TRACKING AND ANALYSIS OF MOVEMENT AT DIFFERENT SCALES

FROM ENDOSOMES TO HUMANS

Dimitra Georgopoulou

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

2020, Swansea University

Summary

Movement is apparent across all spatio-temporal scales in biology and can have a significant effect on the survival of the individual. For this reason, it has been the object of study in a wide range of research fields, i.e. in molecular biology, pharmaceutics, medical research but also in behavioural biology and ecology. The aim of the thesis was to provide methodologies and insight on the movement patterns seen at different spatio-temporal scales in biology; the intra-cellular, the cellular and the organism level. At the intra-cellular level, current thesis studied the compartmental inheritance in Human Osteosarcoma (U2-OS) cells. The inheritance pattern of the endosomal quantum dot fluorescence across two consecutive generations was for first time empirically revealed. In addition, a in silico model was developed to predict the inheritance across multiple generations. At the cellular level, a semi-automated routine was developed that can realize long-term nuclei tracking in U2-OS cell populations labeled with a cell cycle marker in their cytoplasm. A method to extract cell cycle information without the need to explicitly segment the cells was proposed. The movement behaviour of the cellular population and their possible inter-individual differences was also studied. Lastly, at the organism level, the focus of the thesis was to study the emergence of coordination in unfamiliar free-swimming stickleback fish shoals. It was demonstrated that there exist two different phases, the uncoordinated and the coordinated. In addition, the significance of uncoordinated phase to the establishment of the group's social network was for first time evinced. The adaptation of the stickleback collectives was also studied over time, i.e. the effect of group's repeated interactions on the emergence of coordination. Findings at the intra-cellular and cellular level can have significant implications on medical and pharmaceutical research. Findings at the organism level can also contribute to the understanding of how social interactions are formed and maintained in animal collectives.

Declaration and Statements

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 .... 

Date .....30/09/2019.....

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated.
Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed ... 

Date .....30/09/2019.....

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 . 

Date .....30/09/2019.....

Acknowledgements

This thesis is the outcome of a four year effort accompanied by a lot of personal and professional challenges. I would, therefore, like to thank:

My supervisors, Rowan Brown, Andrew King and Ines Fürtbauer, for their understanding, continuing support and their feedback throughout the whole process of the Phd.

Professor Rachel Errington, Rachel Howard Jones and Marie Wiltshire from Cardiff University for providing their time and knowledge, without any suspicion, on cancer research and cell microscopy. Rachel's Errington support and guidance has been of great value for me and her professional attitude will be a good example to follow during my next steps in research. It was a pleasure to work with all of you.

Heather Barnett, for the collaboration and for introducing me to the whole new world of art science communication.

Danis Kiziridis, for contributing in all possible ways he could to the final outcome of this thesis.

My colleage, Mathieu Duteil, who was working on similar projects and could understand, practically help but also share my insecurities.

Claudio de la O and Bibiana Montoya for the nice conversations we had, their feedback and their precious emotional support. Thank you Marta, Suzana, Dania, Yuonqiang, Teja, Jessica, Laura, Irene Chrysovalanto, Kosta, Daniel and Hannah for being always around when needed.

My life-long friends Kalliopi, Rafaela, Lida, but also Marie, Oriol and Tata for cheering me up and be always available to take some of my load. I would also like to thank Odyssea, Anna and Vasili for their support during the writing period of the thesis.

My family, Athina, Konstantinos, Anastasis and my parents for being always here to support me and embrace all my efforts.

Contents

Chapter 1

Introduction

Movement is defined as the change in the spatial location of an individual in time (Nathan *et al.*, 2008) and can be active or passive, or a combination of both (Demšar *et al.*, 2015). Active movement, is apparent in all species of the animal kingdom. If not constantly, all organisms have to move at some time in their lifespan and this process is so fundamental for the functioning and survival of the individuals. Depending on the characteristics and timing of movement, it can affect each individual's fitness in a positive or negative way; by moving, individuals can locate food, avoid predators, react to an environmental disaster and even transfer information (Jones, 1977; Russell *et al.*, 2015, 2017). Movement can consequently have an effect on groups, populations and whole species. Therefore, seeking for proximate and ultimate explanations of the existance of animal movement is an active field of research (Tinbergen, 1963; Nathan *et al.*, 2008; Pyke, 2015).

Movement is apparent across different spatio-temporal scales in biology. Intra-cellular compartments and organelles are constantly moving, either actively or passively, and transfer important elements for the survival to the inner and the outer part of the cell. Intra-cellular movement and the successful distribution of the cellular material is also vital for the survival of the cells and the organisms (Trybus, 2013). At a higher spatio-temporal level, cells have developed a wide variety of mechanisms to actively translocate themselves (Petrie *et al.*, 2009) in order to find appropriate optimal environmental conditions, receive nutrients, avoid predators and also cooperate with neighbouring cells (Bodor *et al.*, 2020). At the organism level, multicellular organisms, and more specifically animals, would not be able to survive if they were not able to move at some stage in their lifespan. Foraging, habitat selection, reproduction and predation avoidance are all vital needs that require movement to be accomplished. A group of individuals, at any of the above scales, could show synchronization in movement resulting in collective movement (Méhes & Vicsek, 2014; Nagy *et al.*, 2013; Ostner *et al.*, 2013; Tunstrøm *et al.*, 2013; Herbert-Read, 2016). This can emerge from the local interactions between the individual entities and can serve to more effectively cope with the previously reported

challenges.

After decades of studying animal movement it is now appreciated that in many cases animal movement must be considered and quantified within a social context (Westley *et al.*, 2018). Even in non social groups, individuals are influenced by the movement decisions of others and often social cues are as important as environmental cues. Collective movement has long been the subject of studies in animal behaviour and questions of its ultimate explanations are under active investigation. They attempt to establish whether specific collective groups are favoured under different ecological conditions, or if specific collective movement types increase members' survival fitness (Westley *et al.*, 2018).

Complex systems science, can study patterns across multiple scales as well as the unpredictable relationship between the actions of individuals and the behaviour of the collective (Torney *et al.*, 2018a). In some cases, interactions at the individual level give rise to properties at the group level, e.g. a collective phenomenon that has not been predicted from observations of individuals in isolation (Parrish & Edelstein-Keshet, 1999a). For example, the spatial structure and the properties of grazing wave fronts of wildebeest far exceed the perception range of an individual. The structure is therefore an emergent phenomenon and the product of both the individual-level behaviour and the inter-individual interactions (Gueron & Levin, 1993). These concepts suggest that observations of individual movement only provide a partial picture of the global movement (Torney *et al.*, 2018a).

Regardless of whether the focus of study is the individual or the group, extracting individual movement paths (a.k.a movement trajectories) is the first step to study the individual or collective motion. The pattern of an individual's movement trajectory is determined by multiple factors. For example, it is determined by the intrinsic properties of the individual, the environmental conditions it experiences (such as climatic or landscape) and the social intercations it is involved in. Figure 1.1 summarizes the important factors (intrinsic, environmental and methodological) that contribute to the final extracted movement trajectory. In the following paragraphs I categorize them into five groups.

The first group refers to the intrinsic attributes of the focal individual, i.e. its internal characteristics (figure 1.1, Individual branch). The size of the individual, its morphological characteristics, its energy state, it's ability to move actively or passively or its genetic profile, are factors that can affect the individual's movement pattern. As an example, individuals of high energy states can explore larger areas and have higher speeds than individuals of low energy state (Krause *et al.*, 2000).

The second group of factors that can affect the movement pattern of an individual is related to the existence of inter-individual interactions (figure 1.1, Interactions branch). The number of individuals the focal individual is interacting with and the type of the

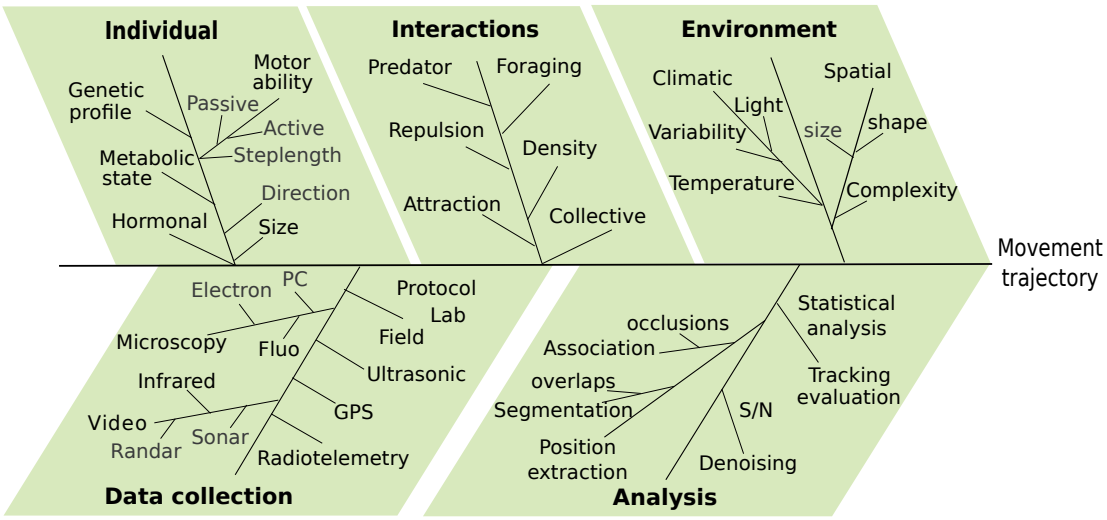

Figure 1.1: Cause and effect diagram (Ishikawa) for variability seen in the extracted movement trajectory.

interaction it forms (i.e. positive or negative) are factors that can shape its motion. Theoretical and experimental works (Czirók & Vicsek, 2000; Tunstrøm *et al.*, 2013) have showed how simple rules of attraction or repulsion can shape an individuals' movement trajectory.

The third group of factors that have a causal effect on the final trajectory referes to the environmental conditions that hold during the realization of motion (figure 1.1, Environment branch). These include the temperature, oxygen levels and illumination conditions, but also the resource availabity, the complexity of the environment and the medium the individual is travelling through (e.g. air or water).

Tthe extend to which a researcher can understand an individual's movement behaviour is constrained by the tools used for the aquisition of the data. For this reason, the fourth group of factors is related to the system that is used for the data collection (figure 1.1, Data collection branch). It includes the equipment and the scientific approach used (if it is a field work or an experiment in the laboratory) but also the methodological protocol that is followed for the data acquisition.

The different devices used for monitoring movement, but also the protocols followed can be error prone under different conditions. For this reason, the initially acquired data need to be filtered out and smoothed to extract the final trajectory. Therefore, the last group of factors is related to the methodology used for the extraction of the trajectories. This includes the algorithms used for extraction of the positions and the pre- and post-processing steps, i.e. the functions used for extracting the position, the denoising techniques that were chosen etc.

It is of great importance to study the extracted movement trajectories and the proximate and ultimate reasons that are responsible for their realization. The inter-individual

heterogeinity in the movement behaviour can serve as a pool where selective forces can act and affect the survival of the individual and subsequently that of the whole population.

The aim of the thesis was to provide methodologies and insight on acquiring and understanding movement data at different biological scales; from the intracellular movement of endosomes to the movement of cells, fish and humans. Movement trajectories, i.e. the collection of the positions an individual had for a time period, are extracted from various sources and are the subject of further analysis. For the extraction of the paths and other data, time-lapse and video recordings were acquired. Subsequently, automated algorithms were developed to extract data. For the analysis, statistical methods and in silico models were the tools to understand and answer questions on the subjects of study.

1.1 Intracellular motion

Movement of the intracellular material could be considered as passive due to the inability of the organelles to produce and consume their own energy. Despite that diffusion plays an important role in the distribution of intracellular material, it has been demonstrated that cells have developed mechanisms for the direct transport of materials. Material motion is most of the times the result of the energetic transportation of the organelles from the cytoskeletal system (actins, dyneins and myosins) of the cells. It can, thus, have different forms (diffusion or directed motion) that could be selected for the fitness benefits they provide to the cells. For this reason, the analysis of movement of intracellular material is necessary and can be studied using to some extent analogous analysis with that used in movement ecology (Trybus, 2013).

For almost two decades researchers have worked to understand and elucidate on intracellular dynamics, organelle motion and the distribution to daughter cells and their ultimate functions. For example, Bergeland *et al.* (2001) showed that endosomal segregation into daughter cells takes place by coordinated movements, and during cytokinesis, these organelles accumulate in the vicinity of the microtubule organization center. Dinh *et al.* (2006) approximated organelle movement as stochastic trajectories of independent discrete particles that transit from one transport state to another. They defined four possible movement/transport states; free diffusion in cytosol, kinesin-driven transport toward Microtubule (MT) plus-ends, dynein-driven transport toward MT minus-end and myosin-driven transport on actin filaments. They believed that each organelle can transition from one state to the other within a movement event.

Trybus (2013) observed long pauses of the early endosomes that dominated their movements and were interspersed by short bursts of directed motion. They further demonstrated different types of endosomal movement, diffusive and directed (motor-driven) and used mean square displacement analysis to recognize and distinguish the

different movement patterns. This analysis is commonly used in the field of movement ecology too (Demšar *et al.*, 2015; Reynolds & Ouellette, 2016). A different perspective was brought from Aoyama *et al.* (2017), where they showed that the intracellular trajectories and motilities of endosomes were influenced by the particle size of the cargo (in this case, nanoparticle) they were carrying.

During the last decade, a significant amount of research has focused on intracellular inheritance, i.e. its distribution to daughter cells after mitosis (Errington *et al.*, 2010; Summers *et al.*, 2011; Brown *et al.*, 2010). The understanding of the distribution of the intracellular material can have important medical and pharmaceutical implications. Initially, it was assumed that the cellular material was equally distributed in to daughter cells (Bergeland *et al.*, 2001). However, recent studies linked stem cell differentiation to the asymmetric inheritance of endosome function in daughter cells. Finally, it has been showed that the components partition asymmetrically at mitosis and, consequently, interest on its biological relevance and its ultimate functions has arisen (Rees *et al.*, 2011).

1.2 Cellular movement

Cells display a very sophisticated and complex locomotory activity. The different types of intracellular motility that were mentioned in the section 1.1 are also responsible for the movement at the cellular level. The cellular movement is the result of the coordinated movement that is taking place inside the cell. Movement at the cellular level is widespread in living organisms and plays a fundamental role in physiological phenomena including neural development, wound healing, and immune function, as well as in disorders such as neurological diseases, fibrosis, and cancer metastasis (Lackie, 1986; Ferguson *et al.*, 2017).

Cells can move either freely in unicellular organisms, or in accordance to the movement of other cells, in multicellular organisms, such as when leucocytes move during inflammation, epithelial sheets in wound healing, embryonic cells in morphogenesis and tumour cells in metastatic phase and malignant invasion. In most of the cases they require a substratum upon which they move, such as protein-coated coverslips (Lackie, 1986). The drivers of cell movements can also be external, for example the cells are adjusting their motion according to the gradients in the concentration of certain chemicals in their environment (a.k.a process of chemotaxis) (Majumdar *et al.*, 2014; Ferguson *et al.*, 2017).

Unlike molecules that move passively due to collisions, cells produce kinetic energy. They have finite speed due to the limitation on the motor functionality and thus, to move longer distances they must limit the number of turns they go through (Lackie, 1986). A lot of studies have characterized the movement patterns different cells show.

Depending on the cell type and the conditions, random walk, correlated random walk, persitent random walk, Levy walk were attributed to the patterns that cell movements show (Svensson *et al.*, 2018).

Initially, the most simplified movement type was attributed to the cell motion was the brownian walk. However, soon it was realized that most cells have directional persistence leading to the Persistent Random Walk (PRW) or Correlated Random Walk (CRW) model (Selmeczi *et al.*, 2008). A popular approach for modeling optimal search patterns of both animals and microbes is the scale free Levy walk. Levy walks are characterized by a stop-start pattern where the cells are more or less stationary for some time followed by a directed walk of length L (Svensson *et al.*, 2018).

1.2.1 Cellular movement in tumour and metastasis

Cell movement and consequently cell migration is a pivotal step in the metastatic process of cancer cell populations (Paul *et al.*, 2016). The cancer cells in a tumour are very heterogeneous and can differ in size and in the genomic content and thus can include bulk cells, stem cells and polyploid cells (Li *et al.*, 2015; Bayani *et al.*, 2003). In the majority of tumors, the bulk of cancer cells are aneuploid, i.e. they have fewer chromosomes than the normal diploid cells (that have two copies of their genes, a.k.a 2n). On the other hand, polyploid cells have a higher than the diploid cells chromosomal number. They are also bigger in size than the bulk of cancer cells and stem cells. Stem cells are also known to show higher motility than the other cells (Griesdoorn, 2014; Mirzayans *et al.*, 2018; Brown *et al.*, 2017; Abarrategi *et al.*, 2016).

Most solid tumours contain few polyploid giant cells. The proportion of these cells increases in response to genotoxic stress, that can also be caused by radiation, or chemotherapy. This was first reported by Puck *et al.* (1956) for the human HeLa cervical carcinoma cell line that was exposed to ionizing radiation. One of the characteristics of these cells is that they cease to proliferate. However, a proportion of these cells remains metabolically active post-treatment (Mirzayans *et al.*, 2017). They are increasingly associated with metastasis and resistance to therapeutic methods. During this state of dormacy they can undergo nuclear budding and produce stem cell-like progeny. Furthermore, they can also promote stemness of the neighbouring cells. Also, it has been demonstrated that these giant cells are more metastatic than their parental cells (Zhang *et al.*, 2015b). Therefore, the proliferation arrest and dormancy of cancer cells should be carefully considered as a therapy-resistance mechanism Mirzayans *et al.* (2018).

It is therefore very important to link stem cell-like progenies with specific movement patterns, and collect the motility footprints of these cells so that predictions and diagnosis of the tumour phase can be achieved. This thesis aims to see if there is any emergence of different movement pattern in Human Osteosarcoma (U2OS) cells that could predict

differentiating behaviour in cancer cells and subsequently the onset of a metastatic behaviour. Targeting cancer cell motility may also be an important therapeutic option (Paul *et al.*, 2016).

Microscopy images and especially time lapse image sequences contain information on the dynamics of cells, the distribution of subcellular components, and the activity of molecules that is inaccessible from other techniques (Danuser, 2011) and they contain all spatial information. Therefore, they are a very useful tool for the study of cellular movement behaviour.

1.3 Collective movement

Collective phenomena in nature result from the local interactions between individual constituents and can be of any form, chemical, physical, social or biological (Wood & Galton, 2009). They include systems consisting of units ranging from macromolecules through metallic rods and robots to groups of animals and humans. They are emergent, i.e. their characteristics differ from the characteristics of the individual constituents and cannot be predicted from the study at the individual level (Tummolini *et al.*, 2006; Bittner *et al.*, 2004). For this reason, they should be approached as a whole, as a multi-component entity.

Phase transitions, i.e. a process in which changes in the state of the collective due to the effect of external parameters on it, are commonly observed in collective systems. They are perceived as a significant modification of the qualitative behaviour of the whole system (Lee & Wurtz, 2019). The units simultaneously change their global behavior to a differentiating behavioural pattern. For example, a group of feeding pigeons randomly oriented on the ground will order themselves into an orderly flying flock when leaving the scene after a big disturbance (Vicsek & Zafeiris, 2012).

Collective phenomena can be observed in different contexts and fields. One example in the social context, is the emergent response that results from individual attitudes and choices that are aggregated in the process of discussions and deliberations (Galam & Moscovlclt, 1991). The emergence of polarization in a ferromagnetic material that results from local interactions, is a typical example in physics. Such phenomena can be seen in every aspect and scale in biological systems. The emergence of multicellularity from the local interactions of cells, the emergence of an intracellular structure are examples where the emergent entity (e.g. an intracellular organelle, an organ or organism, or population) functions differently from its constituent parts (e.g. molecules, cells or individuals). Animal collective movement is another class of collective phenomena that has attracted a lot the interest of scientific community due to its proximate and ultimate consequences (Westley *et al.*, 2018).

The emergence of coordination has an immediate effect on the survival of the indi-

viduals. They can often coordinate to reduce predation risk, to improve foraging success (Bazazi *et al.*, 2011) or to pool information about the direction of new feeding, breeding or nest sites, thereby improving migration efficiency (Codling *et al.*, 2008; Seeley & Buhrman, 1999; GrÜnbaum, 1998). Antagonistic interactions between individuals can also lead to the emergence of coordinated movement. For example, when Mormon crickets or juvenile desert locusts chase and avoid conspecifics in cannibalistic interactions (Simpson *et al.*, 2006; Bazazi *et al.*, 2008; Hansen *et al.*, 2011).

There is a huge number of examples from the living world for the rich patterns the collectives exhibit (Vicsek & Zafeiris, 2012). Flocks of hundreds of starlings can fly forming a uniformly moving group, but can produce turbulent, puzzling aerial displays when they are at their roosting sites. Schools of fish can move in a rather orderly fashion or change direction amazingly abruptly. Under the pressure from a nearby predator the same fish can swirl like a vehemently stirred fluid (Vicsek & Zafeiris, 2012).

Fish have attracted most of the attention in the study of coordinated motion, because of the variable complex patterns they display (i.e. Hemelrijk and Hildenbrandt, 2012; Herbert-Read et al., 2011; Katz, Tunstrøm, Ioannou, Huepe, and Couzin, 2011; Marras et al., 2015; Viscido, Parrish and Grünbaum, 2004). They can macroscopically exhibit three different phases (Tunstrøm *et al.*, 2013) and they can transition from one to another depending on the conditions. At the scale of the individual interactions, it has been theoretically and empirically shown that fish follow some simple interaction rules when they are part of the collective. They are attracted by their close neighbours if they have a higher distance than expected, they are repulsed by their neighbours if they approach closer than a threshold distance and they try to align with them (Couzin *et al.*, 2002; Czirók & Vicsek, 2000; Parr, 1927).

Studies have suggested that variation in individual traits (such as metabolic, personality, information knowledge, etc) also influences the characteristics of the macroscopic collective formation and their temporal dynamics (Bazazi *et al.*, 2011; Gelblum *et al.*, 2015; Jolles *et al.*, 2015; Lord *et al.*, 2016; Nakayama *et al.*, 2016; Watts *et al.*, 2016). More specifically, empirical studies have demonstrated that certain individuals can act as leaders due to inter-individual differences in morphology, state (Briard *et al.*, 2015; Krause *et al.*, 2000), experience (Eskridge & Schlupp, 2014), information (Andrieu *et al.*, 2016; Watts *et al.*, 2016), and/or personality (Jiang *et al.*, 2017; Johnstone & Manica, 2011; Krause *et al.*, 2000).

To have a possibility to deeper understand the proximate and ultimate causes leading to the collective movement in biology a wide range of expertises is needed, from collective animal behaviour researchers to physicists, computer scientists and engineers (Westley *et al.*, 2018).

1.4 Methods for data extraction and analysis

1.4.1 Data acquisition

Cells

Extracting cell migration paths is not an easy task and requires a lot of expertise and a lot of resources to be able to get useful data. Experiments usually try to reproduce the tissue environment, resulting to sophisticated 2D in vitro assays (Kramer *et al.*, 2013) and 3D approaches Doyle *et al.* (2013) which better represent the tissue environment. In 2D cell cultures, cells grow on flat plastic or glass dishes in which they adhere and propagate forming a monolayer. 3D cell cultures, on the other hand, display physiologically relevant phenotypes, such as cell growth and interactions with its surroundings in a multidimensional structure (Souza *et al.*, 2018). In vivo experiments are also useful but are difficult and costly. For this reason, mainly in vitro and ex vivo setups are used Masuzzo *et al.* (2016).

For in vitro experiments, the instrumentation used has been crucial for the progress in cell movement research. The advent of high-throuput and high-content imaging systems has improved massively the quality and quantity of cell movement data. Specifically, the use of long term live cell microscopy provides a powerful tool for the acquisition of useful data and for the characterization of cancer invasion and metastasis. Masuzzo *et al.* (2016).

In vitro live cell microscopy can be conducted using either phase-contrast or fluorescent microscopy. Phase contrast imaging gives the opportunity for long term imaging without any effect on the cell viability due to labeling and pre-treatment. However, images acquired from phase contrast imaging may contain artifacts (such as halos) and could discommode the detection and extraction of cell paths. They cannot provide any other information on the cell state, like the position at the cell cycle the expression of a specific molecule of interest etc.

In contrast, fluorescent microscopy can provide good quality of images, that can highlight specific important parts of the cells and can make the extraction of cell trajectories easier. In addition they can provide very useful information about the state of a cell at each time point. The potential to visualize, track and quantify different molecules and cells has led to the development of a vast collection of fluorophores and fluorescent proteins. Fluorescent dyes exhibit favourable optical properties such as brightness, photostability and narrow band width (Specht *et al.*, 2016). Nanoparticles and more specifically Quantum dots can also be used in fluorescent imaging. A thorough review and analysis of these probes can be found in the reviews of Specht *et al.* (2016); Horan, Paul Karl *et al.* (1990) and Resch-genger *et al.* (2008).

According to Masuzzo *et al.* (2016), typically a study on cell movement includes

live-cell microscopy with image-processing algorithms. Cell populations usually are maintained in culture media and they are prepared and transfered to multi-well plates. They can be labeled with some dye or fluorophores. Then, automated image acquisition is carried out with a digital camera microscope with a motorized xyz-stage an acclimatization chamber. Different imaging can be used, for example phase contrast or fluorescent. The experimental procedure involves time-lapse imaging for several hours to be able to retrieve spatio-temporal information. Image processing then summarizes the acquired image sequences into numerical features.

Animals

The study of animal movement requires different technologies to extract motility information and depend on the size and the mode of motion. Radiotelemetry is used to track land or aerial movement; radiotransmitters, developed since the late 1950s to track animal movements and determine their home ranges. They are based on electronic tags which emit a high frequency (VHF) radio signal that can be used to locate the position of an animal, without the need of being in close contact with the targeted individual (Börger, 2016). There are tags of many different sizes that allow even the successful monitoring of small-sized animals, like bees. Acoustic and ultrasonic telemetry on the other hand is more appropriate for marine animals.

Global Positioning System (GPS) is another widely used tool for animal and human tracking. This is a different telemetric system, in which the antennas are a network of satellites in orbit around the Earth. GPS is only usable if there are no obstacles between the animal and the sky, which excludes indoor use, or use inside a dense forest. One limitation of GPS is that communication with the satellites requires a lot of energy, which imposes a trade-off between the size of the battery, the frequency of readings and the duration of the tracking (Duteil, 2018). Another example that is also used for bird movement analysis is radar; The reader is refered to Krause *et al.* (2013); Demšar *et al.* (2015) for an in-depth comparison of the methods. Finally, there are non-visual methods, such as acoustic waveguide propagation Makris *et al.* (2006) for studying animal movement. This last is a promising tool to use for underwater movement studies.

Video recordings and the subsequent analysis of them is also a widespread way to track motion, that was mainly used, initially, in confined environments, such as artificial environments in the lab, i.e. fish in tanks and insects in appropriate boxes (Dyson *et al.*, 2015; Tunstrøm *et al.*, 2013; Herbert-Read *et al.*, 2011). Three dimensional imaging has recently been used to track animal behaviour within large indoor enclosures (Barnard *et al.*, 2016), but imaging of animals in natural landscapes is also a developing area of research (Robie *et al.*, 2017; Weinstein, 2018). However, with the technological advances and the opportunity to get aerial recordings, with unmanned aerial vehicles (UAVs),

bigger animals could be tracked in the field using video captures (Raoult *et al.*, 2018; Torney *et al.*, 2018b). These methodologies currently provide the most affordable and flexible imaging platforms for obtaining an aerial perspective in the field. UAVs provide the ability to adjust camera positioning on the fly and at distances up to several kilometres from the operator. This capability facilitates truly non-invasive filming of individual or collective animal behaviour when combined with computer vision techniques (Hughey *et al.*, 2018). Finally, biologists can also employ other imaging methods when the previous are not suitable (Dell *et al.*, 2014). These include near infrared, thermal infrared and sonar recordings.

Movement ecology was a field of study that had to base its conclusion more on assumptions, theories and observations rather than real data. The advancement in positional technology and the widespread use of global positioning monitoring has tremendiously eased the acquisition of movement data (Demšar *et al.*, 2015). As time progresses, tracking devices will be even more accessible for almost every case, because they are becoming cheaper, of variable sizes, with a variety of sensor characteristics and with stonger batteries. In addition, data access from these devices is becoming more efficient. Thus, movement ecology is transitioned from a data-poor scientific area into a data-rich discipline, allowing to find new answers to research questions in animal ecology (Demšar *et al.*, 2015).

Challenges someone needs to consider when aims to retrieve animal movement trajectories are the following (see also figure 1.1); First, the spatial scale of the motinoring is important. This is influenced by the size of the individual and the level of detail the study will focus. The temporal scale is also important. The time of the observation is very crucial, because dark and light conditions require different tools for optimized results. Apart from this, the speed of movement of the targeted animals can challenge the correct data acquisition, because it can lead to false sampling rates; The number of individuals to be studied is another aspect that needs to be considered; If many animals are to be studied, then appropriate resolution is required to be able to extract clear paths and interactions between them and the environment of study; Finally, the question of interest regarding animal movement can help to choose the right tool for monitoring movement. Some scientific questions regarding movement would focus on global behavioural patterns without asking details on finer spatial and temporal scales.

1.4.2 Data extraction and analysis

Cells

Once cell images are retrieved and restored, computer vision techniques are performed for cell tracking and quantitative motion estimation. These techniques are ad hoc procedures that depend on the system of study, i.e the motility mode and the microscopy

type that is used. While some cells migrate individually in a single-cell migration mode, cells can also retain cell-cell contacts and move as a single multicellular unit (Trepat *et al.*, 2009) in a collective cell migration mode (Masuzzo *et al.*, 2015).

Previously, manual tracking was used to extract the migration paths. It is still widely used and is also used for the acquisition of groundtruth data but can raise reliability concerns due to the mistakes that can happen because of the bias of the user. In addition, the manual tracking is time and resource consuming (since a person has to spend full time to be able to extract individual cell paths) (Masuzzo *et al.*, 2015).

Automated cell tracking systems that can provide objective and robust migration rates are therefore highly desired (Masuzzo *et al.*, 2015). Many new methodologies have been developed and presented in the last decade. Regarding tracking algorithms, they usually belong to one of the three categories; 'tracking by detection', 'tracking by model evolution', and 'tracking by filtering' (Nketia *et al.*, 2017).

In detection-based approaches, cells must first be separated from the background and from each other in a process termed segmentation. Segmentation refers to the division of an image into different segments that belong either to the objects of interest (groups of cells or single cells) or to the background. More precisely, image segmentation assigns a label to each pixel of an image such that pixels with the same label share particular characteristics (Masuzzo *et al.*, 2015; Nketia *et al.*, 2017). This can be achieved by using simple thresholding techniques, where one intensity value is chosen to be the threshold and the pixels over this value are labeled as foreground and those less are labeled as the background. Edge detection methods, make the use of gradient, horizontally and vertically, to find possible sudden changes in pixel intensities. The detected changes are perceived as the boundaries of the wanted objects (in this case cells).

The drawback of these methodologies is their inability to separate adjecent or overlapping cells. This limitation is overcome by the performance of a widely used method, the watershed algorithm. This methodology considers the image as a topographic relief in which the gray level of a pixel is interpreted as its altitude. 'Flooding' this landscape, starting from the local intensity minima separates the image into regions. This approach can lead to over-segmentation and, thus, it requires pre- and post- processing techniques. It is worth noting that many of the segmentation algorithms currently available rely on a priori knowledge of the typical diameter of the cells to be detected (Masuzzo *et al.*, 2015).

Once cell positions have been identified, the targeted cells need to be associated with the cells in the consecutive frames, and thus, their positions to be connected in order to retrieve the whole trajectory. One way to make the association is to connect each segmented cell with the one closest to it. This is working well if cells are moving with a low speed and are not densely packed. Other ways to associate cells between frames is to find the cells that are more similar with ones in the previous frame. The

similarity can be calculated using feature matching techniques. These methods locate similar cells using an extended list of features such as morphology, volume, surface, size and total curvature that expand the concept of distance beyond spatial location. Typically, these techniques require the user to specify the maximal distance that cells can travel between two consecutive timeframes. Furthermore, feature matching algorithms rely on good segmentation to keep the match as accurate as possible. Moreover, if a cell changes morphology between subsequent timeframes, it can be seen as two different cells, yielding broken trajectories (Masuzzo *et al.*, 2015).

Tracking by model evolution uses a deformable model to describe each tracked object, and segmentation and tracking are performed simultaneously by fitting this model to the image data (Masuzzo *et al.*, 2015; Nketia *et al.*, 2017). Deformable models are given different names (snakes, active contours or surfaces) and come in two types: parametric (Ray & Acton, 2002) and geometric (Li *et al.*, 2008). They involve a contour evolution approach to obtain the boundary of an object in the current frame by evolving the contour from the previous frame. The result of one frame is then used as an initial condition for the analysis of the next frame (Masuzzo *et al.*, 2015; Nketia *et al.*, 2017).

Particle-filtering techniques (also known as sequential Monte Carlo techniques (Liu, 2008)) are widely used in multiple-object tracking systems. The problem to solve is to estimate the state of a system given a set of observations. Particle-filtering methods try therefore to estimate an object's state posterior density function by a set of random particles with associated weights. The algorithm has three major steps, namely selection, prediction, and measurement, which are performed iteratively (Masuzzo *et al.*, 2015).

Once the algorithm is developed, it is very useful to test its performance to understand its strengths and its limitations. Ulman *et al.* (2017) introduced some measure to evaluate the efficiency of the algorithm. They include measures on the paths, the mitotic events and on the cell cycle information that is retrieved from the algorithm. For exanple, complete tracks (CT) is a measure that calculates the fraction of ground truth cell tracks that a given method is able to reconstruct in their entirety, from the frame they appear in to the frame they disappear from. Another measure is the track fractions (TF).This can be interpreted as the fraction of an average cell's trajectory that an algorithm reconstructs correctly once the cell has been detected. Branching correctness (BC) measures how efficient a method is at detecting division events. Finally, the cell cycle accuracy (CCA) measures how accurate an algorithm is at correctly reconstructing the length of cell cycles (that is, the time between two consecutive divisions).

Animals

Animal trajectories and related questions were previously analyzed using manual observations (Elliott *et al.*, 1977). But this way is effort intensive and often leads to poor

spatial and temporal resolution and it can be probe to human bias. Nowadays, advances in technology and image processing have enabled video recordings for experimental and field purposes. Once the videos are retrieved, animal movement trajectories can be extracted using different automated tracking methodologies. The number of individuals and the environmental conditions, i.e. illumination conditions, are important parameters for the appropriate choice of tracking routines applied (Dell *et al.*, 2014).

Automated image-based tracking is a difficult computer vision problem that raises some challenges. First, the captured image of an individual can be distorted or vary widely due to the movement speed of the animal and the complexity and variability of the environment the animal is living (Branson & Belongie, 2005), for example variable illumination conditions (figure 1.1). Second, the increase in the number of individuals studied can worsen the efficiency of the automated tracking, due to the individuals' physical contact, similarity and occlusions. In addition, the speed movement is also an important parameter and together with the time resolution of the camera used, it can produce some deformed objects in the image.

To extract movement trajectories two different subroutines must be applied. In the first, the individuals in the image have to be separated from the background and between each others, i.e they have to be segmented and labeled. In case the number of individuals is small, animals are equipped with individual tags, usually coloured (Webster & Laland, 2009), so that the position of each individual is detected using thresholding techniques (in the colour space) or template matching techniques (to detect the individual tags). If the number of individuals increases (equal or more than hundreds), for example in collective behavioural studies, then animals need to be identified and detected using a combination of thresholding techniques and pre-/post- processing analysis to evercome problems that arise from occlusions and overlapping. More sophisticated methods include artificial intelligence methods where each fish individual is detected using classification methods (Dell *et al.*, 2014).

The next step, after detection, is the association of individual animals between the consecutive frames. In case the individuals are tagged, then this is straightforward, since the coordinates of each individual are the central coordinates of the individual tag. However, constructing trajectories for multiple individuals is much more challenging. It often involves parameterization of a motion model which includes information from previous frames, such as the acceleration of each individual or their preferred direction of motion. In addition to this, methods discussed in previous section (section 1.4.2) can be used for the successful association of individuals across frames (Dell *et al.*, 2014).

Animal movement is linked to behavioural responses, and usually a specific movement can be associated with a specific behavioural response. Foraging, escaping predators, sitting in the nest, soaring in search of prey, all intuitively correspond to different movement patterns. First, the ever increasing availability of movement data provides

the opportunity to infer behaviour from movement types. Behaviour types are often extracted from trajectories with various forms of statistical modelling, including state-space models, various types of random walk models and behavioural change point analysis (Demšar *et al.*, 2015). Network analysis is also used to incorporate interaction between the different moving individuals (Demšar *et al.*, 2015).

Chapter 2

Quantum dot fluorescence inheritance in U2OS cells

The aim of the current study was to experimentally reveal the inheritance of the Quantum Dot (QD) endosomal fluorescence in Human Osteosarcoma (U2OS) cell populations, by using fluorescent time-lapse imaging techniques. In addition, the aim was to develop a in silico model to predict the inheritance across multiple generations.

2.1 Introduction

2.1.1 Quantum dot nanoparticles and their applications

According to the European Commission, nanomaterials are natural, incidental or engineered structures made of all kinds of solid materials and at least one of their dimensions is in the nanoscale. At this scale, the surface-to-volume ratios of materials become large and their electronic energy states become discrete, leading to unique electronic, optical, magnetic, and mechanical properties of the nanomaterials (Biju *et al.*, 2008); In addition, they have size-tunable optical properties, which renders them a very useful tool for physicists, chemists, biologists, medical researchers and technologists.

One of the most intriguing features of QDs is that the particle size determines many of the QD properties, most importantly the wavelength of fluorescence emission. By altering the QD size and its chemical composition, fluorescence emission may be tuned from the near ultraviolet, throughout the visible and into the near-infrared (NIR) spectrum, spanning a broad wavelength range of 400-2000 nm (Smith *et al.*, 2004). In comparison with organic dyes and fluorescent proteins, QDs are emerging as a new class of fluorescent labels with improved brightness, resistance against photobleaching (Smith *et al.*, 2004). QDs have been shown to remain brightly emissive after long periods of excitation, whereas organic dyes are photobleached quickly (Wu *et al.*, 2003). For these reasons, QD provide the possibility of continuous, real-time imaging of single molecules

and single cells over an extended period of time.

More specifically, Quantum dots (QDs) that are mainly used in research (in the current as well), are nanometer-sized semiconductors, clusters of 10 - 20 nm, comprising a core, shell and surface coating. The core is made up of a few hundred to a few thousand atoms of a semiconductor material (cadmium selenide (CdSe) or cadmium telluride (CdTe)). A semiconductor shell (typically zinc sulfide (ZnS)) surrounds and stabilizes the core, improving both the optical and physical properties of the material.

The core–shell assembly is extremely hydrophobic (Peng *et al.*, 1997). An amphiphilic polymer coating is applied to first confer water solubility essential for bioanalytical applications and secondly to provide a platform for covalent functionalization of the DQ nanocrystal with antibodies, oligonucleotides and other affinity reagents that confer targeting specificity for biomolecular detection. Due to their core/shell structure, they absorb more photons than the others due to their higher optical densities at longer wavelengths. This attribute results in an enhanced photostability of the core/shell nanocrystals as compared to other fluoorophores (Peng *et al.*, 1997).

