time-resolved phenotype annotation in high-throughput live cell imaging. Nat Meth. 2010; 7:747–754.

9. Eliceiri KW, et al. Biological imaging software tools. Nat Meth. 2012; 9:697–710.

10. Rasheed S, Nelson-Rees WA, Toth EM, Arnstein P, Gardner MB. Characterization of a newly derived human sarcoma cell line (HT-1080). Cancer. 1974; 33:1027–1033. [PubMed: 4132053]

11. Sakaue-Sawano A, Kobayashi T, Ohtawa K, Miyawaki A. Drug-induced cell cycle modulation leading to cell-cycle arrest, nuclear mis-segregation, or endoreplication. BMC Cell Biol. 2011; 1210.1186/1471-2121-12-2

12. Umesh Adiga PS, Chaudhuri BB. An efficient method based on watershed and rule-based merging for segmentation of 3-D histo-pathological images. Pattern Recognition. 2001; 34:1449–1458.

13. Lin G, et al. A hybrid 3D watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks. Cytometry Part A. 2003; 56A:23–36.

14. Wählby C, Sintorn IM, Erlandsson F, Borgefors G, Bengtsson E. Combining intensity, edge and shape information for 2D and 3D segmentation of cell nuclei in tissue sections. Journal of Microscopy. 2004; 215:67–76. [PubMed: 15230877]

15. Lin G, et al. Hierarchical, model-based merging of multiple fragments for improved three-dimensional segmentation of nuclei. Cytometry Part A. 2005; 63A:20–33.

16. Lin G, et al. A multi-model approach to simultaneous segmentation and classification of heterogeneous populations of cell nuclei in 3D confocal microscope images. Cytometry Part A. 2007; 71A:724–736.

17. Al-Kofahi Y, Lassoued W, Lee W, Roysam B. Improved Automatic Detection and Segmentation of Cell Nuclei in Histopathology Images. IEEE Transactions on Biomedical Engineering. 2010; 57:841–852. [PubMed: 19884070]

18. Vincent L. Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. IEEE Transactions on Image Processing. 1993; 2:176–201. [PubMed: 18296207]

19. Meyer F. Topographic distance and watershed lines. Signal Processing. 1994; 38:113–125.

20. Amat F, et al. Fast, accurate reconstruction of cell lineages from large-scale fluorescence microscopy data. Nat Methods. 2014; 11:951–958. [PubMed: 25042785]

21. Mitchison TJ. The proliferation rate paradox in antimitotic chemotherapy. Mol Biol Cell. 2012; 23:1–6. [PubMed: 22210845]

22. Komlodi-Pasztor E, Sackett D, Wilkerson J, Fojo T. Mitosis is not a key target of microtubule agents in patient tumors. Nat Rev Clin Oncol. 2011; 8:244–250. [PubMed: 21283127]

23. Spencer SL, Sorger PK. Measuring and modeling apoptosis in single cells. Cell. 2011; 144:926–939. [PubMed: 21414484]

24. Jordan MA, Wilson L. Microtubules as a target for anticancer drugs. Nat Rev Cancer. 2004; 4:253–265. [PubMed: 15057285]

25. Janssen A, Beerling E, Medema R, van Rheenen J. Intravital FRET imaging of tumor cell viability and mitosis during chemotherapy. PLoS One. 2013:8.10.1371/journal.pone.0064029

26. Milas L, et al. Kinetics of mitotic arrest and apoptosis in murine mammary and ovarian tumors treated with taxol. Cancer Chemother Pharmacol. 1995; 35:297–303. [PubMed: 7828272]

27. Zasadil LM, et al. Cytotoxicity of paclitaxel in breast cancer is due to chromosome missegregation on multipolar spindles. Sci Transl Med. 2014:6.10.1126/scitranslmed.3007965

28. Gascoigne KE, Taylor SS. Cancer cells display profound intra- and interline variation following prolonged exposure to antimitotic drugs. Cancer Cell. 2008; 14:111–122. [PubMed: 18656424]

29. Shi J, Orth JD, Mitchison T. Cell type variation in responses to antimitotic drugs that target microtubules and kinesin-5. Cancer Res. 2008; 68:3269–3276. [PubMed: 18451153]

