University of Warwick

Erasmus Mundus MSc in Complex Systems Science

M1 project report

Collective behaviour and stigmergy in populations of cancer cells

Author:

Jacopo Credi

Supervisors:

Prof. Jean-Baptiste Cazier
Dr. Sabine Hauert
Dr. Anne Straube

June 18, 2015

Abstract

Investigating and capturing the emergence of collective phenomena in cancer cell migration can advance our understanding of the process of tissue invasion, which is one of the first steps leading to the formation of metastases, or secondary tumours. By reconstructing the trajectories of lung cancer cells populations from microscopy image sequences, we were able to analyse their collective two-dimensional dynamics and measure the system spatial correlation function in different density conditions. This revealed that cancer cells, similarly to other recently studied biological systems, can exhibit a form of collective dynamics without global order. However, the observed density dependence of the correlation function differed completely from the theoretical predictions of standard models of moving particles with mechanisms of local alignment. We propose an explanation for this unexpected finding, supported by an analysis of the role of density in the ability of cells to communicate through the micro-environment (*stigmergy*), which revealed the emergence of a network-like structure of trails when the system density was sufficiently low.

1 Introduction

Increasing experimental and theoretical evidence suggests that malignant tumours can exhibit a range of collective patterns similar to evolved adaptive behaviour found in other biological systems, including collective decision making and collective exploration of the micro-environment [1]. A strongly multidisciplinary approach is required to cope with such a complex and self-organising biosystem, composed of individual mutated cells interacting by some local and stochastic mechanism and giving rise to a seemingly emergent *collective intelligence* [2]. The long-term process of cancer modelling, driven by the massive amount of data produced in molecular and cellular biology experiments, is increasingly regarded as capable of providing valuable qualitative insight into the evolution of this disease as well as predicting its quantitative behaviour. This would ultimately translate into new experiment design guidelines and eventually into innovative therapeutic applications [3, 4, 5].

The single most lethal aspect of cancer, responsible for about 90% of cancer-related deaths worldwide [2] is the formation and growth of secondary tumours, also known as metastases. Metastasis formation is an extraordinarily complex process in which a sequential series of steps has been identified, starting with the separation of cells (isolated

or in groups) from the primary tumour. These cells then invade the surrounding tissues, intravasate or enter the lymphatic system, arrest in a distant target location, extravasate and then survive and proliferate in a new microenvironment, while avoiding apoptosis or anoikis and immune system response [6, 7, 8].

In recent years it has been proposed that collective cell migration could be the main mode of tissue invasion in a wide range of malignant tumours, and many advantages residing in such collective invasion modes over the dissemination of individual cells have been identified [7, 8]. However, little is known on the mechanisms triggering such collective patterns, and although there are few therapies specifically designed to target the motility of cancer cells [9], the possibility of specifically targeting their ability to migrate *collectively* remains unexplored.

The level of complexity in cancerous bio-systems is further increased by heterogeneity in the mutational profiles of cells, which reflects in different morphologies, growth, mobility, adhesiveness and even mutability within the same tumour mass. Some mutations are currently known to exist in specific systems of cells, with various effects [10, 11]. Moreover, both examples competitive and cooperative behaviour, through *commensalism* or *mutualism*, have been observed in genetically heterogeneous tumours, with different effects on the aggressiveness of the colony [12]. Investigating and ultimately modelling the emergence of collective properties in the interactive dynamics of heterogeneous cancer cells are extremely challenging goals, which could shed light on the coupled physical and biological processes leading to the evolution of different metastatic potential in different subpopulations. Ultimately, this would allow us to predict the overall behaviour of a colony under various scenarios of mutational and environmental changes.

In this project image sequences of lung cancer cells *in vitro* were processed into quantitative datasets (Section 2) and analysed (Section 3), in order to identify and quantify collective patterns in the cells' motion (Section 4) with the help of concepts derived from statistical mechanics. A heterogeneous system of cancer cells was also processed into a dataset and a simple classification algorithm was implemented to distinguish between trajectories of cells with different mutational profiles.

2 Methods

Available data

The available data consist of a set of 2D time-lapse microscopy image sequences of PC9 non-small lung cancer cells, incubated at 37^{o}C in 5% Co2/Air in a humidified chamber. Four single-population sequences, denoted by SP1, SP2, SP3, SP4, show stained cancer cells appearing as white objects on a dark background (Fig. 1). In sequence HP2, instead, a heterogeneous population was stained with different fluorescent dyes according to the cells' mutational profiles, and appear as green and red objects on a dark background (Supplementary Material Fig. S1). The main features of the analysed data, including number of frames, total time and magnification, are summarised in Table S1.

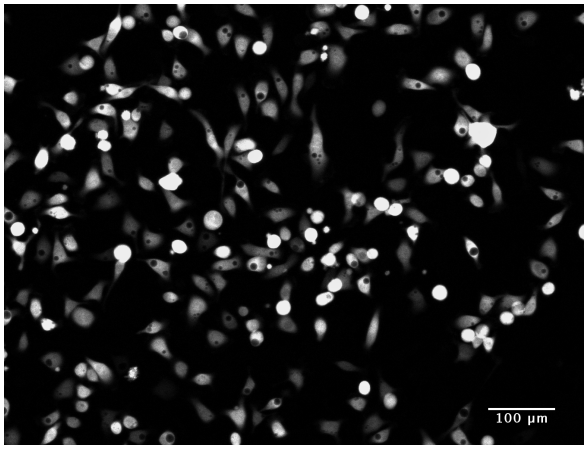


Figure 1: Snapshot from time-lapse sequence SP4. Background removed and contrast enhanced. Scale bar 100μm.

Cell number and size estimation

In order to estimate cell size and density in the systems under investigation, images were first converted into binary using Li's *Minimum Cross Entropy* thresholding method (Li et al. [13]), which is included in the built-in Auto Threshold methods in Fiji [14]. This algorithm works by iteratively finding the threshold value which minimises the cross entropy of the original image and its corresponding segmented version, and was observed to produce the best output among the available thresholding methods.

Next, a *watershed segmentation* algorithm (also available in Fiji) was applied to automatically separate touching objects. Watershed segmentation works by first computing the Euclidean distance

map of a binary image, i.e. by replacing white pixels (corresponding to an object) with grey pixels whose intensity is proportional to their distance from the nearest black pixel (corresponding to the background). Then, the centres of the resulting grey objects, called the ultimate eroded points, are expanded until the edge of the object is reached or they touch a neighbour, and in the latter case a watershed line is drawn in the meeting point.

Finally, the segmented objects were automatically counted using the *Analyse Particles* function in Fiji (see for example Fig. S2), which also measures the area of all detected objects and their circularity. To a first approximation, cells can be considered as spherical objects, with an estimated radius of $r_c = (11.3 \pm 1.2)\ \mu m$. The initial and final density of cells in sequences SP1, SP2, SP3, SP4 are summarised in Fig. 2, whereas the full density time-series of all sequences is reported in the SM.

