FEDERAL UNIVERSITY OF RIO GRANDE DO SUL
MASTER PROGRAM IN PHYSICS

Probing tumor aggressiveness via image analysis *

Paulo Casagrande Godolphim

Dissertation written under the supervision of Professor Leonardo Gregory Brunnet and presented to the Physics Institute of Universidade Federal do Rio Grande do Sul as partial requirement for the degree of Master in Physics.

Porto Alegre, Brazil
February 2021

* Financial support by CAPES-Cnpq and Propesq.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
PROGRAMA DE MESTRADO EM FÍSICA

Sondando a agressividade tumoral por análise de imagem [†]

Paulo Casagrande Godolphim

Dissertação escrita sob a surpervisão do Professor Leonardo Gregory Brunnet e apresentada ao Instituto de Física da Universiade Federal do Rio Grande do Sul como requisito parcial para a obtenção de título de Mestre em Física.

Porto Alegre, RS, Brasil
Fevereiro 2021

[†] Suporte financeiro por CAPES-Cnpq e Propesq.

Acknowledgements

I sincerely thank my mother, my father, and my family; my advisors, professors, and collaborators; my colleagues and friends, my therapists, and my girlfriend. Without their support, love, guidance, and knowledge, this work would not be possible.

We thank our collaborators *PhD. Bibiana F. Matte* and *Prof. Marcelo M. Lamers* from the *Department of Oral Pathology of Universidade Federal do Rio Grande do Sul*. They are responsible for all the experimental part of the work. We also thank the other collaborators *Prof. Rita M.C. de Almeida* and *Prof. Gilberto L. Thomas* from *Physics Institute of Universidade Federal do Rio Grande do Sul* for helping in the development of the work general idea. We thank *Prof. Hélène Delanoë-Ayari*, from *Institut Lumière Matière* of *Claude Bernard Lyon-1 University*, for the enlightening conversations that helped at the beginning and end of this work. We thank *Prof. Olivier Cochet-Escartin* from *Centre national de la recherche scientifique* for the enlightening conversations at the end of this work. We also thank *Ana Susin* and *Tuila Ziliotto* for the general corrections of the text.

"As estrelas estão tão longe que no momento que a luz delas chega até nós, muitas delas já podem estar mortas. ...

... No universo existem quatro forças fundamentais que governam tudo: o eletromagnetismo, a gravidade, a força forte e a força fraca."

My father, Ben Hur Godolphim. In memoriam.

Quotes and citations

"There's no need for being brave if you can copy and paste"

"Ian Donnelly : [upon first meeting] Priority one: What do they want and where are they from? And beyond that, how did they get here? Are they capable of faster-than-light travel? I've prepared a list of questions to go over, starting with a series of "handshake" binary sequences...

Louise Banks : How about we just talk to them before we start throwing math problems at them?"

Arrival, 2016 movie.

"... One of the sailors pointed to the animals that hopped around with their young riding in pouches, and asked an aborigine what they were called. The aborigine replied, *Kanguru.* From then on Cook and his sailors referred to the animals by this word. It wasn't until later that they learned it meant *What did you say?.*

I tell that story in my introductory course every year. It's almost certainly untrue (...) but it's a classic anecdote."

Luise Banks, Story of your life - By Ted Chiang

"Quem mais faz, mais faz."

"Pra que ficar reinventando a roda?"

Ivani Casagrande Rodrigues, My Mother.

Abstract

Cancer is the second leading cause of death worldwide. In this disease, tumor cells have to invade and migrate through various complex structures of the body. It is well-accepted that the physics underlying tissue dynamics play a crucial role in tumor evolution; however, there are not many works bridging clinical applications and physical approaches. Here we analyze the collective migration of three oral human epithelial cells lines - a healthy lineage (HaCat), a less aggressive tumor (CAL27), and a more aggressive tumor (SCC9) - during the formation of *in vitro* 2D tissues, i.e., monolayers. We compare: how fast they move, if the migration is ordered or not, the extent of correlation of their movement, and how these behaviors change during the tissue's growth and maturation. We show that SCC9 monolayers present a close to zero average migration velocity, i.e., a signature of a disordered non-collective system, while HaCat and CAL27 present positive values. Moreover, during tissue growth, the average speed of SCC9 has a monotonic decay compared to what is found for the other two lineages. We believe these results are related to the lower number of cell-cell junctions in SCC9 [1, 2]. We also show that HaCat average velocity correlation is higher in space and with lower fluctuations in time than the two tumor lineages. Finally, using the average speed and average correlation measurements, we build a novel diagram that separates the three lineages into three distinct regions of a single graphical representation. An actual clinical application of our diagram is speculative for the time being, given the limitations associated with the *in vitro* approach and the limited number of samples. Nevertheless, our results indicate a correlation between non-collective behaviors and tumor aggressiveness and pose collective motion quantities as possible tools to classify and diagnose tumor cells. All the analyses were done using an open-source *pipeline* we developed, available on GitHub in an unsupported beta version. In the near future, we hope the evolution of the techniques used here could help the medical community define new malignancy markers; and, in the long term, help pave the way for a new set of physics-based diagnoses for cancer.

Resumo

O câncer está em segundo lugar nas causas de morte no mundo. Nessa doença, as células tumorais precisam invadir e migrar através de várias estruturas complexas do corpo. É bem aceito que a física subjacente à dinâmica do tecido desempenha um papel crucial na evolução do tumor; no entanto, não existem muitos trabalhos ligando aplicações clínicas e abordagens físicas. Aqui, analisamos a migração coletiva de três linhas de células epiteliais humanas orais - uma linhagem saudável (HaCat), um tumor menos agressivo (CAL27) e um tumor mais agressivo (SCC9) - durante a formação de tecidos 2D *in vitro*, i.e., monocamadas. Nós comparamos: quão rápido eles se movem, se a migração é ordenada ou não, a extensão da correlação de seu movimento e como esses comportamentos mudam durante o crescimento e maturação do tecido. Mostramos que monocamadas de SCC9 apresentam velocidade média de migração próxima a zero, ou seja, a assinatura de um sistema não coletivo, enquanto HaCat e CAL27 apresentam valores positivos. Além disso, durante o crescimento do tecido, a velocidade média da SCC9 apresenta uma queda monotônica em relação ao que é encontrado para as outras duas linhagens. Acreditamos que esses resultados estejam relacionados ao menor número de junções célula-célula da SCC9 [1, 2]. Também mostramos que a correlação da velocidade média da HaCat é maior no espaço e tem menores flutuações no tempo. Finalmente, usando as medidas de velocidade média e correlação média, construímos um novo diagrama que separa as três linhagens em três regiões distintas de uma única representação gráfica. Uma aplicação clínica do nosso diagrama é especulativa por enquanto, dadas as limitações associadas com a abordagem *in vitro* e ao número limitado de amostras. No entanto, nossos resultados indicam uma correlação entre comportamentos não coletivos e agressividade de tumores e sugerem que o movimento coletivo poderia ser usado como ferramentas para classificar e diagnosticar células tumorais. Todas as análises foram feitas usando uma *pipeline* de código aberto que desenvolvemos, disponível no GitHub em uma versão beta sem suporte. Em um futuro próximo, esperamos que a evolução das técnicas usadas aqui possam ajudar a comunidade médica a definir novos marcadores de malignidade; e, a longo prazo, ajudar a pavimentar o caminho para um novo conjunto de diagnósticos de câncer baseados em física.

Contents

Chapter 1

Introduction

Acknowledging cancer as a multifaceted disease [3, 4], the scientific community accepts today that the physical properties of the tumor micro-environment represent a crucial aspect of this illness [5, 6, 2]. Today, researchers see cancer also as a physical problem [3, 4]. Worldwide, small independent laboratories, as well as large consortia, are making efforts to incorporate this new paradigm in cancer research believing that new significant advances will come from a better understanding of the physics underlying tumor progression [3]. An important example of these consortia is the *Physical Sciences-Oncology Network* (PS-ON) from the US National Cancer Institute (NCI[1]) started in 2009.

Already on its phase II[2], the PS-ON aims to systematically integrate physical science into cancer biology, with the focus to add new multidisciplinary tools to day-by-day medical decisions of what treatment to choose and what medicine to administrate. In their article reviewing the first ten years of this consortium (*PS-ON phase I 2019* [3]), a critical gap is recognized between the natural science method of observation, description and classification, immortalized by Charles Darwin; and the methods based on mathematical logical steps, as developed in the Physical sciences from Galileo to present days: a gap they are trying to fill.

The field of biological physics has been blooming in the last two decades. Major areas like *Active Mater* and *Cellular Tissue Rheology* delivered publications with new techniques at an intense rate ([7, 8, 9, 10]), principally in the same period that the PS-ON phase I occurred. Those papers often justify their existence with statements of applicability in tumor spreading and similar areas. However, after more than two decades of studies, why we still find sparse works bridging the biological scientists willing to use physics tools with the physicists that are already doing biology? We believe that an answer to this question is the existence of a language problem.

[1] NCI page.

[2] PS-ON phase II page.

The communication gap between biological scientists and physicists is evident. Not only the scientific jargon is significantly different, but also the language and the way of thinking. This opposition leads to papers and theoretical structures too different from each other, making it difficult for the counterpart to see the applicability of some technique or, on the other hand, to understand how to build techniques to be feasible to apply them. Both the physicist approach and the bio-scientist approach have their advantages and disadvantages. The question that stands is: how can we develop works that incorporate the best of these two worlds?

To face this question and add another brick in the road paved by PS-ON and others, we proposed ourselves to build this work in a manner that intersects these two scientific fields. The idea is to use already known tools from biological physics, based on mathematical analyses of collective movement, and apply them to a classification problem, inspired by the natural sciences approach. From the *in vitro* experimental assays with living cells through the *in silico* image analysis and physical calculations, the leading words are simplicity and efficiency, rather than novelty. An adequate way to understand the proposition is to imagine an alternative reality in which cancer was born as a physical problem, and it is up to the physicists to deal with it. In this surrealistic situation, instead of having protein expressions and PCR diagnostics[3], the cancer researchers would have to solve the cancer problem with physical quantities, such as velocity and displacement vector fields. In such a situation, would it be possible to build diagnostic tools? Moreover, if so, would they give new answers to this complex disease? Furthermore, would these tools be viable? This dissertation tries to answer these questions, based on some seminal works in collective movement, cellular jamming, and cancer research.

Vicsek et al. 1995 introduces a simple *in silico* model for the collective motion of active particles. They show that changing the amount in which the particles can see/interact with their neighbors can change the system from a collective migration ordered phase to a disordered phase with no net migration [11]. Adaptations of this work were used to describe the behavior of a variety of complex systems, from flocks of birds to cellular aggregates [12, 13, 14, 7], reveling in common physical rules among them, e.g., the microscopic interaction between particles determine the macroscopic collective behavior of the whole system. When the collective behavior is different from the behavior of the individual parts, it is called emergent. For example, a single cell migrating in a flat protein-coated surface will most probably behave by a Fürth law [15], or, by an extended Fürth law [16], i.e., a combination of diffusive and ballistic motion. Nevertheless, when put together in a wound-healing assay, cells present a high ordered ballistic

[3] PCR explained in NCI page

migration in the wound's direction [17, 18]. This collective migration is an example of emergent behavior. In order to do this, cells need to communicate and transport information through the tissue. The way they communicate is called the law of transport of information. Changing the way they interact changes the transport of information laws that eventually change the collective behavior.

A practical (*in vitro*) application of the previous paragraph discussion is made by *Vedula et al. 2012* in a wound-healing assay with the healthy epithelial lineage MDCK [17]. MDCK cells have high concentration of proteins on their membranes functioning like cell-cell junctions, and allowing them to form a cohesive tissue. They compared the collective migration of wild-type MDCK with MDCK treated to have a lower number, i.e., a down-regulation, of cell-cell junctions in an wound healing experiment where cells are released from a reservoir to a free surface. They show that the wild type MDCK presents a highly ordered ballistic migration, while the treated ones showed a disordered behavior similar to a gas-like state with lower migration velocity and no directionality. The explanation is that the absence of junctions between the cells changed the laws of transport of information in the system, consequently changing the emergent behavior from high speed ordered migration to low-speed disordered migration.

In the context of tumor spreading, the regulation of protein expressions has a crucial role. There are several hallmarks of malignancy, i.e., tumor aggressiveness, related to protein expressions [19, 20], but here we are concerned with only one of them: the down-regulation and up-regulation of the expression levels of E-cadherin and N-cadherin proteins [4]. E-cadherin is a junction protein, concentrated mostly on the cell membrane, that promotes cell-cell adhesion. It works like a velcro, binding cells together through their membranes. N-cadherin is also a membrane protein but has the opposite function. It can be thought of as a chemical scissor for cell-cell junctions [4]. In general, healthy epithelial cells present high levels of expression of E-cadherin and low levels of N-cadherin [2], which assures the strength and the cohesiveness of the epithelial tissue. Tumor cells present the opposite behavior. High invasive tumors, i.e., more aggressive, have high N-cadherin levels and low E-cadherin levels [2, 1], while Low invasive tumors, i.e., less aggressive present intermediate levels of E-cadherin and N-cadherin [2, 1]. This differentiated protein expression facilitates tumor cells to detach from and migrate through the epithelial tissue to search for nutrients and optimal environments to grow.

These collections of biological and physical results discussed in the last three paragraphs made us wonder how this hallmark of malignancy, i.e., the down-regulation of cell-cell junctions, would determine tumor cells' collective behavior. Moreover, are the collective movement of

tumorous and healthy cells sufficiently different from each other so that it could be used to diagnose tumor behavior? In other words, can the collective behavior of tumor cells work as the macroscopic physical equivalent of the down-regulation of junctions proteins so that it could also be considered a hallmark of malignancy? In this work, we give answers to these questions.

In this dissertation, we classify and separate three lineages of oral human epithelial cells based on physical quantities associated with collective movement. What we aimed to compare is: (i) How fast they move, i.e., average speeds of the cells, (ii) is the migration ordered or not, i.e., average vectorial velocity, (iii) what is the extent of correlation of their movement, i.e., spatial velocity correlation functions, and how these behavior changes in time during growth and maturation of the tissue. The lineages we use are: a healthy one called HaCat with high concentration of cell-cell junctions; a less aggressive tumor called Cal27 which is low invasive/high cell-cell junctions; and a more aggressive tumor called SCC9 which is high invasive/low cell-cell junctions [1, 2]. The tumor cells are squamous cell carcinomas.

The more aggressive lineage SCC9 is believed to be less collective than the less aggressive Cal27 [1], probably due to the low level of cell-cell junctions. We hypothesized that if SCC9 is a non-collective lineage, a tissue grown only with this type of cells would show collective behavior compatible with the disordered phase of a Vicsek class system, e.g., present an average vectorial velocity close to zero. Our second hypothesis comes from the idea that high invasive tumors can unjam from the tissue [4]. We hypothesized that this unjamming feature could be related to tumor cells' insensibility to the jamming process. If so, tumor cells should not present a pronounced change in behavior during the density increase and tissue maturation.

To test our hypothesis, we make jamming experiments [21] where each cell lineage are grown separately on Petri dishes for 48h until they reach confluence, i.e., form a monolayer of cells with no open spaces. Microscopy recording starts after confluence and it is taken for 50h. During the experiment, cells will divide, die, and migrate in constant interaction with their neighbors, while the monolayer *ages* in time, i.e., tissue maturation [21] and density increase. The resulting images are analyzed, and the velocity fields are extracted using particle image velocimetry. We calculate the average speed/velocity of the monolayers and the velocity correlation functions. We analyze how these quantities change in time, i.e., versus the monolayer' aging process, for each cell type, and then compare the lineages.

Our results show that the SCC9 monolayer, the more aggressive tumor, has close to zero average vectorial velocity, corroborating the idea that this lineage is non-collective. Concerning insensibility to jamming, all the three lineages present a significant decrease in speed during the experiment, rejecting the hypothesis of jamming insensibility for the more aggressive one.

Nevertheless, the way the speed decrease in time for the SCC9 lineage is quite different from the other two. The healthy lineage HaCat and the less aggressive tumor Cal27 present a fast decay, similar to what is found for healthy epithelial cells [22]. SCC9 lineage, on the other hand, presents a slow decay, similar to what is found for fibroblast cells with close to zero concentration of cell-cell junctions [22]. We believe that the lower number o cell-cell junctions in SCC9 suppress cells' ability to coordinate long-range movement. In high packed system, the lack of long-range coordination hinders the possibility of ordered collective migration [23], making the more aggressive tumor cells locally trapped, what explains the monotonic speed decay. Concerning the malignancy hallmark of cell-cell junctions down-regulation, it seems that to be able to unjam from the healthy tissue, highly invasive tumor cells pay the price of losing long-range coordination in their movement, reinforcing the idea that malignancy is related to a non-collective behavior.

Regarding the idea of cancer diagnosis based on collective movement, we define a parameter called *Activity decay time*, which tells us how fast the speed of the cells decay in time. This parameter enables us to separate the more aggressive cell lineage from the other two. From the analysis of velocity correlation, we defined a parameter called *average positive correlation*. The tumor cells present lower values of this parameter and higher time fluctuations. We are not sure why tumor cells present this behavior, but computing the temporal average and the associated standard deviation of the *average positive correlation* enables us to separate the tumor cells from the healthy lineage. Combining these two quantities, we built a single graph representation that separates the three lineages into three distinct regions. From the treatment of the raw experimental data to the final graphs and diagram, it was all done using an open-source pipeline we developed for this dissertation (link github). Moreover, the experimental techniques used here are simple ones and may be replicated in developing countries. We believe that this work is proof of the concept that a diagnosis tool for tumor aggressiveness based on collective motion is possible. We encourage other parties to search for more physics-based diagnoses, believing that such approaches will help shorten the gap between cancer physics research and real clinical applications.

It is essential to state that the experimental conditions employed here are far from *in vivo* situations or real tumor lesions treated in oncology clinics. A direct translation of the behaviors founded here to real body conditions is speculative for now. The way cells move in a *in vitro* monolayer will probably not be the same if they were in a real 3D body environment. Moreover, the technique proposed here as diagnostic tools is more troublesome them current tools, such as PCR. However, we show that the three lineages, under the same experimental circumstance,

can be identified and separated only by looking at how they collectively move. The result suggests that this physical quantity could also be used to monitor and classify tumor lesions' aggressiveness.

We believe that this work should be thought of as a prototype that needs to be developed, tested, and improved. Another limitation of the work is the low statistical significance. Here we only work with three lineages and with few replicates of each experiment. The right direction would be to apply this pipeline for a greater number of lineages and see if the results hold for a greater number of replicates. *In silico* experiments, similar to what was done by [12] and [21], could also be of good use in the future of this work.

The next chapter is dedicated to literature review presenting basic concepts in active matter physics, in cancer biology and some recent works in tissue movement characterization. Chapter 3 is dedicated to methods, where we explain the experiments and the tools we used to analyse them. We then present the results in chapter 4. Finally, we discuss and conclude in chapter 5.

In case of doubts, contributions or corrections on this work, please, fell free to contact us via **paulocg33@gmail.com** with the subject **dissertation**.

Chapter 2

Literature Review

The human body organization starts with cells, that together form tissues, that form organs, that working together make a functional body. For example, the skin, that has primary functions of protection and sensing the environment (from pain to pleasure), is basically made up of a combination of two tissues: the epithelial and the connective tissues.

A tissue is an assembly of similar cells that together display/execute some specific function individual cells may not be able to do. In the physics of complex systems, a process where microscopic individual parts combine to form a macroscopic system showing behavior not present before the assembling, is called *emergent*. For example, connect together 86 billion neurons[1] in a right way and you get the emergent behavior of *thinking*. The fact that the body organization in tissue levels follows emergent behavior makes a natural path to physicists to enter in this field.

Emergence is a cornerstone of the area of Active Matter (AM), one of the most flourishing areas in biological physics in contemporary times. AM focus studies where active single entities - like birds, soccer players, cars, cells - interact with each other in order to form macroscopic structures like - flocks, soccer teams, traffic, tissue - that can display large scale collective movement such as escape predators, win a match, traffic jam, wound healing.

Interactions among particles may appear in different forms, from visualizing the team mate and passing the ball, to cells sensing their neighbors through membrane proteins. If these interactions change, emergent behavior will change too. The way unities interact inside a system defines information transport laws, which determines the collective emergent behavior. For instance, in an epithelial tissue the concentration and types of cell connections defines the transport information laws, which determine macroscopic emergent modes of the tissue, such as collective cell migration and collective oscillations [17].

[1] More info about the brain in humam-memory page.

2.1 Biology and physics of epithelial tissue

Epithelial cells are the basic unities composing the epithelial tissue, being present in the majority of human organs (lungs, mouth, stomach, skin...). The general function of epithelial tissue is protection. Cancer originated from epithelial cells is called Carcinoma and accounts for 80-90% of all the cancer cases[2]. Figure 2.1 (a) shows a histological section of a skin from the palm of a mouse and (b) shows that same sample in a microscope. The light area (depicted by "III") is the connective tissue and the darker layer is the epithelial tissue. In rats and humans, these two tissues combine to form the organ skin.

Epithelial and connective tissue are separated by a thin layer of a glico-protein called *laminin*, which composes the basal lamina. Epithelial tissue is a confluent tissue, containing only cells. Starting from the basal lamina, the place where cells are generated, they divide growing in stratified layers up to the most external region, where cells in the outer part of the tissue produce keratin (the protein that works as waterproofing and protection).

Inset in figure 2.1 (b) shows epithelial cells (II) and connective tissue (III). The latter is an aggregate of proteins, fibroblasts, defense cells, lymphatic and blood vessels. Epithelial tissue shows no circulatory irrigation, so it depends on connective tissue for nutrients. In figure 2.1 (c) we display a histological section from human skin, showing stratified layers of epithelial cells (dark purple), over the amorphous connective tissue (light purple), the same pattern of the mouse skin.

This stratified structure is maintained in an dynamical equilibrium, where epithelial cells are renewed every $\sim$5 days, therefore having a constant process of death, division and migration to target areas that need replacement. In physics, this homeostatic behavior of tissue maintenance can be interpreted as a stationary state.

The stationary state is broken in the occurrence of a wound. Figure 2.2 illustrates that. A wound is the absence of cells, literally is the act of removing layers of cells from the skin (or other organ) leaving an empty space. In the case of skin, a shallow cut removing only external layers of epithelium will not bleed. A deeper cut puncturing the basal lamina and reaching connective tissue (and its blood vessels) will necessarily bleed. In the occurrence of a cut, chemical and mechanical information will spread over the tissue signaling cells to leave their ground state in order to close the wound. This is showed in Figure 2.2 (b), the empty space is sensed by healthy cells (marked in blue) that respond increasing division rate and migration (green vectors). This elevated rate of division and migration, that we can think, in physics jargon, as an excited state, is maintained until the tissue is repaired. After that, healthy cells

[2] NIH page.

