

Running head: High-content siRNA screening

Authors/Affiliations:

Stockwell, Simon R.

MRC Cancer Unit

University of Cambridge

Hutchison/MRC Research Centre

Cambridge, UK

ss2233@MRC-CU.cam.ac.uk

Mittnacht, Sibylle

Cancer Biology

UCL Cancer Institute

London, UK

s.mittnacht@ucl.ac.uk

High-content imaging and RNAi screens for investigating kinase network plasticity

Summary

High-content imaging connects the information-rich method of microscopy with the systematic objective principles of software-driven analysis. Suited to automation and, therefore, considerable scale-up of study size, this approach can deliver multi-parametric data over cell populations or at the level of the individual cell and has found considerable utility in reverse genetic and pharmacological screens. Here we present a method to screen small interfering RNA (siRNA) libraries allowing subsequent observation of the impact of each knockdown on two interlinked, high-content, G1/S-phase cell cycle transition assays related to cyclin dependent kinase (CDK) 2 activity. We show how plasticity within the network governing the activity of this kinase can be detected by combining modifier siRNAs with an siRNA library. The method uses fluorescent immunostaining of a nuclear antigen, Cyclin A, following cell fixation while also preserving the fluorescence of a stably expressed fluorescent protein-tagged reporter for CDK2 activity. We provide methodology for data extraction and handling including an R script that converts the multi-dimensional data into four simple binary outcomes, on which a hit-mining strategy can be built. The workflow described can in principle be adopted to yield quantitative single cell resolved data and mining for outcomes relating to a broad range of other, similar readouts and signalling contexts.

Key Words

High-content, screening, RNAi, siRNA, kinase networks, network plasticity, Cyclin-dependent kinase

Introduction

A combination of fluorescence microscopy and software-driven image analysis, high-content imaging has proven to be a useful tool in cell based screens (1, 2). Particular strengths of the technique are that it delivers multi-parametric, single cell resolved data and it can measure geometric/morphological

parameters in addition to fluorescence intensities for geometrically-defined subcellular regions. High-content imaging can therefore resolve a rich diversity of possible phenotypic endpoint measurements and it is this that makes it suitable for investigations where cellular signalling networks are systematically disrupted and phenotypic measures are the means by which the contribution of the disrupted nodes can be assessed (3, 4).

This method describes the example of an RNA interference (RNAi) screen using reverse-transfection of siRNA arranged in a 96-well format to knockdown candidate genes in adherent mammalian cells. The impact of the RNAi knockdowns on a cellular phenotype is then determined by high-content image analysis. The option of producing a parallel set of siRNA transfections to address the impact of a candidate epistatic modifier siRNA (siMOD) on marker response is described. This transient inclusion of a modifier may be used to trace network changes such that the siRNA screen can either be employed to look for novel epistatic modifiers of the siMOD target gene or, vice versa, to test a candidate siMOD target for epistatic impact on the various targets in the library. After the RNAi period, cells are fixed, stained and imaged according to the requirements of the assays chosen for the study. The analysis produces a data file listing all complete cells imaged per well and their corresponding assay measurements.

The example phenotype used here is successful exit from G1 phase of the cell cycle measured by two separate complementary measures of CDK2 kinase activity associated with G1/S transit: i) accumulation of the CDK2 regulator Cyclin A, and ii) the nuclear to cytoplasmic translocation of a GFP-tagged reporter of cyclin-dependent kinase 2 (CDK2) activity (5–7). RNAi targets that arrest cells in G1 phase of the cell cycle are expected to yield cells with reduced nuclear Cyclin A and reduced CDK2 activity, indicated by reduced presence of the CDK2 reporter in the cytoplasm. After 48 hours of RNAi and prior to the high-content analysis phase of the method, the transfected cells are fixed and fluorescently stained for both Cyclin A and chromatin. Next, fluorescent microscope images are acquired and organised for analysis with the open-source, high-content image analysis software CellProfiler (8).

Often, multiple independent images need to be collected per well to guarantee a sufficient number of cells are assessed per condition. We detail how systematic image collection along with the provided software settings ensure that images are unambiguously related to the wells they are derived from and, thereby, to any particular treatment on the 96-well plate, while at the same time delivering multiplexed single-cell-resolved information relating to the different fluorophore-channel and assay outcomes.

During the high-content phase of the method, images of chromatin stained with a fluorescent dye (Hoechst 33342) allows the automated identification of individual cells and at the same time generates a ‘mask’ on which automated image segmentation is based. The located nuclei are subsequently used to find associated cytoplasmic regions of each cell in the GFP images. Measurements of each fluorophore in the paired nuclei and cytoplasm regions of each imaged cell are generated during the analysis. These individual cell measurements per knockdown condition allow the impact of each siRNA on the accumulation of Cyclin A in the cell’s nucleus and, in the same cell, the activity state of CDK2, monitored by the nuclear to cytoplasmic ratio of the GFP-reporter protein, to be assessed. Finally we document use of a simple R script that supports cell-resolved extraction, display and analysis of responses to modifier siRNAs using the two phenotypic assays.

Although the high-content image analysis and data mining described here is focussed on delivering quantitative information for nuclear Cyclin A and the nuclear to cytoplasmic translocation of a GFP-based CDK2 reporter, the principles used can be adapted to suit other investigator-chosen phenotypic markers that localise to any combination of these subcellular compartments.

2. Materials

2.1 Tissue culture and siRNA transfection

1. 5x siRNA buffer (Dharmacon/GE Lifesciences, catalogue number B-002000-UB-100).

2. siRNA: Human siGENOME SMARTpool siRNAs were used for Cyclin A, CDK4 and CDK6 knockdowns (Dharmacon/GE Lifesciences). AllStars Negative Control siRNA (Qiagen, catalogue number 1027280) was used for all non-targeting RNAi conditions.
3. Tissue culture plates: Standard sterile, 96-well, transparent plates (Nunc/Thermo Scientific, catalogue number 167008) were used for oligo dilution and lipid-mixing steps. Reverse-transfection, tissue culture incubation and cell imaging were performed on Packard Viewplates (PerkinElmer, catalogue number 6005182).
4. HiPerfect siRNA transfection lipid (Qiagen, catalogue number 301704).
5. Tissue culture medium: DMEM, high-glucose with GlutaMAX supplement (ThermoFisher Scientific, catalogue number 10566-016), 10% foetal calf serum.
6. HCT116 human colorectal carcinoma cell line stably expression GFP-tagged CDK2 reporter (5–7)
7. AeraSeal sterile, breathable plate membranes (Alpha Laboratories, catalogue number LW2783).

2.2 Cell fixation, permeabilisation and fluorescent staining

1. Formalin solution, neutral buffered (Sigma-Aldrich, catalogue number HT5012-1CS).
2. Permeabilisation buffer: 0.1 % Triton X-100 (Sigma-Aldrich, catalogue number X100PC) in 50 mM tris(hydroxymethyl)aminomethane (Tris)-buffered saline pH 8.0.
3. Blocking solution: 5% powdered milk in Tris-buffered saline, 0.1% Tween-20 (Sigma-Aldrich, catalogue number P9416).
4. Antibodies/fluorophores: Mouse anti-Cyclin A (Santa Cruz Biotechnology, catalogue number BF683: sc-239), Goat anti-mouse IgG secondary antibody Alexa Fluor 633-conjugate (Invitrogen, catalogue number A-21052), Hoechst 33342 (bisBenzimide H 33342 trihydrochloride) (Sigma-Aldrich, catalogue number B2261).
5. Wash solution: Tris-buffered saline, 0.1 % Tween-20 (Sigma-Aldrich, catalogue number P9416).
6. Phosphate-buffered saline (PBS) (Gibco/ThermoFisher Scientific, catalogue number 18912-014).

2.3 Microscopy and software analysis

1. Fluorescence or confocal microscope equipped with a 20x dry/LWD objective and filters suitable for FITC, DAPI and 633 nm fluorophores.
2. Software: CellProfiler 2.1.1 (www.cellprofiler.org).
3. Software: R version 3.1.2 or newer (www.r-project.org) with additional module ggplot2 installed (9).

3. Methods

3.1 Transfection of siRNA library in 96-well plates in the presence of modifier RNA

All steps to be carried out at room temperature, pipetting steps and handling of cells to take place in a sterile tissue culture hood.

1. Prepare two tubes of 1X siRNA buffer containing 55.5 nM of either non-targeting (NT) control siRNA or siRNA targeting a candidate epistatic modifier (siMOD) (see **Note 1**)
2. Pipette 63 μ l of the 55.5 nM NT and siMOD modifier siRNA to the wells of separate, new 96-well tissue culture plates.
3. Carefully pipette 7 μ l of the 2 μ M siRNA library to both the NT and siMOD siRNA plates, yielding plates on which the library siRNA concentration now is 200 nM and combined with either NT or siMOD siRNA present at 50nM (**Fig. 1A**. see **Note 2**).
4. Dilute the transfection lipid into 40 volumes of serum-free DMEM media and pipette 105 μ l into each well of the two 96-well plates (**Fig. 1B**). For two full 96 well plates dilute 525 μ l of transfection lipid into 21 ml of serum-free DMEM, each 105 μ l of this master mix dispenses 2.56 μ l of lipid per well. Each well of these plates now contains 175 μ l of siRNA-lipid mixture (see **Note 3**).

5. Mix well contents by gently vibrating plates for ten minutes.
6. Dispense 50 μ l of the well contents onto three opaque 96-well tissue culture plates with transparent bottoms (**Fig. 1C**).
7. Add cells (HCT116) stably expressing the CDK2 reporter by pipetting 8000 cells per well in 150 μ l DMEM containing 10 % serum directly onto the plated 50 μ l siRNA-lipid complexes. No further mixing is necessary at this stage. Using the conditions described here the final concentration of the library siRNA will be 20 nM and the NT or siMOD siRNA will be 5 nM per well (see **Note 4**).
8. Seal the two sets of three plates using sterile, adhesive, breathable membranes (See **Note 5**) and place in a humidified incubator at 37 °C, 5 % CO₂ for 48 hours.

3.2 Cell fixation and staining for fluorescence microscopy

1. After 48 hours following the reverse transfection of the siRNA, tip, or preferably aspirate the medium from the wells of each 96 well plate (see **Note 6**).
2. Immediately after removal of the medium, fix the cells by adding 100 μ l of 4 % buffered formaldehyde per well and incubate at room temperature in a fume hood for exactly 10 minutes (see **Notes 7 & 8**).
3. Carefully remove the fixing solution keeping the plates in the fume hood. At this stage the plates can be stored at 4°C, in the dark with 150 μ l Tris-buffered saline per well for up to two weeks (see **Note 9**). Alternatively, proceed immediately to permeabilisation of the fixed cells.
4. Permeabilise the fixed cells by adding 100 μ l of 0.1 % Triton X-100 in Tris-buffered saline per well to the recently fixed plates from which fixative (or, where used, storage buffer) has been removed. Incubate for 10 minutes without agitation. Empty the plates by tipping off the Triton X-100 solution and repeat two more times (see **Notes 7 & 10**).
5. After emptying the plates following the third incubation with permeabilisation buffer, block the plates with 100 μ l of blocking solution per well and incubate at room temperature for 30 minutes.

6. Remove the blocking agent by tipping the plates ensuring complete removal of the blocking solution. Add 50 μ l of mouse anti-Cyclin A antibody diluted 1:1000 in blocking solution and incubate for 2 hours in the dark at room temperature.
7. Wash the plate three times in Tris-buffered saline and 0.1 % Tween-20, leaving the wash solution on the plate for 5 minutes each time.
8. After tipping away the last of the three washes add 50 μ l of Alexa Fluor 633-labelled anti-mouse secondary antibody diluted 1000-fold in blocking solution. Supplement the diluted secondary antibody with 2 μ M Hoechst 33342. Incubate plates overnight in the dark at 4 °C. (see **Note 11**).
9. The following day, tip away the secondary antibody and wash the plates three times wash solution as before, store under 100 μ l Tris-buffered saline in the dark at 4 °C until ready to acquire images.
10. Use a fluorescent or confocal microscope with a 20x objective to take separate, 16-bit grayscale TIF format images of the same field for the three colour channels corresponding to the Hoechst, GFP and the fluorophore used for immunostaining. A number of non-overlapping sets of three-channel images should be acquired per well to yield an approximate total of 300-400 imaged cells per well, therefore covering close to 1000 cells imaged per triplicate set of wells per RNAi knockdown condition (**Fig. 2A**).
11. Name the image files so that each file name is a unique combination of common experiment name, well address from the plate, frame number and channel identifier (**Fig. 2B**. See **Note 12**).

