Invivo Cytometer Documentation

Release 2.0.0

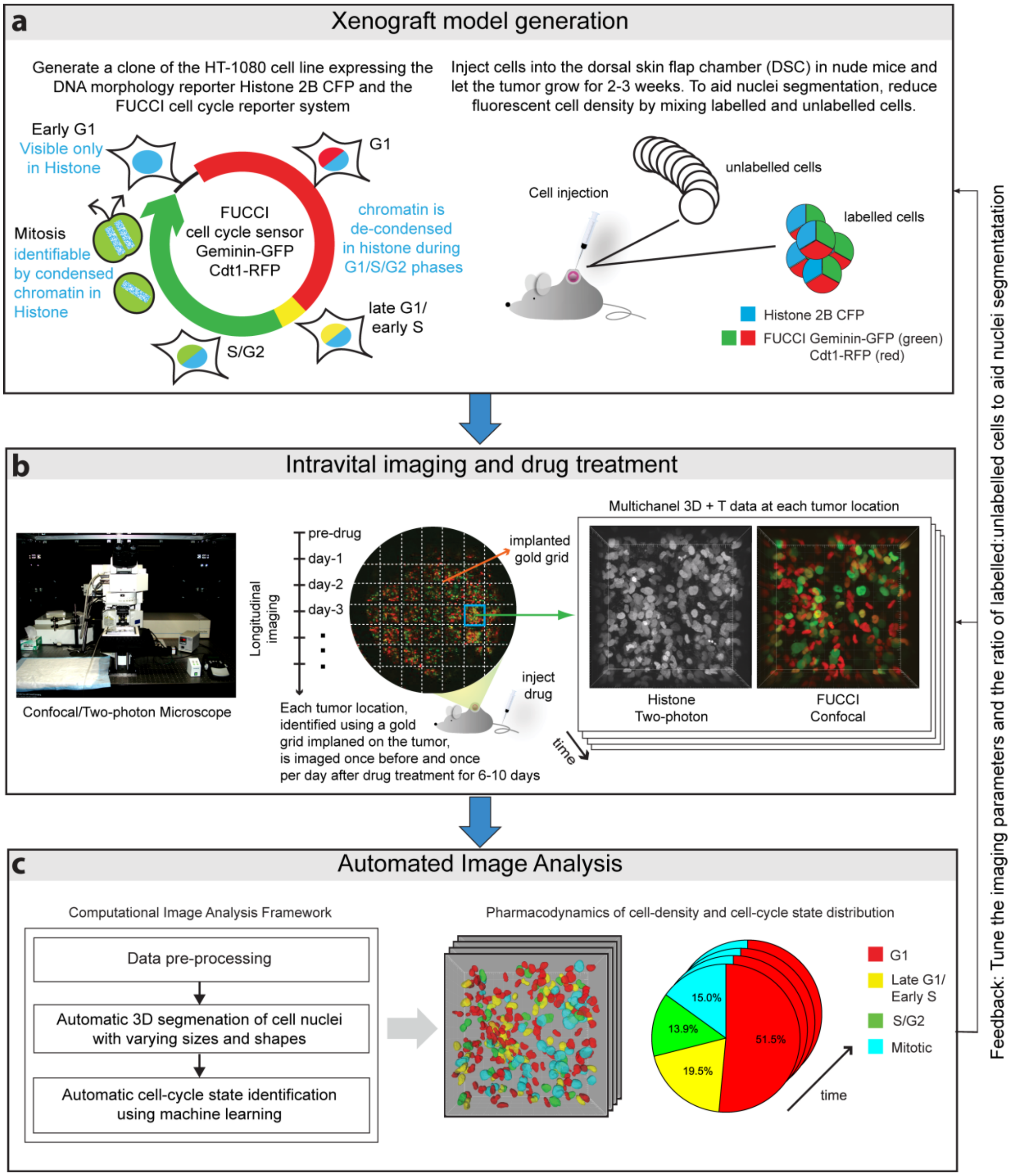
September, 2014

Contents

Overview

InvivoCytometer is a software tool for automated cell cycle profiling from 3D confocal/multi-photon images of HT-1080 cells expressing a fluorescently labeled nuclear marker protein (Histone-2B) and the FUCCI cell cycle reporter system. Given a 3D volume containing the histone and FUCCI channels, the software uses a combination of image analysis and machine learning methods to segment individual cell nuclei and identify their cell cycle state thereby allowing us to study the cell cycle effects of experimental perturbations over time.

Figure 1: Overview of the experimental setup and image analysis



a
Xenograft model generation
Generate a clone of the HT-1080 cell line expressing the DNA morphology reporter Histone 2B CFP and the FUCCI cell cycle reporter system
Inject cells into the dorsal skin flap chamber (DSC) in nude mice and let the tumor grow for 2-3 weeks. To aid nuclei segmentation, reduce fluorescent cell density by mixing labelled and unlabelled cells.
Early G1
Visible only in Histone
G1
FUCCI cell cycle sensor Geminin-GFP Cdt1-RFP
chromatin is de-condensed in histone during G1/S/G2 phases
Mitosis identifiable by condensed chromatin in Histone
late G1/ early S
S/G2
Cell injection
unlabelled cells
labelled cells
🟦 Histone 2B CFP
🟩🟥 FUCCI Geminin-GFP (green) Cdt1-RFP (red)
b
Intravital imaging and drug treatment
Confocal/Two-photon Microscope
Longitudinal imaging
pre-drug
day-1
day-2
day-3
implanted gold grid
inject drug
Each tumor location, identified using a gold grid implaned on the tumor, is imaged once before and once per day after drug treatment for 6-10 days
Multichanel 3D + T data at each tumor location
Histone Two-photon
FUCCI Confocal
time
c
Automated Image Analysis
Computational Image Analysis Framework
Data pre-processing
Automatic 3D segmenation of cell nuclei with varying sizes and shapes
Automatic cell-cycle state identification using machine learning
Pharmacodynamics of cell-density and cell-cycle state distribution
15.0%
13.9%
19.5%
51.5%
time
G1
Late G1/ Early S
S/G2
Mitotic
Feedback: Tune the imaging parameters and the ratio of labelled:unlabelled cells to aid nuclei segmentatation

While the software can be used for a variety of applications in developmental biology and cancer research that involve an in-depth study of the effects of experimental perturbations on cell cycle progression, it was originally developed for quantifying the in vivo cell cycle effects of chemotherapeutic drugs on HT1080 fibrosarcoma xenograft tumors in living mice imaged intravitally using a confocal/two-photon microscope. Below is a brief overview of typical workflow (See Fig 1) that was followed for this application:

1. **Xenograft model generation**
   a. Generate a clone of the HT-1080 fibrosarcoma cell line stably expressing the DNA morphology reporter Histone 2B CFP and the FUCCI cell cycle reporter system to allow the in vivo detection of G1, Late G1/Early S, S/G2, and mitotic cells.

      Late G1/Early S phase is characterized by the overlapping expression of the red (mKO2 - hCdt1) and green (mAG - hGem) FUCCI reporters resulting in a yellow/orange signal. Mitotic cells can be identified by their usually round/spherical shape and their distinct chromatin texture due to chromosome condensation in the H2B-CFP.

   b. Inject cells into the dorsal skin fold chamber (DSC) implanted on the back of a nude mouse and let the tumor vascularize and grow for 2-3 weeks.

      For each experiment, about 2 millions cells were injected. To enhance segmentation accuracy, fluorescent cell density was reduced by mixing fluorescent cells 1:20 or 1:35 with the unlabelled cells from the parental cell line.

2. **Longitudinal spatiotemporal intravital imaging and drug treatment**
   a. To allow for long term repeated imaging of the same tumor regions, place a gold grid on the tumor one day before drug treatment and use it as a reference system.
   b. Use a confocal/two-photon microscope (Olympus FV1000) to acquire 3D stacks at multiple positions in the gold grid before and at periodic time intervals (once a day) after drug injection/treatment. We recommend acquiring the histone channel, used for nuclei segmentation, in the two-photon mode (for better depth resolution) and the FUCCI channels can be acquired in an immediate run in confocal mode.

3. **Image analysis**
   a. *Training phase:* Select a subset of the imaged data to train the machine learning components of the algorithms used for nuclei segmentation and cell cycle state identification. Use the annotation tools provided to generate the training data and build the machine learning models from the annotated training data.
   b. *Automated Analysis phase:* Perform cell cycle profiling by loading the imaged data into the analysis tools with a user-friendly graphical user interface (GUI). Alternatively, an advanced user can use the batch processing scripts provided to deploy the computational framework on a compute cluster to analyze all the datasets in parallel.

General Information

License

This program is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with this program. If not, see http://www.gnu.org/licenses/.

System requirements

1. Hardware specs:

   The software was developed primarily on a desktop workstation with 8 processing cores (2.93 Ghz per core) and 12GB of RAM. For reasonable performance, we recommend using a desktop workstation with atleast 4 processing cores and over 6GB of RAM. To operate the software, we also assume that a mouse with a scroll wheel is available.

2. Operating systems:

   The software was developed primarily on a 64-bit Windows 7 operating system. However, since the software was written in MATLAB, the user should be able to use most of the software's functionality on 64-bit versions of other platforms, including both mac and linux, that are compatible with MATLAB.

3. Additional software
   - MATLAB:

     This software was developed primarily in matlab 2013b. It should work on matlab 2013b or higher. In addition, the following MATLAB toolboxes are required: Image Processing toolbox, Statistics toolbox, Parallel Computing toolbox, Communications Systems toolbox, Curve Fitting toolbox, System Identification toolbox

     In a future software release, we will provide a standalone application for the users who do not have MATLAB.

   - Imaris (optimal): 7.4 or higher

     Imaris is required to use the 3D visualization features of the software. For those, who are just interested in the analysis this is not required.

Software

Installation

1. Download the zip archive containing the software.
2. Create a new directory and extract all the files in the zip archive into this directory.
3. Start Matlab
4. Within matlab, navigate to the directory containing software code (look for the sub-folder named code_package) in the software directory you downloaded.

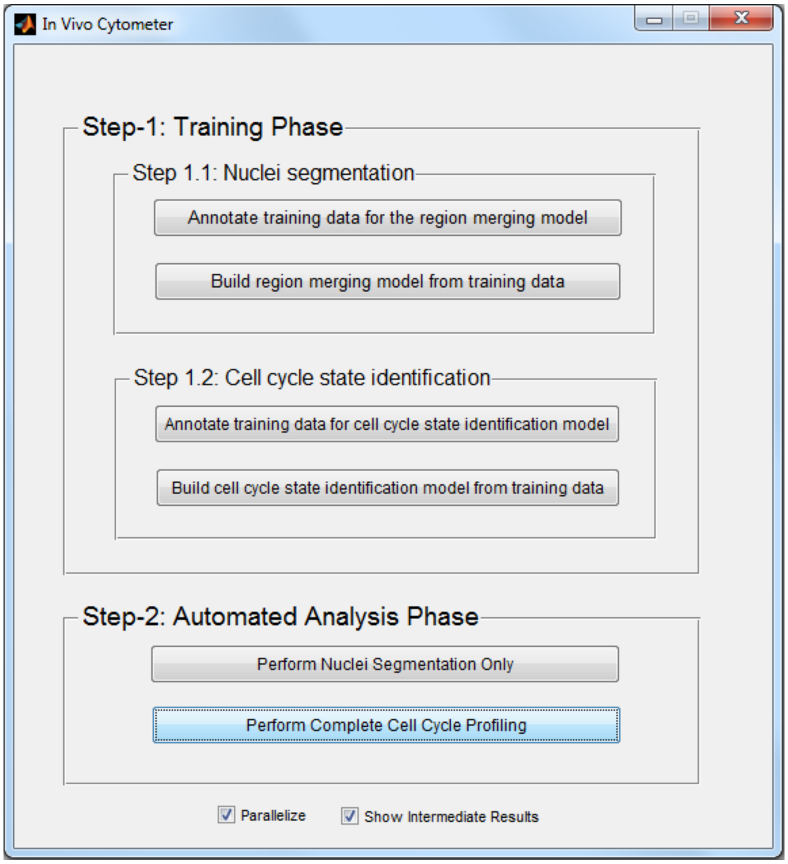
Basic Usage guide

To run the software, open matlab and type the following command at the matlab command prompt

```
>> InvivoCytometer
```

This will bring up the main control panel of the software shown below from which all of its features can be accessed.