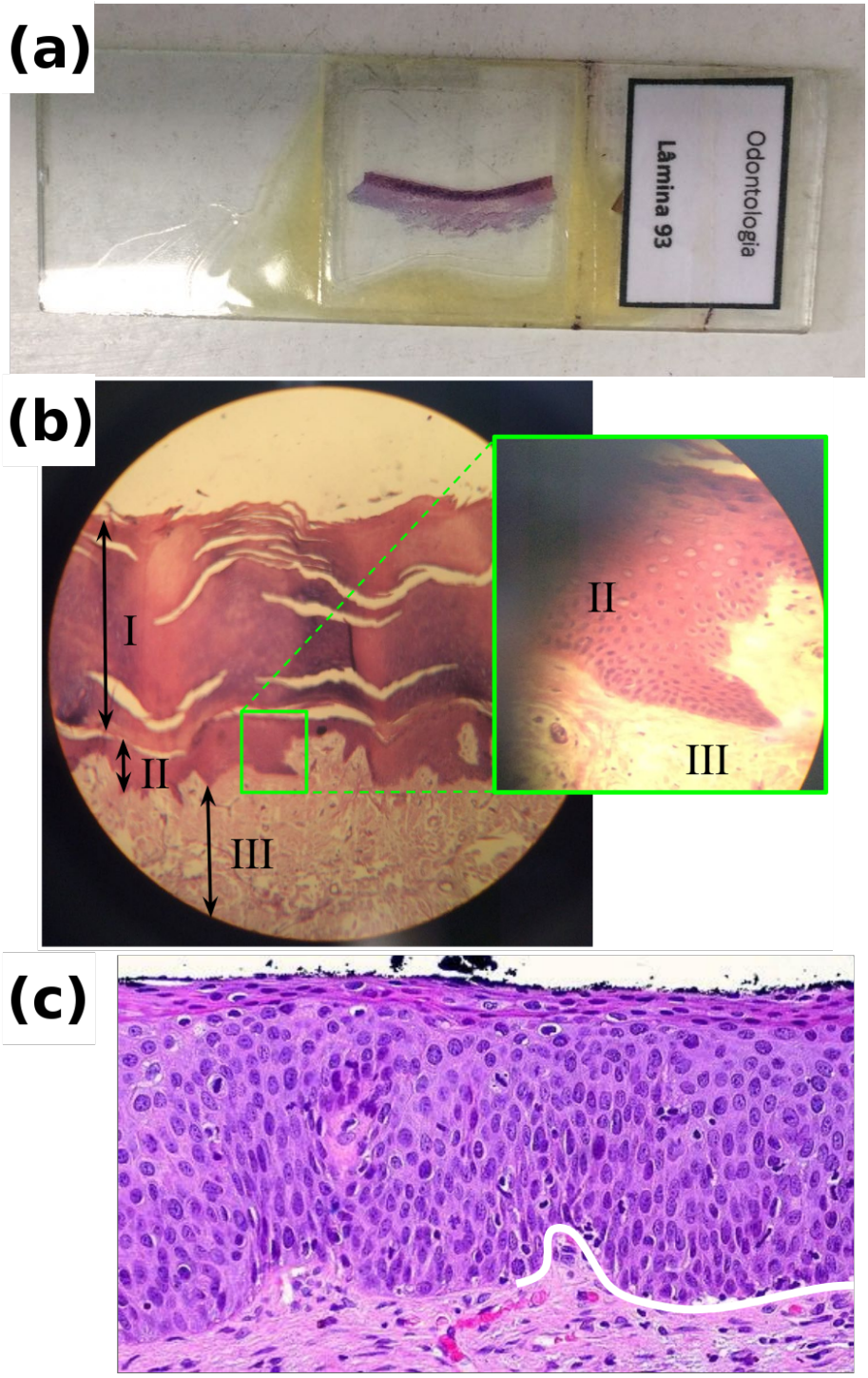

Fig. 2.1: (A) Histological section of a mouse palm skin. (B) Histological section (A) under the microscope: (III) connective tissue, (II) epithelial tissue, (I) keratine layer produced by epithelial cells. (C) Histological section of human skin: stratified confluent behavior of epithelial cells (dark purple) separated by the basal lamina (white) from the connective tissue (light purple). Credits: (c) lifepersona.

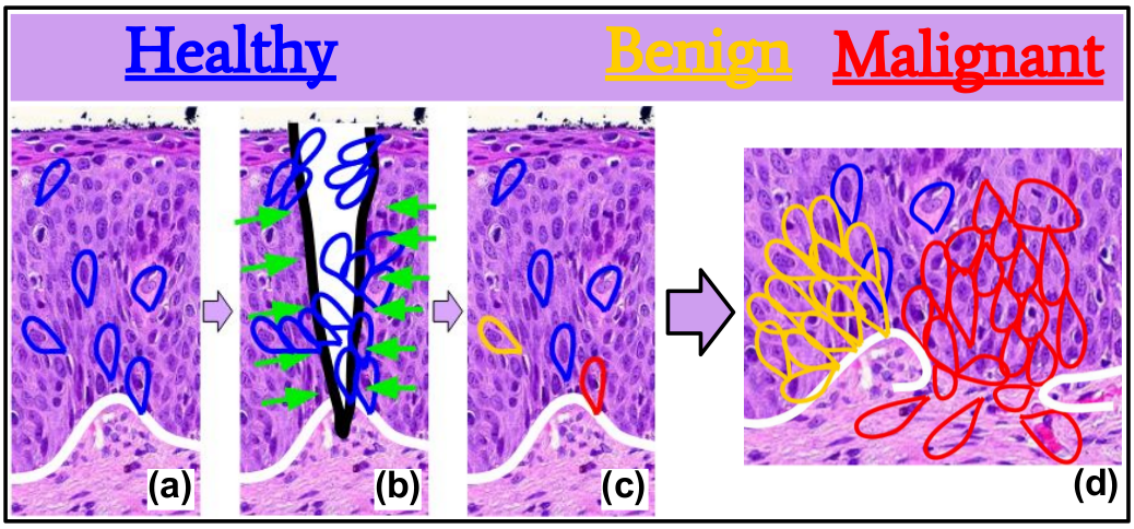

Fig. 2.2: Histological section of human skin overlapped with drawings. (a) epithelial (dark purple) and connective (light purple) tissue separated by basal lamina (white). In blue, healthy epithelium. (b) In the occurrence of a wound, healthy cells enhance division and migrate to close it. (c) Tumor cells (orange, red) are mutated cells that keep dividing (d) even when it is no longer required. (d) Non-invasive (less aggressive) tumor (orange); (red) more aggressive tumor invading the connective tissue. Background histological section credits: lifepersona.

go back to the stationary state by a process, in biology, called contact inhibition ([24]).

Tumor cells (depicted in orange and red in Figure 2.2 (c) and (d)), are healthy cells that have undergone some mutation allowing them to keep dividing, even when this is no longer required. Benign tumors (orange cells in the picture) are cells that divide but do not invade surrounding tissues. Changes in morphology and protein expression resulting from further mutations make malignant tumors (red in the picture) detach from the epithelial tissue and invade the connective tissue to search for optimal environments to grow. This process is called epithelial-mesenchymal transition - EMT [4].

Epithelial phenotype is not well suitable for migration. When required, e.g., in tissue formation or wound healing, epithelial cells changes to a mesenchymal phenotype: they get more similar to fibroblast cells, which are motile cells that live in the connective tissue. A similar process happens for malignant tumors more drastically. High invasive/aggressive tumors will, in general, be morphologically close to fibroblast cells, with a lower number of cell-cell junctions. In contrast, low invasive ones will be morphology close to healthy epithelial cells with a greater number of cell-cell junctions.

Carcinoma happens when malignant tumor cells are able to pierce the basal lamina and colonize the connective tissue. While in connective tissue, which is full of nutrients, they may

reach blood or lymphatic vessels. Metastasis - the turning point on cancer responsible for 90% of fatalities - happens when malignant cells invade those vessels and colonize other parts of the body. To complete a metastasis cycle, they have to migrate through various mechanical and chemical constraints of the body.

A good part of cancer physics research is then focused on cell migration and how the extracellular environments' mechanical constraints modulate it. The video 01, made by [25], shows an immune cell migrating in a 3D *in vivo* situation. Any cell, including tumor cells, will face similar challenges when migrating within the body. How well they will overcome these challenges define the aggressiveness of the tumors. The two tumor lineages studied in this work killed their original patients, but CAL27 is less aggressive, with morphology and cell-cell junction expression close to a healthy epithelial cell [2, 1], while SCC9 is more aggressive, with morphology and cell-cell junction expression close to a fibroblast [2, 1]. Moreover, SCC9 is believed to be less collective lineage than CAL27 [1].

The Appendix 6.1 is a review on velocity vectors, velocity fields and vectorial averages. We encourage the non familiarized reader to proceed to the next section after reading the referred appendix.

2.2 Collective and non-collective motion

The theories of collective motion and the description of self-propelled entities' started in Physics in 1995 in a work by Vicseck [11]. For the first time, the work introduces a physical description of collective motion in the context of phase transitions in statistical physics. The work enabled the physicists to approach the behavior of groups of animal and synthetic self-propelled entities in a novel way, defining one of the cornerstones of the field of Active Matter.

Vicsek presents a simple model of self-propelled bird-like particles that can align their velocity with the neighbor particles. The alignment efficiency depends on how many neighbors a particle has and on an internal noise that simulates how well a particle process the information received from the neighbors. By decreasing the noise, i.e., increasing the ability of the particles to communicate with each other, the same system can change from a disordered phase with no collective migration (Fig. 2.3 (a)) to an ordered collective migration phase (Fig. 2.3 (d)). This model was used to describe the collective motion of birds, fish, insects, and mammals, including humans.

Vicsek proposed a parameter to monitor if a system is at a collective or non-collective phase: the module of the particles' average vector velocity. In a disordered phase (Fig. 2.3

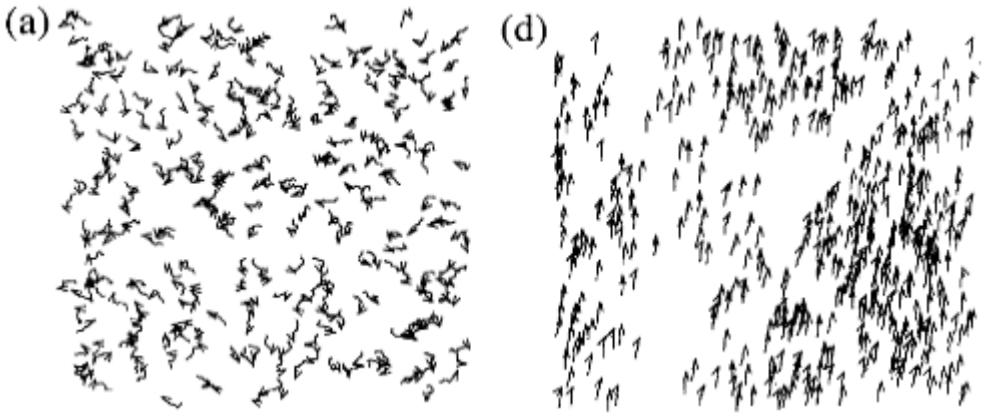

Fig. 2.3: Active matter simulations done by *Vicsek et al. 1995*. The arrowhead indicates the instantaneous direction of the velocity, and the black lines the accumulated trajectory in a previous period. Left: a non-collective disordered system. Right: a collective ordered system with particles migrating in the same direction. Adapted from [11].

(a)), the average vectorial velocity will be close to zero, while in the ordered phase (Fig. 2.3 (d)), it will be close to one, since the velocity of all particles points approximately in the same direction, and because in Vicsek model all particles have their speed equals to one. It is interesting to point out that both systems in Figure 2.3 have the same kinetic energy, i.e., the same average speed, but the velocity vector average is different. So the average speed can tell how much energy particles are spending in movement, regardless if it is ordered or not, and average velocity vector tells how collective the system is.

The transition from non-collective to collective behavior also happens in the Vicsek model when density increases for a fixed noise. With a higher number of particles, the average number of neighbors increases, and then it is easier for a particle to *understand* the neighbors' migration direction. For example, doubling particles' density from video 02 to video 03 changed the system from a non-collective to a collective phase. We did the simulation of these videos using a Vicseck' algorithm written by Cássio Kirch (Cássio Kirch page).

Szabó et al. 2006 [12] show that the increase in density of *in vitro* keratinocytes cells induce a phase transition from a non-collective behavior to a collective one (video 04). In Figure 2.4 (a), (b), and (c), we can see a snapshot of the Szabó experiment. Figure 2.4 (d), (e), and (f) show the velocity of the cells during the experiment. They computed the normalized average velocity vector and found it close to zero value for the disordered phase in (a) and (d) and a value close to one for the ordered phase in (c) and (f). *Szabó et al. 2006* also introduces an adaptation of the Vicscek model that shows phenomenological equivalence with the keratinocytes behavior (video 05).

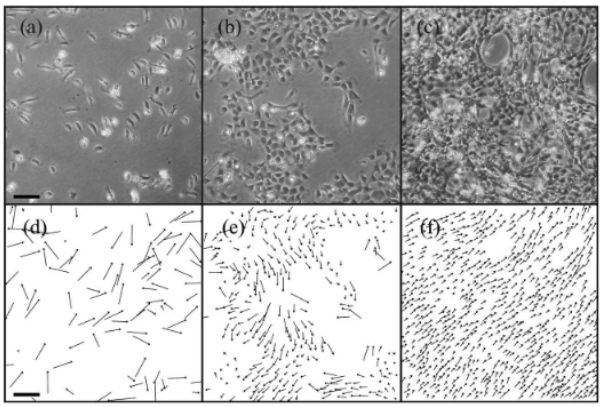

Fig. 2.4: (a, b, c) Migration of keratinocytes from the scales of goldfish in phase-contrast *in vitro* microscopy. (d, e, f) Instantaneous velocities of the cells. Credit: [12].

Gregoire et al. 2003 ([13]) added two new force in the classical Vicseck model: one to enable particles to avoid overlapping and a spring-like force enabling particles to be attached to their next neighbors. The model has the ordered and disordered transition and, depending on the strength of the springs, it has solid, liquid, and gas phases (Fig. 2.5). They show that a system of particles can change from a collective to a non-collective phase for a fixed density and noise by increasing the aggregates' solidification. This work is an excellent example of how microscopic interaction laws between particles can change the system's collective behavior. *Belmonte et al. 2008* [26] did an adaptation of Gregoire work to describe cell behavior on the morphogenesis of hydra.

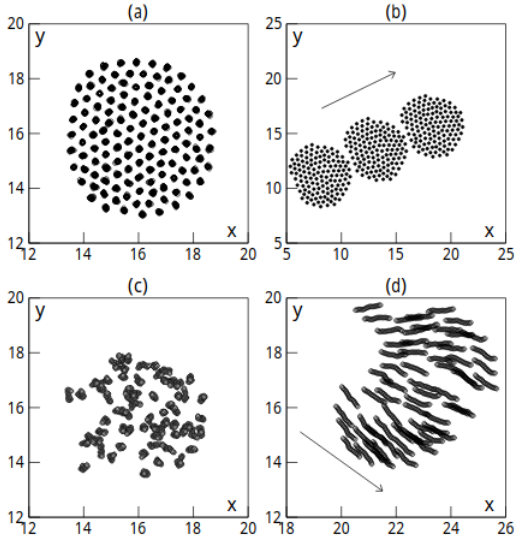

Fig. 2.5: Self-propelled particle simulation done by *Gregoire et al. 2003*. (a) A non-moving solid, (b) moving solid, (c) disordered liquid, and (d) moving droplets. Credit: [13].

Our work's central premise is that the low number of cell-cell junctions in the more aggressive lineages hinders their capability to orchestrate long-range ordered movements, so they should present a disordered phase behavior. In the next section, we discuss results about the influence of cell junctions in collective migration.

2.3 Monolayer experiments and jamming dynamics

Monolayer assays are simple approaches to understanding the physics underling tissue dynamics and tumor spreading (Fig. 2.2). In these experiments, instead of epithelial cells form a 3D stratified tissue with several layers (Fig. 2.1, 2.2), they grow into a single layer forming a *quasi*-2D tissue, since cells are 3D objects, called a *monolayer* (Fig. 2.6 [10]).

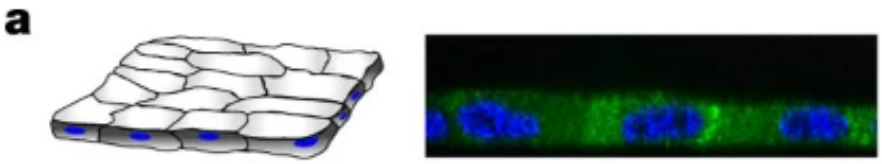

Fig. 2.6: Left: schematic of an epithelial monolayer. Right: lateral view of an epithelial monolayer in fluorescence microscopy. Blue marks the nucleus and green the membrane and cytoplasm of the cells. Adapted from [10].

Epithelial cells are biologically programmed to form tissues. When cells are maintained in an environment with close-to-body conditions, i.e., with chemical composition, temperature, atmosphere, and PH similar to the body, they will inevitably execute their biological program, which they achieve via a combination of mitosis and migration. In video 06, we can see the behavior of the healthy epithelial cell HaCat in an optimal environment inside a Petri dish in low density, i.e., the majority of the substrate is free of cells.

In monolayer assays, the experimentalist places an initial seed of cells in a Petri dish (Fig. 2.7 (a)) and let cells divide until they form a monolayer (Fig. 2.7 (b)). This process takes approximately four days. In video 07, done by [21], we can see Human Bronchial Epithelial Cells - HBEC - forming a monolayer (scale bar is $300\mu m$).

In physics literature, this assay usually comes by the name of *jamming experiment*, since cells get jammed, i.e., locally trapped, once the tissue is formed. There are two principal points of interest during the assay. The first one is the moment of confluence, when there is no more substrate surface free of cells (Fig. 2.8 (a)). After reaching confluence, cells will keep moving and dividing. To migrate, they will have to squeeze through each other or coordinate some

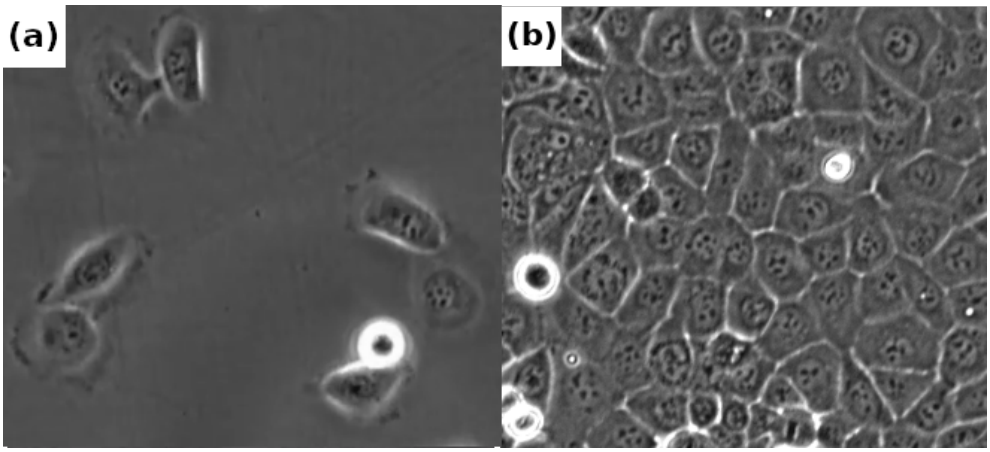

Fig. 2.7: Upper view of healthy epithelial cells from HaCat lineage under a phase-contrast microscope. (a) Four pairs of cells with free surface to move. The very bright spot is a cell under mitosis. (b) The cells organized as a monolayer. All the substrate surface is occupied by the cells.

collective motion. Since there is no more free surface, cell division will increase the monolayer height. The second point of interest is the jamming transition point, where the tissue's density and maturation reaches critical value and division and migrations rates decrease [21]. The tissue will be fully grown after the jamming transition, reaching and stationary state where cells will be highly packed together, and migration and division rates are close to zero or very low.

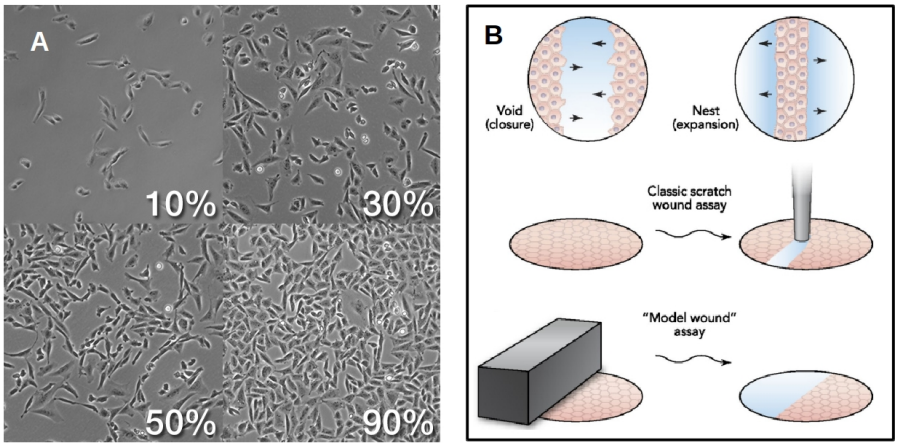

Fig. 2.8: (A) A tissue growing experiment explaining the percentages of confluence. As time goes by, cells divide, density increases to confluence. 100% confluence, or confluence point, correspond to the moment where there is no more empty substrate visible. (B) Some example of wound healing experiments. Credits: (A) biology stack exchange, (B) *Vedula et al 2013* ([18]).

Other approaches to studying collective cell migration are the *wound healing experiments*. In these assays, the experimentalists place an already formed tissue in confront with a free

surface to which cells can migrate (Fig. 2.8 (b)). *Petitjean et al 2010* ([22]), *Vedula et al 2012* ([17]) and *Sepulved et al 2013* ([27]) show that cells migrate in direction to the free space, until they close the wound.

Vedula et al. 2012 [17] show that different geometrical constrains can modulate distinct migration behaviors in epithelial monolayers of MDCK cells (Fig. 2.9). They also elucidate the importance of cell-cell adhesion for the maintenance of the collective patterns. By knocking down cell-cell junctions in the tissue, they show that the migration speed decreases substantially and that migration happens more randomly, with cells in the front detaching in random *gas-like* motion. Video 08 and video 09, done by Vedula [17], shows the ordered migration of wild-type MDCK cells and the gas-like migration of MDCK in a low calcium media, respectively. Low calcium medium disables the normal cell-cell junctions function in MDCK cells [28].

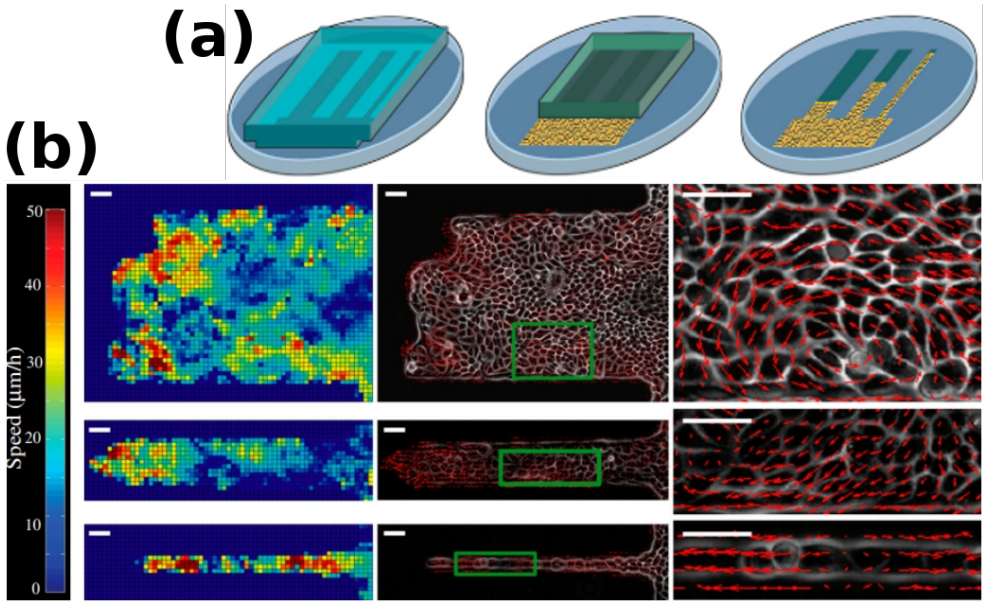

Fig. 2.9: Wound healing assay done by *Vedula et al 2015*. (a) The graphic explanation of the assay of the cells migrating from a reservoir to tracks. (b) Velocity fields associated with each track. The presence of vortex in the wider one shows how constraints can modulate migration patterns. Adapted from *Vedula et al 2012* ([17]).

