# Nanoparticles in medicine: Automating the analysis process of high-throughput microscopy data.

Tonkin, James

# Nanoparticles in Medicine: Automating the Analysis Process of High-Throughput Microscopy Data

James Tonkin

Submitted to the University of Wales in fulfilment of the requirements for the Degree of Doctor of Philosophy

Swansea University

2013



ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

# Abstract

Automated tracking of cells across timelapse microscopy image sequences typically employs complex segmentation routines and/or bio-staining of the tracking objective. Often accurate identification of a cell's morphology is not of interest and the accurate segmentation of cells in pursuit of non-morphological parameters is complex and time consuming. This thesis explores the potential of internalized quantum dot nanoparticles as alternative, bio- and photo-stable optical markers for tracking the motions of cells through time. CdTe/ZnS core-shell quantum dots act as nodes in moving light display networks within A549, epithelial, lung cancer cells over a 40 hour time period. These quantum dot fluorescence sources are identified and interpreted using simplistic algorithms to find consistent, non-subjective centroids that represent cell centre locations. The presented tracking protocols yield an approximate 91% success rate over 24 hours and 78% over the full 40 hours. The nanoparticle moving light displays also provide simultaneous collection of cell motility data, resolution of mitotic traversal dynamics and identification of familial relationships enabling the construction of multi-parameter lineage trees. This principle is then developed further through inclusion of 3 different coloured quantum dots to create cell specific colour barcodes and reduce the number of time points necessary to successfully track cells through time. The tracking software and identification of parameters without detailed morphological knowledge is also demonstrated through automated extraction of DOX accumulation profiles and Cobalt agglomeration accruement statistics from two separate toxicology assays without the need for cell segmentation.

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed .. ..... (candidate)

Date .....13/8/13.....

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed .. ..... (candidate)

Date .....13/8/13.....

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ..... ..... (candidate)

Date .....13/8/13.....

# Contents

*Chapter VI: Additional Application – PEI-QD Toxicology*



*Chapter VII: Cobalt Toxicology Assay*



*Chapter VIII: Conclusions*

# Acknowledgements

I wish to thank my supervisor, Professor Huw Summers, for his guidance and advice throughout the course of this research. I would also like to thank Paul Rees and Rowan Brown for their assistances through productive discussions and interactions along the way.

In addition, I extend my gratitude to Rachel Errington and Sally Chappell of Cardiff University's School of Medicine; Sabeel Shamsudeen and Matt Ware, both of Swansea University and The Methodist Hospital Research Institute and Nicole Hondow of the University of Leeds' Institute for Materials Research for supplying the various raw data analysed throughout this text.

Finally, I wish to thank my parents for their unconditional support, as always, through another of life's journeys.

# Figures List

*Chapter I: Introduction*



*Chapter II: Automatic Tracking of A549 Cells*



*Chapter III: Automated Tracking - Data Mining and Analysis*

*Chapter IV: Tri-Colour Labelling*



*Chapter V: Automated Drug Dosage Tracking*

*Chapter VI: Additional Application – PEI-QD Toxicology*



*Chapter VII: Cobalt Toxicology Assay*

# CHAPTER ONE

# Introduction

## *Microscopy Overview*

Microscopy systems utilise microscopes to visualise samples and objects that are below the resolution range of the unaided eye and can be broadly categorised into three major branches; optical, electron and scanning probe microscopy. Applications and analysis within this text deal with optical microscopy which itself can take multiple forms, some of which shall be discussed in this section. Optical microscopes are typically configured for both bright-field (transmitted), phase contrast and epifluorescence microscopy. Figure 1.1 provides a diagrammatic representation of a typical inverted microscope system and the locations of key elements. The optical pathway directs light by way of mirrors from the illumination source through a condenser to illuminate the specimen.

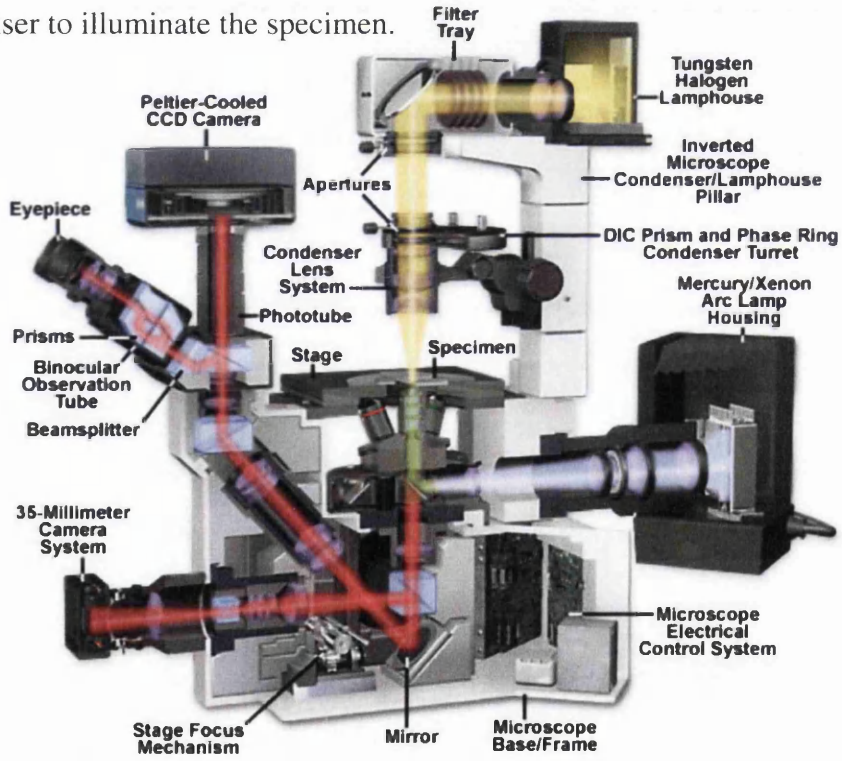



Figure 1.1: Schematic diagram of an Olympus IX70 inverted microscope with the light pathways highlighted. (Source: Olympus Microscopy Resource Centre)

It then passes through the objective lens and onto the image collection medium, whether an eyepiece or a digital camera device. This inverted configuration allows

for the imaging of cells in containers under more natural conditions than is achievable with a more traditional light microscope. Potential formation of an image of the light source in the sample image plane is overcome by using Köhler illumination. First applied by August Köhler, the light path of the illumination beam is exactly controlled to perfectly defocus the light source from the sample plane. Standard microscopic resolution is diffraction limited to approximately 0.2μm and objects in closer proximity than this are indistinguishable from one another [1]. The exact resolution is governed by the equation

$$D = \frac{0.6\lambda}{Nsin(\alpha)} \qquad (1.1)$$

where N is the refractive index of the medium between the specimen and the objective lens, α is the angular aperture or half angle of the light cone entering the objective lens from the specimen and λ is the wavelength of the incident light. Phase contrast microscopy builds on this basic configuration by taking advantage of differences in the refractive index and thickness of cellular material. These methods are a commonly exploited approach for visualization and segmentation of unstained/unlabelled cells. Phase contrast microscopy [2],[3],[4], a Nobel prize winning invention of Frits Zernike [5], generates images within which the degree of brightness of a region of the sample is dependent on the refractive index of that region. This is achieved by differing manipulations of the separated background illumination light and light scattered by the specimen. Figure 1.2 depicts a schematic of the necessary modifications to the optical pathway.

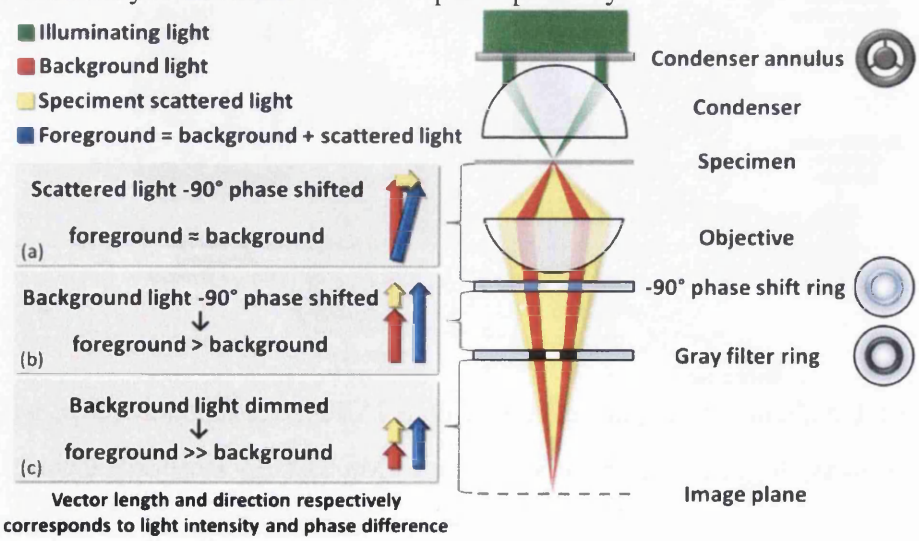



Figure 1.2: Diagram detailing the optical pathway modifications applied to the previously demonstrated microscope optical pathway for phase contrast microscopy. (Source: Wikimedia Commons)

In negative phase contrast microscopy, image contrast is increased by applying a -90° shift to the background light by means of a phase shift ring. This removes the phase difference between the scattered light and the background resulting in an increased intensity difference between the image foreground and background levels (Fig 1.2 (b)). Contrast is further increased by dimming the background using a gray filter ring (Fig 1.2 (c)). The scattered light is not impervious to these two processes but the background light is affected to a much higher extent and this difference creates the phase contrast effect. Positive phase contrast is achieved in the same manner with a phase shift of +90° causing the background light to be 180° out of phase in relation to the scattered light. This results in the scattered light being subtracted from the background light and the resulting image foreground to be darker than the background.

Fluorescence microscopy offers the ability to localize and quantify specific molecules in live cells via fluorescent staining (see *"Image Analysis Applications: Cellular and Intracellular Object Labelling and Tracking Techniques"* for application examples). The fundamental process behind fluorescence microscopy is the absorption and subsequent re-radiation of light, or photoluminescence. Sir George G. Stokes initially made the observation that the mineral fluor-spar exhibited *"dispersive refexion"* when illuminated by ultraviolet light and coined the term "fluorescence" [6]. He observed that the fluorescing wavelengths are longer than those of the excitation source, a phenomenon that is now referred to as Stokes shift [7]. Figure 1.3 shows the basic principle in diagrammatic form.

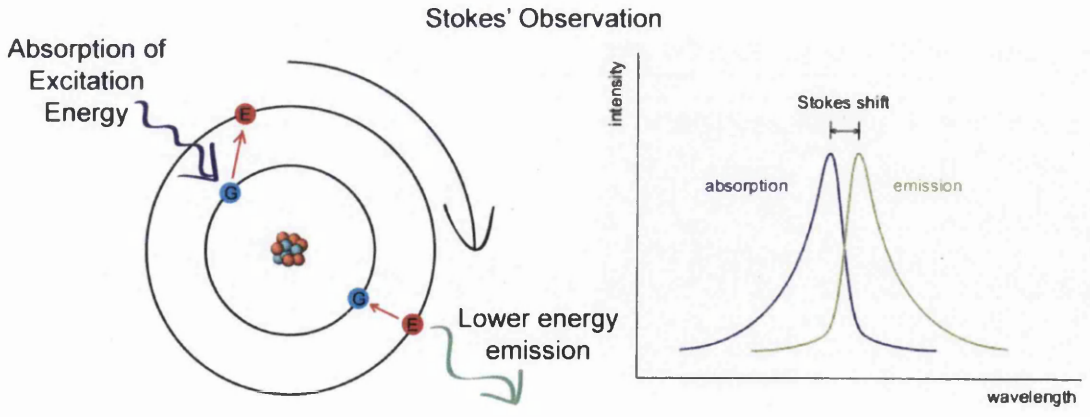



Figure 1.3: Schematic of Stokes' observation. The blue electrons represent the ground state and the red a higher energy excitation state.

Excitation photons interact with an electron, exciting and elevating it to a higher level energy state. Almost instantaneously, the electron returns to the lower level and the expended energy takes the form of a photon with a lower energy than that of the excitation photon. The epifluorescence design shown schematically in figure 1.4 is the most common configuration for fluorescence microscopy in biological applications.

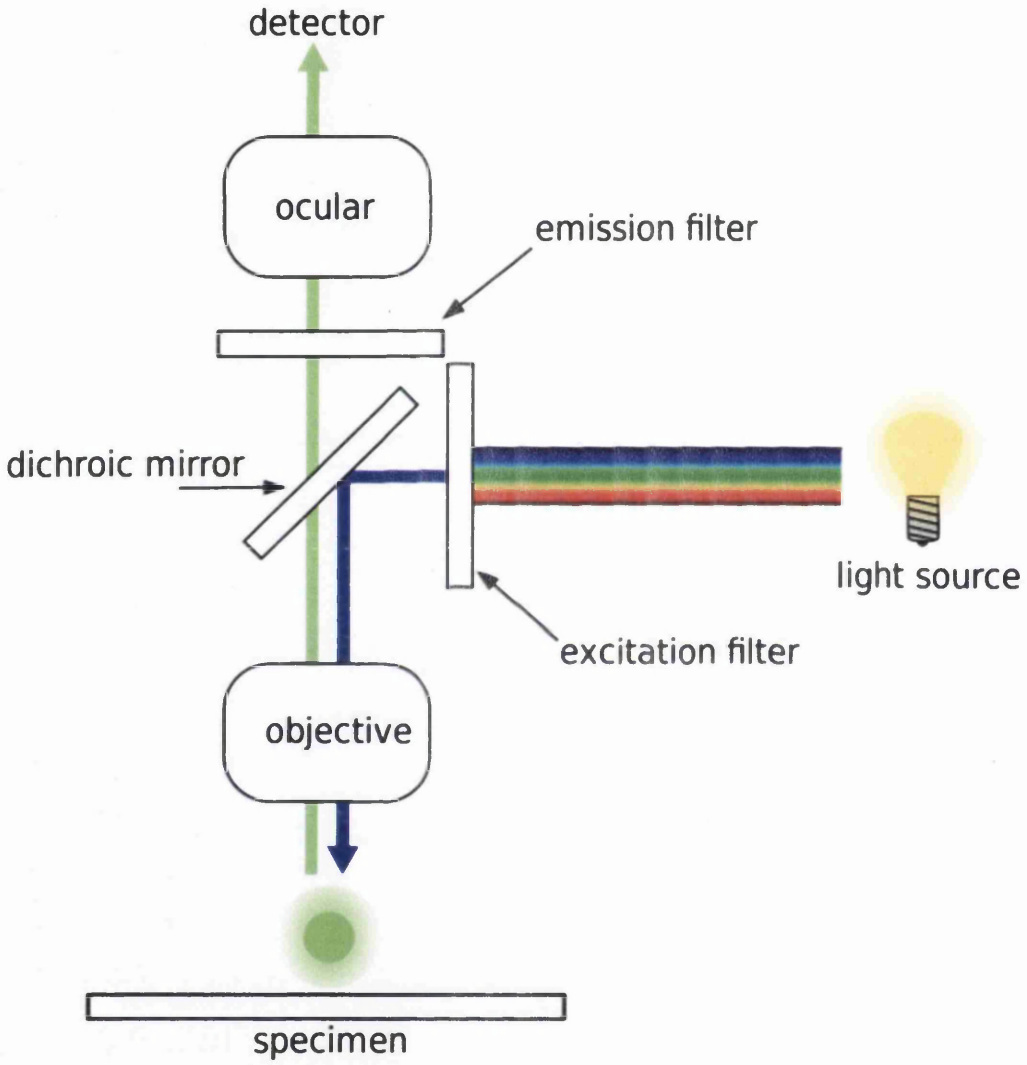



Figure 1.4: Schematic layout of a typical fluorescence microscope.
(Source: Wikimedia Commons)

Illumination light is filtered for all but the excitation wavelengths and is then focused onto the specimen through the objective lens. The same objective focuses fluorescence emissions to the detector and typically has a very high numerical aperture to maximise the sensitivity. Excitation wavelengths are often spectrally filtered between the objective lens and the detector device to further increase the

signal to noise ratio. One of the most common fluorescent sources used in fluorescence microscopy is the green fluorescent protein (GFP) [1] which contains a serine, tyrosine and glycine sequence whose side chain spontaneously cyclise to form a green fluorescing chromophore. Particular proteins can be tagged with GFP and localized using fluorescence microscopy with the GFP-protein fusion frequently retaining normal function even with large GFP polypeptides appended [8].

The basic principles of fluorescence microscopy were further developed by Marvin Minsky [9] who attempted to overcome some of the limitations of the conventional fluorescence microscope. Application of his resulting configuration is known as confocal microscopy where the use of point illumination and a spatial pinhole eliminate out-of-focus light in specimens thicker than the focal plane under observation. Figure 1.5 schematically details the optical configuration of a typical confocal microscope.

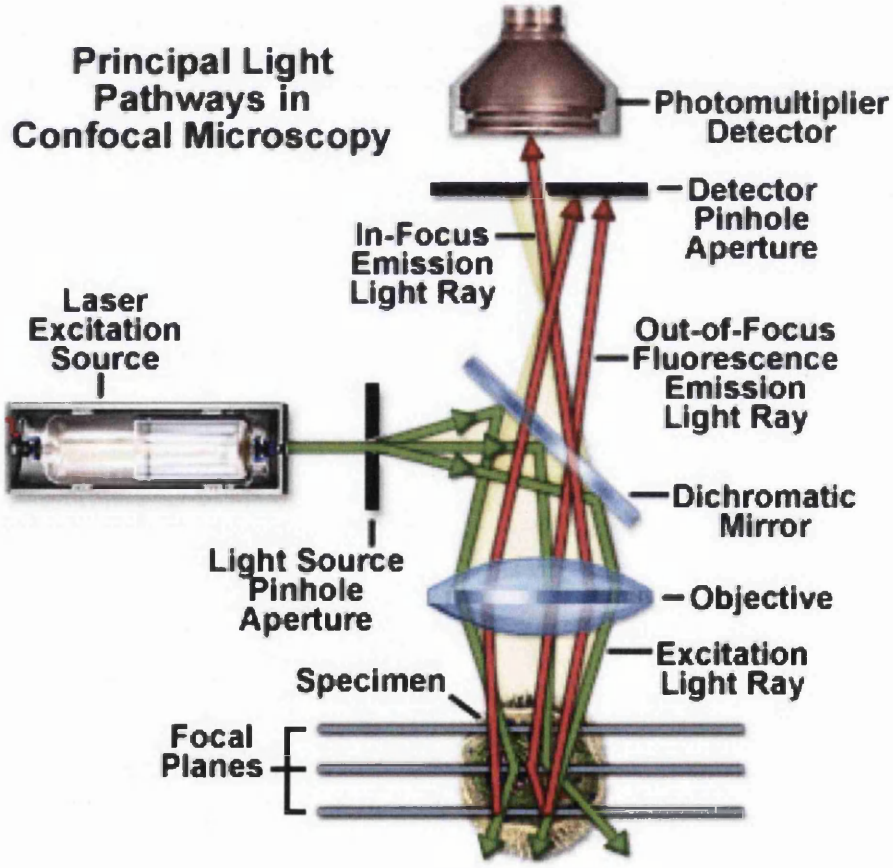



Figure 1.5: Schematic layout of the optical pathway of a typical confocal microscope. (Source: Nikon Microscopy U)

The use of a pinhole located in front of the detector blocks light not originating from the current focal plane resulting in images without blurring from structures outside said focal plane. The majority of confocal systems use lasers as an illumination source. Lasers can provide extremely well defined excitation wavelengths and their optical power levels makes them well suited to penetrating thick sample specimens. Due to singular focus, confocal systems must scan over a regular raster[1] in the sample specimen to generate 2D images. Due to the time taken to perform such scans, confocal microscopy is not well suited to imaging fast biological processes in live cell samples. However, the focal plane specificity allows confocal microscopy to generate 3-dimensional visualizations simply by defining the scanning pattern in all three dimensions and their ability to image thin optical sections makes them useful for both 3D imaging and surface profiling. Brad Amos enhanced the basic confocal microscope concept utilising laser, computer and microelectronic technologies to image biological structures [10]. Further developments were made by C.J Sheppard and Min Gu to compensate for spherical aberrations by altering the effective tube length at which the objective is operated [11]. Immunofluorescence techniques are used to increase the applications of conventional fluorescence microscopy and its variants by employing engineered antibodies linked to fluorochromes. Common fluorochromes include rhodamine and Texas Red which emit red light, Cy3 which emits orange light and fluorescein which emits green light. These fluorochromes can be coupled to purified antibodies specific to almost any macromolecule of interest. The resulting molecular targeting can then be visualized via fluorescence microscopy systems. The technique allows for multiple dyes that fluoresce at different wavelengths to be targeted to multiple proteins as well as DNA which can then be simultaneously visualized within the same cell. (see *"Image Analysis Applications: Cellular and Intracellular Object Labelling and Tracking Techniques"* for examples)

Modern microscopy systems are often computer controlled where possible allowing microscopy setups to automatically perform acquisitions after initial user configuration. Unsupervised acquisition over long periods of time offers the potential for extremely large data sets; consequently these configurations can be setup as high-throughput systems generating vast arrays of 2D and 3D imaging sequences.

[1] A rectangular pattern of parallel scanning lines; originally referred to scanning pattern of electron bean in television screens and computer monitors (Source: Oxford English Dictionary)

## *Image Analysis: Overview of Common Techniques*

Analysis of the output images acquired by the hardware mentioned above was once a monotonous and time consuming task that was performed visually by a manual operator, namely a human being. As the hardware employed more and more automation the task of manually reviewing and analysing the image yields became more and more impractical. This need for computational intervention to lessen the burden on manual operators led to the application of image analysis procedures employed in the field of computer vision to attempt to replace the manual analyser with a virtual one. Resulting software often utilises image augmentation to assist manual users and speed up analysis processes as well as removing manual intervention from the process all together.

The fundamental goal in analysing images is the partitioning of said image into sets of pixels representing objects within the image. The human brain is exceptionally quick and versatile at identifying patterns and grouping features into larger objects but manual interfacing currently prohibits a similarly quick transfer of visualised objects back onto the digital image. For example, after minimal exposure to brightfield images for familiarisation, when shown a brightfield image containing a single cell, most people would be capable of accurately identifying the cell membrane perimeter almost instantaneously. The brain essentially calls on years of experience in image interpretation to isolate the cell by recognising that the pixels within the cell membrane create an object with different visual properties to those of the background media. Such is the versatility of the human neural network that it can accomplish this task even if the image is 'dirty'[2] and of poor contrast. To input that perimeter back into the computer using a typical interface device (a mouse) could take minutes depending on the user's prowess with a mouse. When trying to identify multiple cells per field for multiple fields across multiple time points that interfacing time rapidly makes manual analysis impractical. With systems such as the IN Cell analyzer (GE Healthcare, Albany, N.Y.) readily capturing hundreds of image fields every 15 minutes over multiple days, manually analyzing image data becomes implausible. The equivalent computational partitioning of digital imaging is known in the field of computer vision as image segmentation. Image segmentation aims to

[2] Here 'dirty' refers to lens debris that can cause errors when attempting cell identification automatically

extract objects or object boundaries to create a simpler representation of the original image on which to perform various analysis procedures. A few of the fundamental methods employed by cell segmentation software are outlined below (for further information see bibliography).

Image thresholding is the most basic of segmentation mechanisms setting a threshold value to transform a greyscale image into a binary representation. For a matrix M, of dimensions equal to the image I, each element in M is set to zero if its corresponding element in I is below the threshold and one if the corresponding element in I is above the threshold, or vice versa depending on the objects of interest. The simple thresholding method can also be adapted to include multiple threshold boundaries to generate discrete representations of the original image by grouping pixels into intensity ranges.

Another commonly used method of image partitioning employs clustering algorithms and the most commonly used iterative technique is that of K-means clustering and closely related variants. First described by J. MacQueen [12], the basic steps can be outlined as follows: (Steps correspond to panels in figure 1.6)

1) Select K cluster seed points based on some heuristic measure or randomly generate.
2) For each pixel in an image I, assign the cluster centre (also referred to as a centroid) that minimizes the pixel-centroid distance. This distance could referrer to a Euclidean measure, difference in pixel intensity, colour or a weighted combination of a variety of measures applicable to the system being portioned.
3) Re-calculate the centroid positions by taking the average of all the pixels in the cluster.
4) Iterate steps 2 and 3 until a suitable level of convergence is achieved. (This could be defined as the point where no pixels change clusters or for very large sets, could be defined as the point that all centroid variations are below a defined threshold value)

By definition, this algorithm will converge but it is not guaranteed to converge on the optimal solution. The convergent solution is highly dependent on the selected seed point. Accurate knowledge of 'the right number of clusters' without a priori

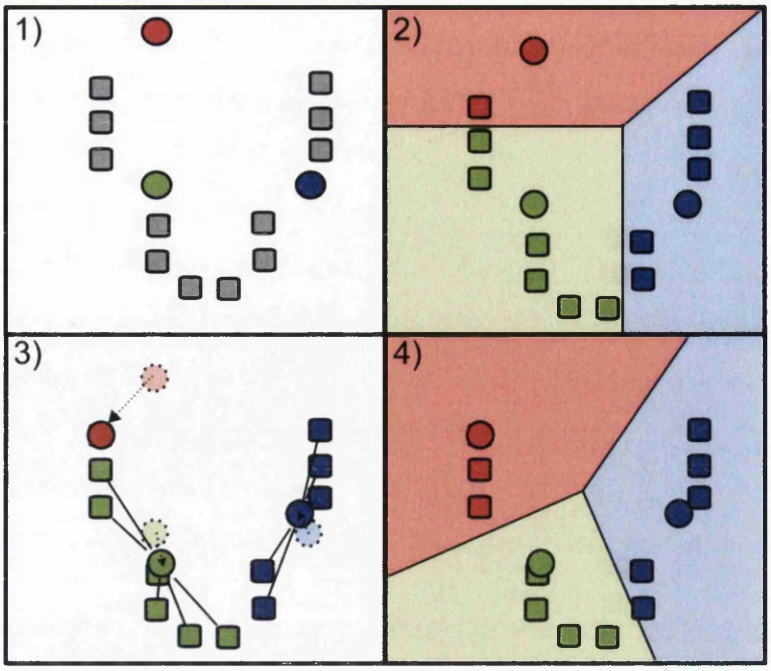



Figure 1.6: Schematic diagram depicting the principles of K-means clustering. (Adapted from Wikimedia Commons)

knowledge can be an extremely complex problem. Automated solutions have been proposed in the field of computer science for use on data structures, [13], and in the field of informatics for use in data mining of gene expression data and chromosome analysis, [14]. All such methods take advantage of implicit traits of the data structure itself, no such exploits are currently known for the automatic identification of seed points in image matrices; however, techniques in biological imaging can provide cell seed points by means of cell or cell nucleus staining with fluorophores, the signals of which are easily identified by basic thresholding in the fluorescence channel. A report by P. Rousseeuw, [15], details a method for comparing the validity of cluster algorithm solutions calculated with different initial seed points. The silhouette method can be used to optimise the k-means seed points by performing multiple k-means clustering iterations with multiple starting numbers of seed points. The best cluster initialization creates the lowest measure of dissimilarity. This process is not impervious to error as, theoretically, an optimal cluster fit is one that has the same number of clusters as data points. However, if the target number of clusters is well below the number of data points to be clustered, imparting sensible ranges of possible seed numbers avoid the possibility of converging to such a result.

Another method widely employed for cell segmentation is the region-growing method. This requires a seed input for each of the objects targeted for segmentation. Regions are grown around these seed locations by comparison of all the unallocated neighbouring pixels to the segmentation regions with the difference in pixel intensity being used as the measure of similarity. An iteration sees the pixel with the lowest intensity difference assigned to the respective region. Each successive iteration continues in this fashion until all pixels are assigned to a region. Region-growing methods are well suited to segmentation of confluent cells and tissue samples and variations with differing similarity measures can be utilised depending on requirements.

The watershed transformation [16] interprets greyscale pixel intensities as altitudes in the image topography. Water placed on pixels enclosed within a watershed line will flow downhill towards a common local intensity minimum and all pixels from which water drains to such a minimum form a topological catch basin. These basins represent segments. The algorithm is best visualised as a given topological relief being flooded via sources situated in the regional minima. The flood level remains uniform across the topology and increases homogeneously until the point at which flooding from two distinct catchment basins start to merge. At this point a division or 'dam' is erected to prevent mixing of the two flood catchments. The union of all resultant dams constitute watershed lines, the boundaries of the segmented regions.

All of these methods and numerous others are employed by software systems to augment and automate the identification process of imaged cells. Programs such as ImageJ[3] provide a platform for viewing and manually segmenting microscopy images and community development of the code has led to a large library of additional software plugins. These plugins take steps towards automating the analysis process by employing segmentation algorithms to automatically identify objects based on user defined parameters. CellProfiler[4] [17] brings computer vision techniques to the hands of biologists to *"quantitatively measure phenotypes from thousands of images automatically"*. The sister software, CellProfiler Analyst, advances the principle further by employing machine learning tools for the

[3] ImageJ: An open-source image processing program developed by the National Institutes of Health
[4] CellProfiler: Another open-source software developed by Broad Institute with an emphasis on bringing the power of computer vision protocols to biologists with no prior training in programming or computer vision.

identification of complex and subtle phenotypes. Machine learning tools make use of systems that can learn from data sets to improve their own criterion for performing tasks; a biological example would be phenotype recognition. Methods for construction of such systems are too numerous to review here and their application to biological analysis is still in its infancy. Hardware manufacturers often supply native analysis software to partner there devices typically providing analysis tools that operate on user inputted parameters. IN Cell analyzer software for example provides tools for image restoration to maximise quantitative accuracy as well as segmentation options for identifying and tracking cells.

Whilst these types of software tools aid in the analysis procedure they typically do not remove the manual element completely. Despite continued developments in the field of computer vision the ability of a manual user to learn and adapt to the identification of such a wide range of complex shapes and patterns is still very much the bench mark for automated systems. Software operating in the machine learning paradigm is beginning to show potential in bridging this gap but for now, selection of parameters for segmentation techniques and reviewing of learner sets in machine learning still requires some human input to retain practicality.

## *Cellular and Intracellular Object Labelling and Tracking Techniques*

The previously mentioned microscopy techniques (*"Microscopy Overview"*) have been utilised in biological image analysis both in their above detailed form and alongside other techniques in more complex variants. The task of automatically segmenting pixels into their parent objects such as cells, nuclei and larger organisms is very challenging in images that typically contain small intensity ranges and few distinguishing features that are extractable by automated means. Brightfield images, whilst being the simplest to acquire images in optical microscopy, yield the most complications for segmentation processes. The transparent nature of cells makes distinguishing them from the background a complex computational task; the human brain utilises too much knowledge when performing image segmentation for current computational systems to mimic. A report by Tse *et al* [18] utilises a combination of

watershed and level set method segmentation applied to brightfield cell images. They comment on the difficulties encountered with low foreground/background contrast and an optical "halo" artefact that is common in brightfield optics. They demonstrate accurate segmentation on single cell images with comparatively high contrast for brightfield images. Tscherepanow *et al* [19] present an automated segmentation protocol for use on brightfield images acquired alongside fluorescence micrographs. They too note the lack of available recognition methods without the need for phase contrast microscopy or additional cell staining and their need for brightfield cell segmentation to associate targeted proteins, labelled with GFP and imaged in the fluorescent channel, to specific cells. The results demonstrate their approach to the segmentation of Spodoptera fugiperda (Sf9) cells and, as with all such routines, the automated process does on occasion miss assign regions and miss some cells altogether. However, the ability to perform such processes without manual intervention enables much larger data sets to be considered and much larger numbers of cells counterbalances the intrinsic errors of automated identification. It is worth noting, neither of these examples are necessarily portable to segmentation of other cell lines, particularly those with poor boundary definition when imaged through brightfield microscopy. Personal experience with brightfield images has demonstrated the cell lines U2OS and A549 often have ill-defined shapes and the available contrast does not always allow for a manual user to accurately identify the cell boundary. Reports of robust segmentation procedures applied to brightfield images do seem to be typically demonstrated operating on somewhat idealised images with optimised contrast levels and/or cell lines with well defined shapes and edges (e.g. T-Cell and stem cells).

Automated segmentation is typically a less complex task when performed on images generated from phase contrast microscopy. The previously mentioned ImageJ and CellProfiler software have become consistently accurate enough in cell identification as to allow for the creation of add-ons to perform cell tracking. Li *et al* [20] propose a cell population tracker based on phase contrast images of MG-63 fibroblast cells. They demonstrated the capacity of their system to construct lineages with reference to both the temporal and spatial planes and cell identification accuracies of ~90%. More recently, Möller *et al* [21] detailed a framework for automated cell tracking for phase contrast microscopy videos. They aimed to overcome problems associated

with intensity variations of observed cells over long time periods where the cell grey-level intensity can vary above and below the background level. They employ normal velocity calculations to predict the location of cells in successive frames and topology preservation using volume constraints in an attempt to resolve problems with under- and over-segmentation. These two scenarios respectively refer to instances where adjacent cells are incorrectly connected and a low contrast cell gets split into multiple parts. Their results show an approximate deviation of 13% between automatic and manual segmentation in terms of overall percentage error with a ~2% deviation in cell centroid position between the two approaches. Such accuracies in automated tracking demonstrate the superior image quality of phase contrast over brightfield for identification of cell objects.

The array of difficulties encountered when segmenting direct-light optics images adds further appeal to the use of fluorescence microscopy where imaged fluorescence sources are often able to be accurately segmented using simple threshold parameters. Cell staining techniques enable cells and cellular molecules to be visualised through the fluorescence modality and many stains are available. Those in common use include DAPI, a fluorescent nuclear stain excited by ultraviolet (UV) light that fluoresces blue when bound to DNA, Rhodamine, a protein-specific fluorescent stain and Safranin, known for colouring cell nuclei red. Hoechst stains, of which two types are used to stain DNA in living cells as well as the Cyanine family of synthetic dyes (Cy2, Cy3, Cy5, Cy7), Texas Red and the DRAQ stain series are also available for DNA staining. The advantages of easy recognition of these fluorescent sources comes at the cost of the potential for phototoxic effects through photobleaching, a process which also places limitations on the potential duration of useful timelapse sequences. The relative simplicity of segmenting fluorescent objects offers the versatility to track fluorescent objects in 3-dimensions using confocal microscopy. Dufour *et al* [22] dynamically track multiple fluorescence sources across 3-dimensions including tracking fluorescent amoebae as they come into contact whilst preserving the identity of each of the adjacent objects through the use of coupled active surface calculations. The group also demonstrates 3-dimensional recreations of the stained nuclei of proliferating epithelial-like MDCK cells, again maintaining individual object identity through collision events. The field of immunofluorescence microscopy takes fluorescence microscopy a step further by targeting specific

biomolecules. Exploiting the specificity of antibodies to their respective antigen, fluorescent dyes can be targeted to molecules enabling the visualization of their locations throughout a sample. Primary immunofluorescence uses a single antibody chemically linked to a fluorophore. This leads to less stages in the staining protocol making the process faster and also avoids issues with cross-reactivity of antibodies. It is however typically less sensitive than secondary immunofluorescence due to limitations on the number of fluorophores that can be bound to a single antibody. Secondary techniques can bind multiple secondary antibodies to an initial primary antibody which amplifies the signal by increasing the number of fluorophores per antigen. Whilst the protocol is more complex than that of the primary technique it offers greater versatility due to the variety of secondary antibodies and detection methods available for use on a single primary antibody. The immunofluorescence technique has been utilised in many experiments. Ghosh *et al* utilised the technique to verify the presence or lack of low density lipoprotein receptor molecules (LDL-R) on the surface of human skin fibroblast cell lines [23]. An early report by Murphy [24] applies K-means cluster analysis to three colour immunofluorescent flow cytometry data demonstrating the potential of immunofluorescence to provide multi-fluorescence source data in multiple acquisition modalities. Many other experiments have been reported utilising immunofluorescence microscopy to visually isolate targeted proteins and investigate molecular dynamics within the intracellular region [25][26][27].

Another up and coming cell labelling technique is that of cellular barcoding in which individual cells are labelled in such a way as to allow them to be distinguished from cells of the same type. Such mechanisms help to simplify the tracking process and in some cases, for example where the dynamics between key time points are of no specific interest, barcoding all but removes the need to conventionally track cells altogether. An example of cellular baroding is reported by Schepers *et al* [28] where T cells are labelled with *"unique identifiers coupled to a microarray- based detection system to analyze family relationships between the progeny of such cells"*. The barcode method employed by this group utilises a retroviral plasmid library constructed to contain a GFP marker gene and a semi-random stretch of DNA, the DNA stretch being their unique barcode. Kinship analysis can be performed on populations of cells having been allowed to proliferate and differentiate. The results

allow each cell to be traced to their respective progenitor and enable the identification of functionally distinct subsets. Viral genetic barcoding has also been demonstrated by Lu *et al* [29] through in vivo studies of the heterogeneity of stem cells and barcoded vector libraries have also been demonstrated [30].

## *Core-Shell Quantum Dots – An Overview*

Quantum dots were first discovered in the early 1980s [31] and over the decades have been studied in multiple fields for their prospective uses in quantum computing, LEDs, diode lasers, and particularly biological imaging, the latter being the focus of this text. Quantum dots are semiconductor objects whose size and shape dictate their electronic characteristics and their resultant observed behaviour. In accordance with the Pauli exclusion principle[5], electrons within a semiconductor crystal lattice cannot occupy the same energy level and as such are squeezed together resulting in their being quantum confined. The energy levels of this formation are best described using an infinite potential well model[6]. Consequently, the quantum dot energy levels are quantized and directly related to the size of the quantum dot particle itself [32]. Quantum dot fluorescence is achieved through excitation of the valence electron with a well defined excitation energy/wavelength creating an electron-hole pair known as an exciton. As the excited electron returns to the ground state and recombines with the hole it expels energy in the form of photons at a lower energy than that of the excitation radiation. Due to the effects of quantum confinement, the energy/wavelength of the emitted photon is determined by the quantum dot size. This process is characterized by a relatively long lifetime (>10ns) [33] and results in emission spectra with a narrow, symmetric energy band in contrast to the short lifetimes and red-tailed emission spectra observed with most fluorophores.

Biocompatibility of quantum dot probes requires that nanoparticles present limited toxicological effects to biological systems and no large scale interruptions to normal

[5] In quantum mechanics, the Pauli exclusion principle states that no two identical fermions may occupy the same quantum state simultaneously.

[6] The infinite potential well model describes a free moving particle within a small space surrounded by impenetrable barriers. As the space narrows into the nano-scale regime, quantum effects have a greater importance and the particles may only occupy specific energy levels governed by its respective wave function.

biological processes within cells. The inherent toxicity of quantum dot probes has been investigated and the main potential source of toxicity has been related directly to oxidation of the core/shell material used to construct the nanoparticle. Cadmium is the most widely utilised element in quantum dot cores and cadmium ions have been shown to bind to thiol groups on functionally critical molecules within mitochondria and cause sufficient damage to induce significant cell death [34][35]. Studies indicate that the oxidation of nanoparticle surfaces, whether induced by exposure to air solubilisation or driven by ultra-violet (UV) light interaction, exposes free cadmium [36][37]. Exposing the nanoparticles to air prior to solubilisation or moderate to extended exposure to UV light post incubation causes enough of an increase in the amount of cadmium within the cells to cause observable cell death. Further studies demonstrate that cells labelled with quantum dots synthesized within stable and inert conditions exhibit no indication of nanoparticle induced toxic effects. Surface oxidation can also occur inside oxidative solutions, a concern as in vivo environments can potentially contain such solutions. A study by Kirchner *et al* [38] reports the use of CdSe/ZnS core-shell quantum nanoparticles in a mercaptoacetic acid solution. They show $Cd^{2+}$ release into cells to toxic levels over a 48 hour time period under standard quantum dot synthesis and microscopy imaging conditions. Examination of the free core materials in solution resulted in the determination of cadmium as the primary cause of cytotoxicity and such levels could be reduced and even eliminated through the addition of surface coatings.

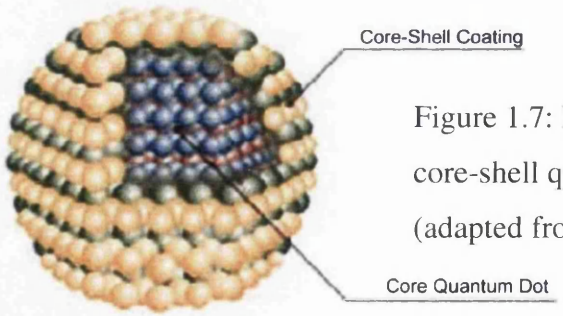



Figure 1.7: Diagram showing the basic core-shell quantum dot construction (adapted from photonics.com) [58]

The core-shell semiconductor quantum dots (qdots) utilised in later chapters of this text consist of a core made of a cadmium telluride (CdTe) with an outer surface coating of Zinc Sulphide (ZnS) to stabilize the nanoparticle (see figure 1.7 for a schematic overview of this structure). The outer ZnS shell electronically isolates the particle creating a passive surface state and also acts to prevent oxidization and other

in vivo environmental factors that contribute to cadmium release. Zinc and sulphur bind with a high level of affinity and form a polymer-like structure encapsulating the quantum dot core, a formation that is much less susceptible to oxidation processes than the core it surrounds [37],[39]. Such a shell structure is shown to nearly eliminate oxidization due to pre-solubilisation exposure to air as shell oxidization is only observed during extended exposure to UV light. Without disruption to the ZnS shell structure the surrounding environment is unable to interact with cadmium in the core and as such an undisrupted core-shell quantum dot is not seen to be toxic to in vivo systems. The additional layer of semiconductor also increases photoluminescence of the quantum dot particle [37],[39].

### *Quantum Dots as Optical Trackers*

QDots of this type offer a potential alternative fluorescence source to those mentioned under previous headings that is both photo and bio-stable for use in fluorescence microscopy imaging. With diameters typically of a few nanometers, the precise size and shape of these nanoparticles can be controlled during the synthesis process by controlling temperature and duration and through the use of differing ligand molecules[33]. Control at these stages enables the creation of qdots with compositional- and size-dependant absorption and emission characteristics. Absorption of a photon increases in probability at higher energies (shorter wavelengths) resulting in a broadband absorption spectrum which is markedly different to that of standard fluorophores.

This tuneable emission combined with photostability makes quantum dot nanoparticles an attractive alternative to organic dyes for use as a fluorescent label in biological imaging. One particular area where this stability is sought after is three dimensional optical sectioning where photobleaching of fluorophores during the accruement of optical z-sections can often compromise the accurate reconstruction of 3D structures imaged. Quantum dots have also been successfully applied to immunofluorescence labelling of fixed cells and tissue samples; more specifically immune-staining of membranous proteins ([40],[41]) and actin, microtubules and nuclear antigens ([42],[43]). Immunofluorescence techniques are capable of utilising

the fact that only a very small number of quantum dots are necessary to provide a detectable signal. Studies have commented on the observation of fluorescence signal fulguration, a phenomenon attributable to quantum dot blinking [44],[45]. Such observations demonstrate the potential to identify single quantum dots within immunocytological conditions and the possibility of a sensitivity limit of one quantum dot per target molecule.

In what is often described as a landmark study in the field of drug delivery, Tada *et al* detail the delivery mechanism of quantum dots into human breast cancer cells [46]. Using a high-speed confocal microscope with a high-sensitivity camera they were able to track a single quantum dot particle conjugated with a tumour targeting antibody in tumours of living mice. The group implemented a cell line whose membrane over-expresses the human epidermal growth factor receptor 2 (HER2) alongside a conjugate of quantum dots to anti-HER2 monoclonal antibodies, namely Trastuzumab. This enabled the visualization of the nanoparticle complex within the blood vessels feeding the tumour cells. These quantum dot fluorescence signals helped the group to determine characteristics such as direction and velocity for the six identified stages of the delivery process:

- Blood vessel circulation
- Extravasation
- Migration in the extracellular region
- HER2 binding on the cell membrane
- Internal migration from cell membrane to perinuclear region
- The perinuclear region itself

The tracking process revealed motions of the quantum dot complex that manifested random orientation and speed, exhibited characteristic stop and go behaviour and ceased directional intracellular movement once around the perinuclear region with continued motion around the perinuclear region being described as Brownian. This study perfectly demonstrates the potential of quantum dot fluorescence markers in studying particle behaviour within complex biological environments.

The tuneable feature of quantum dot emission wavelengths allows for a potentially large array of well separated colours that are excitable by exposure to a single wavelength. Such characteristics can be used to simplify image acquisition and can

also be exploited in confocal microscopy to accomplish nano-scale resolution colocalization of individual quantum dots of multiple colours. Lacoste *et al* [47] utilised two unique fluorophore families in the form of fluorescent beads and semiconductor nanocrystals as sources with differing emission wavelengths that can be excited by a single excitation wavelength. Applications of quantum dots in whole-cell labelling, membrane-bound protein labelling and cytoplasmic and nuclear targeted labelling open up their use in live-cell experimentation. A report by Shroeder *et al* [48] demonstrates the use of quantum dots to successfully target specific receptors in vivo by conjugating quantum dots entrapped within a lipid shell to a nutrient of critical necessity for rapid growth and cell division. Using fluorescence microscopy, these folate-targeted lipodots[7] were shown to be taken up by the tumour cells and an observed increase in lipodot fluorescence intensity compared to non-specific quantum dots in the same cell line was suggestive of folate being the primary factor in bio-recognition and its use provided a highly specific targeting mechanism. Confocal depth scanning also showed substantial internalization which confirmed the specificity of folate-binding and folate-targeted lipodot uptake. Conjugation of quantum dots with antibodies have also been utilised to target cell adhesion molecules related to the vasculature of retina in rats [49]. Employing a polyethylene-glycol (PEG) cross-linking scheme to link antibodies, Jayagopal *et al* were able to distinguish between differing cell adhesion molecules by conjugating specific monoclonal antibodies to quantum dots. The results demonstrate the ability to perform non-invasive, in vivo microscopy imaging of the retinal vasculature with single cell level spatial resolutions. Targeted quantum dot fluorescence intensity was again shown to increase against control and non-specifically labelled quantum dots, both of which showed no evidence of localization of fluorescence intensity in or around the vascular system. Quantum dots tuned to near-infrared (near-IR) emission wavelengths have been used with particular interest towards real-time in vivo imaging due to their ability to act as a fluorescent marker at wavelengths that are minimally scattered by biological specimens. Jiang *et al* utilised biocompatible near-IR quantum dots for multiplexed imaging organs, tissues and cells. By performing deep tissue laser confocal fluorescence imaging they were able

[7] A term used by Shroeder *et al* for the result of mixing hydrophobic quantum dots with 1,2 dipalmitoyl-sn-glycero-3 phosphocholine and methoxy-polythethylene-glycol-distearoyl-phosphatidyl-ethanolamine (mPEG-DSPE)

to visualize a mouse femur injected with near-IR quantum dots. Conjugation of transferrin to quantum dots with an emission wavelength of 650 and 750nm allowed the group to acquire simultaneous fluorescence images of two distinct sections of the chest cavity of a mouse in vivo. The optical signals detected by Jiang et al were found to be at a tissue depth of up to 0.8mm without any laser power or optical gain adjustments. Further use of multiple quantum dot colours is seen in the work by Kobayashi et al [50]. Here five similarly sized quantum dots with differing, resolvable emission spectra are used in fluorescence lymphangiography to simultaneously visualize five lymphatic drainage flows in vivo. Assessing multiple separate lymph drainage basins via alternative techniques[8] requires each basin to be studied separately with appropriate intervals between each examination as they are incapable of differentiating contributions from multiple drainage basins. Current developments in quantum dot synthesis techniques are also showing further potential for quantum dots through enabling the creation of hybrid quantum dots that are capable of imaging via multiple modes, so called multimodal QDots. An example of such a particle is reported by Bakalova et al [51] where they describe:

"*A multimodal QD probe with combined fluorescent and paramagnetic properties of silica-shelled single quantum dot micelles with the paramagnetic substances tris(2,2,6,6-tetramethyl-3,5-heptanedionate)/gadolinium incorporated into the micelle and/or silica coat*".

The results of the application of their multimodal probes indicated no significant change in physiological characteristics such as heart rate and blood pressure or any prolonged effect on structure of the vascular network. Consequently, such probes look viable as tracers for brain vasculature simultaneously imaged by both fluorescence microscopy and MRI systems.

Another approach for utilising quantum dots for cell tracking exploits the endosomal pathway to deliver the nanocrystals into the cytoplasm of live cells. Reagents in cell labelling kits such as the Qtracker® range from Invitrogen employ custom targeting peptides that results in live cell uptake of quantum dots and their long term, stable internalisation encapsulated within endosomes in the cell cytoplasm [52],[53].

[8] More traditional techniques for this type of study include X-ray lymphangiography, MR lymphangiography or radionuclide radio-scintigraphy.

Endocytosed quantum dots of this type have been tracked across all 3-dimensions and through time by analysing image sequences acquired by multidimensional fluorescence microscopy [54]. The report focuses on accurate tracking and characterization of the fluorescent vesicles themselves without attempts to exploit and relate the extracted parameters to their cellular carriers but does demonstrate the potential. Internalised quantum dots have however been used to infer parameters for cell cycle dynamics and cell division modelling by Errington *et al* [55]. Using flow cytometry to quantify the internalized quantum dot fluorescence at 24 hour intervals the group were able to extract numerical values from statistical distributions to enable the creation of cell population models. This concept of inferring cellular characteristics from quantum dots internalised as intracellular fluorescent markers is explored in chapters 2-4 of this text by considering the quantum dot loaded endosomes equivalent to points in a moving-light-display. Moving light displays are now used extensively in the entertainment industry to capture the mechanical motion of the human body and robotic extensions for translation into the motions of computer generated characters. Placing highly visible strategic markers about the body to be translated allows computer systems to track and map essential motions in the captured images (see figures 1.8 and 1.9).

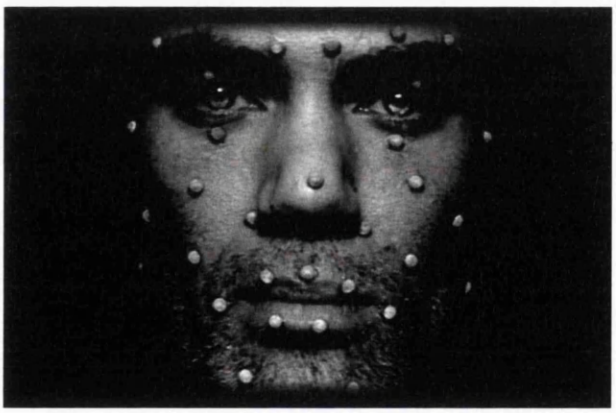


Figure 1.8: Optical markers applied to the face of an actor. This process is utilised widely to create more realistic representations of facial motion in CGI film characters and video game avatars. (Source: http://studiotenenbaum.com/a-video-game-selected-to-a-film-festival-the-silent-revolution)

Human perception of motion does not require complete knowledge of the visualized system, only key points of reference are considered during interpretation [56]. A large field of computer vision has emerged out of this very principle where the computational power is applied to the interpretation of complex input imagery through identification and processing of key points of interest. Moving light displays such as those in figure 1.8 offer easy to isolate points on which to focus tracking algorithms and the motion of the full object can be inferred by the motions of these points [57]. In a similar fashion, internalised quantum dots within cells can be

thought of as a moving light display and their motions can be interpreted to infer dynamics for their carrier cell. Such techniques could offer the ability to track and characterise cells and key dynamic properties without the requirement of accurate knowledge of the cell morphology. Removing the necessity for accurate morphology information could also offer far greater portability across multiple cell lines.

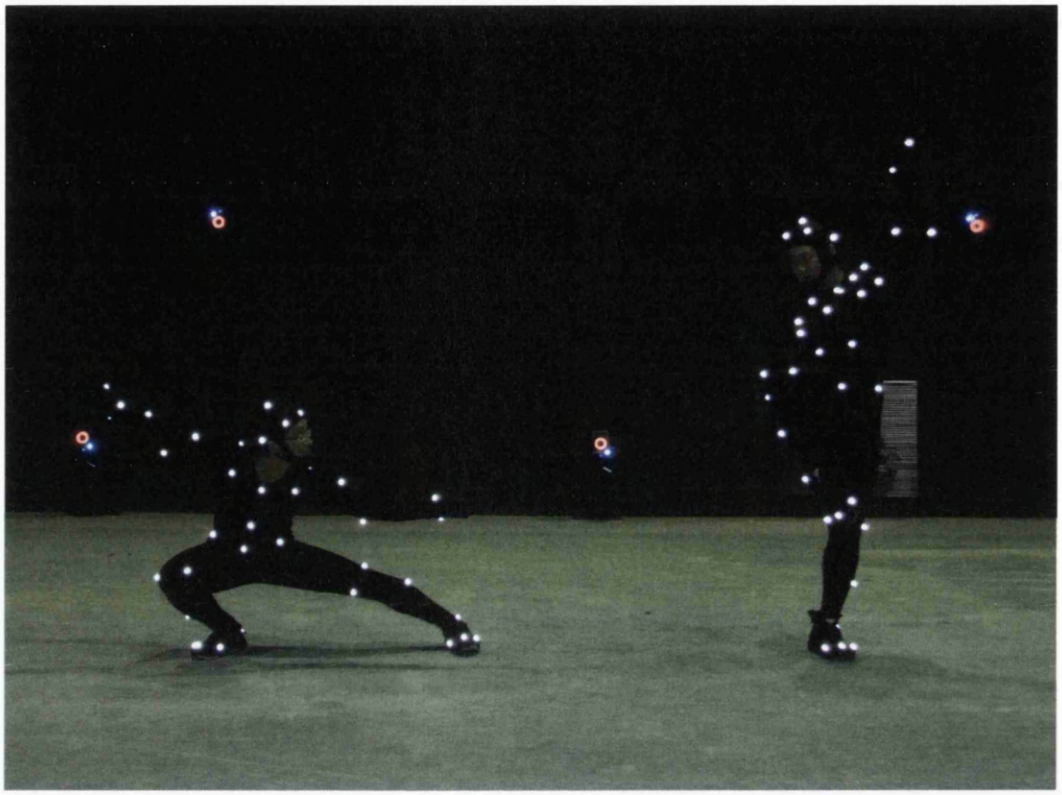


Figure 1.9: Example of full body motion capture suits typically used in the film industry. (Source: deviantart.com)

# CHAPTER TWO

# Automatic Tracking of A549 Cells

This chapter investigates the use of core shell quantum dots as potential optical markers for tracking cells within an acquired timelapse image sequence. Quantum dots internalized within A549 cells form the basis of a moving light display construct representing the field of view as groupings of digitized points. Interpretation of these groups enables the inference of cell locations without the complexity of conventional computer vision algorithms. The objective is to demonstrate the potential for tracking solely on the fluorescence channel using these bio- and photo-stable fluorescence sources.

## *The Data Set*

The data set acquired from Cardiff University consisted of brightfield and fluorescence images of cells loaded with CdTe/ZnS core-shell quantum dots. The cell line itself was A549 (ATCC CCL-185) epithelial lung carcinoma cells maintained under G418 selection in McCoy's 5a medium supplemented with 10% fetal calf serum (FCS), 1mM glutamine and antibiotics incubated at 37°C in an atmosphere of 5% $CO_2$ in air. In preparation for experiments, cells were grown at a density of $1x10^6$ cells $ml^{-1}$ as a monolayer in either coverglass bottomed chambers (Nunc, 2 Well Lab-Tek II, Fisher Scientific) or glass bottomed 24 multi-well Sensoplate, Greiner Bio-one, for 24 h prior to imaging. All cell concentrations were determined using a Coulter Particle Counter (Beckman Coulter, High Wycombe, UK). The cells were loaded with commercially available targeted nanocrystals using the Qtracker® 705 (QTracker705) Cell Labeling Kit (4 nM Invitrogen (Q25061MP)). The reagents in the Qtracker® 705 Cell Labeling Kit use a custom targeting peptide (9-arginine peptide) to deliver near-infrared-fluorescent nanocrystals into the cytoplasm of live cells via the endosomal pathway. Briefly, Qtracker reagent A and B were premixed and then incubated for 5 minutes at room temperature. 1ml of fresh full growth media was added to the tube and vortexed for

30 seconds. This labelling solution was then added to each well of the cells and incubated for 1 hour at 37°C after which they were washed twice with fresh media. 24 hours later, labelled cells were analyzed by time-lapse, confocal laser scanning microscopy over a period of a further 48 hours using a x40 0.75 NA air lens, 488nm excitation and 680-20nm emission filters to collect the Qtracker705 fluorescence every 5 minutes.

*Image pre-processing*

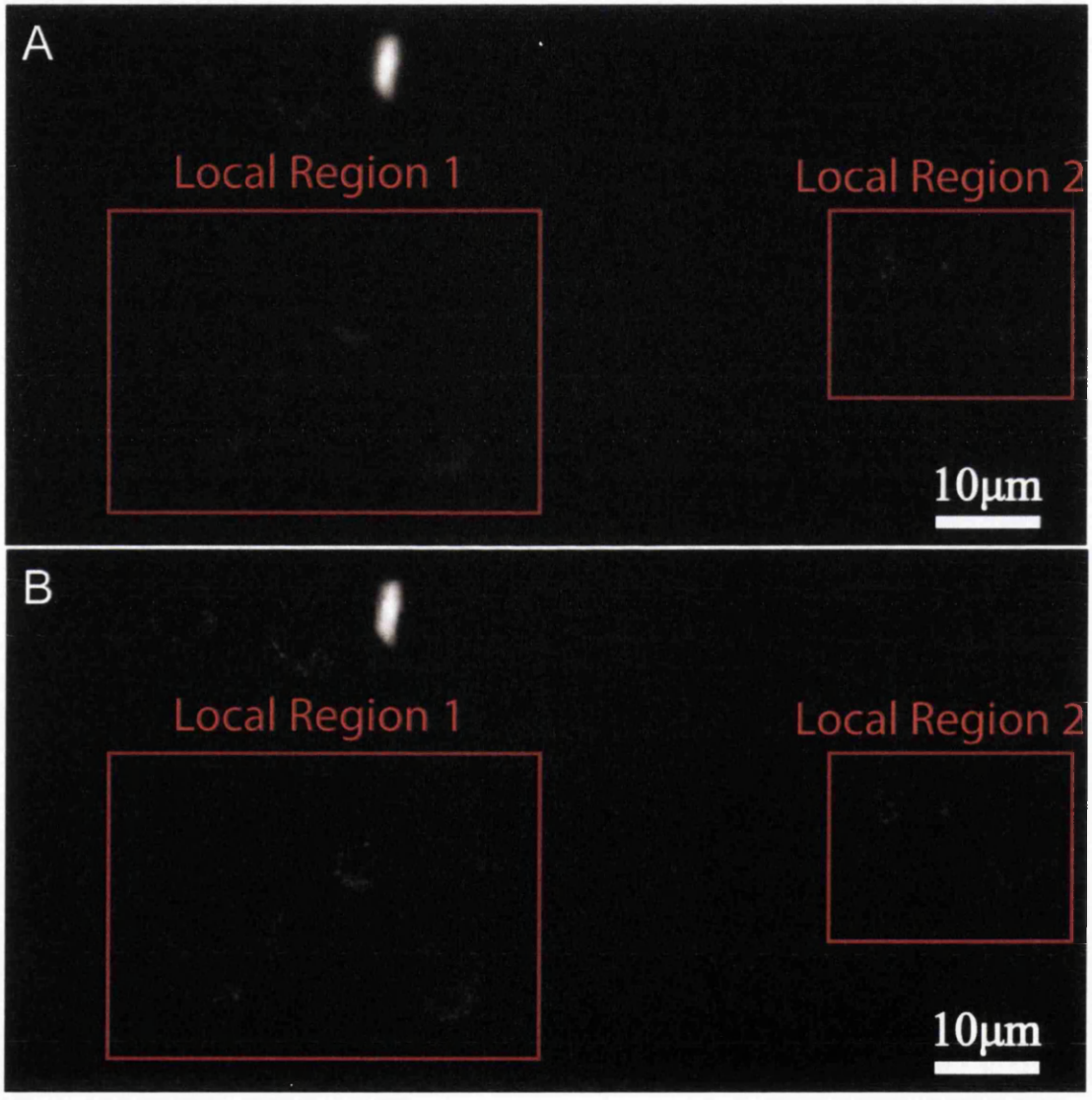



Figure 2.1: A two image sample taken at the same interval at differing focal lengths. Comparing the two highlighted regions in each image shows variation in contrast across the field of view. Image B has the highest overall contrast using an absolute gradient measure however local region 2 has higher contrast in image A.

Both the brightfield transmission and QD fluorescence (x, y, x, t) optical sections are stored in respective multi-layer tiff stacks for every 5 minute interval across the 48 hour experimental range with each file containing 8 z-dimension focal plane images separated evenly across a 5μm focal range. Autofocus protocols on the microscopy system were disabled in favour of storing multiple z-plane images to maximise the amount of raw data available for analysis. Simply selecting the image with the highest overall contrast measure can lead to loss of clarity in localized regions due to cells occupying slightly different spaces in the z-dimension (see Fig. 2.1 opposite). A simple way to alleviate this problem is to segment the image space into smaller regions, select the focal plane with the best contrast measure for each region individually and recombine these regions to create a composite image of localized best focus. The fluorescence image sets are all of dimension 1024x1344 pixels with each pixel covering 0.25μm, each pixel area equating to $0.0625\mu m^2$ making the corresponding field of view 256x336μm. This image space is segmented into an 8x8 grid of 64 regions of size 128x168 pixels and the highest contrast focal plane for each region is found by applying a simple absolute gradient algorithm of the form

$$F_{grad} = \sum_{Height} \sum_{Width} |I(x+1, y) - I(x, y)| \qquad (2.1)$$

where I(x,y) is the intensity of the pixel (x,y). For each region the focal plane with the highest value $F_{grad}$ has its corresponding z-stack number, valued from 1-8, assigned to the respective element in the 8x8 matrix **A** (see Fig. 2.2).

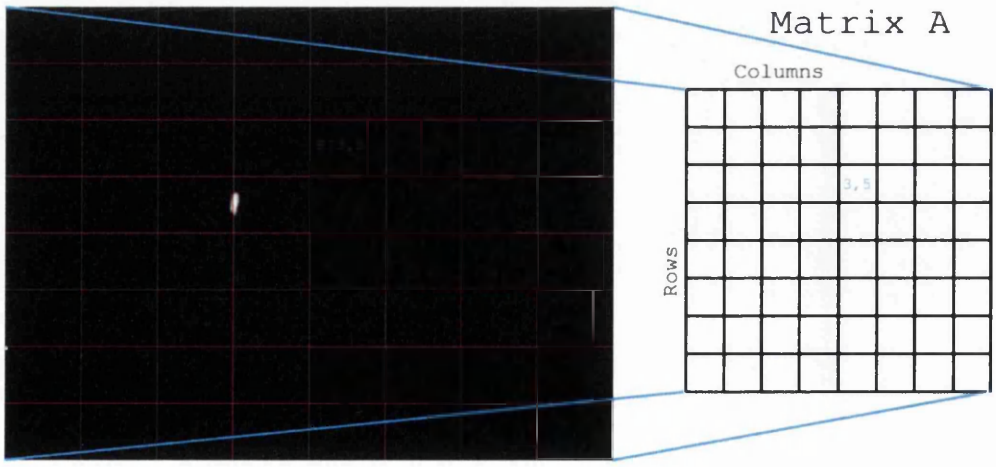



Figure 2.2: Graphic displaying simple method to reduce localised variation in contrast across the full field of view. Using the labelled grid region as an example, the absolute gradient algorithm is applied to a crop of this region on all 8 z-stack images and the index of the z-plane with highest $F_{grad}$ value is stored in element (3,5) of matrix A.

With this process repeated for all 64 grid regions the optimized full field of view is formed by recombining the regions using the z-plane defined in each respective element of the matrix A. The composite image is then created by stitching the regions back together using the z-stack plane labelled within the matrix A as the source image for each region (see Fig. 2.3). While various incarnations of de-convolution could have been employed here to try and remove the blurring effect of differing focal lengths on the internalized QD signal, the problem of applying a single filter to a field of view containing objects at varying depths still remains. With 8 focal planes available to analyse this simple segmentation method provides a computationally inexpensive way to create a full field of view image with locally optimized QD signal contrast and upholds the ethos of maintaining simplicity.

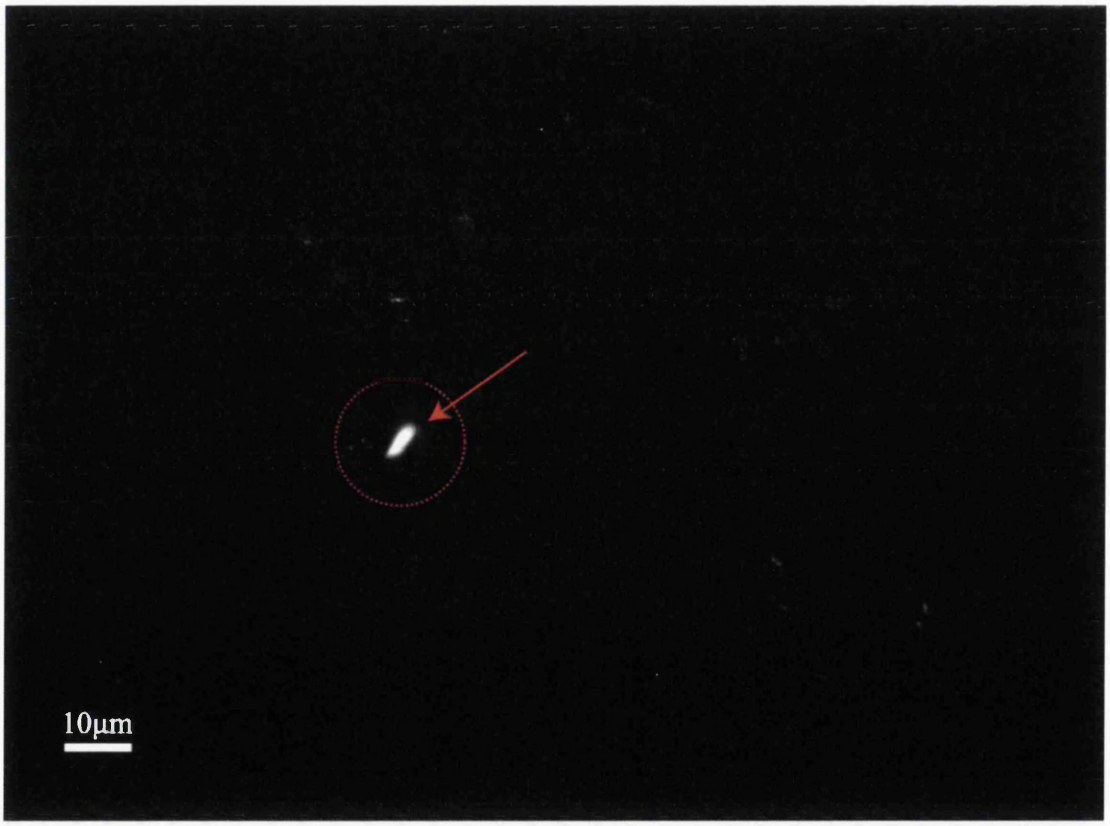



Figure 2.3: An example of a fluorescence composite image. The edge of a grid region can be identified by the linear edge of noise around the large white object circled in red.