QDs can have several applications in different areas. For example, current research in food industry investigates the use of nanoparticles for food packaging (Bajpai *et al.*, 2018) and their use for food-preservation (Kaphle *et al.*, 2018). In agriculture, they can be used as delivery systems for agrochemicals like fertilizers/pesticides and for pathogen detection and plant-disease management (Kaphle *et al.*, 2018). Nanoparticles have also been used for oil extraction due to the rheological properties they exhibit (Krishnamoorti, 2006). In biological research they can be a valuable tool and have been used as biomarkers and to target specific structures and molecules (Biju *et al.*, 2008). In medicine and pharmacology they have great potential to be used for more efficient drug delivery and disease detection and control (Zhang *et al.*, 2017a; Aranda *et al.*, 2018; Gurunathan *et al.*, 2018).

In biological research, nanoparticles are increasingly used for various purposes. QDs can be utilised to target specific molecules, genes and organelles and for cell and tissue engineering (Renukaiah *et al.*, 2018). They can also be used as antimicrobial agents, disrupting different vital microbial functions. As shown above, their fluorescent stability (figure 2.1), their low toxicity and size are ideal attributes required to utilise them as powerful biomarkers (Biju *et al.*, 2008). Because of their properties they can serve as a very useful tool for studies on the intracellular (molecular interactions) or cellular (cell-to-protein or cell-to-cell interactions) scale. Therefore, they can be used in imaging experiments.

In bioimaging, QDs can be used for molecule targeting and for in vivo imaging, tumor imaging, vasculature imaging, lymph node imaging and cell and tissue imaging (Yao *et al.*, 2018). In addition, QDs can be used to trace the cell lineage in embryogenesis (Dubertret *et al.*, 2002). They also have some important advantages compared to other

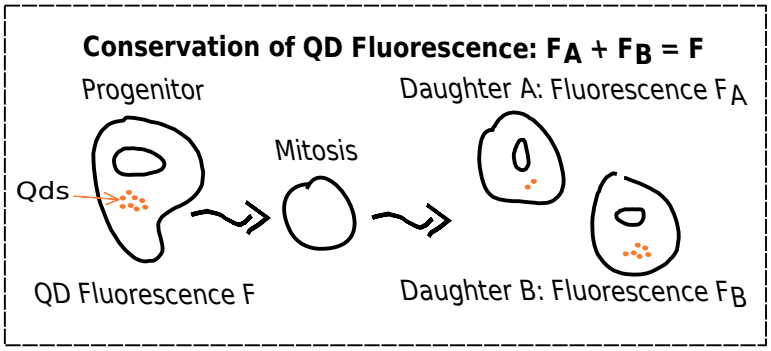

Figure 2.1: Schematic representation of the QD fluorescence redistribution in dividing cells. Due to the property of photostability of QDs the sum of the fluorescence of the two daughter cells should equal to the fluorescence of the parental cell, i.e. the signal should be conserved.

fluorescent dyes and proteins, for example, they represent broad luminescence excitation spectra and narrow symmetrical emission spectra with large Stokes shifts (Bilan *et al.*, 2016). Furthermore, QDs with emission spectra in the near-infrared (NIR) region are especially promising for deep tissue imaging both in vitro and in vivo (Aswathy *et al.*, 2010).

QDs have several applications in medical research and pharmaceutics, such as, in imaging (Jaiswal *et al.*, 2003; Chu *et al.*, 2006), drug delivery (Lin *et al.*, 2011; Adeli *et al.*, 2011), pharmaceutical analysis (Wang *et al.*, 2008), photothermal therapy (Zhang *et al.*, 2017b), biochips (Cui *et al.*, 2007), and targeted surgery (Yao *et al.*, 2018; Mchugh *et al.*, 2018). They are used in pharmacology for drug screening, drug delivery, drug target identification and drug analysis (Xu *et al.*, 2003; Han *et al.*, 2010; Yao *et al.*, 2018). QDs have shown great potential in drug discovery and diagnosis due to their photophysical properties. QDs can be cross-linked to biomolecules such as peptides, antibodies, or small-molecule ligands to make them target specific biological sites or systems and therefore be used for diagnosis and targeted drug delivery. Some pivotal parts are involved in such novel architectures that are fluorescent superparamagnetic NPs, tumor-specific antibodies, and anticancer drugs, and they are used for multimodal imaging and hyperthermia, cell targeting, and local therapy purposes, respectively (Zhang & Feng, 2006).

In cancer research, there is continuing emergence of innovative approaches that use nanotechnology to enable not only the detection and diagnosis of cancer at its earliest stages but also the delivery of anticancer drugs directly to the malignant cells (Gurunathan *et al.*, 2018). The nanomaterials can be conjugated with different kinds of ligands (e.g., proteins, antibodies, small molecules), producing the so-called “actively-targeted material” that favors drug-targeting to specific cell-surfaces and thus to specific cell-populations, leading to a selective reduced toxicity (Jia *et al.*, 2016). Other nano-materials can be multifunctional materials, such as “theranostics”, which allow the

codelivery of a therapeutic and a diagnostic agent in the same nanostructure (Biju *et al.*, 2010; Jia *et al.*, 2016). These nanostructures may improve several technological issues presented by marketed anticancer formulations such as solubility, pharmacokinetic profiles, cellular uptake, biodistribution patterns. Finally, the characterization of the QD load and dynamics per cell, in a proliferative population, can provide valuable information pertaining to cytoplasmic material, pharmacodynamics and efficacy of a delivered drug dose, such as, anti-cancer therapeutics Errington *et al.* (2010); Brown *et al.* (2010).

2.1.2 Human Osteosarcoma Cells as model system

Osteosarcoma is the most widespread bone cancer in children and young adults (Abarrategi *et al.*, 2016). Many different osteosarcoma cell lines have been derived through the years, such as the MG63, U2OS and SAOS-2 (Laitinen *et al.*, 1997). One of the first generated cell lines, the human osteosarcoma U2OS, was established by Ponten & Saksela (1967). The original cells were taken from a differentiated sarcoma of the tibia of a 15-year-old Caucasian female. Chromosomal instability, structural rearrangements and alterations and high incidence of aneuploidy characterizes these cells, as karyotype and cytogenetic analysis has revealed (Bayani *et al.*, 2003). Because of their properties the U2OS cells are widely used in biomedical research, for example see Raile *et al.* (1994); Zhao *et al.* (2015); Chou *et al.* (2018). This cell line is widely used as a model in cancer research for understanding the biology of cancer but also drug delivery in pharmaceutics.

In order to follow their progression through the cell cycle, the U2OS cells can be tagged with a green fluorescent protein (GFP), a stealth reporter (Thomas & Goodyer, 2003) whose expression is driven by the promoter of the cyclin B1 protein and allows for continuous cell cycle readout. This cell cycle marker was developed by Amersham Biosciences. U2OS cells were transfected with the reporter and they were maintained via antibiotic selection. It is commercially available as the 'G2M Cell Cycle Phase Marker' product. Cells labeled with this reporter gradually increase their GFP fluorescence during G1 phase and reach their maximal intensity in G2 phase, just prior to mitosis (figure 2.2).

The U2OS cells can also be loaded with QDs and together with the GFP marker they can be used to study, in high detail, the fluorescent inheritance properties and dynamics of the cells (current study).

2.1.3 Previous studies

In previous work, Errington *et al.* (2010) and Brown *et al.* (2010) investigated the proliferative features associated with QD fluorescence inheritance in human osteosarcoma cancer cell (U-2OS; ATCC HTB-96) populations. In these studies, contiguous long-term flow cytometry (1-8 days) measurements were utilized in conjunction with a stochastic

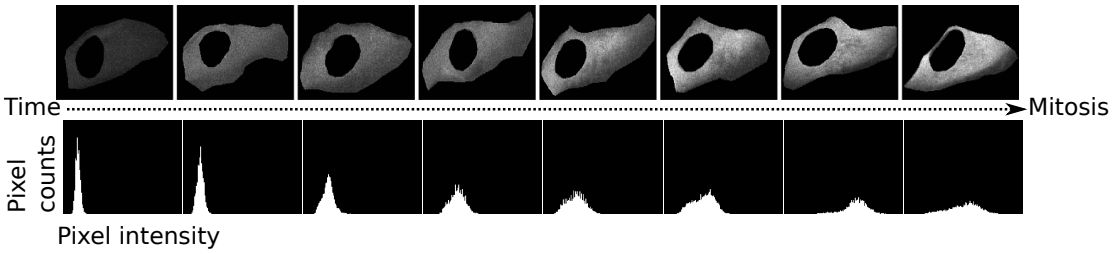

Figure 2.2: Progression of GFP fluorescence intensity along the U2OS cell cycle. Upper row shows the increase of the GFP fluorescent intensity values as cells approach to mitosis. Second row shows the histogam of the intensity values of the GFP signal. The histogram immediately after mitosis is very narrow and peaked around the very low intensity values. It progressively flattens and is translocated towards higher values.

cell cycle model to resolve global proliferation rates and quantify QD fluorescence dilution in control and pharmacologically perturbed systems. More specifically, the U2OS cells were loaded with fluorescent QDs in their cytoplasm, via the endocytotic route (Delehanty *et al.*, 2009), and a sample of cells was acquired every 24 hours and the quantum dot fluorescence for each cell of the sample was calculated by the use of flow cytometry. U2OS cells divide every approximately 22 hours (Ponten & Saksela, 1967), and thus 24 hours-interval sampling provided the QD fluorescence distribution for subsequent generations.

The experiments showed that the QD fluorescence was not redistributed equally (50:50 ratio) to the daughters but instead there was an asymmetric redistribution of Quantum Dot fluorescence. The numerical simulations predicted that the parental QD fluorescence is partitioned following an 80:20 split ratio to the daughters. This finding ultimately suggested that there must be a mechanistic explanation for this apparent preference towards a specific ratio.

However, other studies at the same time (Summers *et al.*, 2011) revealed that there is a random uptake and amalgamation of QDs within the U2OS cellular system. This finding cannot provide a biological rationale for the perceived asymmetry of the QD fluorescence partitioning ratio. Despite the fact that flow cytometry data could give an insight on the dominant generations at each sample, they could not provide clear detailed parent to daughter genealogical information to extract the actual QD fluorescence partitioning ratio.

2.2 Current study

Understanding endosomal QD fluorescence inheritance is a crucial requirement for all nanotherapeutics. This includes the ability to predict the effect of an applied nanomaterial dose and to further quantify how this evolves on a cellular level over time (i.e. how the nanomaterial is diluted and which is the threshold generation after which the nanomaterial does not have any effect on the population). Here, I investigate the fluorescence inheritance pattern of QD loaded endosomes in cells, using image cytometry and I compare the results with the respective experiments previously done in flow cytometry. More specifically, I investigated how the Quantum Dot fluorescence per cell is inherited across U2OS cell generations. To achieve this, time lapse high content microscopy was used to measure the QD fluorescence of cells and an image processing routine was developed to identify genealogical relationships between cells. Next, the QD fluorescence partitioning ratio between parent and daughters and between daughters and granddaughters was quantified (i.e for 2 generations). Finally, using the empirically driven information, a in silico model was developed. The model was built to reproduce the results of both, the microscopy and the flow cytometry experiment. It can also be used to predict the QD fluorescence dilution for more than two generations.

2.3 Methods

2.3.1 Cell culture preparation

U2OS (ATCC HTB-96) cells were maintained in 10 % McCoy's 5a full medium and cells were loaded for 1 hour with commercially available targeted quantum dots (QDs) using the Qtracker 705 Cell Labeling Kit (4 nM). Using Qtracker® Cell Labeling Kits, you can observe labeled cells using extensive continuous illumination, without the photobleaching and degradation problems often associated with organic dyes. Qtracker® labels are distributed in vesicles in the cytoplasm, and are inherited by daughter cells for at least six generations. Fluorescence from the Qtracker® labels can be seen up to a week after delivery in some cell lines. Long-term cellular retention makes Qtracker® Cell Labeling Kits ideal for studying cell motility, migration, differentiation, morphology, and many other cellular function studies. Qtracker® labels do not leak out of intact cells to be taken up by adjacent cells in the population. During the loading process the QDs were attached to the surface and were internalized into the cells via the endocytotic route. After one hour the cells were trypsinized and transfered to a 24-well plate (see figure 2.3). After 24 hours fluorescent time-lapse images, using IN Cell Analyser 2000 (GE Healthcare), were acquired with a 30-minute interval between frames for a period of 67 hours. Two fluorescent channels were used (one for the GFP stealth reporter and one for

the QDs). The experiment was designed and run by Rachel Errington's group at Cardiff University.

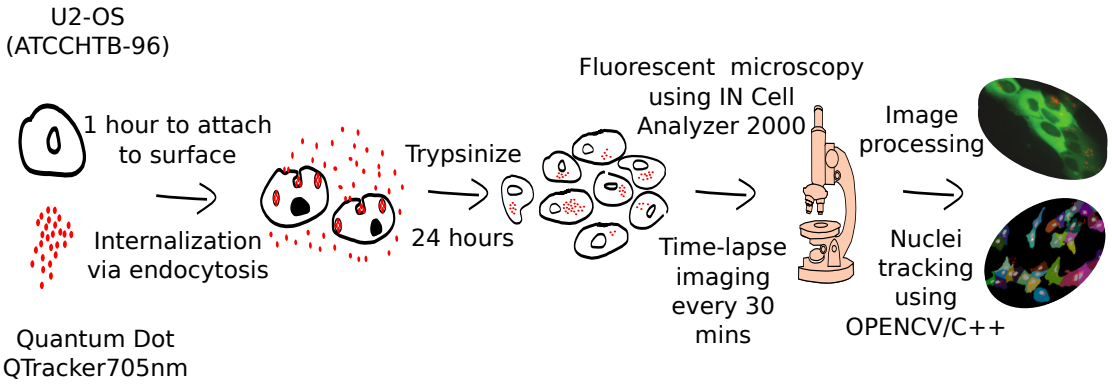

Figure 2.3: Schematic representation of the experimental protocol followed. Initially, U2OS cells are loaded with QDs and they are transferred to well plates. After 24 hours cells are transferred to the microscope and time-lapse images are acquired using two channels, the GFP and QD channel. Finally, the images are processed using Computer Vision techniques to extract the redistribution of QDs and the lineages.

2.3.2 Data extraction and analysis

Time-lapse images consisted of two-channels, the GFP-channel, showing the cytoplasm outline, and the QD-channel, showing the QD fluorescence in the cell population.

First step for the data extraction and analysis was to eliminate existing noise. Figure 2.4 on the left shows the histograms of a background region before the denoising. The histogram indicates that there is Gaussian noise ($\mu = 34$, $\sigma = 4.190$ intensity). The nonlocal denoising approach is used to get rid of the gaussian noise (Buades *et al.*, 2005). The parameter σ was calculated from the intensity histogram of background region, assuming that any variability in the background region is due to noise, since the growing medium is well mixed.

Second step was to check if the conservation of the fluorescent signal during our experiment holds. As previously stated in section 2.1 the QDs are characterized by photostability and thus, the QD signal should be preserved across time (see figure 2.1). However, because the cell images that were acquired from microscopy provide a 2-D representation of a 3-D system, some loss of information is expected, when 2D representations are acquired. Also, the Signal to Noise Ratio (SNR) of the GFP channel is low for cells being in the G1 phase, immediately after mitosis. Under-/over-estimation of the actual QD signal can also be caused from the inaccurate detection of the cell outline. Comparing the calculated fluorescence of the parental cells with the sum of the fluorescence of the daughter cells would show if the conservation of signal holds in our

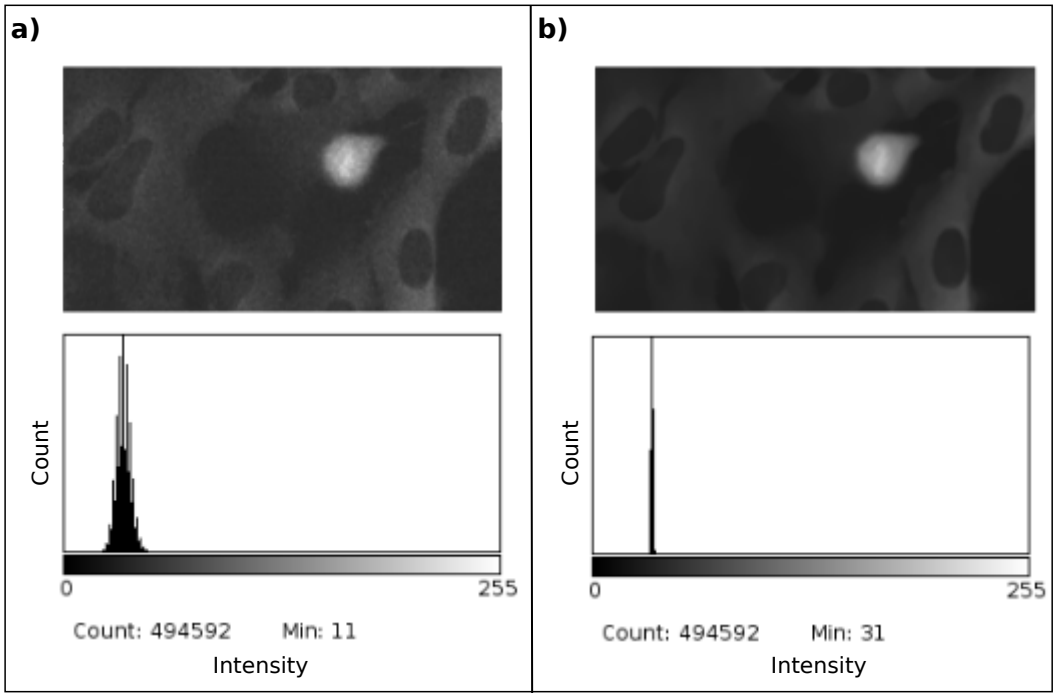

Figure 2.4: Histogram of the intensity values of the background region before (left) and after (right) denoising. The noise is gaussian with $\mu = 34$ and $\sigma = 4.190$.

data.

Figure 2.1 shows the conservation of the signal for different image segmentation techniques (i.e. QD thresholding techniques) applied on the QD image channel. The signal is preserved for all different techniques between parental and generation 1 cells. But is best preserved using Otsu's threshold for the generation 1 to generation two. This is possible because the QDs after 2 generations have been diluted significantly and the noise is more prevalent. Thus, Otsu's threshold can eliminate noise and clear signal in contrast with the other to methods. Another reason why the preservation of QD signal does not strongly hold for generation 2 is that the density of the cells has increased exponentially, rendering the cell outline segmentation process difficult.

Fluorescence inheritance extraction - algorithm

From the images, 500 cell lineages were extracted and analyzed. To extract the data all the images were processed using a bespoke algorithm built in OPENCV/C++ (Itseez, 2015). The algorithm was developed to manually extract lineage information, i.e. the user had to follow the cell from frame to frame until division was realized.

The routine for the data extraction was the following (see also figures 2.6 and 2.7): First, the mouse was activated for the user. The user had to draw the cytoplasm outline of the targeted cell on the GFP channel. Then, the image was denoised using the non-local means denoising technique (Buades *et al.*, 2005) 2.4, and the binary mask of the cell was extracted from the outline, using contour detection algorithm. Once the

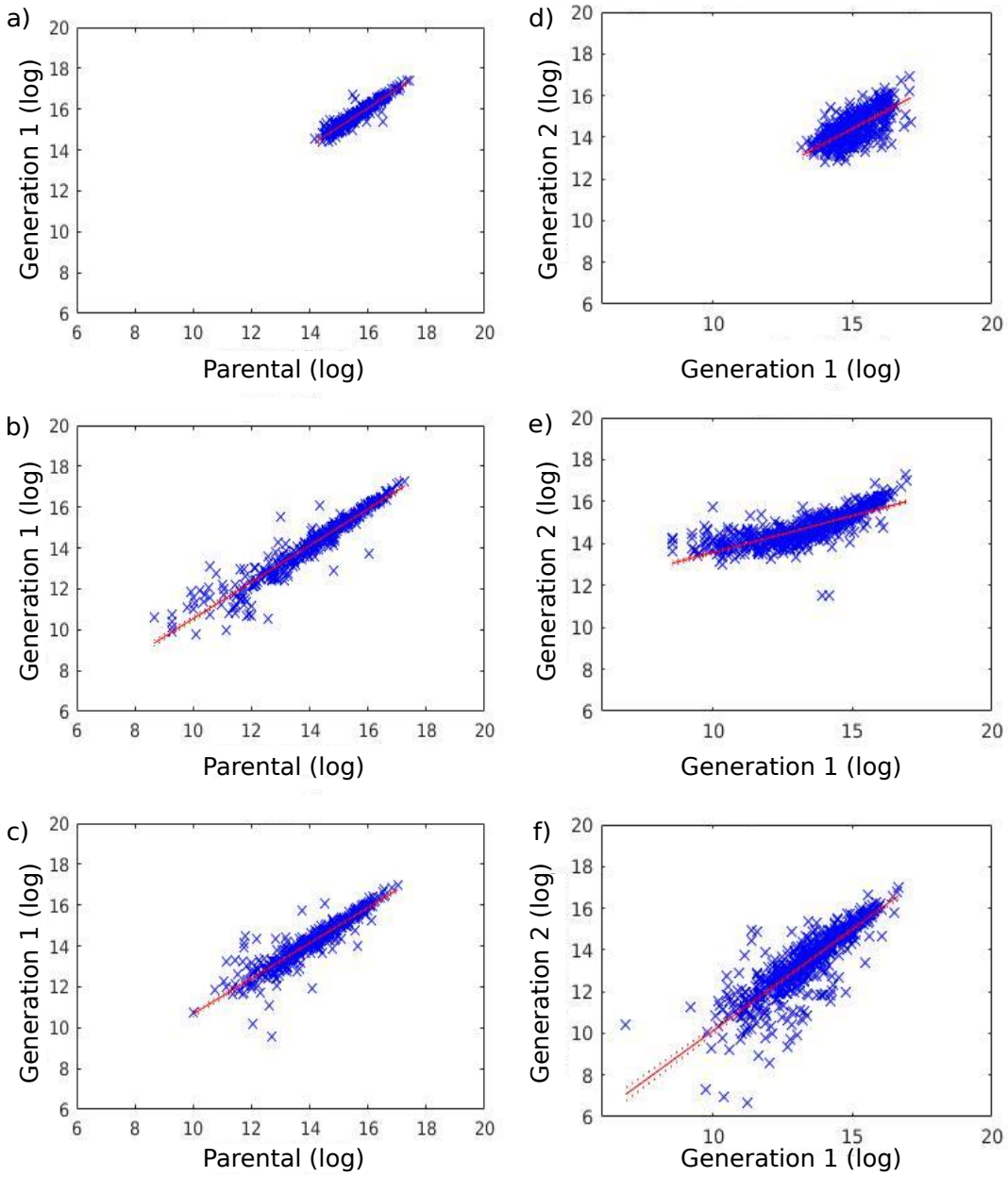

Figure 2.5: The conservation of signal using three different image thresholding techniques: no threshold applied into quantum dot channel and the fluorescence is just the sum of the values of the pixels of the denoised image (a, d), adaptive threshold (b, e) is used and the fluorescence is the sum of the white pixels of the binary QD image and Otsu's (c, f) threshold is used and the fluorescence is the sum of the white pixels of the binary QD image. The scatter plots show the sum of the fluorescence intensity of the daughter cells against that of the parental, for generation 1 (a,b,c) and 2 (d,e,f).

mask was retrieved, intersection (which is a binary operation) between the mask and the QD channel image was applied to extract the area in the QD channel that will be used for the quantification of the QD fluorescence of the targeted cell. A Region of interest (ROI) to include the area of interest in the QD channel and was thresholded using Otsu's threshold to create a binary image with the quantum dots as the foreground. The Otsu's threshold best works in images where distribution of the intensity values is bimodal, like in this case where there are very dark and very bright values and not so many in between. The routine then calculated the optimum threshold intensity value that is located between the two peaks of the distribution. Using the ROI instead of the whole image the white-to-dark ratio increases and the algorithm segments the QDs efficiently. The total QD fluoresence of the cell was the sum of all the white pixels in the segmented QD image. As a final step, the position of the cell, its intermitotic time (i.e. the time interval between two consecutive mitotic events), the total GFP of the cell and the QD fluorescence values were stored into a file. In addition to the above parameters, the cell ID, its parental ID and its generation ID (i.e. if it is a parental cell, a cell of generation 1 or a cell of generation 2) was stored as well.

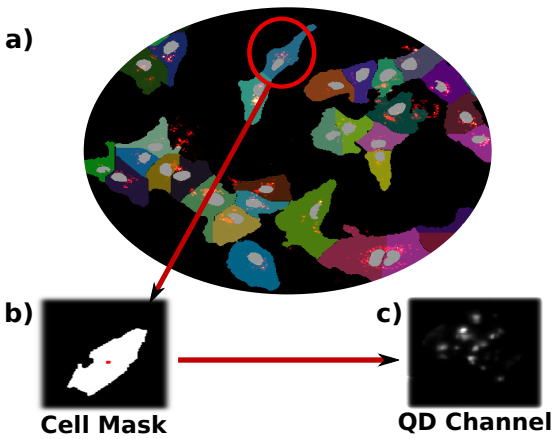

Figure 2.6: Schematic representation of the steps followed for extraction of Quantum dot fluorescence after processing the two images. a)Top image shows the segmented cells. b) Cell mask that is extracted from the channel with the cells. c) The QD fluorescence that has been extracted from the cell mask.

The extraction of the inter-mitotic times (IMT) distribution of the U2OS cells was useful for two reasons: first, because I could check for unexpected cell behaviour during the experiment that would deviate the inter-mitotic times far from the expected (22 *hours*). In addition, by extracting the actual inter-mitotic times and by fitting a theoretical distribution on them, I could validate the in silico simulations of the microscopy experiment and randomly extract relevant IMT, and increase my model's predictive power. The IMT of each cell was defined as the time interval between the two consecutive mitotic events.

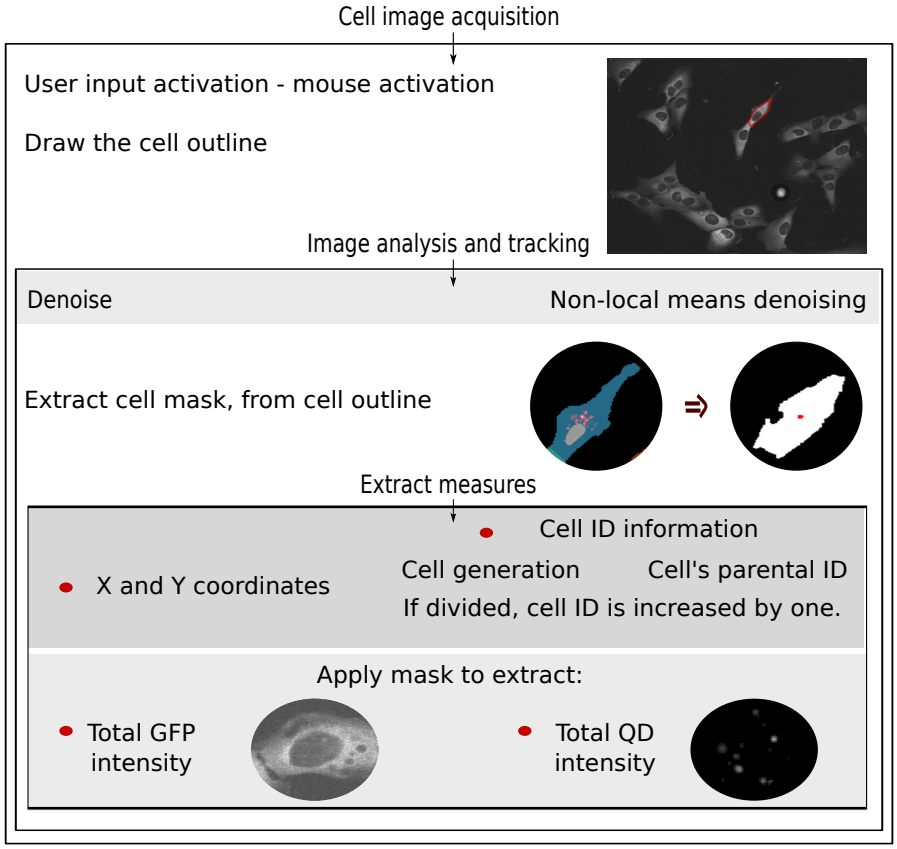

Figure 2.7: Manual lineage extraction software workflow. After acquiring the GFP channel image, the user draws the outline of the targeted cell as indicated by the image on the top-right. Consequently the image is denoised using non-local means denoising technique to get rid of gaussian noise. The next step is to extract the mask of the cell outline and apply it to count the total GFP intensity of the cells from the GFP channel and the total QD intensity of the cell from the QD channel. Finally, all important information is extracted and is stored to a file, including the x and y coordinates of the centre of the cell, its genealogical information (i.e. cell id, its generation and the id of its progeny), intermitotic times and the calculated intensities from the GFP and QD channels.

To acquire the experimental QD fluorescence distributions of each generation I categorized each cell according to the generation it belongs to and according to its genealogical relations, i.e. its parents and daughters. Then, I extracted its QD fluorescence intensity, and categorized the daughters into high and low QD fluorescence loaded.

In silico model

A stochastic model of the redistribution of the QDs was produced, based on the empirical image cytometry data (figure 2.8). The initial population was $N_{in} = 10000$ cells. Each cell was characterized by four parameters, its QD fluorescence (f) and its maximum age before mitosis (a_{mx}), its generation (gen) and its parent id (id_p). The parent id of the initial population was set to the value -1. To get the initial QD load I sampled from the experimental QD fluorescence distribution, and to assign the age in each cell I

sampled from the experimental IMT distribution I discussed earlier. The time step of the simulation was 24*hours*. At each time step, all cells were tested to see if they had undergone mitosis. For those who would divide a QD splitting ratio value was drawn from a uniform distribution (see section 2.4) to decide how the parental QD fluorescence load would be distributed into the daughters. For each cell that divided, two daughter cells were created and the four parameters (f, a_{mx}, gen and id_p) were assigned a value accordingly. In the end of the simulation, the daughters of each generation were grouped as high loaded if they had more than 50% of the parental QD fluorescence load or low loaded if they had less than or equal to 50% of the parental QD fluorescence load. Next, I constructed the QD fluorescence distribution of the high and low loaded daughters of each generation (i.e. I age-sorted the cells).

As a next step, to be able to compare the QD fluorescence data from the simulation with that from flow cytometry experiments I convolved the high and low loaded daughter subpopulations into one population and I removed the genealogical relationships between parent and daughter cells and divided the QD fluorescence data into 24 hours groups.

```
Initialize N numbers of cells:

Create Nx4 vector allparams to store the following data:
for each cell:
    Cell's age before mitosis drawn from distribution (α_mx)
    Cell's QD fluorescence drawn from distribution (f)
    Cell's generation = 0 (gen)
    Cell's parental id = -1 (id_p)
for time = 24, 48 hours:
    while the age of cells is smaller than time:
        find how many cells are dividded
        for each dividing cell
            kill dividing cell
            create 2 new cells
            draw a splitting ratio drawn from a uniform distribution
            for each cell
                assign its age before mitosis, drawn from distribution
                assign its QD fluorescence value from distribution
          update vector allparams
```

Figure 2.8: In silico model algorithm for the redistribution of Qd fluorescence, shown in pseudocode.

Statistical Analysis

Kolmogorov-Smirnoff test was used to compare the empirical distribution from the microscopy experiment and that of the in silico experiment (see table 2.1).

2.4 Results

The distribution of the IMT was extracted and is shown in figure 2.9 a. The best fitted distribution was a Generalized Extreme Value probability density function (GEV, equation 2.1), with shape parameter $k = 0.1851$, scale parameter $\sigma = 4.2044$ and location parameter $\mu = 20.691$ (The fit was realised using custom made routine in MATLAB).

$$f(x) = \begin{cases} \frac{1}{\sigma} exp\left(-\left(1+k\frac{x-\mu}{\sigma}\right)^{-1/k}\right)\left(1+k\frac{x-\mu}{\sigma}\right)^{-1-1/k} & k \neq 0 \\ \frac{1}{\sigma} exp\left(-\frac{x-\mu}{\sigma} - exp\left(-\frac{x-\mu}{\sigma}\right)\right) & k = 0 \end{cases} \quad 2.1$$

The generalized extreme value model combines three simpler distributions into one, and depending on the three parameter values it can allow a continuous range of possible shapes. The empirical distribution is fat tailed on the right and the tail does not decrease exponentially but rather as a polynomial and thus, the GEV distribution of Type II (a.k.a Frechet) distribution was the one to best explain the IMT data. Figure 2.9 b, in addition, reveals that the QD fluorescence partitioning ratio values from parent to daughters are distributed uniformly, irrespective of generation.

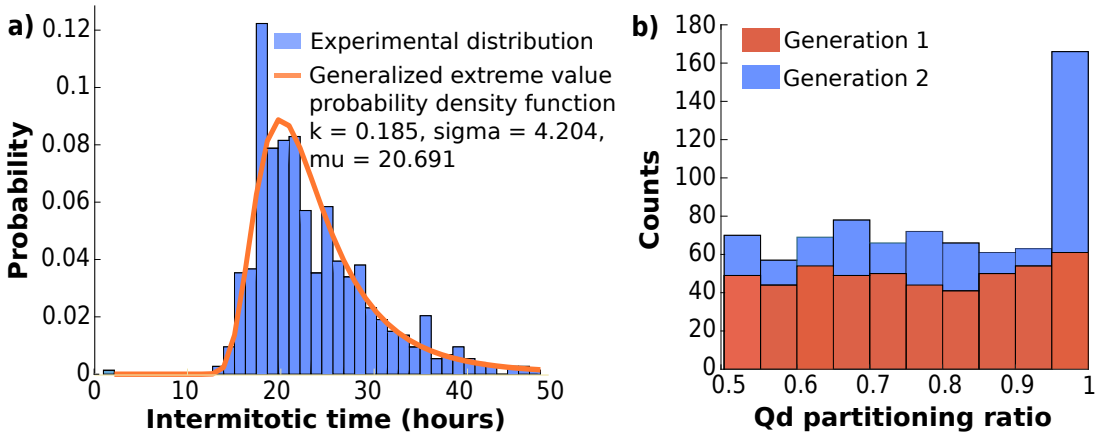

Figure 2.9: Empirical distributions for a) the Intermitotic Times (IMT) and b) the QD fluorescence partitioning ratio. The IMT distribution is fitted with a Generalized Extreme Value (GEV) probability density function with parameters $k = 0.1851$, $\sigma = 4.2044$ and $\mu = 20.691$.

The empirical distribution of the QD fluorescence intensity for the daughters at each generation is shown in Figure 2.10 a,b and c. The distributions of the daughters do not coincide and, thus, the QD fluorescence is distributed asymmetrically into the daughters, where some daughters appear to have higher QD fluorescence intensity values (i.e. distributions where the mode is closer to the parental).

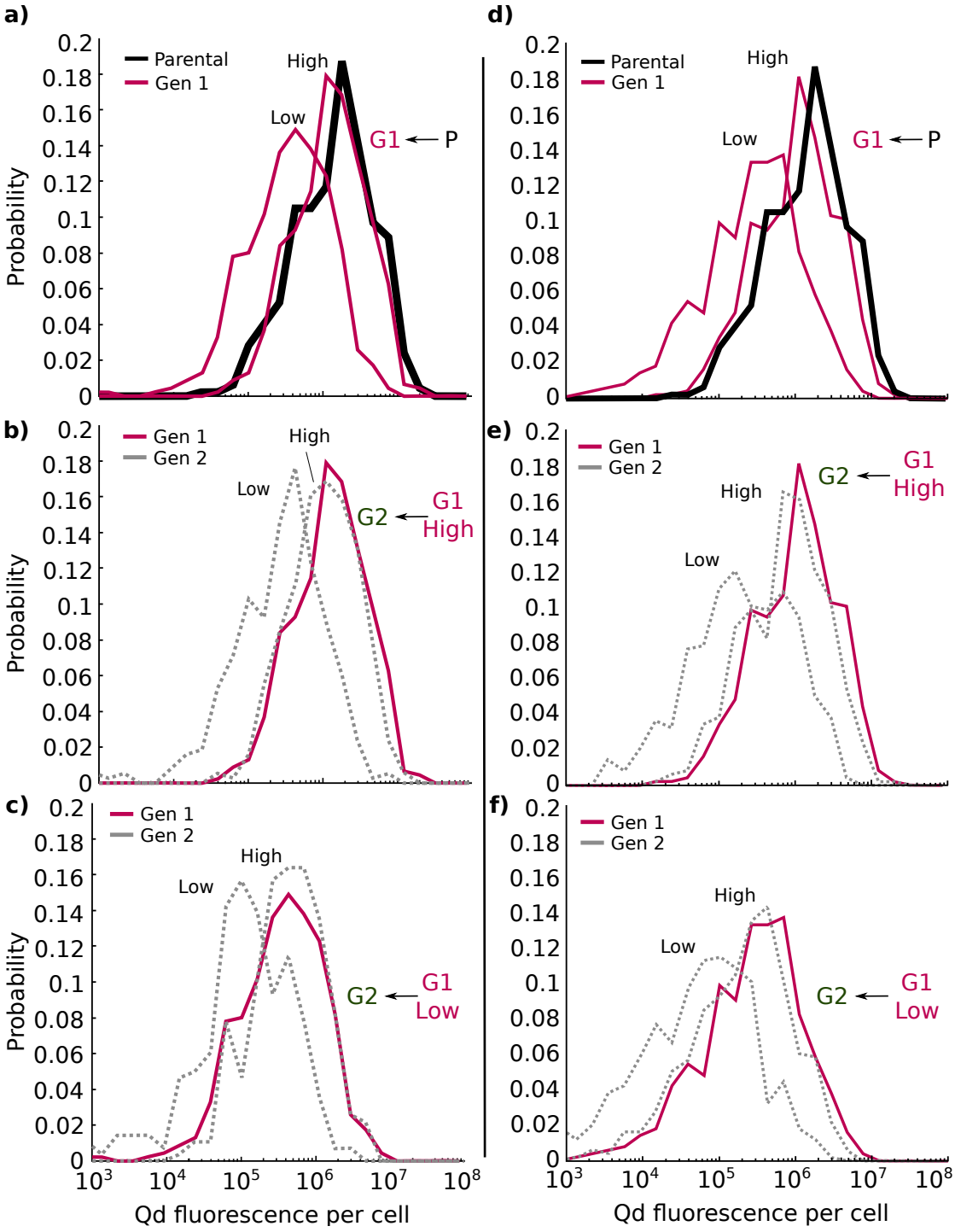

Figure 2.10: The empirical (a, b, c) and in silico (d, e, f) distributions of the QD fluorescence intensity of the cells of 3 consecutive generations; The daughters are sorted into high and low fluorescence load. (a) and (d) show the redistribution from parental to generation 1, (b) and (e) the redistribution from generation 1 (high loaded) to generation 2 and (c), (f) the redistribution from generation 1 (low loaded) to generation 2.

Figure 2.10 d, e and f shows the in silico distribution of the QD fluorescence of each daughter of each generation. Kolmogorov-Smirnov statistical test indicated no difference between the empirical (figure 2.10 a,b,c) and in silico distribution (figure 2.10 d,e,f) as can be seen in table 2.1. This assures that the in silico model is working well and can successfully represent the experimentally driven distribution. Therefore, it can be used for further analysis and prediction.

