Advances in Quantitative Microscopy

Samuel Cooper

A thesis submitted in partial fulfilment of a Doctor of Philosophy in Stratified and Translational Medicine (STRATiGRAD), with the Department of Surgery and Cancer, at Imperial College London. Affiliated with the Dynamical Cell Systems Team, at the Institute of Cancer Research.

Supervised by:

Prof. Robert Glen, Department of Surgery and Cancer, Imperial College London

Dr. Chris Bakal, Dynamical Cell Systems Team, Institute of Cancer Research.

Abstract

Microscopy allows us to peer into the complex deeply shrouded world that the cells of our body grow and thrive in. With the emergence of automated digital microscopes and software for anlysing and processing the large numbers of image that they produce; quantitative microscopy approaches are now allowing us to answer ever larger and more complex biological questions. In this thesis I explore two trends. Firstly, that of using quantitative microscopy for performing unbiased screens, the advances made here include developing strategies to handle imaging data captured from physiological models, and unsupervised analysis screening data to derive unbiased biological insights. Secondly, I develop software for analysing live cell imaging data, that can now be captured at greater rates than ever before and use this to help answer key questions covering the biology of how cells make the decision to arrest or proliferate in response to DNA damage. Together this thesis represents a view of the current state of the art in high-throughput quantitative microscopy and details where the field is heading as machine learning approaches become ever more sophisticated.

Statement of originality

I certify that the intellectual content of this thesis is the product of my own work, or where intellectual content has been generated in collaboration the contribution of others is fully acknowledged at the point of use.

Copyright

Acknowledgements

To Bobby Glen for being a super calm, collected and reassuring supervisor, and providing tons of incredibly valuable life and career guidance. To Amine Sadok, for being a great mentor during my internship pre-PhD in the Bakal lab and during the first year too, and for giving me a Tunisian hair dryer experience if my question wasn't a 'biological question'. Also, to Amine, Vicky Bousgoni, and Chris himself (who always likes to remind everyone he took the photo's) for the experiments in Chapter 2, plus lots of guidance and help on the manuscript. To Patricia Pascual-Vargas, Stefan Heldt, Heba Sailem, and Alex Lu for putting amazing amounts of work into papers, and then putting my name on them as well (Katie Hansel and Spencer Crowder fingers crossed). To everyone else in the Bakal and Marshall lab for putting up with my general arrogance and cockiness, and still being nice, providing useful guidance and feedback, and above all else making the ICR an unbelievably fun place to work. To everyone at Imperial College CSM and on the STRATiGRAD program, especially Goncalo Correia, everyone has been fantastic friends and super supporting. To Jeremy Nicholson and Elaine Holmes, and the STRATiGRAD program for funding me for these four years, despite my work being a little tangential. To Oren Kraus, for a career defining introduction to deep-learning, let's hope the next phase of our careers is as good as the last. Also, to the SBI2 and CytoData community, for being welcoming to new people and voices. To my examiners Ben Glocker and Davide Danovi, for great comments and feedback, and who also may be the only people aside myself and supervisors who will ever read this entire thesis. To Rebecca Rickman, for being the best girlfriend ever over the last two years of this PhD and being great fun and a great friend during our year in Chris's lab, despite there being some rather stressful times for both of us. To Mum, Dad, Esme and Michael for dealing with me over 4 years of very intermittent communication, unfortunately there's no sign of that changing. To Alexis Barr for letting me work with her on an absolutely incredible project that she initiated and did the vast majority of wet-lab experiments for it was super exciting, and the results are definitely a significant step forward in science, which is way more than I was expecting from this PhD. And finally, to Chris Bakal for seeing hope and putting an immeasurable amount of time and effort into a kid who showed up to an interview in a T-shirt and tracksuits, was sick in the lab-sink from a hangover on the second day, and continues to speak largely incomprehensible English. I couldn't' have asked for a better supervisor, mentor and friend over these four years.

Table of Contents

Figure List

1 Introduction and Literature Review

1.1 Background

1.1.1 Microscopy

Microscopy has remained a staple modality for studying biology since the first description of cells by Robert Hooke in his seminal work *Micrographia* (published in 1665 by the Royal Society), that contained intricately detailed hand drawn images of natural objects invisible to the naked eye, Fig. 1-1A. Hand drawn images remained the primary means of disseminating microscopy data, until the emergence of cameras some 200 years later, and development of the first 'photoelectric' microscope by Alfred Francois-Donne. Using this photographic microscope he discovered both the microorganism *Trichomonas vaginalis*, now known to be a parasite responsible for the sexually transmitted disease Trichomoniasis, as well as being the first person to observe the enormous expansion of 'mucus goblet' cells, now termed leukocytes, in leukemia (Thorburn 1974), Fig. 1-1B. Charged coupled devices or CCD chips were developed in the Bell laboratories over the 1960's, that allow for the direct conversion of light into digital data (Smith 2010), and have since spawned a huge growth in the amount of imaging information captured around the world, now estimated to be over 70% of all data traffic (Cisco 2017). This massive growth in imaging data globally has also been paralleled by an increase in the amount of microscopy imaging data we are able to collect. Key factors driving this rise include the wide adoption of digital microscopes, as well as the development of automated microscopy systems, that in

combination with the miniaturization of imaging plates, e.g. 384 wells, now allow imaging of tens to hundreds of thousands of conditions every week (Wollman and Stuurman 2007), Fig. 1-1C&D.

Yet despite this digital revolution in the way we collect and store microscopy data, many analyses of imaging data are still performed using qualitative methods and the intuition of human experts in the field (Sailem et al. 2016; Pepperkok and Ellenberg 2006). This limits both the number of images that can be analysed and the size of experiments that can be performed, as well as introducing potential bias in the observations and conclusions of the researcher. For example, an experimenter may select fields of view, or specific cells that support a hypothesis, rather than making an unbiased assessment of the overall phenotypic effect of a treatment. At best this means important novel findings could be missed, at worst it leads to conclusions and findings that cannot be reproduced, and may mislead future studies (Caicedo et al. 2017). Thus, quantitative approaches to image analysis are critical for performing unbiased analysis of experimental conditions and yielding reproducible results.

The lack of widespread adoption of quantitative methods is likely driven by the fact that collecting quantitative data from microscopy images can often be incredibly tedious, requiring extensive manual input. Even now, biologists will often find themselves using standard digital software packages, such as Image-J (Abràmoff et al. 2004), to manually measure the length of hundreds of yeast (Smith et al. 2014), count the number of thousands of cells displaying mitotic phenotypes in cancer histopathology images (van Diest et al. 1992; Veta et al. 2015), or measure intensity changes in a fluorescently stained protein over imaging time courses that may last

several days (Barr et al. 2016). Thus, there is a desperate need for better tools and workflows that are able to automate complex and time consuming quantitative image analysis tasks and perform unbiased analysis on the ever-increasing volumes of image data now being produced.

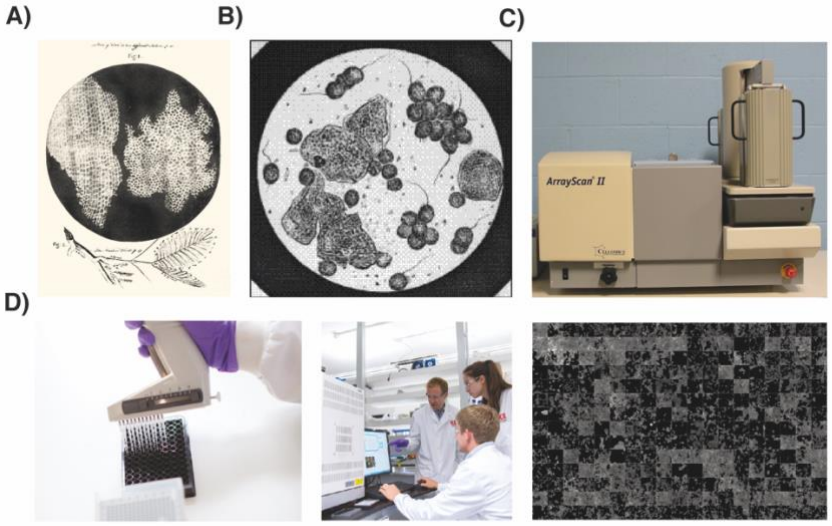

Figure 1-1, Development of high throughput microscopes: *A) Illustrations of cork tree specimens that for the first time revealed the existence of cells, that are now known to be the fundamental living unit in biology; B) The photographs of microscopic specimens, demonstrated significant over proliferation of white-blood cells in patients with blood diseases, now termed leukemias, due to excessive proliferation of leukocytes; C) One of the first automated microscopes, the ArrayScan®, was able to capture hundreds to thousands of fields of view every day, storing data digitally for quantitative analysis. D) Today standardised 96, 384 and 1536 well formats allow advanced robotic platforms, coupled to sophisticated high-content microscopes, to prepare and image stacks of microplates collecting hundreds of thousands of images and equating to several terabytes of data, every day.*

1.1.2 Automated image analysis

The earliest use-cases of automated analysis of microscopy images emerged from studies that looked to drive automation and accuracy improvements in the detection of cancer cells from pappenheim-stained blood, or vaginal swab smears, such that population wide screening of leukemia, or cervical cancers, respectively could be performed (Bostrom et al. 1959; Harms et al. 1979). These approaches progressed over a period of a decade or so, from classification of one-dimensional readouts from electronic scanners (Bostrom et al. 1959; Spencer and Bostrom 1962), through early image based classifiers developed using custom optical scanning techniques (Prewitt and Mendelsohn 1966), to more advanced techniques that made use of computers to segment regions, before subsequent extraction and processing of basic features such as area and intensity from images captured using early digital microscopes (Harms et al. 1979; Harms et al. 1986; Jaggi et al. 1988).

The use of quantitative image analysis applied to molecular biology research first arose in the analysis of hybridisation between fluorescent reporters and DNA for genetic applications, termed fluorescent in-situ hybridisation (FISH) (Nederlof et al. 1992). It wasn't then until the pharmaceutical industry got heavily involved that automated image analysis for molecular biology research became wide spread. Here, researchers began to develop secondary cell-based *in vitro* assays to validate hits emerging from increasingly high-throughput biochemical target-binding screens (HTS), that often had more complex endpoints. For example, Htun et al. labelled the glucocorticoid receptor with GFP fluorescent reporter to identify agonists that would induce this protein to translocate from the membrane to the nucleus, as was known

to occur following binding of glucocorticoids (Htun et al. 1996). This assay required imaging data to be collected and the nuclear and membrane regions to be identified using separate stains and thresholding operations on the images. The ratio of membrane protein intensity to nuclear intensity could then be calculated and the level of glucocorticoid receptor translocation quantified (Htun et al. 1996; Giuliano et al. 1997). This analysis made use of early ArrayScan software developed by Cellomics, Fig. 1-2A. In another key example, voltage sensitive dyes and reporters were also used as an imaging readout of membrane potential. Here, image analysis steps consisted of identifying and quantifying intensity differences between regions on the inside and outside of the membrane (González and Negulescu 1998; Epps et al. 1994). This allowed the identification of a small molecule activator of ATP sensitive K+ ion channels in vascular smooth muscle cells. By interfering with membrane potential this small molecule, in turn, inhibited chloride transport and reduced vascular smooth muscle contraction (Holevinsky et al. 1994). Taken together, over the 1990s, an explosion in the use of high-throughput imaging approaches for research occurred; this was driven by novel screening approaches in the pharmaceutical industry.

Typically however, such early automated image analysis pipelines remained tailored for a specific phenotypic assay, or readout, such as cell proliferation, or viability (Boutros et al. 2004). In key work by Murphy et al. it was shown that by extracting larger numbers of features describing the shape and texture properties of regions defined by fluorescently labeled proteins, accurate classification of the localisation of the protein to a cellular compartment, such as the endoplasmic reticulum, could be performed automatically (Boland and Murphy 2001; Murphy et

al. 2000). This study therefore paved the way towards using high-dimensional morphological profiles to classify and predict biological behaviour from images. For example, Perlman et al. demonstrated that high-content imaging profiles from compounds with a known mechanism of action could be used to predict the mechanism of action of an unknown compound using its phenotypic profile (Perlman et al. 2004). That high-content phenotypic profiles can be used to infer mechanism of action has now been replicated in a number of retrospective studies since this first work (Young et al. 2008; Abassi et al. 2009; Kraus et al. 2016; Towne et al. 2012). Most notably, in recent work by Simm et al., prospective predictive power of such an approach has been shown. Specifically, the authors demonstrated that by cherry picking a set of small molecules from a large HTS compound screening library, 250-fold enrichment in the likelihood target binding could be achieved by using a previous high-content screen to predict target binding (Simm et al. 2018). Relational studies also emerged that looked to infer genetic, physical, or functional interactions between genes and/or proteins. For example, Moffat et al. demonstrated that siRNA gene depletions that resulted in similar image based phenotypic profiles, would likely fall within the same biological signalling pathway (Moffat and Sabatini 2006). In another landmark study, Bakal et al. found that RNAi depletions that resulted in similar morphological characteristics in Drosophila BG2 cells, led to validated predictions as to a protein's function, localisation, and binding partners (Bakal et al. 2007).

This increasing use of large arrays of single cell features extracted from high-content screens, termed phenotypic profiles, also motivated the development CellProfiler, an open-source software package that could be broadly applied to high-

content imaging data (Carpenter et al. 2006; Lamprecht et al. 2007; Jones et al. 2008). This software continues to be used today to analyse both chemical and genetic screens, Fig. 1-2B. For example, Shan et al. more recently performed a high-content screen of small molecules against primary human hepatocytes looking for small molecules that could either induce their proliferation or cause them to differentiate further towards mature hepatocytes as detected by a phenotypic profile extracted from imaging data. The authors identified a handful of compounds that could perform each task and thus represented important tools for generating a ready supply of human hepatocytes for research and potential future cell-therapies (Shan et al. 2013). Commercial software has also kept pace and driven perhaps even greater developments in the high-content imaging field. The Columbus software package is used by a huge number of pharmaceutical and research labs around the world (Zanella et al. 2010), Fig. 1-2C. Enabling such a diverse array of high-content imaging studies as profiling compound toxicity based off morphology and proliferation markers (Martin et al. 2014), screening of potential chemotherapies against live tumour spheroid models (Reid et al. 2014), studying the infection of macrophages by parasites (Aulner et al. 2013), and even the detection and characterisation of biological weapons of mass destruction (Peruski et al. 2002). Thus, by extracting high-dimensional phenotypic profiles from imaging data, screens can now be performed with complex endpoints, that may involve either the determination of multiple phenotypes, or relational studies that look to correlate the similarity between different conditions in order to identify the connection between different conditions. These studies together define the common high-content profiling workflow used by the majority of groups today (Caicedo et al. 2017).

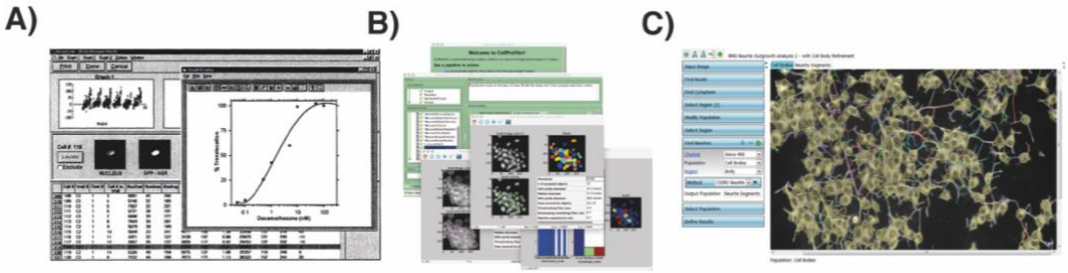

Figure 1-2, Example interfaces developed for analysis of high-content imaging data: *A) Software that accompanied the ArrayScan® microscope allowed research to use simple thresholding operations to identify regions that intensity readings could be collected from. From these measurements, dose response curves and other key assay readouts could be plotted; B) CellProfiler represents perhaps the first open source software for performing high-content image analysis, that continues to be developed today, this allows pipelines to be created that can be used to process and extract thousands of single-cell features from very large, 10,000 conditions/1,000,000 image, screens. C) Perkin Elmer maintain the Columbus™ software that remains perhaps the most popular and easy to use software developed for analysis of high-content imaging data. Here a simple plug and play, format allows you to drag and drop functions to create an imaging pipeline for high-content analysis.*

1.2 The high-content analysis workflow

Currently, high-content image analysis workflows are generally tailored to a given screen being analysed, due to the unique combination of the cell line and fluorescent labels used, as well as the biological questions being asked of the dataset. However, across researchers in the field, a set of common themes and steps emerge. Together these underpin the standard workflow that includes the most common process used by labs today in the field of high-content imaging, Fig. 1-3. Although the exact number of steps is trivial, the process typically involves: 1) image-capture and storage 2) preprocessing of imaging data; 3) object detection and segmentation; 4) feature extraction; 5) feature reduction, processing and aggregation; and 6) Analysis and visualisation of results (Caicedo et al. 2017; Swedlow et al. 2009; Boutros et al. 2015; Carpenter 2007). Prior to studying how different conditions affect phenotypic profiles, analysis of single cell data can also be performed to identify and analyse subpopulations of cells, such as those in different stages of the cell cycle (Neumann et al. 2010; Jones et al. 2008; Mukherji et al. 2006). Additionally in live cell experiments, after extraction of object coordinates and other features, cell tracking and time-series analysis can be performed (Cooper and Bakal 2017). Finally, deep-learning strategies have now been developed that facilitate direct conversion of raw data to visualisations and screening results (Kraus et al. 2017; Godinez et al. 2017).

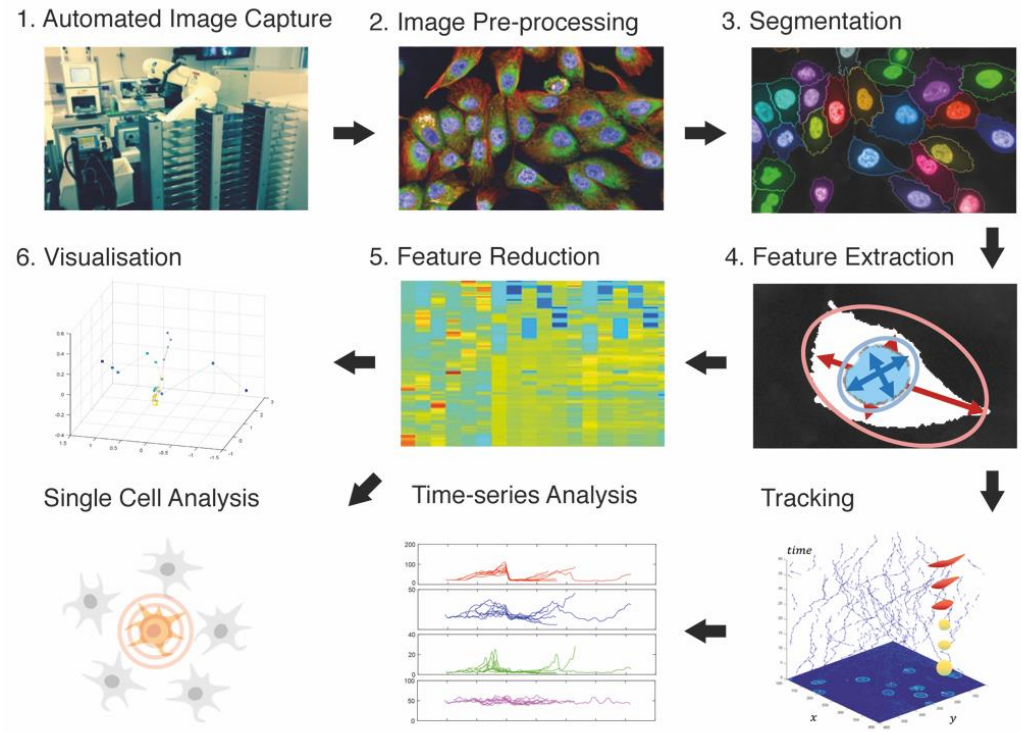

Figure 1-3, A standard high-content microscopy workflow: *1) Cells are automatically prepared and imaged using robotic plate handling and high-content automated microscopes; 2) Image preprocessing steps look to normalise images across microplates and reduce batch effects; 3) Segmentation steps involve detection of objects and identification of unique binary masks that cover each region; 4) Hundreds to thousands of features describing each region are extracted from segmented cells; 5) Features are typically reduced to non-redundant spaces, using either feature selection or transformation algorithms; 6) Extracted features can then be explored and visualised using a range of graphics and clustering approaches.*

1.2.1 Image Capture and Storage

The first step following the fixing and staining of cells in a high-content screen is to image the screen and store images in an accessible format for processing. Typical high-content microscopes can capture tens to hundreds of thousands of images every day using automation to supply microplates for imaging. Managing the huge amount of data that is produced can represent a significant challenge. A number of solutions exist for handling this data and are typically commercial. However, in a notable example Swedlow et al. have developed the Open Microscopy Environment (OME) an open-source software solution that is able to host and visually display image based screening data (Swedlow et al. 2003; Li et al. 2016; Allan et al. 2012). However, setting up and maintaining such servers remains challenging for most labs, thus commercial packages are typically used such as those marketed by PerkinElmer as the Columbus™ platform.

1.2.2 Image preprocessing

Image preprocessing involves clipping and illumination correction algorithms, that seek to ensure all conditions, and cells, being imaged are represented and treated equally over the entire imaging screen. Clipping, or the removal of high-intensity outlier pixels is performed as microscopes will often capture a very wide range of pixel intensities that may include artifacts, such as aggregated clumps of immunofluorescent probes. These can distort subsequent steps such as the calculation of intensity thresholds, aggregation of single cell data, or illumination correction.

Illumination correction, in turn, accounts for variations in the background image intensity over the field of view; these are generally an inherent property of the light path between the illumination source and camera. A number of strategies have therefore been devised to reduce this source of variation: 1) *Prospective correction:* Background images are captured with no sample present, and negated from future images (Singh et al. 2014); 2) *Single image correction:* A very wide filter, such as a gaussian blur, is applied to the image to give a rough intensity map that is then negated from the original image (Likar et al. 2000; Babaloukas et al. 2011); and 3) *Multiple image correction:* Average intensities are taken over the entire screen, and then negated from every individual image (Singh et al. 2014; Babaloukas et al. 2011). Generally, prospective methods and multi-image correction are preferred as adjustment of intensity values within single images can reduce the true range of intensity changes captured over an imaging dataset (Caicedo et al. 2017; Singh et al. 2014).

1.2.3 Object detection and Segmentation

In object detection and segmentation steps, the goal is to identify regions of the image that correspond to objects of interest. A vast amount of research has gone into developing better algorithms for these tasks (Pal and Pal 1993; Haralick and Shapiro 1985). Many of these approaches have been transferred to nuclei and cell segmentation in high-content imaging. In the most straightforward approach objects can be detected with simple thresholding operations, for example Otsu's method (Otsu 1979). However, these will often classify multiple touching nuclei and/or cells as a single object. Cell segmentation is thus typically split into two steps:

1. The centers of nuclei are found. As these appear as round regions of intensity, consecutive radial blurring and erosion/dilation operations are often sufficient to transform the image such that a single intensity peak corresponds to each nuclei center, peak picking can then be used to detect nuclei centers (Carpenter et al. 2006). Other algorithms can also be used to detect nuclei where they may be touching, although are less common. For example, by assuming nuclei are round, a Hough transform, that is designed to detect circular objects, can be used to identify candidate nuclei (Thomas et al. 1992). Model based approaches, have also demonstrated high levels of performance (Molnar et al. 2016).
2. A mask of the nuclei/cell region must then be created based off the detected centers. By far the most common approach here is the watershed algorithm (Beucher and Meyer 1992; Roerdink and Meijster 2000), that treats cell nuclei and background markers as valleys, and cell edges as ridges and then expands regions to cover the space between ridges (Malpica et al. 1997), [described in more detail in Chapter 3]. However, higher levels of performance have been shown (Meijering 2012) by using both graph-cuts (Shi and Malik 2000)(Al-Kofahi et al. 2010) and learnt textures (Sommer et al. 2011). More recently, deep-learning approaches such as fully convolutional neural networks (Long et al. 2015), implemented as U-net for biomedical imaging data (Ronneberger et al. 2015), mask R-CNN (He et al. 2018), and learnt watershed algorithms (Wolf et al. 2017; Bai and Urtasun 2017) have achieved human-level performance in nuclei segmentation challenges.

1.2.4 Feature Extraction

Following detection of cells and segmentation of regions, or masks, that define these objects features are then extracted that describe the shape of the cell, as well as the texture and intensity properties of different fluorescent markers inside the cell mask (Boland and Murphy 2001; Carpenter et al. 2006). Many features important for studying the shape of cells were described by Boland and Murphy, and include measures of size, eccentricity, and membrane/edge smoothness (Boland and Murphy 2001). These measures can also be performed on geometrically defined sub-regions of the overall cell, such as protrusions that extend outward from the main cell body (Sailem et al. 2014; Yin et al. 2013). Features based on channel intensity and texture are also typically extracted and include: *Haralick features;* these analyse the relationship between neighbouring pixels (Haralick et al. 1973); *Gabor features;* that measure the intensity of a region following transformation with kernels tailored to find textures such as edges, spots and ridges (Jain and Farrokhnia 1991); and *Zernike moments;* a type of statistical moment that describes the distribution of intensity over a specific region (Khotanzad and Hong 1990; Zernike 1934). More recently features have also been engineered that describe the environment a cell is in with respect to other cells, such as stretches of cell membrane that are in contact with other cells, termed neighbour fraction, (Sero et al. 2015), or the spatial location of a cell relative to the center of clusters of cells, or other cells in a locally connected neighbourhood (Snijder et al. 2009).

In high-content analysis the emphasis is upon building a broad and effective description of segmented regions, as opposed to specific features that have been

engineered into the assay, such as intensity (Boutros et al. 2015; Caicedo et al. 2017). However, in some cases features must be engineered into the design of the analysis to reflect the proteins being fluorescently labelled; for example a number of proteins translocate into the nucleus on activation. Features that quantify the ratio of nuclear intensity levels versus intensity in a perinuclear ring region, are more robust for example, than those that give the ratio of nuclear to cytoplasmic intensity (Sero and Bakal 2017). Overall tens to hundreds of features can be extracted from each region of the cell; concatenated together these features should form a 1-dimensional vector that effectively describes the morphology of a single cell.

1.2.5 Feature aggregation and reduction

Where very large numbers of features have been extracted, high levels of correlation and redundancy will often exist between features. For many analysis methods, such as determining the similarity between phenotypic profiles of different treatment conditions, having large groups of highly correlated features presents challenges. For example, in identifying effects that may be strong and highly relevant, but only present in a small number of features, versus correlated noise that may be present in tens to hundreds of features. Thus, identifying non-redundant feature spaces represents an important step in preparing high-content data sets for exploratory analysis (Caicedo et al. 2017).

There are two approaches that are typically used to reduce the number of features that are used to describe a high-content imaging dataset, and both lend from large amounts of research that have been carried out on this task in other fields:

1) **Feature selection:** In feature selection approaches, algorithms seek to iteratively select an optimal set of features. In one example of a greedy approach, the most correlated pair of features in the dataset is identified and the feature in this pair that is most correlated with the rest of the dataset is then removed, this is iterated until a minimum correlation threshold is reached (Woehrmann et al. 2013). Minimum-redundancy maximum-relevance selection have also been tested (Ding and Peng 2005; Ng et al. 2010). In more advanced algorithms, combinatorial optimisation methods such as genetic algorithms can be used to optimise against metrics such as Akaike or Bayesian information criterion (Akaike 1998; Schwarz 1978), with a penalty for features number.

2) **Linear transformation:** Linear transformation methods in contrast look to rotate or transform the data such that a reduced number of features capture the majority of information in the dataset. The most common transformation method is principle component analysis (PCA). In PCA variance is maximised in successive orthogonal dimensions (Hotelling 1933). Meaning that, following transformation, dimensions are linearly uncorrelated with the first principle component capturing the most variance in the dataset, followed by the second, and so on. Typically, PCA is performed by singular value decomposition of the matrix of cell or well-level data following mean centering of feature distributions. Factor analysis where non-orthogonal linear combinations of features that represent frequent patterns in the data are found, and linear discriminant analysis where projections of the data that maximise separation between positive and negative controls have also been tested in high-content workflows (Young et al. 2008; Kümmel et al. 2010).

Following reduction of feature spaces aggregation of single cell data to well-level readouts is performed, where each well corresponds to an experimental condition and a unit of a microwell plate. This converts an array of data for each well, where rows represent single cell objects and columns represent features, into a single 1-dimensional vector for that well. Multiple aggregation techniques exist, though typically mean or median feature values are taken across the population of cells (Caicedo et al. 2017). These have consistently performed well (Ljosa et al. 2013), although more advanced measures such as divergence from control population values as measured using the KS-statistic have also been used (Perlman et al. 2004). Typically multiple wells will exist for each condition, however these will not be aggregated prior to downstream analysis, as these constitute technical replicates of each condition and thus contain valuable information on the reproducibility of a given condition. Aggregation may also occur prior to feature reduction in certain cases.

1.2.6 Data set exploration and visualisation

The major goal of most high-content imaging workflows is to understand how the effect of different conditions relate to one another, and to control data. This relational understanding is achieved through similarity metrics that allow comparison and hierarchical clustering of different conditions, a strategy that's typically used with larger non-redundant feature spaces. The two most common similarity-metrics that are used are: 1) Pearson's correlation, this represents how similar two different profiles are based on the features values of both conditions (Schulze et al. 2013; Singh et al. 2015); and 2) Euclidean distance, this gives the total distance between the two data points in Euclidean space, (Adams et al. 2006; Woehrmann et al. 2013). Rank

correlation metrics and other metrics have been tested but are typically less common, and only used in cases where standard normalisation and feature reduction strategies have been used (Reisen et al. 2013; Ljosa et al. 2013).

Where a smaller number of dimensions contain the majority of information, it can often be easier just to plot well-level data points directly and interpret results. High-dimensional visualisation techniques, such as t-Stochastic Neighbor Embedding (Maaten and Hinton 2008) and Isomap (Tenenbaum et al. 2000), that look to identify manifolds and unwrap these into lower dimensional spaces have also been explored, as in (Amir et al. 2013). However, these approaches can generate data structures that may be over-interpreted without care (Balasubramanian and Schwartz 2002). Typically the choice of visualisation technique will depend on the biological question being asked in the experiment or screen.

1.3 Single cell analysis

Aggregating populations of single cell features into readouts that describe the overall effect of a condition typically works very well for identifying hits that are similar or dissimilar to controls, as well as to relate conditions to one another (Caicedo et al. 2017). However, such aggregation techniques disregard the presence of potential subpopulations of cells, that may point towards exciting new biology that can be explored (Snijder et al. 2012), and may also significantly aid biological interpretation of high-content imaging datasets (Snijder and Pelkmans 2011; Altschuler and Wu 2010). This includes understanding shape (Fuchs et al. 2010; Bakal et al. 2007; Yin et al. 2013; Sailem et al. 2014), cell-cycle phase (Neumann et al. 2010;

Jones et al. 2008; Mukherji et al. 2006), and even the likelihood of viral infection (Snijder et al. 2009). As well as developing a better understanding of biological processes, subpopulation identification can also provide important clinical insights. For example, a high-content analysis of specific signaling and morphological subpopulations, in H460 lung cancer cells, demonstrated that some groups were more likely to be specific for resistance to paclitaxel, versus others, due to a distinct signalling state (Singh et al. 2010).

Typically in identifying subpopulations of cells in high-content imaging screens, manual qualitative inspection of conditions is performed and alongside prior knowledge, is used to identify examples for each population (Kiger et al. 2003). These in turn are used to train classifiers for these subpopulations that can then be applied to the dataset as a whole (Fuchs et al. 2010; Eggert et al. 2004; Neumann et al. 2006; Loo et al. 2007). For example Bakal et al. used a weakly supervised approach to identify reference shapes, in an RNAi screen for modulators of drosophila BG-12 morphology. Specifically, 7 treatments were identified that qualitatively produced cell morphologies that were highly visually distinct from control conditions. Neural networks were then trained on extracted features that sought to classify cells into each of these conditions, and thus learnt the dominant morphology in each of these 7 conditions (Bakal et al. 2007). A similar technique was also used to study the organisation of Rho GTPase signalling pathways in drosophila cells, using a single cell approach in which different Rho GTPases were depleted using siRNA (Nir et al. 2010). Classification of cells into distinct morphologies has also been used to map signalling networks using double knockout siRNA screens where network connections were

inferred from non-additive effects, calculated using a pairwise interaction matrix (Horn et al. 2011). Finally, where subpopulations have been manually identified on a limited group of cells, online learning methods have also been used to expand the number of subpopulations to include additional groups not qualitatively described in the original training set (Jones et al. 2009; Yin et al. 2007). Classification of single cells into distinct subpopulations, can therefore be used as an effective strategy to aid interpretation of the effect of experimental conditions and in the inference of genetic interaction and signalling networks.

However, such supervised approaches still introduce bias into dataset analysis, and also require extensive manual input in defining the presence of subpopulations. Unsupervised analysis in turn offers the potential to quickly identify subpopulations, without also introducing potential bias of the experimenter. In analysing a screen of genes that regulate morphology and cytoskeleton dynamics in melanoma cancer cells plated upon collagen matrices that mimic physiological condition I therefore explored how unsupervised approaches could be used to test whether or not discrete subpopulations of cells exist within a population or whether a more continuous set of forms are present. This analysis being the subject of Chapter 2.

1.4 Live cell imaging

Traditionally, whilst high-content workflows were engineered by pharmaceutical companies and academic groups to screen large libraries of drugs and genetic conditions, live-cell imaging approaches have been developed by academics with a focus on understanding the dynamics of specific biological pathways, and how information is propagated through these pathways (Purvis and Lahav 2013). In live cell-imaging experiments, throughput is limited by the fact that tens to hundreds of images may need to be captured, for each and every condition, to build up an accurate profile of how any given gene may be behaving over time (Cooper and Bakal 2017). Moreover, the number of fluorescent reporter genes and non-toxic dyes that can be employed has also been traditionally limited to only a handful of genes within mammalian systems.

However, with advances in gene editing capabilities an increasing number of live cell reporters are now available that can quickly be introduced into cells and provide accurate readouts of the localisation and expression levels of key proteins, using only a single channel (Gaudet and Miller-Jensen 2016). Meaning that, the gap in throughput and diversity of protein readouts captured in a single experiment is closing between high-content analysis and live cell imaging. For example, Stewart-Ornstein et al. used CRISPR gene editing technology to engineer a system, termed eFlut, that allows efficient tagging of proteins with fluorescent reporters at their endogenous loci, thus providing a detailed readout of protein expression levels (Stewart-Ornstein and Lahav 2016), Fig. 1-4A. Meanwhile, Regot et al. developed kinase translocation reporters, that possess a target sequence engineered to be

homologous to the consensus binding sequence of a specific kinase. Upon phosphorylation of this target sequence, these reporters translocate into the nucleus and therefore act as a readout of the upstream kinases' activity (Kudo et al. 2018; Regot et al. 2014), Fig. 1-4B. These one channel reporters are much more amenable to high-content type live imaging screens, as they are able to capture the effects of multiple signaling pathways, whilst leaving channels free for automated nuclei detection and segmentation. This is in contrast to more traditional readouts of signalling activity such as Forster Resonance Energy Transfer (FRET) reporters, Fig. 1-4B. With FRET reporters, signalling activity is quantified by the ratio of emission from a donor fluorescent protein to an acceptor fluorescent protein, that may be connected (Zhang and Allen 2007). When donor and acceptor proteins/domains are in close proximity, the donor absorbs at a higher wavelength, and resonance transfer shifts this energy to the acceptor which in turn can then emit light at a lower wavelength. A number of different strategies can then be used to link signalling activity to proximity changes (Fritz et al. 2013). However, these reporters occupy two different wavelengths, or channels, in imaging experiments. This limits their effectiveness in studies that look to relate multiple reporters. As such, advances in reporter technology have facilitated automated imaging and tracking of entire populations of cells, in combination with the ability to measure multiple readouts of signaling activity.

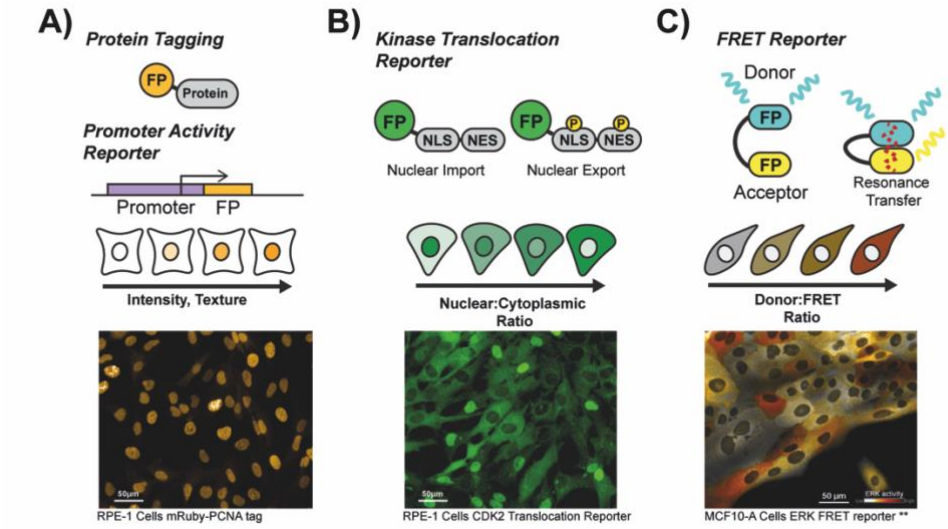

Figure 1-4, Fluorescent Sensors for Signalling Dynamics (Cooper and Bakal 2017): *(A) The levels and localisation of a protein may be determined by linking a fluorescent protein (FP) to the N or C terminus of the protein and measuring fluorescence intensity. Ideally this is done at the endogenous loci, such that it is under control of the natural promoter (B) Förster resonance energy transfer (FRET) reporters make use of resonance energy transfer between donor and acceptor proteins where resonance transfer has been engineered to change as a result of biological signalling. Activity levels are quantified as the ratio of donor emission intensity to acceptor emission intensity. (C) Studies characterising the behaviour of proteins that translocate out of the nucleus upon phosphorylation have led to the development of kinase activity reporters (KTRs). Kinase activity is then recorded as the cytoplasmic: nuclear (C:N) fluorescence ratio (Kudo et al. 2018; Regot et al. 2014). ** Reproduced, with permission, from John Albeck (C).*

Significant improvements in auto-focus, stage control, and incubation technologies have also increased the number of fields of view that we can capture in a single experiment as well as the length of time cells can be imaged for (Pepperkok and Ellenberg 2006). Yet, these improvements have now created major bottlenecks in tracking cells over long periods of time, especially around automating entire tracking pipelines and integrating different steps in data analysis (Coutu and

Schroeder 2013). In development of a tracking workflow, typically the canonical high-content workflow of image preprocessing, object detection and segmentation, and feature extraction is adopted as described previously (Cooper and Bakal 2017). However, notable variations on this pipeline occur, for example, frequently tracking is performed on bright-field images which can introduce additional challenges (Kerz et al. 2016; Buggenthin et al. 2013; Olivier et al. 2010). Following these stages an additional step is then included that seeks to track cells, or more commonly cell nuclei, over the period of imaging, Fig. 1-3. Yet, current solutions to live cell tracking typically require independent software packages to be used for segmentation, tracking of cells, and correction of tracks. This can create issues, and reduce throughput, for the majority of biologists who do not have experience in building software pipelines. This can be especially problematic when different software packages may use different programming languages and/or import and export images data in varying types of file format. To improve throughput and ease of use, new software tools are desperately needed for to enable simple tracking of cells and extraction of time-series data. This motivated the development of NucliTrack, a cross-platform package that allows segmentation, features extraction, tracking, and correction of tracks from live cell datasets in a single application. Development of this application is detailed in Chapter 3.

By using higher throughput live cell approaches, we can also start to ask the question of whether subpopulations of cells exist that display distinct patterns of signaling behaviour. For example, in heterogeneous populations, where does a continuous set of behaviours become two discrete populations, defining a bifurcation

point in dynamics, that can for example determine a cell fate decision. In key studies for example, Spencer et al. identified a bifurcation point following mitosis in CDK2 activity that governs a cell decision to proliferate or quiesce at the restriction point (Spencer et al. 2013). Studying the dynamics of how bacteria enter a competent state Süel et al. also identified a bifurcation point that underpins where a bacteria enters a limit cycle of competency, before transitioning back to a proliferative state (Süel et al. 2006). Finally, previous work by Barr et al. demonstrated hysteresis at the G1/S transition in Hela cells, governing their ability to commit to DNA synthesis or remain in a G1 state (Barr et al. 2016).

Following on from these studies we used a live high-content approach to explore how p21 modulates the decision point between arrest and quiescence in G1 phase in cells in response to DNA damage. Here using a combination of reporters, most importantly p21 tagged with GFP at its endogenous locus in hTert-RPE1 cells. Specifically, we found that two double negative feedback loops, between the cell cycle inhibitor p21 and the E3 ubiquitin ligases CRL4-Cdt2 and SCF-Skp2 underpin a bi-furcation point between proliferation and arrest that occurs at the restriction point in populations of hTert-RPE1 cells in response to endogenous replication stress occurring in mother cell S-phase (Barr et al. 2017). This work being the subject of Chapter 4.

1.5 Deep-learning strategies for analysis of high-content data

Shortly prior to, and during the course of this thesis, the field of deep-learning was rapidly emerging as a major new force in machine-learning and data analysis (LeCun et al. 2015). This was driven by both significant improvements in processors, known as Graphics Processing Units (GPUs), that are able to very efficiently handle large matrix operations such as those required to train and evaluate deep-neural networks, as well advances in software for managing such operations (Abadi et al. 2016; Bergstra et al. 2010). Together, these improvements allowed researchers to go from achieving superhuman performance on toy benchmark challenges such as classifying handwritten digits (LeCun et al. 1998), to such levels of accuracy on large image databases consisting of thousands of images that may have dimensions of hundreds to thousands of pixels (Krizhevsky et al. 2012), as well as segmenting these images (Long et al. 2015), detecting objects within them (Girshick 2015; Ren et al. 2017).