30. Purcell JW, et al. Activity of the kinesin spindle protein inhibitor ispinesib (SB-715992) in models of breast cancer. Clin Cancer Res. 2010; 16:566–576. [PubMed: 20068098]

31. Towle MJ, et al. Broad spectrum preclinical antitumor activity of eribulin (Halaven(R)): optimal effectiveness under intermittent dosing conditions. Anticancer Res. 2012; 32:1611–1619. [PubMed: 22593439]

32. Manfredi MG, et al. Characterization of Alisertib (MLN8237), an investigational small-molecule inhibitor of aurora A kinase using novel in vivo pharmacodynamic assays. Clin Cancer Res. 2011; 17:7614–7624. [PubMed: 22016509]

33. Towle MJ, et al. Eribulin Induces Irreversible Mitotic Blockade: Implications of Cell-Based Pharmacodynamics for In vivo Efficacy under Intermittent Dosing Conditions. Cancer Research. 2011; 71:496–505. [PubMed: 21127197]

34. Mori T, et al. Retention of paclitaxel in cancer cells for 1 week in vivo and in vitro. Cancer Chemother Pharmacol. 2006; 58:665–672. [PubMed: 16534615]

35. Vanhoefer U, Cao S, Harstrick A, Seeber S, Rustum YM. Comparative antitumor efficacy of docetaxel and paclitaxel in nude mice bearing human tumor xenografts that overexpress the multidrug resistance protein (MRP). Ann Oncol. 1997; 8:1221–1228. [PubMed: 9496387]

36. Friedrich J, Seidel C, Ebner R, Kunz-Schughart LA. Spheroid-based drug screen: considerations and practical approach. Nat Protoc. 2009; 4:309–324. [PubMed: 19214182]

37. Tosoni D, Di Fiore PP, Pece S. Functional purification of human and mouse mammary stem cells. Methods Mol Biol. 2012; 916:59–79. [PubMed: 22914933]

38. Fan J. Notes on Poisson distribution-based minimum error thresholding. Pattern Recognition Letters. 1998; 19:425–431.

39. Vapnik, VN. Statistical learning theory. Wiley; 1998.

40. Jordan A. On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. Advances in neural information processing systems. 2002; 14:841–848.

41. Ning J, Zhang L, Zhang D, Wu C. Interactive image segmentation by maximal similarity based region merging. Pattern Recognition. 2010; 43:445–456.

42. Breiman L. Random Forests. Machine Learning. 2001; 45:5–32.

43. Otsu N. A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man and Cybernetics. 1979; 9:62–66.

44. Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. Elastix: A Toolbox for Intensity-Based Medical Image Registration. IEEE Transactions on Medical Imaging. 2010; 29:196–205. [PubMed: 19923044]

45. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. IEEE Transactions on Systems, Man and Cybernetics. 1973; SMC-3:610–621.

