UNIVERSITY OF BATH

Programming in Cell Cycle analysis: An introduction in the context of ROCK inhibition on MCF-10A cells

Bede Young

May 2022
Word Count: 4812

Table of Contents

Lay Summary (220 words)

The progression of a cell through its life cycle is tightly controlled by proteins, including the enzyme ROCK – which performs a variety of key functions throughout this cycle. Previously, ROCK has been eluded as a key factor during cancer development, particularly during metastasis - the long-distance spreading of cancers into new tissues. Use of ROCK inhibitors, which are able to reduce the functions of ROCK, as cancer therapeutics has proved highly encouraging in treatment of many types of cancer – including breast and ovarian. Understanding the full effect of applying ROCK inhibitors on live cells may supply critical knowledge for the future development of effective cancer treatments. In this study, a new method has been proposed which may allow effective and rapid determination of any effect a particular treatment may have, specifically on the cell cycle.

We combined previous analysis methods with a new, coding-orientated approach in order to analyse a larger dataset of cells with reduced selection bias. Thoroughly analysing all cells imaged revealed that inhibition of ROCK appeared to increase the rate of cell cycle progression in several cells, whilst simultaneously hindering progress in other cells within the same condition.

This technique poses to be transferable to many alternate experiments, particularly those involving cell cycle dynamics, although applications may be relevant for various other investigations in the cell biology field.

Abstract (205 words)

Rho-associated protein kinases (ROCK) are multifunctional serine/threonine kinases involved in the regulation of many cellular processes. It is widely appreciated that alterations in ROCK expression and activity are critical in the development of many cancers. However, the effect of Y-27623 ROCK inhibition on proliferation and cell cycle dynamics in cancer development is not fully clear. Furthermore, large-scale analysis of cell data has previously proved challenging in practice. Here, we attempt to produce a novel method of investigating cell cycle stage durations in order to form unbiased comparisons between various conditions of cells from live cell images. MCF-10A breast cancer cell images from Julia Sero's lab (2017) were processed and analysed using machine learning algorithms and image analysis software before analysing data using python. This method generally proved capable of reproducing manual observations over larger quantities of cells. Additionally, cells treated with Y-27623 appeared to produce a greater variance in cell cycle stage durations than those in complete medium. ROCK was therefore indicated to possess multifunctional roles in the regulation of cell proliferation. The use of this method in future data analysis may prove powerful for rapid determination of the effects of various agents on proliferation. However, adjustments and improvements will be required to increase general accuracy.

Introduction (659 words)

Understanding the roles of distinct proteins involved in the regulation of cell cycle dynamics remains an attractive proposition within cell biology. Rho-Rho-associated coiled-coil containing protein kinases (ROCK) have been shown to control many cellular processes, including cell migration, control of cell size and contractility ***(Amano et al. 2000)***. Furthermore, increasing evidence from various studies have hypothesised the dynamic role of this protein in regulation of the cell cycle, hence designing therapeutics to act on this protein may result in the production of effective cancer therapeutics.

ROCK1 and 2 serine/threonine kinases act on a variety of downstream targets. They are regulated by the molecular switches RhoA, RhoB and RhoC, belonging to a subfamily of the Ras superfamily of proteins ***(Goitre et al, 2013)***. These are small GTPases which determine ROCK1 and 2 activity by converting between the GTP-bound active and GDP bound inactive states, mediated by Guanine Exchange Factors and GTP-ase Activation Proteins. Whilst ROCK 1 and 2 only share 65% sequence identity ***(Nakagawa et al, 1996)***, both proteins possess similar kinase, coiled-coil and Rho binding domains - hence share similar mechanisms of activation. Knockout of the ROCK genes in mice have shown to impair interactions between the placenta and embryo, with 90% of ROCK1-knockout mutants dying *in utero* ***(Thumkeo et al, 2003)***. Whilst ROCK1-knockout mutants were often less severely affected, defects in actin cytoskeletal organisation caused failure in eyelid closure and damaged development of the ventral body wall, often preventing survival of mice in early life ***(Shimizu et al, 2005)***.

Growing evidence suggests that ROCK1 and 2 protein kinases play important roles in development of cancers. Overexpression of these proteins has been linked to various cancer hallmarks, including increased invasiveness, migration and proliferation. Rho-GTPase altered activity has previously been linked to metastasis during development of breast cancer ***(Tang et al, 2008)***, and dysregulation of this pathway is further observed in ovarian cancers ***(Cheng et al, 2009)***. In these cases, effects of Rho-GTPase overexpression are linked to increased motility of cells and invasiveness, due to Rho regulation of focal adhesion complex assembly. Furthermore, the downstream ROCK effector adducin promotes cross-linking of

actin filaments with the plasma membrane through increased actin-spectrin interactions ***(Kimura et al, 1998)***, contributing to increased contractility and cellular migration.

Inhibition of the ROCK pathway is viewed as an effective therapeutic for treatments of many cancers. The Y-27623 inhibitor is able to impair function by competing with ATP in the catalytic binding site of ROCK1 and 2. Studies of mice have shown a decrease in cancer cell's invasiveness following treatment with Y-27623, through reductions in expression of c-Myc regulated miR-17-97 clusters, typically seen in higher levels in metastatic cells ***(Liu et al, 2009)***. Whilst this inhibitor has proved effective in many situations, alternate reports of increased proliferation of cancer cells following Y-27623 treatment is evident. Colorectal cell line SW620 showed increased invasiveness upon Y-27623 treatment ***(Vishnubhotla et al, 2012)***, indicating the need for greater context when determining the effectiveness of therapeutic treatments.

Whilst migratory effects on cell lines due to ROCK inhibition are generally well studied, the regulatory function of ROCK on the cell cycle is less understood – with contradicting results produced depending on cellular context and cell line. Increased proliferation is commonly associated with ROCK inhibition, however several downstream effects on key cell cycle signalling pathways have been identified as potential antagonists for proliferation (***(Liu et al, 2009)***.

Investigating these effects can be challenging, since ROCK inhibition has previously been suggested to both promote and inhibit cell cycle progression. Previous experiments attemped to further highlight the effect of Y-27623 on the duration of cell cycle stages in breast cancer cell line MCF-10A by useing live single cell imaging techniques to determine durations. Whilst manual observations provide greater accuracy, this method of single cell analysis may lead to selective bias, either towards cells undergoing hyperproliferation or those having left the cycle completely due to ROCK inhibition. To remove this bias and increase the efficiency of data analysis, attempts were made using the Python language to produce an automated method for determining the duration of each stage of the cell cycle. This project aimed to incorporate the entire set of processed cells within the dataset whilst maintaining high accuracy for each prediction.

4. Methods and Materials (1445 words)

4.1 Experimental Setup

MCF-10A cells were pre-prepared and imaged in Julia Sero's lab and engineered to express endogenous mRuby-tagged PCNA. Cells were starved in base medium (DMEM:F12 + 5% horse serum) for 24 hours in order to enter quiescence, then plated in 20 µl Complete Medium **(CM)** (DMEM:F12 + 5% horse serum + 20 ng/ml epidermal growth factor (EGF) + 0.5 µg/ml hydrocortisone (HC) + 10 µg/ml insulin + 100 ng/ml cholera toxin), or plated in 20 µl **CM** + 10 µl **Y-27623 (Y27)**. Images were taken at 15-minute intervals by an Opera Cell :: Explorer spinning disk confocal microscope with 20x air objective lens (Numerical Aperture = 0.45). Cells were maintained at 80% humidity, 37°C and 5% CO_2 throughout imaging.

The required cell data was extracted from 8-bit movie files using three image processing applications: FIJI; Ilastik 1.3.3 and CellProfiler 4.2.1. FIJI and Ilastik were primarily used to produce four types of input files for CellProfiler **(Figure 1)**, with CellProfiler exporting the cellular data to csv. files. These spreadsheets were then manipulated using excel formulas and analysed with python to produce cell cycle stage durations.

4.2 FIJI

FIJI was first used to adjust the LUT of microscopy image stacks to higher contrast levels. The same values were applied to all files, allowing Ilastik to predict classifications with greater accuracy later in the workflow process. These stacks were subsequently saved as 8-bit tiff sequences with specific filenames. These would later be imported into CellProfiler, with a regular expression (Regex) used to extract metadata from the filenames given.

4.3 Ilastik

Using Ilastik's machine learning tools, Pixel Prediction (PXP) Maps, Object Classification (OBJ) and Cell Tracking (TRK) files were produced.

Contrast adjusted stacks from FIJI were uploaded to Ilastik. File dimensions were converted from zyx to tyx, allowing the stack to be viewed at individual timeframes, each taken at 15-minute intervals. Pixel prediction training was then performed on time frames at 2-hour intervals of the adjusted stack (every eight frames), ensuring a broad representation of the entire time period during training. User-annotation of 'background' and 'nuclei' was painted onto the images, with specific care taken to highlight narrow background spaces **(Figure 2)** between two cells, as well as cells with particularly dim appearances. The algorithm uses the annotated training to assign each pixel a prediction value for 'background' or 'nuclei'. Prediction maps were then exported with pixel intensity values ranging from 1-0, with higher values indicating greater certainty of nuclei and lower values indicating greater certainty of background. Export files were saved as unsigned 8-bit tiff sequences (import files for CellProfiler) and floating 32-bit hdf5 files (these would later be used for object classification and cell tracking steps). This training step must be completed before all other Ilastik-related processes **(Figure 1)**.