This ability to handle larger images such as those captured by microscopes has driven the transfer of such deep-learning approaches to tasks in quantitative microscopy (Kraus and Frey 2016). For example, in early work Ronneberger et al. demonstrated how a fully convolutional neural network architecture, U-net, could be used to achieve state-of-the-art performance for segmenting cell images using brightfield and fluorescence microscopy in a key benchmark challenge (Ronneberger et al. 2015). Moreover, Kraus et. al. demonstrated how correct pooling of information captured from a deep convolutional neural network could be used to accurately classify both protein localisation and compound mechanism of action. This used an approach termed multiple instance learning, where the network was trained on data

that had examples of proteins with known localisation, and compounds with known mechanism of action (Kraus et al. 2017; Kraus et al. 2016). Godinez et al. also applied deep-neural networks to classifying phenotypes using deep-convolutional neural networks in an early approach (Godinez et al. 2017), and transfer of features learned on large imaging datasets has also shown promise as a way of improving our ability to detect cell phenotypes (Pawlowski et al. 2016).

Currently, however deep-learning strategies in microscopy require the use of labelled control data to train neural networks (Ching et al. 2018). However, often in high-content screens new phenotypes may be present. When faced with new phenotypes, neural networks trained on control examples will incorrectly classify them. This motivates the need for deep-learning strategies that can build meaningful hierarchical representations of imaging datasets without prior knowledge of classes. Deep convolutional neural networks, known as autoencoders, learn to compress images into lower dimensional spaces, known as an embeddings, and then reconstruct them in a way that minimises the reconstruction error as compared to the original image (Hinton and Salakhutdinov 2006; Vincent et al. 2010). Such autoencoders can learn embeddings that when clustered lead to results that match ground truth class labels; this demonstrates that meaningful hierarchical representations of imaging datasets can be learnt (Hinton and Salakhutdinov 2006; Vincent et al. 2010). However, in contrast to many images, cells on 2D substrates are translationally and rotationally invariant, display a diverse range of morphologies that may be independent of experimental conditions, and often contain large batch effects. As such when applied to high-content images autoencoders typically learn

embeddings that capture batch effects, or meaningless information on the spatial localisation of cells.

This motivates the current line of research into using weakly-supervised deep-neural networks to learn meaningful embeddings, that can then be clustered using traditional unsupervised approaches (Michael Ando et al. 2017; Caicedo et al. 2018). The ability to correctly classify technical replicates of a condition versus all other conditions also represents a promising weakly-supervised approach to extracting biologically meaningful embeddings that unsupervised clustering can then be performed on. Some of the concepts developed in this thesis as well as the work of Kraus et al. (Kraus et al. 2017; Kraus et al. 2016), have contributed to very recent work by Lu et al. who used such a weakly supervised approach to demonstrate state-of-the-art performance when classifying protein localisation at the single cell level (Lu et al. 2018), using a technique that can be broadly applied to the task of screening for new phenotypes in high-content screens. Thus, the transfer of deep-learning strategies to high-content microscopy, will likely underpin the next wave of innovation in the field, and is discussed briefly in Chapter 5, Summary and Future Directions.

2 Characterising heterogeneity in cell shape dynamics

2.1 Introduction

Melanoma is a metastatic and highly aggressive form of skin cancer that arises from melanocytes. The aggression of melanomas stems in part from their ability to effectively invade neighbouring tissues, enter circulation, and ultimately disseminate to distant sites around the body. At the single cell level melanoma tumorigenesis is associated with transition from an epithelial state to a mesenchymal state, known as epithelial to mesenchymal transition (EMT). In over half of all cases this transition is associated with mutations in the driver genes BRAF and NRAS (Heppt et al. 2017), that together drive proliferation and transformation of melanocytes into a cancerous state (Caramel et al. 2013).

For a long time EMT was associated with a single transition to an adherent motile phenotype. Though we are now aware that melanoma cells, and cancers more generally, display a wide range of migratory modes. Each of these are typically associated with varying substrate and/or stiffness of the extracellular matrix (Friedl and Wolf 2003). For example, melanoma cells cultured in soft 3D collagen matrix will adopt an adhesion independent, amoeboid phenotype, where migration is driven by contraction and membrane blebbing, Fig. 2-1A. This allows the cell to squeeze through gaps in the extracellular matrix (Sahai and Marshall 2003; Sanz-Moreno et al. 2011). In contrast, melanoma cells cultured on stiffer substrates, or *in vitro* on plastic, adopt a more typical mesenchymal form with migration characterised by rounds of

protrusion, adhesion, and retraction (Parsons, Horwitz, and Schwartz 2010; Yin et al. 2013), Fig. 2-1B. Where cancer cells are unable to modify their migratory mode, in changing environments, traction and forward movement are severely compromised (Liu et al. 2015; Tozluoğlu et al. 2013; Ruprecht et al. 2015). The ability of cancer cells to adapt to environments of varying stiffness therefore represents a key factor in their ability disseminate through the body and metastasise.

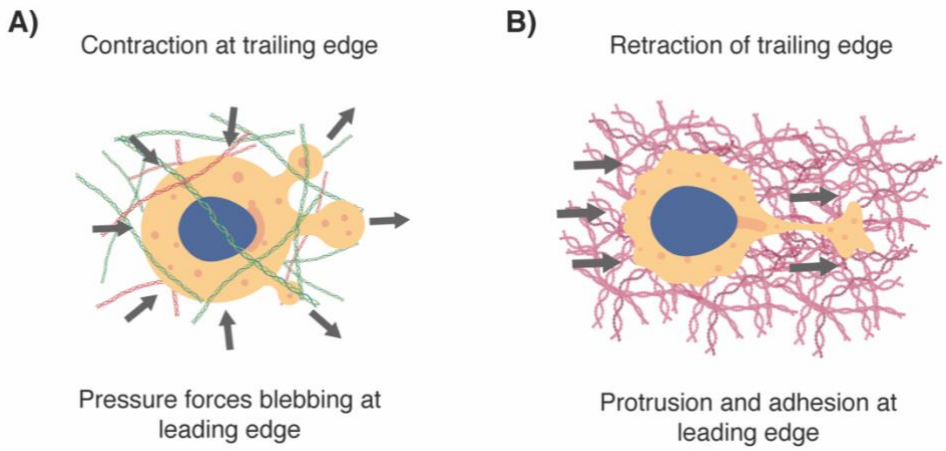

Figure 2-1, Melanoma cells adopt different migratory modes in vivo: *A) In soft extracellular matrix melanoma cells have been found to adopt an amoeboid migratory mode that involves contraction of the cell body. This drives blebbing of the membrane and exploration of new areas that the cell and nucleus can squeeze into. B) On stiffer ECM matrices cells migrate through mesenchymal forms that typically involve cyclic rounds of protrusion at the leading edge, adhesion to the extracellular matrix, and then retraction of the leading edge; this generates force that pulls the nucleus through the matrix.*

At the molecular level different migratory modes are driven by distinct signalling pathways. Chief among the regulators of migration, cell shape, and cytoskeletal dynamics are Rho family GTPases (Sadok and Marshall 2014; Hall 1998; Etienne-

Manneville and Hall 2002). Rho GTPases are activated when bound to GTP and exert their effects through protein-protein interactions. Binding of a guanine activating protein (GAP) induces intrinsic hydrolysis of GTP to GDP and deactivation of the GTPase, subsequent binding of a guanine exchange factor (GEF), catalyses the exchange of GDP for GTP re-activating the Rho GTPase (Hall 1998; Etienne-Manneville and Hall 2002), Fig. 2-2A.

The Rho GTPases, RhoA, Rac1, and Cdc42 represent the best studied Rho family GTPases (Ridley 2012; Nobes and Hall 1995), Fig. 2-2B. RhoA's major role in regulating cell morphology is associated with the induction of contraction, through activation of the effector protein Rho-associated kinase (ROCK) (Matsui et al. 1996; Amano, Fukata, and Kaibuchi 2000). ROCK promotes stabilisation of actin filaments by phosphorylation of the cytoskeleton modulator LIM kinase, an inhibitor of the actin polymerisation protein, cofilin (Maekawa et al. 1999). ROCK also phosphorylates and activates myosin light chains that generate force and drive contraction of actin filaments, overall leading to the development of contractile bundles and creation of tension within the cell (Kimura et al. 1996), Fig. 2-2B. In contrast, Rac1 is associated with the formation of actin protrusions and lamellipodia, through NED9/DOCK3 dependent activation of Arp2/3 WAVE complexes that stabilise formation of new branched actin filaments (Eden et al. 2002), Fig. 2-2B. Finally, CDC42 is associated with induction of polar structures, and filopodia (Nobes and Hall 1995). CDC42 dependent WASP activation of ARP2/3 complexes is considered a major mechanisms that induces filopodia (Rohatgi, Ho, and Kirschner 2000; Rohatgi et al. 1999), Fig. 2-2B. IRSp53/MENA complexes have also been implicated in this process (Krugmann et

al. 2001). Thus, in cells cultured in the lab, distinct morphological characteristics are associated with the activity of specific Rho GTPases proteins, and these in turn, underpin different migratory forms at the single cell level (Friedl and Wolf 2003).

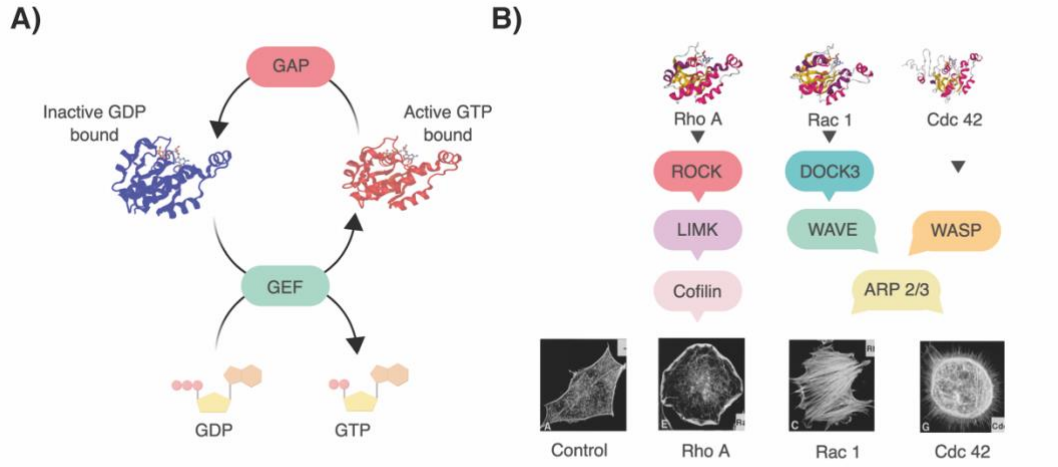

Figure 2-2, Rho GTPase pathways involved in regulation of the cytoskeleton: *A) RhoGTPase's are active when bound to GTP and inactive when bound to GDP. GTPase activating factors (GAPs) catalyse intrinsic hydrolysis of GTP to GDP and lead to inactivation of the GTPase. GTP exchange factors catalyse exchange of GDP to GTP, and thus activate the GTPase. B) Observations from overexpression studies identified key pathways that regulate cytoskeletal dynamics downstream of Rho GTPase activity. Images are from (Hall, 1998).*

In establishing the link between specific Rho GTPases and their respective cell morphologies, typically microinjection or over-expression studies were performed (Hall 1998). However more recently, analysis of endogenous Rho GTPase activity in live cells using förster resonance energy transfer (FRET) reporters, Fig. 1-4C, has shown that in fact discrete spatio-temporal activity of both Rac1 and RhoA is required for correct cytoskeletal regulation in migrating cells (Pertz et al. 2006). Localisation of Rho and Rac signalling is achieved through negative feedback in which Rac1

inhibits RhoA-ROCK activity, and conversely RhoA inhibits Rac1, in part through Rac1 dependent GAP activity (ARHGAP22) (Sanz-Moreno et al. 2008). High-content imaging of *Drosophila* BG-2 cells following siRNA depletion of different Rho GTPases also demonstrated that compartmentalisation of Rho GTPase activity occurs at a cell wide level (Bakal et al. 2007). This led to the proposal of a model where Rac1 activity at the leading-edge drives protrusions, and RhoA activity at the trailing edge drives membrane retraction; this enabling generation of traction and forward movement (Bakal et al. 2007). Chemoattractants in turn can then influence the directionality of this movement through CDC42 activity regulating the localisation of RhoA and Rac1 activity (Pertz et al. 2006). Thus, where studies of individual Rho GTPases were able to identify the key signalling mechanisms required for Rho GTPases to modulate cytoskeletal dynamics, live-cell studies and genetic depletion screens of Rho GTPases *in vitro* have shown how emergent properties such as cell form and migration depend on interactions and feedback between key Rho GTPases such as RhoA, Rac1 and CDC42.

However, with such studies being performed *in vitro* on stiff plastic substrates, they likely only captured a subset of the possible interactions and network behaviours that can occur in conditions that closer match those seen *in vivo*. Understanding how these networks change and adapt in more physiologically relevant conditions remains an open question. Specifically, where multiple migratory modes and forms are available to cells migrating in extracellular matrix is the lack of redundancy in Rho GTPase signalling that has been demonstrated on plastic maintained (Bakal et al. 2007; Sailem et al. 2014). If cells continue to demonstrate reliance on specific Rho

GTPases for certain cell shapes and migration modes, these Rho GTPases would then represent attractive targets for blocking the dissemination of melanoma throughout the body. If in contrast, there is increased levels of redundancy in melanoma cells growing and moving in more physiological conditions, this would raise questions as to how accurate our understanding of Rho GTPase signalling is based upon studies of cells cultured on stiff plastic substrates and suggest that targeting Rho GTPase signalling to block migration may not be an effective clinical strategy.

Looking to answer these questions, as well as exploring the observation that melanoma cells exist in two distinct amoeboid and mesenchymal forms on substrates that resemble those found *in vivo*, we sought to perform an unbiased analysis of the set of shapes that wild-type populations of melanoma cells can adopt in conditions that mimic the stiffness of extracellular matrix *in vivo*. By then systematically depleting Rho family GTPases with small interference RNA (siRNA) we looked to understand how the set and/or distribution of cell shapes would change; this in turn allowing us to build an understanding of the degree of redundancy exhibited by Rho GTPase networks *in vivo*. Could we block the adoption or maintenance of specific cell forms as Rho GTPase depletions on stiff plastic substrates can achieve? A positive result here would indicate a viable strategy for blocking cancer cell migration *in vivo* and potentially suggest therapeutic targets for further exploration. A negative result would indicate greater redundancy in the system exists than we have previously observed *in vitro*, suggesting our observations on plastic don't translate well to *in vivo* conditions?

2.2 Results

To study how Rho GTPases affect the set of shapes that melanoma cells can adopt, WM266.4 cells were cultured upon thick (300-700μM) matrices of Bovine fibrillar collagen I gel. Thick collagen matrices have been shown to approximate *in vivo* conditions and induce amoeboid forms in melanoma cells (Sanz-Moreno et al. 2008). Specifically, the elastic modulus of the gel used is estimated to be ~200Pa (Paszek et al. 2005) a stiffness resembles that of lung alveolar tissue. WM266.4 cells were selected due to previous observations that they exist in a 50:50 mix of amoeboid and mesenchymal forms, and frequently transition between these forms on a scale of minutes to hours (Sahai and Marshall 2003; Yin et al. 2013). In contrast other lines such as A375M2, adopt a largely amoeboid form at such stiffnesses, and rarely transition between different shapes (Sahai and Marshall 2003; Sanz-Moreno et al. 2008). We reasoned that cells that transition at faster rates and have a more even mixture of shapes in wild-type populations, would give us a stronger effect size when exploring their ability to adopt and switch between different shapes.

WM266.4 cells were imaged live, in 3D, in these conditions, over a period lasting several hours to capture shape transitions, using high-throughput confocal microscopy, Fig. 2-3A (Methods). Initial qualitative inspection of the images showed that the majority of cells remained atop of the collagen gel, in the x, y plane over the course of imaging, rather than penetrating the gel and moving in the z-dimension. The design of the optical path in confocal microscopes means that the point spread function is disperse in the z-axis limiting z-axis resolution, versus narrow in the x, y axis giving high resolution in these axis. Together these factors meant that we took

the decision to reduce each z-stack into a single maximally projected 2D image; here each pixel in the maximally projected image corresponds to the pixel with the maximum intensity value across the z-stack at that x-y position, Fig. 2-3B. Therefore, the output of the screen was a time-series of 2D fields of view, capturing melanoma cells plated atop 3D collagen matrix, treated with both control non-targeted siRNA and a library of siRNA targeting different Rho GTPases. For each condition at least two repeats were captured, though limitations in imaging through thick matrix, often prevented more than 2 repeats being captured in the screen.

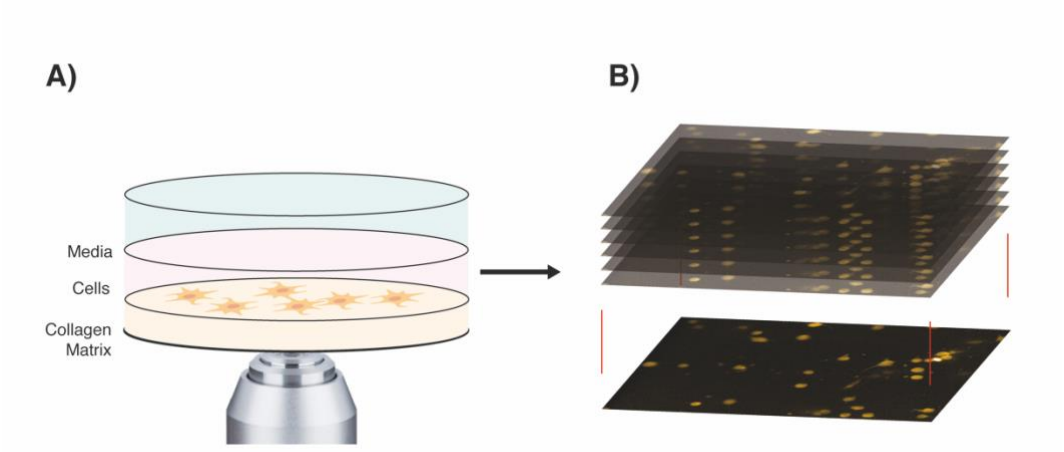

Figure 2-3, Imaging of melanoma cell plated in thick collagen matrix: *A) Cells were imaged by confocal microscopy atop thick collagen matrices, in imaging media, live over time; B) Maximum projections were taken through the z-region corresponding to the top of the collagen gels.*

2.2.1 Characterising the shape space of the screen

Segmentation was performed, and a set of 15 features was recorded for each cell object, filtering was then carried out to remove debris and out of focus cells, finally tracking was performed using a custom script (Methods). Overall after filtering a total

of 62,000 segmented cell objects being captured, with 423 being the minimum number of objects captured in a time series. A random sample of 2000 cells stratified over all conditions was used to identify the set of shapes in the population. By using a pooled sample this typically means all shapes observed in the screen are captured (Bakal et al. 2007; Bakal et al. 2008). Random samples of 400 cells were taken for each well for analysis of the effect of a specific siRNA depletion. We assumed that the speed that WM266.4 cells transition between shape would mean a random sample over time would approximate imaging a larger population of cells.

Shape Clusters

Previous work by the Bakal lab (Bakal et al. 2007; Yin et al. 2013; Sailem et al. 2014), and others (Jones et al. 2008; Fuchs et al. 2010) have shown that identification of distinct cell shapes within a population can: 1) reduce dimensionality, by reducing large multiple features to just a few shapes; and 2) improve the interpretability of results. For example, understanding differences in the statistical moments of pixel intensity over a region is difficult, whilst knowing that more cells in a population have a 'triangular' shape is both easier to understand and to relate qualitatively to raw data, for sanity checks. Here, I quantitatively defined specific cell shapes as spatial clusters in the space of features extracted from single cells. Each 'Shape Cluster' (SC) corresponds to a cell shape seen within the imaged population of cells and is identified from the sample of 2000 cells pooled across the entire screen, Fig. 2-4A.

I defined the 'Shape Cluster Profile' (SCP) as a unit vector giving the fraction of cells in each SC for any given condition. This allows us to quantitatively ask whether

specific conditions almost completely reduce the number of cells in a given shape, lead to cells adopting new shapes, or have weaker effects that correspond to shifting the distribution of cells in different shapes, Fig. 2-4B.

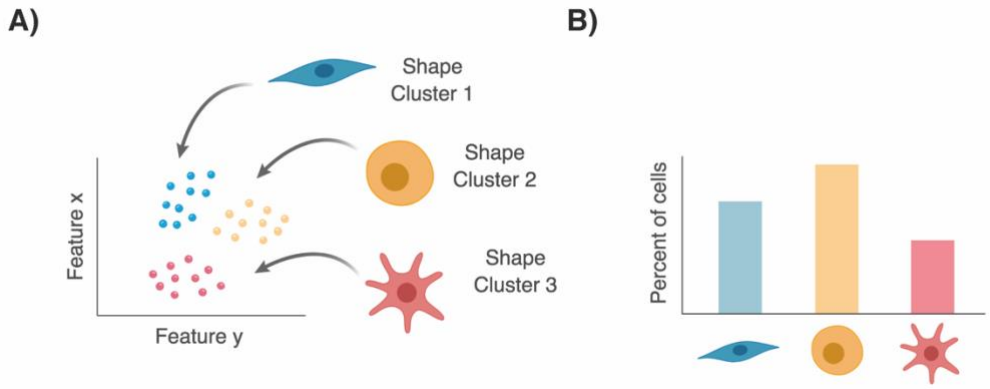

Figure 2-4, Defining the shape space of a population of cells: *A) Shape Clusters (SCs) are defined as spatial clusters of cells in feature space; B) Shape Cluster Profiles (SCPs) give the fraction of a population of cells in a specific SC.*

Identifying Shape Clusters

Typically, to identify distinct SCs within a sample of cells, supervised methods are used (Jones et al. 2009; Rämö et al. 2009; Bakal et al. 2007; Boland and Murphy 2001). Here, a set of reference manual shapes are selected, and classifiers used to map cells to these reference shape classes. However, such methods suffer from two challenges: 1) to fully understand the set of all possible shapes, humans would need to analyse every shape in the sample. Due to time constraints this is difficult; and 2) There is natural bias in the way people select shapes, and this can introduce bias into the analysis. Thus unsupervised machine learning approaches can lead to more complete

and unbiased analysis (Friedman, Hastie, and Tibshirani 2001), and form the basis of the analysis conducted here. However, these unsupervised approaches pose additional challenges in identifying the number of SCs that should be used to describe the population, termed model selection (Tibshirani, Walther, and Hastie 2001).

In previous work Sailem et. al. developed an unsupervised method to characterize SCs based on the mean silhouette statistic (Sailem et al. 2014). In this approach, following z-score normalisation, PCA was performed on shape space to reduce the number of feature dimensions used to describe shapes (Hotelling 1933). Gaussian Mixture Models (GMMs) were fit with expectation maximisation to identify cell SCs in the dataset (Bilmes and Others 1998), and the silhouette score used to measure clustering quality (Rousseeuw 1987). Model number was chosen to maximise the silhouette score. When I used this previously adopted approach, the silhouette score for two clusters was high at 0.53, Fig. 2-5A. However further inspection of single cell data points after PCA transformation suggested the data was log distributed, and k-means clustering was capturing the long tail of the log-distribution, Fig. 2-5B. I therefore log transformed feature values and reapplied k-means clustering for increasing model number. After log transformation the silhouette score never surpassed 0.4 suggesting no natural multi-modality, Fig. 2-5C (Dimitriadou, Dolničar, and Weingessel 2002; Bolshakova and Azuaje 2003; Rousseeuw 1987), visual inspection of individual data points confirmed this Fog. 2-5D. Yet, qualitative analysis also highlighted a divergence from the normal distribution along the first principal component Fig. 2-5D, indeed a Kolmogorov-Smirnov (KS) test for normality confirmed this with $P<0.001$, and this could be clearly seen in a histogram of the PC1

probability distribution plotted alongside a normal distribution of equal mean and variance Fig2-5E. Moreover, structure could also be seen qualitatively when comparing the 1st and 3rd principal components Fig2-5F. This gave an early indication that no natural multimodality existed in the sampled population of melanoma cells, and that more forms than the observed amoeboid and mesenchymal shapes existed. However, often feature transformation and selection, as well as optimisation of clustering technique can improve results (Caicedo et al. 2017). Therefore, I sought to develop an alternative unsupervised approach to analysing cell shape, to further test for the existence of natural structure in the dataset that would aid interpretation.

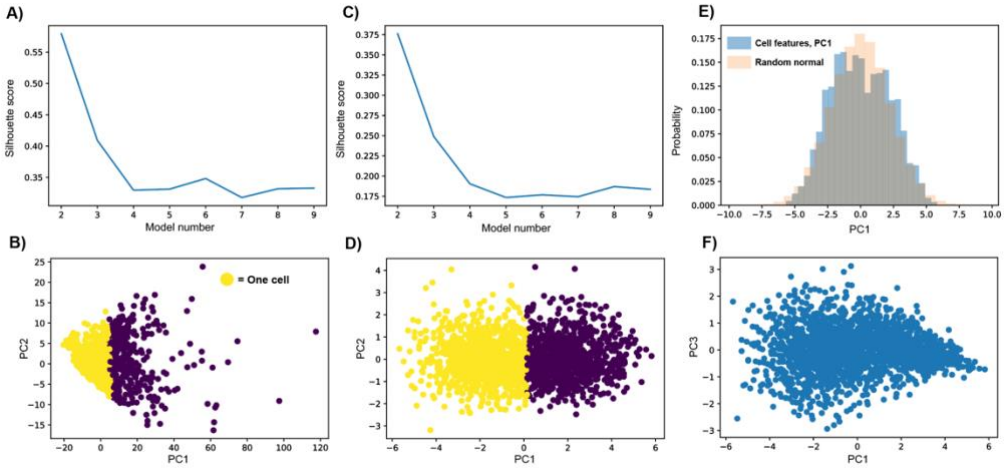

Figure 2-5, The Silhouette score indicated that shape space was continuous: *A) The silhouette score was calculated for increasing model number, using clusters determined by k-means from the first 3 PC's following PCA transformation of raw feature values from a sample of 2000 cells; B) Plotting of single cell PC1 and PC2 values showed a log distribution with no clear multi-modality; C) Applying the same method to log transformed data showed much weaker silhouette scores for all model numbers; D) In the log-transformed feature space no clear clustering could be seen. However, PC1 demonstrated divergence from a normal distribution; E) A probability distribution of single-cell PC1 values, versus an equivalent*

normal distribution further shows this; F) Further structure was seen between principle components 1 and 3.

Optimising a method for unsupervised cell shape detection

I began by looking to transform the feature space into one that would result in the most biologically relevant SCs, and SCPs that could best discriminate between conditions (signal) and would show minimal variation between technical repeats (noise). The Davies-Bouldin index (DBI) is a standard measure of internal clustering quality (Davies and Bouldin 1979). Lower DBI scores correspond to better clustering of the data. More formally in Euclidean space the separation of a cluster S_i can be defined in terms of the cluster center A_i, data points within the cluster X_j, and the number of data points in the cluster T_i, given the set of all clusters,

$$S_i = \left(\frac{1}{T_i} \sum_{j=1}^{T_i} \left(X_j - A_i \right)^2 \right)^{\frac{1}{2}}$$

In turn the distance between two different clusters in Euclidean space $R_{i,j}$, is the Euclidean norm between the two centers, A_i and A_j,

$$R_{i,j} = \left\| A_i - A_j \right\|_2$$

For each cluster the worst overlap between that cluster and another cluster is defined in (Davies and Bouldin 1979) as,

$$D_i = \max\left(\frac{S_i + S_j}{R_{i,j}}\right), j \neq i$$

The DBI then gives the average overlap between a cluster and the nearest cluster to it over the entire dataset,

$$DBI = \frac{1}{N}\sum_{i=1}^{N} D_i$$

By defining SCP vectors of technical repeats as belonging to the same cluster, and different conditions (siRNA gene depletions) as belonging to different clusters, we can use the DBI to determine how effectively any given feature reduction technique and clustering approach is extracting biological information from technical noise. We reasoned that those combinations of approaches result in the best, or lowest, DBI scores would be most likely to reveal morphological sub-populations.

Testing combinations of feature reduction and clustering methods.

Using the DBI to evaluate the performance of feature reduction and clustering techniques, I looked to test several methods for reducing feature space and identifying SCs, against a benchmark DBI value. To obtain a benchmark DBI value, I took a mean-aggregate of features values from the sample of 400 cells from each condition. This has been found to give the most sensitive measure of a conditions effects, however suffers from issue with interpretability, and does not help with identification of subpopulations (Caicedo et al. 2017). The benchmark DBI was then calculated from the Euclidean distance between mean-aggregates of technical

repeats, versus the distance between different conditions. Since dimensionality is known to affect the DBI (Davies and Bouldin 1979), PCA was performed on the mean feature vector as is most commonly performed in high content analysis (Caicedo et al. 2017), to create a vector with a dimensionality that would match the number of extracted SCs a technique was being compared to. This benchmark set of DBI values could then be used to assess the performance of different feature transformation and clustering strategies, Fig. 2-6 (red line).

There are several major approaches to reducing the size of the space describing cell shapes. Firstly, reducing the dimensionality of the dataset, either through linear transformation, or feature selection. Secondly, reducing the space of variables for each feature, for example, through binning all features into discrete values. There are also several well documented approaches to unsupervised clustering of data, notably K-means clustering, Gaussian Mixture Models, fuzzy C-means clustering, and hierarchical clustering (Jain 2010). To test combinations of these techniques, each of the clustering approaches was tested over SC numbers ranging from 2 to 7, in combination with: 1) Raw feature data, Fig. 2-6A; 2) Binning of features into two values, those above and below the mean, Fig. 2-6B; and 3) Reduction of feature space using PCA to 3 PC's (capturing >90% of variance), followed by clustering in this feature space, Fig. 2-6C.

This analysis demonstrated that identifying SC's from both raw data, and reduced feature spaces using PCA both led to greater DBI values, independent of the clustering method used, or number of SCs selected. In contrast, binning of features into two bins, based upon values being above or below the mean, resulted in DBI values that were

similar to those of the benchmark, as long as the SC number was greater than 2. Following feature binning, all clustering methods performed equally well. K-means clustering was significantly faster to run computationally and was therefore used in combination with feature binning. Together this analysis defined the feature transformation strategy and technique for clustering cells by their shapes.

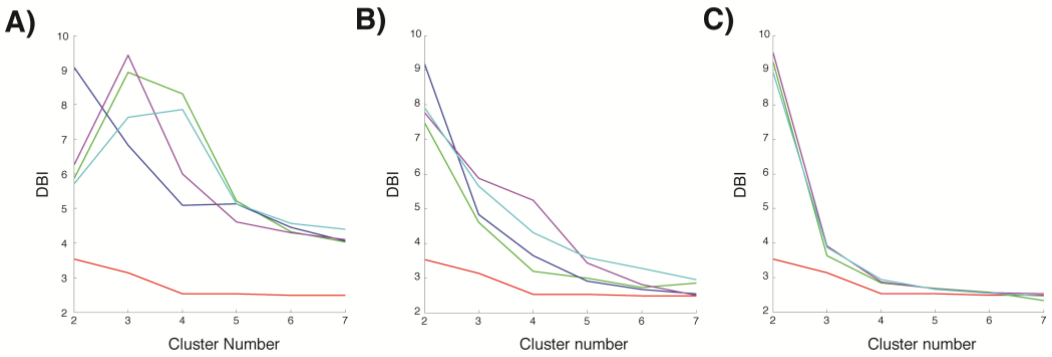

Figure 2-6, Analysis of different combinations of clustering technique and feature space transformation technique: *A) Comparison of different unsupervised clustering approaches on raw data, versus benchmark score (red). B) Comparison of unsupervised approaches on data transformed by binning features into two values, above and below the mean, above 2 clusters scores resemble those of the mean feature benchmark; C) Comparison of clustering methods on the first 3 principal components following transformation of the feature space by PCA. All methods perform significantly worse than the benchmark here.*

Selecting shape cluster number

Choosing the correct number of SCs with which to describe the data, often termed model selection, is a known challenge in unsupervised data analysis (Tibshirani, Walther, and Hastie 2001). Following K-means clustering of the data, calculation of the mean silhouette value on the transformed feature space showed that whilst the

quality of clustering was improved, the ideal value of 1 was still never approached, Fig. 2-7A. Moreover, the silhouette value dropped continuously from 2 clusters, further indicating no natural clustering was present, Fig. 2-7A. The Gap statistic was also considered (Tibshirani, Walther, and Hastie 2001), and compares increasing numbers of clusters compared to a control randomised dataset, here a single peak value indicates a positive results, but instead no peak value was observed instead the statistic plateaued after 4 clusters indicating a negative result, Fig. 2-7B. Finally, Bayesian information criterion (BIC) identified a broad optimum centred on 5 or 6 SC rather than a clear minimum as may be expected for natural clustering (Schwarz 1978), Fig. 2-7C. Together this analysis indicated the data represented a largely continuous distribution with no clear discrete clusters emerging, together supporting the notion that melanoma shape space is largely continuous for cells cultured atop thick collagen matrix.

To convert features space into a set of SCs for improving interpretability and facilitating analysis of dynamics I therefore looked to use measures of clustering stability to determine SC number, and identify structure in the continuous distribution of melanoma cells shapes. Such stability approaches are known model selection methods that seek to identify the number of clusters that leads to most reproducible separation, subject to perturbation of the data (Shamir and Tishby 2008, 2009). I initially explored clustering reproducibility by performing k-means clustering 1000 times on cells resampled from a set of 61,000 cells; with a 1000 cells held out for validation of stability. How near cluster centroids fell to each other over successive iterations of clustering, was analysed using the silhouette score.

Using this approach clear peaks at 2 or 6 clusters emerged with high silhouette values, when clustering was performed with either k-means, Fig. 2-7D, or fuzzy c-means Fig. 2-7E. Moreover, 2 or 6 clusters remained stable on analysis of held out data, giving again similar centroid values to those obtained from the resampled datasets, Fig. 2-7F (large dark points). Given that the DBI score for 2 SCs was significantly worse than the benchmark results following transformation of feature space, Fig. 2-6B, indicating loss of information, 6 SCs were chosen to describe the data. Taken together this unsupervised analysis therefore revealed that whilst melanoma shape space was likely continuous, structure existed that could be captured by clustering. Both 2 SC's, perhaps corresponding to the observed amoeboid and mesenchymal shapes, as well as 6 SC's emerged as reproducible SC numbers that could be used to describe the data.

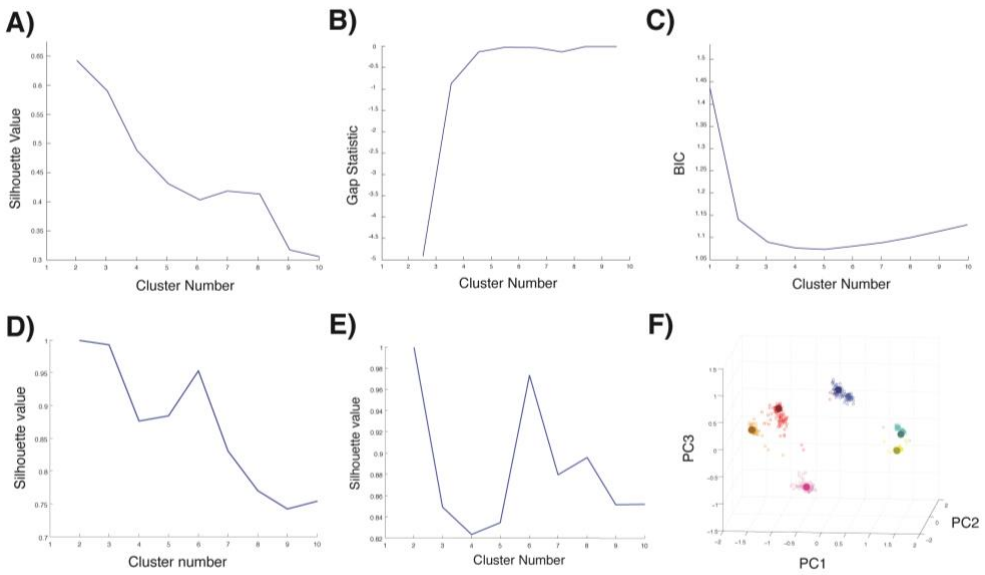

Figure 2-7, Selection of the optimum number of shape clusters with which to describe the data:

A) Silhouette values were calculated on k-means clustering applied features after binning values above and

below the mean; B) Calculation of the GAP statistic from cluster numbers 2 through to 10; C) BIC calculated on the full dataset and cluster number increasing to 1; D) The stability score over cluster numbers 2–10, calculated as the silhouette value applied to k-means clustering of centroids. Centroids are from 100 resamples of initial k-means clustering to identify SCs from a sample of 2000 cells; E) Same as (D) except Fuzzy C-means clustering is used; F) Held out data (large dark points) mapped to clusters centroids sampled from training data (light circles).

Corresponding cell shapes

To interpret SC's, the mean feature values of cells in each SC was calculated for both 2 SCs and 6 SCs, Fig. 2-8A. Classification of cells into two SCs yielded; 1) rounded cells, characterised by low length to width ratio, and poor symmetry caused by cells in this category having no clear lines of symmetry; and 2) elongated/protrusive cells that have a high length to width ratio, and contain typically bipolar cells with distinct axis of symmetry. The identification of these two cell SCs supports the canonical notion that melanoma cells adopt two distinct amoeboid and mesenchymal from when plated in 3D matrix.

When clustered into 6 shapes, we found that three shapes that qualitatively would be classed as mesenchymal emerge, Fig. 2-8A. Specifically, these include the most common bipolar or 'spindle' like shapes, large stellate or 'star' shaped cells that have both a large area in the image and are highly protrusive, and finally mono-polar or 'tear drop' shape cells, typified by a round nuclear body with single extending protrusion. Three distinct clusters also emerge from the amoeboid group, these are 'small elliptical' and 'small round' cell shapes that are largely similar, but differentiated by levels of symmetry, and a 'large round' group of cells that resemble

cells that have lost contractile force. The fact that these allow better separation of conditions, and match benchline performance compared to mean aggregated well values as measured by the DBI, indicates that these shapes also correspond to important forms that are influenced by depletion of different Rho GTPases.

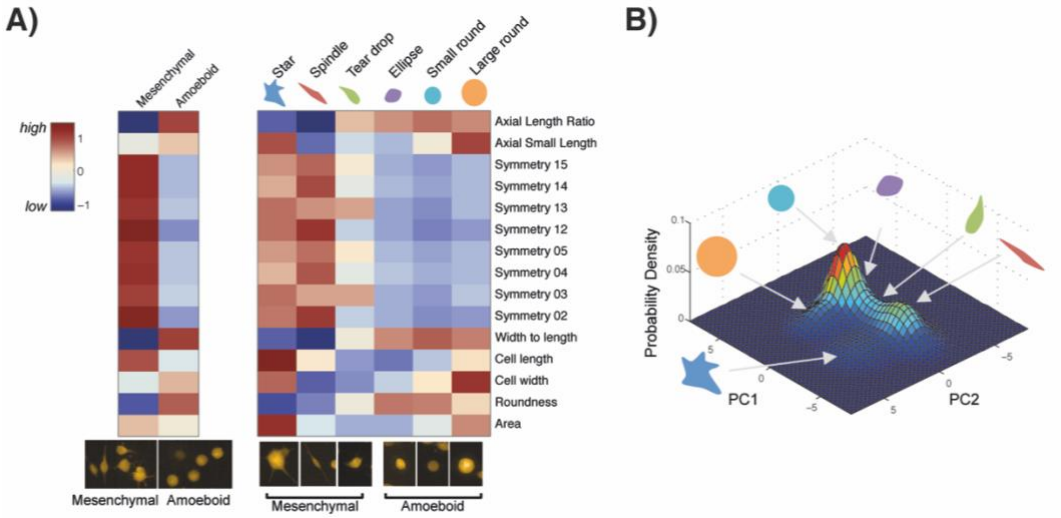

Figure 2-8, Heat maps of normalised feature values were averaged for each SC: *A) Mean feature values when two clusters are chosen to split single cell data, one cluster is enriched for length and is asymmetric, the other is enrich for area and roundness features and is symmetric; B) When split by 6 categories, finer grained sub-populations emerge, Round shapes correspond to low symmetry and high roundness/width:length scores. The teardrop shape scores stronger for odd symmetry measures, whereas the spindle/mesenchymal shape is stronger for even symmetry measures and the large star shape scores high for all measures, except roundness/width:length; C) SCP of wild type cell populations visualised using a contour map, PCA was applied to a pooled sample of 2000 cells; the mean and SD for the first two PCs (77% of variance) are then plotted as a normal distribution scaled to cluster membership for wild-type cells.*

Analysis of the wild-type SCP, or percentage of wild-type cells in each SC demonstrated that the majority of wild-type WM266.4 cells existed in the small round shape. These forms likely represent the contractile amoeboid shape that correspond to the canonical amoeboid form previously observed, Fig. 2-8B. The second most predominant shape is the spindle shape, that would classically be defined as mesenchymal. In wild-type WM266.4 cells, the ellipse, teardrop, large round and star shapes all exist as rarer forms. Thus, in wild-type populations we observed dominance of the canonical amoeboid and mesenchymal forms. However all shapes were present to some degree in wild-type populations, meaning that Rho-GTPase depletions did not give rise to completely new shapes, not-seen in wild-type populations.