46. Hall M, et al. The WEKA data mining software: an update. SIGKDD Explorations. 2009; 11:10–18.
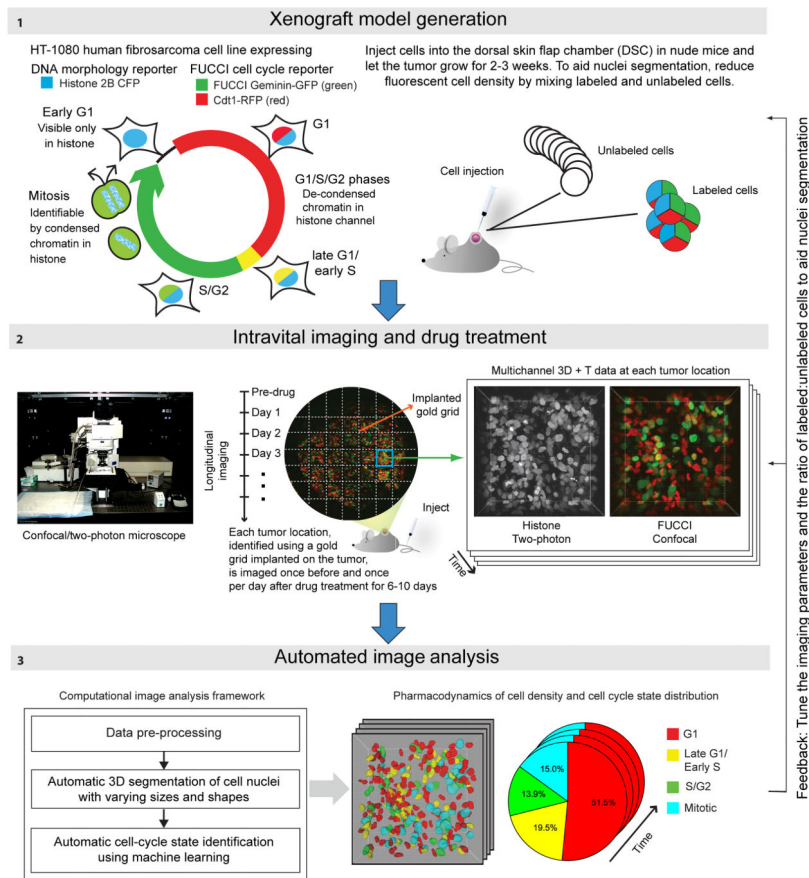
Figure 1. Overview of experimental setup and image analysis Panel 1
Generation of xenograft tumors. HT-1080 cells engineered to stably express the FUCCI cell cycle reporter system and Histone H2B CFP allow *in vivo* detection of G1, Late-G1/Early-S, S/G2 and mitotic cells. In Late-G1/Early-S phase expression of the red and green FUCCI reporters overlaps, resulting in a yellow/orange signal. Two million cells per experiment were subcutaneously injected into DSCs implanted on the back of nude mice. To enhance segmentation accuracy fluorescent cells were diluted with unlabeled cells from the parental cell line. **Panel 2:** Gold grids placed on the tumor were used as a spatial reference system. 3D stacks were acquired at multiple positions before and at varying intervals after drug injection. The histone channel (two-photon microscopy) and FUCCI channels (confocal mode) were acquired in two consecutive runs. **Panel 3:** The computational image analysis framework automatically detects and segments nuclei in the histone channel in 3D and to identifies their cell cycle states based on information in both the histone and FUCCI channels.
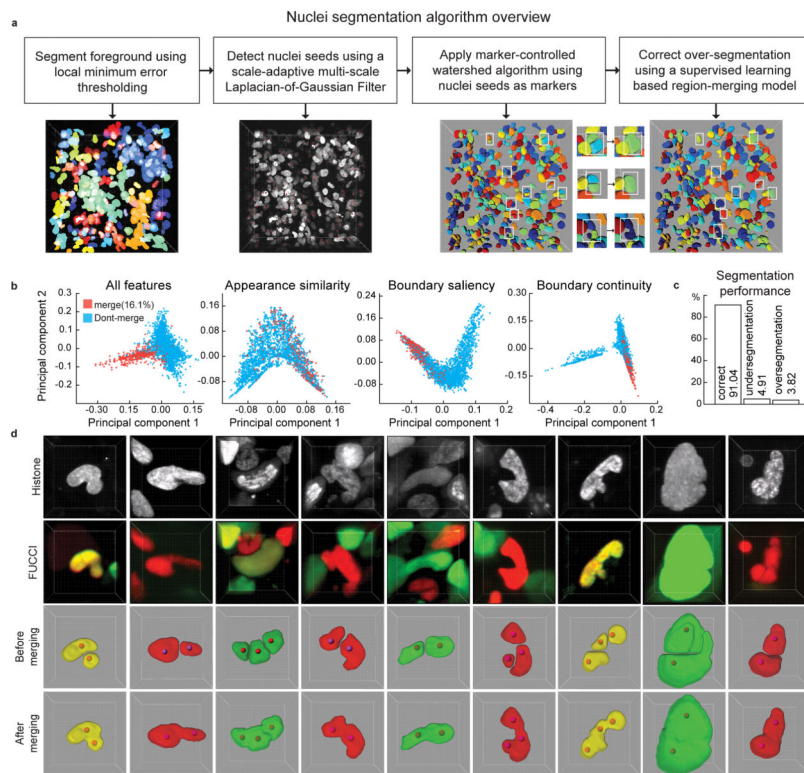
Figure 2. Automatic segmentation of cell nuclei