3.4 High-content analysis – software image file settings

1. Overview: The image files acquired and organised as described above are now processed using the open-source CellProfiler 2.1.1 image analysis software. The following steps describe how to set up the software to interpret the image file names and correctly associate related sets of images with wells and channels. The settings described in this section and in 3.5 can be saved as a ‘Project’ file for repeat analyses in CellProfiler and it is recommended to save the file periodically during the following steps to preserve progress (see **Note 13**).

2. Run the software CellProfiler 2.1.1 (www.cellprofiler.org). Close the welcome box to view the 'File list' drag-and-drop window on the right hand side of the program window.
3. Using the mouse, drag a folder containing the all image files relating to one 96 well plate into the drag-and-drop window. A list of the files relating to these images should now be visible in the program (**Fig. 3**).
4. Select 'View output settings' in the lower left corner of the screen. On the next screen use the box labelled 'Default output folder' to specify where the resulting data from the analysis should be sent.
5. Select the 'Metadata' tab in the Input Modules box on the left hand side of the page (**Fig 4**). On the right hand side of the screen select the 'Yes' option to extract metadata and copy/paste the regular expression shown below and in **Fig. 2** into the box labelled 'Regular Expression'.

Regular expression to define image file metadata relating image files to experimental context:

```
^.*(?P<WellRow>\w?[A-H])(?P<WellColumn>\d{1,2})_(?P<FieldNumber>\d{1,2})(?P<Channel>.*).tif
```

6. Select the 'Update' button below to see the metadata extracted from the image filenames populate a table at the bottom of the software window (**Fig. 4**. See **Note 14**).
7. Select the 'NamesAndTypes' tab from the 'Input Modules' list and in the drop-down box 'Assign a name to' select the option 'Images matching rules'.
8. Next select the 'Add another image' button now at the bottom of the screen twice so that details can be entered for a total of three image types (**Fig. 5**. See **Note 15**).
9. For each of the three image types keep the default criteria rules (i.e. match All: File / Does / Contain), then enter either 'Blue', 'Red', or 'Green' into the box to the right of the 'Contain' statement corresponding with the exact spelling of the channel identifier used in the file-naming convention detailed in **Fig. 2**.

10. Into the box labelled 'Name to assign these images' type DNA (for 'Blue' channel), GFP ('Green' channel) and CyclinA ('Red' channel).

11. The software will now be able to interpret the file name convention to associate images with specific fields from defined well locations on the plate and associate them with the correct fluorophore. This can be checked by selecting the 'Update' button in the lower part of the screen to see if the correct image file names associate with the three named image types, in this example 'Cyclin A', 'DNA' and 'GFP' (**Fig. 5**).

12. The 'Groups' settings in the Input Modules list are not used here and the default setting of 'No' is to be selected.

3.5 High-content analysis – image analysis software settings

1. Overview – the steps above provide the software with a framework to relate sets of images to the structure of the siRNA screen. The following settings further develop the Project file to enable analysis of the cells in each image; the DNA stain data is used to identify individual cells via the position of the nuclei (**Fig. 6A**); the resulting digital nuclei mask is used as a seeding point to establish a cell mask based on the extent of GFP staining in the reporter channel image for each individual cell (**Fig. 6B**); finally the nuclei mask is used again to deliver a cytoplasm mask by subtracting the nuclei mask position from each corresponding cell mask (**Fig. 6C**). Fluorophore intensity data from each channel is collected for related sets of nuclei and cytoplasm masks and additional nuclear/cytoplasm GFP intensity ratios for each cell are calculated from these.

2. Using the mouse, select the '+' symbol next to 'Adjust modules:' in the lower left corner of the software to the 'Add modules' window. Use this interface to find and add the following modules in this order: 1) IdentifyPrimaryObjects 2) IdentifySecondaryObjects 3) IdentifyTertiaryObjects 4) MeasureObjectIntensity 5) CalculateMath 6) ExportToSpreadsheet. Finally, select 'Done' to finish adding new modules and return to the main software window (see **Note 16**).

3. The added modules will appear in the ‘Analysis Modules’ list on the left (**Fig. 7**). Select the IdentifyPrimaryObjects module in this list to open the settings for identifying cell nuclei from the ‘Blue’ channel images.

4. The default settings for the IdentifyPrimaryObjects module should now be adjusted to adopt those shown in **Fig. 7**.

Briefly, ‘DNA’ should be selected in the drop-down menu at the top as input to link the module to the ‘Blue’ channel data. The suggested name ‘Nuclei’ in the next box will serve as the reference for the nuclei-mask in subsequent modules used in the analysis. The remaining settings adopted in **Fig. 7** are revealed by choosing ‘Global’ in the ‘Threshold Strategy’ drop-down menu. Below this, the choice of ‘Otsu’, ‘Shape’ and ‘Propagate’ methods in their respective drop-down menus govern how CellProfiler will distinguish the limits of each nucleus from the image background. Additionally, close neighbours are resolved on the basis of shape and the assumption that apparently touching nuclei are divided at the mid-point between their respective centres.

5. Parameters such as the accepted range of nuclei diameters (‘Typical diameter of objects’), ‘Threshold correction factor’ and ‘Lower and upper bounds on threshold’ will need to be adjusted to suit each individual study. CellProfiler will use these settings to accept or reject nuclei ‘objects’ in ‘Blue’ channel images on the basis of several criteria - being within an anticipated size range; being fully in the image; being within a threshold-defined intensity range (see **Note 17**).

6. Optional: At this stage the chosen settings can be tested by selecting the ‘Analyze Images’ button in the lower left corner of the software to see if the resulting masks displayed in the Primary Objects window agree with the appearance of the nuclei in the Blue channel input images (**Fig. 8**). Select with the mouse to toggle to the ‘open eye’ icon next to the IdentifyPrimaryObjects module if the PrimaryObjects window doesn’t open during analysis. Iterative rounds of trying different values here will allow suitable values to be determined.

7. Select the IdentifySecondaryObjects module in the Analysis Modules list to access the settings with which the software will relate the Green channel (GFP-tagged CDK2 reporter) image data with each corresponding nucleus identified in the previous module.

8. Using the settings in **Fig. 9** will enable the IdentifySecondaryObjects module to create a mask of each cell using the Green channel data.

Briefly, the GFP images and the Nuclei identified in the previous module are selected as ‘input images’ and ‘objects’, respectively. Similar to their use in the IdentifyPrimaryObjects module, the ‘Threshold correction factor’ and ‘lower’ and ‘upper bounds’ are key to the success of this module and are empirically derived as before (see **Step 6 & Note 17**). The aim here is to use threshold values as low as possible to enable even very weak cytoplasmic GFP fluorescence to contribute to each individual cell’s ‘Cells’ mask.

9. Select the IdentifyTertiaryObjects within the Analysis Modules list to access the settings with which the software will create a ‘Cytoplasm’ mask by subtracting each Nuclei mask from its corresponding Cell mask.

10. The Cells mask is to be selected in the drop-down menu for ‘larger identified objects’ and the Nuclei mask in the corresponding drop-down menu for ‘smaller identified objects’ (**Fig. 10**). The option to ‘shrink nuclei prior to subtraction’ is also selected here to ensure there will always be a cytoplasmic mask generated for each cell.

11. Select the MeasureObjectIntensity module in the Analysis Modules list. Select the ‘Add another image’ button twice so that the ‘DNA’, ‘GFP’ and ‘CyclinA’ images can be selected using the resulting drop-down menus to be used for measurements (**Fig. 11**).

12. Next select the ‘Add another object’ button at the bottom of the MeasureObjectIntensity settings and then select ‘Nuclei’ and ‘Cytoplasm’ to be the areas from which fluorescence intensity measurements will be taken (**Fig. 11**. See **Notes 18 & 19**).

13. Select the module CalculateMath in the Analysis Modules list to access the settings used to calculate the ratio nuclear/cytoplasm GFP intensity for each cell. Type 'GFP_Ratio' in the box to name the output measurement and use the 'Operation' drop-down box to select 'Divide' (**Fig. 12**).
14. The options that are now open in the CalculateMath module as a result of the selection of the 'Divide' Operation concern the definition of the numerator and denominator for the ratio calculation. Both the numerator and denominator 'measurement type' drop-down boxes should be set to 'Object' and the subsequent drop-down boxes should be set to 'Nuclei' and 'Cytoplasm', respectively.
15. For both the numerator and denominator sections of the CalculateMath module, the 'Category', 'Measurement' and 'Image' drop-down boxes should be set to 'Intensity', 'MeanIntensity' and 'GFP', respectively.
16. Select the final analysis module 'ExportToSpreadsheet' to access the settings that specify what data is saved and written to the default output folder (determined in **3.4, Step 4**).
17. The following settings of the ExportToSpreadsheet module should be changed from their default positions: 'Add a prefix to filenames' is set to 'No', 'Add image metadata columns to your object data file' is set to 'Yes' and 'Select the measurements to export' is also set to 'Yes' (**Fig. 13A**).
18. A button labelled 'Press to select measurements' will now appear, select this to reveal a 'Select measurements' sub-window displaying a tree diagram of available measurements, which can be found beneath the various object classes defined by the analysis (i.e. Cells, Cytoplasm, Experiment, Image and Nuclei).
19. Inside this sub-window select the 'Nuclei' subclass to expand the measurement options and subsequently select to open the 'Intensity' tree of options. Similarly, open 'MeanIntensity' and 'Math/GFP' and tick the boxes corresponding to 'CyclinA', 'GFP' and 'Ratio' (**Fig. 13B**).
20. Select the 'OK' button to close the 'Select measurements' sub-window. Next, select the 'Analyze Images' button in the lower left corner of the main CellProfiler window to start the analysis. The chosen ExportToSpreadsheet settings will now produce a 'Nuclei.csv' data file in the output folder.

21. When the analysis is complete select 'File' and then 'Save Project As...' in the upper left corner of the CellProfiler window to save the analysis settings.

3.6 Example analysis – scatter plots of single cell siRNA screen data

1. The settings for CellProfiler described above will produce several comma-separated value (.csv) files in the user-defined output location. Each file contains data and measurements corresponding to different image-segmentation or file elements determined using these settings. Of these files, Nuclei.csv contains the single-cell measurements 'Intensity_MeanIntensity_CyclinA', corresponding to the average nuclear intensity of Cyclin A antibody staining for each cell, and 'Math_GFP_Ratio' corresponding to the ratio of nuclear/cytoplasmic mean GFP intensity per cell, respectively (**Fig. 14**. See **Note 20**).

2. Each row of data in Nuclei.csv corresponds to observations of a single cell. Each row of data also carries information indicating that cell's location on the 96-well plate ('Metadata_Well') and which image field from that well it corresponds to ('Metadata_FieldNumber').

3. The organisation of data in Nuclei.csv can thus be used to observe the behaviour of each cell in the Cyclin A expression assay and GFP-CDK2 activity assay for each siRNA knockdown condition tested (i.e. well position in the screen). A direct way to visualise this is to plot an x/y scatter plot of each cell in a given well against its corresponding scores in both assays. The R script in **Fig. 15** automates the generation of these scatter plots on a well-by-well basis.

4. Install and run R on the computer to perform the analysis. After the R terminal opens copy and paste the R script at the command prompt.

5. Amend line 6 in the script (**Fig. 15**) to be give the location of the Nuclei.csv file.

6. Between the quotation marks in line 11, type the well address of the cells to plot (e.g. "B2").

7. On line 12 type a title for the scatter plot. For example this could describe the siRNA condition or the position of the relevant well.

8. Press the Enter key to run the amended R script and generate a plot relating the assay scores for each cell in the chosen well.

Examples are shown in **Fig. 16** demonstrating how cells subjected to knockdown of a candidate modifier target respond differently in relation to the two assays.

4. Notes

1. The protocol assumes the screen will make use of parallel NT and siMOD siRNA transfections of the RNAi library. To remove the epistasis component of the siRNA transfection steps, do not supplement the siRNA buffer with any extra siRNA in **Step 1** and dilute the library plate to yield 200nM. Then follow **Steps 2-7**, producing three replicate transfection plates.