Figure 2: Main control panel of the software


In Vivo Cytometer
Step-1: Training Phase
Step 1.1: Nuclei segmentation
Annotate training data for the region merging model
Build region merging model from training data
Step 1.2: Cell cycle state identification
Annotate training data for cell cycle state identification model
Build cell cycle state identification model from training data
Step-2: Automated Analysis Phase
Perform Nuclei Segmentation Only
Perform Complete Cell Cycle Profiling

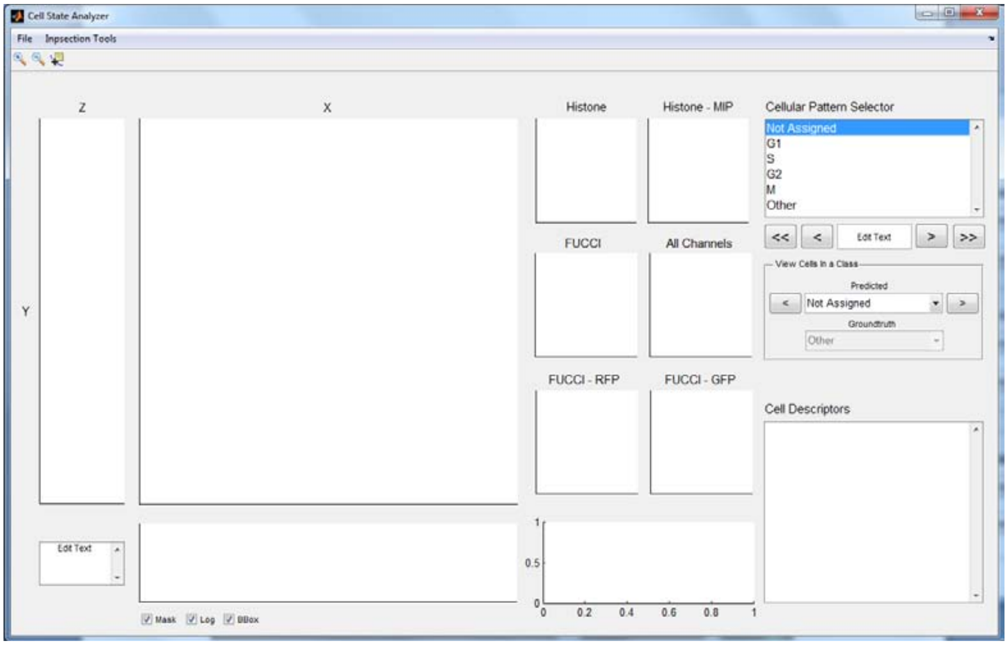
☑ Parallelize ☑ Show Intermediate Results

The checkbox "parallelize" specified whether or not to use parallelization to boost computational performance. The checkbox "show intermediate results" specifies whether or not to show intermediate results of the image analysis algorithms.

Quick sneak peek into the analysis

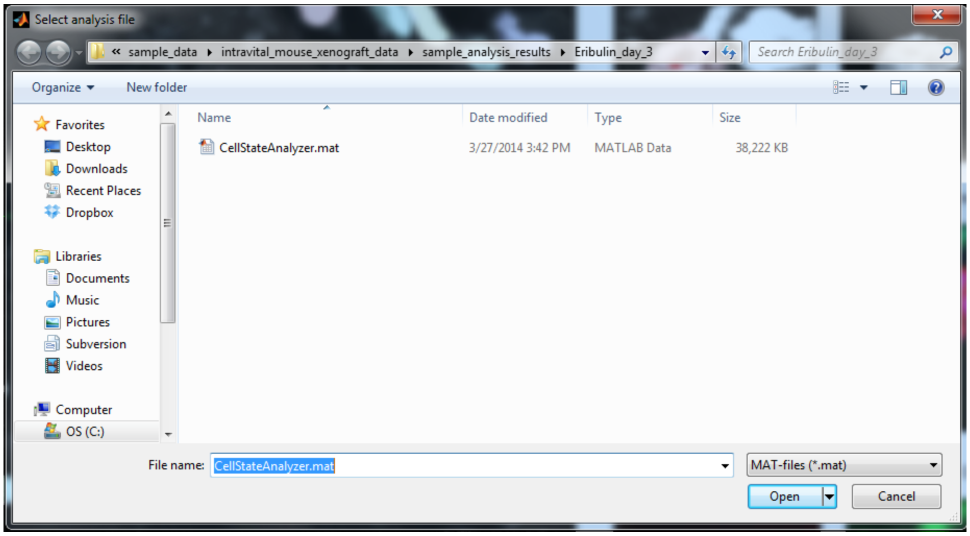
Below is a series of steps to follow to get a sneak peek into the analysis output of the software:

1. Click the button that says "Perform Complete Cell Cycle Profiling" within the Automated Analysis Phase Group of the main control panel of the software (Figure 2). This will bring up the following window named "Cell State Analyzer"



   This is the analysis tool that you will/should be using if you want to perform cell cycle profiling. Now let's look at a sample analysis result.

2. In the menu bar, click File->Load Analysis. Now it will ask you to select an analysis file.


Select analysis file

« sample_data ▸ intravital_mouse_xenograft_data ▸ sample_analysis_results ▸ Eribulin_day_3

Name	Date modified	Type	Size
CellStateAnalyzer.mat	3/27/2014 3:42 PM	MATLAB Data	38,222 KB

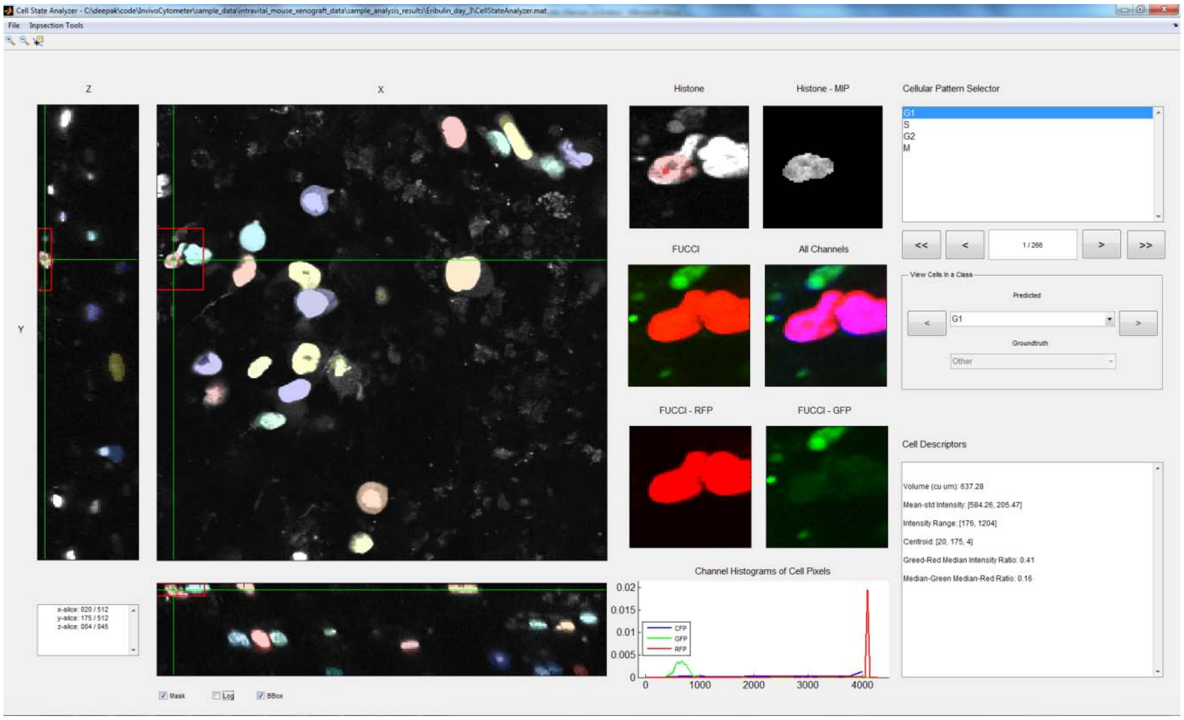
File name: CellStateAnalyzer.mat

MAT-files (*.mat)

To help you understand the workflow, we must have given you some sample data along with the software. This is typically inside a folder called "sample_data" (unless you renamed it) which is either inside the software package you downloaded or provided to you for download separately.

Find the directory containing the sample data and navigate to the sub-folder "*<sample_data_folder>/intravital_mouse_xenograft_data/sample_analysis_results*". There must be a file named "CellStateAnalyzer.mat" in this folder. This file contains the analysis done on some sample image data. Load this file by (i) Double clicking it, or (ii) select it and then press the button which says "open".

In a brief moment, the image data and all the analysis done on it must be loaded into the CellStateAnalyzer tool as shown below:



Cell State Analyzer - C:\deepak\code\InvivoCytometer\sample_data\intravital_mouse_xenograft_data\sample_analysis_results\Eribulin_day_3\CellStateAnalyzer.mat

```
x-slice: 020 / 512
y-slice: 175 / 512
z-slice: 004 / 045
```

1 / 266

Predicted: G1

Groundtruth: Other

```
Volume (cu um): 637.28
Mean-std Intensity: [584.26, 205.47]
Intensity Range: [176, 1204]
Centroid: [20, 175, 4]
Greed-Red Median Intensity Ratio: 0.41
Median-Green Median-Red Ratio: 0.16
```

3. Now, you can browse through the analysis performed.

   Navigating through the cells: The button group with buttons "<<" (show first cell), "<" (show previous cell), ">" (show next cell), ">>" (show last cell) will let you navigate through each cell detected, segmented and classified into one of the cell cycle states by our algorithm. The textbox located in the middle of these buttons show <currently-active-cell-id>/<total-cell-count>. As you navigate through the cells, the cell cycle state of the currently active cell is highlighted in the list named "Cellular Pattern Selector" located in the top right part of the window.

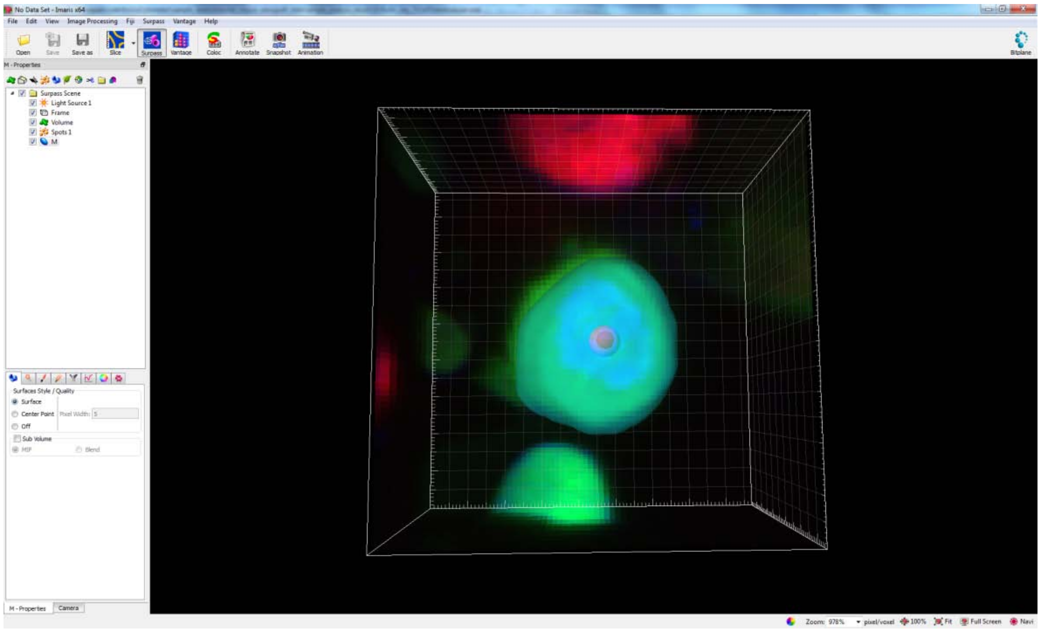
   You can also look at all the cells in each class (or cell cycle state) by selecting the desired class from the drop-down menu named "Predicted" (within the group named "View cells in a class" in the mid-rightmost part of the window) and navigate through them using the buttons "<" (show previous cell in class) and the ">" (show next cell in class).