The selection of the best focus focal plane for each region and re-stitching to form an optimal focus image from the raw data avoids the need for computationally expensive de-convolution techniques and provides sufficient contrast to identify quantum dot loaded endosomes.

With an optimized fluorescence image created, the extraction of the QD loaded endosome locations becomes a simple two stage process. Firstly, as the majority of QD markers provide a fluorescence signal that is consistently 3-4 times higher than the background level, (See Fig. 2.4 A overleaf), there is no requirement for complex image foreground/background thresholding techniques to remove image noise. The background is therefore defined as a straight forward step function of the form

$$F_s(x,y) = \begin{cases} 1 \; if \; I(x,y) \geq \theta \\ 0 \; if \; I(x,y) < \theta \end{cases} \qquad (2.2)$$

where θ is the pixel intensity threshold simply defined as

$$\theta = \mu_{image\ intensity} + \sigma_{image\ intensity}. \qquad (2.3)$$

Here μ is the average image intensity and σ the image standard deviation the addition of which sets the noise filter comfortably clear of the noise floor and its associated intensity variation whilst retaining the QD signal foreground (see Fig. 2.4). The second stage requires extracting the geometric centres of the remaining peaks, which correspond to nanoparticle loaded vesicles within the cell cytoplasm and will be used as cell markers in the tracking process. These points are identified by compiling row and column intensity profiles and cross checking their localized maxima points against each other as a 2-dimensional peak will have a localized maxima in both the x and y dimensions.

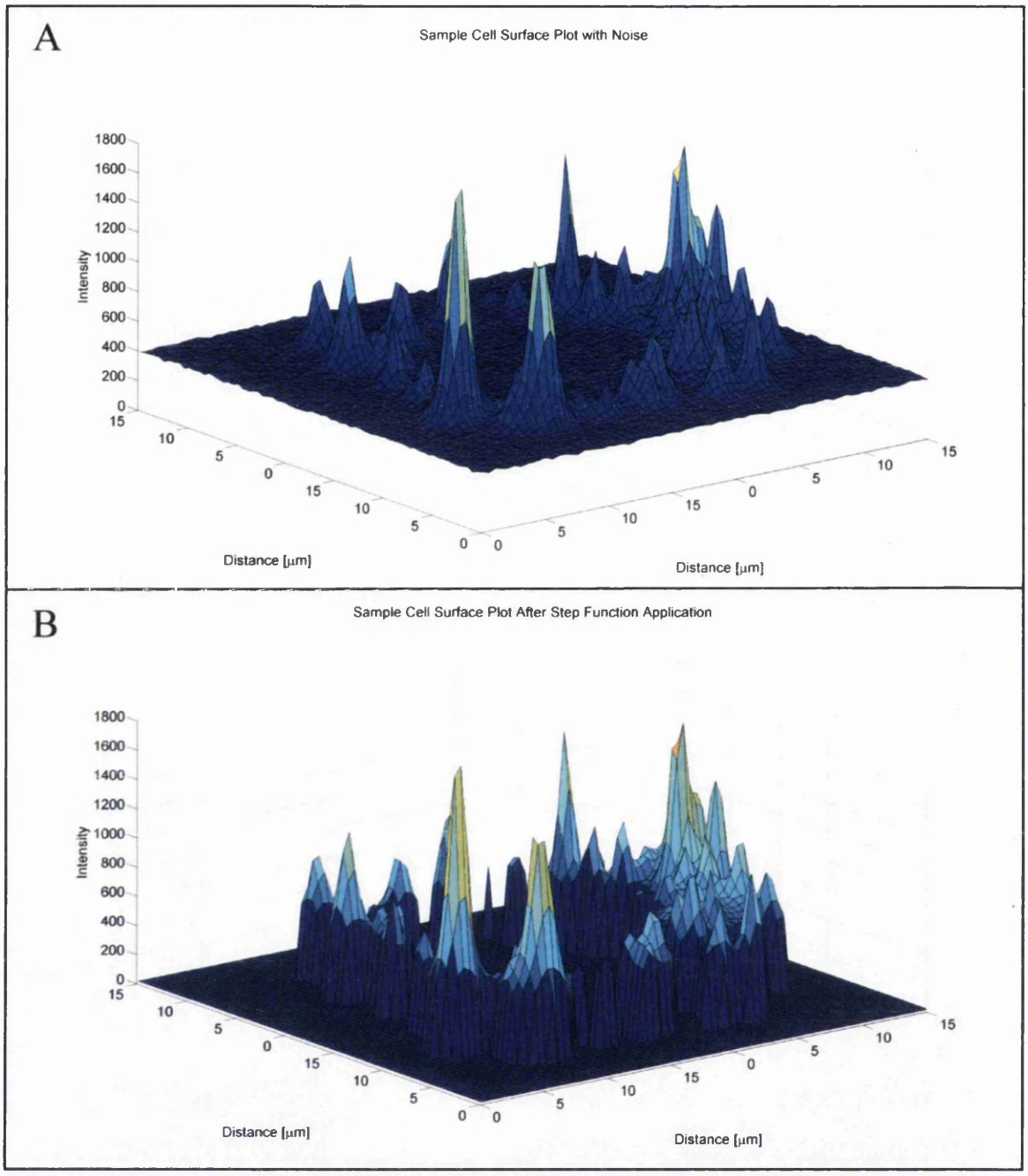



Figure 2.4: Examples of a cellular region of dimensions 15x15μm within the field of view visualized as surfaces. (A) Raw data from the image set demonstrating the clearance of the QD loaded endosome signal from the background fluorescence. (B) The same region after application of the step function to the entire field of view showing how the prominent marker signals pass through the thresholding process.

An example of the principle using 2 Gaussian peaks for ease of visualization is shown in figure 2.5 (opposite) where multiple row and column profiles in panels B and C reveal the two peaks in each dimension separately. Let a single profile, passing

through a row of elements, be labelled as the vector R, an element is a maxima under the following criteria:

$$\left.\begin{array}{c} R_i > 0 \\ R_i > R_{i-1} \\ R_i > R_{i+1} \\ (R_i - R_{i-1}) \times (R_i - R_{i+1}) \geq 0 \end{array}\right\} R_i = maxima \qquad (2.4)$$

Any identified maximal elements in R are flagged with a 1 in the respective elements of a marker matrix $M_R$, similarly maximal column elements are flagged with a 1 in the respective elements of a marker matrix $M_C$. The peak matrix, P, is then simply

$$P = M_R + M_C \qquad (2.5)$$

where peaks are elements containing a value of 2, namely the intersects of the column and row profiles.

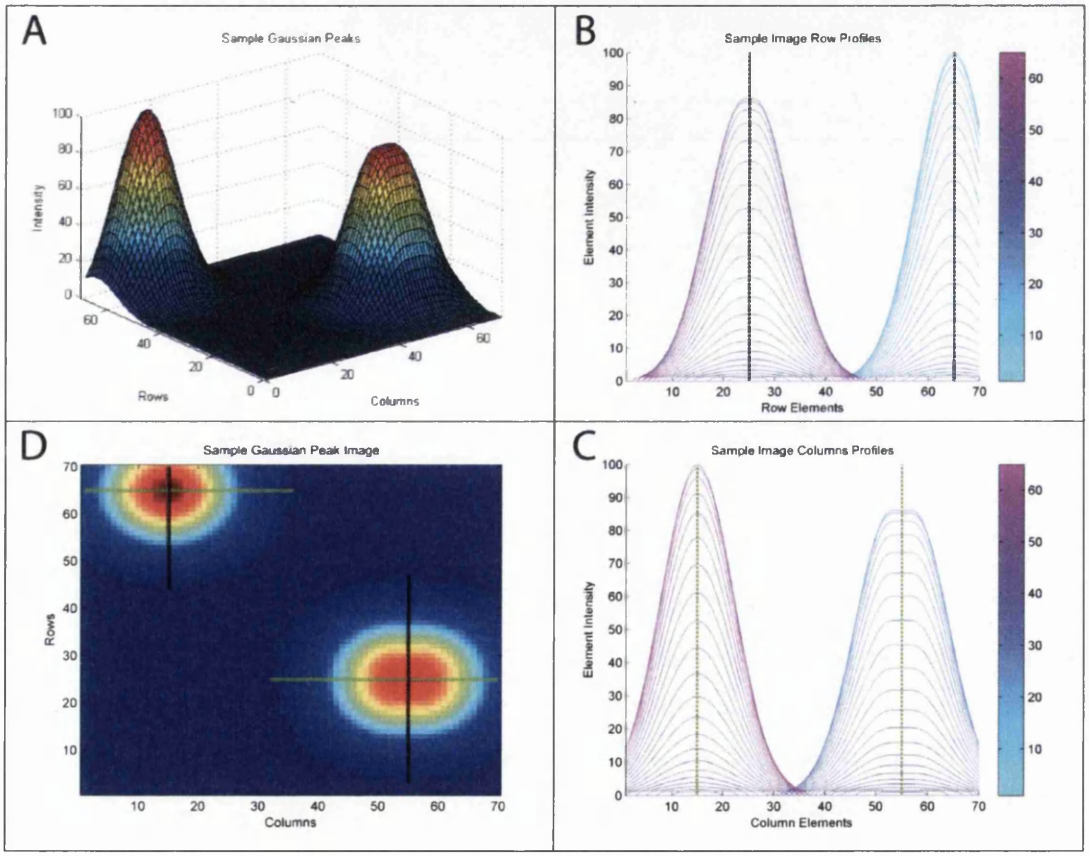



Figure 2.5: Graphic demonstrating principle of peak finding. (A) An example region containing 2 Gaussian peaks on which to demonstrate the peak finder principle. (B-C) contain row and column profiles respectively with a colour bar indicating each plotted lines respective row or column. (D) Displays the peaks in A as a top-view 2-D image.

The marker matrices $M_R$ and $M_C$ have been overlaid onto the top-view image in panel D in black and green respectively to show how the row and column profile maxima form "cross-hairs" intersecting at the peaks of features within the field of view. The peak coordinates, row and column of any element of P containing the value 2, are stored to create a digital map for each fluorescence image frame.

This simplified coordinate representation, referred to from here onward as the binary element display, can be likened to a moving light display utilised in many fields such as the film and gaming industry and sports science analytics to capture key features of motion using the minimum amount of signal input [59][60][61]. A comparison of a brightfield transmission image, a fluorescence image and their corresponding binary element display can be seen in figure 2.6.

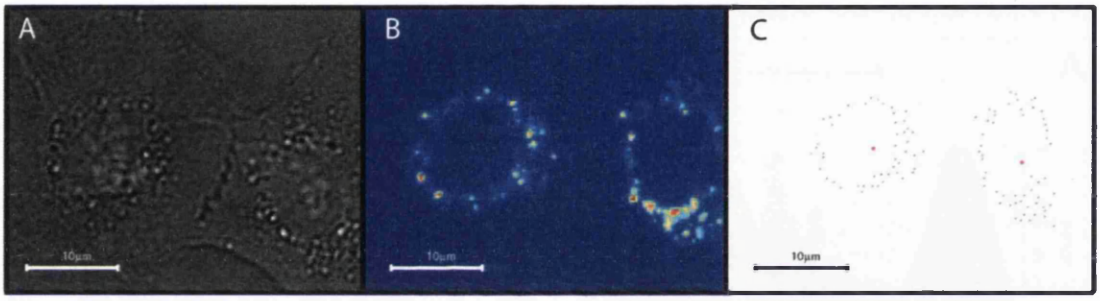



Figure 2.6: Comparison of a two cell region displayed in three paradigms for comparison; (A) brightfield transmission image, (B) quantum dot loaded endosome fluorescence, (C) binary element display.

The stored coordinate files, an example plot of which can be seen in figure 2.7 (opposite), are, in principle, all that is required to track and monitor cell motion through time, consequently, long term storage of the original brightfield and fluorescence data is no longer required. Using the moving light display concept of the binary element paradigm as a representative sub-sampling of the original image rather than fully visualising the raw data represents a major decrease in file storage requirements as key feature of interest from the original data set, ~12GB in size, can be stored as binary elements in a file three orders of magnitude smaller at ~ 8MB. Where high-throughput applications using standard acquisition can produce 1 TB of data per day this provides a data reduction method to aid in avoidance of the time and monetary costs of data analysis. Also the raw experimental information needed

to characterise and investigate cell line motility and potential effects of applied pharmaceuticals can be kept for re-investigation for much longer than is practical for the original images.

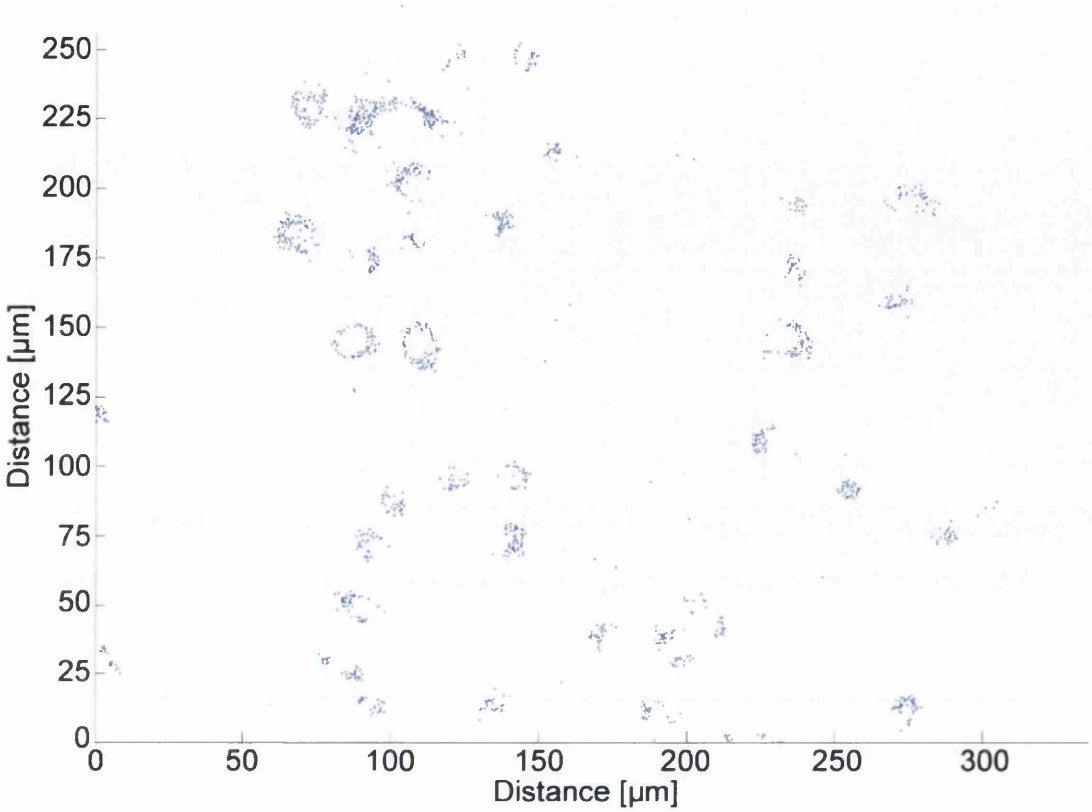



Figure 2.7: Scatter plot displaying the binary element format for the full field of view of the initial time frame and showing the visually identifiable groupings of binary elements.

## *Cluster Analysis: First Pass - Interpreting the Binary Points*

With the binary element coordinates stored the process of interpreting them can begin by identifying groups in the binary element display as these groupings correspond to fluorescent vesicles within the same cell (see Fig. 2.7). One of the simplest methods of doing this is to apply a k-means clustering algorithm to identify nearest neighbour groups. For a set of binary element coordinates $\boldsymbol{B} = \{b_1, b_2, \ldots, b_n\}$, the function

$$argmin_C \sum_{i=1}^{k} \sum_{b_j \in C_i} \|b_j - \mu_i\|^2 \qquad (2.5)$$

is iteratively applied until convergence. Here $\boldsymbol{C} = \{C_1, C_2, \ldots, C_k\}$ is the set of clusters with total cluster number k, $b_j$ is the set of binary elements, a subset of **B**, associated with the cluster $C_i$ and $\mu_i$ is the mean coordinate of the binary elements associated to $C_i$. The k-means protocol has a minimal input requirement of an initial cluster set known as seeds, either k defined cluster locations or k randomly initialised cluster locations. Automatically seeding a clustering algorithm without prior knowledge of the number of clusters and with group numbers as high as are present in this system is a notoriously complex mathematical problem with as yet no clear solution (see introduction section *"Image Analysis: Overview of Common techniques"*). To achieve the goal using QDs as cell tracking markers using simple processes and to avoid getting drawn into highly complex methodologies for attempting to blindly identify clusters, the initial seed points were identified manually using the first brightfield image to locate all cells within the field of view. While this does add a non-automated step, the task of manually identifying the initial frame in order to automatically track the following 600 frames is far quicker than both entering into the computationally expensive task of attempting to automate the seeding process and still favourable to manually tracking all 600 frames. Once the k-means function reaches convergence the cluster centres, or centroids, for the current timeframe are stored. In the successive images beyond the initial frame the cluster seeds are defined as the final cluster centroid locations from the previous timeframe. The cells are generally not motile enough to cause the previous centroid location to lie outside of the cell marker boundary over a 5 minute period.

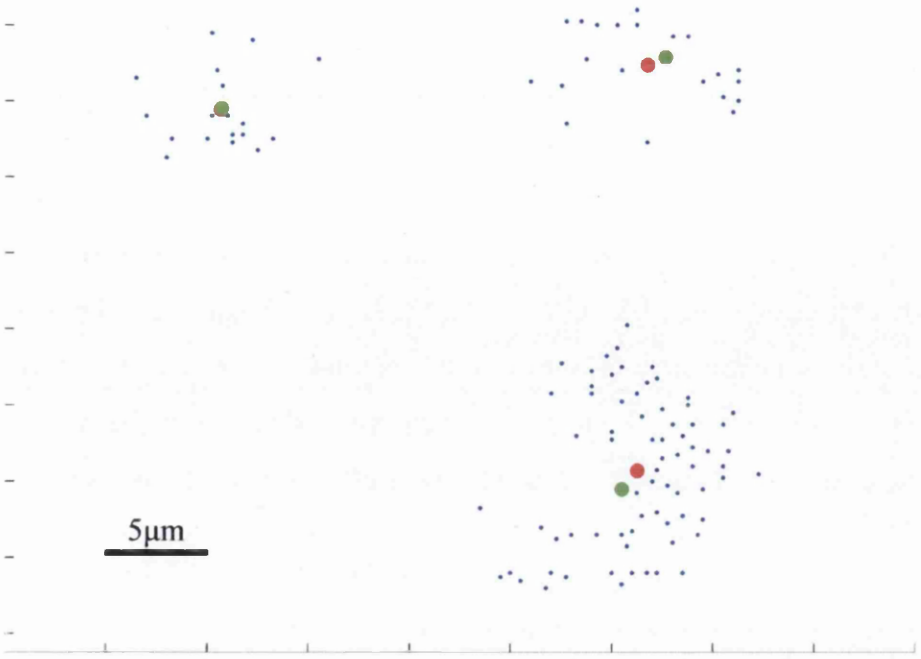



Figure 2.8: Scatter plot demonstrating the evolution of the cluster process on three binary element groupings. The green dot marks the location of the previous 'seed' centroid and the red dot marks the current binary element group centre.

Figure 2.8 shows a sample of 3 binary element clusters with the previous cluster centroid marked in green and the current convergent centroid marked in red. To account for the sporadic occurrence of rogue binary elements, peaks of noise that may have passed through the filter, and binary elements representing cells entering the field of view from beyond the image boundaries, new centroid seeds are assigned to regions containing binary elements that are more than 120 pixels (30µm or ~3 cell diameters) from their nearest cluster centroid. New seeds are assigned to deal with both the occurrences in the same way as they are indistinguishable without time consuming comparisons with previous frames and interpretation of boundary events. It is temporally and computationally more efficient to assign binary elements to new clusters generated according to basic rules and interpret the nature of these groupings afterwards using the temporal plane of the collated data. For example, groups of random binary elements that should be attributed to noise are unlikely to be present in the next frame in the same place and will lead to a binary element group whose lineage will be 5 minutes long and as such will be easy to filter out as a non-cell grouping. At the opposite end of the spectrum, a proximity validation is applied by

identifying any current seeds with no binary elements within a distance of 50 pixels (12.5μm) and removing them from the seed set. Primarily this is to deal with the event of a cluster group moving out of the field of view and leaving a centroid seed from the previous frame with no binary elements to properly define it and secondarily, with centroids previously defined to account for rogue elements whose intensities have dropped back below the noise filter. These modifications to the seed set are made to prevent erroneous miss-weightings in the k-means algorithm that can result in single binary elements causing a cluster centroid to deviate from the true cluster centre. Once a binary element set has been assigned to clusters based on the current seed set a cluster fit parameter is defined using a k-means silhouette algorithm [15]. The silhouette process is performed as follows:

For each binary element i, let $a_i$ be a measure of dissimilarity with elements in the same cluster, in this application the dissimilarity is the average distance from all other cluster member elements.

Perform a similar operation comparing the element i with all elements in all other clusters successively and assign the lowest found dissimilarity measure be $b_i$.

Define the silhouette of the binary element i as

$$\boldsymbol{S_i} = \frac{b_i - a_i}{max\{a_i, b_i\}} \qquad (2.6)$$

where $\boldsymbol{S_i}$ can be seen to obey $-1 \leq \boldsymbol{S_i} \leq 1$. A value close to 1 indicates the element i fits well within its assigned cluster and the average $\mathbf{S_i}$ of all the binary elements is a measure of how well assigned the clusters are. To identify possible mitosis events via the binary elements an average binary element separation parameter is calculated for each cluster sequentially by summing all the inter-element distances and dividing by the number of elements within the cluster. If this parameter is found to be greater than 15μm for a given cluster, ~1.5 average cell diameters, it is categorized as a possible occurrence of a division event. To check the validity of this possibility, the k-means algorithm is run again with an additional centroid seed placed 10 pixels from the centroid associated with the probable division event. A silhouette cluster fit parameter is then re-calculated and compared to the fit parameter before the additional seed was added. If the new value indicates a more favourable fit than the previous configuration the additional centroid is accepted as a seed point to better

interpret the binary element clustering. The silhouette validation step provides a second measure for the assessment of division events lessening the chance of such events being wrongly flagged due to events such as low numbers of outlier binary elements driving the separation parameter high. This process is reiterated until all clusters within the field of view are validated. An example of the resulting centroid allocations is shown in figure 2.9 where the binary elements in a full field of view are plotted alongside the k-means cluster centroids.

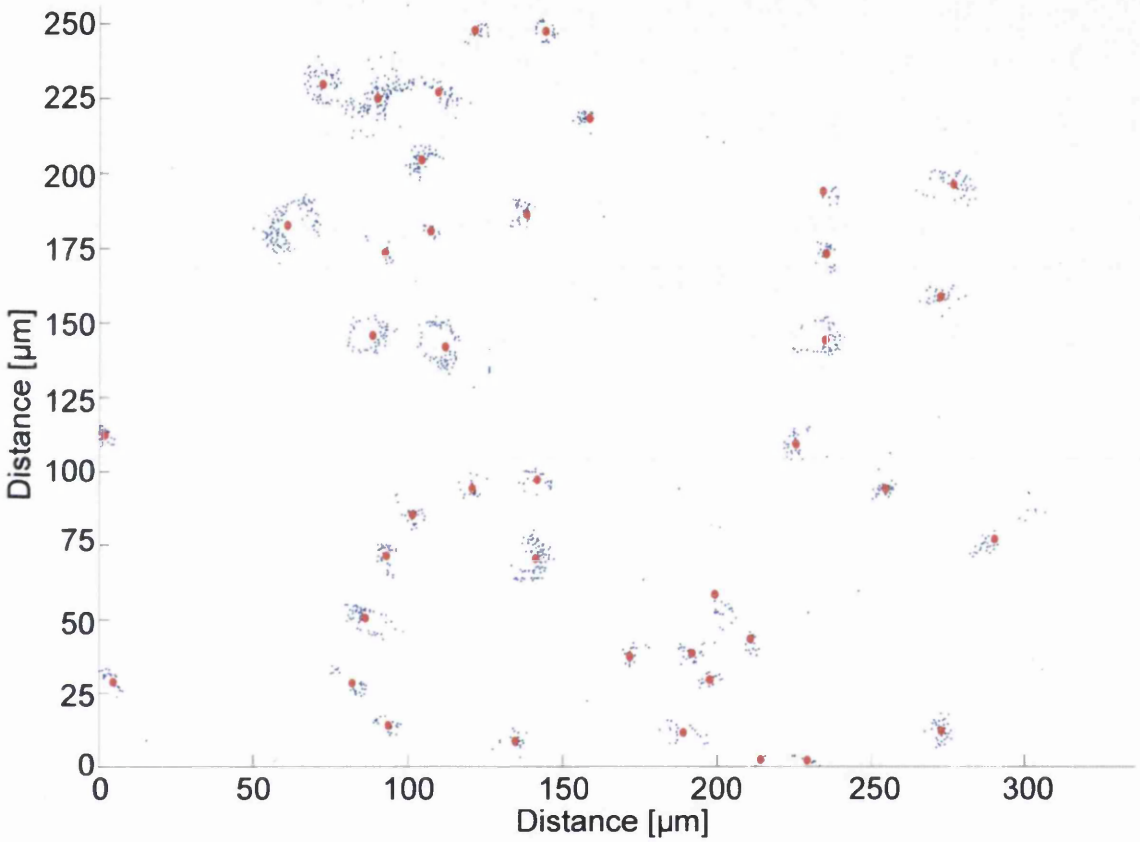



Figure 2.9: Scatter plot displaying binary element groupings in blue and their respective centroids in red for time frame 10, 45 minutes after initial image.

## *Cluster Refinement: Centroid Linkage through the Temporal Plane and the Mitotic Signature*

The previously detailed clustering protocols provide a base centroid set **C** for each time frame in the data set and can be further refined in terms of spatial coordinates by taking advantage of the temporal dimension. Treating time as the z-dimension, each cluster centroid can now be linked to centroids in surrounding time frames through the application of a nearest neighbour map. Two sequential frames with centroid sets $\mathbf{C}_T$ and $\mathbf{C}_{T+1}$ are mapped to each other as follows:

- If $\mathbf{C}_{T+1} > \mathbf{C}_T$

Calculate the Euclidian distance matrix **D** of dimensions J×K where $1 \leq J \leq |\mathbf{C}_T|$, $1 \leq K \leq |\mathbf{C}_{T+1}|$ and $\mathbf{D}_{JK}$ is the Euclidian distance between the coordinates of the $J^{th}$ centroid member in $\mathbf{C}_T$ and the $K^{th}$ centroid member in $\mathbf{C}_{T+1}$. Then iterate the following assignment protocol;

- Identify the minimum entry, $\mathbf{D}_{JK}$, in the matrix **D**. Providing $\mathbf{D}_{JK}$ is less than 50, corresponding to the 12.5μm average diameter of a single cell, assign the $J^{th}$ centroid in $\mathbf{C}_T$ to the $K^{th}$ centroid in $\mathbf{C}_{T+1}$.
- Set all entries of matrix **D** in row J and column K to 50, effectively removing them from the assignment protocol.
- Repeat the process until the minimum of **D** is greater $\geq$ 50.

Any remaining unassigned centroids in $\mathbf{C}_T$ are marked as lineage endpoints, typically a consequence of cells migrating beyond the boundaries of the field of view or previously mentioned rogue features that appear randomly and disappear in a like fashion. Centroids in $\mathbf{C}_{T+1}$ that were unassigned in the first neighbour mapping are subject to a second mapping against the now assigned subset of centroids from the set $\mathbf{C}_T$. Letting $\mathbf{C}_T^A$ be the subset of assigned centroids from set $\mathbf{C}_T$ and $\mathbf{C}_{T+1}^U$ be the subset of unassigned centroids from set $\mathbf{C}_{T+1}$. Calculating another distance matrix of dimensions J×K where $1 \leq J \leq |\mathbf{C}_T^A|$, $1 \leq K \leq |\mathbf{C}_{T+1}^U|$ and sequentially checking each column minima enables the assignment of the remaining centroids in subset $\mathbf{C}_{T+1}^U$. Columns with minima less than 50 correspond to daughter centroids that moved further away from the coordinates of progenitor location than their siblings and columns with minima greater than 50 are either rogue features that have occurred in

the frame T+1 or a QD cluster centroid within a cell that has migrated into the field of view, both scenarios are defined as new progenitors and stored accordingly.

- If $\mathbf{C}_{T+1} \leq \mathbf{C}_T$

The Euclidian distance matrix is defined as before only now, simply pass iteratively through columns 1 to $|\mathbf{C}_{T+1}|$ of the matrix **D**,

- Identify the minimum entry in the $K^{th}$ column and, providing the value is less than 50, assign the $J^{th}$ centroid in $\mathbf{C}_T$, the row of the column minima, to the $K^{th}$ centroid in $\mathbf{C}_{T+1}$.
- Set all entries of matrix **D** in row J to 50.
- Repeat across all columns in the matrix **D**.

In this instance, much like before, any unassigned centroids in $\mathbf{C}_T$ are marked as lineage endpoints and any unassigned centroids in $\mathbf{C}_{T+1}$ are defined as new progenitors. For both instances the centroid data in set $\mathbf{C}_T$ is associated with a vector, $\mathbf{L}^+$, of length $|\mathbf{C}_T|$ containing the index or indices of the nearest neighbour or neighbours[9] in $\mathbf{C}_{T+1}$ for each entry in $\mathbf{C}_T$. Similarly, the data in $\mathbf{C}_{T+1}$ is associated with a vector, $\mathbf{L}^-$, of length $|\mathbf{C}_{T+1}|$ containing the index or indices of the nearest neighbour or neighbours in $\mathbf{C}_T$. End and progenitor points as previously defined are marked with zeros in $\mathbf{L}^+$ and $\mathbf{L}^-$ respectively indicating no temporal connection with other centroid entries, for example, $\mathbf{L}^-$ for the initial frame is a vector of zeros as the initial frame contains only progenitor centroids.

The vectors $\mathbf{L}^+$, looking forward, and $\mathbf{L}^-$, looking backward, map the progression of the cluster centroids through time and in conjunction with the centroids coordinate sets can be used to extract lineages (see Chapter 3). With lineages available, a further clustering refinement can be made by monitoring the x-y spacing of the binary elements in a cluster through time (see fig 2.10 for time series showing a mitosis event in the 3 data formats).

[9] In the case of mitotic linkages

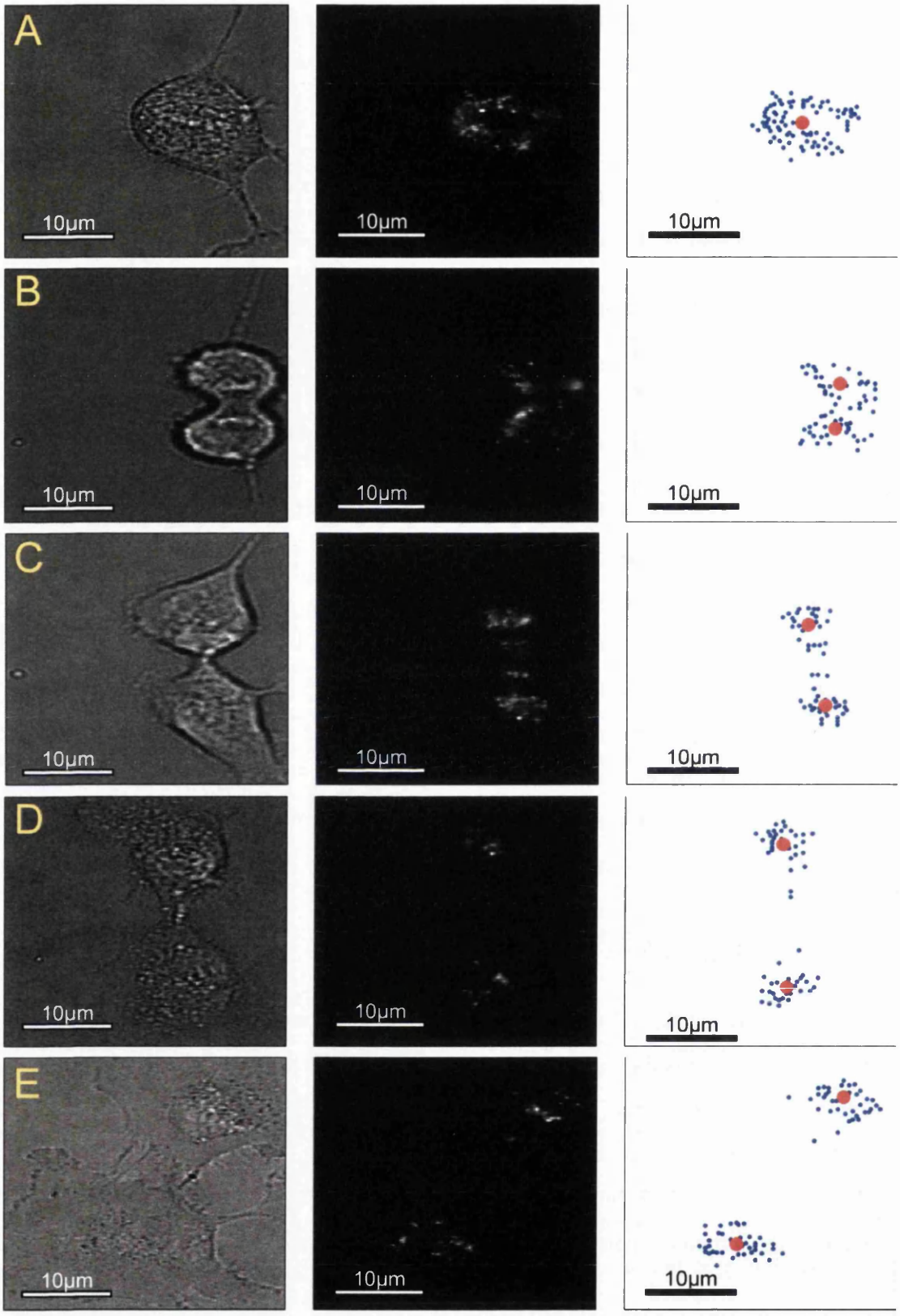



Figure 2.10: Stills showing a mitosis event from the point of contraction, A, through to complete separation, E in the brightfield, fluorescence and binary element display formats.

The average separation of the binary elements within a cluster can reveal some information about the cell phase, specifically, distinct morphological changes can be tracked using the binary element cluster's mean distance as a characteristic parameter to define the 'mitotic curve' shown in figure 2.11.

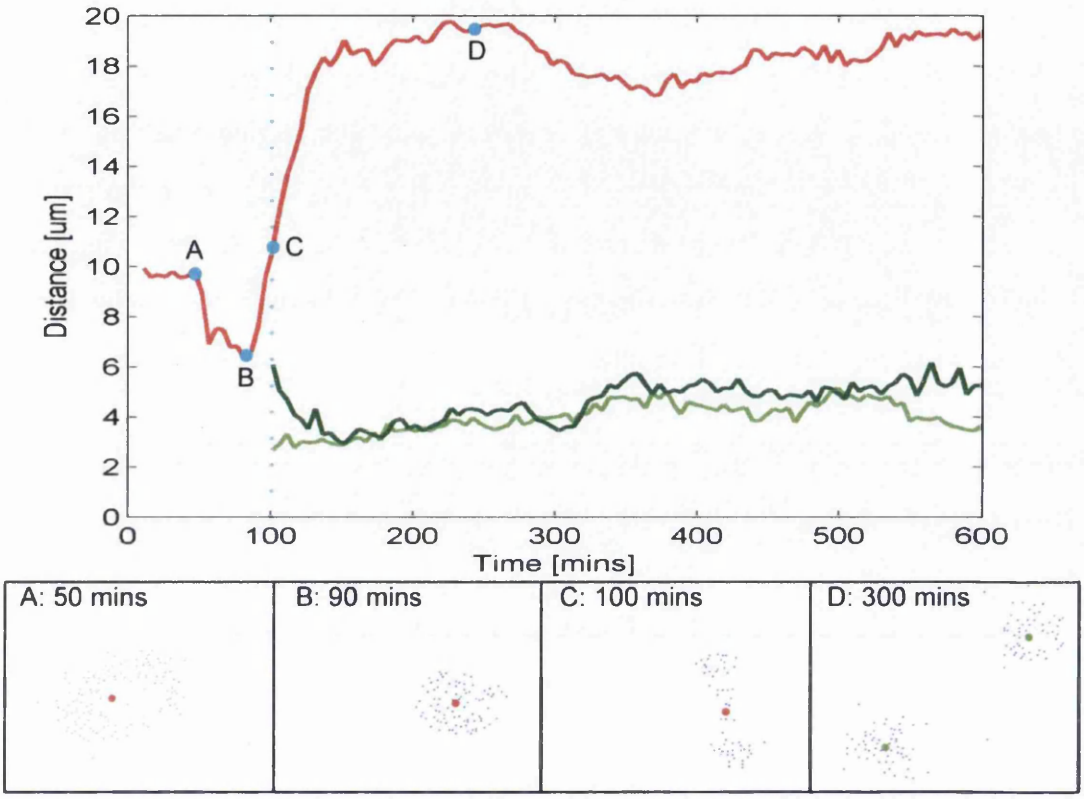



Figure 2.11: An example of a signature mitotic curve for a mitosis event. Panels A-D are marked in their relevant timeline positions on the graph and the point where the binary elements are best described as 2 separate clusters is marked by the blue dotted line.

Defining a binary element group containing n elements as the set $\mathbf{B} = \{b_1, b_2, ..., b_n\}$ the average distance is a simple calculation:

- Define another distance matrix **D** of size $|\mathbf{B}|$ x $|\mathbf{B}|$ where element $D_{1,1}$ is the distance between $b_1$ and $b_1$, zero as it is the same point, $\mathbf{D}_{1,2}$ is the distance between $b_1$ and $b_2$ and so on.
- The mean is then defined as $\mu = \frac{\sum_{i \neq j} \boldsymbol{D}_{i,j}}{n^2 - n}$.

The above definition sums all the elements off of the diagonal then divides by the total number of matrix elements minus the elements present in the diagonal

accounting for double counts of distances between points, i.e. $b_{1,2} = b_{2,1}$, when dividing by the total number of elements, $n^2$, with the diagonal, n, removed. This process is computationally quicker than isolating the upper or lower triangle of **D** to avoid double counting. When cells enter the mitotic phase they detach from the adherent surface and contract into a spherical geometry in preparation for division. This is evident in the binary element paradigm as this morphology change causes the condensation of the quantum dot nanoparticle vesicle distribution resulting in a reduction in the mean binary element separation (fig 2.11 A-B). After chromosomal separation the nanoparticle loaded vesicles disperse as the cell passes the telophase and the mitotic curve begins to increase (fig 2.11 B-C). As the cell completes mitosis the binary element spacing further grows as the now two clusters separate in their respective daughter cells (fig 2.11 C-D); this increase arises from the computational assignment of elements to a single centroid whereas there are clearly two physically separate groupings following a division event. This is therefore a method to consistently define a mitosis event and refine the clustering process by transitioning from a single to a pair of centroids and reassign the binary elements to daughter cells when the average separation of a cluster after the contraction phase surpasses a level defined as the stable separation distance maintained prior to contraction (fig. 2.11 C). This mitotic curve signature provides a clear marker of cell division, a quantitative track through the mitosis phases which can inform of the cell division process and a method for analyzing the shape and motion of cells. The curve itself remains well defined through longer sampling rates as demonstrated opposite in figure 2.12 where an artificially applied sampling shows the general shape of a signature mitotic curve remains under 10 minute, 15 minute and 20 minute image collection time steps demonstrating robustness with regards to the temporal resolution. However, while it is not necessary to have a 5 minute capture rate for the purposes of mitosis identification through the signature curve, the reality is that the slower the image capture rate the more error prone the initial clustering system becomes as objects have the potential to cover more distance, a serious issue in situations of higher confluence, particularly after mitosis when A549 daughter cells tend to move in opposite directions quite rapidly.

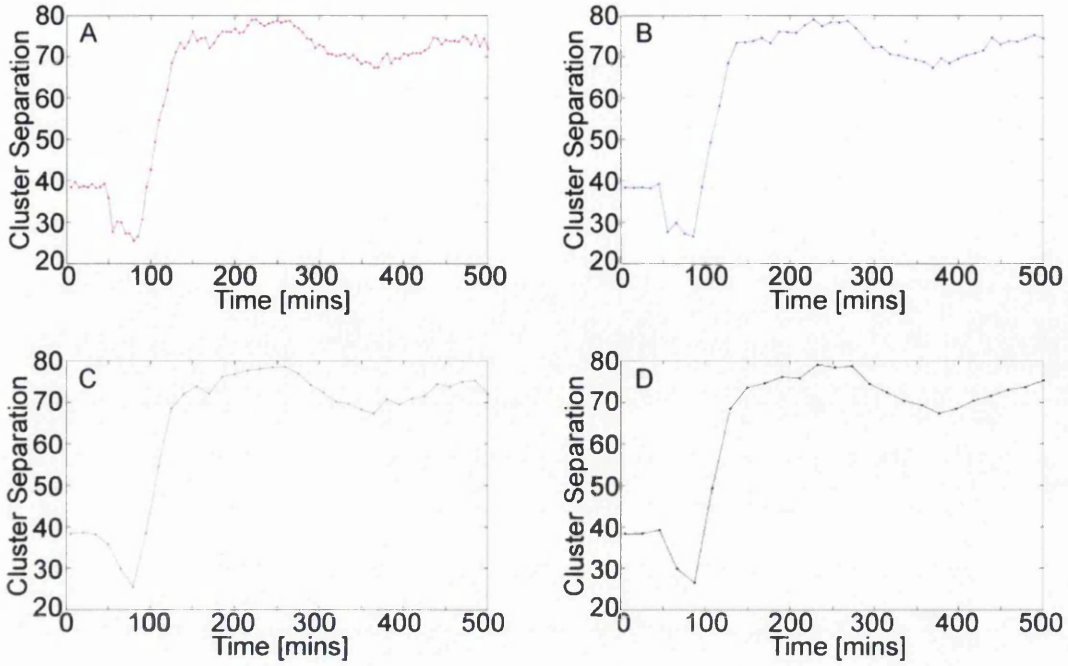



Figure 2.12: Plots comparing the mitotic curve under different sampling rates. (A) original capture intervals of 5 minutes, (B) pseudo 10 minute sample intervals, (C) pseudo 15 minute capture intervals and (D) pseudo 20 minute capture intervals. (Cluster separation is displayed in pixels)

## *Operational Range - System Accuracy and Robustness*

Numerical quantification of the operational range of the quantum dot nanoparticle moving light display concept can be conducted by taking a representative sample image and investigating the effects on cluster identification and centroid assignment when applying a variation in the noise filter threshold. This approach allows for a direct relation between the accuracy of the technique and the number of encoding light points (identifiable nanoparticle loaded vesicles) and corresponds to an experimental measure of fluorescence signal to noise (SNR). For a single sample image the peak finder protocol is applied, without the image being first filtered and all the local peaks are identified. The sample frame chosen was the initial frame as the number of centroids was a known quantity having been manually identified for seeding the clustering. Due to the number of peaks found in background noise this process is time consuming but is done purely for the purpose of investigating the operational range of the tracking system. The process identified approximately 140,000 separate peaks compared to the 1,300 peaks identified with the step function filter applied and stored their coordinates and intensities in vectors sorted in terms of intensity in descending order. A plot of this peak vector is shown in figure 2.13A with the x-axis cropped at 2000 as beyond this point the variation in intensity plateaus around the noise floor. As before, binary elements are representative of these peaks, as such the x-axis is readily interchangeable between binary element number and peak vector number and is labelled accordingly. The minimum number of binary elements needed to define all initial centroid groups is then found by sequentially adding entries from the peaks list to a data set on which the clustering algorithms are applied. Essentially, providing there are at least as many binary elements as there are seed point as it makes little sense to try clustering data with less data points than proposed clusters, if there are no binary elements in proximity of a seed point the clustering algorithm will iteratively move the centroid away from the seed point. Beyond a threshold value, set here as the average single cell diameter for an A549 of 12.5μm or 50 pixels, the centroid is incorrectly assigned. The minimum number of binary points is therefore when they are sufficient in number for all centroids, optimized via k-means clustering, to be within this threshold region of the initial seed (see supplemental media for animated play through). The point where 38 initial clusters, 3 of the manually seeded 41 being ignored here as these cells were on

the boundary with large portions not being visible in the field of view, had at least 1 binary element within the defined threshold region is shown in figure 2.13B by the blue curve against the left-hand y-axis with a red dashed line marking the number of binary elements needed as 520.

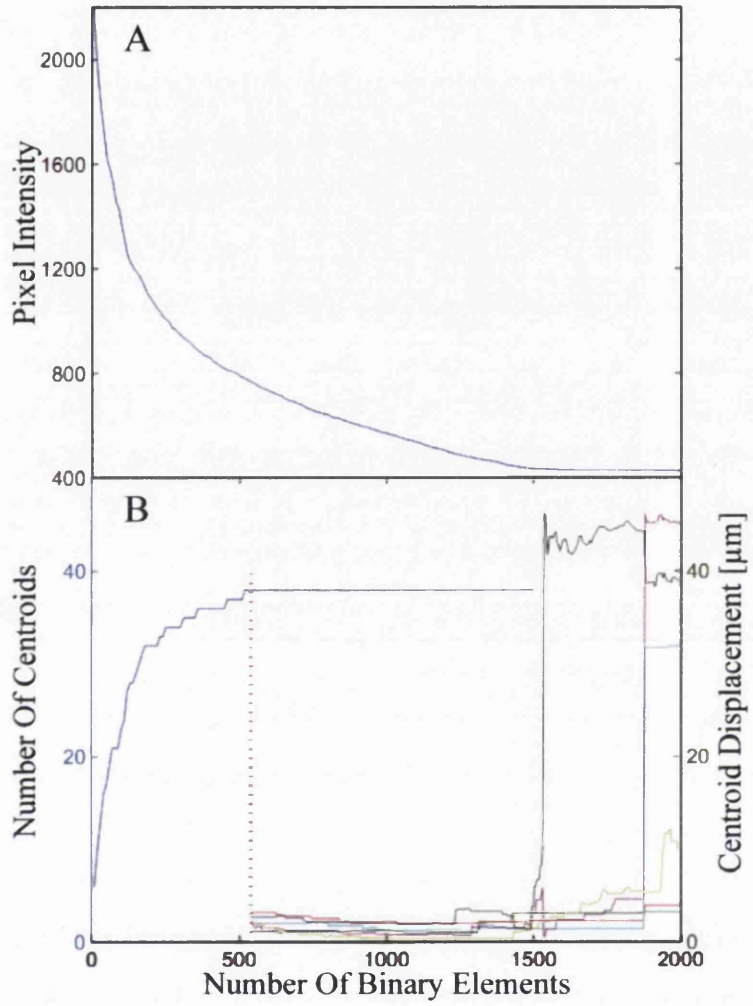



Figure 2.13: (A) Calibration curve of the initial image showing the relationship between the intensity threshold filter cut-off and the number of binary elements peaks identified (see media for animation). (B) Operational range of the protocols through the binary element dependence on centroid (cell) identification (blue curve) and the centroid displacement of a selection of cells (coloured curves). Red dashed line indicates the minimum number of required binary elements for maximum cell identification.

Utilising the conversion plot in figure 2.13A, this can be thought of as corresponding to noise intensity cut-off of 784 units. With the maximal signal reaching approximately 2200 units this gives a signal to noise ratio (SNR) of 2.8 and is deemed a 'worst case' scenario as 30 of the 38 centroids were defined by a minimum

of 5 binary elements by the time the threshold cut-off reached 1100 intensity units (SNR = 2). The 8 centroids that seemingly require 320 further binary element points to be allocated to the field of view before they all acquire one within their respective local regions were all located towards the boundaries of the image. This non-isotropic distribution of binary elements across the field of view is primarily the result of a lens effect on the fluorescence channel images causing a generally higher level of intensity towards the centre. This acts to increase the global average intensity and causes the simple step function filter to cut-off boundary peaks before affecting central ones. There is consequently the prospect of improving the 'worst case' scenario through the use of more complex image pre-processing techniques taking into account and scaling for such local regional variations by, for example, characterising the shape of the lens effect and damping local intensities accordingly.

At the other end of the scale, if the noise threshold filter is reduced too far the binary element number begins to relate to background noise rather than valid QD encoding pixels. As a random noise source has the potential to cause a rogue peak in each cell cluster with equal probability the presence of noise does not immediately invalidate the identification of cells but can affect the centroid coordinates and so lead to an inaccurate determination of a cells position. As the filter approached the noise floor, intensity approximately 400 units, and the total number of binary points exceeds approximately 1500 the centroid positions can show marked deviations, >10μm, from their known seed points (see supplemental media for animation). Figure 2.13B contains 6 sample centroid deviations shown by the coloured lines against the right-hand y-axis. The plot demonstrates the operating range of 520-1500 binary elements before noise begins to adversely affect the accuracy of cell identification.

The minimum number of elements necessary within a cell for successful identification is an important criterion when tracking proliferating cells as the QD loaded vesicles will become diluted during mitosis [62]. Whilst there are fluctuations in binary point numbers per cell from frame to frame the numbers show stability in longer scales. Short term variations can occur due to occasional noise interference, an issue that worsens as the QD signal fades over time and cells move in and out of the field of view. The very nature of the fluorescence image being a 2D planar image of a 3D object creates the possibility for loaded vesicles to be covering each other or hidden behind features such as the nucleus. With the step filter in operation the initial

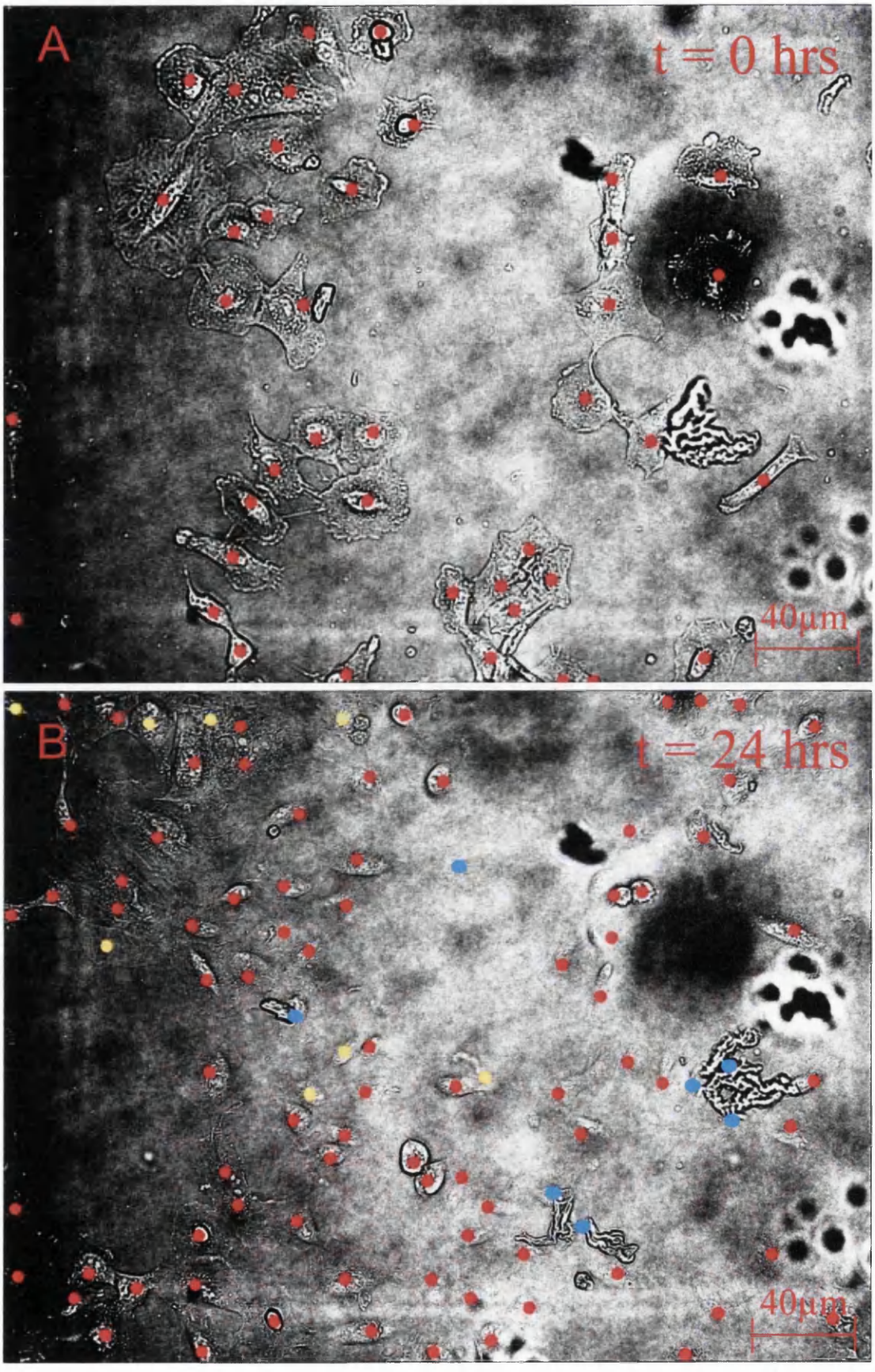
frame contains 1393 binary elements, leading to an average point per cell number of ~34. The 24 hour time point had an average of 18 and at 40 hours the average binary element number per cluster had reduced to ~12. This trend leads to an expected points per cell average of ~5 around the 72 hour mark (3-4 divisions events) meaning the QD markers are viable as tracking points for 3 days before over-dilution would make cell identification unviable. Increasing cell densities cause issues before this dilution limitation making confluency the main hurdle for longer term tracking. The accuracy of the moving light display approach utilizing quantum dot loaded vesicles as an impartial method for identifying and tracking cells over a 24 hour period can be visually assessed later in figure 2.15.

## *Tracking Results*

Automation is useless if it does not provide reasonable levels of accuracy so as to be a useful tool in replacing manual tracking; saving manual work-hours at the cost of meaningful results is a false strategy. A 602 frame data set underwent the previously detailed protocols to identify cell lineages through a 48 hour time span. Figure 2.14 (opposite) shows two brightfield images overlaid with their respective identified centroids. The initial frame, 0 hours, was manually labelled and details the 41 manually identified cells within the field of view[10]. It is clear to see that often images contain miscellaneous foreign bodies amongst the cells which can show in the fluorescence channel and cause confusion and consequently errors in the application of the tracker. Most oddities of this nature are essentially filtered out while performing the final lineage extraction as artifacts (previously referred to as rogue points). Randomly appearing in the fluorescence channel and being assigned centroids, they are discarded as their 'lineages' typically exist for 3-4 frames or about 20 minutes. Even if these were cells they would be of little use when analysing cell lineages as they are far too short for the extraction of meaningful parameters. At time t = 24 hours the actual number of visible cells has risen to 89, with 81 cells being correctly labelled (red) and 8 cells being missed (yellow). This represents an approximate 91% tracking success rate for the first 24 hour image sequence. The 7 blue marks are centroids showing rogue points, having been erroneously tagged. However, 5 of these only exist for a few frames and are as such discarded and the other 2, while appearing and disappearing across the entire data set do not exhibit any x-y motion other than minor oscillations.

Figure 2.14 (opposite): Graphic overlaying the identified centroids onto the relevant brightfield image for t = 0 and t = 24 hours. The initial frame shows the manually identified cells labelled with their respective centroids in red and at 24 hours red indicates correctly assigned centroids, blue indicate wrongly assigned centroids and yellow marks cells missed by the tracking algorithm.

[10] It is provided here to show the somewhat subjective nature of identifying cells by eye. Differences in the perception of what is and is not a cell can cause discrepancies between different manual trackers making it a false claim to say manual tracking is perfect solution.


A
t = 0 hrs
40μm
B
t = 24 hrs
40μm

Towards the 48 hour mark the cells become highly confluent making it difficult even for manual tracking and causing interpretation issues when trying to apply a set of basic rules for automation. Cells can switch places between time points, mitotic daughter cells can remain detached from the surface for multiple frames after division as they try to find enough space to adhere to due to the increased cell density. The number of nanoparticles within cells inevitably drops as they are diluted by mitosis events and the nanoparticle fluorescence signal itself fades over time with a significant drop over a 48 hour period. By the 40 hour mark these factors combined to create a situation where the clustering algorithm in particular struggled to correctly assign centroids with lowering numbers of identifiable QD loaded vesicles and increasing levels of rogue points as a consequence of the lowering signal to noise ratio. Consequently, despite a data range of 48 hours, confidence in the ability to accurately identify individual cells begins to drop after 36 hours with 40 hours being the cut-off point at which most cells have insufficient identifiable QD loaded vesicles for accurate clustering and meaningful inter-binary element distance parameters. At this cut off the manual cell count was 128 with 101 correctly labelled representing a hit rate percentage of ~78% which for a large scale experiment would still provide a large number of potential lineages to analyse. The critical limiting factor of the digitization approach is the ability to identify cells as their confluency increases and they become difficult to resolve without knowledge of their respective boundaries. One possible way to extend this limitation is to apply a nuclear marker to the cells at the end of the experiment and, using the nuclear stained cells as the initial seed points, run the entire analysis process backward in time. This would provide well-identified, initial cluster centroid seeds and simplify the identification of lineage bifurcations as in reverse the daughter centroids would converge to their progenitor removing the need to make a decision of whether binary element groups are best described as multiple or singular groupings. Another prospective solution to highly confluent cell identification is the application of multiple coloured quantum dots which is discussed further in chapter 4. A plot of the number of centroids per time frame is plotted in figure 2.15 which can be viewed as the binary paradigm's growth curve. Fluctuations exist due to random appearance of rogue points causing temporary centroids assignments but regardless of these erroneous allocations the trend does follow that which is expected for an A549 cell line with a near doubling after 22 hours and progression towards a further doubling after 44 (although as

previously mentioned, the analysis only runs to 40 hours). The growth curve itself along with the QD vesicle dilutions predict a cell doubling time of ~ 23 hours within the first 24 hour period which as an overview with no interpretation of the nature of each centroid counted is a good estimate of the true value.

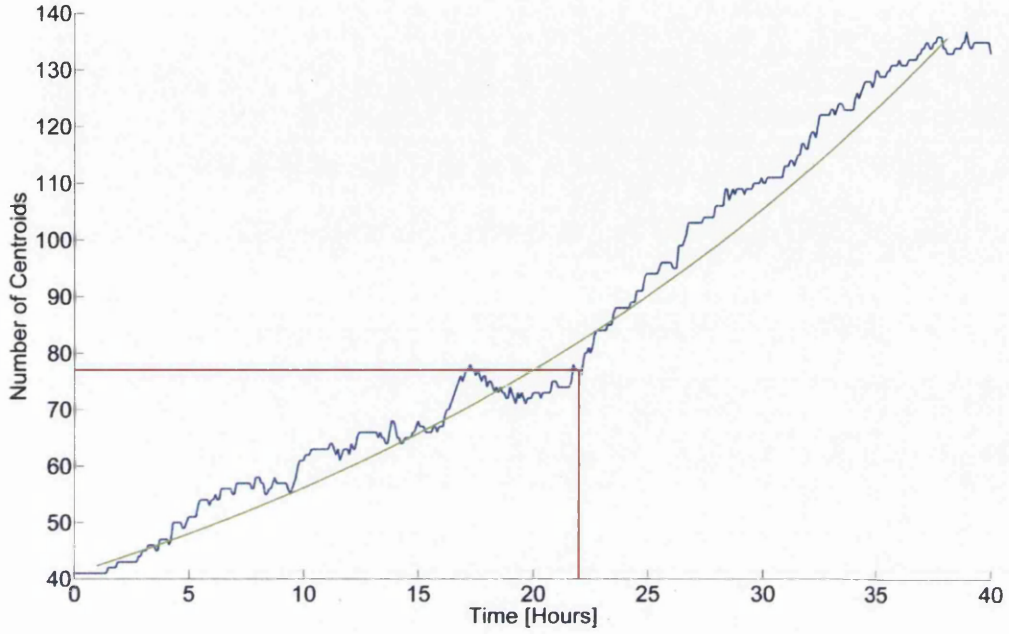



Figure 2.15: Number of centroids present in each frame scaled to time along the x-axis. The red dashed line shows the inter-mitotic time of ~22 hours for A549 cells correlating to close to a doubling from 41 to 77 centroids. The green line shows an exponential growth curve for a population with a doubling time of 22 hours.

# CHAPTER THREE

# Automated Tracking: Data Mining and Analysis

## *Nodal Storage and Raw Lineage Extraction*

The identified cell centroid locations, and therefore the cell locations, can now be utilised to investigate the motility of the A549 sample though the timelapse sequence. The biological interest in this particular sample was to see whether any subpopulations of cells within the population exhibited any extremes in terms of motility. Application of the processes described in chapter two yields centroid vector sets and their respective interaction maps which encode the lineages of all the identified binary element centroids (corresponding to cells). It is at this point that all said lineages can be fully extracted, interpreted and filtered to remove instances of non-cell-like groupings such as the aforementioned rogue element lineages. The general method involves interpreting the linkage vectors, $\mathbf{L}^{+}$ and $\mathbf{L}^{-}$, to temporally connect the centroid locations stored in $\mathbf{C}_T$ for each individual timeframe and generate lineage vectors of the form

$$v_i = (x, y, t) \qquad (3.1)$$

storing the spatio-temporal data for cells between lineage node points[11]. Consequently, each lineage vector stores the cell centroid location from its initial existence, whether from a mitotic division or movement onto the field of view, to its demise, either moving out of the observable range, dividing into two new daughter cells or succumbing to apoptosis. Complete lineage trees can then be constructed by piecing together these lineage vectors and tracing their heritage through the time sequence. (An example of this process is illustrated on page 53)

[11] Here a lineage node points refers equally to new progenitor cells and cells just prior to mitotic division. New daughter cells are considered as progenitor starting points for lineage vector creation as well all cells in the initial timeframe and any cells migrating into the field of observation.

As detailed before, for each time frame, T, a centroid set, $\mathbf{C}_T$, is generated alongside the corresponding vector doublet $\{\mathbf{L}^+, \mathbf{L}^-\}_T$. Each row of the mapping vectors, $\mathbf{L}^+$ and $\mathbf{L}^-$, contains the indices of the centroids temporally associated with the centroid in the corresponding row of $\mathbf{C}_T$, with $\mathbf{L}^+$ indexing associations between $\mathbf{C}_T$ and $\mathbf{C}_{T+1}$ and $\mathbf{L}^-$ associations between $\mathbf{C}_T$ and $\mathbf{C}_{T-1}$. Lineages are extracted by first identifying all centroids labelled as progenitors, P, by association with an $\mathbf{L}^-$ index of zero. The lineages are then found by sequentially walking the timeline of each progenitor by following the relevant indices located in each successive $\mathbf{L}^+$ vector and storing the centroid coordinates into a lineage vector, $\mathbf{v}_1$, until said index reads zero indicating the end of the current lineage as no further temporal associations exist or until $\mathbf{L}^+$ contains two index values indicating a lineage bifurcation event. A bifurcation event triggers the creation of a further two lineage vectors, $\mathbf{v}_{1+bc+1}$ and $\mathbf{v}_{1+bc+2}$, a header vector, **H**, containing references to the source of each daughter lineage and a generation vector, **G**, storing the generation of each daughter lineage. The variable *bc* is then defined as

$$bc = 2 \times (n_b - 1) \qquad (3.2)$$

and accounts for the number of bifurcations that have already occurred, $n_b$ being the bifurcation count number[12]. Any further bifurcation events similarly trigger additional vector stores and increase the length of the header and generation vectors, **H** and **G**, in accordance with the value of $1 + bc$. All of the lineage components are then stored together in the set $\mathbf{A}_j = [\mathbf{H}, \mathbf{G}, \mathbf{V}]$, where $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_N\}$ and $N = 1 + 2^{n_b}$. The lineage of the next identified progenitor, $P_{i+1}$, is then extracted and so on. This sequential process generates a set of lineage vectors, $\mathbf{S} = \{\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_j\}$ where $1 \leq j \leq N_P$ and $N_P$ is the number of identified progenitors.

[12] Each bifurcation event adds to the $n_b$ counter. The variables bc and $n_b$ are only created in the event of a bifurcation event and from that point on are updated on the occurance of further bifurcations.

To better illustrate this procedure, consider the sample lineage diagrammatically displayed in fig 3.1. Here, the lineage vector from the progenitor, P, to the first bifurcation point, B1, is stored as $\mathbf{v}_1$. At the point B1 two new vectors are created; $\mathbf{v}_2$ for the lineage between B1 and B2 corresponding to the first daughter, D1, and $\mathbf{v}_3$ for the lineage from B1 to the final frame shown here as END and corresponding to the second daughter D2. The header for this part would read $\mathbf{H}$ = [0, 1, 1] as the vectors $\mathbf{v}_2$ and $\mathbf{v}_3$ originate from $\mathbf{v}_1$ and $\mathbf{v}_1$ is the progenitor, signified by the zero and the effective lack of an origin and the generation vector would read as $\mathbf{G}$ = [1, 2, 2]. The second bifurcation point, B2, creates another two vectors, $\mathbf{v}_4$ and $\mathbf{v}_5$ corresponding to D11 and D12 respectively. The sample lineage shows $\mathbf{v}_4$ running to the end and $\mathbf{v}_5$ ending prematurely, such occurrences can be due to events such as cells migrating out of the field of view or cellular apoptosis. The header vector would now read $\mathbf{H}$ = [0, 1, 1, 2, 2] with $\mathbf{G}$ = [1, 2, 2, 3, 3]. The vectors $\mathbf{v}_1$-$\mathbf{v}_5$ and $\mathbf{H}$ represent the full lineage associated with the progenitor P in fig 3.1 and would be stored together along with $\mathbf{G}$ in the lineage set $\mathbf{A}_P$.

