ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Physics (31^{st} cycle)

Stochasticity in biological systems

from modelling to experimental validation in cell growth and post-transcriptional gene regulation

By

Chiara Enrico Bena

******

Supervisor:
Prof. Andrea Pagnani, Supervisor
Dr. Carla Bosia, Co-Supervisor

Doctoral Examination Committee:
Prof. Roberto Mulet, Referee, University of Havana
Dr. Matteo Cereda, Referee, Italian Institute for Genomic Medicine (Torino)
Prof. Michele Caselle, University of Torino
Prof. Emiliano Descrovi, Politecnico di Torino
Prof. Andrea Antonio Gamba, Politecnico di Torino

Politecnico di Torino
2019

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Chiara Enrico Bena
2019

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Acknowledgements

First of all, I would like to thank Carla Bosia, for her huge patience, the teaching, support and motivation she constantly gave me during all these years. And thanks also for giving me unique opportunities.
Thanks also to Andrea Pagnani for his advices and, together with the rest of the research group, for their support and encouragement. A particular thank also to Andrea De Martino for his patience, suggestions and for sharing interesting discussions.
I wish to express a sincere gratitude also to Riccardo Zecchina for giving me the unexpected opportunity to do this PhD.

A special thank goes to my PhD companion, Marco Del Giudice, for having shared many hours of work, chats and ideas.
Thanks to Mattia Miotto, Emilia Turco, Thomas Gueudré, Matteo Osella, Alice Grob, Giulia Villari and Guido Serini.

I wish to thank also Francesca Ceroni for her kind hospitality at Imperial College in London and for the interesting discussions we had.

Thanks to Roberto Mulet and Matteo Cereda for their revisions.

Part of the projects of this thesis has received fundings from the European Union's Horizon 2020 MSCA-RISE-2016 under grant agreement No. 734439 (InferNet Project). In this context, I would like to thank also the people of the Centro de Inmunología Molecular I worked with and all the group of Complex Systems of University of Havana for their great hospitality and friendship.

Abstract

A large number of studies published in the last years shows how a population of cells is characterised by a high cell-to-cell variability. Even by considering monoclonal and identically prepared single bacteria, these give rise to highly diverse dynamics: some of them may be able to form colonies, while others do not. Such strong heterogeneity finds its roots in the stochasticity of gene expression, that leads single cell to express given genes into proteins at given levels at a certain time. However, classical experimental techniques typically analyse large population of cells and are not suitable to highlight heterogeneity at the individual level. It is only in the last decade, thanks to the development of innovative experimental techniques (e.g. single cell sequencing, FACS, etc.), that we can quantify and thus model this heterogeneity. Thus, such effects call the development of stochastic models able to both describe and make predictions of the system at a macroscopic scale by taking into account the microscopic diversity.

In this thesis we aim at investigating such heterogeneity of a population of cells under different points of view. On one side we analyse how cell-to-cell variability influences the growth of the entire population, by studying the main aspects that characterise the macroscopic growth. On the other side, we focus on a molecular level and investigate a mechanism of gene regulation that may lead to bimodal phenotypes. In both cases, we adopt an approach of investigation that can be summarised in the following loop: we begin by developing simple minimal mathematical models, then we experimentally test its prediction through quantitative, systematic measurements of variables representative of the system, and finally we modify the initial model hypothesis in light of the empirical results.

The dynamics of growth of a population of cells is the topic of the first part of this thesis. Here, we investigate the dependence on the initial conditions of different phases of growth firstly by developing two mathematical models, one for

each interesting growth phase (i.e. adaptive phase and exponential growth phase). By defining a robust experimental protocol for cell growth, we then test the predictions of the models on a widely studied cancer cell line (Jurkat). Following this approach, we suppose a possible mechanism of communication among cells on which the entire population dynamics may lye. We then investigate a second aspect of cell growth, namely its relation with gene expression. Recent works performed on growing bacterial colonies, by starting from simple empirical relations between the physiological state of a population and its gene expression, have developed an entire theory of bacterial growth. This is a model that assumes proteome partitioning and is able to be predictive even in absence of a complete knowledge of molecular details. Indeed, it has been used for predicting a wide spectrum of bacterial behaviours that range from antibiotic resistance to unnecessary protein production. Given the lack of similar studies for mammalian cells, and based on a large range of analogies between bacteria and cancer cells, we transfer the same approach to cancer cells.

The second part of the thesis is devoted to the study of the heterogeneity of a population of cells under a molecular point of view. In particular, we focus on the role of cell-to-cell variability on a peculiar mechanism of post-transcriptional gene regulation mediated by microRNAs (miRNAs). miRNAs are small non coding RNA molecules able to bind to other messenger RNAs (mRNAs) and downregulate their expression. It has been found that such regulation may lead the system to bimodal distributions in the expression of the target mRNA, usually fingerprint of the presence of two distinct phenotypes. Moreover, the nature of the interaction between miRNAs and their targets gives rise to a complex network of miRNAs interacting with several mRNA targets. Such targets may then cross-regulate each other in an indirect miRNA-mediated manner. This effect, called "competing endogenous RNA (ceRNA) effect", has remarkable properties even in presence of extrinsic noise, i.e. fluctuations that affect all the components of the system. While miRNA-mediated interactions have been widely characterised from a theoretical point of view in the past years, quantitative experiments on the ceRNA effects are much more recent. Here, we first review the stochastic models developed to describe the miRNA-mediated gene expression, pointing out the predictions that may be experimentally investigated. Second, we present the experimental setup we used to test such predictions. Last, we focus on the experiments performed to test the influence of extrinsic noise on miRNA-target interaction.

Contents

List of Figures **x**

List of Tables **xxiv**

1 Introduction and thesis outline
Quantitative biology: where physics meets biology **1**

I Stochasticity and emergent behaviours in cell growth **8**

2 Initial density dependence in cell growth **14**

2.1 The growth of a population of cells: modelling and experiments . . 15

2.1.1 Initial cell density dependence 17

2.1.2 Phases of growth 18

2.2 Mathematical models for cell growth 20

2.2.1 The carrying capacity 21

2.2.2 Population and single cell lag times 21

2.2.3 Logistic growth model 25

2.3 Experimental testing of theoretical predictions 29

2.3.1 Approaches inspired by experimental physics 30

2.3.2 A systematic experimental approach revealed robust growth curves 33

2.3.3 The determination of the growth phases 40
2.3.4 The independence of the carrying capacity on the initial condition 44
2.3.5 Population lag time and the extreme values statistics 45
2.3.6 Exponential growth rate and initial conditions 52
2.3.7 Cell-to-cell communication 63
2.3.8 New insights on the experimental design 67
2.3.9 Qualitative identical features are shown by a further cancer cell line 69
2.4 Conclusions 72

3 Cell growth laws **75**

3.1 The Ohm's laws for bacterial growth 76
3.1.1 Analogy with an electric circuit 81
3.2 Project outline 82
3.3 A systematic experimental approach 84
3.3.1 Cell growth experiments 84
3.3.2 Three sets of experiments 87
3.3.3 RNA and Protein extraction 89
3.4 Experimental results 89
3.4.1 Calibration 90
3.4.2 Optimal sugar concentrations 93
3.4.3 Towards the growth laws 95
3.5 Conclusions and perspectives on growth laws 99

II Modelling and experimental test of miRNA-target interaction 102

4 Overview of microRNA-mediated gene regulation 107

4.1 Overview on microRNAs 107

4.2 MiRNA-target interaction is titrative 109

4.3 The role of intrinsic and extrinsic noise 111

4.4 Competing Endogenous RNAs (ceRNAs) 112

5 Modelling the role of noise in miRNA-target interaction 114

5.1 Stochastic model for miRNA-target interaction with extrinsic noise . 114

5.2 Bimodality as a population-level effect 117

6 Experimental investigation of the model predictions 121

6.1 State of the art 122

6.1.1 One miRNA and one target mRNA 122

6.1.2 One miRNA and two ceRNAs 124

6.2 The role of extrinsic noise in miRNA-target interaction 125

6.2.1 Experimental setup 127

6.2.2 Preliminary results 128

6.2.3 Future perspectives 135

Appendix A The role of sodium acetate on cell growth 136

A.1 Bioproduction and aim of the work 136

A.2 Summary of the experimental methodology 139

A.3 Experiments 139

A.3.1 Image segmentation and growth parameters 141

A.4 Toxic activity of sodium acetate 144

A.4.1 Robust growth curves 144

A.4.2 The tuning of the parameters of growth 145
A.5 Criticalities and improvements 150
A.5.1 New image segmentation algorithm to recognise single cells and clusters 151
A.5.2 Testing of the algorithm on a second set of experiments . . 153
A.6 Conclusions 156

Appendix B The role of STIM1 in endothelial cell endosomes trafficking 159
B.1 Overview on endosome trafficking and project outline 160
B.2 Methods for endosome tracking 161
B.3 Results and discussion 165

References 171

List of Figures

2.1 Figures adapted from [1]. (a) Time-lapse images of *Salmonella* Typhimurium individual cells over time. A. Cell able to form a colony; B. cell filamentation; C. cell lysis precedes cell division. (b) Representative growth curves of individual cell colonies of *Salmonella* Typhimurium. (c) Probability distributions of the number of cells in microcolonies originated from different initial N_0s after 8 hours of growth obtained from Monte Carlo simulations. 16

2.2 Phases of growth. Bottom plot represents the log_2 of bacteria density, the top one the growth rate (a.u.) both as function of time. Dashed vertical lines divide the growth phases. Growth phases identified by numbers: lag phase (1), acceleration phase (2), exponential growth phase (3), deceleration phase (4), stationary phase (5) and death phase(6). Figure adapted from [2]. 20

2.3 (a) Growth curves of a small initial density population of bacteria seeded in growth medium with all nutrients in excess except one which is limiting. Each shade of blue represents different concentration of such growth-limiting nutrient. The population is left free to grow by exhausting nutrients and the total growth (i.e. carrying capacity) is emphasised by vertical arrows. (b) The total growth is plotted as function of the concentration of the growth-limiting nutrient. Both (a) and (b) are adapted from [3]. 21

2.4 Schematics for the lag time model. (a) Cartoon summarising the model. At the population level, a population of N_0 cells (light blue dots in the top-left circle) needs a time t_{lag} before starting to grow exponentially over time ($N(t)$) according to 2.2. The same dynamics is valid at the single-cell level, where each single cell i starts after a time τ_i to give birth to sub-colonies whose sizes grow exponentially over time $n_i(t)$. Each colour of the dots in the bottom circles represents a single cell (on the left) and its corresponding sub-colony (on the right). (b) Schematic representation of the two possible scenarios for the population lag time t_{lag} derived from eq. 2.5. The Gaussian statistic scenario (in blue) and the extreme-values one (in red). This second case has been obtained considering a uniform distribution for the τ_is. 25

2.5 Cartoons representative for the analysis of the logistic growth model. (a) The solution of eq. 2.8 plotted as the logarithm of the population size as a function of time. The initial condition is given by the initial population size N_0. (b) Mapping of the punctual growth rate (dN/dt) vs population size (N). (c) Maximum growth rate (λ_{max}) calculated as the derivative in time $t = 0$ of growth curves obtained by the model with different initial conditions plotted as function of N_0. In all the three plots the carrying capacity level (k) is emphasised by a dashed black line. 28

2.6 Figure adapted from Bio-Rad protocol for cell counting [4]. Example of a counting grid called Neubauer counting grid. It is composed of nine sections of $1\,mm^2$ each indicated by a red number. 32

2.7 Cartoon of the experimental procedure. From left to right, from a flask with cells growing exponentially in their standard growth medium, at time t = 0 h, samples of N_0 cells have been taken and moved into 6-well dishes supplied with the very same nutrient quality present in the flask. From time t=0 on the growth was monitored through automatic cell counting and growth curves similar to the one on the very right have been obtained. 36

2.8 (a) Micrograph of a representative sample of cell population. (b) Result of the image segmentation algorithm applied on image (a). Colours do not have a specific meaning, are simply used to distinguish single objects that are represented with a progressive number. 39

2.9 Examples of experimental growth curves expressed as the logarithm of the number of cells in time. In (a) single replicates at different inoculi (N_0) are plotted. Each colour is relative to a different inoculum. Solid lines represent the fit of the corresponding data. N.B. Cyan and blue curves are two examples of the same inoculum. Since their inflection point coincide, a difference in the slope of the curves can be easily noted. In (b) three technical replicates are considered separately and identified by a shade of blue. Error bars are the propagation f the error over the counting. 40

2.10 Example of growth curve (blue dots) with the corresponding logistic fit (blue line) with lag time t_{lag}, maximum growth rate λ_{max} and saturation level A. The horizontal and tilted red lines emphasise the parameters A and λ_{max} respectively. The two vertical red lines represent the edges of the exponential phase, that correspond to times in t_{lag} and t_{log}. t_{log} is given by the intersection between the tangent at the inflection point and the level of saturation. The error bar on the data are obtained by the error propagation over $ln\frac{N}{N_0}$ by assuming $\sqrt{N}$ as error for the concentration N. 42

2.11 z-score distributions for the comparisons among corresponding fitting values obtained through the modified logistic and the modified Gompertz functions. Distributions of the z-scores of (a) saturation level parameters A, (b) maximum growth rates λ_{max} and (c) lag times t_{lag}. Vertical lines represent the upper value for accepting a compatibility with significativity of 5%, $z_{crit} = 1.96$. 43

2.12 (a) Parameter A vs $ln(N_0)$. Orange dots are experimental data, the solid black lines represents their linear fit. The dashed vertical line is located at the averaged carrying capacity k. (b) Distribution of the carrying capacities for each experiments, obtained by converting the parameter A through the exponential form of eq. 2.18. The solid black line emphasises $k = (8 \pm 2) \cdot 10^6$cell/ml, the average of the distribution for values smaller than $1.5 \cdot 10^7$cell/ml. 45

2.13 (a) Experimental lag times as function of the initial seeding (N_0). Lighter colour dots are experimental data obtained through the fit. Their error bars are the error on the fit. The darkest red dots are the average of the smaller dots binned over N_0. (b) Standard deviation of the averaged values (dark red dots) of plot (a). In both plots the vertical dashed lines are located in correspondence of the carrying capacity value k. 47

2.14 (a) Schematic representation of the microfluidic device used to follow single cells. Adapted from User guide of CellASIC® ONIX M04L-03 Microfluidic Plate. Red numbers represent: 1. The wells where the solution of cells+growth medium is loaded; 2. the effective growth chambers; 3. the outflux chambers. (b) Bright field micrograph at time $t = 0$. Red circles emphazie single cells. 52

2.15 (a) and (c) Exponential growth rates (λ_{max} and λ respectively) as a function of the initial cell density (N_0). Brightest dots represent all the experiments. Error bars are the error of the fits performed to calculate λ and λ_{max}. The darkest dots are the average of the brightest ones binned over N_0. (b) and (d) Standard deviations of the data of (a) and (c) for the growth rates by binning over N_0. The dashed black line present in all the plots is located at the value of N_0 corresponding to the carrying capacity k. 54

2.16 (a) and (b) Three bins on N_0. For each bin, the exponential growth rates λ_{max} in (a) and λ in (b) are plotted as a function of $N_{A/2}$ and N^* respectively. 55

2.17 Same plot as Figure 2.15. The bright dots represent the experimental data for λ_{max} as a function of N_0 with the error bars representing the error of the fit. The darkest dots are the average of the brightest ones binned over N_0. Continuous blue line represents the theoretical λ_{max} according to the logistic growth model. Dashed horizontal blue line emphasises the offset value r. Dashed vertical black line is set at the level of the carrying capacity k. 56

2.18 Representations of the growth rate (dN/dt) of a population of size N as function of its size. (a) Trends for logistic growth (blue solid line), weak (green) and strong (yellow) Allee effect. The value k corresponds to the carrying capacity. (b) Examples of weak Allee effects with different values of parameter α that increases according to the arrow. The blue curve is the logistic case with $\alpha = 0$. The dashed black line represents the carrying capacity k. 58

2.19 (a) Example of mapping of the punctual growth rate ($\frac{dN}{dt}$) vs the size of the population (N). Cyan dots are experimental data. A sub-set of growth curves is here considered. The red line is their fit with a weak Allee effect function; blue line represents a logistic growth. (b) Distribution of the positive values of α obtained by fitting the experimental growth curves with a weak Allee effect function, eq. 2.30. 59

2.20 Scheme of the procedure through which t_{lag} and t_{log} are defined starting from the growth curves obtained through the weak Allee effect dynamics. 60

2.21 Exponential growth rate λ vs initial population density N_0. Light small dots are the experimental data, the error bars represent the error on the fit parameter. Darkest red dots are λs evaluated from the theoretical growth curves. Dashed vertical line is represents the carrying capacity k. 62

2.22 (a-c) Example of growth curves with three different pipetting times: three times a day (a), once every two days (b) and once every five days (c). Blue data represent the control, i.e. counting every day. The legend shown in (a) refers to (a), (b) and (c). (d) Bar chart of the values of the growth rates of the pipetting experiments (orange bars) and the control (blue bars). The values have been obtained by averaging the growth rates of each single replicates. The error bar is the error over such average value. The three groups of data correspond to the three experiments: pipetting three times a day (3/dd), once every two days (1/(2dd)) and once every five days (1/(5dd)). 67

2.23 (a-b) Standard deviations of the maximum growth rate λ_{max} (a) and the lag time t_{lag} (b) within bins over N_0. (c) Same standard deviations (σ) shown in (a) and (b) normalised to their maximum values (σ_{max}) as a function of the inoculum density (N_0). In red is the exponential growth rate and in blue the lag time. The solid lines represent the fits of the two series of data. The black dot at the intersection of the two lines represents the "optimal" experimental condition. The dashed vertical line is located in correspondence of the carrying capacity. . 68

2.24 (a) Cartoon of the experimental procedure. K562 cells were pre-cultured in a flask supplied with the standard growth medium. At time $t = 0$ h populations of N_0 cells were transferred into a new culture dish (a new flask) supplied with a fixed volume of the same standard growth medium. The growth of the populations was daily monitored through the *NucleoCounter*. The last plot shows three representative growth curves obtained through this procedure at the end of each experiment. Each colour represents an inoculum size (N_0), the dots are the experimental data and the solid lines are the sigmoidal shape functions fitting the data. (b) Trend of the fitting parameter A as a function of the inoculum density N_0. The data are compatible with a line with slope -1 as a function of $ln(N_0)$, as emphasised by the black solid line that fits the data. (c) Distribution of the values of the carrying capacities converted from the parameter A through the relation 2.18. The solid blue line represents the normal distribution best fitting the data. The solid black line is the average of the distribution, corresponding to $k = (7.5 \pm 1.8) \cdot 10^5$cell/ml. Such value is emphasised by a dashed vertical line in Figures (b,d,e). (d) Lag time (t_{lag}) trend as a function of N_0. Each bright dot corresponds to a single growth curve, the darkest ones are the average by binning on N_0. The error bars are given by the error of the fit. (e) Exponential growth rate (λ) as a function of N_0. As in (b) the bright dots correspond to all the experimental data, while the darkest ones are their average by binning on N_0. The error bars are the error of the fit. 71

3.1 Growth rate of *Escherichia Coli* as function of glucose concentration. The solid line is obtained with the Monod relation (eq. 3.1). Figure adapted from [2]. 77

3.2 Three figures showing the results obtained by Scott et al. in [5]. (a) Sketches of the empirical growth laws. Figure adapted from [6]. (b) Summarising sketches of the theory of proteome partitioning. Figure adapted from [5]. (c) Analogy with Ohm's laws for electrical circuits. The three resistors correspond (also in colours) to the three proteome sectors of panel (b). Figure adapted from Supplementary Materials of [5]. 80

3.3 Example of A (blue) and B (red) growth curves expressed as the logarithm of the cell concentration (N) normalised to the initial cell concentration (N_0) as a function of time. For each colour there are 3 shades representative of 3 different replicates of the same condition. 85

3.4 An example of K562 growth curve in glucose at 37°C. (a) shows the raw measurements performed by the *Tecan*: absorbance at 430 nm in red and 560 nm in green. (b) shows the ratio of absorbance 430 nm over 560 nm *vs* time after background subtraction. 87

3.5 The left panel relates to Jurkat and the right one to K562. Plots of the row (a) represent the logarithm of the number of cells per ml (N). Plots (b) show measurements of the same wells in (a) performed with the *Tecan*. Plots (c) reveal the relation between cell concentration and absorbance ratio. Here two sets of data have been joined to determine the conversion law. The darkest dots corresponds to the growth curves in (a) and (b), the brightest ones are the second set of data whose growth curves are not shown. 91

3.6 Sugar saturation results for Jurkat (left) and K562 (right) cell lines. Each raw and each colour correspond to a different sugar. Data represented as dark filled circles are the average values of different replicates (brightest circles). The error bars show replicates dispersion. The growth rates are expressed as referred to the reference value for *Tecan*'s measurements λ_b. 94

3.7 (a) Jurkat (on the left) and K562 (on the right) growth rates (λ/λ_a) as a function of the inoculum size. The brightest dots are all the experimental data and their error bars are the errors obtained by the fitting procedure. The darkest circles represent average values obtained by binning over N_0. The data correspond to the experiments at 37°C with glucose (coloured in red, according to the legend of (b)). (b) Growth rates (λ/λ_b) as a function of temperature for Jurkat on the left and K562 on the right. Each colour corresponds to a different sugar. Black solid lines emphasise the trend of the data. . 96

3.8 Contents of RNA per cell, protein per cell and ribosomal fraction (RNA/protein) as function of the growth rate. (a) shows results of Jurkat cell line and (b) of K562, both at 37°C. Black solid lines represent the linear fits of the data. (c) RNA/protein data of Jurkat cells shown in (a) joint with the results obtained in Torino's lab from manual counting experiments shown as red open circles. The x-axis represents the growth rates normalised to the reference ones: λ_b for filled dots (Tecan's measurements), λ_a for open circles (manual counting data). Each colour corresponds to a single sugar, as explicited by the legend that refers to all the shown plots. 98

4.1 MiRNA biogenesis and target binding. Adapted from [7]. 109

4.2 (a) Network of interaction between microRNA (S) and its target mRNA (R). k_R and k_S are the transcription rates of target mRNA and miRNA respectively and g_R and g_S their degradation rates. k_P and g_P are respectively the transcription and the degradation rates of the protein (P). g is the interaction strength between the miRNA and the target and α is the fraction of not-recycled miRNA after the binding. Adapted from [8]. (b) Cartoon of threshold-like behaviour describing miRNA and target mRNA interaction. The amount of free mRNA is shown as a function of its transcription rate. Below the threshold the amount of free mRNA molecules is smaller than that of miRNA (repressed state); the two quantities are roughly equal around the threshold and above the threshold the free amount of mRNA is higher than that of miRNA (expressed state). Adapted from [8]. 111

5.1 (a) Phase diagram of the bimodality in the distribution of free molecules of target mRNA. The extrinsic noise on the miRNA transcription rate (CV) is plotted as a function of the target transcription rate (k_R). The colour map shows the presence of bimodality for different values of the miRNA-target interaction strength g (see the legend). Adapted from [8]. (b) Distribution of free mRNA in the case of: pure intrinsic noise and small miRNA-target interaction strength g (solid black line); pure intrinsic noise and high miRNA-target interaction strength g (blue histogram) and extrinsic noise and small interaction strength g (orange histogram). Adapted from [8]. . 119

6.1 (a) Two colour fluorescent reporter system used in [9]. Bidirectional plasmid engineered to code for eYFP and mCherry genes. The 3'UTR region of the latter gene contains a number N of binding sites for the miR-20a. Adapted from [9]. (c) Relation between fluorescence intensity of the mCherry gene (proxy for the amount of free molecules of target mRNA) as a function of that of eYFP (proxy for the transcription rate of the target). The two series of data represent the case of 0 (black data) and 1 (blue data) binding sites for the miR-20a. Adapted from [9]. 123

6.2 (a) Fluorescent reporter system used in [10]. The upper construct is the same birectional plasmide used by [9] and described in Figure 6.1(a). The lower one is the bidirectional plasmid coding for mCerulean and mKOrange. The 3'UTR sequence of the former gene is engineered to contain a fixed number of N binding sites for the miR-20a. Adapted from [10]. (b) Sketch of the minimal model of miRNA-ceRNA interactions studied in [10]. Adapted from [10]. . . 124

6.3 (a) DNA marker (Hoechst DNA stain) distribution. The cell cycle phases (G0/G1, S and G2) are emphasised by arrows. (b) Distributions of the Hoechst DNA stain for big (in green) and small (in red) cells. The inset shows the determnation of the two groups of cells by constraining on cell granularity and dimension. Both (a) and (b) are adapted from [11]. 129

6.4 Zero miRNA binding sites. Scatter plots (each dot is a cell) of the fluorescence intensity of the mCherry as a function of that of eYFP. The cells are sorted according to the phase of the cycle: from left to right, G0/G1, S and G2 phase. Adapted from [11]. 130

6.5 Scatter plots of the fluorescence intensity of mCherry as a function of that of eYFP. Each dot represents a cell. Each raw corresponds to a different number of binding sites, from up to down 1, 4 and 7 respectively. Each column corresponds to a sorting according to the cell cycle phase, from left to right: G0/G1, S and G2. Adapted from [11]. 132

6.6 (a) Phase diagram for the bimodality in the distribution of the free molecule of mRNA target of a miRNA. The x-axis represents the transcription rate of the target mRNA (k_R); the y-axis shows levels of the extrinsic noise; and the colours are relative to different values of interaction strengths (parameter g). The white arrows parallel to the x- and y- axis represent the direction of variation of the corresponding experimental parameters, i.e. the cell cycle phase as a proxy of the extrinsic noise and the eYFP intensity for k_R. Adapted from [8]. (b) Similar phase diagram of plot of (a) obtained with the experimental data corresponding to the case of strong interaction strength (7 binding sites for the miRNA). The extrinsic noise is shown as a function of the level of eYFP fluorescence intensity. The distributions are relative to the mCherry expression level for cells sorted according to the cell cycle phase and eYFP levels. From high to low noise (i.e. from top to bottom), the cell cycle phases are: unsorted, G0/G1, S and G2. 134

A.1 Two examples of micrographs (a) and (c) and their respective image segmentation results (b) and (d), where each RGB colour represents a different detected element. (a) and (b) are an example of lower concentration while (c) and (d) set of images refers to a high inoculum experiment. 142

A.2 Example of size distribution of object detected by the algorithm. The highest value of area is not visible in the plot and corresponds to a size of 34750 px^2. The red line represents the mean of the distribution, the black line the mode. The mode has been chosen as representative of single cell areas. 143

A.3 Examples of growth curves obtained for two different concentrations of sodium acetate: 0 mM in blue, 20 mM in red. Dots are experimental data expressed as the mean value of 3 replicates. Error bars are given by the replicates semidispersion. (a). Results of low inoculum experiment ($N_0 \sim 3 \cdot 10^5$cell/ml). Solid lines are the sigmoidal shape fit of the experimental data. (b). Growth curves from high inoculum experiment ($N_0 \sim 3 \cdot 10^6$cell/ml). Solid intersected lines represent the trend of the exponential and saturation phases of the experimental data. 145

A.4 Growth parameters analysis. (a) shows the exponential growth rate (λ) as function of sodium acetate concentration. Blue dots are relative to the lower inoculum experiment ($N_0 \sim 3 \cdot 10^5$ cell/ml), orange dots refer to higher inoculum ($N_0 \sim 3 \cdot 10^6$ cell/ml). Each point is a single experiment. Error bars are obtained through the fitting procedures that estimates the parameters. The lines represent the trend of the data. (b) represents the carrying capacity vs sodium acetate concentration. Again here the error bars are given by the fits. In this case, the general trends of the data are relative to sodium acetate concentrations bigger than 0. In (c) the lag time vs sodium acetate concentration only relative to the experiment at low inoculum. Error bars are the errors of the fit to obtain the lag time. The line is the linear fit of the data. 147

A.5 Examples of inputs (a,c) and corresponding outputs (b,d) of the new counting algorithm. The darkest spots in (a) and (b) are dead cells. In (b,d) the single cells are coloured in red and clusters in blue. . . 153

A.6 Results obtained through the use of the new algorithm and the new sample preparation protocol on the second set of experiments. (a) and (b) are representative growth curves at low and high inoculum density respectively. Red stands for a concentration of 0 mM of sodium acetate while blue is 20mM. Dots are experimental data, the lines show their trends. Each curve corresponds to a single replicate. The error bars are the propagation of the error on the counting. (c) Growth rate trend as a function of the sodium acetate concentration. Each dot corresponds to a single replicate, the error bars are the error of the fit. The lines represent the trend of the data (linear fits). Low density data are shown in blue, while high density ones are in orange. (d) Carrying capacity as a function of the sodium acetate concentration. Dots are experimental data and their error bars are the error of the fit. When not visible is because they have the same size than the dots. The lines represent linear fits of the experimental data. It is valid the same colour legend of plot in (d). 154

B.1 Example of the size trend in time of an early endosome detected with the image segmentation and tracking code. The three colours represent the three endosomes in which the first detected one (#174) is split. For further analysis only the green one will be considered, since it appears continuously for more than 21 frames. 164

B.2 Comparisons between features of early (EE, left column) and late endosomes (LE, right). Blue data represent siSTIM1 condition, red/orange ones the control. (a) Shows the average number of endosomes per cell *vs* time for EE and number distribution for LE. In(b) again temporal average trend for EE and distribution at a fixed time for LE of endosome sizes. Line (c) is dedicated to distance to nucleus distributions. For EE the plot of cumulative distribution functions showing the difference between the two populations is shown. Last line (d) represents the persistence versus minimal path of early endosomes, not available for LE. Blue and red lines here represent general trend of correspective data. 169

B.3 Schematic representation of the main results of the project. Each square represents a portion of a cell, where the empty circle is the nucleus and the black line the membrane. Early and late endosomes are represented respectively as green and red filled circles whose sizes give an idea of the differences in endosomes size. From top to bottom the control (siCTL) situation is shown together with the following conditions: (a) STIM1 silencing (siSTIM1); (b) dyneine silencing (siDyn); (c) KIFC1 silencing (siKIFC1) and (d) the combination of silencing of KIFC1 and dynein together. 170

List of Tables

2.1 Logistic and Gompertz modified equations [12] that explicitly express the growth parameters (A, λ_{max} and t_{lag}). 41

2.2 *z-scores* resulting from normal tests performed to compare fit parameters of the experiments (a) and (b) with their control. Each column is relative to the comparison between specific parameters: lag times ($z_{t_{lag}}$), maximum growth rates ($z_{\lambda_{max}}$) and saturation level (z_A). The critical value for a two-tails test with significancy of 5% is $z_{crit} = 1.96$. 67

3.1 Values of growth rates (λ) and relative doubling times ($d.t.$) of manual counts (first two columns) and *Tecan*'s ($3^{rd} - 4^{th}$ columns) measurements. Lines correspond to cell lines. 92

3.2 Summary of optimal sugar concentration chosen for each cell line (rows) and each sugar (columns). 95

Chapter 1

Introduction and thesis outline Quantitative biology: where physics meets biology

The discovery of the double-helix structure of DNA is the first striking instance of meeting between physics and biology. Indeed, the Nobel Prize laureates Watson and Crick proposed the nowadays well-known structure of DNA basing on the experimental evidences of Franklin and Wilkins, who applied the then relatively new technique of X-ray crystallography to DNA [13]. Similarly to X-ray diffraction, several techniques built on physical principles have been developed for decades by physicists and engineers and found a fertile field of application in biology and medicine. Such technical advances resulted as the easiest and more natural way for a physicist to enter biology. One can think for example of nuclear magnetic resonance (NMR) or computer assial tomography (CAT) scan as familiar techniques that allowed to improve diagnostics. Furthermore, from a more basic research point of view, there are advanced tools such as scanning electron mycroscopy (SEM) or confocal microscopy - techniques nowadays widely used for investigating cell functions - that are the result of studies on lenses, optics, image analysis algorithms. Thus, the first intersection between the two disciplines saw physics providing experimental techniques to biology, maintaining the two sciences focused on different subjects of investigation. This has been clearly evidenced by the British physicist Donald in [14], where she states that physicists typically tend to simplify the system under exam as much as possible in order to model it. Given this, for a long time

biology has been seen too complex and disordered to be modeled and thus it has not been considered as a subject of interest for physicists. However, in the last decades, the two sciences have undergone changes that made them approaching from both experimental and theoretical points of view. Indeed, on one hand, the developed modern techniques have recently allowed to sequence the human genome, make microfluidic experiments, investigate single cell functions with high accuracy, just to give some examples. Such new tools give the possibility to biologists to both explore new fields and expecially to acquire an always increasing amount of quantitative data. This, in turn, calls for new quantitative methods to deal with such an increasing amount of data [15]. On the other hand, a new interest towards complex and disordered systems rised among physicists. The peculiar feature of a complex system is the emergence of a collective behaviour from the individual activity of its large number of components. Such emergent behaviour is different than the simple sum of single individuals and presents its own characteristics, that could not be predicted by investigating the dynamics of isolated single components. Furthermore, in such systems, the disorder and the heterogeneity play a fundamental role becoming themselves new and interesting topics in modern physics. Consider for example a flock of birds. This is an out-of-equilibrium system composed by a large number of single birds moving autonomously, each one following its own motion and interacting through each other and with the environment. However, when we zoom out and look at the flock as a whole, we can say that there is a flight direction in which it moves with a coherent amorphous shape, out of the network of single components interactions [16]. Similar examples can be found in biology also at different length scales, such as the molecular motors within cells [17], collective motion of cells in developing embrios, the crawling cells [18] or swarming bacteria. These last ones can swim in association with other cells in a thin film of liquid over a moist surface [19]: such process is called swarming and can produce macroscopic motions that involve millions of cells, reminiscent of birds flocking.

The issue of conciling macroscopic laws (e.g. flock behaviour) with microscopic variability (e.g. single bird motion) is a classical task of statistical physics. Here, starting by the knowledge of (i) the presence of high number of single individuals and (ii) the laws driving the motion of such single components, it is possible to derive several macroscopical features by addressing the system through a probabilistic point of view and avoiding to solve each equation of motion of the single components [20]. A known example of the application of statistical physics to link macroscopic and

microscopic behaviours, is given by the statistical thermodynamics. Let's consider a gas. It is known from classical thermodynamics that the state of the system is given by macroscopical variables like temperature, pressure and volume and thanks to the classical thermodynamical laws it is possible to predict the behaviour of the system when holding all variables fixed but one. However, the gas is composed by a large number of small particles (molecules) that are known to follow Newton's and quantum mechanics. Statistical mechanics enters exactly here, since it aims at deriving classical macroscopical laws by applying probability theory to the mechanic equation of motion for systems composed by large ensemble of particles [21].

Systems like flocks of birds or molecular motors are referred to as "active matter" systems [22] and its features and collective emergent behaviours have been investigated since recent times through the use of statistical physics tools.

Thus, going back to the relation between physics and biology, the merging of (i) the use of advanced technologies in parallel to consolidated biological ones, (ii) the increasing interest of physicists towards complex systems together with (iii) the improvements in computational power, arised the opportunity to develop always more deep mathematical analysis on the experimental results and thus mathematical modelling of biological systems. Such models then do not aim at simply fitting the empirical data, but they point towards the ability of being predictive. This is the framework that goes under the name of "quantitative biology", a recently born interdisciplinary research field characterised by a continuous interplay between physics and biology, not only on the methodology and techniques used but also on the kind of questions arised. As reported by Shekhar and collaborators in [15], quantitative biology adopts accurate measurements in order to refine hypothesis and aims at comparing experimental results with predictions of theoretical models. One of the ambitious challenges of the modern science has been clearly stated by Bialek in [23] and it is the "development of a theoretical physics of biological systems that reaches the level of predictive power that has become the standard in other areas of physics".

This thesis adopts such quantitative biology framework and investigates two different biological issues through a both theoretical and experimental approach that can be summarized in the following way - by rephrasing again what Bialek said in [23]. First, we identify the principles that describe a given question at a system level; second, we express such principles through a mathematical formalism and

finally, we test the quantitative predictions of the mathematical theories by defining the experimentally accessible phenomena and the right methodologies that allow to make the needed measurements. For this last purpose of testing mathematical models, two are the main issues: first, the need of highly controlled experiments, second, the necessity of accurate and innovative statistical methods. To do so, we borrow methodologies and approaches by experimental physics: we fix all the representative variables of the system but one and tune this "free variable" to perturb the system. Its response is then studied by making measurements of different observables characteristics of the system and corresponding to the theoretical variables. Empirical results and theoretical predictions are then compared in order to test the model and relax or modify the eventual falling hypotheses.