(a) Steps of segmentation algorithm. Intermediate results of each step are visualized in 3D. Parameters of the seed point detection filter are tuned to lean towards over-segmentation of the nuclei which are then detected and corrected in the merging step (white boxes). **(b)** 2D projections of the high-dimensional feature-space representations of the cell nuclei using all features, and the features only in appearance similarity group, boundary saliency group, and the boundary continuity group, respectively. The 2D projections were obtained using a non-linear dimensionality reduction technique called Laplacian Eigenmap. **(c)** Performance of the segmentation algorithm on a dataset of 54 volumes (Supplementary Table 2–3) out of which 26 volumes were used to train the classification model that is used to guide the region merging process and the remaining 28 were used to evaluate the performance of the segmentation algorithm. The segmentation quality was scored manually with a software tool that presents each cell one by one to a human observer who then specified whether it was well-segmented, over-segmented or under-segmented. We report the average of per-stack percentages of cells that were well-segmented, under-segmented or over-segmented in the test dataset. (d) A few examples of multi-nucleated cells with non-spherical shapes that were correctly segmented using a two-step strategy of initial over-segmentation followed by region merging. Red spheres: seed-points used in an initial marker-controlled watershed algorithm.
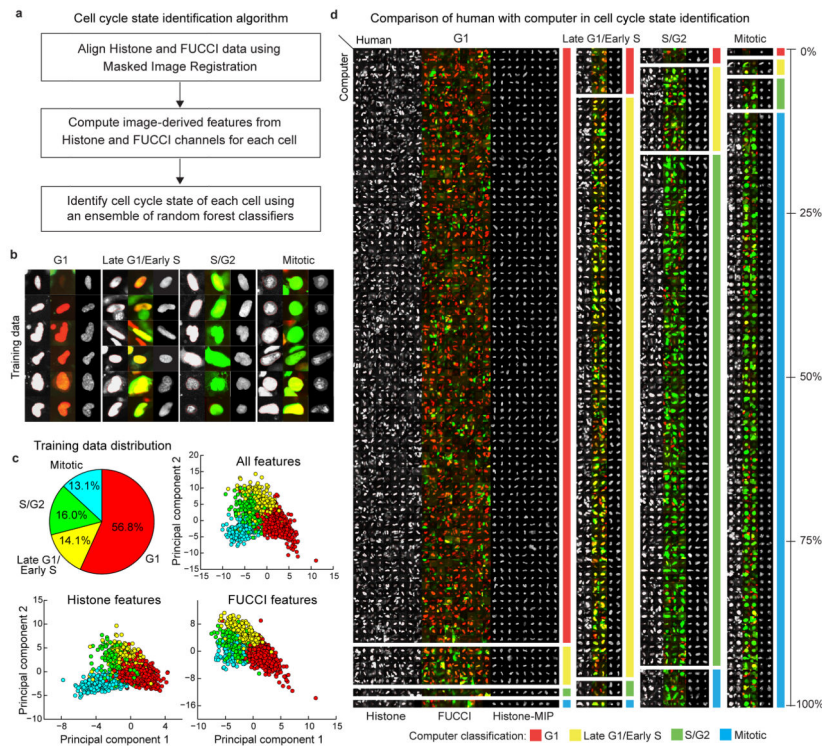
Figure 3. Automatic identification of cell cycle state