Hdf5 PXP maps were subsequently used - along with the contrast adjusted raw files - to create object classification predictions and cell tracking files. Pixel threshold values were universally set at 0.53 during object classification and tracking, meaning only pixel prediction values equal to or above 0.53 were counted as identified cells. This step may have reduced apparent cell size by a small number of pixels but conversely ensured separate identification of cells close to each other, ultimately increasing the accuracy of object classification. Falsely identified cells were also filtered out by setting the minimum diameter of objects to 40 pixels. Classification training was carried out using five distinct coloured labels for identifying each stage of the cell cycle:

- Yellow - Dim (G1);
- Blue - Dim speckles (Early S phase);
- Red - Bright speckles (Late S);
- Cyan - Smooth and bright (G2);
- Green - Dividing (Mitotic).

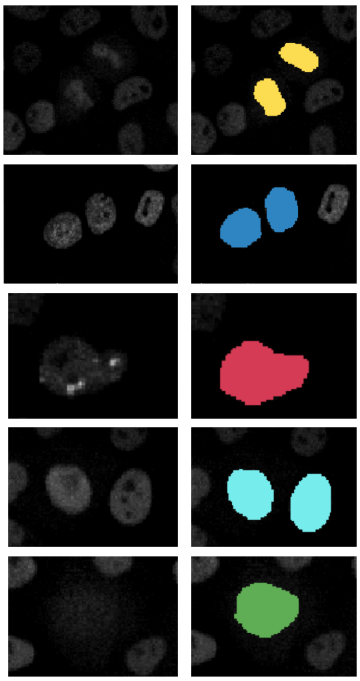

Figure 3 – Object Classification Training. Example annotations **(Top to Bottom)** for G1 **(Yellow)**, Early S **(Blue)**, Late S **(Red)**, G2 **(Cyan)**, Mitotic **(Green)**, within the Ilastik Machine Learning object classification workflow.

A large amount of training was carried out on the test file in order to produce accurate object identity predictions following batch processing. Specific training focus was applied to cells transitioning between cell cycle stages, since this period would likely produce the greatest variance in predictions. Following manual inspection of object predictions, OBJ files were exported as unsigned 8-bit tiff sequences (import files for CellProfiler).

Finally, Ilastik's cell tracking (TRK) function was used to assign each predicted cell a 'Track Number'. The tracking algorithm was first trained to detect scenarios in which mitotic cells divided into two daughter cells, and further to distinguish these from scenarios in which two non-dividing cells merged into each other between frames **(Figure 4)**. This prevented cells from merging track numbers, which would severely reduce the reliability of eventual results produced. The tracking parameters **(Figure 5)** were designed to both prevent cell tracks from disappearing and simultaneously prevent daughter cells from sharing the same track as the parent cell. This process aimed to limit the number of instances in which track numbers appeared multiple times within the same frame. Accomplishing this would allow the user to easily track the progression of each cell through its cycle by simply sorting the output csv file by track number. Manual inspection of tracking results was performed, and outputs were exported as unsigned 8-bit tiff sequences.

Figure 4 – Tracking Training. (Top Left) Dividing cell takes on two new track numbers in following frame **(Top Right).**

Annotation to indicate merging cells **(Bottom Left)** and diverging cells in the next frame **(Bottom Left)** using Ilastik's tracking function.

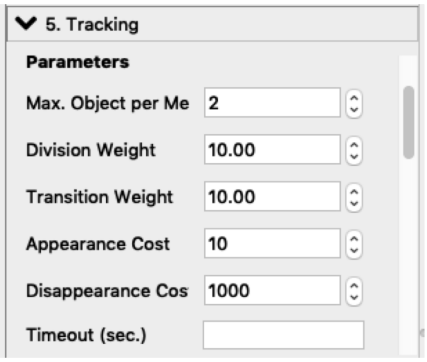

Figure 5 – Tracking Parameters. Disappearance cost was adjusted to be much higher than default, to limit instance of cells changing track numbers between frames.

4.4 CellProfiler

The Ilastik output files were uploaded to CellProfiler as 8-bit tiff sequences. This included 4 folders (RAW, PXP, OBJ and TRK) of 173 tiff sequences, each ordered from frame 1 to frame 173. Metadata from the filenames was extracted using a specific Regular Expression (see appendix). A pipeline previously worked on by Thain, A. (2021) was used as a template and adjusted to produce desired outputs (see appendix). For the purpose of this project, only

data relevant to the cell cycle stage (Intensity_MaxIntensity_Object), track number (Math_TrackNumber) and image number was extracted – however, various other measurements and calculations can be obtained from this pipeline. A single image set was analysed before running full analysis, to ensure cells had been correctly identified, classified and tracked.

4.5 Microsoft Excel

Following CellProfiler analysis, csv. output files were sorted by Math_TrackNumber from largest to smallest. Data was tidied by deleting all rows containing a track number of 255, since these cells had been identified by Ilastik as false detections or as cells that had left the image frame - either by passing the edge of the frame or by fading into the background. The Intensity_MaxIntensity_Object and Math_TrackNumber columns were copied and brought into new xlsx. files. In order to later load these columns into python, the values were standardised and rounded. In this case, object intensity values were universally multiplied by 250, before using Excel's rounding function:

=ROUND('cell_ID':0) ***# selected cell is rounded to 0 decimal places***

This produced a new column of intensity values ranging from 1-5 titled "**stage**", sorted by track number.

****Note – the value of multiplication may vary depending on classification intensities following Ilastik export, however producing a column ranging from 1-5 is essential for the code to run****

Math_TrackNumber values were rounded in a new column titled "**track**" using the rounding formula above. Before importing this file to python, a final check was performed to ensure that no rows contained "stage" values of 0, since this would result in errors whilst running the code. In this case, the entire row was deleted. Whilst this check may have slightly

affected results in a few cases, the total volume of data processed was great enough to overwhelmingly negate these changes.

4.6 Python

A code was written to analyse large amounts of data produced in parts *4.2, 4.3* and *4.4* and prepared in *4.5* in order to return average cell cycle stage durations as well as variance and standard deviations (see appendix for full script). An initial code consisted of a 'for loop' which read through the "**stage**" column by converting this into an 'array' (list of values). For each repeated value, a 'count' was increased by one, which was added to an 'list' of that particular value in any scenario where the current value did not match the previous **(Figure 6).** Once every row in the "stage" column was read, mean values for each 'list' were returned.

This code was expanded to include a dynamic function, allowing "**stage**" values to be mutated in the event of a 'blip'. This is defined as a sequence matching the pattern:

$$[\, a\, ,\, n(a')\, ,\, a\,]$$

Where:

- a = any number
- a' = any number except a
- $n \leq$ 'Frame Blip Width'

A Regular expression was written into the code to perform this function, with the user inputting the desired 'frame blip width' (FBW). If a sequence matched the Regular Expression, the *a*' values, which were sandwiched between identical *a* values, were mutated

to match a. An excel sheet containing the mutated values was produced, and adjusted values for mean and variance were returned. These outputs were used to produce violins plots for variance within each cell cycle stage.