Figure 2.11 a, shows the initial empirical distribution of QD fluorescence from flow cytometry experiment and the 24-hour snap-shot distribution of the QD fluorescence intensity of the in silico model presented here. 1000 realizations of the in silico distribution of the QD fluorescence are shown, together with the empirical distributions from flow cytometry experiments. The empirical and in silico distributions match and thus, my model can be used to understand the content of the QD fluorescence distributions derived from the flow cytometry data. Figure 2.11 b shows the contribution of each generation that is involved in one empirical flow cytometry distribution. In this example, a 24 hour QD fluorescence distribution (thick orange line in figure 2.11 a) is consisted of actually 2 generations, the parental and generation 1. Therefore, the figure demonstrates that snap-shot flow cytometry samples are consisted of more than one genaration of cells and, thus, data from flow cytometry empirical distributions have to be interpreted carefully when genealogical relationships are the subject of the study.

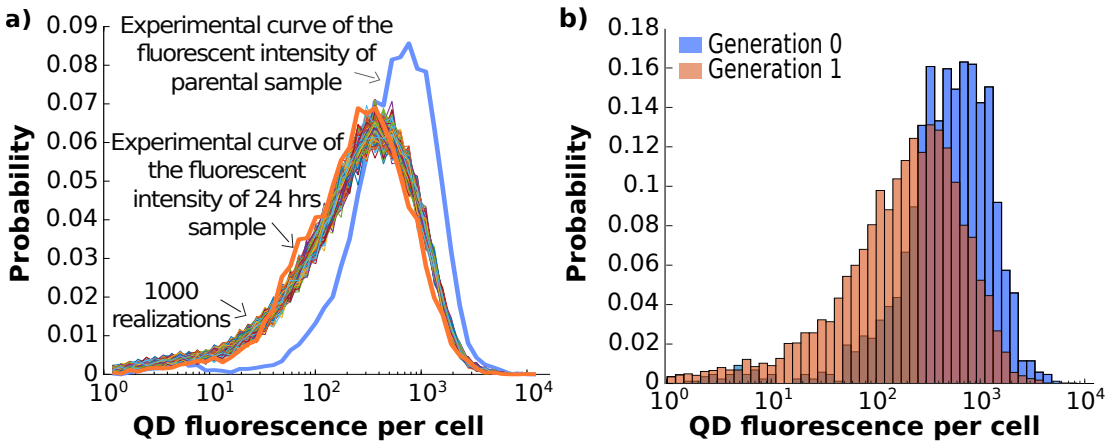

Figure 2.11: Distributions derived from the in silico flow cytometry model. a) 1000 realizations of the QD fluorescence per cell distribution of a 24 hour sample, overlayed with the actual empirical distribution. b) Distributions of the generations that are involved in the 24 hour in silico sample.

2.5 Summary

The current chapter strongly demonstrated that the partitioning of the QD fluorescent load in proliferative Human Osteosarcoma cells is asymmetric and is governed by pure

stochasticity and there is no underlying biological process that creates a preference for a particular asymmetric partitioning value. It also derived empirical inter-mitotic (IMT)splitting ratio distributions. And developed a in silico model that can be used to predict the dilution of the signal for longer times (more than two generations). Lastly, the insilico model can be used to reproduce also flow cytometry experiments and reveal/predict the underlying contribution of generations in each flow cytometry sample.

The current chapter strongly demonstrated that the partitioning of the QD load in proliferative Human Osteosarcoma cells is asymmetric and is governed by pure stochasticity and there is no underlying biological process that creates a preference for a particular asymmetric partitioning value. It also derived empirical IMT splitting ratiodistributions and developed a in silico model that can be used to predict the dilution of the signal for longer times (more than two generations). Lastly, the in silico model can be used to reproduce also flow cytometry experiments and give insight on the underlying contribution of generations in each flow cytometry sample.

The methodological approach I followed (i.e. microscopy) intrinsically age-sorts the QD fluorescence data into generations and the daughter cells are further sorted into a high and low loaded QD fluorescence categories. By simulating the redistribution of the QD fluorescence in an age-sorted in-silico model I can recreate the perceived asymmetry found in flow-cytometry experiments as is shown in figure 2.11 a. By convolving the highand low loaded daughter subpopulations into one population I remove the genealogical relationships between parent and daughter cells and by randomly sampling the measured intermitotic time distribution I remove the age-structured or generational relationship. Under these conditions, the QD partitioning ratios match contiguous flow cytometry measurements.

Overall, the analysis identified two contributing factors that obscure correct interpretation of snap-shot, QD fluorescence flow cytometry experiments; these experiments are incapable of i) identifying genealogical relationships and ii) robustly identifying an age-structured population, i.e. clearly assign the cells to generations. These criteria are essential to quantify and predict how a known QD fluorescent dose evolves over time. For a more in-depth discussion please refer to section 6.1.

Table 2.1: Kolmogorov-smirnov results for the comparison between the empirical and in silico Quantum dot fluorescence distributions.

Generation	K-Statistic	P-value
G1-high	0.073	0.999
G1-low	0.049	1
G2D1-high	0.15	0.738
G2D1-low	0.12	0.901
G2D2-high	0.098	0.985
G2D2-low	0.07	0.999

Chapter 3

Long term tracking and analysis of Human Osteosarcoma fluorescent Cells.

The purpose of this work was twofold, firstly to develop a routine that automatically tracks the nuclei of Human Osteosarcoma (U2OS) fluorescent cells and extracts cell cycle and Quantum Dot (QD) load information. Secondly, the aim was to demonstrate some collective statistical properties of the U2OS cell population, i.e. their mobility patterns (e.g. the distance traveled) and their cell cycle attributes (e.g. intermitotic times).

3.1 Introduction

The advent of high throughput time-lapse microscopy has provided data at relevant length-scales to answer many pertinent biological questions. For example, the dynamical behaviour and, specifically, the mobility patterns in cell populations require the acquisition of long term time-lapse data. Previously, these questions were difficult to study, but now it is possible to extract thousands of images in only few hours.

However, the ability to collect such a big amount of images raised questions on how to effectively extract the information that is contained into the images. Until very recently, the only way to extract the data was to manually segment and/or track the cells (Meijering, 2012). This is a hard task that requires many hours of repetitive work and is prone to human bias. The use of algorithms that can automatically segment and/or track cells could solve the problems that emerge from manual tracking.

However, developing such tools is also challenging for a couple of reasons (Kan *et al.*, 2011; Sbalzarini, 2016; Nketia *et al.*, 2017). First, this task requires a lot of expertise on the field of computer vision and signal processing. In addition to that, because of

the particularities of each cell imaging technique, but also the intrinsic cell variability in the sample within a specific experiment, the aquired images can vary widely in their characteristics (i.e. in their intensity, features and signal-to-noise ratio). This makes it difficult to develop a routine that can successfully detect all cells or cell features in all theses images. For this reason, the tracking methodologies are highly specific and, usually, not widely applicable, i.e. a routine that is used to track one particular cell type cannot be easily applied to track other cell types. For example, phase-contrast imaging techniques are very different from fluorescent images (Meijering, 2012; Sbalzarini, 2016) (where specific parts of cells are labeled by some fluorophores) and produce images of very different attributes (figure 3.1).

Tracking cells from phase contrast images appears to be challenging because of the artifacts that are produced during image acquisition. Cells appear to have halos around the cells, the nucleus is not visible and there are dark areas inside the cells due to the refractive properties of some cell structures (figure 3.1). In addition, there are difficulties in detecting and separating correctly cells that are very close to each other (Debeir *et al.*, 2005; Thirusittampalam *et al.*, 2013; Dewan *et al.*, 2014).

In fluorescent images, depending on the part of the cell that is labeled and the selection of the fluorescent marker, it can be possible that the cell outlines are not detectable, the intensity values can have a high dynamic range and some cells can be invisible because of the absence of fluorescence. In addition, studies have shown that the chemical agents that are used to produce fluorescent images can cause unwanted damages to the cell (Horan, Paul Karl *et al.*, 1990; Resch-genger *et al.*, 2008; Specht *et al.*, 2016; Laissue *et al.*, 2017). This is very common, when the fluorescent marker targets the cell nucleus (Dobrucki *et al.*, 2007). This results in cells that have a limited viability and sometimes modified functionality due to the toxicity caused by the dye (Horan, Paul Karl *et al.*, 1990; Resch-genger *et al.*, 2008; Specht *et al.*, 2016). For these reasons, the researchers compromise, and either track the cells for short time periods, or minimize the fluorescent dose as much as possible to increase cell's life expectancy. This method usually results in a very low signal to noise ratio as long as in the dilution of the signal after some generations.

In addition to the previous, differences in cell images also can appear from the differences in the intrinsic characteristics of the cell types. For example, neural cells have different morphology, shape and behaviour from stem cells. Therefore, to develop an efficient tracking algorithm, it is useful to consider the artifacts that can be caused from each imaging technique, as well as the behavioural particularities of each cell type.

The methodologies that are developed for cell image analysis are also question-oriented, and therefore vary, depending on the topic of interest. For example, questions regarding the quantification of some cell attributes irrespective of time demand algorithms that can accurately segment specific or whole cell areas (Zimmer, 2012). On the other

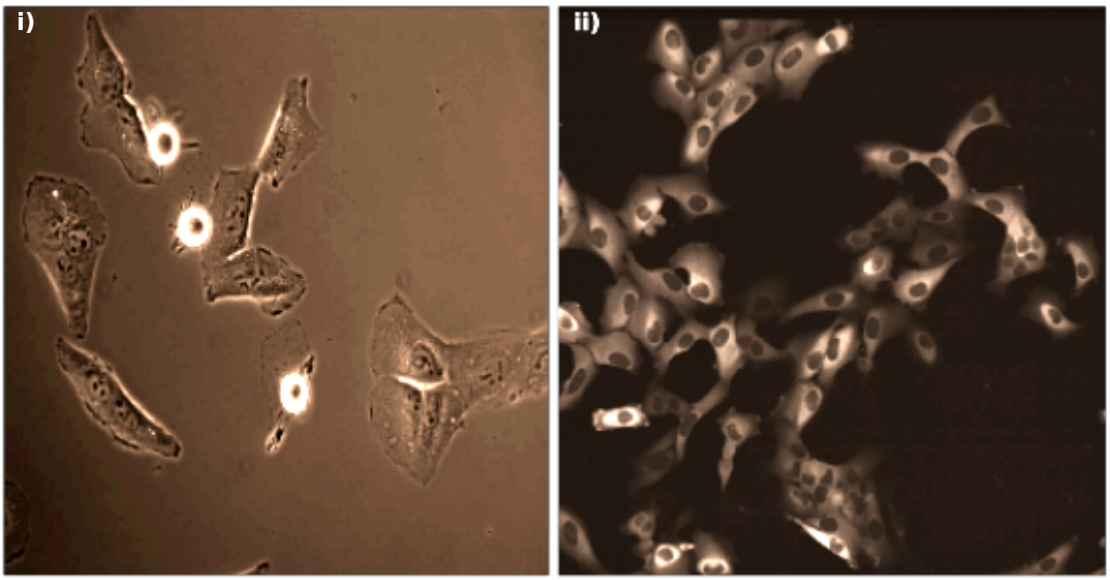

Figure 3.1: Examples of two different imaging systems, i) phase contrast imaging and ii) fluorescent imaging. Halos and dark areas due to the refractive properties of the intra-cellular structures are visible in phace contrast images. Cells show highly variable intensities depending on the concentration of the fluorophore in fluorescent images.

hand, questions on the dynamical behaviour of the cellular population require routines that can successfully associate cells between frames (Li *et al.*, 2008; Kostelec *et al.*, 2015). For example, questions on the characteristics of the cell cycle of a population require a routine that successfully detect two consecutive mitotic events (Li *et al.*, 2010b), but do not require the accurate segmentation of the whole cell outline. The same holds for cases where the extraction of cell lineages is the target of the study (Li *et al.*, 2008).

Despite the challenges and difficulties in the field, there are many studies proposing a wide range of methodologies for cell segmentation and tracking. Masuzzo *et al.* (2016) and Nketia *et al.* (2017) have categorized the existing tracking algorithms in three main groups: independent segmentation of individual frames followed by data association (or tracking by detection as proposed by reference), tracking by model evolution and tracking by filtering.

In the first category, that is the one followed in the current work, each frame is segmented separately and the cells from consecutive time frames are associated by applying different rules and approaches. Wang *et al.* (2010) in their work use fluorescent images and a neighborhood score to associate the cells in different time frames. Youssef *et al.* (2011) use multiple parameter tracking, i.e. they use cost matrices from multiple cell parameters such as object size, position or texture to assign the nuclei in consecutive frames. A more complicated and effective approach is the work of Chatterjee *et al.* (2013), where they use the bipartite networks and cost function minimization approaches to make the linking between timeframes, in fluorescent images. Finally, Thirusittampalam *et al.* (2013) use Delaunay mesh to track the cells, in phase contrast images.

In the second category, model evolution tracking algorithms are used to follow cell positions. The cells are associated between each frame by the use of deformable models. These are curves or surfaces defined within an image domain that can move under the influence of internal forces coming within the model itself and external forces computed from the image data. In case of cell tracking, contours of cells are evolved along time to follow the changes and movements of each cell. Examples of works using model evolution are those applying explicit or implicit active contours such as Chan & Vese (2001); Zimmer *et al.* (2002); Li *et al.* (2010a); Dzyubachyk *et al.* (2010); Yeo *et al.* (2011); Zou & Tomasi (2016); Boukari & Makrogiannis (2016); Sazonov *et al.* (2016). Tracking methodologies that use the mean-shift approach also belong to this category according to the recent paper of Masuzzo *et al.* (2016). Some examples of works applying meanshift for tracking are the works of Debeir *et al.* (2005); Zhang *et al.* (2015b).

As already discussed, labeling the nucleus with fluorescent markers can seriously affect the viability of the cells. This is a limiting factor that can prevent scientists from running long-term microscopy experiments. Labeling other parts of the cell can be more promising for long term tracking. U-2 OS cells can be tagged with a green fluorescent protein (GFP) in the cytoplasm, a stealth reporter (Thomas and Goodyer, 2003) whose expression is driven by the promoter of the cyclin B1 protein and allows for continuous cell cycle readout. This particular marker oscillates according to the cell cycle (figure 2.2) and, thus, works as a cell cycle marker too. However, due to the non-linear nature of the GFP signal identification tracking is non-trivial and is not achievable by recognised cell tracking softwares, such as, Cellprofiler (Jones *et al.*, 2008) and NIST Tracking (Chalfoun *et al.*, 2016). This system is more challenging than tracking nuclei, but is more promising as well for long term tracking and can provide a deeper understanding of the system, as it can provide cell cycle information too. Some of its advantages are that i) the labeling is not invasive and threatening for the life of the cells, ii) the labeling is stable for very long periods and iii) the cell cycle information provided by the marker could shed light on intrinsic attributes of the cells.

Current work

In this work, I developed a semi-automated method for tracking cells, in a system of U2OS cells labeled with a fluorescent marker in their cytoplasm that is activated and deactivated in a nonlinear way according to the cell cycle. The algorithm in principle takes advantage of the intensity from the GFP marker in the cytoplasm to track the nuclei of the cells. It is easier to segment the nuclei (figure 2.2) because of the similar shape and intensity they have along time and because of the low probability of finding two nuclei so close to each other, thus it's easier to separate them. The algorithm can also

detect mitotic events and extract cell cycle properties using the GFP signal. Because it is nucleus label free it has the capability of tracking cells for very long times and extract lineages. The algorithm is implemented in C++ using OPENCV 3.0.0.

3.2 Cell culture and microscopy

3.2.1 U2OS cell lines and the GFP marker

Chapter 2 and section 2.1.2 presents information on the Human Osteosarcoma cell line and the Green fluorescent reporter (GFP) used for labeling the cytoplasm of the cell. Therefore the reader is refered to that chapter for retrieving relevant information.

3.2.2 Microscopy setup and experimental procedure

Cell culture preparation

U-2 OS (ATCC HTB-96) cells were maintained in 10% McCoy's 5a full medium, in culture flasks. Then, commercially available targeted quantum dots (QDs) were added to the medium, using the Qtracker®705 Cell Labeling Kit (4 nM). They were then left for an hour for the particles to land on the cell membrane and to be attached to the cell surface. After that, the cells were trypsinized and consequently transfered and seeded to a 24-well plate (see figure 2.3). 24 hours later the Qdots were internalised into the endosomes via the endocytotic route.

Image acquisition

After 24 hours, fluorescent time-lapse images using IN Cell Analyser 2000 (GE Healthcare) and $\times 40$ magnification, were acquired with a 30-minute interval between frames for a period of 67 hours. Two fluorescent channels were used (one for the GFP stealth reporter, see figure 3.2, and one for the QDs). The images were stored in Tagged Image File Format (TIFF) and $1\,pixel = 0,16\,\mu m$.

3.3 Segmentation and particle tracking

3.3.1 Ground truth data acquisition

From the images 500 cell lineages were extracted manually (figure 2.7). To extract the data all the images were processed using OPENCV/C++ (Itseez, 2015). To track the lineages an algorithm was developed to manually extract the lineages, to collect ground truth data. First, all the images were denoised using the non-local means denoising technique (Buades *et al.*, 2005).

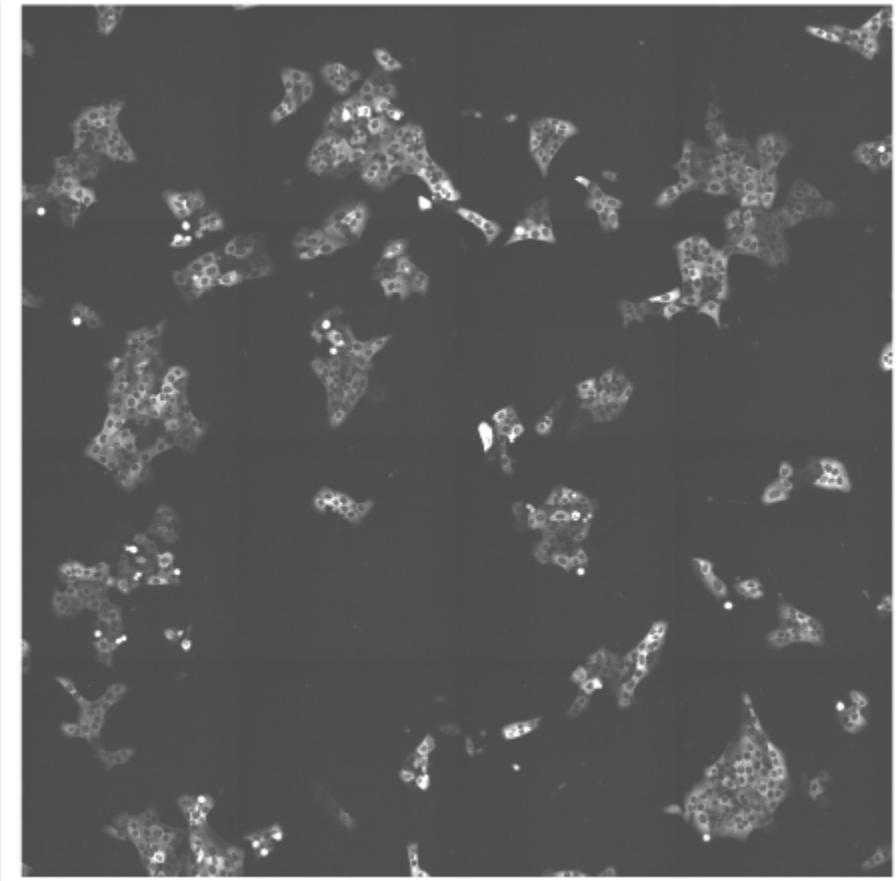

Figure 3.2: An example of a $4\,field of view \times 4\,field of view$ image of the fluorescent cells in a well plate. Most cells are part of colonies and can be either inside of it or at the periphery. Cell density can vary depending on the field of view studied, rendering the effectiveness of the algorithm variable.

The routine for the manual data extraction was the following:

First, read and denoise each channel of each frame (see figure 2.4). Activate mouse for user (figure 2.7). With the mouse, draw the cytoplasm outline of the targeted cell on the GFP channel. Then create a binary cell mask. Store $x-y$ coordinates and lineage information. Use OTSU's threshold to segment the QDs of the QD channel. Finally, count the total QD and GFP intensity.

3.3.2 General outline of the automated tracking algorithm

To choose the appropriate steps for the cell segmentation and tracking the following particularities of the system were taken into account: The foreground objects (i.e. the cells) have a high intensity variation. The signal to noise ratio is relatively small resulting in low contrast images. U2OS cells are densely packed, i.e. one is next to the other and, thus, the outlines of the cells are not very clear. In addition, the cells' speed is low meaning that the searching window of a targeted cells will be relatively small between frames. Finally, the dynamic range across timeframes varies too.

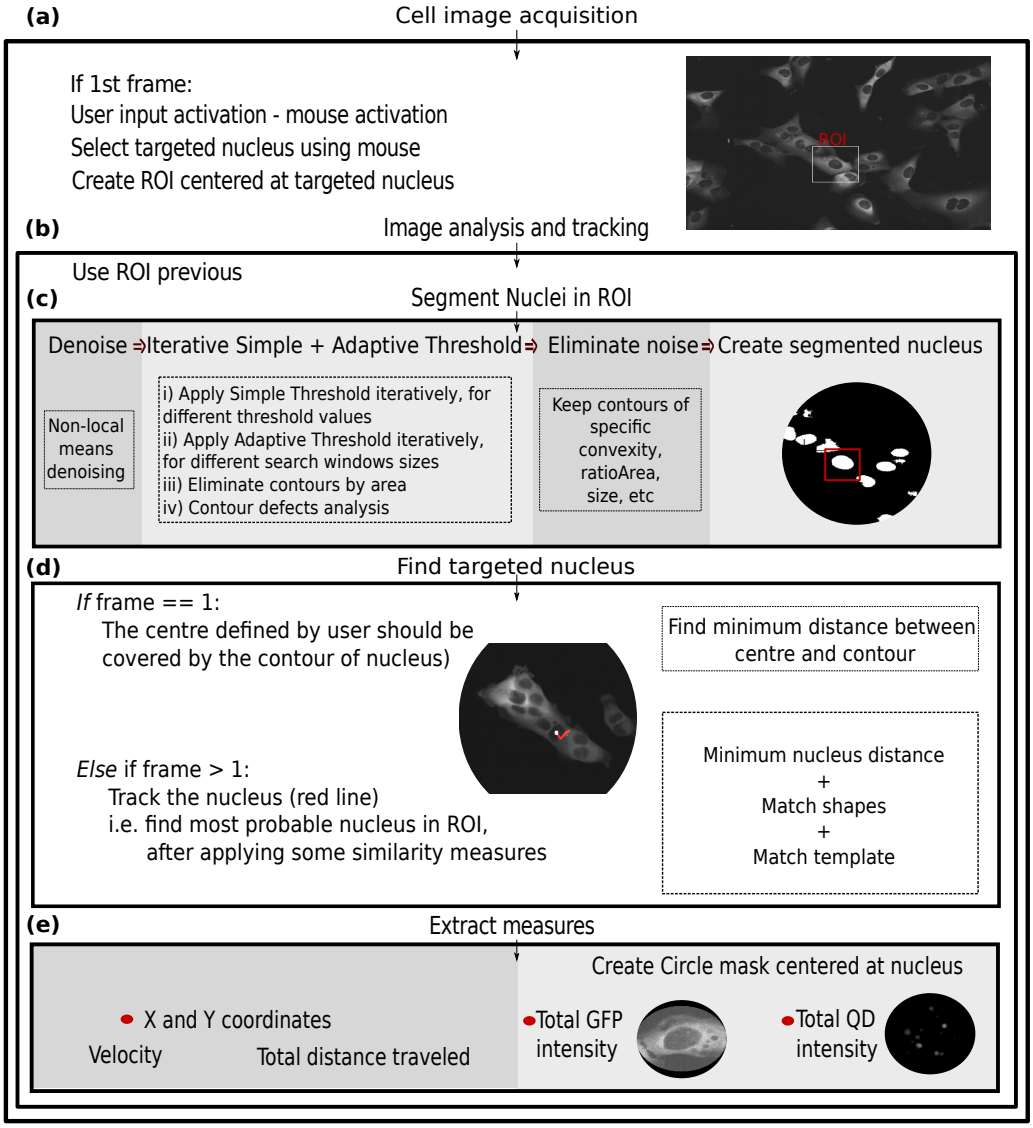

Figure 3.3: Overview flowchart of the proposed algorithm. It is categorized in five subroutines. a) The user is selecting the targeted cells by clicking on the trageted nucleus and a Region Of Interest (ROI) is created centred at the nucleus. b) Then, the processing of the image is followed. The ROI is the new window where all analysis will take place in the following steps. c) Next is the Segmentation subroutine, where the image is first denoised and the Iterative simple and adaptive Threshold techniques are applied to segment the image and extract the nuclei. These techniques are followed by postprocessing step, where the image is filtered to remove noise and irrelevant objects and to seperate touching nuclei. d) Consequently the cell association subroutine is called, where the most probable cell is associated with the trageted cell of the previous time frame. This is the result of the comparison of the probability each segmented object has to be the targeted nucleus. The probability consists of the weighted sum of different similarity metrics. e) After assigning the nucleus, all important measures are extracted (such as the nucleus position, the GFP and QD intensity, the cell cycle information and the genealogical information.

The main objective of the currect work was to retrieve cell paths. Extracting the whole outline of the cell was not necessary, especially if there was another way to get the cell position. Cell nuclei have more distinct outlines and thus is easier to segment them. For this reason, first step was to extract the positions of the nuclei.

The basic steps of the algorithm are, therefore, the following. At the beginning, the user indicates the targeted cell, by pointing to the centre of the nucleus. A Region of Interest (ROI) of specific size ($width = 250\,pixels$, $height = 250\,pixels$) is created centered at the targeted cell in order to decrease the complexity and time needed to track the cells (see figure 3.3). All the analysis will be restricted inside the ROI. The next step is to de-noise the image. Figure 2.4 on the left shows the histograms of a background region before the denoising. The histogram indicates that there is Gaussian noise ($\mu = 34$, $\sigma = 4.190$ intensity). The nonlocal denoising approach is used to get rid of the gaussian noise (Buades *et al.*, 2005). The parameter σ was calculated from the intensity histogram of background region, assuming that any variability in the background region is due to noise, since the growing medium is well mixed. The dynamic range between frames changes. For this reason, to be able to extract meaningful Cell cycle information, I normalized the intensity values, so that the GFP changes in intensity reflect mainly the intrinsic changes during the cell cycle.

After denoising, the image is segmented into the nuclei and postprocessing takes place to eliminate noise. Specific rules are followed to associate previous nucleus at time t with nucleus at time t+1 (figure 3.3). Once the nuclei are extracted, a circle mask of specific radius centred at the nucleus is created to extract cytoplasm and other measures (such as GFP and QD intensities etc).

3.3.3 Cell segmentation (still image)

Overview of segmentation techniques

The simplest way to segment the nuclei is by using Simple thresholding, where pixels of an intensity value higher than a chosen threshold are labeled as foreground. Despite its high performing speed this method does not account for the intensity variations of the signal and thus it is not efficient in cases the foreground's intensity varies, since one threshold value is not sufficient to segment all the cells of interest. In Adaptive threshold, each pixel has its own threshold value calculated by the intensity values at the neighborhood of the focal pixel. The function transforms a grayscale image to a binary image according to the following function

$$dst(x,y) = \begin{cases} maxVal & \text{if } src(x,y) > T(x,y) \\ 0 & otherwise \end{cases}$$

where $src(x,y)$ is the intensity value at x and y pixel position of the source image and the threshold value $T(x,y)$ is the weighted sum of neighbourhood values where weights

are a gaussian window of size $blockSize \times blockSize$.

Adaptive threshold is still a fast and easy to implement methodology and it also considers the variation in the cell intensity. However, in cases cells differ in size and have a low contrast and/or the density of the cells is very high the method could miss some cells (figure 3.4). However, in the current case, the target is to segment the nuclei and thus, adaptive threshold can give a high percentage of the nuclei even in a densely packed population, except in the case the outline of the nucleus is not so clear, i.e there is low contrast between nucleus and cytoplasm.

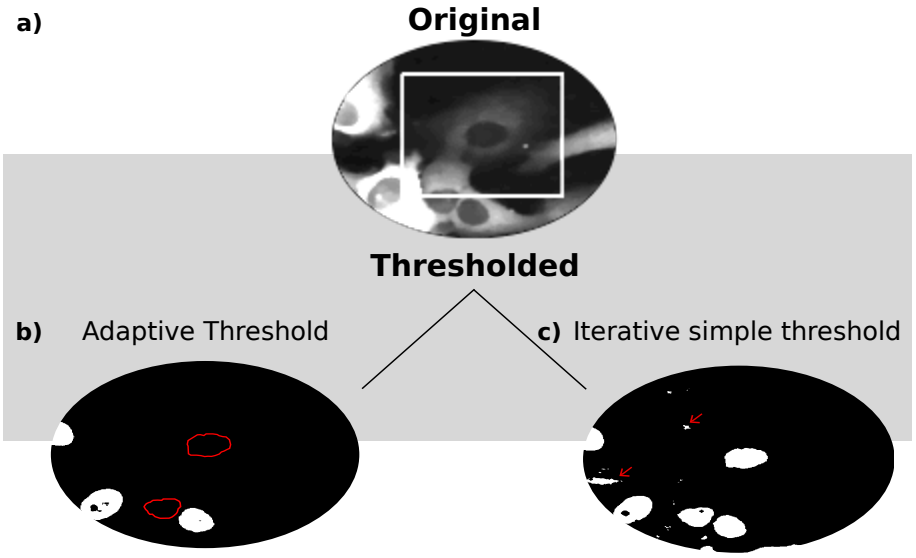

Figure 3.4: Comparison of the resulted segmentation between the Iterative Simple and Adaptive threshold. a) Original GFP channel image shows the low intensity/low contrasted cell (centered at the with rectanle). b) Segmented nuclei using adaptive threshold technique. c) Segmented nuclei using the iterative threshold technique. Red arrows point to the false segmented objects, a.k.a noise.

To increase the percentage of segmented nuclei an Adaptive threshold with further processing could be used. Such post-processing could include morphological operations, isolation of the targeted object (in our case nucleus) by applying some filtering (according to size, shape, area etc) and use of morphological properties of the nuclei (convexity defects, size etc). The advantages of the previous method are that it remains simple and fast to implement, and increases the percentage of correct segmented nuclei considerably. However, it requires adjustment of many parameters and steps for different system, and thus, it cannot be applied immediately to different systems.

Another approach for nuclei segmentation would be to use simple threshold in an iterative mode: If we threshold the image for consecutive threshold values and then apply some morphological filters (such as filtering according to size) to eliminate some of the noise, then we can eventually get an image where most of the nuclei will be represented by a forground object. This method is particularly useful for segmenting the dimmer

cells (figure 3.4). It is also easy to understand and implement. However, this method also segments some wrong objects that need to be removed. However, it can extract part of the nulceus and not all object.

Combination of the above methodologies could result in better segmented images and less false segmented objects. For example, by using the iterative approach only for segmenting the dimmer cells (and determining the threshold values at each time frame from the respective image histogram), and using the adaptive threshold for segmenting the rest of the cells, then we can further minimize the error and maximize the successfully segmented cells.

Some more elaborated methodologies widely used to segment cells include the H-maxima, where the local maxima of the cells/nuclei can be detected and isolated creating the seeds to be used in the following steps (Dewan *et al.*, 2011). H-maxima transformation suppresses any of the regional maxima whose height is less than h. Graph cut algorithms represent the image as a weighted graph where pixels are linked to either a source or a sink node with some weights. There is also a cost function associated with the weights. The final segmented image will be the one that minimizes this cost function (Rother & Blake, 2004; Min & Wan, 2013). Another way to segment nuclei would be to use model evolution and more specifically, implicit active contours. Using a roughly segmented image as initial we can segment the cells and the nuclei. The method is not affected by the noise and by the case of open boundaries of objects. Also, it is not affected by the different intensity values. Thus, it is a very robust method and creates closed curves even if objects of interest do not have distinct boundaries. However, it is more difficult to understand and implement due to its strong mathematical concepts and formulation. It is also slower than the previous methods.

Machine learning techniques that can also be use to segment the nuclei. The simplest one is the K-means clustering, but it requires the a priori knowledge of the number of cluster you need. Other techniques such as artificial neural networks and deep learning, demand a strong computational power and also a good amound of data to train the algorithm.

Current segmentation steps

The steps to segment the nuclei in the current algorithm are the following: First, I create a Region-of-Interest (ROI) centered at the segmented nucleus of 72×72 μm size, so that all segmentation steps are applied on the sub-set area (figure 3.3). Because the ROI is centered at the previously segmented nucleus and ROI is much bigger than the average steplength of the cells, the incorrect segmentation of the objects in the corners is not limiting the successful detection of the nucleus. However, in rare cases that the targeted cell is close to the edges of the initial image the ROI will be truncated and the

segmentation will be affected if the nucleus is not apparent.

Before proceeding to the actual segmentation some image pre-processing steps take place. These include, image denoising (figure 2.4), image normalization(3.5) and image smoothing (i.e. Gaussian Blur).

In more detail, I first denoise the image using Non-local means denoising (Buades *et al.*, 2005). In contrast with previous denoising filters that consider local areas to denoise the image, i.e. take the mean value of a group of pixels surrounding a target pixel to smooth the image, non-local means filtering takes a mean of all pixels in the image, weighted by how similar these pixels are to the target pixel. This results in images with greater clarity, and less loss of detail in the image compared with local mean algorithms (figure 2.4).

After denoising, I apply Mean Shift segmentation, a local homogenization technique that is very useful for damping shading or tonality differences in localized objects (Yizong Cheng, 1995). The output of the function is the filtered "posterized" image with color gradients and fine-grain texture flattened. This creates a locally homogenized image that is easier to segment. The number of parameters I need to adjust is two; the radius of the neigbourhood to be considered and the intensity range in which the averaging is allowed.

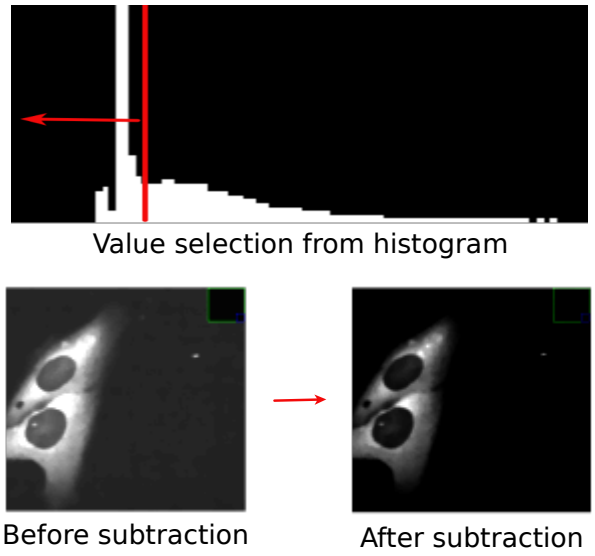

Figure 3.5: Histogram of the intensity values of the GFP fluorescent channel (top row) and the comparison between the original and filtered image after the subtraction of the value from the original image (bottom row, left and right respectively). Red line indicates the value up to which 15 % of the total counts of the histogram are laid.

In addition, I had to take into account that the background value of the image in different frames is not constant, due to the illumination variation in different times. For this, I had to set a minimum intensity value of the background to be the same for all images, i.e. from each image frame to subract a value so that all images have a background value equal to 0. This would limit the intensity variation due to illumination

artifacts and will make consecutive frames more comparable. This step is important for the extraction of cell cycle information, as the progression of cell cycle is represented by the cell intensity variation across time frames in the GFP channel. To achieve this, I get the intensity histogram of each image frame and find where the background values begin. After investigating and comparing the histograms from different frames I decided that the value to be subtracted by the original is the intensity value up to which 15 % of the total counts of the histogram are laid (see figure 3.5). These include the values of the background region too. I then subtract this value from the image.

After pre-processing the image, the next step is to segment the images. I use two complementary methodologies to extract nuclei. First, I apply the Iterative *Simple* threshold to extract the nuclei. The iterative mode is applied by using the simple threshold for a range of threshold values between $0-255$. For each iteration, I use a function to filter the detected objects and keep only those that have a relevant size (i.e. that of a nucleus). As figure 3.4 shows, this methodology extracts most of the nuclei, but also falsely detects as nuclei some objects that belong to the background regions (cytoplasm and medium). Second, I apply Iterative *Adaptive* threshold to extract the nuclei. Again, I use adaptive threshold in an iterative way, where the searching window is the changing parameter. Because a frame is always consisted of cells that largely vary in their intensities, it is very helpful to locally enhance the contrast of the image so that it is easier to segment the nuclei of most of the cells. The parameters to be adjusted are three: the sigma of the Gaussian, the intensity difference to be enhanced and the amount of enhancement to be done.

After applying the main segmentation functions, I post-process the images to eliminate some noise. To do this, I create a subroutine that iterates over the detected objects and discards them, if the objects are not within a specified size range, and show irregular structure. The irregular structure is calculated as the ratio of the area of the object over the area of the elliple and rectangle that bounds the object ($ratio_{ell} = area_{obj}/area_{ell}$ and $ratio_{rec} = area_{obj}/area_{rec}$).

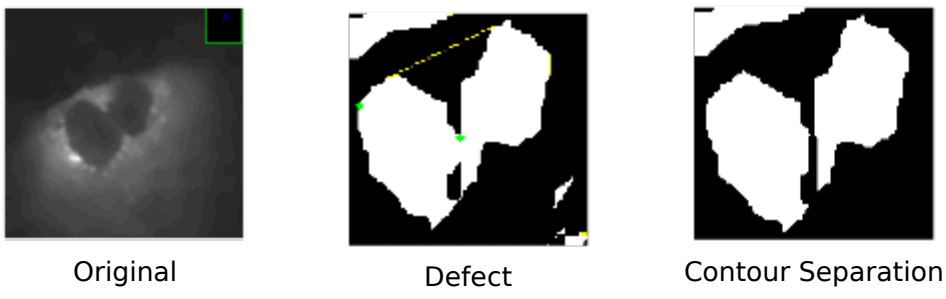

Figure 3.6: Representation of the effect of the contour defect analysis on the cells. The green spot indicates the defect depth and the yellow line indicates the width of the defect.

In addition, I use geometric characteristics of the segmented objects to separate nuclei that were segmented as one object (figure 3.6). To remove false positive objects of small size I apply morphological operations, such as dilation and erosion. To seperate connected contours I developed a custom made routine in C++/OPENCV that realizes contour defect analysis (figure 3.6). Usually contours of more than one nucleus tend to have bigger size and also defects of specific depth and width. This, allows me to find the relevant defects of specific depth and width, as well as their direction and draw a black line perpendicular to the main axis of the defect, to seperate the contour into two.

Dividing cells have different morphology than other cells and thus, a different segmentation method is required for them. Dividing cells appear bright and round or oval (figure 3.8). Simple adaptive threshold followed by feature-specific contour filtering is used to segment the dividing cells. More specifically, segmented objects are kept only if the ratio of the area of the object over the area of the bounding circle is close to 1 ($ratio_{div} = area_{obj}/area_{circle}$).

The final segmented image is the union of the three images segmented in previous steps; i) from the iterative simple threshold, ii) from the iterative adaptive threshold and iii) from the segmentation of the dividing cells.