2.2.2 Gene depletion of Rho GTPases results in distinct groups

In the screen of Rho family GTPases, 17 genes were knocked down. Hierarchical clustering of resulting SCPs identified four distinct groups of genes Fig. 2-9A. Notably, no gene depletions resulted in the formation of a single shape, rather depletions enriched for specific shapes with heterogeneity still present, this supporting the notion that no single Rho GTPase is responsible for adoption of a single form. Gene depletions that enriched cell populations for amoeboid phenotypes including teardrop, ellipse and small round represented the largest group of effects following gene depletion. These cells also had notably few cells in the star shape. Both Rac1 and RhoG depletions enhanced for such transitional shapes. This is in line with studies demonstrating that Rac1 is necessary for mesenchymal protrusive states, for

example, inhibition of Rac1 causes increased rounding in HT1080 cells cultured in 3D matrix (Sanz-Moreno et al. 2008; Yamazaki, Kurisu, and Takenawa 2009), and Rac1 activity has also been shown to induce protusiveness and polarity in a Rho G dependent manner (Damoulakis et al. 2014). RhoA depletion also falls in the same large cluster, although demonstrates weaker enhancement of the small round shape. The less pronounced effect on morphology following depletion of RhoA, compared to similar studies, indicates that in 3D collagen matrix RhoA depletion is either less penetrant, or less involved in generation of contractility in melanoma than previously reported in *in vitro* studies.

The second most prominent group includes those gene depletions that result in enhancement of the large round shape. Notably, this group contains the three genes Rnd2, Rnd3, and RhoB, agreeing with evidence suggesting that both Rnd2 and Rnd3 both activate RhoB in endothelial cells, to promote contractility (Gottesbühren et al. 2013). That we observe depletion of the Rnd2/3-RhoB axis leads to enrichment of large round cells provides support to the hypothesis that this shape is associated with loss of contractility. This contrasts those depletions that enrich for small round cells likely corresponding to the classical highly contractile amoeboid shape. The third group contains cells that suppress rounded forms. The three genes Rac3, RhoH and RhoD all strongly enrich for mesenchymal forms, suggesting a role for these genes in suppressing protrusions and/or formation of adhesions. Of these genes Rac3 has been demonstrated to exert an opposing effect to Rac1, blocking outgrowth of neurites in neuronal cell lines that share a similar developmental lineage to melanocytes (Hajdo-Milasinovic et al. 2009; Hajdo-Milasinović et al. 2007). The final group enriches for

bipolar spindle shapes, though no other shape is notably depleted. Wild-type populations feature in this group indicating that generally in wild type populations the full range of cell shapes is present. Thus overall the static unsupervised analysis that I performed indicated that Rho GTPase depletion by siRNA caused enrichment of certain forms over others, versus, emergence of new shapes, or loss of specific morphologies, although notably some genes did lead to very strong enhancements of the spindle shape vs more amoeboid forms. This meant that sufficient redundancy likely exists in Rho-GTPase signalling in WM266.4 cells plated in thick collagen matrices as to account for loss of most genes, although this assumes effective gene depletion by siRNA.

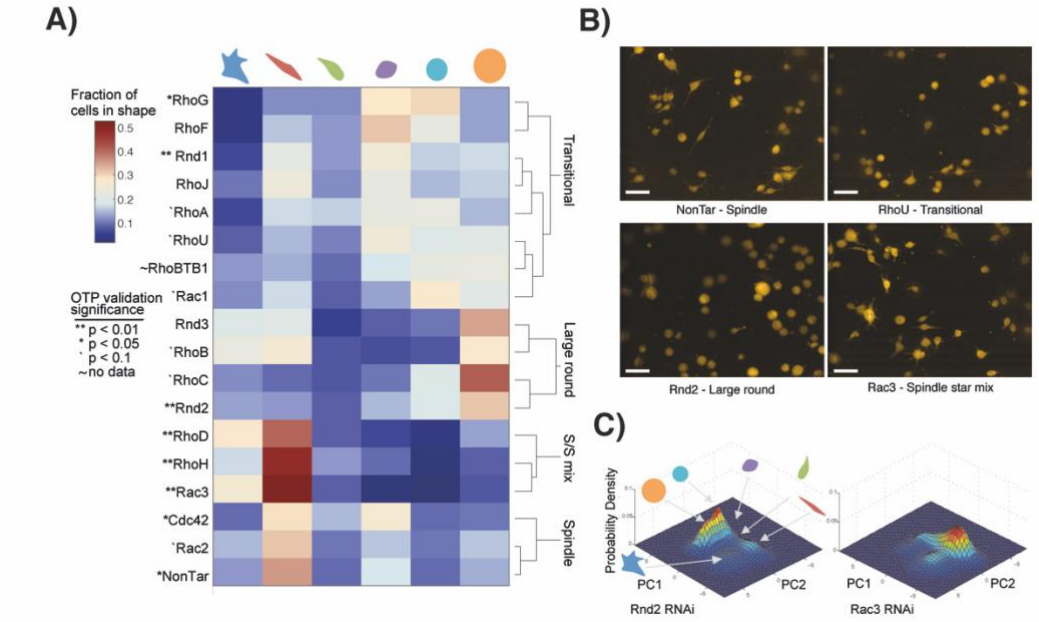

Figure 2-9, Rho-family GTPases regulate the exploration of shape space in 3D matrices. *A) Hierarchical clustering of genes based on SCPs. Wards linkage was used for clustering, with a cut-off value*

of 0.3 (maximum distance 1). Validations with individual OTP siRNAs was performed and p values for the best validating siRNA against a null distribution are displayed. B) Images of representative knockdowns for the four groups. Scale bars, 50 µm. C) Frequency distribution of Rnd1-depleted cells (left) and Rac3-depleted cells (right). Rnd1-depleted cells are enriched in large round cells, and Rac3 is enriched in spindle-shaped cells.

Validation of individual gene depletions

Seeking to understand the degree to which enrichment, versus loss or gain of morphologies, was due to biological redundancy in Rho GTPase signaling, and more generally cytoskeletal signalling, we looked to better understand the effects of the siRNA depletion technology on cell morphology. For the initial screen we used pooled 'siGenome' siRNA. By analysing how the effect of different 'onTarget plus' siRNA depletions for each gene compared, we looked to test whether a statistically significant number of siRNA invoked the same effect as the siGenome pooled knockdown, suggesting a targeted and reproducible biological effect.

We therefore conducted a screen of both pooled 'siGenome' siRNA and four individual 'onTarget plus' siRNA targeting each of the Rho GTPases, together giving 5 different siRNA depletions for each Rho GTPase gene. This validation screen was conducted using WM266.4 melanoma cells plated on 3D collagen matrix, under the same conditions as for the original screen. However, in contrast to the original screen, cells were fixed and imaged statically. Overall 4 wells (2 repeat wells per plate, 2 plates) were captured for each onTarget plus siRNA, alongside the siGenome pools used for the original screen. Imaging was then performed on fixed cells and maximum projections were taken through the thick collagen, to ensure all four well repeats

across two plate repeats were recorded. The same set of 15 features as used in the dynamic analysis were recorded from cell segments, and feature vectors were normalised plate wise to control for batch effects. Binning to binary values was performed and the same SC centroids from the dynamic analysis were used to classify shapes in this validation dataset. This led to SCPs which would be directly comparable to those obtained from static analysis of the first dynamic dataset.

To determine the statistical significance of gene depletion effects, Pearson correlations were calculated between the SCP of the 5 cell populations created following SMARTpool and OTP siRNA gene depletions. This gave total of 20 data points (correlation values between depletion SCPs) per gene. A null distribution was then calculated by randomly drawing a SCP from a different condition, 4 times, for each of the 5 siRNA targeting the same gene and calculating the Pearson correlation between the siRNA SCP and the randomly drawn SCP. Overall this demonstrated that siRNA induced reproducible shapes changes in melanoma cells, Fig. 2-10A.

We then looked at individual siRNA gene depletions for each Rho GTPase gene to determine whether onTarget siRNA gave rise to SCPs for all Rho GTPase gene depletions, or whether for certain gene depletions a higher number of onTarget siRNA validated with significant effects than for others. Meaning, where could we be more confident that the gene depletion had a biologically meaningful effect? Across the screen of deconvoluted siRNA a number of genes validated with high significance, Fig. 2-10B. This was exemplified by the gene RhoD that demonstrated significant reproducibility of the pooled effect across all individual siRNA tested. Many genes had at least one individual siRNA demonstrate a similar effect to the pooled siRNA. Several

genes also failed to validate. Overall this analysis therefore demonstrated that whilst a statistically significant number of depletions are on-target, high levels of off-target effects and overall noise contribute to many of the results. Thus, whilst this screen effectively induced heterogeneity in the population, and this innate heterogeneity produced reproducible results, the results of an individual treatments must be treated with a degree of caution. Thus, in discussing the effect of single treatments, only those with a significant degree of validation are mentioned. Of note though, RhoD was observed to induce both a significant effect size following siRNA depletion, and this phenotype was recreated across all other siRNA tested.

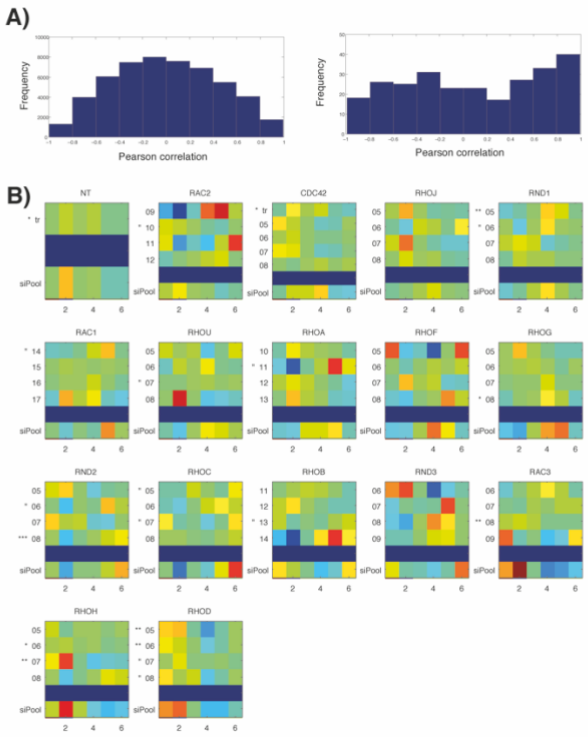

Figure 2-10, Validation of Rho GTPase gene depletions: *A) OTP vs SMARTpool SCP correlation distribution for all wells shows a positive skew versus the null distribution generated by randomly*

permuting OTP SCs and calculating correlation between SCPs and SMARTpool SCPs; B) Heatmaps show plate normalised SCPs for OTP siRNA against SMARTpool siRNA. OTP data is averaged across 4 well repeats. Significance values as shown in the main text are given beside the OTP siRNA numbers, these correspond to the final two digits of the catalogue number in the table of OTP siRNAs used (Cooper et al. 2015).

2.2.3 Quantification of shape dynamics

To understand how cells explore shape space dynamically I quantified the number of transitions that tracked cells (methods) made over time between different SCs. To explore these results matrices were created where rows corresponded to shapes in any given frame, and columns corresponded to shapes in the subsequent frame, Fig. 2-11A. Thus, for every frame that a shape remained the same, the corresponding diagonal value was increased by one, and for a shape transition the off-diagonal value corresponding to the forward transition was increased by one. The value of each matrix element was then divided by the total number of transitions recorded for that condition, thus giving changes between shapes as a percentage of all transitions made. Initial inspection of these matrices revealed that we were likely capturing all of the transitions being made by cells between shapes, since typically we observed transitions between shapes with more similar morphologies, e.g. elliptical to teardrop, versus randomly between shapes with either similar or very different morphologies e.g. small-round to star.

Two routes in shape space

Noticeably two distinct routes between the amoeboid shapes and mesenchymal shapes emerged. One, a polar route, through the elliptical and tear-drop shapes, and

the other through the large-round and mesenchymal star-shaped cell, here termed the 'apolar' route. Although transitions directly between amoeboid and mesenchymal shapes occurred these happened at much lower frequencies, as visualised using a graph of the transition matrix, Fig. 2-11B. I also observed no directionality in the shape transitions being made, this was evidenced by symmetry in the matrices along the diagonal, i.e. in no cases did cells make forward transitions from one shape to another only. Quantitatively the sum of upper triangular elements over lower triangular elements averaged over all matrices was 1.0028; with a standard deviation of 0.0098, indicating no deviation from this symmetry. This indicated that the various different shapes reside in equilibrium, and transitions are likely made due to stochastic variability in the signaling state or microenvironment of the cell, rather than say through evolved cyclic behaviour. Transitions to self were most-likely for cells in the spindle shape and large round shape, indicating these were also more stable shapes than the other transitionary forms, Fig. 2-11C.

The effect of gene depletion on shape transition dynamics

As well as having distinct effects on how cells explore shape space, siRNA depletions of Rho family GTPase genes also affected the dynamics with which cells transition between shapes. To quantify this I calculated a 'dynamic score' giving the number of transitions between different shapes versus a shape staying the same, this being the sum of off-diagonal elements in transition matrix divided by the sum of diagonal elements. Higher values mean that a population is more dynamic whilst

lower values, indicate that cells make less transitions. Across all matrices the average value of this score is 0.5, indicating shape transitions occur once every three frames.

To determine the similarity of different gene depletions on transition dynamics I calculated the Euclidean distance between flattened transition matrices, and performed hierarchical clustering over all conditions in the live-cell screen, Fig. 2-11D. Separating the different effects into five major groups gave: 1) cells in which the polar route is enriched and mesenchymal morphogenesis is weakly compromised (Rnd1 RNAi); 2) cells in which mesenchymal morphogenesis is compromised (Rnd2 RNAi); 3) cells in which the apolar route is enriched and amoeboid morphogenesis is weakly compromised (RhoD RNAi); 4) cells with transition dynamics that are similar to wild type; and 5) cells in which amoeboid morphogenesis is compromised and the dynamic score is notably lower (Rac 3 RNAi). Thus, different depletions not only enrich for distinct morphologies, but also alter the dynamics of how cells transition between shapes. Moreover, where gene depletions enrich for similar shapes statically, for example, RhoD, RhoH and Rac3; dynamically the effect can be quite different, for example RhoH and Rac3 both reduce the number of transitions being made versus wild-type, whilst RhoD depletion caused an increase in the dynamic score. Together indicating certain genes may play a more significant role in regulating transition dynamics, versus controlling adoption of a specific shape.

When looking at the absolute number of transitions being made versus percentage of transitions, through a network graph visualisation, we found that in cases where one dynamic transition is enriched, movements through the alternate polar or apolar routes were not noticeably lower, with the exception of Rac3 and

RhoH depletions, Fig. 2-11E. This indicates the number of transitions being made is generally higher following Rho depletion and is supported by these conditions having an increase in the dynamic score. Thus rather than shifting the way cells transition between shapes, Rho gene depletions appear to reduce the barriers in signalling changes that melanoma cells are required to undergo to transition between shapes.

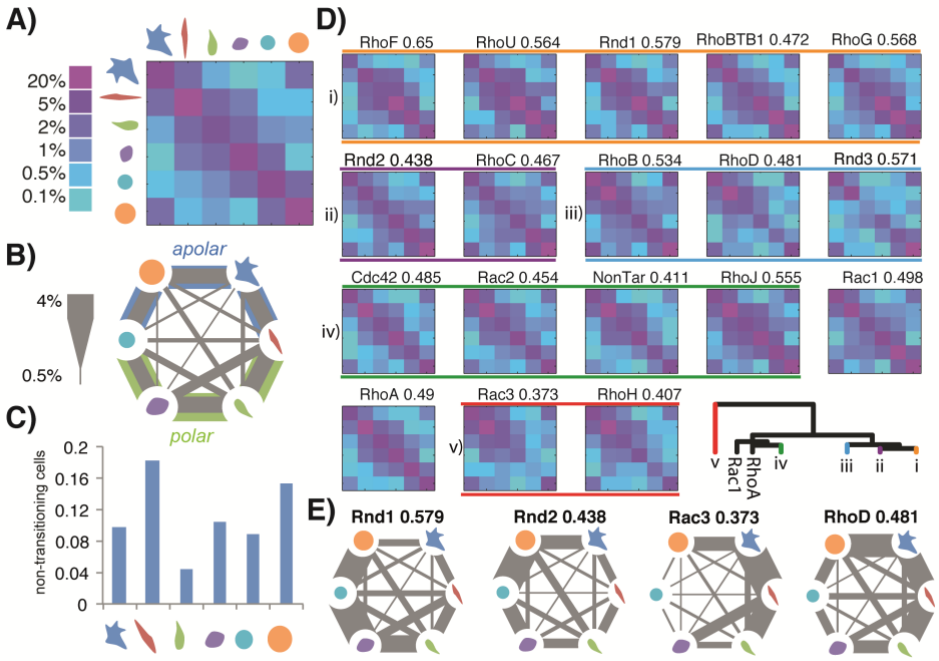

Figure 2-11, Quantifying the dynamics of shape transitions in melanoma cells plated in thick collagen: *A) Heat map showing the percentage of transitions between shapes including to self (diagonal) averaged across all gene knockdowns; log-scale coloring; B) Alternative visualisation of the matrix shown in A. Weighted edges show the mean percentage of transitions made between shapes; the majority of transitions are made between "neighboring" shapes. The total joint probability (shown as percentage) of a cell going from small round to spindle via either route is given for a sense of the overall flux along each route; this is calculated as the sum of joint probabilities for the three possible ways of transitioning along either the polar or apolar route; C) Percentage of cells in a given shape staying in that shape; D) Effect of*

depletion on the number of transitions being made; color as in A. In addition, off-diagonal values divided by diagonal values are given beside the gene name as a statistic of how dynamic a knockdown is, termed the "dynamic score"; E) Weighted graphs of four representative gene depletions, which emphasise how gene depletion affects dynamics. The dynamic score and route percentages based on joint probability are also shown for the example depletions.

Information derived from dynamics

Finally, I looked to determine whether including dynamic information increased the ability to discriminate between different conditions, meaning that there is additional biological signal captured in the dynamic profiles, that is not detected by static imaging alone. To measure this I calculated the Davies Bouldin Index score of the flattened dynamic profiles, reduced to a dimensionality of 6, to match the dimensionality of the static analysis, by using the first 6 PC's obtained following PCA. Strikingly the Davies Bouldin index score decreased from 2.25 (obtained for static clustering using 6 SCs) to 1.88 demonstrating that additional biological effects were being captured that were not detected by static imaging alone. Taking the first 6 PC's following transformation by PCA on the combined static SCs and flattened transition matrices, also gave rise to a DBI of 1.85 indicating that the dynamic transition matrices almost fully capture the information contained in the static profiles, likely through the number of identity transitions being made by a shape to itself. Thus taken together the dynamic analysis demonstrates that distinct and reproducible effects can be detected through analysing live cells that cannot be captured through static imaging alone. Moreover, we identify two distinct routes in features space that cells

can transition through when converting between amoeboid and mesenchymal shapes.

2.3 Discussion

2.3.1 Unsupervised analysis of cell shape

Culturing and imaging of live cells in thick collagen matrices that mimic *in vivo* conditions represents a significant technical challenge, even when conducted in low throughput and analysed using manual qualitative approaches. Here we demonstrated an approach that allowed us to study the behaviour of thousands of live cells imaged in thick collagen matrices over time. Performing such experiments in these 3D conditions, is critical as it allows us to study cytoskeletal and cell shape regulatory pathways that may be masked when cells are cultured on stiff 2D plastic substrates (Yin et al. 2013). Importantly such conditions are also more likely to translate into the clinic (Sachs and Clevers 2014).

By using unsupervised approaches, we designed the analysis to reduce human bias, that could be caused by previous qualitative observations of melanoma cells plated in 3D collagen matrix. This identified structure in the high silhouette sore of 2 SC's that could be likened to the canonical amoeboid and mesenchymal forms (Sahai and Marshall 2003), as well as revealing a sub-division of shapes beyond the canonically described amoeboid and mesenchymal forms. Importantly, this sub-division of shapes improves the DBI, and ability to discriminate between different conditions implying these are true biological forms. Although we did not observe

these shapes to exist in discrete separable clusters in feature space, the evidence that these 6 SCs have biological significance, was strongly supported by the finding that cells transition between shapes that are more morphologically similar, and that two routes between the canonical amoeboid and mesenchymal forms exist. If these 6 shapes were an artefact of the analysis, we would expect to observe random transitions between polar and apolar transitionary shapes, or directly from spindle shaped cells to small round cells. Together this provides important evidence that the method developed for quantifying cell shapes, led to SCs that had biological significance. Such an analysis approach could be generalised to other studies, and we believe is particularly amenable to noisy data sets, where clear clusters do not emerge.

2.3.2 Emergence of polar, and apolar routes between amoeboid and mesenchymal forms

The analysis conducted here identified two distinct routes in shape space that enable cells to transition between amoeboid and mesenchymal forms. The polar route resembles cases where cells either encounter a chemoattractant signal or migrate along 1D fibers in the extracellular matrix (Doyle et al. 2009). Depletions that enriched for transitions through this route likely shifted the balance towards establishment of polarity, and formation of either one or two protrusions that inhibit formation of further protrusions and adoption of the star shape. In the apolar route we observe formation of multiple protrusions and adoption of the star shape. Such a form is consistent with loss of contractility, the expansion of multiple protrusions and

adhesion to extracellular matrix components (Paňková et al. 2010). These properties could be driven by the absence of chemoattractant or ability to respond to it, or alternatively the binding to and exploration of multiple extracellular matrix fibers. Together these two routes demonstrate additional plasticity in the ability of melanoma cells to adopt different forms, that emerges in environments that mimic physiological conditions.

2.3.3 Role of Rho GTPases in regulating melanoma cell shape

It has been demonstrated in mouse knockout lines that loss of the Rho GTPases Rac1 or Cdc42 both cause early embryonic lethality, demonstrating the critical importance of these factors for correct development. However, in the WM266.4 melanoma cells analysed here, effects were less pronounced, perhaps with the exception of RhoD. In no cases did we observe significant levels of cell death or the gain of strikingly different shapes that may be associated with say cytokinesis defects, or apoptosis. Instead we observed enrichment of specific forms present in wild-type populations. We therefore hypothesise that in cancer cells in physiological environment, redundancy in Rho GTPase signaling exists, that allows cells to use multiple cytoskeletal regulatory pathways to maintain cell form and migrate through tissues, in the absence of a single Rho GTPase. Targeting of a specific Rho GTPases may however be more effective where activating mutation are present in that Rho GTPase. For example, multiple studies have implicated activating mutations in Rac1 as a major driver of melanoma progression (Hodis et al. 2012; Krauthammer et al. 2012). In such Rac1 addicted cancers, Rac1 depletion may have a significantly more pronounced effect than we observed in WM266.4 melanoma cells. Yet despite this

redundancy, our finding that the Rho GTPAse, RhoD, significantly supressed amoeboid forms in all tested siRNAs, suggests it may play a role in influencing melanoma's ability to migrate through amoeboid forms, and thus warrants further follow up studies.

2.4 Conclusion

In conclusion, our work demonstrates that melanoma cells dynamically explore a more diverse shape space in 3D environments than has been observed on 2D plastic substrates. Although It remains unclear how transitions between these forms plays a role in migration and ultimately metastasis of melanoma cells, *in vivo*. The ability of cells to transitions between migratory modes through these two different routes may evolve in cancer populations to allow greater plasticity and ultimately increased dissemination capability than either route alone. To shed light on whether targeted inhibitors of specific Rho GTPases could be effective against melanoma cell migration, further studies would need to test how depletion of genes such as RhoD, mapped to the ability of melanoma cells to migrate and metastasise in animal models. Such a study would also provide fascinating insights into how much more predictive results from cell cultured in environments that mimic *in vivo* conditions are of behaviour in animals.

2.5 Methods

Cell cultures were prepared and treated with siRNA by Amine Sadok, as per (Cooper et al. 2015). Imaging was performed using the Opera high-content imaging platform by Chris Bakal, as also described in (Cooper et al. 2015).

2.5.1 Features and data preprocessing

Cells were segmented using a simple threshold using the Acapella software (PerkinElmer). Fifteen features were recorded for each cell, detailed in Table 2.1; many as described in Boland and Murphy (2001). Cells touching the edge were removed, and a linear classifier, manually trained, removed poorly focused/segmented cells. Tracking was performed by searching for the closest centroid in the next frame. A track would stop if a movement between two frames was greater than the cell width. Cells tracked for fewer than four frames were discarded, as these would often be debris or false detections. Tracking and filtering was performed using custom written C++ scripts. In analysis of the static siRNA validation screen, features were standardised, and cells with a feature outside five SDs were removed. All data and code is available at:

bitbucket.org/samocooper/wmpaper-data-and-code/src/d35c7dc716ba?at=master.

Table 2-1, Features extracted from cell segments and used for subsequent data analysis

1	**Axial length ratio:** Eccentricity of the ellipse that is equivalent based on second order moments to the segment (Boland and Murphy 2001).
2	**Axial small length:** The width of the ellipse that is equivalent, based on second order moments to the segment (Boland and Murphy 2001).
3	**Symmetry** n, m: Normalised moments of the segment and set of polynomial functions $R_{n,m}(\rho, \varphi) = ne^{-im\varphi}$
11	**Width to length ratio:** Width divided by length.
12	**Length:** Maximum shortest path between two points on the edge, where the path is bounded by the cell edge.
13	**Width:** Double the minimum distance of the centre point to the cell edge, where the centre point is the point which is furthest away from any point on the cell edge.
14	**Roundness:** defined as $\frac{perimeter^2}{4\pi\, area^2}$ this approaches one as the shape approaches a circle (Boland and Murphy 2001).
15	Area: number of pixels in the segment.

3 Developing the tools to track live cells in higher throughput

3.1 Introduction

The most significant bottleneck in live single-cell analysis is the challenge of accurately tracking live cells over long periods of time (Cooper and Bakal 2017; Coutu and Schroeder 2013). For example in the majority of labs, researchers continue to use manual approaches; these involve making a click-based intensity recording for every individual cell, in every frame. Where hundreds of timepoints may be recorded, for hundreds of cells, in multiple conditions, this task quickly becomes intractable. To resolve this problem computational approaches to cell tracking are therefore essential (Coutu and Schroeder 2013; Hilsenbeck et al. 2016; Maška et al. 2014). However, tracking cells is not trivial. Several key factors contribute to making this a very difficult problem to solve (Hilsenbeck et al. 2016):

1. Cells can be highly motile and often make large jumps between frames. Imaging cells at higher frequencies would resolve this, but this is often not possible due to issues with phototoxicity and throughput.
2. Where multiple fields of view must be captured, typically the microscope stage is not returned to exactly the same position. Here, jitter between frames must be managed effectively.
3. Cells can undergo large morphology changes, divide, and can move into and out of the field of view. As such, there is no guarantee that an object will be present, or look the same, from one frame to the next, Fig. 3-1A.

4. Cells will also often clump together or migrate over the top of other cells. This means that errors in object detection and segmentation are frequently made; tracking must be robust to these errors, Fig. 3-1B.

To address these challenges, and develop software that would allow cells to be tracked over the long periods of time necessary to study cell fate decisions, I experimented with existing approaches to cell tracking and tested a number of software solutions (Hilsenbeck et al. 2016; Klas E. G. Magnusson et al. 2015; Maška et al. 2014). It was immediately apparent that whilst many solutions claimed to address these challenges, no single package covered the entire segmentation and tracking pipeline, and all packages required extensive programming knowledge and understanding of file-formats, even to get examples provided with the programs to work. As a result of these challenges biologists attempting to track cells would: 1) Be forced to navigate multiple programs and fine tune parameters in each of these; 2) Have to install multiple software libraries and dependencies; and 3) Have difficulties in relating tracking data back to the original time-series images. No biologists have this level of computational expertise. Furthermore, to achieve human levels of accuracy manual correction is still needed, but no readily available solutions existed that allowed simple visual correction of tracking data at the time of creating NucliTrack. Since developing NucliTrack, Hilsenbeck et al. developed stand-alone software for track correction (Hilsenbeck et al. 2016).

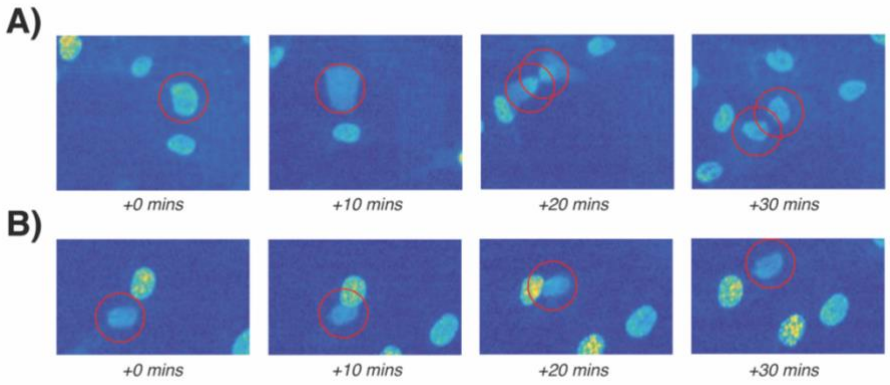

Figure 3-1, Challenges in tracking cell Nuclei: *A) Nuclei frequently undergo divisions and form two daughter cells that continue to continue to move and proliferate; B) Cells will often pass over the top of other cells, occluding nuclei for specific frame.*

I therefore decided to implement the most promising tracking and segmentation algorithms into a standard programming language, Python, and integrate these solutions into a single package that would allow users to go from time-series images all the way through to extracted time-series data in as short a time as possible. The software developed, NucliTrack (Cooper et al. 2017), cuts the time required to extract data from a single track lasting a period of several days from roughly 20 minutes to 2 minutes, giving a 10-fold improvement in throughput. In this Chapter I will outline the algorithms and development of the published version of NucliTrack. Documentation[1] and code[2] can both be found online.

[1] http://nuclitrack.readthedocs.io

[2] https://github.com/samocooper/nuclitrack

3.2 Development of NucliTrack

3.2.1 File loading

The first objective in developing NucliTrack was to construct an interface for loading sequences of image files. Frequently videos can be several hundreds of frames with each image exceeding several megabytes in file size, the memory requirement therefore exceeds that which can be loaded into local memory on most computers. To address these problems, images are read directly from drive storage for all operations, a list of pointers to the image locations is the either loaded in by the user or created by interpolating between the first and last file names in the image sequence. In loading a number of checks are performed to ensure: 1) The files are in the specified locations; 2) are not corrupted; 3) are in a readable format; and 4) do not miss any key timepoints.

Files are also created at this step that store parameters, and data on the video being tracked. The separation of the parameter file and data file, allows NucliTrack to be run from the command line in batch mode. Here chosen parameters can be used to segment and track cells automatically across an entire screen. At this stage pre-segmented label images can also be loaded, allowing the user to use external segmentation software if desired.

3.2.2 Segmentation

Nearly all automated tracking workflows can be separated into two steps: 1) The detection and segmentation of cells or their nuclei; and 2) The construction of tracks linking the detected objects over the period of imaging. Such a two-step approach is adopted in NucliTrack, as it lets the user visually inspect segmentation results, prior to tracking. The key goals for development of NucliTrack's cell segmentation interface were:

1. To develop a natural intuitive user interface that would allow the user to explore different parameter settings and understand how these changes affect the results of the segmentation pipeline.
2. For the interface to be responsive; users would ideally receive real time updates on changes that are made to parameters, and how these affect results.
3. To include the key steps for users to be able to segment cell nuclei that may be imaged in a wide variety of conditions and contain varying levels of heterogeneity between the appearance of nuclei in the same image.

These considerations together motivated the choice of algorithm to use for segmentation, and the interface design.

Watershed Segmentation

The most common method of cell detection and segmentation remains marker-controlled watershed segmentation. This method is both simple, provides intermediate visual outputs e.g. successful detection of nuclei centre, and is

computationally fast enough to be applied to hundreds of images on a typical laptop. The classical marker-controlled watershed segmentation approach involves two major components, firstly detecting foreground objects [nuclei centres] and secondly using the watershed algorithm to expand markers until they cover the whole nuclei region:

1. *Identification of foreground and background markers, Fig. 3-2 steps 1-6:* In this stage the objective is to label the center of every cell nuclei with a single unique marker. This typically involves processing the image such that a single peak of intensity corresponds to a single nuclei center in each image. Following pre-processing where illumination effects are removed, the most common step is to threshold the image into foreground and background masks, based on intensity. This is because nuclei are typically labelled with a fluorescent marker and can be easily separated from background regions of low intensity. Following thresholding, the main challenge is then de-noising and blurring the foreground image such that a single peak of intensity exists at the center of each nucleus. A peak picking algorithm is then used to identify peaks in the foreground region, with filters being used to identify peaks that are separated by either insufficient distance or drop in intensity.
2. *Application of the Watershed Algorithm, Fig 3-2 steps 7&8:* The watershed algorithm works by identifying segments that are separated by ridges of intensity, as per dropping water into a mountain range and identifying river basins. An edge filter is used to transform the image such that nuclei edges correspond to peaks in intensity, or ridges, whilst markers for nuclei are

transferred from the preceding steps. Application of the watershed algorithm then expands each marker region until the entire image is labelled. The result is that the background and each nucleus is assigned a unique integer label.

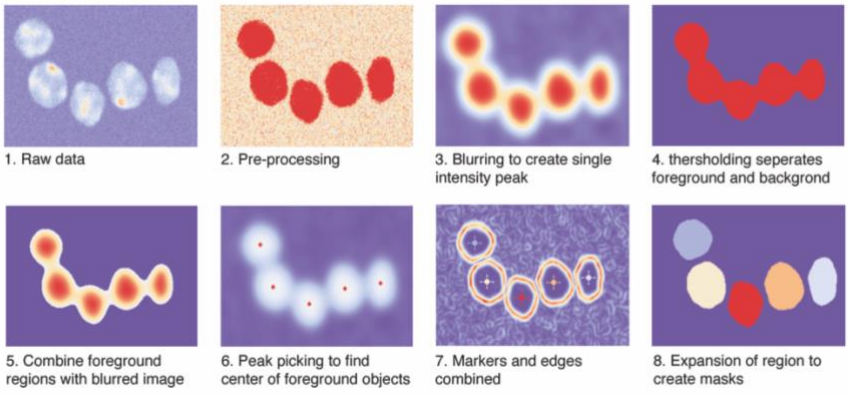

Figure 3-2, A typical nuclei detection and segmentation workflow, implemented in NucliTrack: *Raw data often contains multiple intensity spikes per nuclei and regions of very high intensity (1), clipping allows the peaks to be removed, and highlights the region covered by a nuclei (2), blurring transforms the image so that a single peak of intensity corresponds to the center of each nucleus. Thresholding separates foreground and background regions (3), and the combination of foreground region and blurred image (4) can be used to identify a single peak per nuclei (5). An edge transform is applied to the raw data to detect where the boundaries of each nucleus lie, this is weakly blurred to emphasise the most significant edges (6). By combining the foreground and background markers with the edge image (7), the watershed transform can be used to identify region masks covering each nucleus (8).*

Implementation

To implement the watershed approach in a way that would provide user feedback, I developed a segmentation pipeline that involved a series of sliders that allow the user to adjust parameters controlling each step of a marker-controlled watershed

segmentation workflow Fig. 3-3. On adjustment of the parameter value, an event is triggered that calls all of the previous segmentation steps that have not been calculated, up to the stage that is being adjusted by the user. This means that on adjustment of a slider, the results of that specific change are smoothly returned as an image that demonstrates the impact changing that parameter has, with a minimum amount of computational effort. Once parameters have been chosen in this user interface, the user can then run segmentation across the entire set of images. A loading bar gives visual feedback on progression through the set of images. Importantly, to accelerate segmentation over the entire set of images where multiple CPU's are available, I also implemented an option for the user to run segmentation of different images in parallel, using multiple CPU threads.

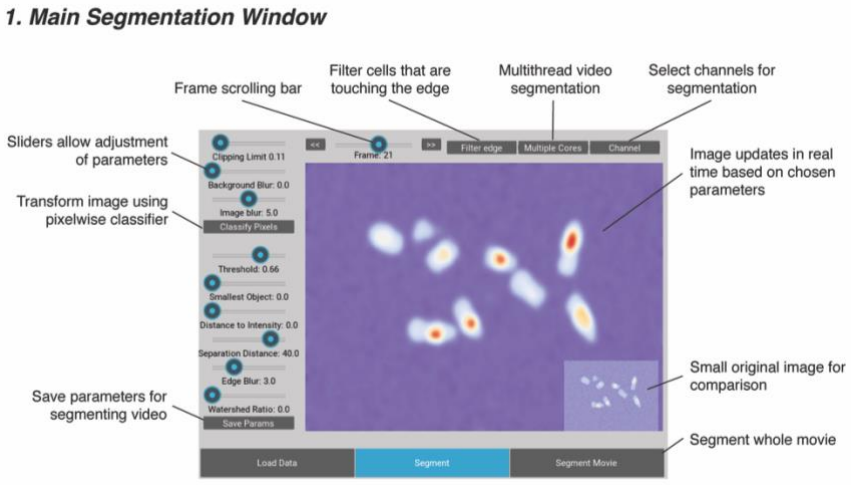

Figure 3-3, Interface for segmenting cell nuclei: *Users can control sliders that adjust the parameters used for the segmentation workflow described in Fig. 2.1, a window also exists that allows the user to compare results to the original image.*

Pixel wise classifier

On top of the initial segmentation pipeline, I also implemented a pixel-wise image classifier based off a shallow 2-layer convolutional neural network, Fig. 3-4. Here the user is able to select example regions of foreground nuclei and background noise. A neural network is then trained to classify pixels, based upon a 15x15 region of interest (ROI) around each pixel. The trained model is then used to assign a probability of foreground or background to every pixel with the transformed image being returned to users. The user is then given the option to adjust the training dataset and rerun classification or continue the segmentation pipeline on the transformed image. Overall, this pixel-wise classifier therefore allows improved segmentation results of nuclei where a high level of heterogeneity in intensity and/or background noise vs. signal exists in the image. Overall the segmentation script thus allows the user to effectively detect and segment cells using a natural and easy to use interface.

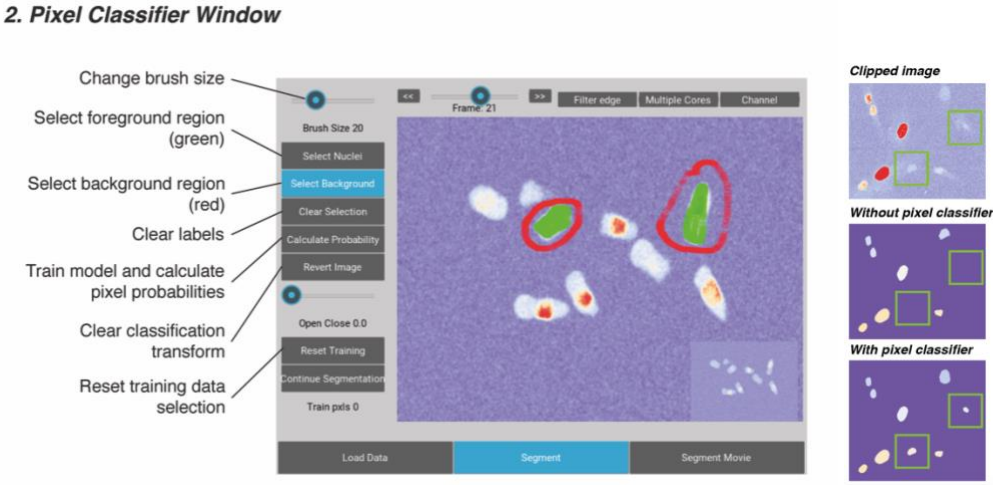

Figure 3-4, Interface for classifying pixels in image: *Users can select training data examples of foreground and background regions. A neural network is then trained to classify pixels based on a region of interest surrounding each pixel. The results are returned to the user for inspection, allowing either more or new training data to be chosen, or segmentation to be continued. This interface allows the user to identify nuclei in cases where high levels of noise exist.*

3.2.3 Feature extraction

The segmentation interface and script produce a set of labelled images, with each label corresponding to an object. To track these objects over time and quantify biological processes, features describing each segmented object must be extracted. Simple shape and intensity features were extracted using a standard library in SciKit image. Additional features were engineered to detect reporters of cell cycle progression (Fig. 4-3C, Chapter 4). Moreover, to accommodate live-cell reporters that translocate to the nucleus as a readout of activity, for example kinase translocation reporters (Fig. 1-4B, Chapter 1), I also implemented a feature that gives the ratio of

mean intensity within the nucleus, versus mean intensity in a ring region around the nucleus, as defined in Sero et al. (Sero and Bakal 2017). The full list of features is described in methods. Overall, the output of the extraction process is a matrix giving feature values for every object detected, in every frame.

3.2.4 Tracking

Tracking algorithm

Multiple object tracking in microscopy has typically lagged behind advancements in segmentation. In the majority of implementations tracks are typically constructed by linking objects from one frame to the next (Meijering, Dzyubachyk, and Smal 2012). This involves the use of algorithms that either look for large overlaps between objects in neighbouring images or proceed through K-nearest neighbour approaches in which the minimum distance between both sets of objects is sought (Hilsenbeck et al. 2016; Cooper and Bakal 2017). Whilst these approaches work well in cases where nuclei do not make large jumps between frames, they struggle in cases where nuclei are more motile, or the frequency of imaging is lower due to technical or photo-toxicity reasons.

In these cases optimisation and machine-learning approaches, developed for general multiple object tracking problems, have been successfully modified for the challenge of tracking cells (Maška et al. 2014). Such optimisation and classical machine-learning approaches remain the gold-standard in object tracking(Milan et al. 2016; Chenouard et al. 2014). In these approaches cost-functions and algorithms are

developed which seek to optimise tracks over the entire video sequence as opposed to between two frames (Bise, Yin, and Kanade 2011).

In a recent challenge focused specifically at tracking cells and nuclei, such global optimisation approaches performed best (Maška et al. 2014). Notably a probabilistic, dynamic programming approach, described by Magnusson *et al.* stood out as outperforming other methods (Klas E. G. Magnusson et al. 2015; K. E. G. Magnusson and Jaldén 2012). This method superseded others through carefully handling cases where mis-segmentation, gaps, mitotic divisions and entry to and exit from the field of view occur. This approach proceeds through the following steps:

1. Training data is selected by the user that contains examples of any erroneous segmentation, as well as normal, mitotic and post-mitotic cells. A model is then trained on this data and used to assign probabilities to every object detection in the entire image sequence, giving it a likelihood of being one of the above-mentioned classes, Fig. 3-5A. A score between nuclei in neighbouring frames is also calculated based on the distance between objects, Fig. 3-5A. This score assumes that cell movements are exponentially distributed, thus small movements are penalised exponentially less than larger movements.
2. In a forward pass, the algorithm determines the highest scoring route to any given object, based on the sum of: 1) the scores of moving from an object in the previous frame to the object in the current frame; 2) the probability of the object in the current frame being a single nuclei; and 3) the scores of objects in the previous frame Fig. 3-5B. Scores exist for the probability of the object

having emerged from a mitotic event, or entering the field of view. Exiting the field of view is handled by an additional object in the graph.