Image panels: The three largest image panels show the ortho-slices/cross-sections along the X-, Y-, and Z- axis of the 3D volume (use checkbox named "log" to apply a log transformation to see it better) overlaid with the segmentation mask (this can be toggled on/off using the checkbox named "mask"). You can use the mouse scroll wheel to navigate through the slices. To navigate through the X- and Y- slices, you have to position the mouse pointer inside the corresponding image panel and move the mouse scroll wheel. Moving the mouse scroll wheel anywhere else in the window will let you navigate through the Z-slices. The green lines shown where the other ortho-slices are located.

The six small image panels show a zoomed-in version of a cropped XY region (defined by the red bounding box shown in the XY slice) around the currently active cell in: (i) Histone channel - useful for seeing the raw texture inside the cell, (ii) Histone MIP - Maximum intensity projection of the cell pixels in the cropped region of Histone channel along the z-axis useful for specifically for identifying mitotic cells, (iii) FUCCI - overlay of the two FUCCI GFP and RFP channels useful specifically for identifying cells in Late G1/Early S phase that appear yellow/orange in color in this view, (iv) All Channels - Overlay of the histone channel and the two FUCCI channels, (v) FUCCI-RFP channel -- this shows cells in G1, (vi) FUCCI-GFP channel - this shows cells in S/G2/M phases of the cell cycle.

3D visualization using Imaris: InvivoCytometer interacts with a commercial software called IMARIS through its COM interface to provide nice and interesting 3D visualization of the analysis results that can be accessed from the "Inspection Tools" Menu. To be able to use these features, you should run the tool in the windows platform and should have already installed Imaris (version 7.4 or higher) before running the tool.

"Inspection Tools-->View Cell Segmentation In Imaris" shows a 3D visualization of the currently active cell in imaris as shown below

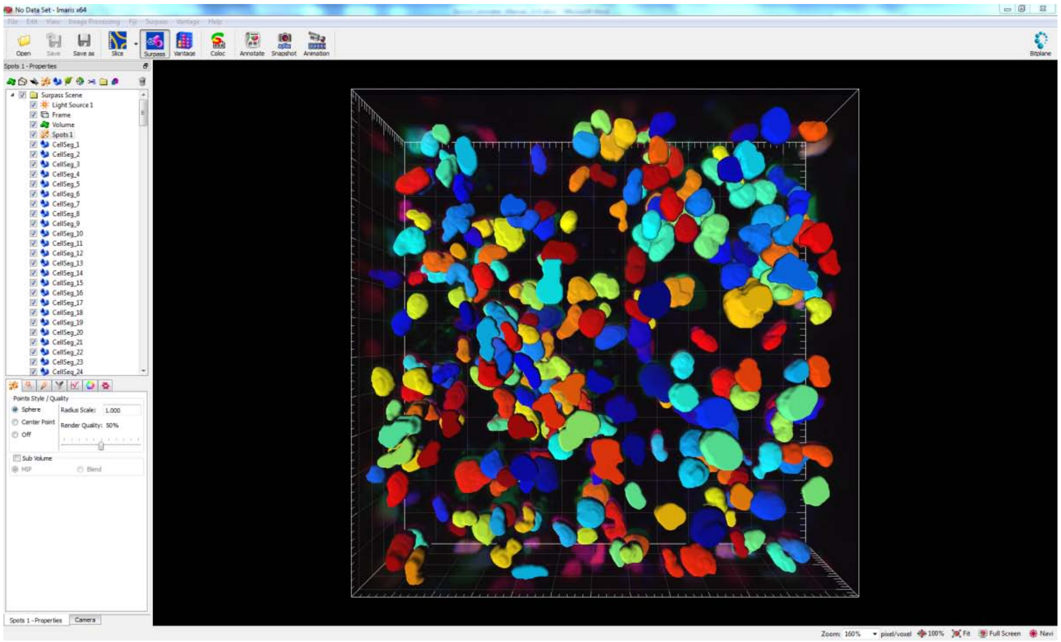


☑ Surpass Scene
☑ Light Source 1
☑ Frame
☑ Volume
☑ Spots 1
☑ M

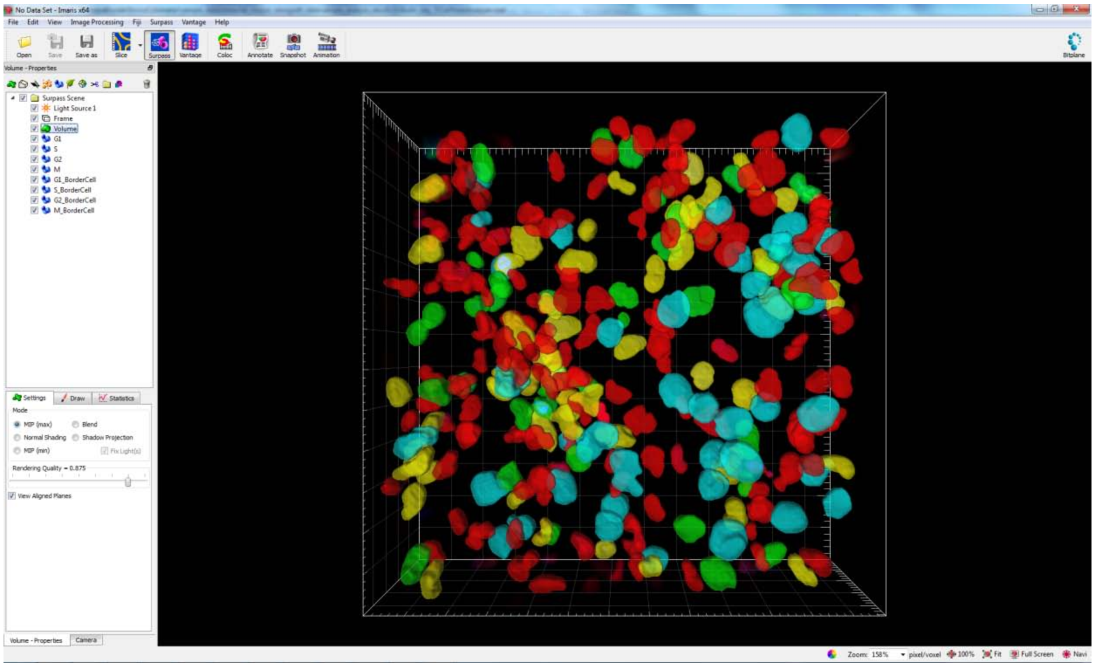
● Surface
○ Center Point
○ Off
Pixel Width: 5
□ Sub Volume

Zoom: 978%

"Inspection Tools-->View Full Segmentation In Imaris" shows a 3D visualization of the cells detected and segmented (colored randomly) by our nuclei segmentation algorithm as shown below. This view lets the user asses the segmentation quality in 3D.



"Inspection Tools-->View Cell Patterns In Imaris" shows a 3D visualization of all the detected/segmented cells color coded and grouped by their cell cycle state as shown below. This yields a quick qualitative assessment of the overall cell cycle state distribution and the morphologies of the cells in each cell cycle state. The rich set of features provided by IMARIS can now be used to do further analysis of the data.



Training phase

In this step, a small subset of the imaged data is selected to train/teach the machine learning components of the algorithms used for nuclei segmentation and cell cycle state identification. Manual annotation tools provided within the software are then used to generate training data that will be used to build the machine learning models.

Pre-trained models built from manual annotations performed on HT1080 fibrosarcoma xenograft tumors in living mice treated with three antimitotic drugs (Paclitaxel, Eribulin, and Ispinesib) and imaged using intravital microscopy for 7-8 days are included with the software. The user can skip the laborious training phase and use these pre-trained models, if the data being analyzed is similar to this setup.

Guidelines for selecting training data

Since the basic idea of machine learning algorithms is to learn a task from manual annotations performed on a set of examples (training data) by a human observer and later be able to mimic him, the data selected for training the machine learning models must be as diverse as the data used in the final analysis phase in terms of possible variations in image quality, imaging conditions, and the morphologies and appearances of the cells being analyzed. In general, machine learning algorithms deteriorate in performance if the data being analyzed is significantly different from the data they were trained on.

For example, in our application involving the quantification of in vivo effects of chemotherapeutic drugs on cell cycle progression, the major factors found to contribute to the heterogeneity of the data are (i) after drug treatment an increasingly significant number of cell nuclei showed up in non-elliptical shapes due to a phenomenon multi-nucleation, (ii) each drug was seen to induce unique characteristic effects on the appearance and morphologies of cells in different cell cycle states, and (iii) the quality of the images seemed to change with time for each experiment, especially at later timepoints where in the images seemed to get a bit blurrier sometimes and also increasing number of dead cell fragments were spotted. To capture all possible variations in our training dataset, we chose to include the images of each drug we wanted to analyze at all timepoints in one/two locations of the gold grid implanted on the tumor.

Training the Nuclei Segmentation Algorithm

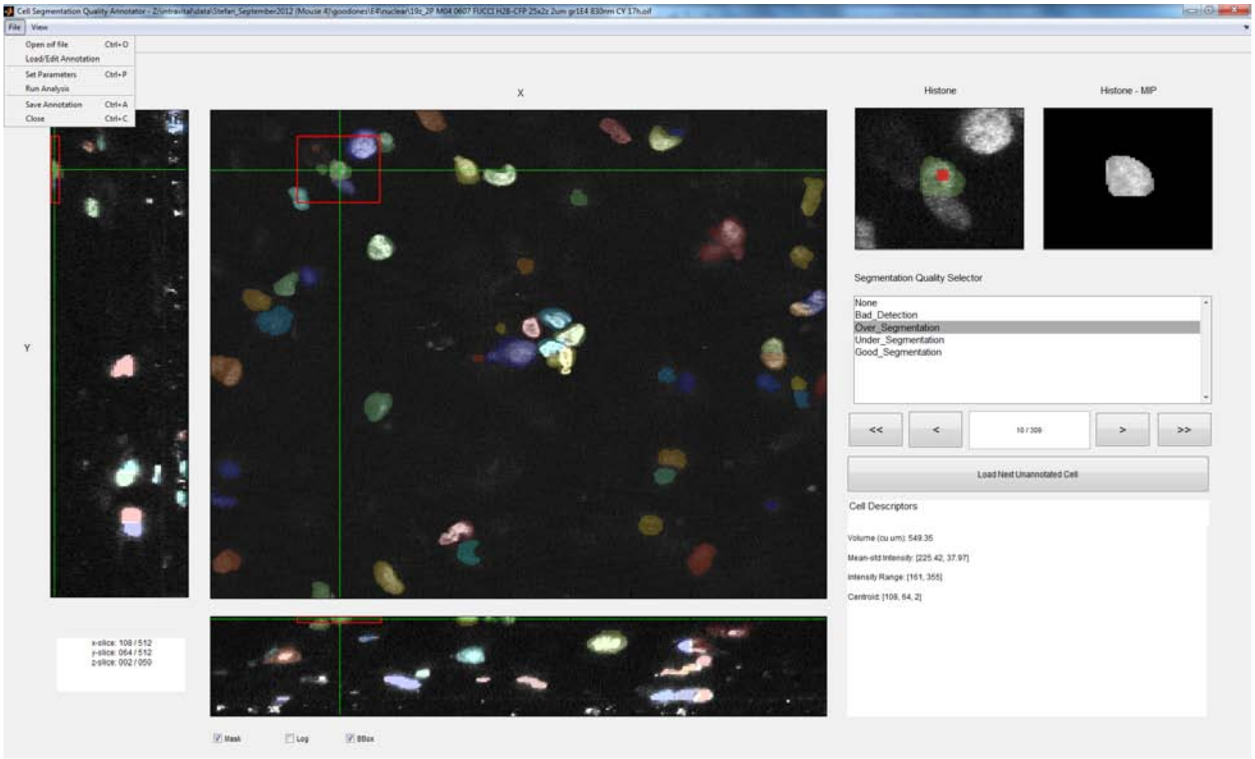
Our nuclei segmentation algorithm uses a machine learning model, henceforth referred to as the region merging model, that is trained to detect and correct over-segmentation errors. Specifically, given a pair of segmented regions touching each other, the region merging model determines whether (i) they belong to same cell nucleus that was over-segmented and hence should be merged, or (ii) they belong to two distinct nuclei. Further details about the algorithm are available in the paper.

Here, we will present the procedure that needs to be followed to (i) Use an annotation tool to generate training data for teaching the region merging model, and (ii) Use the generated training data to build the region merging model that can then be used to detect and correct over-segmentation errors. Each of these steps are described in detail below:

1. Generate training data for the region merging model

To start the annotation tool, click the button named "Annotate training data for the region merging model" in the main control panel (Figure 2) of the software. This will bring up an annotation tool shown below.