The thesis is focused on the study of the heterogeneity of population of cells under two very different points of view. On one hand, we study how such variability impacts on the growth of the population by analysing different aspects of cell population growth. On the other hand, we focus on a molecular level and investigate a mechanism of gene regulation that may lead to bimodal phenotypes. Due to the high diversity of these two topics, the thesis is divided into two main parts.

In Part I, we investigate the growth of population of cells as the emergent behaviour of an ensemble of individual proliferating cells. In particular, we focus on two different aspects of such growth: its dependence on the initial conditions on one side (Chapter 2) and its relation with gene expression, on the other side (Chapter 3).

The former work, presented in Chapter 2, has been driven by the presence in literature of many evidences stating the relevance of the following two elements on the dynamics of the growth of a population of cells: first, the high heterogeneity among single cells within a population of genetically identical cells (see for e.g. [1]) and second, the initial density of the population [24–30].

We develop mathematical models that aim at bridging the gap between macroscopic (population level) and microscopic (single-cell level) behaviours and subsequently test such theoretical predictions on an *in vitro* system composed by a population of cancer cells. To this purpose, we developed a highly controlled and accurate wet-lab experimental protocol through which we measure population-level growth variables. In light of the experimental results we modify then the mathematical models to better interpret and support empirical observations.

Chapter 3, instead, focuses on the study of the same populations of cancer cells and aims at identifying the relation between population growth and gene expression. Such study has been inspired by recent quantitative works carried out in microbiology where the physiological state of cells has been related with their gene expression. From simple quantitative empirical relations called the bacterial Ohm's laws, it has been developed a phenomenological model for proteome partitioning able to be predictive even in absence of a complete knowledge of molecular details [5]. Such theoretical model has been used for predicting a wide spectrum of bacterial behaviours that ranges from antibiotic resistance to unnecessary protein production [31]. Since similar studies are lacking for mammalian cells and based on the presence in literature of many analogies between bacteria and cancer cells [32, 33], we transfer to the above-cited population of cancer cells, the same systematic approach used for bacteria. The work is part of a collaboration with researchers at Imperial College in London.

In Part II of the thesis, we investigate the heterogeneity of a population of cells from a biochemical point of view. In particular, we focus on the role of cell-to-cell variability on a peculiar mechanism of post-transcriptional gene regulation mediated by microRNAs. MicroRNAs (miRNAs) are small molecules of non-coding RNAs that can recognize and bind other molecules of messenger RNA (mRNA). Every miRNA is predicted to regulate many targets and, viceversa, every mRNA targets may be regulated by a combination of different miRNAs. Once the two species are bound, the eventual translation of the mRNA into protein is inhibited. In particular stoichiometric conditions, when the pool of miRNA is limiting, different targets may crosstalk to each other (an effect called "competing endogenous RNA (ceRNA) effect") in an indirect miRNA-mediated manner. This mechanism of indirect interaction has been widely analysed from a theoretical point of view [34–36]. The interaction between miRNAs and targets is titrative [9, 10, 34] and it depends on the relative abundance of the two species. If miRNAs are more abundant than targets, all target mRNAs are expected to be bound to miRNAs and the system is said to be in its repressed regime. In the opposite situation, when mRNA abundance exceedes that of miRNA, then some molecules of mRNA will be not bound to miRNA and thus free to be translated. In this second case, the system is said to be in its unrepressed (or expressed) regime. The regime between the two states, close to the equimolarity point of the two molecular species, is highly sensible to small fluctuations in the amount of both the molecular species: depending on

which prevails, the system can jump either in one state or in the other. Such feature suggests the possibility to observe bimodal distributions in the amount of free target molecules simply due to stochastic fluctuations. Bimodality is of particular relevance in biology since it is usually related to the presence of phenotypic variation: each mode of the distribution is linked to a physiological state [8]. MiRNA-target systems are of peculiar interest in this sense, because aberrant miRNA expression is often related to cancer or other diseases. Thus, the modes of the bimodal distributions may represent different desease states or cancer subtypes. Moreover, the biochemical reactions involved are probabilistic and take place in a noisy environment. These two sources of noise are usually referred to as intrinsic and extrinsic noise [37]. Several stochastic models have been developed in order to investigate miRNA-target interactions in the presence of noise. In particular, it has been found that the extrinsic noise plays a pivotal role in regulating the rising of bimodal distributions even in case of weak interaction strength between miRNAs and targets [8]. Given the biological importance of bimodal distributions, the fact that through the presence of noise the system can be driven in one state or the other is not trivial.

As mentioned, the miRNA-target systems have been widely investigated from a mathematical point of view through both analytical and numerical approaches [10, 34–36]. However, a similar broad experimental characterisation and testing of the model predictions is still lacking. Thus, in Part II, we first give an overview on the existing models able to describe the main features of miRNA-target interactions and ceRNA effect (Chapter 5). Afterwards, by focusing on the interaction between one miRNA and one target, we experimentally test the predictions about the rising of bimodality in presence of extrinsic noise (Chapter 6).

The end of the thesis is devoted to two appendices, Appendix A and B, where we discus two further projects I carried out during the PhD. In both cases, we adopted a systematic approach mainly on experimental development, data acquisition and analysis, similar to those described in Part I and II.

Appendix A, is mostly related to Chapter 2 and has been performed at Centro de Inmunología Molecular (La Habana, Cuba) in collaboration with researchers both of the research institute and the Physics Faculty of the Universidad de La Habana. The project fits in the context of bioproduction, namely the employment of cells as producers of chemicals useful for therapeutical purposes. On this topic, we refer also to the commentary [38] I had the opportunity to work on recently. It is known that the

adding of sub-lethal doses of external chemicals to the growth medium may enhance the bioproductivity of the growing cells [39–41]. In light of this, we investigate the response of a peculiar cell line known for its bioproductivity features, when exposed to different doses of an external chemical, i.e. the sodium acetate. In particular, by taking advantage of the knowledges on cell growth optimization through initial cell density acquired in Part I, we finally aim at optimizing the growth in order to make cell both growing fast and enhancing their productivity.

Appendix B refers to intra-cellular trafficking carried out by endosomes, vesicles responsible for the transport of the extra-cellular molecules within the cells. The project is part of a collaboration with researchers of the Institute for Cancer Research and Treatments (Candiolo, Italy) and concerns the effect of a specific protein on the regulation of endosome trafficking based on molecular motors. Our contribution focused on the data analysis, in particular on the development of an image segmentation and particle tracking algorithm able to extract quantitative information on endosome morphology and dynamics from fluorescent time-lapse videos. In the dedicated appendix B we discuss the main results of the entire work that is object of a pubblication in preparation entitled “STIM1 regulates the dynein/dynactin complex formation contributing to the transport of early and late endosomal traffic on microtubules”.

Part I

Stochasticity and emergent behaviours in cell growth

Cell growth: from single cell dynamics to population behaviour

From everyday experience, it is known that a favourable environment may induce a population of bacteria to grow faster [2]. Such simple observation gives evidence on the feature of cell growth in encoding information on cellular fitness, making the growth a remarkable process of cell physiology. For this reason, since Monod's first quantitative work on bacterial populations [2], an always increasing number of studies have been carried out to investigate microbial proliferation. When considering the typical growth of a population of cells in an environment with finite-nutrients, after a phase of adaptation (called lag phase), the size of the population grows exponentially (linearly in semi-log scale) in time as long as nutrients are available [2]. Of great relevance are the quantification of the two variables representative for such growth phases, called the lag time and the exponential growth rate. The former gives information on the time needed for the population to adapt to the new environment, while the latter represents a measurement of the proliferation. The growth rate is referred to as an asymptotic behaviour since, in an environment with infinite nutrients, the population would keep on growing constantly at such rate. Thanks to its feature of being constant during the entire exponential growth [42], the exponential growth rate is usually considered as the representative observable for defining the physiological steady state of the culture. To further support the pivotal role of the growth rate in encoding information of cellular fitness, several works have recently investigated quantitative laws relating such variable with further cellular features such as population size, macromolecular composition and ribosome functions [5, 43, 44].

The growth described so far concerns the proliferation of the entire population considered as a whole. However, such macroscopic behaviour is the result of the

growth of each single cell that composes the system. Thanks to improvements of modern technologies like single-cell techniques, microfluidics and high resolution microscopy [45], it is nowadays possible to perform quantitative measurements of single cells growth. The parameters that have been typically studied to characterise the growth are division time intervals and single cell sizes. An always increasing number of works has been performed in this direction and revealed the presence of an high heterogeneity even in populations composed by genetically identical cells [46, 47]. An interesting example is given by the work of Koutsoumanis and coworkers [1] who investigated the dynamics of microcolonies originating from identically prepared single bacteria. Through a systematical experimental approach, they found that some cells were able to form microcolonies while others did not. Moreover, the dynamics of the grown microcolonies showed different behaviours both in the time needed by the single cell to replicate the first time and in the effective growth rate of the subsequent colony. Such high heterogeneity is observed also within cells belonging to the same growing colony, as observed for instance by Wang and colleagues in [42]. Here, the authors developed the "mother machine", a high-throughput microfluidic device that allowed to follow cells for a large number of generations and thus quantify such heterogeneity.

The cited evidences shed light on the seemingly paradoxical behaviour of a population of cells: the strong heterogeneity at single-cell level results in a constant behaviour when the whole population is considered. Moreover, the population features are not simply the average of the single cell features: the heterogeneity combines in a more complex way. The population growth rate, for instance, is smaller than the mean single cell division times [46, 48], suggesting then that the population grows faster than the average single cell. Thus, the challenge has become the understanding of how the strong single-cell heterogeneity combines in order to give rise to a macroscopical emergent behaviour with its own features.

This part of the thesis fits exactly here. We consider the growth of a population as the emergent behaviour of an ensemble of single cells growing in a certain environment. Thus, by adopting approaches inspired by both statistical and experimental physics we aim at characterising the different phases of growth of such system. Firstly, we focus on a theoretical point of view and develop mathematical models able to make predictions on the growth of the population by taking into account single cell dynamics. Secondly, we test such predictions on *in vitro* systems consisting of populations of two widely studied cancer cell lines (Jurkat and K562).

To this purpose, we need to perform highly controlled experiments that allow us to have statistically significant data to compare to the theoretical scenario. For this reason, we develop a robust protocol of cell growth inspired by methodologies usually adopted in experimental physics that allows to obtain an high statistics of measurements at the population level. Such measurements need to be representative of the response of the system when all variables are fixed but one. In our case we consider the initial size of the population as the only tuning knob and investigate the dependence on it of the different phases of growth. Finally, we modify the initial models in light of the empirical results. The main body of such topic is discussed in Chapter 2. The Chapter ends with considerations on how the experimental results may be useful for the design of experiments for populations of growing cells. The work is object of a publication in preparation which integrates the topics discussed in [49].

Chapter 3 focuses on evidences on the cell fitness that emerge from quantifications of the growth rate. Robust bacterial growth theory recently emerged by linking the growth rate of the population with the macromolecular composition of the cell [5]. The growth of cells requires (i) the conversion of environmental nutrients into amino acids and (ii) the subsequent synthesis of proteins by the ribosomes that polymerise such amino acids into polypeptide chains [31]. By studying bacteria growing exponentially in different conditions (and with different growth rates) simple empirical relations known as "growth laws" have been found. By combining such relations with a coarse grained partition of the protein expression profile (the proteome) a general theory for bacterial growth raised [5]. Due to an analogy with electrical circuit behaviour, such theory has been known also as bacterial Ohm's laws [5]. Ohm's laws simplified the design of electrical circuits way before the microscopical understanding of the electricity and thus they did not required any detailed atomic-level description. Similarly, the correlations described in [5] are considered as microbial growth laws that can help the understanding and design of complex biological systems with the ability of being predictive without the need of a complete knowledge of molecular-scale details [5, 6]. Such bacterial theory indeed, revealed to be very useful for predicting different aspects of cell behaviours. It has been used, for instance, in investigating the burden of hetherologous protein expression: in agreement with experimental results, the model is able to predict the decrease of the population growth rate when unnecessary proteins are expressed [5]. Moreover, it has been adopted to reveal feedback effects that govern the interaction

between drug and drug resistance [50] and to predict the way biosynthetic pathways balance enzyme cost with product demand [51].

Given its pivotal relevance in microbiology and the lack of similar studies for more complex organisms, in Chapter 3 we intend to extend bacterial growth laws to other systems by investigating the relation between population growth rate and macromolecular cell composition in a different cell type. Since in Chapter I we showed that the initial population size influences the growth rate of the population, in Chapter 3 we consider the same cell types and the same macroscopic observable to tune the growth rate. Moreover, we identified other two additional measurable quantities, sugar quality and temperature, and studied the response of the growth while perturbing one observable at a time. Even though eukaryotic cells are more complex systems than bacteria, we aim at transferring the very same systemic approach used for deriving bacterial growth laws, thanks to a wide range of analogies in the behaviours of bacteria and cancer cells.

Chapter 3 begins with a detailed overview on the bacterial growth theory and on the methodologies adopted in that case. Then, it follows a description of the systematic experimental protocols used for our purposes and the analytical tools for the identification of the interesting observables. Finally, we present the results. The work presented here is part of a collaboration with researchers at Imperial College (London) and is the subject of a publication in preparation.

Chapter 2

Initial density dependence in cell growth

In this Chapter we investigate models of population growth. In particular, we are interested in studying the dependence of growth on the initial condition, mainly the initial density of the population.

For this purpose, we first overview different stochastic models based on single-cell dynamics recently developed for bacterial growth (Section 2.1) pointing out the relevance of the initial concentration of cells on the population dynamics. In Section 2.2, we focus on different growth phases and discuss the models that can be used to predict their trends. Then, in Section 2.3 we test these models on a real *in vitro* system consisting in a population of a widely studied cancer cell line. The cell line is called Jurkat clone E6.1 and is composed of human leukemic T-lymphocites. We describe the experimental methods adopted and discuss the reasons that drove us to develop systematic experimental protocols. We then present the analytical tools applied to analyse the data. In light of the results, we subsequently modify the previously presented models in order to better interpret the experimental results (from Section 2.3.4 to Section 2.3.7). Section 2.3.8 is devoted to the discussion on how the obtained experimental results could drive choices for the design of new experiments of cell growth. At the end of the chapter, in Section 2.3.9 we discuss how the same qualitative features shown by Jurkat cells are reproduced also by a second cancer cell type that we investigated, the K562 cell line.

2.1 The growth of a population of cells: modelling and experiments

The intrinsic heterogeneity of single cells within a population plays a pivotal role in the development of the dynamics of the population as a whole. To better understand this, we discuss here (see Figure 2.1) some of the results of the interesting work on the dynamics of bacterial microcolonies performed by Koutsoumanis and coworkers [1] and mentioned in Chapter I.
Figure 2.1(a) depicts three examples of different dynamics due to single cell heterogeneity: cells in Figure 2.1(a)A were able to form a growing colony, while other ones did not and exhibited filamentations (Figure 2.1(a) B) or lysis before division (Figure 2.1(a) C). Figure 2.1(b) shows that, focusing on the single cells able to form microcolonies, even the dynamics of such colonies showed different behaviours. For instance Cell 1 (diamonds) and Cell 3 (triangles) in the plot show that within the same time range (6 hours), the latter colony reaches higher densities than the former (ln(cell number)$\sim$ 4.5 vs 2.5). Furthermore, the number of cells in microcolony born from Cell 3 increases after a shorter time than that related to Cell 1, that did not replicate (ln(cell number) $= 0$) for more than 3 hours. Such time, called *lag time*, is the time the cells need to adapt to a new environment before being able to replicate exponentially.

Classical deterministic models adopted to describe cell population growth fail when small initial populations are considered, because heterogeneity among cells becomes relevant when dealing with small numbers. For this reason, models for population dynamics have been developed based on stochasticity of the single cells [1, 52–55]. Let's consider again the work of Koutsomanis and colleagues [1], which represents an interesting example for our purposes. Based on the experimental results, the authors developed a stochastic model of growth that takes into account single cell heterogeneity. Through Monte Carlo simulations, they investigated the effect of the initial population size (N_0) on population dynamics. From the simulations they found two main features. Firstly, the variability in population growth decreases with the increasing of N_0, as shown by Figure 2.1(c). As a consequence, in case of large N_0 ($N_0 > 100$ cells) - a usual condition of laboratory growth experiments - the variability becomes negligible and the system has a deterministic behaviour. This supports the use of deterministic models for describing high initial population

sizes, but also the need for stochastic models when interested in smaller N_0s. The second result concerns the lag time. In agreement with findings on *Escherichia Coli* obtained by Pin and Baranyi [55], Koutsoumanis and colleagues showed a decrease in the lag times when N_0 increases.

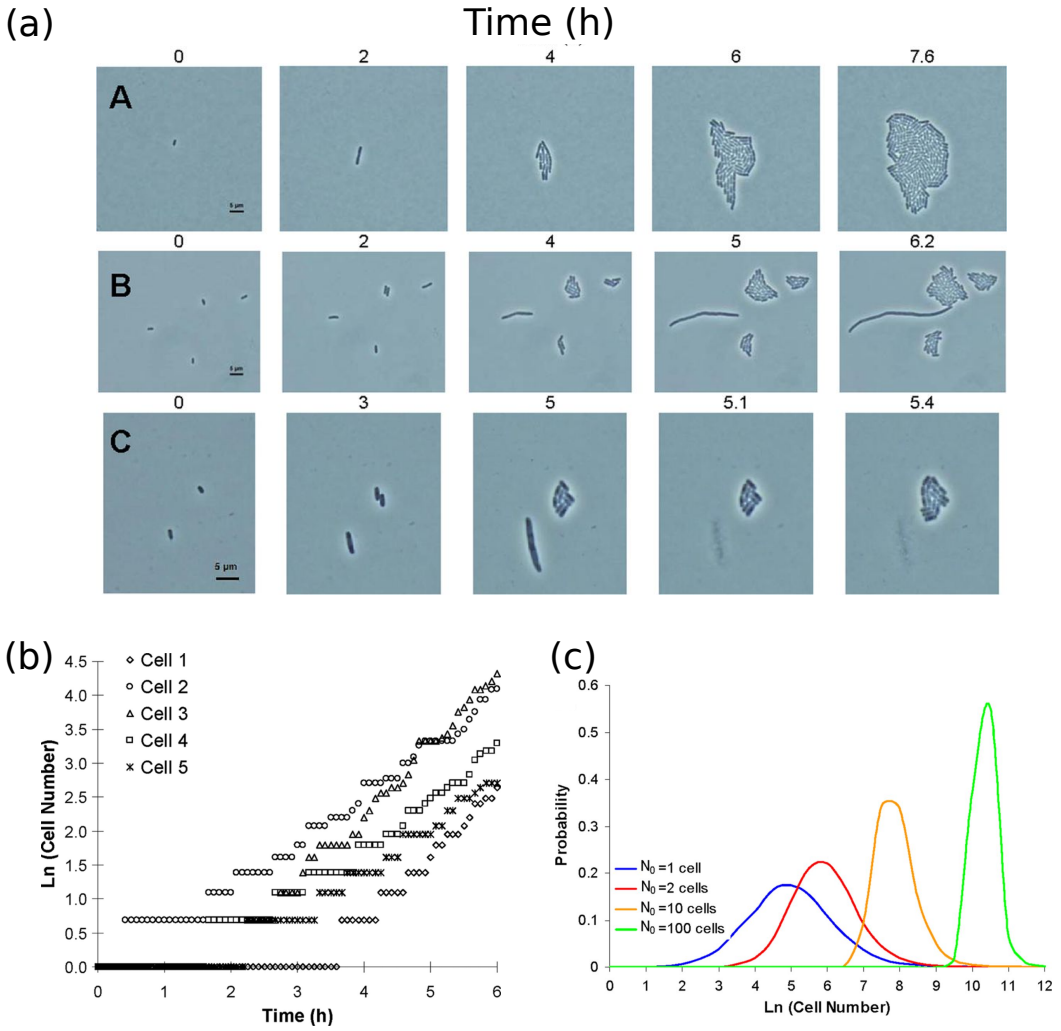

Fig. 2.1 Figures adapted from [1]. (a) Time-lapse images of *Salmonella* Typhimurium individual cells over time. A. Cell able to form a colony; B. cell filamentation; C. cell lysis precedes cell division. (b) Representative growth curves of individual cell colonies of *Salmonella* Typhimurium. (c) Probability distributions of the number of cells in microcolonies originated from different initial N_0s after 8 hours of growth obtained from Monte Carlo simulations.

Such results call the attention to two main intertwined arguments. On one hand, the importance of the employment of stochastic models for the interpretation of cell

population dynamics; on the other hand, the dependence on the initial conditions of such dynamics.

2.1.1 Initial cell density dependence

The initial cell density (N_0, inoculum) has been found to influence the growth of the population in a wide range of organisms, from bacteria [56, 57] to plants [24, 26, 27], insects [28], mammalian and cancer cells [25, 29, 30]. Across the different studies, initial population size showed relevant effects on both metabolic signatures and the physiology of the organism. Metabolic signatures refer to the ability to secrete specific compounds. Ozturk and colleagues [30] for instance, found that the efficiency to produce monoclonal antibodies - example of bioproducts used for therapeutical purposes - in a murine cancer cell line depends on the initial concentration of the cells. The physiological aspects that may reveal an inoculum-density dependence concern instead the different phases of growth of the population, namely the maximum attainable cell density [30] (also called carrying capacity), the adaptation time and the doubling time (for details see next Section 2.1.2). In particular, the growth rate of the population, which is inversely related to the doubling time, can increase [28, 24] or decrease [27, 25] with the initial density depending on the organism and on the environment. In case it remains constant, its fluctuations can drastically change when modulating the initial seeding [57, 26].

Because of the plethora of results, a direct comparison among different organisms and conditions is difficult. However, the very fact that effects of inoculum dependence are evident in such a wide spectrum, suggests the presence of peculiar mechanisms intrinsic of the initial condition whose effects are preserved during the subsequent history of the population. Such initial condition dependencies in cell growth may be due to many different causes. For instance, in the bacterial growth model defined by Baranyi [58], the initial density dependence of the growth rate is imposed by the growth medium through a finite carrying capacity. As quantified in [59], the growth rate is expected to decline when the inoculum density approaches the carrying capacity, since a strong competition among cells occurs in order to overcome starvation.

Evidences of initial condition dependence are also present when considering the lag time, as found for example in *Escherichia Coli* by Pin and Baranyi [55]. As

mentioned in Section 2.1, many population models based on stochasticity of single cells have been used also to predict the growth of a population given different initial number of cells [1, 52].

In light of these observations, we are interested in modelling the population growth and in discussing the growth phase dependencies on the initial population size, N_0. Before going into details, we present a brief introduction on the phases of growth of a population of cells. Afterwards, for each parameter that represents the growth phases, we discuss a mathematical model whose predictions are subsequently tested.

2.1.2 Phases of growth

One of the most used methods for growing cells in wet-laboratories is the so-called batch-cell culture. The method consists in seeding the cells in a new environment with a limited amount of nutrients and leaving them free to grow by consuming the growth medium. The dynamics of a population of cells growing in such condition has been widely studied in microbiology. Of great relevance is the work of Monod [2], where he identified mainly three distinct phases representative of the growth: the *lag*, the *exponential* and the *stationary* (or saturation) phases. Figure 2.2 shows such phases in terms of both the logarithm of bacteria density (bottom) and the growth rate (top) of the population. During the *lag* phase (represented by the number 1 in Figure 2.2) the population adapts to the new environment and the density of cells is constant in time: the growth rate is equal to zero. This phase is followed by the *exponential* phase (number 3 in Figure 2.2) in which the number of cells increases exponentially (linearly in the semi-logarithmic scale in the plot) and thus the growth rate is again constant but with a non-zero value. So, the population density is able to grow until it reaches a saturation phase (*stationary* phase, number 5 in Figure 2.2) and it is not able to increase anymore. The growth rate in this phase is again constant and equal to zero. This regime is followed by the *death* phase (number 6 in Figure 2.2) where the population decreases as well as the growth rate. Furthermore, Monod identified two more phases (numbers 2 and 4 in Figure 2.2) that relate to the transitions from the three main ones and present a non-constant growth rate. In phase (2) the number of cells increases slightly until it enters the exponential phase. This phase is often considered as an *acceleration* phase (Monod) or second lag phase [60], and differs from the first one because both the growth rate and the cell density

are not null. Phase (4) links instead the exponential and the stationary phases and is called a *deceleration* phase, due to the decreasing of the growth rate in time.
For each of the three main growth phases (1, 3 and 5), Monod identified a representative variable: the *lag time*, that represents the duration of the lag phase; the *exponential growth rate*, namely the rate of growth during the exponential phase; and the *carrying capacity*, the maximum achievable cell density in a given growth medium, also known as *total growth*.

The lag phase is the most debated one. It is in general considered an adaptation phase and its total duration may include both the first two phases identified by Monod. For our purposes we will include in the lag time also the acceleration phase.

The growth rate instead is often considered a constant regime, due to the fact that every cell component doubles at a constant rate [2]. Due to this observation, the exponential growth rate (λ) is usually related to the doubling time ($d.t.$), an experimentally widely used variable, according to:

$$d.t. = \frac{ln(2)}{\lambda} \tag{2.1}$$

This equation is obtained by considering that the doubling time represents the time a population in the exponential phase needs to double its size.

The total growth, i.e. the carrying capacity level, is a property of the medium and is controlled by the quantity of nutrients in the medium.

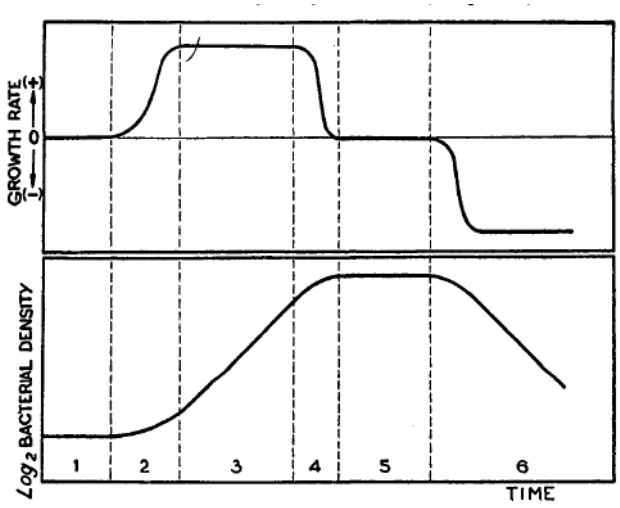

Fig. 2.2 Phases of growth. Bottom plot represents the log_2 of bacteria density, the top one the growth rate (a.u.) both as function of time. Dashed vertical lines divide the growth phases. Growth phases identified by numbers: lag phase (1), acceleration phase (2), exponential growth phase (3), deceleration phase (4), stationary phase (5) and death phase(6). Figure adapted from [2].

2.2 Mathematical models for cell growth

Consider a theoretical batch cell culture experiment. A sample of N_0 cells, belonging to a population growing at a constant rate, is transferred in a new environment with fixed amount of nutrients. In this new condition, cells are left free to grow by consuming nutrients. After an adaptation (lag) phase, the population grows exponentially at a constant rate up to the maximum attainable level where the growth stops and the concentration of cells saturates. In such scenario, we are interested in the following parameters representative for each growth phases: the lag time (t_{lag}), the maximum growth rate in the exponential phase (λ_{max}) and the carrying capacity (k).

This section is organised into three parts. Section 2.2.1 discusses our expectations on the carrying capacity. Section 2.2.2 presents a null model we developed to describe the population lag time by taking into account the single cell variability. Finally, in Section 2.2.3, we present a deterministic model for population growth through which we make expectations on the exponential growth rate of the population.

The theoretical predictions here discussed are then tested in Section 2.3.2.

2.2.1 The carrying capacity

As mentioned in Section 2.1.2, in batch cell cultures the density of the population saturates at a level defined by the amount of nutrients present in the growth medium. Such saturation level can be then adjusted by changing the quantity of a particular nutrient in the medium, as depicted in Figure 2.3(a). In this case, all nutrients but one are left constant and it is possible to see that the level of saturation changes accordingly to the concentration of the growth-limiting substrate (Figure 2.3 (b)).

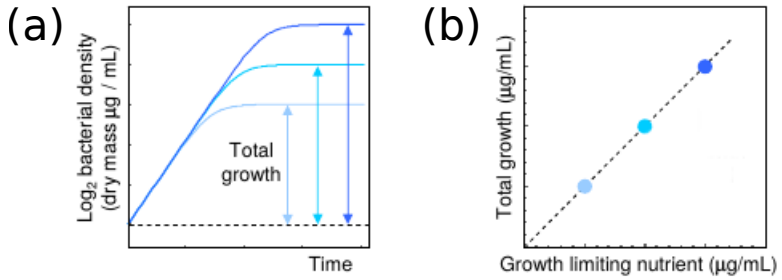

Fig. 2.3 (a) Growth curves of a small initial density population of bacteria seeded in growth medium with all nutrients in excess except one which is limiting. Each shade of blue represents different concentration of such growth-limiting nutrient. The population is left free to grow by exhausting nutrients and the total growth (i.e. carrying capacity) is emphasised by vertical arrows. (b) The total growth is plotted as function of the concentration of the growth-limiting nutrient. Both (a) and (b) are adapted from [3].

For our purposes, we are interested in fixing the environment, thus the quality and quantity of nutrients, and consider the initial concentration of cells as the only variable to tune. For this reason, we expect that the maximum reachable level of concentration is constant with respect to the initial cell density, indeed it represents the maximum amount of cells that a given environment can carry (thus the name "carrying capacity").

2.2.2 Population and single cell lag times

To better understand the importance of the lag time models that take into account cell-to-cell variability, we briefly overview the main principles underlying the existing models.

An overview on bacterial lag time models

When cells are transferred in a new environment, they need a certain time to adapt to the new condition before starting to exponentially proliferate. A clarifying example can be found in [54], where a widely known bacterial metabolic process is described: the growth of *Escherichia Coli* in glucose and lactose. These two represent two sources of carbon that allow cell growth. When bacteria growing exponentially in a medium supplied with glucose, are moved to a new environment with lactose instead of glucose, the population shows a lag phase before the beginning of a new exponential growth. Within this period, cells adapt to the new condition by inducing the production of the specific enzyme (lactase) needed to process the lactose. In particular, the absence of glucose and the presence of lactose induce transcription and translation of a set of three genes (lac operon) coordinately regulated within *E. Coli*. In this way, the cells become able to metabolize the new carbon source and can thus replicate. According to this phenomenon, if the carbon source is not changed, in principle any lag time is expected during the population growth.

The interest in developing mathematical models to predict bacterial lag phase duration has recently enhanced [61]. One of the main reasons lies in an increasing attention of the modern societies for food safety and thus in having an always deeper knowledge of the growth of microorganisms in food. Thus, a broad literature on mathematical models for the prediction of population lag times is available. Methods we are interested in are those able to give predictions on the population lag times, based on the stochasticity of single cell dynamics [53, 62]. In such models, the biological variability of the population is represented by the model parameters which are random variables distributed over the entire population [61]. In such models two different lag times are distinguished and connected: the population and the single cell lag times. The former is sometimes considered as the average of the latter [54]. However, as stated also in Chapter I, this is not always true when careful analysis on the single-cell stochasticity is considered.

Such topic is the core of the next section, where we present our model for lag time that relates population and single cell variables.

Mathematical model for population lag time

Consider a population of N_0 cells moved to a new environment at time $t = 0$. After the adaptation time $t = t_{lag}$ the population grows exponentially according to (by neglecting the carrying capacity)

$$N(t) = N_0 e^{\Lambda(t-t_{lag})}, \tag{2.2}$$

where Λ is the exponential growth rate of the population. The very same dynamics is valid also for every single cell i belonging to the initial population N_0, with $i = 1,2,N_0$. Indeed, each single cell needs a time τ_i to adapt to the new environment. In agreement with the existing models (e.g. [62]), we assume that the single cells are independent and do not have any information on the size of the colony they belong to. Thus, after τ_i, each sub-colony derived from cell i grows exponentially with a constant rate Λ_i. We assume that the growth rates of the sub-colonies are equal to the growth rate of the macroscopical population Λ and that τ_i are identical independent random variables. So, the size of each sub-colony (n_i) evolves in time - again by neglecting the carrying capacity - according to:

$$n_i(t) = e^{\Lambda(t-\tau_i)}. \tag{2.3}$$

The size of the entire population of cells (N) is then a combination of the n_i single colonies and grows as follows:

$$N(t) = \sum_{i=1}^{N_0} n_i(t) = \sum_{i=1}^{N_0} e^{\Lambda(t-\tau_i)}. \tag{2.4}$$

Therefore, from eq. 2.2 and 2.4 it follows that the population lag time t_{lag} can be written in terms of the single cell adaptation time τ_i as:

$$t_{lag} = -\frac{1}{\Lambda} ln(\frac{1}{N_0} \sum_{i=1}^{N_0} e^{-\Lambda\tau_i}). \tag{2.5}$$

Notice how the population lag time encodes information on both the initial cell density N_0, the single cell lag times τ_i and the population growth rate Λ. Depending on the value of Λ, in eq 2.5, two different regimes can be identified.

If Λ assumes "small" values, $\Lambda \ll \frac{1}{\tau_i}$, the population lag time is equal to the average of the single cell first division times. Indeed, through asymptotic expansions of eq. 2.5 it follows:

$$t_{lag} \backsimeq \frac{1}{N_0} \sum_{i=1}^{N_0} \tau_i. \tag{2.6}$$

In this case, by increasing N_0, t_{lag} is driven by a Gaussian statistics and thus the value of t_{lag} is better defined, while its fluctuations decrease.

If we now consider Λ assuming "large" values ($\Lambda \gg \frac{1}{\tau_{min}}$), the term $\sum_{i=1}^{N_0} e^{-\Lambda \tau_i}$ is dominated by the minimum of τ_i for the saddle point method, leading to the following expression for t_{lag}:

$$t_{lag} \backsimeq \tau_{min} \equiv min\{\tau_1, ...\tau_{N_0}\}. \tag{2.7}$$

In other words, in this second case, the population lag time is described by the extreme value statistics. Thus, by increasing the sampling, namely increasing N_0, τ_{min} decreases resulting in a consequent monotonic decrease of both t_{lag} and its fluctuations. The way t_{lag} decreases with N_0 depends on the distribution of the τ_is.

Figure 2.4 summarises the above described model and the two possible scenarios for the population lag time derived from it.

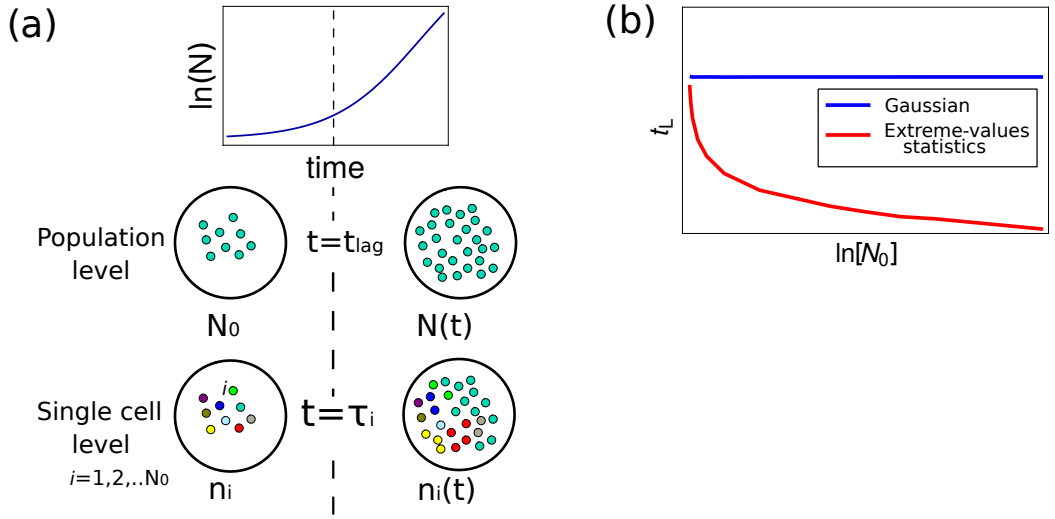

Fig. 2.4 Schematics for the lag time model. (a) Cartoon summarising the model. At the population level, a population of N_0 cells (light blue dots in the top-left circle) needs a time t_{lag} before starting to grow exponentially over time ($N(t)$) according to 2.2. The same dynamics is valid at the single-cell level, where each single cell i starts after a time τ_i to give birth to sub-colonies whose sizes grow exponentially over time $n_i(t)$. Each colour of the dots in the bottom circles represents a single cell (on the left) and its corresponding sub-colony (on the right). (b) Schematic representation of the two possible scenarios for the population lag time t_{lag} derived from eq. 2.5. The Gaussian statistic scenario (in blue) and the extreme-values one (in red). This second case has been obtained considering a uniform distribution for the τ_is.