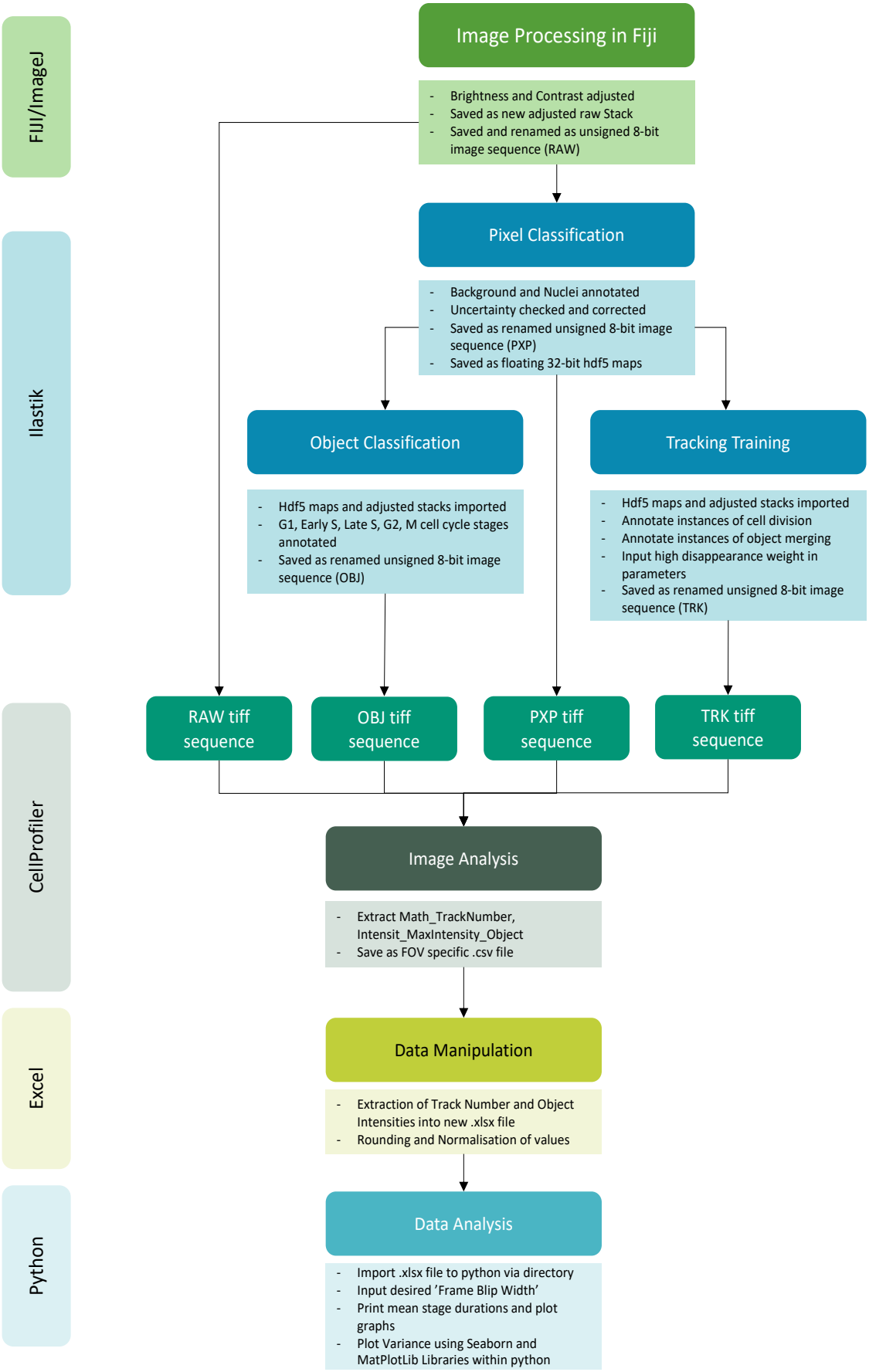

Figure 1 – Flowchart illustrating project workflow. Images produced in FIJI and Ilastik were processed in CellProfiler and analysed using Excel and Python.

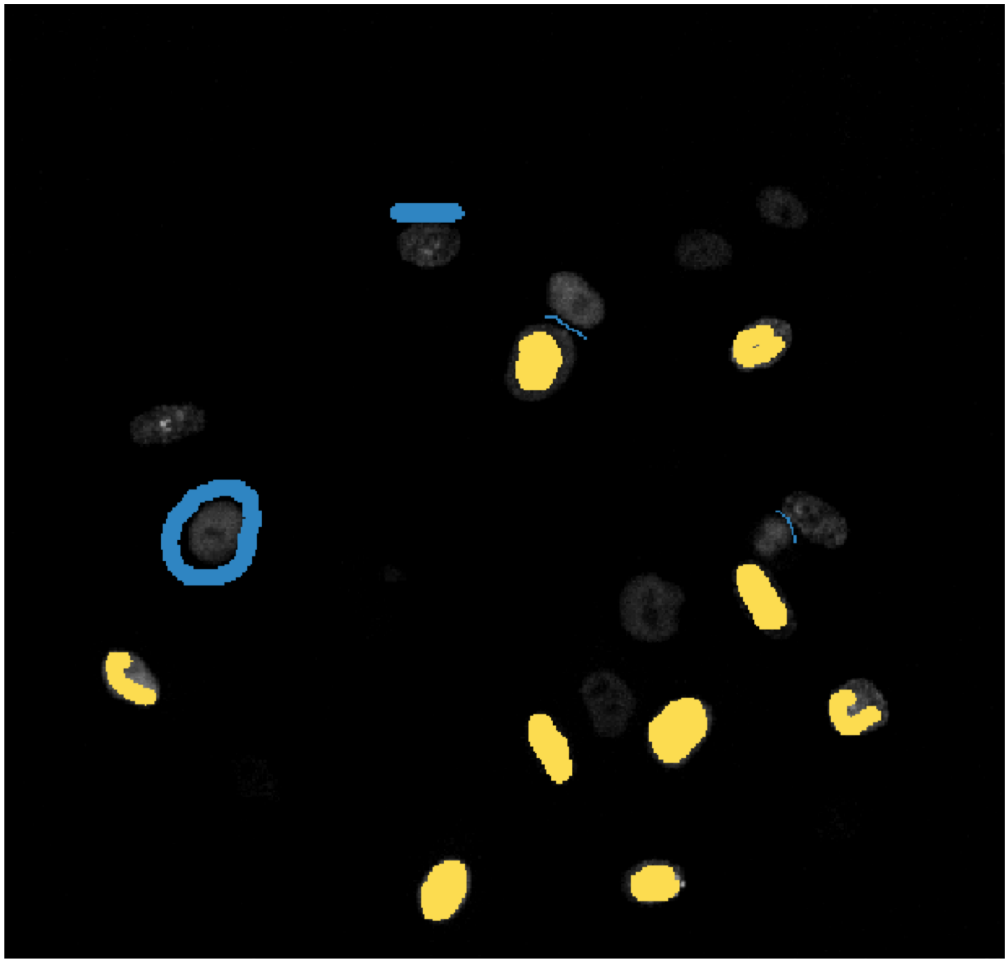

Figure 2 – Pixel Classification Training. Example annotations for Nuclei (Yellow) and Background (Blue) within the Ilastik Machine Learning pixel classification workflow.

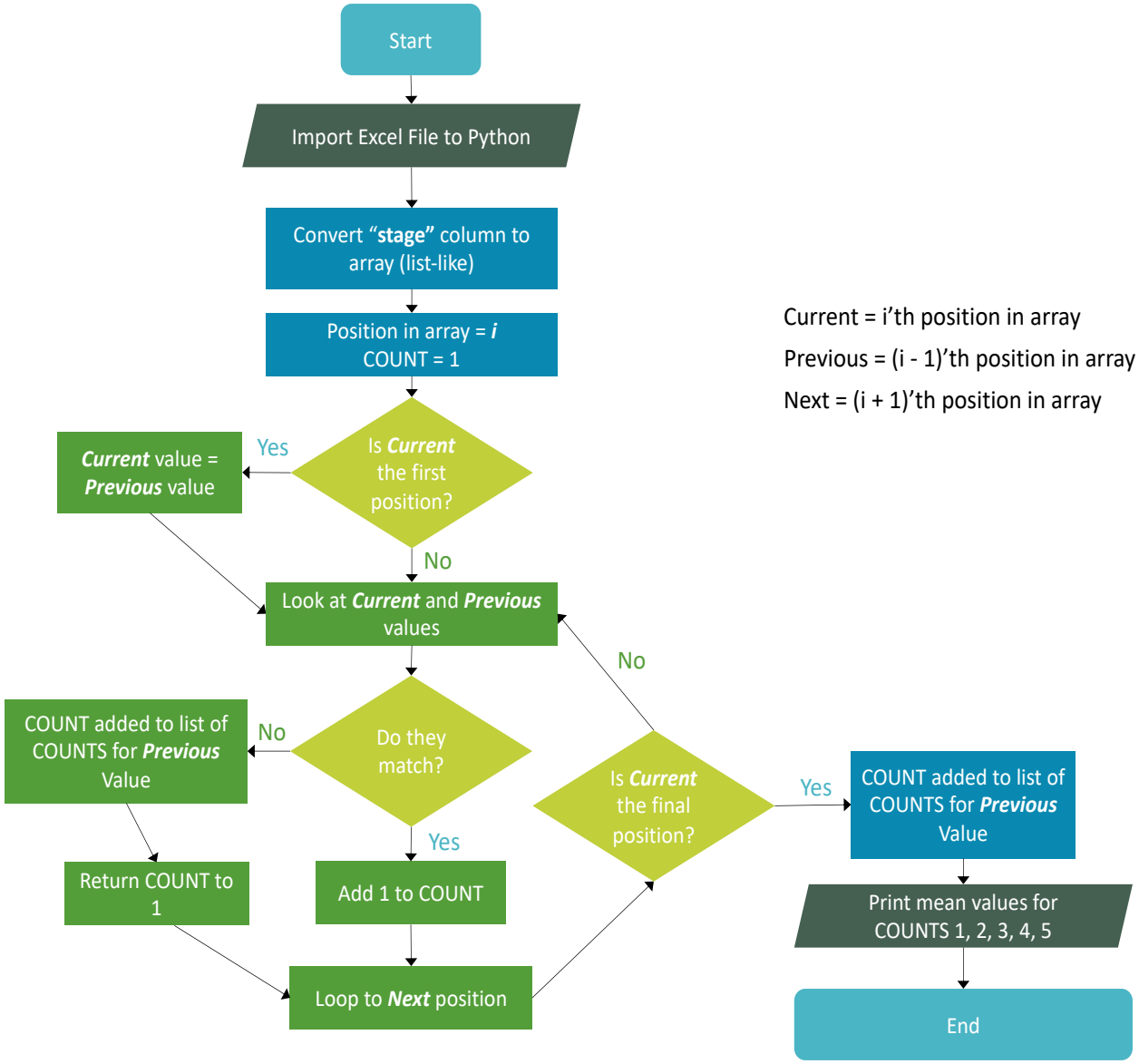

Figure 6 – Flow Chart describing initial 'For' loop.