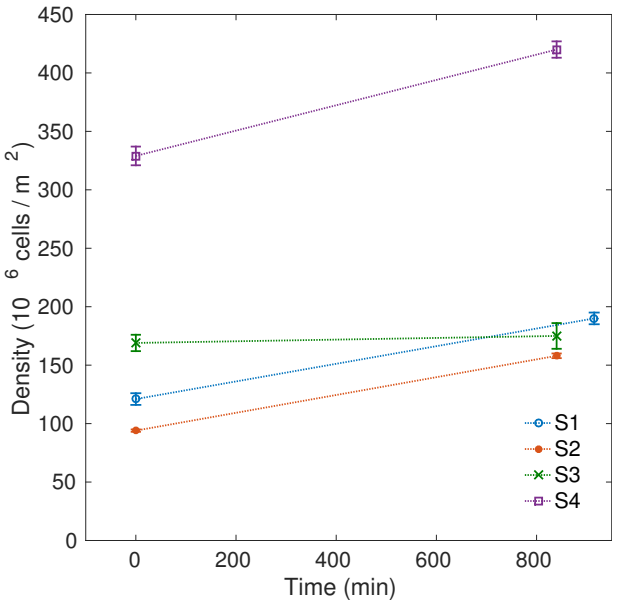


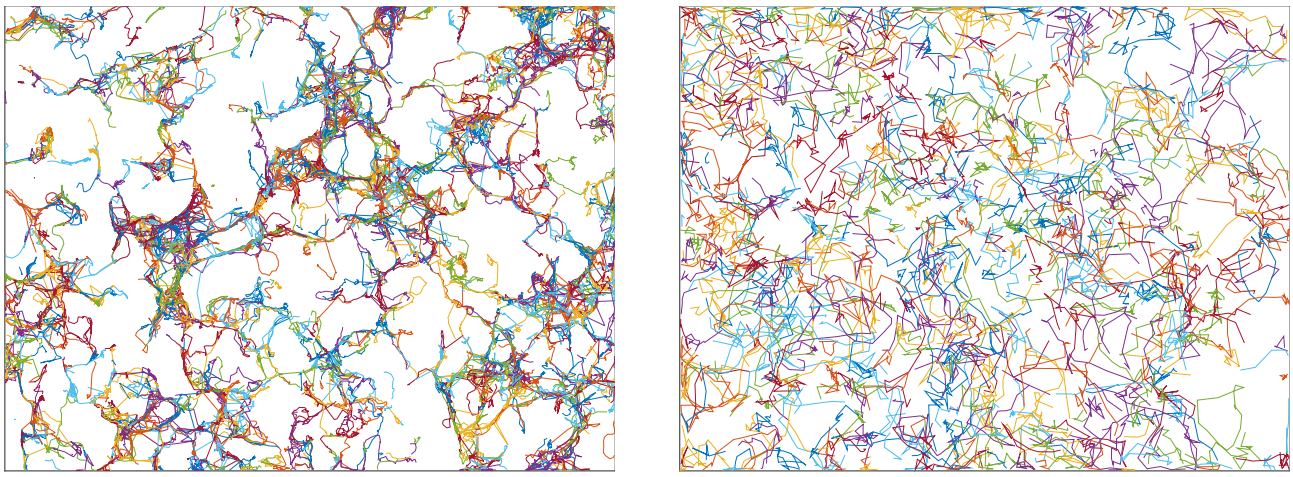
Figure 2: Initial and final cell density in single-population sequences.

Feature point tracking

In order to analyse and model the collective dynamics of the cells under investigation, their trajectories have to be reconstructed from the microscopy image data. This process is called *single-particle tracking* or *feature-point tracking*, and it is usually composed of two independent steps. First, the target object (in this case, the cell) has to be detected in the image and its precise location has to be determined (*detection phase*). Then, the different locations have to be linked in time (*linking phase*). If multiple objects are detected in subsequent frames, the linking process is nontrivial and corresponds to solving an *assignment problem*, with a cost function depending on some features of the object under investigation. Commonly used features in this combinatorial optimisation problem are the degree of overlap between two objects in consecutive frames, their relative displacement, and the similarity of their size.

After testing several 2D tracking software with our image sequences and observing their output, we decided to use the Particle Tracker tool included in the MOSAIC plugin suite for Fiji (based on a work by Sbalzarini *et al.* [15]). After a preliminary image filtering phase, this algorithm computes a first estimate of the feature point locations by finding local intensity maxima: all pixels in the upper i-th percentile (where i can be specified by the user) of intensity in each frame are considered as candidates for object locations. These candidates are then accepted if no other pixel within a distance r (where r is also an input parameter) is brighter than the candidate point. This first guess is then refined by iteratively calculating and minimising the *intensity moment* of order 0 within a distance r of the candidate location (intensity centroid estimation). Possible spurious objects are then identified by using a classification algorithm which assigns to each particle a score, based on the intensity moments of order 0 and 2, and discards particles with score lower than a user-provided threshold value θ.

In the linking phase, the algorithm assigns a cost to each link based on the relative displacement of the two linked objects and on the difference of their intensity moments. The relative weight of the object dynamics and features can be set by the user, and the cost is set to infinity if the displacement is larger than a certain threshold value L, also user-provided. The algorithm then finds an overall optimal association between the object locations by minimising the total cost of all links, using either a *greedy* optimiser, or the *Hungarian optimisation* method. This latter algorithm was observed to considerably improve the linking accuracy over the greedy method, and was therefore chosen in this work. Several sets of values were tested for the MOSAIC Cell Tracker input parameters. The chosen set of values is shown in Table S3, and the resulting trajectories for sequences SP2 and SP4



(a) SP2 **(b)** SP4

Figure 3: Trajectories reconstructed by tracking cells in single-population sequences SP2 **(a)** and SP4 **(b)**, discarding trajectories with 5 time-points or less.

are are respectively plotted in Fig. 3a and 3b (see the SM for SP1 and SP3).

Overall, this tracking procedure produced remarkably good results, given the low frame rate of available image sequences (from a minimum of 3 to a maximum of 6 frames per hour, see SM Table S1). However, in all datasets the observed number of time points per trajectory exhibited a large variance, and a short[1] trajectory is usually a sign of poor tracking accuracy. This is not a major problem as long as the analysis of the cells' dynamics is restricted to instantaneous or local trajectory statistics. Nevertheless, extremely short trajectories (i.e. with 5 time points or fewer) were discarded and not considered in the following analysis of dynamics, except for the determination of nearest neighbours. In other words, the detection output of the algorithm is assumed to be optimal, whereas the linking output is rejected when trajectories are shorter than 5 time-points.

The tracking algorithm described above is designed for grey-level (8-bit) images, thus being unable to distinguish between cells with different staining. In order to automatically classify trajectories extracted from H2 according to their mutational profile, a simple classification algorithm based on the analysis of RGB spectrum of the original images was implemented in MATLAB (see Algorithm 1 in the Supplementary Material for the pseudo-code). The set of trajectories obtained from sequence H2 after automatic tracking and classification is shown in Fig. 4.

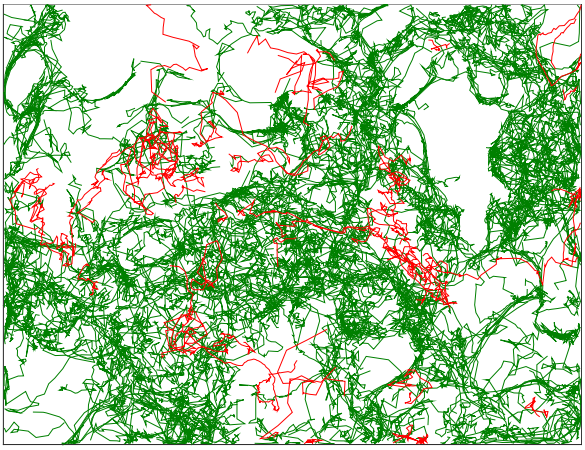


Figure 4: Trajectory data set after tracking and classification (Seq. H2).