In conclusion, the behaviour of the population lag time depends on the initial concentration of cells N_0, the population growth rate Λ and the distribution of single cell lag times τ_i, that can be considered as the first division time of each single cell. As stated above, the main distinction concerning the statistics driving the behaviour of t_{lag} *vs* N_0 is given by the value of Λ with respect to the single cell lag times. In Section 2.3 we test such model on populations of cancer cells, in order to understand which is the statistics that better interprets the data and is able to predict their behaviour.

2.2.3 Logistic growth model

If a population of cells in batch conditions does not undergo a change in the environment, it is expected to have zero-lag time and to grow exponentially up to the saturation level determined by the limited food availability. The simplest determin-

istic model able to describe such growth is the logistic growth introduced by P. F. Vehrulst [63], with the following differential equation :

$$\frac{\dot{N}}{N} = r(1 - \frac{N}{k}), \tag{2.8}$$

where N is the size of the population, r the intrinsic growth rate and k the carrying capacity.

Eq. 2.8 can be understood with the following simple heuristic argument [64]. Population size N at time t can be seen as a balance between births and deaths at that specific N. Being $b(N)$ and $d(N)$ the birth and death rates as function of N, we can write:

$$N = b(N) - d(N). \tag{2.9}$$

Then the population size changes over time according to:

$$\dot{N} = \frac{dN}{dt} = (b(N) - d(N))N. \tag{2.10}$$

The difference in parenthesis, $(b(N) - d(N))$, is the net increasing rate per unit time. By taking into account that the growth occurs in an environment with fixed carrying capacity k, the parenthesis of eq. 2.10 can be expressed as the following linear decreasing function of N:

$$b(N) - d(N) = r(1 - \frac{N}{k}), \tag{2.11}$$

where $r > 0$ is the intrinsic growth rate. By plugging such equation in eq. 2.10, the expression of eq. 2.8 is obtained. To better understand the meaning of the intrinsic growth rate, let's consider eq. 2.8 in the limit for small population sizes, i.e. $N \to 0$. In this case, equation 2.8 can be approximated to the first order and thus

$$\dot{N} = rN, \tag{2.12}$$

which represents an exponential growth. The growth rate r is thus the individual growth rate in an ideal situation of infinite nutrient availability.

Equation 2.8 can be analytically solved. By plugging N_0 as initial condition, the following solution is obtained:

$$N(t) = \frac{N_0 k e^{rt}}{k + N_0(e^{rt} - 1)}. \tag{2.13}$$

A representative cartoon of eq. 2.13, namely the logistic growth curve, expressed as the logarithm of N as a function of time is shown by Figure 2.5(a). Here, it is possible to notice that the population, starting from the initial size N_0 (initial condition), grows exponentially (linearly in the semi-logarithmic scale) up to the level of saturation given by the carrying capacity k.

The map of the punctual growth rate expressed as the derivative of the population size N over time (dN/dt) as a function of N is depicted in Figure 2.5 (b). The growth rate increases when N is far from the carrying capacity, then it reaches a maximum and finally decreases as the population size approaches k. When the size of the population increases, because of the overcrowding of the environment and the consequent shortage of nutrients, also the competition for nutrients among individuals increases. The net effect is a decrease of the overall population growth rate [58].

We are now interested in analysing how the exponential growth rate of such logistic dynamics may depend on N_0.
To this purpose, we firstly need to analytically define the exponential growth rate. Since there is not a standard definition, and assuming that the exponential phase is the regime of maximum speed of growth of a population, we consider as representative growth rate, the maximum slope of the curve. We consider the growth curves in the following form: $y = ln(\frac{N(t)}{N_0})$ and calculate the slope of the curve as $\lambda = \dot{y} = \frac{dy}{dt}$. By analysing λ, it follows that:

(i) From eq. 2.8 and 2.13 the time derivative of y in $t = 0$ is

$$\dot{y}|_{t=0} = r(1 - \frac{N_0}{k}) \tag{2.14}$$

(ii) For $t \to \infty$, $N(t) \to k$ and thus $\dot{y}|_{t=\infty} = 0$

(iii) From eq. 2.8 and 2.13 the second order time derivative of y is

$$\ddot{y} = -\frac{r^2 N}{k}(1 - \frac{N}{k}) \tag{2.15}$$

and it is always negative.

These three observations lead to the conclusion that, since the slope λ decreases when $t > 0$, its maximum value must be in $t = 0$. Thus, the maximum growth rate λ_{max} assumes the form expressed in eq 2.14. By calculating it for different values of N_0, the trend as function of the initial condition N_0 shown by Figure 2.5(c) is obtained. By assuming a logistic dynamics, the maximum growth rate does not depend on N_0 if the initial population size is sufficiently smaller than the carrying capacity k. However, when it assumes values closer to k, the growth rate decreases asymptotically to it (dashed vertical line in Figure 2.5(c)). The latter trend can be understood as discussed above: the more the overcrowding, the more the competition for the limited nutrients and the slower the population grows [58].

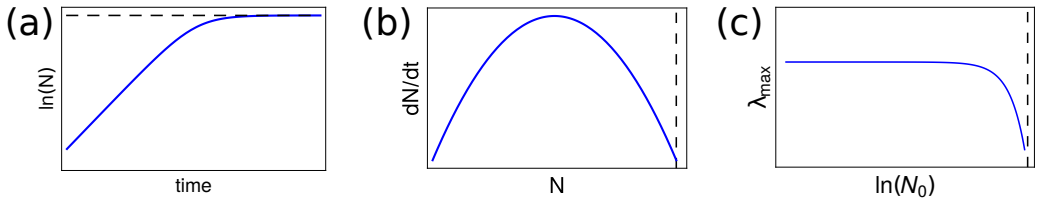

Fig. 2.5 Cartoons representative for the analysis of the logistic growth model. (a) The solution of eq. 2.8 plotted as the logarithm of the population size as a function of time. The initial condition is given by the initial population size N_0. (b) Mapping of the punctual growth rate (dN/dt) vs population size (N). (c) Maximum growth rate (λ_{max}) calculated as the derivative in time $t = 0$ of growth curves obtained by the model with different initial conditions plotted as function of N_0. In all the three plots the carrying capacity level (k) is emphasised by a dashed black line.

In summary, by considering a population of independent cells growing in a fixed amount of nutrients, we predict the following features. (To be noted: when talking about initial condition, we always refer to initial cell density, i.e. the initial size of the population.)

(i) Due to the limited availability of food, the maximum attainable cell concentration (carrying capacity) is constant independently on the initial condition;

(ii) The population undergoes a period of adaptation (lag time) when the quality of nutrients is changed;

(iii) Through a stochastic model based on single cell dynamics we can predict that such population lag time can be related to the single-cell first division times either through a Gaussian or an extreme values statistics. This statistics can in turn predict the population lag time dependence on the initial condition.

(iv) The simplest deterministic null-model for describing the dynamics of the population discarding the lag time, is a logistic growth where an exponential phase is followed by a saturation regime. In the latter, the population size reaches the carrying capacity. In this scenario, the exponential growth rate is not influenced by the initial condition for a wide range of N_0, namely as long as the initial size of the population is far from the carrying capacity. When the population size approaches the carrying capacity, the growth rate decreases.

Our purpose is now to test such hypothesis by measuring the population growth variables on an *in vitro* system consisting of populations of widely studied cancer cells. The next Section 2.3 is dedicated to this topic.

2.3 Experimental testing of theoretical predictions

In the previous section, we developed and discussed models for the dynamics and the initial condition dependence of a population of cells, beginning with simple assumptions inspired by microbiology and population dynamics. In this section we aim at testing such hypothesis on a real sistem, consisting of a population of widely studied cancer cells, called Jurkat clone E6.1, a human leukaemic cell line. In particular, we will experimentally investigate the macroscopical features of the population, i.e. how the three phases of growth of the population depend on its initial size. At the end of this section we will repeat the very same investigation on a further leukaemic cell line, i.e. K562, and we will compare the results.

To our purposes, we need highly controlled and statistically relevant experiments. With the first requirement, we intend to maintain growing conditions as much controlled as possible and only vary the initial population size among experiments. The quantification of the growth phases and their relation with the initial cell density would give us information on the macroscopical response of the system when

changing the initial state. Concerning the second requirement, we need statistically relevant data of cell growth in order to make comparisons with theoretical predictions. To address such issues, we developed an experimental protocol for cell growth inspired by techniques usually adopted in experimental physics, as discussed in the following Section.

2.3.1 Approaches inspired by experimental physics

The very first step in performing quantitative and systematic experiments relies in the definition of the right measurable observables that allow to answer a specific question. Once determined, the experimental design follows with the selection of the best methodology to adopt in order to extract the desired information from the system under exam. An essential (and obvious) point is that the measured quantities must vary according to the sample itself and not to the way measurements have been performed.

In our case, we are interested in testing the predictions of mathematical models on the behaviour of a population of cells. To achieve this, we intend to relate the representative variables of the three distinct growth phases of the population to its initial size. Thus, from an experimental point of view, our main scope is the generation of a robust set of growth curves with different initial conditions. It is important to highlight that our aim is to investigate the influence of a single variable (N_0) on the system, thus, we point towards the design of highly controlled experiments in which we fix all variables and change only N_0. This point is not trivial when performing wet-lab experiments, given the high sample-to-sample variability discussed at the beginning of this chapter. To overcome such issue, we followed the standard procedure in biology. It requires that, a part from the definition of experimental boundary conditions constant in each experiment (i.e. fixed growing conditions), for each initial condition one sets up three identical parallel growth cultures, often referred to as "technical replicates". The very same set of experiments is then repeated in a different day, with a different aliquote of cells. In this way independent measurements are available and the biological variability is taken into account.

The relevant measurable observables we intend to study are related to the growth of the population by ignoring single-cell features and are the initial size of the

population (N_0), the lag time (t_{lag}), the maximum exponential growth rate (λ_{max}) and the carrying capacity (k).

Population growth curves are usually expressed as the variation in time of the number of cells. Thus, for our purposes a precise and accurate determination of the number of cells becomes necessary as well as a sufficient time resolution among the measurements. In this way an accurate estimate of the growth parameters can be obtained, as discussed also in [61]. To achieve this, we need a method that allows us to (i) repeat several times the measurement of the same sample in order to increase the precision; (ii) make a high amount of measurements in a reasonable amount of time; (iii) obtain reproducible and unbiased results; (iv) be confident on the fact that the differences observed among samples are due to samples themselves and not to measuring procedures.

The most popular standard methods used in biology for counting cells are the estimation of the concentration through successive dilutions [65] and the direct manual cell counting of viable cells, usually performed through either an hemocytometer or an automated cell counter. Both methodologies present high criticalities for our scopes. Indeed, we need precise values of cell concentrations. However, with the first method, such information is lost since it estimates the number indirectly and does not measure it. The second method seems to better suit our requirements. The hemocytometer is a counting device firstly introduced for counting blood cells and then became popular for counting other cell types. Different versions of hemocytometers are available nowadays. The most known one consists of a thick glass microscope slide with a chamber formed by an indentation. Such chamber presents a grid of perpendicular lines: the area bounded by the lines and the depth of the chamber are known. By injecting a sample of cells in the chamber, it is then possible to count, by eyes under a microscope, the number of cells within the grid and to calculate their concentration (see Figure 2.6 for an example). However, such method presents many shortcomings such as (i) the lack of statistical robustness especially at low sample concentrations, (ii) the subjectivity of counts among users, (iii) the characteristic of being labour-intense and time consuming, especially in presence of high number of samples and high number of cells per sample. To overcome such issues, automatic cell counters has been introduced in biological labs. Such devices allow to automatically have unbiased and reproducible information on the number of viable cells. However, the most of them require the adding of chemical buffers to the samples in order to detect cells. Such feature causes the method to be

rather expensive if a high number of measurements of the same sample is required, a necessary issue for having a reliable statistics.

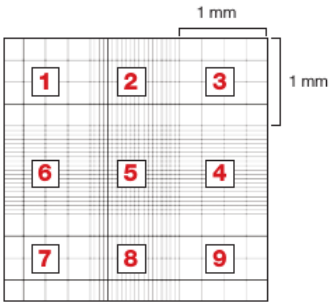

Fig. 2.6 Figure adapted from Bio-Rad protocol for cell counting [4]. Example of a counting grid called Neubauer counting grid. It is composed of nine sections of $1\,mm^2$ each indicated by a red number.

Driven by these reasons, we developed the following systematic method for counting cells. Note that all details are described in the following Section. We inject in a single-use counting chamber composed by a grid of 5 counting fields, a sample representative of the population to be measured. Phase-contrast micrographs of the 5 fields are taken under a common inverted microscope. The number of cells is then given by the application to the images of a custom-made image segmentation algorithm. The measurement is repeated several times in order to reduce the error on the measurement, which is then within $\sim 10-15\%$ of the measurement itself. Indeed, for each sample we count the number of cells within the 5 fields for at least 3 chambers, for a total of 15 measurement. By knowing the area of each square and the height of the chamber we can calculate the concentration of the population of cells by averaging such 15 measurements. In case of very low concentrations, the number of repeats is increased. This method allows to increase the number of measurements and thus its statistical relevance in a reasonable amount of time.

Such protocol results thus less time consuming than hemocytometer counting and allows to improve the automation by reducing the user bias. In this way, statistically relevant and highly reproducible direct measurements of the concentration of cells are obtained with acceptable relative experimental errors. Moreover, the use of an image segmentation algorithm allows us to have information not only on the number, but also on the size of the single cells, giving the possibility to investigate further morphological features and relate them to growth.

2.3.2 A systematic experimental approach revealed robust growth curves

We discuss here the details of the main steps of the experimental protocol we adopted for growing cells.

Each experiment began at time $t = 0$ h by taking a concentration N_0 of cells from a population growing exponentially in its standard growth medium. The N_0 cells were then transferred in a new plate containing a fixed amount of fresh standard growth medium and left free to grow by consuming nutrients. Notice that the quality of the growth medium before and after the seeding was the same. Such information will be relevant for the analysis of the lag times.

The entire growth - from time $t = 0$ h until the saturation phase - was monitored every 24 hours by following the cell counting protocol introduced in the previous section. Briefly, the concentration was calculated by taking micrographs of a representative sample of the population and by segmenting each image with a custom-made image processing algorithm able to extract number and size of alive cells.

In this way, 215 growth curves were obtained (Section 2.3.2) and analysed through fitting tools to extract the growth parameters, as discussed in Section 2.3.3.

The details on the cell line, the cell culturing method and the image segmentation algorithm are discussed in the following paragraphs.

Jurkat cell line

In this paragraph, we report a brief summary of the main features of the cell line we studied, together with an introduction on what a cell line is and which are its advantages with respect to primary cells.

Normal human cells have a limited ability to proliferate in culture, because they undergo senescence. This means that after many proliferations, the entire population is not able to grow anymore despite the availability of space and nutrients in the medium [66]. A cell line is instead a population of monoclonal immortalised cells, that is, identical cells that do not undergo senescence [67]. If in presence of

optimal medium and space, they are able to proliferate indefinitely giving an ideal homogeneous pool of identical cells useful for biochemical and genetic investigation.

The cells we investigate are Jurkat cells, clone E6.1. Jurkat cells are a leukaemic immortalised line of T lymphocytes. The original cell line (called JM) was derived in 1970s from a 14 years old boy with T cell acute lymphoblastic leukaemia. The subclone Jurkat E6.1 was developed in the 1980s [68] and became one of the most used Jurkat cell line, often considered as the standard one [69].

Jurkat cells live in suspension and tend to form clusters when growing in healthy conditions. The clusters are easily dissolvable by pipetting, so there is no need to add any chemical to the medium to dissolve them. In this way the population does not undergo a strong stress, keeps on growing and form clusters again after their disaggregation.

Jurkat cells have a great scientific interest due to their strong resemblance to the primary T lymphocytes [69]. For this reason they are widely used as a model cell for investigating T-cell expansion [70], T-cell leukaemia, T-cell signalling as well as the expression of receptors that can be susceptible to viral entry, like HIV [68]. A very relevant feature of Jurkat is their ability to produce large quantities of interleukin-2 [69], a growth factor that maintains the T-cell *in vitro* proliferation [71, 72].

Since we are interested in analysing the dependence of the growth on the initial density of the population, it is worth to mention that in literature there are some evidences showing the importance of cell density for maintaining certain kinds of T-cells in vitro [73] as well as some kinds of leukaemic ones [74].

As anticipated, we performed the very same kind of investigation on a further leukaemic cell line, the K562. We remind to the dedicated Section 2.3.9 for all the details.

Cell cultures

Growth experiments were performed by culturing Jurkat cells clone E6-1 in their optimal growth medium: RPMI-1640 supplied with 10% of fetal bovine serum, 5% of Penicillin/Streptomycin.

The growth medium contains a pH indicator, the phenol red, that lets the medium change colour according to its pH. (The medium turns from pink (basic) to yellow

(acid)). Its presence in the medium is considered very convenient by biologists because it helps checking quickly whether a culture is contaminated or not. In case of contamination indeed the environment becomes quickly very acid.

To begin each experiment, an initial concentration of cells (N_0) was extracted from a population growing exponentially in the above mentioned optimal growth medium and seeded in 6-well plates with a fixed volume (5 ml) of fresh optimal growth medium. For each initial concentration at least 3 wells were prepared. Cells were then incubated at 37 °C with 5% of CO_2 and left free to grow without adding any new medium for the entire duration of the experiment. The growth of the population was monitored by counting cells every 24 hours. To count cells two main features were necessary: firstly having a homogeneous sample, secondly distinguishing alive from dead cells. To address the first issue, since the tendence to form clusters, the volume with medium and cells was gently pipetted each time a counting measurement had to be performed. At this point, in order to count only alive cells, 30 μl of homogenised cells and medium were mixed with an equal volume of methylen blue (diluted 1% in Phosphate-Buffered Saline). The methylene blue only enters dead cells because of their permeable membranes. Dead cells are thus stained blue while alive ones remain colourless. At this point the homogenised mixture was transferred in 3 single-use hemocytometer chambers. The choice of repeating the measurement of the concentration of cells over three chambers pointed towards a better sampling of the population. In this way, possible irregularities due to sampling are averaged. For each chamber (corresponding to a volume of $10\,\mu l$), 5 phase contrast photos were taken by using an inverted microscope (Axiovert Zeiss) supplied with a $10x$ objective. In this way, at a given time, 15 micrographs of representative samples for each well were taken. Thanks to the methylen blue, the alive cells could be distinguished, as they appear as round-shaped bright spots (see Figure 2.8 (a) in the next paragraph for an example). Given a counting-chamber, the fact of counting over five fields allows to reduce the effects of eventual non-homogeneous distribution of the sample within the chamber.

Micrographs were then analysed through a Matlab-based custom made algorithm that implemented built-in functions in order to have information on the number and sizes of alive cells. Details of the algorithm are described in the next paragraph.

The procedure of counting was applied every 24 hours from the seeding time until the population reached the saturation phase.

Figure 2.7 summarises the experimental protocol used to obtain the growth curves.

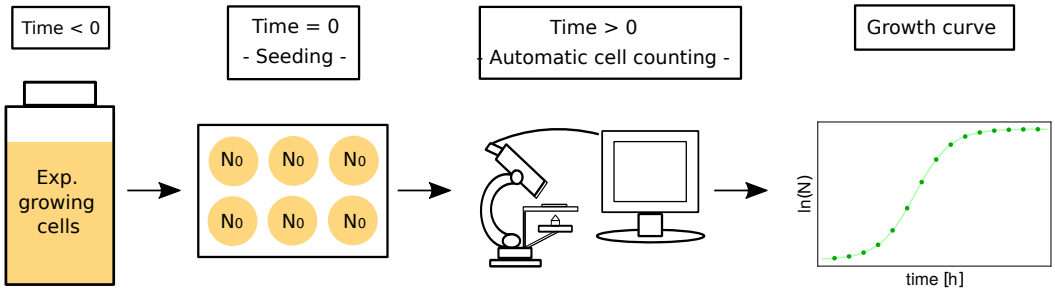

Fig. 2.7 Cartoon of the experimental procedure. From left to right, from a flask with cells growing exponentially in their standard growth medium, at time t = 0 h, samples of N_0 cells have been taken and moved into 6-well dishes supplied with the very same nutrient quality present in the flask. From time t=0 on the growth was monitored through automatic cell counting and growth curves similar to the one on the very right have been obtained.

Cell counting algorithm

In order to have information on cell number, we developed the Matlab code described in this paragraph. The code consists in two parts: the first one for the number detection, the second one for the size determination.

Our code uses Matlab built-in functions of image segmentation. Image segmentation is a general procedure that divides an image into its regions or objects of interest (cells in our case). The number of detected objects in our code is the result of a first thresholding on the pixel intensity (this allows a first detection of the objects in the image), followed by a *watershed* to distinguish touching objects and finally the selection of only round shaped cells.

The code requires as input the phase contrast micrographs taken through the procedure explained in the previous paragraph, see Figure 2.8 (a) for an example.

After a first contrast adjustment of the original image, a Gaussian filter is applied. This is an operator that blurs the image reducing the intensity of structures and it is useful especially for having a first reduction of noise without introducing any artefacts [75]. Afterwards the image is converted in gray scale first and then into a binary image by using the thresholding method. This method requires the definition of a threshold, empirically chosen in our case by analysing the intensity of both the background and of interesting objects. All pixels that have a value lower than the

threshold are labelled 0 in the output image, while those higher than the threshold are labelled 1. In this way the binary image is obtained. At this point we are interested in labelling the objects by using the Matlab function *bwlabel*. This is a function that returns a label matrix containing labels for the 8-connected objects found in the binary image. To be sure that through the binarization of the image, the structures we are interested in remained detectable, we apply three different image manipulation functions before the labelling. Firstly we fill holes by using *imfill* function. A hole is defined as a set of background pixels that cannot be reached by filling in the background from the edge of the image. Secondly we *open* and finally *close* the image with a disk as structuring element. The *opening* smoothes the contour of objects and eliminates thin protrusions while the *closing* smooths sections of contours, eliminates holes and fills eventual gaps in the edge. By applying them subsequently, a better refinement of single objects can be obtained [76]. At this point the function *bwlabel* is applied, therefore every detected object has a number as label (see Figure 2.8(b)). We put then a threshold on the size: all objects with less then 10 pixels are eliminated. The value 10 pixels was chosen by measuring the size of detected objects that cannot be considered as cells when looking at the original image. Despite we pipetted cells before counting to avoid clusters, some cells can eventually be close one to the other. In this case the procedure adopted up to now is not always able to separate two attached cells. For this reason, at this point of the code we apply the morphological watershed to the last labelled generated image. Morphological watershed transform takes the idea (and name) from the geographical concepts of *watershed* and *catchment basin*. In geography, in fact, a *watershed* is a ridge that divides adjacent drainage basins, while the *catchment basin* is the area draining into a reservoir [76]. The watershed transform applies these principles to the image it operates by treating it like a topographic map, with spatial coordinates and intensity as highness: dark pixels are low elevations, bright pixels are high elevations. The watershed transform aims at finding the lines that correspond to the ridges.

After applying the watershed and re-labelling the image, new objects are defined and the last part of the algorithm is performed in order to select only the round shaped elements.

To do this, we compute the circularity of cells. The circularity c is defined as

$$c = \frac{p^2}{4\pi A}, \tag{2.16}$$

where p is the perimeter and A is the area, calculated for each labelled object with the *regionprops* function of Matlab. The area is computed as the actual number of white pixels of the object. The perimeter instead is the distance around the boundary of the region of interest, namely the distance between each adjoining pair of pixels around the border of the object.

According to equation 2.16, an object is circular if $c = 1$. Among all detected objects we selected only those with a circularity $0.51 < c < 1.6$. We select this range after some analysis (not shown) on the circularities of objects corresponding to real cells in the original image. By discarding the objects that did not satisfied the circularity condition and re-labelling the resulted image, the algorithm is finally able to give the total number of cells present in the image. By knowing the volume of one analysed area, we are then able to calculate the cell density.

Figure 2.8(b) shows an original micrograph and the final result of the segmentation, where each cell is labelled with a progressive number. The maximum gives the total number of cells.

Concerning the detection of the sizes of cells, the algorithm follows similar steps. Some differences are present because the interest in this case is to have an information on the size of the object as much reliable as possible. For this purpose, slightly different operations were applied to the original image. The most relevant is the absence of the watershed step. The main drawback of the watershed is indeed to reduce the areas of the objects by defining confining edges. To overcome this issue we avoided the watershed and changed thresholds on minimal accepted size and circularity range. In this way a reliable information on the size of cells is obtained. However the size analysis of cells is not presented here.

To conclude, we would like to underline that the performance of the algorithm decreases when the population size is at the carrying capacity level. In this case, indeed, many cells are dead and the environment contains several residuals. Often, these cannot be easily distinguished from the cells even though the presence of the methylen blue and the application of a threshold on the size of single cells in the segmentation. For this reason, the absolute concentration of cells obtained in such regime fluctuates more than in the regimes with lower concentrations. Such information will be taken into account during the analysis of the growth parameters.

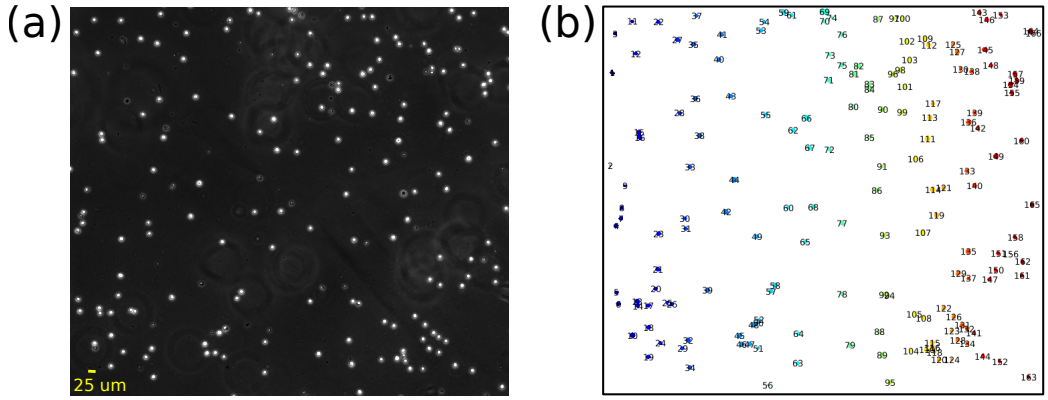

Fig. 2.8 (a) Micrograph of a representative sample of cell population. (b) Result of the image segmentation algorithm applied on image (a). Colours do not have a specific meaning, are simply used to distinguish single objects that are represented with a progressive number.

Robust growth curves

Through the systematic cell counting adopted in the experiments, highly reproducible growth curves have been obtained for each experiment at different initial concentrations (over four orders of magnitude of N_0), for a total of 215 experiments.

Figure 2.9 shows some examples of growth curves plotted as the logarithm of the concentration of cells versus time.

The first noteworthy observation concerns the measurement errors. In both plots, the error bars are given by the propagation of the semidispersion (i.e. the semidifference between the maximum and the minimum value) of the counts performed through the automatic counting. On average, experimental errors obtained through such procedures were about the $15-20\%$ of the measurement. However, two different cases must be distinguished concerning counting dispersion. First, when the concentration of cells is low - in general $N < 10^4$cell/ml - each field of the counting chamber may contain a very low number of cells (i.e. 0 or 1 cell). Thus, even though the counting has been repeated an higher number of times, the error bar often resulted exceeding the 15%. The second case instead focuses at the saturation. As discussed at the end of the previous paragraph, the population in this phase contains an high number of residuals (dead cells), that makes the performance of the algorithm decreasing. As a consequence, the average values on counting show bigger error bars.

By focusing now on Figure 2.9(a), the solid lines are the fit of the corresponding experimental data through the procedure described in Section 2.3.3. From this plot, it is already possible to notice differences among growth curves at different inoculum sizes. For instance, the curves seem to present variable lag times and growth rates (even at fixed inoculum, note the different slope of blue and cyan curves) and a constant level of saturation. Quantitative analysis of each variable will be discussed in the next sections. For an example of growth curve with emphasised parameters, see Figure 2.10.

The reproducibility of the experiments can instead be noted in Figure 2.9(b), where the three shades of blue are representative of identical replicates of the same condition, namely three wells identically prepared at the same time. For each fixed time, the semidispersion among the replicates resulted within the 10% of their mean value and a compatibility test gave a positive answer. Thus, it is possible to conclude that, for the time scales we are investigating, there is an high reproducibility among replicates identically prepared at the same time.

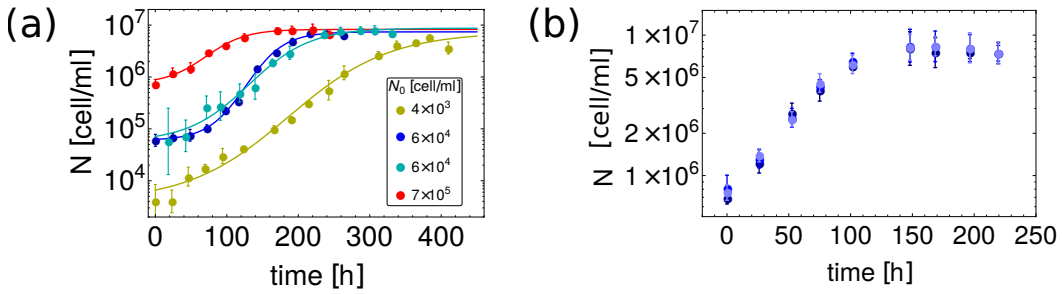

Fig. 2.9 Examples of experimental growth curves expressed as the logarithm of the number of cells in time. In (a) single replicates at different inoculi (N_0) are plotted. Each colour is relative to a different inoculum. Solid lines represent the fit of the corresponding data. N.B. Cyan and blue curves are two examples of the same inoculum. Since their inflection point coincide, a difference in the slope of the curves can be easily noted. In (b) three technical replicates are considered separately and identified by a shade of blue. Error bars are the propagation f the error over the counting.

2.3.3 The determination of the growth phases

Once the growth curves have been obtained, it is necessary to define an analytical tool to quantify the parameters representing the different phases of growth [2], namely the adaptation, the exponential and the saturation phases.

To our purpose we use curve fitting tools. We would like to stress that, inspired by a discussion raised by Baranyi and Roberts in [77], the functions we use here to fit the curves are mere tools adopted for quantifying the macroscopical variables interesting for our studies. They do not have any modelling purpose, namely they do not underlie any mechanistic hypothesis. In support of this claim, in Section 2.3.3 we show that the growth parameters obtained through different fitting curves are compatible among each other.

Typically, when considering growth curves, the logarithm of the number (or concentration) of cells normalised to their initial value N_0 is expressed as a function of time. The typical shape of such curves is sigmoidal, as it can be noticed in Figure 2.9. Several sigmoidal functions have been studied for describing bacterial growth which present features similar to those we are observing for Jurkat cells. Some of these curves are known as the logistic, Gompertz, Richards, Schnute [12, 54] or three-phase linear functions. Even though most of them underlie a model, we are here considering them simply for their mathematical form. Since these are usually not expressed in terms of biological interesting variables, Zwietering and coworkers [12], modified the equations of some of the mentioned sigmoidal functions to explicitly express the desired parameters. We observed that some of the above mentioned functions present either more than three parameters (Richards function, for example) or require further knowledges for the definition of the different phases (three-phase linear function). Since our aim is to simplify the procedure as much as possible, we discard all those functions that present at least one of such criticalities. Driven by these reasons, we fit the data with the modified logistic and Gompertz functions presented in [12] and reported in Table 2.1.

Modified logistic function	$ln(N/N_0) = \frac{A}{1+exp\left[\frac{4\lambda_{max}}{A}(t_{lag}-t)+2\right]}$
Modified Gompertz function	$ln(N/N_0) = Aexp\{-exp[\frac{\lambda_{max}e}{A}(t_{lag}-t)+]\}$

Table 2.1 Logistic and Gompertz modified equations [12] that explicitly express the growth parameters (A, λ_{max} and t_{lag}).

We considered then as fitting parameters: t_{lag}, λ_{max} and A representative for the lag, the exponential and the saturation phase respectively. They are emphasised in Figure 2.10, where the modified logistic curve (solid blue line) is used to fit the data (blue dots). t_{lag} is the intersection between the x-axis and the tangent to the sigmoidal

curve at the inflection point; λ_{max} is the slope of such tangent; A is given by the limit of N towards infinity.

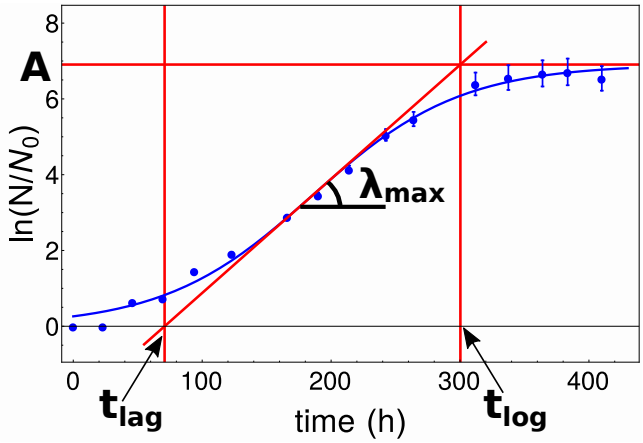

Fig. 2.10 Example of growth curve (blue dots) with the corresponding logistic fit (blue line) with lag time t_{lag}, maximum growth rate λ_{max} and saturation level A. The horizontal and tilted red lines emphasise the parameters A and λ_{max} respectively. The two vertical red lines represent the edges of the exponential phase, that correspond to times in t_{lag} and t_{log}. t_{log} is given by the intersection between the tangent at the inflection point and the level of saturation. The error bar on the data are obtained by the error propagation over $ln\frac{N}{N_0}$ by assuming $\sqrt{N}$ as error for the concentration N.

We tested the compatibility between the fitting parameters obtained by the two curves through a normal test - see Section 2.3.3 for details. Since the result of the test suggested a compatibility among the two fitting functions, we conclude that the values of the parameters are independent on the function used. We chose to fit the data with the modified logistic function. Its equation is here reported:

$$ln(N/N_0) = \frac{A}{1 + exp\left[\frac{4\lambda_{max}}{A}(t_{lag} - t) + 2\right]} \tag{2.17}$$

In Figure 2.10, a fourth parameter, t_{log}, is emphasised. We introduced it in order to define the entire exponential window. t_{log} has been obtained as the intersection between the tangent at the inflection point and the horizontal line with value A. The exponential phase may be relatively large, as in Figure 2.10 for instance, where it lasts almost half of the entire experiment. For this reason, we define here a second growth rate for quantifying the exponential phase. We call it λ and it is the slope

of the linear fit of the empirical data within the exponential window. Such value is, for construction, always smaller than λ_{max} and it may be interpreted as the average growth rate of the exponential phase.

In conclusion, for each growth curve, we calculated four parameters: t_{lag}, λ_{max}, A and λ. We fitted each growth curve with a modified logistic function to extract the first three parameters and to determine the exponential window to estimate the fourth. Concerning λ, we finally highlight that, in presence of narrow exponential windows, we manually extended its edges in order to calculate λ by fitting at least three data points.

Fitting parameters compatibility

For each growth curve, we performed a non-linear fit with both the modified logistic and the Gompertz functions (eq. Table 2.1) to extract t_{lag}, λ_{max} and A. The values of the corresponding parameters obtained with the two fit functions have then been compared through a z-test. Figure 2.11 depicts the distributions of the z-scores of each parameter. Vertical black lines represent the critical value for compatibility with a significance of 5% (two tails test, $z_{crit} = 1.96$). It is worth highlighting how, in all the three cases, the z-scores are mainly distributed towards values lower than the threshold. This allows then to conclude that the parameters obtained through the two fit functions are compatible and thus the use of either one function or the other does not affect the further investigation.