2. The subsequent steps in the protocol will further dilute these siRNA such that the final concentrations the cells will experience during reverse transfection will be 20 nM (library siRNA) and 5 nM (modifier siRNA). The 20nM oligonucleotide concentration is intended to produce a robust level of knockdown across the various targets represented in the library, whereas the 5 nM used for the siMOD target assumes prior characterisation/optimisation has identified this as sufficient to achieve knockdown of this target. Additionally, we have determined a net concentration of 25 nM oligonucleotide is not limiting for the amount of lipid described in this method. Other siRNA concentrations for library or modifiers are possible and for studies with different cells and knockdown targets/conditions further optimisation here is recommended to achieve optimum target knockdown.

3. The choice of transfection lipid and the amount to be used will need to be optimised if different cells are to be studied. Ideally, an amount of lipid that can simultaneously deliver both the library and modifier siRNA without approaching saturation is desired. This will require test transfections with titrations of lipid against fixed concentrations of positive control siRNA, and vice versa, to identify optimal knockdown conditions while avoiding toxicity from excessive reagent concentrations.

4. The number of cells seeded at this point in the protocol has been optimised for the HCT116 cell line. The number of cells must be high enough to yield a practical number of cells for images acquired at the chosen magnification, however the cells must not be allowed to grow to confluency as contact inhibition will likely interfere with the G1 transit assays. Taking these points into consideration, the seeding density will need to be adjusted screens of differing duration and cell lines with differing growth rates. Additionally, we have observed that HCT116 cells tend to clump following trypsinisation, which will impact the accuracy of some counting methods and may prematurely result in contact inhibition of newly seeded experiments. It is recommended that cells likely to clump be passed through a sterile, blunt needle using a syringe to produce a finer mono-cellular aspirate after trypsinisation.

5. Use of the breathable adhesive membranes to seal the newly transfected tissue culture plates is critical to limit 'edge-effect' whereby evaporation from the edges of the plate during the incubation period places cells in these wells at a disadvantage. Use membrane seals in addition to the normal lids supplied with the tissue culture plates.

6. Aspiration of the transfection media from the plates is preferable to tipping away the media. Depending on the cell line used for a study, tipping and shaking a plate may shed cells from the plate creating bias or increasing the variability of plate replicates. Ideally a vacuum-trap will be used with an 8- or 12-well adapter to allow complete rows or columns to be aspirated in a single action.

7. We make the assumption this method is being carried out manually. The fixation and subsequent permeabilisation steps in this method require 10-minute incubations with the respective solutions or buffers. It is important to choose a fixed pattern of processing, such that timelines of incubation are kept for all plates. Thus remove medium from and add fixative to the first plate, then sequentially process subsequent plates using in fixed time intervals (i.e one minute). Return to the first plate to remove fixative and add permeabilisation or storage solution after exactly 10 minutes, and process subsequent plates in turn and after they have had 10 minute exposure to fixative.

8. The use of 4% formaldehyde after aspiration of the media from the plates is a deliberately chosen cell fixation strategy intended to preserve the fluorescence of the GFP-tagged reporter expressed by the HCT116 cells described in this method. Use of higher concentrations of formaldehyde will begin to irreversibly destroy the fluorescent signal. This is an important consideration as alternative formaldehyde fixation pipetting strategies exist, such as adding higher stock concentrations of formaldehyde directly to the medium, thus obviating the need for media aspiration. However, in our experience using 96well formats, mixing formaldehyde with the media to arrive at the target concentration of 4% will subject the cells to a transiently inaccurate concentration with unsatisfactory preservation of the GFP fluorescence. If a study were to not use stably expressed fluorescent proteins, instead making use of post-fixation stains such as antibodies and dyes, the choices for how to fix the cells with formaldehyde become more flexible.

9. If storing the plates in this way it is recommended to use adhesive plate seals or a sealable plastic bag to prevent evaporation from the plates.

10. The three permeabilisation washes can be reduced to one for some immune-staining epitopes. If adapting this method to probe with different antibodies this may be a useful optimisation opportunity to explore. In our experience, a 2nd and 3rd round of permeabilisation has improved the reliability of antibody detection for some nuclear epitopes.

11. Alexa Fluor 633 (AF633) was chosen as the fluorophore for the immunostaining of Cyclin A as it has sufficiently different and non-overlapping excitation/emission spectra to Hoechst 33342 and GFP such that these reagents are unlikely to interfere. Alternative fluorophores that could equally well substitute for AF633 include AF647 and AF546, however, we do not recommend using fluorophores with lower excitation wavelengths than AF546.

12. The file naming structure used in **Fig. 2B** is a suggestion and can be modified to suit studies with different image data arrangements. For example, the renaming of a channel or addition/subtraction from the total number of channels used. However, care will be needed to make sure that the CellProfiler 'Input modules', Metadata and NamesAndTypes reflect the changes applied. This may

also require additional modification of the regular expression shown in **Fig. 2B** if the pattern relating the elements of the file naming convention change substantially. Refer to the Help '?' button to the right of the regular expression box in the CellProfiler 'Metadata' module for a guide how to modify the regular expression to suit a new file naming convention.

13. The CellProfiler software is supported by an extensive user manual and a user forum available at www.cellprofiler.org. These are a great resource to enhance the understanding of the user of the broad scope of capabilities of this software beyond those used here.

14. Selecting the 'Update' button should fill a table at the bottom of the software window with details extracted from the image file names arranged in columns under the appropriate headings. If this doesn't happen, consider selecting the magnifying glass icon to the right. This will allow you to check the compatibility of the regular expression with the selected files' names. The program will colour-code the successfully recognised parts of the file names, thus identifying parts of the regular expression that require attention.

15. The 'NamesAndTypes' module is the point in the software where the number of expected channels is coordinated with the metadata settings and with the filenames given in the earlier modules. So if a study uses more or less fluorophores than the three used in this method, this is where the information entered should be adjusted to reflect the changing number of channels and the names associated with them in later analysis modules.

16. Under the different categories within the 'Add Modules' window there are a large number of modules enabling additional analytic features to be added beyond the scope of this protocol. The reader is referred to the CellProfiler manual to explore these fully. Some modules will require others to have already be implemented. Hence the recommended order of the modules used in the current method is to satisfy the dependencies some modules have for functions of earlier analysis modules to have already been carried out (e.g. IdentifySecondaryObjects requires an upstream IdentifyPrimaryObjects module in the analysis to have completed prerequisite tasks before it can function correctly).

17. We find that once a useful set of values are found for the ‘Lower and upper bounds on threshold’ most of the routine fine-tuning between runs with different image sets can be achieved by adjusting the ‘Threshold correction factor’. Generally, the upper bounds value can stay set to 1.0, however, the lower bounds value should be adjusted if background staining is misdirecting the segmentation apparent when an analysis is run. A good value for the lower bounds setting is a number above the background intensity for the Blue channel images. Typical background intensity values can be observed in the primary objects segmentation window that is produced during analysis (Fig. 8). Holding the mouse pointer over areas outside of the nuclei in the original image will give a pixel intensity value corresponding to those positions in the image in the lower left corner of the window. Suitable values for the Threshold correction factor can be found by trial and error. Larger values used in the IdentifyPrimaryObjects module will shrink the objects produced in the nuclei mask and smaller values will allow them to expand. Using this principle, it should be possible to find a good consensus value for the segmentation of all the Blue channel images for a data set. Too low a value will result in frequent merging of close neighbours and too high a value will begin to miss nuclei with lower intensities of staining. Start with a value of 1 and explore Threshold correction factor values above and below this to determine where the best setting can be found.

18. Although not used in this analysis, a by-product of the settings used in the MeasureObjectIntensity module will be the intensity values of DNA staining measured at the Nuclei and Cytoplasm regions in each image. In this case, only the Nuclei measurements of staining intensity will be useful and the option to collect the integrated DNA intensity for each nucleus can serve as a useful parameter in identifying likely cell cycle phases of the corresponding cell at the time of fixing. This parameter has to be selected later in the ExportToSpreadsheet module (**Fig. 13B**).

19. A useful module similar to MeasureObjectIntensity that may also fit into the analysis is MeasureObjectSizeShape. This can be used to additional geometric parameters are useful to characterise the cells in a study (e.g. size of Nuclei objects or Cell objects).

20. The parameters used in the current analysis are all to be found in Nuclei.csv. Depending on the analysis design and parameters chosen for export in the ExportToSpreadsheet module, some values of interest may only be found in the other output files (e.g. Cells.csv or Cytoplasm.csv will be the files containing parameters such as intensity measurements relating to the ‘Cells’ and ‘Cytoplasm’ masks, respectively). If this is the case, the relevant column corresponding to the measurement of interest can be cut/pasted into the ‘main’ csv file containing the majority of useful data to combine them into one comprehensive file for further analysis. The row-order which the cells adopt will be identical across each of the unmodified csv files produced by the same analysis (e.g. Cells.csv, Cytoplasm.csv and Nuclei.csv).

5. References

1. Krausz E (2007) High-content siRNA screening. *Mol Biosyst* 3:232–40. doi: 10.1039/b616187c
2. Lyman SK, Crawley SC, Gong R, et al. (2011) High-content, high-throughput analysis of cell cycle perturbations induced by the HSP90 inhibitor XL888. *PLoS One* 6:e17692. doi: 10.1371/journal.pone.0017692
3. Santos SDM, Verveer PJ, Bastiaens PIH (2007) Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate. *Nat Cell Biol* 9:324–330. doi: 10.1038/ncb1543
4. Garcia MA, Alvarez MS, Sailem H, et al. (2012) Differential RNAi screening provides insights into the rewiring of signalling networks during oxidative stress. *Mol Biosyst* 8:2605. doi: 10.1039/c2mb25092f
5. Gu J, Xia X, Yan P, et al. (2004) Cell Cycle-dependent Regulation of a Human DNA Helicase That Localizes in DNA Damage Foci. *Mol Biol Cell* 15:3320–3332. doi: 10.1091/mbc.E04
6. Richardson E, Stockwell SR, Li H, et al. (2012) Mechanism-based screen establishes signalling framework for DNA damage-associated G1 checkpoint response. *PLoS One* 7:e31627. doi: 10.1371/journal.pone.0031627
7. Stockwell SR, Mitnacht S (2014) Workflow for High-content , Individual Cell Quantification of Fluorescent Markers from Universal Microscope Data , Supported by Open Source Software 1 . Experimental Perturbation and Cell Labeling for Response Markers (Reverse Transfection siRNA Screen). 1–14. doi: 10.3791/51882
8. Carpenter AE, Jones TR, Lamprecht MR, et al. (2006) CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* 7:R100. doi: 10.1186/gb-2006-7-10-r100
9. Wickham H (2009) ggplot2: elegant graphics for data analysis. Springer New York