3. At the end of the forward pass the highest scoring track is chosen and added to the set of all tracks Fig. 3-5C; this may be selected as the state corresponding to a cell that has left the field of view.

The algorithm then iterates through this procedure until track scores become negative, indicating only incorrectly classified objects are left in the video. Importantly, the algorithm considers swaps between existing tracks, and thus the algorithm performs a greedy heuristic global optimisation.

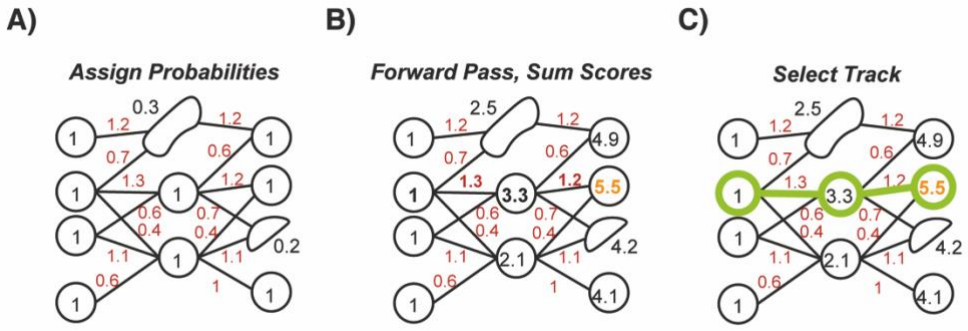

Figure 3-5, Dynamic programming approach: *A) Initially probabilities of objects being correct, as well as the cost of a transition between two locations in neighbouring frames are assigned; B) In a forward past these scores are summed in a greedy manner; C) The highest scoring route to any given object is than calculated and the highest scoring track is then added to the set of all tracks.*

Assigning classes to segmented objects

The probabilistic algorithm developed by Magnusson et al. accounts for different objects that a segmented object can correspond to by first introducing a classification step that assigns a probability of a labelled object corresponding to any one of these occurrences (Klas E. G. Magnusson et al. 2015). Tracking then looks to optimise the probability of finding a correct track over these classes, Fig. 3-5. I utilised the same set of objects as Magnusson et al., additionally introducing a new class for cells exiting mitosis. This improved detection of the overall mitotic event. Thus the object types I defined for classification are:

1. **A Single nucleus:** an accurately captured individual nucleus
2. **Mitotic event:** a detection corresponding to cell undergoing mitosis, the appearance of these can vary dramatically depending on the fluorescent marker.
3. **Mitotic exit:** an object corresponding to a nucleus that has very recently exited mitosis
4. **Noise:** False detection, corresponding to background noise or debris.
5. **Double detection:** A segmentation corresponding to two nuclei captured as one, triple detections occurred but these are infrequent enough to warrant their own class.

In the user interface segmentation results are returned to the user, and the user selects examples of each of these classes, Fig. 3-6. A classifier (support vector

machine) is then trained using the feature vector describing each object, to assign a class probability for every object in every frame.

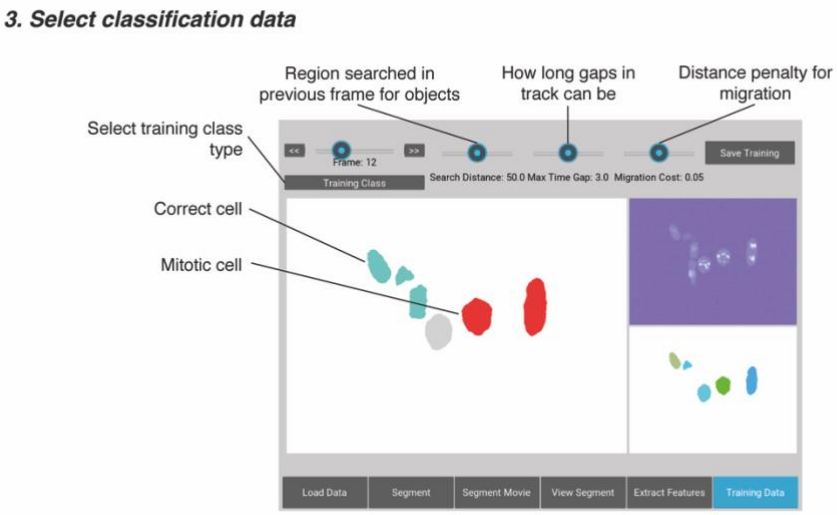

Figure 3-6, Class selection interface: *The user is presented with segmented objects and can then assign a label to the object depending on whether the object is a single cell, multiple cells, error, mitotic cell, or cell that has undergone a mitotic exit. Tracking parameters are also selected in this interface. After saving of training data, the software classifies all objects in the time-series, prior to tracking.*

Tracking Implementation

After selection of data classes, a function is then iterated through the procedure, described in Fig. 3-5, for adding a single track. This function is called by a loop in the main code that that passes the current graph of all tracks, and an array of data describing each objects location and class probability. At the end of tracking, each object in this array is assigned a track ID, corresponding to the track it belongs to, as well as a parent ID if the track emerged from a mitotic event.

3.2.5 Track correction

Whilst the software is capable of automating the majority of the tracking task, errors are still made. In studies over longer periods of times, these errors cause major inaccuracies in results. A final manual correction and inspection step is therefore necessary for allowing the user to edit the set of existing tracks and construct new tracks. The interface that I created for this had several key features:

1. The ability to follow the selected track over the course of imaging. To achieve this I randomly coloured each track individually and rendered an image that allows the user to see where errors are made as the coloured track changes between frames on objects that by eye should be in the same track.
2. The ability to see the raw images, to make decisions by eye as to what the correct track should be where coloured masks are unclear, as well as to see in the raw data effects such as changes in texture or intensity that can be seen in extracted features.
3. A graph showing extracted features from objects (nuclei) over the entire period the object is tracked for. Here, disconnected jumps in the plot highlight potential areas of tracking error. Also the user can get an early understanding of how reporters are behaving over the video and inspect the raw data where interesting changes occur.
4. The ability to add events to tracks such as the start or end of different phases in the cell-cycle, as well as selecting which tracks are accurate and should be exported for further analysis and visualisation. These are enormously helpful

for computationally synchronising time-series data to key transitions (for example, Fig 4-2A, Chapter 4).

On top of developing the visual interface for tracking cells, keyboard shortcuts were also included that allow much more efficient editing of tracks by the user. Overall the interface provides significant throughput advantages over existing software, Fig. 3-7.

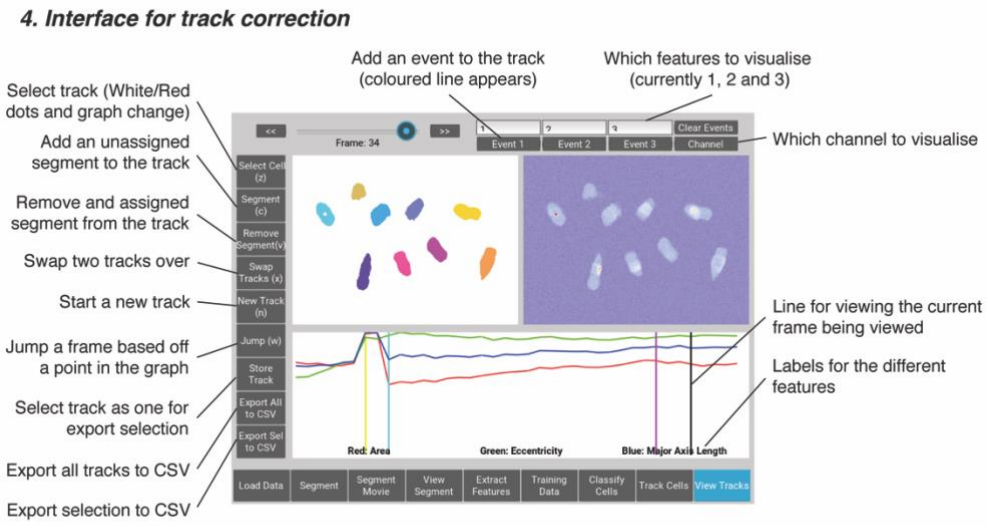

Figure 3-7, Track correction interface: *Here the user is able to inspect the results of tracking. A panel is provided that lets the user inspect the results of tracking and look for key biological changes and events, such as mitosis, or increases in reporter intensities. Functions are also provided to manually correct tracks and select data for export*

3.3 Performance and benchmarking

To ensure that my combination of segmentation approach and automated tracking were on par with existing approaches, I compared the performance of NucliTrack against the benchmark tracking challenge described in Maska et al.(Maška et al. 2014). I automatically segmented and tracked six different videos that were made available as part of the ISBI Cell Tracking Challenge 3. Each video is a time series acquisition of fluorescently-labelled nuclei, from three labelled cell lines. These lines are: GFP-GOWT1 labelled mouse stem cells, H2B-GFP labelled HeLa cells, and simulated nuclei of HL60 cells stained with Hoechst. Exported tracks were then compared to ground truth values to obtain a 'tracking precision of the method' score (TRA)(Maška et al. 2014). I compared the TRA values obtained by NucliTrack on held-out data to those obtained by the top teams in the cell tracking competition. Whilst the comparison is on held-out training data versus test data, the results indicate that NucliTrack is able to obtain tracking results of a similar standard to other top performing algorithms. Overall the performance of NucliTrack was limited by the more basic segmentation procedures I adopted. To allow for higher accuracies based off more accurate segmentation results, I therefore implemented a functionality that allows the user to import segmentation results from other potentially more accurate packages into the file loading step.

[3] http://www.codesolorzano.com/Challenges/CTC/Datasets.html

Table 3-1, Tracking results of NucliTrack: *Due to the test data not being available NucliTrack was benchmarked against available training data, specifically, parameters were chosen for a single video, and then applied to a second held-out time series.*

Competition Performance [undisclosed test data]			
Rank	N2DH-GOWT1	N2DL-HeLa	N2DH-SIM
1	0.976	0.991	0.975
2	0.925	0.986	0.957
3	0.916	0.982	0.948
Performance on Public Datasets public data [training data from the competition]			
NucliTrack	***0.952***	***0.943***	***0.970***
Teams			
KTH-SE	KTH-GE	CUNI-CZ	HD-Har-GE
HD-Gau-GE	FR-Ro-GE	PAST-FR	

3.4 Discussion

Tracking cells represents a significant bottleneck in using live-cell data to understand the dynamics of cell signalling behaviour. By developing NucliTrack as an easy to use package for extracting data from fluorescently labelled cells/nuclei, significantly higher throughput extraction of tracking data can be performed than was previously possible. Importantly, by allowing users to inspect and correct tracking data, near 100% accuracy can be achieved. This is critical in understanding and quantifying the long-term effects that specific patterns and events in signalling dynamics have upon the cell, such as fate decisions.

Within our lab NucliTrack has been used by others to explore the dynamics of Erk activity over time in cancer cells in order to understand how different Erk inhibitors affect Erk activity in a cell cycle dependent manner. Importantly this led to the discovery that Erk inhibitors were only able to downregulate Erk activity in a window that occurs after mitosis and before entry into S-phase, demonstrating cell-cycle phase and quiescence may be a critical factor in determining a cancer cells ability to resist Erk inhibitors(Simpson and Bakal 2018). NucliTrack was also used extensively to study and understand how p21 regulates passage through the restriction point and entry into S-phase of the cell cycle, the subject of the next chapter(Barr et al. 2017). Moreover, a number of other labs have also implemented NucliTrack into their live cell workflow since publication. Overall, by coupling established packages for data processing in Python, with more recent solutions to data management (h5py), graphical user interface development (kivy) and image processing (skimage), this software also provides an example of how user friendly, scientific software may be efficiently developed in an open source environment.

3.5 Methods

3.5.1 Software and libraries

To develop NucliTrack I first explored available options for which language to develop NucliTrack in and which libraries were available. Previously, I have tracked cells using custom software that I have written in C++ and Matlab. Whilst C++ has well developed libraries for image analysis, and fast compute speeds, developing user interfaces in C++ is highly challenging and typically requires allot of experience.

Meanwhile, Matlab suffers from the issue that the software is both proprietary, performance is typically slower than C/C++, and interfacing with other languages is very difficult. In contrast, Python is now the dominant language for data analysis, and additionally has abilities to integrate with web-packages, manage data storage, other languages such as C/C++, and more recently deep-learning libraries such as TensorFlow. Moreover, with a strong community behind it Python continues to be developed and maintained to a high standard, and thus was the language of choice for developing NucliTrack.

However, Python currently suffers from a lack of good cross-platform tools for developing graphical user interfaces (GUI). The most commonly used packages have been PyGame, and tKinter, although without significant effort the GUIs developed using these libraries are both clunky and difficult to interact with. More recently, a more natural and interactive GUI library kivy has been developed by a community of programmers at MIT. Kivy allows user interfaces to be easily created launched from the terminal and packaged into cross-platform applications. For these reasons I therefore chose to develop the GUI using the Kivy library.

Overall in developing NucliTrack this combination of Kivy and Python worked very effectively. Although in several cases the performance of Python was insufficient for specific image processing and rendering tasks. I therefore also included functions written in C and a hybrid C and Python language 'Cython'. These improved the speed of NucliTrack significantly. To create a self-contained distributable application, I used the PyInstaller library, although there were a number of challenges in creating an application that could be run on multiple operating systems, ultimately this goal was

achieved, and NucliTrack can now be downloaded and run on Linux, OSX, and Windows systems. NucliTrack therefore is coded predominantly in Python, contains certain optimised C functions, has a GUI written using the Kivy library, and is packaged and distributed as a standalone application that can be run on many systems.

3.5.2 Features extracted

In total 4 measures of cell shape are extracted as well as 7 measurements of the intensity and texture of each fluorescent channel imaged.

Table 3-2, Features extracted from detected and segmented object in NucliTrack

Area: the number of pixels in the mask
Eccentricity: defined by second order statistical moments
Length: the longest straight line between two points on the convex hull
Perimeter: length of the edge in pixels surrounding the region
Mean intensity: measured for each channel
Median intensity: measured for each channel

Floored mean intensity: all values below the mean are set to the mean, the mean of pixels above this value is calculated, measured for each channel
Standard deviation of intensity: measured for each channel
Floored mean standard deviation: all values below the mean are set to the mean, the standard deviation of pixels above this value is calculated, measured for each channel
Ring-region: what is the intensity of pixels in the perinuclear ring region, optionally measured for each channel
Nuclear to Ring-region ratio: what is the ratio of the perinuclear ring region intensity to the nuclear intensity, optionally measured for each channel

4 DNA damage in the mother cell causes p21 dependent G1 arrest in daughters

4.1 Background

Over the cell cycle, cells grow, replicate their DNA, and then divide into two daughter cells. Cell division, or mitosis represents the most visually striking of the phases that make up the cycle. Following division, proliferating daughter cells enter the first growth phase, termed G1 phase. During G1 phase the necessary machinery for DNA replication is produced and assembled on DNA strands in what are known as DNA replication origins. Following G1 phase, S-phase is initiated by the firing of replication origins, from which replication complexes including DNA polymerase and processivity factors such as proliferating cell nuclear antigen (PCNA) synthesise an identical copy of DNA. After replication completes, the cell enters a second G2 growth phase where the necessary machinery for mitosis is assembled, prior to mitosis occurring and the cell cycle beginning anew.

Across the cell cycle a number of checkpoints exist where cells can either continue to proliferate or enter an arrested state. Canonically, these checkpoints are: 1) The restriction point (RP) that occurs during G1 phase after mitosis. Passage through this point is associated with commitment to the cell cycle (Pardee 1974), whilst arrest typically causes entry into a quiescent G0 phase (Planas-Silva and Weinberg 1997); 2) The G2 checkpoint which occurs prior to chromosome condensation and mitotic entry. This checkpoint is traditionally associated with ensuring DNA is fully and

correctly replicated, and prevents propagation of mutations and damage to daughter cells (Norbury and Nurse 1992); and 3) The spindle assembly checkpoint (SAC) that serves to prevent aneuploidy occurring in daughter cells (Rudner and Murray 1996). Finally, an earlier S-phase checkpoint has also been proposed to exist that would prevent DNA damage being propagated into G2 phase (Katou et al. 2003; Hurley and Bunz 2009). Whilst the RP acts as the major barrier to cells proliferating in the absence of extracellular stimulus, such as the presence of growth factors, all checkpoints play a major role in balancing the need for proliferation with the maintenance of genomic stability (Hartwell and Kastan 1994). Where these checkpoints fail, cells can begin to proliferate excessively and with this acquire instable genomes that together form the basis of tumorigenesis and cancer.

At the molecular level, progression through cell cycle checkpoints is driven by the activity of Cyclin Dependent Kinases (CDKs) that phosphorylate a number of downstream targets and drive transcription of factors associated with the next cell cycle phase (van den Heuvel and Harlow 1993; Graña and Reddy 1995), Fig. 4-1A. CDKs are only active when bound to their respective Cyclins, factors whose levels oscillate over the course of the cell cycle (Nurse 1990; Gautier et al. 1990; Evans et al. 1983). Passage through each checkpoint is associated with the activity of one or more Cyclin:CDK complexes (Norbury and Nurse 1992). CycD:CDK4/6 complexes are the principal mediators of passage through the G1 restriction point (Meyerson and Harlow 1994; Sherr 1993; Matsushime et al. 1991). CDK4/6 activity drives hypo-phosphorylation of Retinoblastoma protein (Rb) (Matsushime et al. 1992), that in turn releases the bound transcription factor E2F (Weinberg 1995; Dyson 1998;

Chellappan et al. 1991). E2F causes upregulation of Cyclin E that binds to CDK2 (Geng et al. 1996), further phosphorylates Rb protein and establishes a positive feedback loop that commits the cell to S-phase, and therefore cell cycle entry (Arthur B. Pardee, Li, and Reddy 2004; Johnson and Skotheim 2013). Progression through the G2 checkpoint meanwhile is associated with the buildup of Cyclin A/B levels that in turn induce CDK1 activity (Jackman et al. 2003). This, in turn, leads to phosphorylation of a number of targets associated with mitotic entry (Ubersax et al. 2003). Critically one of these is the Anaphase Promoting Complex (APC) (Zachariae and Nasmyth 1999). Following correct kinetochore attachment during mitosis, activation of the APC occurs through silencing spindle assembly complex signalling (Acquaviva et al. 2004). A set of events is then set in motion that leads to ubiquitination of CycA/B, as well as a host of other factors that block re-entry into mitosis, by the APC that together ensure robust transition into G1 (Q. Yang and Ferrell 2013). In turn, daughter cells begin the next G1 phase, and again make the decision to pass through the restriction point or enter a quiescent state. Taken together, the activity of different Cyclin:CDK complexes over the course of the cell cycle drives passage through key checkpoints, and entry into mitosis that splits the cell into daughters ready to begin anew.

Negative regulation of Cyclin:CDK activity is also observed at each checkpoint, Fig. 4-1B. Two families of Cyclin:CDK inhibitors (CKIs) have so far been identified. These are the INK4 family of genes that contain p15, p16, p18 and p19, and the CIP/KIP family of genes containing p21, p27, and p57 (Besson, Dowdy, and Roberts 2008). The INK4 genes are primarily associated with inhibition of Cyclin D:CDK4/6 complexes and restriction point passage. Meanwhile, the CIP/KIP genes have been shown to

inhibit all Cyclin:CDK complexes to varying degrees (Sherr and Roberts 1999), and have typically been associated with cell-cycle inhibition in response to a specific regulatory pathway (Besson, Dowdy, and Roberts 2008).

The CIP/KIP protein p21 is best known for its role in the DNA damage pathway. In response to DNA damage, protein expression levels of the tumour suppressor p53 increase, through phosphorylation and inhibition of its ubiquitination and degradation by the E3 ubiquitin ligase MDM2 (Kastan et al. 1991; Shieh et al. 1997). Increased levels of p53 in turn lead to transcriptional upregulation of p21 expression (el-Deiry et al. 1993), which in turn can cause p53 dependent G1 (Waldman, Kinzler, and Vogelstein 1995) and G2 arrests (Bunz et al. 1998). That are hypothesised to occur through p21 inhibiting all Cyclin:CDK complexes to varying degrees (Xiong et al. 1993), as well as p21 inhibiting the DNA replication factor Proliferating Cell Nuclear Antigen (PCNA) (Waga et al. 1994). In contrast, p27 has been implicated in the ability of cells to arrest in response to serum starvation (Coats et al. 1996), and was identified as the major CKI that is upregulated following stimulation of fibroblasts by TGF-b in serum starved conditions (Toyoshima and Hunter 1994). Unlike p21, p27 is generally associated with a G1 specific arrest rather than the ability to arrest the cell cycle beyond the restriction point (Toyoshima and Hunter 1994). Specifically, *in vitro* p27 has been demonstrated to bind CDK1 (Toyoshima and Hunter 1994), but in cells p27 has been found to bind CDK4/6 and CDK2 complexes, but not CDK1 (Soos et al. 1996). Moreover, p21 binds CDK4/6 and CDK2 in animal systems, in contrast p21 inhibition of CDK1 in animals had only been seen where levels have been elevated through depletion of the E3 ubiquitin ligase SCF-Skp2 that is known to

degrade p27 (Pagano 2004). Finally unlike p21 and p27 that are universally expressed in mammalian tissues, the CDK inhibitor p57 displays distinct and differing tissue localisation patterns in developing organisms and adults, thus implicating it in a developmental role (Balint et al. 2002; Georgia et al. 2006; Gosselet et al. 2007; Besson, Dowdy, and Roberts 2008). Again like p27 it is typically implicated in arrest at the restriction point, for example p57 is required for TGF-beta induced G1 arrest in hematopoietic stem cells (Scandura et al. 2004). Thus, whilst evidence supports the fact that p21 inhibits all cell cycle checkpoints in response to DNA damage, p27 and p57 typically act at the restriction point and are influenced by growth conditions and developmental processes respectively. Notably though, experiments demonstrating p21's ability to inhibit all checkpoints used exogenous DNA damage to stimulate cell cycle arrest both in cell and animal models (Bunz et al. 1998; Waldman, Kinzler, and Vogelstein 1995; Xiong et al. 1993; Abbas and Dutta 2009). Thus, the question of which checkpoints and to what degree p21 inhibits cell cycle progression in cells in response to endogenous damage remains open.

Figure 4-1, Cyclin CDK activity drives cell cycle progression: *A) Cyclin levels oscillate over multiple cell cycles. Over a single cell cycle, Cyclin-D activity is highest over the restriction point, afterwards allot of uncertainty surrounds whether its levels stay high or if it is degraded. Cyclin E meanwhile peaks over G1/S transition and is removed shortly after this (Barr et al. 2016). Cyclin A and B levels accumulate over S phase and G2, drive transition through mitotic entry, and play a role in regulating the timing of mitotic events; B) A simplified canonical model for regulation of cell cycle progression. In this dynamical system, Cyclin:CDK activity drives activation of downstream transcription factors such as E2F, that in turn drive expression of subsequent cyclins. Mitogen signalling initiates the cell cycle by driving Cyclin D transcription. The INK4 proteins act primarily against Cyclin D: CDK4/6 complexes whilst, the CIP/KIP proteins p21, p27 and p57, are able to inhibit all Cyclin:CDK complexes, although the timings and relative strength of inhibition is uncertain.*

Correct regulation of the cell cycle checkpoints is critical for balancing the need for proliferation, required for growth and repair, with the build-up of endogenous DNA damage and mutations that can occur over successive rounds of replication and

cause cancer (Hanahan and Weinberg 2011). For example, one of the first genes to be implicated in cancer, and the first example of a tumour suppressor gene was that of Rb, a protein critical in suppressing E2F activity and passage through the restriction point (Weinberg 1995; Murphree and Benedict 1984; Cavenee et al. 1983). Here, mutations in Rb protein were almost exclusively correlated with the onset of pediatric retinoblastomas (Knudson 1971). The tumour suppressor p53 is also perhaps the best studied of all tumour suppressors (Hollstein et al. 1991), and has been implicated in over half of all cancers (Vogelstein et al. 2013). Moreover, familial mutations in p53 are associated a wide range of early onset cancers, generally defined as Li-Fraumeni syndrome (Malkin et al. 1990; Li et al. 1988). Mutations in driver genes, such as Raf and Ras (Hall et al. 1983; Bos 1989; Davies et al. 2002; Wan et al. 2004), also have a vast body of evidence behind them implicating them in tumorigenesis. These driver mutations typically drive Cyclin-D expression Cyc-D:CDK4/6 activity, which together pushes the cell through the restriction point (Massagué 2004). Together, this demonstrates the importance of correctly balancing proliferation against the build-up of somatic mutations in tumour suppressor genes such as TP53 and oncogenes such as Raf and Ras.

Yet despite their major role in cell cycle progression, evidence shows that the dysregulation of CKIs such as p21 and p27 has a much weaker effect on tumourgenesis (Abbas and Dutta 2009). For example, whilst p21 showed is directly responsible for p53 dependent cell-cycle arrest (El-deiry et al. 1993), studies have also shown p21 is not responsible for p53 dependent apoptosis (Deng et al. 1995), Fig. 4-2. Moreover, whilst Caballero et al. showed *CDKN1A*$^{-/-}$ mice spontaneously

develop tumors, this happens with late onset suggesting CDKN1A loss alone is insufficient for tumourgenesis as opposed to p53 (Caballero et al. 2001). The fact that Ras driven cancers are greatly accelerated by CDKN1A loss (Adnane et al. 2000), and CDKN1A null mice are much more susceptible to chemical induced tumorigenesis (Topley et al. 1999; Poole et al. 2004; Jackson et al. 2002), further supports the notion that p21 mutation must be accompanied by mutation in another tumor suppressor or driver gene. Yet, complicating the picture more recent studies that find p21 deletion can actually increase the lifespan of mice with telomerase mutations, without increasing levels of cancer (Choudhury et al. 2007). Taken together, these studies highlight that whilst p21 has tumor suppressive roles, these are not nearly as strong as those observed for TP53, and in certain cases p21 loss can be beneficial.

Mutations in CKI's also have a much weaker association with familial cancer diseases (Rainville and Garber 2008). For example, polymorphisms around CDKN1A, CDKN1B genes have been associated as risk-factors for colorectal, breast and prostate cancer (Dunlop et al. 2012; Ma et al. 2006; Kibel et al. 2003), but compared with TP53 mutations in Li Fraumeni syndrome these effects are very minor (Malkin et al. 1990). Similarly, CDKN1B familial mutations have only recently been implicated in a subset of patients harbouring multiple endocrine neoplasia. Here tumours and excessive growth are observed in multiple endocrine glands (Georgitsi et al. 2007). Familial p57 mutations are the exception in that they induce lethal developmental defects (Yan et al. 1997; Zhang et al. 1997), and mutations in the CDKN1C that encodes p57 are associated with familial Beckwith-Wiedemann syndrome, a severe developmental disease (Lam et al. 1999). Though together the remarkable lack of somatic mutations

in CKIs in cancer stands out. In fact, the CKI p21 is more-frequently over-expressed across all tumours than under-expressed, (Forbes et al. 2011), and the presence of CDKN1A is in fact critical for maintenance of stem-like population of colorectal cancer cells that are able to effectively resist chemotherapy (O'brien et al. 2012). thus, major questions remain in our understanding of the role of CKIs, notably p21, in cell-cycle regulation and the progression of cancer.

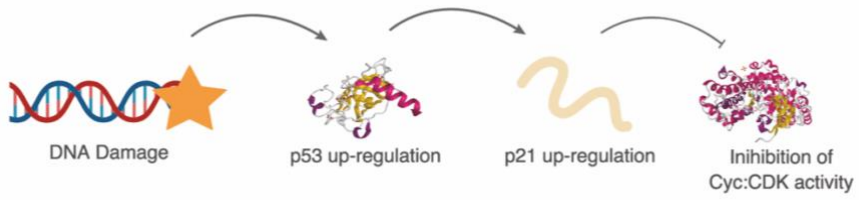

Figure 4-2, DNA damage leads to cell cycle arrest: *In the canonical DNA damage response, damage leads to stabilisation and upregulation of p53, which in turn transcriptionally upregulates p21 protein which in turn inhibits Cyclin CDK complexes.*

These studies collectively highlight that we have identified the major players in cell cycle control and determined whether they positively or negatively regulate checkpoint progression in response to artificial conditions such as the addition of DNA damaging agents. Yet they also show that we have very little understanding of the dynamics of these checkpoints and how they function in normal cycling cells (Nurse 2000). For example how is it that cell cycle checkpoint are robust and irreversible, meaning that factors expressed prior to checkpoint passage cannot be re-expressed after passage (Tyson, Csikasz-Nagy, and Novak 2002), and that transitions are delayed until the cell is ready and able to commit to the next phase of

the cell cycle? Here, mathematical modelling has demonstrated that both positive and negative feedback loops are critical to such decision making, and that each transition can be considered as a bistable state (Tyson, Chen, and Novak 2003; Sha et al. 2003; Chen et al. 2004). Specifically, negative feedback loops create a threshold level of CDK activity that must be surpassed to induce transition into the next phase, whilst positive feedback loops then ensure robust progression, and block passage back into the preceding phase (Tyson, Chen, and Novak 2003). Understanding what these feedback loops are, how they create a threshold level that can be surpassed by positive regulation, and how their dysregulation contributes to disease remains a largely open question (Kolch et al. 2015; Kholodenko 2006).

Live single-cell experiments allow us to quantify how the activity and levels of signalling factors result in cell-fate decisions in individual cells, without the use of exogenous agents needed to synchronise entire populations (J. E. Purvis and Lahav 2013; Cooper and Bakal 2017; Gaudet and Miller-Jensen 2016). As such, they are an ideal tool for understanding the dynamics of cell-cycle processes. Specifically, by using live cell analysis to analyse how normal, unperturbed, cycling cells behave in response to varying p21-levels over the course of the cell cycle we sought to resolve the uncertainties and paradoxes surrounding how p21 is linked to cell-cycle checkpoint control and cancer.

In this line of research, a live-cell study by Spencer et al., in unperturbed cycling cells, demonstrated that a bifurcation in CDK2 activity after mitosis causes a subpopulation of cells to enter a p21-dependent post mitotic quiescent arrest state prior to restriction point passage (Spencer et al. 2013). Here, the authors examined

this arrest state in response to mitogen withdrawal and implicated this p21-dependent arrest as a response to mitogens since they observed after mitogen withdrawal that a number of cells shifted from a CDK2 low state to a CDK2 high state. Yet, this work stopped short of demonstrating that p21 depletion or knockout further sensitises cells to mitogen withdrawal, thus leaving the field open to alternative explanations. Independently and prior to this, Lahav et al. demonstrated that in a live-cell system, in response to DNA damage, p53 displays pulses of expression (Lahav et al. 2004). Where these are transient pulses, cells continue to cycle. However, following more frequent pulses, or sustained activity, cells will either arrest or undergo apoptosis; this being dependent on the rate of p53 increase (Jeremy E. Purvis et al. 2012). Given the canonical pathway, in which p53 induces p21 expression that then causes cell cycle arrest, we hypothesised that p21 would act to integrate pulsatile p53 signalling over time, demonstrating slower non-pulsatile dynamics, and then be responsible for the observed bifurcation of CDK activity and the overall proliferation-quiescence decision that Spencer et al. observed, but attributed principally to mitogen signalling withdrawal. To test this hypothesis, we looked to record p21 levels in live, unperturbed, cycling cells, over time and analyse how levels were regulated and corresponded with cell cycle arrest.

Alexis Barr working in the Bakal lab conceived this project, following previous work looking at G1/S transition in Hela cells (Barr et al. 2016), and conducted all wet-lab experiments. I used the NucliTrack software that I developed to analyze and interpret the live cell imaging data that Alexis collected, including manual inspection and correction of every track. Additionally, I created all figures and graphs relating to

live cell data, with the exception of Fig. 4-7B. My contribution increased the number of cells being tracked by ~10 fold compared with previous projects in the Bakal lab and allowed us to ask questions about the dynamics of p21 levels over entire populations of cells rather than just a select handful of cells (Cooper et al. 2017). We also collaborated with Stefan Heldt from Bela Novak's group, to model the time series data being produced and better understand the role of p21 in the establishment and passage of the restriction point, and other cell cycle checkpoints.

4.2 Results

4.2.1 Quantifying p21 levels over the cell cycle

To understand how p21 levels vary in live cells over time, a GFP tag was introduced into the C-terminus of both alleles of the CDKN1A gene, at the endogenous loci, in hTert-RPE1 cells using a CRISPR system. Visual inspection of the p21 reporter indicated it to be exclusively nuclear, matching immunofluorescent staining of p21, Fig. 4-3A. Population growth was also unaffected, and cell cycle times matched the control cell line. Moreover, *in vitro* experiments demonstrated that p21 binding to CDK was not disrupted either. Together these experimental studies indicated that the GFP tag did not interfere with p21 functionality (Barr et al. 2017).

The GFP reporter was introduced into hTert-RPE1 cells that already contained an mRuby reporter on a single allele of the gene Proliferating Cell Nuclear Antigen (PCNA) (Zerjatke et al. 2017). PCNA is a key protein in DNA synthesis; with levels of PCNA shown to progressively increase over S-phase. Quantitative analysis of another

hTert-RPE1 cell line expressing both mRuby-PCNA, and Cyclin-A GFP, suggested that the point at which PCNA levels begin to increase is precisely timed with the start of S-phase Fig. 4-3B. A very small jump in the foci feature is also observed here Fig. 4-3C. Therefore, by marking where PCNA levels begin to increase G1/S transition can be identified. PCNA develops distinct foci structures towards the end of S-phase that are rapidly disassembled over the S-phase / G2 transition, the feature engineered in NucliTrack is able to quantify the presence or absence of these foci, and allows G2/S transition to be demarcated, Fig. 4-3C. The major changes in both area and intensity, associated with nuclear envelope breakdown, then allow the mitotic event to be identified and G1 and G2 phase to be defined. As such the three growth phases of the cell cycle could all be determined with a single PCNA channel, Fig. 4-3C. This represents an important advance as previously two channels have typically been required, for example, the commonly used FUCCI system uses a combination of GFP-geminin for marking S and G2 phase and RFP-cdt1 for marking G1 phase (Zeilke et al., 2015). Thus, by combining the dual p21, PCNA reporter cell line developed by Alexis with NucliTrack, we were able to quantitatively analyse the relationship between p21 and cell cycle progression across large populations of unperturbed, asynchronously growing cells.

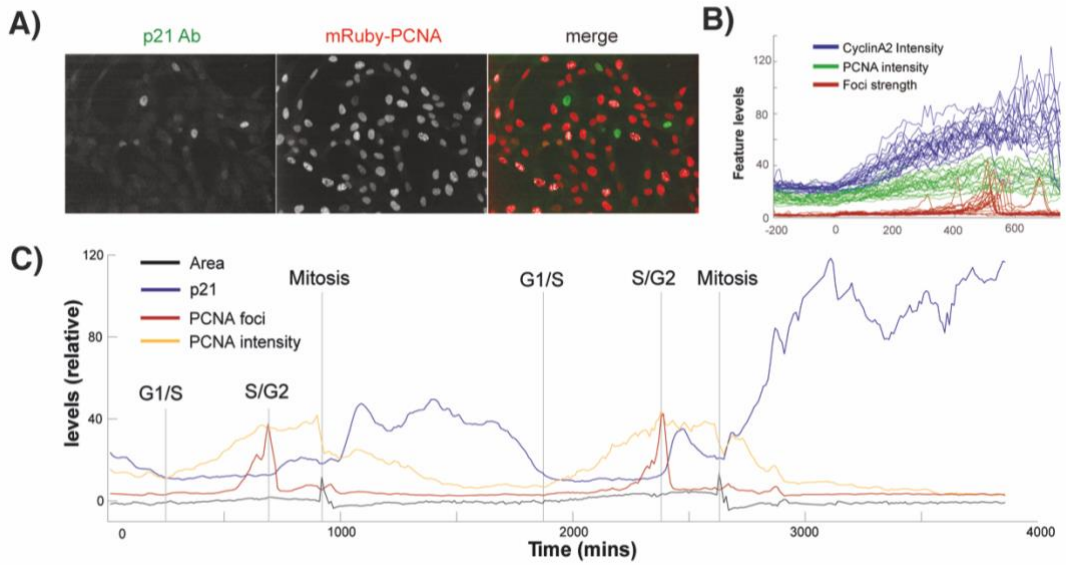

Figure 4-3, Quantifying the cell cycle: *A) p21 and PCNA can only be detected in the nucleus in hTert-RPE1 cells; B) increasing levels of PCNA correlate with the start of S-phase as marked by increasing cyclin-A level; C) Using only a PCNA reporter we are able to capture every phase of the cell cycle, at G1/S PCNA levels increase, at S/G2 the engineered PCNA foci measure peaks prior to disassembly of replication foci, mitosis can be identified by gross change in morphology.*

4.2.2 Heterogeneity in p21 levels

Tracking of wild-type populations of tagged hTert-RPE1 cells with NucliTrack revealed notable variability in p21 levels over the cell cycle, Fig. 4-4A. Levels of p21 were greatest and most variable in G1 phase and demonstrated strong positive correlation with the length of G1 phase (R = 0.62**; Methods), meaning that cells with higher levels of p21 remained in G1 phase for longer periods of time, Fig. 4-4B. In S-phase p21 levels were undetectable in all imaged cells, indicating the presence of strong protein degradation and/or suppression of transcription. G2 phase

demonstrated intermediate levels of p21, and a significant positive correlation between p21 levels and G2 length (R = 0.51**).

Notably, in contrast to previous reports we observed no degradation of p21 over mitosis, as such, p21 levels were inherited from mother to daughter. Indeed we observed strong correlation between the highest p21 levels reached in mother G2 phase and daughter G1 phase (R = 0.75**), Fig. 4-4C. p21 levels were also highly correlated between sister cells (R = 0.81**), as did length of G1 phase between sister cells (R = 0.53**), indicating inheritance of factors corresponding to G1 length and p21 levels. Unusually, given correlation between mother and daughter cell p21 levels, p21 levels in G1 with G1 length, and p21 levels in G2 and G2 length, we observed very little correlation between mother G2 length and G1 length (R = 0.07, P = 0.36, Pearson's Correlation). Instead, very high levels of correlation were observed between the highest p21 levels reached in G1 phase and the combined length of mother G2 phase and daughter G1 phase (R = 0.68**). In fact, this correlation was stronger than that observed in either phase individually (G1: R = 0.61**, G2: R = 0.51**). This indicated that variation in G2 and G1 length was caused by noise in the timing of mitosis, Fig. 4-4D. As such, we found that the maximum p21-GFP level reached in the G1 phase of daughter cells is most reflective of the time period between the end of S-phase in the mother cell and S-phase entry in daughter cells. That we observed much weaker correlations between phase lengths before and after S-phase, therefore suggested that stochastic events occurring in S-phase determine the p21 levels in, and the length of, the combined mother G2 and daughter G1 phase.

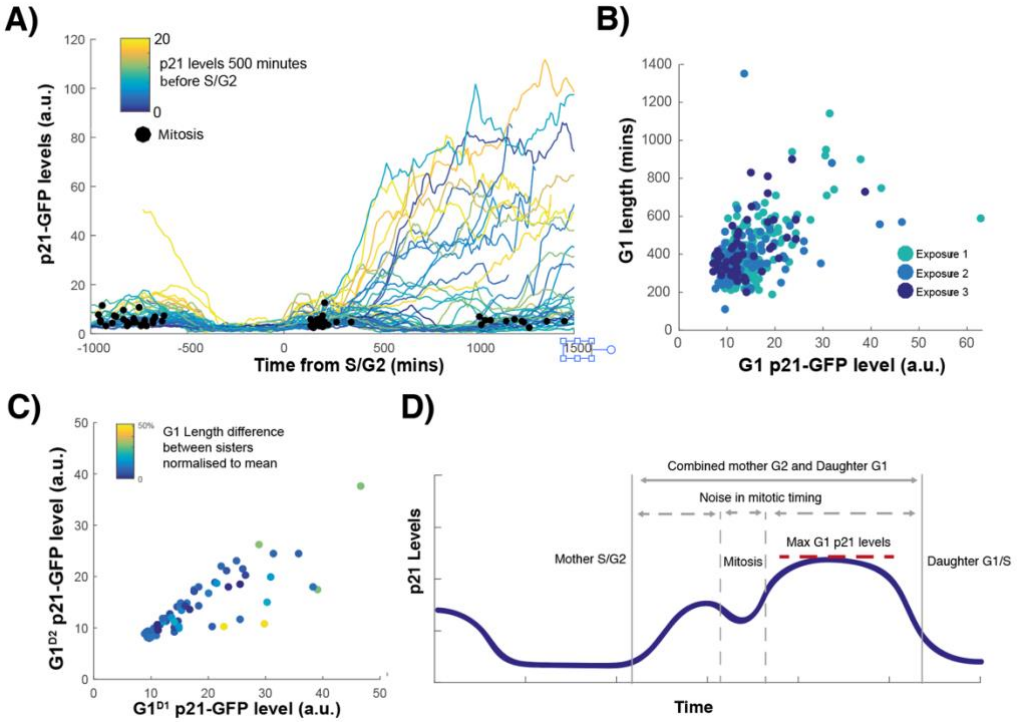

***Figure 4-4*, p21-GFP levels over the cell cycle:** *A) Tracks of p21 levels over the cell cycle are computationally synchronised to S/G2 transition, identified by the loss of PCNA foci. We observed that following S/G2 transition the majority of cells did not express high levels of p21, however, in some cases p21 levels increased prior to mitosis (Black Marker) often these then increased significantly after mitosis. p21 levels were below baseline over S-phase in all cells measured, no G2 arrest were observed; B) The maximum levels of p21 reached in G1 phase demonstrated strong positive correlation with G1-phase length (Pearson's Correlation R = 0.62** (P < 0.01). We saw no change in the distribution of cells at different exposure intensities (Exp1, n = 206 cells; Exp2, n = 90; Exp3, n= 52 (Methods)), indicating exposure levels were below those that caused phototoxic effects on p21 levels and phase length; C) A high level of correlation was observed between sister cells in both p21 levels and phase length (n = 62, Pearson's Correlation R = 0.81** (P < 0.01)). D) The strongest correlation between maximum p21 levels in G1 phase and phase length occurred between the combined mother cell G2 phase and daughter cell G1 phase. This suggests events over mother cell S-phase determine the time period until G1/S transition in daughter cells, whilst a level of uncertainty not dictated by these events exists in the timing of mitosis.*

4.2.3 A p21 high G1 arrest state

Aligning tracks of p21 levels to mitosis immediately reveals that in a fraction of cells p21-GFP accumulates to a high level shortly after mitosis and these cells then enter an arrest state that we define as G1 post mitotic arrest, where G1 length is greater than 600 minutes, Fig. 4-5A. We found these cells also had hypophosphorylated Rb indicating this arrested state to be a block before the restriction point. Consistent with p21 levels being inherited by daughter cells, linear regression between the maximum p21 level in the mother G2 phase and entry to the G1 post mitotic arrest state displayed significant positive correlation (R = 0.43**).