Excel sheets were converted in python into lists, in order to be read by the code. Means were calculated from a dataset of 'COUNTS' for each value 1-5, corresponding to each stage in the cell cycle:

1 = G1;
2 = Early S;
3 = Late S;
4 = G2;
5 = Mitotic

Results (980 words)

Initial mean calculations were significantly lower than expected values.

To determine the effectiveness of the analytical method used, results for mean stage durations were compared to previously observed measurements from a previous study **(Lindley, F. 2021)**. Median times for each stage were observed and plotted against calculated means **(Figure 6)** (S phase was calculated by adding Early S and Late S durations), predominantly indicating vast underestimations produced by the code for each stage, besides G1. Average S phase, G2 and M stage durations were < 50% of the observed times by Lindley, counted in frames.

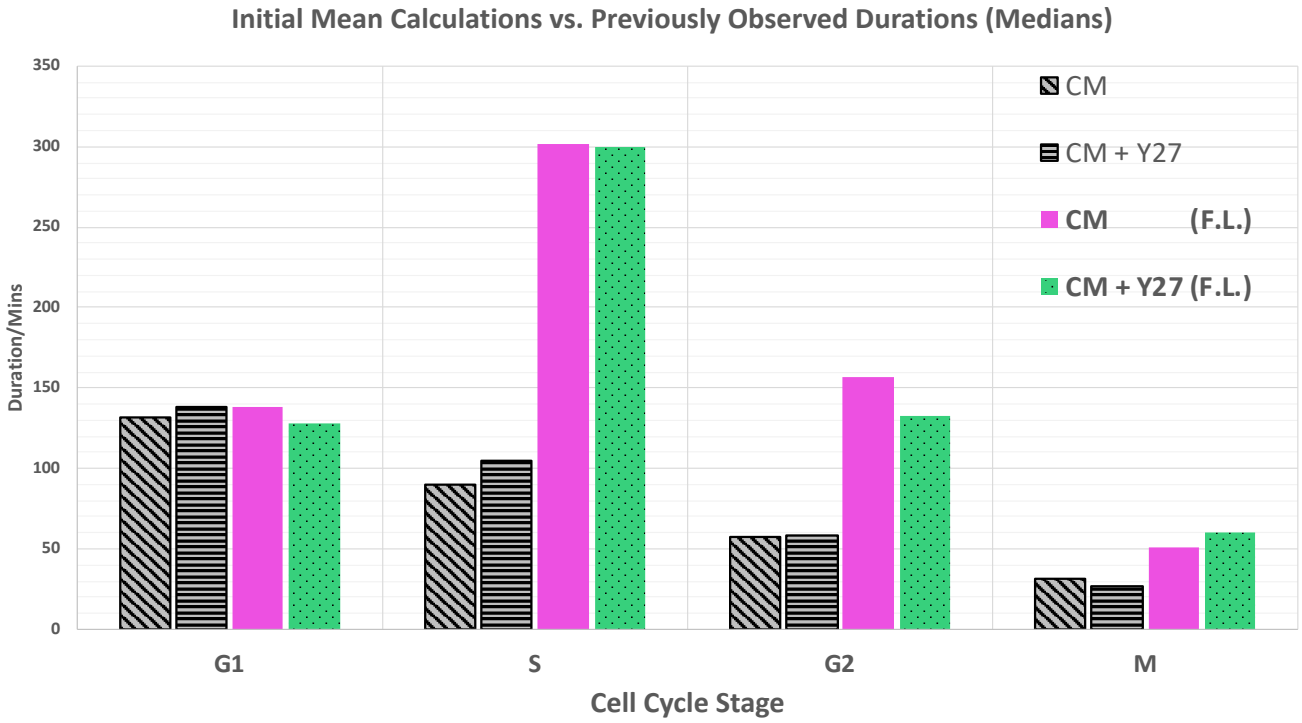

Figure 6 - Mean Calculations from initial coding resulted in lower than expected stage durations.
Median stage durations from single cell analysis (Lindley, F. 2021) were significantly higher than mean calculations before input of the 'Frame Blip Width' function. Averages from five fields of view of one **CM** well and two **CM + Y27** well were collected in Excel to generate an overall average across all fields of view for each medium.

Dynamic 'Frame Blip Width' Function reduced frequency of 'blips' caused by uncertainty during Object Classification.

Underestimates from **Figure 6** were hypothesised to originate from high levels of uncertainty during object classification, particularly during transition periods between stages of the cell cycle. This was confirmed by generating a plot of a single cell over 61 frames against its predicted classification – indicated by the values 1-5 **(Figure 7A)**. High levels of uncertainty created large amounts of 'blips' between stages – likely stemming from similarities in PCNA intensity between these stages. Ultimately, the code would incorrectly register these blips as '1-2 frame durations' for respective stages, producing hugely underestimated means during calculation.

To investigate the effect of applying the dynamic 'Frame Blip Width' (FBW) function to predictions of cell cycle stages, a single cell was tracked through 61 frames before and after applying a FBW of '3' frames into python. Between frames 1 and 11 in **Fig 7A**, the cell appeared to change between G1 and Early S (1 & 2) phases 7 times. These deviations were completely removed for this particular cell upon running the FBW function with a width of 3 inputted into the code **(Fig 7B)**, producing more representative durations for G1 and Early S in this instance. Further smoothening was observed at frame 34-36 in **Fig 7B** when compared to **Fig 7A** between stages 4 and 2 (G2 and Early S).

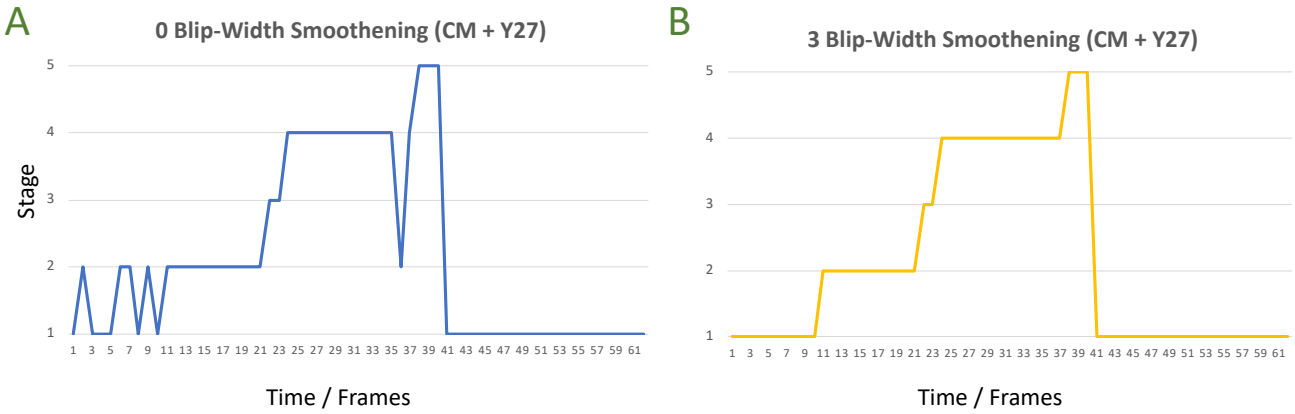

Figure 7. 'Frame Blip Width' (FBW) function successfully reduced "noise" during cell cycle stage predictions.
A single cell from a CM + Y27 well was tracked through 61 frames by the sorting of excel columns by Track Number. Time = 1 refers to the first frame this cell was tracked from (1 frame = 15 min), as opposed to the first frame of the 8-bit tiff file. Stage values 1, 2, 3, 4 and 5 correspond to cell cycle stages G1, Early S, Late S, G2 and Mitosis respectively. **(A)** Time/Stage graph produced on output Excel file following no 'Smoothening' by the FBW function in Python. **(B)** Identical cell tracked over the same timeframe following input of "3" in frame_blip_width.

Increasing the 'Frame Blip Width' Value Led to Higher Calculated Means for Each Cell Cycle Stage Duration.

In order to uncover the most suitable input for the Frame Blip Width (FBW) function in order to produce representative results, a field of view from **CM** and **CM + Y27** was run in python with 0, 1, 2, 3 and 10 FBW to produce average stage durations for each cell cycle stage.

Each increment in FBW produced higher values for mean stage durations in both **CM** and **CM + Y27** **(Fig 8B)**. The trend observed was maintained whilst running 0, 1, 2 and 3 FBW, whilst FBW of 10 produced Late S average duration calculations lower than FBW 3, 2 and 1 in **CM** **(Fig 8A)**. Furthermore, mean values appeared to plateau as higher FBW values were inputted, therefore a value of 3 was considered most appropriate.