Figure Caption

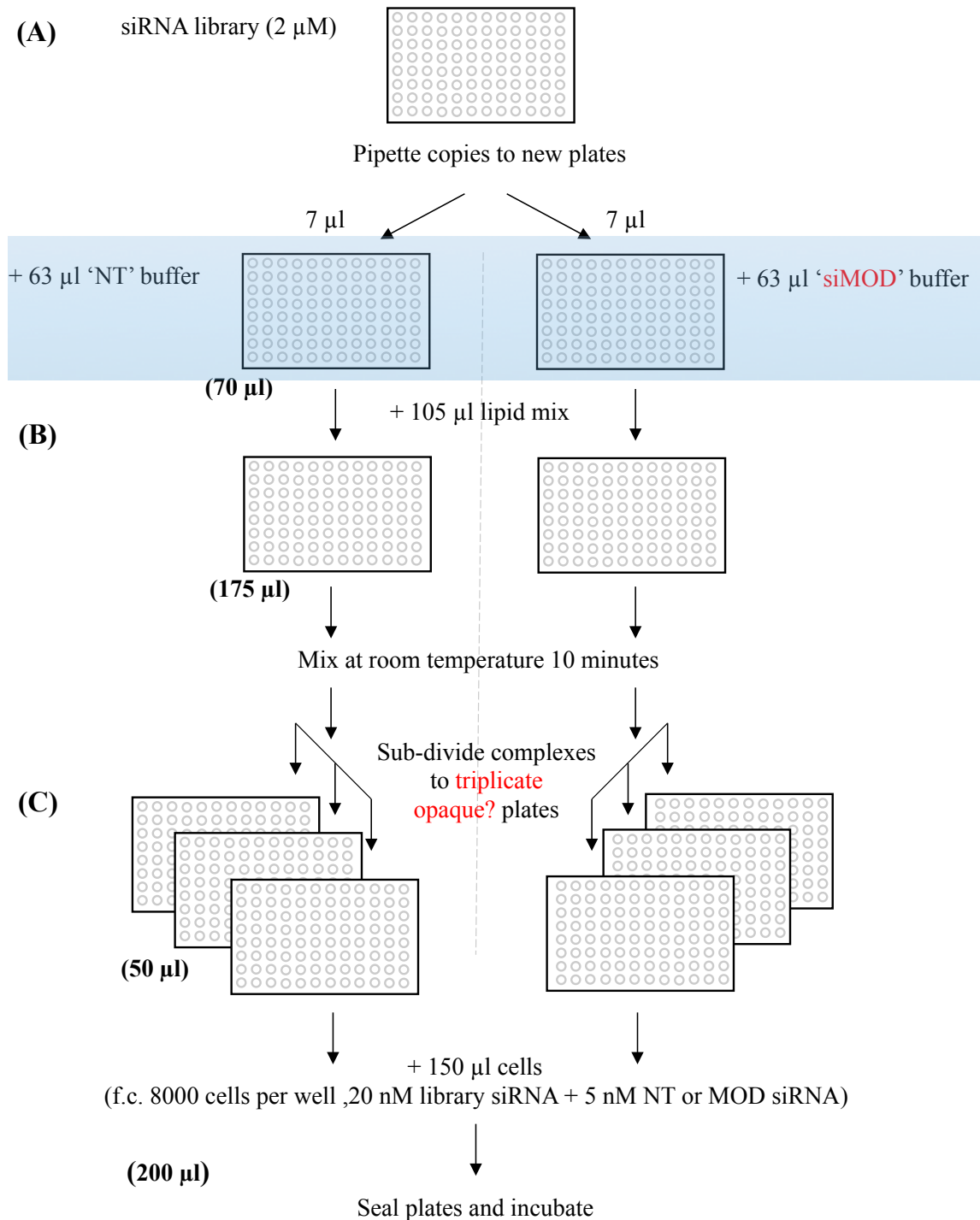


Fig.1. Pipetting scheme for reverse transfection of adherent tissue culture cells with arrayed siRNA oligonucleotides. Steps either side of the dashed line represent preparation of parallel copies of the

library siRNA transfections combined with either non-targeting (NT) siRNA or a specific additional modifier siRNA (siMOD). This siRNA screen can either look for epistatic modifiers of the siMOD target gene or, vice versa, to test a candidate siMOD target for epistatic impact on the various targets in the library. Values in bold and in parentheses indicate the current volume per well of liquid in the plates upon completion of the indicated steps. (A) A plate of an siRNA library arrayed in a 96 well format is diluted in siRNA buffer to yield a 200 nM copy on a new plate containing 70 μ l final volume. The siRNA dilution buffer is supplemented with either 55.5 nM NT or siMOD siRNA, or non-supplemented if the screen does not involve a modifier (see **Note 1**). (B) Lipid-siRNA complexes are generated by adding diluted lipid mix to the plate(s). (C) The resulting complexes are then subdivided into three copies on new plates, each with opaque walls and a transparent base, to which the cells are subsequently added. Final oligonucleotide concentrations once the cells have been added are 20 nM for the library siRNA and 5 nM for the modifier siRNA.

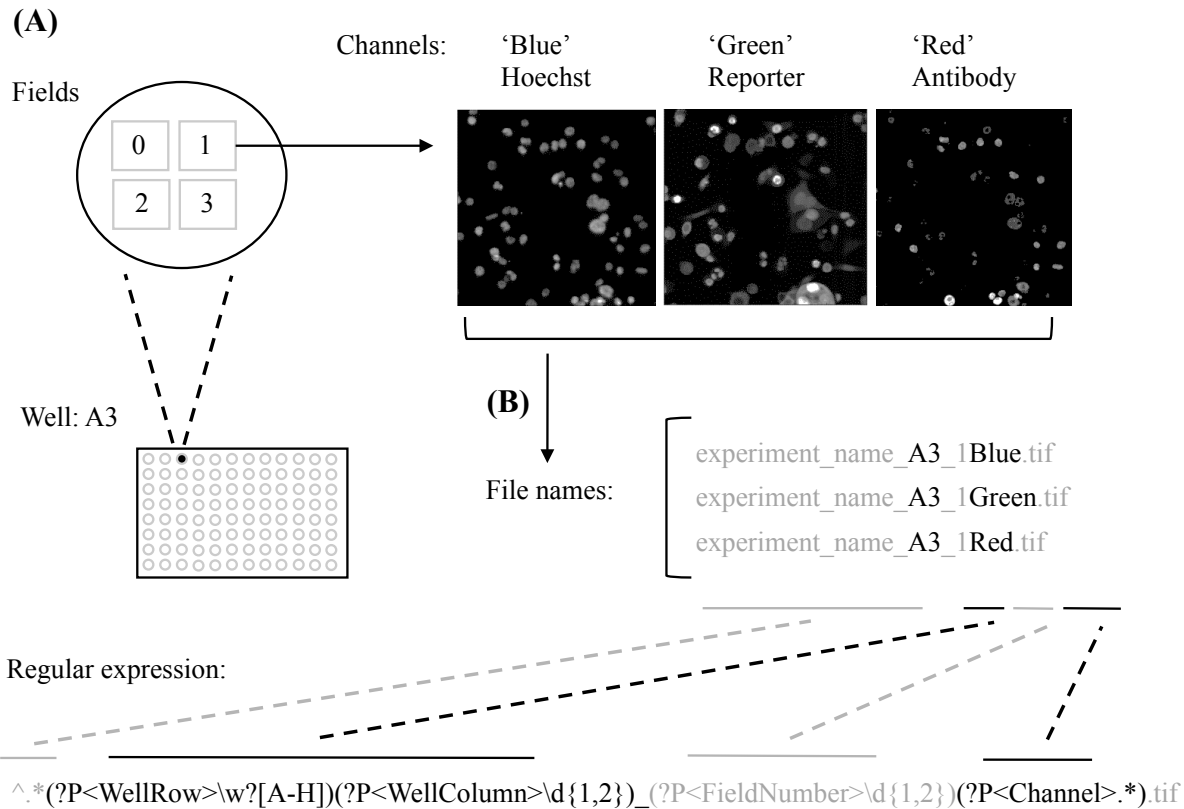


Fig. 2. Organisation of acquired image files for subsequent analysis. (A) Images are acquired for three wavelengths corresponding to emission of the fluorescent DNA dye, the emission of a GFP-tagged CDK2 reporter and emission of the fluorophore used for the detection of bound Cyclin A antibody. Sets of these three channel images are taken for a number of non-overlapping fields per well to collectively image a target number of cells (e.g. 800-1000). (B) Image files from a given screening plate should adopt a naming convention whereby a common experiment name followed by well address, iterative field number and a channel identifier. The example used here gives each file a unique name using this structure with which related sets of channel images can later be traced back to common fields and cells within these for any specific well. Below the example file names is the required 'regular expression' code that CellProfiler can use to interpret this specific example of file name structure.

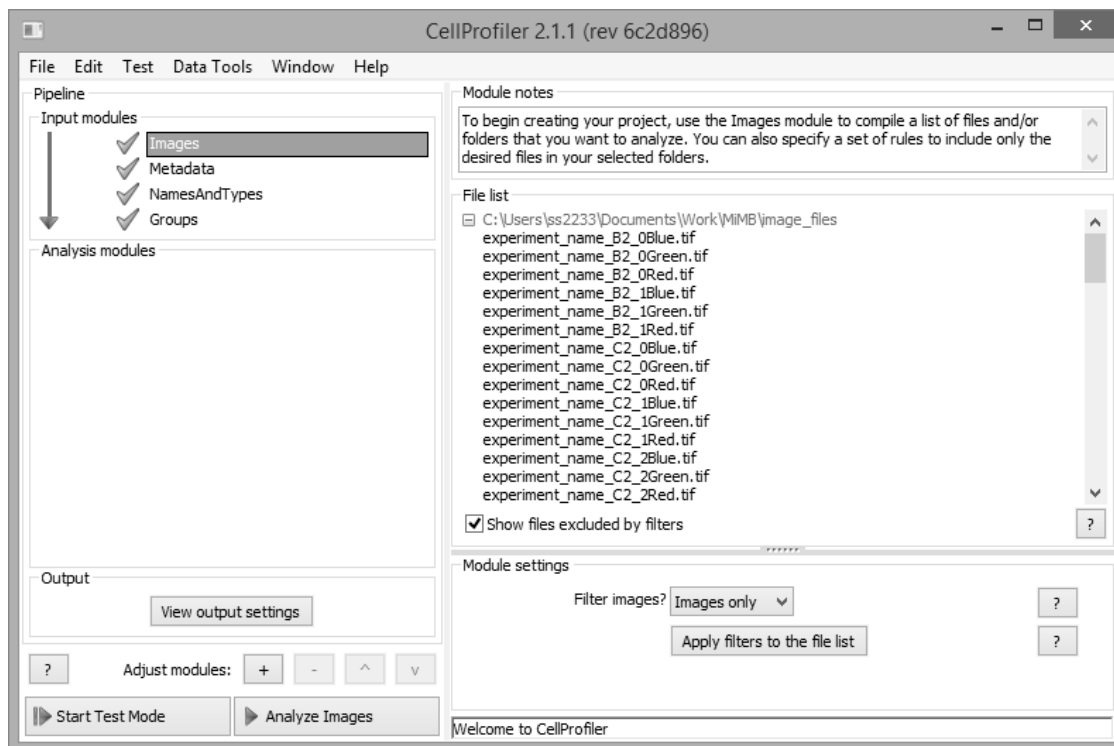


Fig. 3. Screen capture of the CellProfiler main window early in the process of setting up an analysis. Image files have been dragged using the mouse into the large pane on the right where the file names can be viewed. The destination of analysis output files can be set using the 'View output settings' button on the lower left of the screen. At this stage no analysis modules have yet been added in the large pane on the left. If settings from any previous analysis have been saved as a project file this can be loaded now to populate the remaining settings to perform an analysis. This protocol assumes the Project file still needs to be created in the subsequent steps, however, selecting File/Open Project... at the top of the screen allows loading of a previous Project file and its settings.

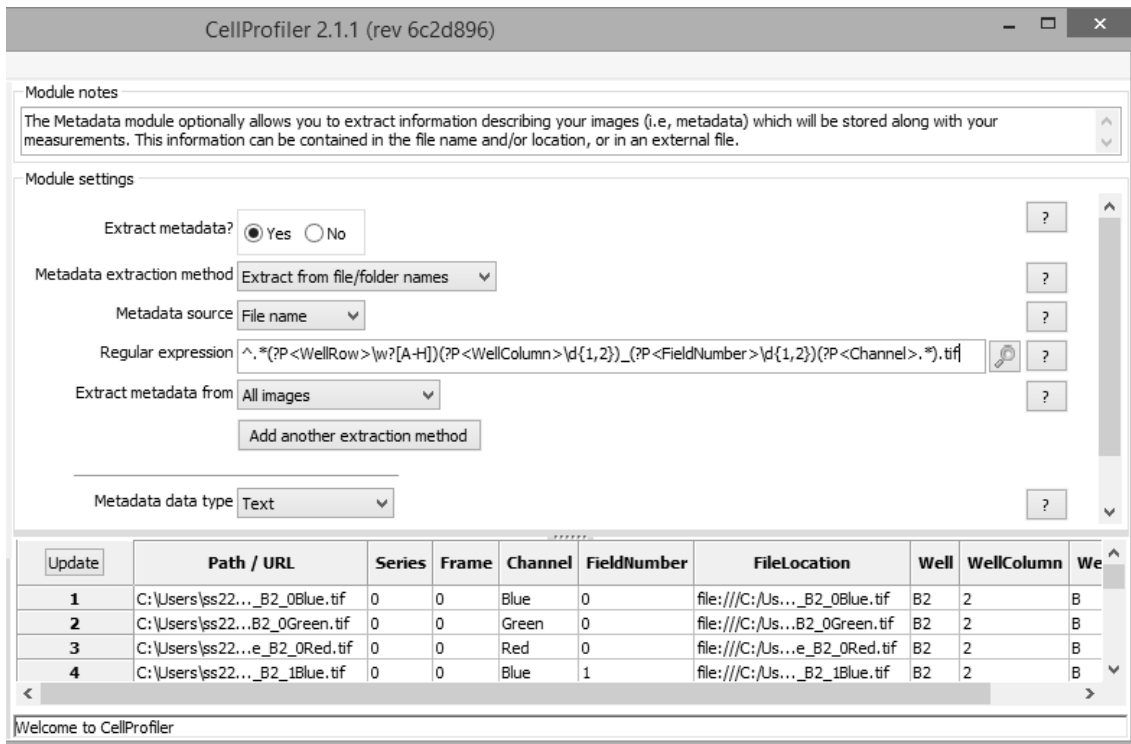


Fig. 4. Screen capture of the Metadata module in CellProfiler. Following the operations described in **Figure 3**, open this window by selecting the 'Metadata' tab in the list of 'Input modules' in CellProfiler. 'Metadata' stipulates conditions for extraction of data from the image file names. Shown are the settings to make use of the 'regular expression' from **Fig. 2** allowing the CellProfiler software to relate each image to the part it performs in the structure of the whole experiment. Selecting the 'Update' button in the lower left of the screen displays the separated metadata elements from each filename and serves as a visual check that the settings are correct.