(a) Steps of cell cycle state identification algorithm. **(b)** Examples of cells from each cell cycle state in our training dataset. **(c)** Class distribution (top-left) in training dataset and 2D projections (obtained by Linear Discriminant Analysis, LDA) of the feature-space representations (**Online Methods**, Supplementary Table 4) using all features (top-right), and features only in the histone (bottom-left), and FUCCI channels (bottom-right), respectively. **(d)** A pictorial depiction of the confusion matrix (Supplementary Table 3) showing the agreement/disagreement between human observer and computer in identifying the nuclei of each cell cycle state in a dataset of 53 volumes (25 used for training the classification model, the remaining 28 to evaluate the performance). The height of each block is proportional to the corresponding element in the row-normalized confusion matrix. The large blocks on the diagonal represent agreement between human and computer, other blocks represent disagreement. Due to space limitations only 30% of all cells in our dataset are shown. Thumbnails in (b) and (d): mid-axial cross-section of the histone channel (left) and FUCCI red-green-overlay (middle), maximum intensity projection (MIP) of the histone channel (right). In all panels, the cell thumbnails represent 35×35 μm by default. Larger cells were scaled to fit.
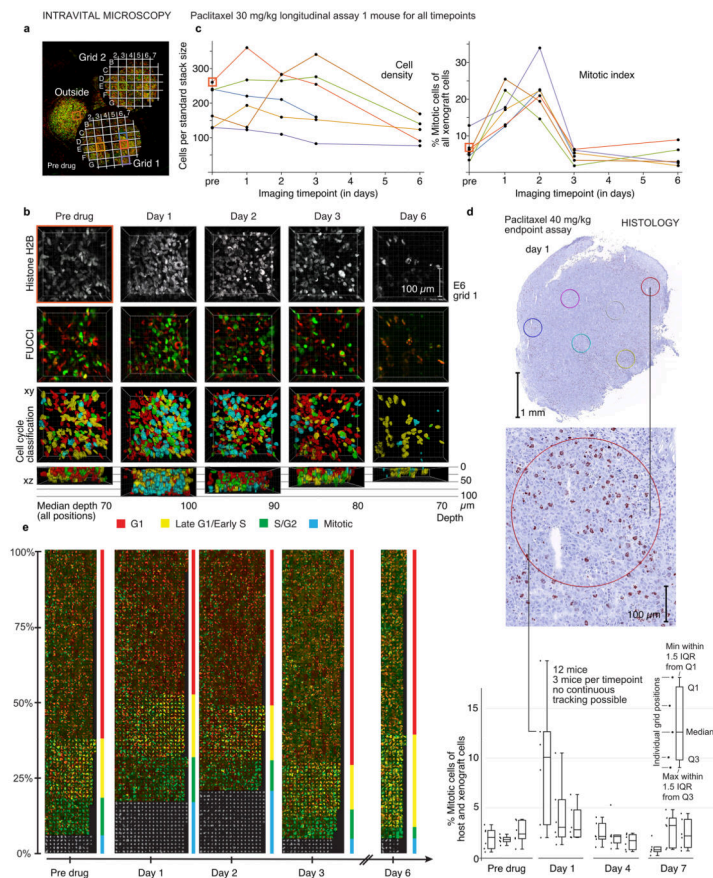
Figure 4. Quantitative analysis of drug response to paclitaxel treatment at a single-cell level
Analysis of an HT-1080 tumor treated with a single dose (30 mg/kg) of the MT-interacting drug paclitaxel and validation by histology. **(a)** Two gold grids allow tracking of tumor positions over time, before drug injection. **(b)** 3D visualizations of the histone and FUCCI channels, along with the cell cycle classification obtained by automated analysis at different time points for position E6 on grid 1 (orange square in (a, b, c)). **(c)** Left panel: cell density at each position over time (as cells per volume of typical size in our acquisition setup, 256×256×80 μM). Right panel: % mitotic tumor cells over time at each individual position. Color coding as in (a). **(d)** Mitotic index quantified by phospho-histone H3 immunoperoxidase staining in HT-1080 xenograft tumors before and after treatment with paclitaxel 40 mg/kg in a single dose. Tumors from 12 mice (3 untreated, three treated for 1, 4, and seven days each were quantified. Example images from a mouse sacrificed 1 day after treatment. IQR is inter-quartile-range. For more details see Supplementary Fig. 4. (e) Montage of all cells from (a)–(c), grouped by time point (all grid positions for each timepoint were pooled) and cell cycle state. In each block, cells are ordered by prediction probability in vertical columns from left to right, i.e. within a block cells on the left have a higher probability to be in the respective cell cycle state. The height of each cell cycle block represents the percentage of cells in that state, column width is proportional to the number of imaged cells(determined not only by density but also by the depth of the imaged stacks, which varies from day to day (see (b)). In mitosis, MIPs of the histone channel are shown, other stages show mid-axial cross-sections. 5931 cells from 29 volumes corresponding to 6