Petitjean et al. 2010 (Fig. 2.10 (b)) show that for densities above the confluence point, division rate and overall monolayer speed decreases drastically. Results like this led to the belief that the physics counterpart of the biological contact inhibition could be of a jamming transition (Fig. 2.10 (a)), which is characterized in granular media physics as an abrupt arrest of movement resulting from density increase above a threshold.

The characterization of this transition's nature is still in debate ([4]). *Bi et al. 2015* ([29]) claim a new type of transition, related to the cell cortex, rather than density. A promising

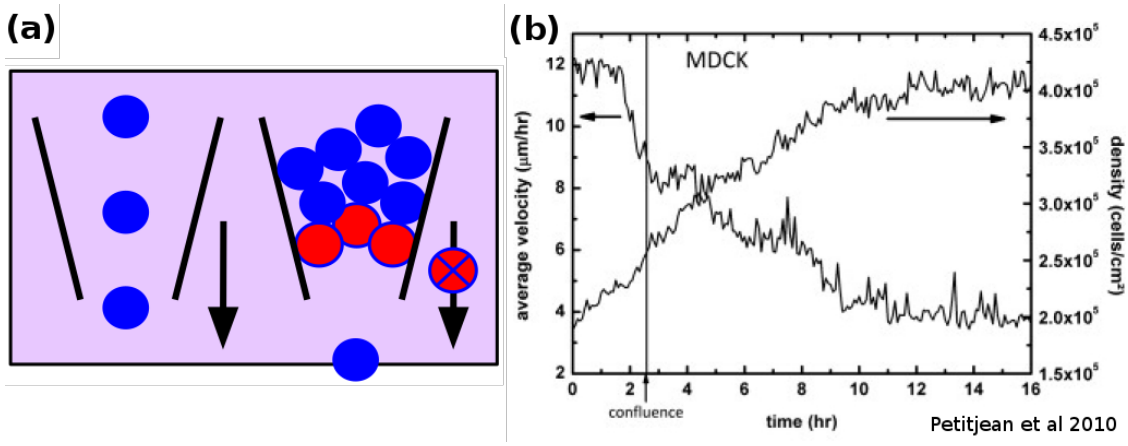

Fig. 2.10: (a) Jamming in granular media physics: the increase of density arresting the flow of particles. (b) Madin-Darby Canine Kidney (MDCK) lineage monolayer experiment: the fast decrease in velocity after reaching confluence point is accompanied by density increase. Both quantities reach a *quasi*-asymptotic state. Very similar behavior is expected for healthy tissue in vivo. (b) credits *Petitjean et al 2010* ([22]).

approach for this question comes from the work *Garcia et al. 2015* ([21]): they argue that the jammed dynamics in epithelial monolayers are controlled by the maturation of the cell-cell and cell-substrate junctions, placing density as a secondary parameter for the transition. As can be seen in figure 2.11 (A) and (B), while velocity drops with time (Fig. 2.14-B), the concentration of cadherins (proteins related to cell-cell adhesion) increases substantially. Figure 2.11 (C) shows the increase in vinculin, proteins related to cell-substrate adhesion.

Garcia's findings place jammed dynamics of epithelial tissue as a solidification problem (like water becoming ice). However, this solidification process occurs in a system of active particles increasing in density, setting this transition as a combination of two features: adhesion maturation and density. Based on these ideas, we call such process *jammed solidification transition* (JST). JST is then considered here the physical equivalent of the biological contact of inhibition.

A decrease in the monolayer's average velocity is a good and straightforward indicator of JST, as shown before. Another tool to analyze JST is correlation functions. Correlations are of significant importance in active matter physics, and the work of *Cavagni et al. 2018* ([30]) does a reliable review about it. The basic idea is to define the extent of similarity between two independent parts of the system. For example, when a cell dies or moves abruptly to the right, how far can neighboring cells receive this mechanical information?

Other relevant works on how cell-substrates and cell-cell interactions modulate collective behaviors is presented by *Angeline et al. 2010* ([23]) and *Petitjean et al. 2010* ([22]), respectively. By the use of soft substrate (with softness similar to real epithelial cells), Angelini and

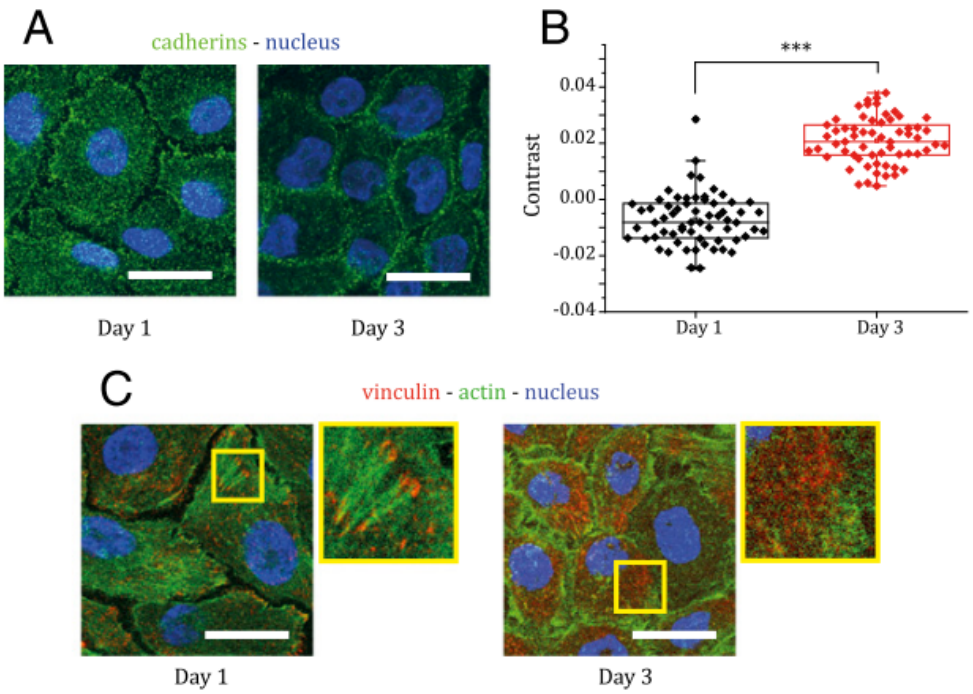

Fig. 2.11: (A, B) Increase in concentration of proteins (green) associated with cell-cell adhesion for days 1 to day 3. (B) Increase in concentration of proteins (red) associated with cell-substrate adhesion for days 1 to day 3. Credits: *Garcia et al 2015* ([21]).

co-workers grow a monolayer of epithelial cells and show that deformation of a soft substrate correlates with the increase of the swirls patterns (Fig. 2.12 A and B).

Using velocity correlation functions (Fig. 2.12 C), they extracted swirl sizes (defined as the minimum value of the correlation function). Figure 2.12 D shows that the swirl sizes increase in time, stressing the importance of cell-substrate interactions for macroscopic behavior. Working with fibroblast monolayers (cells with close to zero number of cell-cell adhesion proteins), Petitjean showed that while the epithelial (more collective) cells presented a correlation length of $\sim 200\mu m$, the fibroblast showed a correlation of $\sim 40\mu m$. Regarding the JST and average velocity, epithelial cells present a well-marked point with a fast decay, while fibroblasts show no marked transition point with a more monotonic decay (Fig. 2.13).

The work by *Garcia et al. 2015*, commented above, also presents a correlation analysis, but in contrast to Angeline, it uses a rigid substrate. The correlation length increases for early times (like the one of Angeline) but decreases at the end of the experiment, presenting a bell shape curve (Fig. 2.14 D).

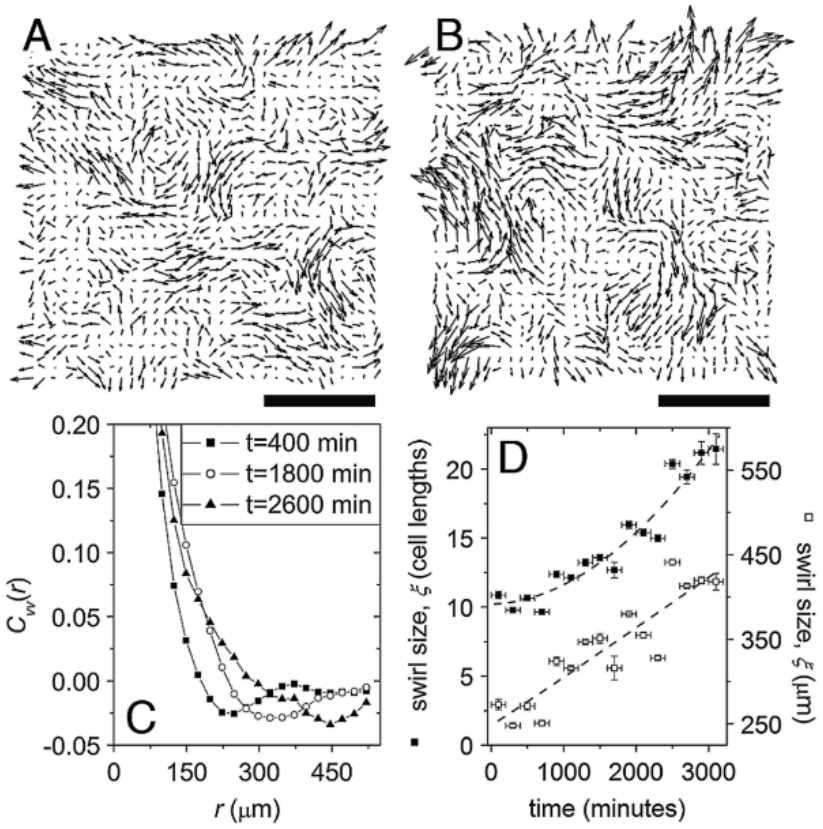

Fig. 2.12: (A) and (B): velocity field at the beginning of the experiment (A), and in the end, (B) showing that the swirl sizes have increased. (C) Velocity correlation functions for the associated vector fields. (D) Swirl sizes in microns and in cell sizes vs. time. Credits: *Angelini et al. 2010* ([23]).

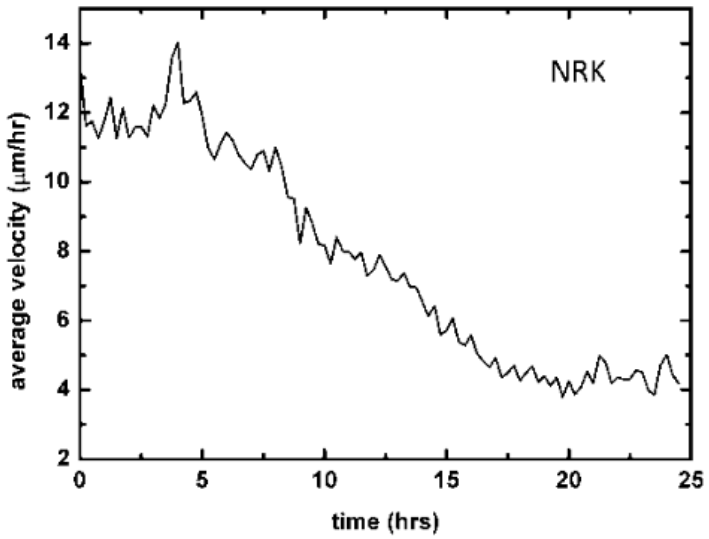

Fig. 2.13: Fibroblast NRK lineage monolayer: comparing with the MDCK from Fig. 2.10 (b), fibroblasts present a more monotonic decay of the average velocity. Credits: *Petitjean et al. 2010* ([22]).

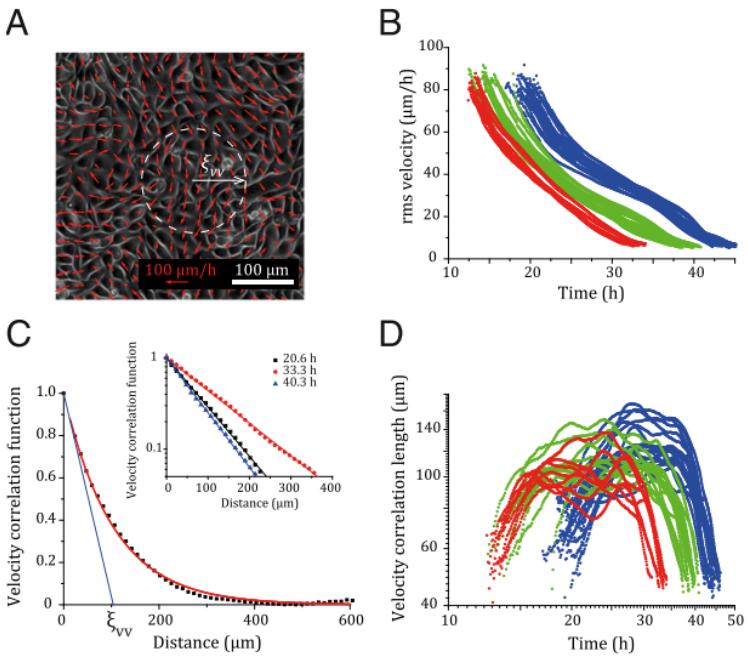

Fig. 2.14: (A) Snapshot of the MDCK monolayer merged with the measured velocity field. The dashed circle represents a typical correlation length for that image. (B) The average velocity decay in time for three different initial densities. (C) Velocity correlation function for the associated velocity fields. (D) Velocity correlation length for three different initial density showing the same behavior of bell curve for all. Credits: *Garcia et al. 2015* ([21]).

Chapter 3

Methodology

In this chapter, we explain the methodology we use in this work. Initially, we describe the experimental cell culture procedure and the time-lapse microscopy. Then we explain the image processing tools, including normalization and binarization. Next, we approach the image analysis technique called Particle Image Velocimetry (PIV), which we use to obtain the movies' velocity fields. Finally, we define quantities related to the velocity field, like activity decay rate and velocity correlation. Figure 3.1 shows in 5 steps the workflow commented above. Our collaborators Bibiana F. Matte[1] and Marcelo M. Lamers[2] did all the experimental part. The end of the workflow (image treatment, image analysis, calculations) are all integrated into a single *open source based* pipeline we developed for this dissertation (link github).

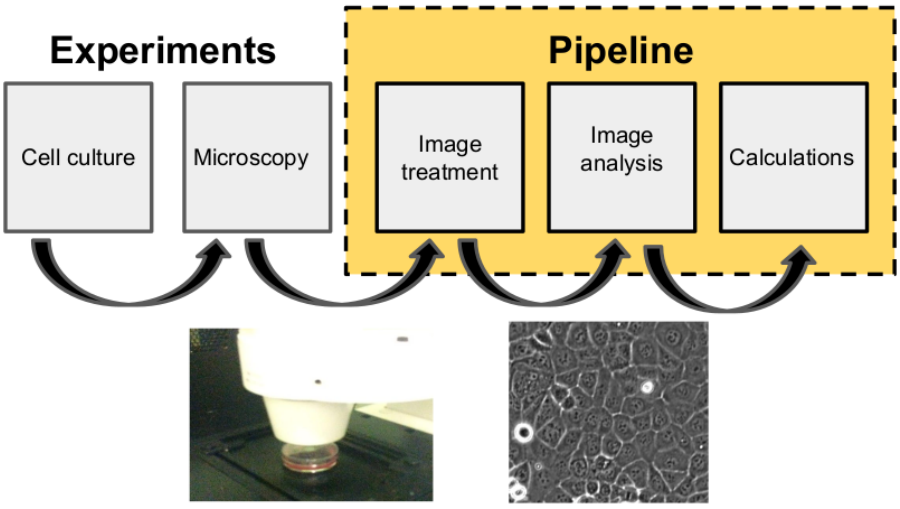

Fig. 3.1: Top: flowchart explaining the work methodology in five steps. Bottom: left, the cell plate under microscopy during data acquisition; right, the resulting image of the cell monolayer after image treatment.

[1] Department of Oral Pathology, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil.

[2] Department of Morphological Sciences, Institute of Basic Health Sciences, Universidade Federal of Rio Grande do Sul, Porto Alegre, RS, Brazil.

3.1 Experimental assay

3.1.1 Cell Culture

HaCat (human keratinocyte cell line) and SCC-9 (OSCC lineage) cell lines are obtained from Rio de Janeiro Cell Bank (BCRJ, Rio de Janeiro, RJ). CAL27 (OSCC lineage) is obtained from the Tissue Culture Facility at School of Medicine of University of Virginia. They are cultivated in Dulbecco's modified Eagle's media (DMEM) with high glucose (Gibco) supplemented with 10% Fetal Bovine Serum (FBS) (Gibco) and 1% penicillin/streptomycin (Gibco). All cells are maintained in incubator at 37°C with 5% CO2.

3.1.2 Time-lapse assay

Cells are plated on collagen-coated dishes ($5\mu g/cm^2$) at a density of 5x10^5 cells per well and maintained at the incubator for 48 hours (time they take to form a monolayer). Once the cells are in a monolayer, they are removed from the incubator and images are captured using phase contrast at 15 min intervals for 50 hours (Axio Observer Z1 microscope - Zeiss) with heat control at 37°C (but without atmosphere control) and a charge coupled device camera (Axiocam mrn, Zeiss) using a 10X objective (Ecplan-Neofluar 10x/0.3 aperture, Zeiss-) and AxioVision Software (Zeiss-). This results in a field of view (FOV) of 1388x1040 pixels^2 with a distance to pixel factor of 0.645 $\mu m/px$. The area of the FOV in μm is 600540.408 μm^2 (895.26x670.80 μm^2).

3.1.3 Replicates

We made two movies (replicates) for the healthy lineage HaCat, three for the less aggressive tumor lineage CAL27, and three for the more aggressive tumor lineage SCC9. All the velocity related results involve averages along the entire FOV. In all cases, these averages are performed in 600 cells or more for each time.

3.2 Image treatment

3.2.1 Normalization

Cell division in a confluent monolayer will necessarily increase its height, making significant changes in image brightness level (Figure 3.2 (a), (c), and (e)). To fix that we normalize the

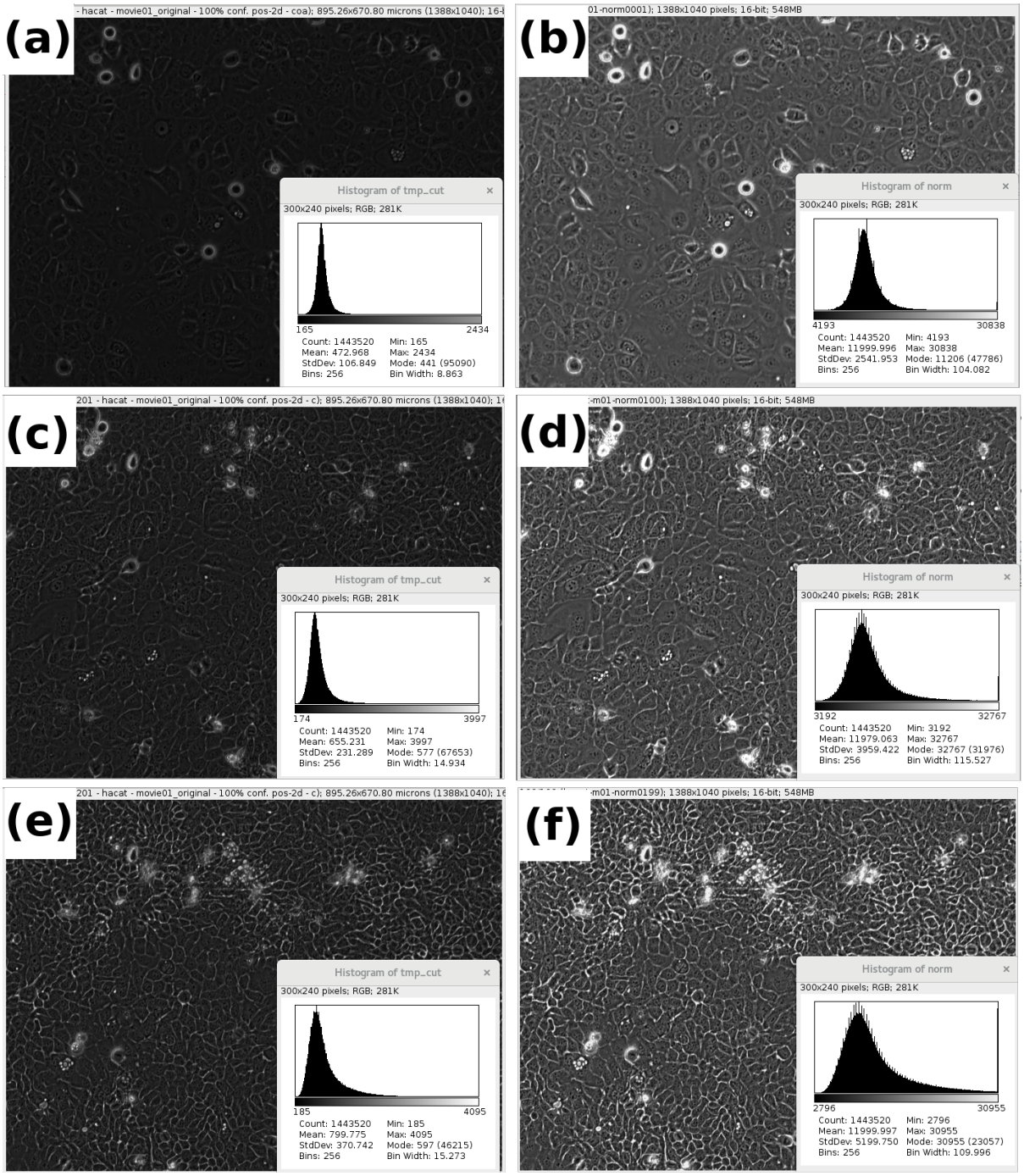

Fig. 3.2: Left, (a),(c) and (e): original monolayer images at t= 0.25h, 25.00h, and 47.75h, respectively. Right, (b),(d) and (f): the corresponded normalized images of (a), (c), and (e), respectively. Insets: pixels intensity histograms, obtained with FIJI, for each image.

images pixels intensity using an octave script. The resulting normalized images can be seen in Figure 3.2 (b), (d) and (e). Pixels intensity histograms are normalized separately for each image. The mean intensity is set to 12000, for original movies with mean bright below 800, or

to 10000, for movies with mean bright greater than 800.