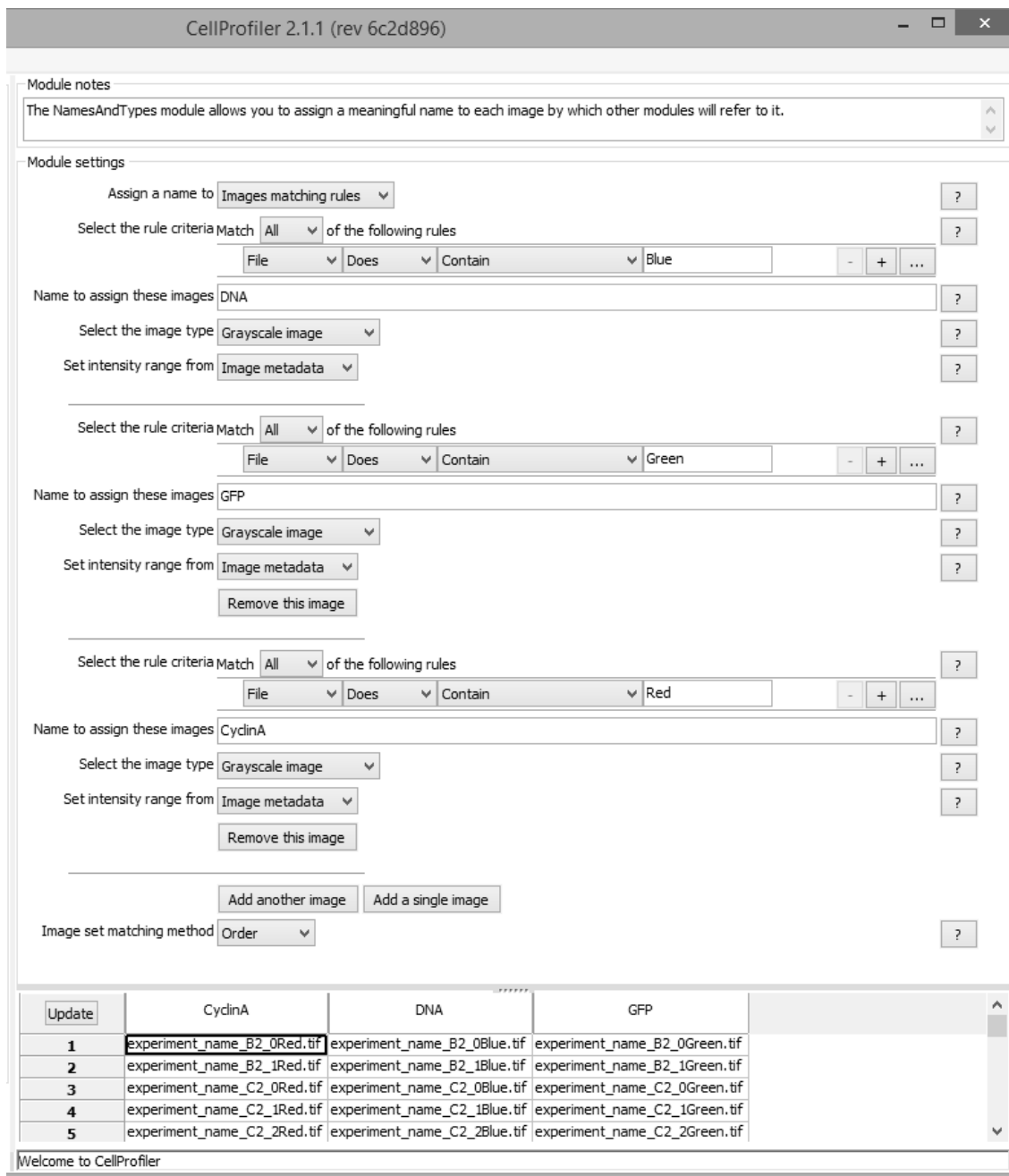


Fig. 5. Screen capture after selecting the 'NamesAndTypes' tab in the list of 'Input modules' in CellProfiler. Here rules are entered such that biologically meaningful names are assigned to the image files bearing the corresponding channel identifiers. Three rule sets are shown here, one for each channel used. At the bottom of the screen use of the 'Update' button has been selected to test the settings with the current image files similar to the check applied in Fig. 4.

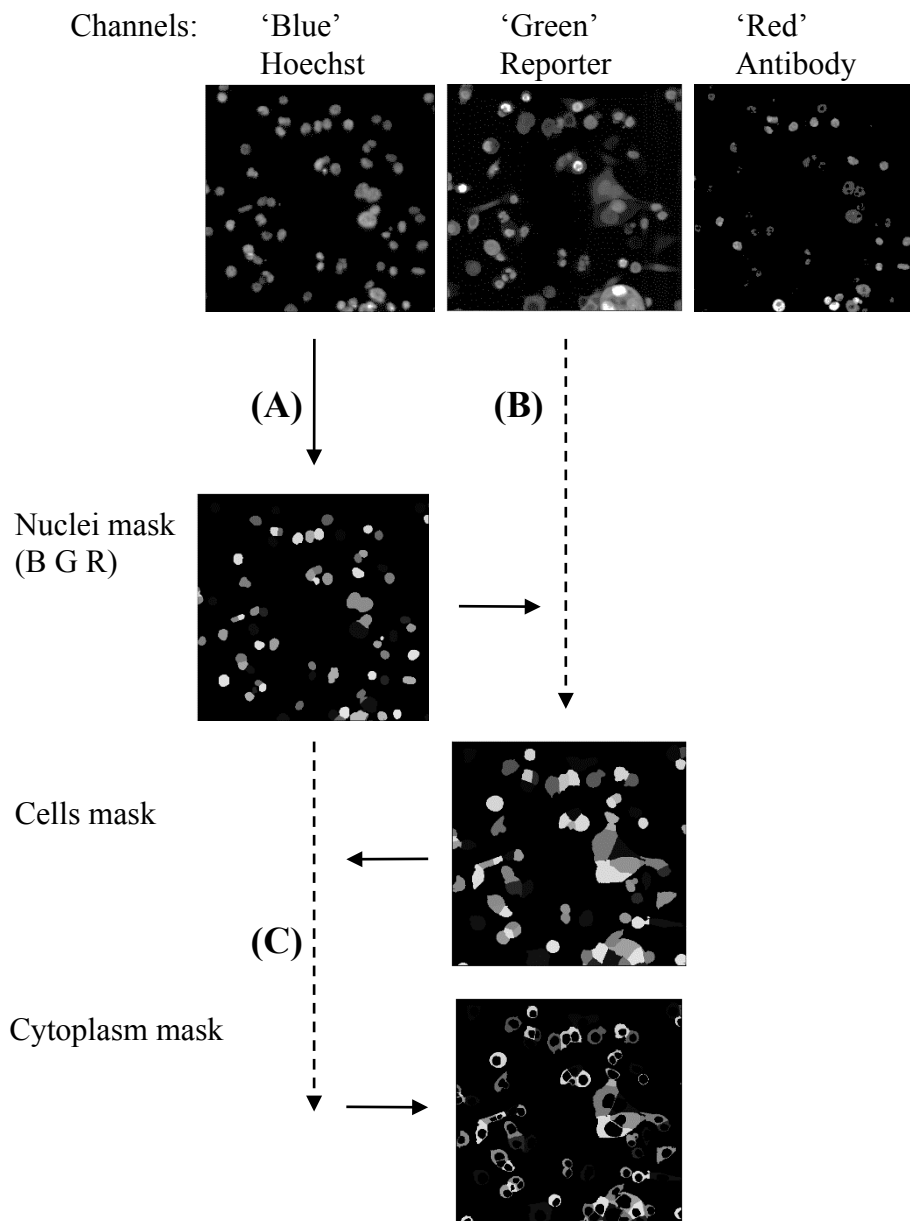


Fig. 6. Image segmentation strategy. The settings described here will guide the CellProfiler software to use the fluorescence intensity information in corresponding 'Blue', 'Green' and 'Red' channel images to identify individual cells and fit to these data-derived measurement areas ('masks') from which subsequent measurements will be taken. (A) The nuclei visualised in the 'Blue' channel representing Hoechst DNA dye fluorescence define a 'Nuclei mask' which locates cells individually and separates closely spaced cells. (B) The Nuclei mask from the 'Blue' channel image and the GFP intensity values from the 'Green' channel image are combined to create a 'Cells mask'. The Nuclei mask is used as the starting points to find zones of GFP positivity that propagate outwards from each

nucleus. User-defined thresholds in CellProfiler will allow this to proceed until either the limits of the cell are reached (i.e. background levels of low GFP are reached) or when the midpoint between two touching cells is detected. (C) The Nuclei mask pattern for each field is subtracted from its corresponding Cells mask to produce a donut-like ‘Cytoplasm mask’ for each cell. Each mask can now be used in the analysis as an area for fluorescence intensity measurements from any channel. The ‘Red’ channel, in this experiment representing immunodetected Cyclin A, does not contribute to the segmentation that produces the masks in this example analysis.

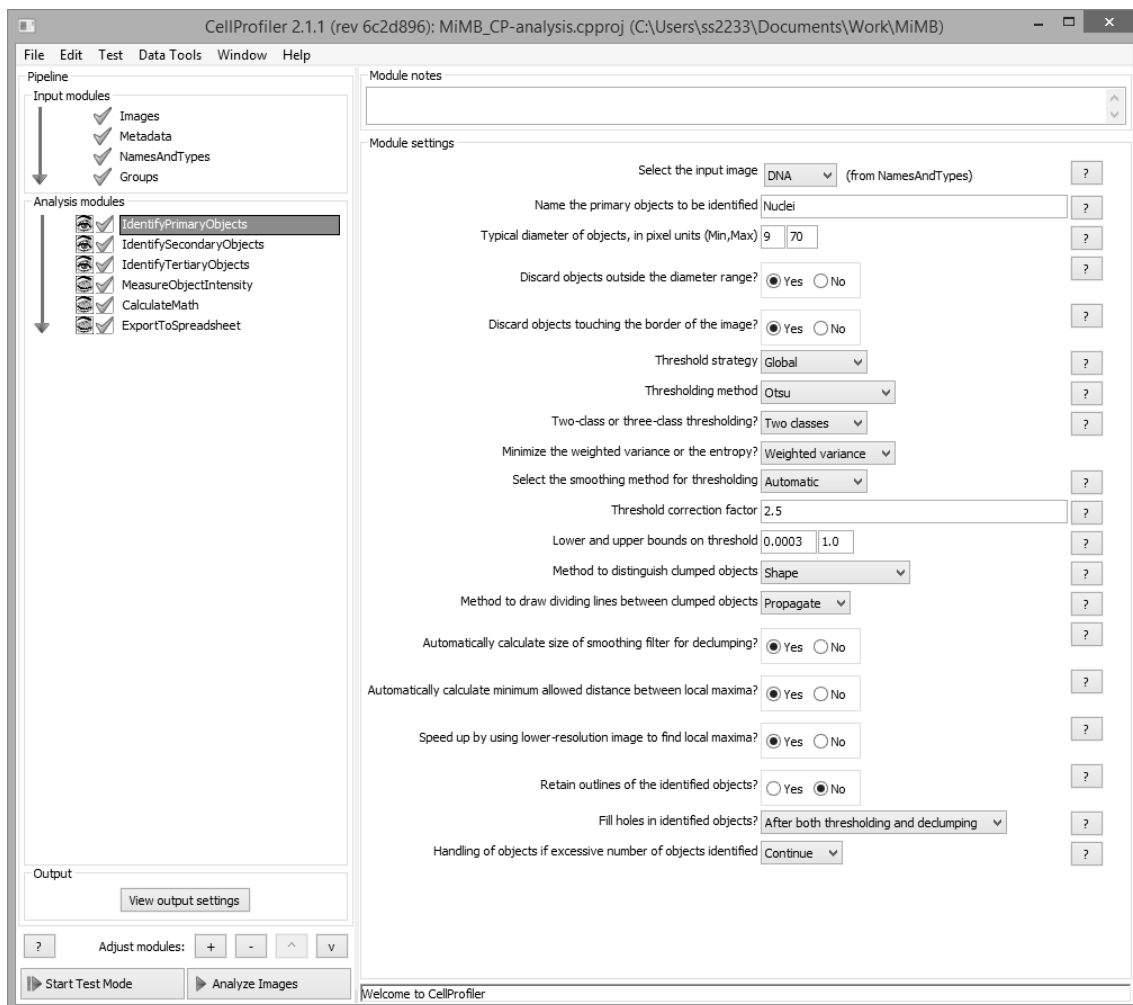


Fig. 7. Screen capture after selecting the ‘IdentifyPrimaryObjects’ analysis module in CellProfiler.