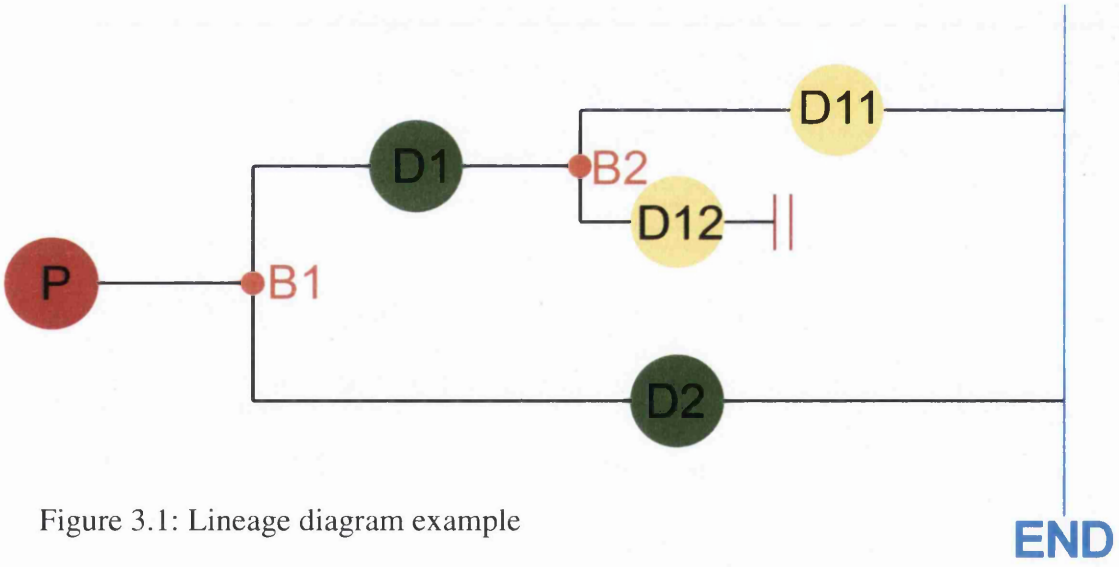



Figure 3.1: Lineage diagram example

Whilst the process seems somewhat elaborate it enables the storage of each branch of the lineage with quick references to its respective generation, a useful feature when investigating cross-generational behaviours, and the header vector, **H**, provides the directions to easily stitch together the branches to form graphical lineage trees for visualization purposes if required. This storage format also provides some simple measures to remove lineages attributed to rogue points by recognition of their non-cell-like characteristics. Groupings of rogue points in early time frames typically only appear for a few frames and as such any false lineages incurred by their presence contain too few centroids to be considered as true cell lineages and are discarded. Whilst it is possible in principle for a cell to appear briefly within the field of view and succumb to cell death in a similarly short time period such an occurrence would have to be as a result of a cell passing over a field of view boundary as any cells initially within the field of view would have an associated initial seed point and would therefore could not be interpreted as a rogue binary element grouping. Consequently, any prospective cell that migrates into view and then rapidly ceases to exist is of little interest in terms of population analysis as nothing is known about it prior to the aforementioned migration and little if anything can be inferred by the insufficient data points stored in the resulting lineage. Such a lineage would act only to falsely weight any statistical representations and as such it is reasonable to accept the possibility of misidentification of a genuine cell grouping with such a small footprint as a rogue binary element grouping and discard it accordingly. As the timelapse sequence analysed in chapter two becomes more confluent the number of rogue binary elements existing over a significant number of time frames, $> 5$ (25 minutes), increases. This is due to the number of cells in the field of view increasing over time and lessening the effect of the proximity filter used to discard low numbers of points occurring at an unrealistic distances from any known centroid or boundary. The majority of these rogue feature lineages however are removable via a characteristic trait resultant of their random nature. As they begin to gain greater temporal stability these rogue binary elements maintain a centroid with multiple binary element associations and are submitted to the same rules as the cell-like binary element groups. The random nature of these rogue elements results in the appearance of points across regions greater than a typical cell diameter boundary and their scarcity means low enough numbers to quickly breach the average separation parameter used as the primary marker for mitosis events. The resulting lineage

profiles of rogue element origins often contain an unrealistic number of generations for the time they exist over, regularly over 5 generations, which is an unrealistic number of divisions over the 40 hour range for which the automated tracker was operational. All such lineages are discarded by filtering for lineages with more than 4 generations by screening the final entry in the generation vector, **G**. All the remaining lineages are deemed to be sufficiently cell-like as to label them as cell lineages. Some lineages attributable to non-cell features may pass through these simplistic filters but the automatic analysis of such large data sets dilutes the impact of such events on the statistical yields as is the benefit of high-throughput experiments.

Whilst it would be possible to create a program to interpret the stored lineage data and create a graphic depiction of each lineage by stitching together the lineage branch vectors contained in each lineage set in **S** it is essentially an exercise in developing a graphic user interface (GUI) to create a user interpretable visualisation. However, the dominant principle throughout has been the automation of analysis procedures on large data sets. Consequently, the objective is for the system to output parameters sought to answer questions and test hypotheses with as little manual intervention as possible. The creation of hundreds of graphics for the purpose of displaying individual lineages to the user is not conducive to this goal. Figure 3.2 (overleaf) provides a sample of two such lineages selected from the set **S** to demonstrate the principle. The lineage in (A) shows 4 generations from the red progenitor through to the $4^{th}$ generation coloured here in blue with the daughters labelled with a N, north, and an S, south, subscript after each mitotic bifurcation point. The north-south naming convention in these graphics is adopted as an arbitrary method of selecting which bifurcation branches are displayed top and bottom with north corresponding to the cell whose centroid has the greater y-axis value in the first image where the two daughter cells are unique. In the event of the y-axis being exactly equal the north label is assigned to the daughter centroid with the greater x-axis value. The relative symmetry throughout this lineage could suggest the daughters remained in a local region with a consistently similar environment offering the offspring similar opportunities in terms of cell cycle requirements and that the daughters themselves are very similar to one another.

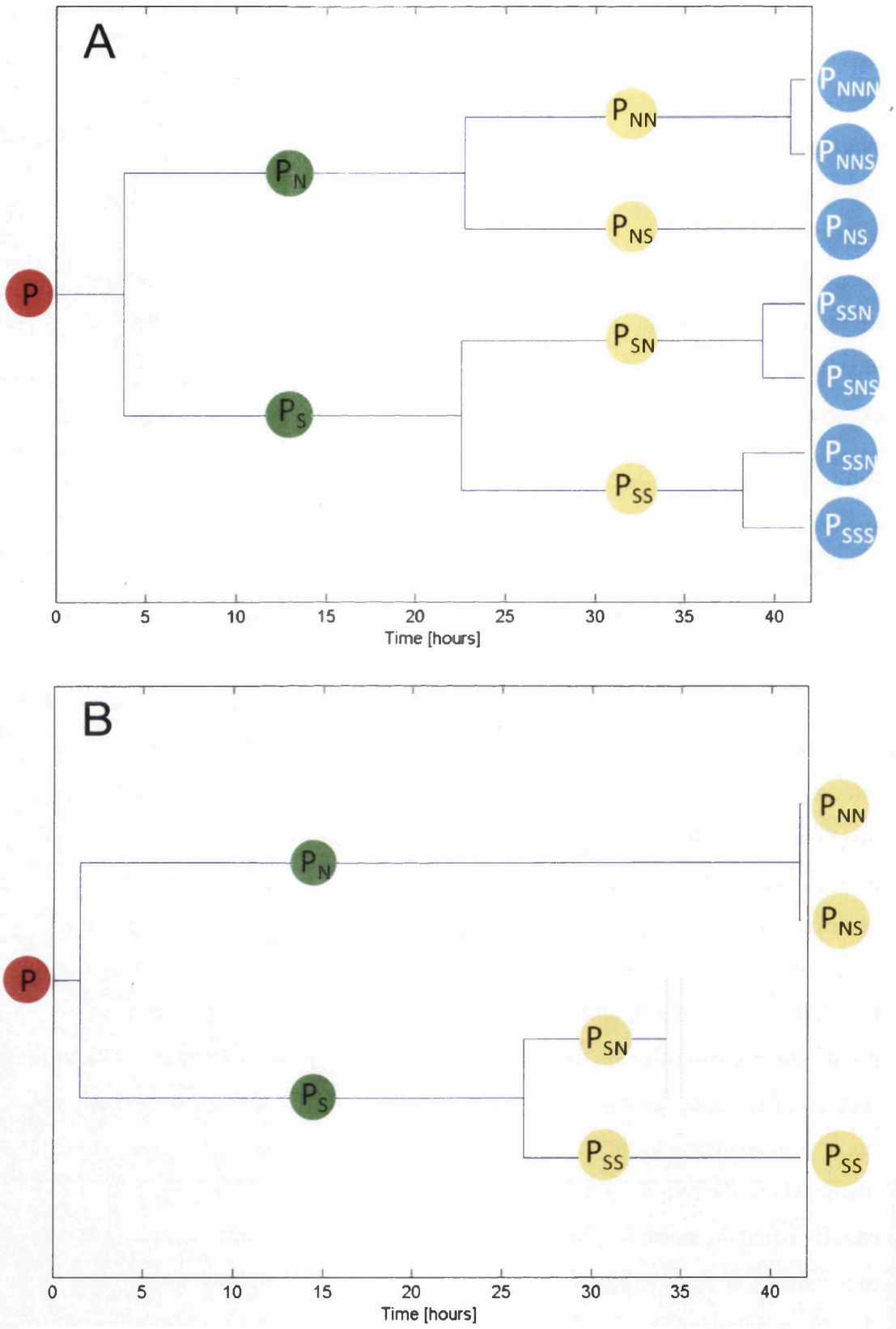



Figure 3.2 (A) Graphical portrayal of an extracted lineage showing 4 cell generations and (B) another extracted lineage with an example of a discontinuous branch where the $P_{SN}$ centroid is lost by the tracking algorithm. (In this particular case, the cell drifted off of the field of view)

In contrast (B) details a lineage with marked differences between its top and bottom halves, the initial north branch, $P_N$, remaining singular until the very end of the tracked range. The southern daughter, $P_S$, underwent a second mitosis event ~22 hours after the initial with the lineage of its respective northern daughter, $P_{SN}$, ending prior to the 35 hour mark. Abrupt endings such as these can be down to cell death, the cell being temporarily lost under lens debris and on its reappearance being recorded as a new cell due to the time gap incurred or a cell simply migrating beyond the boundary of the field of view as in this particular case. Graphic portrayals such as figure 3.2 are relatively easy to create for any given lineage but only really display information relating to the temporal dimension and are unable to readily convey any spatial information. While variations including spatial information are possible (see figures 3.7 for example) this impracticality when trying to represent hundreds of captured lineages proves statistical graphics to be far better suited to displaying meaningful representations of large data sets.

## *Motility Analysis*

With the cell location in the form of centroid coordinates known at every temporal node within each of the identified lineage sets, parameters for describing characteristics of the cell motility can be created dependant on investigative requirements. One simple measure implements Pythagoras' theorem to calculate the centroid migration distance between time frames, the results of which are shown in figure 3.3 (overleaf) in the form of histogram curves. Due to the way the lineages are stored it is a trivial task to separately calculate these profiles for each generation set and here generations 1-3 are displayed in red, green and blue respectively. Whilst instances of $4^{th}$ generation cells existed in the lineage set **S** they were present in insufficient numbers to attain meaningful statistics and as such are excluded from these examples of possible analysis procedures. The profile peaks reveal a consistent inter-frame migration distance across all three generations with averages of 4.62μm, 4.22μm and 4.36μm for the $1^{st}$ $2^{nd}$ and $3^{rd}$ generations respectively. These values equate to average speeds of approximately 15.4, 14.1 and 14.5nms$^{-1}$ between acquisitions demonstrating reasonable motile stability across the generations.

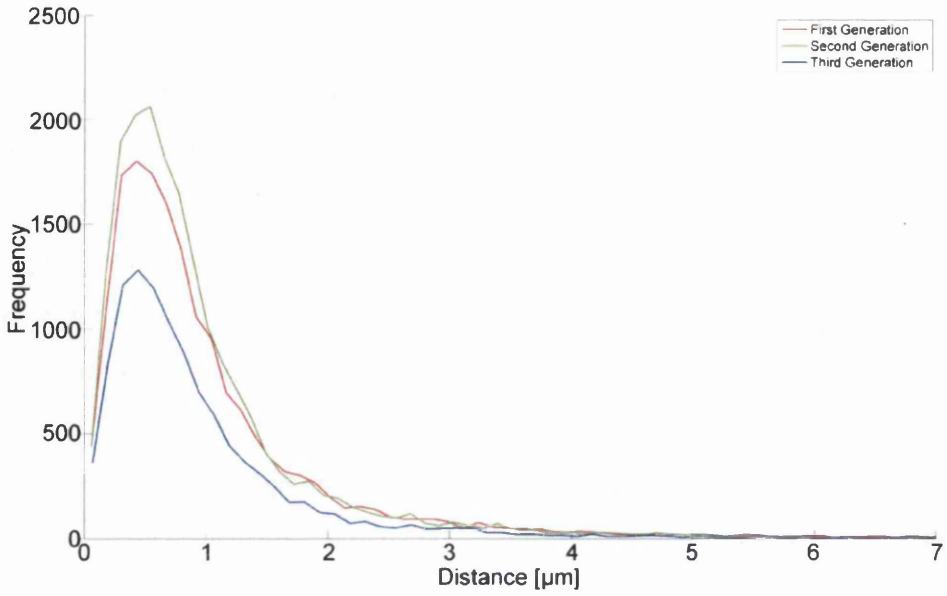



Figure 3.3: Histograms comparing the inter-generational differences in frame to frame motions. The x-axis is cropped at a value of 7 to allow for clear visualisation of the profile peaks.

With the number of cells increasing over time it would be expected that the height of these peaks would increase with each generation as is so when comparing the 1st and 2nd generations, the unexpected drop in numbers across the 3rd generation profile can however be reasoned by considering the duration of the time lapse. The A549 cell line has an average cell cycle duration of 22 hours [63] and as such over a 48 hour timelapse it is reasonable to expect two mitosis events to occur within a lineage tracked across the full temporal range. The actual number of mitosis events observed is dependent on the stage of the cell cycle the progenitor cell occupies at the initial time point, the variation of which can mean the difference between observing only a single bifurcation within a lineage or observing three. Taking 22 hours to be the standard A549 cell cycle duration, within a 48 hour range only the observed 2nd generation tracks can be expected to consistently cover the full cell cycle with all other observed generations varying in length with dependence on the state of the progenitor cell at time zero. With this in mind the lower values seen on the blue profile are most likely a remnant of the fact that typically by the time the 1st and 2nd generations have occurred there are insufficient time frames remaining in the timelapse for the 3rd generation to be fully described. As the 2nd generation tracks are the most likely to describe a full cell cycle they are also the best source from which

to extract cell cycle duration measurements. For each lineage set $\mathbf{A_j} = [\mathbf{H}, \mathbf{G}, \mathbf{V}]$, existing in **S**, extract the value $T_j$ such that

$$T_j(i) = |\mathbf{v_i}| \qquad (3.3)$$

where $i$ represents indices of $\mathbf{G_j}$ where $\mathbf{G_j}(i) \equiv 2$. The cell cycle duration can then be extracted by defining the average as

$$C = \frac{5}{60}\sum_j \sum_i T_j(i) \qquad (3.4)$$

with the fractional term transposing from frame count to hours. The result of this simple calculation yields a value of C as 22.4 hours, a better revision of ~23 hour estimate of the first chapter. This simple calculation could also be made to include all available generations with suitable lineage tracks by filtering for lineage vectors with bifurcation event associations at both their beginning and end points. It is also worth noting that there are more occurrences of larger inter-frame migration distances in the 3rd generation than the 2nd due to the larger number of cell divisions involved in the 2nd to 3rd generational transition and the propensity of the A549 cell line to rapidly separate and quickly migrate away from each other.

Having extracted a generational overview of the full set of lineages another analysis option is to compare the motions of complete lineage tracks, the inter-frame elements of which were displayed previously. For a selected lineage branch, **v**, the total distance travelled is simply defined for each dimension of the stored centroid coordinates as

$$D = \sum_i |\boldsymbol{v}(i) - \boldsymbol{v}(i+1)| \qquad (3.5)$$

where $1 \leq i \leq (|\mathbf{v}| - 1)$. The scatter plots in figure 3.4 (A) (overleaf) show the results of this summation in each dimension with distances depicted in micrometers for all of the identified lineage vectors and labelled with red representing 1st generation lineage vectors, green representing 2nd generation and blue 3rd. The majority of points lie along a linear fit that corresponds to the cells showing no obvious sign of preference towards motion in either the x or y planes.

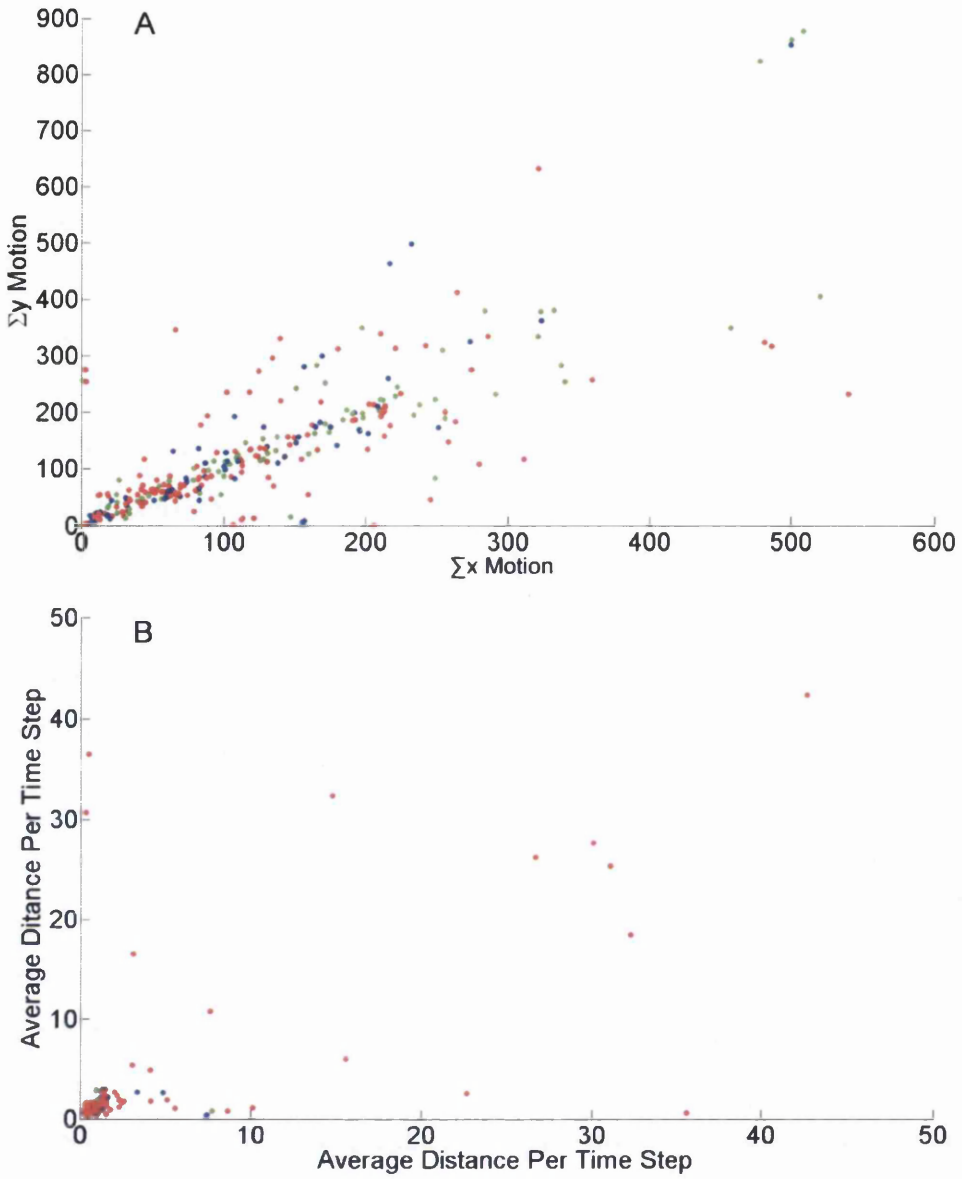



Figure 3.4: Scatter plots depicting the summation of each distance vector components, (A), and the same summation averaged over the number of track nodes, (B), for all identified cell tracks.

A small yet distinctive outlier group of 2$^{nd}$ and 3$^{rd}$ generation vectors can be seen around the (500,900) area. Such high values of D in this representation indicate long lineage vectors that could correspond to either a subpopulation of cells with prolonged cell cycles or a subpopulation of cells with high motility characteristics. Using this measure, automatic screening for cells exhibiting higher than average motility could be performed by creating a radial boundary, dependant on the experimental objectives, beyond which cells are deemed either particularly motile or potentially stalling in their cell cycle. Development of graphic user interfaces to

automatically create a parent lineage diagram, such as those seen in figure 3.2, on selection of motility points on these graphs is also possible using the lineage storage construct used here but as such a system reverts back to manual interaction for operation they are off-topic here. Figure 3.4 (B) shows the same points shown in (A) having been normalized against their respective lineage vector lengths[13]. The majority of points exist in the bottom left corner of this plot in agreement with the profiles seen in figure 3.3. The loss of the outlier group of 2nd and 3rd generation cells present in (A) suggests their inter-frame migration distances are comparable to the main population. Manually tracing the ancestry of the previously mentioned outliers in figure 3.4 (A) using the relevant header vectors revealed them to be of different lineages to one another. Only one of the 2nd generation points underwent mitosis which occurred around 37 hours having maintained steady motion since its progenitor divided at around the 6 hour mark. The other two 2nd generation points existed were created from similarly timed 1st generation mitosis events between 15 and 17 hours. While they did exhibit longer than average inter-mitotic times they also show higher than average migration distances resulting in large D values. The 3rd generation outlier was born of its 2nd generation progenitor after 23 hours which itself was born of its 1st generation progenitor within the first hour. Each of the observed generations of this lineage demonstrated higher than average levels of motility across a near to average inter-mitotic interval. Consequently, the large D values of the outlier group, or subpopulation, seen in figure 3.4 (A) seem to be attributable primarily to greater intrinsic motility and are enhanced in some cases by longer than average inter-mitotic times. In each case the trait appears with consistency throughout the lineage generations. Independent plots of the 2nd and 3rd generations are provided in figure 3.5 (A) and (B) respectively enabling optimised axis scaling to detail the similarities of their respective distributions. Here again the axis are defined in micrometers and the plot can be thought of as an alternative representation of the information shown in figure 3.3. When applied to the multiple acquisition points of a high-throughput experiment the histogram representations are more suited to displaying results with the scatter plots useful for acquiring and displaying subsets of data extracted from said histograms, outlier groups for example.

[13] As the lineage vector stores the centroid coordinates for a given cell for each frame through which it exists, normalizing to the lineage vector length yields the average migration per time step.

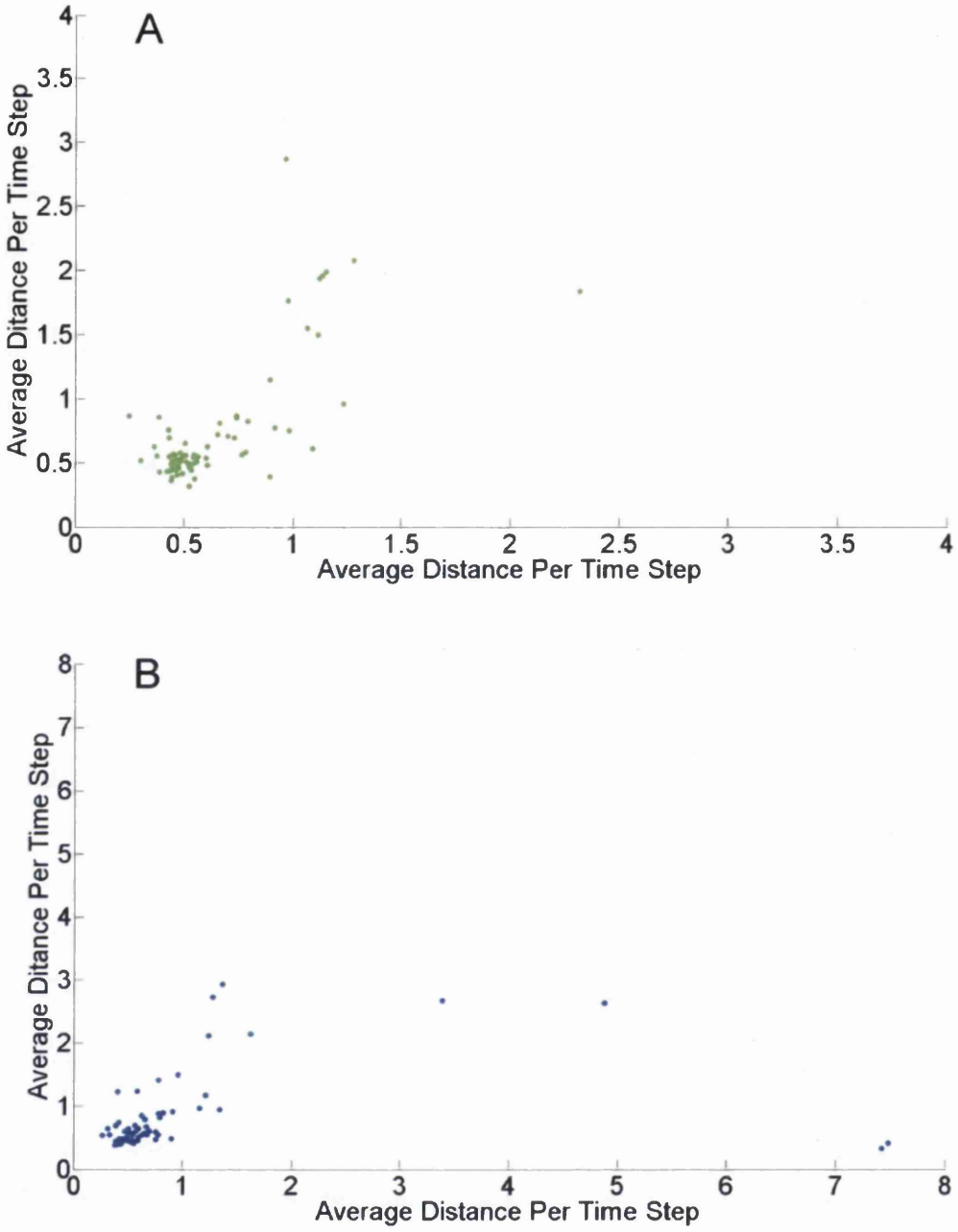



Figure 3.5: Normalized inter-frame migration distances for the 2nd (A) and 3rd (B) generations shown independently of each other and the 1st generation

Another simple measure of motility is provided by means of comparing the displacement from the initial to the final positions in the lineage vectors to provide insight into the long term migration prospects. Here the displacement D is defined as

$$\boldsymbol{D}_{x,y} = \boldsymbol{v}_{x,y}(1) - \boldsymbol{v}_{x,y}(end) \qquad (3.6)$$

and is calculated for each dimension independently and displayed in figure 3.6 with the origin corresponding to the initial position and the generational colour coding remaining consistent with previous plots. This sort of global overview of cell motion could potentially be used as a characterisation of cell lines as differing cell lines exhibit differing tendencies for migration; A549 cells for example tend to migrate around the field of view where as U2OS cells, being endothelial, tend to group together and therefore generate lower migration distances.

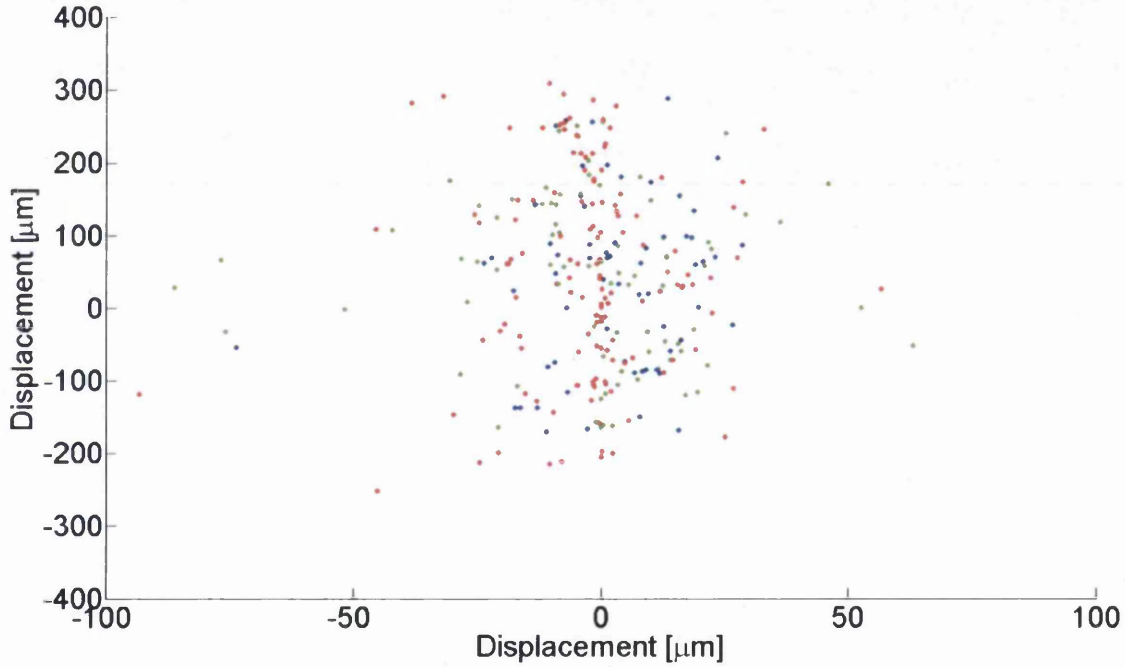



Figure 3.6: Scatter plot detailing the end point of the displacement vector of all identified cell tracks from their respective initial to final frames. Red corresponds to first generation, green to second and blue to third.

Having the cell centroid locations stored for complete lineages offers the opportunity to display a lineage bifurcation point in terms of spatial parameters rather than the temporal visualization shown earlier in figure 3.1. An example of this is shown in figure 3.7 for a lineage bifurcation and the resultant daughter migrations. The initial lineage vector, red, shows the progenitor migrating from region (A) towards the bifurcation point (B). At (B) the cell splits into the north and south daughter cells with the division appearing to occur along a y-axis bisect. The two daughters are then seen to migrate along their own pathways, (CN) and (CS), until seemingly settling in the regions (DN) and (DS). The larger spacing between the north daughter centroid spots around the CN region and the comparative scarcity of the dot spread in the region DN suggest it is more mobile than its southern counterpart.

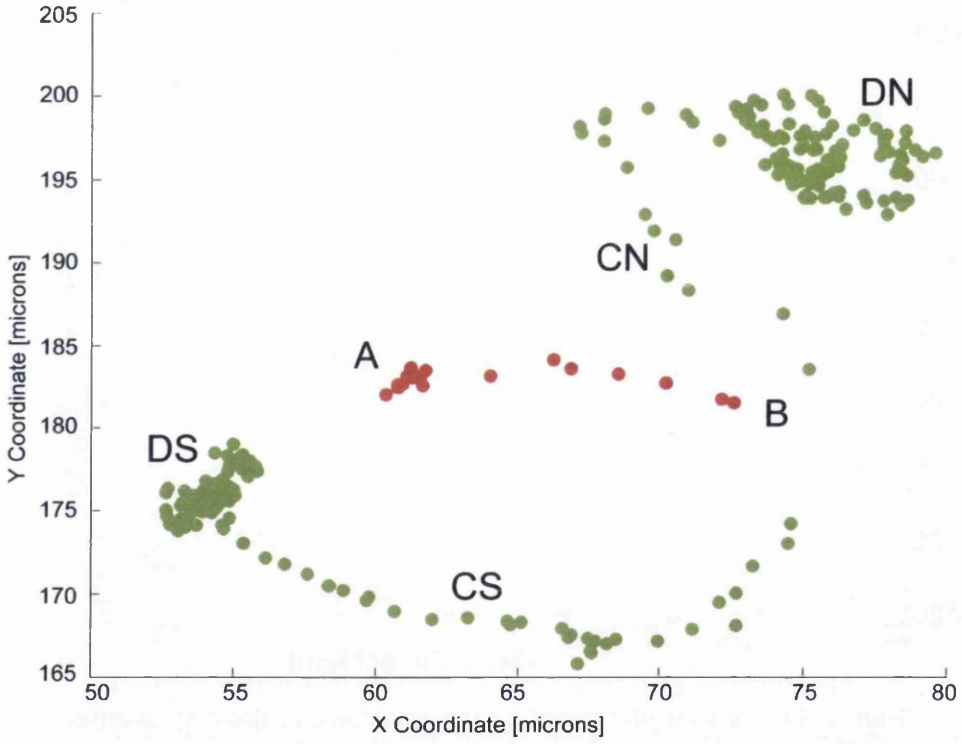



Figure 3.7: Sample scatter plot detailing the migration of a progenitor cell, shown in red, beginning in the region about (A). Around (B) the cell undergoes mitosis and the two daughter cells migrate (CN and CS) away from each other before settling within regions DN and DS respectively.

Such settling could be the result of migration into regions of greater confluence or the result of preparations in the cell cycle prior to further mitosis events. With only a 2 hour interval between the mitosis point (B) and the north daughter reaching region (DN) the likelihood of such behaviour being attributable to mitotic preparations is negligible. The positions of other cell centroids in the area were also non-suggestive of the north daughter being mobility restricted by the presence of surrounding cells. The plot pattern could be interpreted as two daughter cells migrating apart in search of a suitable place to occupy and meet their respective growth needs with the regions (DN) and (DS) representative of these sought locations. Reviewing the position of other cell centroids in the field of view around the region (DS), occupied by the southern daughter, reveal it to be surrounded by a greater number of neighbouring cells. The (DS) region was particularly crowded due to multiple mitosis events in close spatial and temporal proximity. This contrast with the northern daughter could explain the apparent differences in the distribution of points between regions (DN) and (DS) as the cells prepare to divide. The example shown in figure 3.7 also appears to show a noticeable symmetry between the motions of the daughter cells in relation to one another. Observations of other pairings however suggest this to be purely coincidental and that such tandem like motions are not typical.

Displays such as figure 3.7 are useful as a method for investigating and visualising lineages that are defined as outliers by other measurements and could correspond to subpopulations within the global population. Automatically tracking large numbers of lineages and storing a representation of the location of all identified cells provides the potential basis on which to apply analytical methods of characterising group behaviour and cell to cell interactions through time with the potential for creating global parameters to infer details of the inner workings of the cells under observation. The complexity of such methods extend beyond the simplistic approach championed throughout this text and are consequently deemed an area of future development.

# CHAPTER FOUR

# Tri-Colour Labelling

## *Tri-Colour Quantum Dot Loading*

The quantum dot-seeded, moving-light-display concept can be advanced a step further by utilising multiple coloured quantum dots to create cell specific colour codes providing a method of cellular barcoding. The current principle method for creating cellular barcodes for high-throughput analysis utilises fluorescent dyes or proteins either by combining them in differing proportions within different cell groups, labelling samples under differing stimulation conditions with dye concentrations that increase at defined levels or using entirely different substances for different cell groups [64][65][66][67]. These methods are well suited to tagging cell populations as a whole but, along with the photo-toxic effects over longer time periods and unknown effects that modified proteins have when interacting with objects within the intracellular region; they do not readily provide a way to generate individual cell barcodes. Another method creates genetic barcodes in the form of viral DNA which can be applied to virus accessible cell types to track single cells [29]. This procedure is however not readily available outside of the laboratories developing the procedures. Loading cells with multiple coloured quantum dots provides photo-stable and bio-stable cellular barcodes due to the random nature and regional competition of the uptake process. In the data set analysed within this chapter, U2OS cells were sequentially loaded with Qtracker 525, 585 and 705 nm quantum dot nanoparticles and imaged using an InCell Analyzer to acquire 25 separate fields of view that join to form a single 5x5 grid view at intervals of 15 minutes over a period of 5 hours. The sequential loading and washing process provides the singularly coloured endosomes, referred to, from here onwards, as quantum dot markers, necessary for colour coding.

## *Recovering the Quantum Dot Groups*

The main purpose of this chapter's analysis is to investigate the potential of multicoloured quantum dots to create cell specific colour codes rather than using them to track cell lineages as in previous chapters. In accordance with this objective the cell centroids in these image sequences are all manually seeded prior to the quantum dot channel analysis as automated tracking of the cell centroids as done in earlier chapters inevitably incurs some errors; particularly in endothelial cell lines such as the U2OS cell line due to their propensity to exist in close proximity to one another in groups. It is more important here to have the quantum dots correctly assigned to their respective centroids as they are being trialled in the creation of a cell scale identification variable. This contrasts with the previous tracking system where only the general group dynamics were required. Experimental procedures should minimise sources of variation to ensure observations are attributable to the variable under investigation.

Identifying the quantum dots within each colour channel uses the same method used previously in chapter two, namely a simple step function noise filter to isolate the image foreground and a peak finder algorithm to localise the quantum dot signals. The InCell Analyzer makes use of a charge-coupled device, CCD, with a 3 colour filter array to acquire all three channel images in a single pass. This simultaneous acquisition reduces the total image capture time and results in less overall exposure of cells to excitation energy. The use of a single CCD does however incur the penalty of intensity bleeding across the three channel images, a problem that is addressed by an additional filtering step detailed below. Figure 4.1 shows a sample brightfield image, (A), and the result of the quantum dot identification process for the three colour channels (B-D) displayed as scatter plots and coloured accordingly. Whilst a wavelength of 705mn corresponds to red, 585nm to yellow-orange and 525 to green in the light spectrum, the quantum dots are referred to here as red, green and blue respectively[14] in reference to their prospective use in generating RGB colour codes. These identified peak sets can be further refined by applying an additional three stage filter process that utilises set dependent parameters.

[14] Whilst the actual allocation of each quantum dot size to one of the three RGB channels is arbitrary, it is deemed sensible to assign each quantum dot in correct wavelength order. Thus 705mn corresponds to the red channel, 585nm to green and 525nm to the blue channel in the RGB colour-space.

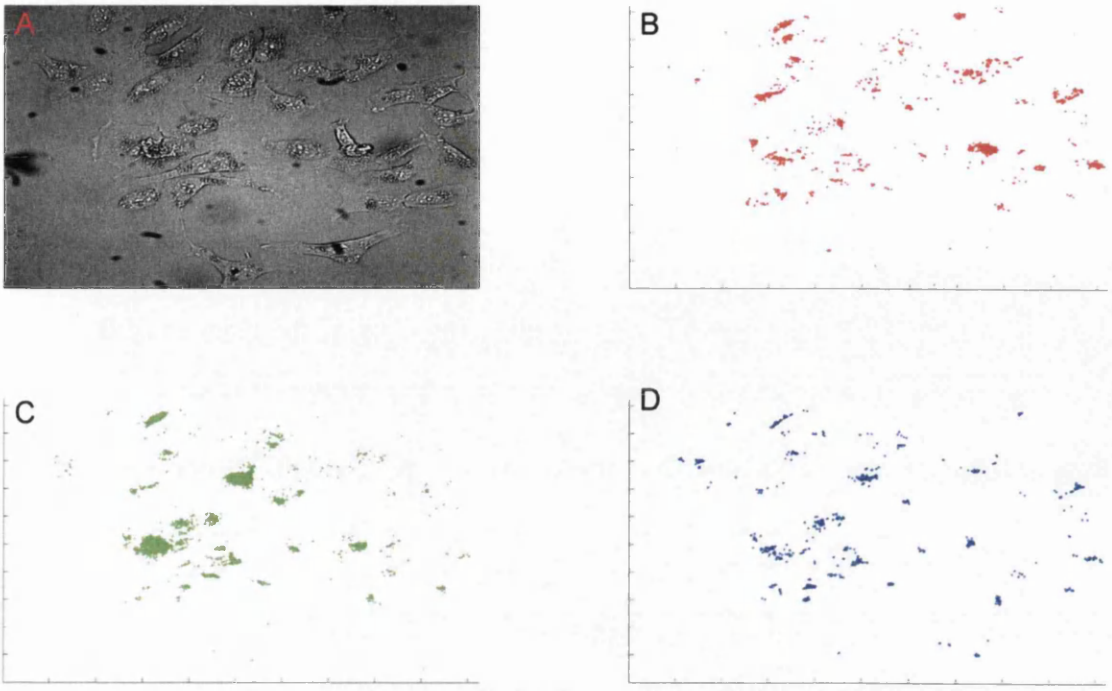



Figure 4.1: (A) Sample full field of view Brightfield image of U2OS cells. (B), (C) and (D) show scatter plots of the quantum dots identified using a simple step filter and peak finder on the red, green and blue channels respectively.

Due to the cell centroids being known a priori at this point, the first obvious filtering method is to discard any peaks that lie beyond a reasonable distance from any centroid locations, defined in this instance as two times the average cell radius, the average being found by manual measurement of 10 cells randomly selected from across all 25 fields of view. This quickly removes any peaks that cannot productively contribute to the colour coding process. The second stage in this additional filter step is to use the intensity distributions of the remaining peaks to create a cut-off below which any identified peaks are discarded. Figure 4.2 (A-C) (*p71*) shows the peak intensity profiles of a sample time frame as histograms coloured according to channel.

The intensity cut-off for each channel is simply defined as the intensity value with the highest frequency. To automate the identification of this value let the intensity set for a given time frame and channel be A and define a vector B such that

$$|B| = \max\{A\} - \min\{A\} + 1. \tag{4.1}$$

Each element in B corresponds to a bin value for the intensities in A with

$$B(n) = \min\{A\} + (n-1) \tag{4.2}$$

for the integer n where $1 \leq n \leq |B|$. Define an additional vector, C, to be the same size as B and sort the entries of A into the vector C in accordance with the corresponding values of B. The intensity cut-off is then simply the value in the element of B corresponding to the element containing the maximum value within C.

The third stage deals with intensity bleeding across the channels by comparing the QDot coordinates of all three colour channels to identify points that exist in multiple channels. As an example, comparing the red and green channels, let the set $C_R = \{X, Y\}$ be the red channel QDot coordinates and similarly the let set $C_G = \{X,Y\}$ be the green channel QDot coordinates. The bleed through sets are defined as

$$B_{R\rightarrow G} = C_R\{X_i, Y_i\} \; where \; C_R\{X_i, Y_i\} \in C_G\{X, Y\}$$

$$B_{G\rightarrow R} = C_G\{X_j, Y_j\} \; where \; C_G\{X_j, Y_j\} \in C_R\{X, Y\} \tag{4.3}$$

where $1 \leq i \leq |C_R|$ and $1 \leq j \leq |C_G$. Peak coordinates in B exist in both red and green channels due to channel bleed through and are assigned to the channel with the highest intensity value at that point. So for the (x, y) coordinate set in $B_{R\rightarrow G}$, the red channel image is defined as a matrix of intensity values $IMG_R$, the green channel as $IMG_G$, and the intensities of the conflicting QDot locations as

$$I_R = IMG_R(x, y) \; and$$
$$I_G = IMG_G(x, y). \tag{4.4}$$

The intensity difference is then simply

$$\Delta I = I_R - I_G. \tag{4.5}$$

Where ΔI is negative, the corresponding coordinate set in $B_{R\rightarrow G}$ is removed from the parent set $C_R$ and where ΔI is positive, the corresponding coordinate set in $B_{G\rightarrow R}$ is removed from its parent set $C_G$. This process is iteratively applied to all channel combinations each time using the most up-to-date parent lists. This step is applied last as most of the bleed effect is dealt with by the previous filtering step where low

intensity peaks below the distribution maxima are removed. As stage two is a bulk process and the resultant reduction in remaining QDots with signals in multiple channels increases the speed of the set membership calculations in equation 4.3, performing this step last optimises the speed of the filtering process.

Removing all peaks with intensity less than the cut-off results in the distributions shown in fig 4.2 (D-F). In this distribution sample the filtering process had seemingly no effect on the red channel but the combination of distance filtering and intensity cut-off reduced the number of peaks from 951 to 609.

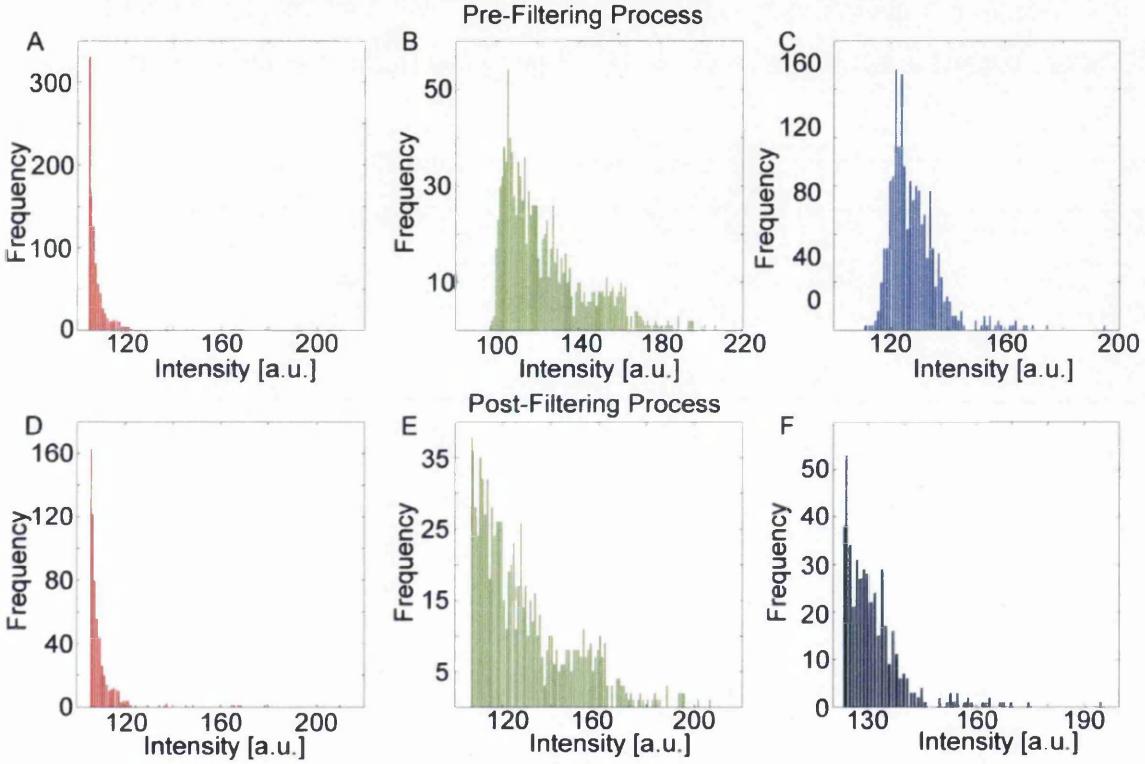



Figure 4.2: Sample quantum dot intensity profiles coloured in accordance with their respective channels. (A-C) show the intensity profiles of the red, green and blue channels respectively prior to application of the filtering process with (D-F) showing the resulting profiles after filtering.

The reduction in the green and blue channels of 1141 to 898 and 632 to 459 respectively also demonstrates how the second filtering process reduces the variability between channels in identified quantum dot peaks levelling the numbers which provides a better foundation for creating the colour codes. This effect also carries over into the temporal plane where after filtering the variation in the number of identified peaks from one frame to the next is subdued. Throughout this additional filtering process an essential factor is also preserved, namely that any effects seen in

reduction of the identified peak numbers affects all three channels equally providing no indication of over estimation or prejudice towards any channel. The results of the filtering process are shown in fig 4.3 with comparison between pre-filtering (A-C) and post-filtering (D-F) for each of the three channels. The reduction in numbers in areas of higher density is difficult to see on the full scale scatter plots but the sparsely distributed rogue points in plots (A) through (C) can be seen to have reduced significantly in plots (D) through (F). Figure 4.4 (opposite) provides a zoom in on a cell doublet in the brightfield with the QDot associated peaks overlaid in their respective colours, (A) before filtering, (B) after. The reduction in the number of blue markers within the left hand cell is visible as well as the resulting thinning of the red marker density within the right hand cell.

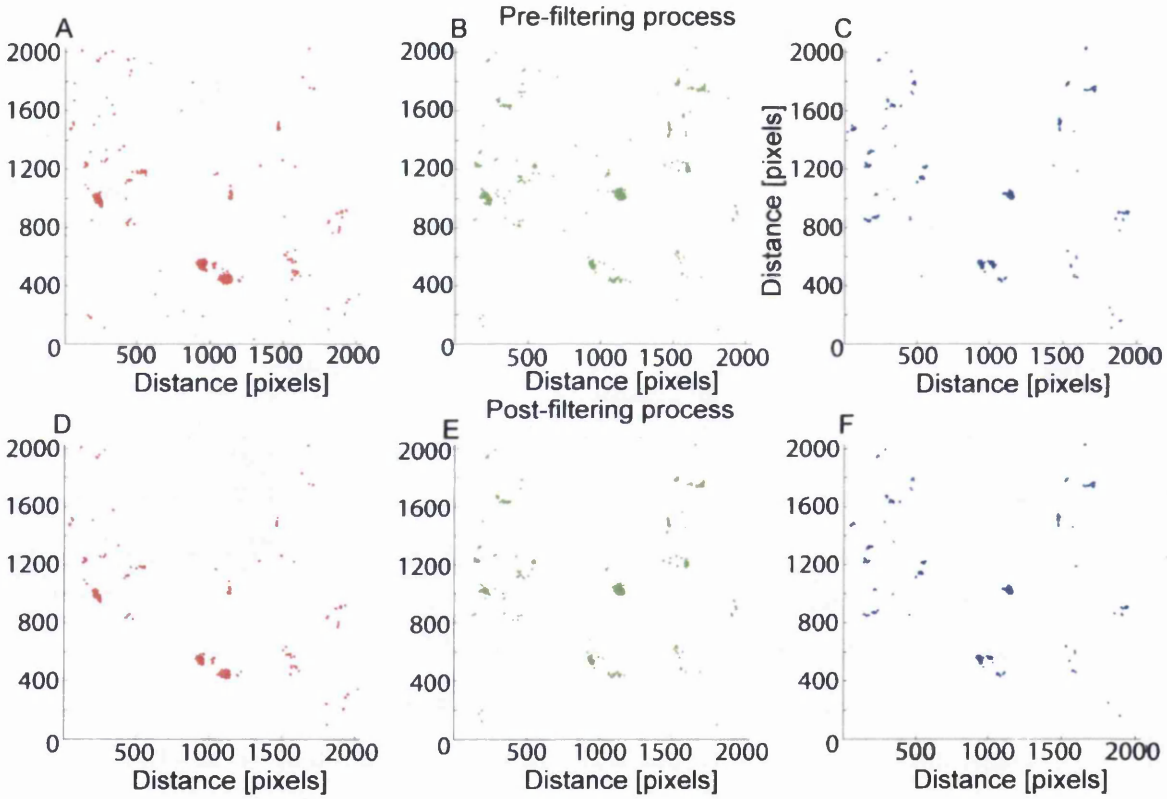



Figure 4.3: Scatter plots comparing pre and post filtering quantum dot spatial distributions. The pre-filtering channels (A-C) show outlier quantum dots and higher numbers within clusters.

The data set, as mentioned, contains 25 fields that form a 5 by 5 grid resulting in a complete field of view of 10,240 by 10,240 pixels. The ability to be able to stitch all these fields together to produce a single large brightfield of view with all the quantum dot locations overlaid would be highly favourable.

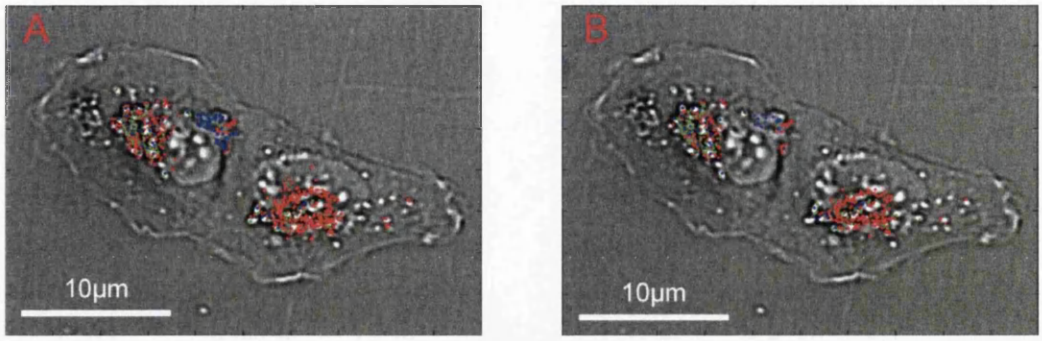



Figure 4.4: Brightfield image of cellular pair with the identified quantum dot markers scatter plots overlaid. (A) shows the pre-filtered distributions and (B) the post-filtered results. The number of blue markers on the left hand cell shows a marked decrease and the density of red markers in the right hand cell can be seen to have thinned.

In reality however this is rather impractical as rendering such an image would be a hardware intensive process and the resulting image of little use as the image resolution would be an order of magnitude larger than that of a typical monitor. Consequently some alternative ways to view the full field are shown in figures 4.5 and 4.6. Figure 4.5 (overleaf) contains a representation of a proximity map coloured in reference to the proximity of each pixel to the locations of the identified QDots and each image is coloured in accordance with its respective colour channel, (A) being red, (B) green and (C) blue. This graphical display provides a way to visualize the quantum dot density patterns across the full field of view that can be still be seen when viewed at a presentable size. The images show how each channel has a similar global distribution which is to be expected as the cells were equally exposed to all three quantum dot types but closer inspection demonstrates the local variations. (see accompanying media for uncompressed images) Figure 4.6 (*p76*) contains a scatter plot of all three channels on the same axes for the full 25 field range and is presented horizontally in an attempt to increase its visibility (an uncompressed image is provided in the accompanying media). It serves to show that cells on the whole take up all three colours to some degree with sporadic occurrences of single colour clusters. It also shows the expected grouping behaviour of U2OS cells with island patches of QDot clusters.

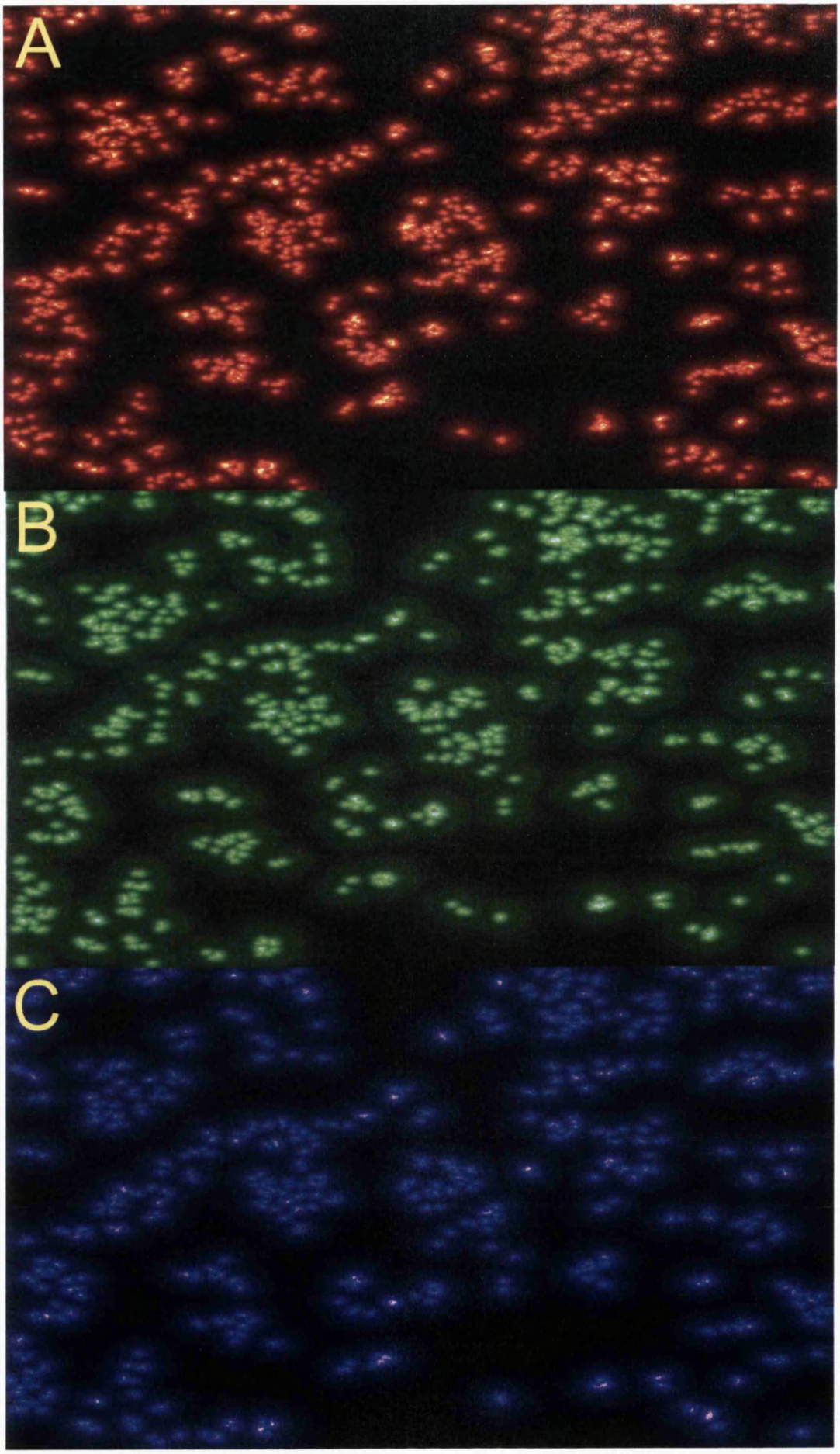
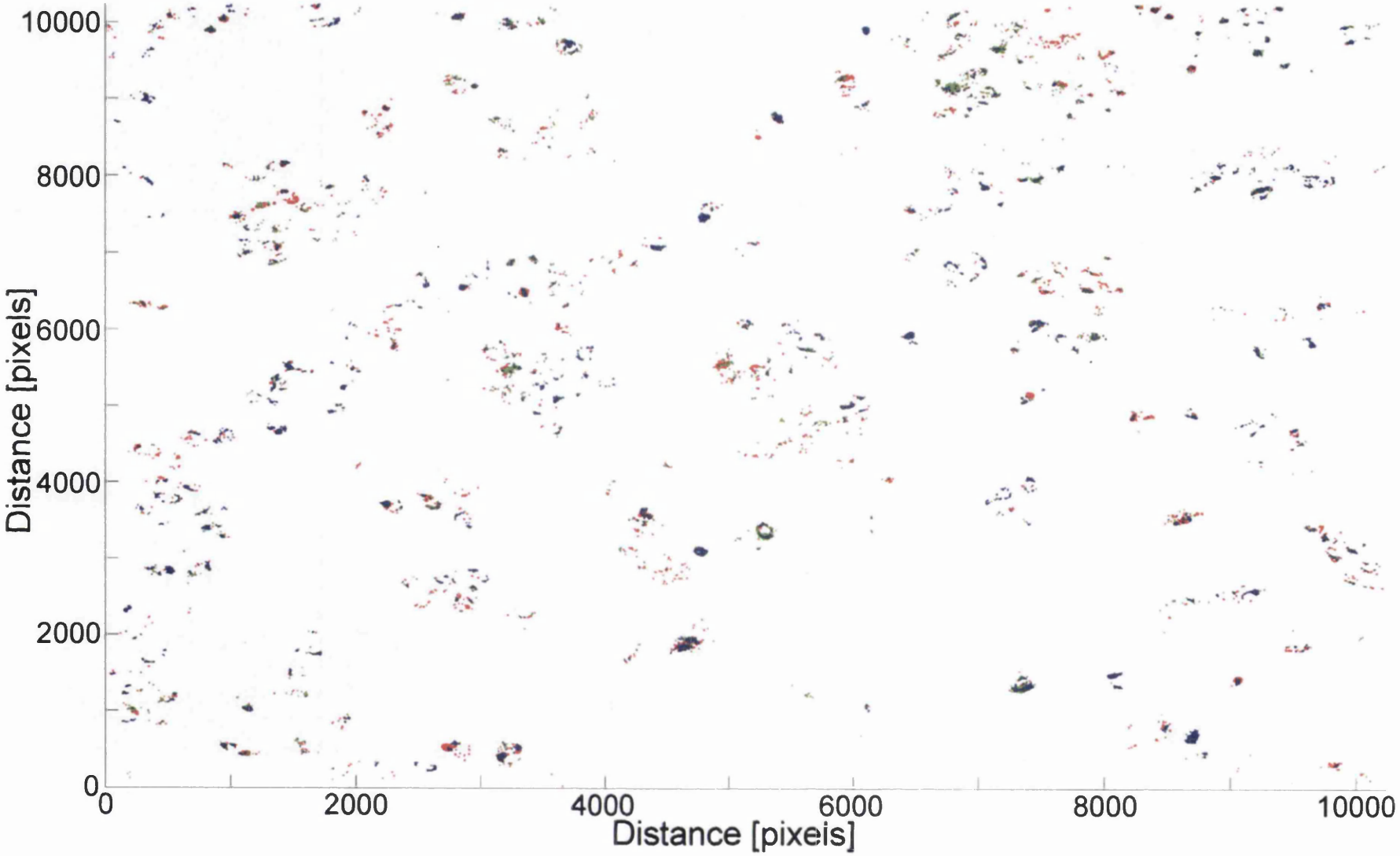

A
B
C

Figure 4.5 (opposite): Graphical display of all 25 fields of for each QDot colour channel, after the filtering processes, artificially coloured accordingly to emphasize the distribution patterns. The differences in these patterns across the three channels can be seen by visually comparing regions. (Full size images are presented within the accompanying media)

Figure 4.6 (overleaf): Complete 25 field view of the three QDot channels displayed as a scatter plot. Practical limitations do not allow for a bright field overlay for the full 25 field composite image. (Displayed here horizontally for maximum visibility, a digital version is provided on the accompanying media)


10000
8000
6000
4000
2000
0
Distance [pixels]
0
2000
4000
6000
8000
10000
Distance [pixels]

## *Centroid Barcoding*

With the quantum dots identified and their locations stored they can now be used as the basis for a unique cell coding system. With the centroids also known the task of assigning dots to cells is simply a case of nearest neighbour analysis. Matlab Image Processing Toolbox provides a useful tool for this task in the form of the function *bwdist( )* which calculates the distance between pixels set to 0 and their nearest nonzero pixel for binary images. Use of this function requires the centroid locations to be represented in a binary matrix which is readily done by defining a logical matrix M of the same size as the source data files, namely 2048x2048, with default values of 0. If the centroid (x, y) coordinates are vector lists labelled $X_C$ and $Y_C$ for their corresponding axis then the desired centroid binary representation is defined as

$$M(Y_C, X_C) = 1. \qquad (4.6)$$

Here the $Y_C$-axis equates to matrix rows and $X_C$-axis to columns and the resulting matrix M, containing 1's at the centroid coordinates, can now be passed onto the *bwdist()* function. Samples of the two outputs of this function, default labelled D and L for Distance and Label, are displayed in fig 4.7. (A) maps the Euclidian distance of each pixel to its nearest centroid in gray-scale and (B) the label matrix where each pixel whose value in the matrix M was false, takes the value of the index of the nearest pixel whose value in M was true. In this case that corresponds the index of each pixels nearest centroid.

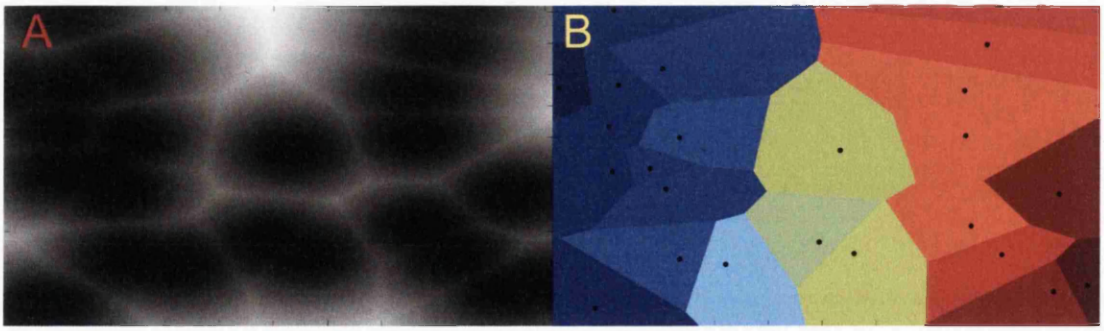



Figure 4.7: Sample output of Matlab's *bwdist()* function. (A) A proximity map, darker equates to closer proximity to a centroid location. (B) A label matrix plotted in the form of a Voronoi diagram to highlight regions.