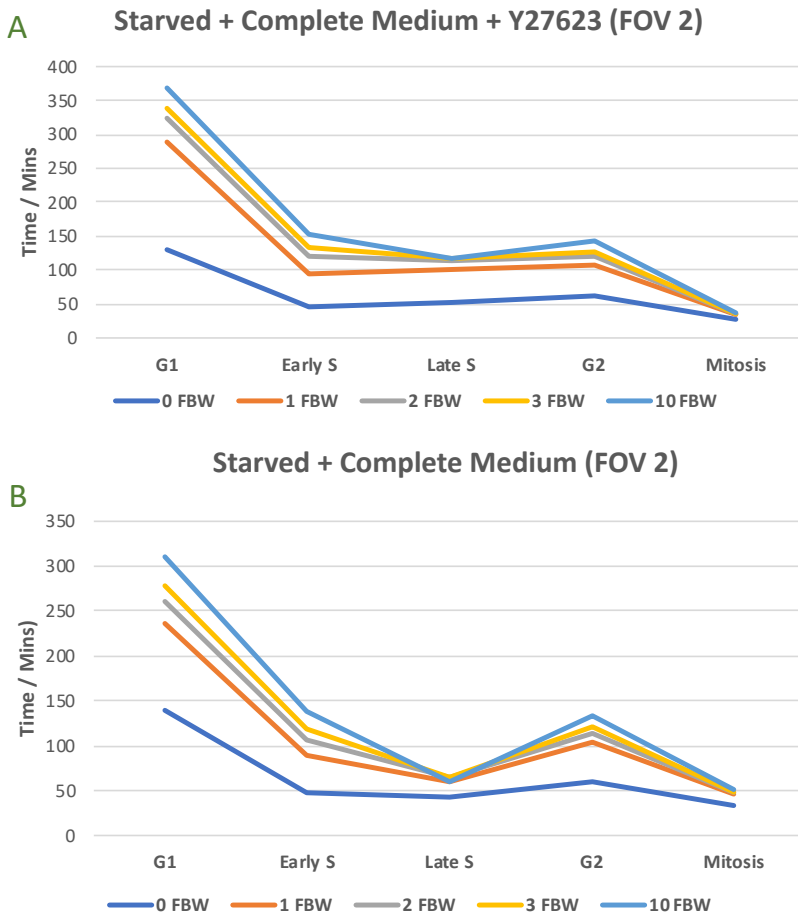

Figure 8. Increasing 'Frame Blip Width' (FBW) input generally resulted in increased cell cycle duration estimates in both conditions.
Mean values for cell cycle stage durations (mins) in MCF-10A cells expressing mRuby-PCNA, **(A)** plated in 20 µl CM in field-of-view 02, **(B)** plated in 20 µl CM + 10 µl Y-27 field-of-view 02. Each stage was tested at FBWs of 0, 1, 2, 3 and 10 by inputting the above values into the FBW input in python.

Applying 'Frame Blip Width' function generated results more comparable to expected values.

To evaluate the success of Frame Blip Width in correcting underestimated means, final calculations were plotted in **Figure 9** against medians as observed by Lindley, F (2021) from the same cell images. Much greater values were observed in G1, whilst lower values were in seen in S phase, although less disparity was observed in G2 and M, despite final calculations remaining generally slightly lower than previous observations. Disparity in G1 and S stages may be attributed to either subjective classification, as well as inaccuracies stemming from the FBW function (see discussion for further evaluation).

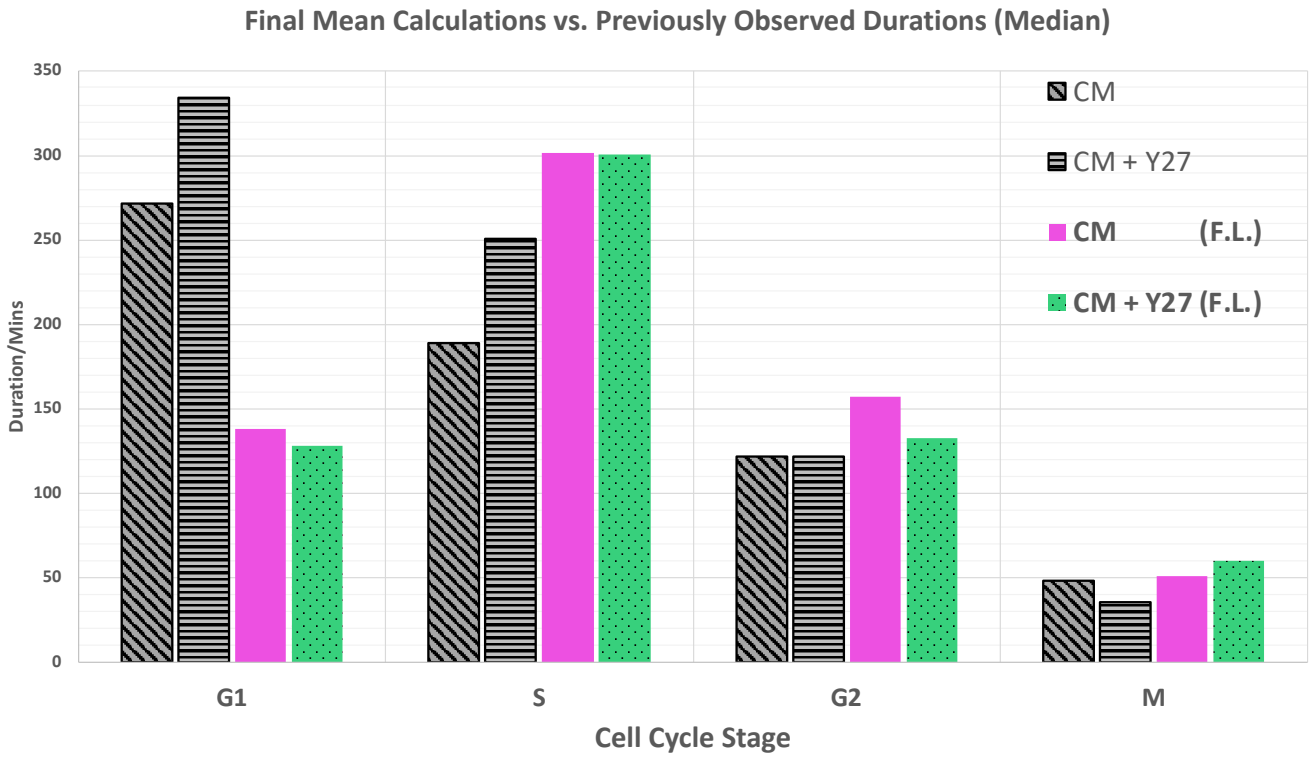

Figure 9 - Mean Calculations following 'blip' correction appeared far more comparable to previously recorded observations.
Averages were recorded after inputting '3' into the 'Frame Blip Width' function and plotted against median results from live single cell analysis (Lindley, F. 2021) for comparison.

Y-27623 Mediated ROCK Inhibition May Affect Average Duration of Certain Cell Cycle Stages in MCF-10A cells.

Following the decision to run '3' as the Frame Blip Width value **(Figure 8)**, Inhibition of ROCK in MCF-10A cells was shown to increase the average duration of G1 **(Figure 10A)** and Late S **(Figure 10C)** stages when compared against five fields of view (FOV) from cells in **CM**. Average means of G1 duration from **CM** FOVs was 272 mins, with **CM + Y27** average durations being 334 mins. Unpaired ***t*** tests were performed for both observations, producing ***p*** values of 0.001 (G1) and 0.00002 (Late S), as well as for results for both repeats of **CM + Y27** in **(Figure 10A)**, producing ***p*** = 0.43. Results for Early S, G2 and Mitosis **(Figure 10B, Figure 10D, Figure 10E)** showed little evidence of effect on stage duration but showed comparable variance between repeat fields of view and wells.

Results for Late S phase from **(Figure 10C)** were challenged by manually counting durations of Late S from the raw microscopy images. FOVs from **Figure 10C** which had produced extreme values were selected for manual checking **(Figure 11A)**. Average durations of Late S phase observed by 10 cells at 15 frame intervals was shown to produce significantly less disparity **(Figure 11B)** between tests than calculated by the code, with **CM** cells' Late S duration averaging 70.0 mins against **CM + Y27** cells' mean duration of 87.0 mins.