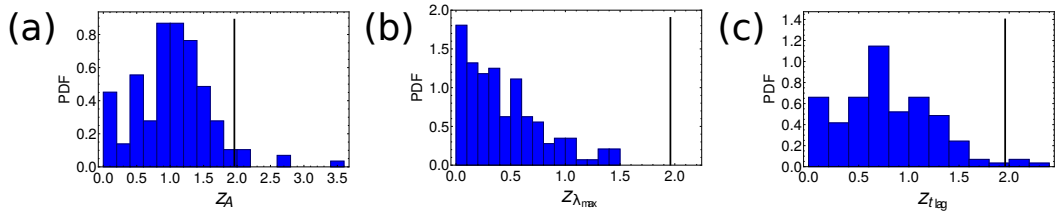

Fig. 2.11 z-score distributions for the comparisons among corresponding fitting values obtained through the modified logistic and the modified Gompertz functions. Distributions of the z-scores of (a) saturation level parameters A, (b) maximum growth rates λ_{max} and (c) lag times t_{lag}. Vertical lines represent the upper value for accepting a compatibility with significativity of 5%, $z_{crit} = 1.96$.

Normal test A normal test is performed by calculating the following score:

$$z = \frac{|var_1 - var_2|}{\sqrt{\sigma^2_{var_1} + \sigma^2_{var_2}}}$$

where var_1 and var_2 are the two variables to be compared, namely the fit parameters and σ_{var_1} and σ_{var_2} their error, i.e. the error obtained through the fit. If a two tail test is considered with a significancy of 5%, two variables can be assumed compatible if $z < z_{critical} = 1.96$ [78].

2.3.4 The independence of the carrying capacity on the initial condition

The saturation phase is characterised by the parameter A described by equation 2.17. From an analytical point of view, A is related to the concentration of cells achieved at an infinite time (carrying capacity, k) in the following way (from eq. 2.17):

$$ln(k) = A + ln(N_0) \tag{2.18}$$

Figure 2.12(a) shows the experimental results of A plotted as function of the logarithm of the initial concentration N_0. It is evident that the two variables are negatively correlated, as emphasised by the black line, with a slope compatible with -1. This, together with eq. 2.18, suggests that the carrying capacity is constant across the experiments and that it does not depend on the initial density of the population.

To support this hypothesis, we converted the values of A into the concentration of cells k by taking the exponential form of eq 2.18. The distribution of all the ks is plotted in Figure 2.12(b). A symmetrical distribution is shown for values smaller than $1.5 \cdot 10^7$cell/ml. The values in the right tail, i.e. $k > 1.5 \cdot 10^7$cell/ml, correspond to growth curves with a few data in the saturation phase. This led to a less precise estimation of the carrying capacity by the fit function and they can be discarded. The remaining data are compatible with a normal distribution (solid blue line in Figure 2.12(b)). A reduced $\chi^2 = 0.34$ is obtained from the fitting with a normal distribution. $\chi^2 < 1$ suggests that the data are normally distributed.

Thus, the mean and standard deviation of such distribution represent the value of the carrying capacity of our system together with its error: $k = (8 \pm 2) \cdot 10^6$cell/ml (Dashed line in figure 2.12(a) and solid black line in Figure 2.12(b)).

We can conclude that the experiments have been carried out in controlled conditions since, in agreement with our expectations, the carrying capacity is constant. This confirms that the carrying capacity is a feature of the growth medium.

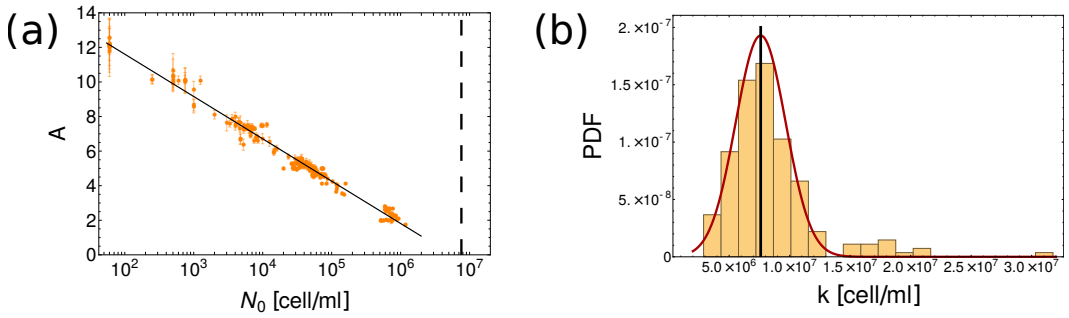

Fig. 2.12 (a) Parameter A vs $ln(N_0)$. Orange dots are experimental data, the solid black lines represents their linear fit. The dashed vertical line is located at the averaged carrying capacity k. (b) Distribution of the carrying capacities for each experiments, obtained by converting the parameter A through the exponential form of eq. 2.18. The solid black line emphasises $k = (8 \pm 2) \cdot 10^6$cell/ml, the average of the distribution for values smaller than $1.5 \cdot 10^7$cell/ml.

2.3.5 Population lag time and the extreme values statistics

In this Section we focus on the expectations on the lag time. We firstly remind the main theoretical results obtained in Section 2.2.2, secondly we present the empirical trend of lag times as function of the initial concentration for all the analysed growth curves and finally we suggest the mathematical model developed.

Model results for the lag time dependence on N_0

We report here a brief summary of the conclusions obtained in Section 2.2.2. By neglecting the carrying capacity, we assumed an initial population of N_0 independent cells. After a period of adaptation τ_i in which each single cell (i) does not replicate, cells start to exponentially proliferate, thus giving birth to new colonies. We developed a model for the population growth that takes into account the single cell

variability and relates the population lag time (t_{lag}) to the single cell first division times through the following expression (eq. 2.5):

$$t_{lag} = -\frac{1}{\Lambda} ln(\frac{1}{N_0} \sum_{i=1}^{N_0} e^{-\Lambda \tau_i}) \ . \tag{2.19}$$

Λ is the population growth rate, N_0 the initial population size and τ_i the single cell lag time. Depending on the value of Λ two different scenarios for t_{lag} are possible:

$$t_{lag} \backsimeq \begin{cases} \frac{1}{N_0} \sum_{i=1}^{N_0} \tau_i & for \Lambda \text{"}small\text{"}, \Lambda \ll \frac{1}{\tau_i} \\ \tau_{min} \equiv min\{\tau_1, ... \tau_{N_0}\} & for \Lambda \text{"}large\text{"}, \Lambda \gg \frac{1}{\tau_{min}} \end{cases} \tag{2.20}$$

Thus, the statistics of the population lag time may be driven either by a Gaussian (Λ "small") or an extreme value (Λ "large") statistics. In the first case, by increasing N_0, t_{lag} is expected to not change its mean value, while its fluctuations are expected to decrease according to the Central Limit Theorem. In the second case instead, through an increasing of the number of samplings (N_0), the minimum value of τ_i is expected to decrease as well as its fluctuations.

We are now interested in defining which is the right statistics able to describe our empirical data. To this purpose in the next section, we relate the experimental lag time to the initial condition N_0.

Experimental testing of the model

Figure 2.13(a) shows the lag time (for details on its estimation, see section 2.10) as function of the initial density N_0. The darkest red dots are mean values of the experimental data (lighter orange dots) within bins on N_0. Two noteworthy observations emerge from the plot: firstly the lag time is non-zero for a large range of N_0 and secondly, it decreases when increasing N_0.

The former is a non-trivial result since, as highlighted in Section 2.2.2, the composition of the growth medium used for growing cells before and after the seeding was the very same (see Section 2.3.2 for details). If the growing environment does not change, cells do not need to adapt and zero lag time is expected. Since this is not the case, we should suppose a difference between the two growth media, namely before and after the seeding. Thus, such difference must be only attributed

to cells themselves which may secrete chemicals during growth. If this is true, cells growing exponentially before the seeding are adapted to grow in an environment full of such secreted chemicals. When they are seeded in the new fresh growth medium, although the basal composition is the same, there is lack of cell products, thus cells need a certain time to adapt before beginning again to exponentially proliferate.

Let's focus now on the behaviour of the lag time. As can be noted by the darkest data in Figure 2.13(a), the mean lag time value decreases when increasing the initial concentration of cells, as well as its fluctuations, which are shown in Figure 2.13(b). Such observations suggest a dependence on the initial condition of the duration of the adaptation time.

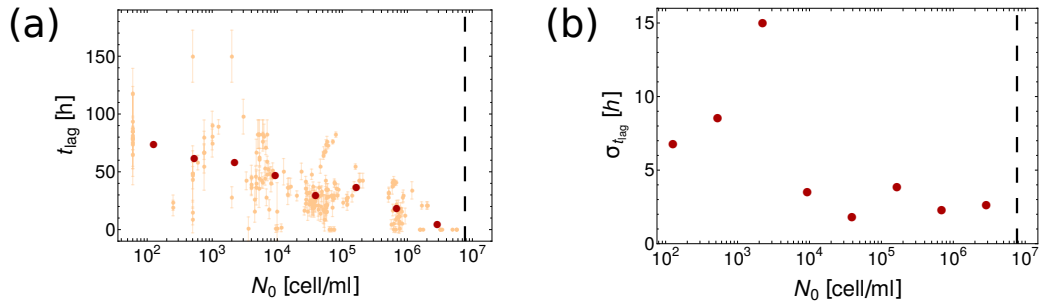

Fig. 2.13 (a) Experimental lag times as function of the initial seeding (N_0). Lighter colour dots are experimental data obtained through the fit. Their error bars are the error on the fit. The darkest red dots are the average of the smaller dots binned over N_0. (b) Standard deviation of the averaged values (dark red dots) of plot (a). In both plots the vertical dashed lines are located in correspondence of the carrying capacity value k.

The simplest trend that describes our experimental data is a linear negative correlation between the population lag time t_{lag} and the initial cell density N_0.

Considering now, the two possible theoretical scenarios summarised in the previous section, we can reasonably say to be in the regime of Λ "large", with the growth rate of the population $\Lambda \gg \frac{1}{\tau_{min}}$. Thus, the extreme values statistics predicts a decreasing trend of the lag time when increasing N_0, which is exactly what our experimental data show.

The decreasing trend of the lag time is driven by the distribution of the single cell lag times. In absence of further information it is not-trivial to estimate such distributions. In the following paragraphs we suggest two different approaches to address such issue. The former is a computational approach through which possible

τ_i distributions are assumed and the trend for the average value of t_{lag} is derived. However, in order to choose the best distribution that fits our data, further information is necessary. The second approach is instead experimental and it implies the direct measurement of the τ_is: we present here the methodology we intend to use in the future in this direction.

Possible single cell first division time distributions

By following the Section 2.2.2 focusing on the derivation of the population lag time through a stochastic model, it is worth noticing that:

- if $\tau_i s$ are independent and identical distributed random variables with probability density $p(\tau)$
- if the argument of the logarithm in eq. 2.5, $x \equiv \frac{1}{N_0}\sum_{i=1}^{N_0} e^{-\Lambda\tau_i}$, can be considered a Gaussian random variable with mean $\mu = \int p(\tau)e^{-\Lambda\tau}d\tau$ and variance σ^2/N_0

then, the lag time of the population presents the following distribution:

$$p(t_{lag}) = \frac{\Lambda e^{-\Lambda t_{lag}}}{\sqrt{2\pi\sigma^2/N_0}} e^{-\frac{(e^{-\Lambda t_{lag}}-\mu)^2}{2\sigma^2/N_0}} \quad (t_{lag\geq 0}) \tag{2.21}$$

We discuss here some cases in the "large" Λ limit by assuming different distribution for τ_is and deriving the expected trends for t_{lag} as a function of N_0.

Exponential distribution

If τ_i are identical distributed random variables taken from an exponential distribution

$$p(\tau) = re^{-r\tau} \tag{2.22}$$

it follows that

$$Prob(t_{lag} \geq x) = \prod_{=1}^{N_0} Prob(\tau_i \geq x) = \prod_{=1}^{N_0} e^{-x\tau_i} = e^{-N_0 xr} \tag{2.23}$$

So, the probability of having t_{lag} smaller than x is $Prob(t_{lag} \leq x) = 1 - e^{N_0 xr}$

and the distribution of t_{lag} is:

$$q(t_{lag}) = N_0 r e^{-N_0 r t_{lag}} \tag{2.24}$$

This implies that both the fluctuations and the average of t_{lag} decrease as $1/N_0$, resulting in a constant *CV*.

Uniform distribution

If $p(\tau)$ is uniform on $[0,T]$, then:

$$Prob(t_{lag} \leq x) = 1 - (1 - \frac{x}{T})^{N_0} \tag{2.25}$$

$$q(t_{lag}) = \frac{N_0}{T}(1 - \frac{x}{T})^{N_0 - 1} \tag{2.26}$$

and

$< t_{lag} >= \frac{T}{N_0+1}$ and $CV = \sqrt{\frac{N_0}{N_0+2}} \simeq const.$

If the carrying capacity is included in the system, then:

$$e^{-\Lambda t_{lag}} \simeq -\frac{1}{\Lambda} ln < e^{-\Lambda \tau} > \equiv \int_0^\infty e^{-\Lambda \tau} p(\tau) d\tau \;\; (N_0 \gg 1) \tag{2.27}$$

i.e., as N_0 increases towards large values, the lag time should converge to

$$t_{lag}^{\infty} \simeq -\frac{1}{\Lambda} ln < e^{-\Lambda \tau} > \tag{2.28}$$

For example, if $p(\tau) = re^{-r\tau}$, then:

$$t_{lag}^{\infty} \simeq \frac{1}{\Lambda} ln \frac{r - \Lambda}{r} \tag{2.29}$$

In order to compare the analytical results here derived with the empirical scenario, further information on the decreasing trend of the lag time with respect to N_0 would be necessary. Since the actual lack of such information, we suggest in the following paragraph an experimental and direct way to derive the τ_i distributions.

Future perspective: measurements of single cells first division times

In this paragraph we present an experimental setup useful for measuring the distributions of single cell first division times. In particular, we aim at directly measuring the single cell lag times, considered as the time that each cell needs to replicate the first time. To this purpose we apply the very same experimental conditions of the cell growth experiments, but in a setup that allows us to follow single cells over time. Indeed, instead of seeding cells in 6-well plates, we use microfluidic devices and take micrographs of desired fields of the micro-wells through bright field time-lapse microscopy.

We began setting up the experiments at Imperial College in London together with Dr. Francesca Ceroni. We present here the methodology and the measurements we intend to carry out in the future.

Time-lapse microscopy can be considered as one of the best techniques for measuring single cell variables since it allows to avoid experimental errors arised by traditional microbiology methods adopted to indirectly monitor cell growth [79]. Through time-lapse microscopy it is possible to follow and count in time the cells originating from each initial seeded single cell [1].

Since Jurkat cells live in suspension, we use microfluidic devices (Onix Plates) with low ceiling in order to trap single cells and avoid on one hand the formation of $3D$ clusters, on the other hand the lost of the focus during the experiment. Furthermore, the microscope is supplied with a programmable motorised stage that allows to take micrographs over time of different areas of the plate, not necessarily continuous in space. Thanks to this feature we are able to sample an high number of initial single cells, even in conditions of low N_0s.

As mentioned above, we intend to replicate the very same experiment of cell growth, but in a different plate. Thus, the only difference from the standard experimental protocol we described in Section 2.3.2, is the quantity of medium injected in the growing well. In the present case, we inject in the microfluidic device $0.1\,\mu l$ of medium with the desired concentration of cells. Figure 2.14(a) shows the schematic of the microfluidic device. Once loaded the mixture composed of cells and medium in the loading chambers (emphasised by the red number 1 in Figure 2.14(a)), the fluid moves for capillarity to the monitored chamber (number 2) until the pressure of the connected empty well (number 3) and the loading channel is balanced. Once the

flux stops the measurements can be performed. To this purpose, the plate is inserted on the motorised stage of the microscope which is supplied with an incubator: in this way cells may live in their optimal condition at $37^{\circ}C$ for the entire duration of the experiment.

Since the value of the single cell first division time is unknown, we take measurements each 15 minutes for 3 consecutive days: at each time, a micrograph of the selected field of the plate is taken. In this way the cells are followed in time. Figure 2.14(b) shows an example of micrograph: the red circles emphasise the single cells. By adapting to such images the segmentation algorithm described in Section 2.3.2 supplied with a further section for tracking single cell, it would be then possible to follow the cells and monitor the time needed before the first division.

By performing such experiments with different N_0s we point towards the determination of the distribution of the single cell first division times that lead the observed trend for the population lag time.

Before reaching the final result, we still need to (i) optimise the experimental protocol for obtaining the best images over time, (ii) adapt the existing image segmentation algorithm to this new purpose. Once these issues will be optimised, we will be able to obtain the τ_i distributions.

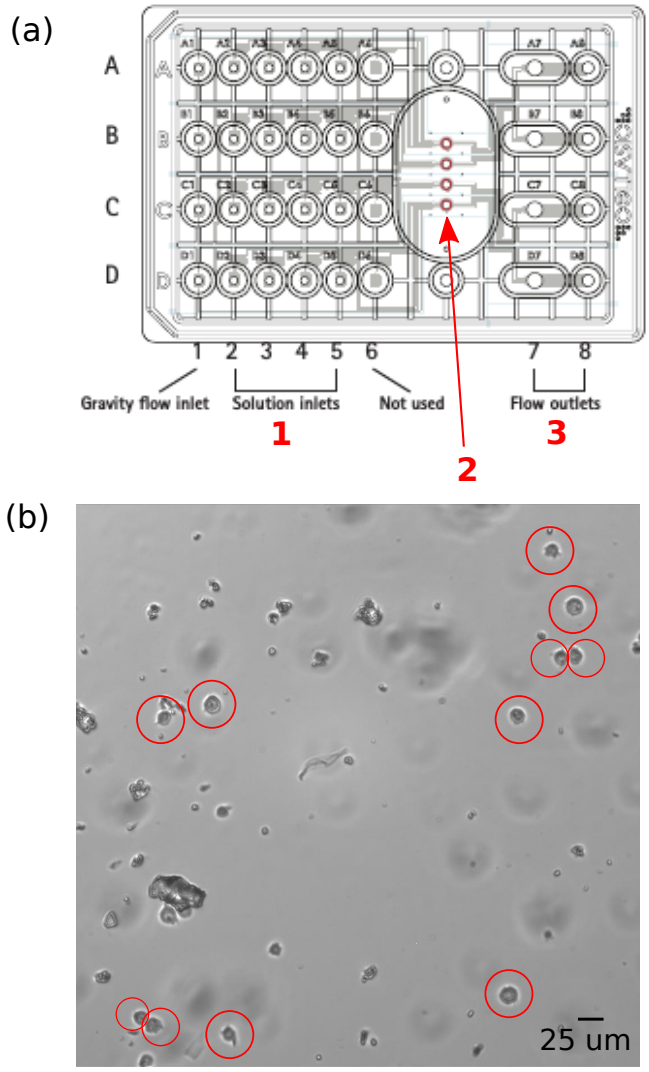

Fig. 2.14 (a) Schematic representation of the microfluidic device used to follow single cells. Adapted from User guide of CellASIC® ONIX M04L-03 Microfluidic Plate. Red numbers represent: 1. The wells where the solution of cells+growth medium is loaded; 2. the effective growth chambers; 3. the outflux chambers. (b) Bright field micrograph at time $t = 0$. Red circles emphazie single cells.

2.3.6 Exponential growth rate and initial conditions

In this paragraph we move the attention towards the exponential phase of growth, in which the population of cells grows with a constant rate. Our aim is to test the logistic growth hypothesis by investigating the dependence on the initial conditions

of the exponential growth rate. Part of the data here reported are an integration of those presented in [38].

The experimental growth rate depends on N_0

Figure 2.15(a-d) show the experimental results. Two sets of data are plotted as function of the initial population density N_0: the maximum growth rate λ_{max} (Figure 2.15(a)) and the average exponential growth rate λ (2.15(c)). Firstly it is important to notice that, as expected, the latter set of data is systematically shifted towards lower values with respect to the former. Secondly, the two growth rates follow the same trend when related to N_0. This suggests that, independently on the definition of the growth rate we consider, a dependence on N_0 rises. Such behaviour can be divided into three different parts: the growth rate remains roughly constant with small fluctuations for values of N_0 sufficiently far from the carrying capacity, it increases together with its fluctuations for intermediate N_0 (roughly between 10^4 and 10^5 cell/ml) and finally decreases when the initial density reaches the carrying capacity (dashed vertical line). The last decrease is expected and it is due to the presence of a fixed amount of nutrients. Here, since the population is seeded at a concentration close to the carrying capacity, cells compete for nutrients and the population fitness decreases.

By binning the data over N_0, we calculated the fluctuations of the growth rates and plot their values in Figure 2.15(b) and (d). It is worth noticing that the fluctuations increase when increasing the population size.

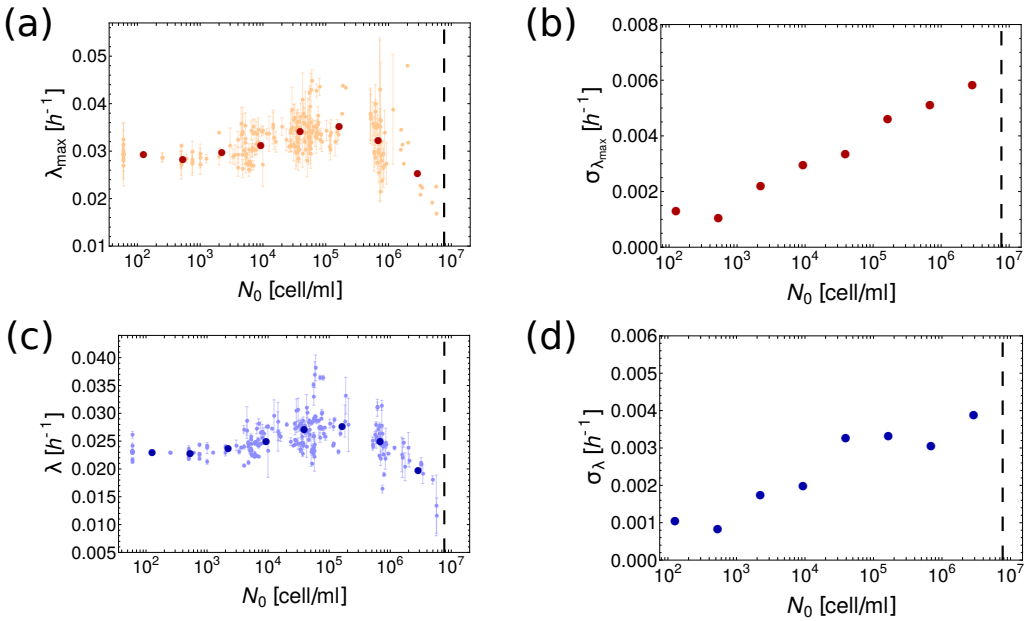

Fig. 2.15 (a) and (c) Exponential growth rates (λ_{max} and λ respectively) as a function of the initial cell density (N_0). Brightest dots represent all the experiments. Error bars are the error of the fits performed to calculate λ and λ_{max}. The darkest dots are the average of the brightest ones binned over N_0. (b) and (d) Standard deviations of the data of (a) and (c) for the growth rates by binning over N_0. The dashed black line present in all the plots is located at the value of N_0 corresponding to the carrying capacity k.

The trend shown for intermediate values of N_0 is not trivial if one thinks that usually at the time the exponential growth rate is evaluated, the population has already undergone several replications and thus, in principle, no dependence on the initial condition is expected. The parameter that most likely may be expected to be related to the growth rate is the size of the population at the time the growth rate is considered ($N^* = N(t)$). By ignoring the growth curves with inoculi close to the carrying capacity ($N_0 > 10^6$), we verified that the dependence we noticed is on N_0 and not on N^* by fixing one of such variables at a time. To this purpose, we consider for both growth rates, the relative N^* and N_0 and investigated the relation between the growth rate and N^* when fixing N_0. N^* has been calculated in the following way. When considering λ_{max}, N^* is the value of the fit function at the inflection point converted in terms of cell concentration, that is $N_{A/2} = N_0 \cdot e^{A/2}$. When considering λ, instead, we evaluated N^* for each experimental growth curve as the average density of the population ($N(t)$) within the entire exponential phase, namely for $t_{lag} < t < t_{log}$. It has been computed by exploiting the line fitting the data within

the exponential phase. N^* has been obtained as the average of $N(t_{lag})$ and $N(t_{log})$ belonging to such line.

Figure 2.16(a) shows the results for the case of λ_{max} vs $N_{A/2}$ and Figure 2.16(b) depicts λ vs N^*: each plot represents a bin on N_0. The binning on N_0 corresponds to the removal of a dependence on it. Except for some data in the last N_0 bin (at the border of the regime where the growth rate λ decreases with N_0), each set of data results compatible with a straight line with zero angular coefficient suggesting thus the absence of a dependence of λ on N^* and of λ_{max} on $N_{A/2}$. While concerning the case of λ_{max} such result should have been expected given the constancy of the carrying capacity k, it was not trivial for the case of λ. However, the obtained result suggests that N_0 is the only variable to resolve the growth rate.

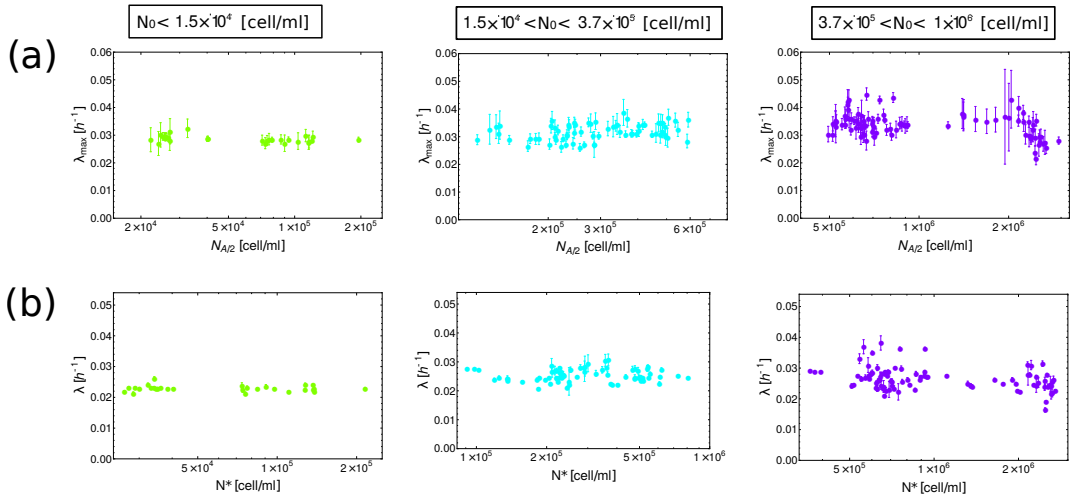

Fig. 2.16 (a) and (b) Three bins on N_0. For each bin, the exponential growth rates λ_{max} in (a) and λ in (b) are plotted as a function of $N_{A/2}$ and N^* respectively.

Disagreement with the logistic model

In order to compare the experimental data with the null logistic model, we evaluated the theoretical λ_{max} as described in Section 2.2.3, namely as the tangent to the logistic growth curve $N(t)$ (eq. 2.13, reported here below) at time $t = 0$.

$$N(t) = \frac{N_0 k e^{rt}}{k + N_0(e^{rt} - 1)}$$

The parameters r (the specific growth rate) and k (the carrying capacity) have been chosen in the following way: r equal to the mean of the growth rates for $N_0 < 3 \cdot 10^3$cell/ml, i.e. the values corresponding to the constant regime; k equal to the experimental k (see Section 2.3.4). We only compare to the experimental data the predictions over λ_{max} and not over λ. This is because, to evaluate the exponential rate λ, it is necessary to identify the exponential time window through the use of the tangent to the curve at the inflection point. However, since the second derivative of the logistic curve is always positive, no inflection point is present and then λ cannot be calculated.

The comparison between the experimental data (reddish dots) and theoretical prediction (blue line) is shown in Figure 2.17. The data suggest that the logistic dynamics is sufficient to describe only the regimes at low and high inoculum sizes, when the population is in conditions of infinite (λ_{max} constant with N_0) and limiting (λ_{max} decreasing with increasing N_0) nutrient quantity respectively. On the other hand, the logistic dynamics does not describe the λ_{max} increasing for intermediate seedings. In this case, a small increase in the density of the initial population, gives benefits to the overall population during the subsequent exponential phase, thus enabling it to grow faster.

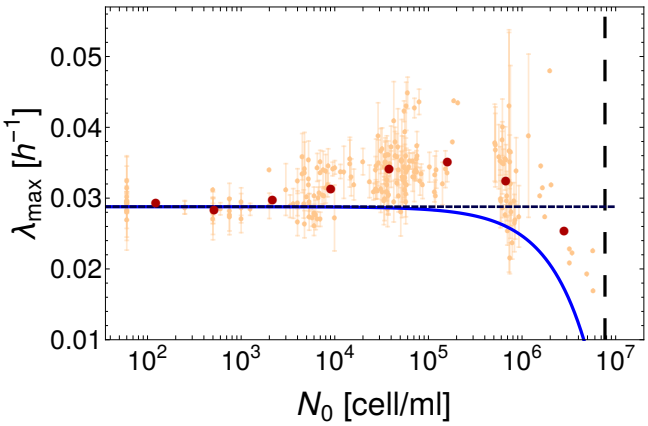

Fig. 2.17 Same plot as Figure 2.15. The bright dots represent the experimental data for λ_{max} as a function of N_0 with the error bars representing the error of the fit. The darkest dots are the average of the brightest ones binned over N_0. Continuous blue line represents the theoretical λ_{max} according to the logistic growth model. Dashed horizontal blue line emphasises the offset value r. Dashed vertical black line is set at the level of the carrying capacity k.

To better understand such scenario, a density dependent growth may occur within the populations under exam. We discuss such topic in the next paragraph.

Density dependent growth

The easiest correction that may be made to the logistic equation in order to take into account a density dependence in the growth, is the adding of a further term $(\frac{N}{k})^{\alpha}$. This leads to the following equation:

$$\frac{\dot{N}}{N} = r(1-\frac{N}{k})(\frac{N}{k})^{\alpha} \tag{2.30}$$

where, as in Section 2.2.3, N is the size of the population, r the intrinsic growth rate and k the carrying capacity. The parameter $\alpha \geq 0$ refers to the density-dependence strength. Note that the case $\alpha = 0$ corresponds to the standard logistic growth with fixed carrying capacity (eq. 2.8). $\alpha > 0$, instead, causes a decreasing of the fitness for low population sizes. Such density dependence effect is known as weak Allee effect [80].

Allee effect has been firstly introduced in 1932 by Allee [81], and it is nowadays widely suggested in ecology to explain deviations from a simple logistic growth for populations of small size. A comparison a simple logistic growth (blue line) and the Allee effects (green and yellow lines) is shown in Figure 2.18(a), where the instantaneous growth rate is plotted as function of the population size (N). We can notice that for the Allee effect, high values of growth rate are only present after the overcoming of a threshold for the population size. On the other hand, the logistic growth shows positive growth rate already for small values of N. Two kinds of Allee effect are known: strong and weak. The former, represented by the yellow line in Figure 2.18(a), is described by the following equation [82]:

$$\frac{\dot{N}}{N} = r(1-\frac{N}{k})(\frac{N}{N_c}) \tag{2.31}$$

Where N, r and k have the same meaning of eq. 2.30, and N_c is the Allee threshold, i.e. the minimum size of the population necessary to have a positive growth. Once such thresholds is overpassed, the colony is able to grow.

The weak Allee effect, whose differential equation is eq. 2.30, shows positive growth rates even at low population sizes, but with smaller values before the threshold. In Figure 2.18(a), the elbow in the green curve is related to the parameter α of eq. 2.30. The stronger the density effect is, the higher is the value of α and thus the elbow is shifted towards higher values. Figure 2.18(b) shows the instantaneous growth rate as a function of the size of the population, where the parameter α increases in the direction of the arrow.

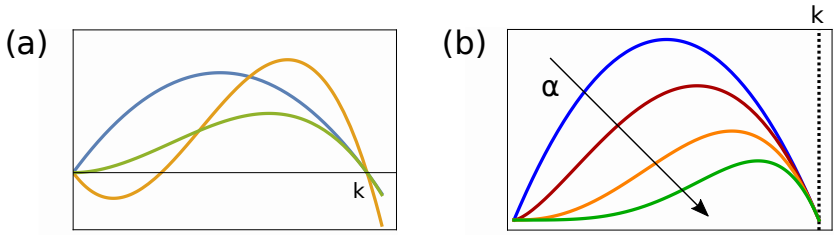

Fig. 2.18 Representations of the growth rate (dN/dt) of a population of size N as function of its size. (a) Trends for logistic growth (blue solid line), weak (green) and strong (yellow) Allee effect. The value k corresponds to the carrying capacity. (b) Examples of weak Allee effects with different values of parameter α that increases according to the arrow. The blue curve is the logistic case with $\alpha = 0$. The dashed black line represents the carrying capacity k.

Data are in agreement with a weak Allee effect dynamics

We tested the hypothesis of the presence of a weak Allee effect within the experimental data. To this purpose, we fit each growth curve with eq. 2.30 whose free parameters are r, k and α and investigated the values of α. The fit has been performed over the following quantities:

$$\frac{dN}{dt} = \frac{N(t+1) - N(t)}{(t+1) - t} \tag{2.32}$$

$$N = \frac{N(t+1) + N(t)}{2} \tag{2.33}$$

where $N(t)$ and $N(t+1)$ are the concentration of cells at two subsequent times, t and $t+1$ respectively.

Figure 2.19(a) shows an example of a sub-set of the experimental growth data (cyan dots) together with their fit function (red) and the logistic case (blue). Figure

2.19(b) depicts instead the distribution of positive α obtained through fitting each single curve. The plots suggest that a weak Allee dynamics is reasonable to describe the data.

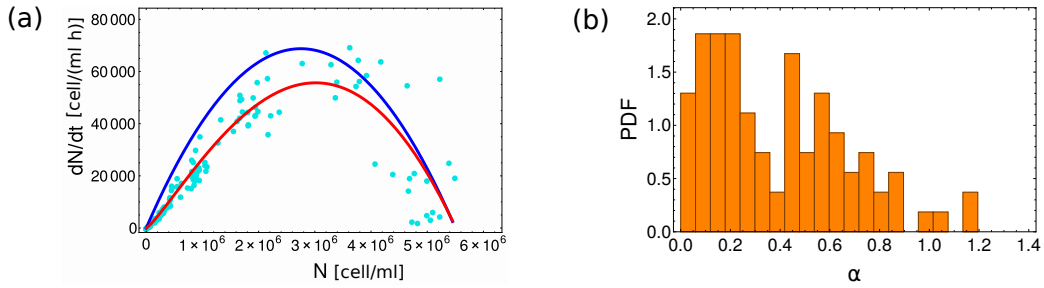

Fig. 2.19 (a) Example of mapping of the punctual growth rate ($\frac{dN}{dt}$) vs the size of the population (N). Cyan dots are experimental data. A sub-set of growth curves is here considered. The red line is their fit with a weak Allee effect function; blue line represents a logistic growth. (b) Distribution of the positive values of α obtained by fitting the experimental growth curves with a weak Allee effect function, eq. 2.30.

The derivation of theoretical growth rates

In light of the previous paragraph, we aim now to compare the trends of the exponential growth rates obtained through the experiments and those of the "theoretical" growth rates evaluated from the growth curves that follow a density-dependent dynamics. To this purpose, we numerically integrate eq. 2.30 in order to obtain the growth curves $N(t)$ with initial condition N_0 at $t = 0$.

The parameters have been chosen as follows. $r = 1$, to match the theoretical with the experimental ones; k is a random number picked from the normal distribution that fits the empirical k (see section 2.3.4 for details) and α is taken from a lognormal distribution that fits the experimental values of $\alpha > 0$ described above. Such distribution is centred in $\alpha = 0.1$ with a constant variance of 30%.

The growth curve so obtained has been then considered in the following form:

$$y = ln(\frac{N(t)}{N_0}) \tag{2.34}$$

We are interested in comparing the two experimental growth rates, λ and λ_{max}.