3.2.2 Hysteresis

In a growing monolayer, some cells eventually die, and their debris accumulates on its top. This debris appears in the microscope as very bright spots and hinders cell recognition by image analyzes[3]. A similar problem appears during mitosis. In order to divide, cells need to detach from the substrate. The confluent high-density environment pushes the detached cells to the monolayer's top, producing a bright spot. This can be seen in Figure 3.2 (b), (d), (f).

To eliminate these bright spots, we use the multi-threshold Fiji ([31]) plugin to cut off the contributions of pixels with intensity higher than the mean normalized level (Fig. 3.3 (a)). After that, we use the hysteresis options of this plugin to binarize pixels intensity to 0 or 255 (Fig. 3.3 (b)). These procedures enhance PIV quality.

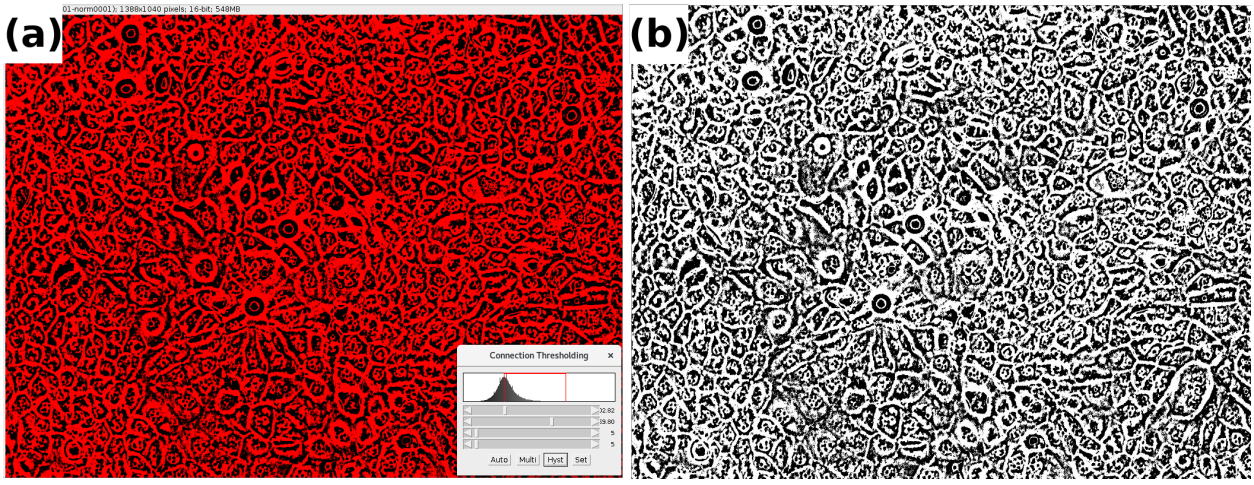

Fig. 3.3: (a) Threshold procedure after bright spots removal. Inset: pixel intensity histogram indicating the excluded region. (b) Same image after binarization.

3.3 Image analysis: measured quantities and data processing

3.3.1 Manual counting, growth rates and mean cell area

The monolayers number density is accessed manually by counting the number of cells inside the FOV. We use the point and click method [21] (see figure 3.4). Count cells in confluent assays

[3] In phase-contrast microscopy, higher objects in the FOV appears brighter than lower objects.

without stain or fluoresce is a slow and troublesome process. Since our main results do not directly depend on the measured density, we counted only a few frames for just one replicate of each lineage (see table 3.1). Visual inspection of the not counted replicates confirmed a similar behavior concerning density. We saved significant labor time using this approach. The downside is the obvious lack of statistical significance of the density-related results.

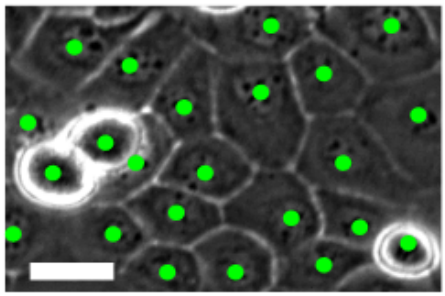

Fig. 3.4: Manual counting by point and click method using FIJI *manual_counting* plugin. Using the mouse courser, point to the cells' nuclei and click on it (green dots).

Lineage movie	(Frame time [h], Number of cells)
HaCat Movie 2	(0.25,708) | (05.25, 790) | (10.25 , 905) | (25.25,1203) | (30.00,1293) | (50.00,2050)
Cal27 Movie 1	(0.25,947) | (21.25,1243) | (42.50,1881) |
SCC9 Movie 1	(2.50,552) | (23.75, 666) | (45.00 , 987) |

Tab. 3.1: Each parenthesis is a data pair containing the time of the counted frame and the number of cells in that frame. For the healthy lineage HaCat, only five frames were counted and only for one replicate. For the tumor lineages, only three frames were counted and only for one replicate each.

We defined the growth rate of the lineages as given by equation 3.1:

$$Growth\_Rate = \frac{N^{t+\Delta t} - N^t}{\Delta t} \tag{3.1}$$

where $N^{t+\Delta t}$ and N^t are the number of cells measured in the end and beginning of the time interval Δt. This quantity tells how fast the number of cells increases during the time interval Δt[4]. The data used to compute the growth rates can be seen in table 3.2.

The growth rate should not be confused with mitotic index or duplication time since it only compares the number of cells of two frames separated by some time interval. No information

[4] This is the way to get the derivative of a linear function.

on the number of mitoses, cell death, and senescent cells is explicitly taken into account to compute it. It is an average time quantity that can be interpreted as the total number of new cells introduced by division subtracted by the number of cells death and added/subtracted by the inflow/outflow (flux) of cells in the FOV during the time interval Δt.

Lineage movie	$N^{t+\Delta t}$	N^t	$t + \Delta t$ [h]	t [h]	Δt [h]
HaCat Movie 2	2050	708	50.00	0.25	49.75
Cal27 Movie 1	1881	947	42.50	0.25	42.5
SCC9 Movie 1	987	552	45.00	2.50	42.50

Tab. 3.2: Data used to compute the lineages growth rate. N is the number of cells, and t and $t + \Delta t$ is the time in hours that the cells were counted. Δt is the time interval between measurements.

The mean cell area of a frame is computed by dividing the total area of the FOV (see section 3.1.2) by the number of cells in that frame. This quantity is different from the growth rate only by a constant (the area of the FOV).

3.3.2 Particle Image Velocimetry (PIV)

To measure the velocity field of the cells, we use a python open source Particle Image Velocimetry (PIV) software from the *openPIV-python* initiative [32, 33, 34]. We use a standard PIV cross-correlation function algorithm, in a single pass, with interrogation window linear size of 32 pixels, 16 pixels overlap, search linear area size of 86 pixels and signal to noise ratio method *peak to peak*.

For an image in 1388x1040px^2, the windows/overlap size used here returns a regular 85x64 lattice of characteristic length of 16 pixels with a total of 5440 vectors. Independent of the number or the size of the cells, the PIV returns a velocity field with always the same number of vectors. This happens because the PIV routine does not track individual cells' displacement, treating the monolayer as a continuum material/fluid, defining velocity vectors to predefined areas of regular sizes rather than individual cells. The PIV usually returns some outlier vectors that need to be removed and replaced in order to have a more accurate velocity description of the samples. The outliers elimination is done by signal to noise validation with threshold = 1.3 and outliers replacement with method = *local mean*, max iteration = 10 and kernel size = 2. More information can be found in openPIV-python page.

3.3.3 Velocity Field Frame maker (VFF)

After the PIV procedure, some outlier vectors with huge anomalous modulus remain (related with debris displacements in the top of the monolayer). We build a python algorithm (VFF) to eliminate those vectors. First, we eliminate the two external borders of the PIV data. In the remaining bulk vectors, we first remove the outliers using a 4-sigma clipping around the global mean module of vectors in the field. After that, we use a local 3-sigma clipping to remove outliers relative to the vectors' mean module in a circular region of 36 pixels radius, i.e., two next neighbors, making it a 13 vectors region.

We replaced the removed outliers with the remaining vectors' vectorial average inside the 36 pixels radius region. We do not replace vectors in the two external borders because it is impossible to define the two next neighbors' radius for them. The two external borders are then eliminated as well. At the end of this process, the four external borders of the 5440 vectors PIV original lattice are eliminated, resulting in a 77x56 lattice with 4312 vectors. In 95% of the cases (for all images in all replicates), the number of outliers is less than 1%. In 95% of these cases, the replaced vectors are averaged over eight or more vectors.

3.4 Calculated quantities

Using the resulting vector field from section 3.3.3, we perform two independent sets of calculations. One with four quantities related to velocities averages extracted directly from the velocity field here refereed as ADV quantities (an acronym for activity and drift velocities); and the second set with three quantities related to velocity correlations of the velocity field, here indicated by VAF (an acronym for velocity auto-correlation function). We define two independent parameters, one from ADV and another from VAF quantities, used to build a diagnostics diagram that separates the different analyzed cell lineages, presented at the end of the work.

3.4.1 Average velocity quantities (ADV)

In the literature, the average monolayer velocity is used to monitor jamming related transitions [21]. In those measurements, a set of N spatially distributed velocity vectors $\vec{v_i}(\vec{r_i})$ (and their associated modules $v_i = |\vec{v_i}|$) are transformed (averaged) into a single value for each frame of the movie. Here we compute three average quantities: (a) the root mean square velocity v_{rms} (Eq. 3.2), as done by *Garcia et al 2015* [21]; (b) the mean cell speed $\bar{v}$, i.e., the average

of the velocity module of each vector, which we define here as *Activity* (Eq. 3.3); (c) the mean drift speed $\bar{v}_{drift}$ of the monolayer (Eq. 3.4).

$$v_{rms} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}|\vec{v_i}|^2} = \sqrt{\left\langle|\vec{v}|^2\right\rangle_{\vec{r}}} \tag{3.2}$$

$$Activity \equiv \bar{v} = \frac{1}{N}\sum_{i=1}^{N}|\vec{v_i}| = \frac{1}{N}\sum_{i=1}^{N}v_i = \langle v\rangle_{\vec{r}} \tag{3.3}$$

$$\bar{v}_{drift} = \left|\overline{\vec{v}}\right| = \left|\frac{1}{N}\sum_{i=1}^{N}\vec{v_i}\right| = \left\langle\vec{v}\right\rangle_{\vec{r}} \tag{3.4}$$

In the above equations, the symbol $\sum$ denotes the summation of all the $N = 4312$ i vectors belonging to the velocity field. Dividing this sum of vectors by N defines the spatial average. The average can also be denoted by $\langle\rangle_{\vec{r}}$; where the suffix $\vec{r}$ denoted that the average is in space, i.e., in the FOV of the microscope.

The $\bar{v}_{drift}$ (the module of the drift velocity) defines how fast the monolayer is migrating (in average), i.e., moving in a collective and ordered direction. This is an equivalent parameter as the one used by Vicsek to monitor collective movement, as explained in section 2.2. Activity ($\bar{v}$) and v_{rms} are related to the average amount of energy that the cells are spending in movement, regardless if it is ordered (collective) or not. There is a difference between *movement* and *migration*, cells can be moving a lot (having then high activity), but not migrating at all (low $\bar{v}_{drift}$). For example, a cell moving in a circle around a fixed point: it is spending energy in movement but is not migrating (out of the circle region).

We also compute the standard deviation of activity ($\sigma_{\bar{v}}$) and defined it as the fluctuation in activity ($\delta Activity$, Eq. 3.5);

$$\delta Activity \equiv \sigma_{\bar{v}} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(v_i - \bar{v})^2} \tag{3.5}$$

the coefficient of variation (Γ, Eq. 3.6) then tells how homogeneous is the distribution of energy spent on movement. $\overline{\Gamma}$ (Eq. 3.7) is the time average of Γ for $0.25h \leq t \leq 45h$.

$$\Gamma = \frac{\delta Activity}{Activity} \tag{3.6}$$

$$\overline{\Gamma} = <\Gamma>_t \tag{3.7}$$

The activity characteristic decay time τ_{act}, is explained in chapter Results section 4.2.1. All the quantities explained here were obtained with a python code written by us.

3.4.2 Velocity correlation quantities (VAF)

The second set of computed quantities came from the velocity fluctuation field. This fluctuation field is obtained by subtracting the mean drift velocity $\left\langle \vec{v} \right\rangle_{\vec{r}}$ from each vector of the velocity field $\vec{v}(\vec{r_i})$:

$$\vec{v^*}(\vec{r_i}) = \vec{v}(\vec{r_i}) - \left\langle \vec{v} \right\rangle_{\vec{r}} \tag{3.8}$$

where $\vec{v^*}$ denotes the velocity fluctuation field and $\langle\rangle_{\vec{r}}$ is a space average (as explained in section 3.4.1). Figure 4.7 (a) and (b), from Results, show the velocity field and its associate fluctuation field obtained using Equation 3.8.

To account for the correlations in the fluctuation field, we use the velocity auto-correlation function C_{vv}, defined by Equation 3.9 below:

$$C_{vv}(\Delta r,t) = \frac{\left\langle \vec{v^*}\left(\vec{r}+\Delta r,t\right) \cdot \vec{v^*}\left(\vec{r},t\right) \right\rangle_{\vec{r}}}{\left\langle \left| \vec{v^*}\left(\vec{r},t\right) \right|^2 \right\rangle_{\vec{r}}} \tag{3.9}$$

where the $\vec{v^*}$ represents the velocity vectors from the fluctuation fields defined by Equation 3.8 and $\langle\rangle_r$ denotes the average over the space (FOV). This function tells us how much the movement of a part of the system is correlated (similar) to another part of the system at distance Δr. In other words, it tells, on average, how much a velocity vector $\vec{v^*}\left(\vec{r},t\right)$ is similar to other velocity vector $\vec{v^*}\left(\vec{r}+\Delta r,t\right)$ when separated a distance Δr from each other.

The velocity correlation function does not have a very friendly formulation, but it is not hard to get intuition over it. The $C_{vv}(\Delta r_i)$ value is a fraction where the numerator is based in a mathematical tool called *dot product*, which is used to calculate how similar are the directions of two vectors. Two parallel vectors will have the maximum value of similarity, and two orthogonal vectors will have a zero value of similarity. C_{vv} can even returns negative correlation values, associated with *anti-parallel* vectors.

The numerator in expression 3.9 takes all possible vectors pair at the same defined distance Δr, applies the dot product to get their similarity, and then computes an average (on space) over all those vectors pairs. The denominator enters just as a normalization factor, eliminating the overall velocity module's contribution and making it a dimensionless quantity. This normalization defines the maximum value for positive correlation as one. The normalization enables us to compare correlation curves for fields with different velocity modules. An example of a correlation curve can be seen in figure 3.5, and more information about correlation and fluctuation tools in the context of biological physics can be found in [30].

For a regular lattice of 77x56 sites (like the ones used here), there are 2015 distinct possible Δr distances between the sites (vectors positions). This results in a $C_{vv}(\Delta r,t)$ function with 2015 points. We binned this 2015 point in space over a window of size 16 μm resulting in a much smoother curve (Fig. 3.5). The fluctuation field and C_{vv} curves are obtained with a python-FORTRAN algorithm written by us.

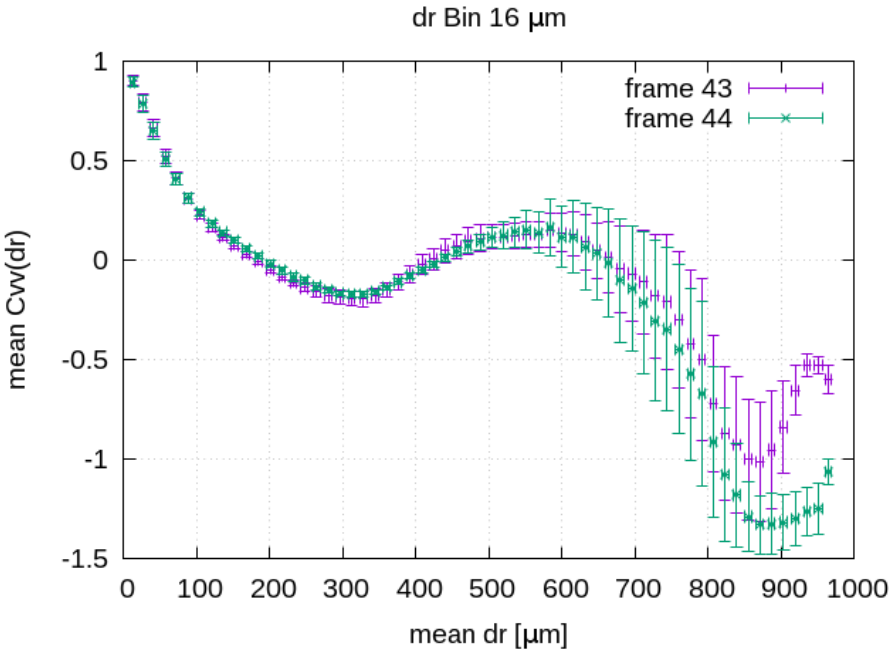

Fig. 3.5: Velocity correlation function plotted as a function of the distance for two images separated by a time-lapse of 15min. The curves are averaged in a 16 μm window. Error bars result from the averaging process.

From the positive part of C_{vv}, we extract three quantities related to the correlation in the system:

- *correlation length*, λ, related to how fast the C_{vv} reaches the zero value (the bigger the λ, the smaller the fall), obtained by fitting an exponential decay function (Eq. 3.10) to the positive part of the C_{vv}, as done by *Garcia et al 2015* [21];
- *effective distance of zero correlation*, ξ, defined as the distance where C_{vv} is equals to zero [30];
- *integrated total degree of correlation*, χ_{int}, defined by Equation 3.11 as the summation of all the positive values of the C_{vv} for a distance smaller than ξ.

$$f(\Delta r) = exp(-\frac{\Delta r}{\lambda}) \tag{3.10}$$

$$\chi_{int} = \int_0^{\xi} Cvv(r)dr \tag{3.11}$$

By construction, these three quantities have units of length and are exemplified in Figure 3.6. The quantity χ_{int} is an integral form presented here for the *total degree of correlation* χ originally presented by *Cavagna et al. 2018*. The Python codes written by us obtain these quantities: λ is obtained using python module *scipy.optimize.curve_fit* and χ_{int} by standard trapezoidal integration (C_{vv} curves in Fig. 3.6).

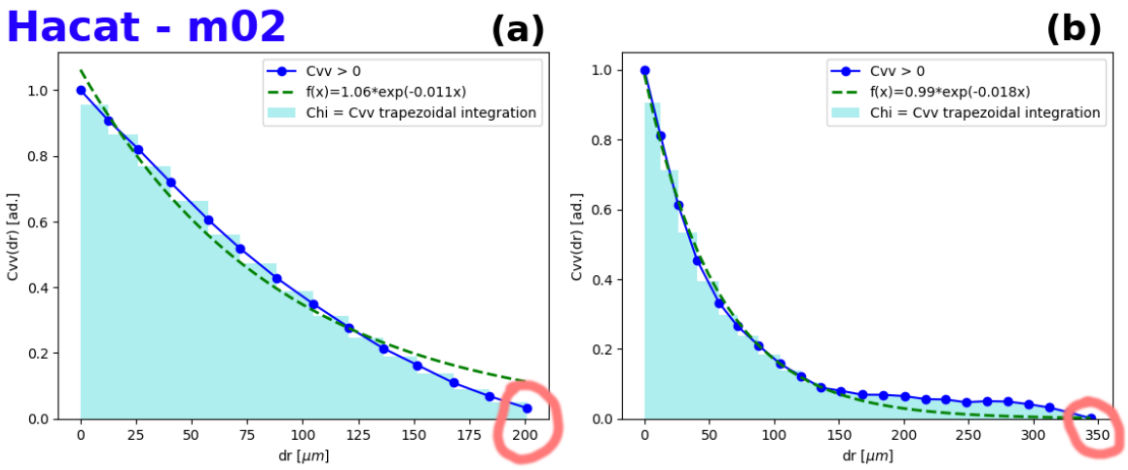

Fig. 3.6: Velocity correlation for the HaCat monolayer at two different times. Chi is the Greek name for the letter χ and corresponds to χ_{int} defined by Eq. 3.11. The green curve is a fit function for the C_{vv} using Eq. 3.10. Marked in pink is the effective distance of zero correlation ξ: (a) $\xi = 200\mu m$ and (b) $\xi = 350\mu m$.

The *mean positive correlation* $\overline{C}^{+}_{vv}$, also presented at the end of section 4.3, is defined as:

$$\overline{C}^{+}_{vv} = \frac{\chi_{int}}{\xi} \ . \tag{3.12}$$

$\overline{C}^{+}_{vv}$ is a dimensionless normalized quantity enabling us to compare C_{vv} curves with different ξ (like the ones from Fig. 3.6). An inspection of Eqs. 3.11 and 3.12 results in the equivalent form:

$$\overline{C}^{+}_{vv} = \frac{1}{\xi}\int_0^{\xi} Cvv(r)dr \tag{3.13}$$

showing a clear interpretation of $\overline{C}^{+}_{vv}$ as the average correlation in the positive C_{vv} region.

The quantities η and $\delta\eta$, presented in section 4.3.1 as parameters to separate tumorous behavior, are defined as follows: η is the *time average of the mean positive correlation*, analytically

given by Equation 3.14 below:

$$\eta \;=\; \frac{1}{40h}\int_{5h}^{45h} \overline{C}^{+}_{vv}(t)dt \;=\; <\overline{C}^{+}_{vv}>_t \,, \tag{3.14}$$

but, since we have finite time intervals among images, the way we actually computed it is given by Equation 3.15 below:

$$\eta \;=\; <\overline{C}^{+}_{vv}>_t \;=\; \frac{1}{N}\sum_{i=0}^{N-1} \overline{C}^{+}_{vv}(t = 5h + i\Delta t) \tag{3.15}$$

where $\Delta t = 0.25h = 15$min (the microscope time lapse) and $N = 161$ points ($40h/0.25h = 160$, $+1$ because is a closed interval); when $i = N - 1$, $t = 5h + 160 * 0.25h = 45h$ and when $i = 0$, $t = 5h + 0 * 0.25h = 5h$. Is not hard to show equivalence to Equation 3.14, but 3.15 have the advantage of a straight forward standard deviation that we used to define the fluctuation $\delta\eta$:

$$\delta\eta \;=\; \sqrt{\frac{1}{N-1}\sum_{i=0}^{N-1}\left(\overline{C}^{+}_{vv}(t = 5h + i\Delta t) - <\overline{C}^{+}_{vv}>_t\right)^2} \tag{3.16}$$

Chapter 4

Results

This chapter has four sections. First, we show an overview of the growth of the analyzed monolayers. Second, we analyze quantities related to the monolayers' overall speed, velocity averages, collective movement, and jamming, defining a parameter that separates the more aggressive tumor from the other two lineages. In the third section, we show results associated with the monolayer's velocity correlation, extracting a parameter that separates the healthy lineage from the tumor ones. Finally, in the last section, we present a diagram that can separate the three lineages into three distinct regions. The results graphs in figures 4.3, 4.4, 4.5, 4.6, 4.9 and 4.10 all follow the same format, explained in section 6.2 of the Appendix.