3 Analysis of dynamics

Following the procedures described in the previous section, datasets of 1718 and 620 trajectories were respectively obtained from sequences SP2 and SP4. The analysis of dynamics and collective behaviour is focussed on these two datasets, as they both correspond to cell cultures observed for a period of 14 h, but with very different densities. The average observed density in sequence SP4 is over 3 times higher than that of SP2, thus allowing to consider negligible the density increase over time due to cell duplication, when compared to the relative difference in the two datasets.

[1]The term *short*, in this context, refers to the number of time points in the trajectory, and not to the total covered distance.

Preliminary analysis

Cell migration appears to be highly stochastic, due to its dependence on complex biophysical mechanisms, tightly coupled with environmental chemical (e.g. chemotaxis) and physical phenomena (e.g. exchange of momentum and shear stresses) [16]. This leads to considering the reconstructed trajectories as realisations of a stochastic process. In this framework, the motion would be completely characterised by determining the conditional probability density function $f(\boldsymbol{x} \mid \boldsymbol{x}_0, \delta t)$, also called transition density of the process, quantifying the probability of finding a cell at position $\boldsymbol{x}$ after a time δt has passed from its previous observation at position $\boldsymbol{x}_0$.

The standard method for single-particle trajectory analysis is based on the calculation of the second moment of displacement or mean square displacement (MSD):

$$\mu_2(\delta t) = \langle \|\boldsymbol{x}(\delta t) - \boldsymbol{x}(0)\|_2^2 \rangle_M \ , \qquad (1)$$

where the average $\langle \cdot \rangle_M$ is taken over an ensemble of M independent realisations of the same process. The time-dependence of the MSD can be analytically derived for stochastic processes whose transition density is known (e.g. normal diffusion or directed motion), thus allowing to identify the type of motion by comparing experimental and theoretical curves (see for example Saxton [17]). In fact, given a dataset of M trajectories observed at discrete time steps $\Delta n = 1, \ldots, N_j$, where N_j is the (finite) length of trajectory j, an experimental estimate of the MSD time-dependence can be obtained by computing the following time-average [18, 19]:

$$\mu_{2(j)}(\Delta n) = \frac{1}{N_j - \Delta n} \sum_{n=1}^{N_j - \Delta n} \|\boldsymbol{x}(n+\Delta n) - \boldsymbol{x}(n)\|_2^2 \ , \qquad (2)$$

for each trajectory j. This corresponds to calculating the mean of a set of non-independent random variables, whose statistical uncertainty must be corrected accordingly. Assuming that all M trajectories are samples coming from the same generating process, these values can then be averaged to obtain an experimental estimate of the MSD curve for the process.

This approach was attempted with poor success, as the MSD curves obtained from the reconstructed datasets (see Fig. S5) were not particularly informative about the cells' motion. This can be in part explained by the limited length of the original image sequences and the low frame rate, both leading to high experimental uncertainty. More importantly, however, two implicit assumptions are made when using equation (2), namely that the generating stochastic process is *stationary* and *ergodic*. Such assumptions can not be considered valid in this case, as cell density is constantly increasing and micro-environmental conditions are likely to being modified by the cells themselves over time. Therefore, this and other standard analytical methods based on global trajectory statistics can not be applied in this context.

Local trajectory statistics are in this case more informative. For example, observing the distribution of the turning angle between consecutive observation of cells' velocity revealed that the motion can not be modelled as a Markov process. In fact, the probability of observing a cell moving with a similar directionality in two consecutive frames is statistically much higher than that of observing extreme turns (see for example Fig. 5, corresponding to sequence SP2). This effect is damped, as expected, when the frame rate decreases, but is still statistically relevant even for the lowest frame rate sequences (SP3 and SP4).

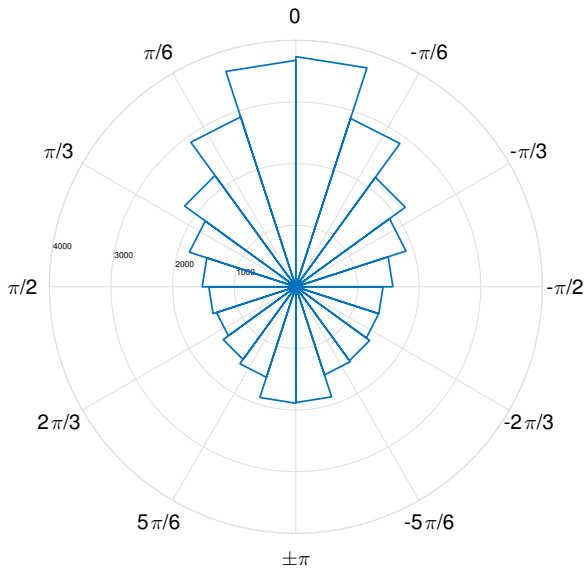


Figure 5: Distribution of observed cell turning angles, measured in rad, between two consecutive frames (sequence SP2).

Contact inhibition of locomotion

One of the key features of cancer cells, strongly correlated with tumour invasiveness, is the partial or total loss of contact inhibition of locomotion

(CIL), which is the ability of a healthy cell to avoid collision with a nearby cell [20, 21]. This phenomenon was investigated by measuring the instantaneous acceleration of each cell i as

$$\vec{a}_i(t) = \vec{v}_i(t) - \vec{v}_i(t-1) \tag{3}$$

and then projecting all these values onto the direction of the nearest cell in frame t. By doing this, we obtained a set of vector components $a_{\mathrm{nc}}(i,t)$ containing information on how each cell i is influenced by its nearest neighbour, denoted by the subsctipt *nc*. A positive value here means that cell i is moving away from its neighbour. From these values, the following quantity was computed:

$$F(r) = \frac{\sum_t^N \sum_i^M a_{\mathrm{nc}}(i,t)\ \delta\big(r - r_{\mathrm{nc}}(i,t)\big)}{\sum_t^T \sum_i^M \delta\big(r - r_{\mathrm{nc}}(i,t)\big)} \tag{4}$$

where

$$\delta\big(r - r_{\mathrm{nc}}(i,t)\big) = \begin{cases} 1 & \text{if } r < r_{\mathrm{nc}}(i,t) < r + dr \\ 0 & \text{otherwise} \end{cases},$$

and $r_{\mathrm{nc}}(i,t)$ is the distance between cell i and its nearest neighbour in frame t, whereas dr is the space binning factor. In Fig. 6, this quantity is plotted for sequences SP2 and SP4 against r/r_c, where r_c is the previously measured average cell radius, also used as binning factor.

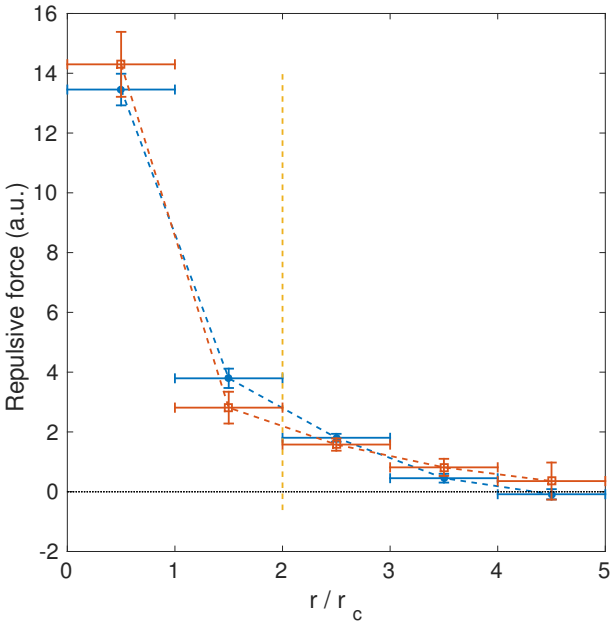


Figure 6: Repulsive force between cells, calculated using Equation 4, for sequence SP2 (blue dots) and SP4 (red squares). The vertical dashed line marks estimated contact distance.