Following both daughter cells after mitosis, we also found that in 70% (45/64) of cases both daughters continue to proliferate. In 16% (10/64) of cases both daughter cells enter a G1 arrest (twin arrest), whilst 14% of the time (9/64), one daughter arrests and the other cycles (single arrest). If arrest were to occur by chance following mitosis, we would expect fewer than 4.4% of mitoses to result in twin arrest (21% chance of any cell arresting, 0.21^2=0.044). As such, conditions for arrest are frequently inherited by both daughter cells. Finally, where either single or twin arrest occurs in daughter cells, mean p21-GFP levels are higher in the mother cell G2, than in the mother cell G2 where daughter cells continue to proliferate (twin arrest = 9.6 a.u.; single arrest = 8.0 a.u.; cycling = 5.9 a.u), Fig. 4-5B. Thus, we propose that factors contributing to G1pm arrest are detected prior to mitosis and the single arrest we observe is caused by asymmetric inheritance of these factors. Although p21 levels can increase further in G1 to promote G1pm arrest. Taken together, these observations suggest that the decision between proliferation and arrest in daughter cell G1 phase,

in unperturbed cells, is regulated by p21 levels, which in turn are determined by events occurring in the mother cell S-phase.

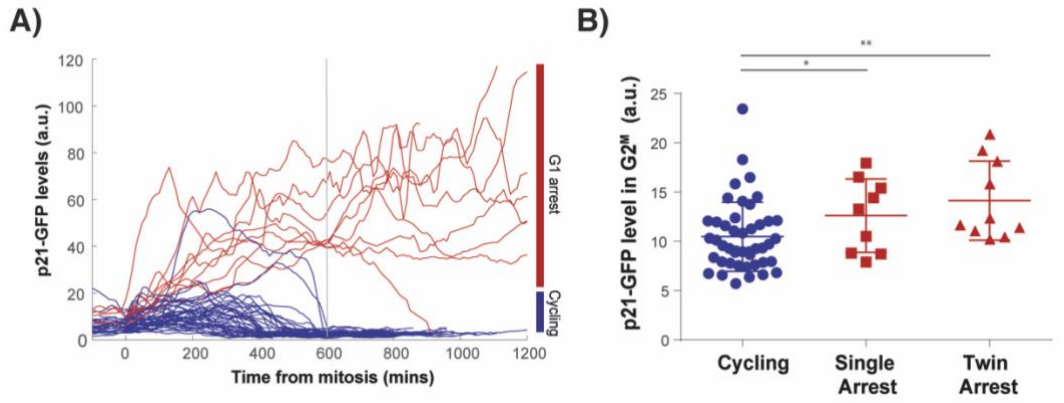

***Figure 4-5,* A p21 high arrest state:** *A) Tracks of p21 levels in unperturbed hTert-RPE1 were computationally aligned to mitosis. Two distinct subpopulations emerged, those where p21 levels increased to very high levels, and cell entered a G1 arrest state that we qualitatively defined as having a G1 length > 600 minutes, and those that continued to cycle (n = 51 cells); B) Average p21 levels in cells that continued to cycle were lower than those where either daughter arrested. Average G2 p21 levels were even higher where both cells arrested. Cycling (n = 49 cells; single arrest, n = 9; twin arrest, n = 10). Significant differences are observed between arrested and cycling states using a two-sample t-test on log-transformed data. Error bar is s.d. (*P < 0.05, **P < 0.01).*

4.2.4 The tumour suppressor p53 drives heterogeneity in p21 levels in cycling cells

The most well studied regulator of p21 levels is the tumour suppressor p53, due to its role in the DNA damage response. To understand the degree to which p53 levels are responsible for p21 transcriptional upregulation we depleted p53 with siRNA in populations of hTert-RPE1 cells. Following p53 depletion we used NucliTrack follow p21 levels in asynchronous cells, and observed a major reduction in basal p21 levels,

Fig. 4-6A. Notably the length of G1 also became similar across all cells tracked and we observed no cells entering a G1 arrested state, Fig. 4-6B. Thus, we found that p53 accounted for all detectable heterogeneity in p21 levels, variation in G1 phase length, and entry into G1 arrest states in unperturbed hTert-RPE1 cells.

To further validate this finding, we monitored p21 following addition of the compound Nutlin-3. Nutlin-3 is an inhibitor of the interaction between p53 and the E3 ubiquitin ligase MDM2 that is responsible for degradation of p53 and suppression of p53 levels in the absence of DNA damage (Vassilev 2004; Kojima et al. 2006). Thus, by adding Nutlin-3, p53 levels increase, and we hypothesised we would see an increase in p21 levels. Indeed, this was the case, as demonstrated by a drastic net increase in p21 levels following addition of nutlin 3 to hTert-RPE1 cells, Fig. 4-6D. However, only cells in G1 and G2 phase on Nutlin-3 accumulated p21 continuously. Cells in S-phase meanwhile maintained low levels of p21 over the course of S-phase before rapidly accumulating on S/G2 transition. In these cells exiting S-phase, we observed that a fraction entered a G2 arrest state, consistent with the observation that following exogenous DNA damage p21 can cause a G2 arrest (Bunz et al. 1998). Most strikingly, by computationally synchronizing cells exiting S-phase we found that G2 arrest was associated with the rate of p21 increase, Fig. 4-6E. Specifically, in those cell with the highest rate of p21 accumulation G2 arrest was observed, whilst in those with lower accumulation rates, cells would undergo mitosis and arrest in the subsequent G1. Potentially indicating competitive feedback, against factors promoting mitotic entry. As such, by analysing how p21 levels and cell-cycle progression is affected by p53 suppression, and up-regulation, we provide very

strong evidence that observable heterogeneity in p21-GFP levels in hTert-RPE1 cells is caused by p53. Moreover, our data supports the notion that factors block p53 dependent, p21 expression, over S-phase whilst DNA replication is still occurring.

4.2.5 G1 arrest requires p21, however G1 delay is p21 independent

Having established that both G1 arrest and delay is dependent on p53 we looked to understand the role of p21 in both scenarios. We depleted p21 using siRNA again, and this time observed that whilst less cells entered a G1 post mitotic arrest (5/75 for p21 siRNA versus 11/64 for NT siRNA, P=0.0536), variability in G1 length remained unaffected. To validate these finding Henriette Stoy, also generated p21 knockout cell lines, using two CRISPR/cas9 induced deletions, termed p21 knockouts 1 and 2. Again, no change in G1 length heterogeneity was observed, Fig. 4-6C, and an even stronger reduction in the number of cells arresting was observed, this time demonstrating a high level of statistical significance, with only 1 out of 172 cells entering an arrested state ($P < 0.001$). Thus, we found for the first time that in hTert-RPE1 cells, G1 delay occurs through p21 independent mechanisms whilst G1 arrest in asynchronous unperturbed cells is p21 dependent.

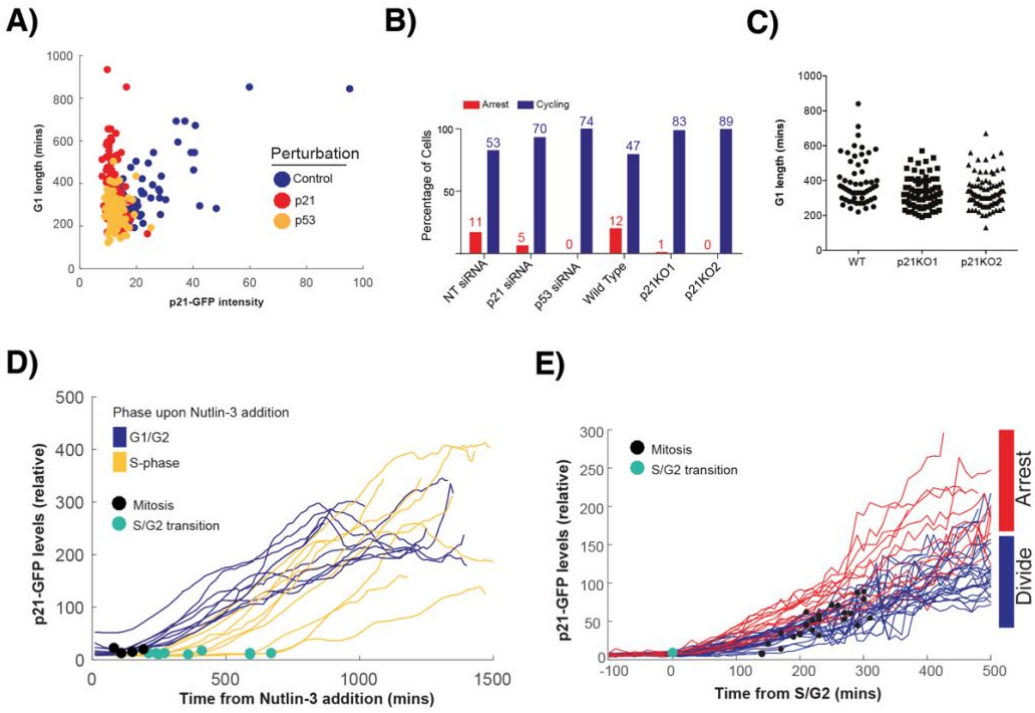

***Figure 4-6*, The effects of p53 and p21 on cell cycle progression:** *A) Following depletion of p53 (yellow, n=81) we observed a significant reduction in both p21 levels and the length of G1 phase versus control populations (blue, n=75). In contrast, following p21-GFP depletion, no change in G1 phase length was observed, despite the expected loss in p21 signal (red, n=90); B) Depletion of both p21 and p53 led to a significant reduction in the number of cells entering a G1 post mitotic arrest. 'N-1' Chi-squared test for divergence between arrest (assumed binomial probability) in NT and Wild Type vs. p21 and p53 siRNA/KO gave; NT:p21 siRNA: P = 0.0536, NT:p53 siRNA: P < 0.001, WT:p21 KO1: P < 0.001, WT:p21 KO2: P<0.001); C) G1 phase length in both p21 knockout lines was not statistically different from control populations; D) Following addition of nutlin-3 that is known to increase p53 signalling, we observed am immediate and dramatic upregulation in p21 levels, in cells in G1/G2 phase, that increased to higher levels than those observed in wild-type populations, (>200 a.u. Versus typical heightened levels of ~100 a.u. However, where cells were in S-phase p21 levels remained low until S-phase exit (teal marker) and entered a G2 arrest indicated by the absence of mitosis (black marker); E) Cells were computationally synchronised to S/G2 transition to highlight the transition dynamics. We found G2 phase arrest to be strongly associated with the rate of p21 increase following S/G2 transition.*

4.2.6 High levels of p21 inhibit CDK2 activity

Cyclin:CDK complexes represent the best characterised inhibitory target of p21. Importantly biochemical studies have demonstrated that p21 is a strong suppressor of Cyclin E:CDK2 activity. For example, Spencer et al. demonstrated that CDK2 activity bifurcates in cells post-mitosis (Spencer et al. 2013). However, in this study the cause of CDK2 activity bifurcation, in multiple unperturbed cycling cell lines was largely attributed to mitogen signalling levels. We hypothesised that in hTert-RPE1 cells we would observe a bifurcation in CDK2 activity post-mitosis in unperturbed cycling cells and that this would be p21 and p53 dependent. To test this hypothesis the CDK2 reporter used by Spencer et al. was introduced into hTert-RPE1 cells expressing mRuby-PCNA only. This reporter translocates out of the nucleus on phosphorylation by CDK2, Fig. 1-4B. To quantify the activity of CDK2 in these cells, an additional feature therefore needed to be developed within NucliTrack that reported the ratio of intensity inside the nucleus, versus the intensity in the perinuclear ring region outside of the nucleus.

In wild-type populations, this feature gave an effective readout of CDK2 activity, indicated by detection of the sudden drop in nuclear reporter intensity post-mitosis in the majority of cells, indicating loss of CDK2 activity associated with APC mediated degradation of Cyclins, Fig. 4-7A. In line with the findings of Spencer et al. we also observed a CDK2 low state to emerge post-mitosis, and the percentage of cells entering this state was insignificantly different to the number of cells p21-GFP tagged hTert-RPE1 cells that we observed entering a p21 high post mitotic arrest state.

Heterogeneity in CDK2 activity levels post mitosis was also observed. Cells that had lower levels of CDK2 activity for longer, had delayed G1 phases, Fig. 4.7A.

To test whether this CDK2 low state was p53 dependent we depleted p53 with siRNA and followed CDK2 activity levels post mitosis. In line with our previous observations we observed that following p53 depletion, G1 phase length became homogeneous, Fig. 4.7A. Moreover, we observed that CDK2 activity levels also has reduced variability, increasing rapidly in all cells following mitosis. Following p21 depletion, we observed that whilst there remained heterogeneity in CDK2 levels, and G1 phase length, significantly fewer cells entered a CDK2 low arrest state. Thus, our data showed that in hTert-RPE1 cells, the CDK2 post mitotic arrest phenotype is both p21 and p53 dependent, consistent with the notion that high levels of p21 inhibit CDK2 activity, and that this activity blocks cell cycle progression.

Finally, to validate that the p21 high state corresponded to cells with inhibited CDK2 activity, hTert-RPE1 cells expressing p21-GFP only were transiently transfected with a CDK2-mRuby reporter. Several cells were identified that were in a p21 high state and had low CDK2 activity for prolonged periods of time, Fig. 4-7B. This provides critical further evidence that the p21 high post-mitotic arrest state corresponds to inhibition of CDK2 activity. Taken together we thus find that a percentage of cells enter a p21 and p53 dependent arrest state post mitosis, and this corresponds to inhibition of CDK2 activity.

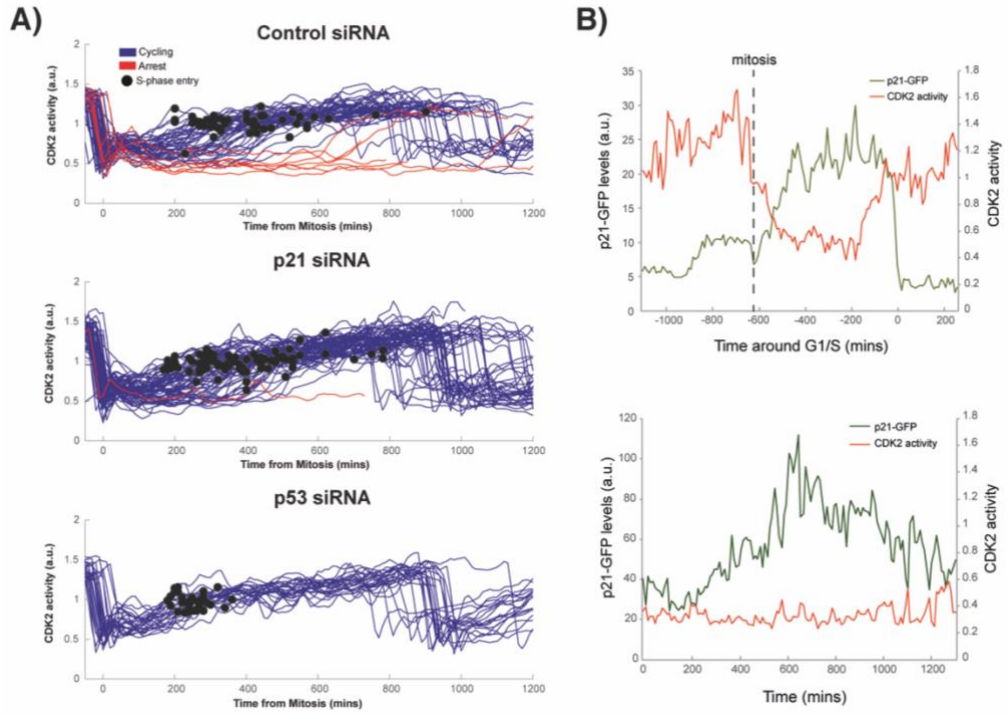

***Figure 4-7,* CDK2 low post mitotic arrest is both p21 and p53 dependent:** *A) Time-series of CDK2 activity in cycling cells show a significant drop in activity over mitosis associated with APC degradation of Cyclins (Peters 2002; King et al. 1996). Moreover in control cells we observe bifurcation after mitosis into CDK high and low states in line with previous observations (Spencer et al. 2013; Overton et al. 2014) (n=51). Following p21 depletion, consistent with our previous observations, variability in CDK2 levels of G1 Phase were present, although only a single cell to enter a CDK2 low state, suggesting that the post-mitotic CDK2 low state is p21 dependent (n=68). Depletion of the tumour suppressor p53 resulted in homogenous behaviour in CDK2 levels following mitosis, with no cells entering a post-mitotic arrest state (n=25); B) Transient transfection of hTert-RPE1 cells expressing p21-GFP only, with a mRuby-CDK2 reporter allowed us to quantify p21 and CDK2 level simultaneously in a select handful of single-cells. This demonstrated that high p21 levels correlate with suppression of CDK2 activity* [Graphs by Alexis Barr].

4.2.7 Expression of p21 correlates with the presence of DNA damage foci

Given that: 1) The G1 post mitosis state is p53 dependent; 2) p21 levels are most reflective of the combined length of G2 in the mother cell and G1 in the daughter cell; and 3) p21 levels accumulate after the S/G2 transition, we postulated that DNA damage occurring over S-phase in the mother cell underpins p21 levels in G2 phase of the mother cell and G1 phase of daughter cells. To test this proposal, we looked to correlate the presence of DNA damage in the mother cell S-phase with DNA damage levels in daughter cells. Much of the work in this section was performed by Frankie Butera working in the Bakal lab.

The protein γH2AX marks DNA damage, and is typically associated with double strand breaks (Kuo and Yang 2008), where it is phosphorylated by the damage sensing protein ATM (Burma et al. 2001). By immunostaining for γH2AX we looked for correlation between γH2AX expression levels and high levels of p21. Although no significant correlation was observed when intensities were related; a striking correlation between the presence of at least one strong γH2AX foci in the nuclear region, G1 phase length, and high p21 levels was observed, Fig. 4-8A,B&C. Immunofluorescence staining of other markers of DNA damage, specifically 53BP1 (Schultz et al. 2000) and phosphorylated ATM (Bakkenist and Kastan 2003), also demonstrated a similar correlation between a strong foci and p21-GFP levels, Fig. 4-8C. It was noted that these single foci looked resembled stretches of incompletely replicated DNA, that have been shown to result in single-stranded DNA gaps being present in G2 cells (Durkin and Glover 2007). Indeed co-localisation of RPA2, a marker for ssDNA (Zou and Elledge 2003), was observed with single γH2AX foci in G2

Fig. 4-8D. Finally, time-lapse imaging revealed that in the majority of cell divisions, sister G1 cells behave similarly such that in 47% (67/142) of mitoses neither sister has a 53BP1 focus, in 24% (34/142) both sister cells have a focus, and in 29% (41/142) only one sister has a focus, Fig. 4-8E. Taken together, this analysis suggested that incompletely replicated DNA and damage that occurs over the mother cell S-phase is inherited by daughter cells. Inheritance of this DNA damage in turn correlates with up-regulation of p21 levels and induction of G1 post mitotic arrest.

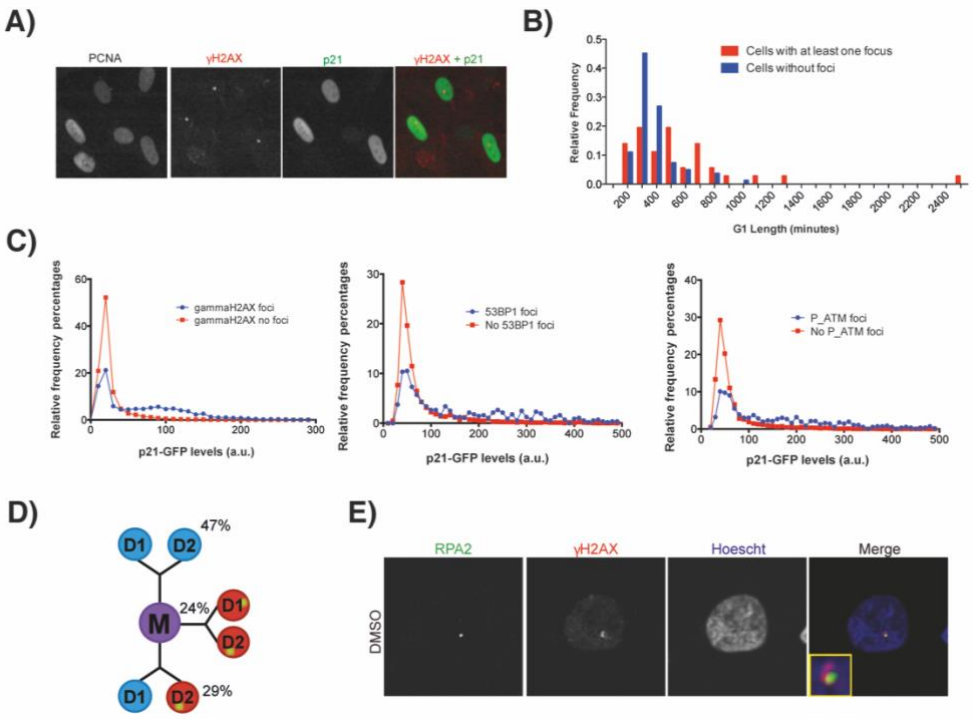

***Figure 4-8,* DNA damage over S-phase correlates with p21 induction:** *A) Immunofluorescent staining γH2AX revealed that the presence of a distinct foci is correlated with high p21 expression; B) The presence of a foci is associated with an increase in G1 phase length (unpaired student's t-test P < 0.001); C) Foci of the DNA damage markers, γH2AX, 53BP1and phosphorylate ATM all correlate with increased p21 levels, indicating that these foci are DNA repair complexes, marking regions of damage; D) After 47% of mitosis both daughter cells continue to proliferate, in 24% of cases a single daughter arrests, whilst in 29% of cases both daughters undergo a post mitotic arrest. The fact that these percentage are higher than those observed for p21-GFP, mRuby-PCNA labelled cells, indicates that the 53BP1 reporter being used may be responsible for low levels of DNA damage or replication stress; E) RPA-2 a marker of single strand DNA localises alongside γH2AX, indicating that the DNA damage is stemming from under replicated DNA originating in S-phase* [Figure by Alexis Barr and Frankie Butera].

4.2.8 Replication stress induces high levels of p21 expression

Taken together our evidence suggested that endogenous DNA damage, occurring during DNA replication in the mother cell S-phase, was responsible for p21 heterogeneity and G1 post mitotic arrests observed in daughter cells. To determine whether induction of replication stress with exogenous agents could further increase p21 levels and the number of cells arresting in G1 phase, we treated populations of cells expressing p21-GFP and mRuby-PCNA with aphidicolin. Aphidicolin is a fungal antibiotic that inhibits the processivity of DNA polymerases in S-phase and thus causes increased levels of replication stress, and DNA damage specifically during S-phase (Ikegami et al. 1978). In hTert-RPE1 cells treated with aphidicolin we observed an increase in the number of cells expressing γH2AX foci, as well as a major increase in the number of foci per cell, Fig. 4-9A. S-phase length was increased and in cells exiting S-phase p21 upregulation occurred versus DMSO treated control populations, Fig. 4-9B. Following mitosis, we found that a much higher fraction of cells entered G1 post mitotic arrest (62.5%, 20/32) when compared to DMSO controls (13.2%, 5/38). As such, by inducing exogenous replication stress with Aphidicolin we were able to significantly shift the fraction of daughter cells entering arrest states, indicating that inheritance of DNA damage is indeed behind the G1 arrest that we have observed. Moreover, a fraction of mother cells also entered a G2 arrest state after S-phase (25%, 8/32), with high levels of p21, a phenotype that we previously hadn't observed in wild-type or control populations.

To test whether this arrest state corresponded to the CDK2 low state we have observed in unperturbed populations of cycling cells, we treated hTert-RPE1 cells

expressing the CDK2 reporter and mRuby-PCNA with aphidicolin, Fig. 4-9C. Consistent with the arrest states equating to one another, we observed an increase in the number of cells entering G1 post-mitotic arrest, as well as cells arresting in G2. In those cells entering a G2 arrest, we found that after a period of time CDK2 activity was downregulated, indicating that once p21 levels are sufficiently high they are able to effectively inhibit CDK2 activity in G2 phase.

We also sought to both validate that both p21 and p53 are required to invoke G1 post-mitotic arrest following exogenous stress and determine whether G2 arrest is p53 and p21 dependent. We replicated the p21 and p53 siRNA gene depletions previously performed and following knockdown cells were treated with aphidicolin. Consistent with the notion that G1 arrest and delay are p53 dependent, we observed no variation in G1 length of G1 arrest following p53 depletion and aphidicolin treatment, moreover no cells entered a G2 arrest indicating this arrest is also p53 dependent, although in a rare case extended G2 was observed, Fig. 4-9C. Following p21 depletion again we observed heterogeneity in G2 and G1 lengths. Notably G2 arrest was absent, and G1 arrest was compromised, Fig. 4-9C. Together this verified that the observed G1 arrest is p53 and p21 dependent, in response to either endogenous or exogenous replication stress in mother cell S-phase. This analysis also revealed the presence of a p21 and p53 dependent G2 arrest at higher levels of DNA damage, and thus also supports a causal link between the rate of p21 increase versus likelihood of G2 arrest.

We also analysed the response of hTert-RPE1 cells to ATR kinase inhibition using the inhibitor AZD6738 (Jones et al. 2013). ATR inhibitors induce additional

replication stress through destabilising stalled replication forks that are formed during DNA synthesis (Couch et al. 2013). Treatment with ATR inhibitors led to an increase in the number of cells with DNA damage foci as marked by γH2AX, Fig. 4-9A. In line with Aphidicolin treatment we also observed a marked increase in the number of cells entering a G1 arrest state versus DMSO treated cells, and an increase in G1 length, Fig. 4-9B. However, in contrast to aphidicolin treatment we noted that both S-phase and G2 phase were shorter, no cells arrested in G2, and the majority of p21 accumulation occurred rapidly and shortly after mitosis, Fig. 4-9B. This supports evidence suggesting that inhibition of progression through G2 phase of the cell cycle is ATR dependent (Cliby et al. 2002; Liu et al. 2000). As such, we find that inhibition of ATR activity, induces increased levels of DNA damage during and after S-phase, and this corresponds to an increase in the number of cells entering a G1 post mitotic arrest. Additionally, we find that ATR may be involved in the p53-p21 G2 dependent arrest that we identified with aphidicolin.

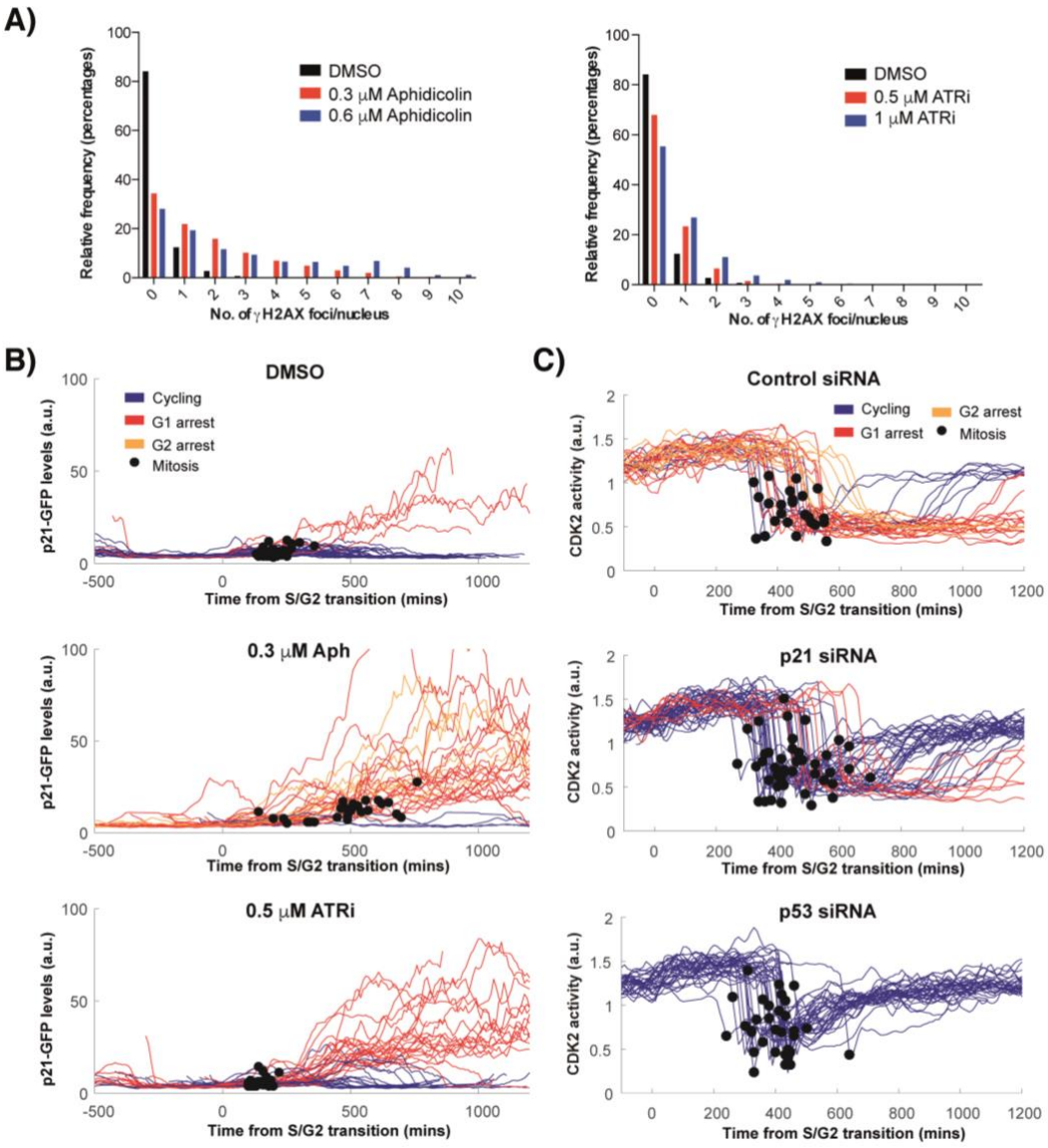
A)
Relative frequency (percentages)
No. of γH2AX foci/nucleus
DMSO
0.3 μM Aphidicolin
0.6 μM Aphidicolin
DMSO
0.5 μM ATRi
1 μM ATRi
B)
DMSO
0.3 μM Aph
0.5 μM ATRi
p21-GFP levels (a.u.)
Time from S/G2 transition (mins)
Cycling
G1 arrest
G2 arrest
Mitosis
C)
Control siRNA
p21 siRNA
p53 siRNA
CDK2 activity (a.u.)
Time from S/G2 transition (mins)
Cycling
G1 arrest
G2 arrest
Mitosis

Figure 4-9, **Exogenous replication increases the number of cells entering a p21 high - CDK2 low post mitotic arrest:** *A) Following treatment of hTert-RPE1 cells for 24hrs with either 0.3uM or 0.6 uM Aphidicolin, there is a dramatic increase in the number of cells that displaying γH2AX foci. ATR inhibitors also increase the number of foci per cell and number of cells exhibiting γH2AX foci, however the effect is less pronounced.* [Graphs by Alexis Barr]; *B) Tracking of cells and computational alignment to S/G2 transition, after addition of 0.3um Aphidicolin revealed that nearly all cells upregulate p21 following S-phase exit compared to DMSO controls cells that typically continue to proliferate (blue). This upregulation of p21 is accompanied by either G1 arrest (red) or G2 arrest (yellow). ATR inhibitors had a similar effect causing the majority of cells to upregulate p21, however no cells were observed to enter a G2 arres; C) Cells that expressed a CDK2 reporter were tracked and computationally aligned to S/G2 transitions following addition of 0.3 uM aphidicolin, either after treatment with non targeting, p21 or p53 siRNA. G2 arrest was completely compromised following p21 depletion, and significantly fewer cells entered a G1 post-mitotic arrest. This effect was ven stronger following p53 depletion where no cells entered either a G1 or G2 arrest, however in one rare case we noticed a very extended G2 phase.*

4.2.9 Skp2 and Cdt2 pathways both co-operate to induce p21 degradation prior to S-phase

Over all tested conditions p21 levels were undetectable over S-phase, and where p21 levels were high in G1, they would drop rapidly prior to S-phase. Together these features suggested that active degradation of p21-GFP was occurring prior to and during S-phase. Further supporting the role of active degradation in p21 loss prior to S-phase was the fact that we observed two rates of p21 loss in wild type cells, a slower rate of degradation that occurred over several hours prior to S-phase entry, and a much faster switch like degradation of p21 that occurred at the G1/S-phase transition. Previous studies had indicated that over G1 and S-phase the E3 ubiquitin ligases SCF-Skp2 and CRL4-Cdt2 both play a role in degrading p21 (Bornstein et al. 2003; Abbas

et al. 2008). However little work has gone into understanding the timing and rate with which these ubiquitin ligases remove p21 protein.

To determine the roles that these two ubiquitin ligases play in p21 degradation, we depleted Cdt2, Skp2 individually and the combination of both Cdt2 and Skp2 using siRNA, and then observed the rates of p21 loss across populations of cells computationally synchronised to G1/S transition. Following loss of Skp2 we observed that all cells entering S-phase, and exhibited switch like loss of p21, Fig. 4-10A. This indicated that SCF-Skp2 was responsible for the slower rate of p21 degradation observed in the run-up to G1/S transition; p21 levels remained undetectable over S-phase. Following Cdt2 depletion we observed that cells entering S-phase demonstrated active degradation of p21 prior to G1/S transition, however no switch like loss of p21 was seen, Fig. 4-10A. After Cdt2 depletion we also saw cells reaccumulate p21 in S-phase, in long slow oscillatory cycles. Nuclei also continued to increase in size beyond what we had previously observed in cell exiting S-phase into G2. Another known role of p21 is as a strong inhibitor of PCNA (Waga et al. 1994), loss of p21 is required to initiate DNA synthesis and S-phase (Arias and Walter 2006), thus we hypothesised that re-accumulation of p21 during S-phase interfered with the reversibility of G1/S transition, leading to incomplete replication occurring. Analysis of total DNA content in cell nuclei using fluorescence activated cell sorting, showed that following Cdt2 depletion cells frequently had DNA content levels that fell between n=2 and n=4 (fully replicated DNA), indicating the presence of incomplete replication and the cell cycle stalling in S-phase. Moreover, by depleting p21 this phenotype was rescued, demonstrating that Cdt2 is required to prevent p21

dependent inhibition of S-phase. Following depletion of both Cdt2 and Skp2 we observed that the rate of p21 depletion prior to G1/S transition was severely compromised, indicating both proteins cooperate to induce degradation of p21, Fig. 4-10A. Importantly, no cell entered S-phase with detectable levels of p21, despite depletion of both of these factors. Thus, indicating that p21 is a highly potent inhibitor of S-phase entry, and that active degradation of p21 is essential for proper transition. Taken together we found that both Skp2 and Cdt2 co-operate to remove p21 prior to G1/S transition in unperturbed cycling hTert RPE1 cells. Whilst SCF-Skp2 complexes are responsible for degrading high levels of p21 in G1 phase, CRL4-Cdt2 complexes account for the switch like loss of p21 of G1/S transition and maintain low levels of p21 over S-phase, where p21 is present both DNA replication and correct temporal control of S-phase are interfered with, and this is likely due to inhibition of PCNA (Waga et al. 1994).

4.2.10 Cdt2 and Skp2 are critical for controlling the length of G1 phase

In line with the notion that Cdt2 and Skp2 degradation are required to remove p21 prior to S-phase, I analysed how the length of G1 phase was affected by depletion of either of these two ubiquitin ligases as well as the combination of both. Following depletion of Cdt2, we observed a small but significant increase in the length of G1 phase, though notably only the G1 length of cells that re-enter S-phase was captured, Fig. 4-10B. A significant proportion of cells entered a permanent stalled S-phase or G1 arrest with increasing p21 levels, as evidenced by the increasing average p21 levels over the course of imaging, Fig. 4-10C. In contrast, following Skp2 depletion we

observed G1 phase to be much longer and for p21 levels to reach significantly higher maximum levels in G1, (Fig. 4-10B. The majority of Skp2 depleted cells also eventually re-entered S-phase through Cdt2 mediated depletion of p21, this likely underpins why average p21 levels remained only slightly above those of control populations, Fig. 4-10C. This data is consistent with the notion that Skp2 suppresses p21 levels over a much longer period in G1 phase and suggests that CRL4-Cdt2 complexes are unable to degrade high levels of p21 either due to direct inhibition of active replication complexes by p21 or through reduced Cyclin:CDK activity. If this were not the case, G1 phase length would not be compromised following Skp2 depletion. Together implying the presence of negative feedback between CRL4-Cdt2 activity and p21 levels.

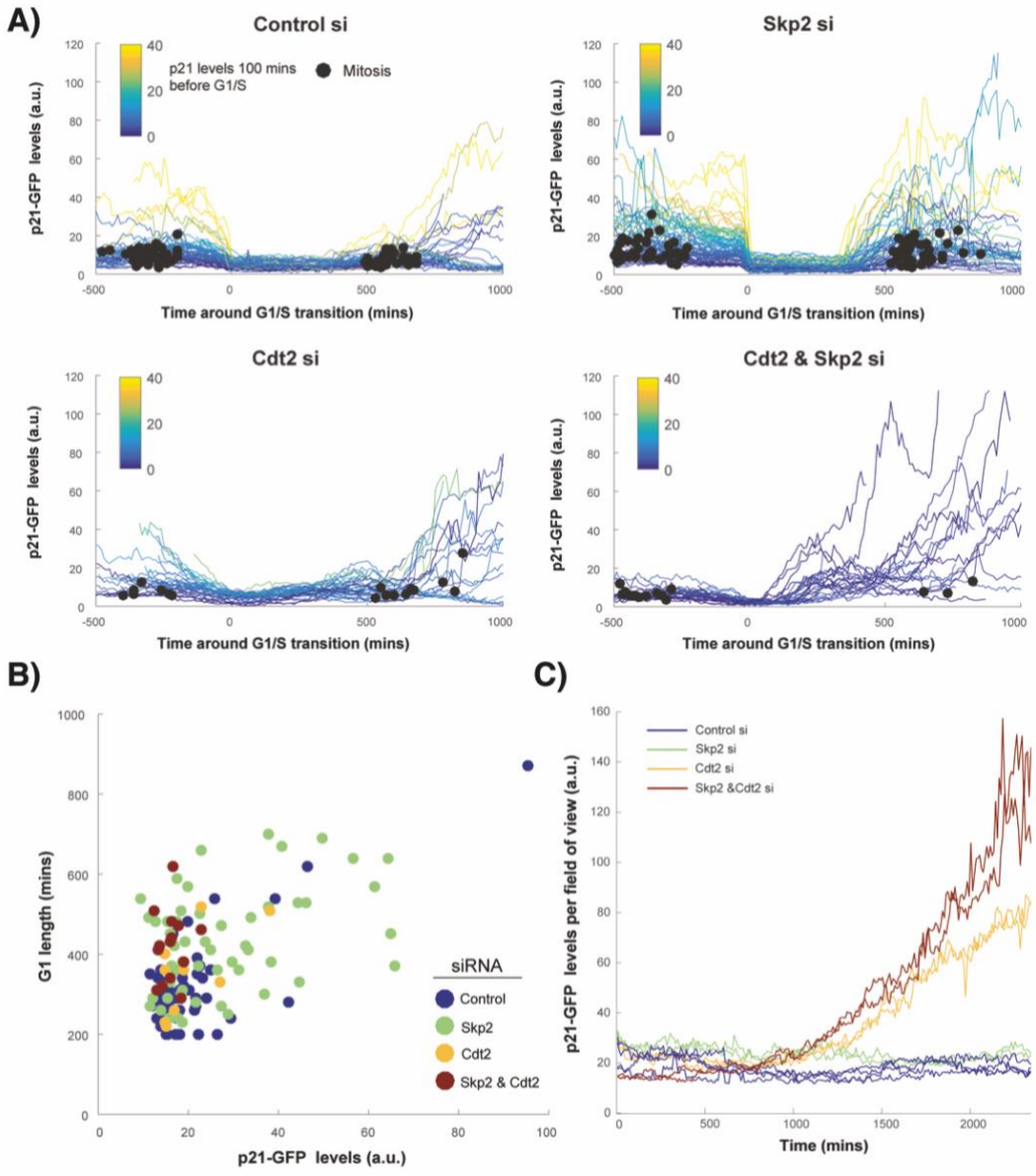
A)
Control si
Skp2 si
Cdt2 si
Cdt2 & Skp2 si
p21 levels 100 mins before G1/S
Mitosis
p21-GFP levels (a.u.)
Time around G1/S transition (mins)
B)
G1 length (mins)
p21-GFP levels (a.u.)
siRNA
Control
Skp2
Cdt2
Skp2 & Cdt2
C)
p21-GFP levels per field of view (a.u.)
Time (mins)
Control si
Skp2 si
Cdt2 si
Skp2 &Cdt2 si

***Figure 4-10*, The E3 Ubiquitin ligases SCF-Skp2 and CRL4-Cdt2 co-operate to target p21 for degradation:** *A) p21-GFP levels in hTert-RPE1 cells that were tracked and computationally aligned to G1/S transition following treatment with four siRNA conditions; non-targeting, Skp2, Cdt2 and the combination of Skp2 and Cdt2. Only cells that underwent G1/S transition and entered S-phase are shown. The number of cells captured entering S-phase following Cdt2 or Skp2 and Cdt2 depletion was significantly less owing to a large number of cells in an arrested state with high p21; B) G1 length versus p21 levels following treatment with the different siRNA conditions. Skp2 depletion led to a significant increase in G1 length, again though it should be noted many cells after Cdt2 or Skp2 and Cdt2 depletion did not have defined G1 lengths as they did not re-enter the cell cycle; C) The effect of Cdt2 depletion alone or in combination on total p21-GFP levels is best observed when all cells in the field of view are considered. Here p21-GFP levels increase ~5 fold over the course of imaging in contrast to p21-GFP levels in Skp2 depleted cells that remain similar to control.*

4.2.11 Serum is not responsible for a p21 dependent arrest

Finally, to exclude the possibility that mitogen signalling is also responsible for a p21 dependent arrest we looked to explore how p21-GFP levels in hTert-RPE1 cells are influenced by the absence of mitogens. To explore this response we serum starved, imaged and tracked hTert-RPE1 cells. Serum contains high levels of mitogens required for maintaining this immortalised cell line in a proliferative state. Following serum withdrawal, we observed that all cells entered an arrested state. However, in contrast to the p21-dependent DNA damage arrest state that we observed, p21 levels in arrested cells following serum withdrawal exhibited high levels of variability, indicating this arrest state was perhaps p21 independent, Fig. 4-11A. To verify this, we withdrew serum from the two p21 KO cell lines that we had previously generated. Indeed, in these cells arrest was not compromised, further providing evidence that following serum withdrawal arrest is p21 independent, Fig. 4-11B.