4.1 The growth of monolayers

Figure 4.1 shows three pairs of images, one for each of the three lineages analyzed: the healthy one, HaCat, the less aggressive tumor CAL27, and the more aggressive tumor SCC9. Left columns show monolayers at $\sim 8h$, and right columns show the respective monolayers 27 hours later ($\sim 35h$). Time = 0h corresponds to the moment the sample was removed from the incubator and taken to the microscope. After confluence cells will keep growing and dividing, as can be seen in Figure 4.2 (a). Since there is no more free space to grow, cells will squeeze each other, increasing monolayer height and decreasing cells visible area, as can be seen in Fig. 4.1 (b), (d), (f) and in Fig. 4.2 (b) ([23] shows similar results). The growth of HaCat, CAL27 and SCC9 monolayers can be seen in video 10, video 11 and video 12, respectively.

Density and growth rate are measured roughly by manually counting only a few frames from only one replicate of each lineage (section 3.3.1). We found that the HaCat monolayer has grown an average of 27 cells per hour during the experiment. CAL27 and SCC9 grown on average 22 and 10 cells per hour, respectively. It is interesting to see that the growth rate is inversely proportional to the aggressiveness of the lineages. This is a surprising result, but an

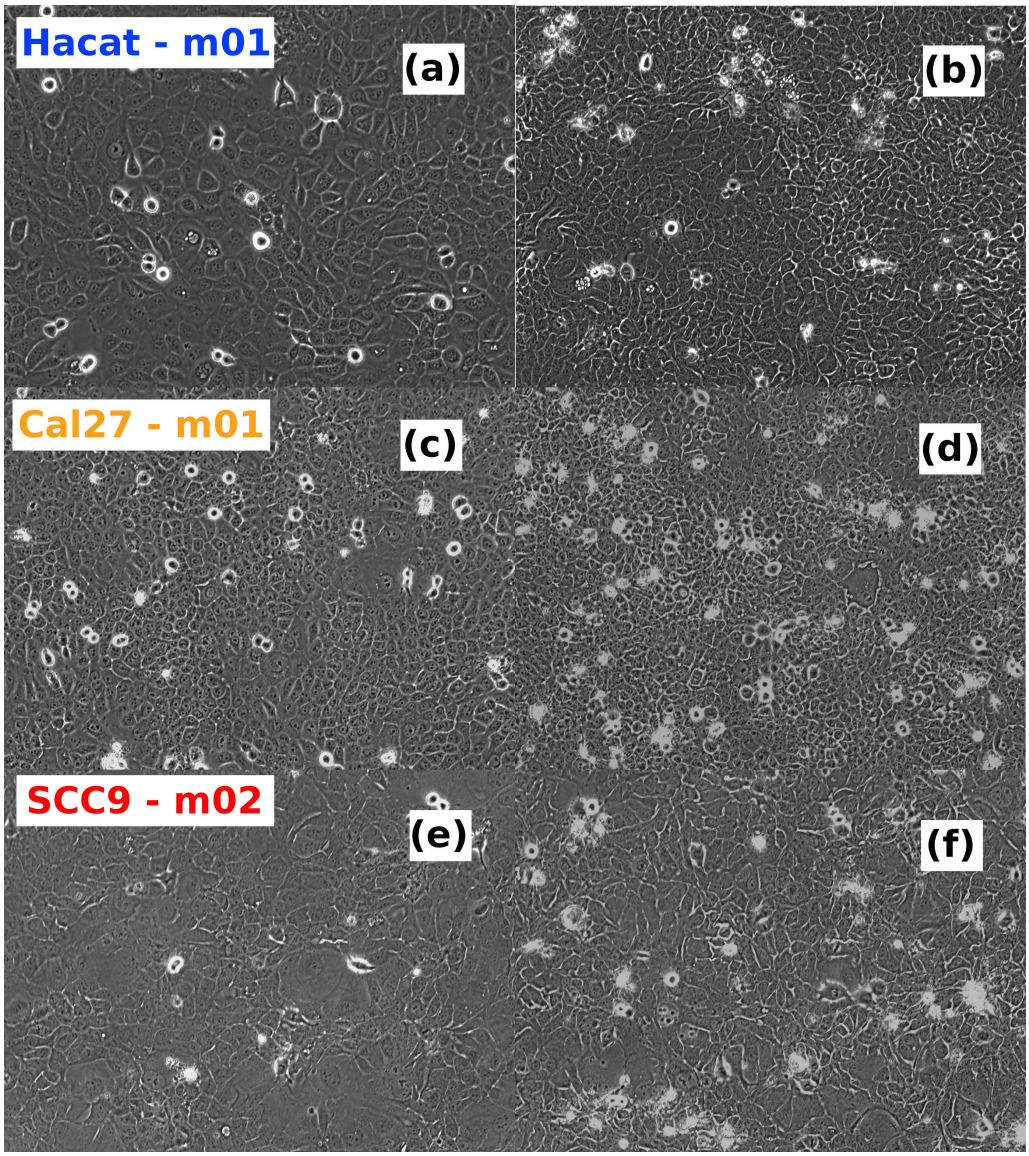

Fig. 4.1: Image of healthy (HaCat) (a), benignant tumor (CAL27) (c), and malignant tumor (SCC9) (e) lineages at ∼8h. (b),(d) and (f) correspond to the same lineages at ∼ 35h: the number of cells has increased significantly, making them appear visually smaller. Brighter and round objects are cells at the monolayer top during a mitoses process. In (b),(d), and (f), we can see a more significant number of white spots with no regular shape. They are cell debris on the top of the monolayer.

accepted characteristic of tumors states is that they keep dividing even when it is no longer necessary. They do not need to divide faster. They need to keep dividing when it is no longer needed. However, it is essential to state that the growth rate is a combination of division rates, death rates, and in/outflow of cells in the FOV. Cells that divide faster but die faster may have the same growth rate as cells with more stable behavior.

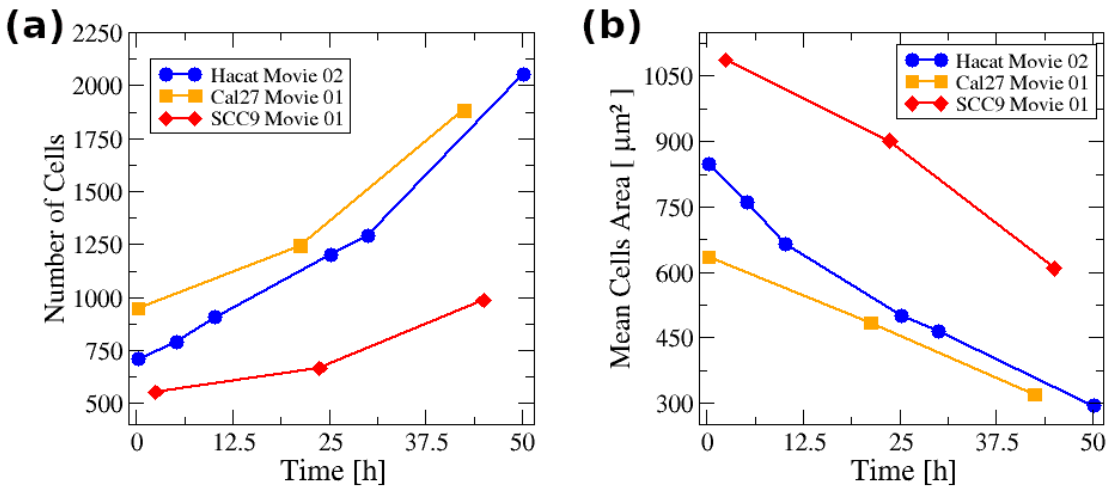

Fig. 4.2: (a) Number of cells (in the FOV) versus time (in hours) for the three lineages. (b) Mean cell area versus time (in hours) for the three lineages.

4.2 Velocity fields, velocity averages, collective movement, and jamming

We used PIV to extract the velocity fields of the monolayers. For the three lineages, there is a significant decrease in the velocity modules, as can be seen in Figures 6.4, 6.5 and 6.6 of the Appendix. Video 13, video 14, and video 15 show how the velocity field changes in time for HaCat, CAL27, and SCC9 monolayers, respectively. Video 16 shows the HaCat' velocity field merged with the microscope images. As explained in section 3.3, PIV always returns the same number of vectors, independent of cell number or cell size. For each square window of $\sim 165\mu m^2$, one velocity vector is attributed. Comparing the window areas with the values for mean cell area from figure 4.2 (b) we see that at the beginning of the experiment, a HaCat cell is described by ∼5 vectors and in the end by ∼2 vectors. For CAL27 is ~ 3.5 and ~ 2, and SCC9 is ~ 6.5 and ~ 3.5.

The down-regulation of cell-cell junctions enables malignant cells to unjam from the epithelial tissue and spread over the body [19, 20, 4]. We hypothesize that the malignant tumors would be insensitive to the JST transition. *Garia et al. 2015* [21] used the v_{rms} speed, i.e., the average value spent in movement regardless if it ordered or not, to monitor JST in epithelial monolayers. If our hypothesis is correct, we should not see significant changes in v_{rms} during the time. We found that the v_{rms} speed of the monolayers change significantly for the three lineages during the experiment (Fig. 4.3, blue curve). This result refutes our hypothesis showing that the two tumor lines are not insensitive to the JST.

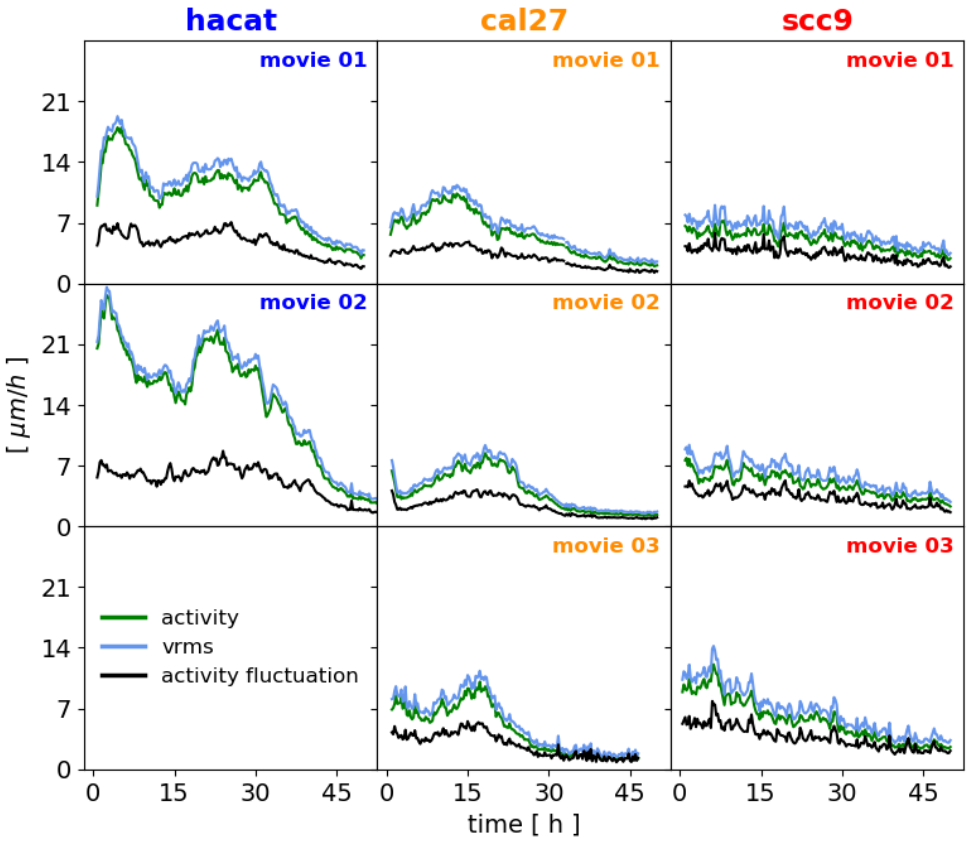

Fig. 4.3: Velocity related quantities (in $\mu m/h$) plotted versus time (in hours) for each lineage and its replicates. Time zero is when cells were taken from cell culture to the microscope. The graph format is the same as 6.3.

Here we propose that *Activity* (Eq. 3.3), i.e., average module of the velocity of the cells, could be used instead of v_{rms} to monitor JST. Both quantities have equivalent physical meaning and approximately the same behavior (Fig. 4.3). However, the more friendly formulation of *Activity* has a straight forward standard deviation in space that can be used to measure the activity fluctuations ($\delta Activity$, Eq. 3.5). The same cannot be attributed to v_{vrms} so easily. The $\delta Activity$ (Fig. 4.3, black curve) has similar values for the three lineages, but when compared with the *Activity* values, in the coefficient of variation (Γ, Eq. 3.6), we see that the tumor lineages have higher values of Γ (Fig. 4.4). This result shows that the distribution of energy spent in movement on the healthy monolayers is more homogeneous than on the tumor monolayers.

The average drift speed ($\overline{v}_{drift}$, Eq. 3.4) of the SCC9 monolayers is close to zero during the experiments, while HaCat and CAL27 present positive values (Fig. 4.5). For later times the $\overline{v}_{drift}$ of HaCat and CAL27 also goes to zero, but this is a consequence of the overall decrease in activity (Fig. 4.3). For SCC9, even when activity is high (Fig. 4.3), the $\overline{v}_{drift}$ is close to

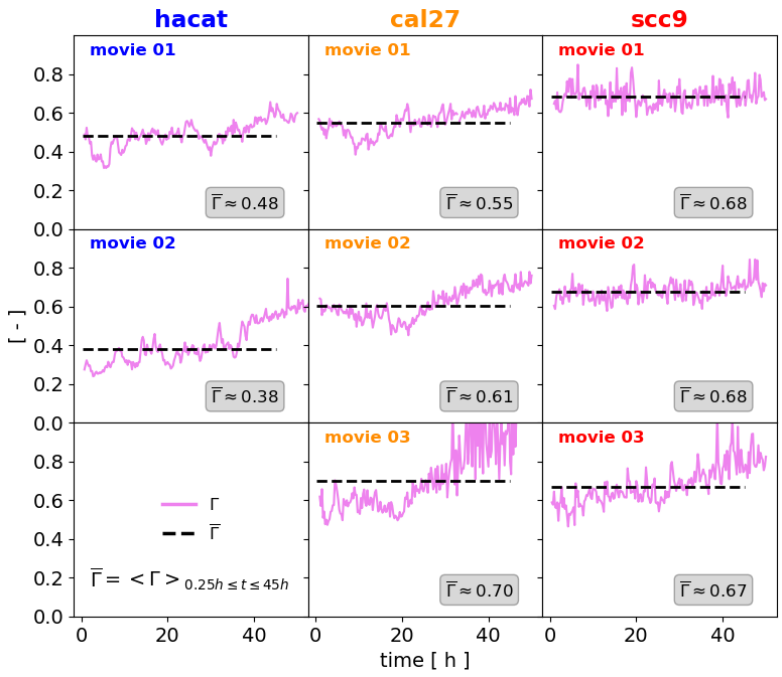

Fig. 4.4: The activity coefficient of variation Γ (no units) and the time average of Γ (dashed line) plotted versus time (in hours) for each lineage and its replicates. Time zero is when cells were taken from cell culture to the microscope. The graph format is the same as 6.3.

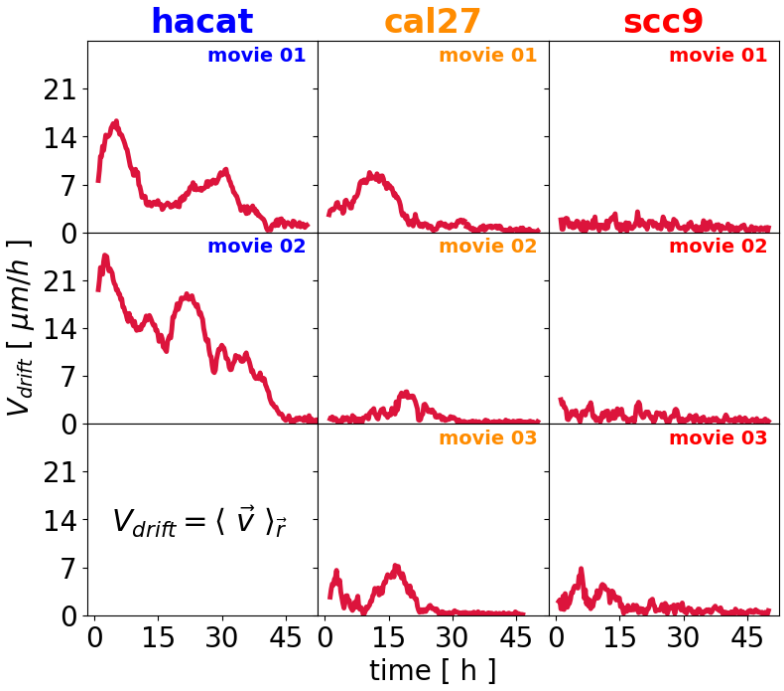

Fig. 4.5: Drift speed (in $\mu m/h$) plotted versus time (in hours) for each lineage and its replicates. Time zero is when cells were taken from the incubator to the microscope. The graph format is the same as 6.3.

zero (Fig. 4.5). Being so, SCC9 shows the signature of a disordered non-collective system [11, 12, 13]. This result corroborates the idea that the more aggressive tumor lineage (SCC9) is non-collective.

4.2.1 Activity characteristic decay time

The three lineages' activity decreases in time, as shown in Figure 4.3, but the activity decrease for the more aggressive tumor is different from the other two lineages. HaCat and CAL27 have a fast decay with marked JST points, starting around 25∼30h and 15∼20h, respectively. The SCC9, differently, has no marked point with a monotonic decay. To quantify this difference, we fitted the activity curves with an exponential decay function ($f(t)$):

$$f(t) = exp(-\frac{t}{\tau_{act}}) \tag{4.1}$$

where τ_{act} is the *Activity*' characteristic decay time.

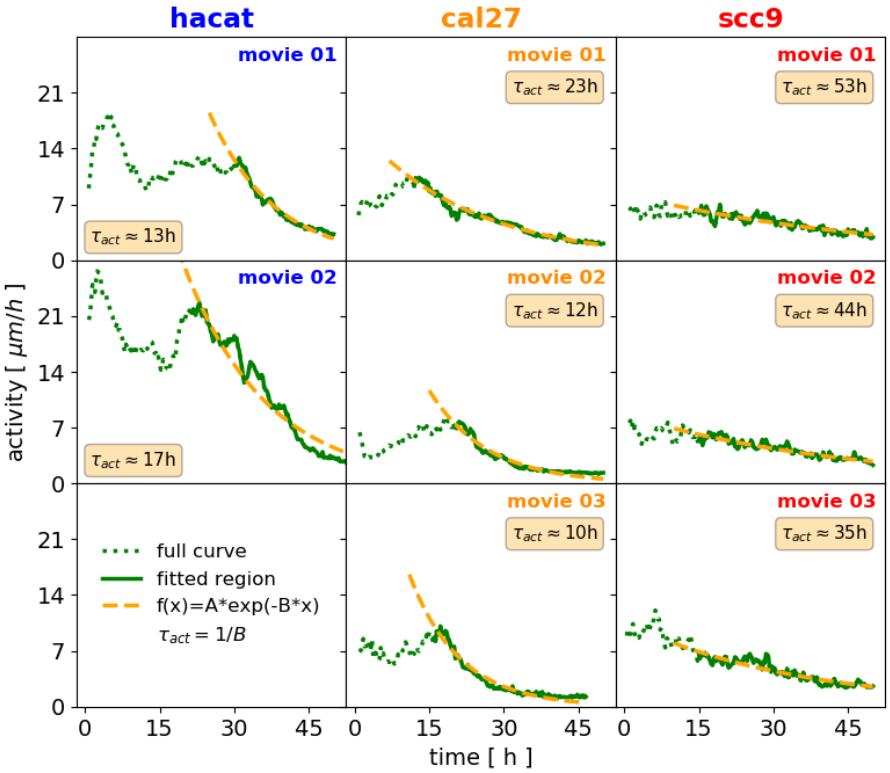

Fig. 4.6: Activity curves (in $\mu m/h$) for each lineage and its replicates. The dotted curve is the full activity curve, the solid curve is the region chosen to be fitted, and the dashed one is the fitted function. Insets show the τ_{act} value for each movie. The graph format is the same as 6.3.

As can be seen in figure 4.6 the values of τ_{act} for SCC9 are significantly larger than HaCat and CAL27. This parameter separates the more aggressive tumor's behavior from the healthy

and the less aggressive tumor. The JST point of each movie defines the fit region for the HaCat and CAL27 lineages. For SCC9, since there is no marked JST point, we chose the arbitrary value of 15h as the starting point of the fitting region. The idea was to eliminate the first hours of the experiment so that the *size* of the fitting region would be similar to the two other lineages.

4.3 Fluctuation fields and velocity correlation

Velocity correlation analysis informs how similar the directions of individual cells' movement are and how far in space these similarities extent[1]. It is essential to state that correlation analysis must be done over the fluctuations of the velocity fields [30], rather than the velocity field itself. So we subtract the average velocity as explained in Methods. A didactic example explaining why correlations are calculated over fluctuation fields can be found in section 6.4 of Appendix.

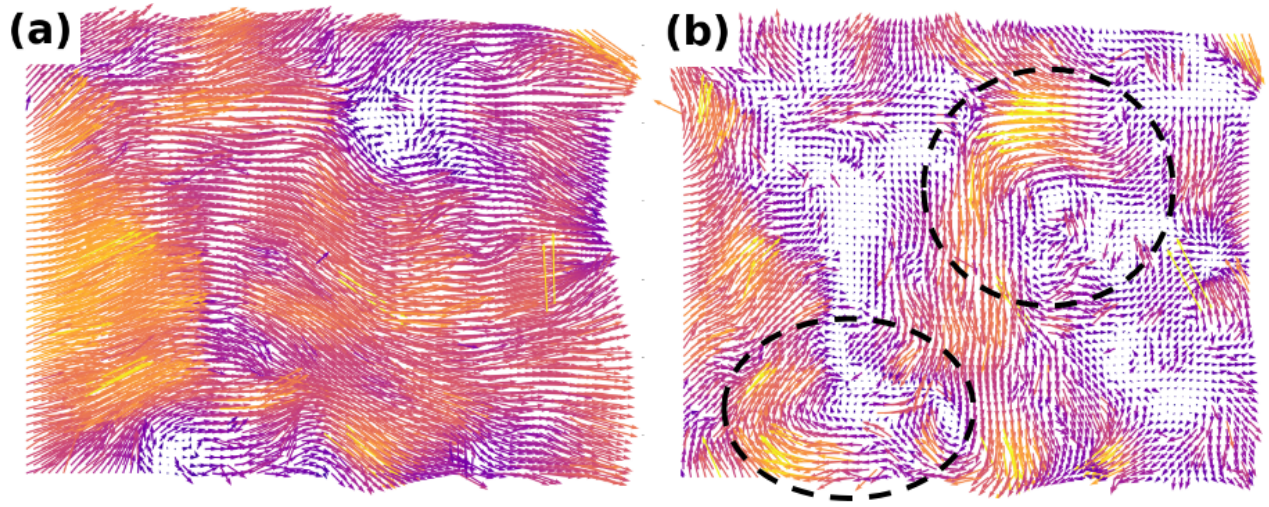

Fig. 4.7: (a) Velocity field for the healthy lineage HaCat (movie 2) one hour after the experiment's start. A significant drift to the right is present. The warmer colors are associated with higher velocity modules. (b) Fluctuation field resulting from subtracting the mean drift velocity from data in (a). Swirls of several cell sizes (depicted by the dashed black lines) emerge from this procedure, showing the underlying correlation behavior behind the velocity field.