3.3.4 Cell association (track cells for consecutive timeframes)

To associate the nuclei of two consecutive time frames I needed to find minimum distance measures between the nuclei of two consecutive time frames. The measures can look for local proximity (i.e Euclidian distance) or other types of proximity (e.g. intensity, orientation, shape and size similarity).

To follow one cell of interest that has been successfully segmented, each segmented object was assigned a probability that was extracted from three different metrics: the simple minimum Euclidian distance rule (in time t+1, the object that has the minimum distance from the targeted cell at time t, will be possibly be the targeted cell at time t+1), the minimum shape distance (i.e. the match of shapes), the minimum intensity distance (the match of the Intensity profile around a segmented object).

First, due to the restricted movement of the cells between time frames, the strongest rule to contribute to the final object selection is the spatial distance rule. The empirically calculated distance between two consecutive time frames usually does not exceed the 20 pxl or $3.2\,\mu m$. Each pixel in the segmented image is assigned a probability value, which is drawn by a 2-D Gaussian distribution with mean value $\mu = the\,centre\,of\,the\,previous\,nucleus$ and $\sigma = 3.2\,\mu m$. Each object's spatial probability P_{sp} is then the probability of the pixel where the centre of the nucleus is located. In addition, each segmented object's shape is compared with the shape of the previously detected nucleus and a shape probability (P_{sh}) is assigned at each object. This is calculated

from the following formula: $P_{sh} = \max_{1...7} \frac{|m_i^A - m_i^B|}{|m_i^A|}$, where $m_i^A = sign(h_i^A) \cdot logh_i^A$, $m_i^A = sign(h_i^B) \cdot logh_i^B$ and h_i^A and h_i^B are the Hu moments (Ming-Kuei Hu, 1962) of A shape and B shape, rspectively. Finally, a template of the previous nucleus from the GFP channel is compared with the corresponding areas around the segmented objects to see which of the object has more similar intensity values with the previously detected nucleus in the GFP channel. Each cell is then assigned a template matching probability (P_{tm}) that is calculated after finding the squared difference between the template of the previously detected nucleus with the window of each of the newly segmented objects.

To find the next most probable nucleus, I first check if there is any segmented nucleus inside a window of $16 \times 16 \mu m$ size (and centered at the previously detected nucleus). If there are detected objects within the window, then the object with the higher spatial probability P_{sp} will be selected.

If there are no detected objects within the window, then the algorithm searches for the existence of segmented objects within a window of $19.2 \times 19.2\,\mu m$ size. If there are objects segmented then the routine will select the object of which the weighted sum of the above probabilities is maximized. The weights can be adjusted but P_{sp} is contributing the most.

There are cases where the segmentation of the nuclei has failed. If this happens, the closest contour detected will be farther than the expected. From observation I know that the studied cells make small steps. Thus, I can use this information to prevent wrong associations. For this reason, a condition that has to be met for the contour to be chosen is that the distance with the previous detected contour cannot exceed a threshold value. If it exceeds, then a random point within a small radius from the previous is set as the new targeted nucleus. Therefore, if there are not segmented objects within the previous searching windows, or the dinstance between the previous nucleus with the most probable one is $> 9.6\,\mu m$, then a random point around the previously assigned nucleus will be the new centered point. This is the way the algorithm treats unsuccessful segmentation and association events.

Division events are the most challenging to associate because: i) during division big jumps can take place (misuse of the minimum rule distance), ii) the targeted cell should be linked with 2 instead of one cell after division.

The tracking steps are the following: 1) associate the cells by finding the nearest neighbour (in distance, in intensity values, the difference in intensity between the frames, orientation, in shape, in size etc). If distance is smaller than a threshold calculate the probability of each object to be the targeted nucleus is the weighted sum of the the probabilities discussed above, i.e. the matching shapes P_{sh}, the matching template P_{tm} and the spatial probability P_{sp}. Else, draw random coordinates within a specific radius from the previous nucleus.

3.3.5 Data extraction

From the algorithm I can extract *x* and *y* coordinates of each nucleus for all time steps between mitotic events. The positions can give meaningful information regarding the movement behaviour of the cell population, (total distance traveled, velocity etc). Because the cells are very densely packed and the cell outlines are not easily distinguished, the algorithm does not extract the outlines of the cells. However, it is possible to extract cell cycle characteristics and Qd information by applying a circle mask centred at the detected nucleus and use only the area within the circle to calculate the GFP signal and the QD load. I can apply this for the QD load because from observations I can see that QDs that are internalized in the endosomes and these are located mainly around the nucleus and not at the periphery of the cell.

All the analysis that follows has taken place at R environment (R Core Team, 2017).

3.4 Performance of proposed algorithm and limitations

Nuclei segmentation

The combination of the two iterative methods for segmenting cells discussed previously increases the percentage of correctly segmented nuclei (figure 3.4). Iterative adaptive threshold can successfully detect high contrasted nuclei despite the possible illumination variation within and around nucleus. In contrast, simple Iterative threshold, cannot successfully segment the whole nucleus area due to the illumination variability but is more sensitive to segment parts of nuclei that belong to low contrasted cells.

However, extracting and selecting objects that are only segments of the whole nucleus of low contrasted cells results in the unavoidable selection of false segmented background objects too. Post-processing the segmented image by filtering out irrelevant objects eliminate most but not all noise. Some of the noise cannot be filtered out because some segmented objects resemble the nuclei. Also, filtering methods have to be relaxed to include these the objects segmented from the dimmer cells, with a cost of keeping some noise.

Noise can decrease the successful nuclei association between frames. This usually happens with background areas that are surrounded by cells (figure 3.7). These areas can be of a size similar to the nucleus, and because they are surrounded by the cells they can also have similar shapes with that of the focal nucleus. Especially, the noisy objects that persist for multiple time frames and can be tracked consecutively, i.e. can behave like an "attractor", can limit the performance of the algorithm in the association step.

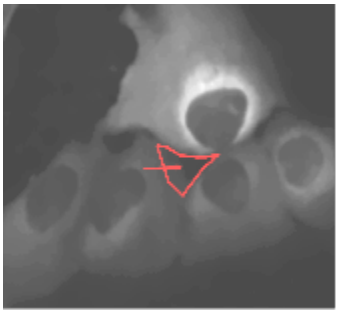

Figure 3.7: The background areas that are surrounded by cells (red outline) can appear to have similar size and shape with the real nuclei and thus be labeled as nuclei. They can also persist for multiple time frames and thus can hinder correct nucleus association.

Mitosis detections

The detection of mitotic cells has been a simple and straightforward task, due to the properties of the dividing cells. They appear to have a round/oval shape and a maximum pixel intensity in comparison with the neighbourhood (figure 3.8).

Mitosis detection failure is observed in the following cases: First, when the cell division is not captured because of the sampling rate of image acquisition. U2OS cells divide around every $20-22$ *hours*. The sampling rate in the current experiment was 2 $frames/hour$. Mitosis usually lasts more than 30 minutes, and thus the capture rate is appropriate for capturing mitosis in most of the cells. Second, the mitosis is not detectable when cell is overlapped with other neighbouring cells. This type of error is minimized after applying some contour post-processing and after detecting and after separating touching cells. Last, but not least, error in mitosis detection can result in the cell association step, not only during segmentation. This mainly takes place when cell's displacement is big, usually more than the length of the nucleus (figure 3.8 c).

30 mins before mitosis Mitosis Mitotic jump

Figure 3.8: Phenomenological behaviour of the cell around mitosis. Left figure shows the cells 30 mins before its division. In the middle figure cell is undergoing mitosis. Its characteristic shape and increased intensity value helps to develop an algorithm that can easily detect mitotic events. However, the third image shows the jump the cell makes in order to get this characteristic morphology. The jump d can be big enough to incommode the correct cell association between the frames.

Cell association

To be able to quantify the performance of the algorithm regarding cell association and the extraction of cell migration routes I had to quantify the error of the routine's detected position relative to the position from groundtruth data. The error ε was calculated as the euclidean distance between the routine's detected position (x_{exp}, y_{exp}) and that of the groundtruth data (x, y): $\sqrt{(x_{exp}-x)^2+(y_{exp}-y)^2}$.

A total number of 85 out of the 203 cells was included in the analysis, since groundtruth data were available only for this number of cells. From those 85, I discarded the cells that were followed for less than 6 time steps. Because groundtruth data are prone to human bias and nucleus segmentation does not always segment the actual outline of the nucleus, ε is approaching but is not reaching 0 in correctly detected cells (figure 3.9). For this reason, a false detection is considered only when *error* exceeds a threshold value.

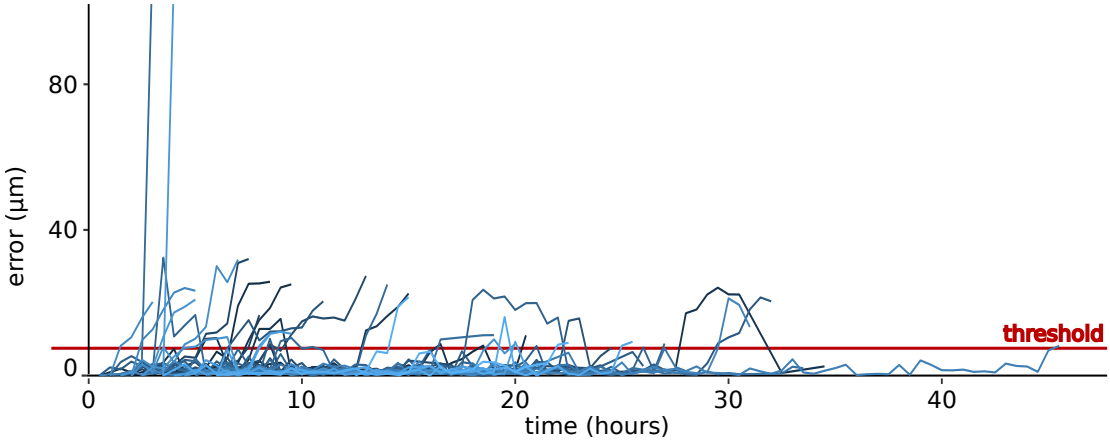

Figure 3.9: The error in the nuclei centre detection across time for all the different cells. The error ε was calculated as the distance between the detected position from the algorithm (x_{exp}, y_{exp}) and that of the groundtruth data (x, y). The red line is pointing to the error value under which the algorithm is assumed to correctly detect the positions.

After calculating the error two performance metrics were computed. The first was the normalized duration of the successfully detected positions before the first error took place (r_{lngt}). If the routine could successfuly identify all positions then the duration would be the total length of the cell's route. Because each cell had different intermitotic times, I normalized it, i.e. took the ratio of the length of the consecutive successfull position detections, l_{emp} over the total time length of the cell's path, l_{tot}, i.e. $r_{lngt} = \frac{l_{emp}}{l_{tot}}$. The metric can thus take values between 0 and 1, i.e. $0 < r_{lngt} < 1$. The second metric was the ratio of the number of successfully detected positions, $N_{success}$, over the total number of occurences, N_{tot}, i.e. $r_s = \frac{N_{success}}{N_{tot}}$.

Figure 3.10 a shows that the algorithm can on average detect consecutively only 50% of the whole path, $mean = 0.51$. The median is 0.45, meaning that half of the times

the algorithm was tested it could not track the cell correctly for more than the 46% of the whole path. However, figure 3.3.4 b also shows that the algorithm can on average detect successfuly 69% total number positions, $mean = 0.69$. Also, half of the times the algorithm was tested, it could successfully detect the 83% of the total positional points. This means, that the algorithm can miss the correct position of the cell along the path, but in case this happen, the error is corrected after some timeframes.

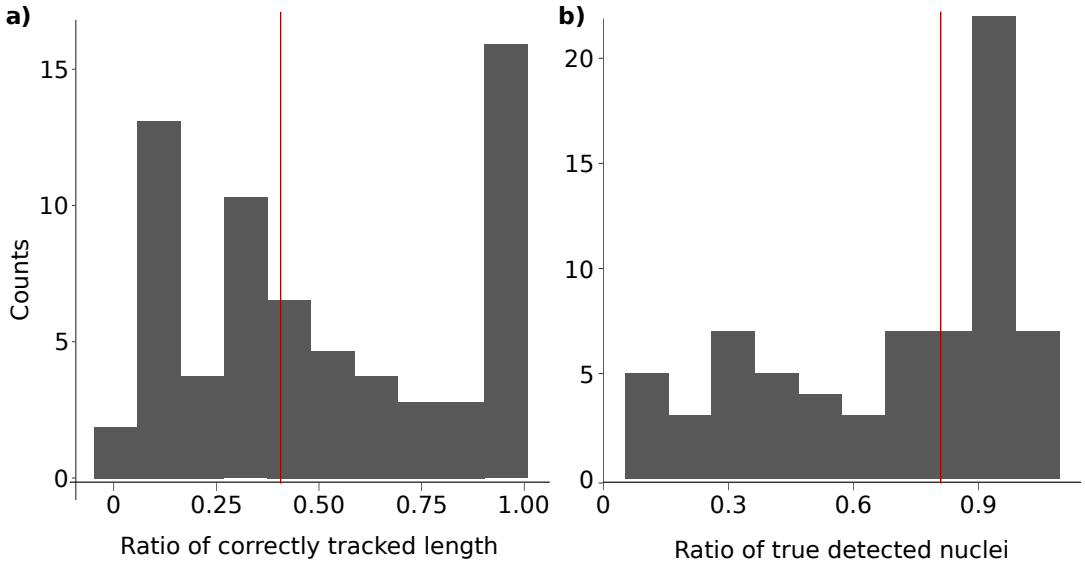

Figure 3.10: a) The distribution of the ratio of the duration of consecutively correctly detection over the total duration of cell displacement. b) The distribution of the ratio of the number of successfully detected positions, $N_{success}$, over the total number of occurences, N_{tot}. Perpendicular red lines correspond to the median values, $median = 0.45$ and $median = 0.83$.

The algorithm presented here can successfully track cells between two mitotic events. The success on the association from a mitotic event to the daughter cells is limited due to the large displacements that take place during this process, something that renders the tracking across generations difficult.

In addition, the algorithm cannot easily distinguish polyploid cells with more than one nuclei. This can be changed and improved if the direction of the detected nuclei is considered.

3.4.1 Cell cycle attributes

Figure 3.11 is presenting the progression of GFP intensity that is calculated by using a circle mask around the nucleus. Manual extraction of the GFP intensity of the whole cell across time (figure 2.2) indicated that the intensity should increase whith the progression of cell cycle and drops immediately after the mitosis. The figure demonstrates similar pattern, where intensity increases up to division and drops after it. Therefore, the

extraction of cell cycle information can be satisfactorily achieved without proper cell segmentation. These curves can be used to predict the cell cycle phase of the cell, for comparison with the same curves under drug treatment, where the cell cycle dynamics can be significantly changed.

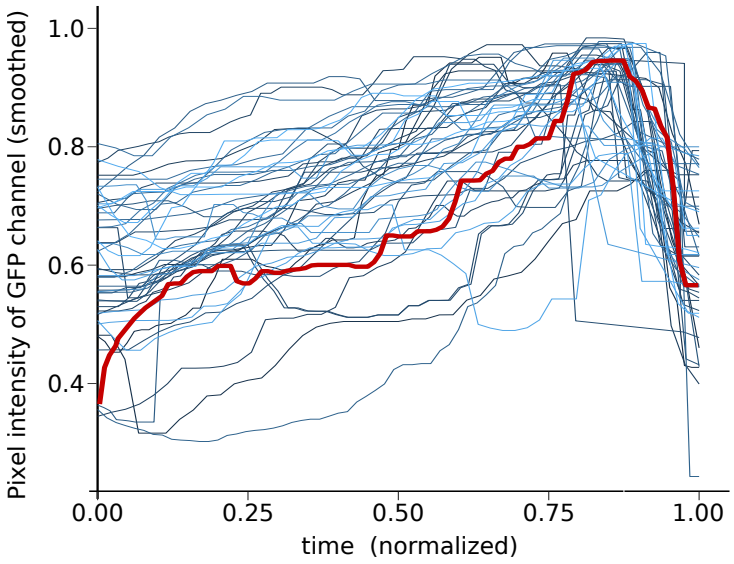

Figure 3.11: Normalized GFP intensity. Due to differences in the life duration of cells, time is also normalized to take values between 0 and 1. It is apparent that the GFP signal could be modeled as a function that increased monotonically and can has a maximum at mitosis.

3.5 Applications

3.5.1 Movement analysis

Figure 3.12 shows the migration routes, between two consecutive division events, of the detected cells. The paths show some different characteristics suggesting two different types of motion; one that is more persistent (red circles) and one that is less (green circles). Regarding first type, cells appear to depart far from their initial position. Regarding the second, cells remain around this initial position throughout the whole time.

The cell paths differ in their length due to the differences in the division times. However, pathlengths should be comparable because for most of them division is taking place after 20 ± 5 hours, as can be seen from the distribution of figure 3.13 a. The cells' total distance traveled is strongly correlated with their intermitotic times (IMT), i.e. their lifespans (figure 3.13 b). This result suggests that the cells with longer lifespans tend to travel longer distances. However, this does not explain the differences in the persistence.

The total distance traveled is shown in figure 3.14. The histogram appears to have a right fat tail. The vast majority of cells travel a distance of $\sim 300 pixels$, while some

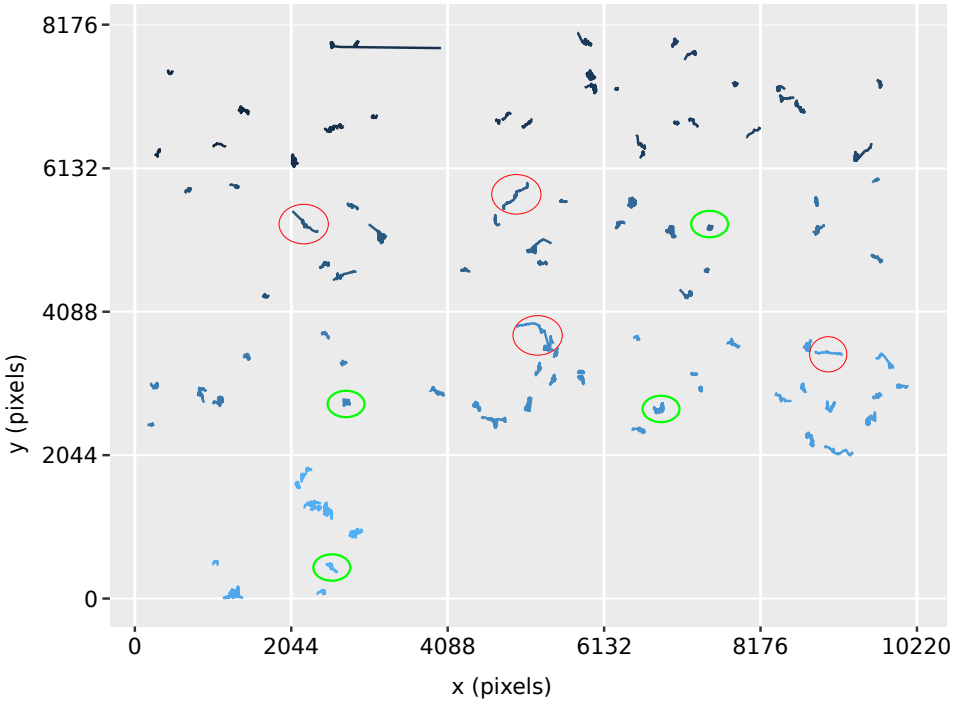

Figure 3.12: Migration routes of the cells. Red cicle indicated cells that show some persistence in their walk while green circles show cells that appear to have difussive − like movement behaviour.

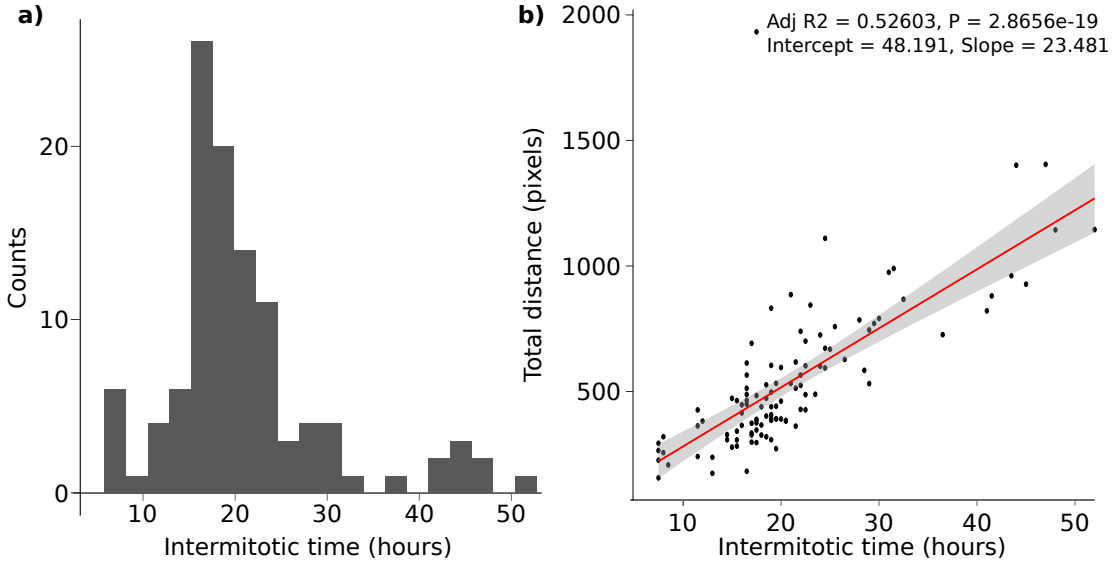

Figure 3.13: a) Histogram of the Intermitotic times. The histogram is peaked around the expected intermitotic time of 21-22 hours and is right-fat tailed. The shape of the distribution is in accordance with the shape of the respective distriburion from the groundtruth data (see chapter 2). b) Scatterplot of the Intermitotic times against the total distance traveled and the fitted linear model showing a positive correlation with $slope = 23.481$.

travel longer distances. To be able to completely understand the reason for this different movement behaviour it is useful to plot also the step length of the whole population (3.14

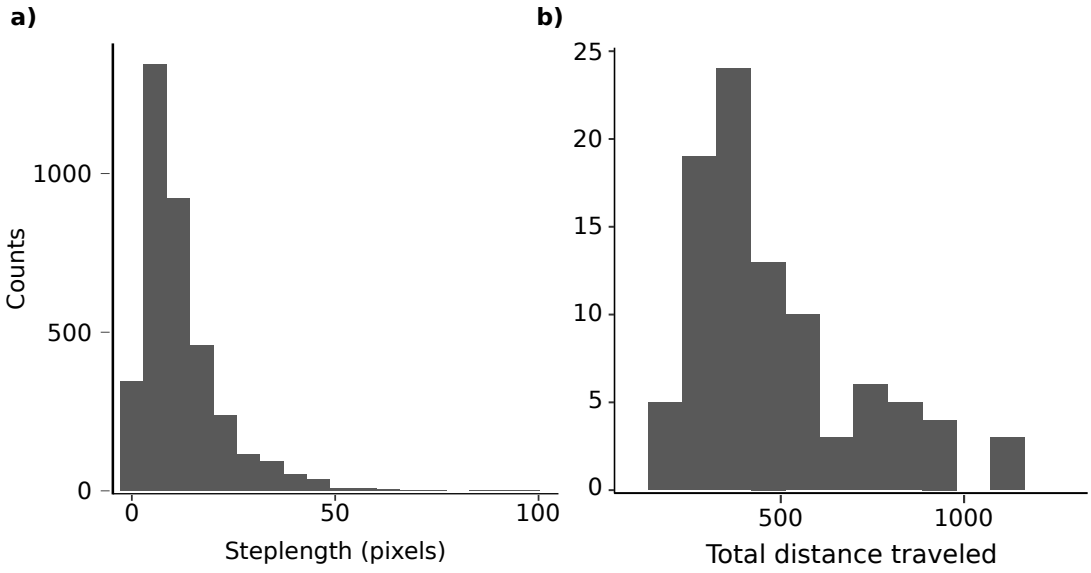

Figure 3.14: Empirical histogram of the steplengths of the cells is shown on the left and the empirical histogram of the total distance traveled is shown on the right. Both distributions appear to be asymmetric with the distribution of the total distance traveled to have a right-fat tail.

left) and the steplength distribution of each cell (figure 3.15). The steplengths of the whole population (figure 3.14 left) are distributed mainly around a value of $10-15 pixels$. However, the distribution has a heavy tail, suggesting that cells can also take some larger steplengths. The question is if these steplength are randomly distributed across the cells or if there are specific cells that siginificantly differ from others and move on average taking larger steps. Figure 3.15 can provide the answer to this question. From the boxplots of the steplengths of each cell it is obvious that there are differences between different cells and therefore the steplength must be closely linked to the cell ID and could be related also to the different types of motion shown in figure 3.12.

Effect of cell density on movement

U2OS cells form colonies and the cell densities can be high. Therefore, understanding how density and cell position in the colony affects the cell trajectory is necessary in order to understand if there are inter-individual differences in cells' movement behaviour.

For this reason, the relative position of each cell to the area of the colony it belongs to should be quantified. Then it should be tested for the existence of any correlation with movement characteristics, such as the steplength and the total distance traveled. If there is some correlation between the position in the colony and the movement characteristics, this would suggest that cell movement is affected by colony properties and is determined by the spatial arrangement. For example, could be that cells in the periphery have more freedom to move towards a direction than cells inside the colony that have a restricted

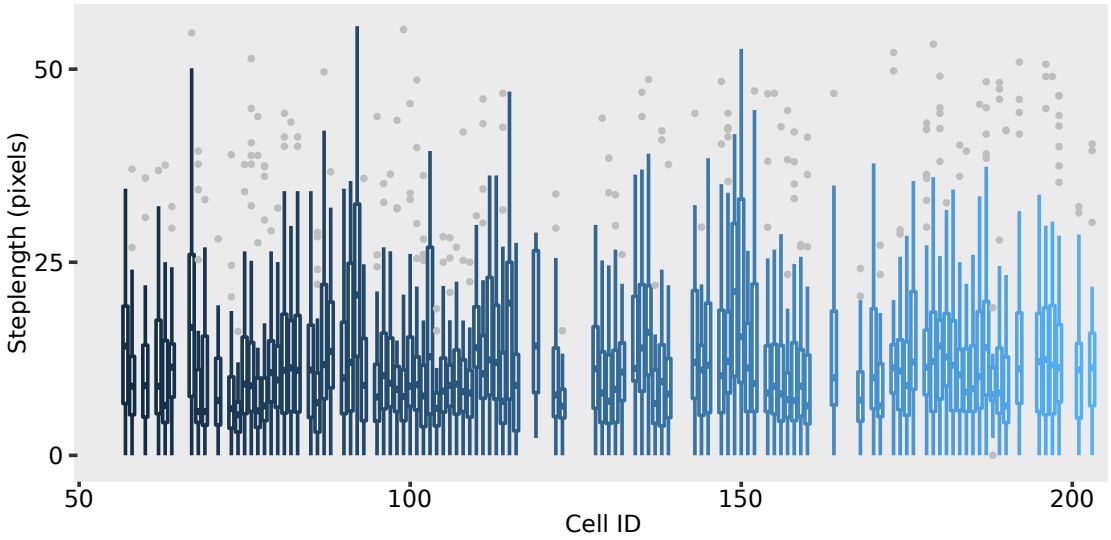

Figure 3.15: Boxplots of the steplength for the different cells.

movement because of their neighbouring cells.

Two colony metrics were calculated for each cell: the closest distance of the cell to the edge of the colony it belongs to (d_{ce}) and the local density of the cell in the colony (d). The last was identified as the ratio of the number of black pixels over the number of the white pixels $d = \frac{N_{dp}}{N_{wp}}$. The area used for calculating the relative coverage is chosen to be the circural area of radius $r = 100$ pixels centred at the centre of targeted cell. Finally, the area of the colony the cell belongs to was also measured.

The colony metrics were tested against the total distance traveled (figure 3.16 b). Because, as I showed above, the total distance traveled is correlated with the cell's lifespan I also plotted the colony metrics against the distance each cell had from its initial position (figure 3.16 a), because this metric is not correlated with the cell's lifespan.

None of the colony metrics was correlated with the total distance traveled and the distance a cell had from its initial point. This means, that at least locally the cell movement does not differ because of the position it has on the colony.

To test if the Human Osteosarcoma cells show on average a diffusive or non-diffusive motion, the Mean squared displacement for each time step was calculated. To get relevant values for the mean squared displacement I had to take into account the difference in paths' timelengths. Therefore, time was normalized between the values 0 and 1, $0 < t < 1$. To take squared displacement averages, the normalized time was sub-divided into 20 intervals. For each interval I calculated the mean squared displacement.

Mean squared displacement was plotted (figure 3.17), together with the respective error bars, against the time intervals and a power function was fitted on the data. The exponent was found to be 0.9406 with a 95% confidence interval $CI = (0.7955, 1.086)$. The value of the exponent indicates that there is an anomalous sub-diffusion, i.e. cells

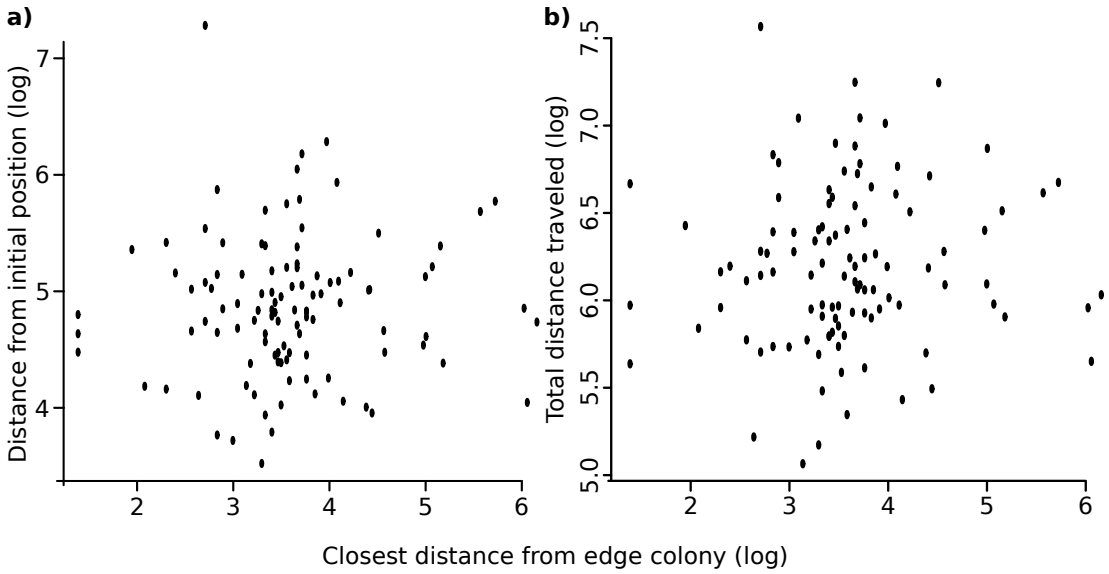

Figure 3.16: Scatterplot of the closest distance of focal cell from the colony edge against a) the final distance of the cell from initial position and b) against the total distance traveled. Both axes are in logarithmic scale.

move slightly slower than how they would have moved if they would follow normal diffusion. This finding suggests that there is a cell crowding effect (Weiss *et al.*, 2004). The value however is close to 1, so normal diffusion processes must hold at least for the initial periods. This is enhanced by the behaviour of the MSD line (figure 3.17), as it appears to be a straight line for the first three to four time points.

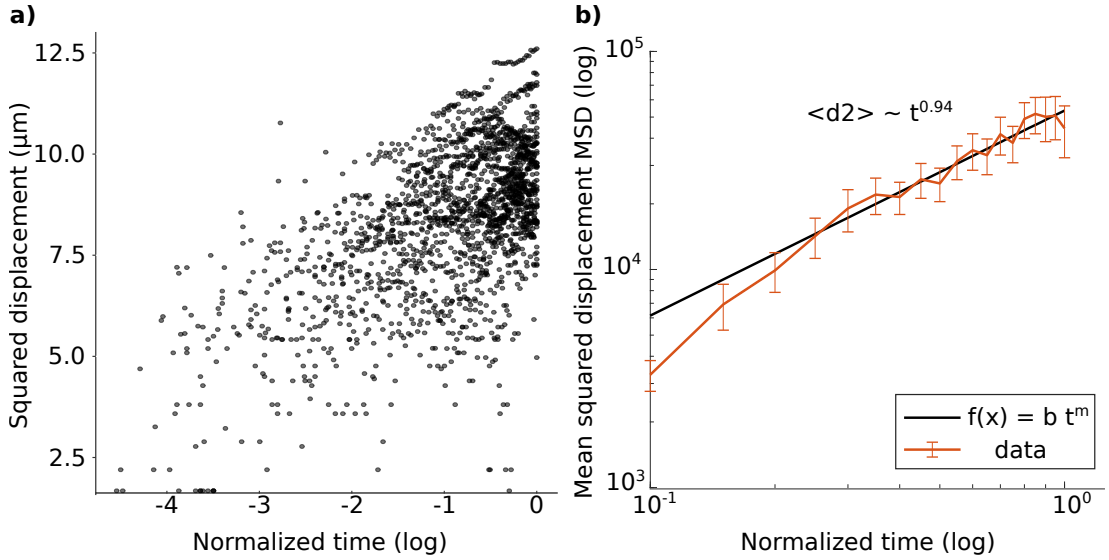

Figure 3.17: Normalized time against Squared displacement for different cells and Mean Squared Displacement (MSD).

In addition to the cell cycle information from the GFP intensity progression, I can extract the intermitotic times of the cell population, as it is shown in figure 3.13. The

histogram is similar with the histogram of Intermitotic times that was extracted manually, suggesting, that the algorithm is working quite well, despite errors on mitosis detection and thus can be used to also extract this type of information.

3.6 Summary

This Chapter presented the first attempt to build an automated long term nucleus tracking routine. In addition, it demonstrated some applications of it for the better understanding of the movement and cell cycle behaviour of Human Osteosarcoma cells in the population.

The advantages of this routine is that it can track cells by following very simple rules, a routine that does not require strong computational power. It is a much preferable way to extrat cell routes from the alternative manual cell tracking. More elaborated tracking methodologies that incorporate Machine Learning and Artificial Inteligence techniques would probably increase the success of the tracking, however they will require a lot of computational resources and expertise.

The capabilities of the routine is to track the nuclei from one mitotic event to the other and to extract meaningful information of the Cell cycle phase without having to totally segment the cell outline. They can also extract lineages, but the success rate of this is lower.

Regarding movement analysis, the current study has for the first time demonstrated at least two different kinds of movement of Human Osteosarcoma cells. From figure 3.12 at least two different movement patterns are apparent, one more diffusive and one more directed. Therefore, the distribution seen in figure 3.14 must be a combination of two different distributions, one Gaussian and one Levy-like.

There is a crowding effect, which means that cells restrict their motion due to the high densities. However, when cell's colony position was tested as a predictor for the distance traveled and the steplength, no correlation was found. This suggests that all cells irrespective of their position in the colony are affected by the high densities, and there is no distinction between cells in different position. The high densities globally affect the cells' movement behaviour.

For a more in-depth discussiosn please refer to section 6.2.

Chapter 4

From disorder to order: Emergence and repeatability of coordinated motion in stickleback fish

The current work studied the emergence of coordination in a group of unfamiliar free-swimming stickleback fish (*Gasterosteus aculeatus*) in the laboratory and aimed to demonstrate the existence of two different phases, the uncoordinated and the coordinated. It also studied the effect of the repeated interactions on the characteristics of the emergence. Finally, it investigated if the initial configuration of the experiment, but also some individual traits such as sex and cortisol levels could predict the attributes of the coordinated state.

4.1 Introduction

How and why complex and coherent behaviours emerge and self organize in systems with large numbers of interacting units has attracted much interest from scientists (Camazine *et al.*, 2003; Nicolis & Prigogine, 1977; Sumpter & Sumpter, 2006). The observation that collective behaviours can arise from simple, local interactions of the constituent parts (Camazine *et al.*, 2003; Clark & Evans, 1954; Herbert-Read *et al.*, 2011; Deneubourg & Goss, 1989; Katz *et al.*, 2011; Parrish & Edelstein-Keshet, 1999b) led scientists to look for universal rules and/or meaningful analogies across different systems; and none more so than studies of animal groups, which typically examine the collective motion and behaviour of bird flocks, fish shoals and insect aggregations (Buhl *et al.*, 2006; Gelblum *et al.*, 2015; Herbert-Read *et al.*, 2011; Nagy *et al.*, 2010; Watts *et al.*, 2016; Viscido *et al.*, 2004).

Fish have attracted most of the attention in the study of coordinated motion, because of the variable complex patterns they display (Hemelrijk & Hildenbrandt, 2012;

Herbert-Read *et al.*, 2011; Katz *et al.*, 2011; Marras *et al.*, 2015; Viscido *et al.*, 2004). Both global and individual-level behavioural observations are necessary for the deeper understanding of the phenomenon (King *et al.*, 2018). Macroscopically, Tunstrøm *et al.* (2013) experimentally showed that coordinated fish motion can go through transitions and can exhibit each of the three states: swarming, milling and polarized. Microscopically, using experimental and theoretical approaches to understand the mechanisms of collective behaviour, it has been suggested that coordination in fish (and other taxa) can be explained by three simple local interaction rules: attraction, repulsion and alignment (Couzin *et al.*, 2002; Czirók & Vicsek, 2000; Parr, 1927). Nevertheless, more simplified models have shown that the same group-level properties can be explained with fewer rules (for example only local attraction) (Strömbom, 2011).

Some of the proposed rules have been experimentally confirmed (Herbert-Read *et al.*, 2011; Katz *et al.*, 2011; Schaerf *et al.*, 2016) but at the same time, empirical work demonstrates asymmetries in the interactions and how these scale to macroscopic behaviour. For example, experiments in the lab have demonstrated that individual fish are attracted more to neighbours in front than those behind them (Herbert-Read *et al.*, 2011; Katz *et al.*, 2011) and suggest that differences in speed and acceleration can also play a role in the group level dynamics, leading to the emergence of leadership, i.e. the initiation of new directions of locomotion by one or more individuals which are then followed by other group members (Krause *et al.*, 2000).

Other studies have shown/suggested that variation in individual traits (such as metabolic, personality and information knowledge) also influence the macroscopic behaviours (Bazazi *et al.*, 2011; Gelblum *et al.*, 2015; Jolles *et al.*, 2015; Lord *et al.*, 2016; Nakayama *et al.*, 2016; Watts *et al.*, 2016). More specifically, empirical studies have demonstrated that certain individuals can act as leaders due to inter-individual differences in morphology, state (Briard *et al.*, 2015; Krause *et al.*, 2000), experience (Eskridge & Schlupp, 2014), information (Andrieu *et al.*, 2016; Watts *et al.*, 2016), and/or personality (Jiang *et al.*, 2017; Johnstone & Manica, 2011; Krause *et al.*, 2000). For example, according to Krause *et al.* (2000), well fed individuals (i.e. an indicator of high energy status), do not need immediate food to survive and thus prefer to remain close to the shoal in order to seek protection and increase their survival probability. Fish with low energy status, in contrast,take greater risks in order to access energy sources (Krause *et al.*, 2000). Thus, low energy fish can act as initiators of movement in a new direction and, if followed by conspecifics, they would then lead the motion at least temporally (Krause *et al.*, 2000; Rands, 2011). Metabolic rate variation, in the context of refuge use and body lengths, is another possible trigger of leader/follower dynamics (Krause *et al.*, 2000). For example, larger fish were found to have higher swimming speeds than smaller conspecifics which is likely to explain their leading position in moving shoals (Krause *et al.*, 1998).