To assign the QDots to their nearest centroid, consider an example set of coordinates from the red channel for the field F containing the centroid set C. Set the vector $X_Q$ as the full list of x-coordinates and similarly the vector $Y_Q$ for the y-coordinates of each identified red channel QDot in field F and generate an index vector IND, defined as

$$IND(n) = n \qquad (4.7)$$

for the integer n where $1 \leq n \leq |X_Q|$ (or $|Y_Q|$ as they are equal). Then the location of the nearest centroid in the set C for each QDot point is simply

$$C_L(IND) = L(Y_Q(IND), X_Q(IND)) \qquad (4.8)$$

where L is the afore mentioned label matrix, an example of which is shown in fig 4.7 (B). The resultant vector $C_L$ contains references, in the form of matrix indices, to one of the centroids within the set C for each entry in the coordinate set ($X_Q$, $Y_Q$). The vectors [$X_Q$, $Y_Q$, $C_L$] now encode the QDot locations and their parent cell centroid location. Extraction of the QDot groupings is simply a matter of identifying the rows in [$X_Q$, $Y_Q$, $C_L$] with the same values of $C_L$, which are indices corresponding to centroid locations, and saving them into an appropriately titled arrays for each centroid coordinate pair in the set C. Performing these actions for each channel successively results in a data set containing the quantum dot allocations for each centroid in each channel from which a cell-specific 3-digit barcode can be created by counting the number of QDot optical markers in each channel that are now associated with the cells centroid. A centroid associated with $N_R$ red, $N_G$ green and $N_B$ QDot markers can be labelled with the barcode [$N_R$, $N_G$, $N_B$]. Figure 4.8 (opposite) shows a sample brightfield field of view image with centroids, coloured using their respective barcodes, overlaid. This visualization is simply done by scaling all the barcodes within the field to the field maximum, $N_{max}$, which is the highest barcode entry for each channel component across all barcodes and acts as the ceiling to the colour range. The RGB colour for a given barcode is then

$$[R, G, B] = \frac{1}{N_{max}} \times [N_R, N_G, N_B] \qquad (4.9)$$

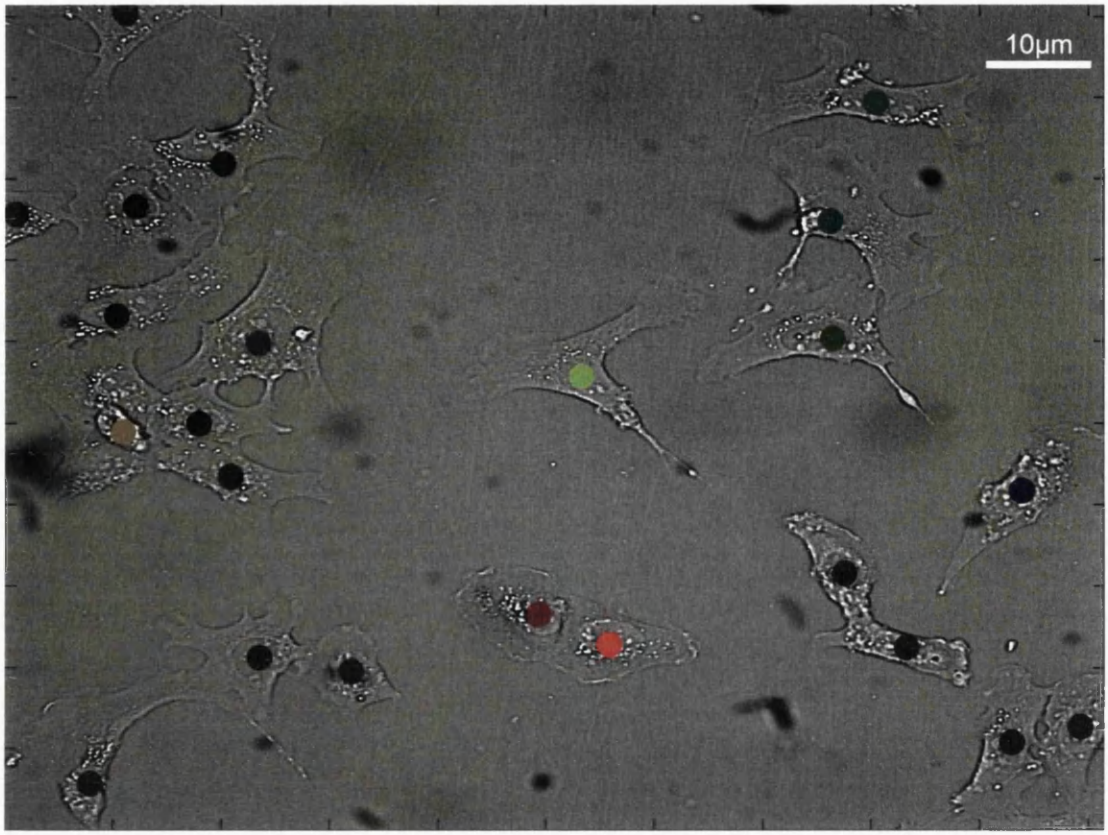



Figure 4.8: Sample brightfield field of view with colour coded centroids overlaid. The colour codes are created using the barcodes created by counting the identified quantum dot markers.

The problem with this simple visualization is that the three barcode parameters $N_R$, $N_G$ and $N_B$ can vary across the field by similar amounts but said variation occurs in different ranges meaning equal changes in the values in different channels are not apparent in the respective output colour. For example the variation of +2 in the red channel between two centroid barcodes with $N_R$s of 5 and 7 is significant to the red component of the colour but the centroid colour may look essentially the same if the green and blue components of each barcode are 50 and 60 respectively; the relative colour difference will be a variation of the hue of a blue-green colour. This effect can be seen in fig. 4.8 where a glancing look identifies four green cells in the upper right quartile, an olive and some dark greens in the upper left quartile, a red and brown at lower middle and the rest appear black. The dark coloured centroids are however shades of blues and purples but are lower in total QDot count than the brighter colours and as such appear to the human eye as essentially black. This does not detract from the barcode as a cell-specific marker as the variation in numbers is

directly quantifiable, for visual identification however the colour codes need to be adjusted to fill the entire colour-space in each of their three components, this issue is addressed later in this chapter. Figure 4.9 contains an image sequence of a single cell with overlaid centroid coloured as previously described covering a time span of 5 hours to visually demonstrate the stability of the barcode through time.

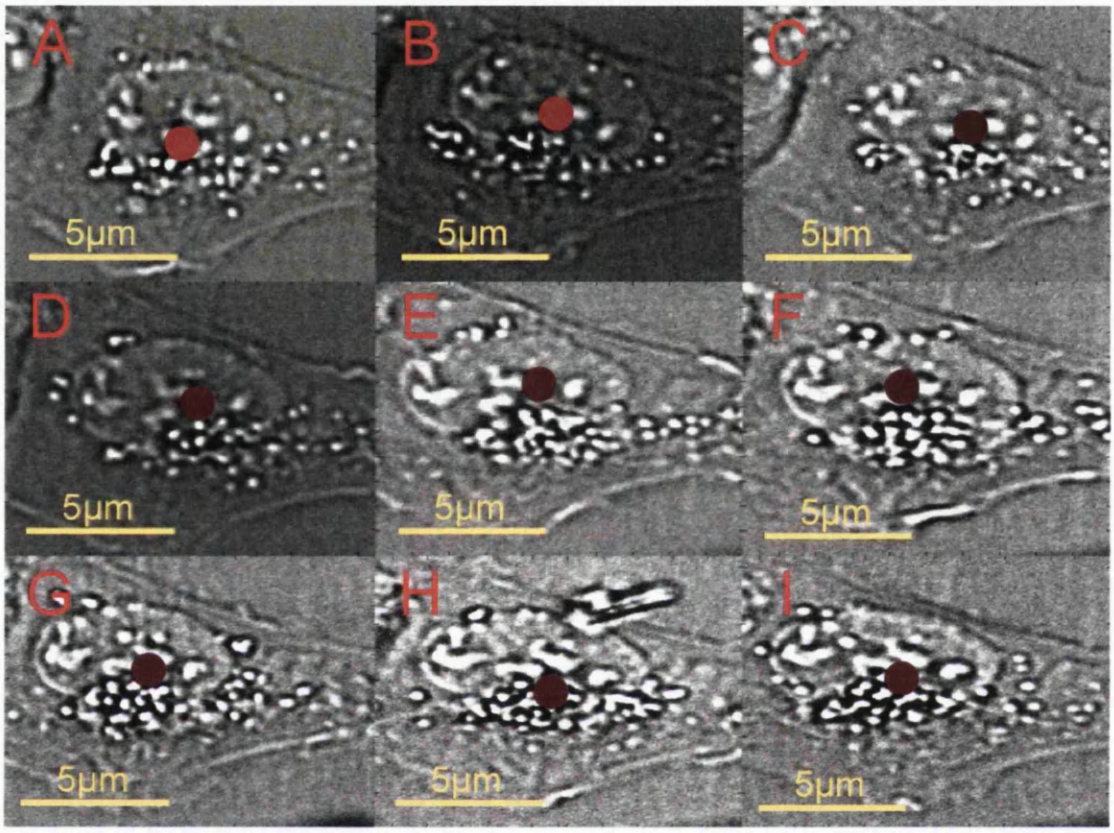



Figure 4.9: (A-I) Sequence following a single cell, interval between images is 30 minutes. Overlaid centroid colour coded using the quantum dot count directly. This visually demonstrates the temporal stability of the barcodes generated using the quantum dot markers.

*Uniqueness*

A cell barcode is only useful as a cell identifier if it is sufficiently different from other cell barcodes, particularly those in closer proximity, to remain unique through the inevitable variations in quantum dot number counts that occur as a result of internal cell dynamics and changes in fluorescence of the quantum dots over time. In an effort to quantify the uniqueness of the generated barcodes a uniqueness parameter is defined as the variation required across all channels of a barcode A such that all the values of its most similar barcode B exist within the said variation of A. The quickest way to attain such a parameter is to define a barcode set F = {$b_1$, $b_2$,... ,$b_n$} where $b_n$ = [R, G, B] are the barcodes for all cells in all fields and generate a difference matrix, D, defined as

$$D(i,j) = \begin{cases} infinity & if\ i = j \\ \max(|b_i - b_j|) & if\ i \neq j \end{cases} \quad (4.10)$$

where $1 \leq i, j \leq |F|$. Each row, and by virtue of symmetry each column, of the matrix D corresponds to the barcode elements in set F and each element in D stores the maximum variation across the three channels between barcode $b_i$ and $b_j$. As an example for the purpose of clarification, let three example barcodes be defined as $b_1$ = [9, 14, 37], $b_2$ = [15, 0, 12] and b3 = [0, 19, 5].

Using equation 4.10, D becomes;

$$D(1,2) = max(abs([9,14,37] - [15,0,12]))$$

$$D(1,2) = \max(abs([-6,14,25]))$$

$$D(1,2) = \max([6,14,25])$$

$$D(1,2) = 25$$

Similarly,

$$D(1,3) = D(3,1) = 32$$

$$D(2,1) = 25$$

$$D(2,3) = D(3,2) = 19$$

$$D(1,1) = D(2,2) = D(3,3) = \infty$$

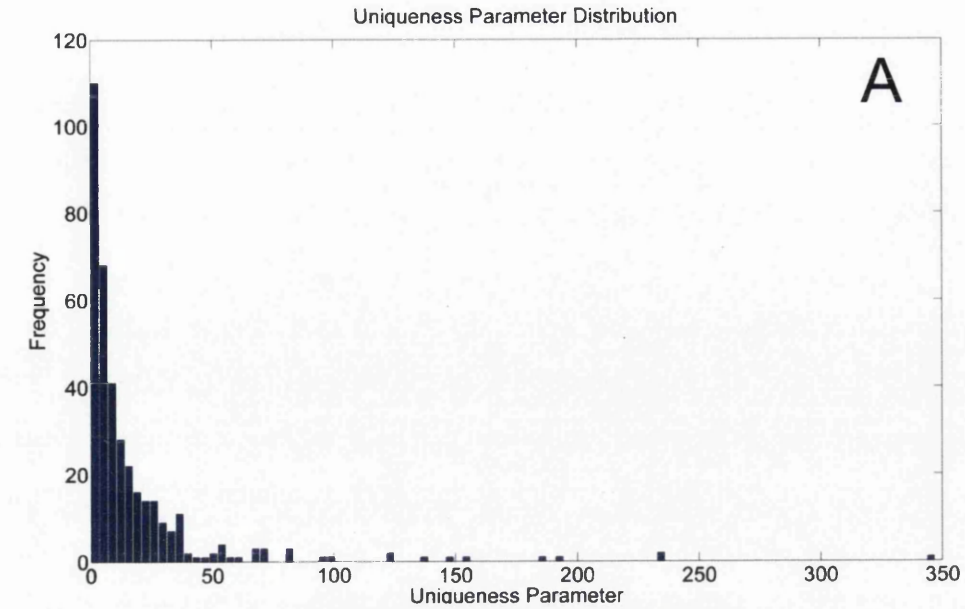
and resultant D matrix looks like

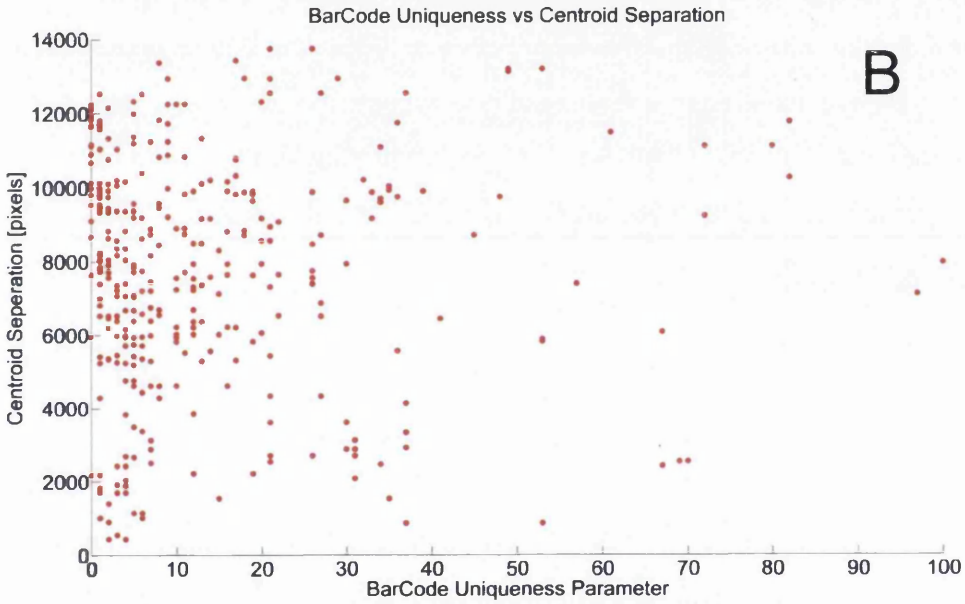
$$D = \begin{matrix} \infty & 25 & 32 \\ 25 & \infty & 19 \\ 32 & 19 & \infty \end{matrix}$$

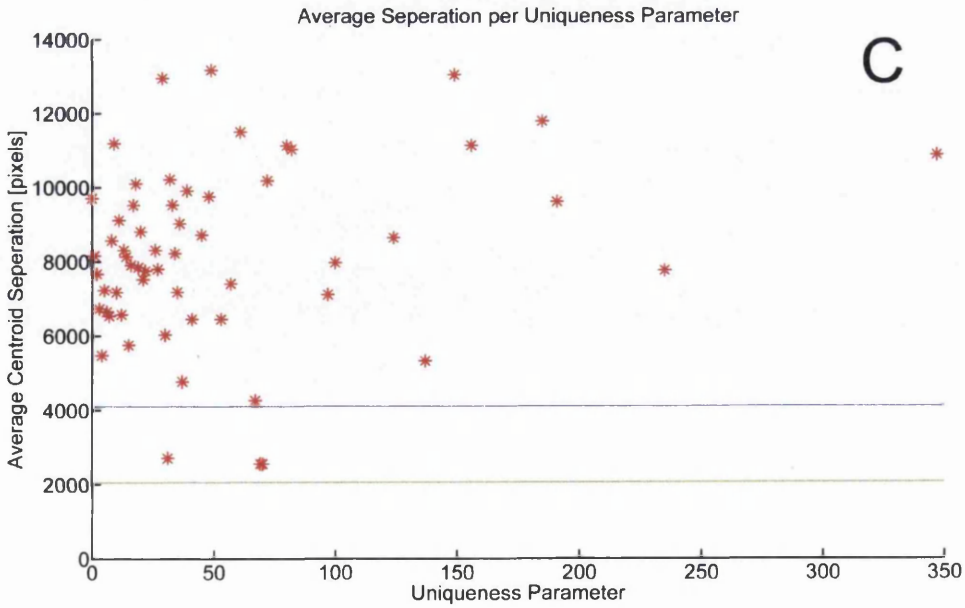
In this small example set the minimum of the first row is 25 in the second column making $b_2$ the most similar to $b_1$ in this set and the uniqueness parameter, UP, for $b_1$ is 25. The minimum on row two is in column three, consequently $b_2$ is most similar to $b_3$ with a UP of 19 and the third row shows that b3 is most similar to b2 with the same UP value of 19. The UP value can be thought of as a margin of error beyond which a barcode is no longer unique.

Figure 4.10 (A) (opposite) shows a histogram distribution of the UP values showing some outliers with very unique barcodes and high UP values. Looking at the UP alone however is not particularly meaningful as two cells with identical barcode values could exist but the issue of confusing them for one another only occurs in the event that they occupy spatial positions in close proximity. As the 25 fields collected form a single large scale image the distances between all centroids can be calculated and it is useful to include the separation of the most similar pairs in an examination of the measure of uniqueness. Figure 4.10 (B) plots the UP values against their respective separation distances and the density of points on the left hand side shows how the cells with lower UP values are generally further from their nearest match than those with higher uniqueness. Figure 4.10 (C) plots the same variables against each other with a single point for each UP value plotted, the single point being the average separation off all the cells with the same UP number. This offers insight into the denser region of points shown in (B) and shows how the majority of low uniqueness cells are separated by enough distance to remove the issue of confusing them with one another. The green line on (C) is plotted to indicate the size in pixels of a single field of view, the blue line to indicate two fields of view.

Figure 4.10 (opposite): Graphical representation of the uniqueness parameter. (A) shows the uniqueness parameter distribution, (B) the uniqueness parameter against the distance to its most similarly coded centroid and (C) the average distance to the same.


Uniqueness Parameter Distribution
A
Frequency
Uniqueness Parameter
0
20
40
60
80
100
120
0
50
100
150
200
250
300
350


BarCode Uniqueness vs Centroid Separation
B
Centroid Seperation [pixels]
BarCode Uniqueness Parameter
0
2000
4000
6000
8000
10000
12000
14000
0
10
20
30
40
50
60
70
80
90
100


Average Seperation per Uniqueness Parameter
C
Average Centroid Seperation [pixels]
Uniqueness Parameter
0
2000
4000
6000
8000
10000
12000
14000
0
50
100
150
200
250
300
350

## *Centroid Colour Coding – Visual Adjustment*

As previously attested, for the purpose of visually differentiating cell colour codes their components need to be scaled to fully utilise the colour space available. This process is purely to improve the spectral variation in the colour codes for the benefit of the human eye and as such there is no requirement to preserve the relative inter-channel relationships of each colour code set only to apply any re-scaling process consistently and with equal measure to all three component channels. The colour re-scaling process makes use of histogram equalization routines to linearly redistribute each colour code component across a range that for each channel is lower bounded by zero and upper bounded by the maximum value of said component across all barcodes. For a given field the complete list of barcodes is placed into an Nx3 matrix, M, where N is the number of cell groups within the field and each column corresponds to each of the colour channel components in the N barcodes. The distribution sample in fig. 4.11 (A) is created by binning each column of the compiled barcode list into a vector, $B_C$, where C is an index reference to the three colour channels and

$$|B_C| = \max(\mathrm{M}(:, C)) - \min(\mathrm{M}(:, C)) + 1. \qquad (4.11)$$

Each element of $B_C$ acts as an integer bin of width 1 spanning the range of each respective colour component and the $B_C$ vector for the red component of a sample field is plotted in fig 4.11 (A). Histogram equalization seeks to linearise the cumulative density of a histogram distribution and was originally defined for use on gray-scale images. While it can be applied to red, green and blue components of an RGB image separately it can affect the colour balance if not converted to an alternative colour space such as CIE 1976 (L*,a*,b*) or HSV. As the objective here is merely to widen the spectrum of colour for visual differentiation where consistency is the key factor it is sufficient to simply apply the algorithm to each channel separately. Consequently, the re-distribution of $B_C$ is then simply a matter of applying the general histogram equalization equation for each colour channel, which with the variables defined here takes the form

$$M_R(i, C) = round\left(\frac{D_C(i) - \min(D_C)}{N - \min(D_C)} \times \left(\max(\mathrm{M}(:, C)) - \min(\mathrm{M}(:, C))\right)\right) \qquad (4.12)$$

where C is again a colour channel index and N is the number of cell groups. The resultant re-distribution is shown in fig 4.11 (C) and its respective cumulative distribution is shown as the blue line in fig 4.11 (B). Comparing (A) and (C) shows how the equalization process provides a wider range of red component values whilst maintaining the range. With all three channels adjusted to better fill the available colour space the adjusted [R,G,B] values are normalized using equation 4.6, the results of which are plotted in fig 4.12 (overleaf).

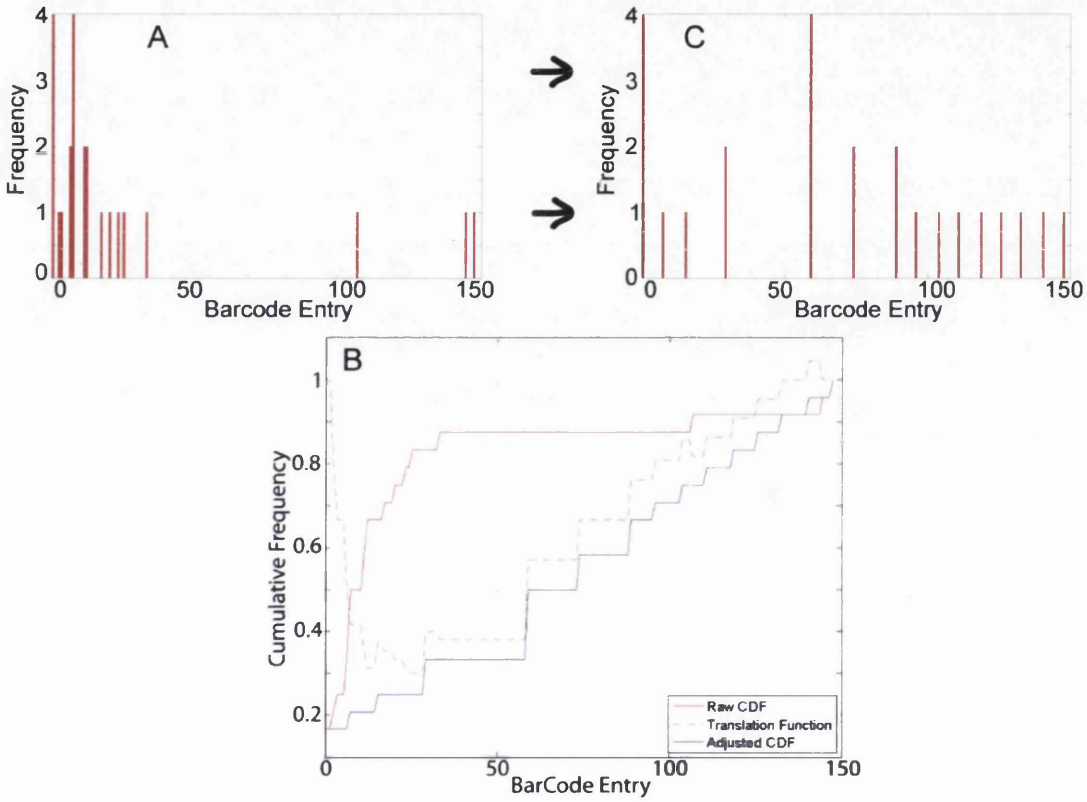



Figure 4.11: Example of the process of re-scaling the barcodes to create centroid colour codes that utilise the full range of the colour. (A) shows the intensity profile of the identified red channel QDots for a single field of view. The translation process is graphically portrayed in (B) resulting in the rescaled distributions in (C). This translation process spreads the distribution in (A) across the available range to create the distribution in (C).

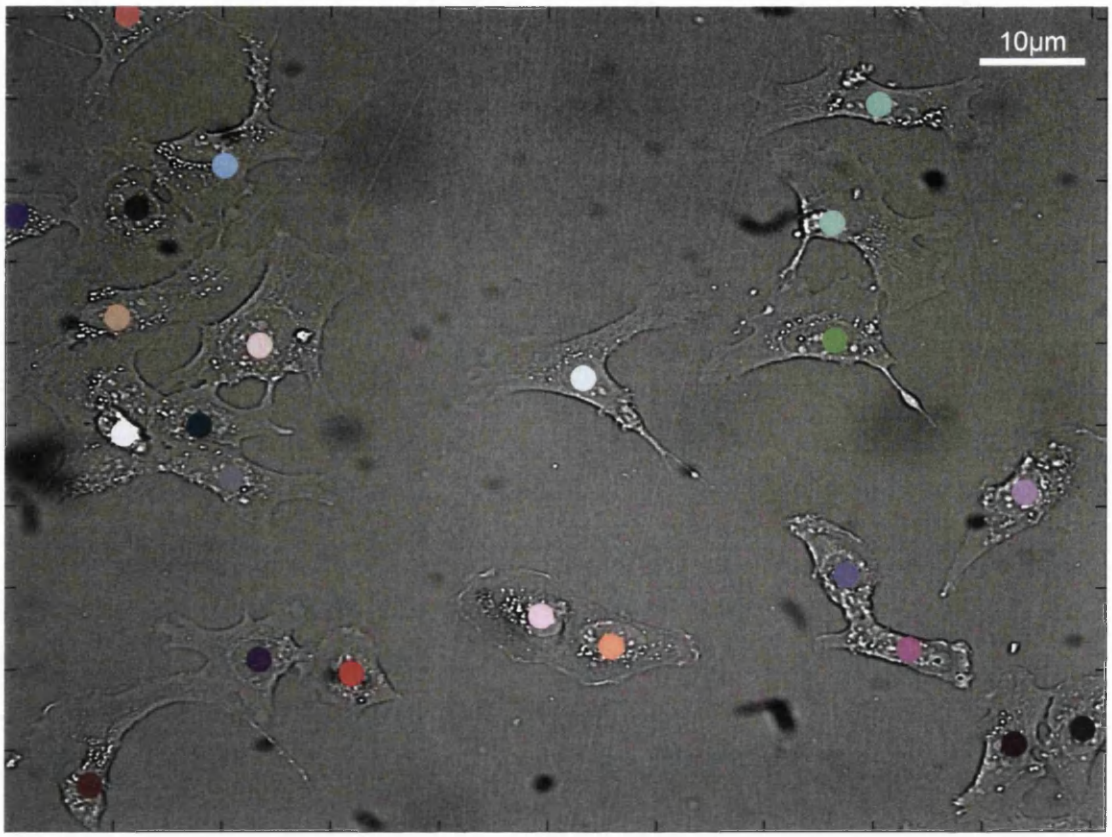



Figure 4.12: Sample brightfield image of a field of view with coloured centroids overlaid. The colour codes used are adjusted to fill the colour space.

The field of view shown is the same as that in fig 4.8 and comparisons of the two clearly show how the artificial colour adjustment visually accentuates the differences between the barcodes assigned to each cell groupings.

## *Temporal Colour Stability*

Unique-to-cell barcodes present an opportunity to identify cells whose position has significantly altered from the last known position therefore increasing the acquisition interval needed to track a cell's progression. In experiments seeking population parameters over longer time periods, where close to real-time dynamics are not the area of interest, this reduces the number of necessary acquisitions and lessens the exposure time of the cell sample to photonic excitation sources. Figure 4.13 shows some sample images of cells in the initial and final frames of the acquisition range.

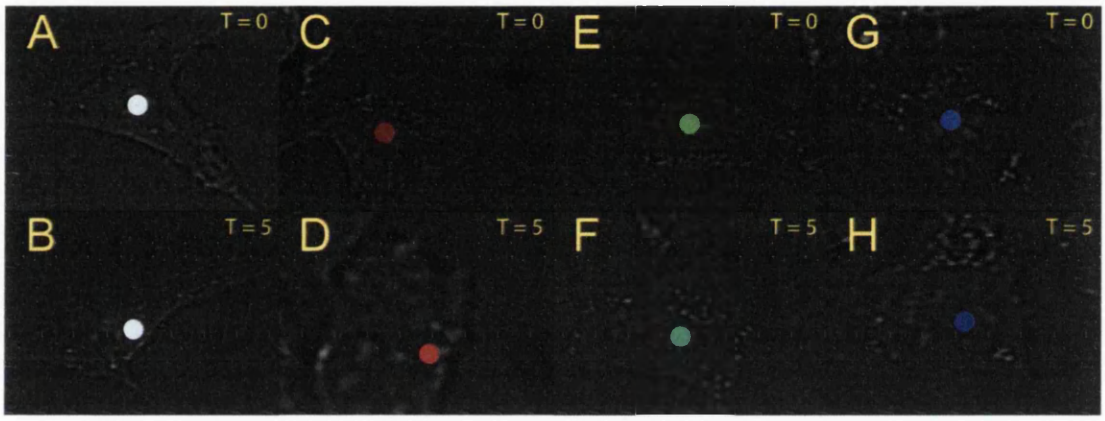



Figure 4.13: Image pairs (A-B), (C-D), (E-F) and (G-H) comparing the initial frame (top row) with the final frame (bottom row) for a selection of colour types. Each centroid can be seen to retain the same hue whilst deviating slightly in colour from their original.

Image pairs, (A-B), (C-D), (E-F) and (G-H) show how the quantum dot markers can create cell barcodes that remains consistent enough to retain the same centroid hue over a 5 hour period with the inevitable variations shown in the subtle changes in the shade of colour seen. The (A-B) colour shade change is not obvious due to the general white hue, (C-D) shows a brighter shade after 5 hours, (E-F) becomes a higher contrasting green and (G-H) shows a marked darkening. Variations in hue do occur through the time frames as a result of changes in the quantum dot marker distributions due to cell motions, close proximity cell interaction, mitosis events and field of view boundary effects. Some of these effects would be lessened by taking sets of images around certain time points to gain barcode averages of cells whilst the variation in cell location is minimised and retain the advantages of less photonic

exposure by leaving the interval between such sets for longer periods. For example, a 12 hour timelapse experiment with 15 minute acquisition interval could instead take three images at 5 minute intervals every two hours, reducing the exposure time by 1/3, or two images with a 10 minute interval every 2 hours for a 5/9 reduction. The exact extent to which these cell specific colourings remain stable over time using the above method is as always dependent on the cell line used. The data analysed here observes the U2OS cell line which, being of an adherent epithelial morphology, typically exist in regional groups in close proximity to one another. This increases the possibility of miss assignment of quantum dots to their respective parent cells during the nearest neighbour process previously described as cells in close proximity often contort to fill the available space resulting in uncharacterised shapes; a problem that increases with confluency until a level forcing a consistent shape pattern is reached. Such issues are lesser in cell lines that actively seek empty regions and separate from neighbouring cells provided sufficient space exists to do so, such as the A549 cell line observed in chapter two. The barcode stability shown in fig 4.13 is therefore all the more striking given that it is yielded from a less than idealised testing ground of the epithelial U2OS cell line.

One insurmountable cause of colour variation comes as a result of cells lying on the boundary of the field of view, a recurrent issue in the analysis of imaged cells. Without the full view of a non-symmetric cell such as the U2OS it is not possible to decipher its complete size and shape, only that it is present. Similarly, if part of a cell is concealed beyond the field boundary then the true barcode values cannot necessarily be ascertained as the unseen region may contain quantum dot markers. An example of such an occurrence is shown in figure 4.14 (opposite) in an image sequence spanning the full 5 hour range with (A-K) at 15 minute intervals and (L) sampled from the final frame for comparison. In (A) a cell can be seen to exist on the left boundary of a field of view where visual extrapolation suggests a sizable region is unseen. The resultant centroid colour using the quantum dot markers in the visible region shows as a reddish brown hue. Following the sequence shows the significant change in colour as a result of the cell migrating into full view and revealing the complete quantum dot marker set attributable to it and by (C) the centroid has seemingly stabilized into a markedly different green hue. (D-E) provide a sample of the variations in hue seen in cell centroids over time as quantum dot markers, being

internalised QDot loaded endosomes, move around the cytoplasmic region and sporadically disappear from view under various objects such as other quantum dot markers, the cell nucleus or simply debris on the lens. This typically only subtly affects the colour shade but does on occasion slightly alter the centroid hue. This effect could be lessened by taking shorter time interval averages as described previously. The (C-K) sequence is however predominantly subtly changing shades of green and the image in (L) is provided as the final frame colour for comparison, shown to retain the green hue. Slight variations in colour are only significant in terms of identifying cells after long acquisition intervals if there are cells in the same region with similar centroid colourings with which to confuse them. The uniqueness parameter detailed previously indicated this to be unlikely as cells with similar barcodes typically existed at distances sufficient to render the prospect of such a migration implausible.

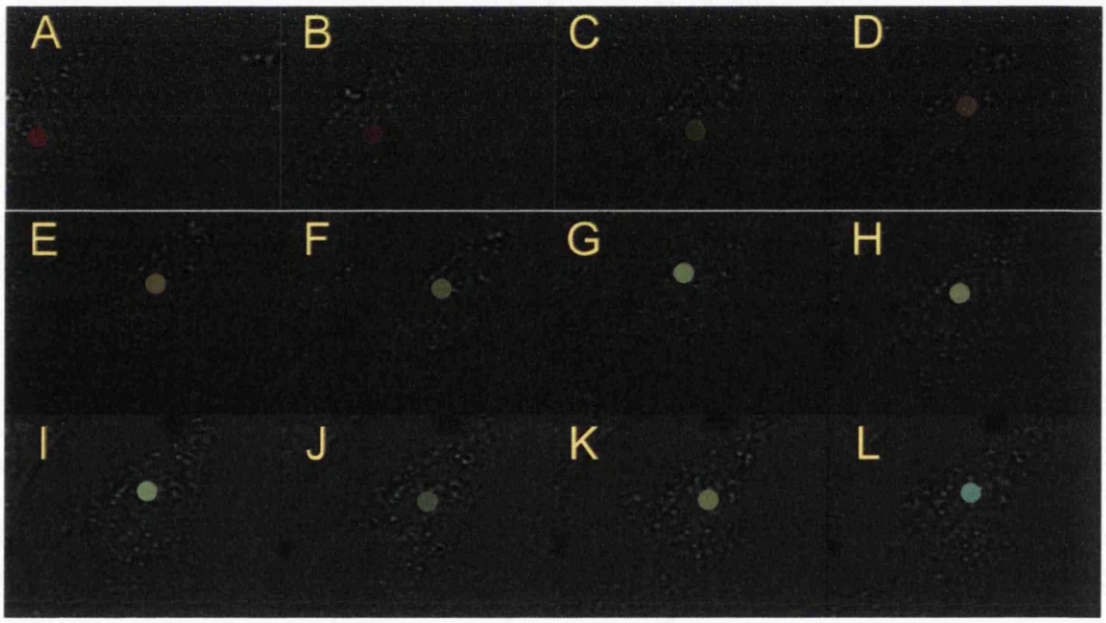



Figure 4.14: Image sequence showing a single cell centroid overlaid onto brightfield images. The centroid can be seen to dramatically change colour early on and then settle into a consistent hue as a result of existing initially on the edge of the field of view and progressing further into the visible range. Whilst at the edge, regions of the cell are not visible causing the centroid to be colour coded without the full number of quantum dot markers.

The other primary source of potential barcode changes and therefore shifts in centroid colours is that of a cell going through mitosis where the transfer of the quantum dot markers from the parent cell to each of the resultant daughters and creates two new centroid colours, both of which have the theoretical potential to be completely different to that of the parent. The image sequence in Figure 4.15 (opposite), details a cell undergoing mitosis with the manually identified, barcode coloured centroids overlaid. The centroid is initially seen to have a green hue (A) with the colour changing to a very bright shade of blue-white (B-E) indicating a major change in the underlying cellular barcode. This occurs as the cell contracts forcing all the quantum dot markers to occupy a smaller region of space which in turn increases the likelihood of said markers being concealed under one another. The morphological changes in the cell as it prepares to split also change its optical geometry, dampening the fluorescence signal and altering the number of peaks identified early in the coding process.

These factors combine to create a cell barcode whose value no longer correlates to that of the same cell before the onset of mitosis but does provide a simple way to identify the mitosis event itself as the resultant centroids take on near white shades of colour distinguishing them from the surrounding population as they divide. The near white centroid colouring could consequently be looked upon as an identifier for the M phase of the cell cycle able to readily provide counts and locations of mitosis events for experiment sets with extended acquisition periods. Continuing through the sequence in fig 4.15, the cell can be seen dividing in frames (F-H) with the daughter centroids maintaining high brightness colours and changing into green and blue-green hues through frames (I-L). This stabilization can be attributed to the process of the cells re-adhering and returning to a focal position similar to their progenitor. By the final frame, (P), the cell centroid colourings have both returned to a green hue, with the north cell taking on a bluer tint, visually demonstrating how the barcodes of the daughter cells relate to their parent cell barcode.

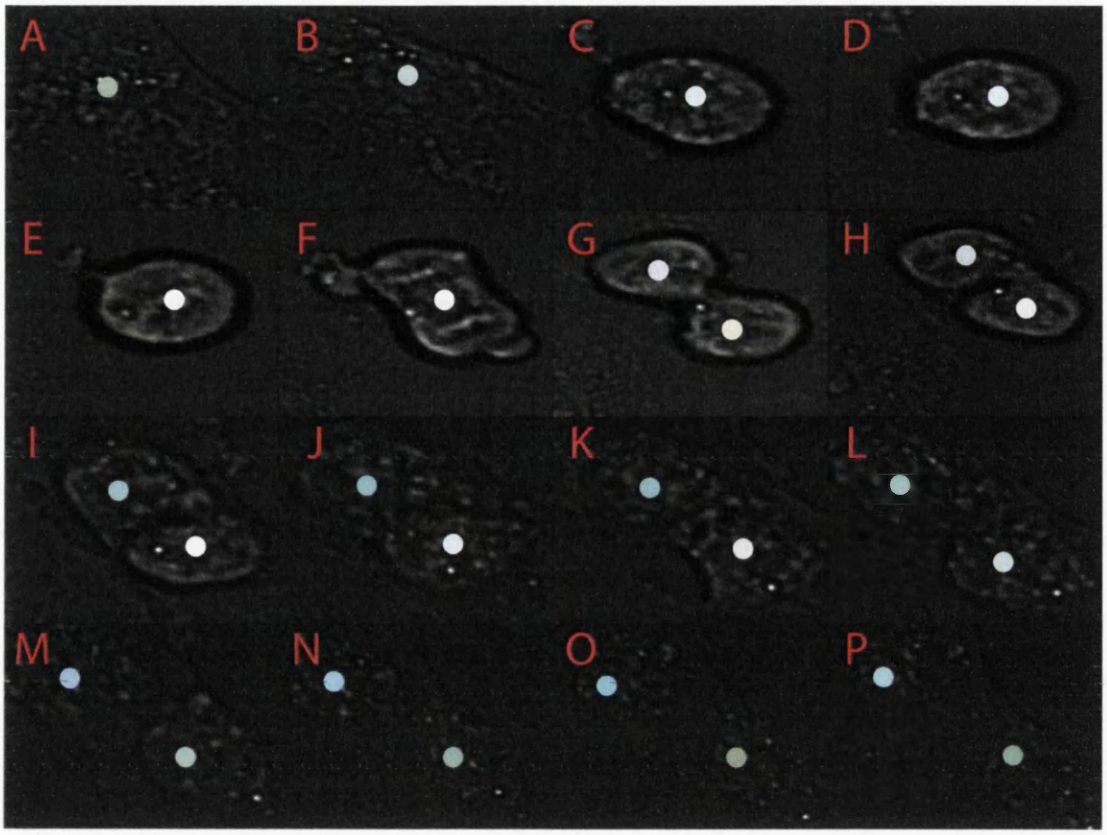



Figure 4.15: Image sequence following a cell through mitosis at 15 minute intervals. The centroid in (A) has a faint green hue and progresses to white as the cell contracts (C-G). As the two daughter cells emerge as individual cells (H) and expand back out (I-L) the centroids begin to regain a similar hue to that seen in the parent cell's centroid with the north cell taking a bluer hue and the south a closely matched green.

Whilst it is theoretically possible for the daughter barcodes to vary completely from that of the parent cell this is highly unlikely. As the quantum dot markers are randomly distributed across the cytoplasmic space without bias towards any of the three colour channels a dilution should maintain the relative ratios between each colour. Consequently, each daughter barcode maintains the relative ratios between each colour channel found in the progenitor and the daughter centroid colours exhibit brightness variations but retain similar hues. This consistency in the centroid hue, or more precisely the inter-channel ratios of the assigned barcode, allows a single cell to be identified and if desired tracked across multiple frames by comparing the barcodes of all cells in the field and matching the most similar pairs.

The data set analysed within this chapter only covers a 5 hour period and as such does not provide enough frames to determine the stability of these barcodes over longer periods of time. However, for the purposes of tracking cells frame to frame, the 5 hour timelapse is sufficient to demonstrate how the colour barcoding process, utilising simple algorithms as detailed in previous sections, can be used to reduce the number of acquisitions needed in a given time range whilst maintaining the ability to correctly track individual cells from frame to frame as performed during tracking in chapter two. The significant changes in the observed quantum dot marker distributions that led to dramatic changes in centroid colouring during the mitosis phase begs the question of whether the more subtle coding variations could be used as an indicator of other cell cycle phases. With the basic method demonstrated, future work to increase the stringency of the peak search protocols when identifying QDot loaded endosomes and investigations of longer timelapse sequences could show tri-colour quantum dot loading, alongside group distribution analysis, as a potential method of cell cycle tagging.

# CHAPTER FIVE

# Automated Drug Dosage Tracking

In accordance with the theme of automation, this chapter explores the application of image foreground signal tracking on a study of human dermal microvascular endothelial cells (HMVEC) exposed to porous silicon (pSi) microparticles loaded with Doxorubicin (DOX). The experiment carried out sought to investigate the potential use of these pSi microparticles as a vector for targeted drug delivery. To this end the accumulation of DOX within cells is tracked through time to investigate the efficacy of the pSi microparticle DOX delivery vector.

## *The Data Set*

The provided data set contains images acquired using an ImageXpress Micro Widefield High Content Screening System taken at 26 acquisition positions across 6-well plates over a time period of 24 hours with positions 1-3 lying in well A1, 4-8 in A2, 9-13 in A3, 14-18 in B1, 19-23 in B2 and 24-26 in B3. $2\times10^5$ human endothelial cells were loaded onto well plates with 3.5µl of DRAQ7 per ml of media and then plated on 5 plates (A1, A2, A3, B1 and B2) of a 6-well plate 48 hours prior to imaging. This experiment makes use of the DRAQ7 membrane impermeant nuclear DNA stain as an easy to identify fluorescent marker for apoptotic cells. Well plates A3, B1 and B2 are of primary interest as they contain HMVEC cells placed in media, DRAQ7 and DOX loaded pSi particles with cell to particle ratios of 1:40, 1:20 and 1:10 respectively.

## *Image Pre-Processing*

Due to the size of the data set the objective of this pre-processing section is to acquire some parameters that can be used to create noise masks for each of the images on-the-fly to avoid the need to run noise filtration analysis procedures during later analysis stages and make redundant the need to store a large set of cleaned images. When analysing large data sets it is prudent to break up the analysis procedure into stages and save resulting outputs along the way where practical to lessen the effect of possible system failure leading to loss of running time. Such

approaches also naturally lessen the number of floating variables which has the additional effect of enhancing performance. The images in the sequences acquired contain a background noise feature with a radial characteristic effect, the nature of which comes as a result of the illumination source lamp, see figure 5.1. To accurately identify objects and measure their respective intensities it is required that this effect be removed.

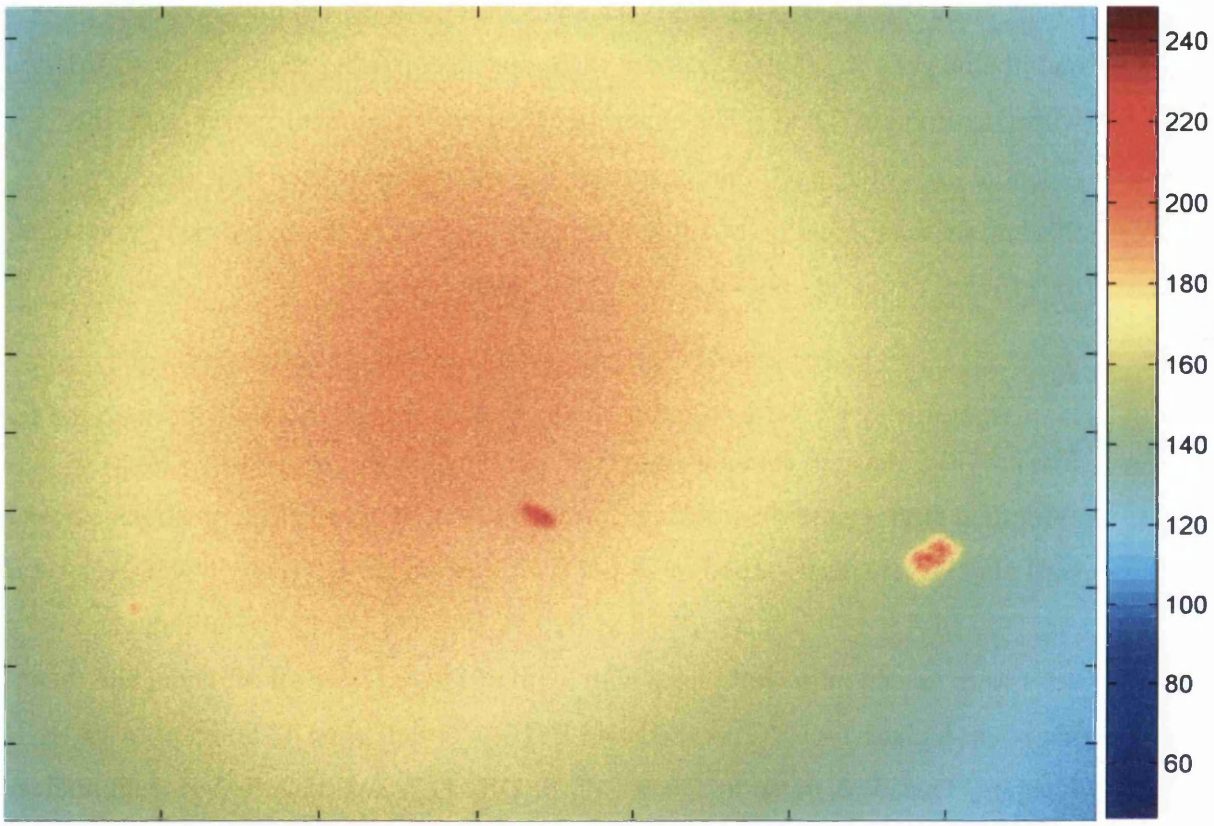



Figure 5.1: Sample DOX channel image at the first time point coloured to highlight the shape of the background noise. The lack of cumulated DOX signal at this early time point accentuates the effect of background noise, the intensity range in this image varies from ~100 (blue region) to ~220 (red).

The image in figure 5.1 shows the effect the noise has on the prospective isolation of foreground objects with the local background around the right-most object showing clear boundary definition and the central object occupying a region of less certainty in terms of edge definition. Such a noise feature prohibits the direct application of simple step-filters in order to isolate foreground objects as the resultant average background calculations would over compensate on the peripheral regions. As such a more complex noise mask needs to be defined. The noise feature here can be seen to be reasonably radially symmetric about a central point or noise centre and

identification of said centre will allow rapid characterisation of the noise as a 2 dimensional Gaussian bell. Identification of the centre follows similar principles to those employed by the peak finder in earlier chapters, the difference being a search for a single global peak rather than multiple localised ones. In order to effectively filter the local fluctuations from the search for the centre of the Gaussian noise bell, two profiles, $P_X$ and $P_Y$, are extracted across all columns and rows respectively in the image matrix, M, about the central point in the y-axis and x-axis respectively such that

$$P_X = M(round(0.5 \times size(M,1)),:)$$

$$P_Y = M(:,round(0.5 \times size(M,1))) \qquad (5.1)$$

Whilst the profile does not need to be extracted from around the central bisecting line, it is a consistent and easily calculable starting point that the noise feature centre is often in close proximity to.

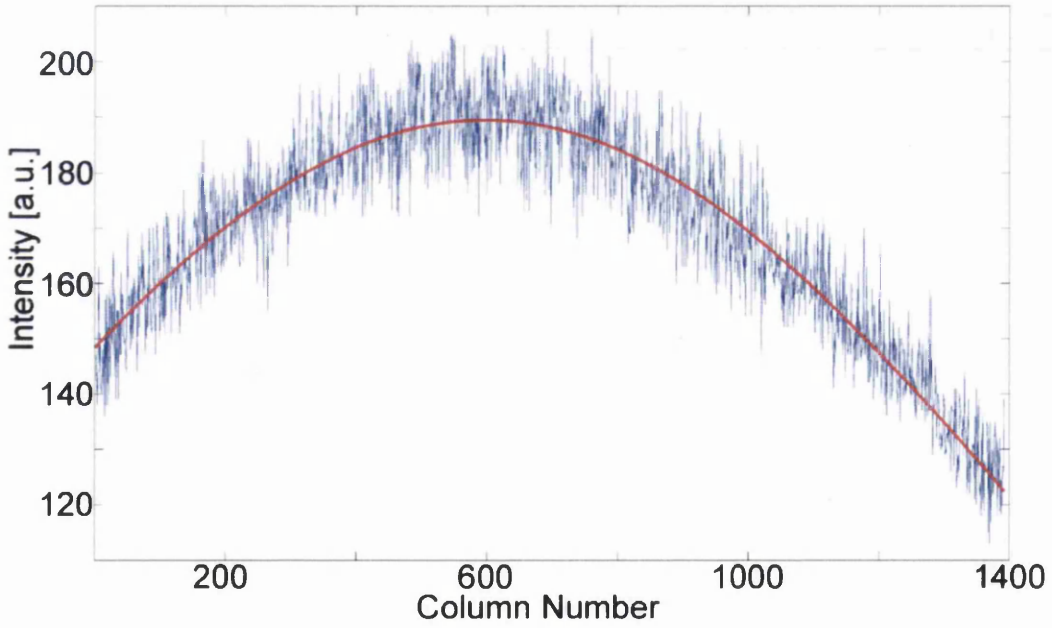



Figure 5.2: Sample profile centrally cored across all columns in the image matrix shown in figure 5.1.

Figure 5.2 contains a sample $P_X$ core profile depicted in blue showing both the intensity fluctuations seen across an image bisection and the overall shape of the noise curve. The red line is the result of a Gaussian fit using the *cftool()* function, an inbuilt Matlab tool for regression fitting curves of selected format to data sets. Replacing the image profile with the profile fit allows the global noise in the full

image noise to be thought of as a single peak and the coordinates of the noise feature centre are then simply the maxima point of $P_X$ and $P_Y$.

$$C = [\max(P_X), \max(P_Y)] \tag{5.2}$$

With the centre of the Gaussian noise bell known, a radial profile is created by identifying the corner pixel with the greatest distance from the noise centre coordinates and walking the shortest path from the noise centre to said corner pixel. An example of the intensities of each pixel in the shortest path are plotted against their Euclidian distance from the noise centre in figure 5.3 to portray the radial profile of the noise curve shown in the previous figures.

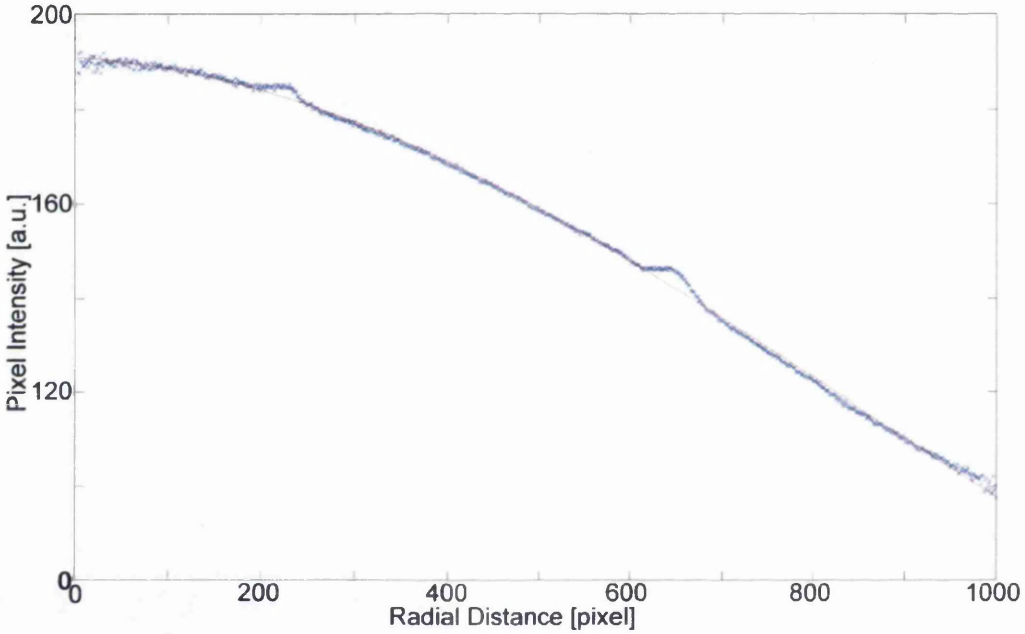



Figure 5.3: Example of a generated radial profile overlaid with the resultant Gaussian fit. The Gaussian coefficients can be used to generate a noise mask and clean the images during later processes.

The furthest corner is selected to provide the largest possible intensity set on which to again apply the *cftool()* function to fit a Gaussian curve of the form

$$f(x) = ae^{-\frac{(x-b)^2}{2c^2}} \tag{5.3}$$

to the radial profile yielding three coefficients, $a$, $b$ and $c$, that, along with the noise curve centre coordinates, can be used to rapidly generate a filter mask to remove the global noise features that inhibit image foreground isolation. Combining these coefficients with the noise centre coordinates creates a set, defined as

$$A_T = [C_X, C_Y, a, b, c], \tag{5.4}$$

that is stored for subsequent use in the reduction of noise in its associated image. Here, the variable T represents the time frame. $A_T$ is calculated for every time point at a given position and assigned to the set $\boldsymbol{B}_P$, such that

$$\boldsymbol{B}_P = \{A_1, A_2, \dots, A_T\}, \tag{5.5}$$

where $P$ refers to the current acquisition position and $1 \leq T \leq 97$. This process repeats for every acquisition point in a given z-plane and all sets of $\boldsymbol{B}_P$ are then similarly assigned to parent sets such that

$$\boldsymbol{C}_Z = \{\boldsymbol{B}_1, \boldsymbol{B}_2, \dots, \boldsymbol{B}_P\} \tag{5.6}$$

where $Z$ refers to the current z-plane number and $1 \leq P \leq 26$. The variable $\boldsymbol{C}_Z$ is separately stored for each of the seven z-plane numbers and the resultant variable sets occupy approximately 50kB of on-disk memory and represent noise masks for a set of images approximately 50GB in size. Details of how these stored numbers are used to remove the noise effects are provided in sections under '*Death Event Identification and DOX Presence Validation*'. Figure 5.4 contains (A), a comparison between a raw image file as seen in previous figures and (B), the result of the cleaning process on (A).

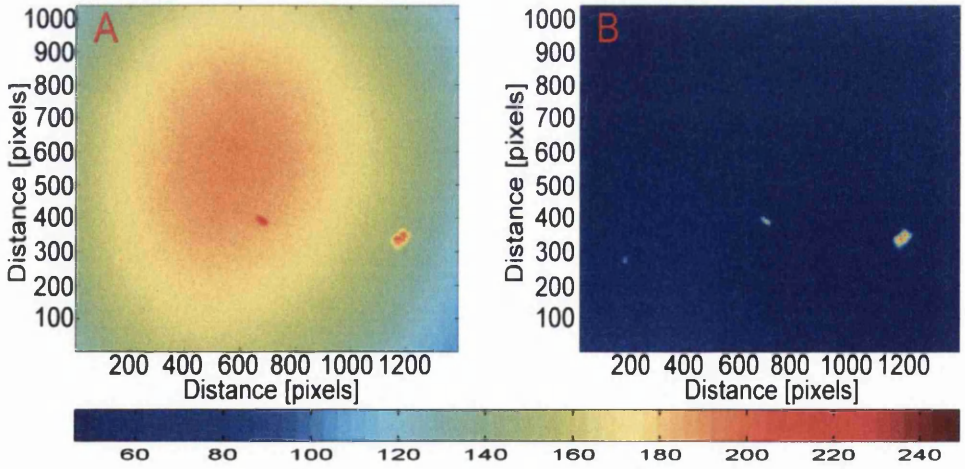



Figure 5.4: Comparison of images before and after applying the Gaussian form noise filter. The two objects clearly seen in (A) remain with similar colouring and a third object has become prominent in (B) which on re-inspection was present in (A) but easily overlooked.

The two objects can now be seen clearly in the foreground and a third object has come into sharp contrast. Whilst it is possible to see the left most object in the raw image (A) it could easily be overlooked and would be susceptible to loss due to low definition in latter stages. The image in (B) can also be seen to be not truly flat as per idealised noise removal with the lower left quadrant displaying regional brightness which comes as an artefact of fitting an ideal Gaussian bell curve as a mathematical approximation to the actual noise shape. A better method of noise reduction would be to record an image from an empty well plate at each time point along the acquired timelapse for use as an experimental noise mask and simply subtract that image from the images needed to be cleaned. This however incurs significant additional file storage penalties and was not provided with this data set and the methods employed here will prove sufficient for extraction of information. The noise present in these images is characteristically multiplicative in nature. As such, a simple subtraction of the background noise does not fully restore the pixel intensities within the object of interest their true measured values. Despite this, the subtraction method is utilised here for speed and simplicity as the main objective is to correctly identify DOX objects and their contained pixel intensities can be adjusted at any point after retrieval.

### *Death Event Identification and DOX Presence Validation*

Doxorubicin (DOX) is one of the most widely used broad spectrum anticancer drugs however, long-term administration in therapies is limited due to the development of dose-related cardiomyopathy [68], [69]. Consequently, methods to reduce the dosages necessary for efficacy are of major interest. To this end, the experiment under analysis aims to utilise pSi microparticles as a drug delivery vector to target the delivery of DOX within the cell membrane. Accumulations of the DOX drug can be identified and tracked by its fluorescence negating the need to acquire morphological knowledge of the host cell. DRAQ7, a membrane impermeable dye which stains the nuclei of dead cells, is used to identify points of cell death [70].

The occurrence of apoptosis results in the fluorescence of DRAQ7 within the cell providing an easy to identify marker for the event in the DRAQ7 channel images.

Figure 5.5 (opposite) provides an overview of the timelapse at position 9 with the brightfield, DRAQ7 and DOX channels labelled as (A), (B) and (C) respectively. The timeline of column (A) shows the cells balling-up over time, a cellular symptom of an environment increasing in hostility as the levels of DOX seen in the corresponding timeline of (C) rise and result in cell death and the triggering of DRAQ7 fluorescence seen in column (B).

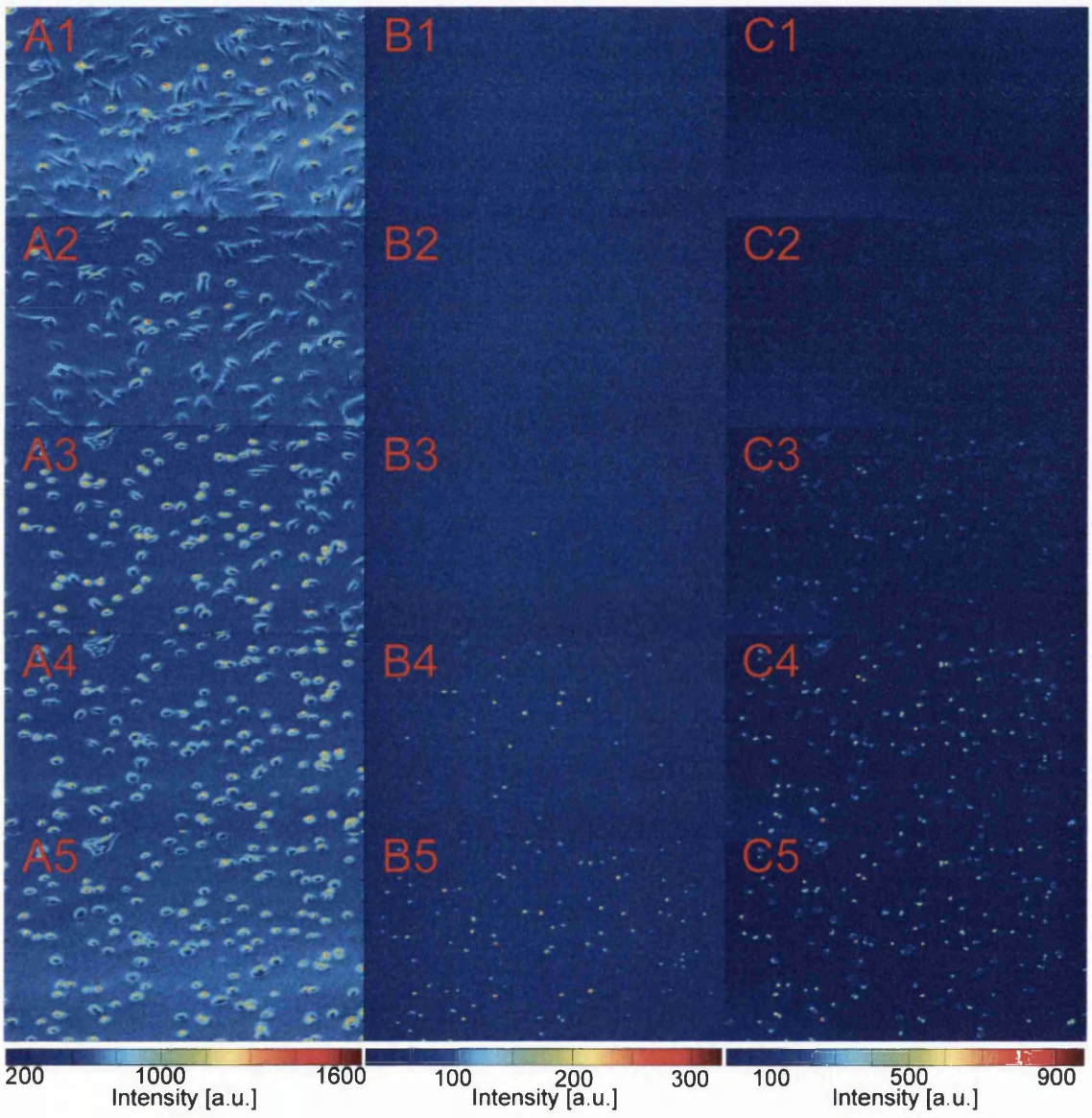



Figure 5.5: An overview of the timelapse in each of the three channels, brightfield (A), DRAQ7 (B) and DOX (C), over intervals of 6.75 hours and coloured in order to accentuate the contrast for visibility and highlight the features of interest. The cells in the brightfield can be seen to react over time to the increasing DOX levels within present and accordingly the DRAQ7 apoptotic marker becomes increasingly prevalent.

For interpretation, these images are coloured using a familiar "heat-map" style colour-map such that the highest pixel value shows in red and the lowest in blue. Taking column (B) as an example, images (B1) and (B2) represent the background fluctuations, having been cleaned using the aforementioned coefficients, where no DRAQ7 signal exists and the variation in the intensity values of red (highest intensity) and blue (lowest intensity) pixels are typically of the order of 10. The arrival of the first DRAQ7 signal in (B3) increases the highest pixel intensity value by 2 orders of magnitude causing a peak point at the DRAQ7 fluorescence region and the background variation to be comparatively insignificant resulting in a flat looking, blue background. A similar effect can be seen in the DOX column (C1-C5) where an increasing DOX signal becomes more prominent over time as it moves clear of the background. The presence of DOX in the media does however reduce the sharpness of contrast when compared to observed time frames in the DRAQ7 channel.

The process of generating DOX profiles begins with the identification of DRAQ7 signal objects. To isolate the foreground of a given DRAQ7 channel image the stored noise coefficients are utilised to clean the image. For an image $I_T$, where T denotes the current time frame, the corresponding coefficients $A_T = [C_X, C_Y, a, b, c]$ are recalled and a noise mask matrix, M, with dimensions of size($I_T$) is initialized as

$$M(:,:) = 0$$

$$M(C_y, C_x) = 1. \qquad (5.7)$$

This creates a matrix of zeros the same size as the current image, $I_T$, with a single entry at the pixel stored in $A_T$ as the centre of the noise feature. Utilising the function *bwdist()*, another inbuilt Matlab function, rapidly generates a matrix D where size(D) ≡ size(M) and each element of D stores said elements Euclidian distance from element point ($C_X$,$C_Y$). Implementing the Gaussian curve equation in (5.3), the mask matrix M is then redefined as

$$M(IND) = ae^{-\frac{(D(IND)-b)^2}{2c^2}} \qquad (5.8)$$

where *IND* references all elements within the matrices M and D. The clean image is then simply

$$I_T^C = I_T - M. \quad (5.9)$$

With the image cleaned, the foreground is isolated by defining a binary image B such that

$$B_T(x,y) = \begin{cases} 1 \; if \; I_T(x,y) \geq \theta \\ 0 \; otherwise \end{cases} \quad (5.10)$$

where θ is again a pixel intensity threshold defined as

$$\theta = \mu_{image\ intensity} + (2 \times \sigma_{image\ intensity}). \quad (5.11)$$

Here the standθ = μ_image intensity + (2 × σ_i ded a greater level of harshness due to the large difference between the foreground DRAQ7 signal and image background and doubling the term lessens the possibility of erroneous identification of spikes in intensity level and simplifies the resultant binary image. As the DRAQ7 channel is analysed only to serve as a yes or no marker signal for death events there is no requirement for knowledge of the signal intensity and as such all further analyses can be conducted directly on the created binary representation, $B_T$.

Within the binary representation, objects are defined as elements or connected groups of elements in $B_T$ with a value of 1 (true). Matlab's image toolbox contains the function *bwconncomp()*, implemented here, which identifies and stores all such objects in a binary matrix into a cell vector, which shall be referred to as **O**. To eliminate erroneous interpretations from early time frames when low intensity variations can result in large numbers of small objects, the cell vector **O** is filtered to remove any identified objects with less than 200 pixels. This cut-off is selected as genuine DRAQ7 signal sources are observed to typically contain approximately 1000 pixels and irrepressible noise induced objects are typically observed as sub 100 pixel objects and often only exist as scarce patches of singular pixels. This leaves objects that can be interpreted as DRAQ7 signal sources and their centre coordinates, centroids stored as a vector list $\mathbf{C_T}$ where T references the current time frame, are used as markers for cell death events. Simply storing these coordinates for every time frame where present is however insufficient as a DRAQ7 signal from a single cell can exist through multiple frames before the cell breaks down and releases the DRAQ7 to dilute into background. This leads to the requirement to effectively track the DRAQ7 markers to avoid miss interpreting their continued existence as further

death events and make sure only the initial instance is used. This procedure is performed by utilising some of the tracking framework developed for use in the cell tracking algorithm detailed in chapters two and three.