Figure 4.7 (a) shows the velocity field for the healthy lineage (HaCat) one hour after the experiment starts, where an apparent drift to the right can be seen. Figure 4.7 (b) shows the

[1] Correlation analysis can also be done in time, but here we are only working with spatial correlation, and seeing how they change in time

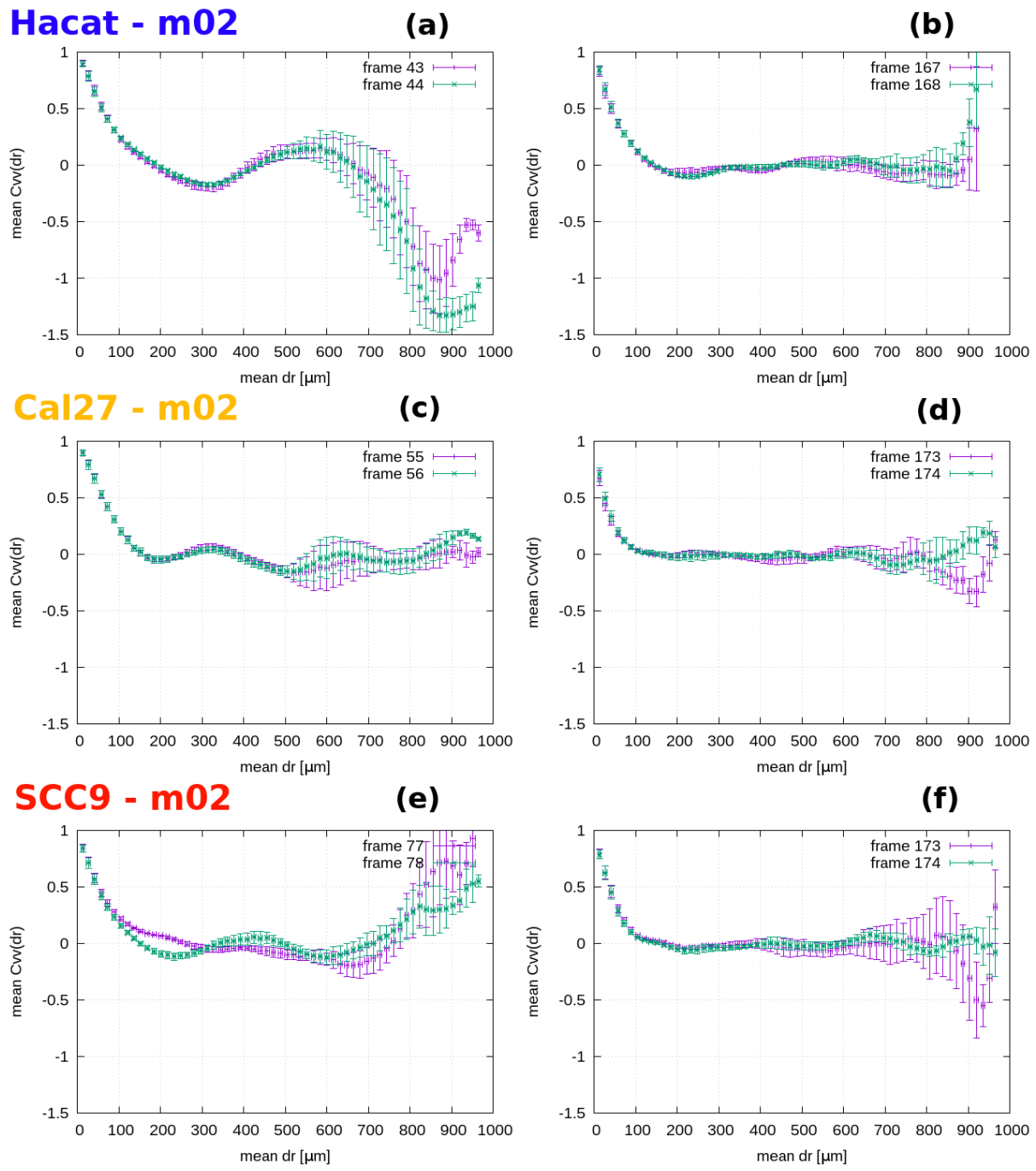

Fig. 4.8: VAF curves for the three different lineages. Frames are taken each 1/4h. For example, frame 40 is taken 10h after the experiment starts. The C_{vv} graphs were binned in a window of 10.32 μm, and the error bars are related to that procedure (see methods).

associated fluctuation field obtained when using Equation 3.8 to data in Figure 4.7 (a). By eliminating the drift bias, the fluctuation field reveals the actual size and patterns of velocity correlation in the monolayer; it shows large size swirls extending to more than 20 cells sizes,

these swirls are depicted by the dashed black lines. In a rough interpretation cell's movement affects other cells in a distance of 20 cell diameters, in this case.

In the fluctuation field, like Figure 4.7 (b), it is easy to visually identify swirls and quantify the extent of velocity correlations by measuring their sizes. However, it is interesting to have mathematical tools to transform those images into numbers without manual or visual procedures. As explained in the Methods, the mathematical tools that enable this are the autocorrelation functions. When applied to a velocity field, it is called the velocity autocorrelation function (VAF) [21, 23]. For convenience, we repeat its definition below (Eq. 4.2).

$$C_{vv}(\Delta r,t) = \frac{\left\langle \vec{v^*}\left(\vec{r}+\Delta r,t\right)\cdot\vec{v^*}\left(\vec{r},t\right)\right\rangle_{\vec{r}}}{\left\langle\left|\vec{v^*}\left(\vec{r},t\right)\right|^2\right\rangle_{\vec{r}}} \tag{4.2}$$

Examples of VAF functions (for the three lineages) can be seen in Figure 4.8. By construction the C_{vv} value for $\Delta r = 0$ is equal to 1.0, i.e., the maximum positive correlation value. For $\Delta r < 600\mu m$, all graphs in Figure 4.8 shows similar behaviour, starting from $C_{vv}(\Delta r = 0) = 1.0$ and then decreasing for increasing Δr values. The point where C_{vv} reaches zero is defined as the effective distance of positive correlation (ξ), i.e. $C_{vv}(\Delta r = \xi) = 0$. The region with values greater than zero, i.e., $\Delta r < \xi$, is used to define the velocity correlation length (λ) and the integrated total degree of correlation (χ_{int}). Both quantities have units of length and can be used to define the extent of correlation, as defined in methods.

In Figure 4.9, λ and χ_{int} present similar behavior and almost the same values for all replicates of the three lineages. For the healthy lineage (HaCat), λ and χ_{int} reach maximum values of $\sim 135\mu m$ at ~25-30h and then decrease to $\sim 45\mu m$ at 45h. Similarly, for the less aggressive tumor (CAL27), λ and χ_{int} reach maximum values of $\sim 90\mu m$ at ~15-25h and then decrease to $\sim 30\mu m$ at 45h. For the more aggressive tumor (SCC9), λ and χ_{int} stay constant around $\sim 45\mu m$ or decrease monotonically from $\sim 90\mu m$ at 0h to $\sim 40\mu m$ at 45h. These results show that λ and χ_{int} are equivalent measurements for the extent of the velocity correlation, and that the tumor lineages have smaller velocity correlation values than the healthy linage.

A limitation of the result in Figure 4.9 is the fact that the cells have different sizes. As shown in Figure 4.2, CAL27 size is smaller than HaCat and SCC9. If the graphs in Figure 4.9 were made in cell length instead of μm, CAL27 would have higher values of λ and χ_{int}. In this situation, the correlation length of CAL27 could be closest to HaCat values rather than SCC9 values. This problem can be fixed by normalizing λ and χ_{int} in order to have an dimensionless quantity. The more natural approach would be to use the cell size as normalization. However, since we do not have substantial data on cell sizes, we divide χ_{int} by the effective distance of

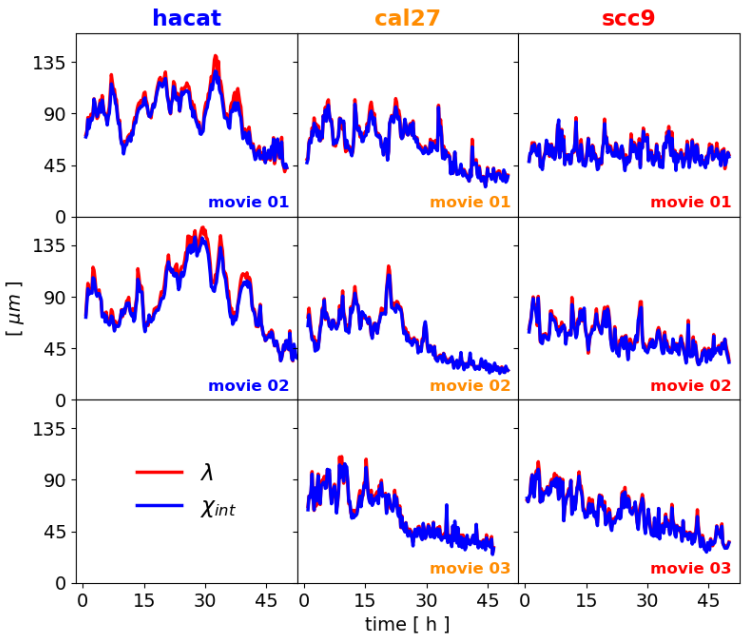

Fig. 4.9: Velocity correlation length, λ (red), and integrated total velocity correlation degree, χ_{int} (blue): note the similarity of the curves, indicating the two measurements' equivalence. Graph format is the same as 6.3.

zero correlation (ξ), defining the dimensionless quantity $\overline{C}^{+}_{vv}$, given by Equation 4.4, which is simply the average positive velocity correlation (this is explained in the methods). The graph of ξ for all the lineages can be seen in Appendix 6.5.

$$\overline{C}^{+}_{vv} \equiv \frac{\chi_{int}}{\xi} = \frac{1}{\xi}\int_{0}^{\xi} Cvv(r)dr \ . \tag{4.3}$$

$\overline{C}^{+}_{vv}$ can be seen in Figure 4.10 (light-blue curve). This quantity tells on average how ordered are the velocity vectors inside the region where the correlation is positive. For example, if all the vectors inside this region are perfectly aligned, then $\overline{C}^{+}_{vv}$ is equal to one. For HaCat the $\overline{C}^{+}_{vv}$ value is ∼0.4 from t=0h to 40h, decreasing to ∼0.20 at 45h. The $\overline{C}^{+}_{vv}$ value for the tumor lineages fluctuates around ∼0.3. Also, for the tumor lineages, $\overline{C}^{+}_{vv}$ fluctuates more in time. These differences in $\overline{C}^{+}_{vv}$ between the lineages could separate the tumor behavior of CAL27 and SCC9 from the healthy behavior of HaCat.

4.3.1 Time average of the mean positive correlation: a parameter to separate tumor behavior

To account for the visual finding in Figure 4.10 and to quantify the difference in $\overline{C}^{+}_{vv}$ into a single number, we defined the parameter η as the temporal average of $\overline{C}^{+}_{vv}$ (see Methods). For

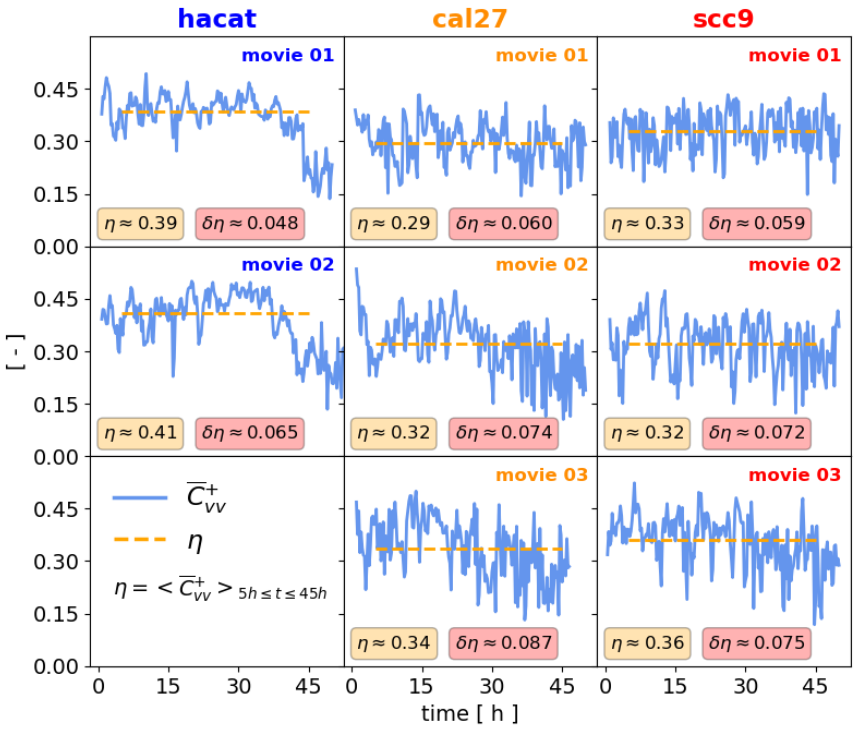

Fig. 4.10: The mean positive correlation $\overline{C}^{+}_{vv}$ (solid blue line) versus time for all lineages and its replicates. Dashed orange line (η) is the temporal average of $\overline{C}^{+}_{vv}$. The insets show that the healthy lineage HaCat have greater values of η and lower values of $\delta\eta$. The graph format is the same as 6.3.

commodity we repeat here its expression:

$$\eta \quad =< \overline{C}^{+}_{vv} >_{5h \le t \le 45h} \quad . \tag{4.4}$$

Its fluctuation, $\delta\eta$, is the standard deviation of $< \overline{C}^{+}_{vv} >_t$. As can be seen in Figure 4.10, the greater values of η, and the lower values of $\delta\eta$, associated with the healthy lineage (HaCat), enables us to separate healthy from tumor behavior.

4.4 Separating the behavior of the healthy lineage, of the less aggressive tumor and of the more aggressive tumor

In section 4.2.1, we show that τ_{act}, i.e., the activity decrease rate of the monolayers, is significantly higher for the more aggressive tumor, while in section 4.3.1, we show that η, i.e., the time average of the monolayers mean positive correlation, is lower for the tumor lineages. In the spirit of providing a tool to separate healthy, less aggressive, and more aggressive behavior

in a single graphical representation, we introduce the τ-η diagram here, as presented in Figure 4.11. In its Y-axis, the τ-η diagram has the dimensionless parameter η that separates tumor behavior and healthy behavior, and in the X-axis, the parameter τ_{act} that separates more aggressive tumor behavior. This diagram shows the three lineages' replicates in three distinct regions, successfully separating all the lineages in a single graphical representation.

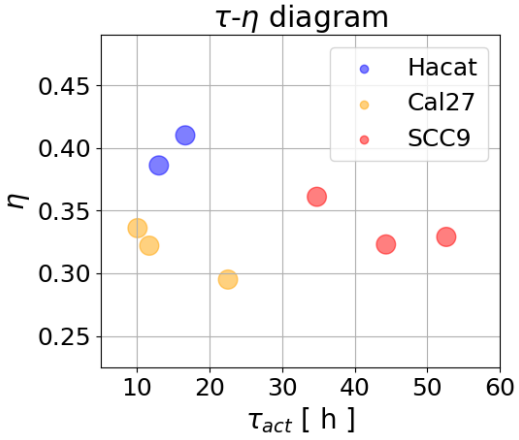

Fig. 4.11: τ-η diagram separates the studied lineages into three different groups. In the Y-axis, the dimensionless parameter η (Eq. 4.4), and in the X-axis, the activity characteristic decay time τ_{act} in hours.

As shown in Figure 4.10, the $\delta\eta$ values, i.e., η's time fluctuation, is relatively smaller for the healthy linage than the tumor ones. We made a variation of the τ-η diagram by introducing the fluctuation $\delta\eta$ as a third dimension. As shown in Figure 4.12, the two tumor lineages' $\delta\eta$ values are higher than the HaCat values, then, the point's size could be used as an auxiliary parameter to identify the tumoral behavior.

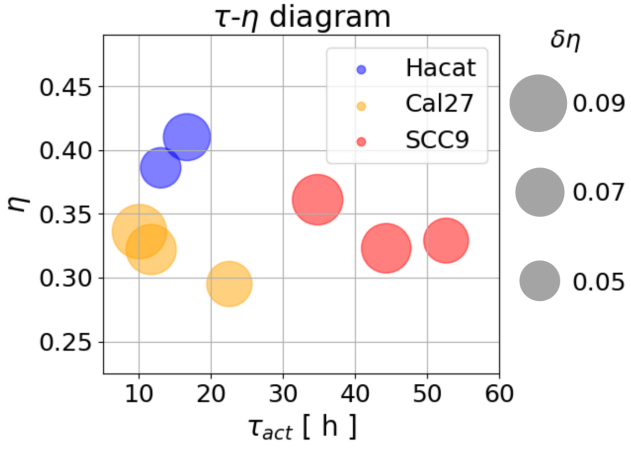

Fig. 4.12: τ-η diagram (Fig. 4.11) with point size defined relative to the $\delta\eta$ values.

Chapter 5

Conclusion, discussion, limitations and perspectives

In this work, we study the collective movement of confluent cell monolayers and how it changes during growth in a jamming experiment. We use image analysis to extract the velocity profile of three different oral human epithelial cell lineages: a healthy one (HaCat) with a high concentration of cell-cell junctions, a less aggressive tumor (CAL27) with a high concentration of cell-cell junctions, and a more aggressive tumor (SCC9) with a low concentration of cell-cell junctions [1, 2]. Starting with *in vitro* confluent cell monolayers, we observe these cell lineages under phase-contrast microscopy for 50h while the cells divide (Fig. 4.2). We treated the images adequately to the particle image velocimetry analysis, where the associated velocity fields are measured. We computed the monolayers' average velocity and average speed from these fields, based on *Vicsek et al. 1995* [11], *Petitjean et al. 2010* [22] and *Garcia et al. 2015* [21]. We extract the monolayers' spatial velocity-velocity correlation profiles from the associated velocity fluctuation fields, based on *Angelini et al. 2010* [23], *Garcia et al. 2015* [21]) and *Cavagna et al. 2018* [30].

We found that the average drift speed (Fig. 4.5), i.e., the module of the average vectorial velocity, of SCC9 is close to zero during the experiment, presenting the signature of a non-collective disordered system [7, 11, 12, 13, 26]. While HaCat and CAL27 present positive values of drift speed, showing the behavior of a collective system. These results corroborate that SCC9 is indeed a less collective lineage than CAL27, as believed by *Ramos et al. 2016* [1].

The three lineages present a significant decrease in activity, i.e., average speed, from the beginning to the end of the experiment (Fig. 4.6), showing the signature of a jamming solidification transition (JST, see section 2.3). High invasive tumors are known to unjam from healthy tissue [4]. Our result indicates that the tumor ability to unjam from the tissue is not coupled

to a JST insensibility since SCC9 also presents a significant decrease in activity. Nevertheless, the way the activity decreases for the more aggressive lineage is monotonic compared with the two other lineages. For HaCat and CAL27, we found well defined JST points followed for a fast decay in activity with $\bar{\tau}_{act} = (15 \pm 2)h$ and $\bar{\tau}_{act} = (15 \pm 6)h$, respectively (Fig. 4.6), averaging over replicates. For SCC9, we found no marked JST point and a $\bar{\tau}_{act} = (44 \pm 7)h$, averaging over replicates. Our results suggest a correlation between aggressiveness and how cells undergo a JST, where high values of τ_{act} could be a marker of more aggressive tumor behavior.

Vedula et al. 2012 [17] shows *in vitro* that the regulation of cell-cell junctions in a healthy epithelial monolayer is necessary to have collective ordered migration. We believe that the lower number of cell-cell junctions in SCC9 [1, 2], i.e., a hallmark of more aggressive behavior [4, 19, 20], hinders this lineage's ability to organize a long-range collective movement. In contrast, the high levels of cell-cell junctions in HaCat and CAL27 [1, 2] enable them to coordinate ordered motion, presenting positive values of drift speed (Fig. 4.5). We conclude that SCC9 drift speed close to zero (Fig. 4.5) results from the downregulation of cell-cell junctions, showing a connection between the microscopic hallmark of malignancy and the macroscopic behavior found here. Possibly, to be able to unjam from the healthy tissue, highly invasive tumor cells pay the price of losing long-range coordination in their movement, reinforcing the idea that malignancy is related to a non-collective behavior.

The lower number of cell-cell junctions can also explain the higher values of τ_{act} in SCC9. Before the JST point, is still possible for the cells to move, but in a confluent system, or the cells squeeze through each other or migrate collectively [23]. The higher concentration of cell-cell junctions in HaCat and CAL27 enable them to communicate with the neighbors and orchestrate collective migration reaching higher speeds than they would have if they were trying to squeeze through each other. Once the critical point of JST is reached, this *fast train* of cells stop abruptly, presenting the low values found for τ_{act}. In SCC9, the inability of large range order makes cells locally trapped since the beginning of the experiment. When density and the maturation of the tissue increase, the decrease in activity is more monotonic because there was no high-speed group of cells to stop in the first place.

The maturation of the cell-cell junctions is an unexplored ingredient in this work that could also play an essential role in SCC9 monotonic decay of activity. *Garcia et al. 2015* [21] shows *in vitro* that the maturation, i.e., increase in concentration or strength, of the cell-cell junctions would be the principal responsible for a jamming transition, and the increase in density plays a secondary hole in a monolayer growth. Here we did not measure the cell-cell junction concentrations nor how they change during the experiment, but we can argue

that a differentiated maturation process of SCC9 could explain the more monotonic decay of activity. If SCC9 cell-cell junctions do not maturate in time or have a rate of maturation significantly smaller than the ones of HaCat and CAL27, then HaCat and CAL27 would arrest by the increase in density and by maturation, while SCC9 would arrest only by the increase in density. To test that hypothesis, we need experiments specific to measure the evolution of the cell-cell junctions or in *silico* experiments like Szabó [12] and Gregoire's [13] models that allow the control of adhesion and density.

Petitjean et al. 2010 [22] and *Garcia et al. 2015* [21] use v_{rms} to monitor the jamming transition. We introduce here activity as an alternative parameter to monitor JST in monolayer. Both v_{rms} and *Activity* have the same behavior (Fig. 4.3), but activity has a more friendly formulation with a direct standard deviation in space ($\delta Activity$) that could be used to monitor the spatial fluctuation of the monolayer speed[1]. Using $\delta Activity$, we build the activity coefficient of variation Γ that tells how homogeneous is the distribution of speed, i.e., energy spent in movement ordered or not, along the monolayer and its temporal average $\bar{\Gamma}$. Averaging over replicates ($< . >_r$), we found for HaCat $< \bar{\Gamma} >_r = 0.43 \pm 0.05$, for CAL27 $< \bar{\Gamma} >_r = 0.62 \pm 0.06$, and for SCC9 $< \bar{\Gamma} >_r = 0.68 \pm 0.005$. $\bar{\Gamma}$'s higher values for CAL27 and SCC9 indicate that the tumor lineages have a less homogeneous distribution of energy along the monolayer. Explaining this result is left for further studies.