Also, bolder fish are more likely to initiate a movement event than shy fish and potentially become leaders according to Nakayama *et al.* (2016) and Wang *et al.* (2017). The coexistence of different personality types has been associated with better performance for specific ratios of bold and shy individuals (Eskridge *et al.*, 2015) indicating that bold fish would act as leaders. Decision making in fish shoals during directed motion is often attributed to the different behaviour of bold individuals that act as leaders, as many studies show (Conradt, 2012; Ioannou *et al.*, 2015; Rosenthal *et al.*, 2015). These inter-individual differences and resultant leader-follower dynamics are also predicted by game-theoretical models of collective action (Johnstone & Manica, 2011).

Finally, leadership can also emerge in homogeneous populations, i.e. in populations that individuals do not differ in personality initially, if knowledge on some useful information related to the target of the motion (i.e. information on food or predator, migration route etc) is acquired (Buhl *et al.*, 2006; Eskridge *et al.*, 2015; Ioannou *et al.*, 2015; Krause *et al.*, 2000).

Previous studies attempted to understand coordination in fish shoals have omitted two important aspects. Firstly, the majority investigate the underlying mechanisms that are responsible for the coordination but for a short-term period. In other words, studies on the underlying rules of fish coordination take into account only a snapshot of the process (Biro & Sasaki, 2016). They study which interactions take place during coordinated motion under specific environmental conditions, such as predation (Herbert-Read *et al.*, 2017; Croft *et al.*, 2003), but do not test if these interactions change over the course of repeated similar tasks (Herbert-Read *et al.*, 2011; Katz *et al.*, 2011). Few theoretical works, such as the agent based model of Quera *et al.* (2016), have accounted for a memory attribute of their agents to incorporate social biases due to previous repetitive interactions. But in experimental works, the interaction properties extracted from a single experiment are assumed to remain constant over repeated coordination events (Herbert-Read, 2016). In the work of Jolles *et al.* (2018) there was a focus on repeatable inter-indivitual differences across different environmental setups, but did not focus on how repeated interactions affect the emergence of coordination and the dynamics of collective movement. In real environments, fish tend to form groups with the same individuals more than once in their lifetime (Biro & Sasaki, 2016) and they have to forage or protect themselves repeatedly (Herbert-Read *et al.*, 2017; Ioannou *et al.*, 2017). Recent works highlight the importance and usefulness of focusing on the time-depth perspective to understand the stability and or adaptation of the hierarchical dynamics in time (Biro & Sasaki, 2016).

Secondly, previous studies trying to find mechanistic and evolutionary explanations that can explain the phenomenon of coordinated motion, typically investigate systems in which coordination has already been established (Tunstrøm *et al.*, 2013). They also try to describe and understand mechanistically the transitions seen between different

collective states, such as between milling and polarized states(Tunstrøm *et al.*, 2013). However, previous studies neglect to study the transitions seen before any coordination is established, i.e. when the system is in a disordered state and establishes for the first time the interactions that will determine later dynamics (Buhl *et al.*, 2006; Dyson *et al.*, 2015; Murakami *et al.*, 2017). Many realistic scenarios can lead individuals within an established group to a disordered state, where the network of interactions and hierarchies is reset (Kelley *et al.*, 2011; Merkle *et al.*, 2015). For example, a sudden predator attack can spread individuals to new locations or environments with unfamiliar conspecifics, resulting in a global disorder in terms of the interactions and collective formations (Herbert-Read *et al.*, 2017). Information conflict can divide groups and interrupt the network of interactions (Merkle *et al.*, 2015). Also, human activities such as fishing can disrupt the stability of shoals and remove members of them. Where individuals start (or re-start) from a random and disordered configuration, where they cannot choose their neighbour or their environment, and have to re-establish their social interactions is thus a common scenario for gregarious fish species (Borner *et al.*, 2015; Nadler *et al.*, 2016). Less extreme, but equally relevant is the behaviour of individuals in fission-fusion animal systems which, in some cases, results in individuals joining and leaving groups and interacting with unknown or rarely encountered individuals at high frequency (Kelley *et al.*, 2011).

The study of this initial and first transition to a coordinated state can therefore provide valuable information on the innate mechanisms by which individuals establish their network of interactions. For example, the position of the fish at the disordered state and the personality differences could determine which individuals will take the front positions and influence others in later dynamics, or if the individuals will show preferential attachment to the fish that were their neighbours at the beginning. An alternative scenario would be that the dynamics during the ordered state (the coordination) would be unaffected by the initial conditions, and that the dynamics of the coordinated motion would be determined by individual personalities and/or environmental cues only.

4.1.1 Current Study

Preliminary observations of stickleback fish (*Gasterosteus aculeatus*) shoals in behavioural experiments indicated that fish 'take a finite time' to become coordinated and start shoaling (Fürtbauer and King, personal communication). I therefore aimed to investigate the existence of distinct phases with respect to coordinated motion and indicative shoaling behaviour – the disordered and the ordered – and to determine the time at which transitions between the two phases take place. I also investigated if leader-follower relations are formed, and if so, if they are consistent across the uncoordinated and coordinated states. Furthermore, I tested if other factors including, i) the initial

configuration at the disordered state (i.e. at the start of an open field test), ii) personality types (activity level), or iii) individual attributes (like sex and cortisol levels prior to experiments) shape the later dynamics of coordinated motion in groups of stickleback fish (*Gasterosteus aculeatus*). Finally, to bring a time-depth perspective, I tested whether the patterns I uncovered were repeatable across two trials.

More specifically, the questions to be targeted in the current study are the following: 1) Are the initial detected activity levels of the fish affected by their physiological characteristics (i.e. the cortisol levels, the weight and the sex of the fish)? Is the initial fish activity level affected by the repeatability of the experiment? Is it affected by the neighbours of each fish? 2) Is there any effect of the repeatability of the experiment on the frequency a fish acts as a leader to others? 3) Is the time at which coordination establishes predicted by the different trials? 4) Do fish that tend to be close to the same fish start coordinating faster than other fish?

The work is divided into the following sections. First, I discuss the methodological pipeline followed, i.e. the experimental setup, the tracking algorithm and the statistical tools used for the analyses. Second, I present the results that are split into three subsections, the initial configuration, the link between initial and post behavioural transition and the time-depth perspective results. Third, I discuss the results and provide relevant questions that could be targeted in future studies.

4.2 Methodology

The experiments were designed and run by dr. Ines Fürtbauer at Swansea University.

4.2.1 Study Subjects

$N = 30$ stickleback fish randomly grouped into six shoals of $n = 5$ individuals were tested twice (trial 1 and 2). In the morning fish were tagged individually using disc-shaped tags (Yellow, Red, Blue, Black and White, see figure 4.1), mounted upon the dorsal spines of the fish (Jolles *et al.*, 2018; Webster & Laland, 2009) and were returned to their individual tanks to habituate to the tags. In the afternoon, social experiments were conducted as follows: five fish were placed in a transparent "starting box" (consisting of five individual compartments) inside the rectangular test arena ($42.5x73cm$; lined with white silica sand) to acclimatise for five minutes. The starting box was then removed and the fish were videotaped for $20minutes$. On the following day, social experiments were conducted at the same time but in reverse order.

4.2.2 Video recordings

Groups of fish were placed in a rectangular tank ($730mm \times 425mm$) filled with water (5 cm depth) in “starting boxes" which consisted of a row of separated transparent divides (figures 4.1 and 4.3). Fish were filmed by a Panasonic HDC-SD60 HD video camera (Panasonic Corporation of North America, Secaucus, NJ, USA) positioned 2 meters above the tank. The tank was surrounded by an aluminum frame and white screen (PhotoSEL BK13CW White Screen) enabling optimum conditions for video recording. The depth of the water was such so that the fish were constrained to move in two dimensions. The camera captured with a rate of $50fps$. During the first minutes the camera was recording the fish in their boxes. After the boxes were removed, the camera recorded fish behaviour for further minutes ($\simeq$ 20 mins).

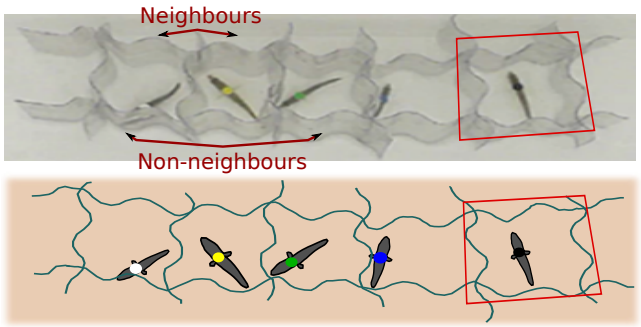

Figure 4.1: Actual (top) and schematic (bottom) representation of the initial setup of the positions of the fish.They are located inside adjacent semi-transparent boxes. Red arrows indicate the neighbouring relationships. Fish in the middle have two neighbours one at each side. Fish at the edges have only one neighbour. The red rectangles show an example of the outline of the box that is used to calculate the fish activity in that box.

4.2.3 Video analysis and Tracking

Individual activity in the box

Fish activity in the starting boxes was measured prior to being released to free-swimming states. To quantify the box activity levels I developed an algorithm using the OPENCV library in C++ (Itseez, 2015). The algorithm works as follows: First, I draw the outline of the region each fish is located (see figures 4.1 and 4.3) and automatically create a grid of points that covers all the area of the box (figure 4.3 right). Second, I apply optical flow algorithm to detect motion/activity at each of the points.

Optical flow is the algorithm that tracks specific features (points) in an image across multiple frames (the documentation follows, from opencv website) (Burton & Radford, 1978). It works on several assumptions: that the pixel intensities of an object do not change between consecutive frames and that the neighbouring pixels have similar motion. Consider a pixel $I(x,y,t)$ in the first frame. It moves by distance (dx,dy) in the next

frame taken after dt time. So since those pixels are the same and intensity does not change, I can say, $I(x,y,t) = I(x+dx, y+dy, t+dt)$. If I take taylor series approximation of right-hand side, remove common terms and divide by dt I get the following equation:

$$f_x u + f_y v + f_t = 0, \tag{4.1}$$
$$where\, f_x = \frac{\partial f}{\partial x}, f_y = \frac{\partial f}{\partial y}\, and\, u = \frac{dx}{dt}, v = \frac{dy}{dt}$$

The above equation is called Optical Flow equation (Lucas & Kanade, 1981). I seek to find image gradients in the x and y directions, f_x and f_y , respectively. Similarly, f_t is the gradient along time. But (u,v) is unknown. I cannot solve this one equation with two unknown variables. So several methods are provided to solve this problem and one of them is Lucas-Kanade.

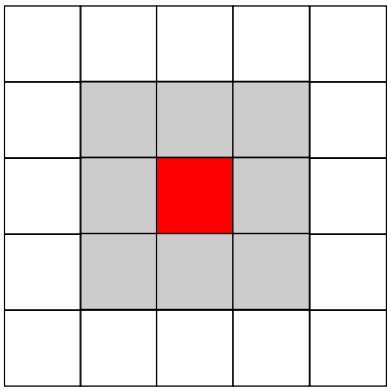

Figure 4.2: A schematic representation of the 3x3 patch of an image. The blocks indicate the pixels of the image. The grey and red blocks indicate the 3x3 patch. The red block is the focal pixel and the grey blocks are the neighbouring pixels (accounting for 8-neighbours).

Lucas-Kanade method takes a 3x3 image patch around the point (figure 4.2). Because of the initial assumptions I can treat the 9 points as following the same motion. I can find (f_x, f_y, f_t) for these 9 points. So now our problem becomes solving 9 equations with two unknown variables which is an over-determined problem. A better solution is obtained with least square fit method. Below is the final solution which is two equations-two unknown problem and solve to get the solution.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum_i f_{xi}^2 & \sum_i f_{xi} f_{yi} \\ \sum_i f_{xi} f_{yi} & \sum_i f_{yi}^2 \end{bmatrix}^{-1} \begin{bmatrix} -\sum_i f_{xi} f_{ti} \\ -\sum_i f_{yi} f_{ti} \end{bmatrix} \tag{4.2}$$

Therefore, using optical flow for the grid points, I count the times each grid point in each box has been moved by the motion of the fish. For the final step, I sum the activated

times of all points to have a total activity index. To reduce redundancy, I also record activity if at least one point is activated. In addition to the two metrics of activity level I create a matrix of the grid points in the box, where each matrix element, corresponding to each point, is filled with $0s$ (no activity) and $1s$ (activity) for each time frame.

To compare activities between groups, each fish activity value was divided by the sum of the activity level of all the fish of its group (normalized box activity, α^i_{tr})

$$\alpha^i_{trnorm} = \frac{\alpha^i_{tr}}{\frac{1}{N}\sum_i^N \alpha^i_{tr}} \tag{4.3}$$

where α^i_{tr} is the box activity of fish i in trial tr .

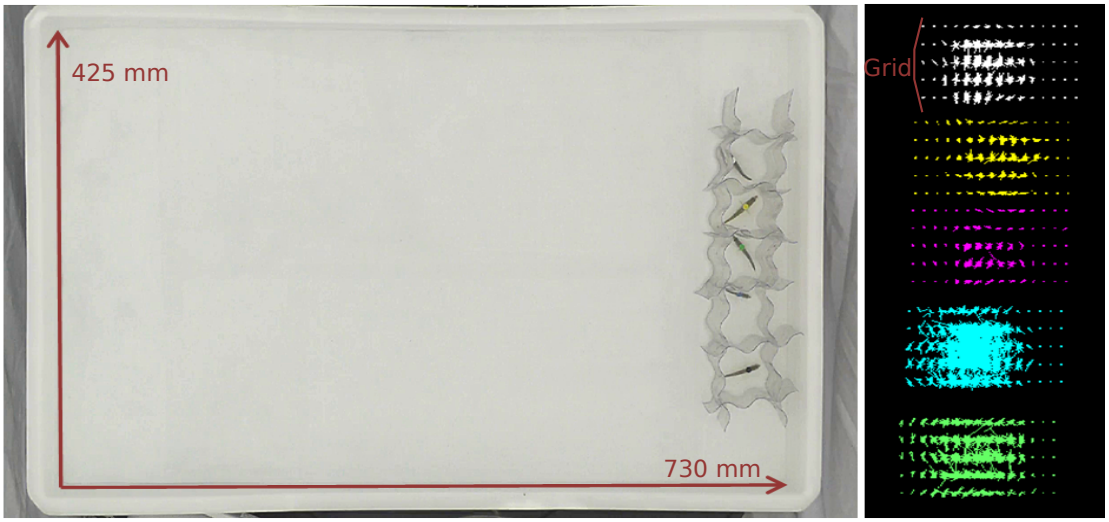

Figure 4.3: The experimental setup (left) where the initial positions of the fish and the size of the tank ($730mm$, $425mm$)are provided. On the right, a schematic representation of the grid points that were used to detect fish activity levels. The fish with the green disk seems to be the least active and in contrast, the fish with the blue disk appears to have activated the grid point more times than its conspecifics.

Tracking algorithm

Segmentation

To extract the trajectories of each fish, I developed a tracking algorithm based on OPENCV in C++. Each fish was tagged with a specific colour. The segmentation of each fish was based on their colour differences. The colours used were black, yellow, green, blue and white. Because the background was white and the illumination conditions were constant, segmenting the black, yellow, green and blue colours was efficient. However, the white tag could not be segmented, but a different methodology used to detect that fish and will be presented in the following paragraph. I worked on the HSV (hue, saturation,

value) color space (an alternative representation of the **R**ed **B**lue **G**reen colour space) to determine the range each color tag has in the video. HSV color space, like RGB, also consists of 3 matrices, HUE, SATURATION and VALUE. HUE represents the color, SATURATION represents the amount to which that respective color is mixed with white and VALUE represents the amount to which that respective color is mixed with black. In OPENCV, value range for HUE, SATURATION and VALUE are respectively $0-179$, $0-255$ and $0-255$. I decided to use HSV colour space, because I know the colour values for each tag (i.e. Hue values) should not vary much, but the Saturation and Value would change more because of changes in water due to the movement of fish or the exact position of the fish in the tank.

After testing for different value ranges of the HSV channels I concluded that the ranges that best segment the colourtags are the following:

Table 4.1: Range of the HSV values used to segment the coloured discs. H stands for Hue values, i.e. the values that characterize the chroma/colour of each tag. Except for the black colour, the range value for this parameter is less variable for the other colours (around $10, 30$ or 90 units). S stands for saturation and V stand for Value. These two parameter values vary a lot (around 200 units).

Fish tags	H		S		V	
	min	max	min	max	min	max
Yellow	19	33	40	255	92	255
Green	43	74	50	255	72	255
Blue	76	179	43	252	30	162
Black	0	179	0	255	0	49
All fish	0	179	0	255	0	120

To segment the fish with the white tag I applied the following:

I created a mask of all the five fish by segmenting the image using Otsu's threshold (left figure 4.4). I applied a $bitwise_{and}$ operation (a.k.a intersection) between the segmented image of each of the colour tags (middle left figure 4.4) and the segmented image of the five fish to extract the mask of the four fish (middle right figure 4.4). Finally, I used the last image as a mask to get only the objects that correspond to the white fish (right figure 4.4). I grouped the remaining contours to get a solid object that corresponds to the white targeted fish.

Association- Coordinates

Following the above procedure, an image with 5 segmented objects is obtained, where each object corresponds to one of the fish. The fish tracks are the collection of the

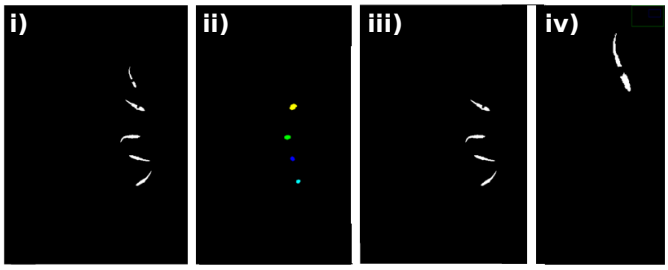

Figure 4.4: Image processing steps using OPENCV to extract the fish with white tag. i) Initial segmentation using Otsu's threshold. ii) The segmented coloured tags, after using simple threshold on the HSV space. iii) The result of the intersection of the two previous thresholded images. iv) Final result of the segmentation of the fish with the white tag.

positions of the segmented object that corresponds to a specific tag. In case of overlapping fish at a time frame t the position of the fish is left empty. Because of the high time resolution of the camera (50 fps) the fish do not show big or unexpected displacements and thus, it is possible to estimate the missing values. For this reaon, once the routine terminates I interpolate the missing values (figure 4.5).

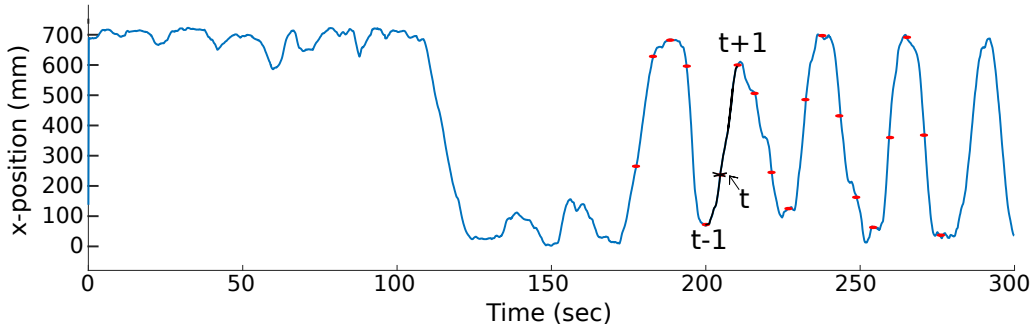

Figure 4.5: The missing entries are calculated by interpolating the missing value, i.e. the missing point will be the average value between the position at time $t-1$ and time $t+1$.

To convert the pixel coordinates to physical quantities I drew two lines that indicated the tank height and width in pixels (see figure 4.3). Due to the 2 meter distance of the camera to the tank and the small size of the tank I assume there is no significant distortion of the coordinate system. The dimensions of the tank are $730 \times 425\,mm$. The origin of our new coordinate system is the bottom left corner of the tank. For each pixel coordinate (x_p, y_p) the physical coordinates are

$$(x, y) = \left(x_p \times \frac{730}{width}, y_p \times \frac{425}{height} \right) \tag{4.4}$$

4.2.4 General Analysis

Global Group Characteristics - Wavelet analysis

The Continuous Wavelet Transform (CWT) is used to decompose a signal into wavelets (Daubechies, 1990). Wavelets are small oscillations that are highly localized in time. While the Fourier Transform decomposes a signal into infinite length sines and cosines, effectively losing all time-localization information, the CWT's basis functions are scaled and shifted versions of the time-localized mother wavelet (Ding *et al.*, 2007). The CWT is used to construct a time-frequency representation of a signal that offers very good time and frequency localization.

The CWT is an excellent tool for mapping the changing properties of non-stationary signals. The CWT is also an ideal tool for determining whether or not a signal is stationary in a global sense. When a signal is judged non-stationary, the CWT can be used to identify stationary sections of the data stream.

In all the experiments the oscillatory movement of the fish was apparent. The matlab function cwt was used to decompose the x- or y- coordinate signal into frequencies. The mother wavelet used for the current analysis was the analytic Morlet (Gabor) wavelet. The wavelet is defined as a constant κ_σ subtracted from a plane wave and then localised by a Gaussian window and its expression is 4.2.4

$$\Psi_\sigma(t) = c_\sigma \pi^{-\frac{1}{4}} e^{-\frac{1}{2}t^2} \left(e^{i\sigma t} - \kappa_\sigma\right) \tag{4.5}$$

where $\kappa_\sigma = e^{-\frac{1}{2}\sigma^2}$ and $c_\sigma = \left(1 + e^{-\sigma^2} - 2e^{-\frac{3}{4}\sigma^2}\right)^{-\frac{1}{2}}$

The wavelet analysis was chosen to confirm the oscillatory behaviour and also to provide the spectral characteristics of the coordinated motion.

Directional Correlation (CV)

To test if a fish is copying another fish, I calculated the directional correlation (CV) between each pair (Nagy *et al.*, 2010):

$$C_{ij}(\tau) = \left\langle \vec{v}_i(t) \cdot \vec{v}_j(t+\tau) \right\rangle \tag{4.6}$$

where $\vec{v}_i(t)$ is the normalized velocity of fish i, $\vec{v}_j(t)$ is the normalized velocity of fish j, and τ is the time delay. For a specific time point and delay time, the dot product calculates how much their directions are correlated after the time delay τ . If it is 1, fish' directions are exactly the same after a delay time τ, if it is 0, their directions are perpendicular after time τ. This value is averaged over a specific time window (t_w). For

this experiment, the selection of the value depended on the question, but it was between $3 < t_w < 5$ secs. Thus, I got CV values for different τ and for all the time period of the experiment. The maximum CV value (CV^*) was selected to represent each time point. The detected CV^* corresponds to a specific value τ. When $\tau > 0$, the focal fish i's direction is copied by fish j and, thus, the focal fish i leads fish j. Otherwise, when $\tau < 0$, the focal fish i copies j's direction, i.e. follows fish j. Therefore, the CV^* time series were used to study the phase transitions.

Frequency of leadership

To detect frequency of leadership of each fish or dyad, I calculated the number of times each focal fish i was leading over fish j , i.e. how many times the delay time τ corresponding to the maximum CV^* (τ^*) is positive. This frequency of leadership was calculated over the whole experimental phase of interest using non overlapping time windows (t_w) for the CV^* estimation. In case I wanted to check the frequency of leadership of each fish, then the average of the dyadic frequencies was calculated.

CV* threshold sensitivity analysis

The CV^* values can range from 0 to 1, where 0 shows no directional correlation at all and 1 shows absolute correlation of the direction. For example, to find the frequency of leadership in the experiment I have to find how many times one fish was copied against its pair (how many times the delay time τ^* at the maximum CV^* of focal fish was positive). But to find meaningful correlations, I should consider taking the frequency of leadership only at the times CV^* was over a threshold value. I assume, that for values of CV^* smaller than 0.4 the correlation is very small and thus there is no real leadership. To choose the appropriate CV^* threshold I run a sensitivity analysis. More specifically, I counted the number of available detected leadership events of each fish for different CV^* threshold values to see how these decrease as the threshold increases. In addition, I calculated each group's variance in the numbers leadership events for different threshold values.

Figure 4.6 shows how the number of available points of leadership events varies for different CV^* threshold values. It also shows the variance of points for the different CV^* thresholds for all fish. The CV^* threshold was chosen to be the value at which the number of available points is high enough, but also and the variance is maximum so that the resolution is not lost. This happened at a threshold around 0.5.

Change point detection

I was interested in detecting the time point of the transition and divide the experiment into two phases, the disordered and the ordered. To do this, I used the 'changepoint'

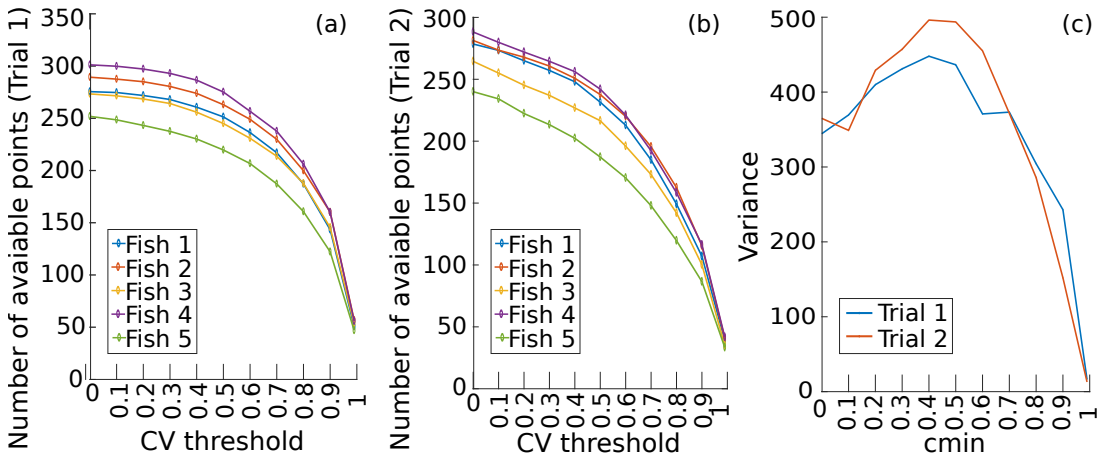

Figure 4.6: The available leadership points of each fish (indicated by different colours) when taking threshold values of the CV^* (a) for trial 1 and (b) trial 2. (c) the variance of the available points for the group for both trials (trial 1 in blue and trial 2 in red).

package in R/ or Matlab (R Core Team, 2013; Matlab team, 2017). This function detects change points in inputted time series (in this case the CV^* time series). It finds the optimal position of the change point according to a specified method. In this case, the selected method detected changes in the mean value of the signal, i.e. significant changes of the CV^* value. It returned the optimal time point at which the signal is divided into two segments of significantly different mean CV^* values.

To extract the CV^* time series I need to decide which is the appropriate t_w that can be used. Thus, I detected the change points for a range of t_w and checked how they are affected by the different t_w.

I could choose the mean change point for each group or the meadian change point to plot against the different t_w values. I chose the median instead of mean because I am interested in the detection of the emergence of group coordination. If most of the fish tend to have strong coordination but one remains distant for a long time, then the mean CV^* will be much affected by the low levels of coordination seen by one fish compared with the others. And thus it can obscure the existence of coordination.

Change point detection

From figure 4.7 below we can see that the different time windows do not significantly affect the changepoint estimation. I thus decided to use the time window of 4 secs, a good interval to detect correlations and also small enough to gain some computational time. The first step before analyzing them is to smooth the data without losing significant information.

In case the time series show a clear change in CV^* values the estimated change point is easily detectable. However, for some cases that the fish are coordinated from the early beginning and the CV^* values do not show a sharp change, or for some cases there is a

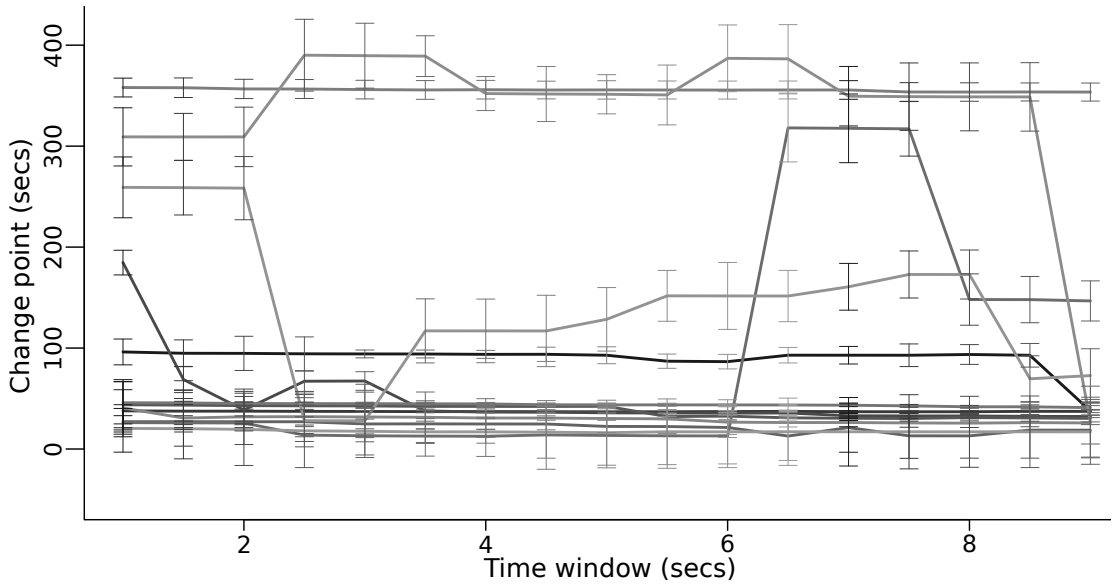

Figure 4.7: The detection of the changepoint of the CV^* for different time windows used to calculate the CV. The changepoint package in R was used to calculate the changepoints of the CV* values.

change at the very beginning of the time series and thus the method is unable to correctly find changepoints. In the second case, this can be improved by changing the length of the inputted signal, so that the detection of the change point is more accurate. The decision of the time series length in this case was decided by checking the sensitivity of the change point estimation for different signal lengths (figure 4.7). The signal length, was decided to be the one at which the median value didn't change significantly by changing the length, and at the same time the variance of the change point was small. I applied this method only in cases where the change point in the signal was locating very early in the signal. This was the case for Group 5 and 6 in trial 2. From the following figures (4.8) we can see that a signal length smaller than $150\,sec$ will give a more accurate change point.

Filtering

Despite the intrinsic noise of the data due to biological variability, there was also noise in the data, due to limitations on video acquisition (e.g. camera resolution) and tracking methodology. Therefore, some filtering methods were applied.

For the positional data, a Gaussian filter over 10 frames was used. For the CV* I used Stavinsky-Golay smoothing technique, which in fact is a low-pass filter. The parameters needed for this technique is the degree of the polynomial and the time window over which the polynomial will be fitted. By testing for different values of the degree I obtained figure 4.9. From there I decided to use a degree of polynomial that would smooth the data but would not give a very high distance between the smoothed data and

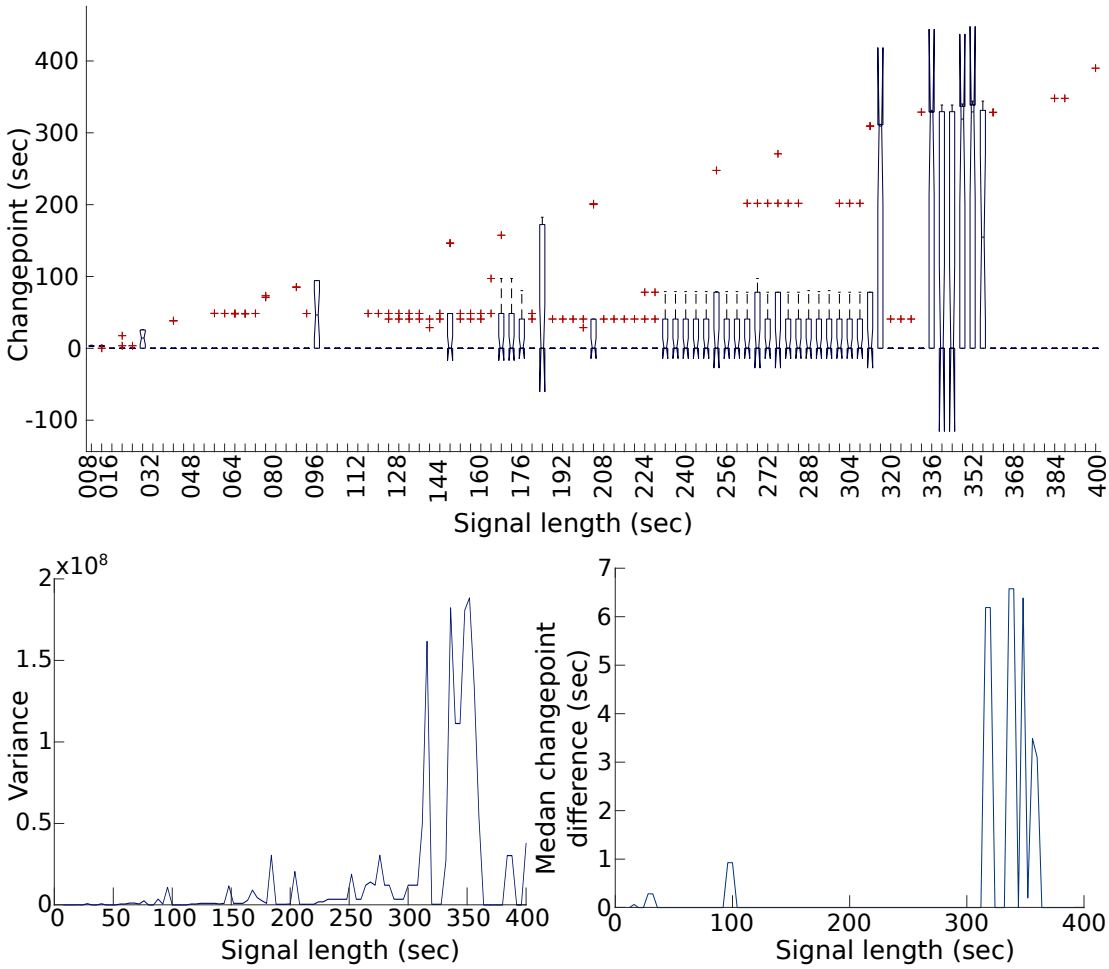

Figure 4.8: Top figure shows the influence of the signal length on the detection of changepoint. The boxplots indicate the different ranges of the changepoints. Bottom figures show the corresponding variance (left) and median difference (right) of the first figure. Changepoint detection is not varied significantly for small lengths.

the original. Thus, I decided to use a degree around 4 - 5. Also, I used a time window for the filter of 41 frames (i.e. 0.8 *secs*).

4.2.5 Statistical Analysis

Permutation – Box Activity

Figures 4.1 and 4.3 show the initial set-up of the experiment where the fish are located in adjacent semi-transparent boxes. Each fish dyad was characterized as neighbours or non-neighbours if the fish were located in adjacent boxes or not, respectively. The maximum number of neighbours a fish can have is two. Thus, when extracting neighbour and non-neighbouring data, a fish can be part of two dyads, resulting in lack of independence in the data (i.e. neighbours may influence each other's' activity). To test this, I used permutation test (custom made code in Matlab, Thompson & Shure (1995)) for differences in activity

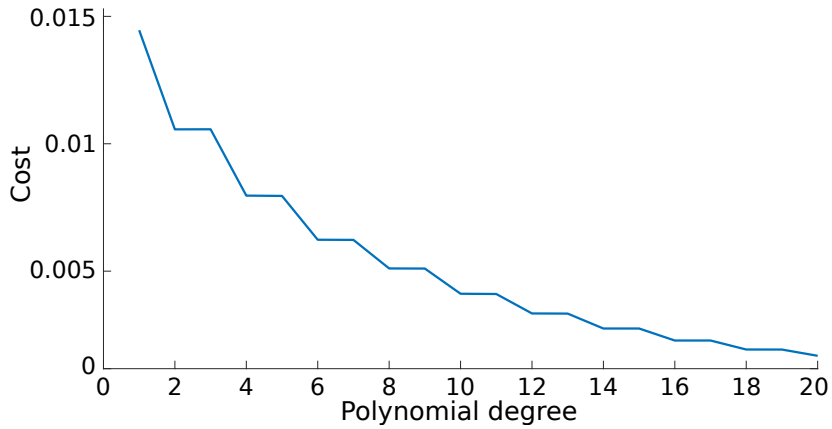

Figure 4.9: The cost of filtering the original CV^* data (i.e. the distance/difference between the original and filtered data) for different values of the polynomial degree parameter.

among neighbours and randomly chosen dyads.

Linear mixed effect models – Metrics

A General Linear Model is a statistical linear model of the form:

$$y_i = \beta_1 \chi_{1i} + \beta_2 \chi_{2i} + ... + \beta_p \chi_{pi} + \varepsilon_i$$
$$\varepsilon_i \sim NID(0, \sigma^2)$$

where $\beta_1, \beta_2, ..., \beta_p$ are the regression coefficients, and ε_i is the error term, with error variance σ^2. This model includes only one random variable to account for noise, i.e. the variation seen in y values that is not explained by the x values. In addition, it cannot account for variation between group observations.

The Mixed Effect models (or just mixed models) include additional random-effect terms, i.e. they can incorporate variation between groups of the sample data and can be expressed in matrix form as (Laird & Ware, 1982):

$$\mathbf{y_i} = \mathbf{X_i}\beta + \mathbf{Z}_i\mathbf{b}_i + \varepsilon_i \quad 4.7$$
$$\mathbf{b}_i \sim \mathbf{N}_q(0, \Psi) \quad 4.8$$
$$\varepsilon_i \sim \mathbf{N}_{ni}(0, \sigma^2\Lambda_i) \quad 4.9$$

where $\mathbf{y_i}$ is the $n_i \times 1$ response vector for observations in the i^{th} group. $\mathbf{X_i}$ is the $n_i \times p$ model matrix for the fixed effects for observations in group i. β is the $p \times 1$ vector

of fixed-effect coefficients. $\mathbf{Z}_i$ is the $n_i \times q$ model matrix for the random effects for observations in group i. $\mathbf{b_i}$ is the $q \times 1$ vector of random-effect coefficients for group i. ε_i is the $n_i \times 1$ vector of errors for observations in group i. Ψ is the $q \times q$ covariance matrix for the random effects. $\sigma^2\Lambda_i$ is the $n_i \times n_i$ covariance matrix for the errors in group i.

The parameters of the mixed model can be estimated using Maximum Likelihood Estimation (MLE) or Restricted Maximum Likelihood Estimation (RMLE).

Mixed Effects Models are seen as especially robust in the analysis of unbalanced data when compared to similar analyses done under the General Linear Model framework (with a restricted covariance structure). They are often appropriate for representing clustered, and therefore dependent data arising, for example, when data are collected hierarchically, when observations are taken on related individuals (such as siblings), or when data are gathered over time on the same individuals.