different positions imaged over a timespan of 7 days are shown. For full-resolution version, see Supplementary Fig. 6. Thumbnails represent 35×35 μm, larger nuclei were scaled to fit.
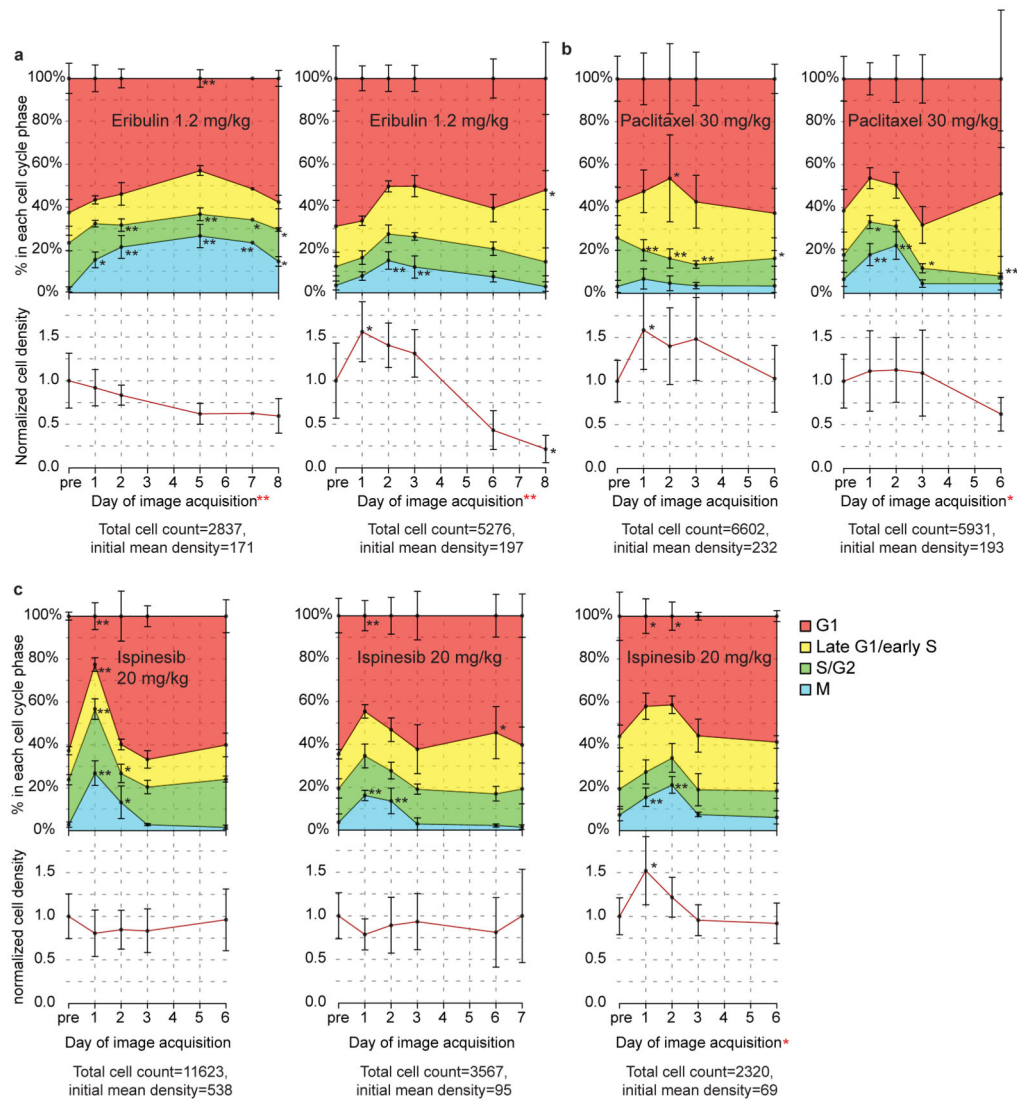
Figure 5. Pharmacodynamics of antimitotic drugs in the HT-1080 xenograft system
Effects of three different antimitotic cancer drugs compared as in Fig. 4. Each plot shows data from one mouse. Upper half: cell cycle state distribution. Lower half: mean cell density over time. **(a)** A single dose of the MT-interacting drug eribulin (1.2 mg/kg) was injected intravenously into two mice. **(b)** A single dose of the MT-interacting drug paclitaxel (30 mg/kg) was injected intravenously into two mice. **(c)** A single dose of the non MT-interacting KSP inhibitor ispinesib (20 mg/kg) was injected intraperitoneally into three mice. Error bars in (a)–(c) show one standard deviation and refer to the variation between different grid positions at the respective timepoint. Density values normalized to the mean density of all positions before drug injection in the respective mouse, the value of which is given below each plot in cells per typical volume of 80 μm depth (5.2 mm³). Late time points with less than 100 imaged cells in total are not shown. The second panel in b represents the same mouse as shown in Figure 4a–c,e. For corresponding images of cell thumbnails for each mouse, see Supplementary Figs. 6–8. Black *, Dunnet's test was used to compare the mean values on each day after drug treatment with the pre-drug (control) time point. Red *, one-