Since no analytical expression can be derived for $N(t)$, λ has been numerically evaluated. We defined it analogously to its experimental definition (see Section

2.3.3), that is the slope of the straight line connecting the edges of the exponential phase, described by eq. 2.35. Once identified the inflection point, the tangent to the growth curve at this point is used to obtain the edges of the exponential phase (t_{lag} and t_{log}). Figure 2.20 summarises the procedure.

$$\lambda = \frac{ln(\frac{N_{log}}{N_0}) - ln(\frac{N_{lag}}{N_0})}{t_{log} - t_{lag}} \tag{2.35}$$

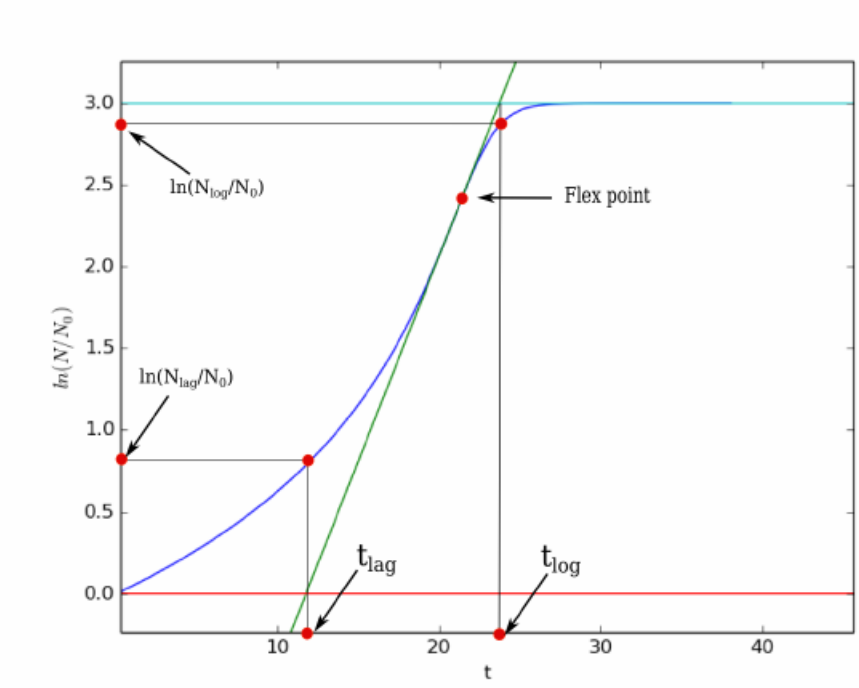

Fig. 2.20 Scheme of the procedure through which t_{lag} and t_{log} are defined starting from the growth curves obtained through the weak Allee effect dynamics.

However, such definition of λ (eq. 2.35) holds until an inflexion point exists, i.e. when $N_0 < \frac{\alpha}{\alpha+1}k$ (see the end of the paragraph for mathematical details on this topic). When $N_0 > \frac{\alpha}{\alpha+1}k$, a new definition of λ is required. To this purpose, we observe that N_{log} does not change when increasing N_0, even though t_{log} does. So we defined the exponential growth rate as

$$\lambda = \frac{1}{t_{log}} ln(\frac{N_{log}}{N_0}) \tag{2.36}$$

If $N_0 > N_{log}$ the growth rate is assumed to be null.

To summarise:

$$\begin{cases} \lambda = \frac{ln(\frac{N_{log}}{N_0}) - ln(\frac{N_{lag}}{N_0})}{t_{log} - t_{lag}} & for N_0 < \frac{\alpha}{\alpha+1}k \\ \lambda = \frac{1}{t_{log}} ln(\frac{N_{log}}{N_0}) & for \frac{\alpha}{\alpha+1}k < N_0 < N_{log} \\ \lambda = 0 & otherwise \end{cases} \tag{2.37}$$

Inflection point of y

Consider y as in eq. 2.34. Its first and second time derivatives are:

$$\dot{y} = \frac{\dot{N}}{N} = r(1 - \frac{N}{k})(\frac{N}{k})^{\alpha} \tag{2.38}$$

$$\ddot{y} = \frac{r^2 N}{k}(1 - \frac{N}{k})(\frac{N}{k})^{\alpha}[\alpha(1 - \frac{N}{k})(\frac{N}{k})^{\alpha-1} - (\frac{N}{k})^{\alpha}] \tag{2.39}$$

respectively. From eq. 2.39 it follows that y admits an inflection point in

$$N_f = \frac{k\alpha}{\alpha + 1}. \tag{2.40}$$

Such point corresponds to the point of maximum slope, so that:

$$\lambda_{max} = \dot{y}(N_f) = r(1 - \frac{\alpha}{1+\alpha})(\frac{\alpha}{1+\alpha})^{\alpha}. \tag{2.41}$$

Thus, concerning the maximum growth rate λ_{max}, we obtained the following expressions:

$$\lambda_{max} = \begin{cases} r(1 - \frac{\alpha}{1+\alpha})(\frac{\alpha}{1+\alpha})^{\alpha} & for N_0 < \frac{\alpha}{\alpha+1}k\,(\exists\, inflection\, point) \\ r(1 - \frac{N_0}{k})(\frac{N_0}{k})^{\alpha} & for N_0 > \frac{\alpha}{\alpha+1}k\,(no\, inflection\, point) \end{cases} \tag{2.42}$$

Aaccording to the derivations above, if $\alpha \to 1$ (our case), then $\lambda_{max} \to \frac{r}{4}$, thus it remains constant if N_0 is changed. Driven by this, we will not compare the experimental and theoretical maximum growth rate but we will only focus on the exponential growth rate λ.

Comparison of model prediction and experimental data

The procedure described above has been applied to growth curves within the same range of N_0 as for the experiments.

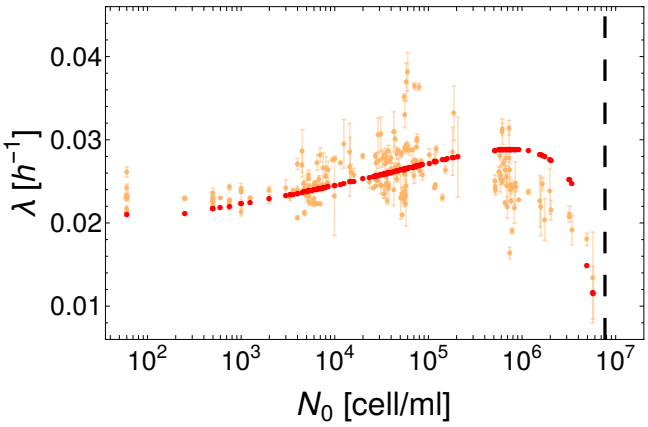

Fig. 2.21 Exponential growth rate λ vs initial population density N_0. Light small dots are the experimental data, the error bars represent the error on the fit parameter. Darkest red dots are λs evaluated from the theoretical growth curves. Dashed vertical line is represents the carrying capacity k.

Figure 2.21 shows the exponential growth rates λ plotted as function of the initial condition for both the experimental data (orange small dots) and the model (red bigger dots). As it can be noticed, a qualitative agreement between the two series of data rises. Even though the maximum of the two data sets are shifted, the increasing trend of the exponential growth rate together with the increasing of initial cell density is reproduced. The result suggests the presence of a mechanism of density dependence that drives the growth of population of cells under study.

Moreover, such results suggest the existence of a cooperative behaviour among the cells. If a population is seeded at the optimal concentration, the result of the cooperation among the single cells increases the fitness of the entire population once they are in the exponential regime of proliferation. In the next Section we discuss this topic by considering two possible ways through which such cooperation may occur.

2.3.7 Cell-to-cell communication

In the previous Section 2.3.6 we have shown how the hypothesis of a logistic model fails in predicting the dependence of the growth rate on the initial condition of a widely studied population of cancer cells. This raised the hypothesis - later on experimentally tested - that the populations of cells under exam may grow in a density dependent manner. Indeed, it has been found that if cells are seeded within a specific range of concentration, their fitness during the asymptotic regime of growth - i.e. the exponential phase - is higher than for different initial concentrations. Such behaviour suggests the presence of a cooperativity among single cells that may start with the beginning of the growth of the population and is maintained up to the total consuming of the available nutrients.

With regards to this, a first observation concerning the lag times is necessary. The first assumption of the stochastic model derived in Section 2.2.2, considers each single cell at time $t = 0$ as independent, namely the fate of the single cells does not depend at all on the presence of the other ones. We found that an extreme values statistics, that takes into account single cell stochasticity, is able to predict the measured empirical trend of the adaptation time of the population while increasing the number of initial cells. In light of the hypothesis of cooperative effects among cells, the assumption of independence falls. However, according to Majumdar and coworkers [83], if weak correlations are considered, the extreme values statistics should hold in general, by providing a rather general renormalisation group type argument (see the next paragraph in this section). Thanks to this, we may still consider valid the model developed to predict the population lag times even though single cells communicate.

There are two ways such cell-to-cell interaction may occur: either mechanical, namely through physical contacts, or chemical, namely by secreting products.

Taking advantage of the feature of Jurkat cells to form clusters (see section 2.3.2), we experimentally tested the first kind of communication by disrupting such clusters with different time frequencies. As described in details in the last paragraph of this section, we found that such mechanical perturbation does not affect the growth curves in the time scales we are interested. In general, it is likely to influence the growth curves but probably on a smaller time-scale. This suggests that the cooperativity among cells that we observe is not due to mechanical interactions.

The most likely mechanism may then be mediated by chemicals. This seems to be reasonable by considering different aspects. First, the fact that most of the measured lag times are not zero (by excluding cases with high inoculum sizes). As already mentioned in Section 2.3.2, when seeding cells at the beginning of the experiments, the only difference between the old and the new growth environment is given by the presence of cells. Before the seeding, cells were growing exponentially, thus they were living in a crowded environment full of secreted chemicals. When seeding them in the new fresh medium, such chemicals are lacking and thus cells may need a certain time to adapt.

A second argument lies in the specific cell line under exam. Indeed, Jurkat cells are widely known for their ability, once activated (namely after an external stimulation), to produce the growth factor Interleukin 2, which is known to maintain T-cell proliferation *in vitro*. However, some variants of Jurkat (JMN and JHAN Jurkat cells) have been found to produce it constitutively without needing any activation process [71]. Even though it is not know whether this may happen for the clone E6 (the one under study) or not, we may consider it as a possibility. Further investigations in this sense would be interesting and, in a positive case, would support the hypothesis of a possible chemical communication among the cells.

Our results thus suggest that a chemical communication among Jurkat cells may occur. Since the dynamics of the entire population is driven by its initial density, it is possible that such chemofactors are produced by the very first cells inoculated and that the subsequent history of the population may be influenced by them.

Extreme values statistics of weakly correlated variables

We summarise here the arguments concerning the extreme value statistics of weakly correlated variables discussed in [83].

Consider a system of N weakly correlated random variables x_i. The weak correlation may be thought as a correlation function with a connected part ($C_{i,j}$) decaying fast (i.e. exponentially) over a finite correlation length l according to:

$$C_{i,j} = < x_i x_j > - < x_i >< x_j > \sim e^{-|i-j|/l} \tag{2.43}$$

where x_i and x_j are the two variables considered. When $|i-j| >> l$, namely when they are separated by a length scale larger than l, they get uncorrelated. Since the correlation is weak, in a sample of size N we may assume that $l << N$. The following simple heuristic argument is used to understand the validity of the extreme values statistics in this case. Consider $N' = l << N$ and divide the entire population of size N in identical groups of size N'. The total number of groups will be then $g = \frac{N}{l}$: within each group k the random variables are strongly correlated, but each group is approximately independent from each other.

Let's consider the local minimum y_k, namely the minimum of all the x-variables belonging to the group k, where $k = 1,2,...,N'$. Given the approximation made, y_ks are uncorrelated and

$$m = min\{x_1, x_2, ..., x_N\} = min\{y_1, y_2, ..., y_{N'}\}. \tag{2.44}$$

Thus, by knowing the PDF of y, the calculation of the minimum of the N weakly correlated variables reduces to the calculation of the minimum of the N' uncorrelated random variables.

In our specific case, if we assume that cells are correlated only with their closest neighbours, we may still consider valid the extreme values statistics for predicting the decreasing of the lag time with the increasing of the initial cell density N_0.

Mechanical way of interaction is unlikely

To investigate the mechanical communication, we take advantage of the necessary disruption of clusters during cell counting (see Section 2.3.2) and quantify whether such perturbation may influence the growth or not.

If we assume that Jurkat cells communicate in a mechanical way, we expect a change in the growth dynamics when the population undergoes a mechanical stress as, for instance, the disruption of clusters.

To test this, we performed three experiments with the same initial concentration corresponding to the regime where we expect the cooperation to occur ($N_0 = 6 \cdot 10^4$ cell/ml). Cells were then pipetted with different time frequencies: (a) three times per day (but counted only once), (b) once every two days and (c) once every five days. In parallel to each condition, a control experiment was run, with cells pipetted

(and counted) once day. The resulting growth curves are shown in Figure 2.22(a), (b) and (c), where the data are the mean over technical replicates and the error bars represent their dispersion. For each condition, it is possible to notice a superposition of the "experiment" and its control curve. In order to compare the growth parameters of the three experiments, a fit of the averaged curves has been performed. Normal tests used to compare the fit parameters of each curve (see details in Section 2.3.3), gave scores lower than the critical value for a significancy of 5%, supporting thus the compatibility among the curves. We performed the test by comparing the fit parameters of the two sets of data for cases (a) and (b). The obtained z-scores are shown in Table 2.2. In experiment (c), due to the lack of data, the usual fitting procedure was not converging. In this case we only estimated the growth rate as the slope of the line between the first two data points. A possible incompatibility may be observed in the level of saturation, but we can attribute such difference more to experimental - decreasing of the counting algorithm performances in saturation phase (Section 2.3.2) - than to biological reasons.

In Figure 2.22(d) instead a finer analysis on the growth rate values is shown. In this case, we considered each replicate separately and calculated its growth rate λ. The plot shows the average of such growth rates over the replicates for each experiment. The error bars associated to the mean values are the dispersion of the single growth rates. For experiments (a)(1/dd) and (b)(1/2dd) it is possible to notice a compatibility among the growth rates of the experiments with their control. Only in the case (c) (i.e. the third group, pipetting time once every five days (1/(5dd))) the growth rates do not seem to be compatible. Again, as for the saturation level, this can be due to analytical reason given by the inaccuracy of the fitting rather than to a biological issue. Moreover, small fluctuations of the growth rates are present if comparing the data across all the experiments and the controls.

Such results suggest that the mechanical perturbation does not influence the growth in the time scale we are interested. This suggests us two main observations. First, the sample-to-sample variations we observed in the growth curves may be attributed to the samples themselves and not to the frequency of the measurements we performed. Second, we may exclude a mechanical communication mechanism and suggest a chemical one.

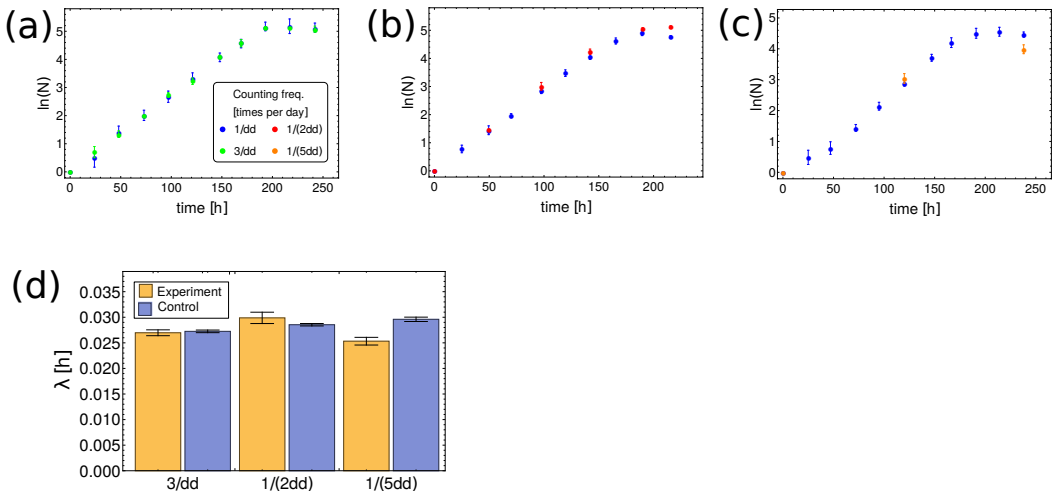

Fig. 2.22 (a-c) Example of growth curves with three different pipetting times: three times a day (a), once every two days (b) and once every five days (c). Blue data represent the control, i.e. counting every day. The legend shown in (a) refers to (a), (b) and (c). (d) Bar chart of the values of the growth rates of the pipetting experiments (orange bars) and the control (blue bars). The values have been obtained by averaging the growth rates of each single replicates. The error bar is the error over such average value. The three groups of data correspond to the three experiments: pipetting three times a day (3/dd), once every two days (1/(2dd)) and once every five days (1/(5dd)).

Exp	$z_{t_{lag}}$	$z_{\lambda_{max}}$	z_A
(a)	0.07	0.27	0.26
(b)	0.66	0.50	0.25

Table 2.2 *z-scores* resulting from normal tests performed to compare fit parameters of the experiments (a) and (b) with their control. Each column is relative to the comparison between specific parameters: lag times ($z_{t_{lag}}$), maximum growth rates ($z_{\lambda_{max}}$) and saturation level (z_A). The critical value for a two-tails test with significancy of 5% is $z_{crit} = 1.96$.

2.3.8 New insights on the experimental design

In the previous sections we analysed the inoculum dependence of carrying capacity, lag time and exponential growth rate of the population. In this section we focus on the fluctuations of the last two variables investigated, shown in Figure 2.23(a) and (b).

According to our results, the fluctuations of the two variables show an opposite behaviour: by increasing the size of the initial population, the lag time fluctuations

linearly decrease, while the growth rate fluctuations increase. In other words, in order to have a sample-to-sample low variability in the growth rate one needs to seed cells at low concentrations. However, in such condition, the precision over the lag times would be lost and the very duration of the adaptation phase would be long-lasting (more than 3 days).

These observations allow to give new insights on the experimental design. First of all, such scenario can give the experimentalists a quantitative argument to design the experiment by choosing the desired initial cell concentration, according to their main purposes of investigation. Second, since the two trends are opposite, an "optimal" initial concentration of cells may be found in correspondence of a trade-off between the two variables. It leads a trade-off between the behaviours of the two variables and it is emphasised by a black dot in Figure 2.23(c). We derived such condition as follows. For both variables, we considered their standard deviation within the same N_0 bin (values shown in 2.23(a) and 2.23(b)) and normalised each of them to its maximum value. In this way we obtain two adimensional comparable variables, as shown in Figure 2.23(c). The "optimal" point is given by the intersection of the two lines (black dot in the plot). Interestingly, N_0 at the intersection point is $N_0 \simeq 2 \cdot 10^4$cell/ml, corresponding to the cooperative phase shown by the growth rate behaviour *vs* N_0 (see Figure 2.15).

Thus, by seeding cells at the "optimal" density, a compromise between the growth rate and the lag time variations is encountered, the lag time has reasonable values and the cells are in the cooperative regime.

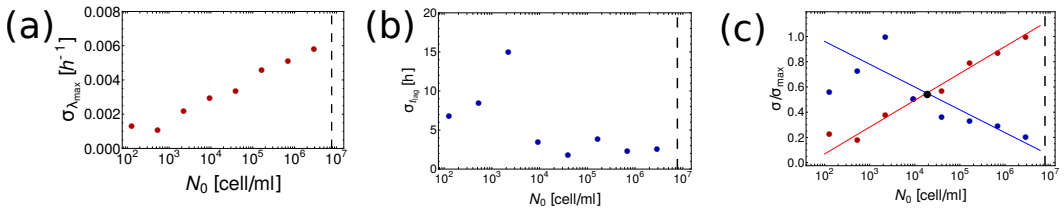

Fig. 2.23 (a-b) Standard deviations of the maximum growth rate λ_{max} (a) and the lag time t_{lag} (b) within bins over N_0. (c) Same standard deviations (σ) shown in (a) and (b) normalised to their maximum values (σ_{max}) as a function of the inoculum density (N_0). In red is the exponential growth rate and in blue the lag time. The solid lines represent the fits of the two series of data. The black dot at the intersection of the two lines represents the "optimal" experimental condition. The dashed vertical line is located in correspondence of the carrying capacity.

2.3.9 Qualitative identical features are shown by a further cancer cell line

In order to investigate whether our findings were specific of Jurkat cells or not, we investigated a second human leukaemia, the K562 cell line, by following the very same approach adopted for Jurkat cells. The experiments have been performed in collaboration with Dr. F. Ceroni at Imperial College (London).

K562 cells, are suspension cells related to a human chronic myelogenous leukaemia. They are slightly bigger than Jurkat ($17\,\mu m$ vs $12\,\mu m$) and do not tend to form clusters while growing. They are widely used in research as highly sensitive *in vitro* targets for the assays to study natural killer cells, immune system cells that once activated can modulate immune responses [84].

We performed 68 experiments (i.e. 68 growth curves) by spanning roughly 3 orders of magnitude of initial cell density. The experimental procedure to obtain the growth curves is analogous to the one adopted for Jurkat experiments and it is summarised in Figure 2.24(a). As for Jurkat cells, at time t=0 h we took a sample of N_0 cells from the original flask where the cells were growing exponentially in their standard growth medium. Such sub-population was then transferred into a new culture dish (75 ml flasks) supplied with a fixed amount of the same standard growth medium of the pre-culture. Differently from Jurkat experiments, K562 were cultured in flasks with 10 ml of growth medium and their growth have been monitored through an automatic cell counter, the *NucleoCounter*.

The *NucleoCounter* is a fluorescent microscope equipped with 3 light sources and a low magnification ($\sim$ 1.75x). It allows to automatically analyse a high number of individual cell properties as for example viability and dimension. Among all, we were interested in the concentration of alive cells in the sample. In order to count with the *NucleoCounter*, the sample had to be prepared by adding to the desired cell+medium a supplied reagent that allows cells to become fluorescent. In this way the instrument is able to detect them and automatically give information on alive and dead cells viability and alive cells size. So, to perform the measurement, the population of cells was firstly pipetted in order to dissolve eventual clusters, then a sample of 100 μl was collected and mixed with $10\,\mu l$ of the supplied reagent. $5\,\mu l$ of the mixture was then analysed by the *NucleoCounter*.

By daily monitoring the cell growth, the growth curves have been obtained. Three examples of so-obtained experimental growth curves (with three different N_0s) expressed as the concentration off cells (N) as a function of time are shown in the plot in Figure 2.24(a), where the dots are the experimental data and the lines their fits.

To estimate the growth parameters, we fit the growth curves with the sigmoidal shape function described in Section 2.3.3 whose parameters are related to the ones of interest, namely the lag time t_{lag}, the exponential growth rate λ_{max} and the parameter A, related to the carrying capacity k. By adopting the analytical approach described in Section 2.3.3, for each growth curve we calculated also the value of the exponential growth rate λ.

Figure 2.24(b-e) show the results of the quantification of the interesting parameters for K562 cells, in analogy to what have been done for Jurkat cells. By comparing the absolute values of K562 parameters with those of Jurkat, it is interesting noticing that the carrying capacity of K562 is smaller than that of Jurkat; the lag time spans smaller values while the growth rate reaches higher values.

However, the trends of such variables as function of N_0 are qualitatively in agreement with those of Jurkat cell line. First of all, the parameter A follows the same relation expressed in eq. 2.18 obtained for Jurkat cells (Figure 2.24(b)). The solid black line that fits the data has a slope compatible with -1. As evidenced by Figure 2.24(c), the distribution of the converted carrying capacity is compatible with a normal distribution (the solid blue line) centred in $k = (7.5 \pm 1.8) \cdot 10^5$cell/ml (solid black line). This suggests then that the carrying capacity is expectedly constant with respect to the initial cell density and thus the experiments have been performed in controlled conditions. Secondly, the lag time shows a decreasing trend together with the increasing of N_0 (Figure 2.24(c)). Finally, the exponential growth rate λ as a function of N_0 decreases when approaching the carrying capacity (dashed black line).

The presence of such qualitative agreement suggests that the discovered features may not be specific to a single cell line, but they can be more generic. Thus, it would be important to enrich the set of experimental evidences by analysing the same issues on other different cell types.

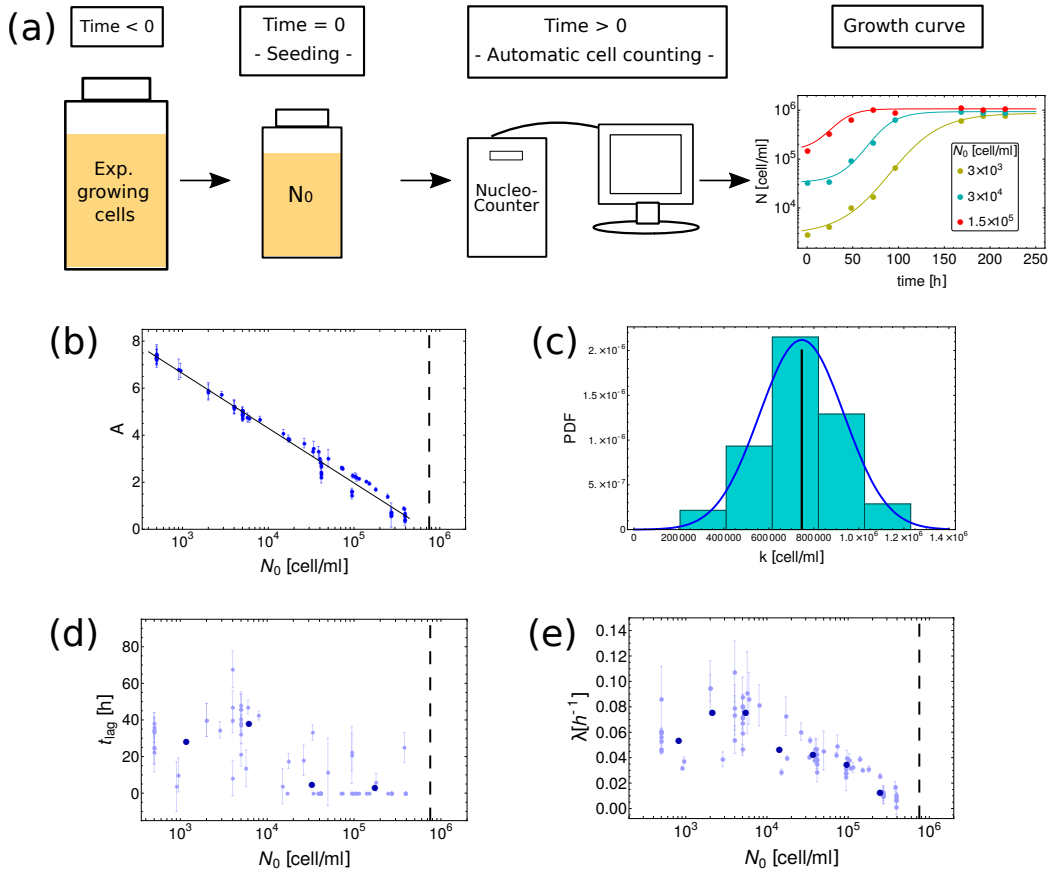

Fig. 2.24 (a) Cartoon of the experimental procedure. K562 cells were pre-cultured in a flask supplied with the standard growth medium. At time $t = 0$ h populations of N_0 cells were transferred into a new culture dish (a new flask) supplied with a fixed volume of the same standard growth medium. The growth of the populations was daily monitored through the *NucleoCounter*. The last plot shows three representative growth curves obtained through this procedure at the end of each experiment. Each colour represents an inoculum size (N_0), the dots are the experimental data and the solid lines are the sigmoidal shape functions fitting the data. (b) Trend of the fitting parameter A as a function of the inoculum density N_0. The data are compatible with a line with slope -1 as a function of $ln(N_0)$, as emphasised by the black solid line that fits the data. (c) Distribution of the values of the carrying capacities converted from the parameter A through the relation 2.18. The solid blue line represents the normal distribution best fitting the data. The solid black line is the average of the distribution, corresponding to $k = (7.5 \pm 1.8) \cdot 10^5$cell/ml. Such value is emphasised by a dashed vertical line in Figures (b,d,e). (d) Lag time (t_{lag}) trend as a function of N_0. Each bright dot corresponds to a single growth curve, the darkest ones are the average by binning on N_0. The error bars are given by the error of the fit. (e) Exponential growth rate (λ) as a function of N_0. As in (b) the bright dots correspond to all the experimental data, while the darkest ones are their average by binning on N_0. The error bars are the error of the fit.

2.4 Conclusions

In this Chapter we investigated how the growth dynamics of a population of cells is influenced by its initial density. Beginning with minimal mathematical models based on very simple theoretical assumptions, we tested then their predictions on a *in vitro* system consisting of populations of widely studied cancer cells, the Jurkat clone E6.1. Finally, we modified those hypothesis that were not confirmed by the empirical proof and suggest new phenomenological mathematical models. The results obtained were qualitatively confirmed by the analysis of a second cancer cell line (K562). To the experimental purpose, we designed a systematic experimental protocol that allowed us to obtain robust growth curves. Through these, we defined the different phases of growth (i.e. adaptation, exponential and saturation) of a population of cells living in an environment with fixed amount of nutrients and analytically quantified the parameters representative of such phases. These prameters were then compared with the corresponding model. Indeed, we studied each growth parameter separately and by unifying the results of each one we drew new hypothesis for a further detailed investigation.

As a first result, we quantitatively proved that the experiments have been carried out in a controlled environment since, as expected, the saturation level of the growth curves (carrying capacity) is independent on the initial condition. The carrying capacity represents the maximum amount of cells that can live in a given environment, thus only depending on the quality of the environment and not on the initial concentration of cells. No matter how many cells give rise to the population, the maximum density they can reach is always the same if the concentration of the nutrients does not change.

The second result concerns the lag time, namely the time a population of cells needs to adapt when moved to a new environment. By taking into account the high heterogeneity of single cells within a population, we developed a stochastic model able to link single cells adaptation times with the lag time of the population. Through simple measurements of the last macroscopical variable on the *in vitro* system, we have been able to conclude that an extreme values statistics may predict the time the population needs to adapt before growing exponentially. Moreover, this depends on the initial size of the population itself. In particular, we found that the fastest single cell to divide after the seeding is the one that prevails over the others and drives the dynamics of the entire population. From such result it is worth highlighting

the complexity of the system we studied here. Indeed, it arises how the behaviour of a population of single growing cells is not given simply by the average of their microscopical behaviours. These are combining in a more complex way in order to give rise to the emergent and robust behaviour at the macroscopic level. Such aspect is of pivotal relevance and allows to consider a population of cells as a complex system whose emergent behaviour has characteristics that differ from those of each single component. Furthermore, in our case, each single component of the system is able to move, reproduce, interact and grow. Thus, by merging all such features, it is evident how the system we have studied fits under the framework of active matter, being it complex, out-of-equilibrium and composed by single individuals able consume energy while and for growing [85, 86].

The last variable we investigated is the exponential growth rate, namely the asymptotic regime at which cells proliferate in a constant rate. To describe the exponential phase of growth in presence of a fixed carrying capacity, we considered the logistic model as the simplest deterministic and population-level mathematical model. In light of the experimental validation, we found that the cells under exam does not follow such dynamics. Indeed, through considerations concerning the exponential growth rate as a function of the initial size of the population, we found the existence of a range of initial population densities in which a cooperative behaviour among cells arises. Such finding has been qualitatively reproduced by a widely known mathematical model for density dependent growth, known as Allee effect. This is not trivial if one thinks that at the time the exponential regime of growth usually occurs, the population has undergone several doublings, so every information of the initial state should be lost. Since this is not the case, we may understand such behaviour in terms of chemical communication among cells. If so, cells secrete chemofactors while growing and this may give rise to possible feedback loops on their own growth. Further investigations would be necessary in order to support this, but it is known that Jurkat cells secrete growth factors while growing. The most known one is Interleukin 2, that has been found to play a pivotal role in *in vitro* T-cell proliferation. Interestingly, in [87], Matera and coworkers show the existence of a further chemical product secreted by Jurkat cells, the prolactin. In their study, they conclude that it plays a central role in influencing Jurkat own growth. Thus, further experimental investigations in this sense would be interesting in order to verify our hypothesis of feedback loops on population growth caused by the production of chemofactors. If this will be the case, then it would be interesting the development of a mathematical

model able to explain such mechanism of communication. This would help to make predictions on the development of the system and on the dynamics of such leukaemic cells. Furthermore, it has been found that some clones of Jurkat cells, are able to express Interleukin-2 even in absence of an external activation. It would be interesting a precise quantification of Interleukin-2 production for the clone we characterised in order to relate it to cell proliferation. Since, as mentioned above, such protein plays a remarkable role in the *in vitro* proliferation of T-cells, it would be interesting defining methods to enhance their productivity by Jurkat cells. The use of cells as factories to produce desired chemicals, goes under the name of bioproduction. Since most of such secreted chemicals, once isolated, are useful for therapeutics purposes, several investigations have been performed aiming at enhancing bioproduction and producer cells growth [38–41]. On this topic, we refer to Appendix A. There, we take advantage of the experimental protocol for cell growth and data analysis that we developed in this Chapter. Through these, we investigate a method for optimising the growth of a population of cells known for its bioproduction features.

Finally, we repeated the very same kind of investigation on a different leukaemic cell line, the K562 cell line. We found that the dependencies of the growth parameters on the inoculum density are qualitatively in agreement with those obtained with Jurkat cells. This suggests the possibility that such qualitative patterns may not be specific of a single cell line. In order to deeply investigate such complex role of the initial condition in mammalian cell population growth, it would be important to expand the analysis discussed here also to further different cell types and identify the mechanisms underlying such features more clearly. In light of this, a more thorough modelling approach would go beyond the presented minimal framework.

Chapter 3

Cell growth laws

In this chapter we present and discuss the second part of the work, which focuses on a collaboration with Dr. Francesca Ceroni at Imperial College in London. The project is still in progress, we therefore discuss here only the preliminary results obtained so far.

The work is closely intertwined with the previous part of this thesis, since it begins from the definition of cell growth and uses such parameter as the tuning variable for investigating mammalian growth physiology. Our main focus is in fact the characterisation of the impact of mammalian cell growth on gene expression. Quantitative studies in this sense have been widely developed in microbiology in the recent years [6]. In particular, a top-down quantitative approach was used in order to determine a phenomenological description of the relation between bacterial growth and gene expression [5]. Since similar studies are lacking for mammalian cells, our aim is to start bridging this gap by transferring the same methodology used for bacteria to cancer cells. Then, from cancer cells we intend to extend it to healthy mammalian cells in order to have a broad characterisation of mammalian growth physiology. With a constant interplay between mathematical modelling, wet-lab experiments and data analysis performed in parallel both in Torino and in London, we point towards a phenomenological large scale description of mammalian cell physiology able to be predictive even in absence of a complete knowledge of molecular details.

In order to better introduce our work, in the following section we discuss the mentioned studies that inspired the project, together with the reasons why we began our investigation from cancer cells.

3.1 The Ohm's laws for bacterial growth

Quantitative physiology plays nowadays a pivotal role in synthetic biology [88] and bioingeneering. Its studies lie at the bottom of the development of artificial genetic circuits by giving the knowledge on how integrating artificial with biological systems. Cellular quantitative physiology has been widely developed in recent years, especially in microbiology [6].

When dealing with the study of biological systems, two different kinds of approach can be followed: bottom-up and top-down. A bottom-up approach usually focuses on microscopic details, like a small regulatory system, and progressively scales up to the more complex regulatory network. This kind of approach becomes difficult to apply when dealing with large complex systems due to the "explosion" of the number of parameters. Therefore a top-down approach seems to be a better strategy. This focuses on global emergent variables and determines the relations among them with simple but quantitative rules, by discarding small scale details. By following this second approach, quantitative empirical relations have recently emerged in microbiology [5]. Starting from the fact that constitutive gene expression is strongly coupled to the growth state of the culture [89], these empirical laws determine a link between the growth rate and the cell composition of exponentially growing bacteria. The exponential phase has been considered, since it is the situation in which every cell and its components double at a constant rate [2]. In [5], the growth rate has been taken as the representative variable for the physiological state of the culture and was adjusted by the quality of the medium the bacteria were growing in. The quality of the medium was defined by the presence of sugars or other components. To give an example on how the medium quality can affect the growth rate, it has been seen that *Escherichia Coli* grows faster in glucose than in glycerol. In order to set the growth conditions, nutrient concentration was properly chosen according to the Monod equation (eq. 3.1) that relates the doubling rate (R) to the concentration of the growth-limiting substrate (C) [2]:

$$R = R_k \frac{C}{C_1 + C} \quad , \tag{3.1}$$

where R_k stands for the maximum growth rate attainable with the increasing of C, and C_1 is the substrate concentration in correspondence of a growth rate $R_k/2$. According to Monod relation, the growth rate increases with the growth-limiting concentration until it reaches a saturation level. In other words, there is a value of growth-limiting substrate concentration from which the growth rate will not increase anymore. Figure 3.1 shows an example of the Monod equation adapted from [2]: experimental data (dots) follow the solid line drawn from the equation 3.1.

The concentration of nutrients in [5] was then chosen in order to be not-limiting for the growth rate.