Here, I used linear mixed models (LMMs) in R (R Core Team, 2013; Kuznetsova *et al.*, 2017) to analyse behavioural patterns across the two different phases and across different trials. The data I am using (e.g. CV*, activity and hormonal) are correlated and show hierarchical structure because each fish can belong into different subgroups (according to trial or to the group it belongs to). Linear mixed models can take into account hierarchical structure of the data and they can also account for random effects, i.e. variation seen between groups or individuals. To allow for individual- and dyad-specific differences, random intercepts were fitted for the "Individual" or the "Dyad" and for the "Group". In all models, the normality of the model residuals was tested and where necessary, transformations were applied, using boxcox() function to determine the most appropriate transformation to meet the model assumptions and to normalize model residuals.

More specifically, to test if box activity is consistent between trials, I included the box activity at trial 1 as the predictor variable and the box activity at trial 2 as the response variable (Table 4.3, model 2). In addition to test if box activity at the start of the experiment is affected by the experimental trial and/or the intrinsic physiological characteristics of the fish (i.e. the cortisol levels, the weight and the sex of the fish), the normalized box activity was included as the response variable and trial and cortisol levels as the covariates (Table 4.3, model 3).

To test if the changepoint is significantly predicted by the different trials, a squared root transformation was applied on the changepoint values because they were skewed. The changepoint was the response variable (Table 4.3, model 4), and trial was chosen to be the independent variable. The group was chosen as a random effect.

I tested if the frequency of leadership was consistent between trials, for both, phase 1 and phase 2 (Table 4.3, model 5 and 6 respectively). In addition, to test the effect of the experimental trial on the frequency of leadership, the frequency of leadership was included as the response variable and the trial was the explanatory (Table 4.3, model 7).

Sex and cortisol levels were chosen to be the covariates, i.e. additional fixed effects.

4.3 Results

Qualitative behaviour

Once start boxes were removed, three distinct motion patterns were detected (figure 4.10): i) minimal 'no motion' period, where fish remain almost still for about few seconds (see cumulative distance plot, figure 5 and x- / y- coordinate plots in figure 4.10). ii) "random" motion where fish change their directions randomly and they travel very small distances compared to the distances they travel later in the experiment. iii) Directed motion i.e. the fish move together, change their direction according to other fish and thus they follow synchronized motion. During phase (iii) fish travel around the edges of the tank, and avoid the centre, i.e. they tend to stay close to the tank wall while moving (see figure 4.10, where the positions of either x- or y- coordinates are close to the high and low end of each coordinate).

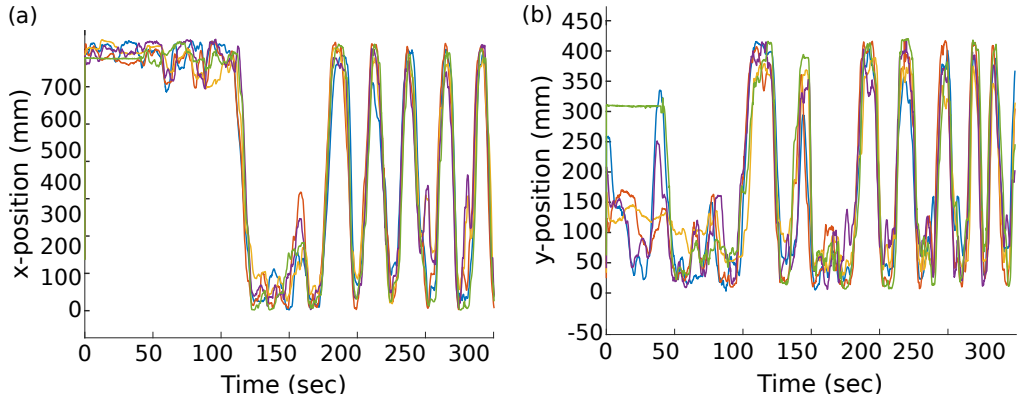

Figure 4.10: X- (a) and Y- (b) coordinates (in mm) of each fish (different lines) of group A in trial 1 across time (in secs). Both x- and y- values alternate between minimum and maximum values across the whole length of the experiment. From the videos and the positional data in the figure, one can distinguish three different motion patterns after the boxes are removed (from the beginning of the experiment).

Figure 4.11 shows each fish's cumulative distance travelled from the initial starting location. This provides a measure of the time period each fish stays "inactive" before starting moving. This period may be related to other individual-based traits, like stress or testosterone levels.

Start box activity

Permutation analysis of mean activity between neighbours and non-neighbours revealed that the fishes'activity in the start box was not influenced by that of its neighbours

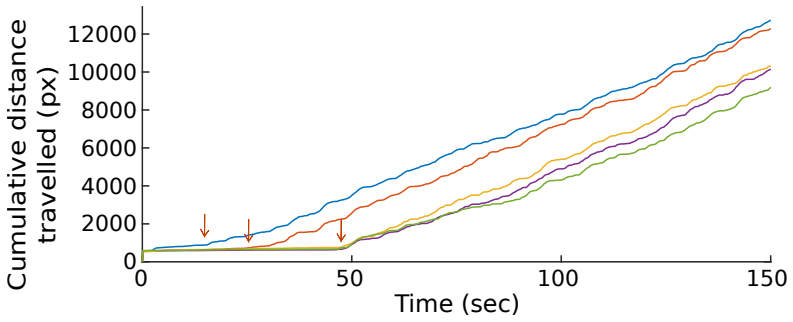

Figure 4.11: The cumulative distance each fish travelled (different fish are indicated with different colours) of group A in trial 2 for the first 150 secs after the boxes were removed. The arrows indicate the different times each fish initializes movement (initiation time lag).

(figure 4.12) ($permu_test : STAT = 1098.2$, $P = 0.442$, *seetable* 4.2). I found a significant positive relationship between normalized box activity (α^i_{trnorm}) in trial 1 and trial 2 (Spearman's $rho = 0.399$, $p = 0.029$, $n = 30$, figure 4.12 Table 4.2). The same was confirmed with a LMM, where 'individual' and 'group' were included as random effects ($LMM : Effect = 0.3245$, $StandardError = 0.1343$, $p = 0.022$; Table 4.3).

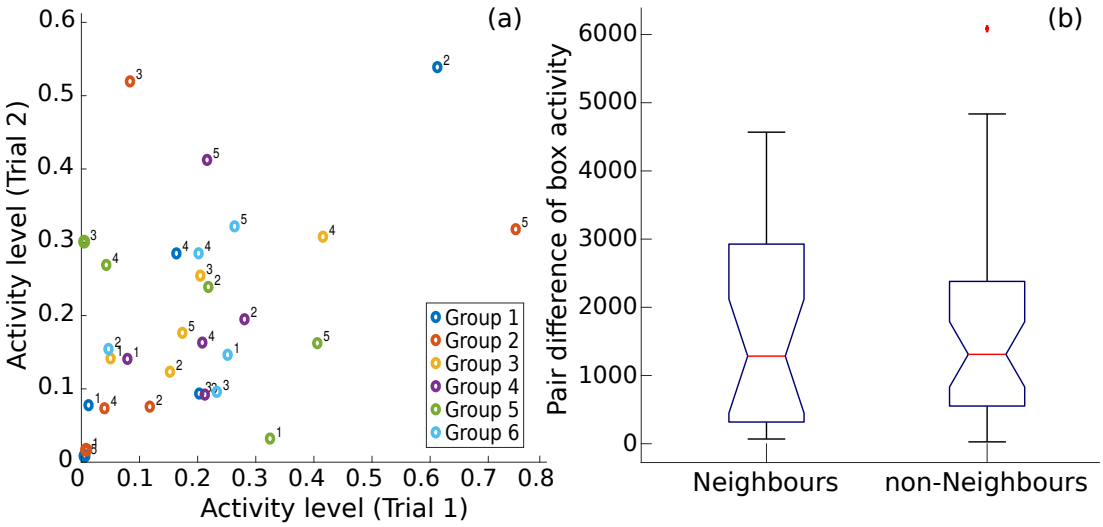

Figure 4.12: a) Normalized box activity levels for each fish across two trials. Each dot indicates a fish, which are grouped by colour (in 6 groups). The figure illustrates a moderate correlation between fish box activity across trials (Spearman's $rho = 0.399$, $p = 0.029$, $n = 30$). b) Boxplots of the difference of the box activity between neighbouring fish (left) and non neighbouring fish (right). The whiskers show the range of values, the boxes represent the interquartile range box, i.e. the middle 50% of the data. The notches overlap widely, indicating non-detectable differences between neighbours and non-neighbours.

Onset of coordination

Maximum directional correlation (CV^*) of each fish dyad overtime (figure 4.14) showed that fish transitioned between uncoordinated motion to coordinated motion. Initially, the values are below 0.4 (i.e. fish do not strongly copy each other's direction) but then the correlation increases, and the dyads' approach 0.9 (see figure 4.14). Change Point detection identified the timing of this change from non-coordinated to coordinated (figure 4.14); min = 1.16, max = 114.76, mean = 38.16secs). Continuous wavelet transform of fish positional data confirmed that the coordinated motion was related to the fish shoals' oscillatory motion around the rectangular tank (i.e. they swim together around edges of the tank) which is around 0.035 Hz (figure 4.13). Mean box activity level of the five fish did not predict the time leader-follower dynamics stabilize (LMM: $Effect = 0.0083$, $Standard\,Error = 0.0083$, $p = 0.33$, see table 4.3), suggesting more active shoals don't coordinate quicker.

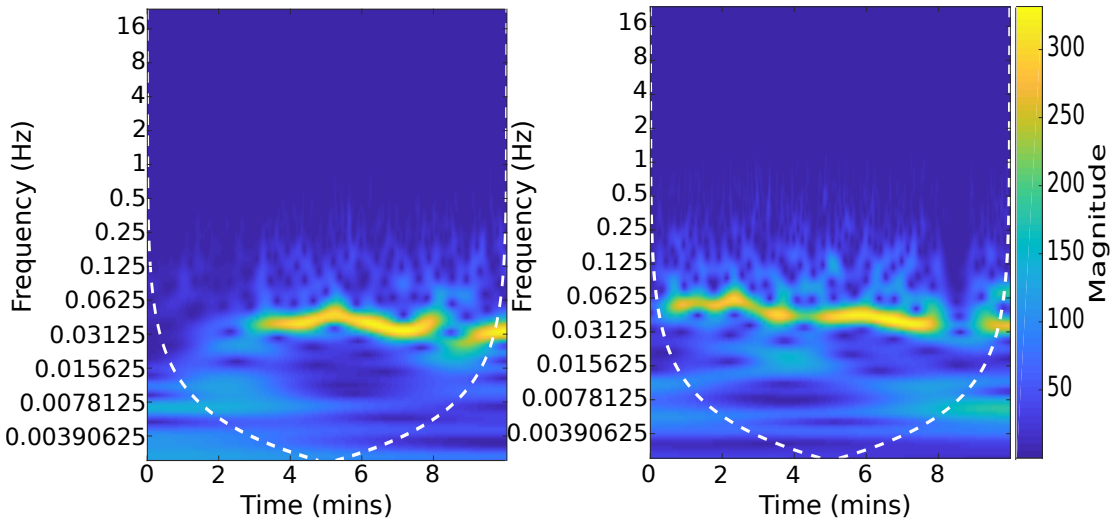

Figure 4.13: Continuous wavelet transform scalograms represents the correlation of the Morlet wavelet of a specific frequency and our signal (the positional data). Yellow colour indicates higher correlation. The yellow pattern shows the dominant frequency of this oscillatory motion which is around 0.035 Hz and indicates the emergence of oscillatory behaviour with period of 25-30 seconds confirming video observations.

Leadership establishment

In the disordered phase, there was no consistent leader-follower dynamics observed between trials (figure 14; LMM: $Effect \pm SE = -0.16 \pm 0.24, p = 0.521$), indicating copying dynamics appear random. In contrast, for the coordinated phase, leadership was observed, and these leader roles were maintained across trials (LMM: $Effect \pm SE = 0.42 \pm 0.11, p < 0.001$) (figure 4.15).

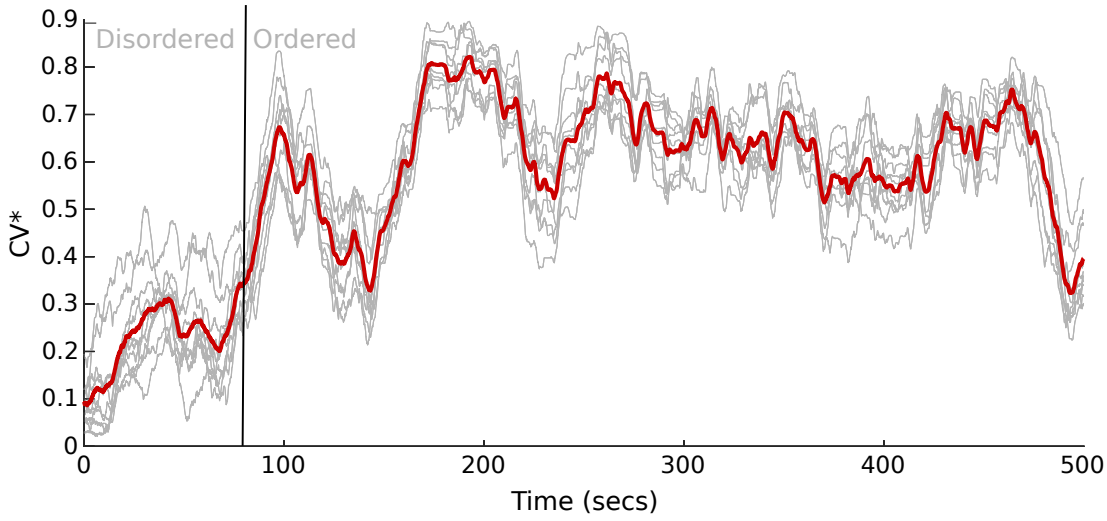

Figure 4.14: Emergence of coordination. Maximum directional correlation (CV^*) of each fish dyad (group A used as example) over time. The vertical line is the detected change point of the mean group Directional Correlation. Initially, the CV^* values are very low, below 0.4, which means that fish do not strongly copy each other's direction. After 2 secs the CV^* increases suddenly with large slope, and the dyads' CV^* increase and gradually approach value 0.9.

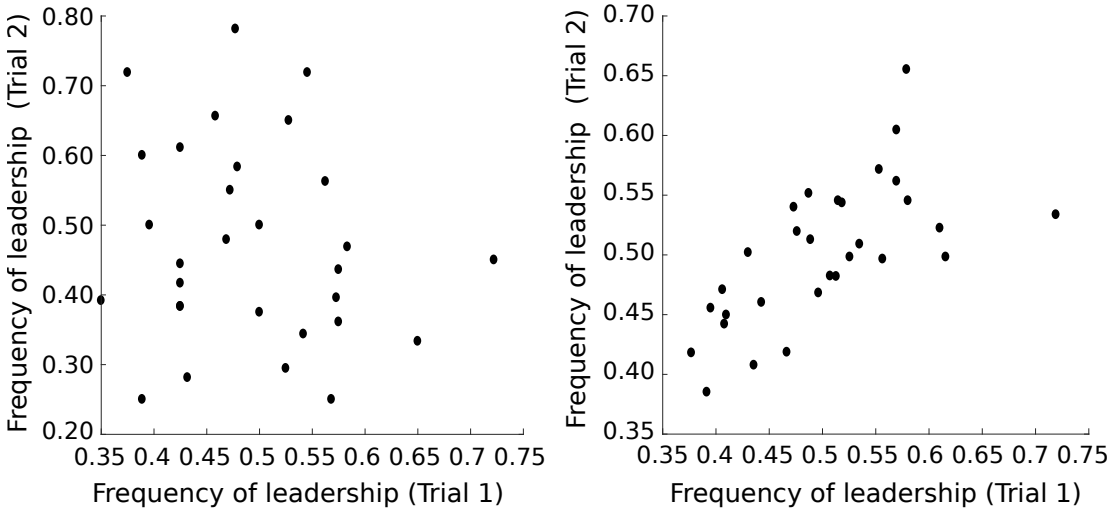

Figure 4.15: Correlation of frequency of leadership of each fish between two trials for the two distinguished detected phases (left: disordered phase, right: ordered phase). Each point represents the frequency a fish i leads over the others in the group, for specific time interval $tw = 2secs$. Each leadership event is calculated using CV formula [4], with $CV_{thr}* = 0.5$.

Time depth perspective - Repeatability

Changepoint analysis revealed that coordinated motion appears significantly faster (figure 4.16) in the second trial than in the first trial (LMM: $Effect = -16.29$, $Standarderror = 3.74$, $p = 3.43e-05$; figure 4.16), suggesting that fish coordinate faster under repeated

coordination events.

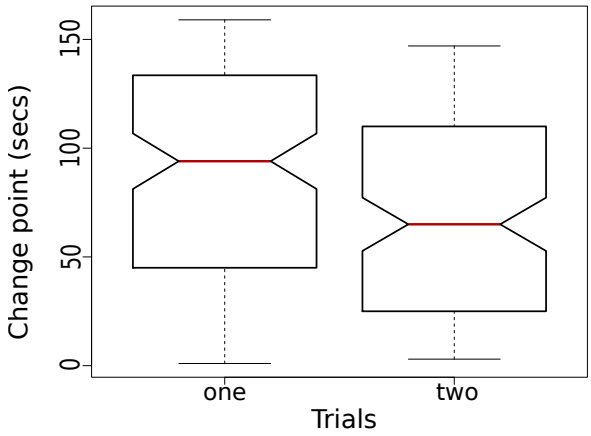

Figure 4.16: Difference in the onset of coordination (change point) in two trials. The whiskers indicate the range of values and the horizontal line inside the box is the median value.

4.4 Summary

This work demonstrated the existence of two different phases, the uncoordinated and the coordinated, and the transition time from one to the other, for shoals of free-swimming stickleback fish in the laboratory. The appearance of coordination came together with the onset of leadership and the establishment of it, since the leadership structure was maintained after the repetition of the experiment. I also demonstrated, for first time, the existence of the process of adaptation of the collective formation over time. I also investigated if some individual traits and the initial configuration determined the characteristics of the coordinated state.

Initially, a new methodology on quantifying fish activity in confined environments was proposed and was used to test for consistency between repeated experiments. The analysis on the fish activity in the box showed that the fish that are more active in the first trial were also active in the second. The activity of the fish was not predicted by their cortisol level and sex type. In this study the high or low activity was not an indication of the stress the fish experience. Other hormones and factors, such as testosterone levels or metabolic states (Ward *et al.*, 2018), could be linked to the activity levels of the fish, and could be tested in future experiments.

The study has evinced the significance the uncoordinated phase to the establishment of the group's network of interaction and suggests that the existence of the uncoordinated phase is a prerequisite for the collective formation to appear in unfamiliar individuals.

For a more in-depth discussion please refer to section 6.3.

Table 4.2: Permutation analysis results for the effect of neighbouring individuals on the box activity.

Box activity	Permutation	Statistic	Observed	P-value
	Trial 1	1098	1179	0.44
		630	454	0.77
		2607	2528	0.59
		1431	1229	0.69
		2361	1963	0.76
		2156	2699	0.14
	Trial 2	3236	4384	0.11
		2114	2523	0.29
		1384	1199	0.71
		1573	1351	0.692
		1722	1361	0.892
		1523	943	0.892
	Spearman's	0.399		0.029*

Table 4.3: Summary of linear mixed effect models investigating predictors for the box activity, leadership and coordination of fish shoals.

Model	Response	Predictor	Estimate ± se	t value	p-value
$1-(RV:group)$	$group\_changepoint$	$log(BoxActivity)$	0.01 ± 0.01	1.03	0.33
$2-(RV:group)$	$Boxactivity_{t2}$	$Boxactivity_{t1}$	0.32 ± 0.13	2.42	0.02*
$3-(RV:group)$	$Boxactivity$	$Trial$	0.01 ± 0.046	0.269	0.789
$4-(RV:group)$	$changepoint$	$Trial$	-24.64± 5.02	-4.91	<0.01***
$5-(RV:group)$ $(phase1)$	$freqleader_{t2}$	$freqleader_{t1}$	-0.16 ± 0.24	-0.65	0.52
$6-(RV:group)$ $(phase2)$	$freqleader_{t2}$	$freqleader_{t1}$	0.42 ± 0.11	3.85	<0.01***
$7-(RV:ID)$	$fleadership$	$Trial$	-0.03 ± 0.03	-0.98	0.34
		Sex	-0.02 ± 0.06	-0.44	0.67
		$log(Cortisol)$	0.04 ± 0.05	0.74	0.46
		Sex	0.06 ± 0.05	1.23	0.23
		$log(Cortisol)$	-0.04 ± 0.07	-0.74	0.46

Chapter 5

Visualisation of movement and public engagement

This chapter is a brief report on how scientific methods that are in principle developed for answering scientific questions can be used and applied for other purposes, for example to engage the public with science.

5.1 Introduction

Public engagement, according to the National coordinating centre for public engagement in UK, "describes the myriad of ways in which the activity and benefits of higher education and research can be shared with the public. It is by definition a two-way process, involving interaction and listening, with the goal of generating mutual benefits".

The above definition clearly demonstrates that the field has shifted its interest from simply contributing to the public understanding of science to focusing on engaging the public actively with Science and Technology. It, therefore, finds importance in maintaining a dialogue and an increasingly more egalitarian communication about science (Schäfer, 2009b,a).

Despite criticisms, public engagement is a necessary part of the process to democratize science (Srinivas, 2017). This is also recognized by organizations such as the US National Academy of Sciences (NAS) and the Organization for Economic Cooperation and Development (OECD). The view of public engagement as a two-way process where, both, the public and the scientists benefit from it is of great importance. Scientists are able to test their ideas and their methodologies, they can face challenges and can boost their creativity by taking part in public engagement projects and events.

This shift in the focus of public engagement has been supported financially for alternative types of science communication, such as consensus or stakeholder conferences, public discussions, or science shops (Dietrich & Schibeci, 2003; Schibeci *et al.*, 2006).

5.2 The project

During the course of my Phd I had the opportunity to get involved in a science communication project called "Animal Collectives". Animal Collectives was a collaboration between the artist Heather Barnett and my supervisors Dr Andrew King and Dr Ines Fürtbauer and their groups and was funded by a Leverhulme Trust Artist in Residence grant and was ran between March – November 2017. My contribution to the project was the devise of an outdoor interactive/collaborative activity regarding ecological concepts, the creation and video projection of the motion from different systems and the implementation of an interactive installation, where participants' motion would be reflected and presented as an abstract motion field on the wall of a dark room. The interactive installation was run at Arebyte gallery in Hackney Wick, East London during the event "Crowd Control" that took place between 1-20 July 2017. The video representations were exposed in both the "Crowd control" event but also for two months in Berlin, i.e. from 30.09.2017 to 26.12.2017, at the Art Laboratory Berlin during the "Nonhuman Networks" exhibition.

5.3 Methodology

Optical flow or optic flow is the pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between an observer (an eye or a camera) and the scene. The method that is introduced in Chapter 4 on the quantification of the fish activity in confined environments is based on the measurement of optical flow. Detailed information on how optical flow algorithms work is presented in section 4.2.3. Briefly, a routine that tracks specific features (points) in an image across multiple frames is developed. This routine can extract the magnitude and direction of the perceived motion of the feature in image sequences.

Based on the build-in function on optical flow of the OPENCV library (Itseez, 2015), I developed a routine that calculates the flow of multiple points in the image frame. More specifically, a grid of points is created and applied on the captured image, covering all the image area. The optical flow algorithm is, then, sequentially applied at each point location, and for each consecutive time frame. If a motion is taking place at a location of the predetermined grid points, i.e. if a change in the intensity values of a predetermined point in the image is detected, the algorithm will calculate and return the flow properties of that point. Assuming that the objects in the image are moving in a smooth and predictable way, the position of the point at the next frame will be updated to the new location.

The above algorithm constitutes the base of a series of routines that have been developed for a particular visualisation purpose. The parameters are tuned and modified,

depending on the type of moving objects, i.e. if they are fish, or humans, or monocellular organisms. They are also modified depending on the video characteristics (e.g. capturing rate, illumination conditions and noise) but also on the desired outcome.

5.4 Results

An example of how videos of human collective motion can be converted to a motion field can be seen in figure 5.1. In this case, participants were asked to experience the rules a slime mould follows in order to explore food. A grid of points is created and applied on the image of the first frame. At each time step the optical flow algorithm was applied and the grid points where updated with their new position in the image. Tracks of the previous positions where used to draw the points trajectories.

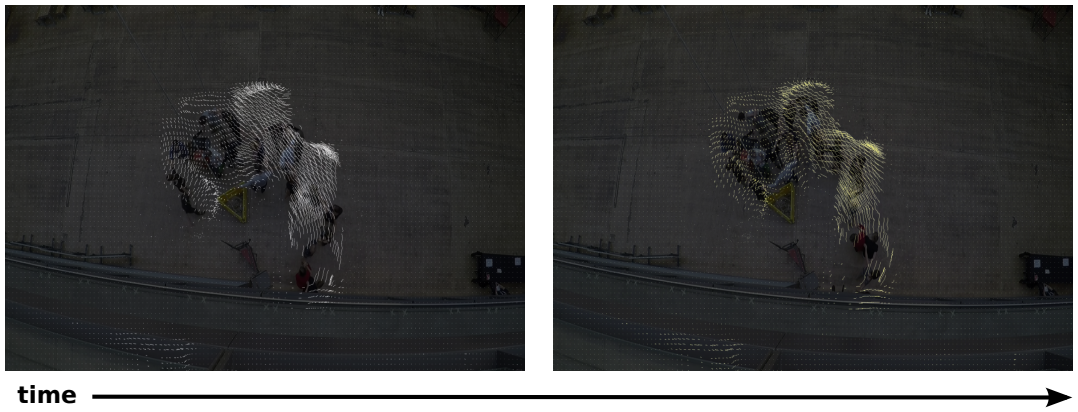

Figure 5.1: Left and right image show the result of applying optical flow on videos of human collective movement

Video representations of movement

Figure 5.2 shows the final set-up of the exhibition at the gallery. The screens were positioned on the floor and were showing different aspects of crowd behaviour and collective movement, from different living and non-living systems. One of the screens was reproducing a video of an abstract motion field of collective human movement. The source videos, were referring to the collective movement of human participants in a previously organized workshop by Heather Barnett. During that workshop participants were asked to become part of a being slime mold scheme, for this they were asked to follow some movement and interaction rules to experience how a slime mould would explore its environment in order to encounter food.

Videos would show the points in the screen to move according to the direction and velocity of the underlying movement. Interesting moving patters emerged as a result of applying the methodology to these videos. For example, points were moving as

following a spiral. All these interesting patterns were not easy to be perveived from the source videos due to the higher level of detail that could easily distract the observer from focusing on these details.

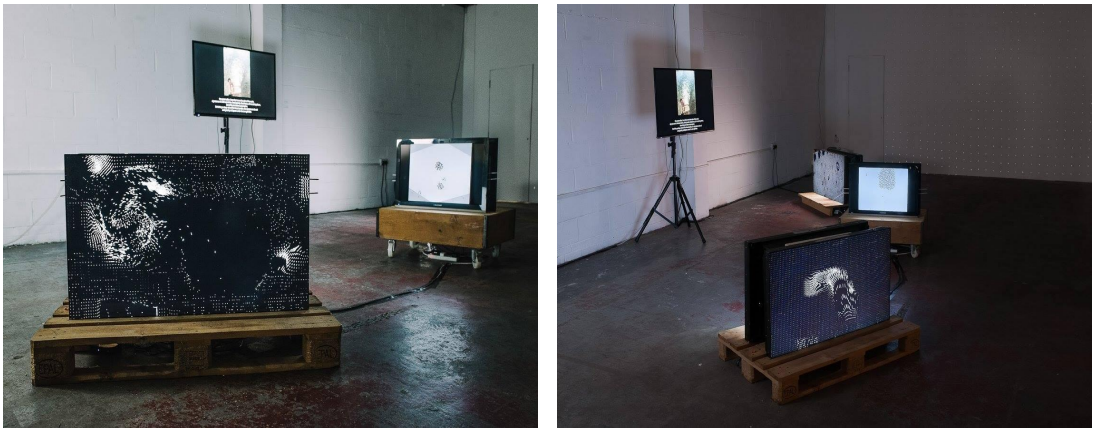

Figure 5.2: The set-up of the gallery during the exhibition period. Motion fields from collective human movements have been displayed on screens placed over the floor.

Interactive

The interactive installation was realized at the *Testing Station* at Arebyte gallery (figure 5.3). Visitors were the targeted subjects. Their movement inside the gallery was captured by a fixed camera, on the ceiling, and was converted and projected into the wall as an abstract motion field, i.e. as a collection of moving dots.

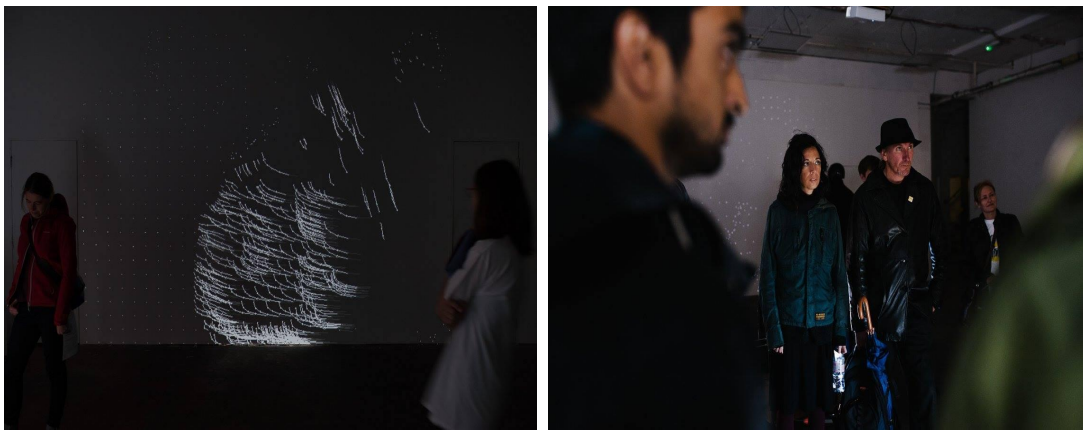

Figure 5.3: The environment where the interactive installation was realized.

The participants were able to perceive how fast they were moving, how their movement modifications would be reflected in the screen. If more than one participants were moving at the same time, then more interesting patterns would emerge, like points that were moving in opposite directions and entounter each others.

5.5 Discussion

In this brief chapter I demonstrated some of the endless possibilities one can choose to visualise motion. In this case, movement was represented as an abstract representation of a vector field. This representation helped participants to realise that motion can be described by very few parameters, the direction and the speed.

The participants benefited from this experience because they could have the chance to get familiar with the notion and the representation of a vector field. The could also be able to perceive that motion can be described by very simple parameter. They could also have a sense of how the area is explored during motion depending on the system, i.e. how much of the total area covered. They finally could be able to see the effect the change in the movement pattern had on the form of motion field.

The reasearchers that participated in the event also benefited. They realized that public engagement projects could potentially provide data for answering questions regarding human movement. The involvement on the public engagement project enhanced the perception that the motion can show similar characteristics across a variety of study systems. In additions, researchers were able to get a feedback from the participants on the impact this methodology can have on other applications. Finally, this experience helped scientists to come up with new ideas and new scientific questions.

Chapter 6

Discussion

The thesis developed tools to track movement (i.e. extract positions and other useful attributes) at three different spatio-temporal scales in biology; the intracellular, the cellular and the organism level (i.e. fish). It also targeted and shed light on relevant questions regarding each scale. Regarding intra-cellular movement, it for first time empirically extracted the inheritance of the endosomal QD fluorescence across two consecutive generations. It demonstrated that QD endosomal inheritance in Human Osteosarcoma (U2OS) cells is random and there is no selection towards a specific splitting ratio but the QD loaded endosomes are distributed uniformly into the daughter cells. This is a finding with significant pharamceutical implications. In addition, it developed in silico models that could be used to predict more accurately QD endosomal inheritance from flow cytometry experiments. Regarding cell movement, the thesis developed a semi-automatic algorithm to track cell positions from long-term microscopy experiments, using a GFP cell cycle marker. It showed, that the global movement resembles brownian walk initially, but is, at later steps, becoming sub-diffusive due to the increase in cell density. In addition, empirical data showed that there must be two different movement patterns within the U2OS populations, one more diffusive and one more directed. The different movement patterns could be related to different behavioural/genetic profiles. Regarding fish movement, it was demonstrated that the emergence of coordination of fish follows after a phenomenically random/disordered state. To my knowledge, it was for first time showed that this disordered phase is of great value and necessary for the establishment of the network of social interactions in a group of unknown fish. The time that is spent in that phase is decreasing after repeated experimental trials, and thus, they seem to habituate to constant environments.

6.1 Intracellular motion

Chapter 2 studied the partitioning of the fluorescence of the loaded QDs in proliferative human osteosarcoma cells (U2OS). The cells were labeled with a GFP marker at their cytoplasm and long-term microscopy experiments took place to extract genealogical information. The use of QDs was decided for their properties (i.e. high photostability, limited photobleaching, narrow fluorescence emission spectra) and for their potential to function as carrier systems for efficient drug and gene delivery. The particles can be provided with multiple functionalities, and their specific characteristics allow the combined application to imaging, therapy and their use as carrier systems for the targeted delivery through one nanoparticle type.

In the current work, it was shown that the QD endosomal fluorescence intensity is not distributed equally into daughter cells. The whole process is purely stochastic and does not select for a specific QD splitting ratio value. Almost two decades ago, Bergeland *et al.* (2001) and Dunster *et al.* (2002) were suggesting that endosomes must be equally distributed to the daughter cells, since no endosomal fusion or fragmentation is observed. But no evidence for a strict mechanism guaranteeing their equal distribution was found. On the other hand, Dunster *et al.* (2002) suggested that specific types of endosomes are distributed equally to daughter cells under an active partitioning process. Therefore, the calculated asymmetry seen in this work must be due to the random uptake of QDs into the vesicles. Indeed, Summers *et al.* (2011) demonstrated that the chance of success or failure of nanoparticle uptake and inheritance is random and thus there is asymmetric nanoparticle segregation in the endosomal compartment.

Previous work of Brown *et al.* (2010) predicted, using in silico models tuned by flow cytometry data, that the QD fluorescence of a cell would be distributed to the daughter cells following a 80:20 splitting ratio value, a conclusion that raised questions on the ultimate causes that lead to the selection for a particular splitting ratio value. The demonstrated heterogeneity in the nanoparticle uptake in cells can work as a pool where selective forces can act and select cells that distribute endosomal material following specific splitting ratio values.

Current work has clearly demonstraded that there is no selection towards a particular splitting ratio value, because QD fluorescence intensity splitting ratio is distributed uniformly and thus there must not be a fitness benefit for the selection of cells towards a particular splitting ratio. Therefore, there is no biological process that creates a preference for a particular asymmetric partitioning value.

The demonstrated heterogeinity of the splitting ratio values, however, is still a very interesting finding and could have significant pharmaceutical implications. The existence of a wide range of QD numbers in the cells and in the endosomes could work in favour of the survival of the cell. A wide variety of drugs that target cancer cells are internalized

into the cells via the endocytotic route (Kirtane *et al.*, 2013). The effectiveness of the administered dose is, therefore, tightly linked with the distribution of the acquired dose into the cells and more specifically into the endosomes. The heterogeinity in the endosomal QD load can, therefore, help the U2OS cellular population, to bypass the effect of the administered drug, by accumulating higher amounts of it in only a percentage of the cells. Future experiments that use drugs that target tumour cells, and subsequent studies on the effect of the administered dose on the population of Human Osteosarcoma cells is required to confirm or reject this hypothesis.

To my knowledge, this was the first study to extract detailed genealogical information from microscopy for the study of the QD fluorescence iheritance, as previous studies attempted to give answers using in silico models and flow cytometry experiments.

The majority of previous works have been based solely on flow cytometry snap-shot, i.e. contiguous time-series, data. These data do not contain detailed spatial and genealogical information. A question of interest, therefore, is to investigate and understand the relationship between the equivalent data from flow cytometry and that extracted from high throughput microscopy. For example, it would be interesting to find ways image microscopy data could contribute to flow cytometry data interpretation and vice-versa (figure 6.1). We need to develop methodologies that can theoretically, but accurately, reveal genealogical information and predict the redistribution of the intracellular QD dose from flow cytometry data, based on the image microscopy experiments.

The methodological approach followed (i.e. microscopy) intrinsically age-sorts the QD fluorescence data into generations and the daughter cells are further sorted into high and low loaded QD fluorescent categories. In contrast with microscopy, it is almost impossible to extract clear genealogical information with flow cytometry, because of the unavoidable variations on the division times in cells. Thus the distributions from flow cytometry include multiple generations. Current work evinced the limitations and concerns someone should have in order to study compartmental inheritance on flow cytometry.

By simulating the redistribution of the QD fluorescence in an age-sorted in-silico model I recreated the perceived asymmetry found in flow-cytometry experiments as is shown in figure 2.11 a. By convolving the high and low loaded daughter sub-populations into one population I remove the genealogical relationships between parent and daughter cells and by randomly sampling the measured intermitotic time distribution I remove the age-structured or generational relationship. Under these conditions, the QD partitioning ratios are asymmetric and match contiguous flow cytometry measurements.

Therefore, current work established a link between flow and image cytometry, with the development of an in silico model for the prediction of the QD fluorescence inheritance from flow cytometric data. The contribution of each generation from flow cytometry samples can now be uncovered. The use of this information can help to more

accurately predict compartmental inheritance from flow cytometry experiments, taking into account the actual percentage of each generation in the sample.

Recent work of Thomas (2017) introduced a theoretical way to study and predict snapshot data, for example data that derived from flow cytometry experiments. In addition to that, he incorporated into his theoretical approach the evolution of chemical reactions of specific molecules in proliferative cells and proposed an analytical way to predict the amount of molecules in time. Based on his work, the next step would be to analytically describe and predict the redistribution of the QDs in proliferative cells from experimental data that are not age sorted.

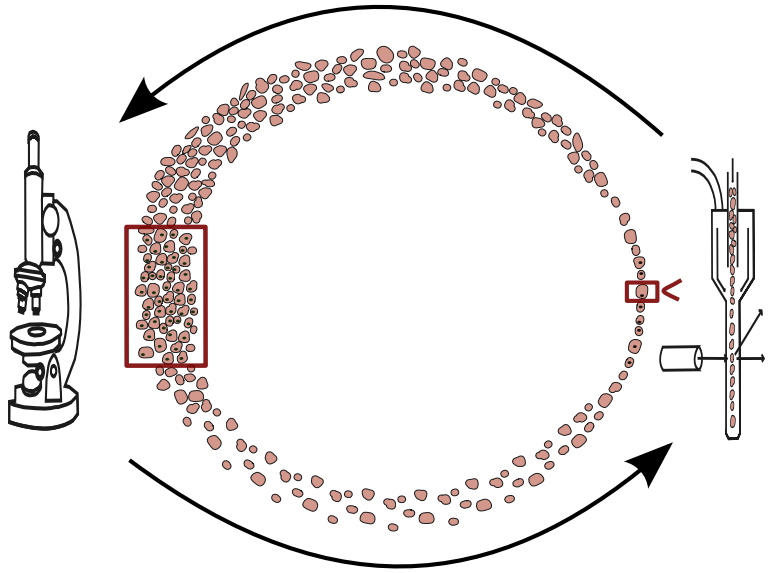

Figure 6.1: Schematic representation of the interplay between flow and image cytometry methodologies. Information is transferred from one methodology to the other and both help to better understand and interpret data on cellular and drug dynamics.