The ‘IdentifyPrimaryObjects’ module is used to produce the Nuclei mask from the Hoechst DNA dye fluorescence data from each field. The settings applied here use the ‘Blue’ channel images named earlier in the analysis as ‘DNA’ (Fig. 5). Limits on acceptable sizes of nuclei are defined here,

followed by the intensity threshold method, the thresholds themselves, rules to separate close nuclei and the exclusion of nuclei not entirely within the image. See **Note 17** for details on adjusting the thresholding settings.

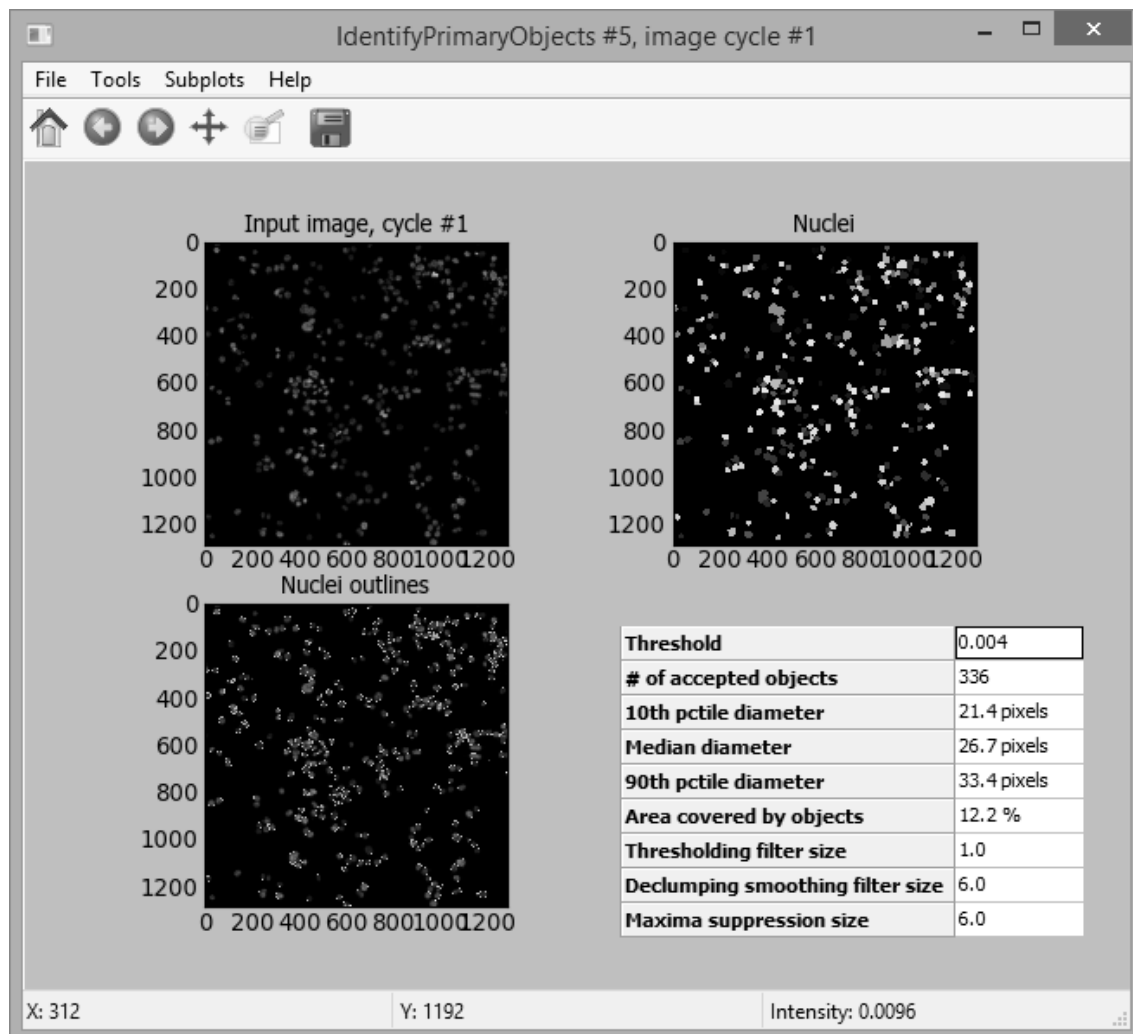


Fig. 8. Screen capture of the PrimaryObjects window displaying images from a test analysis by the selected IdentifyPrimaryObjects settings. To determine the suitability of the settings in the 'IdentifyPrimaryObjects' module in CellProfiler the resulting object identification should be tested visually by observing this window for a representative number of fields during a test analysis. Select the 'Eye' icon to the left of the Analysis modules to 'open' prior to selecting the 'Analyze Images' button at the bottom of the page. The fit of a mask to each source image will be displayed in turn. This allows some trial-and-error amendment of threshold settings under the respective analysis module to

adjust object selection to the image properties, which are likely to vary between different studies and depend on the cell type used and conditions tested

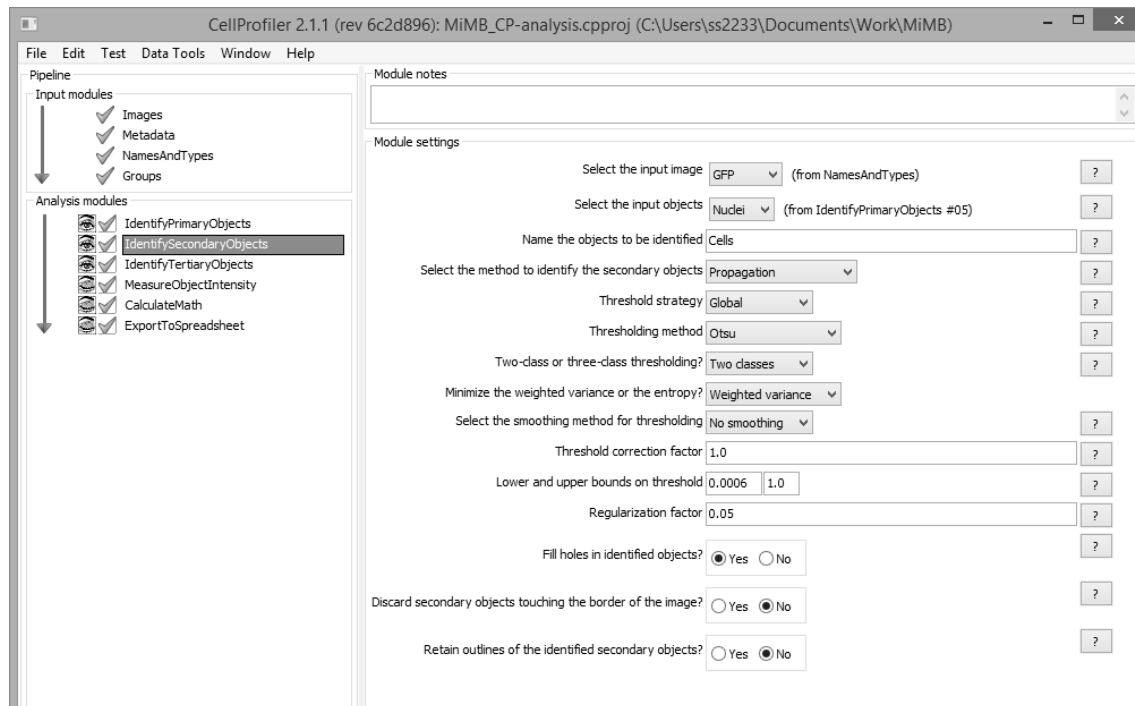


Fig. 9. Screen capture after selecting the 'IdentifySecondaryObjects' analysis module in CellProfiler. The module settings shown here create a 'Cells' mask from the 'GFP' images (this name being applied to the 'Green' channel data earlier in the NamesAndTypes module). Importantly, the 'Nuclei mask' is selected as the 'input objects' to identify the positions of the cells and their nuclear compartment in the GFP images. It is from these nuclear compartment positions that the extent of the 'Cells mask' will be determined using a propagation method. Intensity threshold values are used here to specify background versus GFP fluorescence, similar to those used in the IdentifyPrimaryObjects module. It is advisable to set the threshold correction factor as low as possible to still detect GFP while excluding the space in each image between GFP expressing cells so that even very weak cytoplasmic GFP positivity is detectable. This way the 'Cells' mask approximates as close to a whole cell footprint as possible, even for cells where the majority of the GFP may be nuclear. As for the IdentifyPrimaryObjects module, apply a test analysis by selecting the 'Eye' icon to the left of the IdentifySecondaryObjects module tab to 'open', then select the 'Analyze Images' button at the bottom

of the page to ensure the threshold settings lead to appropriate image dissection in the corresponding SecondaryObjects window. Adjust parameters if necessary as noted under **Fig. 8** and **Note 17**.

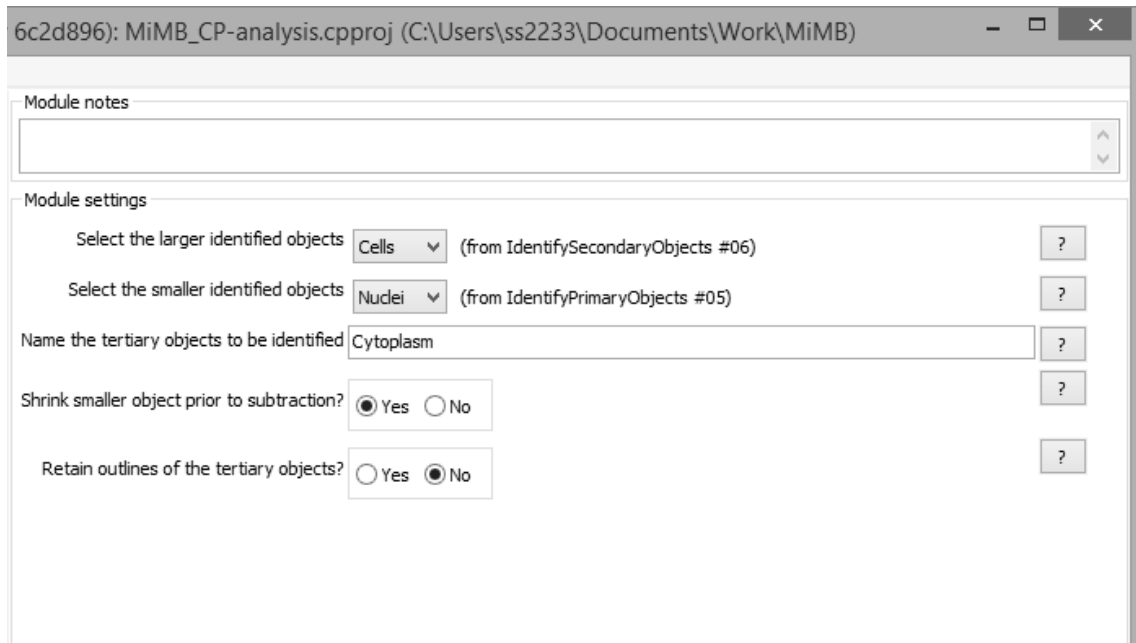


Fig. 10. Screen capture after selecting the ‘IdentifyTertiaryObjects’ analysis module. The IdentifyTertiaryObjects module settings shown will create a ‘Cytoplasm mask’ by subtracting ‘Nuclei mask’ outlines from the corresponding ‘Cells mask’ outlines. As before in the previous ‘Identify...’ modules, the corresponding ‘eye’ icon can be set to open to display the newly created mask in relation to the original images upon selecting the ‘Analyze Images’ button. The accuracy of this mask is entirely dependent on the settings chosen within the IdentifyPrimaryObjects and IdentifySecondaryObjects modules earlier in the analysis. The ‘Cytoplasm’ and ‘Nuclei’ masks will be used in subsequent modules for intensity measurements and nuclear/cytoplasm ratio calculations.