The few acquired data on cell density and cell sizes is a limitation of this work, especially in the speed-related results. As shown in Figure 4.2 (b), or in the videos of section 4.1, CAL27 has a smaller size than HaCat, while SCC9 has a bigger sizer than HaCat. If the activity graph in Figure 4.6, for example, were made in cell lengths instead of μm and normalized by HaCat size, then HaCat curves will be the same, but CAL27 and SCC9 would probably have higher and lower values of activity, respectively. Another limitation is the dimension of τ_{act} ($[time]$). Since we are using τ_{act} to compare different types of cells, i.e., different objects, τ_{act} should be an dimensionless quantity to avoid possible bias. That could be achieved by normalizing activity by the growth rate of each lineage. For that, we need better data on density versus time. These normalization issues will be addressed in the subsequent studies; nevertheless, since our main results depend on the curves' shape rather than on the curves' absolute value, we believe that our results will hold even after the normalization.

Concerning the fluctuation fields, there is more the one way to calculate the spatial extent of correlation in a monolayer from velocity autocorrelation functions (C_{vv}). *Angelini et al.*

[1] Nevertheless, using $(v_{rms})^2$ instead of v_{rms}, one would also get a straightforward standard deviation, but in units of velocity squared.

2010 [23] use C_{vv} minimum negative value to define the size of swirls in the monolayer. *Garcia et al. 2015* [21] fitted an exponential decay function in the positive part of the C_{vv} to extract the correlation length. Here we measured the monolayers' velocity correlation length (λ), as done by Garcia, and compared it to the integrated total degree of correlation (χ_{int}). χ_{int} is the integrated version proposed here for the total degree of correlation explained in *Cavagna et al. 2018* [30]. Our result (Fig. 4.9) indicates that λ and χ_{int} are equivalent quantities, and both can be used as measurements for the extent of correlation. Nevertheless, χ_{int} is easier to compute than λ. Moreover, λ can only be defined for the positive values of C_{vv} curves, while χ_{int} can be extended for the negative values. Negative values in C_{vv} curves are associated with swirls [23]. Based on our results, and in the fact that swirls are relatively common patterns in epithelial monolayers [17, 21, 23], we conclude that χ_{int} should be used instead of λ as the parameter to measure the extent of spatial velocity correlation in monolayers.

HaCat's correlation length has a bell-shaped curve in time (Fig. 4.9). At the beginning of the experiment, HaCat's correlation length is $\sim 70\mu m$, than reaches $\sim 135\mu m$ prior to the JST transition, dropping to $\sim 45\mu m$ in the end. CAL27 also presents a bell-shaped curve (Fig. 4.9) but is less accentuated. At the beginning of the experiment, CAL27's correlation length is $\sim 60\mu m$, than reaches $\sim 90\mu m$ prior to the JST transition, dropping to $\sim 30\mu m$ in the end. *Garcia et al. 2015* [21] also found bell-shaped correlation curves for epithelial monolayers. Contrarily, SCC9 presents a constant correlation length of $\sim 45\mu m$ or a monotonic decay from $\sim 90\mu m$ at the beginning of the experiment to $\sim 40\mu m$ at the end.

As explained before, the v_{drift} positive values of HaCat and CAL27 indicate that these two are collective lineages. A collective behavior should be associated with high correlation lengths values, while SCC9, having close to zero values of v_{drift}, should have lower correlation length values. However, our results show that both the tumor lineages have lower values of correlation when compared to HaCat. A probable reason for that is the different sizes of the cells. Since CAL27 is a smaller cell than HaCat, and SCC9 is bigger than HaCat, if the correlation length were normalized by the cell sizes, as explained before, CAL27 would have higher values, possibly closest to HaCat, while SCC9 values would be even smaller.

To eliminate this bias coming from the size of the cells, we divided the χ_{int} by the effective distance of zero correlation (ξ), creating the dimensionless quantity $\overline{C}^{+}_{vv}$, which we call hear *mean positive correlation* since $\overline{C}^{+}_{vv}$ has the same mathematical structure of the average of the positive part of the C_{vv} curve (sec. 3.4). We found that η values, i.e., the time average of $\overline{C}^{+}_{vv}$, averaged over replicates, for HaCat is $\bar{\eta} = 0.45 \pm 0.04$, for CAL27 is $\bar{\eta} = 0.32 \pm 0.02$, and for SCC9 is $\bar{\eta} = 0.34 \pm 0.02$. The time fluctuation of η ($\delta\eta$), averaged over replicates, for HaCat is

$\bar{\delta\eta} = 0.056 \pm 0.008$, for CAL27 is $\bar{\delta\eta} = 0.074 \pm 0.011$, and for SCC9 $\bar{\delta\eta} = 0.069 \pm 0.007$. The values found for η show that this dimensionless quantity $\overline{C}^{+}_{vv}$ (Fig. 4.10) presents lower values for SCC9, higher values for HaCat, but, surprisingly, it still presents lower values for CAL27. Also, $\delta\eta$'s values show that the $\overline{C}^{+}_{vv}$ curves of the tumor lineages fluctuate more in time than the ones of HaCat.

η's lower values for SCC9 prove that this lineage indeed does not have a long-range correlation, agreeing with its close to zero v_{drift} values. *Petitjean et al. 2010* [22] shows that fibroblast with close to zero cell-cell junctions also presents lower spatial velocity correlation values, indicating that the lower number of cell-cell junctions in SCC9 is responsible for the low correlation values. Nevertheless, we can not explain why CAL27, which is supposed to have a high concentration of cell-cell junctions, presents $\overline{C}^{+}_{vv}$ values similar to SCC9 instead of similar to HaCat. Explain CAL27's and SCC9's $\overline{C}^{+}_{vv}$ similarities and their difference to HaCat is an open question for subsequent studies. However, we used these differences, quantified into the η and $\delta\eta$ parameters, to separate the tumor cells' behavior from the healthy ones.

Finally, we introduce the τ-η diagram (Fig. 4.11) that separates the three lineages into three distinct regions of a single graphical representation. In the Y-axis of the τ-η diagram, we have the parameter η, separating the tumor lineages (with lower average positive correlation values) from the Healthy lineage (with higher values). In the X-axis, we have the parameter τ_{act} that separates the more aggressive tumor with monotonic decay of activity (high values of τ_{act}) from the healthy and less aggressive tumor with fast decay of activity (low values of τ_{act}). We also propose a variation of the τ-η diagram (Fig. 4.12) using the $\delta\eta$ values as a third dimension associated with the points' size. Points with larger sizes, i.e., larger time fluctuation in the average positive correlation, would be associated with tumoral behavior.

It is essential to state what we call here *more* or *less* aggressive behavior can strongly depend on the environmental condition where the cells are grown [4, 1, 2, 17]. We can not know if the collective behavior found here would be the same in other *in vitro* assays or inside the 3D environment of the human body. Nevertheless, we show that these three lineages with different degrees of aggressivity under the same experimental conditions can be separated using the τ-η diagram and with the indication that this separation is associated with one of the hallmarks of cancer: the low number of cell-cell junctions.

From the original phase-contrast movies' treatment to the final graphs, all data was acquired using an open-source semi-automatic pipeline that we developed specifically for this dissertation. The pipeline can be found on GitHub (link github) in a non-supported beta version. Since the pipeline is semi-automatic, a student or researcher using it, after a couple of

weeks of supervised training, can go from an untreated phase-contrast movie of a monolayer to the final graphs and the diagram presented here in less than three hours. This open pipeline, associated with the relatively simple phase-contrast experiments used here, enables our results to be replicated and studied by third parties even in developing countries.

This work indicates that the microscopic hallmark of cancer of the low number of cell-cell junctions is propagated to the macroscopic level in the cells' collective movement. Using current biological physics techniques, we measured the collective behaviors and computed them into single values parameters. Using these parameters we separated three cell lineages with different aggressiveness. We conclude that cells' collective movement can be understood as a macroscopic realization of a microscopic well-known cancer hallmark. We believe that the τ-η diagram is a proof of concept that collective movement physics-based cancer diagnosis is possible and encourage others to pursue similar approaches in the hope they will shed new lights on this troublesome disease called cancer.

5.1 Perspectives

A long shot perspective of applying the τ-η diagram would be in cancer recurrence[2] (Fig. 5.1). After the surgical removal of a tumor, there is a chance that some tumor cells remain. These cells could grow back and form a new tumor that usually is more aggressive than the previous one. In the τ-η diagnose, the oncologist would then extract tissue from the patient's former lesion periodically, execute the pipeline used here, and plot the cells' behavior in the τ-η diagram (Fig. 5.1 (a)). After several months of repeating this test, the doctor could see if the cells' behavior is changing from a healthy behavior to a more aggressive one (Fig. 5.1 (b)) or not.

There is a considerable gap to be filled between the work present here and a real clinical application as the one described in the previous paragraph. We built this pipeline based only on the behavior of three lineages of the same type of cell (epithelial) from a specific part of the body (mouth) of one type of animal (human). This pipeline and our results need to be tested over hundreds of different cell lineages and cells from primary patient tissue cultures. If our result hold and the τ-η diagram regions can separate the hundreds of samples' aggressive behavior, then we can start thinking in real clinical trials. At this point, we would have to compare our pipeline to current techniques, such as PCR, for example: is our pipeline cheaper or faster? Is it reliable? Can our pipeline deliver novel results or work as an additional diagnostic tool? We

[2] NCI page.

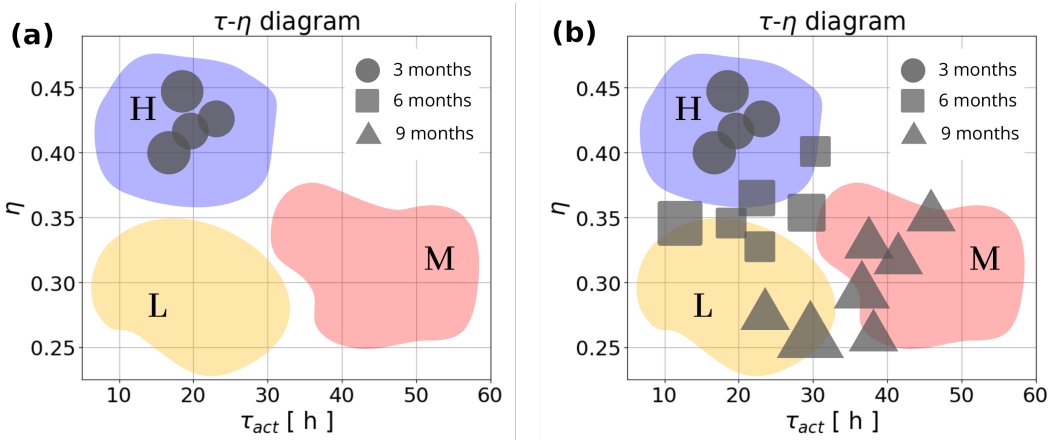

Fig. 5.1: Illustrative examples of application of the τ-η diagram as a diagnostic tool. **These are not real data.** (A) An example of a healthy diagnose. (B) τ-η diagram used to check the evolution of patient lesion. In this case cells are evolving to a more aggressive behavior.

do not know if our pipeline can give new answers that current techniques miss or if it works as a diagnostic tool. We can state that it seems more troublesome than a PCR diagnostic since it requires the extracted patient tissue to be grown in a monolayer inside an incubator for two days and two more days under a microscope equipped with a camera. Only after that, the cells can be analyzed. The good side is that the analyzing process takes less than three hours. The pipeline proposed here and the τ-η diagram must be thought of as a diagnostic prototype that needs yet to be tested, improved, and fully understood.

In the continuity of the work, we will make more robust density measurements, increase the statistical significance of each replicate by acquiring multiples FOV (Fig. 5.2), increase the number of replicates of each lineage to have a better statistical significance, and analyze other cell lineages and cell types, like fibroblast cells. We will use machine learning algorithms to automatically count the cells without the need for fluorescence microscopy, which will help to maintain the experimental assay simple.

To better understand the relation of the results found here with the cell-cell junctions expression of each lineage, we will use mechanistic assays with treated SCC9 cells to have an upper-regulation of cell-cell junctions and treated HaCat cells to have a down-regulation of cell-cell junctions. If the concentration of cell junctions is the principal factor for the distinct behaviors found here, the treated HaCat collective behavior should be similar to the wild-type SCC9 behavior and vice versa. Also, we will use simulations of the class of the Szabó model [12] to understand the relation of density and the strength of the cell adhesions to the collective motion. Algorithms of the Szabó model (like the simulation we did in video 17) enable direct control of particle density and the strength of particles attachment to each other.

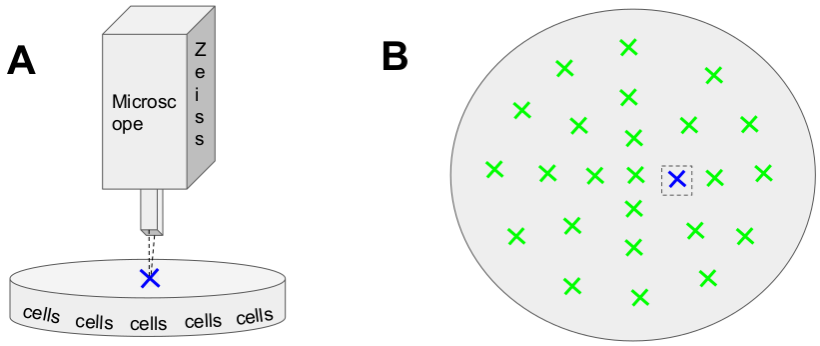

Fig. 5.2: Data acquisition sketch. (A) The current assay. It shows the microscope acquiring only one FOV. (B) Upper view of the cell plate showing a multiple FOV assay. The real assay will probably acquire ∼5 FOV of the same plate, and not that many as showed in the sketch.

Finally, some other pipeline applications' suggestions would be to do co-culture experiments of cells with different aggressivity. Such type of co-culture assays was done by *Castro et al. 2016* [35]. Also, we can seed the cells in different substrates. The effect of migration through different parts of the body can be studied in different substrates experiments. The works of our collaborators *Ramos & Lamers et al. 2019* [1] and *Matte & Lamers et al. 2019* [2] elucidated how substrates of different stiffness modulate single-cell migration of oral carcinoma. Repeated these studies for collective cell migration is then a natural step here.

Bibliography

[1] G. D. O. Ramos, L. Bernardi, I. Lauxen, M. S. A. Filho, A. R. Horwitz, and M. L. Lamers, PLoS ONE **11** (2016), 10.1371/journal.pone.0151338.

[2] B. F. Matte, A. Kumar, J. K. Placone, V. G. Zanella, M. D. Martins, A. J. Engler, and M. L. Lamers, "Matrix stiffness mechanically conditions EMT and migratory behavior of oral squamous cell carcinoma," (2019).

[3] F. M. White, R. A. Gatenby, and C. Fischbach, Cancer Research **79**, 2107 (2019).

[4] L. Oswald, S. Grosser, D. M. Smith, and J. A. Käs, "Jamming transitions in cancer," (2017).

[5] M. J. Bissell and W. C. Hines, "Why don't we get more cancer? A proposed role of the microenvironment in restraining cancer progression," (2011).

[6] O. Chaudhuri, S. T. Koshy, C. Branco Da Cunha, J. W. Shin, C. S. Verbeke, K. H. Allison, and D. J. Mooney, Nature Materials **13**, 970 (2014).

[7] C. P. Beatrici, R. M. De Almeida, and L. G. Brunnet, Physical Review E **95** (2017), 10.1103/PhysRevE.95.032402.

[8] F. Graner, B. Dollet, and P. Marmottant, Eur. Phys. J. E **25**, 349 (2008).

[9] E. Méhes and T. Vicsek, Integrative biology **6**, 831 (2014).

[10] T. B. Saw, A. Doostmohammadi, V. Nier, L. Kocgozlu, S. Thampi, Y. Toyama, P. Marcq, C. T. Lim, J. M. Yeomans, and B. Ladoux, Nature **544**, 212 (2017).

[11] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, Physical review letters **75**, 1226 (1995).

[12] B. Szabo, G. J. Szöllösi, B. Gönci, Z. Jurányi, D. Selmeczi, and T. Vicsek, Physical Review E **74**, 61908 (2006).

[13] G. Grégoire, H. Chaté, and Y. Tu, Physica D: Nonlinear Phenomena **181**, 157 (2003).

[14] S. Ramaswamy, Annual Review of Condensed Matter Physics **1**, 323 (2010), https://doi.org/10.1146/annurev-conmatphys-070909-104101 .

[15] R. Fürth, Zeitschrift für Physik **2**, 244 (1920).

[16] G. L. Thomas, I. Fortuna, G. C. Perrone, J. A. Glazier, J. M. Belmonte, and R. M. de Almeida, Physica A: Statistical Mechanics and its Applications **550**, 124493 (2020).

[17] S. R. K. Vedula, M. C. Leong, T. L. Lai, P. Hersen, A. J. Kabla, C. T. Lim, and B. Ladoux, Proceedings of the National Academy of Sciences **109**, 12974 (2012).

[18] S. R. K. Vedula, A. Ravasio, C. T. Lim, and B. Ladoux, Physiology **28**, 370 (2013).

[19] D. Hanahan and R. A. Weinberg, Cell **100**, 57 (2000).

[20] D. Hanahan and R. A. Weinberg, Cell **144**, 646 (2011).

[21] S. Garcia, E. Hannezo, J. Elgeti, J.-F. Joanny, P. Silberzan, and N. S. Gov, Proceedings of the National Academy of Sciences of the United States of America **112**, 15314 (2015).

[22] L. Petitjean, M. Reffay, E. Grasland-Mongrain, M. Poujade, B. Ladoux, A. Buguin, and P. Silberzan, Biophysical Journal **98**, 1790 (2010).

[23] T. E. Angelini, E. Hannezo, X. Trepat, J. J. Fredberg, and D. A. Weitz, Physical Review Letters **104**, 168104 (2010).

[24] E. Martz and M. S. Steinberg, Journal of Cellular Physiology **79**, 189 (1972).

[25] T. L. Liu, S. Upadhyayula, D. E. Milkie, V. S. ngh, K. Wang, I. A. Swinburne, K. R. Mosaliganti, Z. M. C. ns, T. W. Hiscock, J. Shea, A. Q. Kohrman, T. N. Medwig, D. Dambournet, R. Forster, B. Cunniff, Y. Ruan, H. Yashiro, S. Scholpp, E. M. Meyerowitz, D. H. yer, D. G. Drubin, B. L. Martin, D. Q. Matus, M. Koyama, S. G. Megason, T. Kirchhausen, and E. Betzig, Science **360** (2018), 10.1126/science.aaq1392.

[26] J. M. Belmonte, G. L. Thomas, L. G. Brunnet, R. M. C. de Almeida, and H. Chaté, Phys. Rev. Lett. **100**, 248702 (2008).

[27] N. Sepúlveda, L. Petitjean, O. Cochet, E. Grasland-Mongrain, P. Silberzan, and V. Hakim, PLoS computational biology **9**, e1002944 (2013).

[28] R. C. Brown and T. P. Davis, (2002), 10.1161/01.STR.0000016405.06729.83.

[29] D. Bi, J. H. Lopez, J. M. Schwarz, and M. L. Manning, Nature Physics **11**, 1074 (2015), arXiv:1409.0593 .

[30] A. Cavagna, I. Giardina, and T. S. Grigera, "The physics of flocking: Correlation as a compass from experiments to theory," (2018).

[31] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, Nature methods **9**, 676 (2012).

[32] A. Liberzon, D. Lasagna, M. Aubert, P. Bachant, Jakirkham, Ranleu, J. Borg, and C. Dallas, (2017), 10.5281/ZENODO.580340.

[33] V. Zickus and J. M. Taylor, Biomedical optics express **9**, 2418 (2018).

[34] E. Lång, A. Połeć, A. Lång, M. Valk, P. Blicher, A. D. Rowe, K. A. Tønseth, C. J. Jackson, T. P. Utheim, L. M. C. Janssen, J. Eriksson, and S. O. Bøe, Nature Communications **9**, 3665 (2018).

[35] M. Gamboa Castro, S. E. Leggett, and I. Y. Wong, Soft Matter **12**, 8327 (2016).

Chapter 6

Appendix

6.1 Velocity vectors

Velocity can be attributed to anything that moves, therefore, it can be measured. The basic formula is to divide the amount of displacement by the time interval it took to walk that distance. In mathematical terms we say that the velocity v is given by: $v = \Delta x/\Delta t$, where Δx is the spacial displacement (distance walked) and Δt is the total time to walk it. When there is no direction associated with this displacement, this quantity is called *speed*. When a direction of movement is also associated, then this physical quantity is called *velocity*. On physics, velocities are usually represented using an abstract mathematical object called *vector*. A vector is an arrow that has a magnitude (a size) and a direction (where the arrow points at). If this vector is representing velocity, the bigger the arrow, the faster the object will be moving. It is equivalent to say that, at the same time, the object with grater speed (i.e. velocity module, i.e. vector size) will have the greater displacement. This is exemplified in figure 6.1, where two subsequent photographs of the same cells are displayed.

In figure 6.1 (a), the cells on the left (with the dashed red line), have distinct velocity vectors: one bigger than the other. Added to that, their velocities are in opposite directions. The figure 6.1 (b) shows the final movement of these cells, after executing the displacement defined by its green velocity vectors. In monolayer image analyses, this velocity vectors are defined *a posteriori*, after the comparison of two subsequent images. In these case, the cell closest to the top has moved little from (a) to (b), while the one on the bottom has moved further away from the dashed red line, which is fixed on space. An image analyses algorithm (or a person with a ruler) will compute (attribute) a smaller velocity vector to the cell on the top and a bigger one to the cell on the bottom, as shown in the figure. The orientation of the velocity vectors is pointing to the direction of the movement.

About mean velocities: in the same figure 6.1 (a) we can see two pairs of cells with velocity

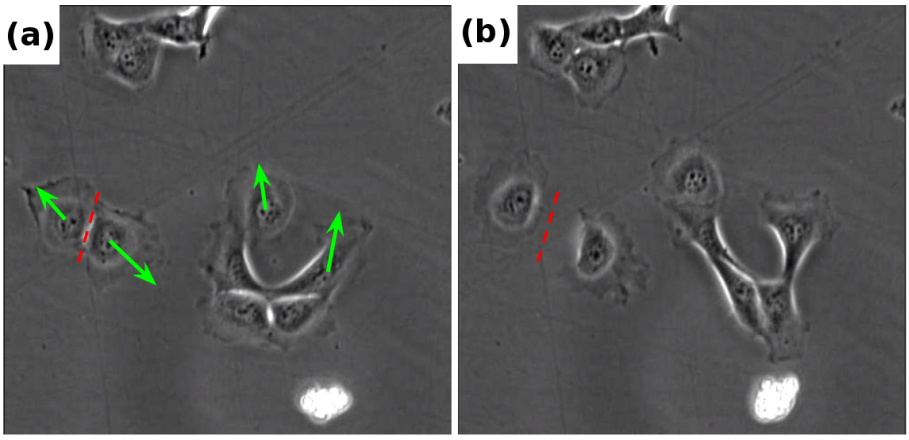

Fig. 6.1: Two subsequent photographs of the same cells: (a) before and (b) after some period of time. (a) Green vectors are the velocity associated with the cell movement from (a) to (b). The red dashed line is fixed on the grid (same position in (a) and (b)).

vectors depicted. Both of them have basically the same *average speed* because the amount of movement, regardless of the direction, is about the same. In other words, the average size of the pairs of vectors is basically the same. However, if we do a vectorial average, which accounts for the direction of movement, we see that the cells on the right have basically the same average velocity than the speed, while the cells on the left have a close to zero vectorial average, because the opposing vectors cancel each other. Imagine a situation in which two cells are firmly attached to each other, unable to detach, and have opposite velocity directions, it does not matter how much energy they spend trying to move: they will not move at all. This means there is a zero vectorial average. Meanwhile, two cells firmly attached to each other, but with velocity vectors parallel to each other (like the cells on the right), will be able to migrate having an average displacement (which is related to the average vectorial velocity) proportional to the amount of energy they spend in movement (which is proportional to the mean speed).