Mitogen signalling is known to stimulate cyclin D CDK4/6 activity, through either driving expression of cyclin D or through phosphorylation and activation of Cyclin D:CDK4/6 complexes. Thus, by depleting cyclin D with siRNA we would also expect to observe a p21 independent cell cycle arrest. We treated hTert-RPE1 cells with cyclin D siRNA and tracked them over time. Following cyclin D depletion in these cells we observed that the majority of the population entered an arrested state, evidenced by no cells entering S-phase. Importantly, this was independent off p21-levels, further supporting the notion that a mitogen withdrawal arrest is p21-independent, and probably occurs through a route that involves downregulation of Cyclin D:CDK4/6 activity as has been traditionally thought (Terada et al. 1999), Fig. 4-11C. However, at intermediate mitogen levels, the threshold p21 levels required to induce arrest would likely be lowered, as observed in Overton et al., due to reduced CDK mediated Skp2 activity (Overton et al. 2014), this would explain the findings of Spencer et al. (Spencer et al. 2013), and is consistent with our model of p21 activity being regulated by two double negative feedback loops.

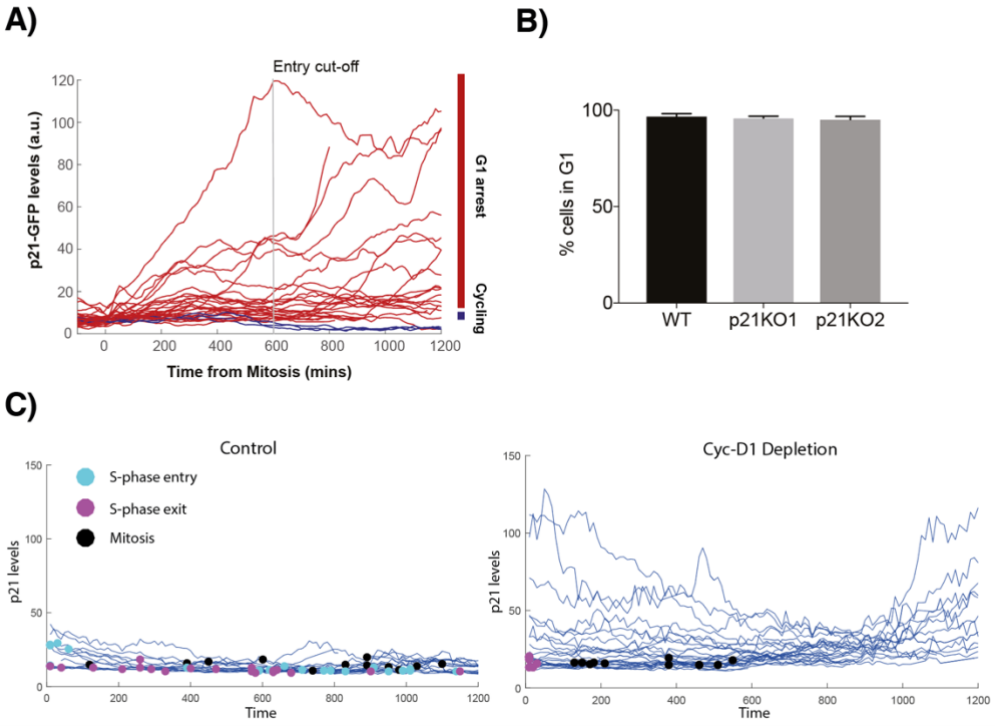

Figure 4-11, Serum withdrawal arrest is p21 independent: *A) Following withdrawal of serum in the previous cell cycle, p21 levels were measured in hTert-RPE1 cells and tracks were aligned to mitosis; n = 29 cells. Grey line marks 600 min time point used to define G1pm arrest. In Fig. 4.5A, red curves represent cells entering G1pm arrest, blue curves represent cells that enter S-phase; B) Graph showing the percentage of cells arresting in G1 in three different mRuby-PCNA hTert-RPE1 cell lines (WT and two p21KO clones) after serum withdrawal for 24 h. The mean and standard deviation (error bars), of two independent FACS experiments are shown* [Graph by Henriette Stoy]*; C) hTert-RPE1 cell were treated with Cyclin D siRNA and tracked over time. Tracks are aligned to 24 hours after treatment with Cyclin D siRNA. G1 arrests are indicated by the absence of cells entering S-phase (blue dots) after S-phase exit (purple) and mitosis (black).*

4.3 Discussion

Prior to this work, our understanding of how p21 is regulated over the cell cycle was limited to static single cell analysis and studies in which dynamics were measured destructively in synchronised populations, for example using western blots. Through endogenous tagging of p21 protein in hTert RPE1 cells, we were able to directly visualise p21-GFP levels in live cells populations. NucliTrack then allowed us to track hundreds of highly motile hTERT-RPE1 and computationally synchronise time series-data on p21-levels, preventing the need for using biological agents to synchronise populations. This is important as multiple studies have demonstrated these can cause unwanted side-effects and behaviour that are not physiologically relevant (Urbani et al. 1995). This pipeline allowed us to also track enough cells that we could quantify and assign statistical significance to behaviours that we observed qualitatively over the cell cycle. In the majority of previous live-cell studies results remain qualitative due to only a handful of cells being tracked. Taken together, this pipeline therefore represents a step towards being able to easily dissect the dynamics of signaling pathways using live-cell reporters, in large populations of live unperturbed single-cells.

Through analysing the time-series data that we collected we were then able to build a picture of how p21 is regulated over the cell cycle. This demonstrated that induction of p21 in hTert-RPE1 cells could be fully accounted for by the activity of p53, consistent with Stewart-Ornstein et al. who identified that p21-CFP levels increased following a peak in p53-YFP activity in live cells (Stewart-Ornstein and Lahav 2016). We also observed that in unperturbed cells, DNA damage originated in

S-phase and was likely due to replication stress. Through p53 dependent upregulation of p21 levels in G2 phase, the transfer of DNA damage to one or both daughter cells during mitosis, and likely continued p53 signalling in G1 phase, cells would then either enter a p21 and p53 dependent quiescent state, or delay progression through a still unknown p53 dependent, p21 independent, mechanism. Importantly, this p21 dependent arrest accounted for the bifurcation of CDK2 activity observed in Spencer et al. (Spencer et al. 2013). Together we therefore quantified the dynamics of p21 over the cell cycle and illustrated how endogenous DNA damage in mother cells influences the proliferation-quiescence decision in daughter cells.

We also characterised how p21 levels are degraded prior to G1/S transition. Where previously studies had implicated both SCF-Skp2 (Bornstein et al. 2003), and CRL4-Cdt2 (Abbas et al. 2008) in the degradation of p21 levels, uncertainty surrounded the rates and timings that these two ligases started p21 for proteolytic degradation (Abbas and Dutta 2009). We identified that both SCF-Skp2 and CRL4-Cdt2 cooperate to facilitate the degradation of p21 over the G1/S transition. SCF-Skp2 targets p21 for degradation at a slower rate, relatively, but is active for a longer period of G1 phase, and likely due to weaker inhibitory feedback appears able to remove higher p21 levels. CRL4-Cdt2, meanwhile, very quickly degrades p21 at the G1/S transition; though we speculate is only active once p21 levels are lower, due to strong inhibition of PCNA by p21 that is essential for formation of the active replication complexes that are required for CRL4-Cdt2 activity (Hayashi et al. 2014; Havens and Walter 2011). Thus, the sequential activity of these two ubiquitin ligases is effective in degrading p21 prior to S-phase, even where levels of p21 may be high. Over S-

phase, we found CRL4-Cdt2 is necessary to maintain low levels of p21, and without it, faulty replication is observed. This is in line with previous reports that CRL4-Cdt2 activity is essential for preventing re-replication (Hayashi et al. 2014; Havens and Walter 2011).

Together, these findings allow us to put forward a working hypothesis that two double negative feedback loops exist, that govern establishment of the proliferation-quiescence decision point and G1/S transition. In the first feedback loop, p21 inhibits CDK activity (likely some combination of CDK4/6 and CDK2 activity), whilst CDK activity in turn drives increasing SCF-Skp2 activity that degrades and thus inhibits the activity of p21. This model is consistent with the theoretical work of Overton et al. who hypothesised that the restriction point is bistable with respect to p21, due to a double negative feedback loop that includes CDK2 and SCF-Skp2 (Overton et al. 2014). In the second double negative feedback system, CDK activity drives the formation and activation of replication complexes that bind CRL4-Cdt2 to induce p21 degradation, whilst p21 inhibits the formation of replication complexes by binding to PCNA, mathematical modelling using systems of ordinary differential equations demonstrated that such a model can fully explain the behaviour of p21-GFP levels that we observed (Barr et al. 2017; Heldt et al. 2018). Thus, the live cell approach has allowed us to develop a more comprehensive model of p21's role in restriction point passage.

Since publication of this work, several papers validating our findings have been published. Notably, Arora et al. and Moser et al. replicated the finding that endogenous replication stress in mother cells causes a p21-dependent arrest in

daughters, using a CDK2 reporter (Arora et al. 2017; Moser et al. 2018). Yang et al. also found that levels of mitogen signalling can compete directly with p53 signalling activity to effect passage through the restriction point, using a live cell Cyclin-D reporter (H. W. Yang et al. 2017). Here, they showed that CycD:CDK4/6 complexes are stabilised and bound by a single protein of p21; where cyclin-D expression increases fast enough it is able to overcome the inhibitory effects p53-p21 signaling, and drive restriction point passage, this work explaining and echoing many of the findings of a similar study examining the behaviour of Cyclin-D in live cells (Zerjatke et al. 2017). Our finding that DNA lesions occurring in S-phase determine G1 phase lengths has also been replicated and published in key work by Lezaja et al. (Lezaja and Altmeyer 2018). Finally, Schwarz et al. quantified CDK activity levels required for restriction point passage in primary fibroblasts, and found variability existed in the timing of passage that could not be explained by mitogen signalling only. The authors postulated that inheritance of DNA damage, as we uncovered may underpin this (Schwarz et al. 2018).

Together this points towards a model where p21 acts downstream of pulsatile p53 signalling to temporally integrate information on DNA damage during the mother G2/M phases and the daughter G1 phases. Such a model is consistent with evidence that cells can exit an arrested state in response to a reduction in total p53 signalling over a distinct period of time (Reyes et al. 2018). In previous work Purvis et al. identified that p53 mediated apoptosis is rate-dependent, with slow induction of a downstream anti-apoptotic factor (Jeremy E. Purvis et al. 2012). Given p21's known anti-apoptotic role (Jänicke et al. 2007; Yu et al. 2005; Attardi et al. 1996; Polyak et al.

1997; Wettersten et al. 2013), we speculate that this may be the factor whose slow build-up blocks apoptosis and instead promotes G1 arrest. Such a hypothesis would be consistent with clinical observations as to the role of p21 in cancer (Abbas and Dutta 2009). Degradation of p21 before S-phase entry would then ensure that information on DNA damage generated during the previous cycle is fully erased before cells commit to a new cell cycle. This would allow the process of gathering information on DNA damage and deciding whether to proliferate, arrest, or apoptose, to begin anew following S/G2 passage. The idea of information-gathering phase and a reset phase leads to the attractive possibility that cells have evolved a DNA damage decision making system that is invariant to time and therefore remains consistent across both populations and generations of live single cells.

4.4 Methods

Wet-lab experimental procedures were performed by Alexis Barr, Frankie Butera, and Henriette Stoy (working in the Bakal lab) and are described in detail in (Barr et al. 2017). Cell segmentation, tracking and correction, was performed by myself using NucliTrack, as described in Chapter II. Additionally, I created all of the figures that made use of live-cell data with the exception of figure 4-7B. Everyone involved aided in interpreting data and conceiving follow up experiments.

4.4.1 Sampling cell tracks for figures and normalisation

To determine correlations, between p21 levels and phase length, for both control data and depletion experiments, a single-cell lineage was used. This was selected based on the daughter cell captured within the FOV for the longest period of time. The likelihood of arrest observed using this sampling regime very closely matched (21% n.47/219) that observed for sampling both daughters, where only mitosis in which both daughters are subsequently tracked for greater than 60 frames are used (23% n.29/128). The single lineage-sampling regime is used outside of calculations comparing daughters, as here n is greater. In all conditions at least two technical replicates were analysed, and tracks from each replicate were visually compared to ensure dynamics were consistent, prior to pooling of the data. Relative p21 levels decreased over imaging, this is due to intensity adjustments made by the microscope caused by increasingly bright images, (more p21-GFP expressing cells as population expands). No adjustment for this decrease was made, such that data remains faithful

to the original imaging data sets, though baseline levels for arrest thresholds are defined at 15 a.u. For all p21-GFP traces except Fig. 4.X, p21-GFP levels were normalised using the lowest level of p21-GFP captured across all tracks. This value (4 a.u. For Fig. 4.X traces) was generally similar to those seen in imaging (p21-GFP background levels were 12 a.u. In the first frame at the 10th percentile, while p21-GFP background levels were 2 a.u. in the last frame at the 10th percentile). The lowest track level was used rather than imaging baseline values, to avoid negative values.

4.4.2 Maximum and mean p21 expression levels and P-values.

p21-GFP values, used for population mean, s.d. and correlation measures in the main text are derived from the maximum p21-GFP value reached within a single cell in the cell cycle phase defined. The maximum p21-GFP value was used rather than the mean p21 value, as this generally gave stronger correlations to cell cycle phase length, though in all cases analysed for both mean and maximum the mean value also provided a statistically significant correlation.

While both maximum G1 p21 levels and G1 length were not normally distributed as determined by a one sample Kolmogorov–Smirnov test (G1 p21 levels, P = 1.8677e-04; G1 length, P = 0.0012). Following log transformation, the sample fell under the normal distribution (G1 p21 levels, P = 0.1467; G1 length, P = 0.2408). Therefore, we calculated the Pearson correlations on the log-transformed data as well, since Pearson's correlation does not fully describe the association of data which does not fall under a bivariate normal distribution. While correlation values did change, at no point did a previously significant result become insignificant. As such, for simplicity

and to match that data presented in figures, correlations determined on the untransformed data are given in the main text. Table 1 gives the values quoted in the main text (derived from the maximum p21 value per cell, per phase) alongside the respected value derived from the mean p21 level per cell. Full P-values are also quoted in this table; otherwise P = 0.01 is marked by **, and P = 0.05 is denoted by *.

Correlations between p21-GFP levels (mean and maximum) with phase lengths following S-phase are also quoted. Results following log transformation of the max p21 levels versus length data are also given. Mitotic timing masks correlation between G2M and G1D lengths. Given that we observe strong correlation between; 1) p21-GFP levels in the mother G2 (G2M) and daughter G1 (G1D), 2) p21-GFP levels in G2M and G2M length, and 3) p21-GFP levels in G1D and G1D length; we were initially surprised to find no correlation between G2M length and G1D length. However, in this comparison noise in the timing of mitosis has the dual effect of increasing G2M length and decreasing G1D length, and vice versa. Thus, if noise in mitotic timing is indeed masking the relationship between cell cycle phase length and p21 levels, we reasoned that the combined length of G2M and G1D would correlate better with the maximum p21-level reached than either length alone. We found this to be the case with R.0.68** for maximum p21 level in G2M/G1D against length of G2M/G1D, versus against length of G1D (R.0.62**) or G2M (R.0.52**) alone. This result was calculated from the merged experimental repeats (n.4). Within each repeat this result held as well (see Supplementary Table 2).

Table 4-1, Summary of R- and P-values for Pearson's correlations.

Correlation	Mean p21 level		Maximum p21 level		Log(time) vs log(max) p21	
	R	P	R	P	R	P
G1 length to p21 level	0.5939	4.1231e-19	0.6163	7.6580e-21	0.5616	5.4876e-17
G2 length to p21 level	0.4528	1.3023e-15	0.5146	2.1324e-20	0.4719	1.6389e-14
S-phase length to p21 level	-0.0741	0.3146	0.0720	0.3285	0.0306	0.6785
G2 Mother p21 level to G1 Daughter p21 level	0.7012	1.2832e-26	0.7482	6.2568e-32	0.7166	3.0996e-28
G2 Mother length to G1 Daughter length	R = 0.0707 P = 0.3580					
G1 Mother p21 levels to G1 Daughter p21 levels	0.5272	0.0023	0.5075	0.0036	0.4247	0.0173
G1 Mother length to G1 Daughter length	R = 0.1258 P = 0.5002					
G1 Daughter p21 levels to G2 Mother + G1 Daughter length	0.6687	1.6250e-23	0.6834	7.2175e-25	0.6445	1.8888e-21
G1 Daughter p21 levels to G2 Daughter p21 levels	0.3308	2.7011e-04	0.3063	7.8206e-04		
G1 Daughter length to G2 Daughter length	R = 0.1607 P = 0.0835					

Table 4-2, R-values for Pearson's correlations between phase length and maximum p21 value for different experimental repeats

	Length of G1 daughter vs p21 max	Length of G2 mother +G1 daughter vs p21 max
Repeat 1	0.3980	0.5993
Repeat 2	0.7671	0.8144
Repeat 3	0.7444	0.7951
Repeat 4	0.6991	0.7254

5 Conclusions and Future Directions

5.1 Summary of work and contributions

Within the field of high-content image analysis, workflows for analysis of simple 2D assays on rigid imaging substrates are now largely defined, and routinely used in drug and target discovery projects. In contrast, we are only just beginning to develop workflows that can robustly handle imaging data generated from more complex physiologically relevant assays, such as cells cultured in extra-cellular matrix, or in combination with other cell-types. Moreover, the use of live-cell imaging for understanding signalling dynamics in cells plated on 2D substrates, is only just transitioning from workflows that require extensive manual effort, to those where the majority of data extraction and analysis is automated.

In Chapter 2, I developed a new data analysis workflow for analyzing an assay where cells were cultured on substrates that much closer match those found in the body. Specifically, I looked at how our understanding of Rho GTPase signalling built up using overexpression studies, biochemical techniques, and systems wide siRNA screens on 2D cultures translates into how melanoma cells behave in collagen matrices (Cooper et al. 2015). Here, we observed that cells adopt a number of morphologies that are not observed on flat substrates, such as the “large round” shape. The siRNA screen also led to weaker phenotypes than others have observed on 2D plastic substrates (Yin et al. 2013). Together this showed that by performing

phenotypic screens on 3D matrices we can observe additional levels of complexity beyond what many have seen on plastic substrates.

However, this also brought with it challenges in both capturing and analysing the dataset. Firstly, the number of technical repeats and fields of view that could be captured vs. a typical high-content screen was limited by the distance that cells had to be imaged through collagen. Secondly, large numbers of artifact and noise meant that extensive effort had to be put into filtering mis-segmented objects and ensuring that features were biologically relevant. This motivated careful consideration of how to transform feature space, using the DBI as a readout, and highlights that as we move away from simple adherent cultures towards 3D matrices (Puls et al. 2018), co-culture systems (Tape et al. 2016; Bourhis et al. 1997), and even miniature organoids (Sachs and Clevers 2014), workflows capable of handling high noise to signal ratios will be necessary. Indeed, since publication, this work has been cited by several studies looking to analyse the behaviour of cells in 3D matrix, (Kimmel et al. 2018; Holmes et al. 2017; Alizadeh et al. 2018; Baniukiewicz et al. 2018). Together this contribution forms part of a larger trend in the pharmaceutical industry towards developing physiologically relevant assays and analysis strategies that improve our ability to predict behaviour in the clinic, and is discussed in 5.2, 'Towards physiologically relevant models'.

In Chapter 3, I developed NucliTrack, that gave myself and others the capabilities required to track hundreds to thousands of live-cells over long periods of time (Cooper et al. 2017). Specifically, where before tools have been widely available for automated segmentation of cell nuclei, as well as automated cell tracking, no tools

combined these two components of the pipeline into a single package. Most importantly, there were no solutions for manually correcting and annotating large number of tracks. This facilitated the studies in Chapter 4 where we analysed how p21 levels vary over the cell cycle (Barr et al. 2017; Heldt et al. 2018). Moreover, in work by Simpson et al. NucliTrack was used to investigate the dynamics of Erk signalling over the cell cycle, revealing a distinct bi-furcation in ERK activity after mitosis, and that BRAF and MEK inhibitors are effective in a cell-cycle dependent manner. As such this software relieves a key bottleneck in the live-cell imaging pipeline and is allowing those studying the cell cycle within the Bakal lab, and elsewhere to increase the number of cells being tracked by and order of magnitude when compared to previous work (Barr et al. 2016).

In Chapter 4 we revealed for the first time how a bifurcation in p21 levels after S-phase in the mother cell, and continuing into mitosis, underpins the previously observed p53 and p21 dependent G1 arrest state that occurs in response to DNA damage signalling. Before this investigation, to the best of our knowledge studies of p21 had been restricted to fixed endpoint assays, where dynamics were inferred from populations of cells that had been synchronised with agents that block cells at specific checkpoints (Urbani et al. 1995), or where arrest is induced by high concentrations of exogenous DNA damaging agents (Waldman, Kinzler, and Vogelstein 1995; Bunz et al. 1998). Here, for the first time we analysed p21 levels over the cell-cycle in unperturbed asynchronous cycling populations. This allowed us to define a p53 and p21 dependent G1 post mitotic arrest state, where elevated levels of p21 suppress the activity of CDK2. We found that this bifurcation of p21 activity is caused by stochastic

DNA damage that occurs during S-phase in the mother cell and is likely the result of endogenous DNA replication stress. In turn this stochastic DNA damage and p21 bifurcation underpins the bifurcation in CDK2 activity previously identified by Spencer et al. in seminal live cell studies of CDK2 activity (Spencer et al. 2013). Finally, we also characterised how p21 is degraded by the two E3-ubiqutin ligases, $CRL4^{Cdt2}$ and SCF^{Skp2}, showing that they cooperate to remove high-levels of p21, and both likely participate in double-negative feedback loops with p21, leading to the robust bifurcation that we observe in p21 levels prior to G1/S. Together this work led to a major step forwards in our understanding of p21 protein. The significance of this is highlighted by the increasing number of studies that our building off these results; several are discussed in Chapter 4 , though now over 50 publications have cited this work in the 2 years proceeding its publication. The significance of this work further highlights a growing trend towards the use of live-cell imaging studies for understanding the dynamics of biological signalling pathways and is discussed in 5.3, 'Towards understanding the dynamics of biological systems'

Over the course of this thesis the informatics tools and biology discovered with them have both advanced our image analysis capabilities and helped answer key questions in fundamental and cancer biology. These studies continue to impact the scientific community, and combined point towards a growing movement to analyse the behavior of cells and systems and a single cell level, using fully automated analysis tools. How the image analysis approached explored in this thesis, and the work I continue to pursue beyond this thesis, is part of an even greater trend towards reducing the cycle time between experiments and analysis is the final future direction

I consider in 5.4, 'Towards unsupervised and weakly supervised to close the loop on discovery', and sets the stage for the next chapter in quantitative microscopy.

5.2 Towards physiologically relevant models

Perhaps the single greatest challenge currently facing pharmaceutical research is that of being able to predict how a drug will behave against a disease in a clinical setting, notably phase 2 clinical trials (Paul et al. 2010). A major aspect of developing a better understanding of how biology will translate into a clinical setting is developing *in vitro* disease models that better approximate physiological conditions (Zanella et al. 2010; Skardal et al. 2016; Li et al. 2016; Breslin and O'Driscoll 2013; Tiscornia et al. 2011; Fatehullah et al. 2016). We also need to understand how the knowledge that we have built up in adherent cultures translates into these models, since where they translate well screening of adherent systems can be much simpler and more efficient (Pollard et al. 2009). Together such models and understanding will drive better selection of targets, and further drive the resurgence in phenotypic screening approaches that is currently happening (Swinney and Anthony 2011).

Quantitative image analysis approaches are now increasingly being developed and used to analyse such physiologically relevant models, such as the method developed in Chapter 2. For example, a 3D watershed algorithm originally used for segmentation of a handful of cell nuclei (Lin et al. 2003), was recently adapted to effectively detect hundreds to thousands of nuclei in a method for analysing fluorescently labeled nuclei in large 3D spheroids in high throughput (Boutin et al. 2018). Updates to CellProfiler now include the ability to extract features from 3D

objects (McQuin et al. 2018). Finally, the development of fully convolutional neural networks (Long et al. 2015), and there transfer to cell segmentation (Ronneberger et al. 2015), have shown incredible performance in segmenting nuclei in complex and cluttered images such as those take from histological stains (Caicedo et al. 2018), and can also be applied to 3D imaging stacks (Çiçek et al. 2016). However, performing unsupervised analysis of microscopy data, as was carried out in Chapter 2, using deep-neural networks remains incredibly difficult, since separating signal from noise is almost in possible in the absence of effective training labels, typically data on a cells position, rotational orientation, or channel intensity dominate more subtle effects. A solution to this is in weakly supervised learning, as discussed in 5.4. As such, trends in pharmaceutical research towards more physiologically relevant models, continue to motivate the development of better algorithms for managing complex images and will remain a major theme as the field progresses over the next five to ten years. If we can effectively analyse images of more translational models and extract valuable results and data from drug screens against these assays, this represents an exciting avenue for potentially reducing the number of phase II failures that occur.

5.3 Towards understanding the dynamics of biological systems

Over the last half century an enormous amount of work has gone into firstly identifying the factors involved in cellular signalling networks, and then understanding the connections between these factors. However, we still have a very limited understanding of how these networks function in time and space. For example, how do emergent properties such as the ability to integrate information on the environment and internal state of the cell in order to respond and make key

decisions (Tyson et al. 2003; Nurse 2000), arise? Live single-cell imaging is emerging as the dominant method for studying signalling dynamics; this being due to the fact it is perhaps one of the only techniques that is non-destructive and provides reliable readouts of signalling behaviour (Purvis and Lahav 2013; Gaudet and Miller-Jensen 2016). However, major challenges remain in processing and analysing live cell data (Cooper and Bakal 2017).

Going forwards, the dynamics of cell cycle regulation continues to be unpicked and dissected by live-single cell studies (Yang et al. 2017; Arora et al. 2017; Moser et al. 2018; Zerjatke et al. 2017). In the pipeline, signalling studies that include three or more live-cell reporters in a single cell promise to shed more light on how the cell cycle is regulated over time. Improvements in tracking are also on the horizon, with deep-neural networks beginning to yield significant performance improvements (Kristan et al. 2015; Wang et al. 2015; Bertinetto et al. 2016), albeit lagging behind advances in segmentation accuracies versus classical models. Moreover tools are now being developed that will allow us to effectively share and analyse others results, as well as perform studies exploring signalling dynamics over multiple datasets (Borland et al. 2018). For example, in recent studies of ERK signalling151 in *Drosophila* embryos, spectacular behaviours were observed. These showed how spatiotemporal orchestration of ERK signalling151 controls cell fate and segmentation patterning, and that optogenetic techniques can be used to alter these spatiotemporal signalling patterns and artificially determine cell fate and segmentation (Toettcher et al. 2013; Johnson et al. 2017; Johnson et al. 2018). Together, these trends underpin what will be a surge in the number of studies using

live-cell analysis to dissect dynamics across large populations of cells. With live cell signalling studies also standing poised to uncover the signalling152 behaviours that underpin developmental and complex multicellular behaviours, that to date have remained largely a complete mystery.

5.4 Towards unsupervised and weakly supervised deep-learning to close the loop on discovery

A topic of research explored in this Chapter 2 of this thesis was that of using unsupervised machine learning strategies to perform unbiased analysis of high-content imaging data. Here, the DBI was used as a readout for the performance of a given feature reduction strategy, and quantified how close members of the same cluster were versus how close different clusters are to each one another, where a low DBI indicated successful feature reduction. Machine learning models can also be trained to directly optimise feature space against the DBI. In this case the primary objective is minimising the DBI. The auxiliary task then contains the result that we are interested in, which is creating a feature space where the distance between different biological conditions represents how similar the phenotypes are to one another. As such, an approach in which we optimize directly against a measure such as the DBI can be considered a weakly supervised learning problem, where we train a model to effectively separation conditions, but are interested in the auxiliary task of where conditions lie in the feature space that the model learns. The DBI can also be replaced by a simple classification challenge, where a model learns to predict which condition an image or cell belongs to. For example, if an effective classification between cells or sample images from two conditions can be made in a given feature space, then they

are likely different and will appear distant in the learnt feature space. If errors are made then they are likely indistinguishable and will appear as close or overlapping in the feature space. As such an effective weakly supervised strategy is to train models to classify images of single cells, regions of interest, or whole fields of view, as belonging to a certain conditions. The feature space learnt in turn can be used to effectively map phenotypic similarity between different conditions.

Critically, this formulates the problem in a way that is amenable to deep-learning and solves many of the challenges faced by unsupervised methods. As discussed in 1.5, convolutional neural networks learn features directly from raw imaging data. By training these networks on such a weakly supervised task as in (Caicedo et al. 2018), we can extract meaningful embeddings of different conditions directly from raw imaging data without any prior knowledge. This means, that there is a much higher theoretical limit on accuracy since features are learnt and not engineered, that in practice has repeatedly been shown to materialize (LeCun et al. 2015). Moreover, it also lets us apply such techniques to complex images where segmentation and feature extraction techniques fail. As such, weakly supervised deep learning has been used successfully to rapidly profile and detect variations between conditions such as compounds with different mechanism of action; results here demonstrated marked improvements on traditional engineered feature based methods (Ando et al. 2017; Caicedo et al. 2018). Moreover, in work by Lu et al., that I was involved in over the last year of my PhD, weakly supervised learning was used to create an effective distance map that described how similar the localisation patterns are of 12,000 different proteins inside the human cell (Lu et al. 2018). This revealed, amongst other findings,

13 proteins that localised to the nucleolar rim and previously had been un-documented (Lu et al. 2018). Taken together weakly supervised learning is therefore letting us move towards analysis pipelines that enable us to rapidly convert raw high-content imaging data into actionable results, that crucially include detection of previously unseen behavior, and overcome issues with pure unsupervised learning of embeddings.

Through training models to jointly embed different data types, for example HTS and high-content data, allow conversion from imaging data to other data modalities. For example, in recent work, Simm et al. used extracted features from high-content images to learn an embedding between high-throughput biochemical binding (HTS) data, that allowed them to predict target binding profiled (Simm et al. 2018). Prospectively, this allowed them to screen for new compounds against a target and demonstrate 250-fold enrichment (Simm et al. 2018), representing perhaps the first significant demonstration of the power of machine and specifically deep-learning tools to pharmaceutical research. With trends towards generative models that allow novel compounds to be created from embeddings of compounds structure, such jointly learnt embeddings have exciting prospects for being able to automatically generate new compound structure that could be synthesised and would have a high likelihood of being functionally active in a given assay. As has been demonstrated in key studies looking at the design of novel materials using fully automated approaches (Duvenaud et al. 2015; Er et al. 2015; Gómez-Bombarelli et al. 2018; Tabor et al. 2018). These generative approaches are the final tool in the puzzle for developing systems that can fully automatically explore and optimize real world spaces, such as

chemical space, thus closing the loop on the cycle of discovery. The question is now how we can practically integrate such machine learning approaches with automated platforms for designing and setting up experiments? and use these platforms to automate the search for new targets against physiologically relevant disease models, reporters that demonstrate interesting dynamic behavior, and the next blockbuster drug. Together, such closed loop discovery, driven by advances in data analysis and machine learning, is looking set to be the next revolution in scientific exploration that will guide advancements over the coming century.

6 Bibliography

Abadi, Martín, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. 2016. "Tensorflow: A System for Large-Scale Machine Learning." In *OSDI*, 16:265–83.

Abassi, Yama A., Biao Xi, Wenfu Zhang, Peifang Ye, Shelli L. Kirstein, Michelle R. Gaylord, Stuart C. Feinstein, Xiaobo Wang, and Xiao Xu. 2009. "Kinetic Cell-Based Morphological Screening: Prediction of Mechanism of Compound Action and off-Target Effects." *Chemistry & Biology* 16 (7): 712–23.

Abbas, Tarek, and Anindya Dutta. 2009. "p21 in Cancer: Intricate Networks and Multiple Activities." *Nature Reviews. Cancer* 9 (6): 400–414.

Abbas, Tarek, Uma Sivaprasad, Kenta Terai, Virginia Amador, Michele Pagano, and Anindya Dutta. 2008. "PCNA-Dependent Regulation of p21 Ubiquitylation and Degradation via the CRL4Cdt2 Ubiquitin Ligase Complex." *Genes & Development* 22 (18): 2496–2506.

Abràmoff, Michael D., Paulo J. Magalhães, and Sunanda J. Ram. 2004. "Image Processing with ImageJ." *Biophotonics International* 11 (7): 36–42.

Acquaviva, Claire, Franz Herzog, Claudine Kraft, and Jonathon Pines. 2004. "The Anaphase Promoting Complex/cyclosome Is Recruited to Centromeres by the Spindle Assembly Checkpoint." *Nature Cell Biology* 6 (9): 892–98.

Adams, Cynthia L., Vadim Kutsyy, Daniel A. Coleman, Ge Cong, Anne Moon Crompton, Kathleen A. Elias, Donald R. Oestreicher, Jay K. Trautman, and Eugeni Vaisberg. 2006. "Compound Classification Using Image-Based Cellular Phenotypes." *Methods in Enzymology* 414: 440–68.

Adnane, J., R. J. Jackson, S. V. Nicosia, A. B. Cantor, W. J. Pledger, and S. M. Sebti. 2000. "Loss of p21WAF1/CIP1 Accelerates Ras Oncogenesis in a Transgenic/knockout Mammary Cancer Model." *Oncogene* 19 (47): 5338–47.

Akaike, Hirotogu. 1998. "Information Theory and an Extension of the Maximum Likelihood Principle." In *Selected Papers of Hirotugu Akaike*, edited by Emanuel Parzen, Kunio Tanabe, and Genshiro Kitagawa, 199–213. New York, NY: Springer New York.

Alizadeh, E., Xu, W., Castle, J., Foss, J. and Prasad, A., 2018. TISMorph: A tool to quantify texture, irregularity and spreading of single cells. *bioRxiv*, p.372755.

Al-Kofahi, Yousef, Wiem Lassoued, William Lee, and Badrinath Roysam. 2010. "Improved Automatic Detection and Segmentation of Cell Nuclei in Histopathology Images." *IEEE Transactions on Bio-Medical Engineering* 57 (4): 841–52.

Allan, Chris, Jean-Marie Burel, Josh Moore, Colin Blackburn, Melissa Linkert, Scott Loynton, Donald Macdonald, et al. 2012. "OMERO: Flexible, Model-Driven Data Management for Experimental Biology." *Nature Methods* 9 (3): 245–53.

Altschuler, Steven J., and Lani F. Wu. 2010. "Cellular Heterogeneity: Do Differences Make a Difference?" *Cell* 141 (4): 559–63.

Amano, M., Y. Fukata, and K. Kaibuchi. 2000. "Regulation and Functions of Rho-Associated Kinase." *Experimental Cell Research* 261 (1): 44–51.

Amir, El-Ad David, Kara L. Davis, Michelle D. Tadmor, Erin F. Simonds, Jacob H. Levine, Sean C. Bendall, Daniel K. Shenfeld, Smita Krishnaswamy, Garry P. Nolan, and Dana Pe'er. 2013. "viSNE Enables Visualization of High Dimensional Single-Cell Data and Reveals Phenotypic Heterogeneity of Leukemia." *Nature Biotechnology* 31 (6): 545–52.

Arias, Emily E., and Johannes C. Walter. 2006. "PCNA Functions as a Molecular Platform to Trigger Cdt1 Destruction and Prevent Re-Replication." *Nature Cell Biology* 8 (1): 84–90.

Arora, Mansi, Justin Moser, Harsha Phadke, Ashik Akbar Basha, and Sabrina L. Spencer. 2017. "Endogenous Replication Stress in Mother Cells Leads to Quiescence of Daughter Cells." *Cell Reports* 19 (7): 1351–64.

Attardi, L. Donatella, S. W. Lowe, J. Brugarolas, and T. Jacks. 1996. "Transcriptional Activation by p53, but Not Induction of the p21 Gene, Is Essential for Oncogene-Mediated Apoptosis." *The EMBO Journal* 15 (14): 3693–3701.

Aulner, Nathalie, Anne Danckaert, Eline Rouault-Hardoin, Julie Desrivot, Olivier Helynck, Pierre-Henri Commere, Hélène Munier-Lehmann, et al. 2013. "High Content Analysis of Primary Macrophages Hosting Proliferating Leishmania Amastigotes: Application to Anti-Leishmanial Drug Discovery." *PLoS Neglected Tropical Diseases* 7 (4): e2154.

Babaloukas, Georgios, Nicholas Tentolouris, Stavros Liatis, Alexandra Sklavounou, and Despoina Perrea. 2011. "Evaluation of Three Methods for Retrospective Correction of Vignetting on Medical Microscopy Images Utilizing Two Open Source Software Tools." *Journal of Microscopy* 244 (3): 320–24.

Bai, Min, and Raquel Urtasun. 2017. "Deep Watershed Transform for Instance Segmentation." In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2858–66. IEEE.

Bakal, Chris, John Aach, George Church, and Norbert Perrimon. 2007. "Quantitative Morphological Signatures Define Local Signaling Networks Regulating Cell Morphology." *Science* 316 (5832): 1753–56.

Bakal, Chris, Rune Linding, Flora Llense, Elleard Heffern, Enrique Martin-Blanco, Tony Pawson, and Norbert Perrimon. 2008. "Phosphorylation Networks Regulating JNK Activity in Diverse Genetic Backgrounds." *Science* 322 (5900): 453–56.

Bakkenist, Christopher J., and Michael B. Kastan. 2003. "DNA Damage Activates ATM through Intermolecular Autophosphorylation and Dimer Dissociation." *Nature* 421 (6922): 499–506.

Balasubramanian, Mukund, and Eric L. Schwartz. 2002. "The Isomap Algorithm and Topological Stability." *Science* 295 (5552): 7.

Balint, Eva, Andrew C. Phillips, Serguei Kozlov, Colin L. Stewart, and Karen H. Vousden. 2002. "Induction of p57(KIP2) Expression by p73beta." *Proceedings of the National Academy of Sciences of the United States of America* 99 (6): 3529–34.

Baniukiewicz, P., Collier, S. and Bretschneider, T., 2018. QuimP: analyzing transmembrane signalling in highly deformable cells. *Bioinformatics*, *34*(15), pp.2695-2697.

Barr, Alexis R., Samuel Cooper, Frank S. Heldt, Francesca Butera, Henriette Stoy, Jörg Mansfeld, Béla Novák, and Chris Bakal. 2017. "DNA Damage during S-Phase Mediates the Proliferation-Quiescence Decision in the Subsequent G1 via p21 Expression." *Nature Communications* 8 (March): 14728.

Barr, Alexis R., Frank S. Heldt, Tongli Zhang, Chris Bakal, and Béla Novák. 2016a. "A Dynamical Framework for the All-or-None G1/S Transition." *Cell Systems* 2 (1): 27–37.

Bergstra, James, Olivier Breuleux, Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Desjardins, Joseph Turian, David Warde-Farley, and Yoshua Bengio. 2010. "Theano: A CPU and GPU Math Compiler in Python." In *Proc. 9th Python in Science Conf*. Vol. 1.

Bertinetto, Luca, Jack Valmadre, João F. Henriques, Andrea Vedaldi, and Philip H. S. Torr. 2016. "Fully-Convolutional Siamese Networks for Object Tracking." In *Computer Vision – ECCV 2016 Workshops*, 850–65. Springer International Publishing.

Besson, Arnaud, Steven F. Dowdy, and James M. Roberts. 2008. "CDK Inhibitors: Cell Cycle Regulators and beyond." *Developmental Cell* 14 (2): 159–69.

Beucher, Serge, and Fernand Meyer. 1992. "The Morphological Approach to Segmentation: The Watershed Transformation." *Optical Engineering-New York-Marcel Dekker Incorporated-* 34: 433–433.

Bilmes, Jeff A., and Others. 1998. "A Gentle Tutorial of the EM Algorithm and Its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models." *International Computer Science Institute* 4 (510): 126.

Bise, R., Z. Yin, and T. Kanade. 2011. "Reliable Cell Tracking by Global Data Association." In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 1004–10.