The interpretation of this plot is not straightforward, as the net force producing motion is the sum of a large number of phenomenologically different components. However, a remarkable fraction (15% and 10%, respectively) of the 34784 and 7065 data points used to produce the plots lies below the estimated contact distance $2r_c$. This clearly suggests that contact inhibition of locomotion has partially been lost by these cells, although a positive (i.e. repulsive) force is observed at short range, which may be due to a partial conservation of CIL or to purely physical (e.g. elastic collision) effects.

Correlation function

For the aims of the project, however, the most interesting statistics are those revealing the existence of collective patterns in the cells' behaviour. The standard method to characterise the emergence of collective phenomena is usually based on the identification of an order parameter able to distinguish between ordered and disordered phases in the system. The concept of emergent collective order is in fact commonly identified as the hallmark of collective behaviour, as it is observed in a variety of biological systems over a huge span of spatial and temporal scales, from the formation of bird flocks and fish schools to the aggregation of bacteria colonies moving in an ordered and synchronised fashion. However, it has been argued that even seemingly disordered systems can exhibit important collective properties, and that strong spatial correlation, rather than order, should perhaps be considered as the true hallmark of collective behaviour. Attanasi *et al.* [22] recently studied collective patterns in swarms of midges, showing that a strong spatial correlation allows information to propagate rapidly in the swarm, despite the lack of collective order, thus enabling it to quickly react to external perturbations.

In order to estimate the spatial degree of correlation in our swarms of cells, we applied the same statistical methods used by Attanasi *et al.* to our trajectory datasets. First, it is convenient to transform the measured cell velocities into dimensionless quantities, as this allows to easily compare experimental data with numerical simulations. This can be done by defining the following vectors:

$$\vec{\varphi}_i(t) = \frac{\vec{v}_i(t)}{\sqrt{\frac{1}{M}\sum_k \vec{v}_k(t)\cdot\vec{v}_k(t)}}\ . \tag{5}$$

The spatial correlation function can then be

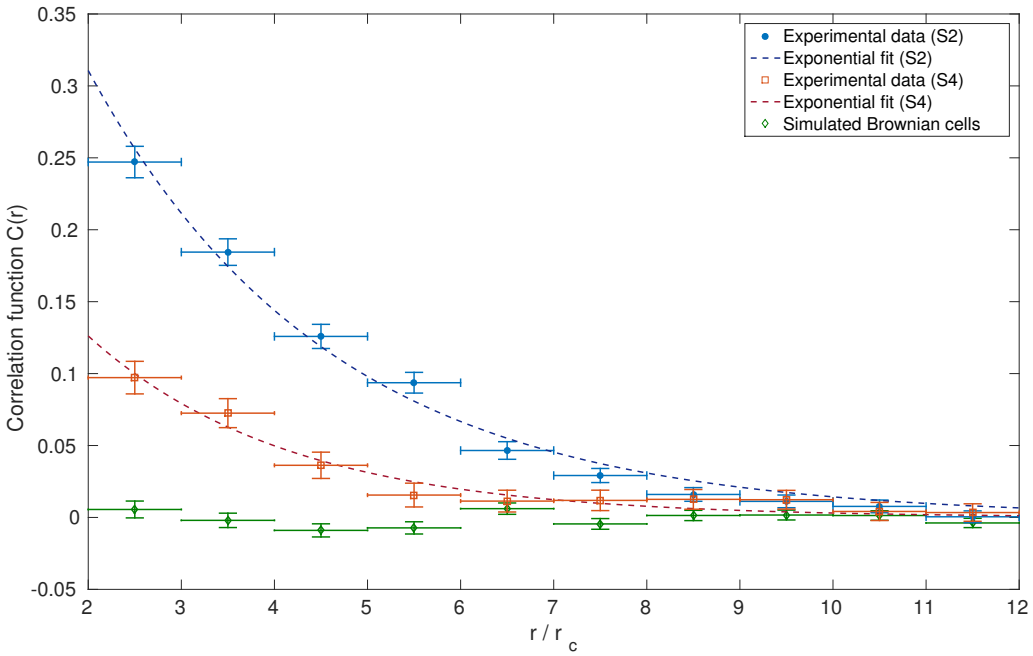


Figure 7: Correlation function calculated using Eqn. 6 for cells in SP2 (blue dots) and SP4 (red squares). Dashed lines are least-squares exponential fits to the data. Green diamonds: correlation function of a set of 2000 simulated random walk trajectories. Values below $r = 2r_c$ are not considered, as more complicated repulsive effects occur below the contact distance (see Contact inhibition of locomotion).

calculated as

$$C(r) = \frac{\sum_t^N \sum_{i \neq j}^M \ \vec{\varphi}_i(t) \cdot \vec{\varphi}_j(t) \ \delta(r - r_{ij}(t))}{\sum_t^N \sum_{i \neq j}^M \ \delta(r - r_{ij}(t))} \ , \quad (6)$$

where

$$\delta(r - r_{ij}(t)) = \begin{cases} 1 & \text{if } r < r_{ij}(t) < r + dr \\ 0 & \text{otherwise} \end{cases} \ ,$$

and $r_{ij}(t)$ is the distance between cells i and j in frame t, whereas dr is again the space binning factor. The obtained experimental correlation function for datasets SP2 (in blue) and SP4 (in red) are reported in Fig. 7, alongside the average correlation function of a simulated set of 2000 random walk trajectories. The plot reveals the existence of unexpected correlation in the alignment of cells, decaying exponentially but persisting up to a distance 4 times greater than the contact distance $2r_c$, for dataset SP2.

Some exploratory modelling attempts seem to suggest that short-range repulsion and noise in directed motion are not sufficient to produce such strong correlations, which actually recall the well known curves typically exhibited by systems of particles with some form of local alignment mechanism. It has been proposed that migrating cells could actually exhibit a tendency to align their travel direction with neighbours, due to adhesion and exchange of shear forces between cells in contact [1]. However, such a mechanism would imply an increasing correlation length as the density of the system increases, as predicted by statistical mechanics for systems of locally interacting particles and experimentally observed, for example, in nematics [23] and biological systems with active alignment mechanisms, such as midges [22] and birds [24].

Remarkably, the experimental data extracted from these swarms of cells appear to be against the hypothesis that orientational correlation is mainly due to a contact alignment mechanism, since the measured correlation function is much stronger when the cell density is relatively low. This rather unexpected behaviour, for which no match exists in the literature to the best of the author's knowledge, is further investigated in the next Section.

4 The role of stigmergy in cancer cell swarming

From a quick look at the acquired sets of trajectories (Fig. 3a and 3b) it is clear that cells are not moving randomly in the space. In low density populations (SP1, SP2, SP3), cell tracks have a tendency to travel through paths that have already been explored, drawing structured network-like patterns. This phenomenon, however, vanishes

when the density is considerably higher (SP4).

Several well-known signalling mechanisms, either chemical (*haptotaxis*, *chemotaxis*) or mechanical (*durotaxis*, *mechanotaxis*, *plithotaxis*) may be responsible for the formation of these patterns. Understanding and characterising the underlying microscopic mechanism leading to the observed phenomena falls outside the goals of this project. However, a quantification of the emergence of this complex collective phenomenon would provide a basis for claiming that there exist a relationship of cause and effect, and not just a correlation, between stimergic communication between cells and the observed density dependence of orientational correlation.