Figure 3: Annotation tool for generating training data for the region merging model used for nuclei segmentation


Cell Segmentation Quality Annotator - Z:\intravital\data\Stefan_September2012 (Mouse 4)\goodones\E4\nuclear\19z_2P M04 0607 FUCCI H2B-CFP 25x2z 2um gr1E4 830nm CY 17h.oif
File
View
Open oif file Ctrl+O
Load/Edit Annotation
Set Parameters Ctrl+P
Run Analysis
Save Annotation Ctrl+A
Close Ctrl+C
X
Y
Histone
Histone - MIP
Segmentation Quality Selector
<<
<
>
>>
Load Next Unannotated Cell
Cell Descriptors

Segmentation Quality Selector: None / Bad_Detection / Over_Segmentation (selected) / Under_Segmentation / Good_Segmentation

10 / 309

Volume (cu um): 549.35

Mean-std Intensity: [225.42, 37.97]

Intensity Range: [161, 355]

Centroid: [108, 64, 2]

x-slice: 108 / 512
y-slice: 064 / 512
z-slice: 002 / 059

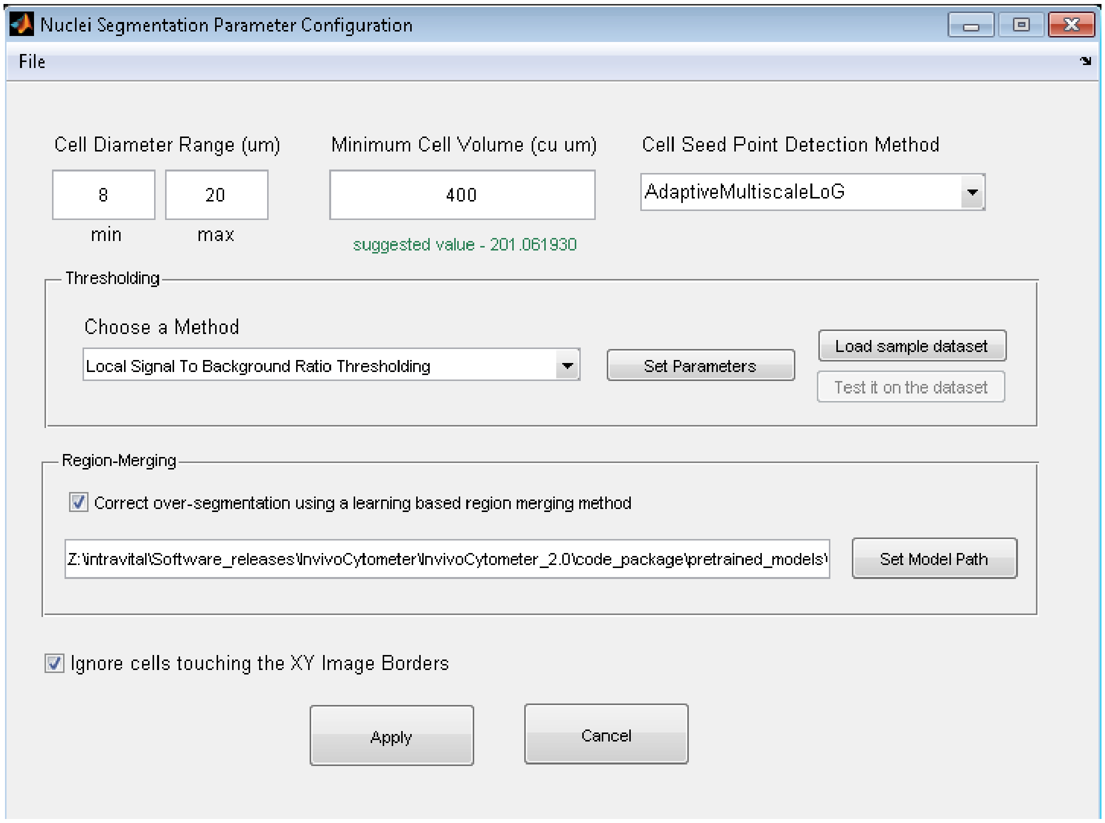
☑ Mask ☐ Log ☑ BBox

This tool allows you to (i) load the raw image data containing a nuclear marker (Histone 2B) channel for annotation or load a pre-existing annotation file to resume/revisit/edit/just-browse it, (ii) run the segmentation algorithm with the region merging model (for incremental training) or without using the region merging model (for training from scratch), (iii) go through each segmented cell nucleus in the dataset one-by-one and annotate its segmentation quality into one of four categories, and (iv) save the annotation which can later be loaded into the tool for editing.

Below is a typical workflow that we used/recommend for using this annotation tool:

a. ***Set Parameters:*** The first thing to do is to set the parameters for the nuclei segmentation algorithm.

To do this, go to the File Menu and click "File-->Set Parameters". This will bring up the following window that will allow you to set/alter the parameters for the nuclei segmentation algorithm.



Cell Diameter Range (um): 8 min, 20 max

Minimum Cell Volume (cu um): 400

suggested value - 201.061930

Cell Seed Point Detection Method: AdaptiveMultiscaleLoG

Thresholding

Choose a Method: Local Signal To Background Ratio Thresholding

Region-Merging

☑ Correct over-segmentation using a learning based region merging method

Z:\intravital\Software_releases\InvivoCytometer\InvivoCytometer_2.0\code_package\pretrained_models\

☑ Ignore cells touching the XY Image Borders

By default, the parameters will be set to values that were empirically found to work well on our data - HT-1080 fibrosarcoma xenografts in living mouse.

For convenience, we have created presets of the parameters for different tasks in our data that can be loaded by clicking "File->Parameter presets->Load a preset". You can also create additional presets using "File->Parameter Presets->Save as preset" and delete a preset using "File->Parameter Presets->Delete a preset".

If your data is similar to ours then leave the default settings. If you want to try our software on a different kind of data, below is the general rationale for setting each of these parameters.

- *Cell Diameter Range (cu um):* Our nuclei segmentation requires a rough estimate of the range of nuclei diameters (use minimum or mean diameter for ellipsoidal nuclei) to be expected on the data in microns. The min diameter is the most sensitive of all parameters and requires a bit of tuning. If it is set to a very small value you will (i) pick up noisy blob-like regions/objects (such as fragments of dead cells) thereby increasing the misdetection rate, or (ii) over-segment the nuclei that are ellipsoidal in shape with high eccentricities. If you set it to a very high value (larger than the diameter of many nuclei found in your data), you will reduce the over-segmentation and misdetection errors but run into the risk of increased under-segmentation errors.

Considering the high variability in the size and shape (most are ellipsoidal and many are multinucleated) in our data and since it is hard to come up with a principled way to resolve under-segmentation errors, we adopted the strategy of intentionally over-segmenting the nuclei (with very little under-segmentation errors) and then using a machine learning based region merging model to detect and correct over-segmentation errors. To intentionally over-segment the nuclei, we set the min value of the cell diameter range to a low value (making sure we do not increase misdetection errors by picking up noisy objects). A minimum diameter of 8-10 um and the maximum diameter of 20 um worked best for our data.

- *Minimum Cell Volume (cu um):* All cell nuclei whose volume is smaller than this value are discarded. This lets you filter out small blob-like noisy objects (ex: fragments of dead cells). We usually set this to a value between 50-75% of the volume of the smallest nuclei expected in the data. For annotating training data for the region merging model, we found a value of 200 cu um to be a reasonable value.

- *Cell Seed Point Detection Method:* This list includes different methods that we experimented with for detecting seed points within the cell nuclei. The method named AdaptiveMultiscaleLoG seemed to work best for our data. Further details about this method are available in our paper.

- *Thresholding:* Since a single thresholding algorithm may not work on all types of data, we provide a set of thresholding algorithms that the user can try and pick from. Use the popup menu underneath “Choose a Method” to pick a thresholding method. Once a method is selected, the user can set/alter its parameters by clicking on the “Set Parameters” button. To try it on a sample data set, first load the dataset by clicking on the “Load sample dataset” button. Once the dataset is loaded, click the “Test it on the dataset” button to try the currently selected thresholding method on it.

- *Ignore cells touching the XY image borders:* This specifies whether or not to discard the cell nuclei that touch the X or Y borders of the image. We decided not to discard cells that touch the Z border because that was resulting in the loss of a significant number of cell nuclei but this option can be easily provided if needed.

- *Region-Merging -> Correct over-segmentation using a learning based region merging method:* This specifies whether or not to use the region merging model to detect and correct over-segmentation errors. If you check this box, then you have to select the model by clicking the button named "Set Model Path".

  For our application, we did not use the region merging model for annotating the training data for the region merging model. However, one can use this option to train the region merging model in an incremental/online fashion where in you can first annotate a few datasets without the using any region merging model for nuclei segmentation, build the model, and then use the generated model to segment the nuclei on some datasets, annotate them, rebuild/update the model and repeat this process until the model yields a satisfactory segmentation. By doing so, the model gets incrementally trained on the errors and thereby might ease out laborious annotation effort by reduced the number of datasets that need to be annotated. We still haven't experimented with this approach but we believed it is a nice idea to try out.

b. ***Load data and run analysis**:* You can either load the raw image data containing a nuclear marker (Histone-2B) channel or load a pre-existing annotation file.

To load raw image data, go the File Menu, and click "File --> Load oif file". This will bring up a window that will allow you to select the image data file. Currently the tool supports *.oif and *.oib files that come out of the Olympus FV1000 confocal/two-photon microscope. In a future, release we will be adding support for other file formats. If the data contains multiple channels, you will be prompted to enter the channel-id that contains the nuclear marker histone-2B.

Once the image data is selected, the tool will automatically run the analysis and show you the result (this might take somewhere between 3-10 minutes depending on size of the volume, the density of the nuclei in your dataset, and the number of processing cores available for parallel processing). You will also be shown the results of all intermediate steps of the nuclei segmentation algorithm.

To load an annotation file (named CellSegmentationQualityAnnotation.mat), go the File Menu, and click "File --> Load/Edit Annotation ". This will bring up a window that will allow you to select the annotation file.

If you want, you can alter the parameters as described above by clicking "File-->Set Parameters" and re-run the analysis with the new parameters by clicking "File-->Run Analysis".

What do the Image panels show?

The three largest image panels show the ortho-slices/cross-sections along the X-, Y-, and Z- axis of the 3D volume (use checkbox named "log" to apply a log transformation to see it better) overlaid with the segmentation mask (this can be toggled on/off using the checkbox named "mask"). You can use the mouse scroll wheel to navigate through the slices. To navigate through the X- and Y- slices, you have to position the mouse pointer inside the corresponding image panel and move the mouse scroll wheel. Moving the mouse scroll wheel anywhere else in the window will let you navigate through the Z-slices. The green lines shown where the other ortho-slices are located.

The two small image panels on the top-right show a zoomed-in version of a cropped XY region (defined by the red bounding box in the larger image panels) around the currently active cell in: (i) Histone channel, (ii) Histone MIP - Maximum intensity projection of the cell pixels in the cropped region of Histone channel along the z-axis.

c. ***Annotate the quality of each segmented cell:*** Go through each cell and annotation its segmentation quality by selecting one of the categories shown in the list named "Segmentation Quality Selector" located in the mid-right portion of the window: (i) Bad_Detection: region does not belong to a cell nucleus, (ii) Over_Segmentation: region belongs to a part/fragment of an over-segmented cell nucleus, (iii) Under_Segmentation: region belongs to two or more cell nuclei, (iv) Good_Segmentation: region represents the whole of a single well-segmented cell-nucleus.