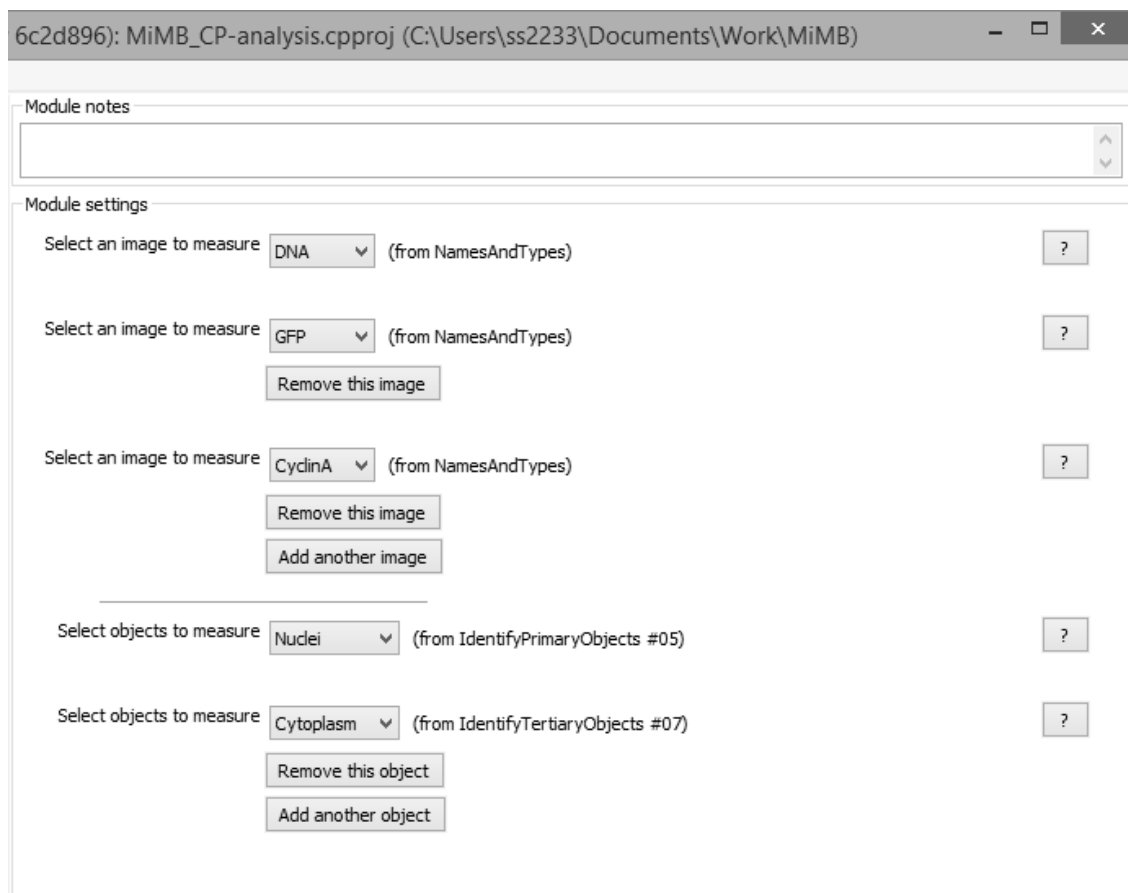


Fig. 11. Screen capture after selecting the MeasureObjectIntensity analysis module. This module is used to specify which channel images are to be used for fluorophore intensity measurements. In the lower part of the module, the particular mask-objects defined earlier in the analysis are selected as the areas within each image from which the intensity measurements are to be taken. The settings shown here will generate intensity measurements in all three channels (DNA, GFP and CyclinA) and be taken for each cell's corresponding 'Nuclei mask' and 'Cytoplasm mask'. Only Nuclei mask measurements will be used for the Cyclin A staining, whereas GFP intensity measurements from both Nuclei and Cytoplasm masks per cell will be used in the next step of the analysis to calculate nuclear/cytoplasm ratios for the CDK2 reporter.

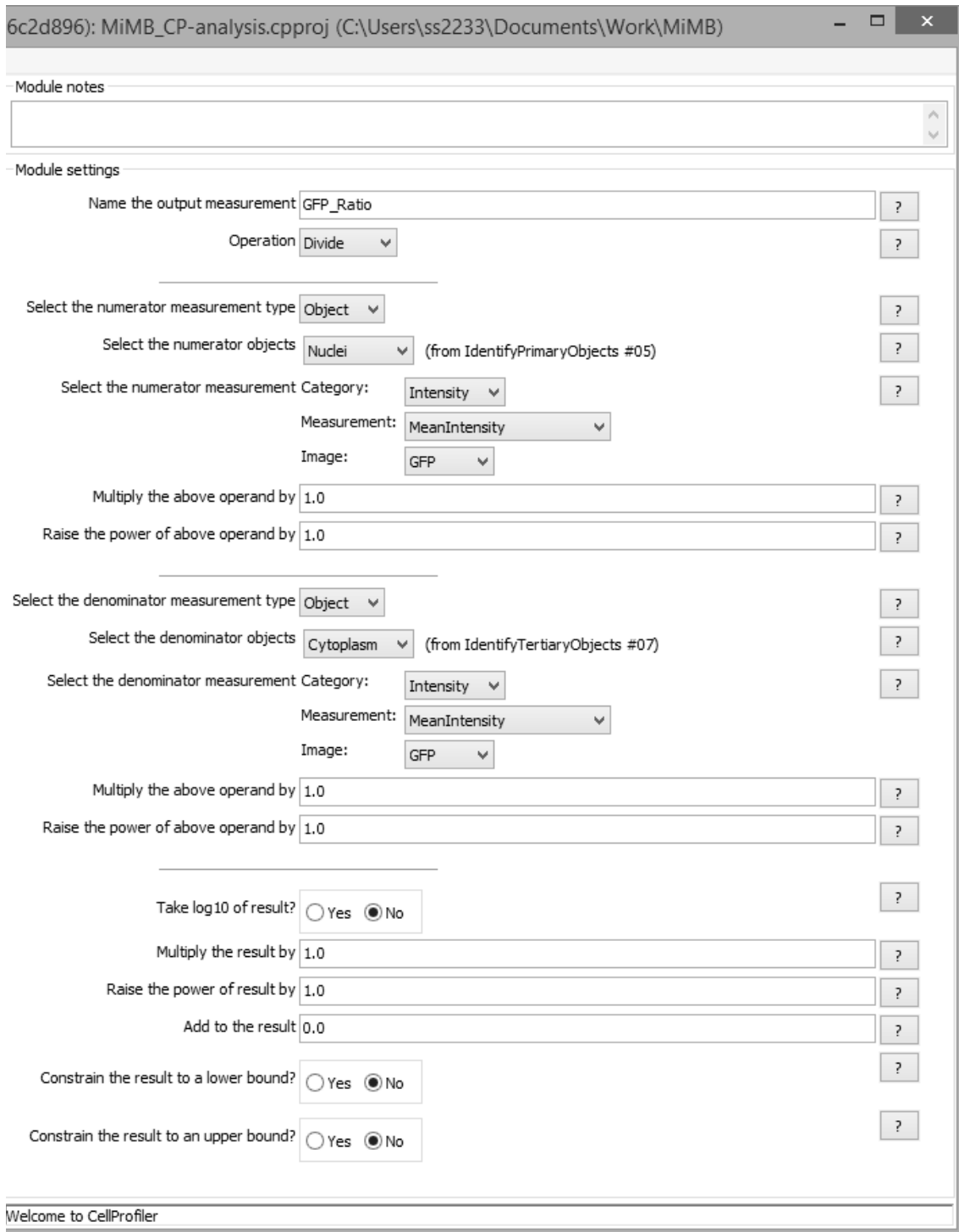


Fig. 12. Screen capture after selecting the CalculateMath analysis module. This module can be used to determine mathematical relationships using specified measurements. For this analysis CalculateMath is used to produce a nuclear/cytoplasmic ‘GFP_Ratio’ parameter for each cell using a combination of specific measurement, image and mask values generated in the earlier modules of the analysis. The

'Divide' operation is used in this analysis with a combination of 'MeanIntensity' values taken from the GFP images of the 'Nuclei' and 'Cytoplasm' mask areas of each cell.

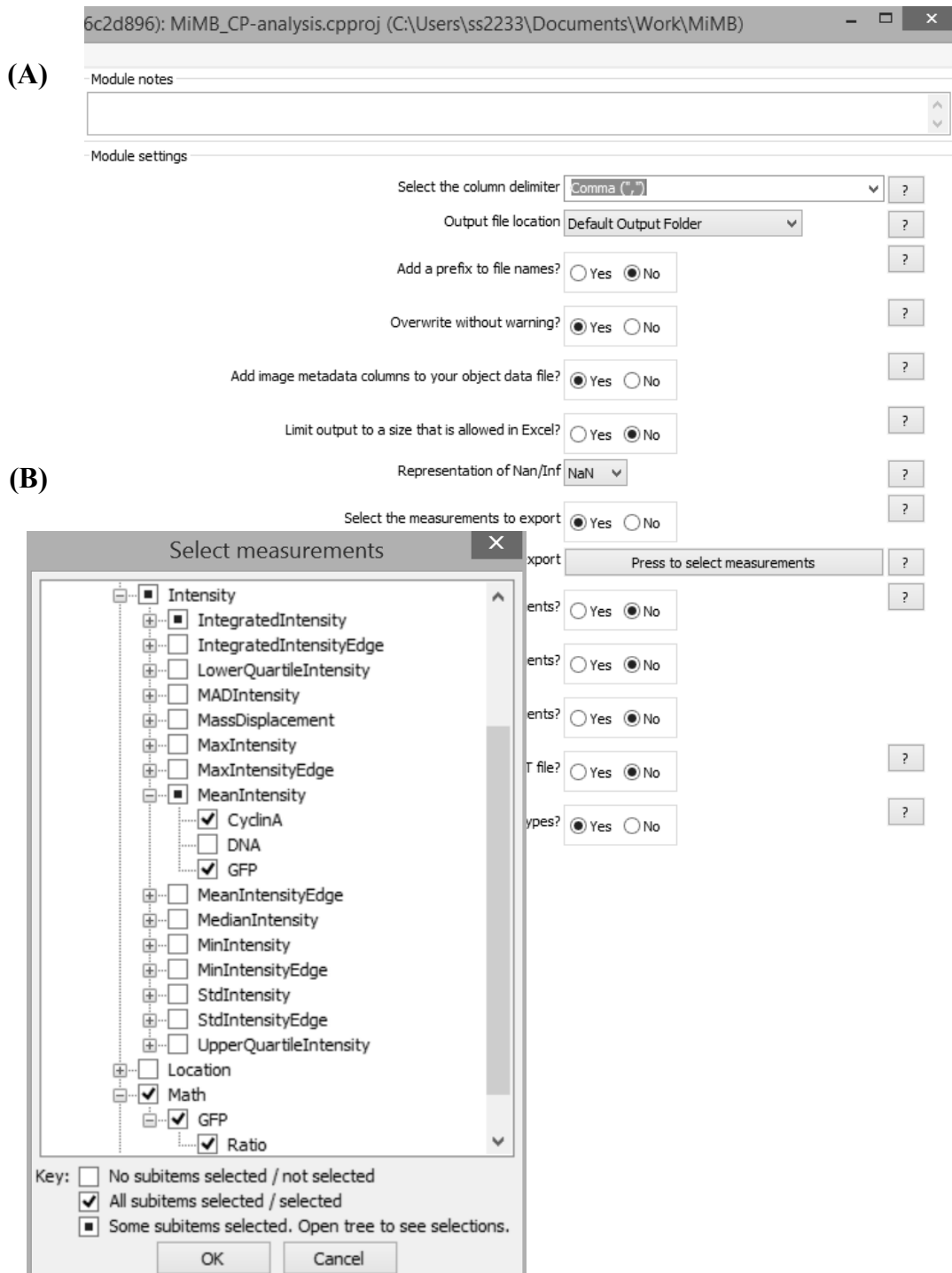


Fig. 13. Screen captures after selecting the ExportToSpreadsheet analysis module. The ExportToSpreadsheet module is used to specify which parameters from those generated in the

analysis should be saved to an output file for subsequent data analysis. (A) Screen capture of the main screen of the module once the ‘Select measurements to export’ option has been set to ‘Yes’ and the ‘Press to select measurements’ button has been revealed. Other key settings here are the choice of comma-delimiter, also ‘Add image metadata columns...’ which have been set to ‘Yes’. (B) Selecting the ‘Press to select measurements’ button opens this sub-window where a tree diagram allows selection of specific measurements for the various output files. The selected check-boxes shown here, for example, are ticked such that the Nuclei.csv output file used in this method will contain mean-intensity values for nuclear CyclinA and GFP as well as the nuclear/cytoplasmic GFP ratio for each cell.