Intensity standard deviation maps

A first step in this respect can be made using a method introduced by Yang *et al.* [25] in a recent study of *trail networks* formed by brain immune cells. In this work, the authors analysed time-lapse microscopy sequences by computing the standard deviation of the local (pixel) intensity time-series, arguing that a pulse-like signal is introduced in the time-series of a fixed site every time a cell passes through, thus allowing to use the intensity standard deviation as an estimator of the cell transit frequency.

This method was used on image sequences SP2 and SP4, producing Figures 8a and 8b, respectively. Indeed, a clear network-like structure emerges in the low density case, but not when the density is higher. Note that the colorbar scale is the same in the two cases. The distribution of the measured values in the two systems is summarised by the histogram in Fig. 8c, which shows that the maximum observed intensity SD is below 100 for sequence SP4, whereas a large number of observations lie above 140 for sequence SP2.

Transit frequency maps

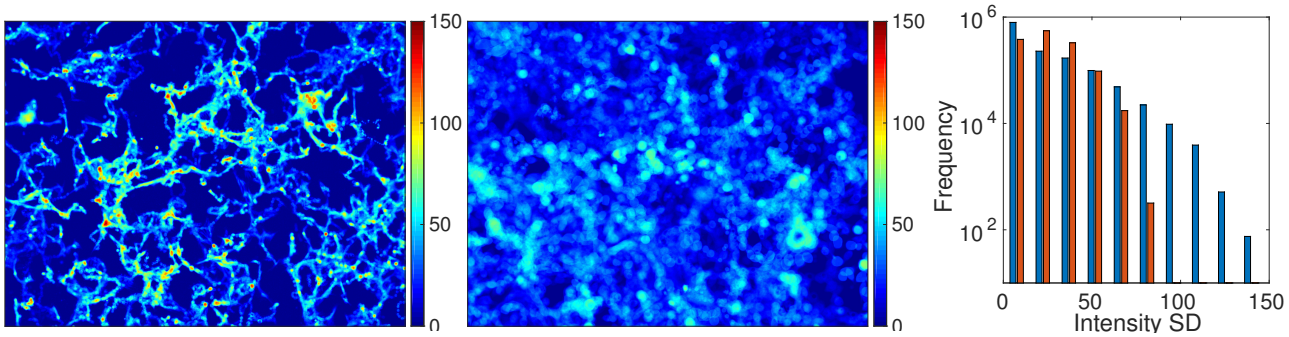
The intensity standard deviation method can be used regardless of the tracking process, as it is designed to quantify the cell transit frequency in each region of the 2D space from the raw image sequence data. However, since in this case the cell positions in each frame have been tracked, transit frequency maps can also be constructed using the obtained trajectory datasets. This was done by dividing the space into a discrete square lattice, with a step size equal to the estimated cell radius r_c, and by counting the number of times that each trajectory intersects each site. All reconstructed trajectories were used regardless of their length, as the linking accuracy is in this case irrelevant.

The obtained maps, presented in Fig. 9a and 9b, match the intensity SD maps remarkably well, showing that trail structures only emerge in the low density case. This is further confirmed by computing the fraction of sites visited at least f times for all observed values of f (Fig. 10a and 10b). The experimental curve resulting from the reconstructed cell trajectories can then be compared with numerical simulations of randomly moving cells, obtained by iteratively randomising the instantaneous angle of the velocity vectors in the real data. The visit frequency curve of high density trajectories (SP4) is statistically indistinguishable from the corresponding curve obtained from simulated random data, whereas the difference between experimental and numerical data is significant for the low density cell population.

At this stage, there is enough evidence supporting the hypothesis that cancer cell dynamics is highly regulated by stigmergy (i.e. communication through perception and modification of environmental conditions) and that the formation of a network-like trail structure is suppressed when cell density is high enough. A reasonable explanation for this behaviour can be inferred from the argument that the ability of a cell to successfully sense a signal, regardless of the particular signalling mechanism, can be modelled as some increasing function of the signal gradient in the portion of the micro-environment directly accessible to the cell's sensing apparatus. With this natural assumption, the inter-cellular signalling effectiveness would be hampered by a high cell density or, in other words, increasing the system density would correspond to an increased level of noise in the cellular signalling network, which directly affects the system correlation.

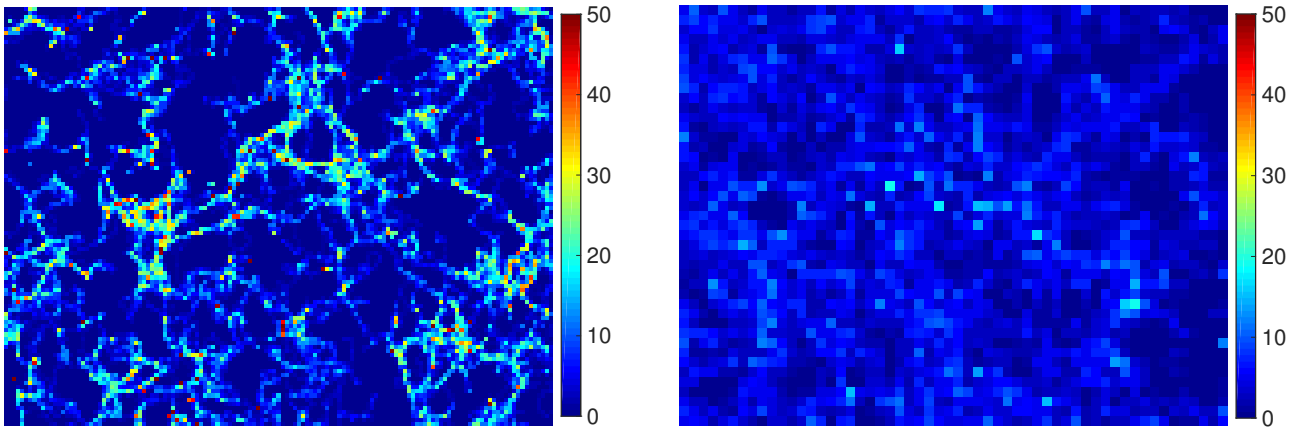
5 Conclusions and further work

The analysis of the reconstructed trajectories of lung cancer cells revealed the emergence of a form of collective behaviour without order, characterised by a relatively long-range spatial correlation, albeit exponentially decaying, in the alignment of



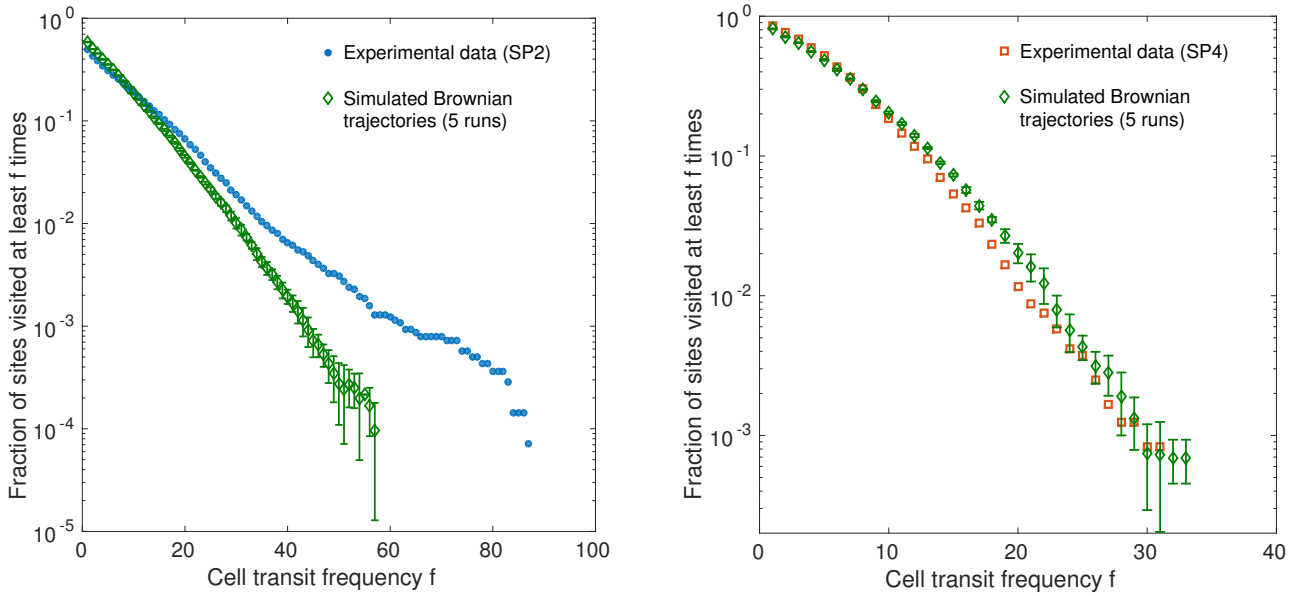
(a) SP2 (low density) **(b)** SP4 (high density) **(c)**