Essentially, for a given position at a given timeframe, $P_T$, identified DRAQ7 signal object centroids are cross referenced with stored DRAQ7 centroids in $P_{T-1}$. For each identified DRAQ7 centroid in $P_T$, if the smallest inter-centroidal distance exceeds 50 pixels, observed to be a typical cell diameter, the centroid is declared as a new event and assigned to a new vector. If the smallest inter-centroidal distance is less than or equal to 50 pixels, it is deemed to be the same death marker as that which it is closest to from the previous frame. Consequently it is added to the vector of its closest neighbour in much the same way as the lineage vectors where compiled in earlier chapters. By iterating through all of the identified DRAQ7 marker signal objects through all time frames a complete set of DRAQ7 lineage vectors are created, the first entries of which can be used as a marker for a death events. These DRAQ7 lineages are catenated into a single storage matrix labelled $\mathbf{S_D}$.

The automated analytical system here aims to extract DOX profiles over time and uses these DRAQ7 death events markers as the end point of such profiles. Consequently, we seek to verify the presence of DOX in the region about all identified DRAQ7 fluorescence events to distinguish between cell death occurrence with probable DOX causation and cell death with no specifically attributable cause. This is done by cross referencing the coordinates stored at the head of the columns in $\mathbf{S_D}$ with the same coordinate position in the corresponding DOX channel images. For each column of the matrix $\mathbf{S_D}$, first identify the row in which the first non-zero entry lies, this coordinate pair is the head of the coordinate set and is denoted along with the row number as $\mathbf{H} = \{(x, y), T\}$, where T references the row in which the coordinate (x, y) exists and can be thought of as the time of death. The corresponding DOX channel image is then loaded and cleaned using the stored noise coefficients in the same manner previously described in application to the DRAQ7 channel images (equations (5.7) through to (5.9)). A sample image is shown in figure 5.6 (opposite) after the application of the cleaning process with three regions highlighted within which DRAQ7 signals were found to first appear. To better deal with the regional variations across the DOX image, prior to isolating the foreground to identify

potential DOX objects, for each set **H** an image crop is taken creating a smaller field of view of size 201×201 about the centre defined by (x, y) stored in **H**.

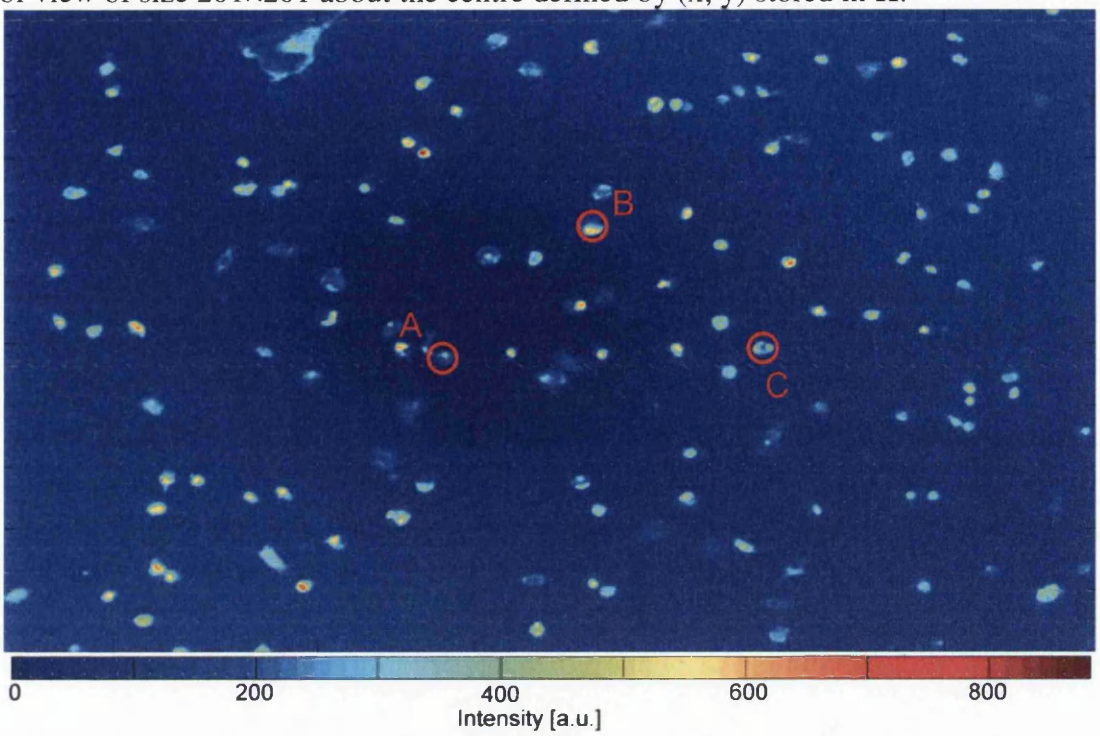



Figure 5.6: Sample DOX image at 4 hours when three DRAQ7 signals appeared within the highlighted regions. To verify cell death occurred as a result of internalised DOX exposure each DRAQ7 signal is checked for proximity to a DOX signal in the respective frame.

A binary image of this crop is then created as before (equations (5.10) and (5.11)) and the remaining foreground objects identified as DOX signal objects. Figure 5.7 provides the respective DOX channel binary images of the highlighted regions in figure 5.6 with the centre of the image crop marked with a black spot.

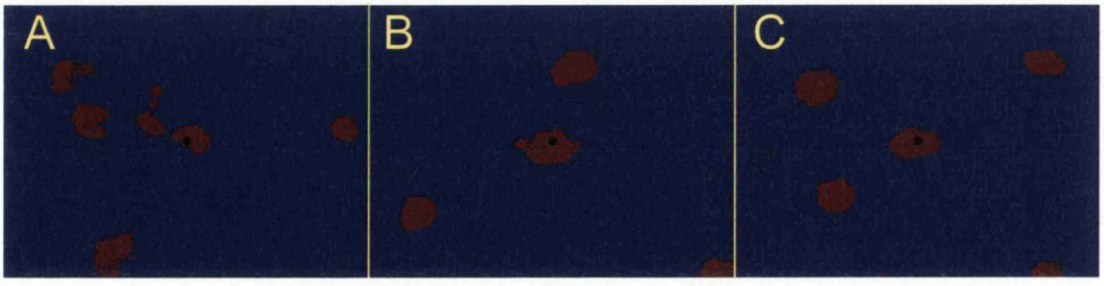



Figure 5.7: Digitized representations of the regions highlighted in fig 5.6. The black dots indicate the coordinates of the centres of the identified DRAQ7 signal and the red objects are the isolated foreground of the DOX image with (A), (B) and (C) corresponding accordingly.

These three examples show instances where the DRAQ7 object centroid lies on top of a DOX object validating each apoptotic event as plausibly DOX causal. The DRAQ7 centroid can however occupy a pixel that is not covered by a DOX object and still be of the same cell origin as said DOX object. This can be as a result of using a single centroid to represent a DRAQ7 object where the complete object would share pixel indices with those of the local DOX objects but the two objects can also be of the same cell origin and not occupy any shared pixels. DRAQ7 is a cell nuclei stain where as the DOX released from internalized micro-particles occupies regions in the cytoplasm; this difference in operational locations can lead to the two objects being seemingly unconnected when represented in binary form. To account for both these possibilities, the DRAQ7 marker is verified as having DOX present if a DOX object exists within a 20 pixel radius from the DRAQ7 centroid. Full-filling this radial proximity criterion places two objects at a distance well within the observed average cell diameter of 50 pixels. This is deemed sufficiently large as the presence of any DOX signal objects within an average cell diameter to the DRAQ7 death marker is indicative of the cell walls having not yet deteriorated sufficiently to allow the DOX to fully escape cellular confinement and should therefore still be in close proximity to the DRAQ7 stained nucleus. While scenarios in defiance of this assessment could theoretically exist they are not readily observed and any resultant loss of potential profiles due to selected automation parameters is deemed to be outweighed by the eventual size of the yielded profile set. A radial proximity allowance can lead to multiple DOX objects being associated to a single DRAQ7 death marker. Such occurrences are more prevalent in earlier time frames when apparent compartmentalisation of the DOX signal can be attributed to the presence of multiple internalised pSi-particles. In these events the multiple objects are combined into a single object pixel list and are treated as a single DOX object group from then onwards. Once the DOX object or objects are assigned to the DRAQ7 marker, the centroid of the DOX object, $\mathbf{C_{DOX}}$, is calculated as

$$\boldsymbol{C_{DOX}} = [\tfrac{1}{N}\textstyle\sum_i x_i \ , \ \tfrac{1}{N}\sum_i y_i] \qquad (5.12)$$

where $1 \leq i \leq N$, N is the number of pixels in the DOX object group and x and y correspond to the coordinates of said pixels. It is useful to store the pixels

corresponding to the DOX object or DOX object group in an alternative format for later use and therefore we define the vector

$$\boldsymbol{L} = sub2ind(x_i, y_i) \qquad (5.13)$$

to store the (x,y) coordinate list as a single column vector $\boldsymbol{L}$. Each DRAQ7 marker **H**, verified to exist in proximity to a DOX object, generates a structural array in which the marker **H** is stored along with its respective DOX channel binary image, such as those seen in figure 5.7, the centroid, $\mathbf{C_{DOX}}$, of the associated DOX object for use as the seed points for the DOX profile generator and the object pixel list $\boldsymbol{L}$.

Some examples of non-ideal scenarios encountered during the validation stage are given in figure 5.8. The image in (A) shows a previously mentioned scenario where the DRAQ7 centroid, shown in yellow, lies within an encapsulating DOX object but does not occupy a pixel within that object. The DRAQ7 marker would be assigned to such an object through application of the aforementioned proximity criterion without implementing the need to store the complete DRAQ7 signal object. The event shown in (B) could be as a result of a cell dying through mechanisms other than those directly related to DOX presence, a rogue feature passing through the filters in the DRAQ7 channel or potentially a cell rapidly breaking up after death causing the DOX to dilute back into the background and traces of DRAQ7 to remain in sufficient quantities to allow its identification as a death marker. The latter two are unlikely. If the cell dies between acquisition points with sufficient time to break up and loose all associated DOX signal before being observed again it is likely the nucleus would also degrade sufficiently to allow DRAQ7 dilution to occur. Rogue noise features are also extremely unlikely to occur at high enough intensities and in large enough groups to be miss-interpreted. Image (C) shows what looks to the eye like two rounded objects in contact with one another. When verifying the presence of DOX such an object is of no consequence as no interpretation of the object is made at this point, only its existence is of relevance. When coming to use the seed points calculated in this section to create DOX profiles however, such an object can cause erratic results depending on the motions of the constituent parts. The object could be a mitosis pair with one of the daughters succumbing to the effects of DOX or simply two cells in close contact with one another. The final image, figure 5.8 (D), details another potentially problematic situation to be aware of when creating DOX profiles. The yellow DRAQ7 centroid exists in the region of a single large object, the image containing only 3 unique objects with the central feature being minimally linked together. Such an object is unlikely to be a single cell and is more likely to be at least two cells in close proximity. Application of image analysis procedures such as object dilutions could be implemented to isolate large objects into their bulk constituents by removing minor pixel linkages but are not used here as the occurrence of such situations is infrequent and typically only cause temporary variations in the extracted profiles as the cells migrate around the available space. The problems of distinguishing possible paired objects, interpreting unusual shapes and the consequences imparted by such situations will be further discussed in later sections.

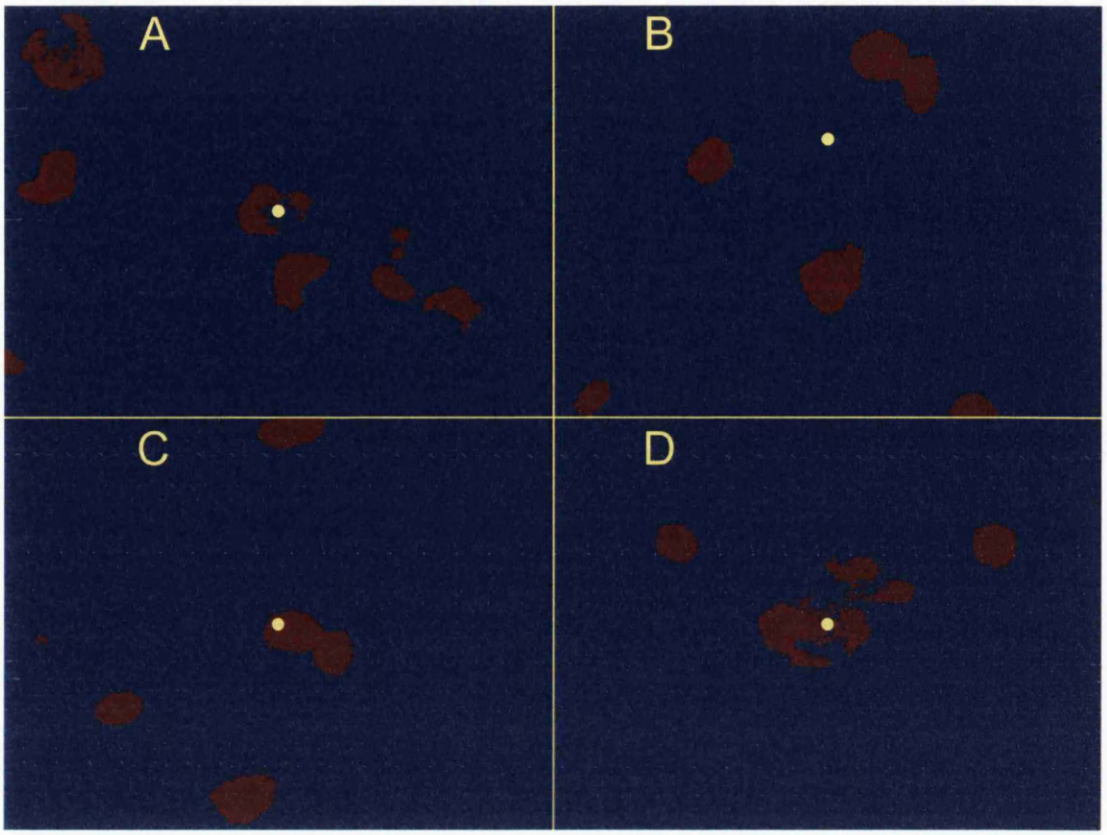



Figure 5.8: Examples of non-idealised scenarios. Here (A) shows an example of a DRAQ7 centroid occurring near a fragmented DOX object, (B) depicts the occurrence of a cell death event without DOX validation, (C) shows a possible close proximity pair which would result in a miss-count of DOX intensity and (D) depicts a mixture of fragmented DOX objects and possible multiple cells in close proximity.

## *DOX Tracking – Generation of DOX Profile Masks*

With the coordinates of the DRAQ7 events identified, verified and stored the process of extracting the profile of the DOX build up to these cell death events can begin. Each stored seed point file contains a centroid starting point, $\mathbf{C_{DOX}}$, the time frame at which the death event occurred, a DOX object pixel list, $\boldsymbol{L}$, and the binary mask of the corresponding DOX channel at that time point. The profile generator seeks to create similar structural arrays from each of the time frames preceding that of the death event by tracking the DOX channel objects associated with the DRAQ7 event backwards through time until the signal fades into the background, or more accurately, until the signal drops below the foreground threshold filter. This is achieved by again implementing the tracking algorithm framework developed in chapter two running in reversed time to track the DOX signal object as though it were a cell object, which it essentially is. As an example, label a structural array containing the seed information as **A**. This array already contains the binary mask for the DOX object at the time frame when death occurred, labelled as T, the mask for the final point of the profile is therefore already defined. To create the DOX object mask for the time frame $F_{T-1}$ the DOX channel image corresponding to $F_{T-1}$ is loaded, cleaned using the stored coefficients as previously described, the foreground is isolated in the same manner as before and the image is then cropped about the centroid, $\mathbf{C_{DOX}}$, that was defined for the frame $F_T$. The interval between acquisition points of 15 minutes was observed to be sufficiently low enough to prevent dramatic migration of cells from frame to frame and as such the DOX objects in the current image frame, $F_{T-1}$, remain in relatively close proximity to that of their respective DOX objects in the successive image frame, $F_T$. Consequently we can identify preceding DOX objects in $F_{T-1}$ via their proximity to the centroid coordinate $\mathbf{C_{DOX}}$ identified from frame $F_T$ and generate another structural array in the same form as **A** for the frame $F_{T-1}$.This process can be repeated for each frame preceding $F_T$ until the earliest time frame acquired, $F_{T=0}$, is reached or the DOX object itself drops below a size of 100 pixels. This object size threshold is chosen as a result of manual observation as objects that do not contain at least 100 pixels are often noise features and it is therefore deemed reasonable to define such a limit below which the DOX signal object is too similar to noise (similar to the DRAQ7 channel object filter previously mentioned). While this size limit could be more rigorously selected it only

affects the last couple of identified DOX object intensity values that exist in the early stages of the resultant profile when the time direction is restored. It is not therefore significant to the overall shape of the extracted profile or to the values extracted towards the temporal end of the profile which are the main points of quantitative interest. The resultant set of structural arrays $\mathbf{A}_{TD}$ through to $\mathbf{A}_{TI}$ contain binary masks that represent the complete DOX profile currently being tracked. Here TD, Time of Death, references the time frame where the DRAQ7 death signal was initially identified and TI references the initial frame of the DOX profile under creation. In real terms, a profile may end as a result of reaching the first acquired time frame, the DOX object containing less than 100 pixels or the cell containing the DOX being tracked migrating beyond the boundary of the field of view. This process is then repeated to generate structural arrays containing binary masks of the DOX channel at each time point in the tracked profile for all previously identified DRAQ7 events.

A sample set of 4 images is provided in figure 5.9 (overleaf) to demonstrate typical sequence showing the build up of DOX within a cell over the space of 10 hours with (A1-A4) revealing the DOX channel image crop after noise reduction and (B1-B4) their respective binary representations. The arrow of time progresses here from 1 to 4 through 3.5, 6.5,10 and 13.5 hours respectively and the DOX signal can be seen increasing over time. The profile mask generator would therefore start with image (B4) and track the DOX objects back through to the final object in (B1) beyond which the object drops below the defined object size limit. The binary image (B4) is the final frame of the sample sequence showing a yellow spot to represent the centroid associated with the DRAQ7 signal that identified the instance of cell death and can be seen to be well centred in the associated DOX object. The yellow spots in (B1-B3) represent the centroid of the DOX object identified in the frame preceding the frame in the profile generation sequence, which is the successive frame in real time as the profile extraction algorithm runs backwards through time. The compartmentalization of significantly higher than background DOX levels provides evidence of targeted DOX accumulation within cells confirming the functionality of the pSi microparticle delivery mechanism.

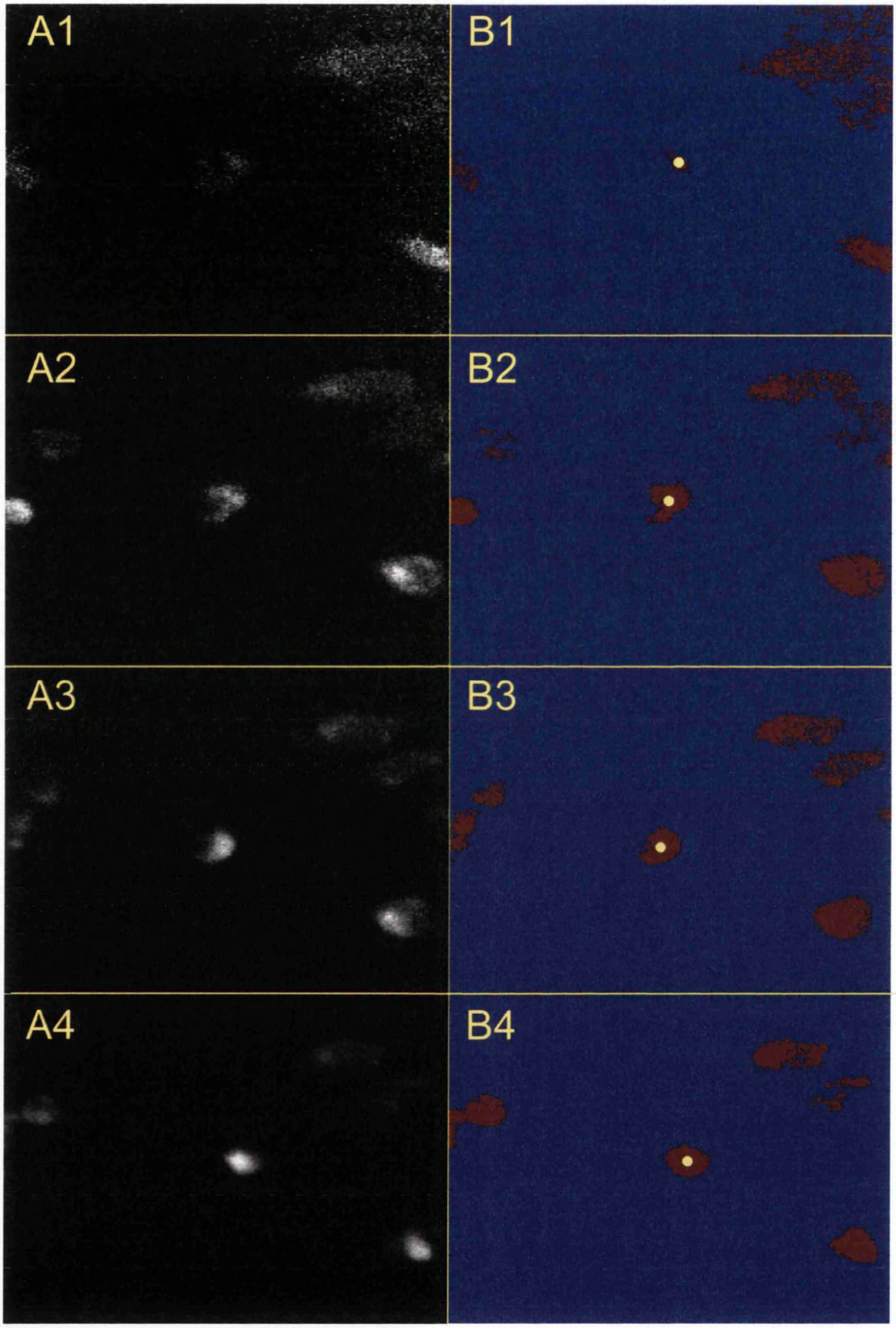

A1
B1
A2
B2
A3
B3
A4
B4

Figure 5.9 (opposite): DOX accumulation sequence spanning 10 hours from 3.5 (1) to 13.5 (4) hours. (A) displays DOX channel images of the local region around the objective cell and (B) the respective digitization of the image foreground. The objective can be seen to fade from the foreground if viewed in reverse and the lack of any foreground object in the precursor frame to (A1) and (B1) results in the declaration of frame one as the initial profile point.

Comparisons of the binary masks and the DOX channel images show pixels that visually appear white in the DOX images but are of insufficient intensity to pass through the foreground threshold, equation (5.11). This is particularly clear when comparing (A1) and (B1) where the DOX mask shows a comparatively small area of red when compared to the region seen the respective region of (A1). Whilst this may seem like a fundamental flaw in the foreground threshold process, defining a global rule for the intensity cut-off and then examining localised regions within it will always result in a compromise between attaining every pixel associated with the objects of interest and filtering unwanted noise. The object that is clearly visually identifiable about the centre of (A1) contains intensities that are similar to the noise pixels that encircle the upper right quartile of the image and is consequently removed by the foreground filter. Such errors could potentially be resolved by incorporating more complexity into the foreground isolation process to include regional analysis procedures. However, for the requirements of this particular experimental assay, it is sufficient to proceed with the simplistic foreground threshold method due to such effects only occurring as the DOX signal approaches intensities similar to background. As stated before in relation to the imposed object size limit, such issues are insignificant relative to the overall gained profile and quantitative attention being aimed at the latter points in the profile rather than the earlier time points. Again, such apparent shortcomings are also counter-balanced by the generation of result sets of magnitudes outside the realm of practicality when considering manual processing. As the DOX builds to levels where the intensity comfortably clears the background levels, (B2), the DOX mask provides a more accurate representation of the DOX channel image and this accuracy increases as the DOX signal intensity increases.

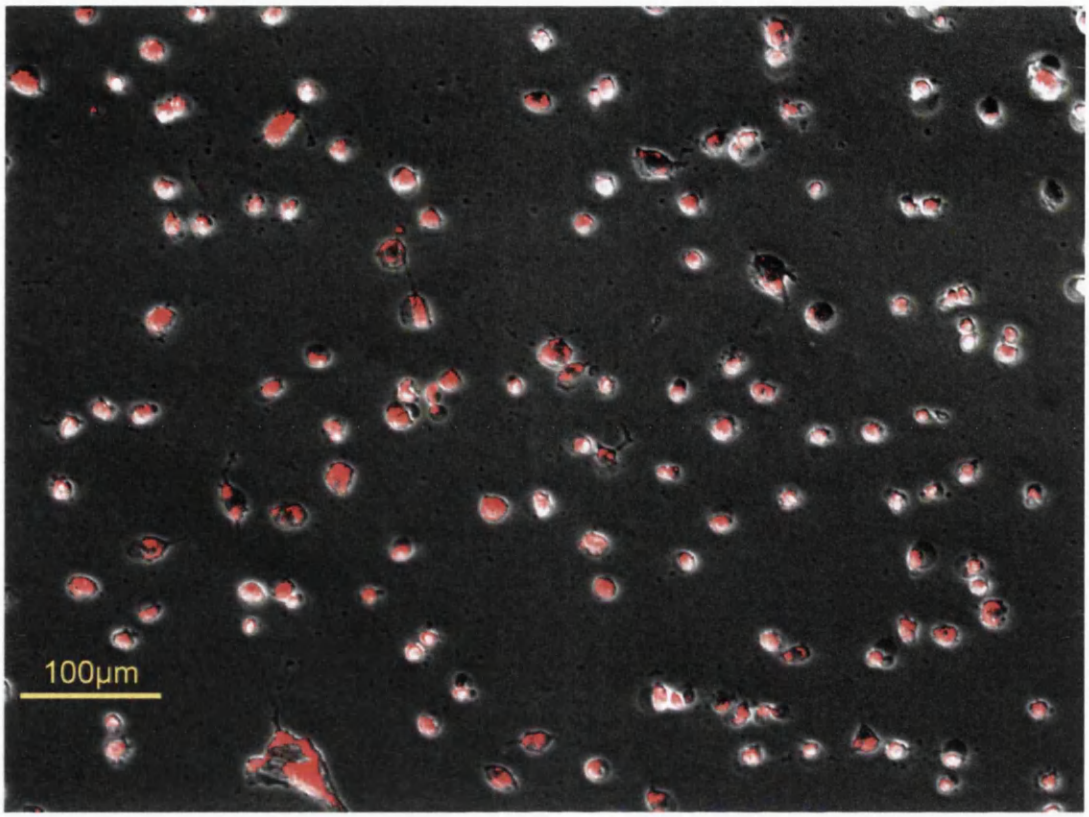



Figure 5.10: Sample final timeframe brightfield image with foreground DOX signal overlaid in red.

The image in figure 5.10 depicts the final timeframe of a brightfield sequence and the general lack of healthy cells. Visually, it is clear that the identified DOX objects are accumulating within cellular regions significantly above background DOX levels. Dark, particulate spots, most clearly visible hear in open media, are visible in the brightfield channel throughout the time sequence. Their locations cross correlate with dark spots seen in the DOX fluorescence channel at early time points (as seen previously in fig 5.9 and later in fig 5.15). This observation is discussed further under the next section.

*Data Mining – An Overview of Collated Data*

With the computationally expensive tasks of DRAQ7 interpretation and DOX object tracking complete the resulting outputs, stored in compact formats to maximise output write speed and minimise the necessary storage space, can now be interpreted. Figure 5.11 is provided as a visual reference and reminder of the experimental configuration.

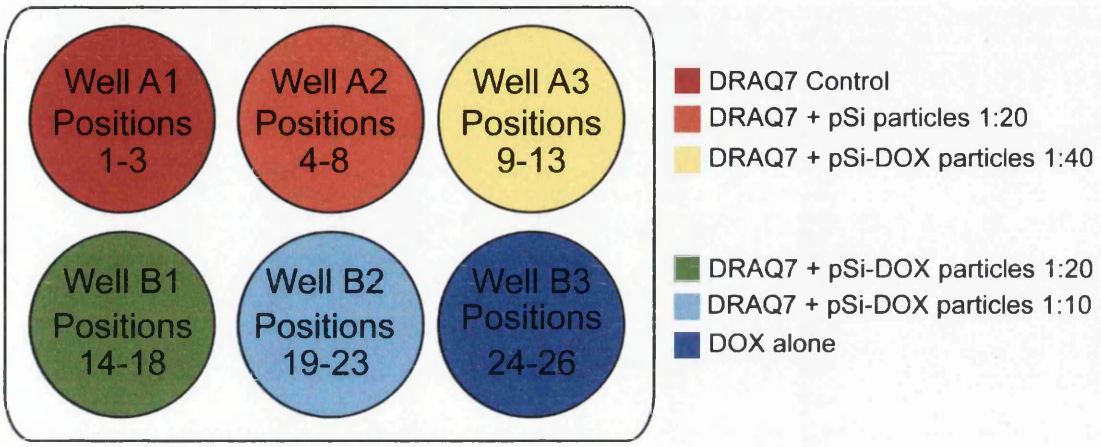



Figure 5.11: Schematic of the experimental configuration. Ratios are displayed in the format of cell:pSi-microparticle. Characterisation of the effects of DOX loaded pSi particles is objective of this analysis and as such plates A3, B1 and B2 are of primary interest.

The 26 acquisition positions are, as stated in earlier sections and detailed in the above figure, taken over 6 well plates each with differing contents. The initial stage of the DRAQ7 event identification procedure produces a variable **H** for each identified DRAQ7 signal before being verified for DOX presence. Figure 5.12 (overleaf) depicts a count of these variables for each well plate against a visual cell count estimate. This estimate is made by utilising the brightfield channel to manually count the number of cells in the same quartile of the initial time frame at each position, multiplying by a factor of four and averaging across the associated positions for each of the 6 well plates. The purpose of such a comparison is purely to provide a check that the significant death count seen in the A3 well plate compared to B1 and B2 is not simply a result of a much larger number of cells being present. Well plates A1, A2 and B3 can be considered as control plates with A1 and A2 containing DRAQ7 and (DRAQ7 + pSi particles) respectively with no DOX present in either, explaining

the lack of cell death in these two plates. B3 contains DOX without HMVEC cells and therefore no DRAQ7 signal so the lack of any count there only serves as confirmation that no noise is being misinterpreted as a genuine DRAQ7 signal. Well plates A3, B1 and B2 all contain HMVEC cells with DRAQ7 and DOX loaded pSi-particles with cell-to-particle dosages of 1:40, 1:20 and 1:10 respectively.

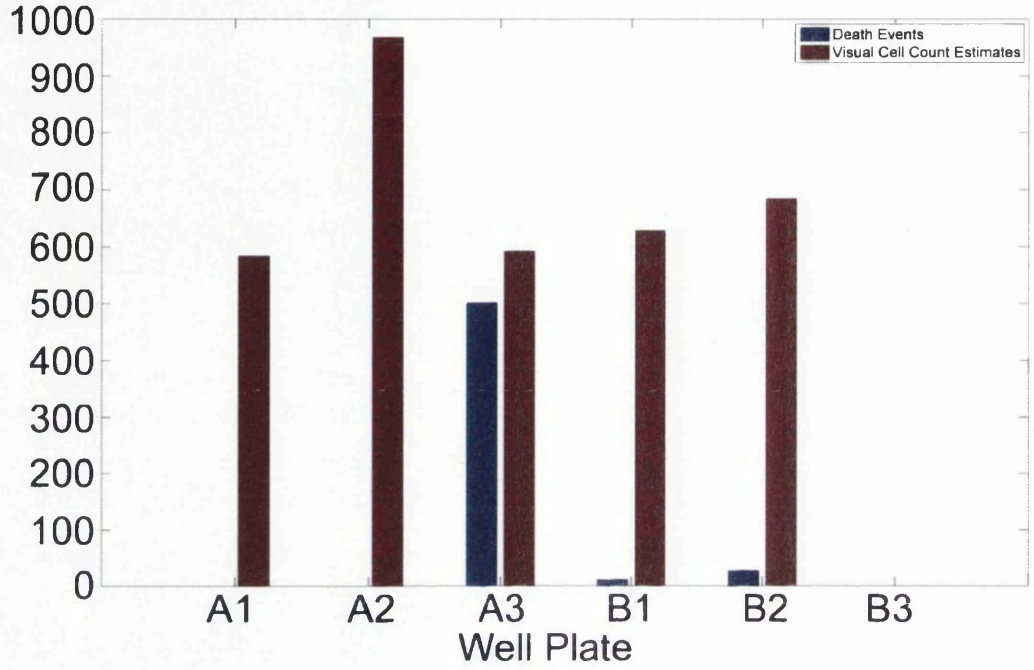



Figure 5.12: Overview of the apoptotic events across all six well plates.

The visual cell count estimates serve to show that the numbers of cells present in these plates are comparable on average and that the significant increase in cell death associated with plate A3 cannot be assumed as a result of higher cell presence and is likely to be down to the increased DOX presence. Figure 5.13 (opposite) similarly presents the comparison of initial cell count estimate against observed DRAQ7 event counts for each of the 26 acquisition positions. Position 1-8 and 24-26 are again void of death counts on account of no DOX or DRAQ7 presence respectively and comparison of the 9-13 and 14-23 regions with their respective cell count estimates again shows the significant death count increase in well plate A3 to not be down to increased number of cells. The unexpected spike in death counts in position 20 however can be seen to correlate with a significantly higher numbers of cells present from the outset.

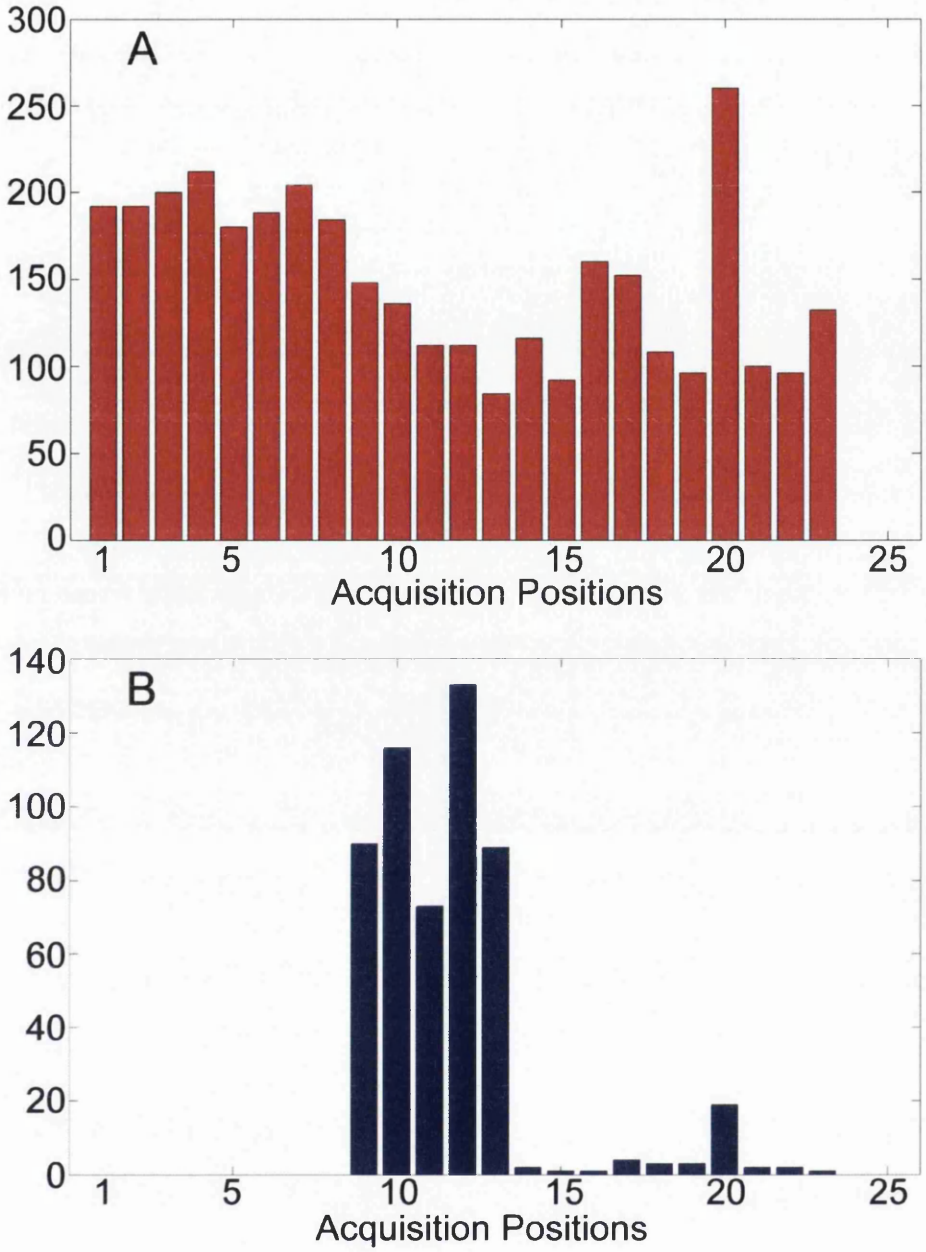



Figure 5.13: Cell population profiles, with (A) showing the visually estimated cell count per position and (B), the cell death profiles across all 26 acquisition positions. The cell number estimate is provided to show how the spike in apoptotic events to the across positions 9-13 is not attributable to higher numbers of cells being present when compared to positions 14-21 and that the spike at position 20 can be attributed to a significantly higher initial cell population.

To make use of the DOX control set at positions 24-26 a DOX background audit is performed to compare the resultant background DOX presence of the different delivery systems. The overall DOX level of an image I at a given acquisition position P and time T is defined as

$$\mu_P^T = \frac{1}{i \times j} \sum_y \sum_y I_P^T(x, y) \qquad (5.14)$$

where $i$ and $j$ are the height and width respectively of the image $I_P^T$. The average DOX level for a well plate at time T is then

$$M_W^T = \frac{1}{N_w} \sum_k \mu_k^T \qquad (5.15)$$

where $N_w$ is the number of positions in the well W and $k$ references those positions. The resultant values of M for each well plate across all available times are plotted in figure 5.14 (opposite). Plates A1 and A2, shown in red and yellow respectively, contained no DOX and as such their intensity values are purely channel background noise and can be used as a reference for comparison. Well plates B1 and B2, shown here in light and dark blue respectively, contain DOX loaded particles in doses of 20:1 and 10:1 particles per cell and as per the expected outcome, the higher dosage corresponds to a higher overall DOX level with no noticeable change across time other than a sharp initial dip which is also seen in the A1 and A2 background noise and can thus be interpreted as an intrinsic feature of the imaging equipment. A3, shown here in green, is the main plate of interest as it contains the highest DOX particle dosage and is where most of the cell death occurs. It is of note then that the DOX image background level at a 40:1 particle per cell dosage bares considerable similarity to that of the B3 control well, shown in burgundy, that contains DOX on its own. This would bring in to question whether the cells are dying purely due to the internalised DOX loaded micro-particles or whether the accumulation of DOX in the media over time is causing too harsh an environment for the cells to exist. If the latter is the case, the micro-particle delivery mechanism could be deemed ineffective. There is also a visible rise in the DOX level in well plate A3 beyond the 12.5 hour mark, a possible cause of which is identified later in this text.

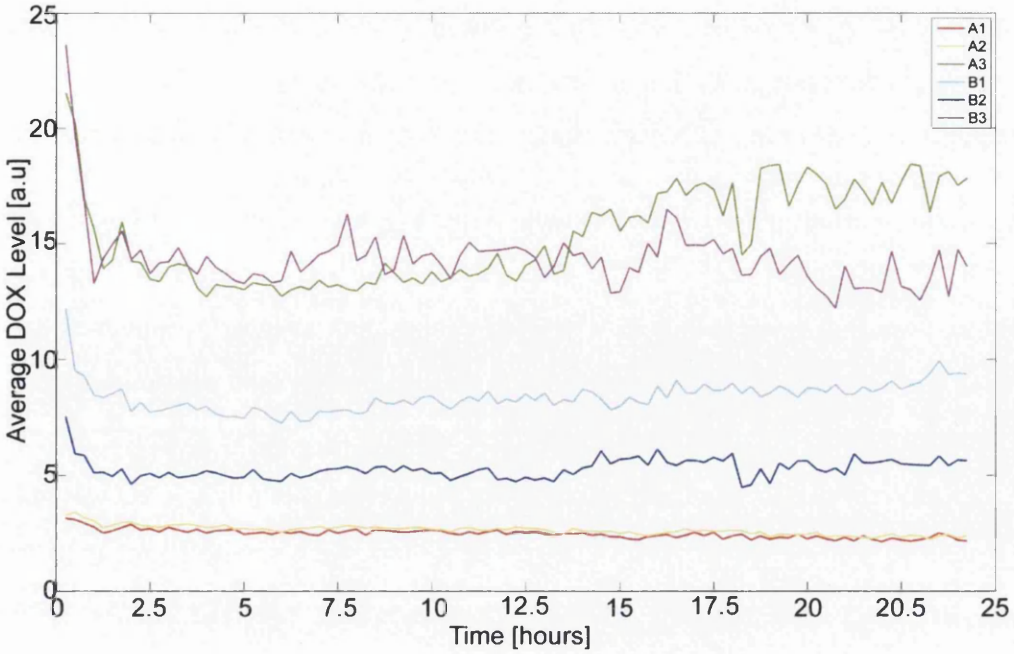



Figure 5.14: Comparison of average DOX background level across all observed wells. Background levels can be seen to increase with higher dosages as per expectation.

Figure 5.15 compares a brightfield image region (A) to its corresponding DOX fluorescence channel region (B). Dark regions in the fluorescence channel (B), as noted previously in figures 5.9 and 5.10, can be seen to correspond to pSi microparticles visible in the brightfield channel.

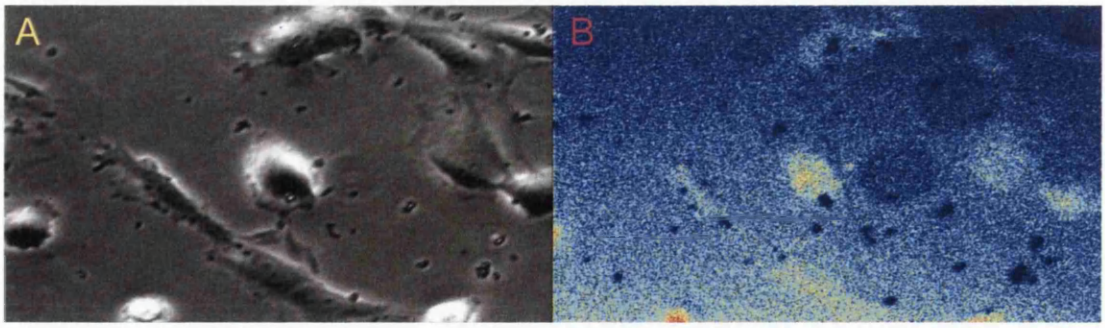


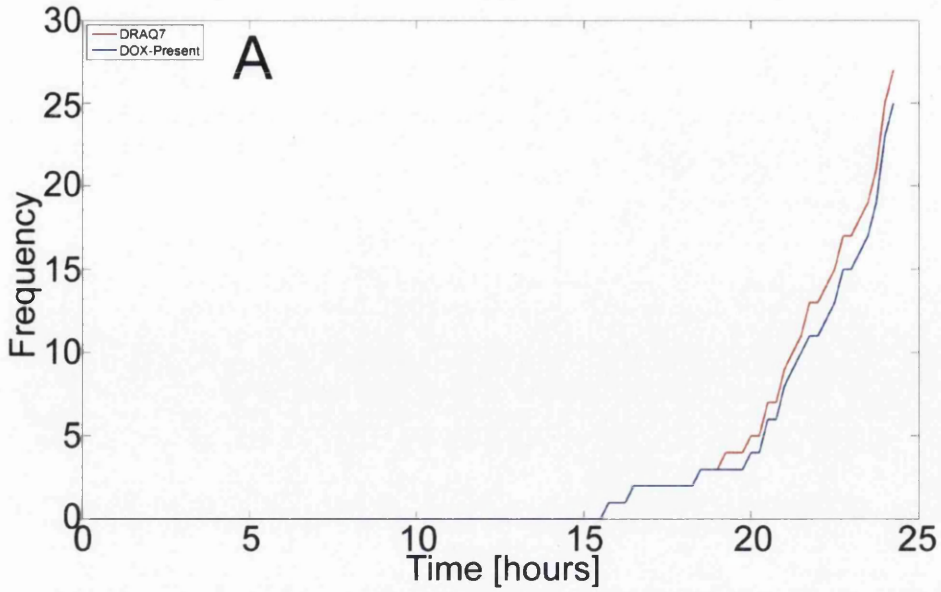

Figure 5.15: Comparison of a brightfield region (A) with the corresponding DOX fluorescence region (B). Dark patches in the fluorescence channel can be visually cross referenced against small particles in the brightfield image.

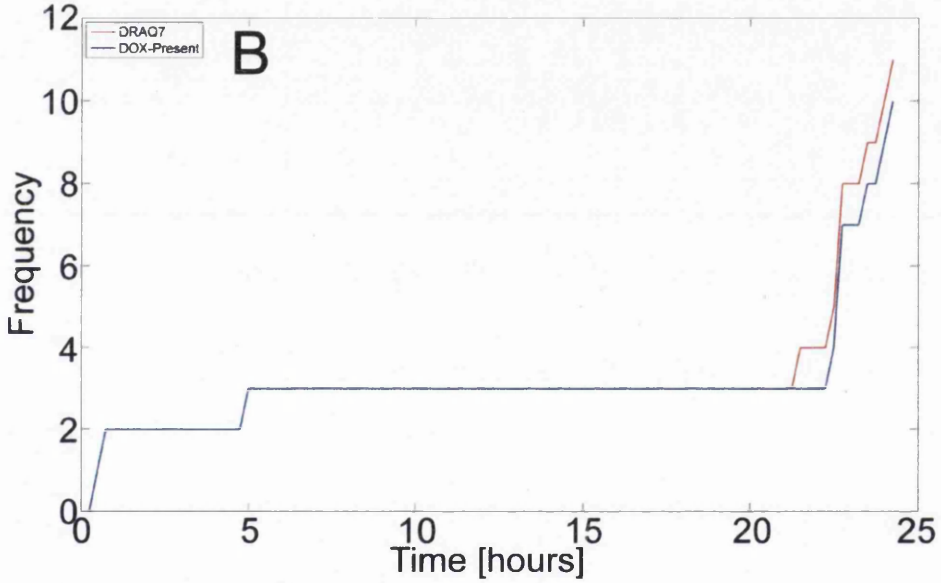
Their apparent blockage of background fluorescence can be taken as evidence that all measured background signal is sourced from released DOX and not DOX internalized within pSi particles. Visual tracking of these dark objects also reveal occasional disappearances when near or around cells with no way of knowing whether they have been internalized or have simply dropped beneath a source of

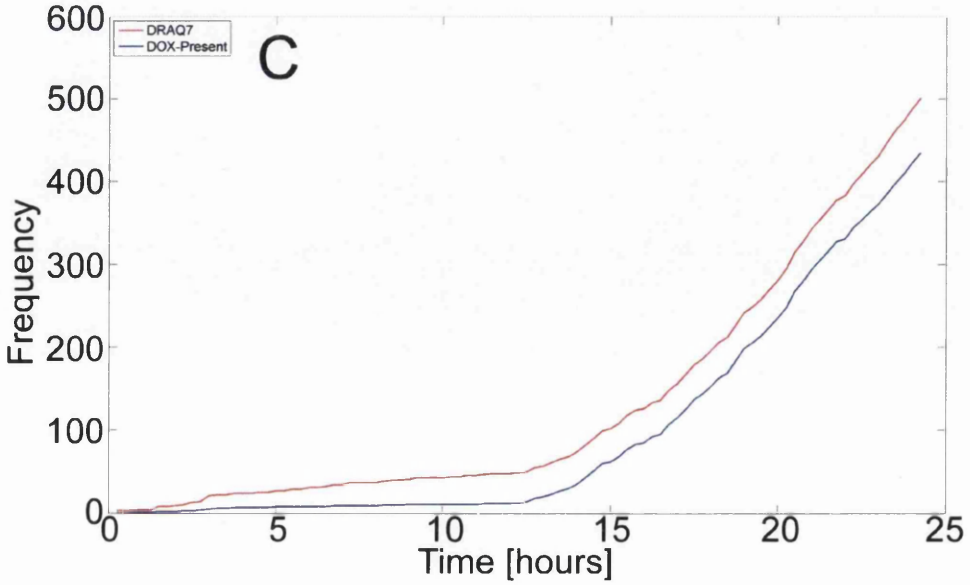
fluorescence. The fact they are only 'visible' in the fluorescence channel at early time points suggests that once internalised, the dispersion of DOX into the cytoplasm conceals their location as they are encapsulated by a source of fluorescence.

The earlier validation of DOX presence in the region of identified DRAQ7 signals results in the identification of DRAQ7 signals with no associated DOX signal. These DRAQ7 marked death events are declared natural death events as they have no apparent causal relation to DOX presence. Consequently, they are omitted from the profile extraction process that is the subject of the next section. Figure 5.16 (opposite) comprises three plots comparing the total number of DRAQ7 signals counted against those with a recognised DOX presence with (A) showing plate B2, a 1:10 cell-to-particle dosage, (B) showing plate B1, a 1:20 dosage and (C) showing plate A3, a 1:40 dosage and the predominant source of cell death. These plots are provided to demonstrate that the majority of cell death events identified using the DRAQ7 channel were attributable to the presence of DOX in the local region. Interestingly, the discrepancies between the red and blue curves begin at different times with the curves in (C) following each other in parallel beyond frame 20 or so and those in (A) and (B) matching exactly until much later in the timelapse sequence. The fact this disparity occurs early on before general stabilization, much like the other dose plots, suggests the occurrence of an aggressive event. As Doxorubicin is a mitotic inhibitor, a cell's susceptibility is highly dependent on its position in the cell cycle and at any given time point only a small percentage of a cell population will be in G2-phase. The early discrepancy in the DRAQ7 and DOX counts in the 1:40 dosage could therefore be a result of a larger number of early G2-phase cells dying as they are in a particularly vulnerable state and have had no time to adapt to the DOX presence. In such conditions cells could succumb to apoptosis triggered by levels of DOX that are overlooked by the protocols described here. The resulting observations would then indicate higher than normal natural cell death, as seen, despite the events being DOX casual.

Figure 5.16 (opposite): Comparison between identified DRAQ7 death events, red, and those then attributed to the presence of DOX, blue. All wells contain DRAQ7 and DOX loaded pSi particles with 1:10 (A), 1:20 (B) and 1:40 (C) cell to particle ratios.


DRAQ7
DOX-Present
A
Frequency
Time [hours]
0
5
10
15
20
25
30


DRAQ7
DOX-Present
B
Frequency
Time [hours]
0
2
4
6
8
10
12


DRAQ7
DOX-Present
C
Frequency
Time [hours]
0
100
200
300
400
500
600

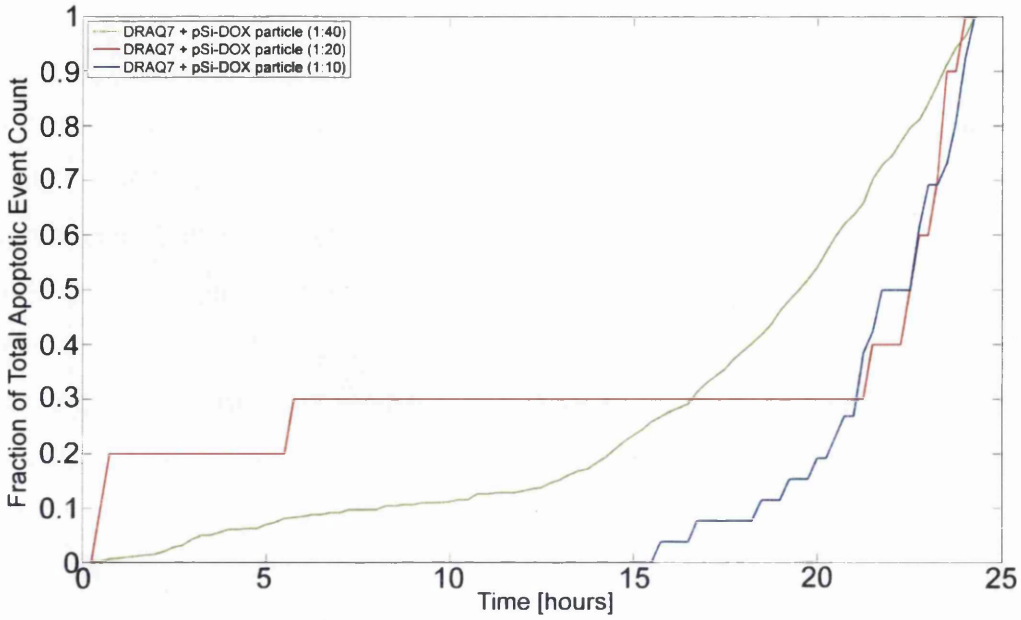



Figure 5.17: A comparison of DRAQ7 signals where DOX was also present for each of the three DOX loadings. Each curve is normalized to its respective maximum value to allow visual comparison of their shape.

The three blue curves in figure 5.16 are normalized against their respective total accumulated cell death counts and displayed alongside one another in figure 5.17 with green depicting the 1:40 dosage, red the 1:20 and blue the 1:10. The green curve can be seen to show two distinct gradient rises, one around the 2.5 hour time frame and another around 14 hours. As well as accumulated DOX levels, these two points of increasing gradients are also suggestive of a susceptibility to DOX induced apoptosis at particular time points in the experiment, namely cell cycle dependence. Although the lower numbers of overall cell death events makes the curve more discrete and therefore less useful for quantitative parameterization, this double gradient increase feature is also inferred in the 1:20 dosage curve, shown here in red, around the $5^{th}$ and $21^{st}$ hours. The lack of death events in early frames excludes the possibility of a full comparison against the 1:10 dosage however, as the blue curve comes into existence it does show similar characteristics to that of the 1:40 dosage when normalized again suggesting a particular region in the cell cycle where the cell is particularly susceptible to the effects of DOX.

Figure 5.18 represents the integral of 1:40 dose cumulative count and demonstrates a general two population shape typical of DNA histogram plots. The 'hump' feature around 7 hours could be interpreted as the G2/M-phase population in which DOX triggered apoptosis occurred early due to the susceptibility of their position in the cell cycle. The steady increase from 12 hours onwards is then representative of the rest of the population entering this G2/M phase and succumbing to DOX accumulations.

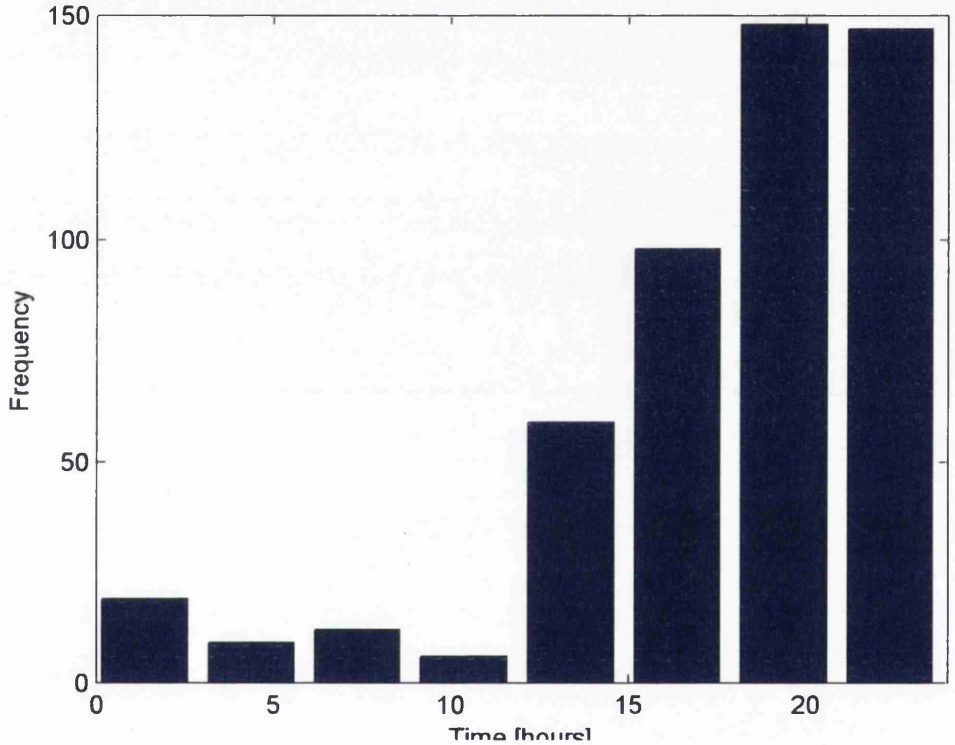



Figure 5.18: Alternative representation of the 1:40 dose cumulative death count depicted in figures 5.16(C) and 5.17. Each bar represents the total cumulative count over 3 hour intervals.

This DNA histogram representation ties back in with figure 5.17 where the 1:40 dose curve is seen to undergo the second phase of apoptotic events earlier than the two lower doses. Although the cells in each well plate are subject to the same susceptibilities and average cell cycle times, the presence of more accumulated DOX at higher doses could be expected to force apoptosis earlier in cells approaching G2-phase. It would appear from these plots that greater levels of internalised DOX widen its range of efficacy and vice versa.

## *Data Mining – Extraction of DOX Profiles*

To further investigate the effect of DOX in cells the previously generated and stored DOX profile masks are now called upon to create DOX intensity profiles. Each stored profile mask set contains a pixel list $\boldsymbol{L}$ for each time frame in the profile sequence with which the extraction of the DOX object intensities is a simple task. For a given DOX object pixel list $\boldsymbol{L_T}$, where T references the time frames within the stored profile mask sequence, the corresponding DOX channel image, $\mathbf{IMG_T}$, is loaded and the intensity of the DOX object is defined as

$$I_T = \sum_i \mathbf{IMG_T}(\boldsymbol{L_T}(i)) \qquad (5.16)$$

where $1 \leq i \leq |\boldsymbol{L}|$. It is worth noting that, by virtue of the fact that all profiles were previously tracked backwards from a DRAQ7 signal, all obtained profiles end at the point of cell apoptosis. The intensity profile vector $I_T$ is calculated for each time point T in the stored DOX object mask sequences and saved into a single vector set IP such that

$$\boldsymbol{I} = \{I_T^1, I_T^2, \ldots, I_T^N\} \qquad (5.17)$$

where N is the total number of intensity profiles generated from the n stored profile mask sequences. A compatriot set labelled **J** is generated alongside **I** taking the form

$$\boldsymbol{J} = \begin{Bmatrix} P_1, P_2, \ldots, P_N \\ T_D^1, T_D^2, \ldots, T_D^N \\ |I_T^1|, |I_T^2|, \ldots, |I_T^N| \end{Bmatrix} \qquad (5.18)$$

in which P stores the acquisition position of its corresponding intensity profile, $T_D$ refers to the time frame in which death occurred , the final time point in the profile, and $|I_T|$ is the profile length or duration. The set **J** allows for easy selection of profiles in terms of the parameters stored within it for display purposes. Figure 5.19 (opposite) contains a sample of 5 profiles randomly selected from acquisition position 9, a position in well plate A3 with a 1:40 cell to pSi particle dosage. The profiles displayed are smoothed using a 7-point moving average to remove the visual distraction of low level variation and better convey the overall curve features. A common feature in these profile curves is a reduction in DOX intensity immediately

prior to the occurrence of cell death. These reductions in DOX levels could be the result of the cell beginning to break up thus allowing internalized DOX to escape and dilute into the media background prior to complete cell death triggered DRAQ7 expression. Such a process could also account for the rise in background seen previously in figure 5.14.

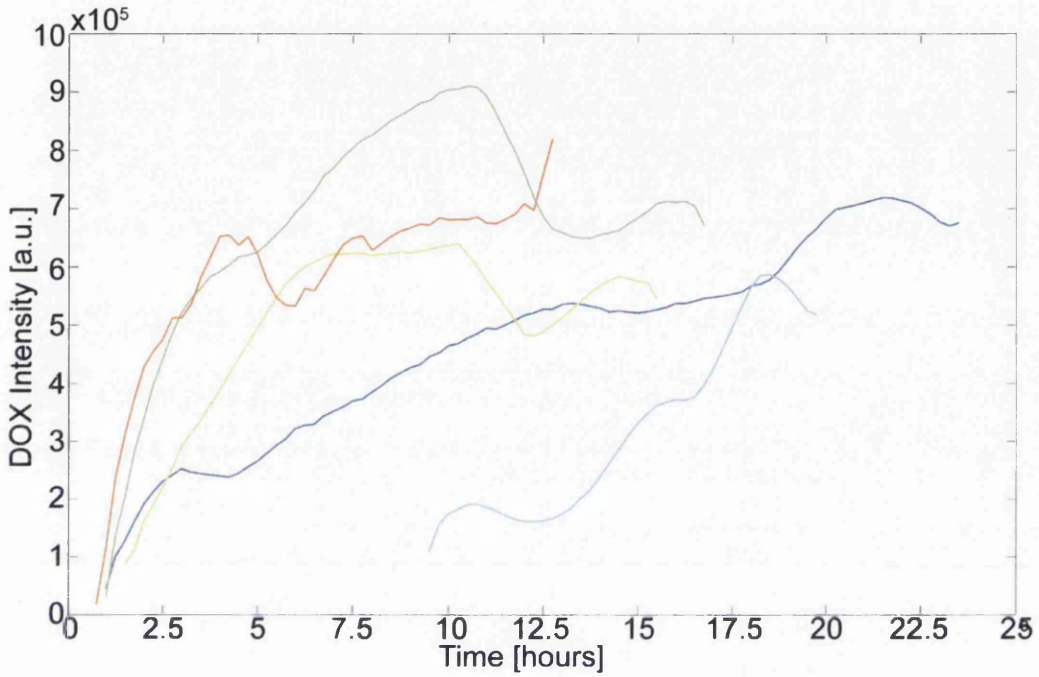



Figure 5.19: Sample of DOX profiles extracted from the full set over the entire time range.

Manual observations to check for such phenomena provide some further insight. Most cells exhibiting a profile dip show a visible blurring a the edges of DOX object shape indicative of DOX dispersion into the media. Some however provide no such indications suggesting in some cases there may be an alternative cause. Cells that did not exhibit a signal blurring often noticeably decreased in size and cross referencing against the corresponding brightfield images showed a significant 'balling-up' of the cell in the final frames. It is therefore possible that some of the fluorescence signal is lost through morphological changes condensing the fluorescence volume and shielding the lower levels from view.

Figure 5.20 (overleaf) provides further profile samples with selections made based on the temporal point of death over the time frames 1-20 (A), 20-60 (B) and 60-97 (C). The plots in (A) reveal two very short profiles, shown in red and light blue. Such instances might invoke the interpretation of particularly rapid cell death on exposure

to DOX loaded micro-particles however their short duration and proximity to the initial time point results in too few data points for meaningful conclusions to be drawn. The longest duration curve in (A), shown in dark blue, demonstrates an extremely sharp decrease in DOX level between the final observation point and the one preceding it.

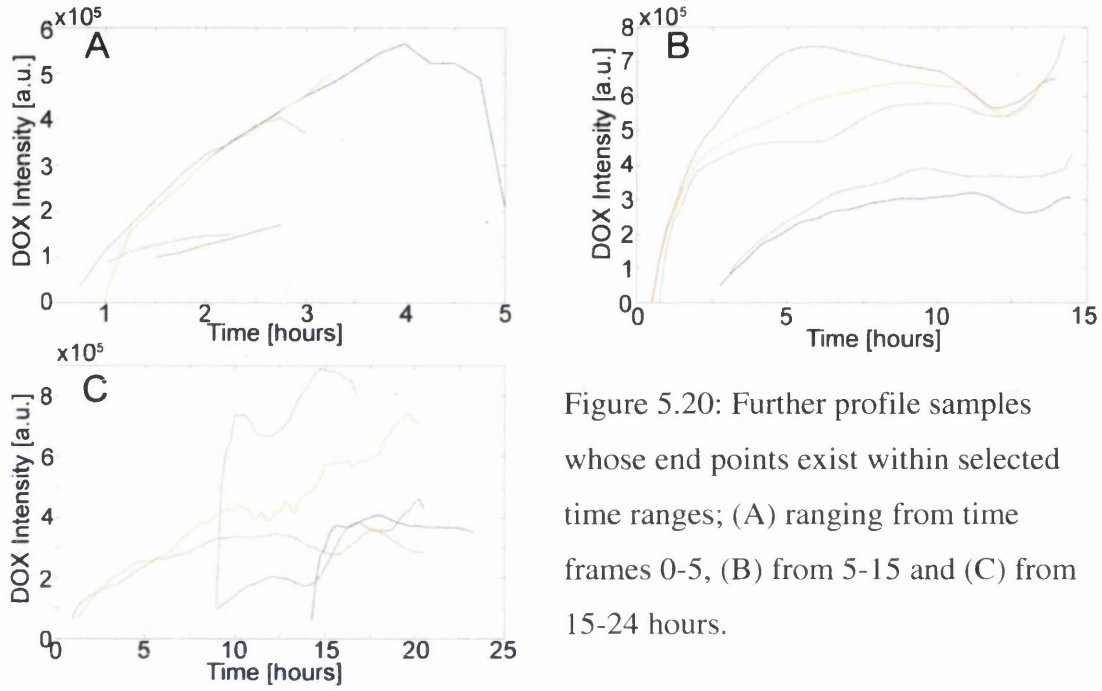



Figure 5.20: Further profile samples whose end points exist within selected time ranges; (A) ranging from time frames 0-5, (B) from 5-15 and (C) from 15-24 hours.