Boland, M. V., and R. F. Murphy. 2001. "A Neural Network Classifier Capable of Recognizing the Patterns of All Major Subcellular Structures in Fluorescence Microscope Images of HeLa Cells." *Bioinformatics* 17 (12): 1213–23.

Bolshakova, N., and F. Azuaje. 2003. "Cluster Validation Techniques for Genome Expression Data." *Signal Processing* 83 (4): 825–33.

Borland, David, Hong Yi, Gavin D. Grant, Katarzyna M. Kedziora, Hui Xiao Chao, Rachel A. Haggerty, Jayashree Kumar, Samuel C. Wolff, Jeanette G. Cook, and Jeremy E. Purvis. 2018. "The Cell Cycle Browser: An Interactive Tool for Visualizing, Simulating, and Perturbing Cell-Cycle Progression." *Cell Systems* 7 (2): 180–84.e4.

Bornstein, Gil, Joanna Bloom, Danielle Sitry-Shevah, Keiko Nakayama, Michele Pagano, and Avram Hershko. 2003. "Role of the SCFSkp2 Ubiquitin Ligase in the Degradation of p21Cip1 in S Phase." *The Journal of Biological Chemistry* 278 (28): 25752–57.

Bos, J. L. 1989. "Ras Oncogenes in Human Cancer: A Review." *Cancer Research* 49 (17): 4682–89.

Bostrom, R. C., H. S. Sawyer, and W. E. Tolles. 1959. "Instrumentation for Automatically Prescreening Cytological Smears." *Proceedings of the IRE* 47 (11): 1895–1900.

Bourhis, Xuefen Dong-Le, Yolande Berthois, Guy Millot, Armelle Degeorges, Monique Sylvi, Pierre-Marie Martin, and Fabien Calvo. 1997. "Effect of Stromal and Epithelial Cells Derived from Normal and Tumorous Breast Tissue on the Proliferation of Human Breast Cancer Cell Lines in Co-Culture." *International Journal of Cancer* 71 (1): 42–48.

Boutin, Molly E., Ty C. Voss, Steven A. Titus, Kennie Cruz-Gutierrez, Sam Michael, and Marc Ferrer. 2018. "A High-Throughput Imaging and Nuclear Segmentation Analysis Protocol for Cleared 3D Culture Models." *Scientific Reports* 8 (1): 11135.

Boutros, Michael, Florian Heigwer, and Christina Laufer. 2015. "Microscopy-Based High-Content Screening." *Cell* 163 (6): 1314–25.

Boutros, Michael, Amy A. Kiger, Susan Armknecht, Kim Kerr, Marc Hild, Britta Koch, Stefan A. Haas, Renato Paro, Norbert Perrimon, and Heidelberg Fly Array Consortium. 2004. "Genome-Wide RNAi Analysis of Growth and Viability in Drosophila Cells." *Science* 303 (5659): 832–35.

Breslin, Susan, and Lorraine O'Driscoll. 2013. "Three-Dimensional Cell Culture: The Missing Link in Drug Discovery." *Drug Discovery Today* 18 (5-6): 240–49.

Buggenthin, Felix, Carsten Marr, Michael Schwarzfischer, Philipp S. Hoppe, Oliver Hilsenbeck, Timm Schroeder, and Fabian J. Theis. 2013. "An Automatic Method for Robust and Fast Cell Detection in Bright Field Images from High-Throughput Microscopy." *BMC Bioinformatics* 14 (October): 297.

Bunz, F., A. Dutriaux, C. Lengauer, T. Waldman, S. Zhou, J. P. Brown, J. M. Sedivy, K. W. Kinzler, and B. Vogelstein. 1998. "Requirement for p53 and p21 to Sustain G2 Arrest after DNA Damage." *Science* 282 (5393): 1497–1501.

Burma, S., B. P. Chen, M. Murphy, A. Kurimasa, and D. J. Chen. 2001. "ATM Phosphorylates Histone H2AX in Response to DNA Double-Strand Breaks." *The Journal of Biological Chemistry* 276 (45): 42462–67.

Caicedo, J. C., J. Roth, A. Goodman, and T. Becker. 2018. "Evaluation of Deep Learning Strategies for Nucleus Segmentation in Fluorescence Images." *bioRxiv*. https://www.biorxiv.org/content/early/2018/06/16/335216.abstract.

Caicedo, Juan C., Sam Cooper, Florian Heigwer, Scott Warchal, Peng Qiu, Csaba Molnar, Aliaksei S. Vasilevich, et al. 2017. "Data-Analysis Strategies for Image-Based Cell Profiling." *Nature Methods* 14 (9): 849–63.

Caicedo, Juan C., Claire McQuin, Allen Goodman, Shantanu Singh, and Anne E. Carpenter. 2018. "Weakly Supervised Learning of Single-Cell Feature Embeddings." *bioRxiv*, 293431.

Caramel, Julie, Eftychios Papadogeorgakis, Louise Hill, Gareth J. Browne, Geoffrey Richard, Anne Wierinckx, Gerald Saldanha, et al. 2013. "A Switch in the Expression of Embryonic EMT-Inducers Drives the Development of Malignant Melanoma." *Cancer Cell* 24 (4): 466–80.

Carpenter, Anne E. 2007. "Image-Based Chemical Screening." *Nature Chemical Biology* 3 (8): 461–65.

Carpenter, Anne E., Thouis R. Jones, Michael R. Lamprecht, Colin Clarke, In Han Kang, Ola Friman, David A. Guertin, et al. 2006. "CellProfiler: Image Analysis Software for Identifying and Quantifying Cell Phenotypes." *Genome Biology* 7 (10): R100.

Cavenee, W. K., T. P. Dryja, R. A. Phillips, W. F. Benedict, R. Godbout, B. L. Gallie, A. L. Murphree, L. C. Strong, and R. L. White. 1983. "Expression of Recessive Alleles by Chromosomal Mechanisms in Retinoblastoma." *Nature* 305 (5937): 779–84.

Chellappan, S. P., S. Hiebert, M. Mudryj, J. M. Horowitz, and J. R. Nevins. 1991. "The E2F Transcription Factor Is a Cellular Target for the RB Protein." *Cell* 65 (6): 1053–61.

Chen, Katherine C., Laurence Calzone, Attila Csikasz-Nagy, Frederick R. Cross, Bela Novak, and John J. Tyson. 2004. "Integrative Analysis of Cell Cycle Control in Budding Yeast." *Molecular Biology of the Cell* 15 (8): 3841–62.

Chenouard, Nicolas, Ihor Smal, Fabrice de Chaumont, Martin Maška, Ivo F. Sbalzarini, Yuanhao Gong, Janick Cardinale, et al. 2014. "Objective Comparison of Particle Tracking Methods." *Nature Methods* 11 (3): 281–89.

Ching, Travers, Daniel S. Himmelstein, Brett K. Beaulieu-Jones, Alexandr A. Kalinin, Brian T. Do, Gregory P. Way, Enrico Ferrero, et al. 2018. "Opportunities and Obstacles for Deep Learning in Biology and Medicine." *Journal of the Royal Society, Interface / the Royal Society* 15 (141). https://doi.org/10.1098/rsif.2017.0387.

Choudhury, Aaheli Roy, Zhenyu Ju, Meta W. Djojosubroto, Andrea Schienke, Andre Lechel, Sonja Schaetzlein, Hong Jiang, et al. 2007. "Cdkn1a Deletion Improves Stem Cell Function and Lifespan of Mice with Dysfunctional Telomeres without Accelerating Cancer Formation." *Nature Genetics* 39 (1): 99–105.

Çiçek, Özgün, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 2016. "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation." In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, 424–32. Springer International Publishing.

Cisco, V. N. I. 2017. "Cisco Visual Networking Index: Forecast and Methodology 2016--2021.(2017)."

Cliby, William A., Kriste A. Lewis, Kia K. Lilly, and Scott H. Kaufmann. 2002. "S Phase and G2 Arrests Induced by Topoisomerase I Poisons Are Dependent on ATR Kinase Function." *The Journal of Biological Chemistry* 277 (2): 1599–1606.

Coats, S., W. M. Flanagan, J. Nourse, and J. M. Roberts. 1996. "Requirement of p27Kip1 for Restriction Point Control of the Fibroblast Cell Cycle." *Science* 272 (5263): 877–80.

Cooper, Sam, and Chris Bakal. 2017a. "Accelerating Live Single-Cell Signalling Studies." *Trends in Biotechnology* 35 (5): 422–33.

Cooper, Sam, Alexis R. Barr, Robert Glen, and Chris Bakal. 2017a. "NucliTrack: An Integrated Nuclei Tracking Application." *Bioinformatics* 33 (20): 3320–22.

Cooper, Sam, Amine Sadok, Vicky Bousgouni, and Chris Bakal. 2015. "Apolar and Polar Transitions Drive the Conversion between Amoeboid and Mesenchymal Shapes in Melanoma Cells." *Molecular Biology of the Cell* 26 (22): 4163–70.

Couch, Frank B., Carol E. Bansbach, Robert Driscoll, Jessica W. Luzwick, Gloria G. Glick, Rémy Bétous, Clinton M. Carroll, et al. 2013. "ATR Phosphorylates SMARCAL1 to Prevent Replication Fork Collapse." *Genes & Development* 27 (14): 1610–23.

Coutu, D. L., and T. Schroeder. 2013. "Probing Cellular Processes by Long-Term Live Imaging–historic Problems and Current Solutions." *Journal of Cell Science*. http://jcs.biologists.org/content/126/17/3805.short.

Damoulakis, George, Laure Gambardella, Kent L. Rossman, Campbell D. Lawson, Karen E. Anderson, Yoshinori Fukui, Heidi C. Welch, Channing J. Der, Len R. Stephens, and Phillip T. Hawkins. 2014. "P-Rex1 Directly Activates RhoG to Regulate GPCR-Driven Rac Signalling and Actin Polarity in Neutrophils." *Journal of Cell Science* 127 (Pt 11): 2589–2600.

Davies, D. L., and D. W. Bouldin. 1979. "A Cluster Separation Measure." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1 (2): 224–27.

Davies, Helen, Graham R. Bignell, Charles Cox, Philip Stephens, Sarah Edkins, Sheila Clegg, Jon Teague, et al. 2002. "Mutations of the BRAF Gene in Human Cancer." *Nature* 417 (6892): 949–54.

Deiry, W. S. el-, T. Tokino, V. E. Velculescu, D. B. Levy, R. Parsons, J. M. Trent, D. Lin, W. E. Mercer, K. W. Kinzler, and B. Vogelstein. 1993. "WAF1, a Potential Mediator of p53 Tumor Suppression." *Cell* 75 (4): 817–25.

Deng, C., Zhang, P., Harper, J.W., Elledge, S.J. and Leder, P., 1995. Mice lacking p21CIP1/WAF1 undergo normal development, but are defective in G1 checkpoint control. *Cell*, *82*(4), pp.675-684.

Diest, P. J. van, J. P. Baak, P. Matze-Cok, E. C. Wisse-Brekelmans, C. M. van Galen, P. H. Kurver, S. M. Bellot, J. Fijnheer, L. H. van Gorp, and W. S. Kwee. 1992. "Reproducibility of Mitosis Counting in 2,469 Breast Cancer Specimens: Results from the Multicenter Morphometric Mammary Carcinoma Project." Human Pathology 23 (6): 603–7.

Dimitriadou, Evgenia, Sara Dolničar, and Andreas Weingessel. 2002. "An Examination of Indexes for Determining the Number of Clusters in Binary Data Sets." *Psychometrika* 67 (1): 137–59.

Ding, Chris, and Hanchuan Peng. 2005. "Minimum Redundancy Feature Selection from Microarray Gene Expression Data." *Journal of Bioinformatics and Computational Biology* 3 (2): 185–205.

Doyle, Andrew D., Francis W. Wang, Kazue Matsumoto, and Kenneth M. Yamada. 2009. "One-Dimensional Topography Underlies Three-Dimensional Fibrillar Cell Migration." *The Journal of Cell Biology* 184 (4): 481–90.

Dunlop, Malcolm G., Sara E. Dobbins, Susan Mary Farrington, Angela M. Jones, Claire Palles, Nicola Whiffin, Albert Tenesa, et al. 2012. "Common Variation near CDKN1A, POLD3 and SHROOM2 Influences Colorectal Cancer Risk." *Nature Genetics* 44 (7): 770–76.

Durkin, Sandra G., and Thomas W. Glover. 2007. "Chromosome Fragile Sites." *Annual Review of Genetics* 41: 169–92.

Duvenaud, David K., Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarell, Timothy Hirzel, Alan Aspuru-Guzik, and Ryan P. Adams. 2015. "Convolutional Networks on Graphs for Learning Molecular Fingerprints." In *Advances in Neural Information Processing Systems 28*, edited by C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, 2224–32. Curran Associates, Inc.

Dyson, N. 1998. "The Regulation of E2F by pRB-Family Proteins." *Genes & Development* 12 (15): 2245–62.

Eden, Sharon, Rajat Rohatgi, Alexandre V. Podtelejnikov, Matthias Mann, and Marc W. Kirschner. 2002. "Mechanism of Regulation of WAVE1-Induced Actin Nucleation by Rac1 and Nck." *Nature* 418 (6899): 790–93.

Eggert, Ulrike S., Amy A. Kiger, Constance Richter, Zachary E. Perlman, Norbert Perrimon, Timothy J. Mitchison, and Christine M. Field. 2004. "Parallel Chemical Genetic and Genome-Wide RNAi Screens Identify Cytokinesis Inhibitors and Targets." *PLoS Biology* 2 (12): e379.

Epps, D. E., M. L. Wolfe, and V. Groppi. 1994. "Characterization of the Steady-State and Dynamic Fluorescence Properties of the Potential-Sensitive Dye Bis-(1,3-Dibutylbarbituric Acid)trimethine Oxonol (Dibac4(3)) in Model Systems and Cells." *Chemistry and Physics of Lipids* 69 (2): 137–50.

Er, Süleyman, Changwon Suh, Michael P. Marshak, and Alán Aspuru-Guzik. 2015. "Computational Design of Molecules for an All-Quinone Redox Flow Battery." *Chemical Science* 6 (2): 885–93.

Etienne-Manneville, Sandrine, and Alan Hall. 2002. "Rho GTPases in Cell Biology." *Nature* 420 (6916): 629–35.

Evans, T., E. T. Rosenthal, J. Youngblom, D. Distel, and T. Hunt. 1983. "Cyclin: A Protein Specified by Maternal mRNA in Sea Urchin Eggs That Is Destroyed at Each Cleavage Division." *Cell* 33 (2): 389–96.

Fatehullah, Aliya, Si Hui Tan, and Nick Barker. 2016. "Organoids as an in Vitro Model of Human Development and Disease." *Nature Cell Biology* 18 (3): 246–54.

Forbes, Simon A., Nidhi Bindal, Sally Bamford, Charlotte Cole, Chai Yin Kok, David Beare, Mingming Jia, et al. 2011. "COSMIC: Mining Complete Cancer Genomes in the Catalogue of Somatic Mutations in Cancer." *Nucleic Acids Research* 39 (Database issue): D945–50.

Friedl, Peter, and Katarina Wolf. 2003. "Tumour-Cell Invasion and Migration: Diversity and Escape Mechanisms." *Nature Reviews. Cancer* 3 (5): 362–74.

Friedman, Jerome, Trevor Hastie, and Robert Tibshirani. 2001. *The Elements of Statistical Learning*. Vol. 1. Springer series in statistics New York, NY, USA:

Fritz, Rafael D., Michel Letzelter, Andreas Reimann, Katrin Martin, Ludovico Fusco, Laila Ritsma, Bas Ponsioen, et al. 2013. "A Versatile Toolkit to Produce Sensitive FRET Biosensors to Visualize Signaling in Time and Space." *Science Signaling* 6 (285): rs12.

Fuchs, Florian, Gregoire Pau, Dominique Kranz, Oleg Sklyar, Christoph Budjan, Sandra Steinbrink, Thomas Horn, Angelika Pedal, Wolfgang Huber, and Michael Boutros. 2010a. "Clustering Phenotype Populations by Genome-wide RNAi and Multiparametric Imaging." *Molecular Systems Biology* 6 (1): 370.

Gaudet, Suzanne, and Kathryn Miller-Jensen. 2016. "Redefining Signaling Pathways with an Expanding Single-Cell Toolbox." *Trends in Biotechnology* 34 (6): 458–69.

Gautier, J., J. Minshull, M. Lohka, M. Glotzer, T. Hunt, and J. L. Maller. 1990. "Cyclin Is a Component of Maturation-Promoting Factor from Xenopus." *Cell* 60 (3): 487–94.

Geng, Y., E. N. Eaton, M. Picón, J. M. Roberts, A. S. Lundberg, A. Gifford, C. Sardet, and R. A. Weinberg. 1996. "Regulation of Cyclin E Transcription by E2Fs and Retinoblastoma Protein." *Oncogene* 12 (6): 1173–80.

Georgia, Senta, Rosemary Soliz, Min Li, Pumin Zhang, and Anil Bhushan. 2006. "p57 and Hes1 Coordinate Cell Cycle Exit with Self-Renewal of Pancreatic Progenitors." *Developmental Biology* 298 (1): 22–31.

Georgitsi, Marianthi, Anniina Raitila, Auli Karhu, Rob B. van der Luijt, Cora M. Aalfs, Timo Sane, Outi Vierimaa, et al. 2007. "Germline CDKN1B/p27Kip1 Mutation in Multiple Endocrine Neoplasia." *The Journal of Clinical Endocrinology and Metabolism* 92 (8): 3321–25.

Girshick, Ross. 2015. "Fast R-Cnn." In *Proceedings of the IEEE International Conference on Computer Vision*, 1440–48.

Giuliano, Kenneth A., Robbin L. DeBiasio, R. Terry Dunlay, Albert Gough, Joanne M. Volosky, Joseph Zock, George N. Pavlakis, and D. Lansing Taylor. 1997. "High-Content Screening: A New Approach to Easing Key Bottlenecks in the Drug Discovery Process." *Journal of Biomolecular Screening* 2 (4): 249–59.

Godinez, William J., Imtiaz Hossain, Stanley E. Lazic, John W. Davies, and Xian Zhang. 2017. "A Multi-Scale Convolutional Neural Network for Phenotyping High-Content Cellular Images." *Bioinformatics* 33 (13): 2010–19.

Gómez-Bombarelli, Rafael, Jennifer N. Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, Ryan P. Adams, and Alán Aspuru-Guzik. 2018. "Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules." *ACS Central Science* 4 (2): 268–76.

González, J. E., and P. A. Negulescu. 1998. “Intracellular Detection Assays for High-Throughput Screening.” *Current Opinion in Biotechnology* 9 (6): 624–31.

Gosselet, Fabien P., Thierry Magnaldo, Raphaël M. Culerrier, Alain Sarasin, and Jean-Claude Ehrhart. 2007. “BMP2 and BMP6 Control p57(Kip2) Expression and Cell Growth Arrest/terminal Differentiation in Normal Primary Human Epidermal Keratinocytes.” *Cellular Signalling* 19 (4): 731–39.

Gottesbühren, Undine, Ritu Garg, Philippe Riou, Brad McColl, Daniel Brayson, and Anne J. Ridley. 2013. “Rnd3 Induces Stress Fibres in Endothelial Cells through RhoB.” *Biology Open* 2 (2): 210–16.

Graña, X., and E. P. Reddy. 1995. “Cell Cycle Control in Mammalian Cells: Role of Cyclins, Cyclin Dependent Kinases (CDKs), Growth Suppressor Genes and Cyclin-Dependent Kinase Inhibitors (CKIs).” *Oncogene* 11 (2): 211–19.

Hajdo-Milasinović, Amra, Saskia I. J. Ellenbroek, Saskia van Es, Babet van der Vaart, and John G. Collard. 2007. “Rac1 and Rac3 Have Opposing Functions in Cell Adhesion and Differentiation of Neuronal Cells.” *Journal of Cell Science* 120 (Pt 4): 555–66.

Hajdo-Milasinovic, Amra, Rob A. van der Kammen, Zvezdana Moneva, and John G. Collard. 2009. “Rac3 Inhibits Adhesion and Differentiation of Neuronal Cells by Modifying GIT1 Downstream Signaling.” *Journal of Cell Science* 122 (Pt 12): 2127–36.

Hall, A. 1998. “Rho GTPases and the Actin Cytoskeleton.” *Science* 279 (5350): 509–14.

Hall, A., C. J. Marshall, N. K. Spurr, and R. A. Weiss. 1983. “Identification of Transforming Gene in Two Human Sarcoma Cell Lines as a New Member of the Ras Gene Family Located on Chromosome 1.” *Nature* 303 (5916): 396–400.

Hanahan, Douglas, and Robert A. Weinberg. 2011. “Hallmarks of Cancer: The next Generation.” *Cell* 144 (5): 646–74.

Haralick, Robert M., K. Shanmugam, Its’hak Dinstein, and Others. 1973. “Textural Features for Image Classification.” *IEEE Transactions on Systems, Man, and Cybernetics* 3 (6): 610–21.

Haralick, Robert M., and Linda G. Shapiro. 1985. “Image Segmentation Techniques.” In *Applications of Artificial Intelligence II*, 0548:2–10. International Society for Optics and Photonics.

Harms, H., U. Gunzer, and H. M. Aus. 1986. “Combined Local Color and Texture Analysis of Stained Cells.” *Computer Vision, Graphics, and Image Processing* 33 (3): 364–76.

Harms, H., U. Gunzer, H. M. Aus, A. Rüter, M. Haucke, and V. ter Meulen. 1979. “Computer Aided Analysis of Chromatin Network and Basophil Color for Differentiation of Mononuclear Peripheral Blood Cells.” *The Journal of Histochemistry and Cytochemistry: Official Journal of the Histochemistry Society* 27 (1): 204–9.

Hartwell, L. H., and M. B. Kastan. 1994. “Cell Cycle Control and Cancer.” *Science* 266 (5192): 1821–28.

Havens, Courtney G., and Johannes C. Walter. 2011. “Mechanism of CRL4(Cdt2), a PCNA-Dependent E3 Ubiquitin Ligase.” *Genes & Development* 25 (15): 1568–82.

Hayashi, Akiyo, Naohiro Suenaga, Yasushi Shiomi, and Hideo Nishitani. 2014. “PCNA-Dependent Ubiquitination of Cdt1 and p21 in Mammalian Cells.” *Methods in Molecular Biology* 1170: 367–82.

He, Kaiming, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. 2018. "Mask R-CNN." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, June. https://doi.org/10.1109/TPAMI.2018.2844175.

Heldt, F. S., A. R. Barr, and S. Cooper. 2018. "A Comprehensive Model for the Proliferation–quiescence Decision in Response to Endogenous DNA Damage in Human Cells." *Proceedings of the*. http://www.pnas.org/content/early/2018/02/15/1715345115.short.

Heppt, Markus V., Timo Siepmann, Jutta Engel, Gabriele Schubert-Fritschle, Renate Eckel, Laura Mirlach, Thomas Kirchner, et al. 2017. "Prognostic Significance of BRAF and NRAS Mutations in Melanoma: A German Study from Routine Care." *BMC Cancer* 17 (1): 536.

Heuvel, S. van den, and E. Harlow. 1993. "Distinct Roles for Cyclin-Dependent Kinases in Cell Cycle Control." *Science* 262 (5142): 2050–54.

Hilsenbeck, Oliver, Michael Schwarzfischer, Stavroula Skylaki, Bernhard Schauberger, Philipp S. Hoppe, Dirk Loeffler, Konstantinos D. Kokkaliaris, et al. 2016. "Software Tools for Single-Cell Tracking and Quantification of Cellular and Molecular Properties." *Nature Biotechnology* 34 (7): 703–6.

Hodis, Eran, Ian R. Watson, Gregory V. Kryukov, Stefan T. Arold, Marcin Imielinski, Jean-Philippe Theurillat, Elizabeth Nickerson, et al. 2012. "A Landscape of Driver Mutations in Melanoma." *Cell* 150 (2): 251–63.

Holevinsky, K. O., Z. Fan, M. Frame, J. C. Makielski, V. Groppi, and D. J. Nelson. 1994. "ATP-Sensitive K+ Channel Opener Acts as a Potent Cl− Channel Inhibitor in Vascular Smooth Muscle Cells." *The Journal of Membrane Biology* 137 (1): 59–70.

Hollstein, M., D. Sidransky, B. Vogelstein, and C. C. Harris. 1991. "p53 Mutations in Human Cancers." *Science* 253 (5015): 49–53.

Holmes, W.R., Park, J., Levchenko, A. and Edelstein-Keshet, L., 2017. A mathematical model coupling polarity signalling to cell adhesion explains diverse cell migration patterns. *PLoS computational biology*, *13*(5), p.e1005524.

Horn, Thomas, Thomas Sandmann, Bernd Fischer, Elin Axelsson, Wolfgang Huber, and Michael Boutros. 2011. "Mapping of Signaling Networks through Synthetic Genetic Interaction Analysis by RNAi." *Nature Methods* 8 (4): 341–46.

Hotelling, Harold. 1933. "Analysis of a Complex of Statistical Variables into Principal Components." *Journal of Educational Psychology* 24 (6): 417.

Htun, H., J. Barsony, I. Renyi, D. L. Gould, and G. L. Hager. 1996. "Visualization of Glucocorticoid Receptor Translocation and Intranuclear Organization in Living Cells with a Green Fluorescent Protein Chimera." *Proceedings of the National Academy of Sciences of the United States of America* 93 (10): 4845–50.

Hurley, Paula J., and Fred Bunz. 2009. "Distinct Pathways Involved in S-Phase Checkpoint Control." In *Checkpoint Controls and Targets in Cancer Therapy*, edited by Zahid H. Siddik, 27–36. Totowa, NJ: Humana Press.

Ikegami, S., T. Taguchi, M. Ohashi, M. Oguro, H. Nagano, and Y. Mano. 1978. "Aphidicolin Prevents Mitotic Cell Division by Interfering with the Activity of DNA Polymerase-Alpha." *Nature* 275 (5679): 458–60.

Jackman, Mark, Catherine Lindon, Erich A. Nigg, and Jonathon Pines. 2003. "Active Cyclin B1–Cdk1 First Appears on Centrosomes in Prophase." *Nature Cell Biology* 5 (January): 143.

Jackson, R.J., Adnane, J., Coppola, D., Cantor, A., Sebti, S.M. and Pledger, W.J., 2002. Loss of the cell cycle inhibitors p21 Cip1 and p27 Kip1 enhances tumorigenesis in knockout mouse models. *Oncogene*, *21*(55), p.8486.

Jaggi, B., S. S. Poon, C. MacAulay, and B. Palcic. 1988. "Imaging System for Morphometric Assessment of Absorption or Fluorescence in Stained Cells." *Cytometry* 9 (6): 566–72.

Jain, Anil K. 2010. "Data Clustering: 50 Years beyond K-Means." *Pattern Recognition Letters* 31 (8): 651–66.

Jain, Anil K., and Farshid Farrokhnia. 1991. "Unsupervised Texture Segmentation Using Gabor Filters." *Pattern Recognition* 24 (12): 1167–86.

Jänicke, Reiner U., Dennis Sohn, Frank Essmann, and Klaus Schulze-Osthoff. 2007. "The Multiple Battles Fought by Anti-Apoptotic p21." *Cell Cycle* 6 (4): 407–13.

Johnson, Amy, and Jan M. Skotheim. 2013. "Start and the Restriction Point." *Current Opinion in Cell Biology* 25 (6): 717–23.

Johnson, Heath E., Yogesh Goyal, Nicole L. Pannucci, Trudi Schüpbach, Stanislav Y. Shvartsman, and Jared E. Toettcher. 2017. "The Spatiotemporal Limits of Developmental Erk Signaling." *Developmental Cell* 40 (2): 185–92.

Johnson, Heath E., Stanislav Y. Shvartsman, and Jared E. Toettcher. 2018. "Signaling Dynamics Control Cell Fate in the Early Drosophila Embryo." *bioRxiv*. https://doi.org/10.1101/342998.

Jones, Clifford D., Kevin Blades, Kevin M. Foote, Sylvie M. Guichard, Philip J. Jewsbury, Thomas McGuire, Johannes W. Nissink, et al. 2013. "Abstract 2348: Discovery of AZD6738, a Potent and Selective Inhibitor with the Potential to Test the Clinical Efficacy of ATR Kinase Inhibition in Cancer Patients." *Cancer Research* 73 (8 Supplement): 2348–2348.

Jones, Thouis R., Anne E. Carpenter, Michael R. Lamprecht, Jason Moffat, Serena J. Silver, Jennifer K. Grenier, Adam B. Castoreno, et al. 2009. "Scoring Diverse Cellular Morphologies in Image-Based Screens with Iterative Feedback and Machine Learning." *Proceedings of the National Academy of Sciences of the United States of America* 106 (6): 1826–31.

Jones, Thouis R., In Han Kang, Douglas B. Wheeler, Robert A. Lindquist, Adam Papallo, David M. Sabatini, Polina Golland, and Anne E. Carpenter. 2008. "CellProfiler Analyst: Data Exploration and Analysis Software for Complex Image-Based Screens." *BMC Bioinformatics* 9 (November): 482.

Kastan, M. B., O. Onyekwere, D. Sidransky, B. Vogelstein, and R. W. Craig. 1991. "Participation of p53 Protein in the Cellular Response to DNA Damage." *Cancer Research* 51 (23 Pt 1): 6304–11.

Katou, Yuki, Yutaka Kanoh, Masashige Bando, Hideki Noguchi, Hirokazu Tanaka, Toshihiko Ashikari, Katsunori Sugimoto, and Katsuhiko Shirahige. 2003. "S-Phase Checkpoint Proteins Tof1 and Mrc1 Form a Stable Replication-Pausing Complex." *Nature* 424 (6952): 1078–83.

Kerz, Maximilian, Amos Folarin, Ruta Meleckyte, Fiona M. Watt, Richard J. Dobson, and Davide Danovi. 2016. "A Novel Automated High-Content Analysis Workflow Capturing Cell Population Dynamics from Induced Pluripotent Stem Cell Live Imaging Data." *Journal of Biomolecular Screening* 21 (9): 887–96.

Kholodenko, Boris N. 2006. "Cell-Signalling Dynamics in Time and Space." *Nature Reviews. Molecular Cell Biology* 7 (3): 165–76.

Khotanzad, A., and Y. H. Hong. 1990. "Invariant Image Recognition by Zernike Moments." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12 (5): 489–97.

Kibel, Adam S., Brian K. Suarez, Jay Belani, Joe Oh, Raul Webster, Michele Brophy-Ebbers, Chan Guo, William J. Catalona, Joel Picus, and Paul J. Goodfellow. 2003. "CDKN1A and CDKN1B Polymorphisms and Risk of Advanced Prostate Carcinoma." *Cancer Research* 63 (9): 2033–36.

Kiger, A. A., B. Baum, S. Jones, M. R. Jones, A. Coulson, C. Echeverri, and N. Perrimon. 2003. "A Functional Genomic Analysis of Cell Morphology Using RNA Interference." *Journal of Biology* 2 (4): 27.

Kimmel, J.C., Chang, A.Y., Brack, A.S. and Marshall, W.F., 2018. Inferring cell state by quantitative motility analysis reveals a dynamic state system and broken detailed balance. *PLoS computational biology*, *14*(1), p.e1005927.

Kimura, K., M. Ito, M. Amano, K. Chihara, Y. Fukata, M. Nakafuku, B. Yamamori, et al. 1996. "Regulation of Myosin Phosphatase by Rho and Rho-Associated Kinase (Rho-Kinase)." *Science* 273 (5272): 245–48.

King, R. W., R. J. Deshaies, J. M. Peters, and M. W. Kirschner. 1996. "How Proteolysis Drives the Cell Cycle." *Science* 274 (5293): 1652–59.

Knudson, A. G., Jr. 1971. "Mutation and Cancer: Statistical Study of Retinoblastoma." *Proceedings of the National Academy of Sciences of the United States of America* 68 (4): 820–23.

Kojima, Kensuke, Marina Konopleva, Teresa McQueen, Susan O'Brien, William Plunkett, and Michael Andreeff. 2006. "Mdm2 Inhibitor Nutlin-3a Induces p53-Mediated Apoptosis by Transcription-Dependent and Transcription-Independent Mechanisms and May Overcome Atm-Mediated Resistance to Fludarabine in Chronic Lymphocytic Leukemia." *Blood* 108 (3): 993–1000.

Kolch, Walter, Melinda Halasz, Marina Granovskaya, and Boris N. Kholodenko. 2015. "The Dynamic Control of Signal Transduction Networks in Cancer Cells." *Nature Reviews. Cancer* 15 (9): 515–27.

Kraus, Oren Z., Jimmy Lei Ba, and Brendan J. Frey. 2016. "Classifying and Segmenting Microscopy Images with Deep Multiple Instance Learning." *Bioinformatics* 32 (12): i52–59.

Kraus, Oren Z., and Brendan J. Frey. 2016. "Computer Vision for High Content Screening." *Critical Reviews in Biochemistry and Molecular Biology* 51 (2): 102–9.

Kraus, Oren Z., Ben T. Grys, Jimmy Ba, Yolanda Chong, Brendan J. Frey, Charles Boone, and Brenda J. Andrews. 2017. "Automated Analysis of High-content Microscopy Data with Deep Learning." *Molecular Systems Biology* 13 (4): 924.

Krauthammer, Michael, Yong Kong, Byung Hak Ha, Perry Evans, Antonella Bacchiocchi, James P. McCusker, Elaine Cheng, et al. 2012. "Exome Sequencing Identifies Recurrent Somatic RAC1 Mutations in Melanoma." *Nature Genetics* 44 (9): 1006–14.

Kristan, Matej, Jiri Matas, Ales Leonardis, Michael Felsberg, Luka Cehovin, Gustavo Fernandez, Tomas Vojir, Gustav Hager, Georg Nebehay, and Roman Pflugfelder. 2015. "The Visual Object Tracking vot2015 Challenge Results." In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 1–23.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2012. "ImageNet Classification with Deep Convolutional Neural Networks." In *Advances in Neural Information Processing Systems 25*, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, 1097–1105. Curran Associates, Inc.

Krugmann, S., I. Jordens, K. Gevaert, M. Driessens, J. Vandekerckhove, and A. Hall. 2001. "Cdc42 Induces Filopodia by Promoting the Formation of an IRSp53:Mena Complex." *Current Biology: CB* 11 (21): 1645–55.

Kudo, Takamasa, Stevan Jeknić, Derek N. Macklin, Sajia Akhter, Jacob J. Hughey, Sergi Regot, and Markus W. Covert. 2018. "Live-Cell Measurements of Kinase Activity in Single Cells Using Translocation Reporters." *Nature Protocols* 13 (1): 155–69.

Kümmel, Anne, Hanspeter Gubler, Patricia Gehin, Martin Beibel, Daniela Gabriel, and Christian N. Parker. 2010. "Integration of Multiple Readouts into the Z′ Factor for Assay Quality Assessment." *Journal of Biomolecular Screening* 15 (1): 95–101.

Kuo, Linda J., and Li-Xi Yang. 2008. "Gamma-H2AX - a Novel Biomarker for DNA Double-Strand Breaks." *In Vivo* 22 (3): 305–9.

Lahav, Galit, Nitzan Rosenfeld, Alex Sigal, Naama Geva-Zatorsky, Arnold J. Levine, Michael B. Elowitz, and Uri Alon. 2004. "Dynamics of the p53-Mdm2 Feedback Loop in Individual Cells." *Nature Genetics* 36 (2): 147–50.

Lamprecht, Michael R., David M. Sabatini, and Anne E. Carpenter. 2007. "CellProfiler™: Free, Versatile Software for Automated Biological Image Analysis." *BioTechniques* 42 (1): 71–75.

Lam, W. W., I. Hatada, S. Ohishi, T. Mukai, J. A. Joyce, T. R. Cole, D. Donnai, W. Reik, P. N. Schofield, and E. R. Maher. 1999. "Analysis of Germline CDKN1C (p57KIP2) Mutations in Familial and Sporadic Beckwith-Wiedemann Syndrome (BWS) Provides a Novel Genotype-Phenotype Correlation." *Journal of Medical Genetics* 36 (7): 518–23.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." *Nature* 521 (7553): 436–44.

Lecun, Y., L. Bottou, Y. Bengio, and P. Haffner. 1998. "Gradient-Based Learning Applied to Document Recognition." *Proceedings of the IEEE* 86 (11): 2278–2324.

Lezaja, Aleksandra, and Matthias Altmeyer. 2018. "Inherited DNA Lesions Determine G1 Duration in the next Cell Cycle." *Cell Cycle* 17 (1): 24–32.

Li, F. P., J. F. Fraumeni Jr, J. J. Mulvihill, W. A. Blattner, M. G. Dreyfus, M. A. Tucker, and R. W. Miller. 1988. "A Cancer Family Syndrome in Twenty-Four Kindreds." *Cancer Research* 48 (18): 5358–62.

Likar, B., J. B. Maintz, M. A. Viergever, and F. Pernus. 2000. "Retrospective Shading Correction Based on Entropy Minimization." *Journal of Microscopy* 197 (Pt 3): 285–95.

Li, Linfeng, Qiong Zhou, Ty C. Voss, Kevin L. Quick, and Daniel V. LaBarbera. 2016. "High-Throughput Imaging: Focusing in on Drug Discovery in 3D." *Methods* 96 (March): 97–102.

Lin, Gang, Umesh Adiga, Kathy Olson, John F. Guzowski, Carol A. Barnes, and Badrinath Roysam. 2003. "A Hybrid 3D Watershed Algorithm Incorporating Gradient Cues and Object Models for Automatic Segmentation of Nuclei in Confocal Image Stacks." *Cytometry. Part A: The Journal of the International Society for Analytical Cytology* 56 (1): 23–36.

Li, Simon, Sébastien Besson, Colin Blackburn, Mark Carroll, Richard K. Ferguson, Helen Flynn, Kenneth Gillen, et al. 2016. "Metadata Management for High Content Screening in OMERO." *Methods* 96 (March): 27–32.

Liu, Q., S. Guntuku, X. S. Cui, and S. Matsuoka. 2000. "Chk1 Is an Essential Kinase That Is Regulated by Atr and Required for the G2/M DNA Damage Checkpoint." *Genes*. http://genesdev.cshlp.org/content/14/12/1448.short.

Liu, Yan-Jun, Maël Le Berre, Franziska Lautenschlaeger, Paolo Maiuri, Andrew Callan-Jones, Mélina Heuzé, Tohru Takaki, Raphaël Voituriez, and Matthieu Piel. 2015. "Confinement and Low Adhesion Induce Fast Amoeboid Migration of Slow Mesenchymal Cells." *Cell* 160 (4): 659–72.

Ljosa, Vebjorn, Peter D. Caie, Rob Ter Horst, Katherine L. Sokolnicki, Emma L. Jenkins, Sandeep Daya, Mark E. Roberts, et al. 2013. "Comparison of Methods for Image-Based Profiling of Cellular Morphological Responses to Small-Molecule Treatment." *Journal of Biomolecular Screening* 18 (10): 1321–29.

Long, Jonathan, Evan Shelhamer, and Trevor Darrell. 2015. "Fully Convolutional Networks for Semantic Segmentation." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–40.

Loo, Lit-Hsin, Lani F. Wu, and Steven J. Altschuler. 2007. "Image-Based Multivariate Profiling of Drug Responses from Single Cells." *Nature Methods* 4 (5): 445–53.

Lu, Alex, Oren Z. Kraus, Sam Cooper, and Alan M. Moses. 2018. "Learning Unsupervised Feature Representations for Single Cell Microscopy Images with Paired Cell Inpainting." *bioRxiv*. https://doi.org/10.1101/395954.

Maaten, Laurens van der, and Geoffrey Hinton. 2008. "Visualizing Data Using T-SNE." *Journal of Machine Learning Research: JMLR* 9 (Nov): 2579–2605.

Maekawa, M., T. Ishizaki, S. Boku, N. Watanabe, A. Fujita, A. Iwamatsu, T. Obinata, K. Ohashi, K. Mizuno, and S. Narumiya. 1999. "Signaling from Rho to the Actin Cytoskeleton through Protein Kinases ROCK and LIM-Kinase." *Science* 285 (5429): 895–98.

Magnusson, K. E. G., and J. Jaldén. 2012. "A Batch Algorithm Using Iterative Application of the Viterbi Algorithm to Track Cells and Construct Cell Lineages." In *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, 382–85.

Magnusson, Klas E. G., Joakim Jalden, Penney M. Gilbert, and Helen M. Blau. 2015. "Global Linking of Cell Tracks Using the Viterbi Algorithm." *IEEE Transactions on Medical Imaging* 34 (4): 911–29.

Ma, Hongxia, Guangfu Jin, Zhibin Hu, Xiangjun Zhai, Wensen Chen, Shui Wang, Xuechen Wang, et al. 2006. "Variant Genotypes of CDKN1A and CDKN1B Are Associated with an Increased Risk of Breast Cancer in Chinese Women." *International Journal of Cancer. Journal International Du Cancer* 119 (9): 2173–78.

Malkin, D., F. P. Li, L. C. Strong, J. F. Fraumeni Jr, C. E. Nelson, D. H. Kim, J. Kassel, M. A. Gryka, F. Z. Bischoff, and M. A. Tainsky. 1990. "Germ Line p53 Mutations in a Familial Syndrome of Breast Cancer, Sarcomas, and Other Neoplasms." *Science* 250 (4985): 1233–38.

Malpica, N., C. O. de Solórzano, J. J. Vaquero, A. Santos, I. Vallcorba, J. M. García-Sagredo, and F. del Pozo. 1997. "Applying Watershed Algorithms to the Segmentation of Clustered Nuclei." *Cytometry* 28 (4): 289–97.

Martín-Caballero, J., Flores, J.M., García-Palencia, P. and Serrano, M., 2001. Tumor susceptibility of p21Waf1/Cip1-deficient mice. *Cancer research*, *61*(16), pp.6234-6238.

Martin, Heather L., Matthew Adams, Julie Higgins, Jacquelyn Bond, Ewan E. Morrison, Sandra M. Bell, Stuart Warriner, Adam Nelson, and Darren C. Tomlinson. 2014. "High-Content, High-Throughput Screening for the Identification of Cytotoxic Compounds Based on Cell Morphology and Cell Proliferation Markers." *PloS One* 9 (2): e88338.