Figure 8: In **(a)** and **(b)**, intensity standard deviation maps obtained from the original image sequences SP2 and SP4, respectively. In **(c)**, comparison of intensity standard deviation distribution. Frequency y-axis is in logarithmic scale.



(a) SP2 (low density) **(b)** SP4 (high density)

Figure 9: Transit frequency maps obtained from the reconstructed cell trajectories for SP2 **(a)**, with $r_c = 7$ pixels, and SP4 **(b)**, with $r_c = 18$ pixels. Values are normalised with respect to system densities.



(a) **(b)**

Figure 10: Fraction of sites visited at least f times plotted against transit frequency f for SP2, blue dots in **(a)**, and SP4, red squares in **(b)**. In both figures, green diamonds are values obtained from averaging 5 runs of simulated random walk trajectories obtained by randomising the instantaneous angles in experimental datasets. y-axis is in logarithmic scale.

cell motion. Such correlations are not believed to be solely explainable in terms of contact inhibition of locomotion or other contact interactions. Furthermore, contrary to what is normally observed for interacting self-propelled particles, the correlation length of the system was observed to decrease as the density of the population increases. A natural explanation for this behaviour followed from the observation that the cells' ability to communicate through the micro-environment also exhibits a strong dependence on the system density. In particular, we claim that the reduced level of noise in inter-cellular stigmergic communication at low density, which reflects in the emergence of trail network-like patterns, may be the main cause of the observed strong correlation.

Further steps are required in order to understand the role of the observed phenomenology in the ability of cancer cells to collectively invade surrounding tissues, known to be directly related to tumour invasiveness. These steps could include a thorough comparison with systems of healthy lung cells, but also with cultures of cancer cells treated with drugs known to hamper specific growth factors or other specific pathways correlated with tumour metastatic potential. All the developed analytical methods could then be integrated with the parallel creation and simulation of (agent-based) mathematical models of cancer cells, which in recent years proved invaluable in providing new perspectives on the complexity of this bio-system and in predicting its evolution.

6 Acknowledgements

I would like to thank my supervisors Jean-Baptiste Cazier, Sabine Hauert and Anne Straube for proposing this interesting project and guiding me through its evolution. Many thanks to all students and staff in the Centre for Complexity Science for providing a friendly and stimulating working environment.
This work was supported and funded by the Erasmus Mundus programme of the EU.

References

[1] T. S. Deisboeck and I. D. Couzin. "Collective behavior in cancer cell populations". In: *Bioessays* 31.2 (2009), pp. 190–197.

[2] M. Tarabichi et al. "Systems biology of cancer: entropy, disorder, and selection-driven evolution to independence, invasion and "swarm intelligence"". In: *Cancer and Metastasis Reviews* 32.3-4 (2013), pp. 403–421.

[3] Z. Wang et al. "Simulating cancer growth with multiscale agent-based modeling". In: *Seminars in cancer biology*. Vol. 30. Elsevier. 2015, pp. 70–78.

[4] S. Hauert and S. N. Bhatia. "Mechanisms of cooperation in cancer nanomedicine: towards systems nanotechnology". In: *Trends in biotechnology* 32.9 (2014), pp. 448–455.

[5] S. Hauert et al. "A computational framework for identifying design guidelines to increase the penetration of targeted nanoparticles into tumors". In: *Nano today* 8.6 (2013), pp. 566–576.

[6] H. Yamaguchi et al. "Cell migration in tumors". In: *Current opinion in cell biology* 17.5 (2005), pp. 559–564.

[7] P. Rørth. "Collective cell migration". In: *Annual Review of Cell and Developmental* 25 (2009), pp. 407–429.

[8] P. Friedl et al. "Collective cell migration in morphogenesis and cancer". In: *International Journal of Developmental Biology* 48 (2004), pp. 441–450.

[9] T. D. Palmer et al. "Targeting tumor cell motility to prevent metastasis". In: *Advanced drug delivery reviews* 63.8 (2011), pp. 568–581.

[10] J.-B. Cazier et al. "Whole-genome sequencing of bladder cancers reveals somatic CDKN1A mutations and clinicopathological associations with mutation burden". In: *Nature communications* 5 (2014).

[11] M. Gerlinger et al. "Intratumor heterogeneity and branched evolution revealed by multiregion sequencing". In: *New England Journal of Medicine* 366.10 (2012), pp. 883–892.

[12] A. Ashworth et al. "Genetic interactions in cancer progression and treatment". In: *Cell* 145.1 (2011), pp. 30–38.

[13] C. Li and P. K.-S. Tam. "An iterative algorithm for minimum cross entropy thresholding". In: *Pattern Recognition Letters* 19.8 (1998), pp. 771–776.

[14] J. Schindelin et al. "Fiji: an open-source platform for biological-image analysis". In: *Nature methods* 9.7 (2012), pp. 676–682.

[15] I. F. Sbalzarini and P. Koumoutsakos. “Feature point tracking and trajectory analysis for video imaging in cell biology”. In: *Journal of structural biology* 151.2 (2005), pp. 182–195.

[16] A. J. Ridley et al. “Cell migration: integrating signals from front to back”. In: *Science* 302.5651 (2003), pp. 1704–1709.

[17] M. J. Saxton. “Single particle tracking”. In: *Fundamental Concepts in Biophysics*. Springer, 2009, pp. 1–33.

[18] J. A. Helmuth. “Computational methods for analyzing and simulating intra-cellular transport processes”. PhD thesis. Diss., Eidgenössische Technische Hochschule ETH Zürich, Nr. 19190, 2010, 2010.

[19] I. F. Sbalzarini. “Analysis, modeling, and simulation of diffusion processes in cell biology”. PhD thesis. Diss., Technische Wissenschaften, Eidgenössische Technische Hochschule ETH Zürich, Nr. 16440, 2006, 2006.

[20] J. R. Davis et al. “Emergence of embryonic pattern through contact inhibition of locomotion”. In: *Development* 139.24 (2012), pp. 4555–4560.

[21] R. Mayor and C. Carmona-Fontaine. “Keeping in touch with contact inhibition of locomotion”. In: *Trends in cell biology* 20.6 (2010), pp. 319–328.

[22] A. Attanasi et al. “Collective behaviour without collective order in wild swarms of midges”. In: *PLoS Comput Biol* (2014).