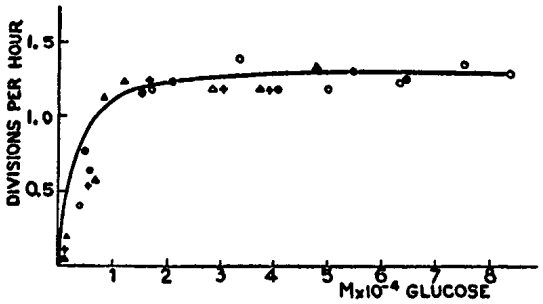

Fig. 3.1 Growth rate of *Escherichia Coli* as function of glucose concentration. The solid line is obtained with the Monod relation (eq. 3.1). Figure adapted from [2].

Going back to the main aim of the work of interest [5], the physiological state (described by the growth rate) was related to the gene expression. Cell composition, i.e. RNA and protein content, was considered as the representative variable for gene expression. The reason is the following. Since $\sim 85\%$ of total RNA is ribosomal RNA [90], the ratio between RNA and proteins is a proxy for the ribosomal content of the cell. Since ribosomes are required for protein synthesis, their number and rate-of-function determine the rate of protein synthesis, and therefore gene expression.

The authors found very simple empirical relations linking the exponential growth rate and cell composition. This was the starting point for the development of the bacterial growth theory, a phenomenological theory that assumes the partitioning of the proteome (i.e. cellular repertoir of proteins) [5]. These findings are briefly described in the following and sketched in Figure 3.2.

Figure 3.2(a) shows a sketch of the mentioned empirical relations. It was found that in *E. Coli*, by changing the growth rate through the nutrient quality, RNA/protein (r) linearly increases with the growth rate (λ) by following the relation:

$$r = r_0 + \frac{\lambda}{\kappa_t} \tag{3.2}$$

where r_0 is the offset and κ_t is related to the translation rate and it is called *translational capacity*. This is represented by the solid line of the left plot in Figure 3.2(a). By tuning the growth rate inhibiting the translation, instead, an inverse linear trend is obtained (dashed lines in the left plot of Figure 3.2(a)). The relation found is:

$$r = r_{max} - \frac{\lambda}{\kappa_n} \ , \tag{3.3}$$

where r_{max} is the maximum allocation fraction for the ribosomes in the limit of total translational inhibition, i.e. $\lambda \to 0$; κ_n is called *nutritional capacity* and it is related to the nutrient quality of the medium. If instead of the ribosomal content, the mass fraction of constitutive proteins is considered (see plot on the right in Figure 3.2(a)), symmetric linear relations are observed. This implies then the presence of a linear constraint between ribosome-affiliated and constitutive proteins within the cell. Starting from these relations and constraints, the authors defined a minimal-model for proteome partitioning. See Figure 3.2(b) for a scheme of the main results of the model. In the simplest case, the bacterial proteome can be partitioned into three components: a core sector (Q, of fraction Φ_Q), a ribosome-related sector (R, Φ_R) and a third one (P, Φ_P) that includes constitutive proteins. While the first sector is expressed in every growing condition (i.e. growth rate), the other two are coordinately reshaping depending on the growth rate.

If $\Phi_P \to 0$ then $\Phi_R \to \Phi_R^{max}$. Given that

$$\Phi_P + \Phi_Q + \Phi_R = 1 \ , \tag{3.4}$$

it follows: $\Phi_R^{max} = 1 - \Phi_Q$ with

$$\Phi_P = \Phi_R^{max} - \Phi_R = \rho(r_{max} - r) \ , \tag{3.5}$$

where ρ is a conversion factor between r and Φ. Eq. 3.5 represents a constraint between Φ_P and Φ_R. Together with eq. 3.3, the model predicts that

$$\Phi_P = \rho \frac{\lambda}{\kappa_n} \ . \tag{3.6}$$

This describes the linear relation between constitutive proteins (P-sector) and growth rate λ for fixed nutritional capacity, and can be interpreted as a manifestation of those proteins that provide the nutrients needed for the growth. In a similar way, eq. 3.2 may be seen as the R-class proteins providing the protein synthesis necessary for growth:

$$(\Phi_R - \Phi_0) = \rho \frac{\lambda}{\kappa_t}. \tag{3.7}$$

For different combinations of the nutritional (κ_n) and translational (κ_t) capacities, an efficient allocation of resources requires that the abundance of proteins of R- and P-classes must be adjusted so that the influx of nutrients provided by the P-sector is balanced by the protein synthesis carried out by the R-sector. Thus, the growth rate dependence of constitutive gene expression results as a balance between protein synthesis and nutrient uptake/processing.

Such physiological laws were called by the authors "bacterial Ohm's laws" due to an analogy with the behaviour of an electrical circuit [5] (see Figure 3.2(c) and Section 3.1.1 for further details). Moreover, the solution of eq. 3.6 and 3.7 together, leads to the Michaelis-Menten relation empirically known for describing the dependence of the cell growth on the nutrient level, namely eq. 3.1:

$$\lambda(\kappa_t, \kappa_n) = \lambda_c(\kappa_t) \frac{\kappa_n}{\kappa_n + \kappa_t} \tag{3.8}$$

where $\lambda_c(\kappa_t) = \kappa_t(r_{max} - r_0)$ is the value of the maximal growth rate in a rich medium.

Through this theory it has been possible to predict and explain global changes in bacterial transcriptional program without the need of knowledge of many molecular-scale details. For instance, they have been adopted to predict a wide spectrum of microbial behaviours that range from antibiotic resistance to the decreasing of the growth rate when unnecessary proteins are expressed [6].

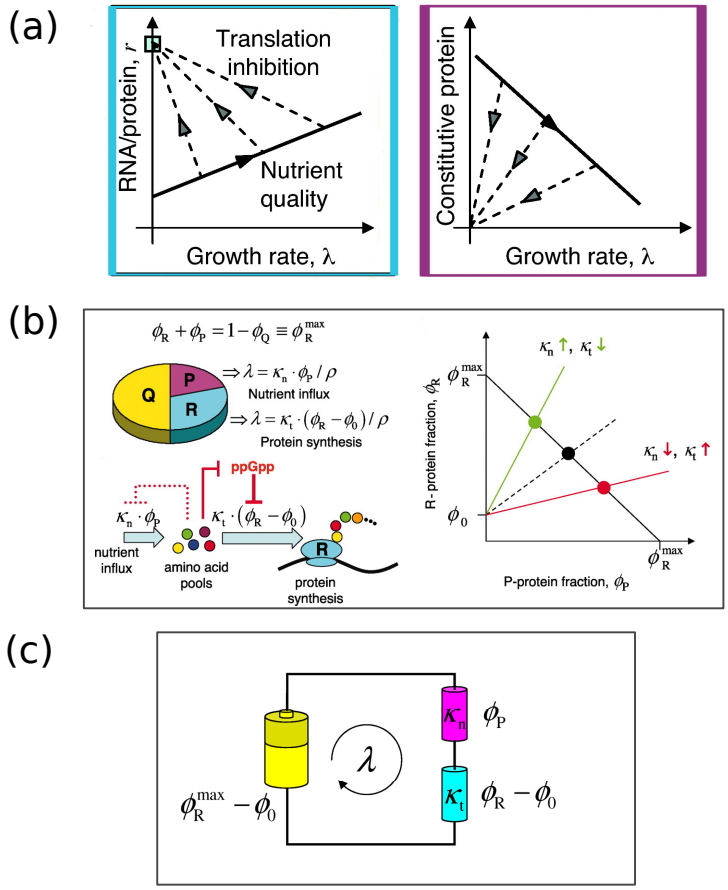

Fig. 3.2 Three figures showing the results obtained by Scott et al. in [5]. (a) Sketches of the empirical growth laws. Figure adapted from [6]. (b) Summarising sketches of the theory of proteome partitioning. Figure adapted from [5]. (c) Analogy with Ohm's laws for electrical circuits. The three resistors correspond (also in colours) to the three proteome sectors of panel (b). Figure adapted from Supplementary Materials of [5].

Even if less broadly than in bacteria, similar studies connecting growth and gene expression have been performed also on some eukaryotic organisms such as fungi [91], algae [92] and yeast [93, 94]. Having a similar systematic study on eukaryotic cells coming from more complex organisms like mammalian cells, would be very important since it could find many useful applications in bioengineering and therapeuthics.

One first way of addressing this issue by adopting the same approach used for bacteria, is to consider cancer cells. The choice is driven by a wide range of analogies between behaviours of bacteria and cancer cells [32, 33]. To give some examples, bacteria, through an advanced communication technique known as quorum

sensing, are able to form complex communities (biofilms) and it has been found that cancer cells may use similar communication mechanisms during metastatic invasion [95]. Also the formation itself of biofilms (bacteria) and metastasis (cancer) can be compared as well from different points of view, from the motility of cells towards a certain substrate, to the adhesion and the formation of the structure, i.e. biofilm or metastasis [96]. Another interesting analogy, useful for our purposes, is the presence in both cases of an exponential phase of growth. As described in the previous Chapter in Section 2.1.2, in his work of 1949 [2], Monod defined three important phases of growth for bacteria (i.e. lag, exponential and stationary phase) and it is interesting to underline how the same phases can be recognised in the case of cancer cells when allowing them to exhaust the limiting nutrient. This is relevant for our purposes since the approach we are interested in transferring to cancer cells was valid for exponentially growing bacteria [5]. A further motivation for our choice is the discovery that several tumor suppressors and oncogenes are related to protein synthesis machinery [97], therefore possible relations between cancer growth and proteome partitioning may exist.

In Section 3.2 we present an outline of the work. To this, it follows the detailed description of the methods (Section 3.3) we used for obtaining the preliminary results discussed in Section 3.4.

3.1.1 Analogy with an electric circuit

The authors of [5] draw an analogy between the proteome partitioning model and an electrical circuit, by giving rise to the so-called Ohm's bacterial growth laws. We summarise here such similarities by referring to Figure 3.2 (c).

The main equations lying at the base of the bacterial growth theory are eq. 3.2, 3.5 and 3.6. Through these, the described theory is identical to the mathematical description of the flow of electric current in a circuit composed by two resistors connected in series to a battery with voltage $(\Phi_R^{max} - \Phi_0)$. The current flowing in the circuit is given by the growth rate λ. The resistors have a conductance equal to κ_n/ρ (in magenta in Figure 3.2(c)) and κ_t/ρ (in cyan). The voltage drop corresponding to the resistors is given by the mass fractions Φ_P and $(\Phi_R - \Phi_0)$. In this way, by rewriting equations 3.6 and 3.7 as:

$$\lambda = \frac{\kappa_n}{\rho}\Phi_P \ , \ \lambda = \frac{\kappa_t}{\rho}(\Phi_R - \Phi_0) \ ;$$

the analogy with the Ohm's laws is evident: the current is equal to the product between the conductance and the voltage drop upon the resistor. The nutrient- and translation- modes for limiting the growth correspond to changes in the conductance of one of the resistors, while the expression of unnecessary proteins is the analogous of changing the applied voltage, namely by decreasing R_{max}.

To conclude the analogy, we refer to eq. 3.8. This corresponds to the relation between current and voltage in presence of two resistors connected in series, as it can be easily seen by rewriting the eq. 3.8 in the following form:

$$\lambda = \frac{(\Phi_R^{max} - \Phi_0)}{\rho}\frac{\kappa_n \kappa_t}{\kappa_n + \kappa_t} \ . \tag{3.9}$$

3.2 Project outline

The aim of this project is to transfer the same top-down quantitative approach adopted for bacteria in [5] to cancer cells for relating their growth to gene expression. The project has been performed part in Torino and part at Imperial College in London. In order to distinguish the two sets of results and methods, from now on we will refer to them as Torino's Lab and London's Lab respectively.

For reaching the goal of the project, the following steps are required:

1. The experimental determination of the quantities we are interested in studying, namely the ones describing the growth state of the culture and its gene expression.
2. Define how to change the growth state.
3. Quantify the two chosen variables in order to look for possible empirical relations between them.
4. Define a phenomenological theory based on the empirical relations of step 3..
5. Repeat the same approach with healthy mammalian cells.

At the present time, steps 1. and 2. were broadly investigated and first data for the realisation of step 3. are also available. Once reached statistically satisfying amount of data of step 3., it will be then possible to continue with steps 4 and 5. Therefore all the description that follows in this chapter will be relative to steps 1, 2 and 3.

As a first point for reproducing the same approach adopted by Scott et al. [5], we decided to define the same variables they used, even though we do not know if these can be representative for mammalian cells as well as they are for bacteria. Therefore, we consider as empirical quantitaties to study: the growth rate, as representative of the growth physiology, and RNA/protein contents as a proxy for gene expression, since also for mammals, the $\sim 85\%$ of RNA is ribosomal [98].

In order to vary the growth rate, we consider as variables the nutrient quality (like for bacteria studies), the inoculum size (taking advantage of the results discussed in Chapter 2) and the temperature. The choice of the temperature as a knob for changing the growth rate, was driven by evidences that show that it effectively has an impact on the growth rate both for mice [99] and for some human cancer cells [100].

In Torino's Lab we fixed one nutrient quality (glucose), the growing temperature at 37°C and changed the inoculum. In London instead, the initial concentration of cells was fixed, while the nutrient quality (i.e. sugars in a minimal growth medium) and the temperature were changed once at a time.

Concerning the nutrient quality, for each sugar, different experiments were performed in order to determine the optimal non-limiting nutrient concentration, reproducing in this way the Monod relation between the growth rate and the growth limiting substrate concentration [2].

Once defined the growing conditions, we left the cells free to grow by exhausting the nutrients and checked the growth day by day until the population reached the exponential phase. At this point the total RNA and total protein content of the population was extracted and quantified and then related to the growth rate.

Slightly different procedures were used in the two labs concerning the growth experiments. All details will be exposed in the following Section 3.3.

3.3 A systematic experimental approach

This section is dedicated to the discussion of the methods adopted in the project.

The cell lines considered here are Jurkat clone E6.1 (used in both labs) and K-562 (only in London's lab). For a description of Jurkat and K562 cell lines see sections 2.3.2 and 2.3.9 respectively.

The general experimental procedure followed for growing cells in both labs can be summarised as follows: cells were seeded with the defined condition (inoculum, sugar quality, temperature) and their growth was monitored day by day. When the cell population reached the exponential phase, it was freezed at $-80°C$ and then total RNA and total proteins were extracted from the entire population.

As mentioned before, slightly different procedures were followed in the two labs for growing cells and following their growth. In the next Section 3.3.1 we describe the experimental procedures for cell growth and in Section 3.3.2 we describe the sets of experiments that we performed. In the last Section 3.3.3 instead, we briefly present the experimental protocol for RNA and Protein extractions used in both labs.

3.3.1 Cell growth experiments

The methodologies used for the experiments in the two labs are described in the following paragraphs.

Torino's Lab

Growth experiments in Torino's lab, were performed by following the general procedure described in Section 2.3.2. To briefly summarise it, an initial concentration of Jurkat cells (N_0) was seeded in a fixed amount (5 ml) of their optimal medium, i.e. RPMI which constitutively contains 25mM of Glucose. Cell population was allowed to growth at constant temperature of 37°C and with 5% of CO_2. The growth was monitored day by day until it reached the exponential phase. At this point the cells were freezed to later extract RNA and proteins (see Section 3.3.3 for RNA and protein extraction protocol). In order to be sure to stop the growth at the desired time point, another identical experiment (called *A* from now on) was started three days before and left run until it reached the saturation phase. In this way, the *A* experiment

resulted to be a "pilot" for checking the growth phases of the experiment needed for the extraction (called *B* from now on). For each condition, in both *A* and *B* cases, at least three replicates were performed. See Figure 3.3 for an example of *A* (blue shaded circles) and *B* (red shaded circles) curves ($N_0 = 60 \cdot 10^3$ cell/ml). Each single shade of the two colours represents one replica. The data were shifted in order to start from time 0 h. It is possible to see how *B* data (red) follow the same trend as *A* (blue) ones.

A curves were also used to estimate the growth rate of *B* ones. To this purpose, for each *B* replica the sigmoidal fit described in Section 2.3.3 was performed on a set of data composed by the joining of the desired *B* replica and all the corresponding *A* data. In this way a complete curve was available and all the growth parameters could be determined, namely the lag time, the saturation level (carrying capacity) and the maximum growth rate. The last time point of the red curve corresponds to the time of extraction of RNA and proteins.

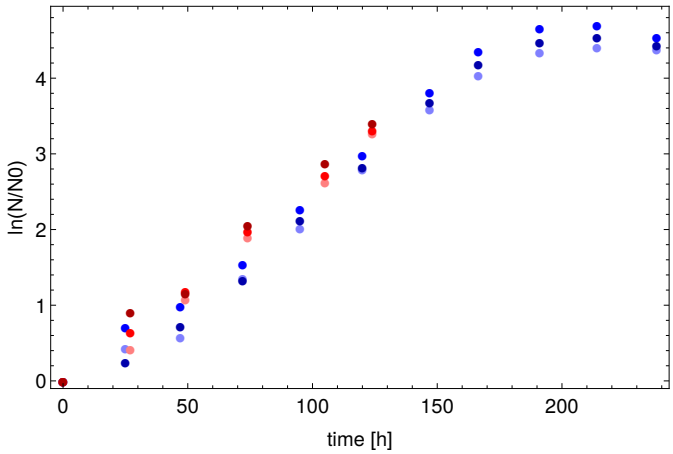

Fig. 3.3 Example of *A* (blue) and *B* (red) growth curves expressed as the logarithm of the cell concentration (N) normalised to the initial cell concentration (N_0) as a function of time. For each colour there are 3 shades representative of 3 different replicates of the same condition.

London's Lab

As mentioned before, in London's lab we analysed two different human cancer cell lines in parallel: Jurkat and K562. The "knobs" used for tuning the growth rate were the nutrient quality and the growth temperature, while the inoculum size was fixed at $N_0 = 10^5$ cell/ml.

The sugars here considered were: glucose (the "standard" used to compare London and Torino results), fructose, mannose, maltose, galactose and xylose. For varying the nutrient quality, different growth media were prepared. The minimal medium lacks sugar so that the desired sugar at the given concentration was added.

Temperature was set at 37°C, 35°C, 33°C and 32°C.

Cells were seeded in 48-multiwells in fixed volume (1 ml) of growth medium. To monitor the growth, we used the *Tecan* plate reader, an instrument that allows high throughput. The instrument was supplied with temperature and CO_2 controllers. In this way, at the beginning of each experiment we set the desired temperature and 5% CO_2 (value suggested for an optimal growth). These two quantities remained constant for the whole duration of the experiment. For each experiment at least two wells were filled with 1 ml of minimal medium and left free from cells. The trend of these wells over time gave the background measurements.

The *Tecan* allows to automatically reconstruct growth curves by measuring the absorbance of the medium at two fixed wavelengths, 430 nm and 560nm, that correspond to the acid and basic peaks of the phenol red respectively [101]. To remind what discussed in Section 2.3.2, the phenol red is a pH indicator present in the growth medium. Therefore the measure of the absorbance of its acid and basic peaks, gives an information on pH changes of the medium. Since the pH changes in time accordingly to the cell density, the ratio between the two peaks ($abs430nm/abs560nm$) can be used as a proxy for the growth of the population.

The protocol followed by the instrument to perform the measurement is the following: it shakes the plate for 5 s, waits few seconds, then it measures absorbance at 430 nm in each well, shakes again for 5 s, waits and measures absorbance at 560 nm in each well.

Since the measurements were very quick and automatic, we set the time step every 12 hours. In this way we obtained growth curves with a resolution in time higher than the one of Torino's lab.

The panel in Figure 3.4 shows an example of *Tecan* measurements. In Figure 3.4 (a) we show the raw measurements performed by the instrument as absorbance values *vs* time. Figure 3.4 (b) instead shows their ratio over time after background subtraction. Indeed for each experiment, we also followed the absorbances of the clean medium, whose ratio is representative of the background. Thanks to

this, we subtract to every absorbance ratio the corresponding background thus obtaining curves like the one shown in Figure 3.4(b). Data shown in this example are measurements of K562 grown at 37 °C in glucose.

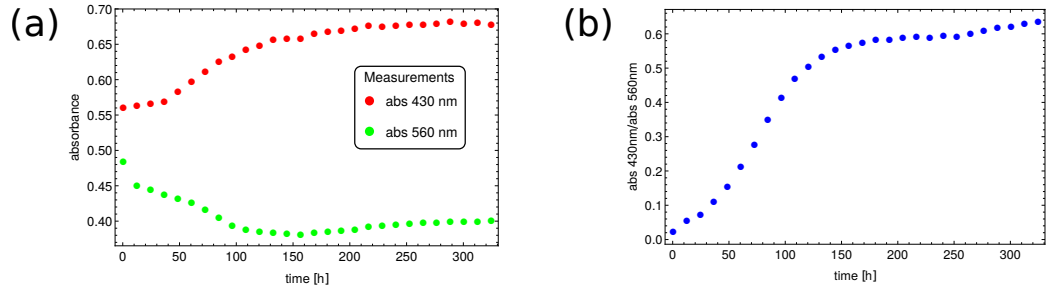

Fig. 3.4 An example of K562 growth curve in glucose at 37°C. (a) shows the raw measurements performed by the *Tecan*: absorbance at 430 nm in red and 560 nm in green. (b) shows the ratio of absorbance 430 nm over 560 nm *vs* time after background subtraction.

Three different kinds of experiments were performed by following the above protocol: (i) calibration, (ii) sugar saturation experiments and (iii) experiments for the RNA and protein extractions. Each of them will be described in the following section and the preliminary results will be discussed in Section 3.4.

It is worthy to mention also that the growth rates were calculated by fitting the *Tecan* curves expressed as $abs430nm/abs560nm$ with the sigmoidal-shaped function described in Section 2.3.3. In this way the three phases of growth were determined. The variable we are interested in is the growth rate of the exponential phase (λ), which was obtained by the slope of the linear fit of data within the exponential phase.

3.3.2 Three sets of experiments

The experiments we performed in the direction of the determination of the mammalian growth laws can be divided in three sets described in the following paragraphs: calibration, sugar saturation and growth laws experiments.

Calibration

Since we need to compare London growth curves with Torino's ones, we need a way to convert results obtained in terms of absorbance (*Tecan* measurements in London) to those in terms of cell density (direct manual counting in Torino). To this

purpose in London's lab we performed a parallel detection of cell growth both with *Tecan* and with direct cell counting, for cells growing in glucose at 37°C with an initial seeding $N_0 = 10^5$ cell/ml. All the results obtained for the other conditions of sugar and temperature are then expressed in relation to this.

Cells were seeded in the growth medium containing 25 mM of glucose. The chosen concentration of sugar was exceeding the growth limiting value (see discussion in Section 3.4). All the wells were measured by the *Tecan* day by day, and every day two of them were also measured by directly counting the cells. Thus two growth curves were obtained with both counting methodologies and used to make the conversion.

The counting was performed through the use of *NucleoCounter* (see Section 2.3.9 for details about the instrument).

The results of this experiment are discussed in Section 3.4.1.

Sugar saturation experiment

The aim of this second set of experiments was to select the optimal concentration of sugar in order to assure that it was exceeding the limiting value for the growth of cells. In this way nutrients are not growth-limiting and any change in the behaviour of the growth is mainly driven by the cells. (See Figure 3.1 for literature on this topic).

To this purpose, different concentrations of sugars were added to the minimal medium and cell growth was monitored with the *Tecan* from the beginning up to the saturation phase. The temperature was set constant at 37°C. Section 3.4.2 is devoted to the results of such experiments.

Growth laws experiment

Once the right concentration of the sugar has been determined, the experiments for the extraction of RNA and proteins could be performed. In order to be sure to stop the growth of the population at the right moment of the growth, we followed the same protocol described for Torino's experiments. Some wells were followed by reconstructing the whole growth curve ("pilot" wells), while others were freezed in three different moments of the exponential phase. In this way, the time point corresponding to the desired moment of the exponential phase, possibly the half,

could be determined a posteriori, once the entire curve was obtained from the "pilot" wells. Then, RNA and proteins were extracted from the desired time point, by following the protocol described in Section 3.3.3.

Results in Section 3.4.3, show the growth rate trends obtained with the chosen sugars and temperatures.

3.3.3 RNA and Protein extraction

In both labs, the *NucleoSpin RNA protein* assay (MACHEREY-NAGEL) was used to extract the total RNA and protein content of each selected population of cells. The major steps of the protocol are here reported.

The freezed sample of cells was firstly unfreezed and joined with a buffer to obtain cell lysate. Cell lysate was subsequently filtrate through a column equipped with a membrane which allowed to separate RNA from proteins. In fact, the RNA, together with DNA, remained bound to the membrane, while the proteins flew through. At this point the two components were purified in order to eliminate all unnecessary elements and then quantified. By knowing the total number of cells from which the RNA and proteins have been extracted, it was possible to calculate the amount of such components per cell.

The preliminary results of the RNA and protein trends by changing the growth rates are presented in Section 3.4.3.

3.4 Experimental results

First of all we will discuss the calibration method adopted to convert *Tecan*'s absorbance data into cell density (Section 3.4.1); secondly we will present the results of the sugar saturation experiments (Section 3.4.2) which will be relevant for the following Section 3.4.3, dedicated to the trend of growth rates when different sugar, temperature and inoculi are considered and the first results of the RNA and protein quantification.

3.4.1 Calibration

As already mentioned, in order to compare the growth rates obtained through the *Tecan* with those coming from direct cell counting, we need a conversion. This has been obtained by following the Calibration experiments described in section 3.3.2.

The calibration has been performed in London's Lab for cells growing in glucose at 37 °C and starting from an inoculum $N_0 = 10^5$ cell/ml. Figure 3.5 shows an example of the growth curves obtained by direct counting (a), by *Tecan* with no background (b) and then (c) the logarithm of the concentration of cells of (a) as function of its corresponding absorbance ratio from (b). In Figure 3.5(c) two series of data are shown (each with a different shade of colour): they correspond to the data of both the growth curves measured to perform such calibration. The growth curves of the darkest dots are those shown in (a) and (b); those of the brightest ones are not shown.

Figure 3.5 (c) shows the relation between the two measurements. For both cell lines, two regimes can be recognised: a first one with positive slope (solid lines in Figure 3.5 (c)), and then a second regime where such linear dependence is lost. The first regime corresponds to the exponential phase of growth, while the second one to the saturation phase. This suggests that the quantity measured by the *Tecan* scales linearly with the logarithm of the number of cells as long as the saturation phase is not reached. Within this range, the *Tecan* measurements can be considered as a proxy for the logarithm of the number of cells. Since such relation does not hold for all the phases of growth, it is not possible to convert the entire growth curve obtained with the *Tecan* in terms of concentration of cells. This can be possible only for data points corresponding to the exponential phase. This is exactly the procedure that we follow to calculate the number of cells from which we extracted RNA and proteins. Indeed, by plugging into such linear relation the absorbance value corresponding to the time point at which the extraction is performed, it is possible to know the corresponding number of cells the population is composed of. This is useful in order to calculate the amount of RNA or protein per cell. To be noted that this conversion has been performed after the background subtraction of the absorbance data.

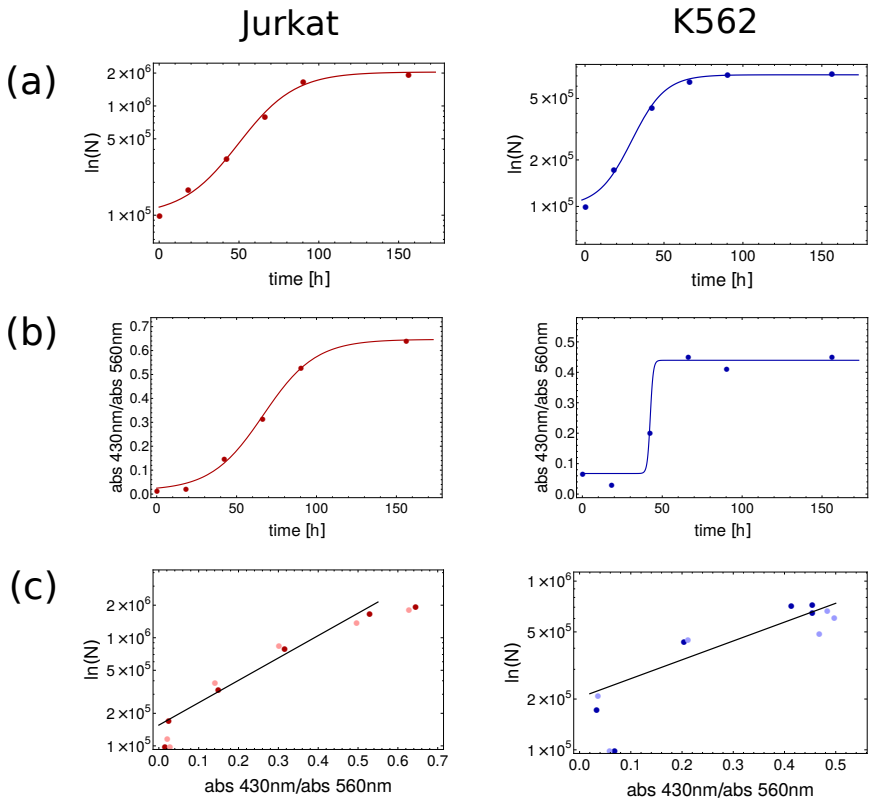

Fig. 3.5 The left panel relates to Jurkat and the right one to K562. Plots of the row (a) represent the logarithm of the number of cells per ml (N). Plots (b) show measurements of the same wells in (a) performed with the *Tecan*. Plots (c) reveal the relation between cell concentration and absorbance ratio. Here two sets of data have been joined to determine the conversion law. The darkest dots corresponds to the growth curves in (a) and (b), the brightest ones are the second set of data whose growth curves are not shown.

Moreover, we are interested in comparing exponential growth rates λ of the two sets of experiments (manual count and *Tecan*'s ones), thus we need a way to convert them. The presence of a positive linear trend between the absorbance and the logarithm of the number of cells (Figure 3.5(c)) is a crucial point for such conversion. Since we are interested in the growth rates of the exponential phase - linear in presence of a logarithmic scale - , we can calculate them, in both sets of data, by exploiting the fitting procedure discussed in Section 2.3.3 and then simply convert the λ values through a proportion.

According to the fitting procedure in 2.3.3, the growth rates were calculated, for both sets of growth curves Figure 3.5 (a) (λ_a) and (b) (λ_b), as the slope of the data within the exponential phase of growth. In order to perform the fits in the same way,

we needed to normalise also (b) curves to the initial value, as we did for curves in (a) where $ln(N/N_0)$ is the observable. Thanks to the fact that the absorbance scales logarithmically with the number of cells, the normalization was performed simply by subtracting the initial value (absorbance in $t = 0$ h) to each point. The values of the obtained growth rates are reported in the following Table 3.1, where we report also the values of the doubling times ($d.t.$) corresponding to the growth rates:

$$d.t. = \frac{ln(2)}{\lambda} \tag{3.10}$$

In order to justify why we are not considering the *Tecan's* growth rate directly as the growth rate of the population, we report the doubling time conversion also for λ_b ($d.t._b$). It is worth noticing how the $d.t._b$ values result almost compatible with the duration of the experiments, therefore they can not be used as representative of the growth rates, and a conversion is needed. As mentioned above, the exponential growth rates have been calculated by applying the fitting procedure described in Section 2.3.3 to both manual counting and *Tecan*'s growth curves. In presence of data with a very steep sigmoidal fit function (see for example K562 curve in Figure 3.5(b)), the exponential window determined analytically was too narrow. Therefore, in order to calculate the exponential growth rate λ, the slope of the line interpolating the first four points of the growth curve has been considered.

Cell line/parameters	λ_a (h^{-1})	$d.t._a$ (h)	λ_b(h^{-1})	$d.t._b$(h)
Jurkat	0.037	~ 18.9	0.0077	~ 90
K562	0.035	~ 19.5	0.0092	~ 75

Table 3.1 Values of growth rates (λ) and relative doubling times ($d.t.$) of manual counts (first two columns) and *Tecan*'s ($3^{rd} - 4^{th}$ columns) measurements. Lines correspond to cell lines.

The two values λ_a and λ_b are thus representative of the same quantity. Since the growth rates are calculated in the exponential phase, they are in the linear regime of conversion between the logarithm of the concentration and absorbance ratios. Therefore the following simple relation is valid:

$$\frac{\lambda_N}{\lambda_a} = \frac{\lambda_T}{\lambda_b} \tag{3.11}$$

where λ_N and λ_T are the growth rates calculated for manual counting curves and *Tecan*'s ones respectively. To note that Each cell line has its own λ_a and λ_b.

Since the direct conversion among the growth rates has been obtained only for cells growing in glucose (25mM) and temperature of 37°C, we can not assume it being the same if changing the growing conditions. Thus, in order to compare all the results, we express all the growth rates (λ_T and λ_N) as relatives to their corresponding reference values λ_b and λ_a.

A last important observation concerns Jurkat growth curves obtained by direct cell counting in London's lab. From the analysis, the growth rates obtained (see λ_a, Jurkat in Table 3.1) are consistent with those of Torino's curve grown in the same condition ($N_0 = 10^5$cell/ml, 25 mM of glucose and 37 °C). Thus, since the same cells, grown in two labs and in geometrically different plates showed consistent behaviours, on one side we can claim a good reproducibility of our experiments, on the other side we can be confident in comparing the results obtained in the two labs.

3.4.2 Optimal sugar concentrations

In order to identify the optimal non-limiting sugar concentration, we performed cell growth experiments by testing different concentrations for each chosen sugar. In Figure 3.6, for both cell lines and each sugar, the growth rates are plotted as function of the sugar concentration. In light of the discussion of the previous Section 3.4.1, the growth rates are expressed in terms of the reference one (λ_b). The darkest dots with the error bars represent the average and the dispersion of different replicates, whose values are shown by the brightest dots within the error bars.

The optimal sugar concentrations have been chosen as the values at which the growth rate is either maximum or saturated. In other words, for levels of sugar in the medium higher than the chosen ones, the fitness of the population of cells does not increase: either it remains constant (see for example the case of Jurkat in Glucose for concentrations higher than 25mM) or it decreases (Jurkat in more than 25mM of Mannose). Such optimal values are listed in Table 3.2.

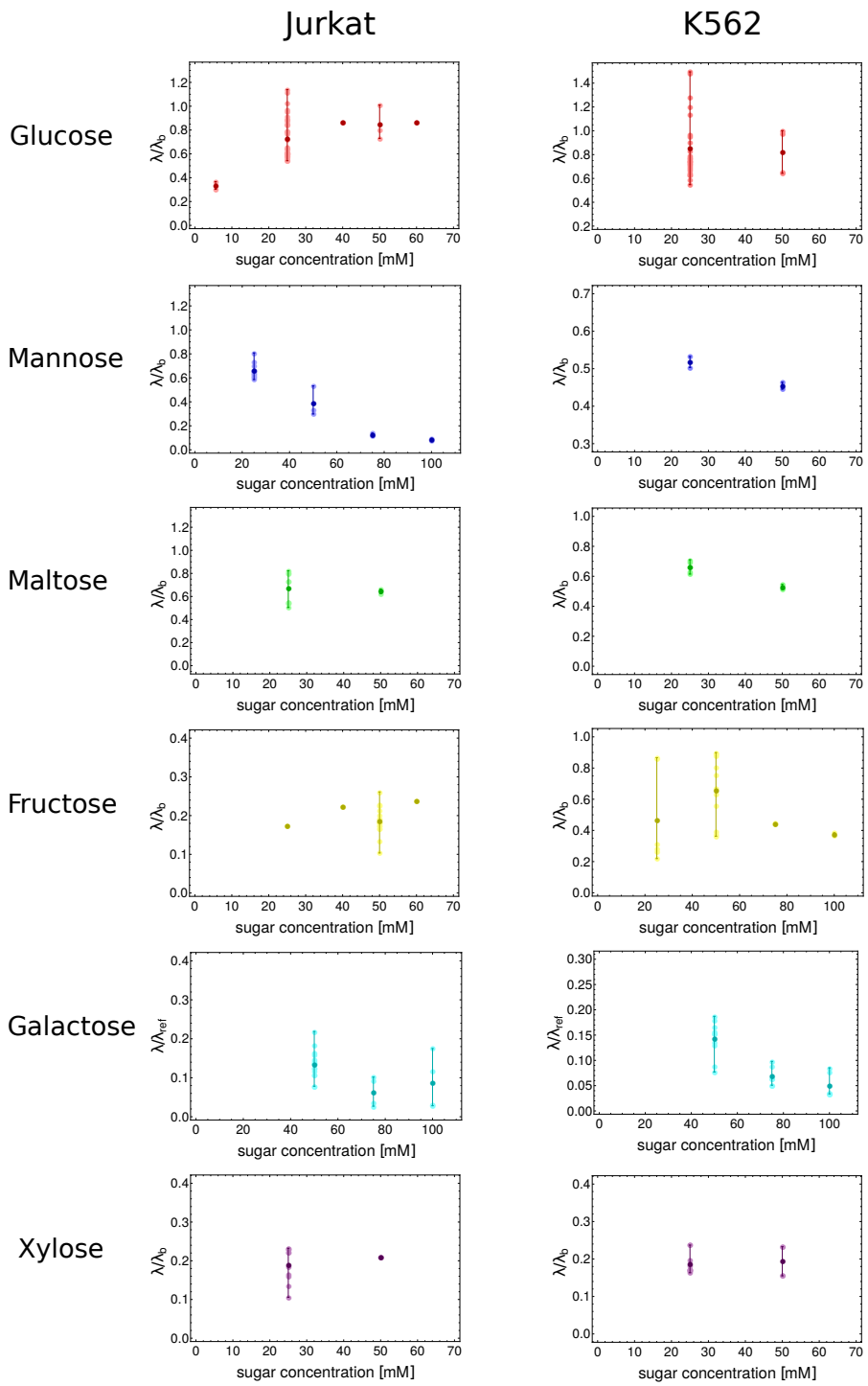

Fig. 3.6 Sugar saturation results for Jurkat (left) and K562 (right) cell lines. Each raw and each colour correspond to a different sugar. Data represented as dark filled circles are the average values of different replicates (brightest circles). The error bars show replicates dispersion. The growth rates are expressed as referred to the reference value for *Tecan*'s measurements λ_b.