Maška, Martin, Vladimír Ulman, David Svoboda, Pavel Matula, Petr Matula, Cristina Ederra, Ainhoa Urbiola, et al. 2014. "A Benchmark for Comparison of Cell Tracking Algorithms." *Bioinformatics* 30 (11): 1609–17.

Massagué, Joan. 2004. "G1 Cell-Cycle Control and Cancer." *Nature* 432 (7015): 298–306.

Matsui, Takeshi, Mutsuki Amano, Takaharu Yamamoto, Kazuyasu Chihara, Masato Nakafuku, Masaaki Ito, Takeshi Nakano, Katsuya Okawa, Akihiro Iwamatsu, and Kozo Kaibuchi. 1996. "Rho-Associated Kinase, a Novel Serine/threonine Kinase, as a Putative Target for Small GTP Binding Protein Rho." *The EMBO Journal* 15 (9): 2208–16.

Matsushime, H., M. E. Ewen, D. K. Strom, J. Y. Kato, S. K. Hanks, M. F. Roussel, and C. J. Sherr. 1992. "Identification and Properties of an Atypical Catalytic Subunit (p34PSK-J3/cdk4) for Mammalian D Type G1 Cyclins." *Cell* 71 (2): 323–34.

Matsushime, H., M. F. Roussel, R. A. Ashmun, and C. J. Sherr. 1991. "Colony-Stimulating Factor 1 Regulates Novel Cyclins during the G1 Phase of the Cell Cycle." *Cell* 65 (4): 701–13.

McQuin, Claire, Allen Goodman, Vasiliy Chernyshev, Lee Kamentsky, Beth A. Cimini, Kyle W. Karhohs, Minh Doan, et al. 2018. "CellProfiler 3.0: Next-Generation Image Processing for Biology." *PLoS Biology* 16 (7): e2005970.

Meijering, E. 2012. "Cell Segmentation: 50 Years Down the Road [Life Sciences]." *IEEE Signal Processing Magazine* 29 (5): 140–45.

Meijering, Erik, Oleh Dzyubachyk, and Ihor Smal. 2012. "Methods for Cell and Particle Tracking." *Methods in Enzymology* 504: 183–200.

Meyerson, M., and E. Harlow. 1994. "Identification of G1 Kinase Activity for cdk6, a Novel Cyclin D Partner." *Molecular and Cellular Biology* 14 (3): 2077–86.

Michael Ando, D., Cory McLean, and Marc Berndl. 2017. "Improving Phenotypic Measurements in High-Content Imaging Screens." *bioRxiv*. https://doi.org/10.1101/161422.

Milan, Anton, Laura Leal-Taixe, Ian Reid, Stefan Roth, and Konrad Schindler. 2016. "MOT16: A Benchmark for Multi-Object Tracking." *arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/1603.00831.

Moffat, Jason, and David M. Sabatini. 2006. "Building Mammalian Signalling Pathways with RNAi Screens." *Nature Reviews. Molecular Cell Biology* 7 (3): 177–87.

Molnar, Csaba, Ian H. Jermyn, Zoltan Kato, Vesa Rahkama, Päivi Östling, Piia Mikkonen, Vilja Pietiäinen, and Peter Horvath. 2016. "Accurate Morphology Preserving Segmentation of Overlapping Cells Based on Active Contours." *Scientific Reports* 6 (August): 32412.

Moser, Justin, Iain Miller, Dylan Carter, and Sabrina L. Spencer. 2018. "Control of the Restriction Point by Rb and p21." *Proceedings of the National Academy of Sciences of the United States of America*, August. https://doi.org/10.1073/pnas.1722446115.

Mukherji, Mridul, Russell Bell, Lubica Supekova, Yan Wang, Anthony P. Orth, Serge Batalov, Loren Miraglia, et al. 2006. "Genome-Wide Functional Analysis of Human Cell-Cycle Regulators." *Proceedings of the National Academy of Sciences of the United States of America* 103 (40): 14819–24.

Murphree, A. L., and W. F. Benedict. 1984. "Retinoblastoma: Clues to Human Oncogenesis." *Science* 223 (4640): 1028–33.

Murphy, Robert F., Michael V. Boland, Meel Velliste, and Others. 2000. "Towards a Systematics for Protein Subcellular Location: Quantitative Description of Protein Localization Patterns and Automated Analysis of Fluorescence Microscope Images." In *ISMB*, 251–59.

Nederlof, P. M., S. van der Flier, N. P. Verwoerd, J. Vrolijk, A. K. Raap, and H. J. Tanke. 1992. "Quantification of Fluorescence in Situ Hybridization Signals by Image Cytometry." *Cytometry* 13 (8): 846–52.

Neumann, Beate, Michael Held, Urban Liebel, Holger Erfle, Phill Rogers, Rainer Pepperkok, and Jan Ellenberg. 2006. "High-Throughput RNAi Screening by Time-Lapse Imaging of Live Human Cells." *Nature Methods* 3 (5): 385–90.

Neumann, Beate, Thomas Walter, Jean-Karim Hériché, Jutta Bulkescher, Holger Erfle, Christian Conrad, Phill Rogers, et al. 2010. "Phenotypic Profiling of the Human Genome by Time-Lapse Microscopy Reveals Cell Division Genes." *Nature* 464 (7289): 721–27.

Ng, Alvin Y. J., Jagath C. Rajapakse, Roy E. Welsch, Paul T. Matsudaira, Victor Horodincu, and James G. Evans. 2010. "A Cell Profiling Framework for Modeling Drug Responses from HCS Imaging." *Journal of Biomolecular Screening* 15 (7): 858–68.

Nir, Oaz, Chris Bakal, Norbert Perrimon, and Bonnie Berger. 2010. "Inference of RhoGAP/GTPase Regulation Using Single-Cell Morphological Data from a Combinatorial RNAi Screen." *Genome Research* 20 (3): 372–80.

Nobes, C. D., and A. Hall. 1995. "Rho, Rac, and cdc42 GTPases Regulate the Assembly of Multimolecular Focal Complexes Associated with Actin Stress Fibers, Lamellipodia, and Filopodia." *Cell* 81 (1): 53–62.

Norbury, C., and P. Nurse. 1992. "Animal Cell Cycles and Their Control." *Annual Review of Biochemistry* 61: 441–70.

Nurse, P. 1990. "Universal Control Mechanism Regulating Onset of M-Phase." *Nature* 344 (6266): 503–8.

Nurse, Paul. "A long twentieth century of the cell cycle and beyond." *Cell* 100.1 (2000): 71-78.

Olivier, Nicolas, Miguel A. Luengo-Oroz, Louise Duloquin, Emmanuel Faure, Thierry Savy, Israël Veilleux, Xavier Solinas, et al. 2010. "Cell Lineage Reconstruction of Early Zebrafish Embryos Using Label-Free Nonlinear Microscopy." *Science* 329 (5994): 967–71.

'Brien, C.A., Kreso, A., Ryan, P., Hermans, K.G., Gibson, L., Wang, Y., Tsatsanis, A., Gallinger, S. and Dick, J.E., 2012. ID1 and ID3 regulate the self-renewal capacity of human colon cancer-initiating cells through p21. *Cancer cell*, *21*(6), pp.777-792.

Otsu, Nobuyuki. 1979. "A Threshold Selection Method from Gray-Level Histograms." *IEEE Transactions on Systems, Man, and Cybernetics* 9 (1): 62–66.

Overton, K. Wesley, Sabrina L. Spencer, William L. Noderer, Tobias Meyer, and Clifford L. Wang. 2014. "Basal p21 Controls Population Heterogeneity in Cycling and Quiescent Cell Cycle States." *Proceedings of the National Academy of Sciences of the United States of America* 111 (41): E4386–93.

Pagano, Michele. 2004. "Control of DNA Synthesis and Mitosis by the Skp2-p27-Cdk1/2 Axis." *Molecular Cell* 14 (4): 414–16.

Pal, Nikhil R., and Sankar K. Pal. 1993. "A Review on Image Segmentation Techniques." *Pattern Recognition* 26 (9): 1277–94.

Paňková, K., D. Rösel, M. Novotný, and Jan Brábek. 2010. "The Molecular Mechanisms of Transition between Mesenchymal and Amoeboid Invasiveness in Tumor Cells." *Cellular and Molecular Life Sciences: CMLS* 67 (1): 63–71.

Pardee, A. B. 1974. "A Restriction Point for Control of Normal Animal Cell Proliferation." *Proceedings of the National Academy of Sciences of the United States of America* 71 (4): 1286–90.

Pardee, Arthur B., Chiang J. Li, and G. Prem Veer Reddy. 2004. "Regulation in S Phase by E2F." *Cell Cycle* 3 (9): 1091–94.

Parsons, J. Thomas, Alan Rick Horwitz, and Martin A. Schwartz. 2010. "Cell Adhesion: Integrating Cytoskeletal Dynamics and Cellular Tension." *Nature Reviews. Molecular Cell Biology* 11 (9): 633–43.

Paszek, Matthew J., Nastaran Zahir, Kandice R. Johnson, Johnathon N. Lakins, Gabriela I. Rozenberg, Amit Gefen, Cynthia A. Reinhart-King, et al. 2005. "Tensional Homeostasis and the Malignant Phenotype." *Cancer Cell* 8 (3): 241–54.

Paul, Steven M., Daniel S. Mytelka, Christopher T. Dunwiddie, Charles C. Persinger, Bernard H. Munos, Stacy R. Lindborg, and Aaron L. Schacht. 2010. "How to Improve R&D Productivity: The Pharmaceutical Industry's Grand Challenge." *Nature Reviews. Drug Discovery* 9 (3): 203–14.

Pawlowski, Nick, Juan C. Caicedo, Shantanu Singh, Anne E. Carpenter, and Amos Storkey. 2016. "Automating Morphological Profiling with Generic Deep Convolutional Networks." *bioRxiv*. https://doi.org/10.1101/085118.

Pepperkok, Rainer, and Jan Ellenberg. 2006. "High-Throughput Fluorescence Microscopy for Systems Biology." *Nature Reviews. Molecular Cell Biology* 7 (9): 690–96.

Perlman, Zachary E., Michael D. Slack, Yan Feng, Timothy J. Mitchison, Lani F. Wu, and Steven J. Altschuler. 2004. "Multidimensional Drug Profiling by Automated Microscopy." *Science* 306 (5699): 1194–98.

Pertz, Olivier, Louis Hodgson, Richard L. Klemke, and Klaus M. Hahn. 2006. "Spatiotemporal Dynamics of RhoA Activity in Migrating Cells." *Nature* 440 (7087): 1069–72.

Peruski, Anne Harwood, Linwood Hill Johnson 3rd, and Leonard Francis Peruski Jr. 2002. "Rapid and Sensitive Detection of Biological Warfare Agents Using Time-Resolved Fluorescence Assays." *Journal of Immunological Methods* 263 (1-2): 35–41.

Peters, Jan-Michael. 2002. "The Anaphase-Promoting Complex: Proteolysis in Mitosis and beyond." *Molecular Cell* 9 (5): 931–43.

Planas-Silva, M. D., and R. A. Weinberg. 1997. "The Restriction Point and Control of Cell Proliferation." *Current Opinion in Cell Biology* 9 (6): 768–72.

Pollard, Steven M., Koichi Yoshikawa, Ian D. Clarke, Davide Danovi, Stefan Stricker, Roslin Russell, Jane Bayani, et al. 2009. "Glioma Stem Cell Lines Expanded in Adherent Culture Have Tumor-Specific Phenotypes and Are Suitable for Chemical and Genetic Screens." *Cell Stem Cell* 4 (6): 568–80.

Polyak, K., Y. Xia, J. L. Zweier, K. W. Kinzler, and B. Vogelstein. 1997. "A Model for p53-Induced Apoptosis." *Nature* 389 (6648): 300–305.

Poole, A.J., Heap, D., Carroll, R.E. and Tyner, A.L., 2004. Tumor suppressor functions for the Cdk inhibitor p21 in the mouse colon. *Oncogene*, *23*(49), p.8128.

Prewitt, J. M., and M. L. Mendelsohn. 1966. "The Analysis of Cell Images." *Annals of the New York Academy of Sciences* 128 (3): 1035–53.

Puls, T. J., Xiaohong Tan, Mahera Husain, Catherine F. Whittington, Melissa L. Fishel, and Sherry L. Voytik-Harbin. 2018. "Development of a Novel 3D Tumor-Tissue Invasion Model for High-Throughput, High-Content Phenotypic Drug Screening." *Scientific Reports* 8 (1): 13039.

Purvis, J. E., and G. Lahav. 2013. "Encoding and Decoding Cellular Information through Signaling Dynamics." *Cell*. https://www.sciencedirect.com/science/article/pii/S0092867413001530.

Purvis, Jeremy E., Kyle W. Karhohs, Caroline Mock, Eric Batchelor, Alexander Loewer, and Galit Lahav. 2012. "p53 Dynamics Control Cell Fate." *Science* 336 (6087): 1440–44.

Rainville, Irene, and Judy E. Garber. 2008. "Familial Cancer Database Online." *The Lancet Oncology* 9 (10): 925–26.

Rämö, Pauli, Raphael Sacher, Berend Snijder, Boris Begemann, and Lucas Pelkmans. 2009. "CellClassifier: Supervised Learning of Cellular Phenotypes." *Bioinformatics* 25 (22): 3028–30.

Regot, Sergi, Jacob J. Hughey, Bryce T. Bajar, Silvia Carrasco, and Markus W. Covert. 2014. "High-Sensitivity Measurements of Multiple Kinase Activities in Live Single Cells." *Cell* 157 (7): 1724–34.

Reid, Brian G., Taleen Jerjian, Purvi Patel, Qiong Zhou, Byong Hoon Yoo, Peter Kabos, Carol A. Sartorius, and Daniel V. Labarbera. 2014. "Live Multicellular Tumor Spheroid Models for High-Content Imaging and Screening in Cancer Drug Discovery." *Current Chemical Genomics and Translational Medicine* 8 (Suppl 1): 27–35.

Reisen, Felix, Xian Zhang, Daniela Gabriel, and Paul Selzer. 2013. "Benchmarking of Multivariate Similarity Measures for High-Content Screening Fingerprints in Phenotypic Drug Discovery." *Journal of Biomolecular Screening* 18 (10): 1284–97.

Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. 2017. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (6): 1137–49.

Reyes, José, Jia-Yun Chen, Jacob Stewart-Ornstein, Kyle W. Karhohs, Caroline S. Mock, and Galit Lahav. 2018. "Fluctuations in p53 Signaling Allow Escape from Cell-Cycle Arrest." *Molecular Cell* 71 (4): 581–91.e5.

Ridley, Anne J. 2012. "Historical Overview of Rho GTPases." In *Rho GTPases: Methods and Protocols*, edited by Francisco Rivero, 3–12. New York, NY: Springer New York.

Roerdink, Jos Btm, and Arnold Meijster. 2000. "The Watershed Transform: Definitions, Algorithms and Parallelization Strategies." *Fundamenta Informaticae* 41 (1, 2): 187–228.

Rohatgi, R., H. Y. Ho, and M. W. Kirschner. 2000. "Mechanism of N-WASP Activation by CDC42 and Phosphatidylinositol 4, 5-Bisphosphate." *The Journal of Cell Biology* 150 (6): 1299–1310.

Rohatgi, R., L. Ma, H. Miki, M. Lopez, T. Kirchhausen, T. Takenawa, and M. W. Kirschner. 1999. "The Interaction between N-WASP and the Arp2/3 Complex Links Cdc42-Dependent Signals to Actin Assembly." *Cell* 97 (2): 221–31.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–41. Springer International Publishing.

Rousseeuw, Peter J. 1987. "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis." *Journal of Computational and Applied Mathematics* 20 (November): 53–65.

Rudner, A. D., and A. W. Murray. 1996. "The Spindle Assembly Checkpoint." *Current Opinion in Cell Biology* 8 (6): 773–80.

Ruprecht, Verena, Stefan Wieser, Andrew Callan-Jones, Michael Smutny, Hitoshi Morita, Keisuke Sako, Vanessa Barone, et al. 2015. "Cortical Contractility Triggers a Stochastic Switch to Fast Amoeboid Cell Motility." *Cell* 160 (4): 673–85.

Sachs, Norman, and Hans Clevers. 2014. "Organoid Cultures for the Analysis of Cancer Phenotypes." *Current Opinion in Genetics & Development* 24 (February): 68–73.

Sadok, Amine, and Chris J. Marshall. 2014. "Rho GTPases: Masters of Cell Migration." *Small GTPases* 5 (June): e29710.

Sahai, Erik, and Christopher J. Marshall. 2003. "Differing Modes of Tumour Cell Invasion Have Distinct Requirements for Rho/ROCK Signalling and Extracellular Proteolysis." *Nature Cell Biology* 5 (8): 711–19.

Sailem, Heba, Vicky Bousgouni, Sam Cooper, and Chris Bakal. 2014a. "Cross-Talk between Rho and Rac GTPases Drives Deterministic Exploration of Cellular Shape Space and Morphological Heterogeneity." *Open Biology* 4 (January): 130132.

Sailem, Heba Z., Sam Cooper, and Chris Bakal. 2016. "Visualizing Quantitative Microscopy Data: History and Challenges." *Critical Reviews in Biochemistry and Molecular Biology* 51 (2): 96–101.

Sanz-Moreno, Victoria, Gilles Gadea, Jessica Ahn, Hugh Paterson, Pierfrancesco Marra, Sophie Pinner, Erik Sahai, and Christopher J. Marshall. 2008. "Rac Activation and Inactivation Control Plasticity of Tumor Cell Movement." *Cell* 135 (3): 510–23.

Sanz-Moreno, Victoria, Cedric Gaggioli, Maggie Yeo, Jean Albrengues, Fredrik Wallberg, Amaya Viros, Steven Hooper, et al. 2011. "ROCK and JAK1 Signaling Cooperate to Control Actomyosin Contractility in Tumor Cells and Stroma." *Cancer Cell* 20 (2): 229–45.

Scandura, Joseph M., Piernicola Boccuni, Joan Massagué, and Stephen D. Nimer. 2004. "Transforming Growth Factor β-Induced Cell Cycle Arrest of Human Hematopoietic Cells Requires p57KIP2 up-Regulation." *Proceedings of the National Academy of Sciences of the United States of America* 101 (42): 15231–36.

Schultz, L. B., N. H. Chehab, A. Malikzay, and T. D. Halazonetis. 2000. "p53 Binding Protein 1 (53BP1) Is an Early Participant in the Cellular Response to DNA Double-Strand Breaks." *The Journal of Cell Biology* 151 (7): 1381–90.

Schulze, Christopher J., Walter M. Bray, Marcos H. Woerhmann, Joshua Stuart, R. Scott Lokey, and Roger G. Linington. 2013. "'Function-First' Lead Discovery: Mode of Action Profiling of Natural Product Libraries Using Image-Based Screening." *Chemistry & Biology* 20 (2): 285–95.

Schwarz, Clayton, Amy Johnson, Mardo Kõivomägi, Evgeny Zatulovskiy, Carolyn J. Kravitz, Andreas Doncic, and Jan M. Skotheim. 2018. "A Precise Cdk Activity Threshold Determines Passage through the Restriction Point." *Molecular Cell* 69 (2): 253–64.e5.

Schwarz, Gideon. 1978. "Estimating the Dimension of a Model." *Annals of Statistics* 6 (2): 461–64.

Sero, Julia E., and Chris Bakal. 2017. "Multiparametric Analysis of Cell Shape Demonstrates That β-PIX Directly Couples YAP Activation to Extracellular Matrix Adhesion." *Cell Systems* 4 (1): 84–96.e6.

Sero, Julia E., Heba Zuhair Sailem, Rico Chandra Ardy, Hannah Almuttaqi, Tongli Zhang, and Chris Bakal. 2015. "Cell Shape and the Microenvironment Regulate Nuclear Translocation of NF-κB in Breast Epithelial and Tumor Cells." *Molecular Systems Biology* 11 (3): 790.

Shamir, Ohad, and Naftali Tishby. 2008. "Model Selection and Stability in K-Means Clustering." http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.228.1429&rep=rep1&type=pdf#page=375.

D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, 2009. "On the Reliability of Clustering Stability in the Large Sample Regime." In *Advances in Neural Information Processing Systems 21*, edited by 1465–72. Curran Associates, Inc.

Shan, Jing, Robert E. Schwartz, Nathan T. Ross, David J. Logan, David Thomas, Stephen A. Duncan, Trista E. North, Wolfram Goessling, Anne E. Carpenter, and Sangeeta N. Bhatia. 2013. "Identification of Small Molecules for Human Hepatocyte Expansion and iPS Differentiation." *Nature Chemical Biology* 9 (8): 514–20.

Sha, Wei, Jonathan Moore, Katherine Chen, Antonio D. Lassaletta, Chung-Seon Yi, John J. Tyson, and Jill C. Sible. 2003. "Hysteresis Drives Cell-Cycle Transitions in Xenopus Laevis Egg Extracts." *Proceedings of the National Academy of Sciences of the United States of America* 100 (3): 975–80.

Sherr, C. J. 1993. "Mammalian G1 Cyclins." *Cell* 73 (6): 1059–65.

Sherr, C. J., and J. M. Roberts. 1999. "CDK Inhibitors: Positive and Negative Regulators of G1-Phase Progression." *Genes & Development* 13 (12): 1501–12.

Shieh, S. Y., M. Ikeda, Y. Taya, and C. Prives. 1997. "DNA Damage-Induced Phosphorylation of p53 Alleviates Inhibition by MDM2." *Cell* 91 (3): 325–34.

Shi, Jianbo, and J. Malik. 2000. "Normalized Cuts and Image Segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (8): 888–905.

Simm, Jaak, Günter Klambauer, Adam Arany, Marvin Steijaert, Jörg Kurt Wegner, Emmanuel Gustin, Vladimir Chupakhin, et al. 2018. "Repurposing High-Throughput Image Assays Enables Biological Activity Prediction for Drug Discovery." *Cell Chemical Biology* 25 (5): 611–18.e3.

Simpson, C. M., and C. Bakal. 2018. "The Dynamics of ERK Signaling in Melanoma, and the Response to BRAF or MEK Inhibition, Are Cell Cycle Dependent." *bioRxiv*. https://www.biorxiv.org/content/early/2018/04/23/306571.abstract.

Singh, Dinesh Kumar, Chin-Jen Ku, Chonlarat Wichaidit, Robert J. Steininger 3rd, Lani F. Wu, and Steven J. Altschuler. 2010. "Patterns of Basal Signaling Heterogeneity Can Distinguish Cellular Populations with Different Drug Sensitivities." *Molecular Systems Biology* 6 (May): 369.

Singh, S., M-A Bray, T. R. Jones, and A. E. Carpenter. 2014. “Pipeline for Illumination Correction of Images for High-Throughput Microscopy.” *Journal of Microscopy* 256 (3): 231–36.

Singh, Shantanu, Xiaoyun Wu, Vebjorn Ljosa, Mark-Anthony Bray, Federica Piccioni, David E. Root, John G. Doench, Jesse S. Boehm, and Anne E. Carpenter. 2015. “Morphological Profiles of RNAi-Induced Gene Knockdown Are Highly Reproducible but Dominated by Seed Effects.” *PloS One* 10 (7): e0131370.

Skardal, Aleksander, Thomas Shupe, and Anthony Atala. 2016. “Organoid-on-a-Chip and Body-on-a-Chip Systems for Drug Screening and Disease Modeling.” *Drug Discovery Today* 21 (9): 1399–1411.

Smith, Carl, Andrew Pomiankowski, and Duncan Greig. 2014. “Size and Competitive Mating Success in the Yeast Saccharomyces Cerevisiae.” *Behavioral Ecology: Official Journal of the International Society for Behavioral Ecology* 25 (2): 320–27.

Smith, George E. 2010. “Nobel Lecture: The Invention and Early History of the CCD.” *Reviews of Modern Physics* 82 (3): 2307–12.

Snijder, Berend, and Lucas Pelkmans. 2011. “Origins of Regulated Cell-to-Cell Variability.” *Nature Reviews. Molecular Cell Biology* 12 (2): 119–25.

Snijder, Berend, Raphael Sacher, Pauli Rämö, Eva-Maria Damm, Prisca Liberali, and Lucas Pelkmans. 2009. “Population Context Determines Cell-to-Cell Variability in Endocytosis and Virus Infection.” *Nature* 461 (7263): 520–23.

Snijder, Berend, Raphael Sacher, Pauli Rämö, Prisca Liberali, Karin Mench, Nina Wolfrum, Laura Burleigh, et al. 2012. “Single-Cell Analysis of Population Context Advances RNAi Screening at Multiple Levels.” *Molecular Systems Biology* 8 (April): 579.

Sommer, Christoph, Christoph N. Straehle, Ullrich Koethe, Fred A. Hamprecht, and Others. 2011. “Ilastik: Interactive Learning and Segmentation Toolkit.” In *ISBI*, 2:8.

Soos, T. J., H. Kiyokawa, J. S. Yan, M. S. Rubin, A. Giordano, A. DeBlasio, S. Bottega, B. Wong, J. Mendelsohn, and A. Koff. 1996. “Formation of p27-CDK Complexes during the Human Mitotic Cell Cycle.” *Cell Growth & Differentiation: The Molecular Biology Journal of the American Association for Cancer Research* 7 (2): 135–46.

Spencer, C. C., and R. C. Bostrom. 1962. “Performance of the Cytoanalyzer in Recent Clinical Trials.” *Journal of the National Cancer Institute* 29 (August): 267–76.

Spencer, Sabrina L., Steven D. Cappell, Feng-Chiao Tsai, K. Wesley Overton, Clifford L. Wang, and Tobias Meyer. 2013. “The Proliferation-Quiescence Decision Is Controlled by a Bifurcation in CDK2 Activity at Mitotic Exit.” *Cell* 155 (2): 369–83.

Stewart-Ornstein, Jacob, and Galit Lahav. 2016a. “Dynamics of CDKN1A in Single Cells Defined by an Endogenous Fluorescent Tagging Toolkit.” *Cell Reports* 14 (7): 1800–1811.

Süel, Gürol M., Jordi Garcia-Ojalvo, Louisa M. Liberman, and Michael B. Elowitz. 2006. “An Excitable Gene Regulatory Circuit Induces Transient Cellular Differentiation.” *Nature* 440 (7083): 545–50.

Swedlow, Jason R., Ilya Goldberg, Erik Brauner, and Peter K. Sorger. 2003. “Informatics and Quantitative Analysis in Biological Imaging.” *Science* 300 (5616): 100–102.

Swedlow, Jason R., Ilya G. Goldberg, Kevin W. Eliceiri, and OME Consortium. 2009. “Bioimage Informatics for Experimental Biology.” *Annual Review of Biophysics* 38: 327–46.

Swinney, David C., and Jason Anthony. 2011. "How Were New Medicines Discovered?" *Nature Reviews. Drug Discovery* 10 (7): 507–19.

Tabor, Daniel P., Loïc M. Roch, Semion K. Saikin, Christoph Kreisbeck, Dennis Sheberla, Joseph H. Montoya, Shyam Dwaraknath, et al. 2018. "Accelerating the Discovery of Materials for Clean Energy in the Era of Smart Automation." *Nature Reviews Materials* 3 (5): 5–20.

Tape, Christopher J., Stephanie Ling, Maria Dimitriadi, Kelly M. McMahon, Jonathan D. Worboys, Hui Sun Leong, Ida C. Norrie, et al. 2016. "Oncogenic KRAS Regulates Tumor Cell Signaling via Stromal Reciprocation." *Cell* 165 (7): 1818.

Tenenbaum, J. B., V. de Silva, and J. C. Langford. 2000. "A Global Geometric Framework for Nonlinear Dimensionality Reduction." *Science* 290 (5500): 2319–23.

Terada, Y., S. Inoshita, O. Nakashima, M. Kuwahara, S. Sasaki, and F. Marumo. 1999. "Regulation of Cyclin D1 Expression and Cell Cycle Progression by Mitogen-Activated Protein Kinase Cascade." *Kidney International* 56 (4): 1258–61.

Thomas, A. D., T. Davies, and A. R. Luxmoore. 1992. "The Hough Transform for Locating Cell Nuclei." *Analytical and Quantitative Cytology and Histology / the International Academy of Cytology [and] American Society of Cytology* 14 (4): 347–53.

Thorburn, A. L. 1974. "Alfred François Donné, 1801-1878, Discoverer of Trichomonas Vaginalis and of Leukaemia." *The British Journal of Venereal Diseases* 50 (5): 377–80.

Tibshirani, Robert, Guenther Walther, and Trevor Hastie. 2001. "Estimating the Number of Clusters in a Data Set via the Gap Statistic." *Journal of the Royal Statistical Society. Series B, Statistical Methodology* 63 (2): 411–23.

Tiscornia, Gustavo, Erica Lorenzo Vivas, and Juan Carlos Izpisúa Belmonte. 2011. "Diseases in a Dish: Modeling Human Genetic Disorders Using Induced Pluripotent Cells." *Nature Medicine* 17 (12): 1570–76.

Toettcher, Jared E., Orion D. Weiner, and Wendell A. Lim. 2013. "Using Optogenetics to Interrogate the Dynamic Control of Signal Transmission by the Ras/Erk Module." *Cell* 155 (6): 1422–34.

Topley, G.I., Okuyama, R., Gonzales, J.G., Conti, C. and Dotto, G.P., 1999. p21WAF1/Cip1 functions as a suppressor of malignant skin tumor formation and a determinant of keratinocyte stem-cell potential. *Proceedings of the National Academy of Sciences*, *96*(16), pp.9089-9094.

Towne, Danli L., Emily E. Nicholl, Kenneth M. Comess, Scott C. Galasinski, Philip J. Hajduk, and Vivek C. Abraham. 2012. "Development of a High-Content Screening Assay Panel to Accelerate Mechanism of Action Studies for Oncology Research." *Journal of Biomolecular Screening* 17 (8): 1005–17.

Toyoshima, H., and T. Hunter. 1994. "p27, a Novel Inhibitor of G1 Cyclin-Cdk Protein Kinase Activity, Is Related to p21." *Cell* 78 (1): 67–74.

Tozluoğlu, Melda, Alexander L. Tournier, Robert P. Jenkins, Steven Hooper, Paul A. Bates, and Erik Sahai. 2013. "Matrix Geometry Determines Optimal Cancer Cell Migration Strategy and Modulates Response to Interventions." *Nature Cell Biology* 15 (June): 751.

Tyson, John J., Katherine C. Chen, and Bela Novak. 2003. "Sniffers, Buzzers, Toggles and Blinkers: Dynamics of Regulatory and Signaling Pathways in the Cell." *Current Opinion in Cell Biology* 15 (2): 221–31.

Tyson, John J., Attila Csikasz-Nagy, and Bela Novak. 2002. "The Dynamics of Cell Cycle Regulation." *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 24 (12): 1095–1109.

Ubersax, Jeffrey A., Erika L. Woodbury, Phuong N. Quang, Maria Paraz, Justin D. Blethrow, Kavita Shah, Kevan M. Shokat, and David O. Morgan. 2003. "Targets of the Cyclin-Dependent Kinase Cdk1." *Nature* 425 (6960): 859–64.

Urbani, L., S. W. Sherwood, and R. T. Schimke. 1995. "Dissociation of Nuclear and Cytoplasmic Cell Cycle Progression by Drugs Employed in Cell Synchronization." *Experimental Cell Research* 219 (1): 159–68.

Vassilev, Lyubomir T. 2004. "Small-Molecule Antagonists of p53-MDM2 Binding: Research Tools and Potential Therapeutics." *Cell Cycle* 3 (4): 419–21.

Veta, Mitko, Paul J. van Diest, Stefan M. Willems, Haibo Wang, Anant Madabhushi, Angel Cruz-Roa, Fabio Gonzalez, et al. 2015. "Assessment of Algorithms for Mitosis Detection in Breast Cancer Histopathology Images." *Medical Image Analysis* 20 (1): 237–48.

Vincent, Pascal, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. 2010. "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion." *Journal of Machine Learning Research: JMLR* 11 (Dec): 3371–3408.

Vogelstein, Bert, Nickolas Papadopoulos, Victor E. Velculescu, Shibin Zhou, Luis A. Diaz Jr, and Kenneth W. Kinzler. 2013. "Cancer Genome Landscapes." *Science* 339 (6127): 1546–58.

Waga, S., G. J. Hannon, D. Beach, and B. Stillman. 1994. "The p21 Inhibitor of Cyclin-Dependent Kinases Controls DNA Replication by Interaction with PCNA." *Nature* 369 (6481): 574–78.

Waldman, T., K. W. Kinzler, and B. Vogelstein. 1995. "p21 Is Necessary for the p53-Mediated G1 Arrest in Human Cancer Cells." *Cancer Research* 55 (22): 5187–90.

Wang, Lijun, Wanli Ouyang, Xiaogang Wang, and Huchuan Lu. 2015. "Visual Tracking with Fully Convolutional Networks." In *Proceedings of the IEEE International Conference on Computer Vision*, 3119–27.

Wan, Paul T. C., Mathew J. Garnett, S. Mark Roe, Sharlene Lee, Dan Niculescu-Duvaz, Valerie M. Good, C. Michael Jones, et al. 2004. "Mechanism of Activation of the RAF-ERK Signaling Pathway by Oncogenic Mutations of B-RAF." *Cell* 116 (6): 855–67.

Weinberg, R. A. 1995. "The Retinoblastoma Protein and Cell Cycle Control." *Cell* 81 (3): 323–30.

Wettersten, Hiromi I., Sung Hee Hwang, Cuiwen Li, Eunice Y. Shiu, Aaron T. Wecksler, Bruce D. Hammock, and Robert H. Weiss. 2013. "A Novel p21 Attenuator Which Is Structurally Related to Sorafenib." *Cancer Biology & Therapy* 14 (3): 278–85.

Woehrmann, Marcos H., Walter M. Bray, James K. Durbin, Sean C. Nisam, Alicia K. Michael, Emerson Glassey, Joshua M. Stuart, and R. Scott Lokey. 2013. "Large-Scale Cytological Profiling for Functional Analysis of Bioactive Compounds." *Molecular bioSystems* 9 (11): 2604–17.

Wolf, Steffen, Lukas Schott, Ullrich Köthe, and Fred A. Hamprecht. 2017. "Learned Watershed: End-to-End Learning of Seeded Segmentation." In *ICCV*, 2030–38.

Wollman, Roy, and Nico Stuurman. 2007. "High Throughput Microscopy: From Raw Images to Discoveries." *Journal of Cell Science* 120 (Pt 21): 3715–22.

Xiong, Y., G. J. Hannon, H. Zhang, D. Casso, R. Kobayashi, and D. Beach. 1993. "p21 Is a Universal Inhibitor of Cyclin Kinases." *Nature* 366 (6456): 701–4.

Yamazaki, D., S. Kurisu, and T. Takenawa. 2009. "Involvement of Rac and Rho Signaling in Cancer Cell Motility in 3D Substrates." *Oncogene* 28 (13): 1570–83.

Yang, Hee Won, Mingyu Chung, Takamasa Kudo, and Tobias Meyer. 2017. "Competing Memories of Mitogen and p53 Signalling Control Cell-Cycle Entry." *Nature* 549 (7672): 404–8.

Yang, Qiong, and James E. Ferrell Jr. 2013. "The Cdk1-APC/C Cell Cycle Oscillator Circuit Functions as a Time-Delayed, Ultrasensitive Switch." *Nature Cell Biology* 15 (5): 519–25.

Yan, Y., J. Frisén, M. H. Lee, J. Massagué, and M. Barbacid. 1997. "Ablation of the CDK Inhibitor p57Kip2 Results in Increased Apoptosis and Delayed Differentiation during Mouse Development." *Genes & Development* 11 (8): 973–83.

Yin, Zheng, Amine Sadok, Heba Sailem, Afshan McCarthy, Xiaofeng Xia, Fuhai Li, Mar Arias Garcia, et al. 2013. "A Screen for Morphological Complexity Identifies Regulators of Switch-like Transitions between Discrete Cell Shapes." *Nature Cell Biology* 15 (7): 860–71.

Yin, Zheng, Xiaobo Zhou, Chris Bakal, Fuhai Li, Youxian Sun, Norbert Perrimon, and Stephen T. C. Wong. 2007. "Online Phenotype Discovery in High-Content RNAi Screens Using Gap Statistics." *AIP Conference Proceedings* 952 (1): 86–95.

Young, Daniel W., Andreas Bender, Jonathan Hoyt, Elizabeth McWhinnie, Gung-Wei Chirn, Charles Y. Tao, John A. Tallarico, et al. 2008. "Integrating High-Content Screening and Ligand-Target Prediction to Identify Mechanism of Action." *Nature Chemical Biology* 4 (1): 59–68.

Yu, Fang, Judit Megyesi, Robert L. Safirstein, and Peter M. Price. 2005. "Identification of the Functional Domain of p21WAF1/CIP1 That Protects Cells from Cisplatin Cytotoxicity." *American Journal of Physiology-Renal Physiology* 289 (3): F514–20.

Zachariae, W., and K. Nasmyth. 1999. "Whose End Is Destruction: Cell Division and the Anaphase-Promoting Complex." *Genes & Development* 13 (16): 2039–58.

Zanella, Fabian, James B. Lorens, and Wolfgang Link. 2010a. "High Content Screening: Seeing Is Believing." *Trends in Biotechnology* 28 (5): 237–45.

Zerjatke, Thomas, Igor A. Gak, Dilyana Kirova, Markus Fuhrmann, Katrin Daniel, Magdalena Gonciarz, Doris Müller, Ingmar Glauche, and Jörg Mansfeld. 2017. "Quantitative Cell Cycle Analysis Based on an Endogenous All-in-One Reporter for Cell Tracking and Classification." *Cell Reports* 19 (9): 1953–66.

Zernike, von F. 1934. "Beugungstheorie Des Schneidenver-Fahrens Und Seiner Verbesserten Form, Der Phasenkontrastmethode." *Physica* 1 (7): 689–704.

Zhang, Jin, and Michael D. Allen. 2007. "FRET-Based Biosensors for Protein Kinases: Illuminating the Kinome." *Molecular bioSystems* 3 (11): 759–65.

Zhang, Pumin, Nanette J. Liegeois, Calvin Wong, Milton Finegold, Harry Hou, Janet C. Thompson, Adam Silverman, J. Wade Harper, Ronald A. DePinho, and Stephen J. Elledge. 1997. "Altered Cell Differentiation and Proliferation in Mice Lacking p57KIP2 Indicates a Role in Beckwith--Wiedemann Syndrome." *Nature* 387 (6629): 151–58.

Zielke, N. and Edgar, B.A., 2015. FUCCI sensors: powerful new tools for analysis of cell proliferation. *Wiley Interdisciplinary Reviews: Developmental Biology*, *4*(5), pp.469-487.

Zou, Lee, and Stephen J. Elledge. 2003. "Sensing DNA Damage through ATRIP Recognition of RPA-ssDNA Complexes." *Science* 300 (5625): 1542–48.

7 Contributions

First/joint first author publications

Cooper, S., Sadok, A., Bousgouni, V., & Bakal, C. (2015). Apolar and polar transitions drive the conversion between amoeboid and mesenchymal shapes in melanoma cells. Molecular biology of the cell, 26(22), 4163-4170.

Barr, A. R.*, **Cooper, S.***, Heldt, F. S.*, Butera, F., Stoy, H., Mansfeld, J., ... & Bakal, C. (2017). DNA damage during S-phase mediates the proliferation-quiescence decision in the subsequent G1 via p21 expression. *Nature communications*, *8*, 14728. ***[Top 5% cited new paper in Nature Communications 2017]***

Cooper, S., Barr, A. R., Glen, R., & Bakal, C. (2017). NucliTrack: An integrated nuclei tracking application. Bioinformatics, 33(20), 3320-3322.

Cooper, S., & Bakal, C. (2017). Accelerating live single-cell signalling studies. Trends in biotechnology, 35(5), 422-433.

Additional publications

Heldt, F. S., Barr, A. R., **Cooper, S.**, Bakal, C., & Novák, B. (2018). A comprehensive model for the proliferation–quiescence decision in response to endogenous DNA damage in human cells. Proceedings of the National Academy of Sciences, 201715345.

Caicedo J.C., **Cooper S.,** Heigwer F., Warchal S., Qiu P., Molnar C., Vasilevich A.S., Barry J.D., Bansal H.S., Kraus O., Wawer M. (2017). Data-analysis strategies for image-based cell profiling. Nature methods. 849.

Pascual-Vargas, P., **Cooper, S.**, Sero, J., Bousgouni, V., Arias-Garcia, M., & Bakal, C. (2017). RNAi screens for Rho GTPase regulators of cell shape and YAP/TAZ localisation in triple negative breast cancer. Scientific data, 4, 170018.

Sailem, H. Z., **Cooper, S.,** & Bakal, C. (2016). Visualizing quantitative microscopy data: History and challenges. Critical reviews in biochemistry and molecular biology, 51(2), 96-101.

Sailem, H., Bousgouni, V., **Cooper, S.**, & Bakal, C. (2014). Cross-talk between Rho and Rac GTPases drives deterministic exploration of cellular shape space and morphological heterogeneity. Open biology, 4(1), 130132.

Hoffman, A.F., Nolan, J., Gebhard, D.F., Nickischer, D., Omta, W., **Cooper, S.**, Presnell, S., Wardwell-Swanson, J. and Fennell, M., 2018. Society of Biomolecular Imaging and Informatics High-Content Screening/High-Content Analysis Emerging Technologies in Biological Models, When and Why?. Assay and drug development technologies, 16(1), pp.1-6.

Lu A., Kraus Z. O., **Cooper S**., Moses A. M., Learning unsupervised feature representations for single cell microscopy images with paired cell inpainting bioRxiv 395954 [Preprint, Under Review]

Hansel C. S., Crowder S. W., **Cooper S.,** Gopal S., Pardelha da Cruz J. M., Martins, L. O., Keller D., Rtoehry S., Becce M., Cass A. E. G., Bakal C., Chiappini C., Stevens M. M., Nanoneedle-mediated stimulation of individual organelles reveals non-canonical regulation of mechanotransduction machinery [Submitted]