6.2 Cellular movement

Chapter 3 aimed to unravel the movement behaviour and characteristics of the Human Osteosarcoma cellular population and to provide information on the cell cycle profile of the cells.

During the initial stages of the cell cycle cells move following on average normal diffusion process. This is demonstrared from the study of the slope of the MSD curve for the initial time points. However, as the cell cycle time progresses, the average movement behaviour is best described as being sub-diffusive, i.e. the increase in the average discplacement does not increase linearly with time. In other words, cells move slower than they would if were realising normal diffusion. This change in movement behaviour seems to happen due to the crowding effect (Weiss *et al.*, 2004), i.e. the

increase of the cell density in the well plate as time progresses.

The current work has also demonstrated the existence of different movement strategies in human osteosarcoma (U2OS) cell populations. The heterogeinity in tumour cells has been widely studied, mainly at the molecular, biochemical and intracellular level (Zhang *et al.*, 2015a; Ekdawi *et al.*, 2016; Brown *et al.*, 2018; Yan *et al.*, 2016). The current study demonstrated that the heterogeinity can also be expressed as a difference in the movement strategies of the cells. Further studies that will link movement trajectories with the genetic profile and persistence individual differences of the cells are required to enhance this observation.

Mean squared displacement analysis has been thoroughly used across multiple fields in biology, from molecular biology to movement ecology (Singh *et al.*, 2016; Chenouard *et al.*, 2014; Fagan & Calabrese, 2014; Differential *et al.*, 2008). It can serve as a tool to measure the mode of displacement of individuals followed over time. Because it is an averaging method, it is unable to unravel the existence of different movement classifications within the population. For example, in the current case, the MSD plot does not reveal the heterogeneity seen in the movement trajectories within the population (figure 3.12). Net squared displacement (NSD) plots are usually used to classify differences in movement trajectories (Singh *et al.*, 2016; Bastille-Rousseau *et al.*, 2016). NSD measures the Euclidean distance between the starting location of a movement path and each subsequent location (Singh *et al.*, 2016). Time-series of NSD values are characteristic of individual movement trajectories. In contrast, Mean squared displacement is used to quantify the diffusive spread of particles, or animals, over time and space. Future analysis and the use of NSD would be prefered in order to study the inter-individual mobility patterns of the U2OS cells.

Osteosarcoma is a bone tumour of mesenchymal origin displaying significant heterogeneity. Although evidence suggests the presence of cancer stem cells (CSCs) in osteosarcoma a consensus on their genetic and phenotypic characterization is still missing. Identify cancer stem-like cell populations in osteosarcoma cell lines, Characterize osteosarcoma CSCs at the genetic and phenotypic levels.

Nowadays, there is a lot of concern on the effectiveness of some therapeutic methods used to attack tumours. An increasing amount of studies focuses on the existence and the behavioural response of stem-like cancer cell types under therapeutic pressure. Detecting, predicting and understanding the cellular heterogeneity is a crucial step to achieve efficient cancer treatment protocols. This heterogeneity can be demonstrated also in the movement pattern of cells. Being able to, therefore, classify different movement types and assign them to specific cell behaviours would be a very useful diagnostic tool for the early detection and prevention of the emergence of metastatic events.

The extracted distribution of the distance traveled of the cells is not bell shaped, but it has rather a fat right tail. Taking into account figure 3.12 and the distribution of the total

distance traveled, it is anavoidable to suspect that the cells, if not all, follow a different than the brownian walk. From figure 3.12 at least two different movement patterns are apparent, one more diffusive and one more directed. Therefore, the distribution seen in figure 3.14 must be a combination of two different distributions, one Gaussian and one Levy-like.

The future aim would be to be able to distinguish stem like cells from their change in movement and thus calculate the percentage of those in a U2OS population of specific density. This could have implications on the time and type of therapeutic method used to attack tumours. We could also study the movement characteristics of the cells and try understand their interactions with other cell types. Targeting the interaction network of the cancer cells and their heterogeinity we could find means of controling tumors more effectively.

In future work, I will classify cells within a population according to their cell movement patterns. Detection and prediction of emergence of metastatic behaviours by the movement patters and cell cycle characteristics of cells. In addition, I will aim to answer if the percentage of motile cells found from image cytometry is correlated to the percentage of stem-like cells from flow cytometry. The above analysis would be repeated in population where real drug has been administered.

6.3 Collective movement

In this work, I demonstrated that the initial detected activity levels of the fish are neither affected by their physiological characteristics (i.e. the cortisol levels, the weight and the sex of the fish) nor by the activity of their neighbours. However, I found that the activity levels were consistent across repeated experiments. I also demonstrated the existence of a transition from a disordered to an ordered phase. The time at which the emergence of coordination takes place was significanlty decreased in the second trial. Finally, I quantified the frequency of leadership of each fish and I demonstrated that the frequency of leadership was maintained across trials for the ordered phase, but not for the disordered.

In more detail, I initially proposed a new methodology on quantifying fish activity (i.e. fish box activity) in confined environments and I used it to test the activity of each fish and if the activity was consistent between repeated experiments. The analysis on the fish activity in the box showed that the fish that are more active in the first trial were also active in the second. The fact that the activity prior to start of the experiment showed consistent inter-individual differences indicates that box activity may represent personality trait (Sih *et al.*, 2015; Mazzamuto *et al.*, 2018) and can be used to test if it determines the behavioural rules established during the coordinated state (Ward *et al.*, 2018). The activity of the fish was not predicted by their cortisol level and sex type.

Cortisol levels indicate the stress state of the fish (Baker *et al.*, 2013). In this study the high or low activity was not an indication of the stress the fish experience. Other hormones and factors, such as testosterone levels or metabolic states (Ward *et al.*, 2018), could be linked to the activity levels of the fish, and could be tested in future experiments.

The basic metric used for understanding the onset of coordination was the directional correlation delay time (CV^*), a metric that has been previously introduced for the detection of coordination and leadership hierarchy in pigeon flocks (Nagy *et al.*, 2010) and recently used for detecting influential neighbours in U-turns (Jiang *et al.*, 2017). In the current study, this metric was used in order to show the existence of the transition to coordination and also to quantify leadership events. The existence of a transition between low and high coordination was perceived as the transition from low to high directional correlation CV^* values (figure 4.14). The initial observation/hypothesis that the fish do not start coordination immediately when they first meet each other is confirmed by the CV^* values. The detected time this change takes place varies depending on the group and the trial, but is apparent in all the cases. The variation on the times of the onset of coordination needs itself investigation, as it can be affected by different factors, such as group composition, initial positioning in the tank, the time and day of the experiment, but also the relationship status (such as familiarity) between the members of the groups. The latter has been investigated in this study.

Once the emergence of coordination was confirmed and demonstrated, questions regarding the importance of the existence of a disordered phase for the determination of the later dynamics were investigated. CV^* was used to quantify leadership for the different phases. The frequency of leadership between trials was consistent in the coordinated state, but not in the uncoordinated. There must be some underlying rule that keeps the hierarchies unchanged in the ordered state. In the current study the fish that participated did not have any prior interaction, and thus they were unfamiliar. These observations lead to the hypothesis that the initial disordered phase is driven by stochasticity, while fish try to react to the new environment and the perceived leadership is of no real meaning. Previous work of Jiménez-morales *et al.* (2018) studied the formation and maintenance of hierarchical organization in unfamiliar crayfish triads under agonistic interactions and demonstrated that the first encounter is necessary for the hierarchy to be established and, after this the hierarchy is maintained. This finding enhances the hypothesis that fish need some time to establish their network of interactions and if conditions remain unchanged so do their hierachical structure. Hobson & DeDeo (2015) studied how does dominance emerges in groups of unfamiliar monk parakeets and demonstrated the existence of a transition from a disordered structure to an ordered one. They found that during the first quarter of the study period the dominance structure of the groups was consistent with the null model, i.e the structure that would be expected if interactions were random and not established yet. However, they found that a transition

towards more structured aggression occurred rapidly, about a week after initial group formation. Therefore, the current and previous studies suggest that the existence of an uncoordinated phase must be a prerequisite to the achievement of collective motion as it could be the period where the group establishes the hierarchical network of interactions.

None of the personality traits. i.e. cortisol levels and sex, available for the study could predict leadership or the time of coordination. These traits have been previously showed to play an important role on the form of the interaction in collective behaviours. However, the results here indicate that other individual traits may be more important for the establishment of coordination and the hierarchies, for example testosterone or other hormonal level, acceleration profile etc. It is also possible that hierarchy does not emerge primarily from differences in intrinsic qualities of individuals but rather as a self-organizing process in which a hierarchy arises as a result of many interactions between the members of the society (Hickey & Davidsen, 2019). In the latter case, the shape of the onset of coordination could also be affected by initial conditions, such as relative positions of the fish at the beginning, the number of participants, the level of intensity of interactions and others. Future studies could work on them.

One of the components of our experiment was the repeatability, i.e. the time-depth perspective. The demonstration of the emergence of coordination and the respected underlying properties that were stated above is an important first step to understand how the phenomenon of fish coordination is established. However, most real situations can be repeated multiple times and the same group of individuals needs to coordinate more than one time (Biro & Sasaki, 2016).

Changepoint analysis showed that coordinated motion appears significantly faster (figure 4.16) in the second trial. This is an interesting result and raises questions regarding the reasons that this is happening. Figure 4.15 shows that during the disordered phase the leader-follower relationships are random, due to the lack of consistency between trials. In contrast, during the ordered phase leader-follower dynamics between trials are strongly correlated. Therefore, a reason that could explain the differences in the time of coordination would be that fish in first trial have to construct their social network from the beginning, and this would require some time (see figure 4.16). Once they have established their network then they can start coordinating faster, thus the duration of the disordered phase would decrease. This suggests that the initial phase can be a fundamental period, very important and necessary that shapes the form of later coordinated phase.

It is important to note that even in the simplest experimental setup, where environmental conditions are stable and there are no external stimuli (like predation), there is a significant change of behaviour of the fish in the time-depth (such as faster coordination). It would be possible that in stable environments like the one studied here, the dynamics on the collective behaviour would stabilize sooner than in more complicated environments, where the external forces could probably act constantly. There are few

studies that targeted the adaptation of collective formations and the links between altered environmental conditions and social organization (Jiménez-morales *et al.*, 2018; Hobson & DeDeo, 2015; Flood & Wong, 2017). However, future studies are needed to understand the changes in the social structure and the collective dynamics under a dynamic environment. The results lead to the conclusion that the time of coordination is not a random event and is related to the interactions that take place during the disordered phase. A reason why frequency of leadership in the ordered phase strongly predicts the fish pair coordination times would be that fish pairs that follow each other more often probably share some attributes (physiological, visual or others) that trigger the establishment of their interactions and hierarchies faster than other dyads. It seems that analysis at the dyadic level can unravel some important information to better understanding the complex process of emergence of coordination.

Regarding fish, it is worth noting that these experiments performed on small group sizes. But most collective actions consist of a much higher number of individuals (Croft *et al.*, 2003) and this can change significantly the relevant times of the onset of coordination and the leadership dynamics in one, and more trials. Therefore, it is important to test the emergence of coordination and the adaptation in larger groups, where the group composition may play a very important role.

The results of the study bring insights on two main components. First, they demonstrate the existence and importance of the uncoordinated state after a disturbance event (after box removal). They suggest that this phase is crucial for the establishment of the coordination. Second, it proves that even in the simplest, and most stable environment, there is an adaptation in time, suggesting that this would be much stronger and of higher importance, in more variable and dynamic environments.

In conclusion, this work demonstrated the existence of two different phases, the uncoordinated and the coordinated, and the transition time from one to the other, for shoals of free-swimming stickleback fish in the laboratory. The appearance of coordination came together with the onset of leadership and the establishment of it, since the leadership structure was maintained after the repetition of the experiment. In addition, it was for first time demonstrated that the fish shoal formation and its dynamics are adapting over time, indicating and enhancing previous statements that collective formations act as super-organisms that are shaped by environmental conditions that drive them to evolve and adapt over time (Biro & Sasaki, 2016).

6.4 Methods for data extraction and analysis

This work also developed three routines to track and extract data from the three different study systems. These routines ranged from fully automatic (fish), semi-automatic (cells), and manual tracking (intracellular movement).

For manual tracking (Chapter 2), the input that is needed from the user is the cell outlines and the cell generation status. Then the routine's task is to store the cell mask, the location and the total count of intensity of the area that is under the cell mask in the QD channel. The algorithm uses simple image analysis tools to extract and store the required information. The usefulness of the algorithm developed here is not its automation but the specificity for the particular task, since there is no other pre-existing routine that could be used unchanged to target this question.

For the semi-automatic tracking, the user's responsibility is to only point to the centre of the targeted cell. The routine then will track and extract the positions of the cells between mitotic events, count the QD intensity for each cell at each point in time, detect the mitotic events and, finally, extract and store the genealogical relations between cells.

The algorithm proposes a simplified but quite effective way to extract migration routes for cells that are not labeled in their nucleus. In fact, it introduces two simple ways to improve nuclei segmentation: the simple threshold and the adaptive threshold in an iterative way. Each image was thresholed for a range of threshold values (for the simple threshold) and window size parameter (for the adaptive threshold). Each thresholded image was then filtered to keep objects of nucleus-specific characteristics. This iterative mode of nucleus thresholding increased the effectiveness of the segmentation.

This routine also introduced an index that assigns a value of probability at each segmented object in the image. This index is the weighted sum of three similarity indices, the spatial, the shape and the feature similarity. The contribution of each terms was decided depending on the question.

Lastly, the routine introduces a way to extract the QD intensity of each cell, without having to segment the whole outline of the cell. This was due to the observation that most of the QDs are distributed around the nuclei. Therefore the calculation of the total QD fluorescence within a circle that is centered at the centre of the nuclues, could provide a good estimate of the total fluorescence per cell.

This routine enabled us to run live cell analysis and get cell cycle information without labelling the nucleus, keeping the spatial information and relationships of cells, in contrast with imaging flow cytometry (paul rees 2016). Future work, would require to improve and achieve a complete automated algorithm to track cell movement and extract also lineages of cells treated with non-nucleus label.

This routine needs to be improved in the following; First, it needs to increase the successful detection of mitotic events. This can be achieved by incorporating the Cell cycle phase of the cell, that will predict how possible is the targeted cell to go to mitosis. In addition, it can be achieved by relaxing the distance rule when a cell is really possible to undergo mitosis. With this, the algorithm will be able to look for mitotic cell taking into account the big jumps that usually take place. Second, the lineage extraction can be improved. After mitosis cells become very dim and small and due to limited space, one

of the daughter cells could have a larger distance to their parental cell than other cells, in a densely packed colony. Therefore, to be able to correctly segment both cells, would be important to introduce a size measure. This means that the algorithm should look for cells that are of much smaller size than other cells. In addition to that, we should incorporate the cell cycle information and ask the algorithm to look for cells of very low intensity.

Finally, for the automated tracking routine, the user has to only specify the range of the HSV values in order to correctly segment the coloured disks of the fish. It then, uses the minimum distance rule to associate each fish between frames. The simplicity of this routine renders it as a generalized tool for tracking any collective of individuals that are tagged with coloured rings. Its efficiency will decrease rapidly in cases where the number of individuals is large, i.e.$>$ 10.

6.5 Concluding remarks

In conclusion, the current thesis studied movement at different spatio-temporal scales in biology. Movement has been the focus of study in many research fields, because it is a vital process and can have an important effect on the survival of the individuals and consequently populations. Evolutionary processes act constantly on individuals and can select for those that have developed more successful movement strategies. Previously, movement processes were treated as being the subject of study of each particular research field and the inter-disciplinary communication was very limited. In recent years, due the increase in inter-disciplinary collaboration and the use of statistical physics methodologies in biology, it is becoming clearer that movement can be approached and studied using similar tools irrespective of the scale and the object of the study. Therefore, there is an increasing attempt to adopt a unifying framework regarding movement in biology and bridge the gap between the studies across different scales. Works of Hansson *et al.* (2014), Schick *et al.* (2008), Hansson *et al.* (2014), Börger (2016), Jacoby & Freeman (2016), Torney *et al.* (2018a) and (Pyke, 2015) are few of the attempts researchers make to bridge the gap between different scales on movement in biology. For the above reasons, this study attempted to use similar methodologies to study movement at different spatio-temporal scales. The empirically extracted movement trajectory is the basic form of information a researcher has to study movement. The current study discussed and categorized the factors a researcher should consider and that could have an effect on the pattern of the extracted trajectory irrespectively of the scale of interest.

References

Abarrategi, A., Tornin, J., Martinez-cruzado, L., Hamilton, A., Martinez-campos, E., Rodrigo, J. P., González, M. V., Baldini, N., Garcia-castro, J. & Rodriguez, R. (2016). Osteosarcoma : Cells-of-Origin , Cancer Stem Cells , and Targeted Therapies. 2016. 6, 20

Adeli, M., Hakimpoor, F., Parsamanesh, M., Kalantari, M., Sobhani, Z. & Attyabi, F. (2011). Quantum dot-pseudopolyrotaxane supramolecules as anticancer drug delivery systems. *Polymer*, 52, 2401–2413. ISSN 0032-3861. URL http://www.sciencedirect.com/science/article/pii/S0032386111001522. 19

Andrieu, J., Henry, S., Hausberger, M. & Thierry, B. (2016). Informed horses are influential in group movements, but they may avoid leading. *Animal Cognition*, 19, 451–458. ISSN 1435-9456. URL https://doi.org/10.1007/s10071-015-0945-2. 8, 60

Aoyama, M., Yoshioka, Y., Arai, Y., Hirai, H. & Ishimoto, R. (2017). Intracellular tra ffi cking of particles inside endosomal vesicles is regulated by particle size. *Journal of Controlled Release*, 260, 183–193. ISSN 0168-3659. URL http://dx.doi.org/10.1016/j.jconrel.2017.06.007. 5

Aranda, P. R., Messina, G. A., Bertolino, F. A., Pereira, S. V., Baldo, M. A. F. & Raba, J. (2018). Nanomaterials in fl uorescent laser-based immunosensors : Review and applications. 141, 308–323. 18

Aswathy, R. G., Yoshida, Y., Maekawa, T. & Kumar, D. S. (2010). Near-infrared quantum dots for deep tissue imaging. *Analytical and Bioanalytical Chemistry*, 397, 1417–1435. ISSN 1618-2650. URL https://doi.org/10.1007/s00216-010-3643-6. 19

Bajpai, V. K., Kamle, M., Shukla, S., Mahato, D. K., Chandra, P., Hwang, S. K., Kumar, P., Huh, Y. S. & Han, Y.-k. (2018). ScienceDirect Prospects of using nanotechnology for food preservation , safety , and security. *Journal of Food and Drug Analysis*, 1–14. ISSN 1021-9498. URL https://doi.org/10.1016/j.jfda.2018.06.011. 18

Baker, M. R., Gobush, K. S. & Vynne, C. H. (2013). Review of factors influencing stress hormones in fish and wildlife. *Journal for Nature Conservation*, 21, 309–318. ISSN 1617-1381. URL http://dx.doi.org/10.1016/j.jnc.2013.03.003. 95

Barnard, S., Calderara, S., Pistocchi, S., Cucchiara, R., Podaliri-Vulpiani, M., Messori, S. & Ferri, N. (2016). Quick, accurate, smart: 3d computer vision technology helps assessing confined animals' behaviour. *PLOS ONE*, 11, 1–20. URL https://doi.org/10.1371/journal.pone.0158748. 10

Bastille-Rousseau, G., Potts, J. R., Yackulic, C. B., Frair, J. L., Ellington, E. H. & Blake, S. (2016). Flexible characterization of animal movement pattern using net squared displacement and a latent state model. *Movement Ecology*, 4, 15. ISSN 2051-3933. URL https://doi.org/10.1186/s40462-016-0080-y. 93

Bayani, J., Zielenska, M., Pandita, A., Al-Romaih, K., Karaskova, J., Harrison, K., Bridge, J. A., Sorensen, P., Thorner, P. & Squire, J. A. (2003). Spectral karyotyping identifies recurrent complex rearrangements of chromosomes 8, 17, and 20 in osteosarcomas. *Genes, Chromosomes and Cancer*, 36, 7–16. 6, 20

Bazazi, S., Buhl, J., Hale, J. J., Anstey, M. L., Sword, G. A., Simpson, S. J. & Couzin, I. D. (2008). Collective motion and cannibalism in locust migratory bands. *Current Biology*, 18, 735–739. ISSN 0960-9822. URL http://www.sciencedirect.com/science/article/pii/S0960982208005216. 8

Bazazi, S., Romanczuk, P., Thomas, S., Schimansky-Geier, L., Hale, J. J., Miller, G. A., Sword, G. A., Simpson, S. J. & Couzin, I. D. (2011). Nutritional state and collective motion: from individuals to mass migration. *Proceedings of the Royal Society B: Biological Sciences*, 278, 356–363. URL https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2010.1447. 8, 8, 60

Bergeland, T., Widerberg, J., Bakke, O. & Nordeng, T. W. (2001). Mitotic partitioning of endosomes and lysosomes. *Current Biology*, 644–651. 4, 5, 90

Biju, V., Itoh, T., Anas, A., Sujith, A. & Ishikawa, M. (2008). Semiconductor quantum dots and metal nanoparticles: syntheses, optical properties, and biological applications. *Analytical and Bioanalytical Chemistry*, 391, 2469–2495. ISSN 1618-2650. URL https://doi.org/10.1007/s00216-008-2185-7. 17, 18, 18

Biju, V., Mundayoor, S., Omkumar, R. V., Anas, A. & Ishikawa, M. (2010). Bioconjugated quantum dots for cancer research: Present status, prospects and remaining issues. *Biotechnology Advances*, 28, 199 – 213. ISSN 0734-9750. URL http://www.sciencedirect.com/science/article/pii/S073497500900192X. 20

Bilan, R., Nabiev, I. & Sukhanova, A. (2016). Quantum dot-based nanotools for bioimaging, diagnostics, and drug delivery. *ChemBioChem*, 17, 2103–2114. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cbic.201600357. 19

Biro, D. & Sasaki, T. (2016). Bringing a Time – Depth Perspective to Collective Animal Behaviour. *Trends in Ecology & Evolution*, 31, 550–562. ISSN 0169-5347. URL http://dx.doi.org/10.1016/j.tree.2016.03.018. 61, 61, 61, 96, 97

Bittner, T., Donnelly, M. & Smith, B. (2004). Individuals , Universals , Collections : On the Foundational Relations of Ontology, 37–48. 7

Bodor, D. L., Pönisch, W., Endres, R. G. & Paluch, E. K. (2020). Of cell shapes and motion: The physical basis of animal cell migration. *Developmental Cell*, 52, 550 – 562. ISSN 1534-5807. URL http://www.sciencedirect.com/science/article/pii/S153458072030109X. 1

Borner, K. K., Krause, S., Mehner, T., Uusi-Heikkilä, S., Ramnarine, I. W. & Krause, J. (2015). Turbidity affects social dynamics in trinidadian guppies. *Behavioral Ecology and Sociobiology*, 69, 645–651. ISSN 1432-0762. URL https://doi.org/10.1007/s00265-015-1875-3. 62

Boukari, F. & Makrogiannis, S. (2016). Joint level-set and spatio-temporal motion detection for cell segmentation. 9. 38

Branson, K. & Belongie, S. (2005). Tracking multiple mouse contours (without too many samples). In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1. ISSN 1063-6919. 14

Briard, L., Dorn, C. & Petit, O. (2015). Personality and affinities play a key role in the organisation of collective movements in a group of domestic horses. *Ethology*, 121, 888–902. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/eth.12402. 8, 60

Brown, H. K., Schiavone, K., Gouin, F., Heymann, M. F. & Heymann, D. (2018). Biology of Bone Sarcomas and New Therapeutic Developments. *Calcified Tissue International*, 102, 174–195. ISSN 1432-0827. URL http://dx.doi.org/10.1007/s00223-017-0372-2. 93

Brown, H. K., Tellez-gabriel, M. & Heymann, D. (2017). Cancer stem cells in osteosarcoma. *Cancer Letters*, 386, 189–195. ISSN 0304-3835. URL http://dx.doi.org/10.1016/j.canlet.2016.11.019. 6

Brown, M. R., Summers, H. D., Rees, P., Chappell, S. C., Silvestre, O. F., Khan, I. A., Smith, P. J. & Errington, R. J. (2010). Long-Term Time Series Analysis of Quantum Dot Encoded Cells By Deconvolution of the Autofluorescence Signal, 925–932. 5, 20, 20, 90

Buades, A., Coll, B. & Morel, J.-M. (2005). A non-local algorithm for image denoising. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2. IEEE. 23, 24, 39, 42, 45

Buhl, J., Sumpter, D. J. T., Couzin, I. D., Hale, J. J., Despland, E., Miller, E. R. & Simpson, S. J. (2006). From disorder to order in marching locusts. *Science*, 312, 1402–1406. ISSN 0036-8075. URL https://science.sciencemag.org/content/312/5778/1402. 59, 61, 62

Burton, A. & Radford, J. (1978). *Thinking in perspective: critical essays in the study of thought processes*, vol. 646. Routledge. ISBN 0416858406. 64

Börger, L. (2016). Editorial: Stuck in motion? reconnecting questions and tools in movement ecology. *Journal of Animal Ecology*, 85, 5–10. URL https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/1365-2656.12464. 10, 99

Camazine, S., Deneubourg, J.-L., Franks, N. R., Sneyd, J., Bonabeau, E. & Theraula, G. (2003). *Self-organization in biological systems*. Princeton university press. 59, 59

Chalfoun, J., Majurski, M., Dima, A., Halter, M., Bhadriraju, K. & Brady, M. (2016). Lineage mapper: A versatile cell and particle tracker. *Scientific Reports*, 6, 36984. URL https://doi.org/10.1038/srep36984http://10.0.4.14/srep36984https://www.nature.com/articles/srep36984{#}supplementary-information. 38

Chan, T. F. & Vese, L. A. (2001). Active Contours Without Edges. 10, 266–277. 38

Chatterjee, R., Ghosh, M., Chowdhury, A. S. & Ray, N. (2013). Cell tracking in microscopic video using matching and linking of bipartite graphs. *Computer Methods and Programs in Biomedicine*, 112, 422–431. ISSN 0169-2607. URL http://dx.doi.org/10.1016/j.cmpb.2013.08.001. 37

Chenouard, N., Smal, I., Chaumont, F. D., Maška, M., Sbalzarini, I. F., Gong, Y., Cardinale, J., Carthel, C., Coraluppi, S., Winter, M. & Cohen, A. R. (2014). Objective comparison of particle tracking methods. 11, 281–290. 93

Chou, W.-H., Liu, K.-L., Shih, Y.-L., Chuang, Y.-Y., Chou, J., Lu, H.-F., Jair, H.-W., Lee, M.-Z., Au, M.-K. & Chung, J.-G. (2018). Ouabain Induces Apoptotic Cell Death Through Caspase- and Mitochondria-dependent Pathways in Human Osteosarcoma U-2 OS Cells. *ANTICANCER RESEARCH*, 38, 169–178. ISSN 0250-7005. 20

Chu, T. C., Shieh, F., Lavery, L. A., Levy, M., Richards-Kortum, R., Korgel, B. A. & Ellington, A. D. (2006). Labeling tumor cells with fluorescent nanocrystal–aptamer bioconjugates. *Biosensors and Bioelectronics*, 21, 1859 – 1866.

ISSN 0956-5663. URL http://www.sciencedirect.com/science/article/pii/S0956566305004033. 19

Clark, P. J. & Evans, F. C. (1954). Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology*, 35, 445–453. URL https://esajournals.onlinelibrary.wiley.com/doi/abs/10.2307/1931034. 59

Codling, E. A., Plank, M. J. & Benhamou, S. (2008). Random walk models in biology, 813–834. 8

Conradt, L. (2012). Models in animal collective decision-making: information uncertainty and conflicting preferences. *Interface Focus*, 2, 226–240. URL https://royalsocietypublishing.org/doi/abs/10.1098/rsfs.2011.0090. 61

Couzin, I. D., Krause, J., James, R., Ruxton, G. D. & Franks, N. R. (2002). Collective Memory and Spatial Sorting in Animal Groups. *Journal of Theoretical Biology*, 218, 1–11. ISSN 0022-5193. URL http://www.sciencedirect.com/science/article/pii/S0022519302930651. 8, 60

Croft, D. P., Krause, J., Couzin, I. D. & Pitcher, T. J. (2003). When fish shoals meet: outcomes for evolution and fisheries. *Fish and Fisheries*, 4, 138–146. ISSN 1467-2960. URL https://doi.org/10.1046/j.1467-2979.2003.00113.x. 61, 97

Cui, R., Pan, H.-C., Zhu, J.-J. & Chen, H.-Y. (2007). Versatile immunosensor using cdte quantum dots as electrochemical and fluorescent labels. *Analytical Chemistry*, 79, 8494–8501. URL https://doi.org/10.1021/ac070923d. PMID: 17927140. 19

Czirók, A. & Vicsek, T. (2000). Collective behavior of interacting self-propelled particles. *Physica A: Statistical Mechanics and its Applications*, 281, 17 – 29. ISSN 0378-4371. URL http://www.sciencedirect.com/science/article/pii/S0378437100000133. 3, 8, 60

Danuser, G. (2011). Essay Computer Vision in Cell Biology. *Cell*, 147, 973–978. ISSN 0092-8674. URL http://dx.doi.org/10.1016/j.cell.2011.11.001. 7

Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory*, 36, 961–1005. ISSN 0018-9448. 69

Debeir, O., Ham, P. V., Kiss, R. & Decaestecker, C. (2005). Tracking of Migrating Cells Under Phase-Contrast Video Microscopy With Combined Mean-Shift Processes. 24, 697–711. 36, 38

Delehanty, J. B., Mattoussi, H. & Medintz, I. L. (2009). Delivering quantum dots into cells: strategies, progress and remaining issues. *Analytical and Bioanalytical Chemistry*, 393, 1091–1105. ISSN 1618-2650. URL https://doi.org/10.1007/s00216-008-2410-4. 21

Dell, A. I., Bender, J. A., Branson, K., Couzin, I. D., Polavieja, G. G. D., Noldus, L. P. J. J., Pe, A., Perona, P., Straw, A. D., Wikelski, M. & Brose, U. (2014). Automated image-based tracking and its application in ecology, 1–12. 11, 14, 14, 14

Demšar, U., Buchin, K., Cagnacci, F., Safi, K., Speckmann, B., Weghe, N. V. D., Weiskopf, D. & Weibel, R. (2015). Analysis and visualisation of movement : an interdisciplinary review. *Movement Ecology*, 1–24. 1, 5, 10, 11, 11, 15, 15

Deneubourg, J. L. & Goss, S. (1989). Collective patterns and decision-making. *Ethology Ecology & Evolution*, 1, 295–311. URL https://doi.org/10.1080/08927014.1989.9525500. 59

Dewan, M. A. A., Ahmad, M. O. & Swamy, M. N. S. (2011). Tracking Biological Cells in Time-Lapse Microscopy : An Adaptive Technique Combining Motion and Topological Features. 58, 1637–1647. 44

Dewan, M. A. A., Ahmad, M. O. & Swamy, M. N. S. (2014). A Method for Automatic Segmentation of Nuclei in Phase-Contrast Images Based on Intensity , Convexity and Texture. 8, 716–728. 36

Dietrich, H. & Schibeci, R. (2003). Beyond public perceptions of gene technology: Community participation in public policy in australia. *Public Understanding of Science*, 12, 381–401. URL https://doi.org/10.1177/0963662503124004. PMID: 14971401. 83

Differential, T. H. E., Of, E., Cell, E., In, O. N., Motility, V. & Malignant, O. F. (2008). THE DIFFERENTIAL EFFECT OF ENDOTHELIAL CELL FACTORS ON IN VITRO MOTILITY OF MALIGNANT AND NON-MALIGNANT. 36, 958–969. 93

Ding, W., Wu, F., Wu, X., Li, S. & Li, H. (2007). Adaptive directional lifting-based wavelet transform for image coding. *IEEE Transactions on Image Processing*, 16, 416–427. 69

Dinh, A.-t., Pangarkar, C., Theofanous, T. & Mitragotri, S. (2006). Theory of Spatial Patterns of Intracellular Organelles. *Biophysical Journal*, 90, L67–L69. ISSN 0006-3495. URL http://dx.doi.org/10.1529/biophysj.106.082875. 4

Dobrucki, J. W., Feret, D. & Noatynska, A. (2007). Scattering of exciting light by live cells in fluorescence confocal imaging: Phototoxic effects and relevance for

frap studies. *Biophysical Journal*, 93, 1778 – 1786. ISSN 0006-3495. URL http://www.sciencedirect.com/science/article/pii/S0006349507714339. 36

Doyle, A. D., Petrie, R. J., Kutys, M. L. & Yamada, K. M. (2013). Dimensions in cell migration. *Current Opinion in Cell Biology*, 25, 642–649. ISSN 0955-0674. URL http://dx.doi.org/10.1016/j.ceb.2013.06.004. 9

Dubertret, B., Skourides, P., Norris, D. J., Noireaux, V., Brivanlou, A. H. & Libchaber, A. (2002). In vivo imaging of quantum dots encapsulated in phospholipid micelles. *Science*, 298, 1759–1762. ISSN 0036-8075. URL https://science.sciencemag.org/content/298/5599/1759. 18

Dunster, K., Hock, B., Sentry, J. W., Service, A. P. & Hospital, A. (2002). Early endosomes , late endosomes , and lysosomes display distinct partitioning strategies of inheritance with similarities to Golgi-derived membranes. 124, 117–124. 90, 90

Duteil, M. (2018). *Collective Behaviour: From Cells to Humans*. Ph.D. thesis, College of Engineering. 10

Dyson, L., Yates, C. A., Buhl, J. & McKane, A. J. (2015). Onset of collective motion in locusts is captured by a minimal model. *Phys. Rev. E*, 92, 052708. URL https://link.aps.org/doi/10.1103/PhysRevE.92.052708. 10, 62

Dzyubachyk, O., Cappellen, W. A. V., Essers, J., Niessen, W. J., Member, S., Meijering, E. & Member, S. (2010). Advanced Level-Set-Based Cell Tracking in Time-Lapse Fluorescence Microscopy. 29, 852–867. 38

Ekdawi, S. N., Jaffray, D. A. & Allen, C. (2016). Nano Today Nanomedicine and tumor heterogeneity : Concept and complex reality. *Nano Today*, 11, 402–414. ISSN 1748-0132. URL http://dx.doi.org/10.1016/j.nantod.2016.06.006. 93

Elliott, J. P., Cowan, I. M. & Holling, C. S. (1977). Prey capture by the african lion. *Canadian Journal of Zoology*, 55, 1811–1828. URL https://doi.org/10.1139/z77-235. 13

Errington, R. J., Brown, M. R., Silvestre, O. F., Njoh, K. L., Chappell, S. C., Khan, I. A., Wilks, S. P., Smith, P. J. & Summers, H. D. (2010). Single cell nanoparticle tracking to model cell cycle dynamics and compartmental inheritance, 121–130. 5, 20, 20

Eskridge, B. & Schlupp, I. (2014). Effects of personality distribution on collective behavior. *The 2019 Conference on Artificial Life*, 908–915. URL https://www.mitpressjournals.org/doi/abs/10.1162/978-0-262-32621-6-ch148. 8, 60

Eskridge, B. E., Valle, E. & Schlupp, I. (2015). Emergence of leadership within a homogeneous group. *PLOS ONE*, 10, 1–17. URL https://doi.org/10.1371/journal.pone.0134222. 61, 61

Fagan, W. F. & Calabrese, J. M. (2014). The Correlated Random Walk and the Rise of Movement Ecology. *Bulletin of the Ecological Society of America*, 95, 204–206. ISSN 0012-9623. URL http://doi.wiley.com/10.1890/0012-9623-95.3.204. 93

Ferguson, E. A., Matthiopoulos, J. & Insall, R. H. (2017). Statistical inference of the mechanisms driving collective cell movement. *Applied Statistics Series C*, 66, 869–890. 5, 5

Flood, C. E. & Wong, M. Y. L. (2017). Social stability in times of change : effects of group fusion and water depth on sociality in a globally invasive fi sh. *Animal Behaviour*, 129, 71–79. ISSN 0003-3472. URL http://dx.doi.org/10.1016/j.anbehav.2017.05.003. 97

Galam, S. & Moscovlclt, S. (1991). Towards a theory of collective phenomena : Consensus and attitude changes in groups. 21, 49–74. 7

Gelblum, A., Pinkoviezky, I., Fonio, E., Ghosh, A., Gov, N. & Feinerman, O. (2015). Ant groups optimally amplify the effect of transiently informed individuals. *Nature Communications*, 6, 1–9. ISSN 20411723. URL http://dx.doi.org/10.1038/ncomms8729. 8, 59, 60

Griesdoorn, V. (2014). *Occult Routes to Chromosomal Instability in Osteosarcoma Cells*. Ph.D. thesis, School of medicine, Cardiff University. 6

GrÜnbaum, D. (1998). Schooling as a strategy for taxis in a noisy environment. *Evolutionary Ecology*, 12, 503–522. ISSN 1573-8477. URL https://doi.org/10.1023/A:1006574607845. 8

Gueron, S. & Levin, S. A. (1993). Self-organization of front patterns in large wildebeest herds. *Journal of Theoretical Biology*, 165, 541 – 552. ISSN 0022-5193. URL http://www.sciencedirect.com/science/article/pii/S0022519383712063. 2

Gurunathan, S., Kang, M.-h. & Qasim, M. (2018). Nanoparticle-Mediated Combination Therapy : Two-in-One Approach for Cancer. 2030, 1–37. 18, 19

Han, Q.-B., Yu, T., Lai, F., Zhou, Y., Feng, C., Wang, W.-N., Fu, X.-H., Lau, C. B.-S., Luo, K. Q., Xu, H.-X., Sun, H.-D., Fung, K.-P. & Leung, P.-C. (2010). Quick identification of apoptosis inducer from isodon eriocalyx by a drug discovery platform composed of analytical high-speed counter-current chromatography and

the fluorescence-based caspase-3 biosensor detection. *Talanta*, 82, 1521 – 1527. ISSN 0039-9140. URL http://www.sciencedirect.com/science/article/pii/S0039914010005667. 19

Hansen, M. J., Buhl, J., Bazazi, S., Simpson, S. J. & Sword, G. A. (2011). Cannibalism in the lifeboat — collective movement in australian plague locusts. *Behavioral Ecology and Sociobiology*, 65, 1715–1720. ISSN 1432-0762. URL https://doi.org/10.1007/s00265-011-1179-1. 8

Hansson, L.-A., Akesson, S. & Åkesson, S. (2014). *Animal movement across scales*. Oxford university press. 99, 99

Hemelrijk, C. K. & Hildenbrandt, H. (2012). Schools of fish and flocks of birds: Their shape and internal structure by self-organization. *Interface Focus*, 2, 726–737. ISSN 20428901. 59

Herbert-Read, J. E. (2016). Understanding how animal groups achieve coordinated movement. *The Journal of experimental biology*, 219, 2971–2983. ISSN 1477-9145. URL https://www.ncbi.nlm.nih.gov/pubmed/27707862https://www.ncbi.nlm.nih.gov/pmc/PMC5091654/. 1, 61