Image	Object	Meta	Meta	Meta	Meta	Meta	Metadata_Well	Met	Met	Intensity_Mean	Intensity_CyclinA	Intensity_Mean	Intensity_GFP	Math	GFP_Ratio
1	1	nan	0	nan	0	0	B2	2	B	0.00139633	0.004198448	0.004198448	1.533248198		
3	1	2	nan	0	nan	0	B2	2	B	0.001498223	0.00478161	0.00478161	1.87220256		
4	1	3	nan	0	nan	0	B2	2	B	0.003066692	0.008932908	0.008932908	2.02092852		
5	1	4	nan	0	nan	0	B2	2	B	0.002666373	0.007158832	0.007158832	2.714209119		
6	1	5	nan	0	nan	0	B2	2	B	0.002299002	0.016080596	0.016080596	3.300196081		
7	1	6	nan	0	nan	0	B2	2	B	0.000947178	0.00529508	0.00529508	1.343685931		
8	1	7	nan	0	nan	0	B2	2	B	0.003167216	0.013874296	0.013874296	3.546157208		
9	1	8	nan	0	nan	0	B2	2	B	0.006044659	0.005122665	0.005122665	1.916539498		
10	1	9	nan	0	nan	0	B2	2	B	0.001240038	0.004647001	0.004647001	1.791802609		
11	1	10	nan	0	nan	0	B2	2	B	0.003987835	0.014322417	0.014322417	4.922882256		
12	1	11	nan	0	nan	0	B2	2	B	0.002181585	0.011921725	0.011921725	2.677402614		
13	1	12	nan	0	nan	0	B2	2	B	0.006916687	0.004344801	0.004344801	1.045273484		
14	1	13	nan	0	nan	0	B2	2	B	0.003318747	0.01756579	0.01756579	3.503633436		
15	1	14	nan	0	nan	0	B2	2	B	0.002715556	0.002330896	0.002330896	0.84093317		
16	1	15	nan	0	nan	0	B2	2	B	0.006974713	0.002230089	0.002230089	1.103462836		
17	1	16	nan	0	nan	0	B2	2	B	0.000247419	0.00193663	0.00193663	0.675552178		
18	1	17	nan	0	nan	0	B2	2	B	0.002294218	0.005137039	0.005137039	1.697352713		
19	1	18	nan	0	nan	0	B2	2	B	0.001674343	0.005906545	0.005906545	2.092818787		
20	1	19	nan	0	nan	0	B2	2	B	0.001792017	0.008351426	0.008351426	2.428334599		

Fig. 14. Screen capture of the output file Nuclei.csv when opened in Microsoft Excel. Each row corresponds to an individual cell identified using CellProfiler. Each column is either a metadata element associated with that cell (e.g. Well, field number) or quantitative data as specified in the ExportToSpreadsheet module (e.g. Intensity_MeanIntensity_CyclinA, Math_GFP_Ratio).

```

1 #install plotting package
2 require(ggplot2) || install.packages(ggplot2)
3 library(ggplot2)
4
5 #set directory and name of analysis data file
6 setwd("C:/Users/ss2233/Documents/work/MiMB/output_files")
7 path <- "Nuclei.csv" #adjust to match chosen filename
8 data <- read.csv(file=path,header=TRUE)
9
10 #choose well to plot
11 set_wells <- c("B2")
12 title <- ("Non-Targeting siRNA")
13 x <- ("GFP-CDK2 Reporter (nuc/cyt) ratio")
14 y <- ("Nuclear CyclinA\nantibody intensity")
15 set <- data[data$Metadata_Well %in% set_wells,]
16
17 #assign thresholds for assays
18 CycAthresh <- c(2.5e-04)
19 GFPthresh <- c(1.5)
20
21 #calculate %cells per plot quadrant subpopulation relative to thresholds
22 set_dim <- dim(set)
23 u1 <- 100*(sum(set$Intensity_MeanIntensity_CyclinA >= CycAthresh & set$Math_GFP_Ratio < GFPthresh)/set_dim[1])
24 l1 <- 100*(sum(set$Intensity_MeanIntensity_CyclinA < CycAthresh & set$Math_GFP_Ratio < GFPthresh)/set_dim[1])
25 ur <- 100*(sum(set$Intensity_MeanIntensity_CyclinA >= CycAthresh & set$Math_GFP_Ratio > GFPthresh)/set_dim[1])
26 lr <- 100*(sum(set$Intensity_MeanIntensity_CyclinA < CycAthresh & set$Math_GFP_Ratio > GFPthresh)/set_dim[1])
27
28 #create x/y scatter plot of cell assay data from selected well displaying thresholds and subpopulation percentages
29 set_xypplot <- qplot(Math_GFP_Ratio, Intensity_MeanIntensity_CyclinA, data = set, xlim = c(0,6), alpha = I(1/10), main = title,
xlab = x , ylab = y)
30 set_xypplot + theme(panel.grid.minor = element_blank()) +
31 geom_vline(aes(xintercept = GFPthresh),size = 1.5, data = set, color = "white") +
32 geom_hline(aes(yintercept = CycAthresh),size = 1.5, data = set, color = "white") +
33 annotate("text", x = 0.25, y = 0.8*(range(set$Intensity_MeanIntensity_CyclinA)[2]),size = 8, label = signif(u1, digits = 3))+
34 annotate("text", x = 0.25, y = 1e-04,size =8, label = signif(l1, digits = 3)) +
35 annotate("text", x = 5, y = 0.8*(range(set$Intensity_MeanIntensity_CyclinA)[2]),size = 8, label = signif(ur, digits = 3)) +
36 annotate("text", x = 5, y = 1e-04,size = 8, label = signif(lr, digits = 3))

```

Fig. 15. R-script to plot annotated scatter plots using the Nuclei.csv file. Installing R onto a computer and running this code will produce a scatter plot of individual cell values from Nuclei.csv. The x-axis displays GFP-ratio values calculated from the GFP-CDK2 reporter and the y-axis the mean nuclear fluorescence intensity of CyclinA antibody staining. User-specified threshold values for both assays can be entered to gate the population and the position of these gates will be superimposed as lines crossing the plot. The percentages of cells within each gate will be automatically displayed over the resulting quadrants. Only those cells in Nuclei.csv corresponding to the specified well-address (e.g. A1) will be plotted and the title of the plot may be user-adjusted to indicate either the well used or the relevant siRNA knockdown condition used in that well

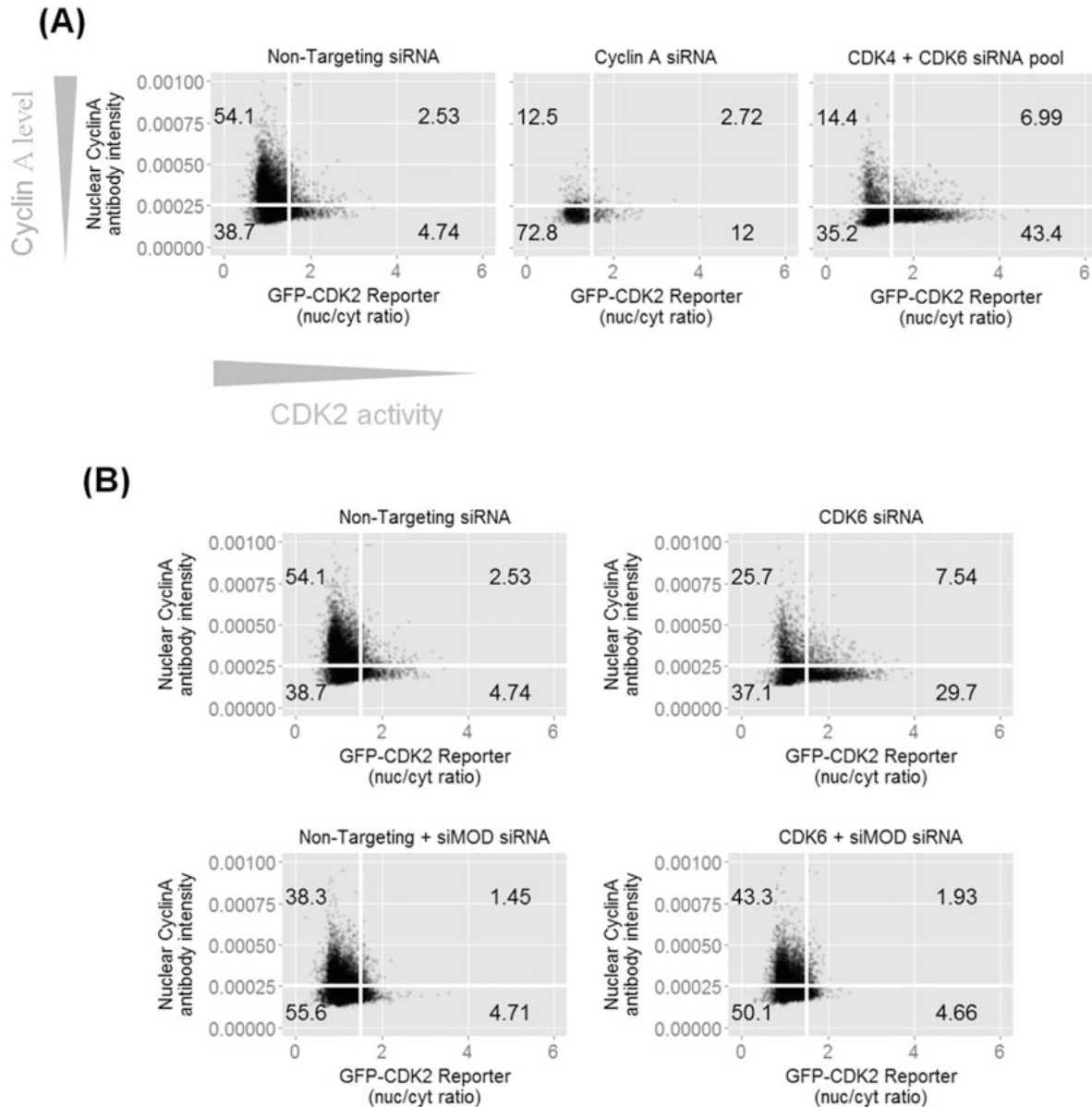


Fig. 16. Example scatter plots of data produced using this analysis strategy. The R-script in **Fig. 15** has been used in conjunction with an output Nuclei.csv file from a siRNA screen. The x-axes display GFP-ratio values calculated from the GFP-CDK2 reporter and the y-axes the mean nuclear fluorescence intensity of CyclinA antibody staining in arbitrary units. Positions of the assay thresholds (bold white lines) are set based on observations of a relevant test set of wells present on each plate of the screen and numbers in each of the quadrants represent the percent cells of the plotted population contained therein. (A) Individual cell assay values from a test set of siRNA representing NT oligonucleotide, siRNA targeting Cyclin A and a siRNA mixture targeting the G1 phase cyclin-dependent kinases CDK6 and CDK4. The Cyclin A knockdown data was used to set an arbitrary

threshold for Cyclin A expression positivity/negativity (horizontal white line) whereas the rightwards trend of the double CDK4 and CDK6 knockdown data, resulting in enrichment of CDK2 negative G1 cells, was used to estimate a suitable value for the vertical threshold of the GFP-CDK2 reporter. (B) Example screen data where a candidate epistatic modifier (siMOD) is shown cancelling loss of CDK2 activity and Cyclin A in cells, otherwise driven by loss of CDK6. Non-targeting siRNA control and CDK6 siRNA data values in the absence (upper panels) and presence (lower panels) of a candidate epistatic siMOD siRNA are shown, respectively. NT siRNA substitutes for the absent siMOD siRNA in the upper panel data such that the total siRNA concentration is constant across all transfection conditions in the screen.