	Gluc. [mM]	Mann. [mM]	Malt. [mM]	Fruct. [mM]	Gal. [mM]	Xyl. [mM]
Jurkat	25	25	25	50	50	25
K562	25	25	25	50	50	25

Table 3.2 Summary of optimal sugar concentration chosen for each cell line (rows) and each sugar (columns).

3.4.3 Towards the growth laws

With the optimal concentrations of sugar defined, it has been possible to run the first experiments for the quantification of RNA and proteins. The first step in this direction is to show how the growth rates are changing with the different conditions: sugar, temperature and inoculum. Afterwards, the very preliminary results of the extraction values are presented.

Tuning the growth rates

In this paragraph we present the dependencies of the exponential growth rates on the different conditions. The results of both Torino and London's labs are shown in Figure 3.7.

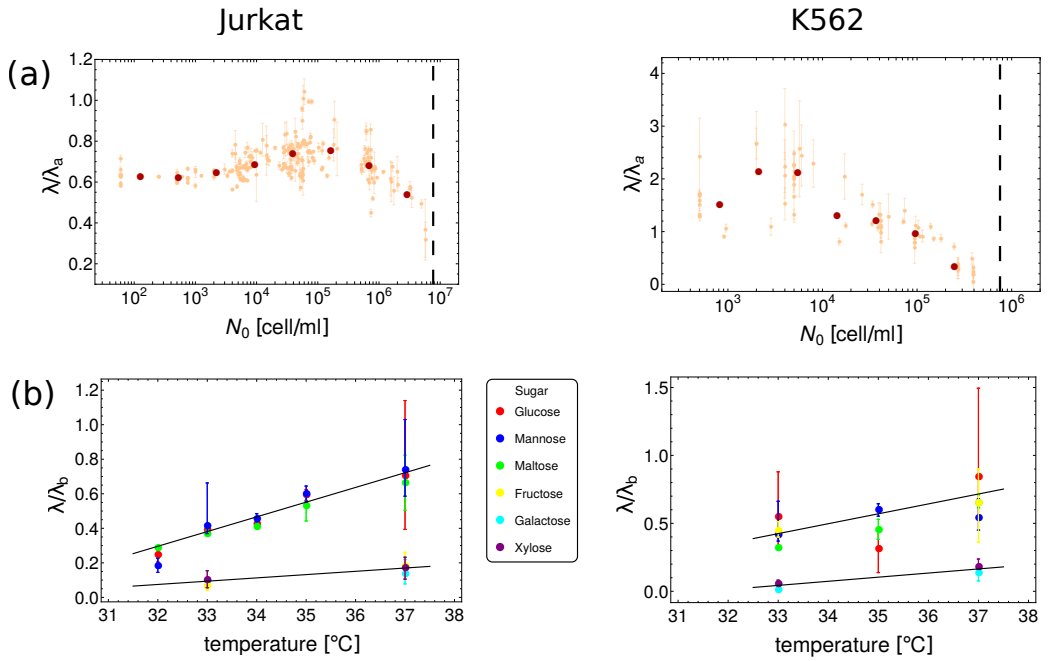

Fig. 3.7 (a) Jurkat (on the left) and K562 (on the right) growth rates (λ/λ_a) as a function of the inoculum size. The brightest dots are all the experimental data and their error bars are the errors obtained by the fitting procedure. The darkest circles represent average values obtained by binning over N_0. The data correspond to the experiments at 37°C with glucose (coloured in red, according to the legend of (b)). (b) Growth rates (λ/λ_b) as a function of temperature for Jurkat on the left and K562 on the right. Each colour corresponds to a different sugar. Black solid lines emphasise the trend of the data.

Figure 3.7(a), shows the growth rate dependence on the inoculum density for both cell lines. The plots are the very same presented in Figure 2.15(a) and Figure 2.24(e), with the values of the growth rates re-scaled (λ/λ_a) in order to make them comparable with results shown in Figure 3.7(b) obtained from Tecan's growth curves. The re-scaling has been performed by using the reference value for manual counts in the case of Figure 3.7(a) and that of Tecan for Figure 3.7(b). The trends shown by Figure 3.7(a) have been broadly discussed in sections 2.3.6 and 2.3.9, thus we remind there for details.

Figure 3.7(b), instead, shows the dependence of the growth rate on sugar quality and temperature for both Jurkat and K562.

First of all, focusing the attention (in all the plots of Figure 3.7) on the values of growth rates of cells seeded at initial concentration of $N_0 \sim 10^5$ cell/ml in glucose at 37 °C (red data), the values obtained with Tecan's measurements (Figure 3.7(b))

are in agreement with those obtained through manual counting (Figure 3.7(a)). This again shows the high reproducibility of our experiments.

By comparing now the growth rates of different sugars at fixed temperatures (Figure 3.7(b)), for both cell lines the data seem to divide into two main groups, namely with "high" and "low" values of growth rates. As emphasised by the solid black lines, both groups show an increasing trend with the temperature.

For both cell lines, at fixed temperature, the growth rates associated to glucose, maltose and mannose belong to the group of "high" growth rates and do not show significant differences among each other. Similarly, galactose and xylose that belong to the second group of sugars, present compatible results. Fructose is the only sugar that shows a different behaviour between Jurkat and K562. While for the first cell line it determines a slow growth rate, it is responsible for an higher growth rate when supplied to K562. As it can be noticed in the plots, results for growth rates are not yet available for all the desired temperatures, therefore it would be necessary to perform further experiments in order to drive better conclusions.

Despite of this, the so-far obtained results suggest the presence of a temperature dependence on the growth rates for both cell lines, i.e. Jurkat and K562. We can then argue that also for these cell lines the temperature may influence the growth, as it has been found for other human cancer cell lines in [100]. It would be interesting then to perform further experiments with other temperatures, higher for example, in order to deeper investigate such issue.

Furthermore, from the above discussed results, it seems that a population of Jurkat or K562 growing in an environment supplied with either glucose or mannose or maltose, grows with a growth rate that is only modulated by the temperature and not by the sugar quality. A similar conclusion can be driven when comparing cells living in galactose or xylose. Thus in order to have a modulation of the growth rates given by the sugars, it is necessary to choose sugars in one of the two groups that cause low and high growth rates. It would be interesting looking for additional sugars in order to enlarge the growth rate modulation due to sugar qualities.

Internal partitioning of resources

In this paragraph we show the results obtained from RNA and protein quantification. The experiments are still in progress, therefore Figure 3.8 (a) and (b) show the very

preliminary results obtained in London's Lab, while 3.8 (c) presents those of Torino's experiments(red circles) joint with the London's ones. All the shown results refer to cells growing at 37°C.

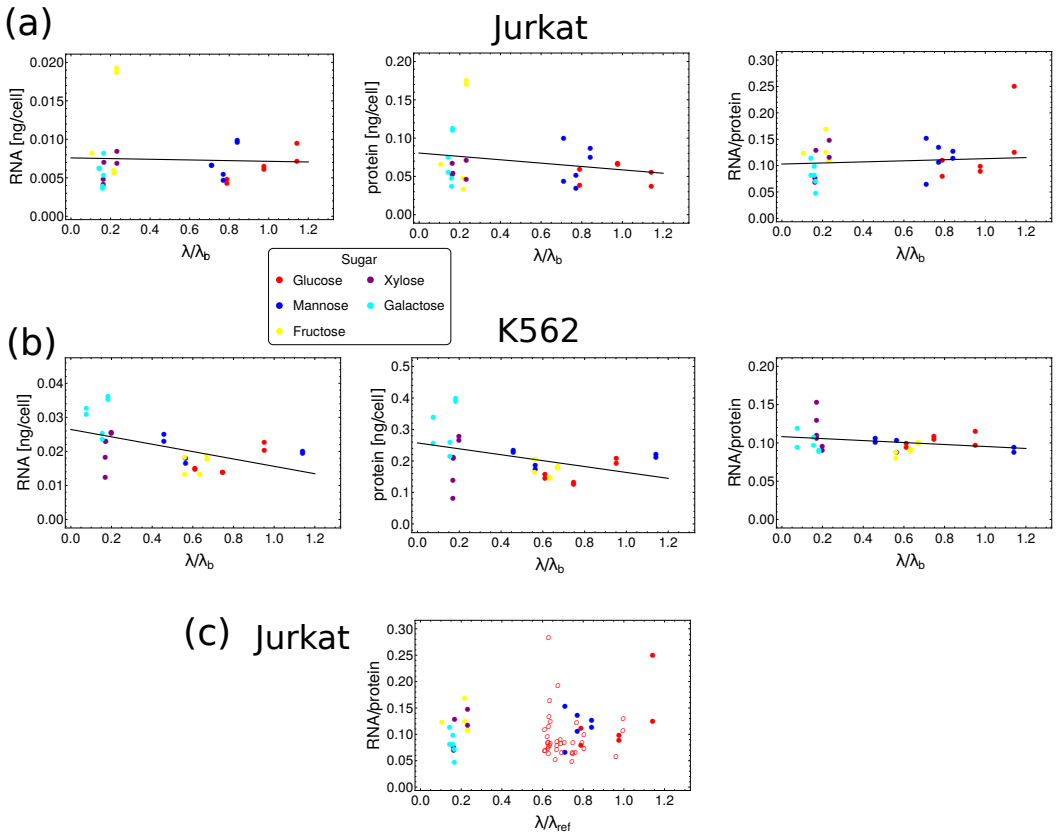

Fig. 3.8 Contents of RNA per cell, protein per cell and ribosomal fraction (RNA/protein) as function of the growth rate. (a) shows results of Jurkat cell line and (b) of K562, both at 37°C. Black solid lines represent the linear fits of the data. (c) RNA/protein data of Jurkat cells shown in (a) joint with the results obtained in Torino's lab from manual counting experiments shown as red open circles. The x-axis represents the growth rates normalised to the reference ones: λ_b for filled dots (Tecan's measurements), λ_a for open circles (manual counting data). Each colour corresponds to a single sugar, as explicited by the legend that refers to all the shown plots.

First of all, focusing the attention on the first two plots of Figure 3.8 (a) and (b), it is worth noticing that, since Jurkat cells are smaller than K562 cells (see descriptions in 2.3.9), they present smaller values of both RNA and protein per cell.

Secondly, in order to drive proper conclusions on the general trend of RNA, protein and ribosomal (i.e. RNA/protein) contents, further experiments would be

necessary to increase the statistics. However, from these first results, in both cell lines, the protein content per cell seems to decrease with the increasing of the growth rate, while the RNA seems to remain roughly constant for Jurkat and decreasing for K562. This results then in two slightly different trends in RNA/protein contents. In Jurkat cells it seems to be very slightly increasing since the decreasing trend of protein per cell prevails over the constancy of RNA per cell. Conversely, K562 cells show a slightly decreasing trend of RNA/protein, suggesting then that the RNA content decreases lightly faster than the protein one.

In the end, in Figure 3.8(c) we highlight the high reproducibility of our experiments. Indeed, the open red circles in the plot refer to the ribosomal contents of Jurkat cells grown and quantified in Torino's lab where the growth rate was modulated through the inoculum density. The very fact that the results are compatible with those obtained in London's lab gives a further proof of the fact that the experimental results are robust with respect to the used protocol for growing and counting cells.

3.5 Conclusions and perspectives on growth laws

In previous sections we presented the preliminary results obtained for the study of the impact of cell growth on gene expression for two human cancer cell lines, i.e. Jurkat and K562. Before reaching the final preliminary results that would allow us to relate experimental values representative for gene expression to those representative for cell growth, analogously to what have been performed for bacteria [5], several necessary steps have been discussed.

Firstly, similarly to several studies in bacteria, we determined the variables to look at: the RNA/protein as representative of gene expression and exponential growth rate as a proxy for cell fitness.

Secondly, we found a method for comparing the interesting growth curve parameter, i.e. the growth rate, obtained with two different experimental strategies. Indeed, on one side we directly counted the cells, on the other side we used the *Tecan* plate reader, an instrument that indirectly estimates cell growth through measurements of the absorbance of the growth medium. The use of such instrument has the advantage to allow high throughput but, as drawback, it lacks a direct measurement of cell concentration. Knowing the number of cells at a given time is a necessary information

when the extraction of RNA and protein occurs, in order to estimate the amount of such quantities per cell. Thanks to the conversion method that we developed, we have been able to estimate the number of cells in the colony in correspondence of the exponential phase of growth, starting from absorbance measurements.

Thirdly, we performed sugar saturation experiments for the identification of the optimal concentration of sugar to add to the minimal growth medium. In this way cell populations were able to grow in a non-limiting sugar environment, thus possible differences in their growth rates could be attributed to differences in the sugar qualities, and not in their concentrations.

Then, fixed the optimal sugar concentration, experiments at different temperatures (and fixed sugar quality and inoculum) have been carried out for both cell lines. It emerged that, among the chosen sugars, two main growth rates could be spanned: for both cell lines, cells growing in media supplied with glucose, mannose and maltose showed growth rates faster than those supplied with galactose and xylose. A difference among the two cell lines have been identified by the fructose. Indeed, K562 cells supplied with fructose grew as fast as those supplied with mannose, while Jurkat cells in fructose presented growth rates compatible with those of galactose and xylose. In this way, two groups of sugars have been identified: one that determines "slow" growth rates and the other "high" ones. In order to better modulate the growth rate with sugar quality, it would be interesting to explore a larger set of sugars and hopefully find more "groups" of sugars that define different growth rates.

An interesting finding concerns instead the modulation of the growth rate with the temperature. Indeed, we found that, for both the above cited groups of sugars, the increasing of the temperature gives an increase of the growth rates. Such result is valid for both the investigated cell lines and suggests a possible growth dependence on the temperature, as already found for other cancer cell lines in [100]. All the results discussed so-far are relative to one inoculum density, however it would be worthy investigating the influence of sugars and temperature on cells seeded at different inocula.

The last part of this chapter was dedicated to the very preliminary results of RNA and protein extractions, namely the experiments through which we investigate RNA and protein contents of cells in order to look for possible empirical relations between RNA/protein ratio and growth rate. We presented the results of RNA and protein quantification by fixing the temperature (37°C) and the inoculum size

(N_0=10^5cell/ml) and modulating the growth rate with the sugar quality. The work is still in progress, but the preliminary results suggest that for both cell lines the amount of protein per cell decreases as the growth rate increases, while the RNA one either remains constant (Jurkat) or it decreases (K562). These result in a constancy or slightly increasing trend of the ribosomal content with the increasing of the growth rate. The main aim of this work is to make comparisons with the bacterial growth laws where the RNA/protein content significantly increases with the growth rate. In order to reach such goal and drive statistically relevant conclusions for the chosen cell lines, we need to increase the number of experiments and conditions especially in the direction of trying to enlarge the range of investigated growth rates. However, such preliminary results seem to be promising in showing that even for such mammalian cells possible relations between cell contents and cell fitness could be found.

In conclusion, in light of the results discussed in this Chapter, we can claim that the experiments we performed were highly reproducible and the results we obtained are robust independently on the methodology used for monitoring the growth of the population of cells.

The results presented in this Chapter, will be part of a publication in preparation together with Dr. Francesca Ceroni and Dr. Carla Bosia.

Part II

Modelling and experimental test of miRNA-target interaction

Gene expression and post transcriptional regulation

In eukaryotes the expression process of a coding gene involves three main steps: transcription (from DNA to mRNA), splicing (mRNA maturation) and translation (from mature mRNA to protein). Gene expression is known to be tightly regulated: all cells of an organism carry the entire set of genes, but they only express less than the 2% of their genes at any particular time. The process through which a cell increases or decreases the production of a specific gene product (protein or RNAs) goes under the general name of *gene regulation*. Among the various gene regulatory steps, the most important ones are (i) the transcriptional control, that occurs in the nucleus and (ii) the post-transcriptional control, that instead takes place in the cytoplasm. The former is one of the best characterised regulatory mechanisms and it is carried out by transcription factors, proteins that through binding to the promoter of a target gene can enhance or impede its transcription into RNA. Since transcription factors are themselves proteins produced by other genes, from the cooperative and combinatorial features of such regulation a complex network of interactions among genes and between genes and their products arises.

The second gene regulatory mechanism is instead the main topic of this part of the thesis. Post-transcriptional regulation has been discovered and characterised more recently than transcriptional regulation [102], and among other players, it is mediated by small (22 nucleotides-long) non-coding RNA molecules, called microRNAs (miRNAs) [103]. MiRNAs act in the cytoplasm and, after transcription, bind to other RNA molecules through Watson-Crick base pairing. Once bound, the miRNA - which is here in its mature stage, that is uploaded into a macromolecular complex called RISC (RNA Induced Silencing Complex) - may either induce the

degradation of the RNA molecule or reduce its translation into protein. This leads to a downregulation of the expression of the corresponding gene.

The interaction between miRNA and their target RNAs, have been investigated in several studies and is based on molecular sequestration [9–11, 34]. Once bound, the two molecules are sequestered: the formation of the complex miRNA-target prevents any further interaction until dissociation. Such interaction is titrative and induces a threshold-like response in the mean amount of free molecules as a function of some key parameters, like for instance, their production rate. Depending on the relative amount of the two species, two different regimes may be recognised: repressed and unrepressed. If the amount of miRNA (the sequester) is larger than the amount of its target mRNA, most of the target molecules are expected to be bound and thus their free amount is close to zero (repressed state). In the opposite case, when the abundance of target exceeds that of miRNAs, some mRNA molecules will be free and thus can be translated, leading the system to the unrepressed state. In this case, the amount of free molecules increases linearly with their total abundance. The region close to the equimolarity point is the so-called *ultrasensitive* region [104]. Here, a threshold between the repressed and the unrepressed state arises and thus small changes in the relative amount of the two species may cause a big fold change in their free amount.

A further important feature of miRNA-mediated regulation is the fact of being combinatorial: one single miRNA can regulate many different targets and one target can be regulated by several miRNAs [105]. This suggests a whole new layer of post-transcriptional cross-regulation named *Competing Endogenous RNA (ceRNA) effect* [106]. The idea at the base of such theory is that the amount of a gene product may be tuned by varying the concentration of another transcript that shares with the first one the same miRNAs. The increasing, for instance, of the expression of one target has the net effect of enhancing the expression of the other targets of the same miRNAs. Indeed, this acts as a sponge, sequesters the miRNAs and prevents them to bind to other targets. Such system has been widely studied from a theoretical point of view [34–36, 107]. Both the description of the crosstalk among the targets and the prediction of the coupling of their fluctuations have been investigated through the use of stochastic models. By taking into account the transcription of both miRNA and targets, their degradation and their interaction, a master equation can be associated to the system. At the steady state, it is possible to obtain approximated solutions for the moments of the distribution whose variation in time is given by the master equation,

namely mean values and variances for the molecular species in the system. These are predictions that have been then validated experimentally on an *in vitro* system [10]. In particular, a synthetic circuit consisting of two genes sharing the same miRNA binding sites was transfected in a population of mammalian cells that endogenously expressed the desired shared miRNA. By measuring the relative expression of the two transfected targets it was possible to measure their miRNA-mediated crosstalk. The results of such experiments were in agreement with the model predictions and gave rise to new questions concerning the appearance of distinct phenotypes in the distribution of the expression of the targets.

A remarkable feature of molecular networks lies in their intrinsic stochasticity [108]. Since chemical reactions are probabilistic and the number of molecules involved is often small, fluctuations in the abundance of such molecules become relevant. Moreover, a second source of noise usually influences such systems. This arises from all those processes "external" with respect to the system under study and it is usually referred to as extrinsic noise. The combination of both intrinsic and extrinsic noise gives rise to phenotypic heterogeneity in a population of identical cells. Such variability may organise in bimodal distributions, whose modes are representative of different phenotypes. Bimodal distributions may be found in different processes like cell differentiation and cell-decision making or even underline the existence of a healthy and a deseased phenotype [109].

This part of the thesis fits the above described framework. Indeed, we aim at investigating the rising of bimodal distributions in a miRNA-target system in presence of extrinsic noise. In particular, we continue the work presented in [11] by experimentally testing the predictions of the mathematical model discussed in [8]. Before going into details of the theoretical results, in Chapter 4 we give a brief overview of what the microRNAs are, how they interact with their targets, what is the role of the noise in such system and then we briefly present the ceRNA effect. Afterwards, in Chapter 5, we present the results discussed in [8] by going into details of the stochastic model that describes the system and how it can be treated in presence of extrinsic noise. Finally, Chapter 6 is devoted to the experimental activity performed in order to test and reproduce the main results obtained in [8]. The work is still in progress, therefore after a description of the experimental setup, we discuss the preliminary results obtained so far.

Chapter 4

Overview of microRNA-mediated gene regulation

In this chapter we discuss the main features of microRNA-mediated gene regulation by dedicating a section to each main subject. By starting from the definition of what microRNAs are (Section 4.1), we move to the description of their interaction with targets (Section 4.2). The considered reactions are probabilistic and the environment in which they occur is noisy. Since this has a pivotal influence on the system under exam, we devote Section 4.3 to discuss the ways different sources of noise may affect such system. In the end, we describe the effect of competing endogenous RNAs that rises from the presence of several mRNA targets sharing the same pool of miRNA (Section 4.4).

4.1 Overview on microRNAs

microRNAs (miRNAs) are 22 nucleotides-long non-coding RNAs that operate post-transcriptionally via sequence-specific binding to target RNAs and play a role in the regulation of eukaryotic gene expression [110, 111]. Figure 4.1, adapted from [7], shows miRNA biogenesis and its mechanism of binding to target RNAs. Following transcription and maturation in the nucleus, miRNAs get incorporated into specialised, multiprotein complexes known as RISCs (short for RNA-induced silencing complexes) [112]. Once within a RISC, the miRNA provides the pattern to bind specific sites called miRNA response elements (MREs) found on its target RNAs [112].

Effective base pairing typically requires 6- to 9-nucleotides of complementarity. As long as the miRNA is bound to its target, the former may promote the degradation of the latter or reduce its translation into protein. This is the reason why miRNAs are considered as post-transcriptional down-regulators of gene expression. As it will be discussed in the following Section 4.2, the interaction between miRNA and target is based on the mechanism of molecular sequestration [8–10, 34, 107].

More than 25000 miRNAs have been identified up to now in viruses, plants and animals and about 2500 in human cells [113]. MiRNAs are well conserved in animals and plants and take an active part in a wide range of regulatory mechanisms involved in cell proliferation [114], organogenesis and differentiation [115] and apoptosis [116]. Moreover their dysregulation can lead to diseases, among which also cancer [117] is included.

A further feature of miRNA-target systems is its combinatorial nature: every single miRNA can regulate several targets and one target can be regulated by different miRNAs [110]. More than the 60% of the human genome has been predicted being targeted by microRNAs [118]. The variety of targets the microRNAs regulate is large and important for different signalling pathways and developmental stages, indeed it is thought that an alteration of their expression may contribute to tumor development and metastatisation [119]. A last remarkable consequence of such combinatorial feature together with the nature of interaction between miRNA and targets, is the rise of competition among mRNAs to bind to the same miRNA. Such competition is in turn a source of crosstalk among the competing species [106]. This crosstalk is also known as Competing Endogenous RNA (ceRNA) effect and it is described in Section 4.4.

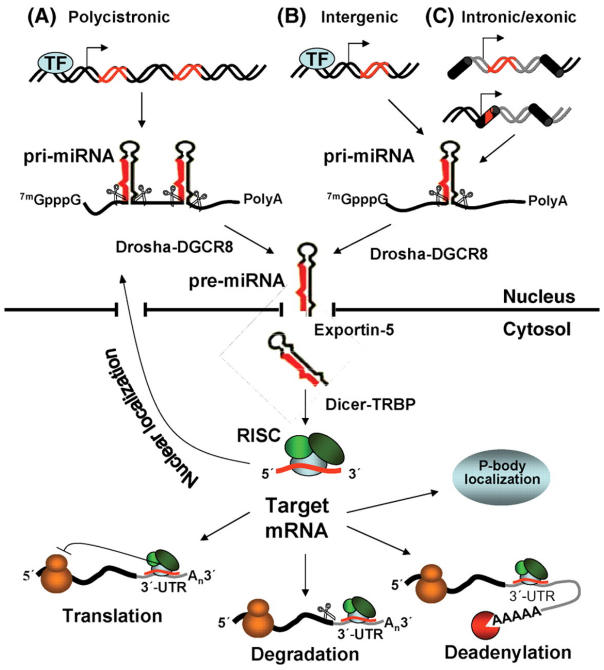

Fig. 4.1 MiRNA biogenesis and target binding. Adapted from [7].

4.2 MiRNA-target interaction is titrative

Several studies suggested that the interaction between miRNA and their targets is based on the mechanism of molecular sequestration [8–11, 34, 35, 107]. This can be easily understood by thinking that when a miRNA is bound to its target, the complex that is formed avoids the possibility of further interactions with other molecules until its dissociation. Furthermore, as long as a mRNA is bound to a miRNA, the former cannot be translated. An example of such interacting network is shown in Figure 4.2(a) adapted from [8], where R and S are the mRNA of the miRNA and the target respectively. Only those targets that are not bound to the microRNAs are free to be translated into proteins (P).

It has been found that the behaviour of the system depends on the relative abundances of the two species, in other words it is a titrative interaction [9]. Indeed, if the quantity of miRNA is higher than that of mRNA, then all mRNA are expected to be bound and not expressed. In this case, the system is said to be in the bound or repressed state (see Figure 4.2(b)). On the contrary, if the amount of miRNA is lower than that of mRNA, the latter would be free for translation and thus the system will be in its unbound or expressed state. So, it is possible the definition of a threshold for the transcription rate of the mRNA (k_r in Figure 4.2(a)) such that the system is either in one or in the other state, characterised by low or high level of expression of mRNA. In particular, such threshold behaviour of the expression of mRNA is strongly dependent on three parameters: the transcription rates of (i) miRNA and of (ii) mRNA and (iii) the interaction strength between miRNA and mRNA. The last parameter is related to the affinity between the two species and the number of biding sites for the miRNA present on the target. This determines the steepness of the threshold: the higher the interaction strength, the sharper the transition between the two states. By varying the above cited three parameters, the two regimes may be explored. The presence of the threshold underlines the non-linear transition between the two regimes. Such threshold is located close to the equimolarity point, namely where the abundance of mRNA molecules roughly matches that of miRNAs.
In the proximity of the threshold both the free miRNA and targets are low in number and their fluctuations are strongly coupled. Thus, small fluctuations in their amounts may lead the system to jump from the repressed to the expressed regime.

Such threshold-like behaviour has been studied from both experimental and theoretical point of view in [9]. Moreover the authors explain the mechanism in terms of molecular titration [104]. Given the goal of our work, we dedicate Section 6.1.1 to the description of the experimental protocol adopted by the authors.

In Chapter 5 we will go into the details of the network depicted in Figure 4.2(a) and discuss the mathematical framework that leads to such behaviour.

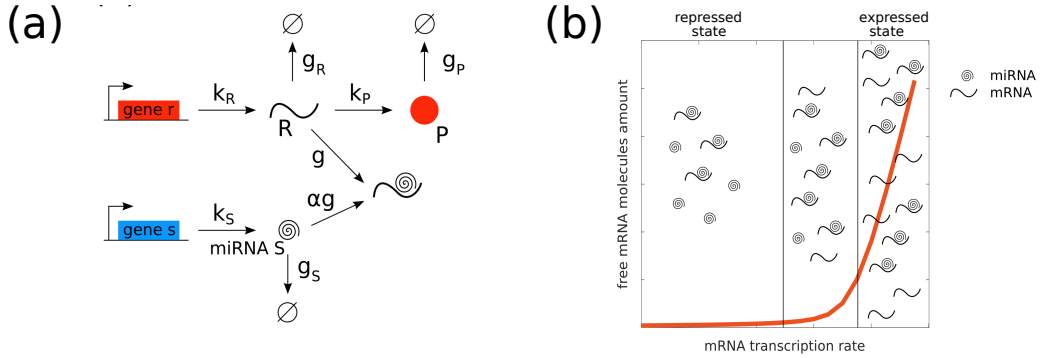

Fig. 4.2 (a) Network of interaction between microRNA (S) and its target mRNA (R). k_R and k_S are the transcription rates of target mRNA and miRNA respectively and g_R and g_S their degradation rates. k_P and g_P are respectively the transcription and the degradation rates of the protein (P). g is the interaction strength between the miRNA and the target and α is the fraction of not-recycled miRNA after the binding. Adapted from [8]. (b) Cartoon of threshold-like behaviour describing miRNA and target mRNA interaction. The amount of free mRNA is shown as a function of its transcription rate. Below the threshold the amount of free mRNA molecules is smaller than that of miRNA (repressed state); the two quantities are roughly equal around the threshold and above the threshold the free amount of mRNA is higher than that of miRNA (expressed state). Adapted from [8].

4.3 The role of intrinsic and extrinsic noise

As anticipated, a molecular network is characterised by an intrinsic stochasticity [108] given by the probabilistic nature of chemical reactions. Such intrinsic stochasticity is referred to as intrinsic noise and it can be considered as negligible in macroscopic systems - where large number of molecules are involved in the reactions -, but it becomes relevant in biological systems where the numbers of involved molecules are small. The miRNA-target system enters such latter scenario. In such system, close to the threshold between the unbound and bound state (see Section 4.2 for details), due to intrinsic fluctuations of the amount of both species, a subset of the targets will be bound to the miRNA and another one will be unbound. By considering such situation in terms of the distribution of the target, it leads to bimodal distributions: the two modes are associated with the two states of the system, i.e. bound and unbound. It is important to highlight that such bimodality is not due to molecular mechanisms that induce multiple deterministic stable states in the system, but it simply rises from the presence of noise.

A second source of noise is known to play a pivotal role in such systems. It is the so-called extrinsic noise and refers to all those noise sources external to the system itself. Some examples are the fluctuations of the external environment that may affect identical systems in different ways or even fluctuations in the space where the biochemical reactions occur. Moreover, also variations of the amount of molecular machineries or cellular components can be considered as an extrinsic source of noise. This is known to influence gene expression and regulation at different levels resulting in a large variability across a cell population. However, such variability seems also to be dominated by the dynamics of the population itself [120]. For example, if a monoclonal population of cells is considered, it presents cells in different phases of their cell cycle, due to the processes of growth and division. It is known that cell-cycle regulators are controlled by miRNAs, miRNA regulation can be in turn cell-cycle dependent [121] and so its expression may depend to the specific phase of the cell cycle [122]. In such a scenario then, the levels of miRNA may fluctuate from cell to cell depending on the specific cell cycle phase. So, in light of this, the heterogeneity with respect to the cell cycle within a population of cells can be considered an example of extrinsic noise. Such feature will be accounted in the experimental investigation of the role of extrinsic noise in miRNA-target interaction discussed in Section 6.2.

As it will be broadly discussed in Chapter 5, it has been found that such extrinsic noise on miRNA expression may induce bimodality on the targets for certain levels of noise and stoichiometric conditions [8].

The bimodality raised by the sole intrinsic noise can be considered as a "single-cell" effect, since if certain constraints are fulfilled then every single cell can jump between the expressed and the repressed target state. Conversely, in presence of the extrinsic noise, the amount of miRNA may be heterogeneous through the different cells within a population (as discussed in [8]), thus bimodality can be seen at a "population level". Such topic will be broadly discussed in the following Chapter 5 since it is the main result we addressed experimentally.

4.4 Competing Endogenous RNAs (ceRNAs)

The combinatorial action of miRNAs means that one miRNA can regulate several targets and one target can be regulated by several miRNAs. Moreover, miRNAs

can target both protein coding and non-coding transcripts. This gives as a result the existence of a potentially enormous miRNA-target interaction network. From here, rised the idea that the miRNA-target interaction network could be based on the competition between targets for binding to the same pool of miRNAs and that this competition could be a source of crosstalk between the competing targets. Such crosstalk between targets of a common pool of miRNAs is known as *competing endogenous RNA* (ceRNA) effect [106]. Together with the combinatorial nature of the system, the mechanism of molecular sequestration underlies such phenomenon: when interacting with a target mRNA, a single miRNA cannot bind to any other target. Thus, the miRNAs available for the binding are the limiting factor in a system composed of potentially interacting mRNAs. To better understand such phenomenon, we refer here to a clarifying example described in [34]. Consider a gene A that shares a miRNA with gene B. If A is upregulated, it means that its concentration is increased, thus the miRNA preferentially binds to it. As a consequence, the mRNA of gene B will result being less repressed and thus its concentration will result higher. Thus, thanks to the presence of a shared pool of miRNA, it is possible to indirect regulate the expression of one target by tuning another one that shares the same pool of miRNA with the former.

Such cross talk effect has been observed in bacteria, where the role of miRNA is played by small RNAs (sRNAs) and a titrative interaction among sRNA and their target is present [123]. Moreover, the importance of such ceRNA effect has been also experimentally proved in several pathological ad physiological conditions, including in cancer [106, 124].

Several theoretical studies characterised such ceRNA effect through the use of mathematical modelling [34, 107]. In particular, among other results, in [34], the authors developed a mathematical model taking into account the intrinsic stochasticity of the system and recover the threshold-like behaviour of both miRNAs and ceRNAs as a function of the transcription rates of both the species. Moreover they also showed the above described crosstalk induced by the competition.

Chapter 5

Modelling the role of noise in miRNA-target interaction

This chapter is devoted to the main results published in [8] and partially described in [125], focusing on bimodal distributions in presence of extrinsic noise. These results will be then addressed expreimentally in the next Chapter 6.

5.1 Stochastic model for miRNA-target interaction with extrinsic noise

The cartoon depicted in Figure 4.2, which involves a miRNA (S), its target mRNA (R) and the corresponding translated protein (P) is the prototypical miRNA-target system considered in [8]. The species S and R are assumed to be transcribed from independent genes at constant rates k_R and k_S respectively. The intermediate reactions that lead to the synthesis of the mRNA are neglected. Moreover, the authors assume that miRNAs and mRNAs can be degraded by reactions governed by mass action-laws with rates g_S and g_R. After their transcription, miRNA and mRNA can bind into a complex (RS). The interaction between miRNA and target is quantified by the parameter g that takes into account the strength of the coupling between the two species, that is, from a biochemical point of view, the affinity of the two molecular species and the number of miRNA binding sites on the target. The mRNAs are always degraded due to the titrative interaction, while the microRNAs can be recycled with

probability $1-\alpha$ (α is the not-recycling probability). Finally, the mRNA not bound to miRNA can be translated into protein (P) with a translation rate k_P and the protein can in turn be degraded with degradation rate g_P.