The cell membrane could have severely deteriorated between observations with the final point only existing to bear witness to the resulting DRAQ7 fluorescence. The curves in (B) all follow a similar characteristic pattern with the major difference being that of overall intensity level between the lower two and the upper three. Such discrepancies could be a result of differing numbers of DOX loaded micro-particles being internalized by the cells, 3 particles in the lower curves and 8 in the higher ones for example, resulting in differing DOX accumulation rates. While pSi particles appear visible in the brightfield throughout and can be inferred in the early time frames in parts of the DOX fluorescence images, identifying the number in a given cell directly is not possible with any level of accuracy. Objects visible in the brightfield images appear to be on the surface of cells rather than being internalised (see figure 5.15(A)) and suggest the same to be true of the small number that can be inferred from the fluorescence channel. Detailed analysis of the dispersion patterns in the early stages of DOX accumulation could theoretically infer the number of dispersion sources within a region. However, as the optimal time to observe the

diffusion of DOX from internalized pSi particles coincides with early, low levels of DOX signal, observing the diffusion of DOX directly is not possible. The time intervals of this particular data set also make it difficult to decipher anything meaningful and the fluorescent objects appear to simply increase in brightness across their entire area rather than diffuse from particular sources. The potential uptake of additional pSi particles throughout the timelapse could somewhat nullify the impact of knowing the number of internalised DOX sources at the beginning. Observed differences in profiles could reflect differences in uptake patterns between cells, a more motile cell for example may expose itself to more opportunities for particle uptake than one that remains localised. Curiously however, such differences do not appear to result in a more rapid capitulation to DOX induced cell death with all 5 profiles existing across the selected time range. In both profile sets the DOX signal appears to reach a saturation plateau well in advance of the final apoptotic profile point. This can be explained by the finite number of pSi particles present in the environment and the fact that this plateau occurs at multiple levels suggest the majority of pSi uptake does in fact occur early on. Future timelapse sequences with shorter acquisition intervals at early time points could potentially infer the pSi content within a cell through analysis of the dispersion patterns that, as previously mentioned, are not identifiable over 15 minute intervals. The delay after apparent DOX saturation prior to cell death is further indication of the focused efficacy of DOX towards a specific stage of the cell cycle. The profiles depicted in figure 5.20 (C) show similar style increases with two profiles in particular, both shown in shades of blue, showing rapid initial increases in DOX intensity. Remembering that these profiles were identified by tracking in reverse, such an increase could arise as a result of cell migration across the boundary of the field of view, causing a tracked cell to seemingly loose, and hence gain with the arrow of time restored, DOX signal as it passes out of sight. Through manual checking the light blue profile shown in figure 5.20 (C) is found to correspond to a cell passing into the field of view.

Figure 5.21 (opposite) shows a single profile above a corresponding sample set of DOX channel images coloured using a heat style colour map for clarity with their respective times labelled in hours. This specific profile was chosen to demonstrate the possible effects of a mitosis event on an extracted DOX intensity profile. The profile curve rapidly increases initially up until around 2.5 hours beyond which the DOX level increases more steadily. At approximately 8 hours the curve undergoes another sharp increase, this feature coincides with the images labelled with T = 8, T = 9 and T = 10 where two distinct objects can be seen forming whilst retaining contact with one another. As the two daughter cells emerge from the mitosis event they gradually expose a surface area with a total summation greater than that of their progenitor cell providing more area in which internalized DOX fluorescence can be directly observed. This causes an apparent increase in the DOX level of the extracted profile until the point where the two daughter cells fully separate, seen in the image labelled T = 10.5. At this point the profile curve dips and the DOX intensity is seen to drop between 10 and 11 hours as the two daughter cells share the internalized DOX and only one is tracked. Figure 5.22 (next page) shows the manually acquired DOX intensities of both daughter cells after mitosis. Addition of the intensities of the manually tracked cell to the automatically profiled one reveals a continuing rise in the cumulative DOX levels. The increasing levels of DOX in both daughters with almost immediate effect suggest both acquired pSi DOX sources during the progenitor division. The differing levels show the dilution of DOX across the daughter pair to be uneven with the automatically profiled cell reporting approximately double the DOX content of the manually retrieved daughter. The near parallel increase of DOX intensity in each daughter cell after 12 hours is interesting, suggesting that similar numbers of DOX release vectors are present in both. These daughter cells provided another interpretation obstacle by making contact with one another at time T = 14, seen at the end of the profile as a sudden rise in intensity levels. Such events in this regime make it impossible to automatically verify the continued tracking of the cell that triggered DRAQ7 fluorescence through the time sequence.

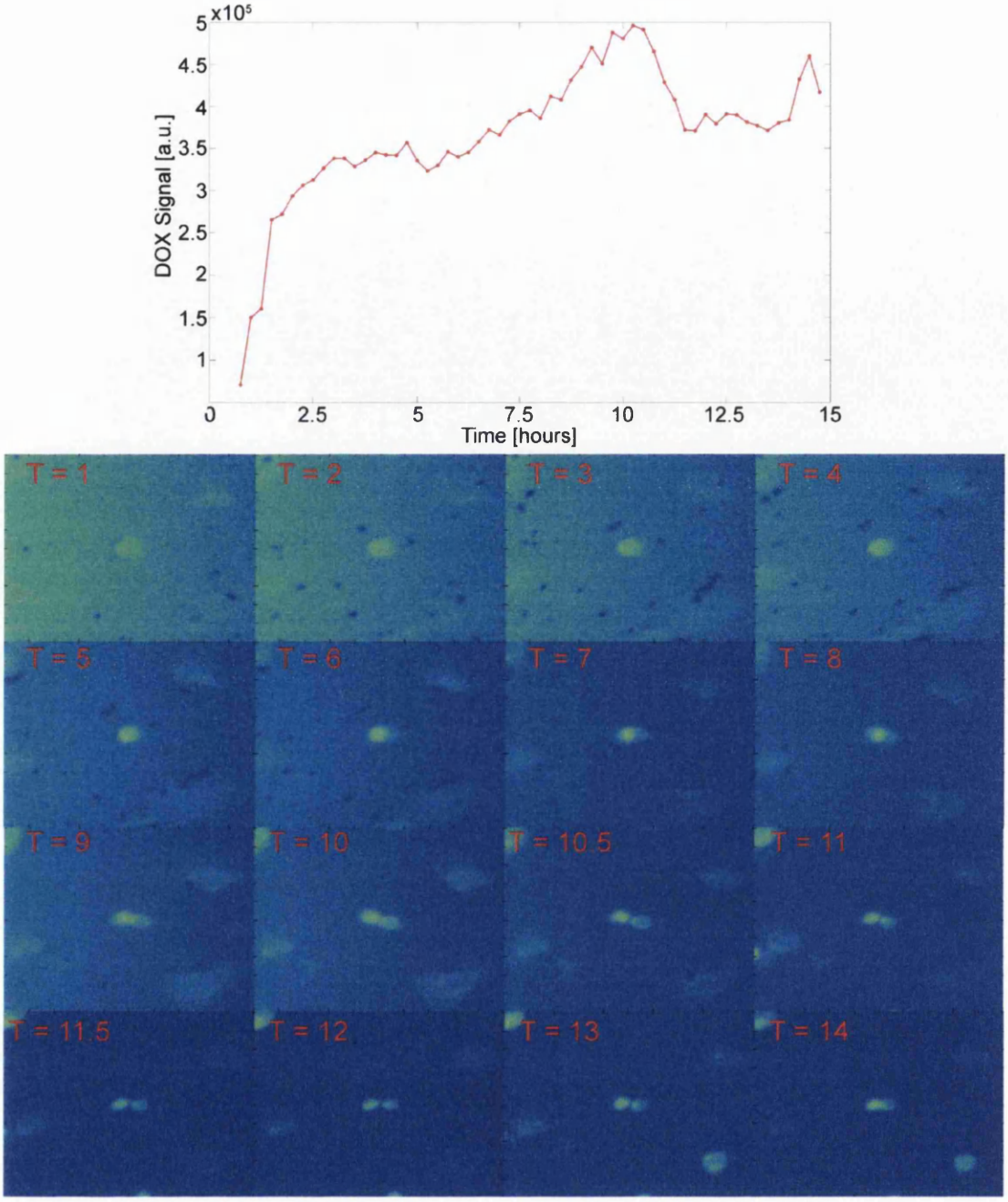



Figure 5.21: Example profile with image sequence of the corresponding DOX track showing the occurrence of a mitosis event and its consequential effects on the shape of the DOX profile. (Displayed in auto-scaled colour-map to improve clarity of centralised object)

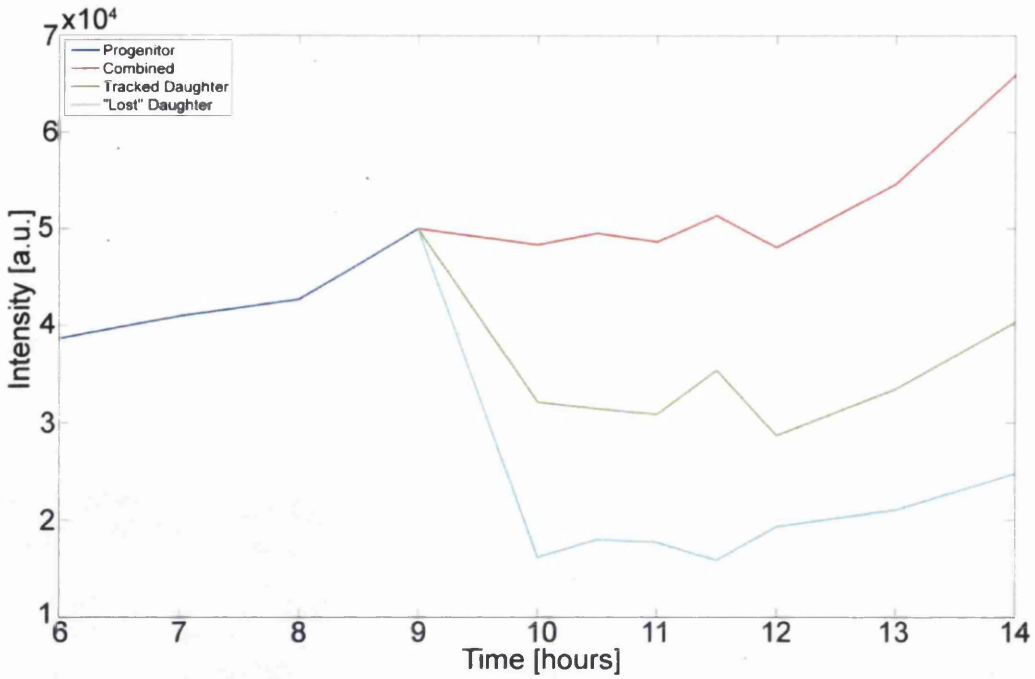



Figure 5.22: Complete trace of the division event shown in figure 5.21. At the point of division, the 'lost' daughter is manually traced to capture the dilution ratio and combined DOX levels.

The profile shown in figure 5.23 (opposite) depicts another example of the resulting effects of paired cells separating. In this case it is not clear whether the final stages of a mitosis event are observed early in the time sequence or whether two unrelated cells in close proximity are migrating apart. Manual cross referencing against the relevant brightfield images revealed this particular event to be two separate cells interacting with one another; visually distinguishing them was a non trivial task. One of the cells appears to attempt to move over the top of the second cell before 'balling-up' and succumbing to apoptosis. The almost discrete drop in the profile coincides with the two cell objects separating from one another with the tracked object maintaining a consistent DOX level until approximately 12.5 hours when the levels again begin to rise. Comparing the two sample profiles against one another it is worth noting that although the profile maxima differ significantly there is a similarity in the final observed DOX levels suggesting the presence of an intensity threshold beyond which the cells cannot cope with the level of internalized DOX during susceptible phases of the cell cycle.

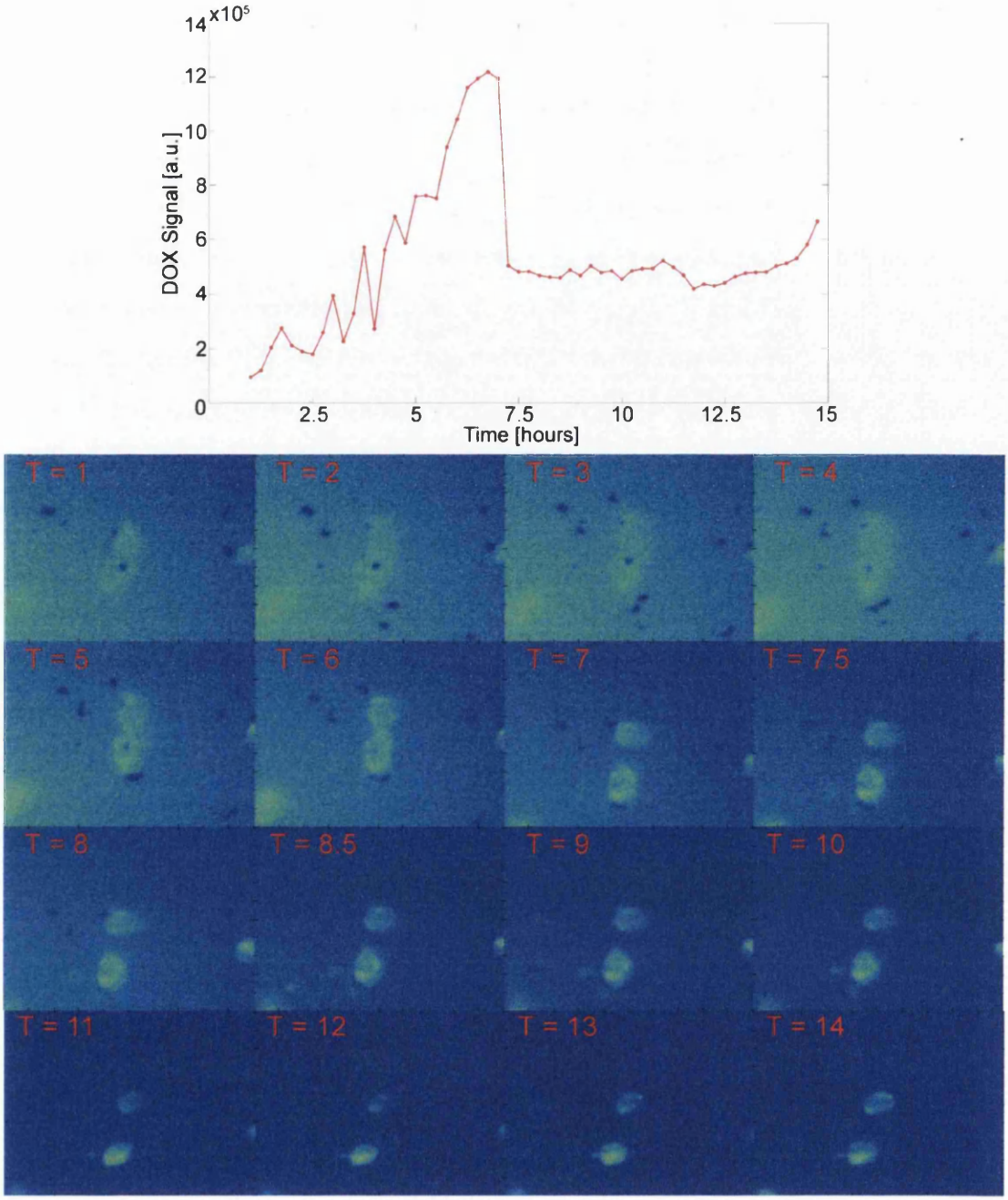



Figure 5.23: Another profile example showing the effects of close proximity neighbours migrating apart causing a near discrete-like drop in the profile. (Displayed in auto-scaled colour-map to improve clarity of centralised object)

With the apparent similarities in final DOX levels it is instructive to compile a data set of the final nodes of all the extracted profiles to compare apoptosis inducing DOX levels against the time of appearance of DRAQ7 fluorescence. Using the data structures previously defined, it is a trivial task to compile such lists based on acquisition position and/or well plate. The results of the data set compiled for the A3 well plate, the one exhibiting the vast majority of DOX induced cell death, are displayed as scatter plots in figure 5.24 (opposite). The scatter plot in (A) contains all the end points of the profiles extracted from positions 9-13 (1:40 dosage). A dense band of points in the lower right quartile conforms to the idea of a threshold beyond which the majority of cells are unable to tolerate the DOX level. The single outlier point with a DOX level far in excess of all others shown was manually verified to be representative of a DOX object contained within a single cell. With an apparent DOX tolerance significantly higher than the majority of the cell population, said cell demonstrates the potential for subpopulations within the data set. The points plotted in (B) represent the average DOX intensity values for each time point. This dampens the visual effect of variation and the inability to meaningfully interpret the patches of high density scatter points in (A) and reveals the potential for a smooth curve fit to characterise the correlation between the time of the profile end point (time of cell death) and intensity of the accumulated DOX. This distribution also hints at two groups existing across the temporal dimension, one up to around 5 hours and the other beyond 12.5 hours. This could again be resultant of the differing susceptibilities to DOX throughout the different stages of the cell's cycle and could be further investigated by fixing the cell cycle starting point in a future experimental assay. The accrued DOX levels are further displayed in figure 5.25 (next page) to detail the distribution of the final intensity values. Outlier objects can be seen on the trailing edges of the overall bell curve shape with the aforementioned outlier object visible on the far right, well clear of the next highest observed intensity. Here, objects in the band between the red boundaries account for 90% of the 435 total profiles extracted across positions 9-13 with the most common value of DOX intensity associated with cell death lying around $1\times10^5$ intensity units.

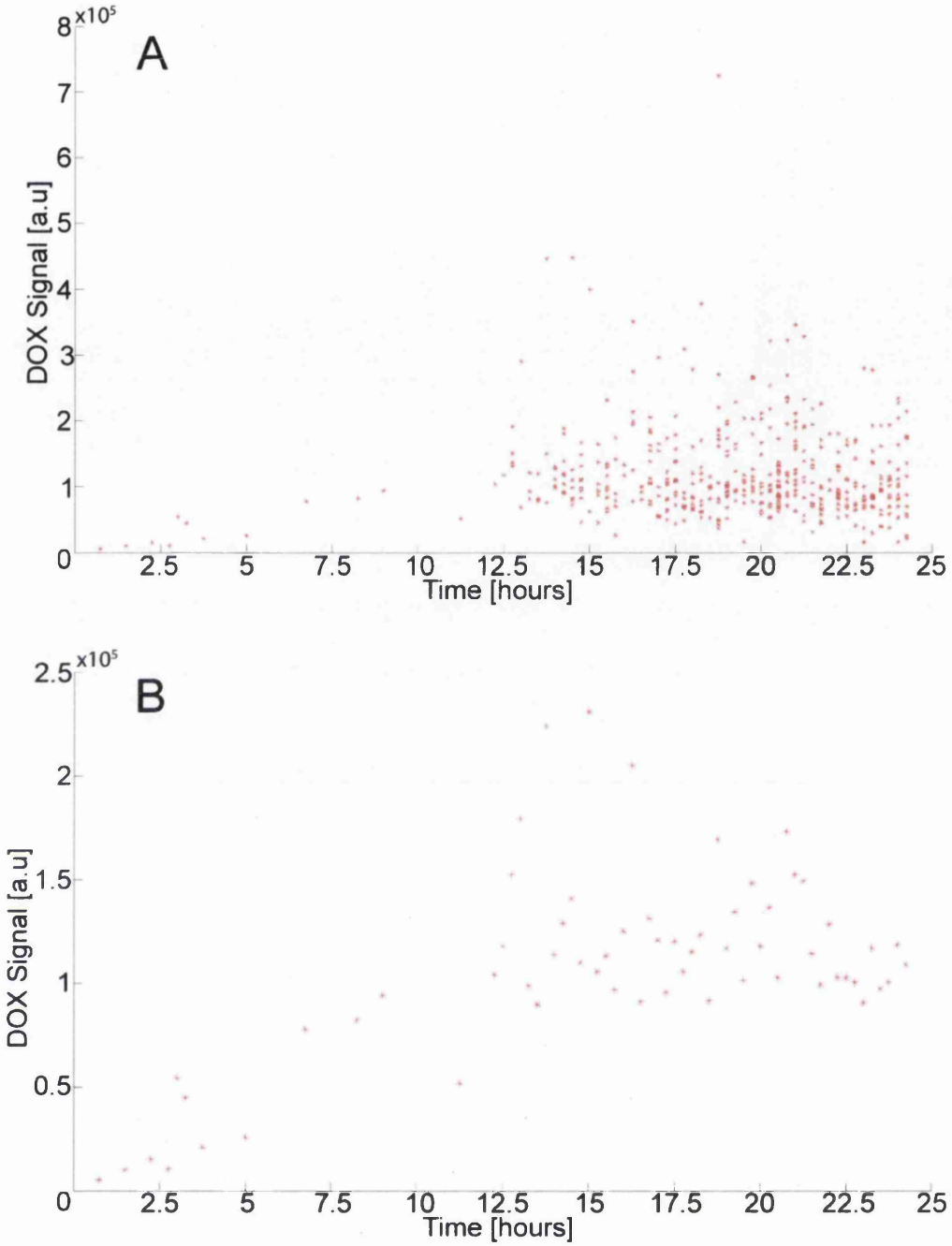



Figure 5.24: Scatter plots comparing the identified DOX signal level to time of cell death for the A3 well plate (1:40 cell to pSi particle dosage). (A) depicts the end points of the complete set of identified profiles while (B) represents the average DOX signal for each time point where death events occurred.

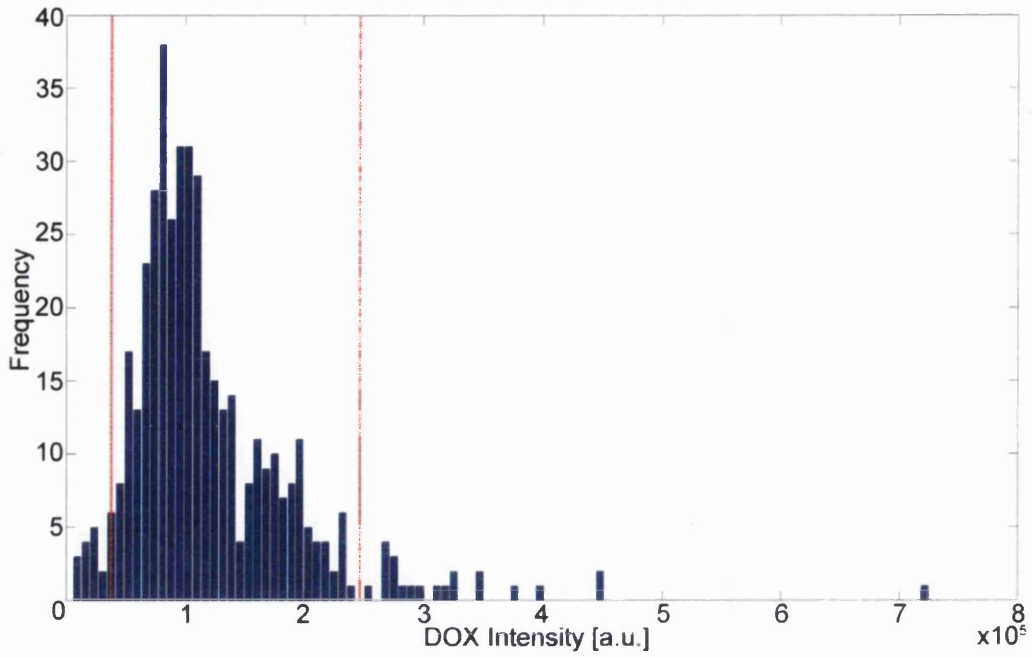



Figure 5.25: Histogram of DOX levels at the final point in the extracted profiles. The region between the red lines contains 90% of the endpoints in the profile set compiled for the A3 well plate (1:40 cell to particle dosage).

Similarities in the final DOX levels across the profile sets could reflect a generally even allocation of pSi particles across the cell population. Evenly distributed numbers of internalised pSi particles containing a finite amount of DOX would have an upper bound on the total accruable DOX level they can provide. As the DOX, contained in concentration in the pSi particles, reaches equilibrium through diffusion the observed fluorescence signal would plateaux, as seen. With the maximal DOX level reached, it would be expected that the cell would continue through the cell cycle until reaching the point where the presence of DOX is particularly cytotoxic, namely the G2-phase. At this stage the cells would rapidly succumb to apoptosis and the death count would rise in accordance with the population's procession through the cell cycle. Additional experiments to characterise the cell cycle of this cell line could be used to add further weight to this hypothesis.

# CHAPTER SIX

# Additional Application – PEI-QD Toxicology

Following the analysis principles laid out in the previous chapter, a second data set is similarly analysed to extract information relating to the dosage of the assay variable at the time of cell death; in this particular case the experimental run was conducted to establish the toxicity over time of polyethylenimine (PEI), and quantum dots (QDs), used here as a fluorescence source for imaging. The level of QDs present in and around the HFF cells at the point of cell death is used to quantify the toxicity of the dosages provided. This experimental run was also a basis for optimising the methods used to gain such data in the future and the automatic analysis conducted here offered the opportunity to investigate a complete set of images that would otherwise have been disregarded in favour of newer runs.

## *The Data Set*

The provided sample data set consisted of images across 3 channels for 58 fields of view acquired from a single well plate containing human foreskin fibroblast (HFF) cells and polyethylenimine (PEI) coated quantum dots (QDs), at a concentration of 6nM, and DRAQ7 for use as a marker for cellular death events much the same as the subject of the previous chapter. One of the three channels records the DRAQ7 signal, used to identify cell death, another the quantum dot fluorescence signal, the subject of dosage analysis, and the remainder a brightfield image record which is not utilised during this analysis procedure. For each time point in the timelapse sequence a control image was also provided for the DRAQ7 and QD fluorescence channels taken from an empty well plate. The timelapse sequence covered 20 hours at single hour intervals between 1 and 7 hours and two hour intervals from 7 to 21 hours resulting in 14 temporal acquisition points in the sequence. This change of intervals was necessary to continue the timelapse for the full duration whilst reducing the number of stored files, a requirement induced by a lack of available storage space on the hardware at the time of the experimental run.

### *Image Pre-Processing and Foreground Isolation*

Previous image sets have required an image cleaning stage to characterise and remove background noise to enable extracted signal values to accurately represent true fluorescence levels. Conveniently for this data set a corresponding set of control images can be utilised to remove the system hardware imparted background noise by simply subtracting the pixel values of the control images from those of the corresponding data image.

$$ {}^{T}_{j}IMG_{cleaned} = {}^{T}_{j}IMG_{raw} - {}^{T}_{j}IMG_{control} \qquad (6.1) $$

Here j refers to the two image channels, DRAQ7 and QD, and T references the time point indices. Such a calculation is computationally cheap and is therefore dynamically performed as required throughout the DRAQ7 identification and QD signal isolation processes. DRAQ7 identification process is performed exactly as detailed previously (see *"Chapter V – Death Event Identification and DOX Presence Validation"*) with differences in the parameters relating to cell characteristics such as average cell diameter and an interest in the presence of QD fluorescence signal rather than DOX. The QD signal objects are identified as the image foreground using the same simple step function filter utilised previously to isolate the QD signal image foreground in chapter two, namely for the QD signal image I,

$$ F_S(x,y) = \begin{cases} 1 \; if \; I(x,y) \geq \theta \\ 0 \; if \; I(x,y) < \theta \end{cases} \qquad (6.2) $$

where $F_S$ is the image foreground and θ is again the pixel intensity threshold defined here as

$$ \theta = \mu_{image\ intensity} + \sigma_{image\ intensity}. \qquad (6.3) $$

With the interest here being to count the total quantum dot fluorescence signal present within a cell at the time of its apoptotic reaction it is unnecessary to make attempts to isolate individual QD signal objects as was the case in chapter two when identifying QD loaded endosomes. The isolated foreground is treated from this point

on much as the DOX foreground objects were in the previous chapter, simply as objects for potential use as pixel intensity masks when compiling the total QD signal associated with a cell death event. The panels in figure 6.1 detail an image crop sample of a DRAQ7 signal event, (A), alongside the corresponding region in the QD fluorescence channel, (B) and a binary representation of the QD signal foreground, (C).


Figure 6.1: Sample image crops comparing a DRAQ fluorescence event (A) with the corresponding region in the quantum dot fluorescence channel image (B) and the resultant binary representation post-filtering process (C). The image in (C) demonstrates the sparse spread of QD fluorescence sources surrounding the cell.

Here the binary representation demonstrates the apparent randomness of the QD distribution across the region, more precisely it is clear that the quantum dot fluorescence sources cannot be seen to define the basic morphology of the cell making the prospective extraction of a QD signal profile akin to the DOX profiles of chapter five far more complex. Consequently only the QD signal at the point of death is obtained in this chapter. For each identified DRAQ7 event, a circular perimeter is drawn on the QD channel image about the DRAQ7 object centroid coordinates with a radius of 30 pixels. This radial choice is based on the manually observed average cell diameter of 50 pixels plus 20% to allow for the variation in cell sizes (see *"Analytical Outputs"* for further discussion). Any QD signal objects with pixels within this perimeter are deemed associative to the cell and are stored along with the DRAQ7 centroid coordinates, $C$, the field number, $F$, and the time frame index, T, into a death signal set $\boldsymbol{D_S}$ such that

$$\boldsymbol{D_S} = \{\boldsymbol{O}, C, F, T\} \tag{6.4}$$

where

$$\boldsymbol{O} = \{O_1^{QD}, O_2^{QD}, \ldots, O_n^{QD}\}. \tag{6.5}$$

$O_i^{QD}$ references the pixel lists of the individual QD signal objects identified within the perimeter, $1 \leq i \leq$ n and n is the number of QD signal objects deemed within the aforementioned perimeter. A set in this form is constructed for every identified DRAQ7 signal event and the values within these stored variables are displayed throughout in the next section. Referring back to figure 6.1 (C), it is noted here that this simplistic approach can be deemed less accurate than counting the intensities of all pixels within the cell membrane but it is as consistent. Given the complex nature of the cell boundary of these fibroblast cells, the only accurate way of segmenting such cells is manually and consequently, this method offers the realistic possibility of operating on large, high throughput data yields, something beyond the practical reach of a manual operator.

## *Analytical Outputs*

Figure 6.2 (opposite) compares the total resultant DRAQ7 counts across time with those found to have QD fluorescence signal within the observationally acquired average cell diameter of 50 pixels. The final cumulative sum of the identified DRAQ7 reaches 3,328 events across the 58 fields over the 20 hour temporal range with 83% of these, 2,768, occurring within the proximal range of a single cell diameter to quantum dot fluorescence objects isolated to the image foreground. The separation of these two counts occurs steadily and consistently across the 20 hour time range which is suggestive of cells with no associated QD signal succumbing to overall environmental conditions that are unrelated to the localized build up of QD particles. Some of the cell death events with seemingly no associated QD signal presence could however have been potentially affected by QD particles existing in or around their occupied space by virtue of two possibilities. Firstly, it is feasible that within the 1 and 2 hour intervals a cell could succumb to the toxicity of localized QD particles, trigger the apoptotic cascade and begin to break down releasing said

internalised particles. By the time the cell is again observed the QD particles could have released in sufficient quantities to drop the total fluorescence level below the foreground filter whilst retaining enough traces of DRAQ7 to be identified as a cell death event. The likelihood of such eventualities could be lessened with smaller intervals between acquisitions pending the availability of storage capacity. Alternatively, some internalized particles could potentially exist within a cell but outside the search radius due to the variation in morphology of these fibroblast cells. This is far less likely, particularly at later time points, as the majority of internalized particles are observed to exist near the cell nucleus and the numbers of particles that are on occasion observed away from the nucleal centre would not equate to a QD intensity of significance.

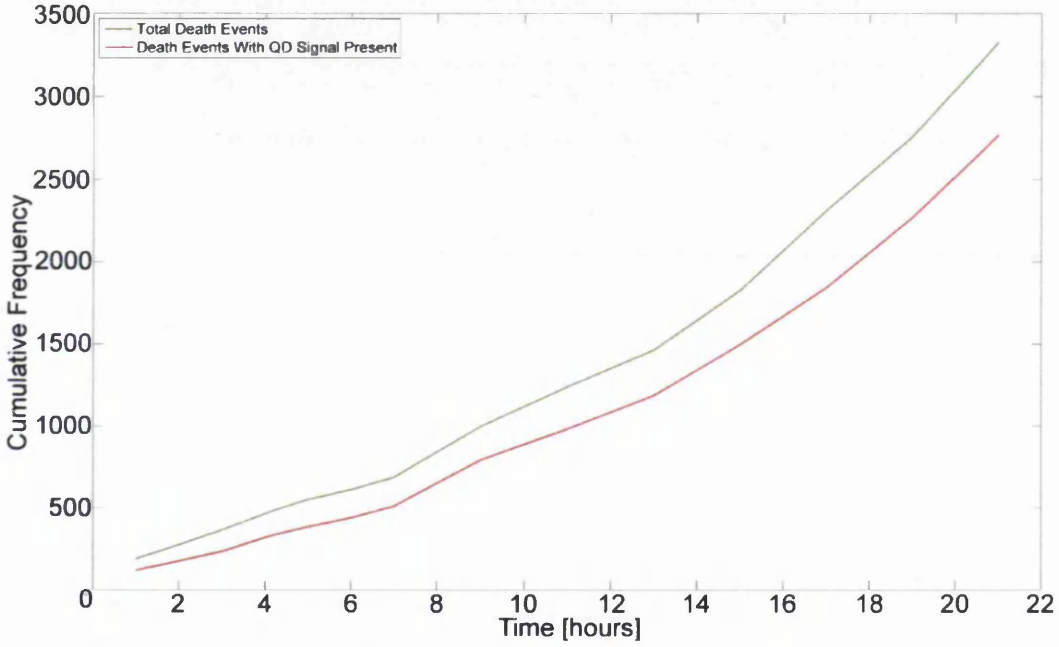



Figure 6.2: Comparison of total counted DRAQ7 events and DRAQ7 events with associated QD signal within an observed average cell width.

Figure 6.3 contains three panels showing three cells as seen through the brightfield channel with the overlaid scale in pixels.

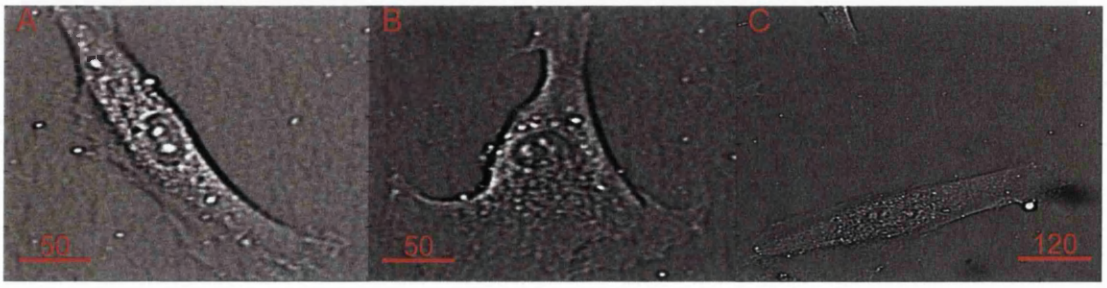



Figure 6.3: Sample of brightfield cell images showing the variation in morphological configuration and overall size.

The cells in (A) and (B), whilst having rather different overall morphology, exhibit an observed diameter that can be estimated as the generally observed average of 50 pixels, (A) being on the thinner and (B) the thicker ends of that scale. The cell in image (C) however details the possible variation of overall cell size achievable by fibroblast cells as they migrate around the field of view. With DRAQ7 being a nuclear stain, the radial search criterion acts about a central coordinate located within the nuclear region of the cell. The elongated shape of the cell in panel (A) could potentially harbour QD particles outside the 30 pixel radial search and the cell in panel (C), with an end to end length in the order of 360 pixels, has an even larger area in which QD particles could potentially evade the search protocol. The choice of the average cell radius is made by opting to ignore the extended cell membrane lengths seen in (A) and (C) and instead focus on the width of the cellular "bulk" where the nucleus is located. This is deemed valid as extending the search radius to account for the stretched cell morphologies observed would result in large overestimates of internalised QD particles for the following reasons:

1) The area of an elongated cell is grossly overestimated by a circle encompassing its two furthest points
2) The majority of internalised objects are observed to accumulate near and around the nucleal centre
3) Prior to apoptosis, a cell under environmental strain typically condenses into a smaller size and exhibits a reduction in motility. (extensions being the trait of a mobile, healthy cell)

Figure 6.4 shows the observed propensity for internalised objects to occupy the near nucleal region by compiling the accumulated QD fluorescence signal within varying radial distances of 45 to 120 pixels about the DRAQ7 signal centres identified in a single field of view across the 20 hour time range.

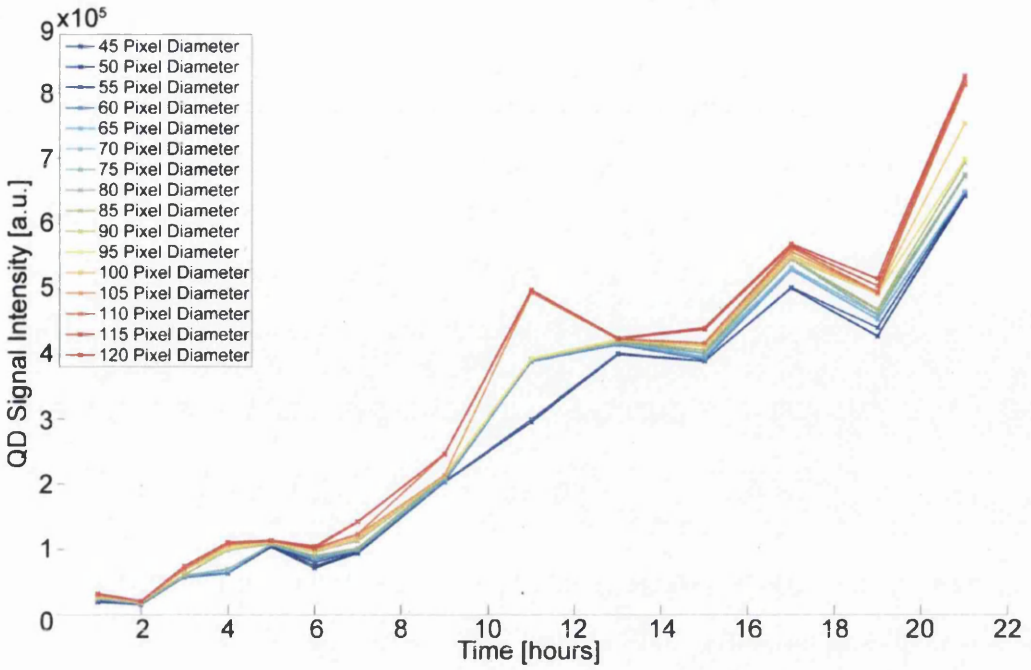



Figure 6.4: Plot showing consistent rise in the average accumulated QD signal at each time frame across the whole observed field range.

While there is an increase in total signal intensity as the search window is dilated, it can be seen to be disproportionate in comparison to the respective radial changes. Doubling the radius from 60 to 120 pixels for example does not result in anything near a doubling of the compiled intensity. The differences in extracted QD fluorescence intensity across the range of radial dilations can be seen to increase over time with the initial points being tightly grouped and the final points less so. This is accentuated by their differing y-axis ranges, with the relative proportions of accumulated QD fluorescence intensity compiled for each of the radial dilations being more consistent across time than is visually apparent, but is also partly down to a general increase in background fluorescence. As cells undergo apoptosis and break down particles that had been internalised escape back into the media. Previously internalised groups of quantum dots can remain together in numbers sufficient to allow their resultant fluorescence signal to penetrate the step function noise filter and be identified as QD signal objects. These objects would occupy the regions outside remaining cells; consequently increasing the radial search window would result in counting more of these objects and the effect on the total compiled QD signal intensity becomes more apparent at larger radii and insignificant at radii enveloping

only the nucleal core of the cell. It is also of note that there is no apparent plateau in the increase of QD signal intensity as the radial search window is dilated which is indicative of external QD fluorescence objects, those outside the cellular membrane, having no obvious propensity to localize around the cell itself. This validates the somewhat arbitrary selection of adding 20% to the observed average cell diameter in so much as such a low level increase is unlikely to falsify the resulting values of QD signal intensity per cell death event by virtue of gross overestimation.

With all the DRAQ7 events and their associated quantum dot signals compiled into multiple $\boldsymbol{D_S}$ sets it is possible to display a statistical overview of the complete 58 field, 20 hour timelapse. The histogram distribution in figure 6.5 (opposite) shows the spread of associated quantum dot intensities for all identified death events across all fields and all time frames. The significantly larger proportion of counts on the low end of the intensity scale could be indicative of a subpopulation of cells with a lower tolerance to the presence of internalised QD particles. They could also potentially be a subpopulation of cells that contained QD particles but which underwent apoptosis for unrelated reasons resulting in death with a low QD signal level although if such a high proportion of cells were dying of non-QD related causes it could be expected that there would be a bigger difference in the total number of DRAQ7 events and those with QD association (see figure 6.2). The staggered and somewhat discrete nature of the decline in numbers as the QD intensity rises could also be suggestive of subpopulations of cells with differing tolerances to levels of internalised QD particles. The general distribution however shows the majority of cells succumbing to apoptosis with QD intensity levels below a value of $10^6$ units and the numbers able to sustain beyond that decreasing with higher QD levels as would be expected.

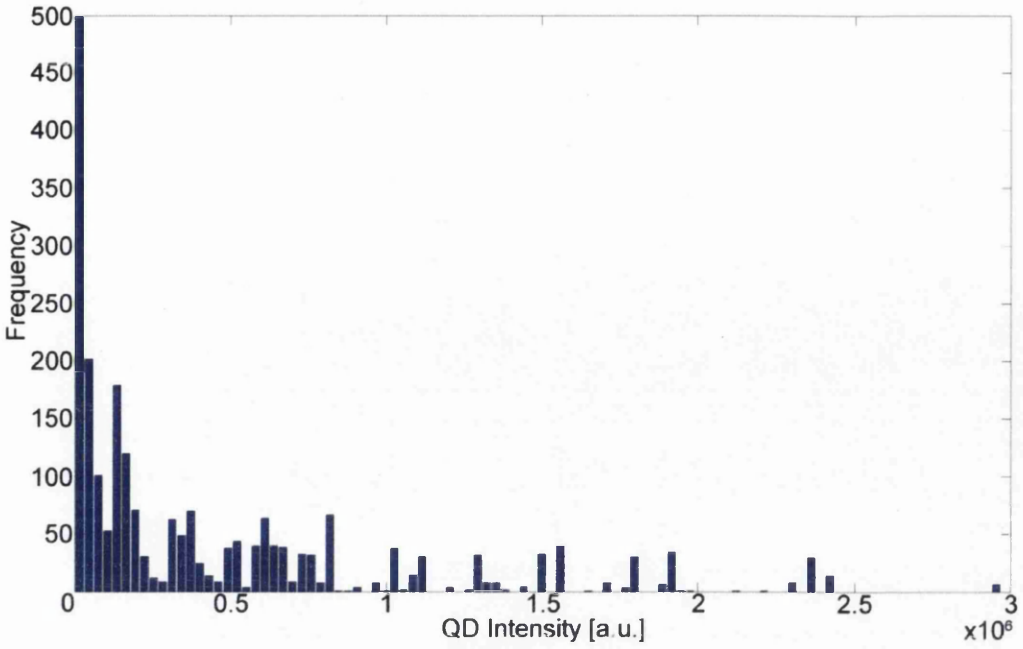



Figure 6.5: Complete set of quantum dot signal intensities within search windows at the time of cell death for all observed DRAQ7 events across all observed fields. Here zero values, those which correspond to a DRAQ7 signal with no localised QD signal present, have been removed.

Figure 6.6 (overleaf) collates all the identified quantum dot signal intensities linked to DRAQ7 events and displays their values against their respective time of occurrence and offers an insight into the proliferation of outlier values at differing time points throughout the timelapse sequence. As was explained earlier, the time interval increases from 1 to 2 hours beyond the $7^{th}$ timepoint in the sequence. At 11 hours there is a noticeable rise in the number of higher intensity values with the distribution of the prospective outlier points being far more continuous than the typically discrete nature of the outliers at other time points, the outliers being well clear of the main point cluster at times 15 and 19 for example. This feature of the $11^{th}$ time point correlates with an apparent spike in the total accumulated QD signal witnessed in figure 6.4 which could indicate a temporary fluctuation in the source of excitation energy within the hardware.

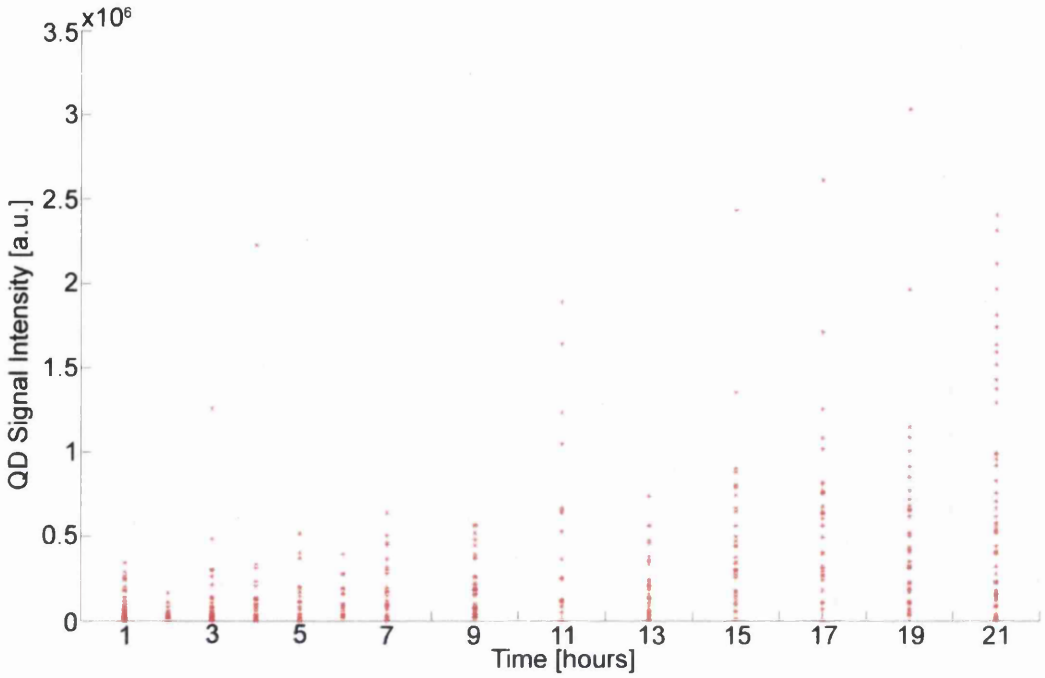



Figure 6.6: Alternative representation showing the quantum dot signal intensity found in the local region of all identified DRAQ7 events across all observed time points.

The majority of the data points in figure 6.6 can be seen to exist in group clusters where the red markers combine to create a red mass rather than individual cross marks. Consequently, much as in the DOX profile end points in the previous chapter, it is more instructive to take averages across all the extracted intensities for each of the 14 time-points in the sequence and subdue the outlier data points. The result of this calculation is plotted in figure 6.7 (opposite) and shows the average intensity of QD signal objects within a 30 pixel radius of the centre of all identified DRAQ7 signal objects for each time point in the timelapse sequence. These averages reveal a steady increase in QD signal with time which would correspond as expected to cells with more resilience to the presence of QD particles lasting longer into the time sequence. A noticeable change in gradient occurs around the $7^{th}$ time point suggesting a possible subpopulation of lower tolerance cells succumbing to the effects of internalized QDs prior to the 7 hour mark. It would be favourable to infer a plateauing effect at the latter stages implying a maximum tolerance to localized QD intensity and therefore QD particle levels however due to the two hour intervals beyond 7 hours there are insufficient data points to make any definitive statements. Time points 7,9,11 and 13 do however form a steeper gradient to a line of best fit

than that formed by the points beyond 13 hours which could suggest the approach to a toxicological limit of the resistance this cell line can exhibit to the presence of quantum dot particles.

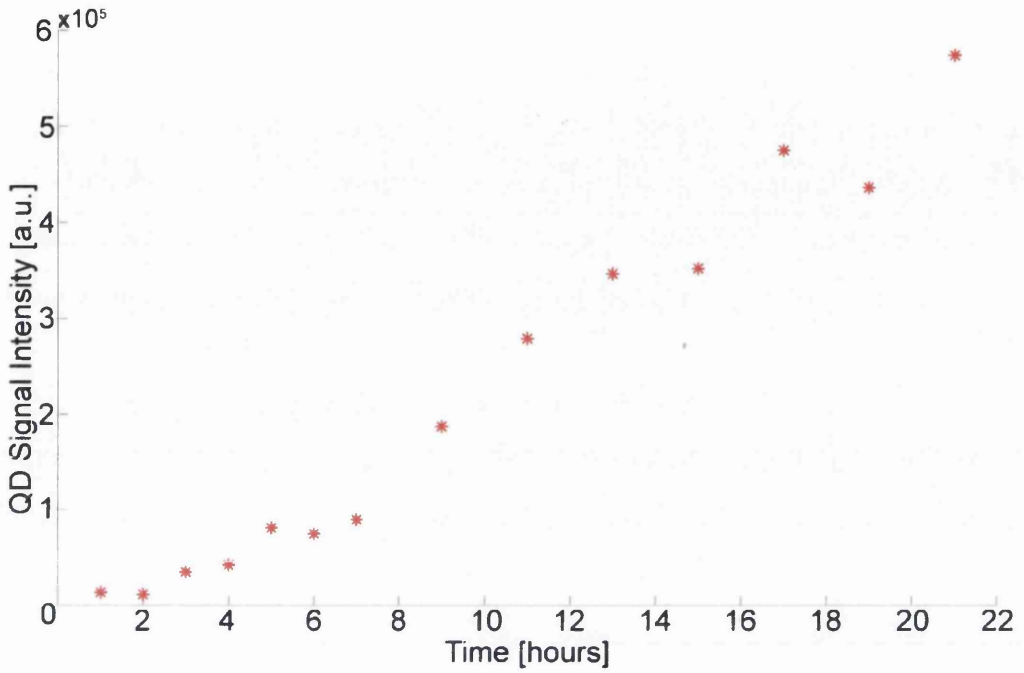



Figure 6.7: Graph showing the per time point average quantum dot signal isolated in the local region within the defined search window for all observed fields.

# CHAPTER SEVEN:

# Cobalt Toxicology Assay

Similar applications of the object identification principles outlined in chapters five and six are applied here to study an experimental assay containing macrophage cells amongst differing concentrations of Cobalt nanoparticles. The experimental goals were to investigate the toxicological effects of Cobalt through statistical analysis of the local region around cells at the point of apoptosis. Such particles can be created via friction during normal wear of internal Cobalt prostheses such as hip replacements. The cells were stained with the fluorescent dye Calcein and loaded into well plates with the fluorescent molecule propidium iodide (PI) which is utilised as a fluorescence source to differentiate between normal and apoptotic cells much as DRAQ7 had been in previous chapters.

## *The Data Set*

The provided data set again contained image sets for 3 channels, brightfield, Calcein fluorescence and PI fluorescence, for 72 acquisition fields across 6 well plates each containing macrophage cells stained with Calcein dye. Plates A1 and A2 contained only these cells and act as controls while A3, B1, B2 and B3 contained Cobalt particle concentrations of $10^8$, $10^{10}$, $10^{11}$ and $10^{12}$ particles per ml respectively. Each timelapse sequence covered a 16 hour range with single hour intervals throughout and the 72 observed fields of view were evenly distributed across the 6 plates with 12 fields randomly initialised in each. The resulting timelapse output contains a total of 3,456 images to be analysed.

## *Image Pre-Processing and PI Identification*

Assay data sets in previous chapters required removal of characteristic noise features to accurately extract pixel intensity values. In this case the first background-foreground isolation process differentiates PI fluorescence signal from the background noise. The panels in figure 7.1 (*page 146*) compare the same image

region in Calcein (A) and PI (B) fluorescence channels and show the significant intensity difference between the PI fluorescence signal and the background noise levels. Worthy of note here are the darker objects seen scattered about the image (A), also present in (B) but harder to see, which are the silhouettes of Cobalt particle agglomerations present in this sample image taken from a field within well plate B2. These features will be discussed further under the heading *"Cobalt Particle Identification and PI Association"* later in the text. A typical PI fluorescence object has an intensity value 5-6 times greater than that of the relevant channel background noise and with no global background noise shape, as was the case with the data set in chapter five, they are easily distinguishable from background via the application of a step function filter. These two channels do however suffer on occasion from cross-channel bleed through which can potentially lead to the misidentification of a Calcein fluorescence object as a PI apoptotic marker and as such the PI objects are identified by applying a step function filter onto an image $I$ defined as

$$I = I_{PI\ channel} - I_{Calcein\ channel}. \qquad (7.1)$$

The resulting image I is shown for the respective region in figure 7.1 (C) where the variation across the background is reduced, any remnant Calcein fluorescence bleeding into the PI channel has been removed and the remaining PI signal objects are still easily differentiated. The applied step function takes the same form as that used in chapter five to create a binary foreground image $F_S$:

$$F_S(x,y) = \begin{cases} 1\ if\ I(x,y) \geq \theta \\ 0\ if\ I(x,y) < \theta \end{cases} \qquad (7.2)$$

where, as before

$$\theta = \mu_{image\ intensity} + (2 \times \sigma_{image\ intensity}). \qquad (7.3)$$

Again the second term in equation 7.3 can be afforded a harsher definition than that used in early chapters due to the significant difference in PI fluorescence object maximal values relative to background levels and the fact they are simply being utilised as markers for apoptotic cells and accurately determining their shape is of no consequence. The isolated PI image foreground is then treated in the same manner as described previously in *"Chapter V: Death Event Identification and DOX Presence Validation"* to extract coordinates for use as apoptotic markers in the next section.

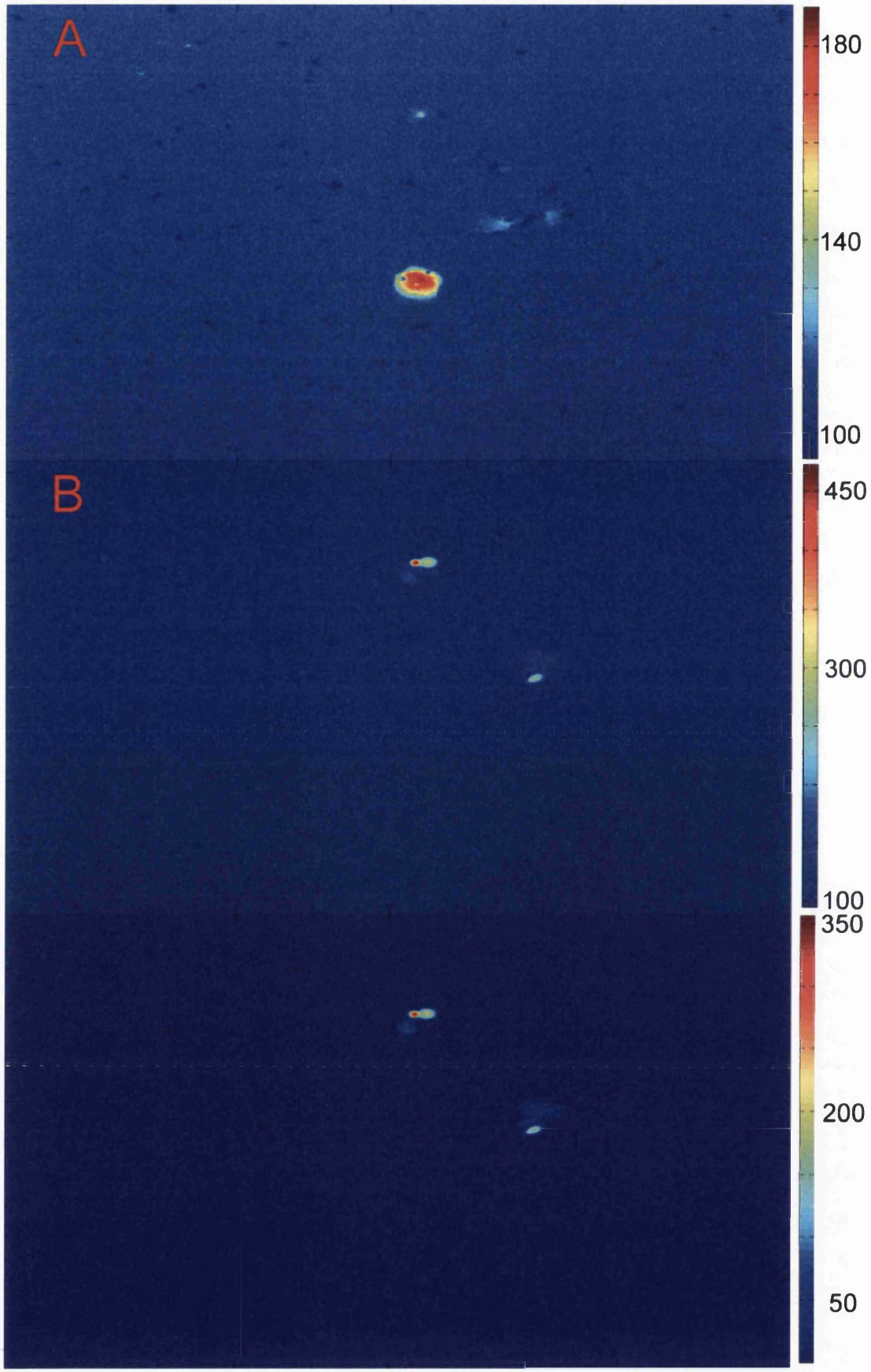



Figure 7.1: Comparison of a region in the Calcein fluorescence channel, (A), the PI fluorescence channel, (B), and the result of (A) subtracted from (B).

## *Cobalt Particle Identification and PI Association*

In each of the 3 imaging channels Cobalt particles reveal themselves through apparent dark spots or silhouettes where the particle agglomeration prevents light from passing through to the imaging lens. These darker regions, whilst faintly perceivable to the human eye in both the Calcein and PI fluorescence channels, differ insufficiently in intensity relative to the background levels to be consistently identified automatically and can be concealed completely by each channel's respective fluorescence signals. Brightfield images however reveal these dark Cobalt objects in relatively high contrast without the possibility of concealment from a fluorescent source and as such it is on this channel that they are identified. Figure 7.2 contains a regional sample of brightfield images corresponding to a field of view within the B2 well plate; the particles seen in the four time panels relate to a Cobalt particle concentration of $10^{11}$ $ml^{-1}$.

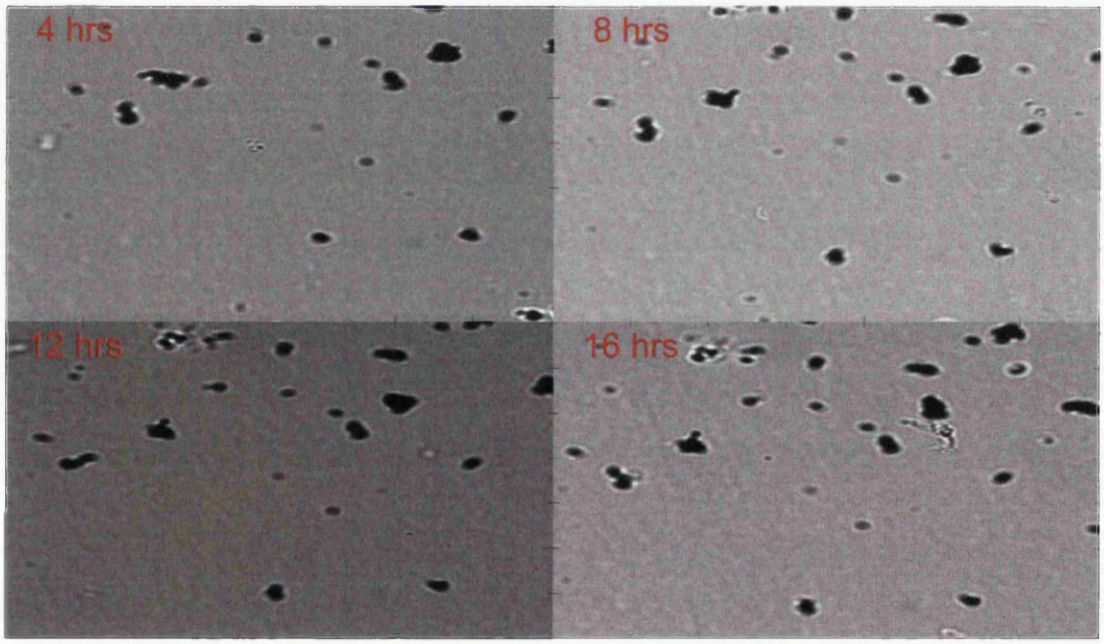



Figure 7.2: Image crop of a single region of the sample brightfield images corresponding to a Cobalt particle concentration of $10^{11}$ $mL^{-1}$ consistently followed across the full timelapse duration.

The eye can distinguish the smaller constituent parts of the dark agglomeration objects and they can be seen to exhibit no characteristic morphology due to the random nature of their formation. The images are cropped on the same region for each time panel and show the relative static nature of these formations away from the perturbations induced by cellular motions.

Figure 7.3 contains a similar display for the B3 well plate, corresponding to a $10^{12}ml^{-1}$ Cobalt particle concentration. They are noticeably larger in size as is expected given the higher number of source particles available for agglomeration into larger structures. This sequence also shows their relative immobility as the images at 4 and 8 hours remain largely unchanged until the presence of a foreign object at 12 hours changes the visible distribution.

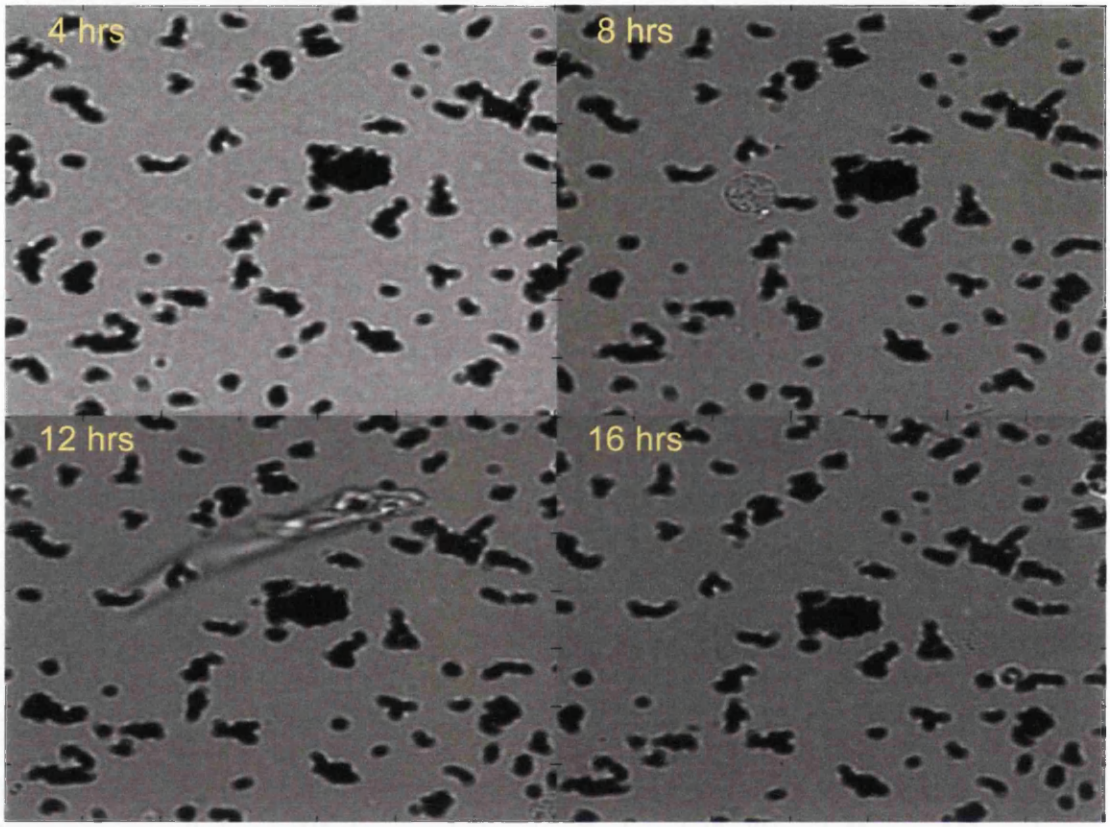



Figure 7.3: Another regional sequence sample of the brightfield channel this time corresponding to a Cobalt particle concentration of $10^{12}$ $mL^{-1}$.

As with previous foreground isolation procedures a step function filter is applied this time filtering all foreground objects and isolating extreme background features due to the presence of Cobalt agglomerations causing an effective absence of light intensity. In the context of a brightfield image source, background intensity can be a somewhat subjective term as an object typically considered as foreground such as a cell contains granular pixel patches in and around the nucleus that often contain dark, low intensity pixel objects. In this chapter background is considered to represent empty space void of cellular and particulate objects. Consequently the objects sought here

are sub background objects and therefore, the resulting filter, applied to the brightfield image $I_{BF}$, takes the form

$$L_S(x,y) = \begin{cases} 1 \; if \; I_{BF}(x,y) \leq \theta \\ 0 \; if \; I_{BF}(x,y) > \theta \end{cases} \qquad (7.4)$$

where $L_S$ represents the resultant low signal objects in the $I_{BF}$ image and the pixel intensity threshold, θ, is defined as

$$\theta = \mu_{image\ intensity} - (2 \times \sigma_{image\ intensity}). \qquad (7.5)$$

The second term of the threshold can again be afforded a harsher value, this time below the average image intensity, as the dark spots of interest are consistently well below the previously defined background level. Particularly small objects associated with noise are then filtered out when initializing the object list, **O**, from the binary image $L_S$ by configuration of the inbuilt Matlab function *bwconncomp()*. The image $L_S$ contains clusters of connected pixels with a 'true' value and the *bwconncomp()* function generates a pixel object list for each of those pixel clusters. Parameterizing this function to exclude groups with a connectivity of 16, which was observed to be much lower than the typical Cobalt object of over 50, is sufficient to omit their inclusion as a valid object. The result object list, **O**, takes the form

$$\boldsymbol{O} = \{p_1, p_2, \ldots, p_n\} \qquad (7.4)$$

where $p_i$ contains a list of pixels for the $i^{th}$ identified pixel grouping, $1 \leq i \leq n$ and n is the total number of isolated pixel groups in the image $L_S$. These pixel lists are plotted alongside their parent brightfield image in figure 7.4 where (A) shows a full field of view brightfield image with a $10^{11}$ $ml^{-1}$Cobalt concentration and (B) plots all the pixels contained in the resultant set **O**.