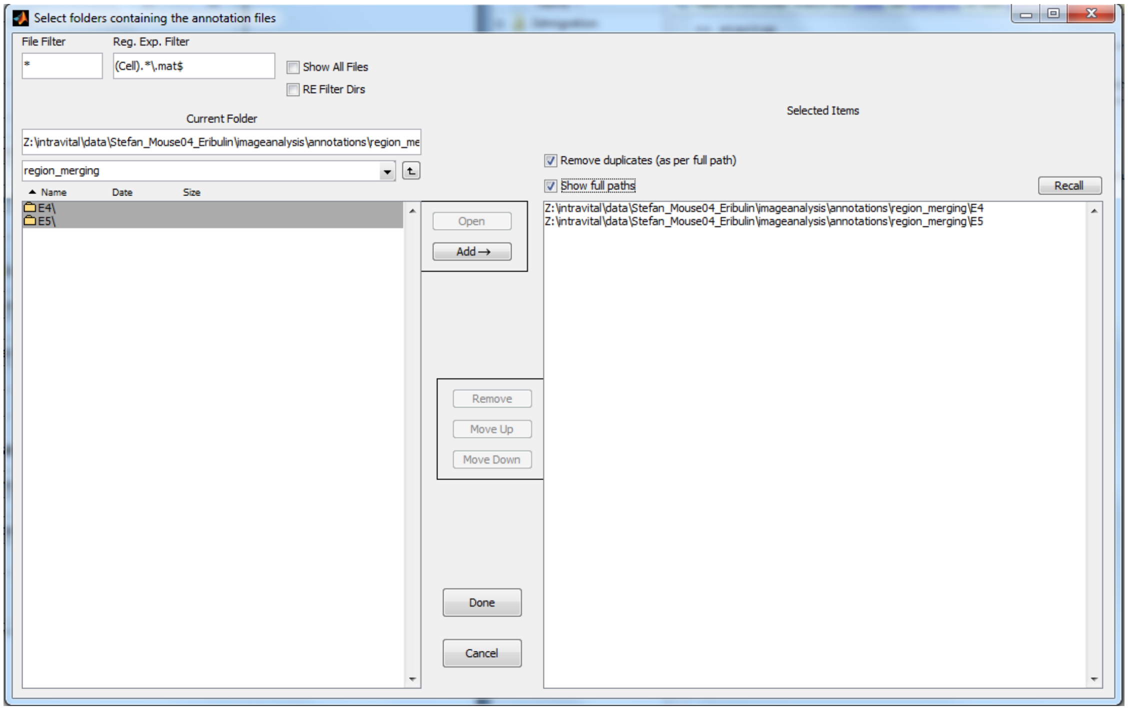
How to navigate through the cells?

The buttons "<<" (show first cell), "<" (show previous cell), ">" (show next cell), ">>" (show last cell) will let you navigate through each cell. The textbox located in the middle of these buttons show <currently-active-cell-id>/<total-cell-count>. The button named "Load Next Unannotated Cell" will jump to the next unannotated cell.

d. ***Save Annotation:*** To save the annotation, go the File Menu and click "File-->Save Annotation". This will bring up a window that will allow you to select the directory in which the annotation file will be saved. The annotation performed on each image data should be stored in a separate directory. Also, you can save an incomplete annotation and resume it later by clicking "File-->Load/Edit Annotation".

2. Build the region merging model

Once you have generated enough training data using the annotation tool described above, you can build the region merging model from the generated training data by clicking "Build region merging model from training data" button in the main control panel (Figure 2) of the software. This will bring up a window that will allow you to select all the "directories" (not files) containing an annotation file as shown below.



File Filter: *
Reg. Exp. Filter: (Cell).*\.mat$
□ Show All Files
□ RE Filter Dirs
Current Folder: Z:\intravital\data\Stefan_Mouse04_Eribulin\imageanalysis\annotations\region_me
region_merging
E4\
E5\

Selected Items
☑ Remove duplicates (as per full path)
☑ Show full paths
Z:\intravital\data\Stefan_Mouse04_Eribulin\imageanalysis\annotations\region_merging\E4
Z:\intravital\data\Stefan_Mouse04_Eribulin\imageanalysis\annotations\region_merging\E5

You can add as many folders as you want by navigating to it and clicking the button named "Add" that will add it to the list of selected folders/directories (containing an annotation files) shown on the right. After selecting all the directories, click the button named "Done" to generate the region merging model. The model generation process will take quite some time (60-80 seconds per volume/dataset).

To facilitate an organized storage of directories, the tool will recursively search for annotation files up to two directory-levels deeper below each selected directory (ex: You can store/organize the directories containing the annotations performed on the image data corresponding to different timepoints acquired at a particular tumor location of a mouse under one root directory).

For advanced users, in addition to the model, the tool also generates a log file for debugging and additional features files (as csv and arff files) that can be loaded into third-party machine learning software for exploration. We used a library called WEKA (http://www.cs.waikato.ac.nz/ml/weka/) to explore, design, conceptualize, train, and validate all the machine learning models used in our framework.

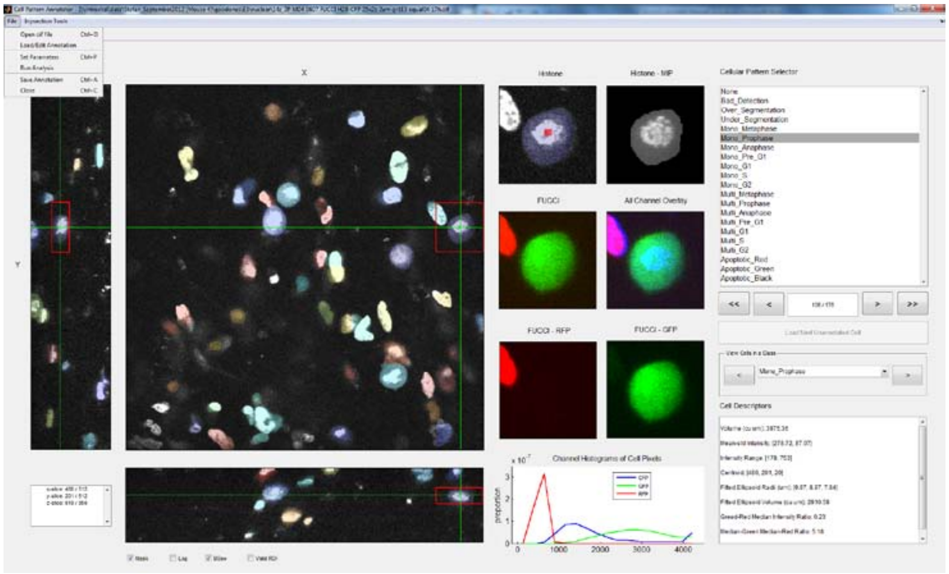
Training the Cell Cycle State Identification Algorithm

Our cell cycle state identification algorithm uses a machine learning model, henceforth referred to as the cell cycle state identification model, that is trained to identify the cell cycle state of a given cell. Further details about the algorithm are available in the paper.

Here, we will present the procedure that needs to be followed to (i) Use an annotation tool to generate training data for teaching the cell cycle state identification model, and (ii) Use the generated training data to build the cell cycle state identification model that can then be used to identify the cell cycle state of any given cell. Each of these steps are described in detail below:

1. Generate training data for the cell cycle state identification model

   To start the annotation tool, click the button named "Annotate training data for cell cycle state identification model" in the main control panel (Figure 2) of the software. This will bring up the annotation tool shown below.



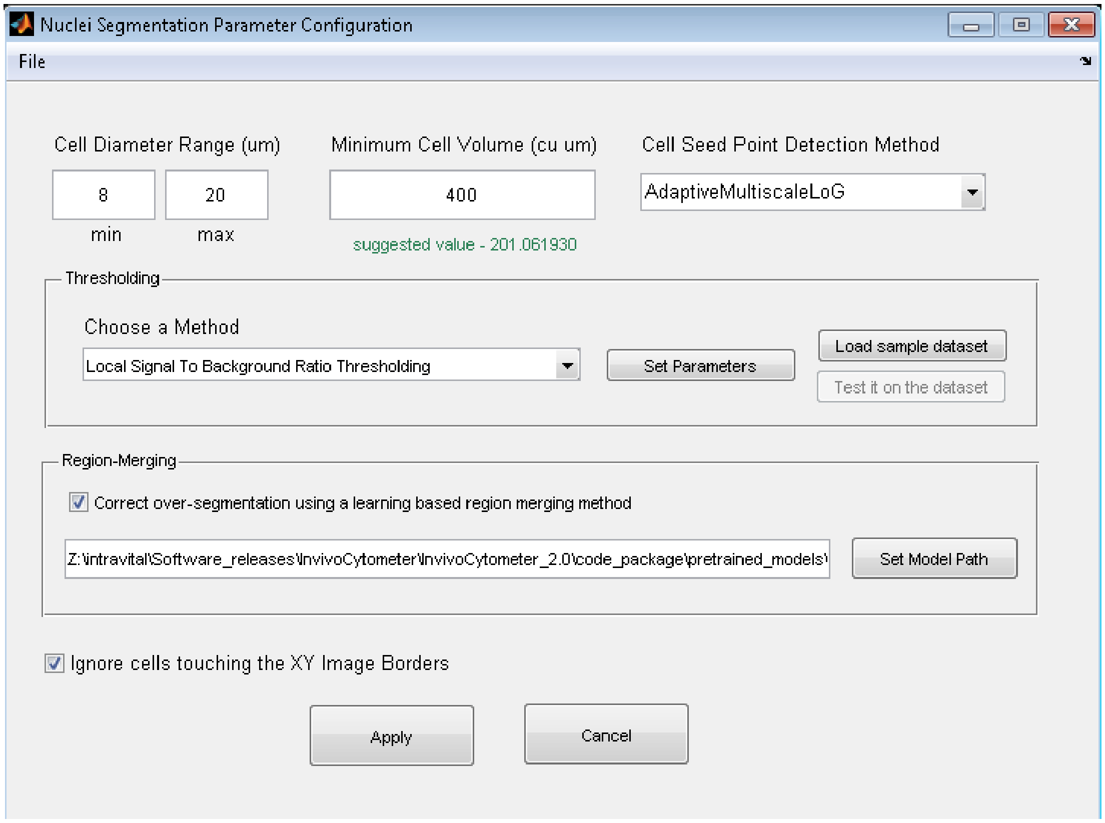
Volume (cu um): 3875.36
Weighted Intensity: [278.72, 87.07]
Intensity Range: [178, 753]
Centroid: [400, 201, 20]
Fitted Ellipsoid Radii (um): [9.87, 8.87, 7.84]
Fitted Ellipsoid Volume (cu um): 2810.28
Green-Red Median Intensity Ratio: 0.23
Median-Green Median-Red Ratio: 5.18

This tool allows you to (i) load the raw image data containing the nuclear marker (Histone 2B) and the FUCCI cell cycle reporter for annotation or load a pre-existing annotation file to resume/revisit/edit/just-browse it, (ii) run the nuclei segmentation algorithm, (iii) go through each segmented cell nucleus in the dataset one-by-one and annotate its cell cycle state, and (iv) save the annotation which can later be loaded into the tool for editing.

Below is a typical workflow that we used/recommend for using this annotation tool:

a. ***Set Parameters:*** The first thing to do is to set the parameters for the nuclei segmentation algorithm.

To do this, go to the File Menu and click "File-->Set Parameters". This will bring up the following window that will allow you to set/alter the parameters for the nuclei segmentation algorithm.



Cell Diameter Range (um): 8 (min), 20 (max)

Minimum Cell Volume (cu um): 400

suggested value - 201.061930

Cell Seed Point Detection Method: AdaptiveMultiscaleLoG

Choose a Method: Local Signal To Background Ratio Thresholding

☑ Correct over-segmentation using a learning based region merging method

Z:\intravital\Software_releases\InvivoCytometer\InvivoCytometer_2.0\code_package\pretrained_models\

☑ Ignore cells touching the XY Image Borders

By default, the parameters will be set to values that were empirically found to work well on our data - HT-1080 fibrosarcoma xenografts in living mouse.

For convenience, we have created presets of the parameters for different tasks in our data that can be loaded by clicking "File->Parameter presets->Load a preset". You can also create additional presets using "File->Parameter Presets->Save as preset" and delete a preset using "File->Parameter Presets->Delete a preset".

Make sure you use the same cell diameter range and seed point detection algorithm that you used to annotate data for training the region merging model.

If your data is similar to ours then use the default settings. If you want to try our software on a different kind of data, below is the general rationale for setting each of these parameters.

- *Cell Diameter Range (cu um):* Our nuclei segmentation requires a rough estimate of the range of nuclei diameters (use minimum or mean diameter for ellipsoidal nuclei) to be expected on the data in microns. The min diameter is the most sensitive of all parameters and requires a bit of tuning. If it is set to a very small value you will (i) pick up noisy blob-like regions/objects (such as fragments of dead cells) thereby increasing the misdetection rate, or (ii) over-segment the nuclei that are ellipsoidal in shape with high eccentricities. If you set it to a very high value (larger than the diameter of many nuclei found in your data), you will reduce the over-segmentation and misdetection errors but run into the risk of increased under-segmentation errors.