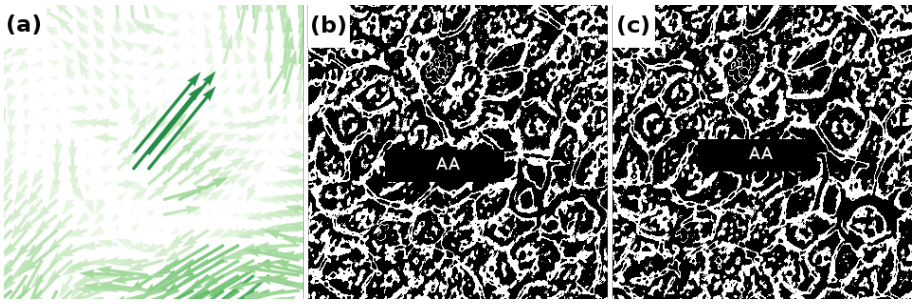

Fig. 6.2: Measured velocity field (a) of the two subsequent pair of images (b) and (c).

A *velocity field* is a higher order mathematical object that, instead of associating a velocity

vector to a single particle, it associates a velocity vector to every position in space, generally in the form of a regular lattice (Fig. 6.2). Velocity fields are usually applied to continuum materials, like fluids, were one can not distinguish the individual particles from the whole. Figure 6.2 (a) show the measured velocity field - using Particle Image Velocimetry (section 3.3.2) - associated with movement of the subsequent images (b) and (c). These are real cell monolayer images after the hysteresis treatment (section 3.2.2). In order to visualize how the velocity field of the image pair works, we overlapped a rectangular object with subscript **AA**. After that, we made a displacement orders of magnitude higher than the average movement of the cells in that region, diagonally up and to the right. The result are the big vectors pointing to the up right corner. The bigger the velocity vectors, the greater the monolayer speeds, which means greater energy spent in movement. Concerning the directions of the vectors: the more ordered they are, the higher is the collective migration velocity.

6.2 How to read the graphs of results

Here we explain the general format of figures 4.3, 4.4, 4.5, 4.6, 4.9 and 4.10 from Results. All these figures presents eight graphs plotted side by side in a common structure. As it can be seen in figure 6.3, there are three columns of graphs. The left column corresponds to data from two healthy lineage replicates **HaCat**; the middle column presents three replicates of the benignant tumorous lineage **Cal27**, and the right column presents three replicates of the malignant tumorous lineages **SCC9**. The colour code is a reminder: **blue**, **healthy**; **orange**, **benignant**; **red**, **malignant**. All eight graphs share the same Y-axis and X-axis units and scale. The lower left is a region used for extra information. Sometimes the Y-axis presents only units, in these cases, the plotted quantities are discriminated in the lower left extra information region.

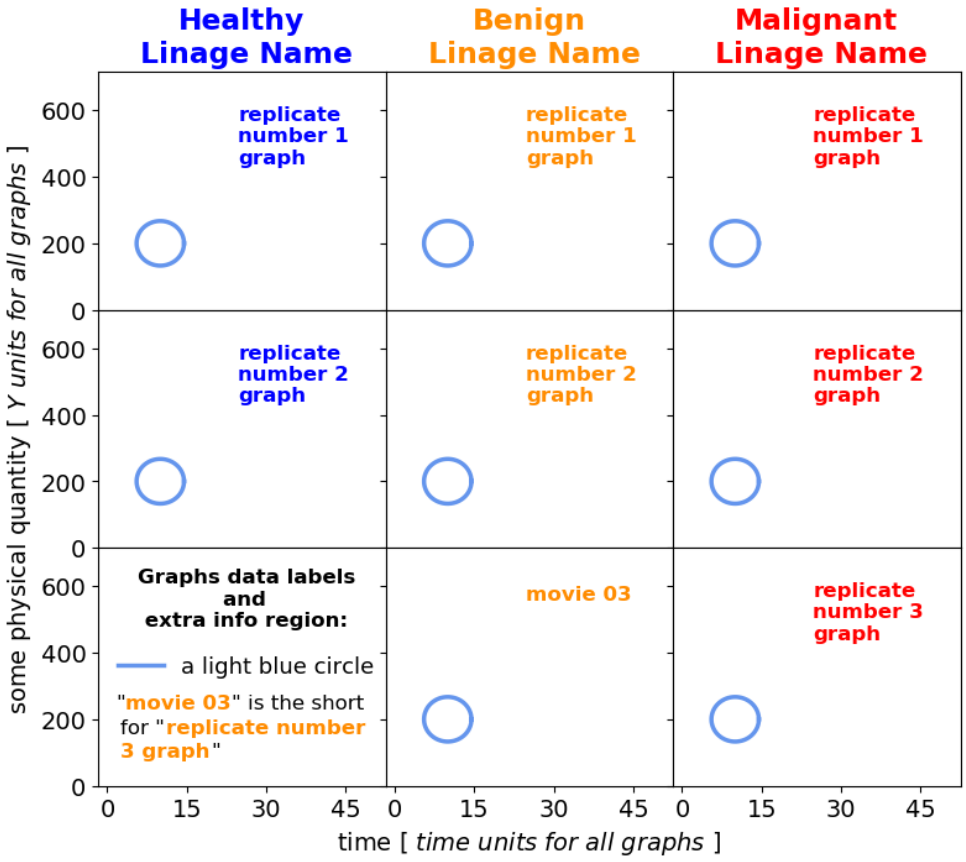

Fig. 6.3: An example of the general format of figures 4.3, 4.6, 4.9 and 4.10. Three columns: left, healthy lineage replicates (movies) graphs; middle, benignant tumor replicates; right, malignant tumorous replicates. All graphs share the same Y and X axis values and scale. The lower left region is used for extra information.

6.3 Velocity fields results

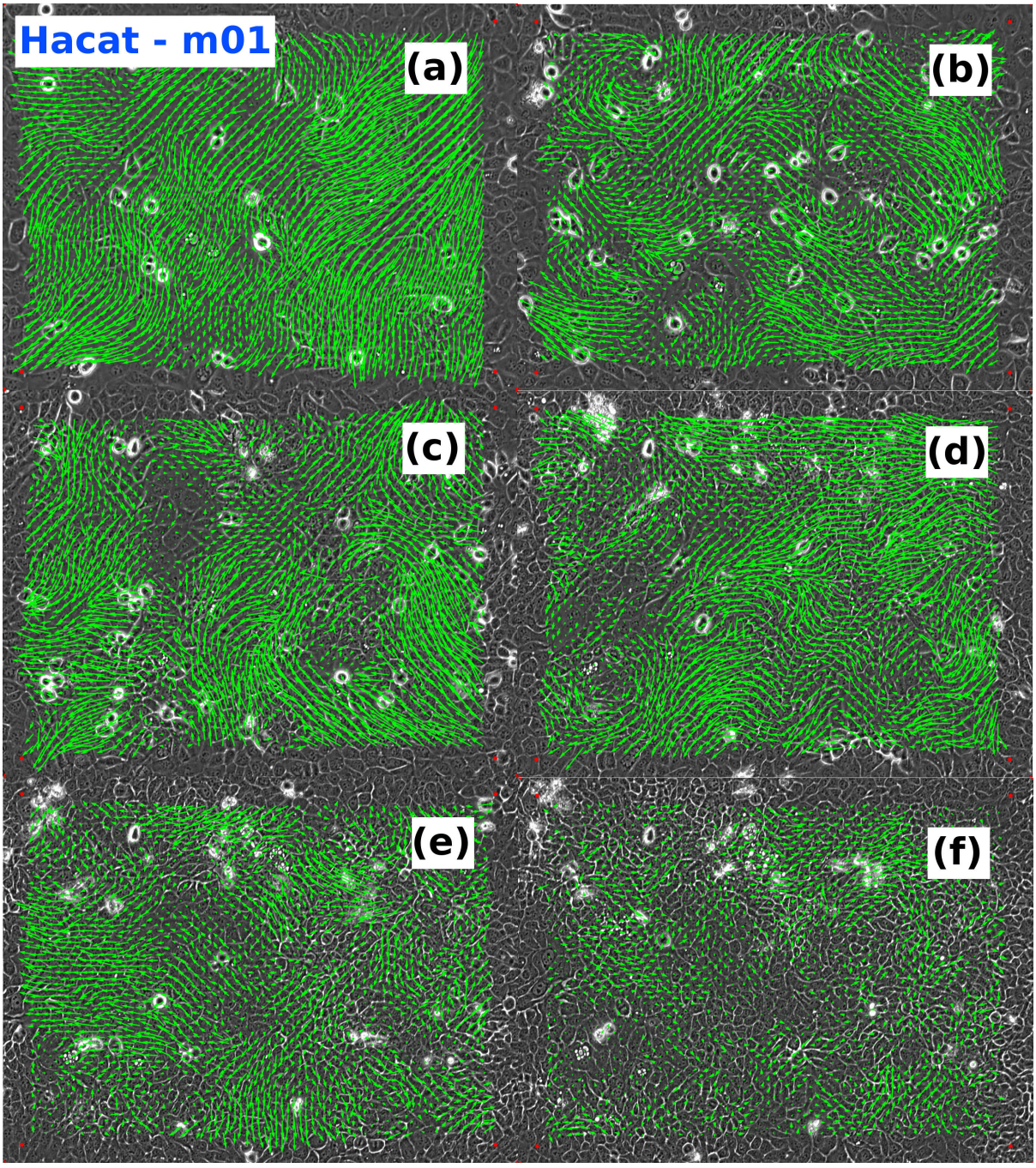

Fig. 6.4: Healthy lineage HaCat images overlapped by the velocity field vectors (in green) for times: (a) 6h, (b) 13h, (c) 20h, (d) 28h, (e) 35h and (f) 42h. The difference in magnitude (size) of the vectors from 6h to 42h is an example of the typical decrease in movement expected for epithelial cells undergoing a JST. All images have the same 2156 number of vectors. Only 50% of all 4312 vectors measured were plotted in the images.

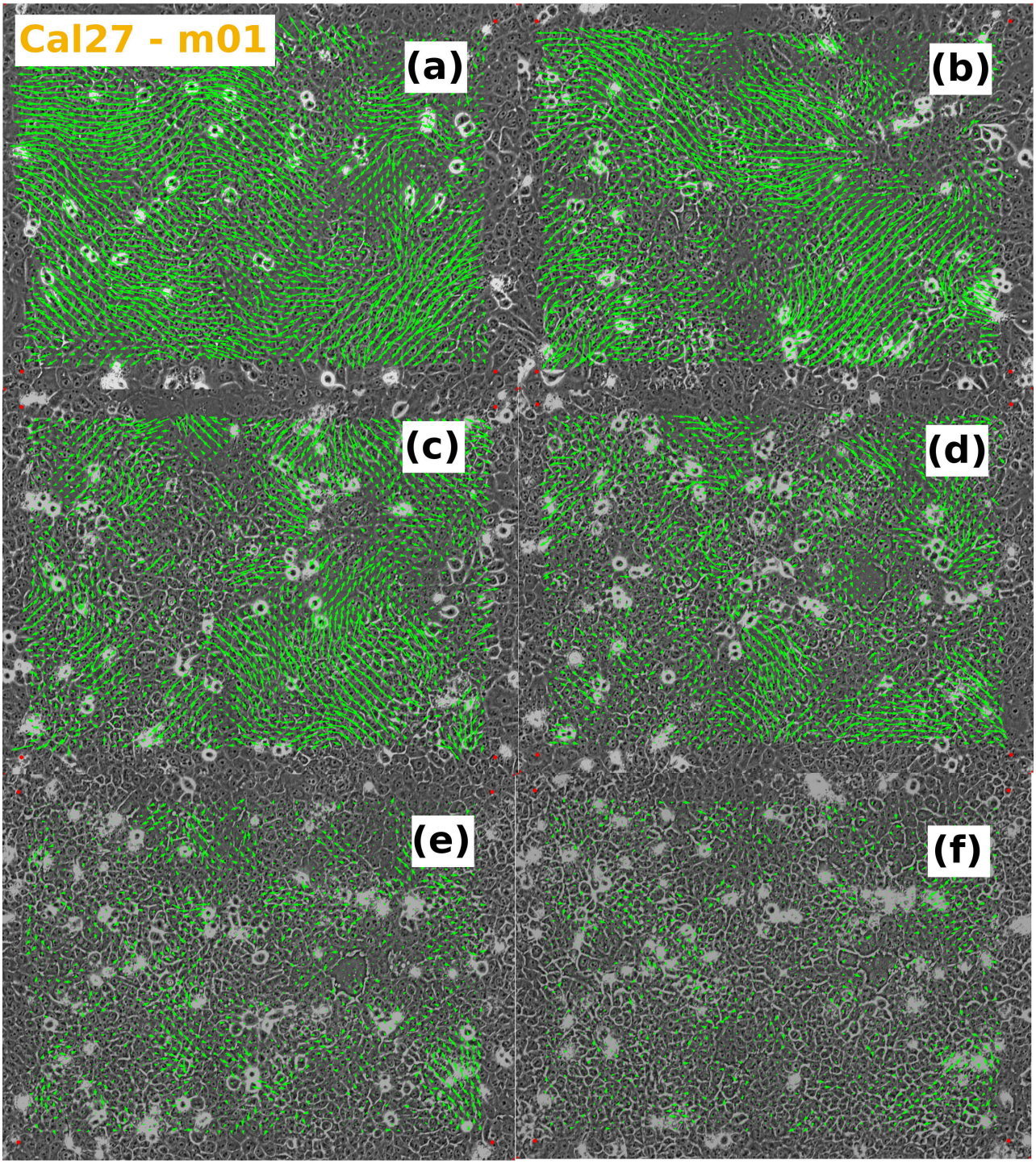

Fig. 6.5: The less aggressive tumor linage CAL27 images overlapped by the velocity fields vectors (in green) for times: (a) 8h, (b) 16h, (c) 22h, (d) 28h, (e) 35h and (f) 42h. The difference in magnitude (size) of the vectors from 8h to 42h is an example of the typical decrease in movement expected for epithelial cells undergoing a JST. All images have the same 2156 number of vectors. Only 50% of all 4312 vectors measured were plotted in the images.

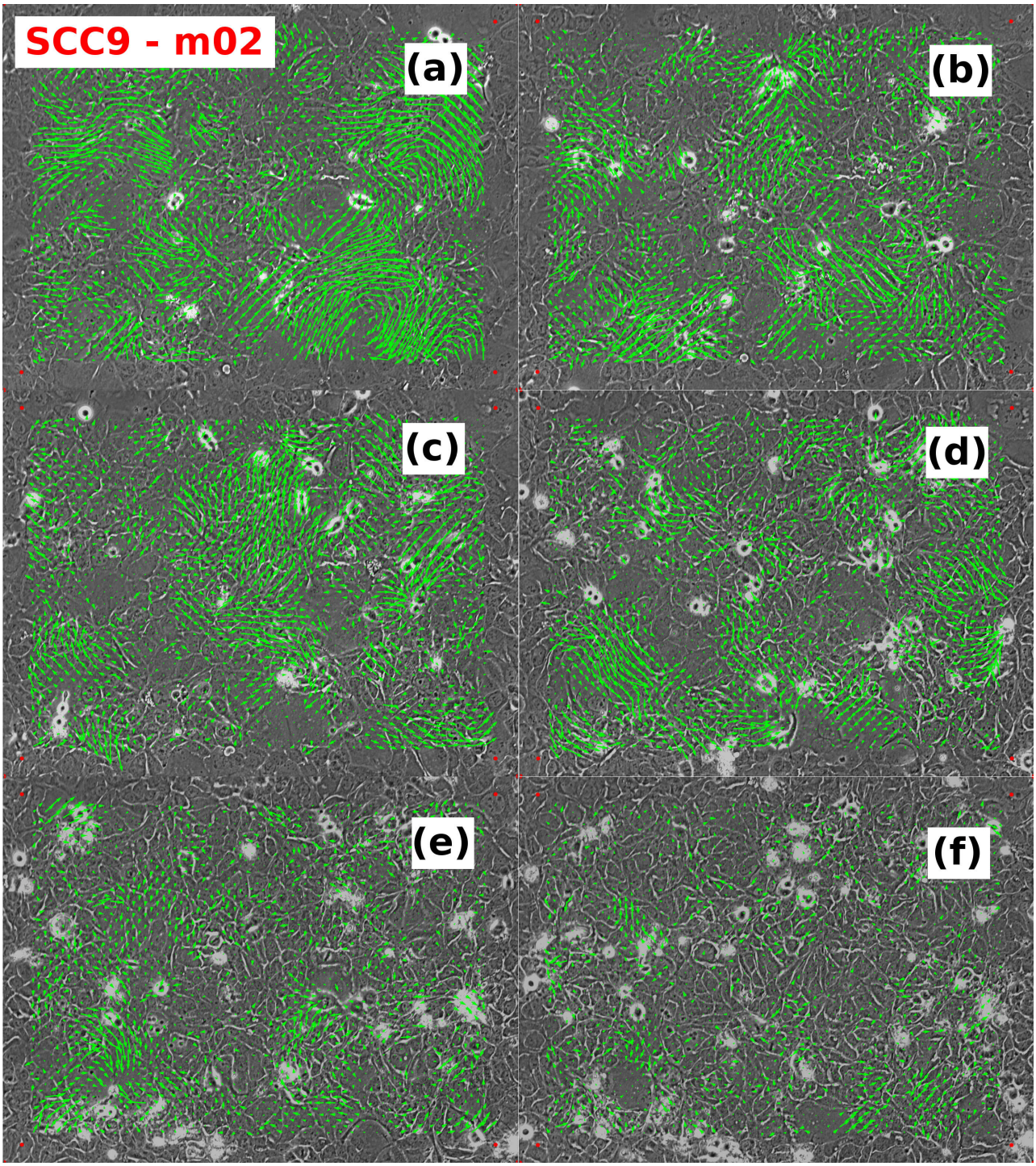

Fig. 6.6: The more aggressive tumor linage SCC9 images overlapped by the velocity fields vectors (in green) for times: (a) 8h, (b) 16h, (c) 20h, (d) 28h, (e) 35h and (f) 42h. The difference in magnitude (size) of the vectors from 8h to 42h is an example of the typical decrease in movement expected for epithelial cells undergoing a JST. All images have the same 2156 number of vectors. Only 50% of all 4312 vectors measured were plotted in the images.

6.4 Explaining fluctuations

Imagine a city bus with glass sealing and no seats with a dozen of people dancing waltz two by two around the entire bus. The proposition is to find out (define) the dancing behavior of the people by only analysing their velocities. Then you, standing over a bridge with a high precision camera, take several photos of the top of the bus when it is passing below the bridge. Using a Particle Image Velocimetry software, you extract the velocity field of the people in the bus. The bus being in a constant but fast speed of 80km/h results in a velocity field with well aligned set of vectors, with almost the same magnitude and directions, similar to the one from Figure 6.7 (a). Judging by this velocity field, the conclusion would be that all the people are correlated, being impossible to identify the *dancing waltz* behavior of them. Nevertheless, by subtracting the drift velocity from the vector field, we get an fluctuation field, as the one in Figure 6.7 (b), that revels another story. The fluctuation field shows the swirls coming from the spin of the dancing couples. This is equivalent to be taking photographs standing over the roof of the bus (and therefore having its same velocity), instead of being on the bridge.

From the fluctuation field is them possible to extract physical information about the waltz behavior of the system. For example, smaller swirls sizes (smaller correlation length), in this case, could be associated with couples dancing cheek to cheek, while bigger swirls (bigger correlation length) could be related with people dancing three by three (in a big circle), instead of two by two. Change the bus to Petri dish, the people dancing to cells migrating and you get a similar problem to the one presented in this work.

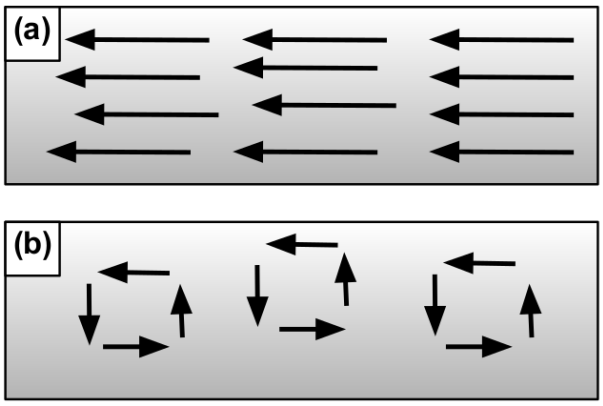

Fig. 6.7: Illustrative figure explaining the difference in behavior that emerges (in a fluctuation field) when the drift bias is eliminated (from the velocity field). (a) Velocity field with a major drift to the left. (b) The resulting fluctuation field after subtracting the drift bias showing the underling correlated behavior of the system.

6.5 Effective distance of zero correlation

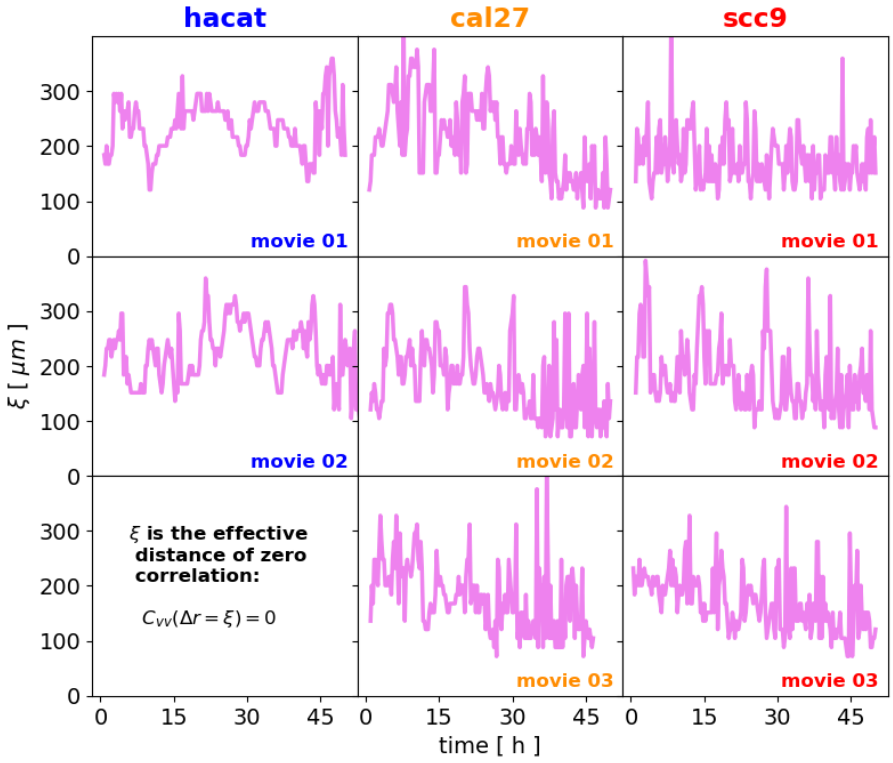

Fig. 6.8: Effective distance of zero correlation (ξ) in μm plotted versus time in hours for each lineage and its replicates. Time zero is when cells were taken from the incubator to the microscope. The graph format is the same as 6.3.