Herbert-Read, J. E., Perna, A., Mann, R. P., Schaerf, T. M., Sumpter, D. J. T. & Ward, A. J. W. (2011). Inferring the rules of interaction of shoaling fish. *Proceedings of the National Academy of Sciences*, 108, 18726 LP – 18731. URL http://www.pnas.org/content/108/46/18726.abstract. 10, 59, 59, 60, 60, 60, 61

Herbert-Read, J. E., Rosén, E., Szorkovszky, A., Ioannou, C. C., Rogell, B., Perna, A., Ramnarine, I. W., Kotrschal, A., Kolm, N., Krause, J. & Sumpter, D. J. T. (2017). How predation shapes the social interaction rules of shoaling fish. *Proceedings of the Royal Society B: Biological Sciences*, 284, 20171126. URL https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2017.1126. 61, 61, 62

Hickey, J. & Davidsen, J. (2019). Self-organization and time-stability of social hierarchies. *PLOS ONE*, 14, 1–30. URL https://doi.org/10.1371/journal.pone.0211403. 96

Hobson, E. A. & DeDeo, S. (2015). Social feedback and the emergence of rank in animal society. *PLOS Computational Biology*, 11, 1–20. URL https://doi.org/10.1371/journal.pcbi.1004411. 95, 97

Horan, Paul Karl, Melnicoff, M. J., Jensen, B. D. & Slezak, S. E. (1990). Fluorescent cell labeling for in vivo and in vitro cell tracking. *Methods in Cell Biology*, 33. 9, 36, 36

Hughey, L. F., Hein, A. M., Strandburg-Peshkin, A. & Jensen, F. H. (2018). Challenges and solutions for studying collective animal behaviour in the wild. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373, 20170005. URL https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2017.0005. 11

Ioannou, C. C., Ramnarine, I. W. & Torney, C. J. (2017). High-predation habitats affect the social dynamics of collective exploration in a shoaling fish. *Science Advances*, 3. URL https://advances.sciencemag.org/content/3/5/e1602682. 61

Ioannou, C. C., Singh, M. & Couzin, I. D. (2015). Potential Leaders Trade Off Goal-Oriented and Socially Oriented Behavior in Mobile Animal Groups. *The American Naturalist*, 186, 284–293. ISSN 00030147, 15375323. 61, 61

Itseez (2015). Open source computer vision library. https://github.com/itseez/opencv. 24, 39, 64, 84

Jacoby, D. M. & Freeman, R. (2016). Emerging Network-Based Tools in Movement Ecology. *Trends in Ecology and Evolution*, 31, 301–314. ISSN 01695347. URL http://dx.doi.org/10.1016/j.tree.2016.01.011. 99

Jaiswal, J. K., Mattoussi, H., Mauro, J. M. & Simon, S. M. (2003). Long-term multiple color imaging of live cells using quantum dot bioconjugates. *Nature Biotechnology*, 21, 47–51. ISSN 1546-1696. URL https://doi.org/10.1038/nbt767. 19

Jia, X., Pei, M., Zhao, X., Tian, K., Zhou, T. & Liu, P. (2016). PEGylated Oxidized Alginate-DOX Prodrug Conjugate Nanoparticles Cross-Linked with Fluorescent Carbon Dots for Tumor Theranostics. *ACS Biomaterials Science & Engineering*, 2, 1641–1648. URL https://doi.org/10.1021/acsbiomaterials.6b00443. 19, 20

Jiang, L., Giuggioli, L., Perna, A., Escobedo, R., Lecheval, V., Sire, C., Han, Z. & Theraulaz, G. (2017). Identifying influential neighbors in animal flocking. *PLOS Computational Biology*, 13, 1–32. URL https://doi.org/10.1371/journal.pcbi.1005822. 8, 60, 95

Jiménez-morales, N., Mendoza-ángeles, K., Porras-villalobos, M., Ibarra-coronado, E., Roldán-roldán, G. & Hernández-falcón, J. (2018). Neurobiology of Learning and Memory Who is the boss ? Individual recognition memory and social hierarchy formation in cray fi sh. *Neurobiology of Learning and Memory*, 147, 79–89. ISSN 1074-7427. URL https://doi.org/10.1016/j.nlm.2017.11.017. 95, 97

Johnstone, R. A. & Manica, A. (2011). Evolution of personality differences in leadership. *Proceedings of the National Academy of Sciences*, 108, 8373–8378. ISSN 0027-8424. URL https://www.pnas.org/content/108/20/8373. 8, 60, 61

Jolles, J. W., Fleetwood-wilson, A., Nakayama, S., Stumpe, M. C., Johnstone, R. A. & Manica, A. (2015). The role of social attraction and its link with boldness in the collective movements of three-spined sticklebacks. *Animal Behaviour*, 99, 147–153. ISSN 0003-3472. URL http://dx.doi.org/10.1016/j.anbehav.2014.11.004. 8, 60

Jolles, J. W., Laskowski, K. L., Boogert, N. J. & Manica, A. (2018). Repeatable group differences in the collective behaviour of stickleback shoals across ecological contexts. *Proceedings of the Royal Society B: Biological Sciences*, 285, 20172629. URL https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2017.2629. 61, 63

Jones, R. E. (1977). Movement Patterns and Egg Distribution in Cabbage Butterflies. *The Journal of Animal Ecology*, 46, 195–212. 1

Jones, T. R., Kang, I. H., Wheeler, D. B., Lindquist, R. A., Papallo, A., Sabatini, D. M., Golland, P. & Carpenter, A. E. (2008). CellProfiler Analyst: data exploration and analysis software for complex image-based screens. *BMC Bioinformatics*, 9, 482. ISSN 1471-2105. URL https://doi.org/10.1186/1471-2105-9-482. 38

Kan, A., Chakravorty, R., Bailey, J., Leckie, C., Markham, J. & Dowling, M. R. (2011). Automated and semi-automated cell tracking : addressing. *Journal ofMicroscopy*, 244, 194–213. 35

Kaphle, A., Akhela, P. N. N., Hemant, U. & Daima, K. (2018). Nanomaterials for agriculture , food and environment : applications , toxicity and regulation. *Environmental Chemistry Letters*, 16, 43–58. ISSN 1610-3661. 18, 18

Katz, Y., Tunstrøm, K., Ioannou, C. C., Huepe, C. & Couzin, I. D. (2011). Inferring the structure and dynamics of interactions in schooling fish. *Proceedings of the National Academy of Sciences*, 108, 18720 LP – 18725. URL http://www.pnas.org/content/108/46/18720.abstract. 59, 60, 60, 60, 61

Kelley, J. L., Morrell, L. J., Inskip, C., Krause, J. & Croft, D. P. (2011). Predation risk shapes social networks in fission-fusion populations. *PLOS ONE*, 6, 1–10. URL https://doi.org/10.1371/journal.pone.0024280. 62, 62

King, A. J., Fehlmann, G., Biro, D., Ward, A. J. & Fürtbauer, I. (2018). Re-wilding collective behaviour: An ecological perspective. *Trends in Ecology & Evolution*, 33, 347 – 357. ISSN 0169-5347. URL http://www.sciencedirect.com/science/article/pii/S0169534718300557. 60

Kirtane, A. R., Kalscheuer, S. M. & Panyam, J. (2013). Exploiting nanotechnology to overcome tumor drug resistance : Challenges and opportunities . *Advanced Drug*

Delivery Reviews, 65, 1731–1747. ISSN 0169-409X. URL http://dx.doi.org/10.1016/j.addr.2013.09.001. 91

Kostelec, P. D., Carlin, L. M. & Glocker, B. (2015). LEARNING TO DETECT AND TRACK CELLS FOR QUANTITATIVE ANALYSIS OF TIME-LAPSE MICROSCOPIC IMAGE SEQUENCES Biomedical Image Analysis Group , Department of Computing Leukocyte Biology Section , National Heart & Lung Institute Imperial College London , UK, 1544–1547. 37

Kramer, N., Walzl, A., Unger, C., Rosner, M., Krupitza, G. & Hengstschla, M. (2013). Mutation Research / Reviews in Mutation Research In vitro cell migration and invasion assays. *Mutation research*, 752, 10–24. 9

Krause, J., Hoare, D., Krause, S., Hemelrijk, C. K. & Rubenstein, D. I. (2000). Leadership in fish shoals. *Fish and Fisheries*, 1, 82–89. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-2979.2000.tb00001.x. 2, 8, 8, 60, 60, 60, 60, 60, 60, 60, 61

Krause, J., Krause, S., Arlinghaus, R., Psorakis, I., Roberts, S. & Rutz, C. (2013). Reality mining of animal social systems. *Trends in Ecology & Evolution*, 1–11. ISSN 0169-5347. URL http://dx.doi.org/10.1016/j.tree.2013.06.002. 10

Krause, J., Reeves, P. & Hoare, D. (1998). Positioning behaviour in roach shoals: The role of body length and nutritional state. *Behaviour*, 135, 1031–1039. ISSN 00057959. URL http://www.jstor.org/stable/4535577. 60

Krishnamoorti, R. (2006). Extracting the Benefits of Nanotechnology for the Oil Industry. 18

Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82, 1–26. 75

Lackie, J. M. (1986). *Cell movement and cell behaviour*. Allen & Unwin. ISBN 9780045740352. 5, 5, 5

Laird, N. M. & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38, 963–974. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2529876. 74

Laissue, P. P., Alghamdi, R. A., Tomancak, P., Reynaud, E. G. & Shroff, H. (2017). Assessing phototoxicity in live fluorescence imaging. *Nature Methods*, 14, 657–661. ISSN 1548-7105. URL https://doi.org/10.1038/nmeth.4344. 36

Laitinen, M., Sorsa, T., Salo, T., Halttunen, T., Konttinen, Y. T., Lindqvist, C. & Teronen, O. (1997). Osteosarcoma cell lines MG63, U-2 OS and SAOS-2 produce latent and proteolytically activated MMP-2. *JOURNAL OF BONE AND MINERAL RESEARCH*, 12, S313. ISSN 0884-0431. 20

Lee, C. F. & Wurtz, J. D. (2019). Novel physics arising from phase transitions in biology. 7

Li, C., Xu, C., Member, S., Gui, C. & Fox, M. D. (2010a). Distance Regularized Level Set Evolution and Its Application to Image Segmentation. 38

Li, F., Zhou, X., Ma, J. & Wong, S. T. C. (2010b). for Quantitative Cancer Cell Cycle Analysis. 29, 96–105. 37

Li, K., Miller, E. D., Chen, M., Kanade, T., Weiss, L. E. & Campbell, P. G. (2008). Cell population tracking and lineage construction with spatiotemporal context. *Medical Image Analysis*, 12, 546 – 566. ISSN 1361-8415. URL http://www.sciencedirect.com/science/article/pii/S1361841508000650. Special issue on the 10th international conference on medical imaging and computer assisted intervention - MICCAI 2007. 13, 37, 37

Li, Q., Rycaj, K., Chen, X. & Tang, D. G. (2015). Ac ce p te cr t. *Seminars in Cancer Biology*. ISSN 1044-579X. URL http://dx.doi.org/10.1016/j.semcancer.2015.07.002. 6

Lin, Y., Zhang, L., Yao, W., Qian, H., Ding, D., Wu, W. & Jiang, X. (2011). Water-soluble chitosan-quantum dot hybrid nanospheres toward bioimaging and biolabeling. *ACS Applied Materials & Interfaces*, 3, 995–1002. URL https://doi.org/10.1021/am100982p. PMID: 21388158. 19

Liu, J. S. (2008). *Monte Carlo strategies in scientific computing*. Springer Science & Business Media. 13

Lord, W. M., Sun, J., Ouellette, N. T. & Bollt, E. M. (2016). Inference of Causal Information Flow in Collective Animal Behavior. *arXiv e-prints*, arXiv:1606.01932. 8, 60

Lucas, B. D. & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In: *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI'81. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. URL http://dl.acm.org/citation.cfm?id=1623264.1623280. 65

Majumdar, R., Sixt, M. & Parent, C. (2014). New paradigms in the establishment and maintenance of gradients during directed cell migration. *Curr Opin Cell Biol*, 33–40. 5

Makris, N. C., Ratilal, P., Symonds, D. T., Jagannathan, S., Lee, S. & Nero, R. W. (2006). Fish Population and Behavior Revealed by Instantaneous Continental Shelf – Scale Imaging. 311, 660–663. 10

Marras, S., Killen, S. S., Lindström, J., McKenzie, D. J., Steffensen, J. F. & Domenici, P. (2015). Fish swimming in schools save energy regardless of their spatial position. *Behavioral Ecology and Sociobiology*, 69, 19–226. ISSN 03405443. 60

Masuzzo, P., Troys, M. V., Ampe, C. & Martens, L. (2015). Taking Aim at Moving Targets in Computational Cell Migration. *Trends in Cell Biology*, xx, 1–23. ISSN 0962-8924. URL `http://dx.doi.org/10.1016/j.tcb.2015.09.003`. 12, 12, 12, 12, 12, 13, 13, 13, 13

Masuzzo, P., Troys, M. V., Ampe, C. & Martens, L. (2016). Taking Aim at Moving Targets in Computational Cell Migration. *Trends in Cell Biology*, 26, 88–110. ISSN 0962-8924. URL `http://dx.doi.org/10.1016/j.tcb.2015.09.003`. 9, 9, 9, 37, 38

Matlab team (2017). *MATLAB version 9.3.0.713579 (R2017b)*. The Mathworks, Inc., Natick, Massachusetts. 71

Mazzamuto, M. V., Cremonesi, G., Santicchia, F., Preatoni, D., Martinoli, A., Wauters, L. A., Vittoria, M., Cremonesi, G., Santicchia, F., Preatoni, D., Martinoli, A. & Wauters, L. A. (2018). Rodents in the arena : a critical evaluation of methods measuring personality traits. *Ethology Ecology & Evolution*, 00, 1–21. ISSN 0394-9370. URL `https://doi.org/10.1080/03949370.2018.1488768`. 94

Mchugh, K. J., Jing, L., Behrens, A. M., Jayawardena, S., Tang, W., Gao, M., Langer, R. & Jaklenec, A. (2018). Biocompatible Semiconductor Quantum Dots as Cancer Imaging Agents. 1706356, 1–18. 19

Meijering, E. (2012). Cell Segmentation: 50 Years Down the Road. *IEEE Signal Processing Magazine*, 29, 140–145. 35, 36

Merkle, J. A., Sigaud, M. & Fortin, D. (2015). To follow or not? How animals in fusion–fission societies handle conflicting information during group decision-making. *Ecology Letters*, 18, 799–806. ISSN 1461-023X. URL `https://doi.org/10.1111/ele.12457`. 62, 62

Min, S. & Wan, J. W. L. (2013). A multiscale graph cut approach to bright-field multiple cell image segmentation using a Bhattacharyya Measure. 8669, 1–8. 44

Ming-Kuei Hu (1962). Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory*, 8, 179–187. 48

Mirzayans, R., Andrais, B. & Murray, D. (2018). Roles of Polyploid / Multinucleated Giant Cancer Cells in Metastasis and Disease Relapse Following Anticancer Treatment. *Cancers*, 10. 6, 6

Mirzayans, R., Andrais, B., Scott, A., Wang, Y. W. & Kumar, P. (2017). Multinucleated Giant Cancer Cells Produced in Response to Ionizing Radiation Retain Viability and Replicate Their Genome. 6

Murakami, H., Niizato, T. & Gunji, Y.-p. (2017). Emergence of a coherent and cohesive swarm based on mutual anticipation. *Nature Publishing Group*, 1–9. URL `http://dx.doi.org/10.1038/srep46447`. 62

Méhes, E. & Vicsek, T. (2014). Collective motion of cells: from experiments to models. *Integrative Biology*, 6, 831–854. ISSN 1757-9708. URL `https://doi.org/10.1039/c4ib00115j`. 1

Nadler, L. E., Killen, S. S., McCormick, M. I., Watson, S.-A. & Munday, P. L. (2016). Effect of elevated carbon dioxide on shoal familiarity and metabolism in a coral reef fish. *Conservation Physiology*, 4. ISSN 2051-1434. URL `https://doi.org/10.1093/conphys/cow052`. Cow052. 62

Nagy, M., Ákos, Z., Biro, D. & Vicsek, T. (2010). Hierarchical group dynamics in pigeon flocks. *Nature*, 464, 890–893. ISSN 00280836. 59, 69, 95

Nagy, M., Vásárhelyi, G., Pettit, B., Roberts-mariani, I., Vicsek, T. & Biro, D. (2013). Context-dependent hierarchies in pigeons. 1

Nakayama, S., Harcourt, J. L., Johnstone, R. A. & Manica, A. (2016). Who directs group movement? leader effort versus follower preference in stickleback fish of different personality. *Biology Letters*, 12, 20160207. URL `https://royalsocietypublishing.org/doi/abs/10.1098/rsbl.2016.0207`. 8, 60, 61

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D. & Smouse, P. E. (2008). movement research. 105, 19052–19059. 1, 1

Nicolis, G. & Prigogine, I. A. (1977). Self-organization in nonequilibrium systems: From dissipative structures to order through fluctuations. 59

Nketia, T. A., Sailem, H., Rohde, G., Machiraju, R. & Rittscher, J. (2017). Analysis of live cell images : Methods , tools and opportunities. *Methods*, 115, 65–79. ISSN 1046-2023. URL http://dx.doi.org/10.1016/j.ymeth.2017.02.007. 12, 12, 13, 13, 35, 37

Ostner, J., Seltmann, A., Majolo, B. & Schu, O. (2013). The Organization of Collective Group Movements in Wild Barbary Macaques (Macaca sylvanus): Social Structure Drives Processes of Group Coordination in Macaques. 8. 1

Parr, A. (1927). *A Contribution to the Theoretical Analysis of the Schooling Behavior of Fishes*. Occasional Papers of the Bingham Oceanographic Collection. URL https://books.google.gr/books?id=kAAEtwAACAAJ. 8, 60

Parrish, J. K. & Edelstein-Keshet, L. (1999a). Complexity, pattern, and evolutionary trade-offs in animal aggregation. *Science*, 284, 99–101. ISSN 0036-8075. URL https://science.sciencemag.org/content/284/5411/99. 2

Parrish, J. K. & Edelstein-Keshet, L. (1999b). Complexity, Pattern, and Evolutionary Trade-Offs in Animal Aggregation. *Science*, 284, 99 LP – 101. URL http://science.sciencemag.org/content/284/5411/99.abstract. 59

Paul, C. D., Mistriotis, P. & Konstantopoulos, K. (2016). Cancer cell motility: lessons from migration in confined spaces. *Nature Reviews. Cancer*. ISSN 1474-175X. URL http://dx.doi.org/10.1038/nrc.2016.123. 6, 7

Peng, X., Schlamp, M. C., Kadavanich, A. V. & Alivisatos, A. P. (1997). Epitaxial growth of highly luminescent cdse/cds core/shell nanocrystals with photostability and electronic accessibility. *Journal of the American Chemical Society*, 119, 7019–7029. URL https://doi.org/10.1021/ja970754m. 18, 18

Petrie, R. J., Doyle, A. D. & Yamada, K. M. (2009). Random versus directionally persistent cell migration. *Nature Reviews Molecular Cell Biology*, 10, 538–549. ISSN 1471-0080. URL https://doi.org/10.1038/nrm2729. 1

Ponten, J. & Saksela, E. (1967). Two established in vitro cell lines from human mesenchymal tumours. *International Journal of Cancer*, 2, 434–447. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ijc.2910020505. 20, 21

Puck, T. T., Marcus, P. I. & Cieciura, S. J. (1956). Clonal growth of mammalian cells in vitro. *Journal of Experimental Medicine*, 103, 273–284. ISSN 0022-1007. URL http://jem.rupress.org/content/103/2/273. 6

Pyke, G. H. (2015). Understanding movements of organisms : it ' s time to vy foraging hypothesis abandon the L e, 1–16. 1, 99

Quera, V., Beltran, F. S. & Gimeno, E. (2016). Modelling the emergence of coordinated collective motion by minimizing dissatisfaction. *Mathematical Biosciences*, 271, 154 – 167. ISSN 0025-5564. URL http://www.sciencedirect.com/science/article/pii/S0025556415002394. 61

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.r-project.org/. 71, 75

R Core Team (2017). R: {{a language and environment for statistical computing, version 3.4.0}}. \url{https://www.R-project.org/}. 49

Raile, K., Hoflich, A., Kessler, U., Yang, Y., Pfuender, M., Blum, W. F., Kolb, H., Schwarz, H. P. & Kiess, W. (1994). Human osteosarcoma (U-2 OS) cells express both insulin-like growth-factor-I (IGF-1) receptors and insulin-like growth-factor-II mannose-6-phosphate (IGF-II/MGP) receptors and synthesize IGF-II -autocrine growth-stimulation by IGF-II via the IGF-I receptor. *Journal of Cellular Physiology*, 159, 531–541. ISSN 0021-9541. 20

Rands, S. A. (2011). The effects of dominance on leadership and energetic gain: A dynamic game between pairs of social foragers. *PLOS Computational Biology*, 7, 1–10. URL https://doi.org/10.1371/journal.pcbi.1002252. 60

Raoult, V., Tosetto, L. & Williamson, J. E. (2018). Drone-based high-resolution tracking of aquatic vertebrates. *Drones*, 2. ISSN 2504-446X. URL https://www.mdpi.com/2504-446X/2/4/37. 11

Ray, N. & Acton, S. T. (2002). Active contours for cell tracking. In: *Proceedings Fifth IEEE Southwest Symposium on Image Analysis and Interpretation*. 13

Rees, P., Brown, M. R., Summers, H. D., Holton, M. D., Errington, R. J., Chappell, S. C. & Smith, P. J. (2011). A transfer function approach to measuring cell inheritance. 5

Renukaiah, G., Mallappa, R. & Swamy, K. (2018). Potential applications of engineered nanoparticles in medicine and biology : an update. *JBIC Journal of Biological Inorganic Chemistry*. ISSN 1432-1327. URL https://doi.org/10.1007/s00775-018-1600-6. 18

Resch-genger, U., Grabolle, M., Cavaliere-jaricot, S., Nitschke, R. & Nann, T. (2008). Quantum dots versus organic dyes as fluorescent labels. *Nature Methods*, 5, 763–775. 9, 36, 36

Reynolds, A. M. & Ouellette, N. T. (2016). Swarm dynamics may give rise to Lévy flights. *Nature Publishing Group*, 1–8. URL http://dx.doi.org/10.1038/srep30515. 5

Robie, A. A., Seagraves, K. M., Egnor, S. E. R. & Branson, K. (2017). Machine vision methods for analyzing social interactions. *Journal of Experimental Biology*, 220, 25–34. ISSN 0022-0949. URL https://jeb.biologists.org/content/220/1/25. 10

Rosenthal, S. B., Twomey, C. R., Hartnett, A. T., Wu, H. S. & Couzin, I. D. (2015). Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioral contagion. *Proceedings of the National Academy of Sciences*, 112, 4690–4695. ISSN 0027-8424. URL https://www.pnas.org/content/112/15/4690. 61

Rother, C. & Blake, A. (2004). " GrabCut " — Interactive Foreground Extraction using Iterated Graph Cuts. 1, 309–314. 44

Russell, J. C., Hanks, E. M. & Haran, M. (2015). Dynamic Models of Animal Movement with Spatial Point Process Interactions. 1

Russell, J. C., Hanks, E. M., Modlmeier, A. P. & Hughes, D. P. (2017). Modeling Collective Animal Movement Through Interactions in Behavioral States. *Journal of Agricultural, Biological and Environmental Statistics*. ISSN 1537-2693. 1

Sazonov, I., Xie, X. & Nithiarasu, P. (2016). Computer Methods in Biomechanics and Biomedical Engineering : Imaging & Visualization An improved method of computing geometrical potential force (GPF) employed in the segmentation of 3D and 4D medical images ABSTRACT. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 1163, 1–10. ISSN 2168-1163. URL http://dx.doi.org/10.1080/21681163.2015.1124027. 38

Sbalzarini, I. F. (2016). *Seeing Is Believing: Quantifying Is Convincing: Computational Image Analysis in Biology*. Springer International Publishing, Cham. ISBN 978-3-319-28549-8, pp. 1–39. URL https://doi.org/10.1007/978-3-319-28549-8_1. 35, 36

Schaerf, T. M., Herbert-Read, J. E., Myerscough, M. R., Sumpter, D. J. T. & Ward, A. J. W. (2016). Identifying differences in the rules of interaction between individuals in moving animal groups. *arXiv e-prints*, arXiv:1601.08202. 60

Schäfer, M. S. (2009a). From Public Understanding to Public Engagement: An Empirical Assessment of Changes in Science Coverage. *Science Communication*, 30, 475–505. URL https://doi.org/10.1177/1075547008326943. 83

Schäfer, M. S. (2009b). *to Public Engagement An Empirical Assessment of Changes in Science Coverage*. ISBN 1075547008. 83

Schibeci, R., Harwood, J. & Dietrich, H. (2006). Community involvement in biotechnology policy?: The australian experience. *Science Communication*, 27, 429–445. URL https://doi.org/10.1177/1075547005285066. 83

Schick, R. S., Loarie, S. R., Colchero, F., Best, B. D., Boustany, A., Conde, D. A., Halpin, P. N., Joppa, L. N., McClellan, C. M. & Clark, J. S. (2008). Understanding movement data and movement processes: Current and emerging directions. *Ecology Letters*, 11, 1338–1350. ISSN 1461023X. 99

Seeley, T. D. & Buhrman, S. C. (1999). Group decision making in swarms of honey bees. *Behavioral Ecology and Sociobiology*, 45, 19–31. ISSN 1432-0762. URL https://doi.org/10.1007/s002650050536. 8

Selmeczi, D., Li, L., Nrrelykke, S. F., Hagedorn, P. H., Mosler, S., Larsen, N. B., Cox, E. C. & Flyvbjerg, H. (2008). Cell motility as random motion : A review Cell motility as random motion. 15, 1–15. 6

Sih, A., Mathot, K. J., Moirón, M., Montiglio, P.-O., Wolf, M. & Dingemanse, N. J. (2015). Animal personality and state–behaviour feedbacks: a review and guide for empiricists. *Trends in Ecology and Evolution*, 30, 50 – 60. ISSN 0169-5347. URL http://www.sciencedirect.com/science/article/pii/S0169534714002432. 94

Simpson, S. J., Sword, G. A., Lorch, P. D. & Couzin, I. D. (2006). Cannibal crickets on a forced march for protein and salt. *Proceedings of the National Academy of Sciences*, 103, 4152–4156. ISSN 0027-8424. URL https://www.pnas.org/content/103/11/4152. 8

Singh, N. J., Allen, A. M. & Ericsson, G. (2016). Quantifying migration behaviour using net squared displacement approach: Clarifications and caveats. *PLOS ONE*, 11, 1–20. URL https://doi.org/10.1371/journal.pone.0149594. 93, 93, 93

Smith, A. M., Gao, X. & Nie, S. (2004). Quantum dot nanocrystals for in vivo molecular and cellular imaging. *Photochemistry and Photobiology*, 80, 377–385. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1751-1097.2004.tb00102.x. 17, 17

Souza, A. G., Silva, I. B. B., Campos-Fernandez, E., Barcelos, L. S., Souza, J. B., Marangoni, K., Alonso-Goulart, L. R. G. & Vivian (2018). Comparative Assay of 2D and 3D Cell Culture Models: Proliferation, Gene Expression and Anticancer Drug Response. URL http://www.eurekaselect.com/node/160997/article. 9

Specht, E. A., Braselmann, E. & Palmer, A. E. (2016). A Critical and Comparative Review of Fluorescent Tools for Live-Cell Imaging. *Annual Review of Physiology*, 79, 93–117. 9, 9, 36, 36

Srinivas, K. R. (2017). Why Public Engagement Matters in Science. *Trends in Biotechnology*, xx, 1–2. ISSN 0167-7799. URL http://dx.doi.org/10.1016/j.tibtech.2017.02.004. 83

Strömbom, D. (2011). Collective motion from local attraction. *Journal of Theoretical Biology*, 283, 145 – 151. ISSN 0022-5193. URL http://www.sciencedirect.com/science/article/pii/S002251931100261X. 60

Summers, H. D., Rees, P., Holton, M. D., Brown, M. R., Chappell, S. C., Smith, P. J. & Errington, R. J. (2011). Statistical analysis of nanoparticle dosing in a dynamic cellular system. 6. 5, 21, 90

Sumpter, D. J. T. & Sumpter, D. J. T. (2006). The principles of collective animal behaviour The principles of collective animal behaviour. 59

Svensson, C.-m., Medyukhina, A., Belyaev, I., Al-zaben, N. & Figge, M. T. (2018). Untangling Cell Tracks : Quantifying Cell Migration by Time Lapse Image Data Analysis. *Cytometry Part A*, 93A, 357–370. 6, 6

Thirusittampalam, K., Hossain, M. J., Ghita, O., Whelan, F. & Member, S. (2013). A Novel Framework for Cellular Tracking and Mitosis Detection in Dense Phase Contrast Microscopy Images. 17, 642–653. 36, 37

Thomas, N. & Goodyer, I. (2003). Stealth sensors: real-time monitoring of the cell cycle. *Targets*, 2, 26–33. 20

Thomas, P. (2017). Making sense of snapshot data : ergodic principle for clonal cell populations. 92

Thompson, C. & Shure, L. (1995). *Image Processing Toolbox: For Use with MATLAB;[user's Guide]*. MathWorks. 73

Tinbergen, N. (1963). On aims and methods of Ethology. *Zeitschrift für Tierpsychologie*, 20 (4), 410–433. 1

Torney, C. J., Hopcraft, J. G. C., Morrison, T. A., Couzin, I. D. & Levin, S. A. (2018a). From single steps to mass migration: the problem of scale in the movement ecology of the Serengeti wildebeest. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 373, 20170012. ISSN 1471-2970. URL https://www.ncbi.nlm.nih.gov/pubmed/29581397https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5882982/. 2, 2, 99

Torney, C. J., Lamont, M., Debell, L., Angohiatok, R. J., Leclerc, L.-M. & Berdahl, A. M. (2018b). Inferring the rules of social interaction in migrating caribou. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373, 20170385. URL https://royalsocietypublishing.org/doi/abs/10.1098/rstb.2017.0385. 11

Trepat, X., Wasserman, M. R., Angelini, T. E., Millet, E., Weitz, D. A., Butler, J. P. & Fredberg, J. J. (2009). Physical forces during collective cell migration. *Nature Physics*, 5, 426. URL https://doi.org/10.1038/nphys1269http://10.0.4.14/nphys1269https://www.nature.com/articles/nphys1269{#}supplementary-information. 12

Trybus, K. M. (2013). Intracellular Transport : The Causes for Pauses. *CURBIO*, 23, R623–R625. ISSN 0960-9822. URL http://dx.doi.org/10.1016/j.cub.2013.06.005. 1, 4, 4

Tummolini, L., Castelfranchi, C., Bottazzi, E., Catenacci, C., Gangemi, A. & Lehmann, J. (2006). From collective intentionality to intentional collectives : An ontological perspective. 7, 192–208. 7

Tunstrøm, K., Katz, Y., Ioannou, C. C., Huepe, C., Lutz, M. J. & Couzin, I. D. (2013). Collective states, multistability and transitional behavior in schooling fish. *PLOS Computational Biology*, 9, 1–11. URL https://doi.org/10.1371/journal.pcbi.1002915. 1, 3, 8, 10, 60, 61, 62

Ulman, V., Maška, M., Magnusson, K. E. G., Ronneberger, O., Haubold, C., Harder, N., Matula, P., Matula, P., Svoboda, D., Radojevic, M., Smal, I., Rohr, K., Jaldén, J., Blau, H. M., Dzyubachyk, O., Lelieveldt, B., Xiao, P., Li, Y., Cho, S.-y., Dufour, A. C., Reyes-aldasoro, C. C., Solis-lemus, J. A., Bensch, R., Brox, T., Stegmaier, J., Mikut, R., Wolf, S., Hamprecht, F. A., Esteves, T., Quelhas, P. & Demirel, Ö. (2017). An objective comparison of cell-tracking algorithms. *Nature Publishing Group*. ISSN 1548-7091. URL http://dx.doi.org/10.1038/nmeth.4473. 13

Vicsek, T. & Zafeiris, A. (2012). Collective motion. *Physics Reports*, 517, 71–140. ISSN 03701573. URL http://dx.doi.org/10.1016/j.physrep.2012.03.004. 7, 8, 8

Viscido, S. V., Parrish, J. K. & Grünbaum, D. (2004). Individual behavior and emergent properties of fish schools: A comparison of observation and theory. *Marine Ecology Progress Series*, 273, 239–249. ISSN 01718630. 59, 60

Wang, C., Chen, X., Xie, G. & Cao, M. (2017). Emergence of leadership in a robotic fish group under diverging individual personality traits. *Royal Society Open Science*,

4, 161015. URL https://royalsocietypublishing.org/doi/abs/10.1098/rsos.161015. 61

Wang, Q., Niemi, J., Tan, C.-m., You, L. & West, M. (2010). Image Segmentation and Dynamic Lineage Analysis in Single-Cell Fluorescence Microscopy. 37

Wang, Y.-Q., Ye, C., Zhu, Z.-H. & Hu, Y.-Z. (2008). Cadmium telluride quantum dots as ph-sensitive probes for tiopronin determination. *Analytica Chimica Acta*, 610, 50 – 56. ISSN 0003-2670. URL http://www.sciencedirect.com/science/article/pii/S000326700800086X. 19

Ward, A. J. W., Herbert-Read, J. E., Schaerf, T. M. & Seebacher, F. (2018). The physiology of leadership in fish shoals: leaders have lower maximal metabolic rates and lower aerobic scope. *Journal of Zoology*, 305, 73–81. URL https://zslpublications.onlinelibrary.wiley.com/doi/abs/10.1111/jzo.12534. 80, 94, 95

Watts, I., Nagy, M., Burt de Perera, T. & Biro, D. (2016). Misinformed leaders lose influence over pigeon flocks. *Biology Letters*, 12, 20160544. URL https://royalsocietypublishing.org/doi/abs/10.1098/rsbl.2016.0544. 8, 8, 59, 60, 60

Webster, M. M. & Laland, K. N. (2009). Evaluation of a non-invasive tagging system for laboratory studies using three-spined sticklebacks Gasterosteus aculeatus. 44, 1868–1873. 14, 63

Weinstein, B. G. (2018). A computer vision for animal ecology. *Journal of Animal Ecology*, 87, 533–545. URL https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/1365-2656.12780. 10

Weiss, M., Elsner, M., Kartberg, F. & Nilsson, T. (2004). Anomalous subdiffusion is a measure for cytoplasmic crowding in living cells. *Biophysical Journal*, 87, 3518 – 3524. ISSN 0006-3495. URL http://www.sciencedirect.com/science/article/pii/S0006349504738163. 57, 92

Westley, P. A. H., Berdahl, A. M., Torney, C. J., Biro, D. & Berdahl, A. M. (2018). Collective movement in ecology : from emerging technologies to conservation and management. 2, 2, 7, 8

Wood, Z. & Galton, A. (2009). A taxonomy of collective phenomena. 4, 267–292. 7

Wu, X., Liu, H., Liu, J., Haley, K. N., Treadway, J. A., Larson, J. P., Ge, N., Peale, F. & Bruchez, M. P. (2003). Immunofluorescent labeling of cancer marker Her2 and other cellular targets with semiconductor quantum dots. *Nature Biotechnology*, 21, 41–46. ISSN 1546-1696. URL https://doi.org/10.1038/nbt764. 17

Xu, H., Sha, M. Y., Wong, E. Y., Uphoff, J., Xu, Y., Treadway, J. A., Truong, A., O'Brien, E., Asquith, S., Stubbins, M., Spurr, N. K., Lai, E. H. & Mahoney, W. (2003). Multiplexed SNP genotyping using the Qbead system: a quantum dot-encoded microsphere-based assay. *Nucleic acids research*, 31, e43–e43. ISSN 1362-4962. URL https://www.ncbi.nlm.nih.gov/pubmed/12682378https://www.ncbi.nlm.nih.gov/pmc/articles/PMC153755/. 19

Yan, G.-n., Lv, Y.-f. & Guo, Q.-n. (2016). Advances in osteosarcoma stem cell research and opportunities for novel therapeutic targets. *Cancer Letters*, 370, 268–274. ISSN 0304-3835. URL http://dx.doi.org/10.1016/j.canlet.2015.11.003. 93

Yao, J., Li, P., Li, L. & M, Y. (2018). Biochemistry and biomedicine of quantum dots: from biodetection to bioimaging, drug discovery, diagnostics, and therapy. *Acta Biomaterialia*, 74, S313. ISSN 1742-7061. 18, 19, 19

Yeo, S. Y., Xie, X., Sazonov, I. & Nithiarasu, P. (2011). Geometrically Induced Force Interaction for Three-Dimensional Deformable Models. 20, 1373–1387. 38

Yizong Cheng (1995). Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17, 790–799. 45

Youssef, S., Gude, S. & Ra, J. O. (2011). Integrative Biology, 1095–1101. 37

Zhang, B., Hu, Y. & Pang, Z. (2017a). Modulating the Tumor Microenvironment to Enhance Tumor Nanomedicine Delivery. 8, 1–16. 18

Zhang, L., Wu, C. & Hoffman, R. M. (2015a). Prostate Cancer Heterogeneous High-Metastatic Multi-Organ-Colonizing Chemo- Resistant Variants Selected by Serial Metastatic Passage in Nude Mice Are Highly Enriched for Multinucleate Giant Cells, 4–11. 93

Zhang, T., Jia, W., Zhu, Y. & Yang, J. (2015b). ScienceDirect Automatic tracking of neural stem cells in sequential digital images. *Integrative Medicine Research*, 36, 66–75. ISSN 0208-5216. URL http://dx.doi.org/10.1016/j.bbe.2015.10.001. 6, 38

Zhang, Y., Xiu, W., Sun, Y., Zhu, D., Zhang, Q., Yuwen, L., Weng, L., Teng, Z. & Wang, L. (2017b). Rgd-qd-mos2 nanosheets for targeted fluorescent imaging and photothermal therapy of cancer. *Nanoscale*, 9, 15835–15845. URL http://dx.doi.org/10.1039/C7NR05278B. 19

Zhang, Z. & Feng, S.-S. (2006). Nanoparticles of poly(lactide)/vitamin e tpgs copolymer for cancer chemotherapy: Synthesis, formulation, characterization and in

vitro drug release. *Biomaterials*, 27, 262 – 270. ISSN 0142-9612. URL http://www.sciencedirect.com/science/article/pii/S0142961205004977. 19

Zhao, X., Ma, S., Liu, N., Liu, J. & Wang, W. (2015). A polysaccharide from Trametes robiniophila Murrill induces apoptosis through intrinsic mitochondrial pathway in human osteosarcoma (U-2 OS) cells. *TUMOR BIOLOGY*, 36, 5255–5263. ISSN 1010-4283. 20

Zimmer, C. (2012). Technoreview From microbes to numbers : extracting meaningful quantities from images. 14, 1828–1835. 36

Zimmer, C., Labruyère, E., Meas-yedid, V. & Guillén, N. (2002). Segmentation and Tracking of Migrating Cells in Videomicroscopy With Parametric Active Contours : A Tool for Cell-Based Drug Testing. 21, 1212–1221. 38

Zou, R. S. & Tomasi, C. (2016). Deformable Graph Model for Tracking Epithelial Cell Sheets in Fluorescence Microscopy. 35, 1625–1635. 38