  Considering the high variability in the size and shape (most are ellipsoidal and many are multinucleated) in our data and since it is hard to come up with a principled way to resolve under-segmentation errors, we adopted the strategy of intentionally over-segmenting the nuclei (with very little under-segmentation errors) and then using a machine learning based region merging model to detect and correct over-segmentation errors. To intentionally over-segment the nuclei, we set the min value of the cell diameter range to a low value (making sure we do not increase misdetection errors by picking up noisy objects). A minimum diameter of 8-10 um and the maximum diameter of 20 um worked best for our data.

- *Minimum Cell Volume (cu um):* All cell nuclei whose volume is smaller than this value are discarded. This lets you filter out small blob-like noisy objects (ex: fragments of dead cells). We usually set this to a value between 50-75% of the volume of the smallest nuclei expected in the data.

- *Cell Seed Point Detection Method:* This list includes different methods that we experimented with for detecting seed points within the cell nuclei. The method named AdaptiveMultiscaleLoG seemed to work best for our data. Further details about this method are available in our paper.

- *Thresholding:* Since a single thresholding algorithm may not work on all types of data, we provide a set of thresholding algorithms that the user can try and pick from. Use the popup menu underneath “Choose a Method” to pick a thresholding method. Once a method is selected, the user can set/alter its parameters by clicking on the “Set Parameters” button. To try it on a sample data set, first load the dataset by clicking on the “Load sample dataset” button. Once the dataset is loaded, click the “Test it on the dataset” button to try the currently selected thresholding method on it.

- *Ignore cells touching the XY image borders:* This specifies whether or not to discard the cell nuclei that touch the X or Y borders of the image. We decided not to discard cells that touch the Z border because that was resulting in the loss of a significant number of cell nuclei but this option can be easily provided if needed.

- *Region-Merging -> Correct over-segmentation using a learning based region merging method:* This specifies whether or not to use the region merging model to detect and correct over-segmentation errors. If you check this box, then you have to select the model by clicking the button named "Set Model Path".

b. ***Load data and run analysis:*** You can either load the raw image data or load a pre-existing annotation file.

To load raw image data, go the File Menu, and click "File --> Load oif file". This will bring up windows that will allow you to select the image data file. Since, in our application, the histone marker channel was acquired in two-photon mode (for better quality along depth) and FUCCI cell cycle reporter was acquired in the confocal mode, we assume that the histone and fucci data reside in two separate files. So, you will first be asked to select the file containing the histone marker after which another window will popup asking you to select the file containing the fucci cell cycle reporter. Currently the tool supports *.oif and *.oib files that come out of the Olympus FV1000 confocal/two-photon microscope. In a future, release we will be adding support for other common file formats. After selecting the files containing the histone and FUCCI markers, you will be given the option of specifying a slice range of interest. You can use this to optionally tell the software to crop out some slices from the top and bottom of the volume where the image quality is often poor.

Once the image data is selected, the tool will automatically run the analysis and show you the result (this might take somewhere between 3-10 minutes depending on size of the volume, the density of the nuclei in your dataset, and the number of processing cores available for parallel processing). You will also be shown the results of all intermediate steps of the nuclei segmentation algorithm.

If you want, you can alter the parameters as described above by clicking "File-->Set Parameters" and re-run the analysis with the new parameters by clicking "File-->Run Analysis".

To load an annotation file (named CellPatternAnnotation.mat), go the File Menu, and click "File -->Load/Edit Annotation ". This will bring up a window that will allow you to select the annotation file.

What do the Image panels show?

The three largest image panels show the ortho-slices/cross-sections along the X-, Y-, and Z- axis of the 3D volume (use checkbox named "log" to apply a log transformation to see it better) overlaid with the segmentation mask (this can be toggled on/off using the checkbox named "mask"). You can use the mouse scroll wheel to navigate through the slices. To navigate through the X- and Y- slices, you have to position the mouse pointer inside the corresponding image panel and move the mouse scroll wheel. Moving the mouse scroll wheel anywhere else in the window will let you navigate through the Z-slices. The green lines shown where the other ortho-slices are located.

The six small image panels show a zoomed-in version of a cropped XY region (defined by the red bounding box shown in the XY slice) around the currently active cell in: (i) Histone channel - useful for seeing the raw texture inside the cell, (ii) Histone MIP - Maximum intensity projection of the cell pixels in the cropped region of Histone channel along the z-axis useful for specifically for identifying mitotic cells, (iii) FUCCI - overlay of the two FUCCI GFP and RFP channels useful specifically for identifying cells in Late G1/Early S phase that appear yellow/orange in color in this view, (iv) All Channels - Overlay of the histone channel and the two FUCCI channels, (v) FUCCI-RFP channel -- this shows cells in G1, (vi) FUCCI-GFP channel - this shows cells in S/G2/M phases of the cell cycle.

c. ***Annotate the cell cycle state of each segmented cell:*** Go through each cell and annotate its cell cycle state by selecting one of the categories shown in the list named "Cell Pattern Selector" located in the top-right portion of the window:
(i) Bad_Detection: region does not belong to a cell nucleus,
(ii) Over_Segmentation: region belongs to a part/fragment of an over-segmented cell nucleus,
(iii) Under_Segmentation: region belongs to two or more cell nuclei,
(iv) Mono_ Metaphase or Multi_Metaphase: mono/multi-nucleated cell in metaphase,
(v) Mono_Prophase or Multi_Prophase: mono/multi-nucleated cell in prophase,
(vi) Mono_Anaphase or Multi_Anaphase: mono/multi-nucleated cell in anaphase,
(vii) Mono_Pre_G1 or Multi_Pre_G1 - mono/multi-nucleated cell in Early G1 phase appearing only in the histone channel,
(viii) Mono_G1 or Multi_G1: mono/multi-nucleated cell in G1 phase appearing predominantly in FUCCI-RFP and hence showing up red in color in the FUCCI GFP-RFP overlay,
(ix) Mono_S or Multi_S: mono/multi-nucleated cell in Late G1/Early S phase (here called S) appearing in both FUCCI-GFP and FUCCI-RFP channels and hence showing up orange-yellow in color in the FUCCI GFP-RFP overlay,
(x) Mono_G2 or Multi_G2: mono/multi-nucleated cell in S/G2 phases (here called G2) without any chromosome condensation seen in mitotsis and appearing predominantly in FUCCI-GFP showing up green in color in the FUCCI GFP-RFP overlay,
(xi) Apoptotic_Red - apoptotic cell appearing red in color in FUCCI GFP-RFP overlay,
(xii) Apoptotic_Red - apoptotic cell appearing green in color in FUCCI GFP-RFP overlay,
(xiii) Apoptotic_Black - apoptotic cell appearing only in the histone channel.

The mitotic phases are identified by the presence of condensation of chromosomes that can be seen readily in the Histone-MIP image panel. Distinguishing cells in the Late-G1/Early-S from cells in G1 and S/G2 and mono vs multi-nucleation was quite confusing for our biologists. For our application, we (i) clubbed all the intra-mitotic phases into one class, (ii) clubbed together both mono- and multi-nucleated cells of each cell cycle state together, and (iii) ignored the Apoptotic cells. For your guidance, shown below are some examples manually annotation by our biologist in each class.

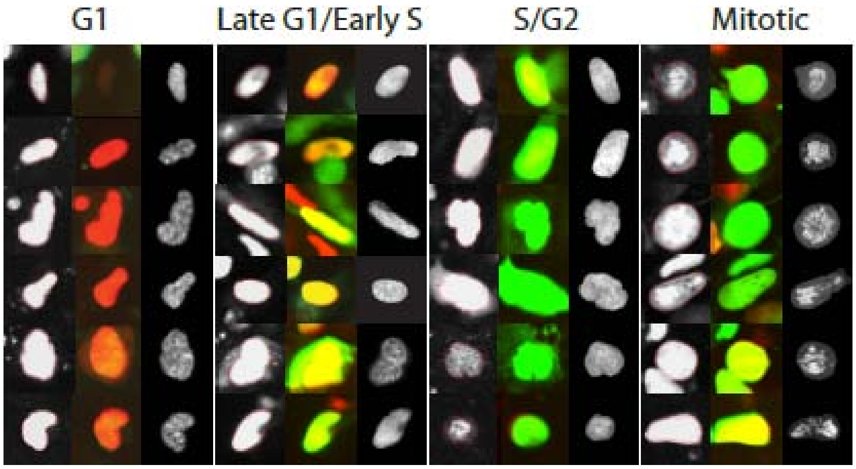


Figure 4 Manually annotated examples in each cell cycle state

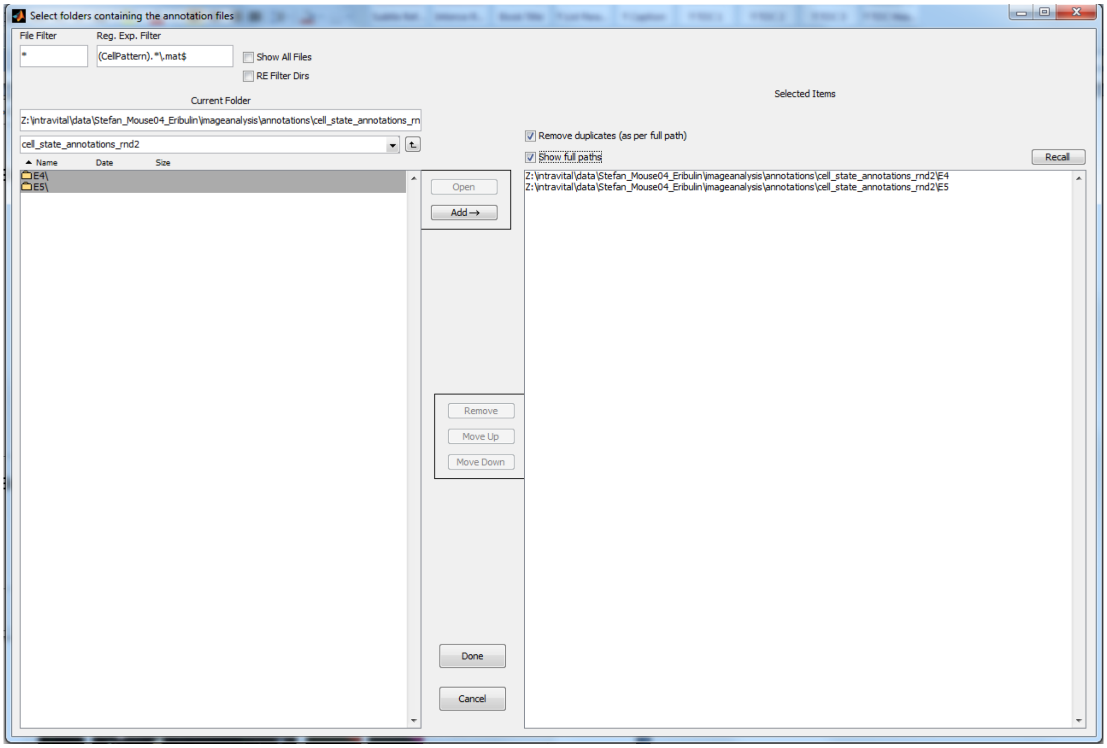
How to navigate through the cells?

The buttons "<<" (show first cell), "<" (show previous cell), ">" (show next cell), ">>" (show last cell) will let you navigate through each cell. The textbox located in the middle of these buttons show <currently-active-cell-id>/<total-cell-count>. The button named "Load Next Unannotated Cell" will jump to the next unannotated cell.

d. ***Save Annotation:*** To save the annotation, go the File Menu and click "File-->Save Annotation". This will bring up a window that will allow you to select the directory in which the annotation file will be saved. While saving the annotation you will be asked whether or not to save cropped thumbnails of cells in each cell cycle state for further inspection and exploration. The annotation performed on each image data should be stored in a separate directory. Also, you can save an incomplete annotation and resume it later by clicking "File-->Load/Edit Annotation".

2. Build the cell cycle state identification model

Once you have generated enough training data using the annotation tool described above, you can build the cell cycle state identification model from the generated training data by clicking the button named "Build cell cycle state identification model from training data" in the main control panel (Figure 2) of the software. This will bring up a window that will allow you to select all the "directories" (not files) containing an annotation file as shown below.