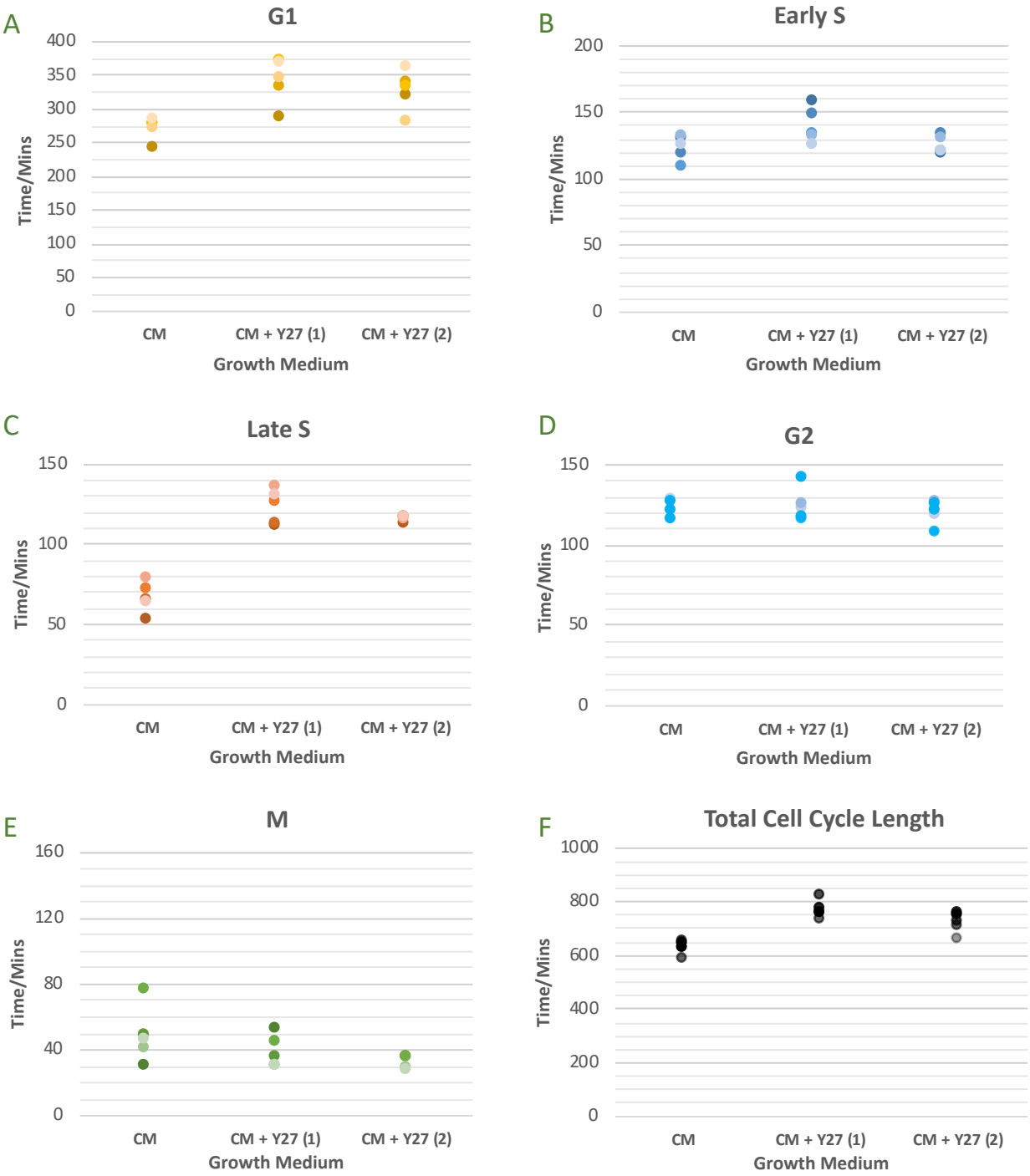

Figure 10. Calculated means for G1 and Late S phase cell cycle stage durations are increased in Complete Medium + Y-27623 cells in comparison to Complete Medium cells.

Five fields-of-view were processed from **CM** and 2 wells of **CM** + **Y27**. Data was run through code with Frame Blip Width of **3** to produce corrected mean values for each stage (G1 **(A)**, Early S **(B)**, Late S **(C)**, G2 **(D)**, M (Mitosis) **(E)** and Total cell cycle length **(F)**) of the MCF-10A cell cycle in each condition. Mean times were plotted by multiplying the duration in frames by 15 to produce time in minutes.

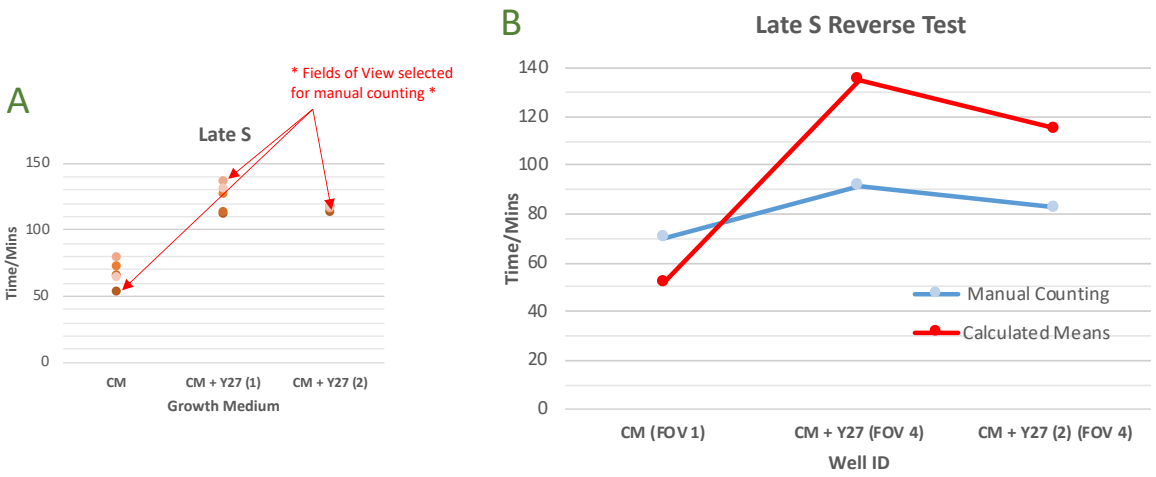

Figure 11. Manual counts from extreme results produced far more comparable results following reverse testing.
(A) Extreme mean calculations from Fig 10C were identified and reverse tested. Ten cells were chosen at intervals of 15 frames, with late S phase durations counted in frames and multiplied by 15 to calculate time estimations in mins. **(B)** Counted average durations in **CM** and **CM + Y27** were plotted against previously calculated means using Microsoft Excel.

Variance Observations in CM and CM + Y27 mediums suggest ROCK Inhibition Decreases Duration of G2 Phase.

Initial cell cycle duration averages failed to identify any change in G2 duration due to Y-27623 mediated inhibition of the ROCK pathway **(Fig 10D)**. To investigate the effect of ROCK inhibition more closely, datasets from **CM** and **CM + Y27** were further analysed using the Seaborn library within python to produce Violin Plots for one FOV in each cell condition **(Fig 12)**.

Kernel density estimations from G2 duration data **(Fig 12D)** in **CM** sharply increased at approximately 11-12 frame duration, whilst a similar spike was observed in **CM + Y27** around the 8-9 frame mark. Increased densities may correlate to higher frequencies of these stage durations within each FOV tested. Large spikes at 1-2 frames were observed in both mediums of G2, as well as G1 **(Fig 12A)**, likely highlighting errors in object classification which the python code was unable to filter out.

Variance patterns were largely consistent across other cell cycle stages, with the exception of low-density predictions of very high duration values, seen in almost all plots of **Figure 12**. Late S variance was also greater in the **CM + Y27** test than **CM** **(Figure 12C)**, supporting results from **Fig 10C**. Lowest variance in results was observed in the M stage **(Fig 12E)**, although anomalies were nevertheless produced at values above 25 frames in duration in both mediums.