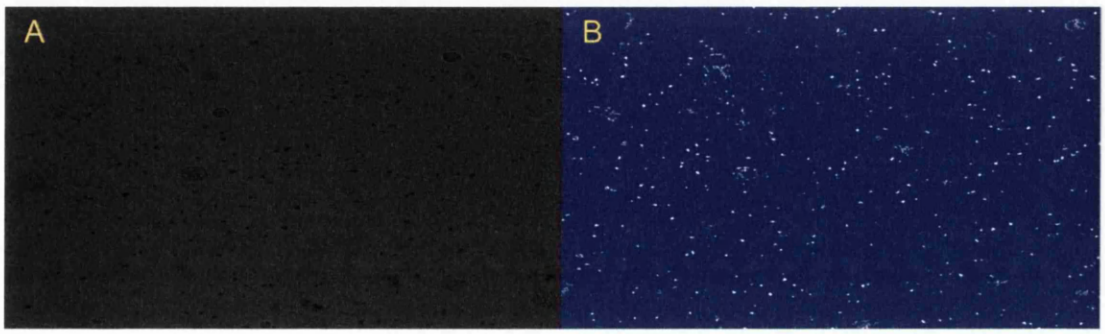



Figure 7.4: Sample brightfield image, (A), alongside a binary representation of the identified Cobalt agglomerations, (B).

Large objects can be seen in some regions where in their respective position in the brightfield image cells can be seen to exist. The problem with isolating Cobalt particles from brightfield images is that cells viewed through brightfield microscopy often contain granular regions of pixels in and around the nucleus with low level intensities resulting from the cell also inhibiting the passage of light. Figure 7.5 contains image samples showing the prospective difficulties in distinguishing Cobalt particle objects from these dark spots typically present within cell nuclei.

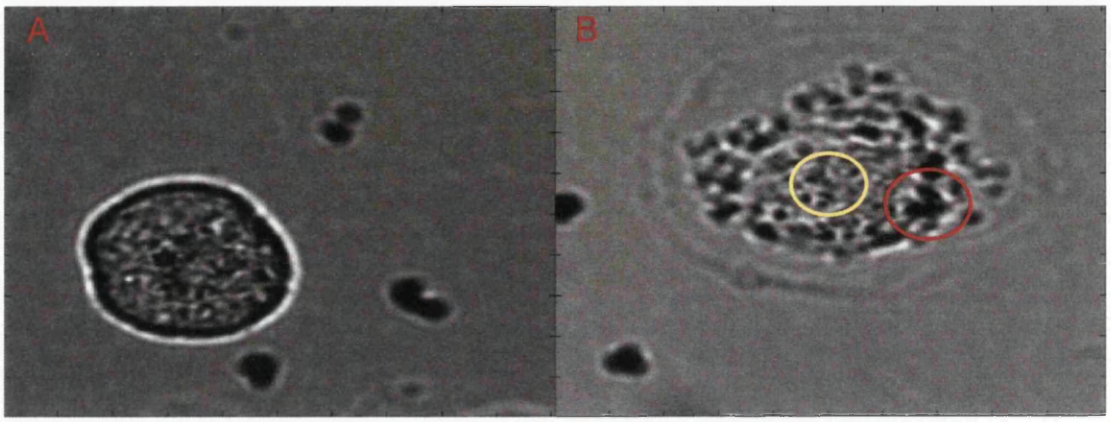



Figure 7.5: Brightfield images of two cells exhibiting features than can lead to misidentification of dark regions as Cobalt particle agglomerations.

The gray-scale image in (A) shows three distinct dark objects around the cell perimeter with a dark ring of pixels of similar intensity encircling the cell itself. Whilst such a well defined edge is ideal for segmentation here a circular object would pass through the step filter and be misidentified as a large Cobalt object. Conceivably such an object could be distinguished from actual Cobalt agglomerations by virtue of the presence of non object pixels within it. More precisely the cellular ring object can be described as hollow, a characteristic unlikely of a Cobalt agglomeration due to their cluster formation. Figure 7.5 (B) contains another cell image in the brightfield channel detailing the granular effect of the nucleus with two regions highlighted. Within the red circle is a large, low intensity pixel object that could well be Cobalt particle but such objects have been witnessed on cells in previous chapters with no affiliation to this Cobalt experiment. The yellow region would be described as having no large objects but does contain much smaller groupings of low level pixels. Such objects are also commonly linked together via single pixel bridges to form much larger objects with patches of higher intensity pixels within them; such objects can be described as being visually

"sponge-like" or porous. Both eventualities lead to a compromising solution of filtering the sub-background objects based on size and, through observation and some trial and error, a pixel limit of 200 per object is applied to successfully isolate the majority of the observed Cobalt particle agglomerations whilst dropping larger, more cell-like objects. A sample of the resultant sub-background objects for each of the four particle concentrations are shown in the panels of figures 7.6A and 7.6B (overleaf), displayed in pairs on separate pages for clarity. Comparing the four panels it is plain to see the effect of higher particle concentrations resulting in larger numbers of Cobalt agglomeration objects, what is less clear from these displays is the propensity for larger formations at higher concentrations, as would be expected. At low concentrations the objects are highly clustered around cell locations. At the lowest concentration this is more a result of misidentification of internal cell features as Cobalt particles rather than directed object orientation as at these concentrations there are simply too few agglomerations of resolvable size to form a well distinguished sub-background level on which to threshold. As the concentrations increase this problem lessens although the particle formations are drawn towards cells as is indicated by the inconsistent regional variation across the image for the $10^{12}$ Cobalt particles per ml well. It is objects such as these that are subject of analysis in the next section.

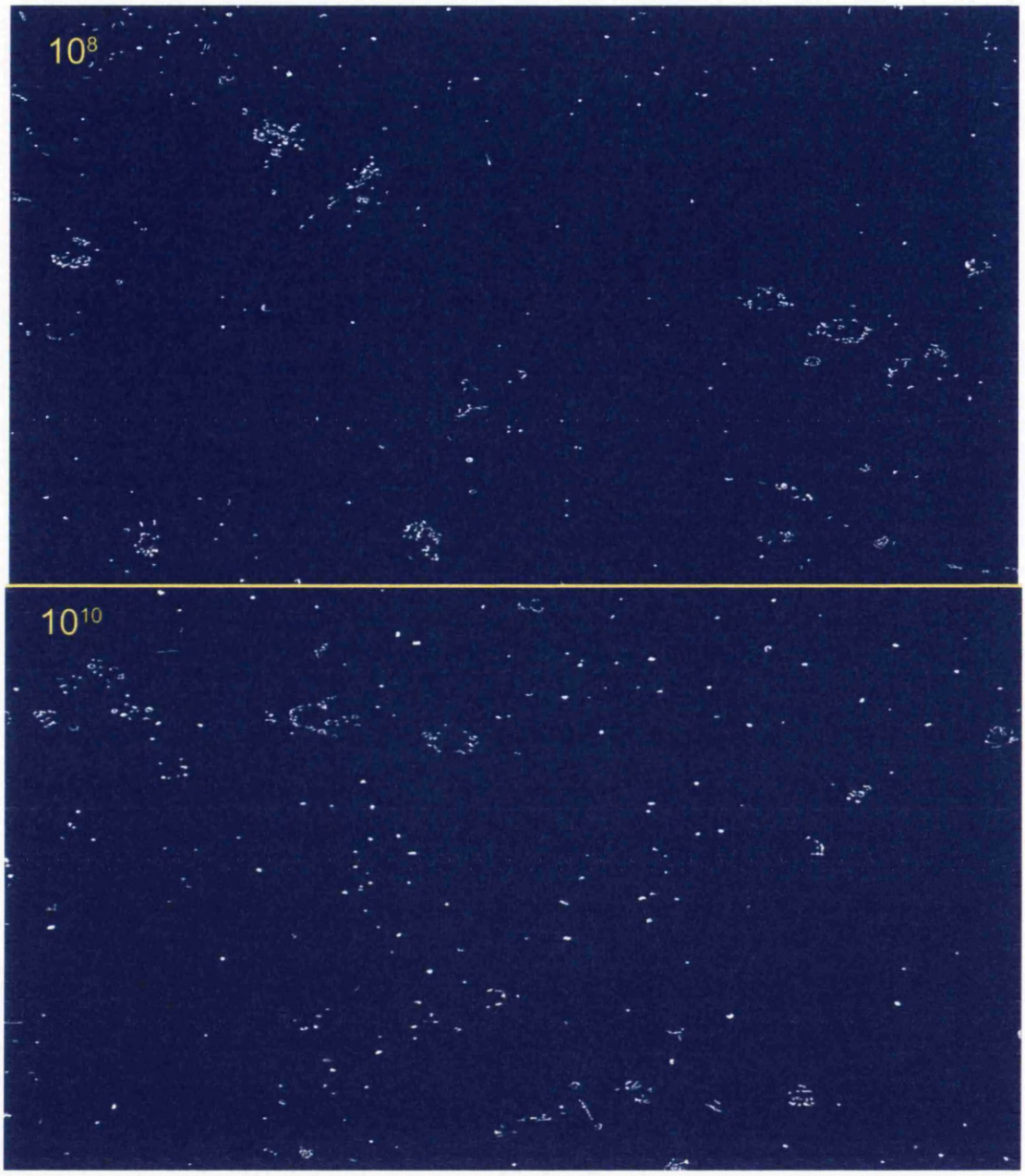



Figure 7.6 A: Full field comparison of the identified Cobalt objects across all four concentration levels. The higher concentrations lead to the increasing number of larger agglomeration objects seen across the images. Concentrations $10^{11}$ and $10^{12}$ $ml^{-1}$ are provided in figure 7.6 B.

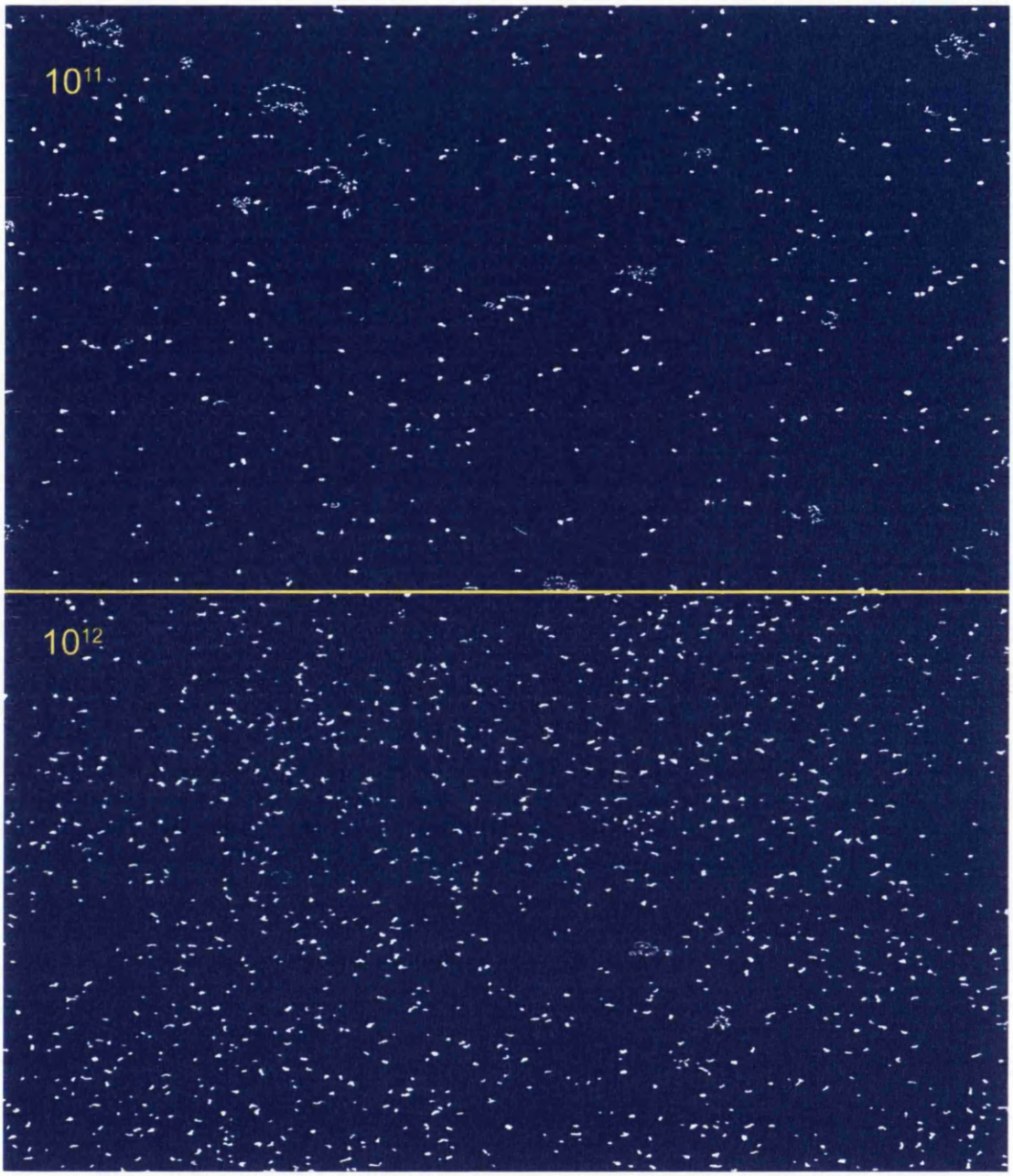



Figure 7.6 B: Second concentration pairing. (Displayed in this manor for clarity)

## *Analysis Outputs: Cobalt Particle Overview*

With the Cobalt particle objects isolated and stored for each time frame across all available wells in the data set statistical representations are utilised in this section to visualise and characterise there distributional properties. It is informative to conduct a particle audit to look at the object spatial distributions across the differing Cobalt particle concentrations of well plates A3 through B3 at the final time point of 16 hours; A1 and A2 having no particles in them are of little interest to this analysis. The left hand side of the graphic in figure 7.7 (opposite) details the differing numbers of Cobalt agglomeration objects identified in each of the four plates by way of a colour coded proximity map. Proximity maps offer a clear visual perspective on the number of particles across a whole field of view and here dark blue regions represent close proximity to an identified Cobalt object centroid and red regions indicate regions of space void of said particles. This visualisation serves to show how the number of Cobalt particle objects increases with higher concentrations of Cobalt particles, as would be expected. The right hand side of this graphic shows another representation of these points detailing the Cobalt ion density where each Cobalt agglomeration centroid is overlaid with a radial function that maps the object area, $A$, as a point mass based on an object volume estimate, $V$, where

$$V = A^{3/2} \qquad (7.5)$$

and calculates a $1/r^2$ potential; the resulting plot can be thought of as a 'region of influence' potential map with a colour map normalized to its individual potential range. These potential maps show the effect of the increasing agglomeration sizes alongside the increasing numbers seen in their counterpart displays on the left. In order to compare the potentials across all four well plates, figure 7.8 (A) and (B) show the same density fields displayed in figure 7.7 with their colour maps normalized across the potential range of all four fields of view (they are displayed on the next opposing pages to allow easy comparison).

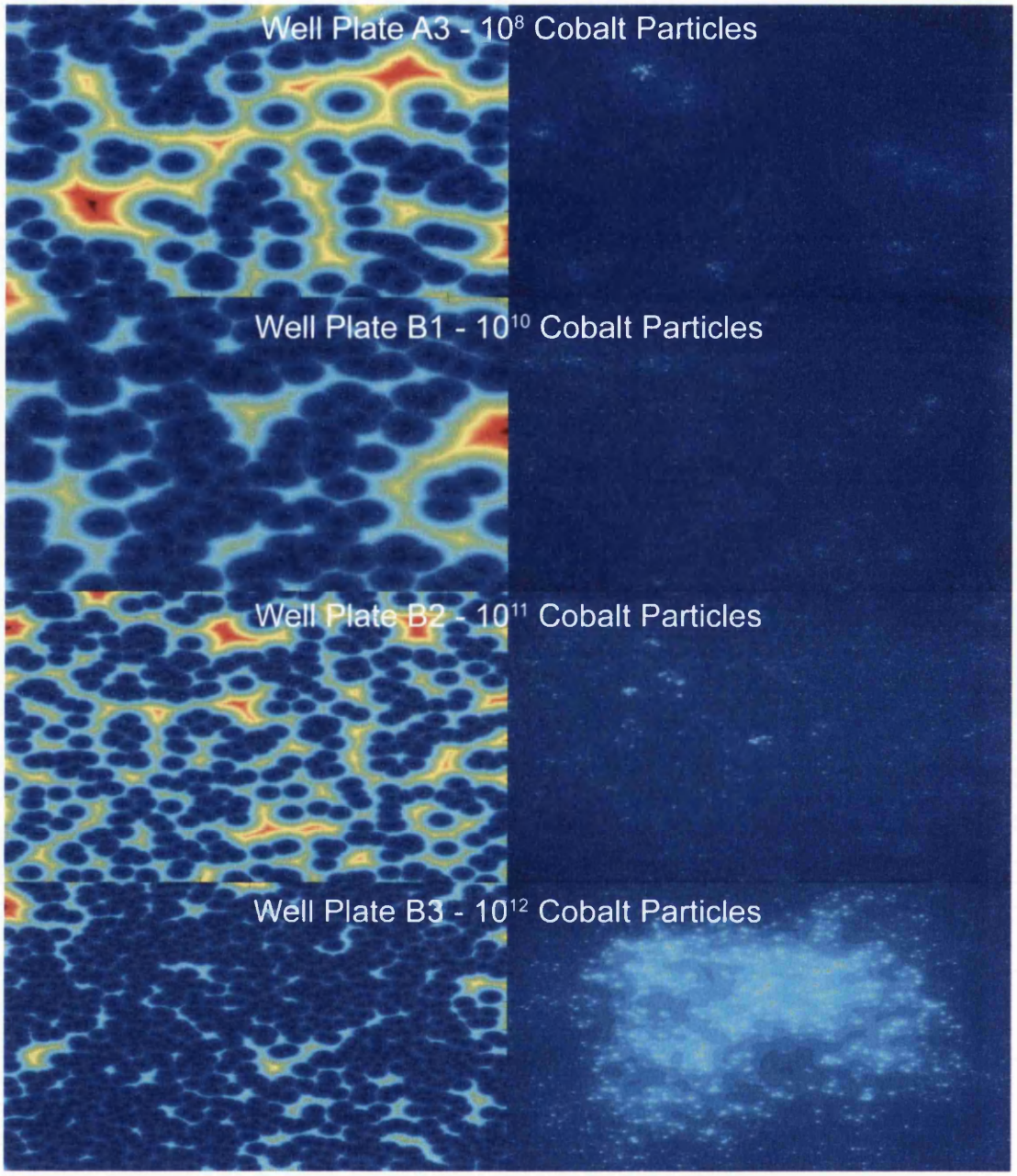



Figure 7.7: Comparison of Cobalt agglomerations across the differing concentrations. The left hand column contains the proximity maps used to create the density graphics shown on the right.

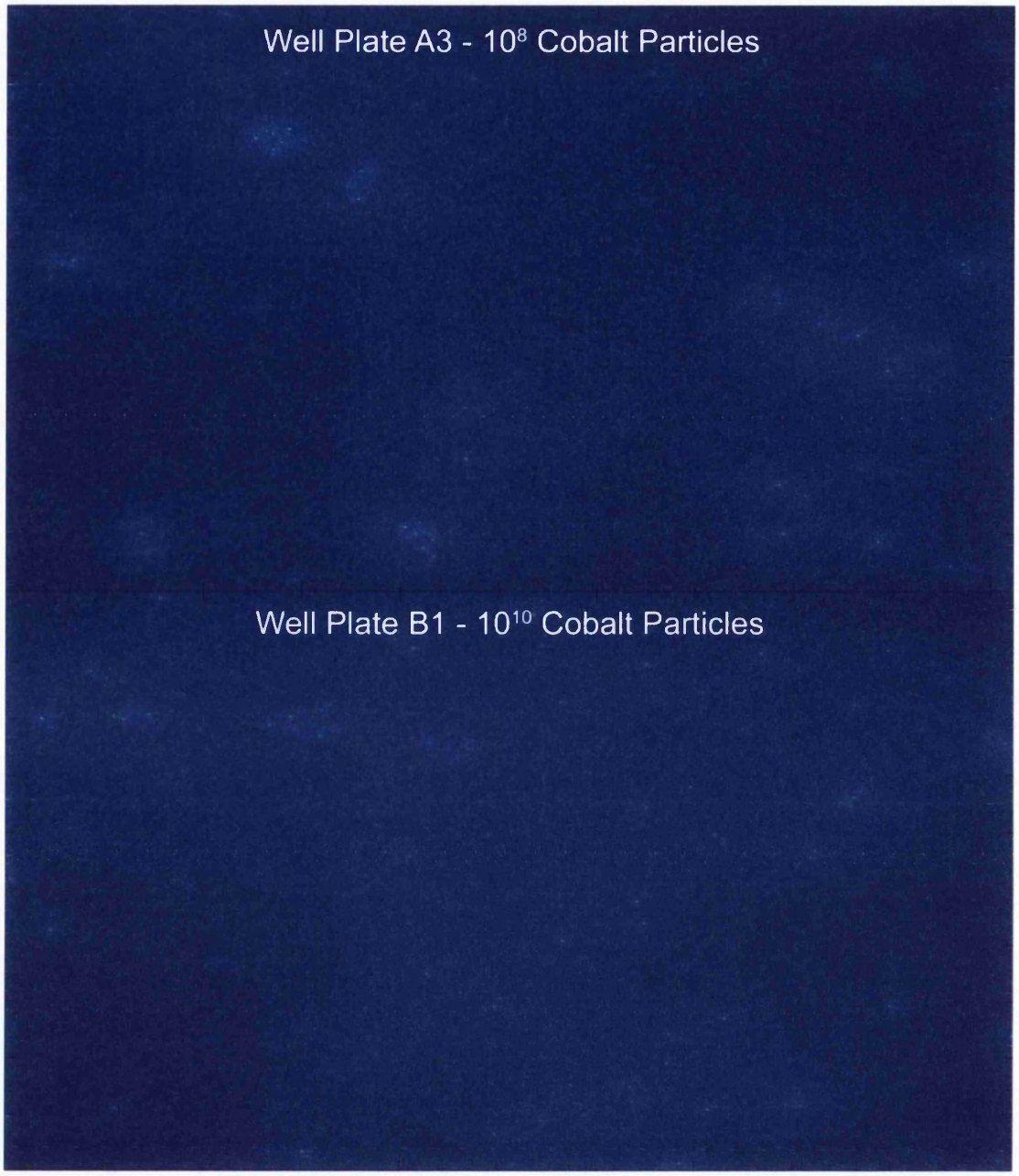



Figure 7.8 A: Density map graphics scaled, unlike previous graphic, to a single colour-map across all for sample fields. Shown above are field samples taken from well plates A3 and B1 with respective $10^8$ and $10^{10}$ Cobalt particles per ml concentrations.

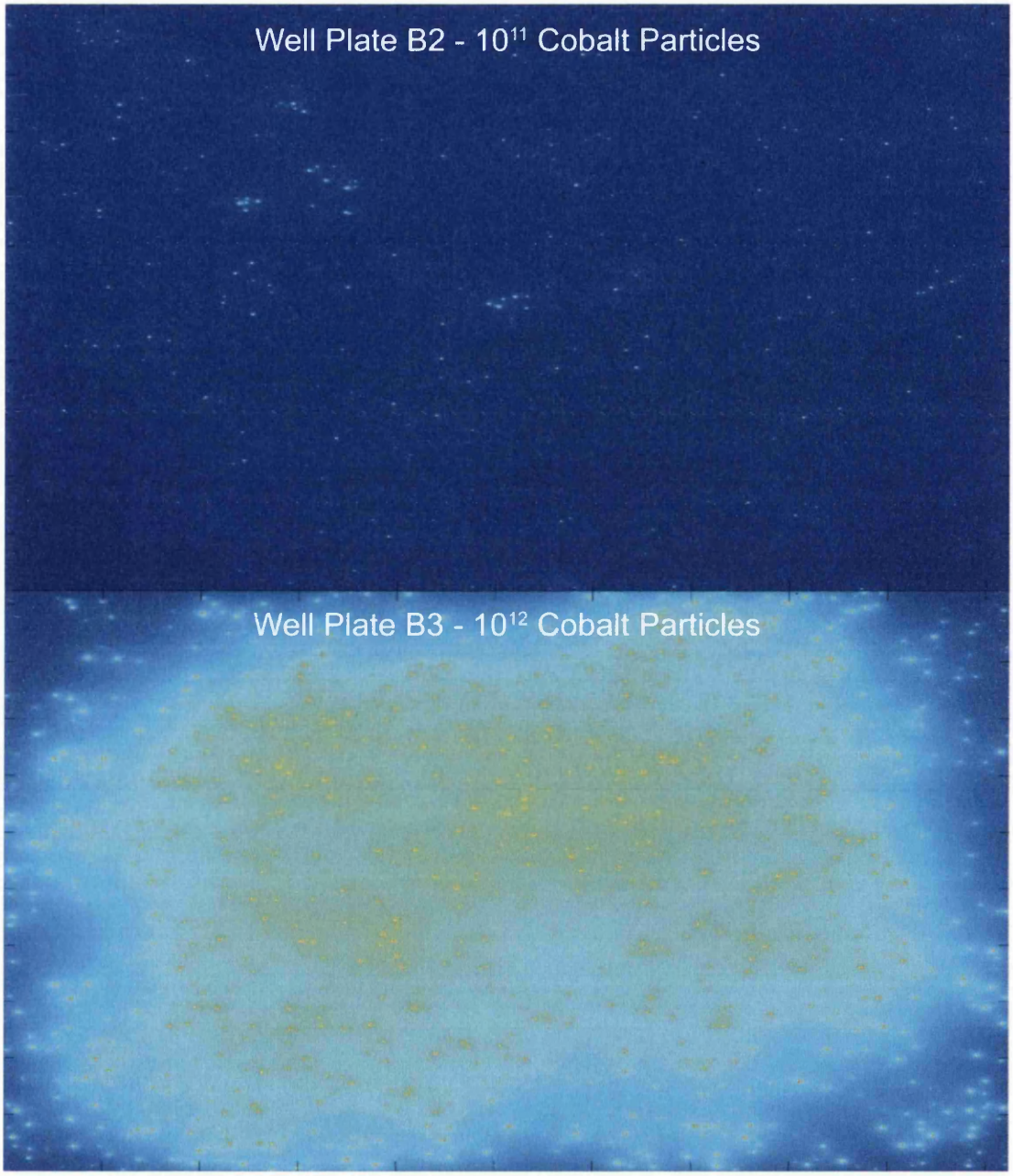



Figure 7.8 B: Second pair of field samples taken from well plates B2, $10^{11}$ Cobalt particles per ml, and B3, $10^{12}$ Cobalt particles per ml. With the same colour-map scaled across all four images the effect of particle agglomeration becomes clear.

This normalization contrasts the difference in agglomeration size and object density as the Cobalt particle concentration increases with the accumulated effect in well B3 being starkly greater than that seen at the other three concentrations. The apparent clustering and resultant potential trough in the image centre, most notable for well plate B3, is merely a boundary effect where particle agglomerations beyond the edge of the field of view cannot be audited and so the accumulated potentials in areas near the edge are underestimated.

Figure 7.9 (A) provides an overview of the four wells of interest across all 12 available fields of observation detailing the average number and average area of all identified Cobalt objects.

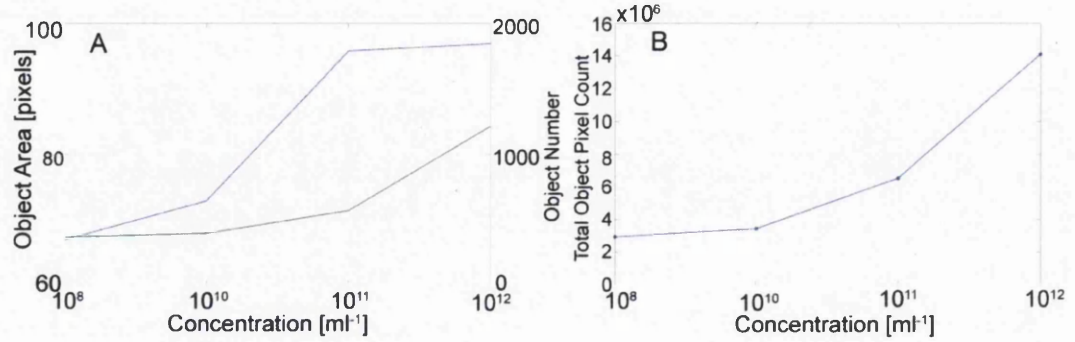


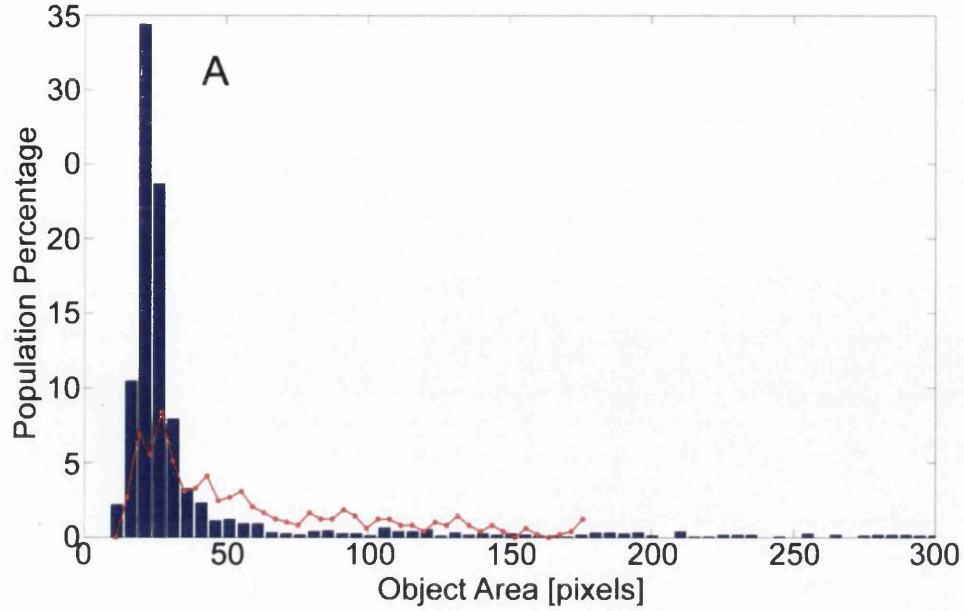

Figure 7.9: (A) Comparing average Cobalt object area and the average number of objects across the A3, B1, B2 and B3 well plates. (B) Total Cobalt object pixel count across all 12 fields in each of the four well plates.

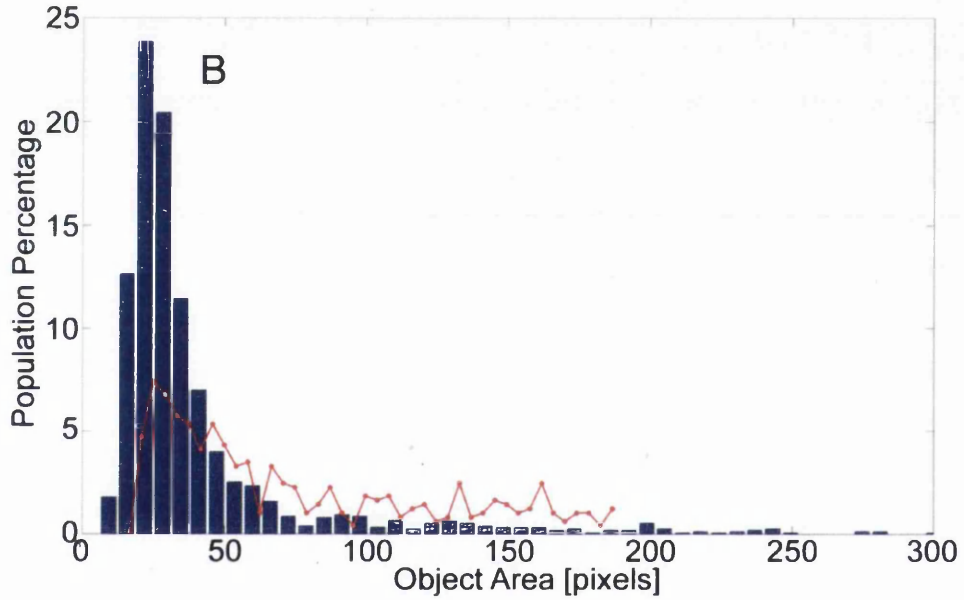
The concentration difference between A3 ($10^8 ml^{-1}$) and B1 ($10^{10} ml^{-1}$) appears to impact more on the size of the Cobalt formations rather than the total number of objects and whilst the numbers show a marked increase between wells B1 and B2 ($10^{11} ml^{-1}$), again the object area increase would seem to account for the additional Cobalt present. Beyond this however, the object area plateaus with the concentration jump between B2 and B3 ($10^{12} ml^{-1}$) appearing to have little additional effect on the agglomerate sizes but greatly increasing the number of overall agglomerations. Panel (B) plots the total count of pixels attributed to Cobalt particle objects similarly across all 12 fields for each of the four well plates. Each sequential well plate increases the Cobalt concentration by a factor of 10; the graph in (B) shows steadily rising pixel counts of approximately 2.9, 3.4, 6.5 and 14 x$10^5$ across the four wells. With a 10 fold increase in Cobalt concentration between wells it could be expected that a count of Cobalt object pixels would reflect a similar multiple increase, however the images only observe a 2-D plane through a complex 3-D Cobalt object. The 'silhouettes' are only a cross section of a Cobalt agglomeration that can feasibly, due to the random nature of its formation, occupy any configuration and within the media can also be positioned in an infinite list of orientations. The previously defined object volume estimate (equation 7.5) assumes some variation of spherical symmetry when in fact no such constraint bounds the Cobalt formations. Providing better volume estimates based on cross sectional area requires structural statistics yielded from further study

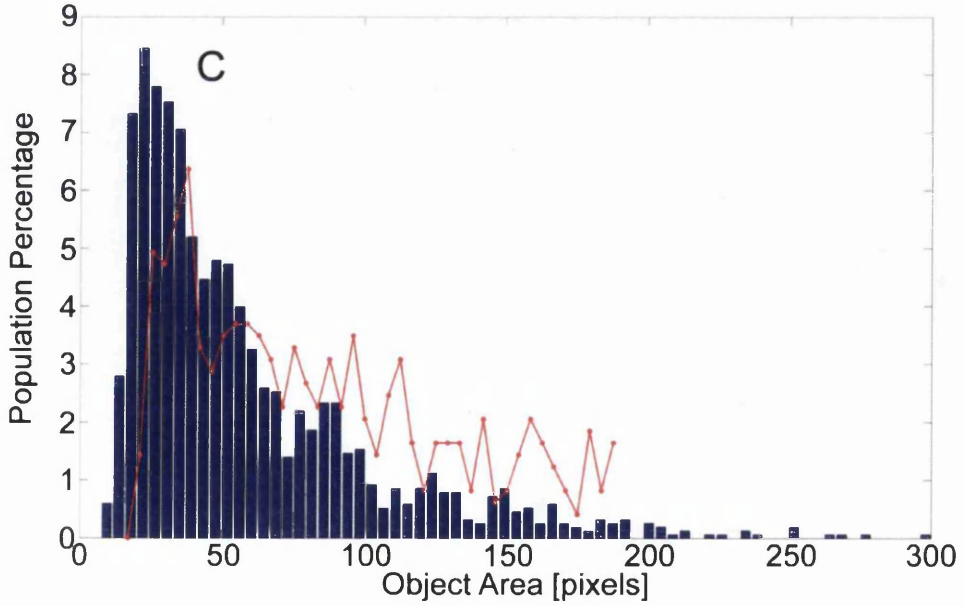
and modelling of the formation of Cobalt particle agglomerations. Such limitations do not inhibit the investigation of the effects of Cobalt in proximity to the macrophage cells in this assay as utilising the agglomeration cross sections is suitable in inferring agglomeration effects. Whilst it is possible for a Cobalt particle formation to be, for example, tubular and orientated such that it displays a cross sectional area deceptively small for its total volume, randomly forming clusters are not likely to form such 'engineered' structures and are more likely to take on globular shapes whose volume and cross section are non-orientation specific and representative of the Cobalt object mass.

These area extractions are compared to the results of an ImageStream data analysis in figure 7.10 (opposite) for the three concentrations looked at in previous figures. The ImageStream flow cytometry system is able to image each Cobalt particle individually to yield large numbers of accurately acquired Cobalt agglomeration cross sections and is used here to verify the statistical observations obtained from the data set analysed in this chapter obtained via fluorescence and brightfield microscopy. Each panel in figure 7.10 shows the ImageStream results with blue bars and the results of the analysis detailed in this text are overlaid in red. Each panel is randomly sampled to compare the same number of Cobalt particle objects at each concentration to enable a comparison of the object size distributions across the three well plates. All three plots in red are visually verified to have comparable distributions to those seen from the ImageStream data set with the apparent underestimate of the peak region. This is predominately due to the fact that the ImageStream system is readily capable of visualising much larger numbers of objects, often to order of $10^4$, and here that is leading to increased definition in the peak regions. ImageStream is however incapable of associating the Cobalt agglomerations with specific cells as is the objective of this chapter and is therefore only used as a verification step to validate the statistical yield of the image analysis procedures outlined in this text. Despite the relative lack of data, 12 image fields per well containing identified particle agglomerations of numbers in the order of $10^2$ compared to ImageStream particle count of $10^4$, the distributions can be seen to match the general shape changes of the ImageStream histograms across the differing concentration levels and validates the image analysis procedures used to identify Cobalt particle agglomerations. Both methods exhibit evidence of larger agglomerations at higher concentrations and a lowering of the number of smaller formations seen in the lower counts and widening of the overall spread.

Figure 7.10 (opposite): Comparison of the Cobalt object areas observed using InCell analyser (Red) and ImageStream (blue) systems. (A) $10^{10}ml^{-1}$, (B) $10^{11}ml^{-1}$ and (C) $10^{12}ml^{-1}$.


A
Population Percentage
Object Area [pixels]
0
5
10
15
20
30
35
0
50
100
150
200
250
300


B
Population Percentage
Object Area [pixels]
0
5
10
15
20
25
0
50
100
150
200
250
300


C
Population Percentage
Object Area [pixels]
0
1
2
3
4
5
6
7
8
9
0
50
100
150
200
250
300

## *Analysis Outputs: PI-Cobalt Association*

Isolated Cobalt agglomeration objects can now be associated with local PI death signals in much the same way as the previous two chapters linked their respective toxicological assay variables to cell death markers. As in chapter six, application of a radial search window centred about the centroid of an identified PI signal object associates any surrounding particulate objects to that particular cell death event. Again, the base radius of this search window is defined as

$$R_B = \frac{1}{2} \times \mu_{cell\ diameter} \qquad (7.6)$$

where $\mu_{cell\ diameter}$ is the average observed cell diameter which for macrophage cells, as was the case for the fibroblast cells in the previous chapter, measures the average width of the cell around the nucleal centre. This omits the far reaching membranous structures that can take a variety of shapes and dimensions as a cell migrates around the field of view and provides a stable base value on which to begin. Unlike previous chapters, the Cobalt ions can theoretically induce detrimental effects to a cell without being in direct physical contact, consequently in this analysis the radial window is also dilated in discrete steps to investigate the possible toxic effects at a distance. For each identified PI signal event the Cobalt objects within proximity of the base radial distance and 5 further dilations taken at 10% intervals are stored into the object pixel lists of the form

$$\boldsymbol{O_N^R} = \{\boldsymbol{l}_1, \boldsymbol{l}_2, \ldots, \boldsymbol{l}_n\} \qquad (7.7)$$

similar to those created previously to store the Cobalt objects where here, $\boldsymbol{l}_i$ represents the pixel list of Cobalt particle objects found to be within a radius **R** of the PI signal centroid, n is the total number of Cobalt objects found within said radius and **N** is the total number of PI events previously identified. These lists are compiled and labelled accordingly for each of the 16 timeframes at each of the 48 fields of view across well plates A3-B3 (as mentioned before, A1 and A2 contain no Cobalt particles so are of no interest here). With all these lists compiled it is then possible to extract data counts for graphical display.

Figure 7.11 details a compilation of the associated Cobalt object sizes found within the base radius and five increasing dilations for each well.

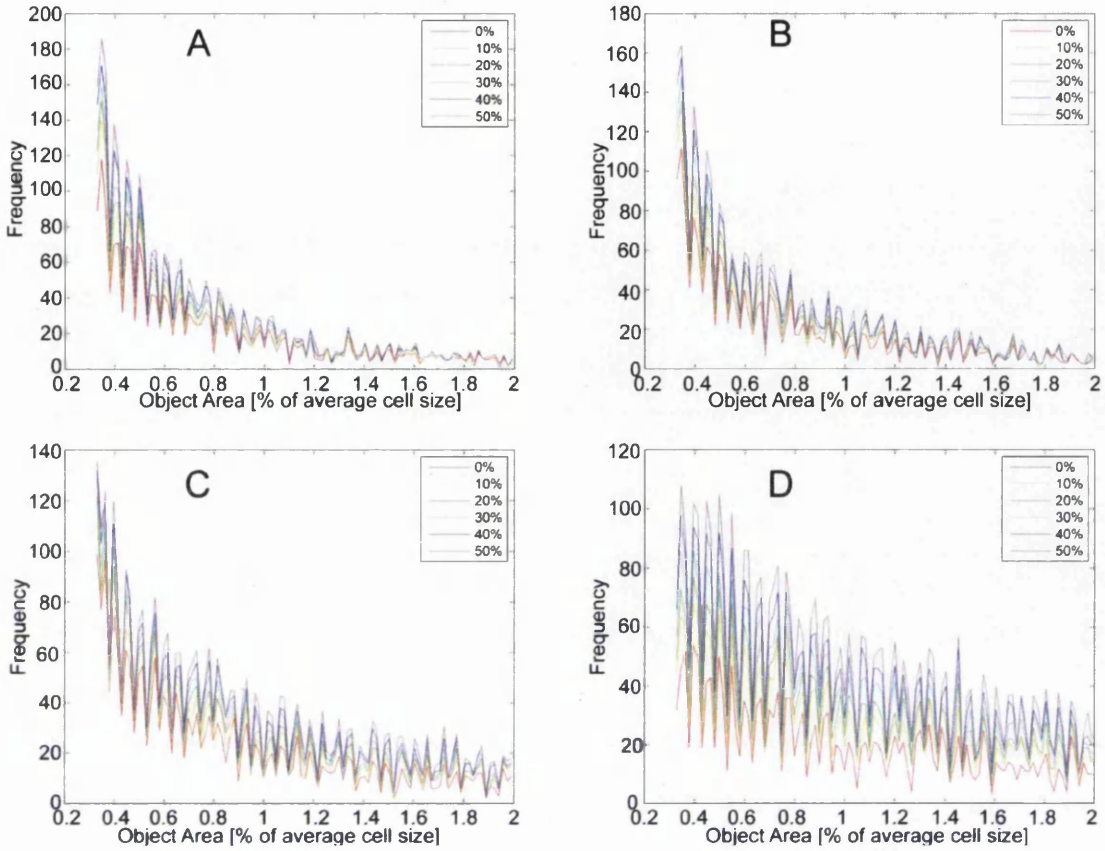



Figure 7.11: Comparison of the identified object sizes as a percentage of an average cell size across the three Cobalt loaded well plates. Each plot shows the resulting identified objects over radial dilations of 0-50% for $10^8 ml^{-1}$ (A), $10^{10} ml^{-1}$ (B), $10^{11} ml^{-1}$ (C) and $10^{12} ml^{-1}$ (D) concentrations.

These object sizes are obtained by counting the length of each object pixel list, $\boldsymbol{l}_i$, and are displayed in terms of percentage of an average cell area which is calculated using the base radius defined in equation 7.6 and the standard circular area equation

$$Area = \pi r^2. \qquad (7.8)$$

Whilst this area is a simplistic estimate of the overall macrophage cell area which can at times greatly exceed such an estimate, it is observed that the bulk region of the cell containing the nucleus lies within this area. It is also noted that when a macrophage cell extends in a particular dimension typically it also contracts in the dimension orthogonal to that extension which acts to conserve the total area. Consequently, coupled with the fact that consistent use of such an x-axis scaling does not affect any extracted parameters relative to one another, it is noted as not idealised

but deemed sufficient for the purposes outlined here. The four panels in figure 7.11 display an overview of the localised Cobalt agglomeration sizes for wells A3-B3 with Cobalt particle concentrations $10^8$, $10^{10}$, $10^{11}$ and $10^{12}$ $ml^{-1}$ respectively. Each plotted line represents a histogram plot displayed in this manner to allow for visual comparison across radial dilations with the percentages in each plot corresponding to the percentage dilation of the peri-cellular search region and 0% relating to the observed average cell diameter of 110 pixels. Increasing the dilation of the localized region of interest from 0%-50% appears to cause no change in the overall trend across all 4 wells suggesting the Cobalt particles are generally isotropic in their distribution across the field of view. Progressing through the concentrations, A-D, shows a marked shift towards objects with a larger area which is a consequence of the greater number of available particles at higher concentrations resulting in larger agglomerate formations and therefore a larger silhouette cross section. This characteristic was implied in figure 7.9 and is also present in a second data set (not analysed in this text) taken in parallel to this one on an ImageStream system. Such an effect can also be down to larger concentrations forcing particles to form heteroscale agglomerations, or agglomerations joining together to form larger structures, again an effect witness in partnering ImageStream data that would not be explicit in the data analysed here.

Total object area changes seen in figure 7.9 (B) and the relative similarity between the plots in figure 7.10 (A) and (B) suggest the $10^8 ml^{-1}$ Cobalt particle concentration is below a threshold beyond which large numbers of agglomerations begin formation and toxicological effects become visible. Consequently, all successive graphs exclude well plate A3 and focus on plates B1-B3. Figure 7.12 (opposite) compares the Cobalt object histograms for the concentrations $10^{10}$ $ml^{-1}$ (blue), $10^{11} ml^{-1}$ (red) and $10^{12}$ $ml^{-1}$ (green) at the first and last time points for the peri-cellular regions within a 10% (A-B) and 20% dilation of an average cell diameter. The plots shown here depict randomly sampled subsets of the total number of cell death events per concentration and count the number of entries in their respective $\boldsymbol{O_N^R}$ list to allow comparison of the same number of death events from each well. At both dilations the number of Cobalt particle objects can be seen to increase over time whilst maintaining similar distributions. This peri-cellular localization inferred by the apparent similarity between the two dilations corresponds to the cell itself being very

active in particle uptake. The total object counts detailed in the plot legends show the previously seen increase in objects at higher concentrations and the relative indifference of the counts across the two dilations that infer a localizing of objects on and around the cell.

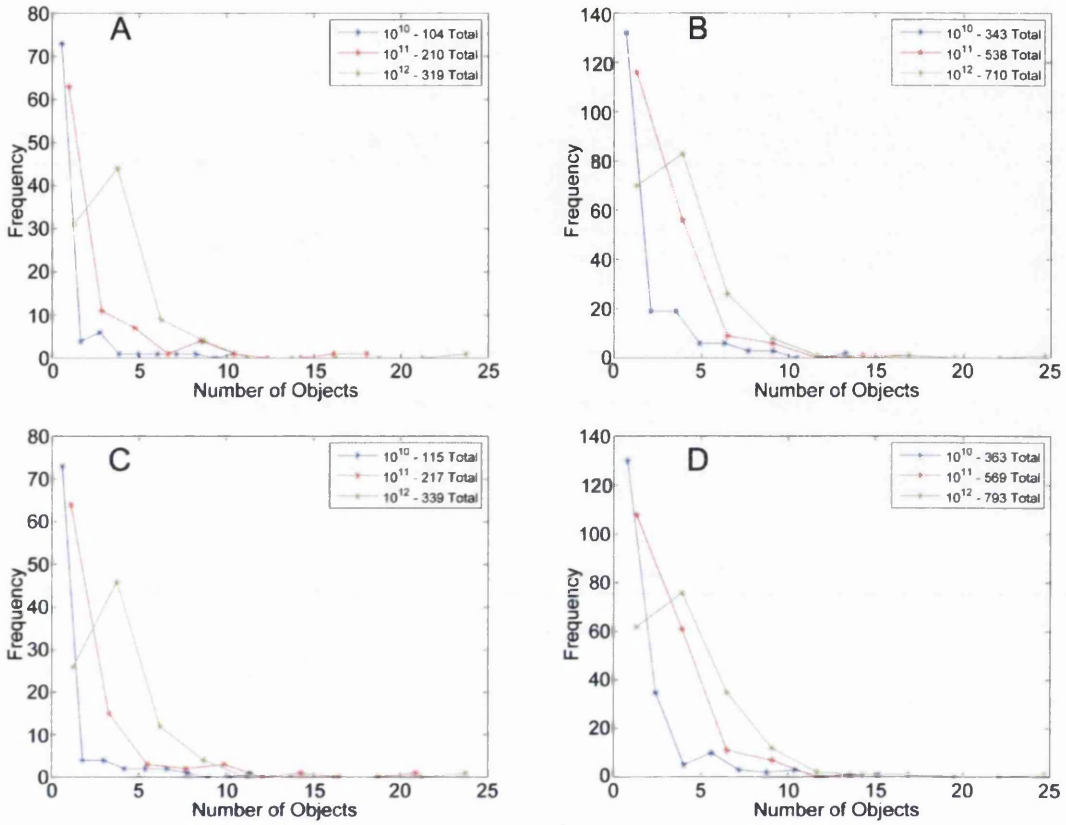



Figure 7.12: Cobalt object counts comparing the Cobalt concentrations of $10^{10}ml^{-1}$, $10^{11}ml^{-1}$ and $10^{12}ml^{-1}$ in blue, red and green respectively. (A) and (B) counts are taken with a 10% radial dilation at a time of 1 hour and 16 hours respectively. (C) and (D) counts are similarly plotted for a radial dilation of 20%.

Similarly, figure 7.13 (overleaf) compares the histograms of the area of cell associated objects for each of the aforementioned Cobalt concentrations and dilations and also randomly samples the total $\boldsymbol{O_N^R}$ set to compare the same number of cell death events per well. The area is again defined as $|\boldsymbol{li}|$ (see *equation 7.7*) for each of the entries within each of the sample subsets of $\boldsymbol{O_N^R}$ variables. Fluctuations caused by this random sampling procedure are more prominent in the first hour plots due to lower numbers of cell death events and results in the inconsistency of the plotted line hierarchy in plots (A) and (C) making the latter time points more insightful.

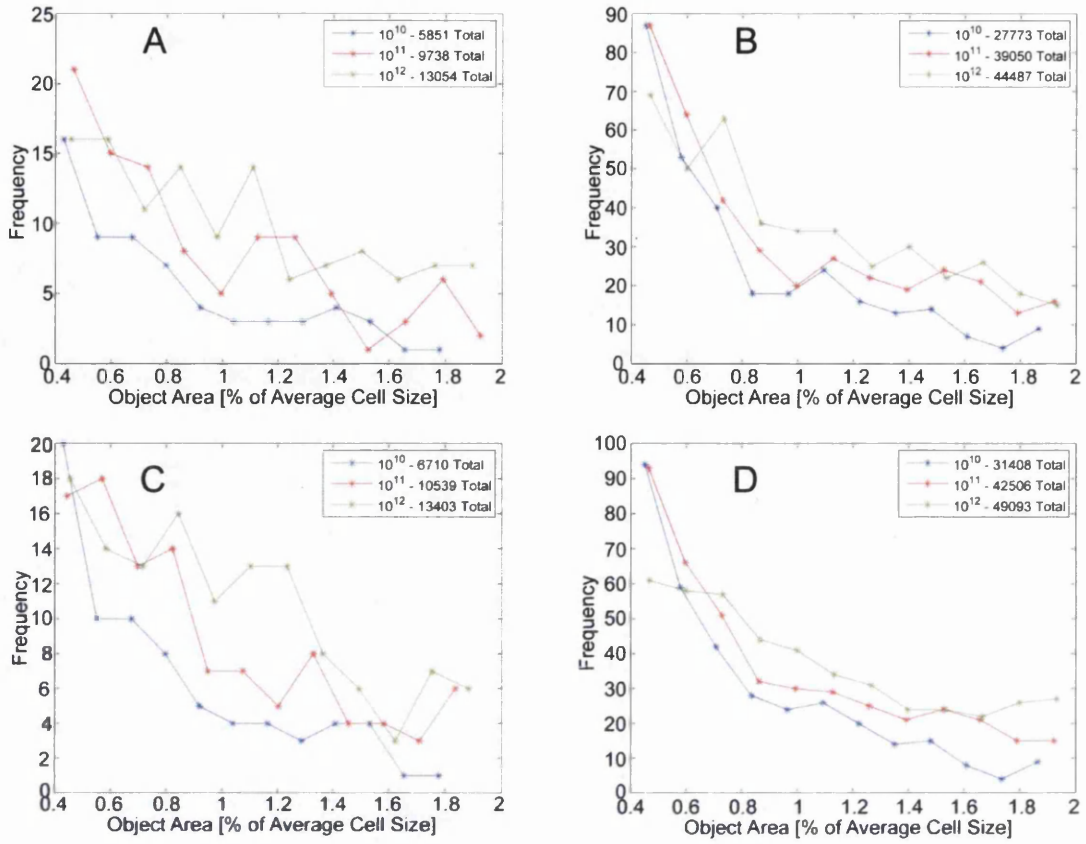



Figure 7.13: Cobalt object areas compared for the Cobalt concentrations $10^{10}ml^{-1}$, $10^{11}ml^{-1}$ and $10^{12}ml^{-1}$ in blue, green and red respectively. (A) and (B) again show the 1st and 16th hours respectively for a 10% dilated search radius and (C) and (D) the same for a 20% dilated search radius.

The $10^{12}$ particles per well concentration induces larger particle formations, as seen previously, and is seen here by the corresponding green line existing predominantly above the lower concentrations lines. The lower concentrations (blue and red) however would be expected to have higher numbers of smaller agglomerate objects which is clear in (B) and (D) where the lower concentration lines peak well above the $10^{12}$ (green) line. The apparent 'stacking' of the three curves is a remnant of agglomerations being more readily formed at higher concentrations; the total object area per concentration is noted in the legend and confirms the overall increase in Cobalt particle area associated to cell death events with increasing concentrations. The change in distribution shape, with a "flattening" of the curve and lower levels of smaller objects and greater levels of larger ones when comparing the initial to final frames also indicates a general increase in particle size over time as Cobalt agglomerations form.

The final figure, figure 7.14, plots a parameter defined here as apoptotic event dosage which is calculated for each cell by summing the total number of pixels of all the Cobalt agglomeration objects associated to the cell resulting in a measure of the total level of Cobalt present within a given peri-cellular search window.

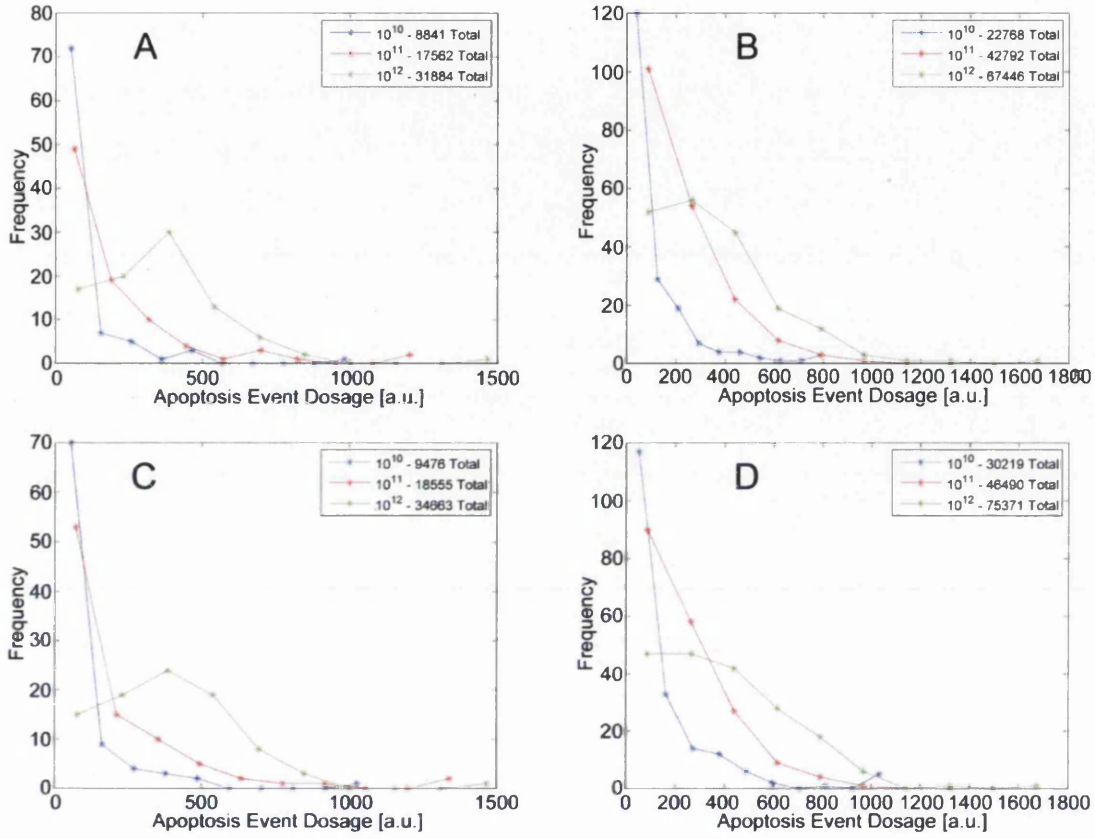



Figure 7.14: A comparison of Cobalt dosages at the time of cell death at concentrations of $10^{10}$ (blue), $10^{11}$ (red) and $10^{12}$ (green) particles per ml. (A) and (B) detail the results of a 10% dilation of the radial search perimeter at 1 and 16 hours respectively and (C) and (D) the same for a 20% dilation.

This parameter can be considered more relevant than those preceding it as it quantifies the total level of Cobalt present around the cell at the time of cell death; it represents the toxicity level a cell could seemingly not withstand. As before the four panels compare the three particle per ml concentrations, $10^{10}$ (blue), $10^{11}$ (red) and $10^{12}$ (green), at times of 1 hour (A & C) and 16 hours (B & D) for radial search windows of 10% (A-B) and 20% (C-D) of the base radius (1/2 average cell diameter). The blue curve ($10^{10}$) shows lesser events at both time frames indicating lower apoptosis rates due to the lower Cobalt concentration level and minimal

particle agglomeration associated with it. In contrast, for the $10^{12}$ concentration (green), there are higher levels of apoptosis as a consequence of the higher Cobalt particle numbers and resultant vigorous agglomeration formation. The red curve ($10^{11}$) shows a greater similarity to the $10^{10}$ concentration curve in terms of overall form suggesting that while agglomerations occur at all the available concentrations it is a more aggressive process at concentrations of $10^{12}$ particles per ml. This similarity between the $10^{10}$ and $10^{11}$ concentrations could also be interpreted as evidence for the agglomerations being the primary trigger for apoptosis rather than sheer quantity of localised Cobalt particles; the size of the Cobalt particles could be decisive in whether a cell does or does not succumb to apoptosis. Over the 16 hour interval the red curve can be seen to shift towards a more linear shape with higher levels of apoptosis around 400 to 600 dosage units. This shift is also seen in the object number counts (figure 7.12) and suggests that the $10^{11}$ concentration provides sufficient particles for the formation of apoptosis inducing agglomerations but that their formation takes longer; it makes rational sense for higher numbers of particles to be capable of forming larger agglomerations at quicker rates. It also corresponds to the lower numbers of Cobalt particles requiring more time to accumulate to toxic level in and around the cells as seen in the object number counts of figure 7.12. The main rise in the dose curve of the $10^{11}$ well at 16 hours also coincides with the peak "hump" of the higher concentration 1012 well at the 1$^{st}$ hour suggesting the region around 400 dosage units to be a significant level of toxicity.

# CHAPTER EIGHT

# Conclusions

The current standard for the automation of image analysis is still the human being and the ability of human mind to rapidly call upon years of experience to interpret complex visual input. Chapters two through four describe methods of visualizing microscopy image data using quantum dot nanoparticles to create a simplified, digital format that is more readily processed by computer systems. Chapters two focused on extracting cell location markers, or centroids, via cluster analysis of internalized quantum dots. Chapter three then details the utilisation of these centroids to automatically generate cell lineages and goes on to demonstrate the extraction of some possible population motility parameters. This method was shown to be capable of automatically tracking cells through mitosis events over a 24 hour time period with an accuracy of approximately 90%. The use of quantum dots instead of fluorescence imaging dyes here offers greater photo- and bio-stability inducing less intrusive effects to the inner workings of the cells under investigation. Transposing the computer vision problem into a more computer friendly format and avoiding complex image interpretation algorithms allows even basic desktop machines to analyse cell populations and extract parameters in real time during the image acquisition stage lessening the overall time expenditure each imaged assay. Real time analysis also offers the potential for smart acquisition where parallel analysis of the already imaged cells affects the parameters of future acquisitions. Such a system could, for example, be used to automatically zoom in on mitosis events as they occur throughout a timelapse sequence and image the event in greater detail. Chapter four details the further potential of multiple quantum dots with differing colours in labelling cells and enabling their successful identification over time intervals of up to 5 hours. While alternative cell labelling techniques are available, the relative ease of the quantum dot uptake protocol and simplicity in which the colour code is generated from the internalized nanoparticles opens the technique to less specialized laboratories. Chapter five explored the extraction of DOX accumulation profiles using only the DOX and DRAQ7 fluorescence channels. Cell death events were identified through the presence of DRAQ7 and the associated DOX object in the

DOX fluorescence channel was tracked until it faded into background noise. Extraction of the resulting accumulation profile did not require the accompanying phase contrast images which would conventionally be used to segment cellular regions to provide perimeters for use as masks to apply to the DOX channel. The automated extraction of profiles associated with 496 cellular apoptosis events provided sufficient data for statistical analysis to infer a cell cycle profile of the imaged population. A similar approach is summarised in chapter six where the level of quantum dot presence at the point of cell death was of interest in quantifying the toxicity of the sample dosage. Chapter seven sought to quantify the toxicological effect of localised Cobalt particle agglomerates through statistical analysis of their silhouettes seen in brightfield images. Again the avoidance of the conventional segmentation approach is employed, this time utilising DAPI as the cell centroid marker at the point of cell death. Automatic identification of Cobalt objects within a defined region about the DAPI marker allows for the accruement of large sample data sets for statistical analysis.

### *Future Developments*

Throughout this text an emphasis was placed on trying to maintain simplicity in terms of the utilised algorithms and computational processes. This approach was taken in an effort to demonstrate the potential of quantum dot nanoparticles as optical markers for cell tracking without the need for performing the difficult task of cell segmentation. With the concept demonstrated, system robustness and versatility can now be developed via the incorporation of more complex computations to improve digital visualization. For example, identification of quantum dot loaded endosomes during the creation of the binary element representations could potentially gain from replacing the simple foreground/background threshold filter with something customised to locally search for point source objects. Fitting point spread functions at this stage could also provide more accurate marker locations on lower resolution images lessening the resolution requirement for operation. Further research into the characterisation of the abstract binary representations could yield methods of cell characteristic extraction as well as opportunities to quantify multiple population dynamics across entire fields of view.

Additionally, an area for continued investigation is the potential for multicolour quantum dot cell-barcoding as a method of cell identification across longer acquisition intervals. Visualizing the three constituent parts of the barcode as a single colour appears to offer a labelling convention with resilience to cross channel variations in quantum dot numbers. Such fluctuations can be attributed to intracellular dynamics and whole-cell motion in and out of ideal focus. A manual operator is very competent at matching similar colours but the automatic linkage of colours in 3-dimensional colour-space is a non-trivial task. Mapping the RGB format colour code into a suitable set of colour-space coordinates and application of dimensional weighting could serve to automate the process.

Another potential improvement comes in the form of high speed imaging at each acquisition point, a procedure that has been looked at but not fully explored. Acquiring a rapid burst of images, of the order of 10Hz, at a time point enables visualization of quantum dot optical marker dynamics. Tracking the motion of individual quantum dot markers allows for their characterisation and markers in the same cell may exhibit differing characteristics to those in another cell. Such differences could be used in conjunction with clustering algorithms, such as k-means, to better identify quantum dot marker clusters and consequently to better identify cell centroid locations. The motions of these markers as observed in high speed image acquisition may also provide a way to infer details relating intrinsic cell properties such as cell cycle position. Combinations of all these ideas applied to the current analytical framework offer signs of extended mileage for the use of quantum dots as moving-light-display optical markers in cellular tracking.

Alongside developments aimed at the quantum dot systems, the process employed to extract DOX profiles to investigate the pSi microparticle delivery vector is to be ported to an operator friendly user interface with the intent on it being used on further timelapse sequences. Feedback and requests from the end user can then be turned into further analysis tools for application on other drug delivery assays.

# References


[1] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of The Cell*, Fifth Edit. .

[2] "Nikon Microscopy U." [Online]. Available: http://www.microscopyu.com/articles/phasecontrast/phasemicroscopy.html.

[3] S. Bradbury and P. . Evennett, *Contrast Techniques in Light Microscopy*. 1996. ISBN-13: 978-1859960851

[4] "Nobeprize - The Phase contrast Microscope." [Online]. Available: http://www.nobelprize.org/educational/physics/microscopes/phase/.

[5] "Nobelprize.org." [Online]. Available: http://www.nobelprize.org/nobel_prizes/physics/laureates/1953/.

[6] G. G. Stokes, "On the Change of Refrangibility of Light," *Philosophical Transactions of the Royal Society of London*, vol. 142, no. 1852, pp. 463–562, 2013.

[7] R. Paschotta, *Encyclopedia of Laser Physics and Technology*. 2008, p. 856. ISBN-13: 978-3527408283

[8] Lodish, Berk, Kaiser, Krieger, Scott, Bretscher, Ploegh, and Matsudaira, *Molecular Cell Biology*, Sixth Edit. ISBN-13: 978-0716776017

[9] M. Minsky, "Memoir on inventing the confocal scanning microscope," *Scanning*, vol. 10, no. 4, pp. 128–138, Aug. 1988.

[10] J. Macqueen, "Some Methods For The Classification and Analysis of Multivariate Observations," vol. 233, no. 233, pp. 281–297, 1967.

[11] W. B. Amos, J. G. White, and M. Fordham, "Use of confocal imaging in the study of biological structures.," *Applied optics*, vol. 26, no. 16, pp. 3239–43, Aug. 1987.