File Filter: *

Reg. Exp. Filter: (CellPattern).*\.mat$

☐ Show All Files

☐ RE Filter Dirs

Current Folder: Z:\intravital\data\Stefan_Mouse04_Eribulin\imageanalysis\annotations\cell_state_annotations_rn

cell_state_annotations_rnd2

E4\
E5\

☑ Remove duplicates (as per full path)

☑ Show full paths

```
Z:\intravital\data\Stefan_Mouse04_Eribulin\imageanalysis\annotations\cell_state_annotations_rnd2\E4
Z:\intravital\data\Stefan_Mouse04_Eribulin\imageanalysis\annotations\cell_state_annotations_rnd2\E5
```

You can add as many folders as you want by navigating to it (on the left) and clicking the button named "Add" that will add it to the list of selected folders/directories (containing an annotation files) shown on the right. After selecting all the directories, click the button named "Done" to generate the region merging model. The model generation process will take quite some time (~2 minutes per volume/dataset), so you may want to get some coffee or do something else while this runs.

To facilitate an organized storage of directories, the tool will recursively search for annotation files up to two directory-levels deeper below each selected directory (ex: You can store/organize the directories containing the annotations performed on the image data corresponding to different timepoints acquired at a particular tumor location of a mouse under one root directory).

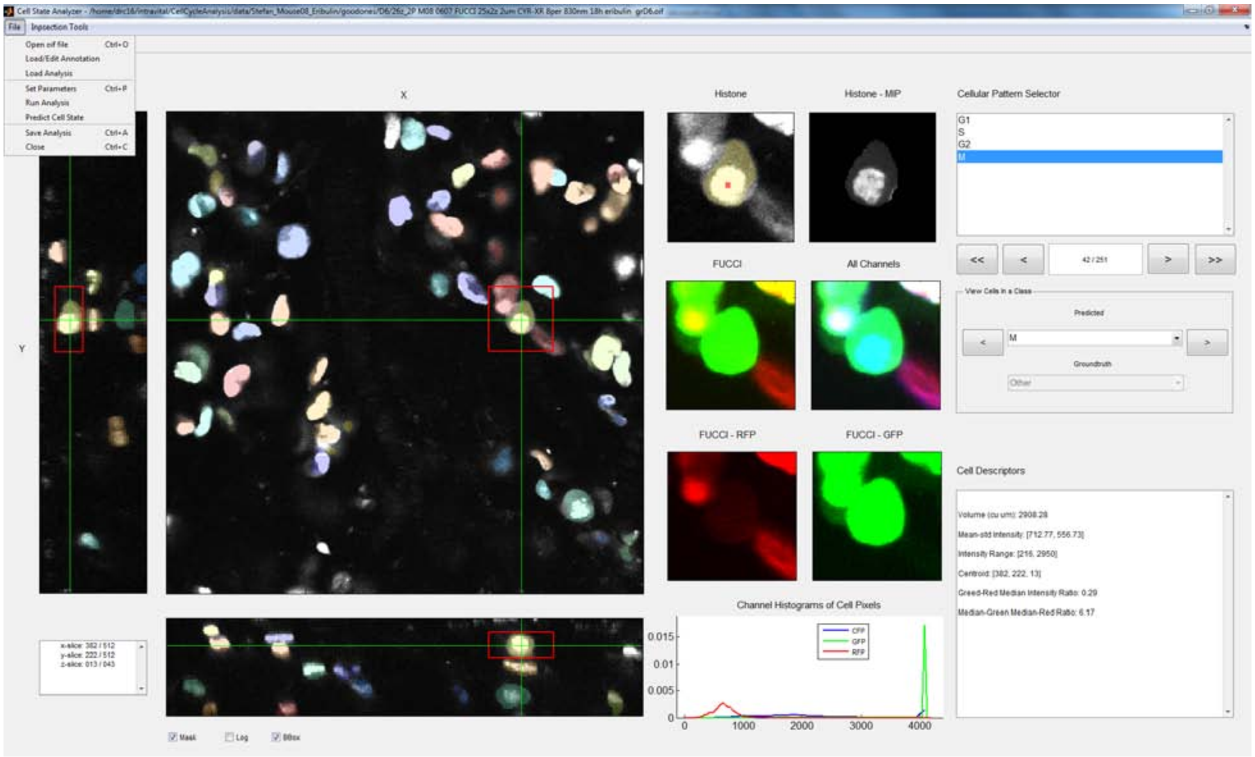
For advanced users, in addition to the model, the tool also generates a log file for debugging and additional features files (as csv and arff files) that can be loaded into third-party machine learning software for exploration. We used a library called WEKA (http://www.cs.waikato.ac.nz/ml/weka/) to explore, design, conceptualize, train, and validate all the machine learning models used in our framework. So, you can load these feature files into WEKA, try different classification strategies, pick the best one, save the classification model to a file, and use it as the model in our software. This gives you an immense amount of flexibility in building the model.

Automated Analysis Phase

Once the training phase is completed and the machine learning models have been built, the software can now be used for fully automated analysis.

Perform Complete Cell Cycle Profiling

To perform automated cell cycle profiling, click the button named "Perform Complete Cell Cycle Profiling" in the main control panel of the software (Figure 2). This will bring the analysis tool named "Cell State Analyzer" shown below.



x-slice: 382 / 512
y-slice: 222 / 512
z-slice: 013 / 043

42 / 251

Predicted: M
Groundtruth: Other

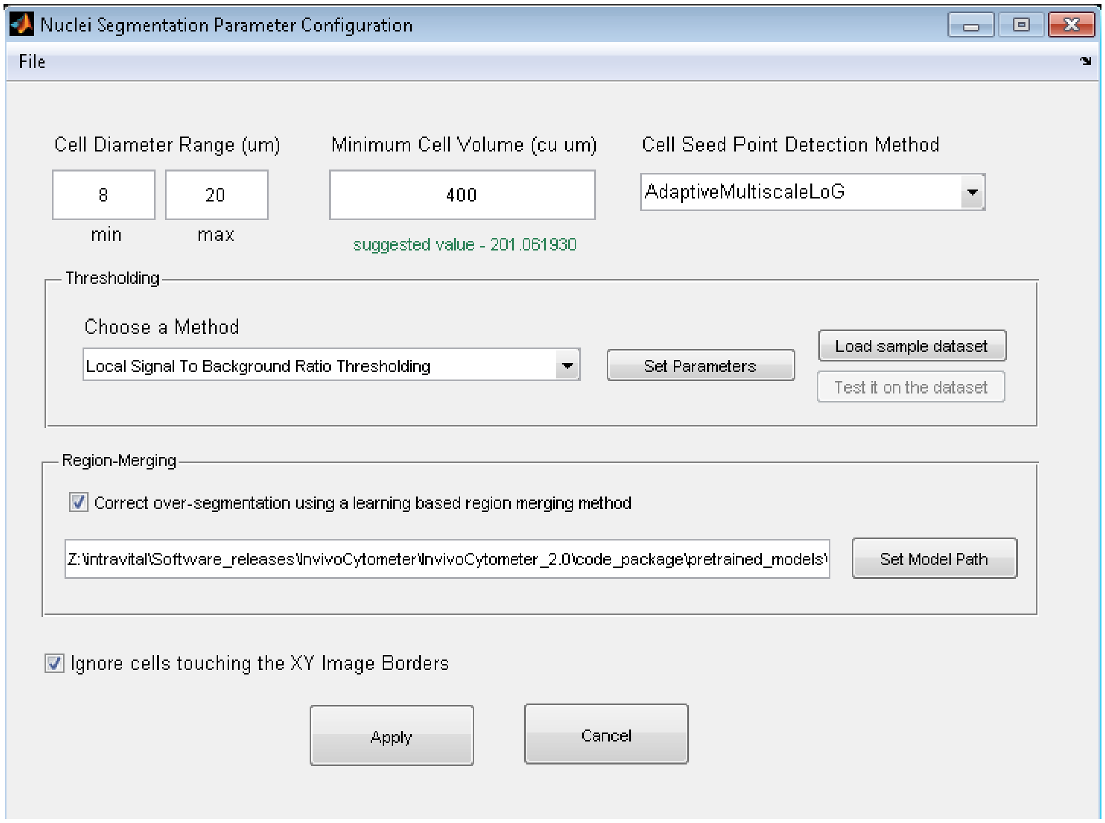
Volume (cu um): 2908.28
Mean-std Intensity: [712.77, 556.73]
Intensity Range: [216, 2950]
Centroid: [382, 222, 13]
Greed-Red Median Intensity Ratio: 0.29
Median-Green Median-Red Ratio: 6.17

This tool allows you to (i) load the raw image data for analysis or load a pre-existing cell cycle state annotation file to compare the performance of the cell cycle state identification model with groundtruth or load a pre-existing analysis file created using this tool for reanalyzing/just-browsing, (ii) run cell cycle analysis workflow, (iii) check the analysis result at a single cell level, and (iv) save the analysis file.

Below is a typical workflow that we used/recommend for using this analysis tool:

a. ***Set Parameters:*** The first thing to do is to set the parameters for the nuclei segmentation and the cell cycle state identification algorithm .

   To set the parameters for the nuclei segmentation algorithm, go to the File Menu and click "File-->Set Parameters->Nuclei segmentation". This will bring up the following window that will allow you to set/alter the parameters for the nuclei segmentation algorithm.



Nuclei Segmentation Parameter Configuration

Cell Diameter Range (um): 8 min, 20 max

Minimum Cell Volume (cu um): 400

suggested value - 201.061930

Cell Seed Point Detection Method: AdaptiveMultiscaleLoG

Thresholding

Choose a Method: Local Signal To Background Ratio Thresholding

Region-Merging

☑ Correct over-segmentation using a learning based region merging method

Z:\intravital\Software_releases\InvivoCytometer\InvivoCytometer_2.0\code_package\pretrained_models\

☑ Ignore cells touching the XY Image Borders

By default, the parameters will be set to values that were empirically found to work well on our data - HT-1080 fibrosarcoma xenografts in living mouse.

For convenience, we have created presets of the parameters for different tasks in our data that can be loaded by clicking "File->Parameter presets->Load a preset". You can also create additional presets using "File->Parameter Presets->Save as preset" and delete a preset using "File->Parameter Presets->Delete a preset".

If your data is similar to ours then use the default settings. If you want to try our software on a different kind of data, below is the general rationale for setting each of these parameters:

- *Cell Diameter Range (cu um):* Our nuclei segmentation requires a rough estimate of the range of nuclei diameters (use minimum or mean diameter for ellipsoidal nuclei) to be expected on the data in microns. The min diameter is the most sensitive of all parameters and requires a bit of tuning. If it is set to a very small value you will (i) pick up noisy blob-like regions/objects (such as fragments of dead cells) thereby increasing the misdetection rate, or (ii) over-segment the nuclei that are ellipsoidal in shape with high eccentricities. If you set it to a very high value (larger than the diameter of many nuclei found in your data), you will reduce the over-segmentation and misdetection errors but run into the risk of increased under-segmentation errors.

  Considering the high variability in the size and shape (most are ellipsoidal and many are multinucleated) in our data and since it is hard to come up with a principled way to resolve under-segmentation errors, we adopted the strategy of intentionally over-segmenting the nuclei (with very little under-segmentation errors) and then using a machine learning based region merging model to detect and correct over-segmentation errors. To intentionally over-segment the nuclei, we set the min value of the cell diameter range to a low value (making sure we do not increase misdetection errors by picking up noisy objects). A minimum diameter of 8-10 um and the maximum diameter of 20 um worked best for our data.

- *Minimum Cell Volume (cu um):* All cell nuclei whose volume is smaller than this value are discarded. This lets you filter out small blob-like noisy objects (ex: fragments of dead cells).
  We usually set this to a value between 50-75% of the volume of the smallest nuclei expected in the data.

- *Cell Seed Point Detection Method:* This list includes different methods that we experimented with for detecting seed points within the cell nuclei. The method named AdaptiveMultiscaleLoG seemed to work best for our data. Further details about this method are available in our paper.