[23] V. Narayan et al. “Long-lived giant number fluctuations in a swarming granular nematic”. In: *Science* 317.5834 (2007), pp. 105–108.

[24] T. Vicsek et al. “Novel type of phase transition in a system of self-driven particles”. In: *Physical review letters* 75.6 (1995), p. 1226.

[25] T. D. Yang et al. “Trail networks formed by populations of immune cells”. In: *New Journal of Physics* 16.2 (2014), p. 023017.

[26] A. Masoudi-Nejad et al. “Cancer systems biology and modeling: Microscopic scale and multiscale approaches”. In: *Seminars in Cancer Biology* 30 (July 2015), pp. 60–69.

[27] A. Anderson and K. Rejniak. *Single-Cell-Based Models in Biology and Medicine*. 1st. Birkhäuser Basel, 2007. ISBN: 978-3-7643-8101-1.

[28] B. Franz and R. Erban. “Hybrid modelling of individual movement and collective behaviour”. In: *Dispersal, Individual Movement and Spatial Ecology*. Springer, 2013, pp. 129–157.

[29] U. Theisen et al. “Directional persistence of migrating cells requires Kif1C-mediated stabilization of trailing adhesions”. In: *Developmental cell* 23.6 (2012), pp. 1153–1166.

[30] E. T. Roussos et al. “Chemotaxis in cancer”. In: *Nature Reviews Cancer* 11.8 (2011), pp. 573–587.

[31] M. Rubenstein et al. “Kilobot: A low cost robot with scalable operations designed for collective behaviors”. In: *Robotics and Autonomous Systems* 62.7 (2014), pp. 966–975.

[32] M. Rubenstein et al. “Programmable self-assembly in a thousand-robot swarm”. In: *Science* 345.6198 (2014), pp. 795–799.

[33] T. R. Geiger and D. S. Peeper. “Metastasis mechanisms”. In: *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* 1796.2 (2009), pp. 293–308.

[34] K. W. Hunter et al. “Mechanisms of metastasis”. In: *Breast Cancer Res* 10.Suppl 1 (2008), S2.

[35] C. Palles et al. “Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas”. In: *Nature genetics* 45.2 (2013), pp. 136–144.

[36] E. Meijering et al. “Methods for cell and particle tracking”. In: *Methods Enzymol* 504.9 (2012), pp. 183–200.

[37] A. Cavagna et al. “Scale-free correlations in starling flocks”. In: *Proceedings of the National Academy of Sciences* 107.26 (2010), pp. 11865–11870.

[38] R. Ferrari et al. “Strongly and weakly self-similar diffusion”. In: *Physica D: Nonlinear Phenomena* 154.1 (2001), pp. 111–137.

[39] H.-P. Zhang et al. “Collective motion and density fluctuations in bacterial colonies”. In: *Proceedings of the National Academy of Sciences* 107.31 (2010), pp. 13626–13630.

[40] W. K. Chang et al. “Tumour–stromal interactions generate emergent persistence in collective cancer cell migration”. In: *Interface focus* 3.4 (2013), p. 20130017.

Supplementary Material

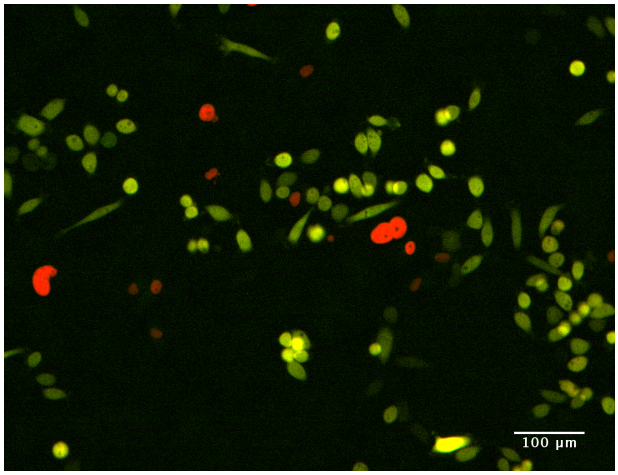


Figure S1: Snapshot from time-lapse sequence H2, in which cells with two different mutational profiles are stained with a green and a red fluorescent dye. Background removed and contrast enhanced. Scale bar $100\mu m$.

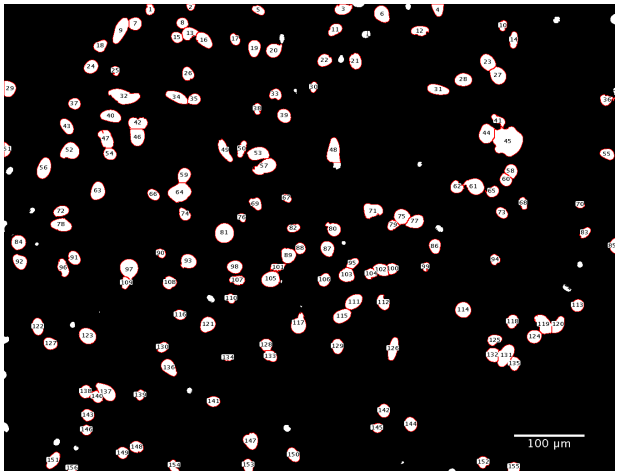


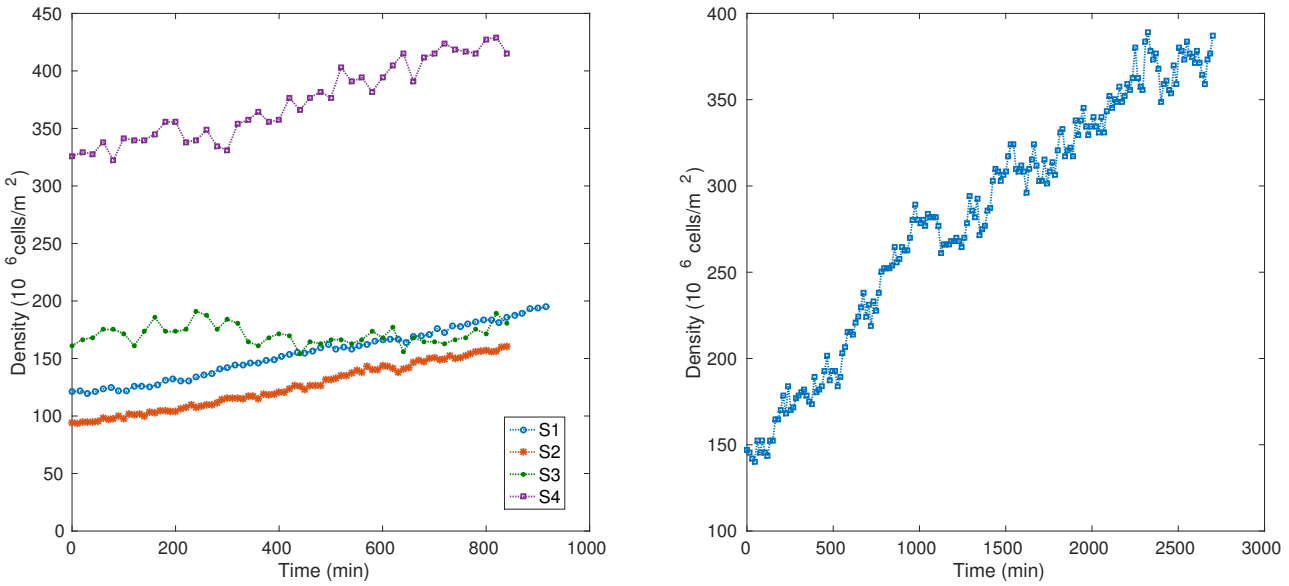
Figure S2: Segmentation of image in Fig. 1 with Li's Minimum Cross Entropy thresholding method (Li et al. [13]) and analysed with Fiji's Analyse Particles tool. Objects with a radius lower than $6\mu m$ were considered as noise particles and therefore ignored. Scale bar $100\mu m$.