sided regression slope t-test was used to check if there is a significant downward trend in cell density over time. *$P<0.05$,**$P<0.005$
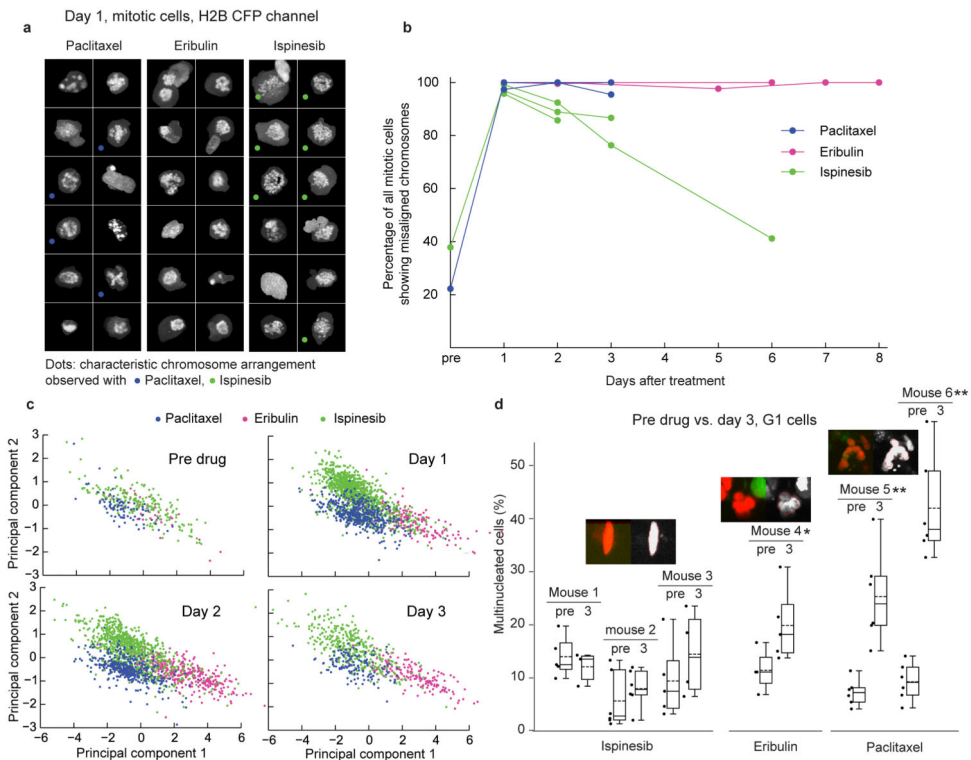
Figure 6. Antimitotic drugs induce different mitotic and interphase phenotypes
Examples of cell morphology induced by antimitotic drugs (same datasets as Fig. 5) **(a)**
Random sample of cells imaged on day 1 after drug injection and identified by the analysis
framework as mitotic. MIPs of histone channel, representing condensed chromosomes. Note
the ability to automatically detect cells with different characteristic morphology for each
drug. **(b)** Percentage of spindles with misaligned chromosomes over time. Each line
represents one mouse. All grid positions of a given mouse at an imaged timepoint were
pooled. Datapoints with fewer than 9 mitotic cells in total (due to very low mitotic index)
are not shown. **(c)** 2D projections (obtained by LDA) of a high-dimensional space of texture
features of mitotic cells at four consecutive time points. The mitotic phenotypes induced by
the three drugs form distinct clusters, confirming the morphologic differences shown in Fig.
6a. **(d)** Representative examples of phenotypes of G1 cells in drug-treated tumors from Fig.
5. Note that ispinesib-treated and untreated cells are mononucleated in morphology while
eribulin induces a mild and paclitaxel a severe multinucleated phenotype. In (a) and (d),
single-cell thumbnails are by default 35×35 μm in size. Larger cells were scaled to fit. A
one-sided t-test was used to compare the mean percentage of multinucleated cells on day 3
after drug treatment with the pre-drug time point. *$P<0.05$, **$P<0.005$. Box plots as in Fig.
4d, dashed lines show the mean.