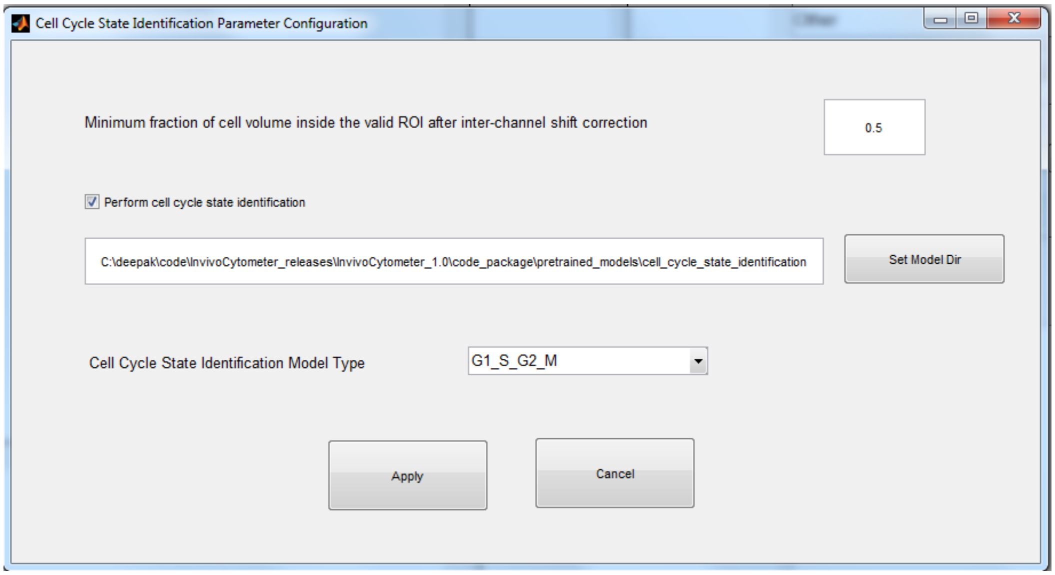
- *Thresholding:* Since a single thresholding algorithm may not work on all types of data, we provide a set of thresholding algorithms that the user can try and pick from. Use the popup menu underneath “Choose a Method” to pick a thresholding method. Once a method is selected, the user can set/alter its parameters by clicking on the “Set Parameters” button. To try it on a sample data set, first load the dataset by clicking on the “Load sample dataset” button. Once the dataset is loaded, click the “Test it on the dataset” button to try the currently selected thresholding method on it.

- *Ignore cells touching the XY image borders:* This specifies whether or not to discard the cell nuclei that touch the X or Y borders of the image. We decided not to discard cells that touch the Z border because that was resulting in the loss of a significant number of cell nuclei but this option can be easily provided if needed.

- *Region-Merging* -> *Correct over-segmentation using a learning based region merging method:* This specifies whether or not to use the region merging model to detect and correct over-segmentation errors. If you check this box, then you have to select the model by clicking the button named "Set Model Path".

To set the parameters for the cell cycle state identification algorithm, go to the File Menu and click "File-->Set Parameters->Cell Cycle State Identification". This will bring up the following window that will allow you to set/alter the parameters for cell cycle state identification.

By default, the parameters will be set to values that were empirically found to work well on our data - HT-1080 fibrosarcoma xenografts in living mouse. If you are running the software on data similar to ours, use the default parameters. If not, below is the general rationale for setting each of these parameters:



Cell Cycle State Identification Parameter Configuration

Minimum fraction of cell volume inside the valid ROI after inter-channel shift correction: 0.5

☑ Perform cell cycle state identification

C:\deepak\code\InvivoCytometer_releases\InvivoCytometer_1.0\code_package\pretrained_models\cell_cycle_state_identification

Cell Cycle State Identification Model Type: G1_S_G2_M

- *Minimum fraction of cell volume inside the valid ROI after inter-channel shift correction:*In our data, there is often a misalignment between the histone and the fucci channels which is corrected using masked image registration algorithm (more details are in the paper). After correcting the misalignment, at the borders of the volume, there will be areas where in data from all three channels (histone + fucci) is not available. We refer to the region of the volume where data from all three channels is avaialble as the valid ROI (region of interest). This parameter specifies the minimum amount of overlap each segmented cell nucleus must have with the valid ROI to be considered as a valid cell nucleus.
- *Perform cell cycle state identification:* This specifies whether or not to do cell cycle state identification after segmenting nuclei when you click the "File->Run Analysis" button. If you check this on, then you have to select the directory containing the cell cycle state identification models by clicking the button named "Set Model Dir".
- *Cell cycle state identification model type:* This allows you to select the type of cell cycle state identification model to use.
    - *G1_S_G2_M:* This is a four-class model that classifies cells into one of G1, S (Late G1/Early S), G2 (S + G2), or Mitotic phases
    - *Interphase_Mitotic:* This is a two-class model that classifies cells into one of Interphase or Mitotic classes.
    - G1_SG2M: This is a two-class model that classifier cells into one of G1, or SG2M (S + G2 + M) classes.

Make sure you use the same parameters used for annotating data in the training phase of the cell cycle state identification model.

b. Load data and run analysis: There are three ways to load data into this tool: (i) load the raw image data for analysis, (ii) load a pre-existing annotation file created using the tool used for cell cycle state annotation to compare the performance of the cell cycle state identification model with groundtruth, or (iii) load a pre-existing analysis file created using this tool for reanalyzing/just-browsing.

To load raw image data, go the File Menu, and click "File --> Open oif file". This will bring up windows that will allow you to select the image data files. Since, in our application, the histone marker channel was acquired in two-photon mode (for better quality along depth) and FUCCI cell cycle reporter was acquired in the confocal mode, we assume that the histone and fucci data reside in two separate files. So, you will first be asked to select the file containing the histone marker after which another window will popup asking you to select the file containing the fucci cell cycle reporter. Currently the tool supports *.oif and *.oib files that come out of the Olympus FV1000 confocal/two-photon microscope. In a future, release we will be adding support for other common file formats.

Once the image data is selected, the tool will automatically run the analysis and show you the result (this might take somewhere between 3-10 minutes depending on size of the volume, the density of the nuclei in your dataset, and the number of processing cores available for parallel processing). You will also be shown the results of all intermediate steps of the nuclei segmentation algorithm if you requested it. We have included sample image data along with the software (check the folder named sample_data/sample_image_data).

To load the cell cycle state annotation file (named CellPatternAnnotation.mat), go the File Menu, and click "File -->Load/Edit Annotation ". This will bring up a window that will allow you to select the annotation file.

To load a pre-existing analysis file (named CellStateAnalyzer.mat) created using this tool, go the File Menu, and click "File -->Load Analysis". This will bring up a window that will allow you to select the analysis file. We have included a sample analysis file along with the software (check the folder named sample_data/sample_analysis_results).

If you want, you can alter the parameters as described above by clicking "File-->Set Parameters" and re-run the analysis with the new parameters by clicking "File-->Run Analysis".

c. *Browse through the analysis result at a single-cell level:*

Navigating through the cells: The button group with buttons "<<" (show first cell), "<" (show previous cell), ">" (show next cell), ">>" (show last cell) will let you navigate through each cell detected, segmented and classified into one of the cell cycle states by our algorithm. The textbox located in the middle of these buttons show <currently-active-cell-id>/<total-cell-count>. As you navigate through the cells, the cell cycle state of the currently active cell is highlighted in the list named "Cellular Pattern Selector" located in the top right part of the window.

You can also look at all the cells in each class (or cell cycle state) by selecting the desired class from the drop-down menu named "Predicted" (within the group named "View cells in a class" in the mid-rightmost part of the window) and navigate through them using the buttons "<" (show previous cell in class) and the ">" (show next cell in class).

Image panels: The three largest image panels show the ortho-slices/cross-sections along the X-, Y-, and Z- axis of the 3D volume (use checkbox named "log" to apply a log transformation to see it better) overlaid with the segmentation mask (this can be toggled on/off using the checkbox named "mask"). You can use the mouse scroll wheel to navigate through the slices. To navigate through the X- and Y- slices, you have to position the mouse pointer inside the corresponding image panel and move the mouse scroll wheel. Moving the mouse scroll wheel anywhere else in the window will let you navigate through the Z-slices. The green lines shown where the other ortho-slices are located.

The six small image panels show a zoomed-in version of a cropped XY region (defined by the red bounding box shown in the XY slice) around the currently active cell in: (i) Histone channel - useful for seeing the raw texture inside the cell, (ii) Histone MIP - Maximum intensity projection of the cell pixels in the cropped region of Histone channel along the z-axis useful for specifically for identifying mitotic cells, (iii) FUCCI - overlay of the two FUCCI GFP and RFP channels useful specifically for identifying cells in Late G1/Early S phase that appear yellow/orange in color in this view, (iv) All Channels - Overlay of the histone channel and the two FUCCI channels, (v) FUCCI-RFP channel -- this shows cells in G1, (vi) FUCCI-GFP channel - this shows cells in S/G2/M phases of the cell cycle.

3D visualization using Imaris: InvivoCytometer interacts with a commercial software called IMARIS through its COM interface to provide nice and interesting 3D visualization of the analysis results that can be accessed from the "Inspection Tools" Menu. To be able to use these features, you should run the tool in the windows platform and should have already installed Imaris (version 7.4 or higher) before running the tool. Read the sneak peak section (Quick sneak peek into the analysis, page 7) for an overview of the 3D visualization features.

d. ***Save Analysis:*** To save the analysis result, go the File Menu and click "File-->Save Analysis". This will bring up a window that will allow you to select the directory in which the annotation file will be saved. While saving the analysis result you will be given the option of whether or not to save cropped thumbnails of cells in each cell cycle state for further inspection and exploration. The analysis performed on each image data should be stored in a separate directory.

Perform Nuclei Segmentation Only

If you are only interested in the nuclei segmentation functionality of the software, after finishing the training phase of the nuclei segmentation algorithm click the button named "Perform Nuclei Segmentation Only" in the main control panel (Figure 2) of the software.

This will bring up the analysis tool that will allow you to: (i) load the raw image data containing a nuclear marker (Histone 2B) channel for annotation or load a pre-existing annotation file to resume/revisit/edit/just-browse it, (ii) run the segmentation algorithm with or without the region merging, (iii) browse through the nuclei segmentation result at a single cell level, and (iv) save the result.

Advanced Usage Guide

This section is intended for advanced users who are well-versed with matlab programming.

Batch analysis on a high-performance compute cluster

This section describes how to deploy our computational framework on a high-performance compute clusters to analysis a large batch of data in parallel.

The code package provided to you contains a matlab script named batchCellCycleAnalysis.m. This is the script that we used to deploy our framework on a compute cluster to analyze a large batch of datasets in parallel. At the beginning of this script are a bunch of parameters (explained with comments) that need to be configured before running it on the cluster. The main parameters are the paths to the region merging and cell cycle state identification model files and an inventory file (created in excel) that contains the paths to the files containing the histone and fucci image data for each dataset you want to analyze.

Extending the functionality of the software

The software is written in a fairly modular fashion allowing the user to extend or customize its functionality if needed. In this section, we intend to point out where the core functionality of the software resides and mention some places that the user can modify to affect/enhance the performance of some key aspects of our computational framework.

Firstly, the core functionality of our computational framework is in two functions included in code package:

- segmentCellsInIntravitalData.m Given a 3D volume of the nuclear marker channel, this function segments the nuclei with or without the region merging model.

- PerformCellCycleAnalysis.m Given the files (*.oif, *.oib) containing the raw image data of the nuclear marker and FUCCI cell cycle reporter, the region merging model file, and the cell cycle state identification model file, this function performs the complete cell cycle profiling.

For both, the region merging model used in the nuclei segmentation algorithm and the cell cycle state identification model, we currently use an ensemble of five random forest classifiers each of which is trained on a balanced subset of the training data. And, we use the well-known machine learning library called WEKA to build the models. If the user wants to use a different classification scheme, this can be done by modifying the following functions:

- BuildWekaModelForRegionMerging.m: Given a weka dataset containing the training data generated (trainRegionMergingModel_3D.m) for the region merging model, this function builds the weka model that is later used to determine whether or not to merge a given pair of regions to resolve an over-segmentation error.

- BuildWekaModelForCellCycleStateIdentification.m: Given a weka dataset containing the training data generated (trainCellStateClassificationModel_3D.m) for the cell cycle state identification model, this function builds the weka model that is later used to identify the cell cycle state of any given cell.

The image-based features used for the region merging model are computed by the function named ComputeRegionMergingFeatures.m.

The image-based features used for the cell cycle state identification models is computed by the following functions:

- ComputeCellStateClassificationFeatures_G1_S_G2_M.m: for the four-class model that classifies cells into G1, Late G1/Early S, G2, or Mitotic

- ComputeCellStateClassificationFeatures_Interphase_Mitotic.m: for two-class model that classifies cells into Interphase (G1 + Later G1/Early S + G2), and Mitotic
- ComputeCellStateClassificationFeatures_G1_SG2M.m: for two-class model that classifies cells into G1 or Later G1/Early S + G2 + Mitotic.