	Number N of frames	Time Δt between frames	Total time T	Resolution (pixel)	Image scale (μm/pixel)	Number of sub-populations
SP1	62	15 min	15 h 15 min	1344×1024	1.61	1
SP2	85	10 min	14 h	1344×1024	1.61	1
SP3	43	20 min	14 h	1344×1024	0.644	1
SP4	43	20 min	14 h	1344×1024	0.644	1
H2	181	15 min	45 h	1344×1024	0.644	2

Table S1: Main features of available time-lapse microscopy sequences.

	Estimated cell radius (μm)	Initial density (10^6 cells/m^2)	Final density (10^6 cells/m^2)	Average density (10^7 cells/m^2)
SP1	12 ± 3	121 ± 5	190 ± 5	15 ± 2
SP2	12 ± 3	94 ± 1	158 ± 2	12 ± 2
SP3	11 ± 2	169 ± 7	175 ± 11	17.1 ± 0.8
SP4	11 ± 3	329 ± 8	420 ± 7	37 ± 3
HP2	11 ± 3	145 ± 6	372 ± 14	28 ± 7

Table S2: Estimated size and density evolution of cells.

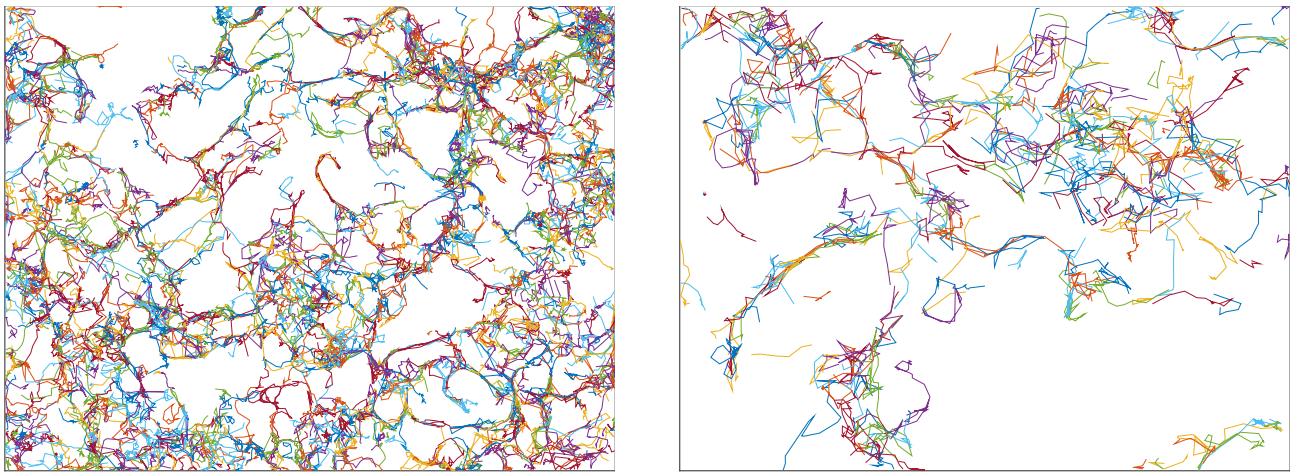


(a) Sequences SP1, SP2, SP3, SP4. **(b)** Sequence H2.

Figure S3: Time series of cell number density in analysed image sequences.

	Object radius r (pixel)	Absolute percentile i (%)	Cutoff threshold θ	Max displacement L (pixel)
SP1, SP2	10	0.05	0	50
SP3, SP4, HP2	24	0.05	0	60

Table S3: Parameters used for cell tracking.



(a) SP1 **(b)** SP3

Figure S4: Trajectories reconstructed by tracking cells in single-population sequences SP1 **(a)** and SP3 **(b)**, discarding trajectories with 5 time-points or less.

```
for i from 1 to N do
    load i-th image as a 3D array;
    \\where 3-rd dimension contains the RGB spectrum of the image
    for each trajectory j do
        if j has a point P at time i then
            compare R intensity and G intensity at point P of current image;
            if R ≥ G then
                color(i, j) = 1;
            else
                color(i, j) = 0;
            end
        end
    end
end
average color array over rows (i.e. time), removing index i;
for each trajectory j do
    if color(j) ≥ threshold then
        trajectory j is considered red;
    else
        trajectory j is considered green;
    end
end
```

Algorithm 1: Trajectory red/green classification algorithm. The threshold value can be set by the user (a value of 0.9 was found to produce good results for sequence H2).

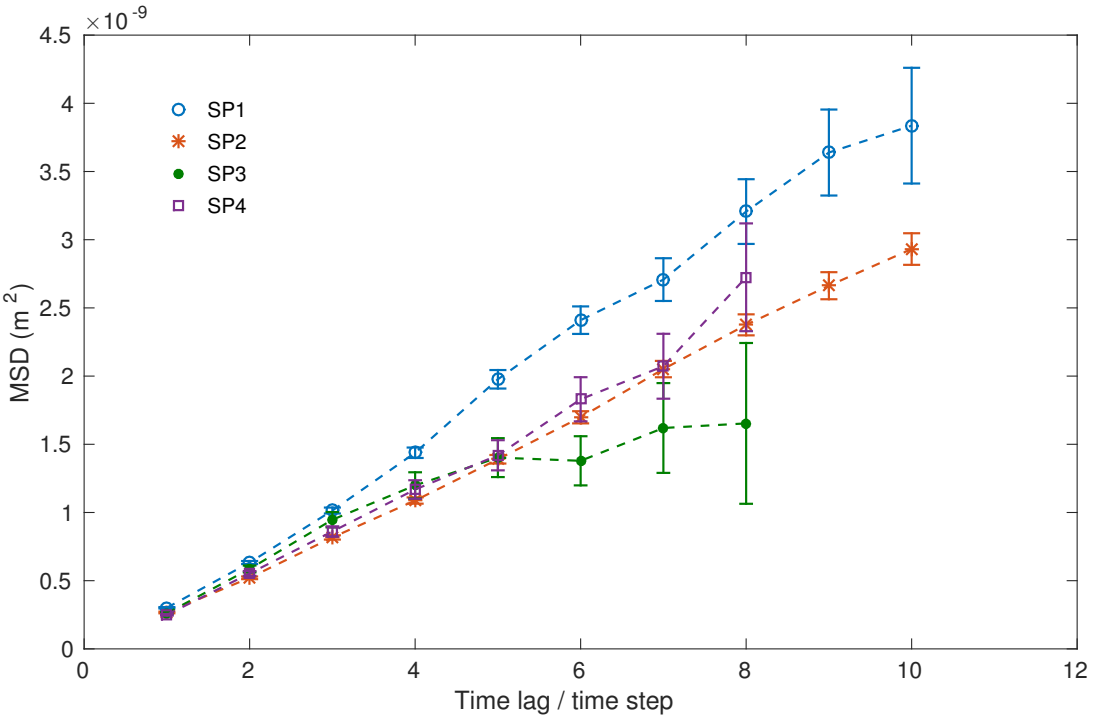


Figure S5: Mean square displacement calculated as an ensemble and time average (see Eqn. 1) using the reconstructed trajectories for sequences SP1 (blue circles), SP2 (red stars), SP3 (green dots) and SP4 (purple squares). Values are corrected according to image scaling. Error bars are ensemble standard deviations.