[12] C. J. Sheppard and M. Gu, "Aberration compensation in confocal microscopy.," *Applied optics*, vol. 30, no. 25, pp. 3563–8, Sep. 1991.

[13] S. L., J. B. Simha, and R. Velur, "Seeding Cluster Centers of K-means Clustering through Median Projection," *2010 International Conference on Complex, Intelligent and Software Intensive Systems*, pp. 217–222, Feb. 2010.

[14] W. Kwedlo and P. Iwanowicz, "Using Genetic Algorithm for Selection of Initial Cluster Centers for the K-Means Method," no. 1, pp. 165–172, 2010.

[15] P. Rousseeuw, "Silhouettes - A Graphical Aid to the Interpretation and Validation of Cluster Analysis (1987).pdf," *Computational and Applied Mathematics*, 1987.

[16] F. Meyer, "Topographic distance and watershed lines," *Signal Processing*, vol. 38, no. 1, pp. 113–125, Jul. 1994.

[17] A. E. Carpenter, T. R. Jones, M. R. Lamprecht, C. Clarke, I. H. Kang, O. Friman, D. a Guertin, J. H. Chang, R. a Lindquist, J. Moffat, P. Golland, and D. M. Sabatini, "CellProfiler: image analysis software for identifying and quantifying cell phenotypes.," *Genome biology*, vol. 7, no. 10, p. R100, Jan. 2006.

[18] S. Tse, L. Bradbury, J. W. L. Wan, H. Djambazian, R. Sladek, and T. Hudson, "A Combined Watershed and Level Set Method for Segmentation of Brightfield Cell Images," p. 72593G–72593G–10, Feb. 2009.

[19] M. Tscherepanow, "Automatic Segmentation of Unstained Living Cells in Bright-Field Microscope Images," pp. 158–172, 2008.

[20] L. Kang, M. Chen, and T. Kanade, "Cell Population Tracking and Lineage onstuction with Spatiotemporal Context," *Medical Image Computing*, vol. 12, no. 5, pp. 546–566, 2009.

[21] M. Möller, M. Burger, P. Dieterich, A. Schwab, A. Mathematik, M. Fakultät, C. Gustav, T. U. Dresden, P. Ii, and W. Wilhelms-universität, "A Framework for Automated Cell Tracking in Phase Contrast Microscopic Videos based on Normal Velocities," 2012.

[22] A. Dufour, V. Shinin, S. Tajbakhsh, N. Guillén-Aghion, J.-C. Olivo-Marin, and C. Zimmer, "Segmenting and tracking fluorescent cells in dynamic 3-D microscopy with coupled active surfaces.," *IEEE transactions on image processing□: a publication of the IEEE Signal Processing Society*, vol. 14, no. 9, pp. 1396–410, Sep. 2005.

[23] R. N. Ghosh and W. W. Webb, "Automated detection and tracking of individual and clustered cell surface low density lipoprotein receptor molecules.," *Biophysical journal*, vol. 66, no. 5, pp. 1301–18, May 1994.

[24] R. F. Murphy, "Automated identification of subpopulations in flow cytometric list mode data using cluster analysis.," *Cytometry*, vol. 6, no. 4, pp. 302–9, Jul. 1985.

[25] C. Lo Celso, H. E. Fleming, J. W. Wu, C. X. Zhao, S. Miake-Lye, J. Fujisaki, D. Côté, D. W. Rowe, C. P. Lin, and D. T. Scadden, "Live-animal tracking of individual haematopoietic stem/progenitor cells in their niche.," *Nature*, vol. 457, no. 7225, pp. 92–6, Jan. 2009.

[26] C. Brooks, Q. Wei, L. Feng, G. Dong, Y. Tao, L. Mei, Z.-J. Xie, and Z. Dong, "Bak regulates mitochondrial morphology and pathology during apoptosis by

interacting with mitofusins.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 28, pp. 11649–54, Jul. 2007.

[27] P. B. Deming, C. a Cistulli, H. Zhao, P. R. Graves, H. Piwnica-Worms, R. S. Paules, C. S. Downes, and W. K. Kaufmann, "The human decatenation checkpoint.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 21, pp. 12044–9, Oct. 2001.

[28] K. Schepers, E. Swart, J. W. J. van Heijst, C. Gerlach, M. Castrucci, D. Sie, M. Heimerikx, A. Velds, R. M. Kerkhoven, R. Arens, and T. N. M. Schumacher, "Dissecting T cell lineage relationships by cellular barcoding.," *The Journal of experimental medicine*, vol. 205, no. 10, pp. 2309–18, Sep. 2008.

[29] R. Lu, N. F. Neff, S. R. Quake, and I. L. Weissman, "Tracking single hematopoietic stem cells in vivo using high-throughput sequencing in conjunction with viral genetic barcoding.," *Nature biotechnology*, vol. 29, no. 10, pp. 928–33, Oct. 2011.

[30] A. Gerrits, B. Dykstra, O. J. Kalmykowa, K. Klauke, E. Verovskaya, M. J. C. Broekhuis, G. de Haan, and L. V Bystrykh, "Cellular barcoding tool for clonal analysis in the hematopoietic system.," *Blood*, vol. 115, no. 13, pp. 2610–8, Apr. 2010.

[31] A. Ekimov, "Quantum Size Effect in Three-Dimensional Microscope Semiconductor Crytals," *Journal of Experimental and Theretical Physics Letters*, vol. 34, p. 345, 1981.

[32] A. P. Alivisatos, "Perspectives on the Physical Chemistry of Semiconductor Nanocrystals," vol. 3654, no. 95, pp. 13226–13239, 1996.

[33] X. Michalet, F. F. Pinaud, L. a Bentolila, J. M. Tsay, S. Doose, J. J. Li, G. Sundaresan, a M. Wu, S. S. Gambhir, and S. Weiss, "Quantum dots for live cells, in vivo imaging, and diagnostics.," *Science (New York, N.Y.)*, vol. 307, no. 5709, pp. 538–44, Jan. 2005.

[34] A. M. Derfus, W. C. W. Chan, and S. N. Bhatia, "Probing the Cytotoxicity of Semiconductor Quantum Dots," *Nano Letters*, vol. 4, no. 1, pp. 11–18, Jan. 2004.

[35] L. E. Rikans and T. Yamano, "Mechanisms of cadmium-mediated acute hepatotoxicity.," *Journal of biochemical and molecular toxicology*, vol. 14, no. 2, pp. 110–7, Jan. 2000.

[36] C. B. Murray, D. J. Noms, and M. G. Bawendi, "Synthesis and Characterization of nearly monodisperse CdE Semiconductor Nanocrystallites," *American Chemistry Society*, no. 4, pp. 8706–8715, 1993.

[37] B. O. Dabbousi, F. V Mikulec, J. R. Heine, H. Mattoussi, R. Ober, K. F. Jensen, and M. G. Bawendi, "( CdSe ) ZnS Core-Shell Quantum Dots□: Synthesis and Characterization of a Size Series of Highly Luminescent Nanocrystallites," *Journal of Physcial Chemistry*, vol. 9463, no. 97, pp. 9463–9475, 1997.

[38] C. Kirchner, T. Liedl, S. Kudera, T. Pellegrino, A. Muñoz Javier, H. E. Gaub, S. Stölzle, N. Fertig, and W. J. Parak, "Cytotoxicity of colloidal CdSe and CdSe/ZnS nanoparticles.," *Nano letters*, vol. 5, no. 2, pp. 331–8, Feb. 2005.

[39] M. a. Hines and P. Guyot-Sionnest, "Synthesis and Characterization of Strongly Luminescing ZnS-Capped CdSe Nanocrystals," *The Journal of Physical Chemistry*, vol. 100, no. 2, pp. 468–471, Jan. 1996.

[40] A. Sukhanova, J. Devy, L. Venteo, H. Kaplan, M. Artemyev, V. Oleinikov, D. Klinov, M. Pluot, J. H. . Cohen, and I. Nabiev, "Biocompatible fluorescent nanocrystals for immunolabeling of membrane proteins and cells," *Analytical Biochemistry*, vol. 324, no. 1, pp. 60–67, Jan. 2004.

[41] D. S. Lidke, P. Nagy, R. Heintzmann, D. J. Arndt-Jovin, J. N. Post, H. E. Grecco, E. A. Jares-Erijman, and T. M. Jovin, "Quantum dot ligands provide new insights into erbB/HER receptor-mediated signal transduction.," *Nature Biotechnology*, vol. 22, no. 2, pp. 198–203, 2004.

[42] X. Wu, H. Liu, J. Liu, K. N. Haley, J. a Treadway, J. P. Larson, N. Ge, F. Peale, and M. P. Bruchez, "Immunofluorescent labeling of cancer marker Her2 and other cellular targets with semiconductor quantum dots.," *Nature biotechnology*, vol. 21, no. 1, pp. 41–6, Jan. 2003.

[43] M. Bruchez Jr., "Semiconductor Nanocrystals as Fluorescent Biological Labels," *Science*, vol. 281, no. 5385, pp. 2013–2016, Sep. 1998.

[44] Y. Xiao and P. E. Barker, "Semiconductor nanocrystal probes for human metaphase chromosomes.," *Nucleic acids research*, vol. 32, no. 3, p. e28, Jan. 2004.

[45] W. C. Chan, "Quantum Dot Bioconjugates for Ultrasensitive Nonisotopic Detection," *Science*, vol. 281, no. 5385, pp. 2016–2018, Sep. 1998.

[46] H. Tada, H. Higuchi, T. M. Wanatabe, and N. Ohuchi, "In vivo real-time tracking of single quantum dots conjugated with monoclonal anti-HER2 antibody in tumors of mice.," *Cancer research*, vol. 67, no. 3, pp. 1138–44, Feb. 2007.

[47] T. D. Lacoste, X. Michalet, F. Pinaud, D. S. Chemla, a P. Alivisatos, and S. Weiss, "Ultrahigh-resolution multicolor colocalization of single fluorescent probes.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 17, pp. 9461–6, Aug. 2000.

[48] J. E. Schroeder, I. Shweky, H. Shmeeda, U. Banin, and a Gabizon, "Folate-mediated tumor cell uptake of quantum dots entrapped in lipid nanoparticles.," *Journal of controlled release□: official journal of the Controlled Release Society*, vol. 124, no. 1–2, pp. 28–34, Dec. 2007.

[49] A. Jayagopal, P. K. Russ, and F. R. Haselton, "Surface engineering of quantum dots for in vivo vascular imaging.," *Bioconjugate chemistry*, vol. 18, no. 5, pp. 1424–33, 2007.

[50] H. Kobayashi, Y. Hama, Y. Koyama, T. Barrett, C. a S. Regino, Y. Urano, and P. L. Choyke, "Simultaneous multicolor imaging of five different lymphatic basins using quantum dots.," *Nano letters*, vol. 7, no. 6, pp. 1711–6, Jun. 2007.

[51] R. Bakalova, Z. Zhivko, A. Ichio, M. Kazuto, and M. Milka, "Multimodal Silica-Shelled Quantum Dots□: Direct Intracellular Delivery ," *Bioconjugate Chemistry*, vol. 19, no. 6, pp. 1135–1142, 2008.

[52] H. D. Summers, M. D. Holton, P. Rees, P. M. Williams, and C. a Thornton, "Analysis of quantum dot fluorescence stability in primary blood mononuclear cells.," *Cytometry. Part A□: the journal of the International Society for Analytical Cytology*, vol. 77, no. 10, pp. 933–9, Oct. 2010.

[53] H. D. Summers, P. Rees, M. D. Holton, M. R. Brown, S. C. Chappell, P. J. Smith, and R. J. Errington, "Statistical analysis of nanoparticle dosing in a dynamic cellular system," 2010.

[54] A. Genovesio, T. Liedl, V. Emiliani, W. J. Parak, M. Coppey-Moisan, and J.-C. Olivo-Marin, "Multiple particle tracking in 3-D+t microscopy: method and application to the tracking of endocytosed quantum dots.," *IEEE transactions on image processing□: a publication of the IEEE Signal Processing Society*, vol. 15, no. 5, pp. 1062–70, May 2006.

[55] R. J. Errington, M. R. Brown, O. F. Silvestre, K. L. Njoh, S. C. Chappell, I. a Khan, P. Rees, S. P. Wilks, P. J. Smith, and H. D. Summers, "Single cell nanoparticle tracking to model cell cycle dynamics and compartmental inheritance.," *Cell cycle (Georgetown, Tex.)*, vol. 9, no. 1, pp. 121–30, Jan. 2010.

[56] G. Johansson, "Visual Perception of Biological Motion and a Model for its Analysis," *Perception & Psychophysics*, vol. 14, no. 2, pp. 201–211, 1973.

[57] a. G. Kirk, J. F. O'Brien, and D. a. Forsyth, "Skeletal Parameter Estimation from Optical Motion Capture Data," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, pp. 782–788, 2005.

[58] "Photonics Spectra." [Online]. Available: http://www.photonics.com/Article.aspx?AID=29421.

[59] "Qualysys Motion Capture System." [Online]. Available: www.qualysys.com.

[60] "PhaseSpace Motion Capture." [Online]. Available: www.phasespace.com.

[61] "Vicon." [Online]. Available: www.vicon.com.

[62] H. D. Summers, P. Rees, M. D. Holton, M. R. Brown, S. C. Chappell, P. J. Smith, and R. J. Errington, "Statistical analysis of nanoparticle dosing in a dynamic cellular system.," *Nature nanotechnology*, vol. 6, no. 3, pp. 170–4, Mar. 2011.

[63] T. G. Graves, M. W. Harr, E. L. Crawford, and J. C. Willey, "Stable low-level expression of p21WAF1/CIP1 in A549 human bronchogenic carcinoma cell line-derived clones down-regulates E2F1 mRNA and restores cell proliferation control.," *Molecular cancer*, vol. 5, p. 1, Jan. 2006.

[64] "Molecular Probes - Cell Labeling and Detection." [Online]. Available: www.invitrogen.com.

[65] "CellVue Cell Labeling Kits." [Online]. Available: www.licor.com.

[66] "Fluorescence Kits." [Online]. Available: www.sigmaaldrich.com.

[67] F. Increase, "BD Phosflow Violet Fluorescent Cell Barcoding Kit Increase throughput with fluorescent cell barcoding technology," 2006. [Online]. Available: http://www.bdbiosciences.com/eu/documents/X23-12946-00.pdf.

[68] T. Murata, H. Yamawaki, R. Yoshimoto, M. Hori, K. Sato, H. Ozaki, and H. Karaki, "Chronic effect of doxorubicin on vascular endothelium assessed by organ culture study.," *Life sciences*, vol. 69, no. 22, pp. 2685–95, Oct. 2001.

[69] D. von Hoff, M. Layard, P. Basa, H. Davis, A. von Hoff, M. Rozencweig, and F. Muggia, "Risk Factors for Doxorubicin-Induced Congestive Heart Failure," *Annals of Internal Medicine*, no. 91, pp. 710–717, 1979.

[70] "ImmunoChemistry Technologies LLC." [Online]. Available: www.immunochemistry.com.

# Glossary

**Antibody** – protein used by the immune system to identify foreign objects

**Apoptosis** – the process of programmed cell death (variants; apoptotic)

**Background** – in computer vision, the intensity region of pixels of non-interest (see also *Foreground*)

**Barcoding** – process of labelling an object with a unique identifier

**Biocompatibility** – ability to function without harmful effects of biological host

**Biomolecule** – a molecule produced by a living organism (plural. biomolecules)

**Bio-stable** – unchanged when in contact with biological systems

**Binary Element** – Digital representation of a quantum dot loaded endosome

**Binary Element Display** – A moving light display construct using binary elements as markers

**Brownian Motion** – random motion of particles suspended in a fluid

**Calcein** – fluorescent dye with 495nm excitation and 515nm emission wavelengths

**Catenate** – to joined along dimensions of the same magnitude (variants; catenated, catenation)

**Centroid** – geometric centre

**Conjugate** – join together

**Chromophore** – part of a molecule responsible for its colour

**Chromosome** – an organised DNA-protein structure within cells

**Chromosomal** – of the chromosome

**Confluency** – measure of cell coverage of a disk or flask (variants; confluent, confluence)

**Cytoplasm** – fluid enclosed within the cell membrane; contains cytosol, organelles and inclusions

**Cytotoxic** – of, relating to or producing a toxic effect on cells

**DOX** – Doxorubicin, a cancer drug used in chemotherapy

**DRAQ7** – a far-red fluorescent DNA dye the only stains the nuclei in dead and permeabilized cells

**Endocytosis** – mechanism of cell absorption of molecules by engulfing them (variants; endocytosed – having been through endocytosis; endocytic/endosomal – of or pertaining to the endocytosis process, e.g. endosomes are endocytic vessels)

**Endosome** – membrane-bound compartment of the endocytic membrane transport pathway from plasma membrane to the lysosome

**Endothelial** – of the endothelium, cell layer lining of certain body cavities, e.g. blood cells

**Epithelial** – of the epithelium, membranous tissues composed of cell layers separated by very little intercellular substance.

**Extravasation** – passage of objects from blood cells into tissue

**Fluorochrome** – fluorescent chemical compound that can re-emit light upon excitation

**Fluorophore** – see *fluorochrome*

**Flourescein** – synthetic organic compound used as a fluorescent tracer

**Folate** – organic compound of folic acid

**Foreground** – in computer vision, the intensity region of the object of interest

**Fulguration** – flashing of light

**Immunofluorescence** – fluorescence microscopy technique that utilises antibodies to target biomolecules with fluorescent dyes

**Impermeant** – unable to pass through a given semi-permeable membrane

**Informatics** – study of processing, management and retrieval of information

**Intercalation** – reversible inclusion of a molecule/group between two other molecules/groups

**Internalization** – see *endocytosis*

**Intracellular** – occurring within a cell or cells

**In Vitro** – taking place outside of a living organsim

**In Vivo** – taking place inside a living organism

**Micelle** – aggregate of surfactant molecules dispersed in a liquid colloid

**Microparticle** – particle of dimensions between 0.1 and 100 micrometers

**Morphology** – form of an organism and their specific structural features (variants; morphological; morphologically)

**Motility** – parameter of motion; cell with zero motility remain static

**Nanoparticle** – particle of dimensions less than 100 nanometers

**Nucleal** – of, or pertaining to, a nucleus

**PEI** – polyethyleneimine

**Peptide** – short chain of amino acid monomers linked by peptide bonds

**Perinuclear** – situated around a nucleus

**Photoluminescence** – process where a substance absorbs photons then re-radiates photons

**Photonic** – of or pertaining to photons

**Photo-stable** – unchanged under exposure to light

**PI** – Propidium Iodide

**pSi** – porous Silicon

**QD** – (var. QDot) Quantum Dot

**Radionuclide** – an atom with an unstable nucleus and consequent excess energy (often referred to as radioactive isotopes or radioisotopes)

**Raster** – rectangular pattern of parallel scanning lines

**Toxicology** – study of the adverse effects of chemicals on living organisms

**Transferrin** – iron-binding blood plasma glycoproteins

**Trastuzumab** – monoclonal antibody that interferes with the HER2/neu receptor

**Vasculature** – blood vessels or arrangement of blood vessels

**Vesicle** – type of organelle; compartment within a cell enclosed by a lipid bilayer

**Vernoi Diagram** – representation of the division of a given space into regions

# Bibliography

'MathWorks' <www.mathworks.co.uk>

Alberts, Bruce, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter, *Molecular Biology of The Cell*, Fifth Edit

Ali, Rehan, Mark Gooding, Martin Christlieb, Michael Brady, and F R S Feng, 'PHASE-BASED SEGMENTATION OF CELLS FROM BRIGHTFIELD MICROSCOPY Wolfson Medical Vision Labs , Dept of Engineering Science , University of Oxford Nuffield Dept of Obstetrics and Gynaecology , University of Oxford', 2007, 57–60 <doi:10.1109/ISBI.2007.356787>

Anas, Abdulaziz, Tetsuya Okuda, Nagako Kawashima, Kenichi Nakayama, Tamitake Itoh, Mitsuru Ishikawa, and others, 'Clathrin-Mediated Endocytosis of Living Cells', 3 (2009), 2419–2429

Anoraganingrum, Dwi, 'Cell Segmentation with Adaptive Region Growing', 1999 <doi:10.1.1.46.7183>

Awodey, Steve, Notes on *Algebraic Set Theory*, 2005, pp. 1–31

Belongie, S., J. Malik, and J. Puzicha, 'Shape Matching and Object Recognition Using Shape Contexts', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24 (2002), 509–522 <doi:10.1109/34.993558>

Bradley, Paul S, and P S Bradley, 'Refining Initial Points for K-Means Clustering', *Microsoft Research*, 1998, 91–99

Buehler, Patrick, Mark Everingham, Daniel P. Huttenlocher, and Andrew Zisserman, 'Upper Body Detection and Tracking in Extended Signing Sequences', *International Journal of Computer Vision*, 95 (2011), 180–197 <doi:10.1007/s11263-011-0480-9>

Canny, J, 'A Computational Approach to Edge Detection.', *IEEE transactions on pattern analysis and machine intelligence*, 8 (1986), 679–98 <http://www.ncbi.nlm.nih.gov/pubmed/21869365>

Cedras, C, and M Shah, 'A Survey of Motion Analysis From Miving Light Displays', 1994, 214–221 <doi:10.1109/CVPR.1994.323832>

Chakraborty, Papiya, 'Histogram Equalization by Cumulative Frequency', *Scientific and Research*, 2 (2012), 1–4

Dente, J A, 'An Autonomous Approach to the Mountain-Clustering Method', *Uncertainty Modelling And Analysis*, 1995, 649–654 <doi:10.1109/ISUMA.1995.527771>

Devlin, Keith, *The Joy of Sets*, 2nd edn, 1993; ISBN-13: 978-0387940946

Gmip, Cvgip, and Terrance E Boult, 'Local Image Reconstruction and Subpixel Restoration Aglorithms', 55 (1993), 63–77 <doi:http://dx.doi.org/10.1006/cgip.1993.1005>

Gu, Lei, and Fuchun Sun, 'Two Novel Kernel-Based Semi-Supervised Clustering Methods by Seeding', *2009 Chinese Conference on Pattern Recognition*, 2009, 1–5 <doi:10.1109/CCPR.2009.5344157>

Guillaumin, Matthieu, Thomas Mensink, Jakob Verbeek, and Cordelia Schmid, 'Face Recognition from Caption-Based Supervision', *International Journal of Computer Vision*, 96 (2011), 64–82 <doi:10.1007/s11263-011-0447-x>

Gultepe, Evin, Dattatri Nagesha, Srinivas Sridhar, and Mansoor Amiji, 'Nanoporous Inorganic Membranes or Coatings for Sustained Drug Delivery in Implantable Devices.', *Advanced drug delivery reviews*, 62 (2010), 305–15 <doi:10.1016/j.addr.2009.11.003>

Hathaway, Richard J., and James C. Bezdek, 'Recent Convergence Results for the Fuzzy C-means Clustering Algorithms', *Journal of Classification*, 5 (1988), 237–247 <doi:10.1007/BF01897166>

Healthcare, G E, and Life Sciences, *IN Cell Analyzer High-Content Cellular Analysis System*

Kalyanaraman, B, Joy Joseph, Shashi Kalivendi, Suwei Wang, Eugene Konorev, and Srigiridhar Kotamraju, 'Doxorubicin-induced Apoptosis: Implications in Cardiotoxicity.', *Molecular and cellular biochemistry*, 234-235 (2002), 119–24 <http://www.ncbi.nlm.nih.gov/pubmed/12162424>

Mao, Zhihong, Lizhuang Ma, Mingxi Zhao, and Xuezhong Xiao, 'SUSAN Structure Preserving Filtering for Mesh Denoising', *The Visual Computer*, 22 (2006), 276–284 <doi:10.1007/s00371-006-0005-7>

Marr, D., and E. Hildreth, 'Theory of Edge Detection', *Proceedings of the Royal Society B: Biological Sciences*, 207 (1980), 187–217 <doi:10.1098/rspb.1980.0020>

Menotti, David, Laurent Najman, Jacques Facon, and Arnaldo De A Araújo, 'Multi-Histogram Equalization Methods for Contrast Enhancement and Brightness Preserving', 53 (2007), 1186–1194 <doi:10.1109/TCE.2007.4341603>

Mirkin, Boris, 'Choosing the Number of Clusters', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1 (2011), 252–260 <doi:10.1002/widm.15>

Mizutani, Hideki, Saeko Tada-Oikawa, Yusuke Hiraku, Michio Kojima, and Shosuke Kawanishi, 'Mechanism of Apoptosis Induced by Doxorubicin

Through the Generation of Hydrogen Peroxide.', *Life sciences*, 76 (2005), 1439–53 <doi:10.1016/j.lfs.2004.05.040>

Padfield, Dirk, Jens Rittscher, Nick Thomas, and Badrinath Roysam, 'Spatio-temporal Cell Cycle Phase Analysis Using Level Sets and Fast Marching Methods.', *Medical image analysis*, 13 (2009), 143–55 <doi:10.1016/j.media.2008.06.018>

Pan, Jiyan, Takeo Kanade, and Mei Chen, 'Learning to Detect Different Types of Cells Under Phase Contrast Microscopy', 2009, pp. 1–8

Pawley, James, *Handbook of Biological Microscopy*, 3rd edn, 2006; ISBN-13: 978-0387259215

Petrou, Maria, and Costas Petrou, *Image Processing - The Fundamentals*, 2010; ISBN-13: 978-0470745861

Roitman, Judith, *Introduction to Modern Set Theory*, 2011; ISBN-13: 978-0982406243

Scita, Giorgio, and Pier Paolo Di Fiore, 'The Endocytic Matrix.', *Nature*, 463 (2010), 464–73 <doi:10.1038/nature08910>

Smith, Stephen M, and J Michael Brady, 'SUSAN — A New Approach to Low Level Image Processing', 23 (1997), 45–78 <doi:10.1023/A:1007963824710>

Sun, Xue, Kunlun Li, Rui Zhao, and Xikun Hu, 'Global Optimization for Semi-supervised K-means', *2009 Asia-Pacific Conference on Information Processing*, 2009, 410–413 <doi:10.1109/APCIP.2009.237>

Suri, Jasjit, Handbook of Biomedical Image Anlysis, Volume II: *Segmentation Models*, 2005; ISBN-13 978-0306486050

Yager, Ronald, and Dimitar Filev, 'Approximate Clustering via Mountain Method', 24 (1994), 1279–1284

Yin, Zhaozheng, Ryoma Bise, Mei Chen, and Takeo Kanade, 'CELL SEGMENTATION IN MICROSCOPY IMAGERY USING A BAG OF LOCAL BAYESIAN CLASSIFIERS Carnegie Mellon University , 2 Intel Labs Pittsburgh', 2010, I, 125–128 <doi:10.1109/ISBI.2010.5490399>

# Appendix A

Paper submitted to the journal PLOS ONE, January 2012.
Title: ***"Automated Cell Identification and Tracking using Nanoparticle Moving-light-displays"***
Published in July 2012
doi:10.1371/journal.pone.0040835

**Authors:**
Tonkin JA, Summers H, Rees P, Brown MR, Errington RJ, Smith PJ

**ABSTRACT**
An automated technique for the identification, tracking and analysis of biological cells is presented. It is based on the use of nanoparticles, enclosed within intra-cellular vesicles, to produce clusters of discrete, point-like fluorescent, light sources within the cells. Computational analysis of these light ensembles in successive time frames of a movie sequence, using k-means clustering and particle tracking algorithms, provides robust and automated discrimination of live cells and their motion and a quantitative measure of their proliferation. This approach is a cytometric version of the *moving light display* technique which is widely used for analyzing the biological motion of humans and animals. We use the endocytosis of CdTe/ZnS, core-shell quantum dots to produce the light displays within an A549, epithelial, lung cancer cell line, using time-lapse imaging with frame acquisition every 5 minutes over a 40 hour time period. The nanoparticle moving light displays provide simultaneous collection of cell motility data, resolution of mitotic traversal dynamics and identification of familial relationships allowing construction of multi-parameter lineage trees.

**INTRODUCTION**
Computerized identification, discrimination and tracking of biological cells, in microscopy images, is vital to modern high throughput platforms that deliver automated scanning and capture of millions of images per day [1-3]. Rapid, machine-based image analysis is now essential as the data generation rate far exceeds human processing capacity and many of the key challenges in cell biology demand knowledge of all individuals within large cell populations, e.g. understanding the role of heterogeneity and division asymmetry in cancer [4-7] or stem cell proliferation and differentiation [8]. Through the use of ever-increasing processing speed and capacity and evolving microscopy techniques, automated cell identification and spatio-temporal tracking is now widely used [9-11]; however it is far from straightforward to implement and requires computational algorithms and imaging science beyond that common to standard microscopy. Thresholding and segmentation routines used to identify cell outlines are often complex, reflecting the intrinsic problem of poor optical contrast within epi-illuminated or bright-field images, caused by the minimal refractive index differences between cells and their surrounding environment. Phase contrast or fluorescence imaging modalities alleviate some of these problems [12-13] but have varying applicability across cell-types due to changing optical density in the case of phase-based techniques or necessitate intervention in the cell biology to introduce fluorescence markers, e.g.

GFP transfection, antibody loading or DNA staining; this can interfere with natural cell function and so application to live cells is limited [14]. Even when successful cellular image analysis has been implemented there often remains a fundamental imbalance between data acquired and information processed: large data-set images are taken at sub-cellular resolution and then processed to produce much simpler, whole cell parameters such as cell identity, type, position etc. This is in-efficient processing of information and imposes an overhead on hardware performance, computational power and data analysis time.
These computerized approaches mimic human visual perception of form and motion where dense and complex image information is processed to obtain much simpler, abstract representations of objects and their position. However, through early studies by Wertheimer and others on the relationship between perception and simplified abstractions, such as points or lines, it is now known that human perception can operate directly at the level of the abstract object and so does not require detailed information – the human form of a 'stick-person' is recognizable despite consisting only of straight lines and a circle. This is the *Gestalt* ("unified whole") theory of visual perception [15] and its consequence to image analysis is that acquisition need not incorporate the full spatial detail of the object. This realization was put to practical use in the early 1970's by Johansson who utilized our ability to accurately discriminate and track objects with minimized information by studying human motion using moving light displays (MLD), created from video sequences of high contrast optical sources attached to the joints of a moving person or animal [16]. The technique has been widely adopted in the computer image community and is now routinely used for optical motion capture and animation through imaging of dark suited actors with bright optical sources or reflectors, positioned at key points, which describe the mechanics of movement [17-19]. In the context of imaging cytometry the MLD technique demonstrates that accurate identification and quantification of cell motion does not require high spatial resolution of cellular form and structure and can be performed with a low number of binary optical markers associated with the cell.

In this paper we report on the implementation of the moving-light-display technique in cells by the use of lysosome-encapsulated quantum dots (QDs) to create clustered points of light, through which cells can be identified, tracked and analyzed. The creation of the optical sources through endocytosis of fluorescent markers provides a generic and innate encoding mechanism, applicable across multiple cell types. The nanoparticles provide robust, bio-stable and photo-stable fluorescence that can be tracked over multiple generations, and at nanomolar concentrations do not perturb cellular function [20]. Under typical loading conditions there are between 10-50 vesicular light sources; these are resolvable at low magnification and provide good signal to noise discrimination due to particle concentration within the vesicles. A typical panel of images of quantum dot labeled cells is shown in figure 1; this depicts the acquired bright-field (1A) and fluorescence (1B) images and a binary map created from the fluorescence data (1C, see materials and methods for details of the image processing). The fluorescently-labeled lysosomes are observed as clusters of light sources, the cellular encapsulation of which provides a localization of the cluster to the intra-cellular domain; this is sufficient to discriminate between clusters and hence identify individual cells. Cell movement is analyzed through identification of the cluster centroid and tracking of its spatial translation in successive time frames. The use of point tracking in biological image analysis is well established [21-

23]; here we are adapting the approach to obtain integrative information at the level of the cell rather than the individual points. As in human MLD the light points allow recognition of the form and motion of the gross object (the cell) through a reduced information set of 10-50 binary-valued pixels (the cluster centroids). The inter-relationship of light sources within a cluster is determined by the relative motion of the labeled lysosomes and hence it carries biological information relating to cellular state and function. For example, the local-level metric of mean point separation within a cluster provides a quantitative measure of a cell's traversal through mitosis with associated spatial arrangements of the QD-loaded lysosomes providing specific, easily recognizable geometrical motifs of cell division.
(details presented in section on *analysis of cell division*)

In the following sections a demonstration of the moving-light-display concept within cells is presented; starting with details of the microscopy and image analysis techniques used and then expanded through reference to the biologist's requirement for i. identification and discrimination of cells, ii. spatio-temporal tracking of their motion and division, iii. analysis of division events and iv. visualization of time dependent relationships in cell lineage maps. We conclude with a discussion on the applicability of the technique and a general summary of the nanoparticle MLD approach.

## MATERIALS AND METHODS

*Cell culture*

A549 (ATCC CCL-185) cells were maintained under G418 selection in McCoy's 5a medium supplemented with 10% fetal calf serum (FCS), 1mM glutamine, and antibiotics and incubated at 37°C in an atmosphere of 5% $CO_2$ in air. For imaging experiments, cells were grown at a density of $1x10^6$ cells $ml^{-1}$ as a monolayer in either coverglass bottomed chambers (Nunc, 2 Well Lab-Tek II, Fisher Scientific) or glass bottomed (24 multi-well Sensoplate, Greiner Bio-one for 24 h prior to imaging. All cell concentrations were determined using a Coulter Particle Counter (Beckman Coulter, High Wycombe, UK).

*Nanoparticle loading*

Cells were loaded with commercially available targeted nanocrystals using the Qtracker® 705 (QTracker705) Cell Labeling Kit (Invitrogen (Q25061MP) at 4 nM concentration. The reagents in the Qtracker® 705 Cell Labeling Kit use a custom targeting peptide (9-arginine peptide) to deliver near-infrared-fluorescent nanocrystals into the cytoplasm of live cells via the endosomal pathway. Briefly, Qtracker reagent A and B were premixed and then incubated for 5 minutes at room temperature. 1ml of fresh full growth media was added to the tube and vortexed for 30 seconds. This labeling solution was then added to each well of the cells and incubated for 1 hour at 37°C after which they were washed twice with fresh media. Subsequently 24 hours later, labeled cells were then analyzed by time-lapse, confocal microscopy.

*Image acquisition*

Confocal laser scanning microscopy (Radiance CLSM, BioRad Ltd) was used to track quantum dot labeled A549 cells over a 48 hour period. The Qtracker705 fluorescence was collected using 488 nm excitation and 680-20 nm emission filters; x,y,z,t optical sections (using x40, 0.75 NA air lens) were collected every 5 minutes.

*Image pre-processing*

All image processing was done within the MATLAB programming environment. The fluorescence image data was provided in the format of multi-layer tiff stacks each containing 8 focal plane images for every 5 minute interval across the 48 hour experimental range. The capture of multiple focal planes allowed creation of a composite image using the highest contrast regions of all available images. This composite was created by segmenting the image space into smaller regions and applying the standard absolute gradient algorithm of the form:

$$F_{grad} = \sum_{Height} \sum_{Width} |I(x+1, y) - I(x, y)| \quad (1)$$

where I(x,y) is the intensity of the pixel (x,y).

Conversion to a binary representation is a simple two step process.

i. A step function of the form

$$F_s(x, y) = \begin{cases} 1 \; if \; I(x, y) \geq \theta \\ 0 \; if \; I(x, y) < \theta \end{cases} \quad (2)$$

where θ is a pixel intensity threshold defined as

$\theta = \mu_{image\ intensity} + \sigma_{image\ intensity}$ (3)

was applied to the composite fluorescence image to remove background noise. Here μ is the image intensity average and σ the standard deviation added to account for variability of background.

ii. A simple peak finder algorithm locating the maximum pixel intensity within a localized area by cross-referencing the linear profiles of the pixel rows and columns was applied to locate the maximum intensity points of the fluorescent signals corresponding to nanoparticle loaded vesicles and their x and y coordinates stored.

*Cluster analysis – the first pass*

A standard k-means clustering analysis algorithm was used to identify groups of pixels in the binary image (binary elements) corresponding to fluorescent vesicles within the same cell. For the binary elements coordinates $(b_1, b_2, \ldots, b_n)$ we iteratively used the function:

$$\arg\min_C \sum_{i=1}^{k} \sum_{b_j \in C_i} \left\| b_j - \mu_i \right\|^2 \quad (4)$$

to find the cluster centroid location. Here $C = \{C_1, C_2, \ldots, C_k\}$ is the set of clusters with total cluster number k, $b_j$ is the set of binary elements associated with the cluster $C_i$ and $\mu_i$ is the mean of the binary points in $C_i$. The seeding of the centroid locations in the initial timeframe was done manually and the k-means algorithm then applied to refine their coordinates to the true cluster centers. The in-built k-means algorithms within Matlab proved to be accurate provided the initial seed centroid locations were accurately assigned by the user.

In successive images the centroid locations from the previous timeframe were used as a seeding set. To deal with the sporadic occurrence of rogue binary elements (noise) and binary elements representing cells entering the field of view, new centroids are assigned to regions containing binary elements that are more than 120 pixels (30μm or ~3 cell diameters) from their nearest centroid. New seeds are

assigned to deal with both these occurrences as they are indistinguishable without time consuming comparisons with previous frames and interpretation of boundary events. It is temporally and computationally more efficient to assign binary elements to new clusters generated according to basic rules and interpret the nature of the groupings during later processes. A proximity validation was applied by identifying current seeds with no binary elements within a distance of 50 pixels (12.5μm) and removing them. This primarily deals with the event of a cluster moving out of the field of view therefore leaving a centroid seed from the previous frame with no binary elements to define it and secondarily with centroids previously defined to account for rogue elements whose intensities have dropped back below the noise filter. The k-means algorithm was then run using the modified centroid set and a measure of the cluster fit taken by implementing a native Matlab silhouetting algorithm to acquire a cluster fit parameter. The Silhouette process provides a validation of the clustering by determining how well each binary element fits within its assigned cluster. For each binary element *i*, let $a_i$ be a measure of dissimilarity with elements in the same cluster, in our case the average distance from all other cluster members. Perform the same operation comparing element *i* with all elements in all other clusters successively and assign the lowest dissimilarity measure as $b_i$. Now define the silhouette of the binary element *i* as

$$s_i = \frac{b_i - a_i}{\max\{a_i, b_i\}} \qquad (5)$$

where $-1 \le s_i \le 1$. A value close to 1 means the element fits well with its assigned cluster and the average $s_i$ of the all binary elements is a measure of how well assigned clusters are. Following optimal assignment of cluster centroids an average binary element separation parameter was calculated for each cluster sequentially. If this parameter was found to be greater than 15μm, ~1.5 average cell diameters, it was likely that a division event had occurred and so the k-means algorithm is run again with an additional centroid seed placed 10 pixels from the centroids associated with probable division events. The silhouette cluster fit parameter was then re-calculated and if the value found to be more favorable the additional centroid is accepted. The process was reiterated until all clusters were validated.

*Cluster refinement – centroid linkage and the mitotic signature*

Refinement of the spatial tracking was accomplished by temporal tracking of the centroids and by identification of cell division events through further analysis of the separation of binary elements within clusters.

The centroids in each time frame were linked to centroids in surrounding time frames through application of a nearest neighbor map. For two sequential frames with centroid sets $C_i$ and $C_{i+1}$ where $C_i > C_{i+1}$ we calculated the Euclidian distance between each member of $C_i$ and $C_{i+1}$ and then assigned nearest neighbors in ascending order until each centroid of $C_{i+1}$ was linked to a corresponding centroid in $C_i$. The remaining centroids in $C_i$ were then linked to their respective closest centroids in $C_{i+1}$ provided they are within the average single cell diameter of 10μm otherwise they were deemed to be new progenitor groupings. The centroid lineages were then stored as vectors of length $|C_i|$ whose entries are the indices of the corresponding closest centroid in $C_{i+1}$. Division events revealed themselves at this point when a single centroid in one frame had two nearest neighbors in the next. With lineages initially defined the temporal placement of division events was further

refined by tracking each of the lineages individually through all timeframes and compiling mean distance profiles to search for the characteristic cluster contraction and expansion dynamic that is a signature of cell mitosis.
(see *analysis of cell division*)

## RESULTS

### Cell identification

A typical example of the embedding of point light sources within cells, through the use of QDs, is shown in figure 1. The nanoparticle loaded vesicles form a perinuclear ring of punctate fluorescence (fig 1B). It is the location of the vesicles rather than the overall light distribution that we require. The fluorescence image is therefore filtered to identify points of peak intensity which are then used to create a binary image, in which, discrete pixels form a digital map of points (fig 1C) representative of the nanoparticle vesicles (see materials and methods section for full detail). We refer to these point sources as 'binary elements'. It is the spatial relations and temporal motion of the binary elements that provide identification of cells and analysis of their motion, function and proliferation. The localization of nanoparticle loaded vesicles within individual cells produces clustering of the binary elements which can be recognized through a k-means clustering analysis, and represented by, a centroid marker which corresponds to the geometrical centre of the binary elements (fig 1C). Thus the complex graphical information of the bright field image (fig 1A) is reduced to a set of Cartesian co-ordinates (the centroid markers), each of which represents a cell, and through which the spatio-temporal behavior of the cell population is analyzed. Previous studies have shown the QTracker705 QDs to be photo-stable over many days and that the QD loaded lysosomes are conserved upon cell division [24]. Identified cells can therefore be tracked across the cell cycle and the proliferating population mapped through detection of division events and the associated daughter cells.

### Spatio-temporal tracking of cells

Once an initial field of cells has been identified and cluster centroids assigned, a full spatio-temporal track can be obtained through linkage of the centroids through successive time frames. Centroid seeds for each successive frame are defined to optimize the speed of the k-means process, account for the possibility of losing and gaining clusters off the edge of the plane of view and deal with rogue binary element points occurring sporadically through time in otherwise empty regions. The general approach taken can be summarized as follows: 1. generate a foundation set using centroid locations from the previous time frame, 2. validate proximity of all centroids to binary elements to identify centroids whose binary element cluster has moved out of view and 3. validate proximity of binary elements to centroids to account for any noise elements that passed through the filter. Images were taken with a 5 minute time interval; this minimizes the probability of large changes in cell position and thus ensures accurate correlation of centroid locations from frame to frame.

A movie sequence of this centroid motion and the 'birth' of new centroids in daughter cells, taken over a 48 hr period, is shown in Movie S2; single images, taken at the 0 and 40 hr time points, are shown in figure 2. A composite, overlay image shows the relation of the QD-encoded binary elements to the cluster centroid and to the detailed topology of the cell, as seen in the bright field image (figure 2A). The

A549 cell-line has a mean inter-mitotic time of ~ 22 hours [25] and so there are 2-3 rounds of cellular division during the time of the experiment. The images in figure 2B and 2C clearly show this proliferation, the centroid number increases from 41 to 133 and the average number of binary elements per cluster reduces from 34 to 12. These cluster statistics are consistent with a cell doubling (QD vesicle number halving) time of ~23 hours.

Analysis of cell division

As cells go through mitosis they undergo distinct morphological changes that can be tracked via the binary elements. The mean distance of the binary elements from each cell centroid provides a characteristic parameter, from which a 'mitotic curve' may be constructed (figure 3). As cells enter the mitotic phase they detach from the adherent surface and round-up to a spherical geometry, this condenses the nanoparticle vesicle distribution and produces a reduction in the mean binary element separation (fig 3A-3B). After chromosomal separation the nanoparticle vesicles disperse during cell telophase and the mitotic curve begins to increase (fig 3B-3C). Upon the completion of mitosis the binary element spacing continues to grow as two clusters separate in the daughter cells (fig 3C-3D); this latter increase arises from the computational assignment of elements to a single centroid whereas physically there are clearly two separate groupings following cellular division. We therefore make a transition from a single to a pair of centroids and reassign the binary elements to daughter cells when the average separation of a cluster after the contraction phase surpasses a predefined level, defined as the stable separation distance maintained prior to contraction (fig. 3C).

The mitotic curve provides not only a clear digital marker of cell division events but also a quantitative, analogue track through the mitotic phases which informs on the kinetics of the cell division process (see supporting media for further examples). Here the moving point display technique provides a tool with which the shape and motion of cells can be analyzed. The linkage of cell division events through spatio-temporal tracking provides a functional lineage analysis capable of describing the clonal relationships of a wide range of morphological and motility measures. As an example, a motility lineage map is shown in figure 4 in which the inter-generational relationships are mapped by velocity vectors, displaying the mean speed and net direction of inter-mitotic cellular motion.

Accuracy and robustness

In order to numerically quantify the operational range of the nanoparticle-encoded MLD we take a representative sample image and investigate the effects on cluster identification and assignment of centroid position due to variation in the noise filter threshold (figure 5). This approach is chosen from a wide range of alternatives as it directly relates the accuracy of the technique to the number of encoding light points (nanoparticle loaded vesicles) and corresponds to the experimental measurable of fluorescence signal to noise (SNR). Increasing the intensity of the thresholding step function reduces the number of binary elements identified, as noisy, rogue pixels are filtered out (fig. 5A). From the conversion plot in fig 5A we calculate the number of cluster centroids identified and the displacement of a selection of these, as a function of the total number of binary elements within the image frame (fig. 5B). There is an operational range between ~ 500-1500 elements over which the technique provides

accurate identification of cell identity and position independent of the QD signal SNR. The accurate identification of cells is determined by the number of binary elements available to define a cell cluster. Using the initial timeframe for this operational range quantification, it was found that 520 binary elements were needed for all 38, manually identified, cluster centroids to have at least 1 binary element assigned to them (the red dashed line in fig. 4B). This corresponded to a noise intensity cut-off of 784 units (SNR = 2.8) and is a 'worst case' in that sufficient binary elements had to be defined to locate all centroids. Relaxing this criterion still identifies the majority of centroids by a minimum of 5 binary elements (30 centroids located from 200 binary pixels) using a threshold cut-off of 1100 (SNR = 2). The minimum number of elements necessary to identify a cell is an important criterion in studying proliferating cells as the QD loaded vesicles will be diluted upon division [24]. Over the 40 hour time series shown in figure 2 the mean cluster number per cell reduces to 12, we would anticipate therefore that 72 hours of tracking could be achieved (3-4 cell divisions) before over-dilution makes cell identification impossible.

If the noise threshold filter is reduced too far the binary element number begins to relate to background noise rather than valid QD encoding pixels. Whilst the presence of these noisy pixels does not immediately invalidate the identification of cells (a random noise source adds pixels to each cell cluster with equal probability) it does affect the centroid co-ordinates and so leads to inaccurate determination of cell position. As the threshold filter approaches the noise floor (intensity ~ 400 units) and the total number of binary elements increase beyond 1500 the centroid positions show marked deviations in excess of average cell diameters (> 10μm). However SNR values as low as 1.1 can be tolerated before this noise weighting produces a cluster position offset of 10μm.

To assess the accuracy of centroid seeding (cell identification) and spatio-temporal tracking two image frames corresponding to the 0 and 24 hour time points were chosen. An initial centroid set was chosen at the intial timepoint by visual inspection of the bright field image; the number of centroids in the timeframe captured 24 hours later was then automatically assessed using the MLD algorithms and compared to direct visual identification. The 24 hour imageframe is shown in figure 5C with automatically identified centroids overlaid on the bright field image; 81% of manually identified cells are correctly mapped.

## DISCUSSION

Quantum dots are widely used for cellular labeling as they provide both a photo and bio stable fluorescence marker that can be spectrally tuned. Here we show that the processing of these nanoparticles by the cell is as important as their innate photonic properties; they are naturally taken up and concentrated into vesicles that are dispersed throughout the cytoplasm and as such provide bright, point like light sources across the majority of the cell cytoplasm. This bio-processing thus enables image analysis by these discrete points which are the fundamental elements of a moving light display. Concentration on representative sub-sampling of the image rather than full visualisation has two major benefits: i. major data reduction and ii. access to rapid and efficient image examination based on cluster analysis techniques. The processing of a bright field image to produce a binary map of QD light-points

(fig. 1A-C) leads to a reduction in data size from ~2.5MB to 20 kB per image. In high-throughput applications, where standard acquisition may produce 1 TB of data per day, there is a pressing need for such data reduction to avoid the ever increasing penalty in cost and time of data analysis. Throughout this work we have knowingly adopted simple algorithms to demonstrate the robustness of the moving light display approach and to highlight the advantage of undertaking cluster based analysis. Further development of these algorithms can only increase the accuracy of tracking and cellular event interpretation. For example, cell definition in terms of marker clarity and resolution significantly degraded towards the edge of the field of view and the addition of pre-processing routines on the original raw fluorescence image to account for the lens effects and variability in background noise across the image would help in enhancing these regions as would the application of more sophisticated noise filtering. The critical limiting factor of the MLD technique is cell identification as cells become increasingly confluent and harder to resolve without knowledge of their respective boundaries. One possible solution to increase the ability to resolve individuals at high confluency is to apply a nuclear marker at the end of the experiment. Using the final nuclear stained cells as the starting point and running a time reversed analysis would provide a well defined initial seeding of centroids and makes identification of the bifurcation points within a lineage far simpler due to the centroids of the daughter cells converging to their progenitor thus removing the decision of when the binary elements groups are best described as dual rather than singular clusters.

To summarise, we have used quantum dot nano-particles as point like markers within a cell to demonstrate the concept of moving light display to resolve cellular mitosis events and lineages without the need for complex interpretation of a complete visual picture of the cellular field. In the study of cell movement and proliferation the knowledge we seek can be resolved at the whole cell level and so the reduction of the data set to a minimal set of Cartesian co-ordinates provides a much greater efficiency of information processing, distilling the experimental measurements down to match just that needed to understand the biology. In future applications this approach will allow real-time data processing during image acquisition and the direct storage of biological knowledge, i.e. cell position and familial relationships.

## REFERENCES

Ljosa V, Carpenter AE. (2009) Introduction to the Quantitative Analysis of Two-Dimensional Fluorescence Microscopy Images for Cell-Based Screening. PLoS Comp. Biol. 5: e1000603.

Lee S, Howell BJ. (2006) High-content screening: emerging hardware and software technologies. Methods Enzymol. 414:468–83.

Swedlow JR, Goldberg I, Brauner E, Sorger PK. (2003) Informatics and quantitative analysis in biological imaging. Science 300:100–102.

Lang P, Yeow K, Nichols A, Scheer A. (2006) Cellular imaging in drug discovery. Nat. Rev. Drug Discov. 5:343–356.

Nigg EA. (2002) Centrosome aberrations: cause or consequence of cancer progression? Nature Rev. Cancer 2:1-11.

Cheng G, Youssef BB, Markenscoff P, Zygourakis K. (2006) Cell Population Dynamics Modulate the Rates of Tissue Growth Processes. Biophys. Journal 90:713–724.

Singhvi A, Garriga G. (2009) Asymmetric divisions, aggresomes and apoptosis. Trends in Cell Biol. 19:1-7.
Morrison SJ, Kimble J. (2006) Asymmetric and symmetric stem-cell divisions in development and cancer. Nature 441:1068-1074.
Lia K, Chenb M, Kanadea T, Millera ED, Weissa LE, et al. (2008) Cell Population Tracking and Lineage Construction with Spatiotemporal Context. Med Image Anal. 12: 546–566.
Padfield D, Rittscher J, Thomas N, Roysam B. (2009) Spatio-temporal cell cycle phase analysis using level sets and fast marching methods. Med Image Anal. 13: 143–155.
Baatz M, Arini N, Schäpe A, Linssen B. (2006) Object-oriented image analysis for high content screening: detailed quantification of cells and sub cellular structures with the Cellenger software. Cytometry A 69: 652–658.
Boland MV, Markey MK, Murphy RF. (1998) Automated Recognition of Patterns Characteristic of Subcellular Structures in Fluorescence Microscopy Images. Cytometry 33:366–375.
Dufour A, Shinin V, Tajbakhsh S, Guillén-Aghion N, Olivo-Marin JC, et all. (2005) Segmenting and Tracking Fluorescent Cells in Dynamic 3-D Microscopy with Coupled Active Surfaces. IEEE Trans. Image Process. 14:1396-1410.
Stephens DJ, Allan VJ. (2003) Light microscopy techniques for live cell imaging. Science 300:82-86.
Wertheimer M. (1923) Laws of Organization in Perceptual Forms. Psychologische Forschung 4:301-350.
Johansson G. (1973) Visual perception of biological motion and a model for its analysis. Perception & Psychophysics 14:201-211.
Cedras C, Shah M. (1994) A Survey of Motion Analysis from Moving Light Displays. Proc. IEEE Conf. Computer Vision and Pattern Recognition 21-23.
Guerra-Filho GB. (2005) Optical Motion Capture: Theory and Implementation. Journal of Theoretical and Applied Informatics 12:61-89.
Hoffman DD, Flinchbaugh BE. (1982) The Interpretation of Biological Motion. Biol. Cybern. 42:195-204.
Brown MR, Summers HD, Rees P, Chappell SC, Silvestre OF, et al. (2010) Long-term time series analysis of quantum dot encoded cells by deconvolution of the autofluorescence signal. Cytometry A 77: 925-32.
Genovesio A, Liedl T, Emiliani V, Parak WJ, Coppey-Moisan M, et al. (2006) Multiple Particle Tracking in 3-D+t Microscopy: Method and Application to the Tracking of Endocytosed Quantum Dots. IEEE Trans. Image Process. 15:1062-1070.
Bacher CP, Reichenzeller M, Athale C, Herrmann H, Eils R. (2004) 4-D single particle tracking of synthetic and proteinaceous microspheres reveals preferential movement of nuclear particles along chromatin – poor tracks. BMC Cell Biol. 5:45.
Gao Y, Kilfoil ML. (2009) Accurate detection and complete tracking of large populations of features in three dimensions. Optics Express 17:4685-4704.
Summers HD, Rees P, Holton MD, Brown MR, Chappell SC, et al. (2011) Statistical analysis of nanoparticle dosing in a dynamic cellular system. Nature Nanotechnology 6:170-174.
Graves TG, Harr MW, Crawford EL, Willey JC. (2006) Stable low-level expression of $p21^{WAF1/CIP1}$ in A549 human bronchogenic carcinoma cell line-derived clones down-regulates E2F1 mRNA and restores cell proliferation control. Molecular Cancer 5:1.

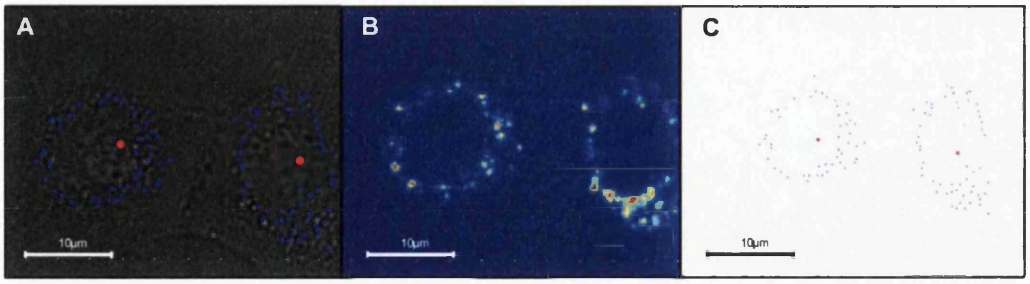



**Figure 1. Typical cell images.**

(A) Bright field image with binary element overlay; (B) fluorescence image of two neighboring A549 cells; (C) the light point cluster map derived from the fluorescent signal. The binary elements within the light point map are displayed as blue points and represent the locations of the centres of the quantum dot labeled vesicles; the centroids of the binary element clusters are displayed as red points.

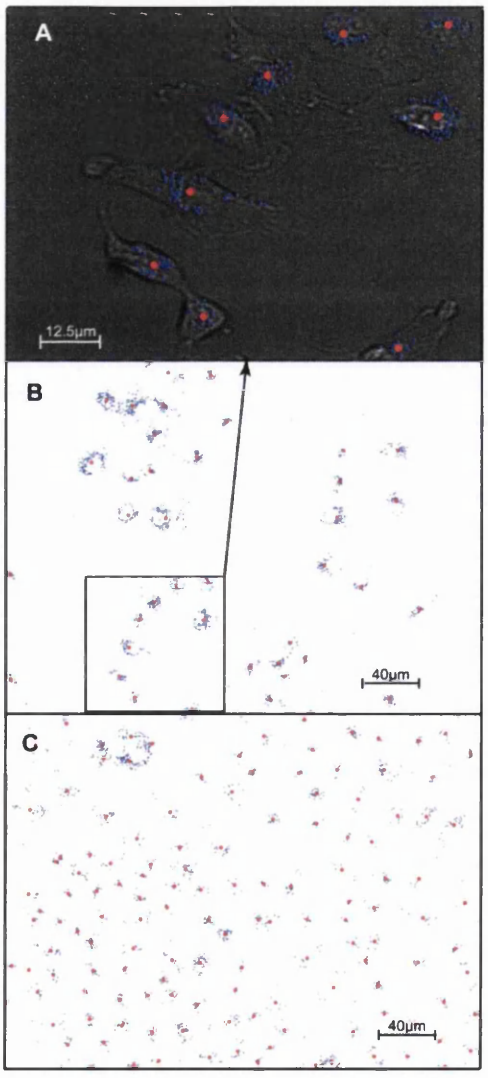



**Figure 2. Spatio-temporal cell tracking.**

(A) Bright field image of A549 cells overlaid with binary elements (blue) and assigned centroids (red); (B) full-field binary element representation of a 336×256 µm fluorescence image taken at the initial time point of experiment and (C) representation of the same field at the t = 40 hour time point.

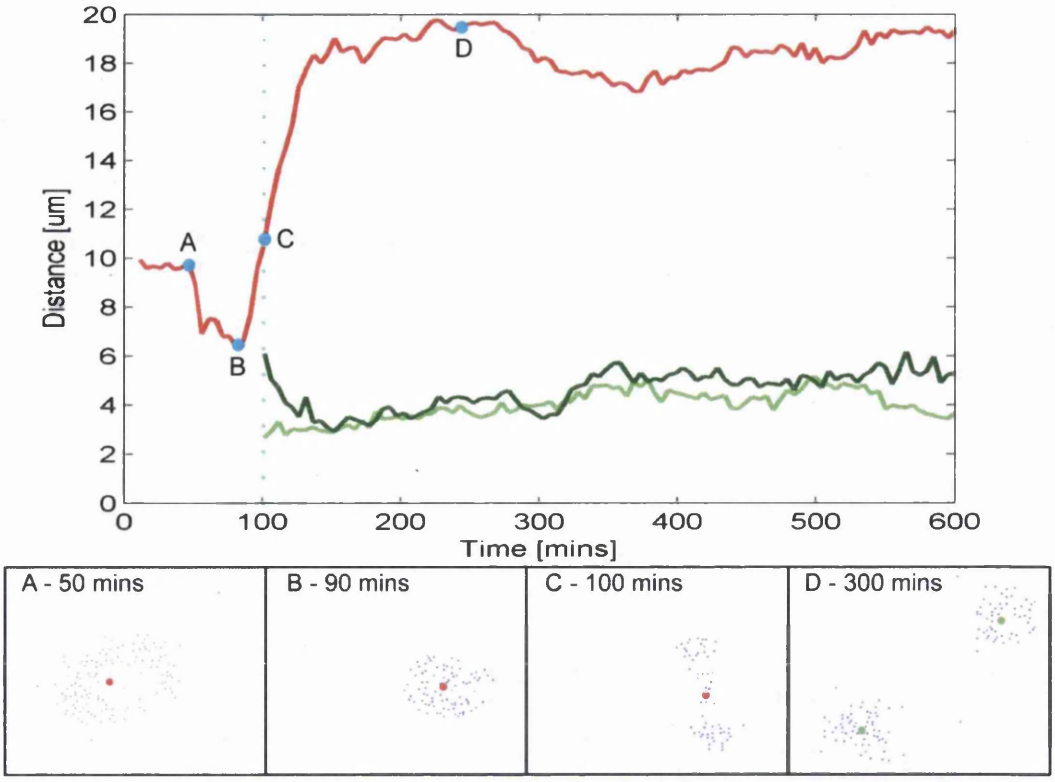



**Figure 3. Analysis of mitosis.**

(main figure) An example of the characteristic curve of mean binary element separation as an individual cell undergoes a mitotic event and divides into two daughters cells. (sub-panels) The corresponding stills below show the binary element distribution, in a fixed frame position, at four time points spanning key stages of the process: (A) Cell prior to extracellular signs of mitotic committal; (B) localized contraction of the light-point markers as the cell prepares to divide; (C) marker distribution indicating telophase stage; (D) two daughter cells identified as independent clusters (frame scale is 40×30 μm).

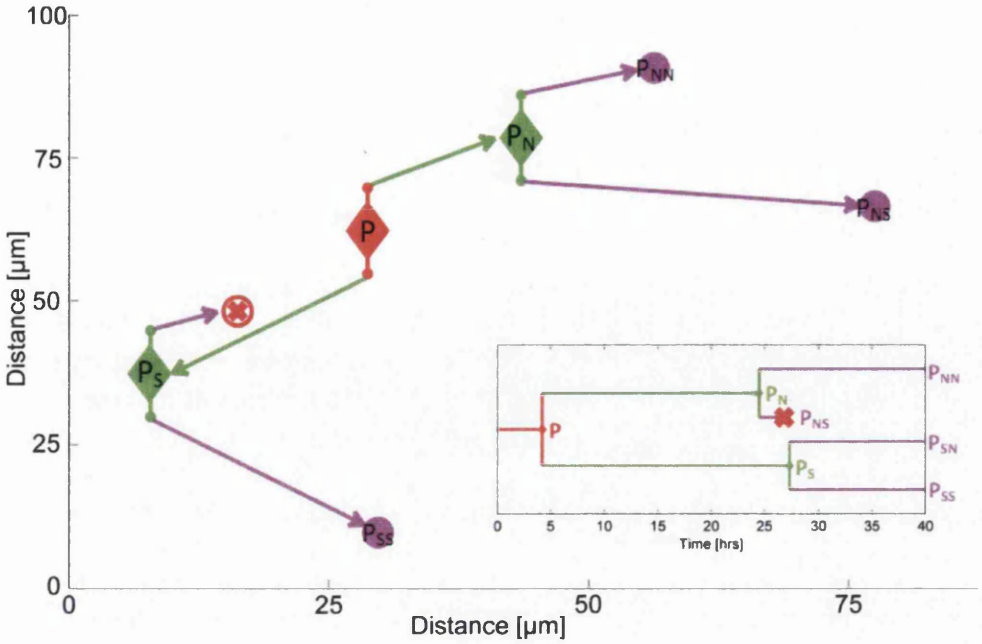



**Figure 4. Lineages.**

Graphical representation of a single cell's lineage evolution through space alongside conventional lineage portrait as defined by the automated centroid tracking. Diamonds mark the spatial and temporal location of mitosis events, red being the progenitor cell, green the second generation cells and magenta the location of the cell centroids in the final frame of the time-lapse sequence. The cross marks the loss of a cell behind an unidentified object in the image (non-cell). The arrows in the main diagram may be viewed as motility vectors reporting the mean velocity of a cell between mitotic events.

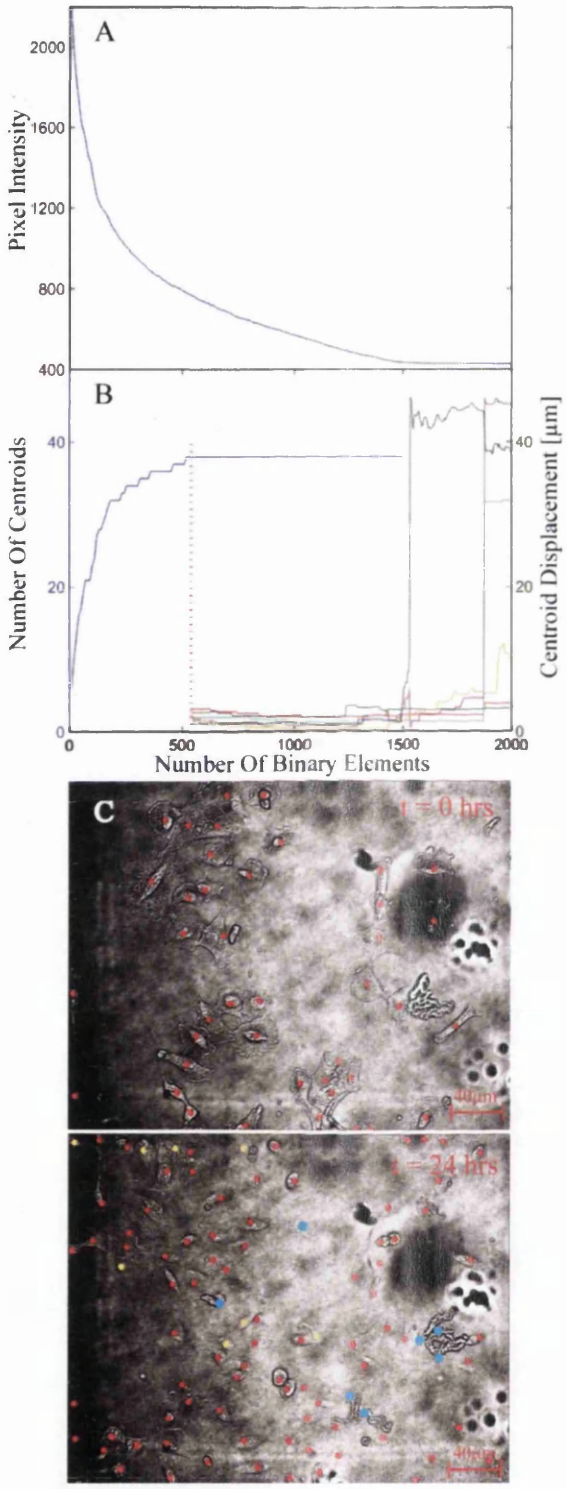



**Figure 5. Operational range.**

(A) Calibration curve of initial image showing the relation between the intensity threshold filter cut-off and the number of binary elements identified. (see Movie S4 for animation) (B) Plot showing the operational range of the system through the binary element dependence of the number of centroids (cells) identified (black curve), and the centroid displacement of a selection of cells (colored curves). The vertical, red dashed line indicates the minimum point number requirement for maximization of the number of cells automatically identified. (C) Bright field images at t = 0 and t = 24 hours with centroids overlaid. Red spots indicate automatically identified centroids, blue spots correspond to incorrectly assigned centroids and yellow spots represent manually identified cells which are unidentifiable by the automated analysis.