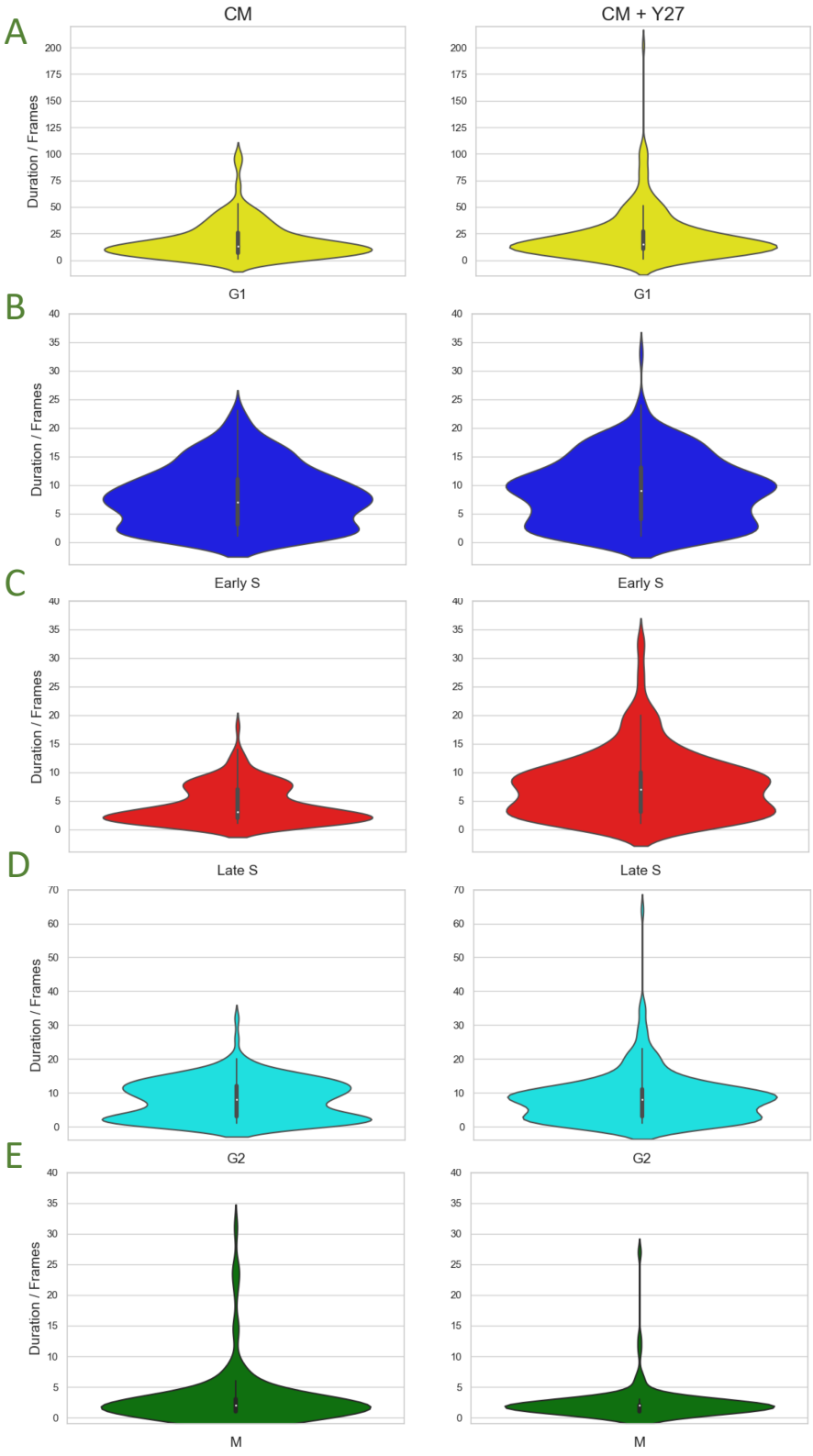

Figure 12. Different variance patterns observed in Late S and G2 stage durations in CM + Y27 compared to CM.

Seaborn violin plots produced in python show variance of cell cycle stage durations for cells plated **CM** or **CM + Y27**. Protrusion of the coloured area indicates higher kernel density estimations, derived from greater occurrence of a particular value within the dataset. Mean values are indicated by white dots, and upper and lower quartiles by thicker black bars. A single field of view was selected from **Fig 10** for both conditions.

Discussion (1728 words)

In this study, we propose a novel method for non-biased, large-scale analysis of cell data in this case specifically targeting the duration of cell cycle phases. Previous use of open source imaging software has proved capable of generating huge quantities of cell data through machine learning and image analysis **(Thain, A. 2021)**. However, since interpretation and analysis of this data proved challenging in practice, attempts were made using the python language to increase efficiency of data analysis whilst maintaining representative and accurate results. We found that despite shortcomings in the algorithm training steps in Ilastik, this method was able to generate mostly comparable results to those observed in previous studies using manual observations. Despite this, adaptations to this code will be required in order to clean up results for increased accuracy of future results.

Successes of the Code

Initial targets of generating average durations of cell cycle stages from large volumes of intensity data were successfully met, despite values appearing significantly lower than expected based on findings from previous studies **(Lindley, F. 2021)**. Uncertainty in object classification proved to be the largest cause of undervalued means for each stage of the cell cycle, particularly noticeable around transition points between each stage. Flickering in predictions between two stages (or 'blips') resulted in both the underestimation of stage durations as well as the code falsely adding results of 1 frame to each growing list of stage durations, from which the mean was calculated.

Results were largely improved following correction of 'blips' using the 'Frame Blip Width' (FBW) function (see appendix). Since multiple cell cycle stages often only lasted 2-3 frames, (eg. Mitosis and Late S) 3 FBW was selected for use in general data collection, reducing the risk of completely overwriting entire stages by recognising these entire stages as 'blips'. Despite this, massively increasing the FBW input value had less of an effect on average durations than expected **(Figure 8)**, indicating that the largest contributors to underestimation of means were 'blips' of one or two frames. Furthermore, comparable

results were observed in M phase regardless of the value of FBW inputted. This indicated that the code had successfully avoided regarding short stage durations as 'blips', despite the inputted FBW value being larger than the average duration of M in frames.

Testing of this code on cells with the same mRuby-tagged PCNA with and without the addition of Epidermal Growth Factor (EGF) would challenge the ability of this code to detect general changes in cell cycle duration, since significant increases in duration of growth phases G_1 and G_2 would be anticipated in EGF absent tests.

Limitations of the Code

One major issue with this method of cell cycle stage classification is the subjective nature of algorithm training within Ilastik. Whilst comparisons can be reliably drawn from results within a single study, attempts to compare findings with previous studies which have trained object classification slightly differently will inherently produce altered predictions. This may be the cause of great disparity in mean durations seen in **Figure 9** between observations by Lindley, F. and final mean measurements from G1 and S phase.

Whilst general errors in classification were largely overwritten **(Figure 7)**, inaccuracies that did not match the Regular Expression written were ignored, and furthermore treated as one or two frame recordings for each respective stage. This occurred most often during transitions - particularly G_1/S - in which one frame predictions of G_2 between these stages will have recorded G_2 duration as 1 frame within the dataset. The sudden condensing of DNA in early S may have caused increased predictions for G_2 for a small number of frames. This may explain the very high frequency of 1 frame G_2 durations observed in Figure 8D CM cells.

Further inaccuracies may have stemmed from cells leaving the frame or field of view before finishing the current stage – leading to lower durations for that stage. Additional code could be written to consider changes in the Track Number column simultaneously, with current 'count' value reset and discarded in the instance of changes in Track Number. This will

reduce the quantity of durations recorded, but slightly improve accuracy of results nonetheless.

The decision to mutate 'blips' to the previous cell cycle stage made for simpler programming, however may have reduced accuracy of G_1 durations in particular – since appearance between the first frames of S and the entirety of G_1 share the closest resemblance. 'Blips' between G_1 and Early S, whilst in this study mutated to G_1, would likely indicate prior passing of the G_1/S checkpoint. In general, G_1 classification proved to generate the greatest variance, and engineering code to accurately predict G_1 lengths may be a challenge. Further difficulty arose from failure to distinguish between G_2 cells deeper into the field of view, and G_1 cells closer to the shot. Greater user-input during object classification training may improve predictions in these scenarios.

Increased G_1 duration in Y-27623 treated cells

Several studies have suggested correlations between Y-27623 **(Y27)** specific inhibition of the Rho-associated protein kinases ROCK1 & 2 and decreased duration of the G1 cell cycle phase ***(Zhiyuan et al 2012)***. Mechanisms of cyclins A, D1 and E regulation by ROCK ***(Croft, Olsen, 2006)*** have been linked to cell cycle progression, in particular the G1/S phase transition ***(Makaewa et al 1999)*****.** Further evidence suggests that ROCK inhibition may play a key role in increased phosphorylation and activation of the ERK1/2 transcription factor ***(Croft, Olsen, 2006)***, a key promoter of G_1/S transition ***(Westbrook et al, 2007)***. Results from this study fail to support previous findings, with recorded G1 durations appearing higher following ROCK inhibition by Y27 **(Figure 10A)**. We speculated that increased variance in G1 values were a potential cause of these findings, perhaps caused by reduced motility caused by disruption of ROCK function in actin cytoskeleton organisation ***(Lloyd et al, 1999)***. This may have resulted in cells that had exited the cell cycle staying within the frame dimension and field-of-view for longer time periods, producing higher frequency of greater G1 durations. This theory is supported by variance graphs for G_1 duration in ROCK inhibited cells, with higher proportion of G_1 durations of 75-120 frames than in ROCK functioning cells. Furthermore, inhibition of ROCK has been related to increased expression of p27Kip1 ***(Croft, Olsen, 2006)***, which down regulates cell cycle progression by binding to the cylin-E-

Cdk-2 complex ***(Meloche, Pouyssegur, 2007),*** inhibiting initiation of DNA duplication during the G_1/S transition. Due to image data being processed un-biasedly, cells which have exited the cell cycle may have partially skewed G_1 durations negatively in both mediums. Hence a multitude of variables leading to cell cycle arrest or cell dynamics may be affecting results in this case. Due to only one well of Complete Medium cells being analysed, further testing is critical before hard conclusions can be drawn from these findings.

G_2 durations in Y-27623 treatment

Observations from variance plots highlighted potential decreases in G_2 duration as a result of ROCK inhibition. Previous observations from Lindley, F. **(2021)** indicated a statistically significant decrease in G_2 length in cells plated with Y-27623 ROCK inhibitor. Initial findings in this study failed to reproduce these results, although further analysis of violin plots **Figure 12D** indicated greater variance in G_2 values in ROCK inhibited cells. Increased proportions of ROCK inhibited cells remained in G_2 for a greater duration compared to cells with ROCK function, possibly indicating roles of the Rho-associated protein kinase during the G_2/M checkpoint. ROCK inhibition has been previously linked to inhibition of Cdk proteins, essential for regulation of cell cycle progression at G_1/S transitions ***(Croft, Olsen, 2006).*** Additionally, silencing of the Cdk regulatory subunits Cks1b and Csk2 has proven to both reduce G_1 duration ***(Street et al, 2010)*** as well as decrease proliferation by arrest of cell cycle at the G_2/M checkpoint ***(Martinsson-Alzhen et al, 2008).*** Increased occurrence of Y27 treated MCF-10A cells with much longer G_2 durations supports this claim. Conversely, greater proportion of Y27 treated cells with shorter durations of G_2 when compared to non-treated cells was also observed. Whilst evidence exists for increased rate of cell cycle by Y27 inhibition of ROCK, conflicting reports suggesting reduced cell proliferation in ROCK depleted cells may explain the increased variance observed in G_2 durations from ROCK inhibited cells. Use of an alternative ROCK inhibitor, such as H1152, with greater specificity for ROCK ***(Fuentes et al, 2008)*** may further highlight the role of ROCK in G_2 and G_2/M stages, since Y27 possesses specificity for additional kinases such as Protein Kinase C (PKC). PKC is shown to also play a role in regulating cell cycle progression ***(Black & Black, 2013)***, which may explain varied results seen in this and other studies using Y27 as an inhibitor of ROCK.

Late S observations in Y-27623 treated cells

Perhaps the most unexpected finding from this experiment would be the disparity between Late S durations due to ROCK inhibition. Manual testing of Late S durations from extreme results revealed less of a correlation than the code initially suggested. There is no evidence that ROCK inhibition may affect the S/G_2 transition, hence the most likely explanation for increased Late S duration would be differences in appearance between mediums, in particular in the G2 phase. Algorithms within Ilastik were trained to identify cells with a strong contrast in PCNA intensities as Late S, hence it could be speculated that ROCK inhibition may affect PCNA localisation within G_2 cells, resulting in incorrect classification of cells within the S/G_2 transition period. The complete mechanisms underlying the S/G_2 transition are not fully understood, although given the wide range of downstream effects caused by ROCK inhibition, the possibility of S/G_2 effects due to Y-27623 inhibition of ROCK may be worth considering. Further analysis of additional wells containing both medium conditions whilst using the same Ilastik training files would be required in order to test these findings.

Future Direction and Summary

This method has proved useful for general predictions of each stage duration, however currently lacks the accuracy required to draw hard conclusions. Further additions to this code will be required in order to increase reliability of results, such as consideration of Track Number as a method to prevent underestimation of final stage durations for each cell tracked.

Whilst this technique may be specific to Cell cycle stage durations, alternate methodologies can be generated following this approach in order to suit experimental needs. A huge amount of cell data can be produced and processed in a variety of ways. The aim in this project was to convert this messy data into easily interpretable values and develop a method within python that is easily replicable by future scientists. This principle may be applicable to various other dynamic processes, such as cell migration and crowding. Understanding the factors that affect these processes could provide critical data for

development of pharmaceuticals, particularly in development of cancer treatments. Increasing the efficiency of data analysis using such methods may unlock this information in the future.

Acknowledgements

Firstly, a huge thank you to Julia Sero for guidance and feeback throughout the process, and for kindly supplying the data used in this project. Furthermore, for providing tutorials and help during familiarisation with the image analysis software used. Special thanks to Michael Zachariadis for kindly supplying a workstation during image analysis, and to Vinusan Jeyananthan for his fantastic support in troubleshooting during the coding process.

Appendix

File repository containing code:

https://github.com/boay00/code-for-ROCK-analysis

References

Amano, M., Fukata, Y., Kaibuchi, K., 2000. Regulation and functions of Rho-associated kinase. *Exp Cell Res*. 25(261) pp. 44-51

Black, A., Black, J., 2012. Protein kinase C signaling and cell cycle regulation, *Front Immunol.,* 17(3), pp. 423

Cheng, K., Agarwal, R., Mills, G., 2009. Ras-superfamily GTP-ases in ovarian cancer. *Cancer Treat Res*, 149 pp. 229–240

Croft, D., Olson, M., 2006. The Rho GTPase effector ROCK regulates cyclin A, cyclin D1, and p27Kip1 levels by distinct mechanisms, *Mol. Cell Biol,* 26, pp. 4612-4627

Fuentes, E.O., Leemhuis, J., Stark, G.B. Lang, E.M., 2008. Rho kinase inhibitors Y27632 and H1152 augment neurite extension in the presence of cultured Schwann cells. Journal of brachial plexus and peripheral nerve injury, 3, pp. 19-19

Giotre, L., Trapani, E., Trabalzini, L., Retta, S., 2013. The Ras superfamily of small GTPases: the unlocked secrets. *Methods in Molecular Biology,* 1120, pp. 1-18

Kimura, K., Fukata, Y., Matsuoka, Y., Bennett, V., Matsuura, Y., Okawa, K., Iwamatsu, A., Kaibuchi, K., 1998. Regulation of the association of adducin with actin filaments by Rho-associated kinase (Rho-kinase) and myosin phosphatase. *J Biol Chem*, 273, pp. 5542–5548

Lindley, F., 2021. The Effect of ROCK inhibition on Cell Cycle Entry and dynamics of MCF-10A cells, *BB30042,* Univerisity of Bath.

Liu, S., Goldstein, R.H., Scepansky, E.M. and Rosenblatt, M., 2009. Inhibition of Rho-Associated Kinase Signaling Prevents Breast Cancer Metastasis to Human Bone. *Cancer Research*, 69(22), pp. 8742-8751.

Liu, Y., Suzuki, Y., Day, R., Fanburg, B., 2004. Rho kinase-induced nuclear translocation of ERK1/ERK2 in smooth muscle cell mitogenesis caused by serotonin, *Circ. Res.,* 95, pp. 579-586

Lloyd, R., Erickson, L., Jin, L., Qian, X., Cheville, J., Scheithauer, B., 1999. p27kip1: a multifunctional cyclin-dependent kinase inhibitor with prognostic significance in human cancers, *Am J Pathol,* 154(2) pp. 313-323

Maekawa, M., Ishizaki, T., Boku, S., Watanabe, N., Fujita, A., Iwamatsu, A., Obinata, T., Mizuno, K., Narumiya, S., 1999. Signalling from Rho to the actin cytoskeleton through protein kinases ROCK and LIM-kinase, *Science,* 285(5429) pp. 895-898

Martinsson-Alzhen, H., Liberal, V., Grunenfelder, B., Chaves, S., Spruck, C., Reed, S., 2008. Cyclin-dependent kinase-associated proteins Cks1 and Cks2 are essential during early embryogenesis and for cell cycle progression in somatic cells, *Molecular and Cellular Biology,* 28, pp. 5698-5709

Meloche, S., Pouyssegur, J., 2007. The ERK1/2 mitogen-activated protein kinase pathway as a master regulator of the G1- to S-phase transition, *Oncogene,* 26, pp. 3227-3239

Nakagawa, O., Fujisawa, K., Ishizaki, T., Saito, Y., Nakao, K., and Narumiya, S. 1996. ROCK-I and ROCK-II, two isoforms of Rho-associated coiled-coil forming protein serine/threonine kinase in mice. *FEBS Lett*. 392, pp. 189–193

Resnitzky, D., Gossen, M., Bujard, H., Reed, S., 1994. Acceleration of the G1/S phase transition by expression of cyclins D1 and E with an inducible system, *Mol. Cell Biol.,* 14, pp. 1669-1679

Sero, J.E., and Bakal, C., 2017. Multiparametric Analysis of Cell Shape Demonstrates that β-PIX Directly Couples YAP Activation to Extracellular Matrix Adhesion. *Cell systems*, 4(1), pp. 84-96.e86

Shimizu, Y., Thumkeo, D., Keel, J., Ishizaki, T., Oshima, H., Oshima, M., et al. 2005. ROCK-I regulates closure of the eyelids and ventral body wall by inducing assembly of actomyosin bundles. *J. Cell Biol*. 168, pp. 941–953

Street, C., Routhier, A., Spencer, C., Perkins, A., Masterjohn, K., Hackathorn, A., Montalvo, J., Dennstedt, E., Bryan, B., 2010. Pharmacological inhibition of Rho-kinase (ROCK) signaling enhances cisplatin resistance in neuroblastoma cells, *Int J Oncol.,* 37(5), pp. 1297-1305

Tang, Y., Olufemi, L., Wang, M., Nie, D., 2008. Role of Rho GTPases in breast cancer. *Front Biosci,* 13, pp. 759–776

Thain, A., 2021. Development of an open-source software pipeline for the investigation of nucleocytoplasmic shuttling of transcription factors, *BB30042,* University of Bath.

Thumkeo, D., Keel, J., Ishizaki, T., Hirose, M., Nonomura, K., Oshima, H., et al. 2003. Targeted disruption of the mouse rho-associated kinase 2 gene results in intrauterine growth retardation and fetal death. *Mol. Cell. Biol*. 23, pp. 5043–5055

Vishnubhotla, R., Bharadwaj, S., Sun, S., Metlushko, V., & Glover, S., 2012. Treatment with Y-27632, a ROCK Inhibitor, Increases the Proinvasive Nature of SW620 Cells on 3D Collagen Type 1 Matrix. *International journal of cell biology*, 259142

Westbrook, L., Manuvakhova, M., Kern, F., Estes, N., Ramanathan, H., Thottassery, J., 2007. Cks1 regulates cdk1 expression: a novel role during mitotic entry in breast cancer cells, *Cancer Research,* 67, pp. 11393-11401

Zhiyuan, Y., Miao, L., Peicai, F., Minjie, X., Wei, W., Xiang, L., 2012. ROCK inhibition with Y27623 promotes the proliferation and cell cycle progression of cultured astrocyte from spinal cord. *Neurochemistry International,* 61(7), pp. 1114-1120