The molecular reactions associated to the above processes and the production and degradation of miRNAs are the following

$$\emptyset \xrightarrow{k_S} S, \qquad S \xrightarrow{g_S} \emptyset \tag{5.1}$$

The production and degradation of mRNA are

$$\emptyset \xrightarrow{k_R} R, \qquad R \xrightarrow{g_R} \emptyset; \tag{5.2}$$

The interaction between miRNA and target mRNA can either lead to the degradation of both the molecules:

$$R + S \xrightarrow{g\alpha} 0; \tag{5.3}$$

or to the degradation of the mRNA and the recycling of the miRNA:

$$R + S \xrightarrow{g(1-\alpha)} S. \tag{5.4}$$

The translation of the mRNA into protein and the protein degradation are:

$$R \xrightarrow{k_P} R + P, \qquad P \xrightarrow{g_P} \emptyset. \tag{5.5}$$

Both the intrinsic and the extrinsic noise are accounted in the model. The intrinsic noise is defined as the fluctuations due to the stochasticity of the chemical reactions that occur with constant rates. The extrinsic noise instead is taken into account by considering fluctuations over the transcription rate of miRNA (k_S). The system is described by the following master equation that takes into account all the above mentioned reactions:

$$\begin{aligned}
\frac{dP(n,t)}{dt} &= k_R[P(n_R-1,t)-P(n,t)] \\
&+ \frac{g_R}{V_{cell}}[(n_R+1)P(n_R+1,t)-n_R P(n,t)] \\
&+ k_s[P(n_S-1,t)-P(n,t)] \\
&+ \frac{g_S}{V_{cell}}[(n_S+1)P(n_S+1,t)-n_S P(n,t)] \\
&+ \frac{k_P n_R}{V_{cell}}[P(n_P-1,t)-P(n,t)] \\
&+ \frac{g_P}{V_{cell}}[(n_P+1)P(n_P+1,t)-n_P P(n,t)] \\
&+ \frac{g\alpha}{V_{cell}^2}[(n_S+1)(n_R+1)P(n_R+1,n_S+1,t)-n_R n_S P(n,t)] \\
&+ \frac{g(1-\alpha)n_S}{V_{cell}}[(n_R+1)P(n_R+1,t)-n_R P(n,t)].
\end{aligned} \tag{5.6}$$

Where V_{cell} is the cell volume and $n=(n_S,n_R,n_P)$ is the number of molecules of the three X species under exam that relate to the concentration ρ_X according to $n_X = V_{cell}\rho_X$.

The authors obtained approximated expressions for the first and second moments of $P(n,t)$ by using Van Kampen's system-size expansion [126].

To investigate the role of extrinsic noise, they considered a fluctuating miRNA transcription (k_S) drawn from a Gaussian distribution $P(k_S)$ with mean $\overline{k_S}$ and variance $\sigma^2_{k_S}$. The intensity of the noise is quantified through the coefficient of variation, defined as $CV = \frac{\sigma_{k_S}}{\overline{k_S}}$. With the introduction of such additional source of noise, equation 5.6 did not hold anymore. The issue has been solved by considering the law of total probability, namely the probability distribution of the entire system $P(n)$ has been written in terms of the conditional probabilities as follows:

$$P(n) = \int P(k_S)P(n|k_S)dk_S \tag{5.7}$$

where $P(k_S)$ is the above mentioned Gaussian distribution of the miRNA transcription rate k_S and $P(n|k_S)$ is the conditional probability of observing a given system configuration n, given k_S. Such probability distribution is a solution of eq. 5.6 for any known k_S. At this point the van Kampen's expansion can be applied on

$P(n|k_S)$ to obtain the moments of such distribution. The full solution is finally given by the average of the results over the values of k_S.

5.2 Bimodality as a population-level effect

Both the intrinsic and the extrinsic noise affecting the system may give rise to bimodal distributions of the target. Given the threshold-like behaviour between miRNA and target and depending on the relative abundance of the two species, the system may be either in the repressed or in the expressed regime (see Section 4.2). The threshold appears at the quasi-equimolarity condition. Here, the intrinsic noise allows the system to jump from one regime to the other in a stochastic manner giving rise to bimodal target distributions that may appear for small ranges of values of the target transcription rate k_R, as discussed in [34]. Such bimodality is a characteristic at the single cell level: every single cell may jump between the two states if the interaction strength (i.e. the coupling constant) between miRNA and target is strong enough. In presence of extrinsic noise, when the population of cells is heterogeneous with respect to the miRNA transcription rate, every cell presents rates either above or below the threshold. In this case the bimodal distribution of the targets arises at the population level. Such bimodal distribution is given by the superposition of single cell unimodal distributions obtained with different values of transcription rate of the target (k_R). So, in presence of the extrinsic noise, the range of target transcription rate for which the bimodality appears is wider than the case with only intrinsic noise. Indeed, in the latter case a fine tuning of the miRNA and target transcription rates is necessary, while in presence of extrinsic noise the effect is visible even with high values of mRNA transcription rates and small miRNA-target interaction strengths. Moreover, also the separation between the two peaks - corresponding to the bound and unbound targets - is greather than in case of pure intrinsic noise.

The so far discussed widening of the range of the parameters is clearly shown by Figure 5.1(a) adapted from [8]. Figure 5.1(a) represents a phase diagram of the bimodality in the distribution of the free molecules of target mRNA obtained through Gillespie's simulations. The three variables considered in the plot are the following: (i) on the x-axis, the transcription rate of the mRNA (k_R); (ii) on the y-axis, the quantification of the extrinsic noise, namely the coefficient of variation of the distribution of the transcription rate of the miRNA (CV); (iii) the colour

map, the presence of the bimodality for different values of the strength of the interaction strength (parameter g of the model). Such last value is represented from a biochemical point of view, by the number of binding sites of the miRNA present on the target mRNA.
In presence of pure intrinsic noise, a bimodal distribution is obtained only for high values of the interaction strength between the two species and the bimodality region enlarges when increasing g. Therefore, the interactions strength influences the range of k_R for which bimodality is present.
When in presence of the extrinsic noise the constraint on the strength of the interaction relaxes. Indeed, it is the heterogeneity of the miRNA transcription rates that leads some cells being in the repressed state while other in the expressed one. Moreover, when increasing the intensity of noise (i.e. $\sigma_S > 0$) the bimodality is present for a wider range of values of the mRNA transcription rate.

In light of the so far discussion, we can say that both the interaction strength g and the extrinsic noise influence the range of values of k_R for which a bimodal distribution is present, thus they can be considered two "knobs" to control the bimodality. Indeed in [8], it is evidenced how the two act at similar levels: in order to obtain bimodal expression distributions, a high level of extrinsic noise can compensate for a weak coupling. Such features are shown in Figure 5.1(b) through results of Gillespie's simulations [127]. Here the authors highlight the possibility of obtaining bimodal distributions either by increasing the extrinsic noise(orange histogram) or through increasing the interaction strength between miRNA and target (blue histogram) with respect to the case of reference with only intrinsic noise (no extrinsic one) and a low interaction strength between the two species (i.e. low number of miRNA binding sites on the target mRNA).

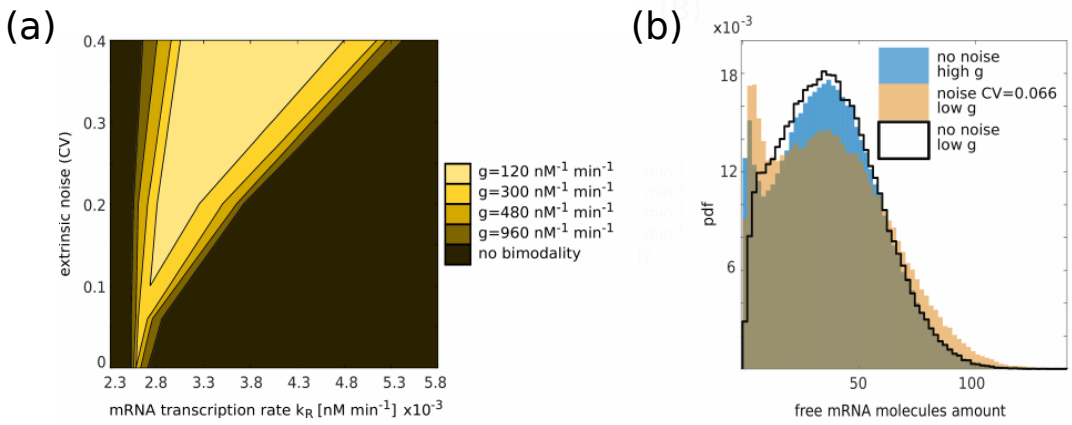

Fig. 5.1 (a) Phase diagram of the bimodality in the distribution of free molecules of target mRNA. The extrinsic noise on the miRNA transcription rate (CV) is plotted as a function of the target transcription rate (k_R). The colour map shows the presence of bimodality for different values of the miRNA-target interaction strength g (see the legend). Adapted from [8]. (b) Distribution of free mRNA in the case of: pure intrinsic noise and small miRNA-target interaction strength g (solid black line); pure intrinsic noise and high miRNA-target interaction strength g (blue histogram) and extrinsic noise and small interaction strength g (orange histogram). Adapted from [8].

A further interesting result discussed in [8] concerns the role of extrinsic noise in the ceRNA effect. As described in the dedicated Section 4.4, given the combinatorial nature of the miRNA-target interaction, a change in the expression level of one target may induce a change in that of another target mRNA sharing the same pool of miRNA [34]. Given such crosstalk between competitors, if one of the target mRNA is bimodally distributed, then the expression of other competing target mRNA is influenced. Thus, the authors modelled a system with two ceRNA and one shared miRNA in presence of extrinsic noise. They found that the extrinsic noise increases the range of values of the interaction strength of the two targets for which the cross regulation is possible - through a comparison with the case without extrinsic noise.

Recalling that the presence of bimodality is associated to the presence of two different phenotypes, all the discussed results allow to say that the combination of extrinsic noise together with the threshold behaviour allow to obtain bimodal phenotypes without the need of fine tuning the rates of the reactions, as it happens in a system with only intrinsic noise. Thus, the strong heterogeneity that characterises a population of cells may be seen as an advantage in view of the role of such bimodality.

In light of the findings of [8] discussed here, in the next Chapter 6, we intend to experimentally investigate the role of the extrinsic noise in the rising of bimodality in a miRNA-target system, by exploiting, as source of extrinsic noise, the heterogeneity of a population of cells with respect to the cell cycle.

Chapter 6

Experimental investigation of the model predictions

So far, we discussed the theoretical results obtained in investigating systems consisting of microRNAs that bind to their targets mRNAs. Moving from the simplest system (one miRNA interacting with one target mRNA [9]) towards more complex one (one miRNA shared by two target mRNAs [10]) experimental evidences have proved the predictions of the models. Our aim is now to test the theoretical predictions obtained in [8], namely on the role of the extrinsic noise on a miRNA-target system. In particular, we aim at investigating the hypothesis concerning the rising of bimodal distributions in the free amount of targets.

To this purpose, in this chapter we firstly describe the previous works [9, 10] in order to introduce the experimental set-up we used (Section 6.1) and then, in Section 6.2 we move into the discussion of the experiments we performed. It has to be said that the work is still in progress, therefore we will present here the very preliminary results we obtained so far.

6.1 State of the art

6.1.1 One miRNA and one target mRNA

The remarkable property of the miRNA-target interaction is the rise of a threshold like behaviour. This has been shown in the work of Mukherji and coworkers [9] through synthetic biology experiments. The authors developed the two-colour fluorescent reporter system whose schematic is represented in Figure 6.1(a). The system consists of a plasmid, i.e. a circular filament of DNA, that contains two genes coding for two fluorescent proteins: mCherry (red) and eYFP (yellow). The two genes are controlled by a bidirectional promoter, namely the RNA polymerase may transcribe them with equal probability. On the 3'UTR (the codon stop sequence) of the gene coding for mCherry, a fixed number N ($N = 0$, 1, 4 and 7) of binding sites for a specific miRNA (miR-20a) has been engineered, while the gene coding for eYFP has been left untouched. The number of binding sites represents a proxy for the strength of interaction between the miRNA (miR-20a) and its mRNA target (mCherry mRNA).

Through a transient transfection, these plasmids were then inserted in human embrionic kidney cells 293 (HEK293) that endogenously express miR-20a. With the transfection the plasmids can enter the cells and be transcribed. To be noted that they are not integrated into the genome, thus when cells duplicate the plasmids are diluted since they do not replicate. Given the randomness of the number of plasmids that enter each cell during the transfection, the fluorescent proteins present an heterogeneous expression level among the cells. However, thanks to the feature of the promoter being bidirectional, both mCherry and eYFP genes are always expressed at the same level inside the cell. The authors measured the fluorescent intensity of both fluorophores at the steady state after the transfection in each cell through two kinds of measurements: fluorescence microscopy and flow cytometry. The latter is a technique that exploits microfluidic channels to measure the fluorescence of each single cell, by exciting them with lasers at the required wavelengths. In this way up to $10^5 - 10^6$ cells can be measured in a short amount of time.

The system composed by the two fluorescent reporters inserted inside the cells, is an empirical representation of the miRNA-target system. In presence of non-zero binding sites (1, 4, 7) the mCherry is targeted by the miR-20a and, once this binds, the translation of the corresponding protein (red) is repressed. Thus the red fluorescence can be considered as a proxy for the amount of free molecules of the

target mRNA. The yellow fluorophore (eYFP) instead, being not regulated by the miRNA, has been used as a control to monitor the level of transcriptional activity of the target mCherry, i.e. its expression level in absence of miRNA regulation. Since, as mentioned above, the number of plasmids entering single cells is random, there is an heterogeneity of the transcriptional activity among the cells and this allows to explore a wide range of expression levels.

In this way, the typical profile of free amount of target molecules as a function of their transcription rate can be reproduced by considering the scatter plot between the fluorescences of mCherry vs eYFP (see Figure 6.1(b)). As expected, in the case with 0 binding sites (i.e. absence of miRNA regulation), the scatter plot mCherry *vs* eYFP fluorescence, presents a linear dependence, see black data in Figure 6.1(b)). By increasing the number of biding sites N, a threshold-like behaviour appears, as shown by the blue data in Figure 6.1(c). Through such system other two theoretical predictions have been evidenced: firstly the position of the threshold in correspondence of the equimolarity point, secondly the fact that the steepness of the threshold could be controlled through the interaction strength between miRNA and target, namely through the number of binding sites for the miRNA.

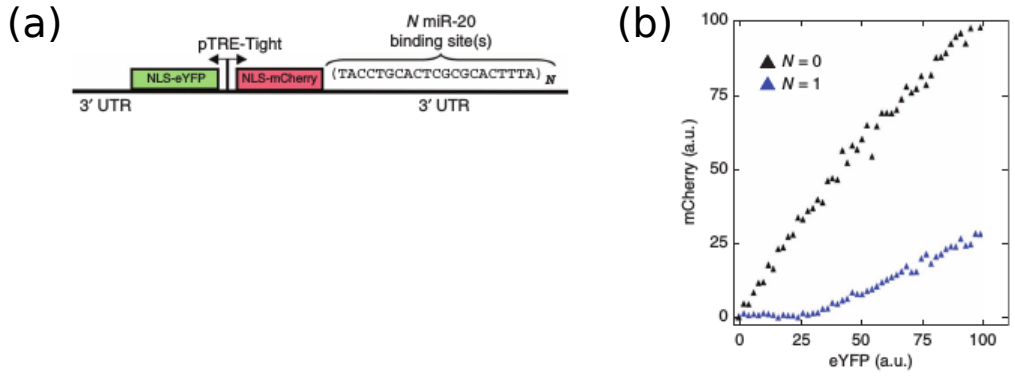

Fig. 6.1 (a) Two colour fluorescent reporter system used in [9]. Bidirectional plasmid engineered to code for eYFP and mCherry genes. The 3'UTR region of the latter gene contains a number N of binding sites for the miR-20a. Adapted from [9]. (c) Relation between fluorescence intensity of the mCherry gene (proxy for the amount of free molecules of target mRNA) as a function of that of eYFP (proxy for the transcription rate of the target). The two series of data represent the case of 0 (black data) and 1 (blue data) binding sites for the miR-20a. Adapted from [9].

Such experimental approach is efficient to experimentally quantify the predictions of the mathematical models giving evidences to the molecular sequestration

mechanism and the threshold-like behaviour of the miRNA-target interaction. Given the efficiency of the method, this has been improved by Bosia and coworkers, in order to experimentally investigate the theoretical predictions of the ceRNA effect [10]. Next section is devoted to such topic.

6.1.2 One miRNA and two ceRNAs

An experimental setup (Figure 6.2(a)) similar to the one described in the previous section has been adopted by Bosia and coworkers [10] in order to experimentally test the predictions obtained in [34] on the ceRNA effect. The synthetic system exploited in [10] is composed of two plasmids: one was the very same previously described and used in [9], the other one instead had the same structure as the first, but coding for two different fluorescent proteins, i.e. mCerulean and mKOrange. Similarly to the case of eYFP-mCherry, the 3'UTR sequence of mCerulean was engineered in order to contain 0, 1, 4, 7 binding sites for miRNA miR-20a, while the gene of mKOrange was left unchanged in order to monitor the transcriptional activity. The theoretical model the author tested is the one sketched in Figure 6.2(a).

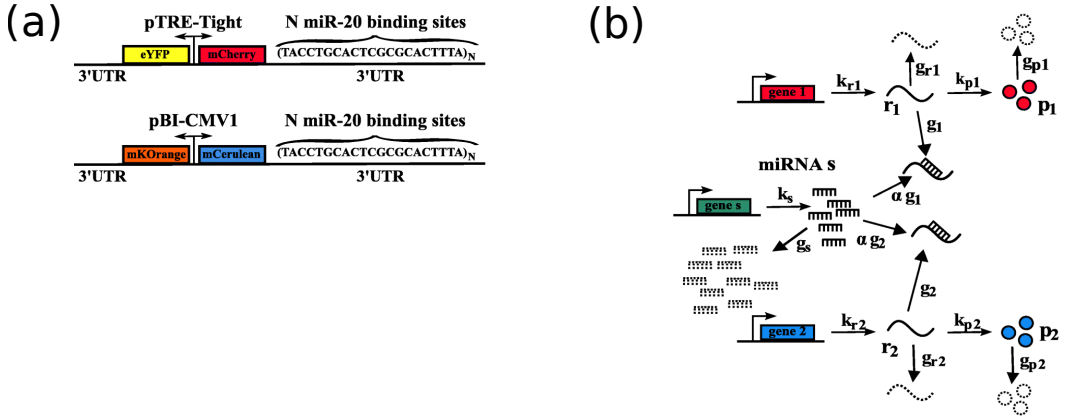

Fig. 6.2 (a) Fluorescent reporter system used in [10]. The upper construct is the same birectional plasmide used by [9] and described in Figure 6.1(a). The lower one is the bidirectional plasmid coding for mCerulean and mKOrange. The 3'UTR sequence of the former gene is engineered to contain a fixed number of N binding sites for the miR-20a. Adapted from [10]. (b) Sketch of the minimal model of miRNA-ceRNA interactions studied in [10]. Adapted from [10].

Such system represents then the case of two targets (synthetic, mCherry and mCerulean) that can bind to the same miRNA (miR-20a). In this way the ceRNA

effect was investigated by performing experiments similar to those discussed in Section 6.1.1. The plasmids were transfected inside HEK293 cells with different combinations of miRNA binding sites. The fluorescence intensity of each cell was then measured through flow cytometry.

In this way the authors verified the theoretical predictions discussed in [34]. Indeed among other results, the predictions on the cross regulation among ceRNAs have been confirmed. The experiments showed that the level of expression of one target and the steepness of the threshold response could be modulated through a variation on the strength of the interaction between the miRNA and the other target, i.e. by increasing the number of binding sites on the other target with respect to the one investigated. Moreover, the correlations between the targets revealed being maximal around the threshold and here, an optimal range of values of miRNA-target interaction strength has been identified. Such correlation raised from the coupling between the fluctuations of one target and those of the other one through the extrinsic noise. Such phenomenon is called retroactivity and discussed for instance in [128]. A last worth to mention result concerns the case of strong repression of one target. In such case, near the threshold bimodal distributions of fluorescent intensity appeared. This finding supported the hypothesis that the gene regulation mediated by microRNAs may increase cell-to-cell variability and the fact that such variability is organised in two distinct phenotypes allows to relate it to the presence of extrinsic noise.

6.2 The role of extrinsic noise in miRNA-target interaction

In Chapter 5 we discussed the role played by the extrinsic noise in the interaction between microRNAs and their targets from a theoretical point of view. In particular we highlighted how a source of extrinsic noise on the miRNA may induce the rising of bimodal distributions on its target. More specifically, the extrinsic noise can compensate for a low miRNA-target interaction strength in order to obtain differentially expressed phenotypes across a population of cells.

In this chapter, we focus on such predictions and aim at testing them on an *in vitro* system, by continuing the work presented in [11].

By adopting the same setup described in Section 6.1, we investigate the expression of the synthetic target of a miRNA (miR-20a), being the latter endogenously expressed in the cell line examined (HEK293 cell line). The transcription rate of such endogenous miRNA depends on the cell cycle phase [121].

During its life, a cell undergoes different phases called G0/G1, S, G2 and M that lead the cell to its division. In the first G1 phase, the cell grows and copies the organelles. Cells in G1 may enter a resting state, the G0 phase, where they do not actively prepare themselves to divide but carry out specific tasks. After phase G1, the cell enters the S phase in which a complete copy of the DNA is synthesised in the nucleus. When such DNA synthesis is finished, the G2 phase occurs, where the cell grows more and makes organelles, proteins and organises its contents for the last phase, the M phase, the mitosis. In such phase the cell divides [129]. So, during such cycle and before dividing the cell increases its size, the number of organelles (like ribosomes), duplicates the chromosomes and modulates the expression of its genes. As a result, the amount of DNA present in the cell varies according to the cell cycle phases.

By considering the fact that the expression of the miRNA we are interested is related to the cell cycle phases, within a population of cells heterogeneous from the point of view of the cell cycle, the levels of transcription of such miRNA are highly variable. Driven by this, we assume such cell cycle heterogeneity as a source of extrinsic noise on miRNA levels.

Our aim is then to modify the intensity of such extrinsic noise from an experimental point of view in order to check the predictions of the model described in Chapter 5. The easiest way to tune the noise has revealed being not an increase rather a decrease of such cell-to-cell heterogeneity. We monitored the different cell cycle phases through a cell-cycle marker, i.e. a fluorescent dye that binds to DNA. The fluorescence level of such dye is a proxy for the amount of DNA present in the cell. Since the amount of DNA is related to the cell cycle phases, by measuring fluorescence emission of the dye, the cycle phase of the cells can be determined at a single cell level. Assuming a relation between extrinsic noise and heterogeneity of the cell cycle, sorting the cells with respect to their specific cycle phase should give a change in the intensity of the extrinsic noise.

We devote the next sections to the description of the experimental setup adopted (Section 6.2.1) and the discussion of the results obtained so far (Section 6.2.2).

6.2.1 Experimental setup

To perform the experiments [11] we used the same bidirectional plasmid developed by Mukherji and coworkers in [9]. To summarise here what described in Section 6.1.1, it is a bidirectional plasmid, namely a circular DNA that hosts two genes that can be transcribed with the same probability thanks to the presence of a bidirectional promoter. The genes code for two different fluorescent proteins: mCherry (red) and enhanced yellow fluorescent protein (eYFP). The stop codon (3'UTR region) of the mCherry gene has been engineered in order to host a fixed number N (with $N = 0, 1, 4$ and 7) of miR-20a binding sites. The gene coding for the eYFP has been instead left unchanged: it represents a control to monitor the constitutive expression of the target, i.e. its expression when no regulation occurs.

Such plasmid is then transfected in epithelial cells of human kidney embryo (HEK293), a cell line that endogenously expresses miRNA miR-20a. Cells were transfected with Effectene Transfection Reagent (QUIAGEN). 48 hour after the transfection, the steady state measurements of the expression of the target across the population of cells have been performed with the flow cytometer. In this way, for each cell it has been measured the fluorescent level of both the mCherry and the eYFP fluorophores. As described in Section 6.1.1, the fluorescence intensity of the mCherry represents a proxy for the expression of the target -i.e. a proxy for the amount of free target molecules -, while that of eYFP is a proxy for the constitutive expression. The latter indicates the transcriptional activity of the system and is proportional to the number of plasmids that enter the cell during transfection. Since we do not have any information on the number of plasmids that enter each cell, we are able to explore a wide range of values of transcriptional activity. This, in turn, can be considered a proxy for the transcription rate of the target, i.e. the parameter k_R of the model discussed in Chapter 5.

Since the amount of DNA within a cell depends on the cell cycle phases, the monitoring of such phases has been performed through a DNA staining. Immediately before performing the measurements at the flow cytometer, a DNA marker (Hoechst 33342) has been injected in the cell culture. Hoechst33342 is a blue fluorescent dye that stains the DNA in living cells. Thanks to the fact that the emission spectrum of such dye does not interfere with those of mCherry and eYFP, it was possible to measure the intensity of fluorescence of all the three fluorophores in each cell. The

level of blue fluorescence has been used as a proxy for the DNA content and thus of the phase of the cycle the specific cell was.

Through flow cytometer measurements, for each transfected cell we obtained information on the fluorescence intensities of mCherry, eYFP and Hoechst3342, together with two morphological information. These are known to be related to the size and the granularity of the cells and have been used to distinguish alive from dead cells. The very same kind of measurements have been performed also on non-transfected cells in order to have an estimation of the background fluorescence intensities relative to each detecting channel.

The raw data obtained through the flow cytometer have then been analysed through Matlab software. We dedicate the next Section 6.2.2 to the preliminary results we obtained.

6.2.2 Preliminary results

In this section we discuss the experimental results obtained so far concerning the investigation of the role of the extrinsic noise on the interaction between miRNA and its target mRNA.

The flow cytometer measurements have been processed through Matlab software. First of all, the background intensity has been subtracted to all measurements. Secondly, we set constraints on the granularity and the size of the cells in order to select only the alive and healthy cells. Finally, according to the specific purpose of the analysis, we sorted the fluorescent values with respect the eYFP intensity (i.e. the target transcription rate) and the Hoechst33342 fluorescence intensity (i.e. the cell cycle phase proxy for the level of noise).

Cell cycle phases

As a first result, we show that the DNA staining allowed us to determine the cell cycle phases. Figure 6.3(a) [11] depicts the distribution of the fluorescence intensity of the DNA marker (Hoechst33342) across the entire population of cells. It is worth noticing that two peaks emerge from the distribution: one is centred at a value roughly twice the other one. Such two peaks correspond to the phases G0/G1 (left peak) and G2 (right peak) of the cell cycle, while the region in the middle is relative

to the S phase. The two peaks present a different abundance of cells due to the different time duration of the phases. As expected, cells in G2 phase (preparing for mitosis) present a double amount of DNA with respect to those in G0/G1 phase. Thus, sorting the cells according to the fluorescence of the DNA marker, allows to divide the cells in sub-populations belonging to specific phases of the cell cycle. Since the expression of the endogenous miR-20a is known to be related to the cell cycle, we exploited the sorting of cells with respect to the cell cycle to reduce the cell to cell variability and thus to decrease the extrinsic noise of the system.

A last observation concerning differences of cells within the cell cycle is depicted in Figure 6.3(b). Here, cells have been sorted with respect to the parameters of the flow cytometer measurements related to size and granularity, as shown by the red ("small" cells) and green ("big" cells) rectangles in the inset. By plotting (in the main plot) thus the distribution of the DNA marker, it is evident that small cells (in red) contain half of the DNA amount than the big ones (in green). Thus, combining this, with the information of Figure 6.3(a), cells in G0/G1 phase are smaller than those in G2. According to this, the discrimination of the cell phases (G0/G1 and G2 at least) can be roughly done even by sorting the cells according to their size.

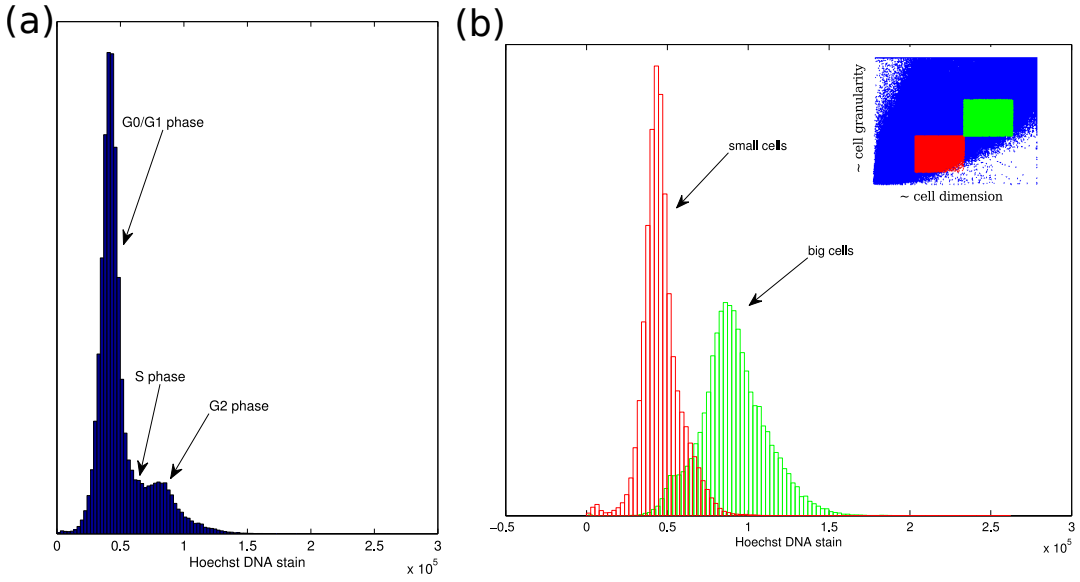

Fig. 6.3 (a) DNA marker (Hoechst DNA stain) distribution. The cell cycle phases (G0/G1, S and G2) are emphasised by arrows. (b) Distributions of the Hoechst DNA stain for big (in green) and small (in red) cells. The inset shows the determnation of the two groups of cells by constraining on cell granularity and dimension. Both (a) and (b) are adapted from [11].

Zero miRNA binding sites. The cell cycle does not influence the transcriptional activity

The transcriptional activity can be investigated in the case of zero binding sites for the miRNAs. Figure 6.4 shows that the transcriptional activity is not influenced by the cell cycle phases. In the scatter plots each dot represents a single cell and for each cell the fluorescent level of the mCherry is expressed as a function of that of eYFP. mCherry fluorescent level is a proxy for the free amount of target molecules while the eYFP fluorescence gives an information on the transcriptional activity of the construct. The three plots represent sub-populations of cells selected according to their cycle phases (i.e. G0/G1, S and G2).
Since the considered case is that with 0 binding sites for the miRNA, this does not regulate the mCherry thus, as expected, in each plot a linear relation between the two variables is shown. Moreover, such linear dependence does not change with respect to the cell cycle phase. In light of this, it is possible to claim that the expression of the construct is independent on the phase of the cell cycle if the miRNA does not regulates its target. A direct consequence is then that in the cases of non-null miRNA regulation, the eventual differences shown among the cell cycle phases may be due to the variability of the miRNA expression across the cell cycle.

Such conclusion supports the idea that the heterogeneity of the cells with respect to the cell cycle can be considered a source of extrinsic noise affecting the miRNA-target system and in particular the transcription levels of the miRNA. By constraining then on specific cell cycle phases such noise can be modulated. Starting from this observation, in the next paragraph we discuss the results obtained so far in this direction.

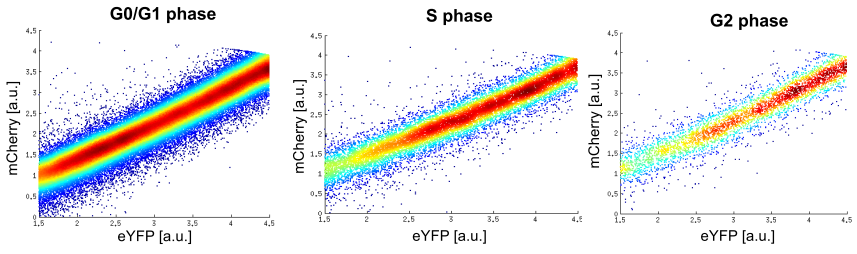

Fig. 6.4 Zero miRNA binding sites. Scatter plots (each dot is a cell) of the fluorescence intensity of the mCherry as a function of that of eYFP. The cells are sorted according to the phase of the cycle: from left to right, G0/G1, S and G2 phase. Adapted from [11].

The extrinsic noise plays a role in the rising of bimodal distributions

Let's consider now the cases of non-zero binding sites for the miRNAs. Figure 6.5 shows the scatter plots analogous to those depicted in Figure 6.4, corresponding to the transfections with constructs with 1, 4 and 7 miR-20a binding sites. Each raw corresponds to a different number of binding site and each column to a cell cycle phase.

A first observation concerns the cell cycle dependence of the miRNA expression. By fixing the eYFP level (i.e. the transcriptional activity), the level of mCherry changes both through the number of miRNA binding sites (i.e. the repression strength) and the cell cycle phase. Since we showed in the previous paragraph that the translational activity of the construct does not change with the cell cycle phase, the variability in the expression of the mCherry that we observe here can be attributed to the variability of the expression of the miR-20a within the cell cycle.

The plots of the first column in Figure 6.5 are relative to the G0/G1 phase. From these it is possible to notice the presence of bimodal distributions in the expression of the mCherry for a wide range of eYFP. While for low and high values of eYFP the data seem to have a monomodal population with respect to the expression of mCherry, there is an intermediate region of eYFP in which the two populations can be seen at the same time. Here is the bimodality. Such effect is more evident by enhancing the miRNA repression, i.e. by increasing the number of miRNA binding sites. In such cases, the repressed and expressed states can be recognised. The mean value of the former is characterised by a threshold-like behaviour as a function of the transcriptional activity. The latter, instead (i.e. the expressed state) shows a linear relation with respect to the translational activity. The role of the cell cycle is that of reshaping such expression distributions. By fixing the number of binding sites and observing plots relative to S and G2 phases, a strong difference arises with respect to the case of G0/G1. By sorting the cells according to the cell cycle one of the peaks of the population distribution can be selected.

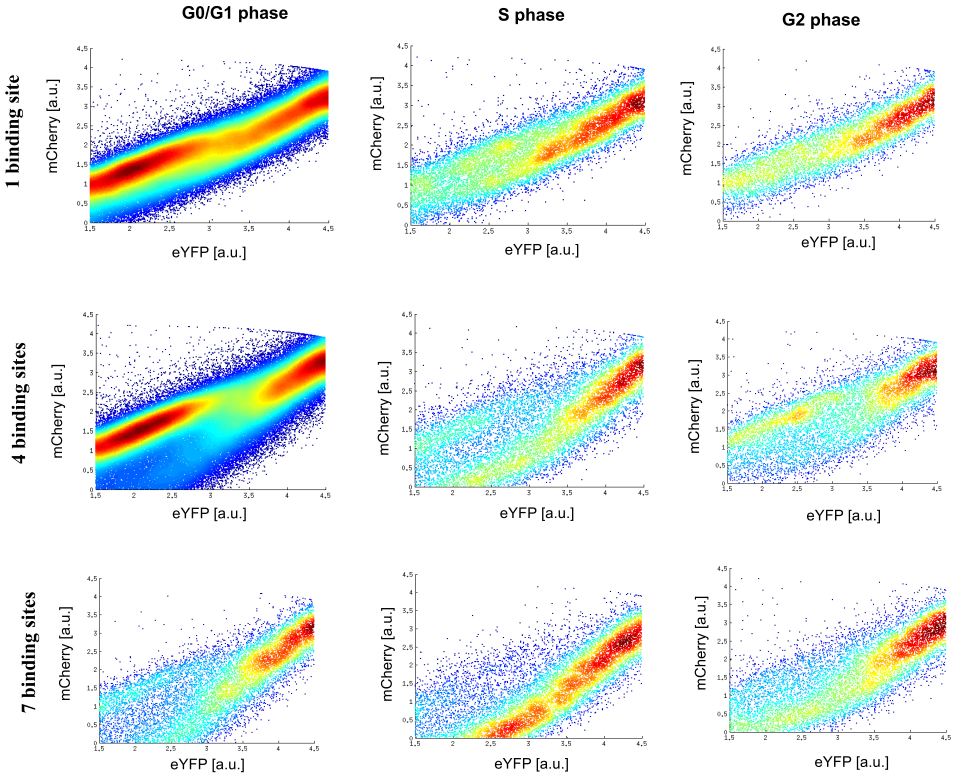

Fig. 6.5 Scatter plots of the fluorescence intensity of mCherry as a function of that of eYFP. Each dot represents a cell. Each raw corresponds to a different number of binding sites, from up to down 1, 4 and 7 respectively. Each column corresponds to a sorting according to the cell cycle phase, from left to right: G0/G1, S and G2. Adapted from [11].

If we consider the extrinsic noise as related to the heterogeneity of the cell population with respect to the cell cycle, the discussed results show that selecting the cell cycle phases corresponds to decrease the level of noise. Thanks to this, we can recover the theoretical phase diagram of [11] (Figure 5.1(a) and 6.6(a)) with the experimental data.

To remind what discussed in Section 5.2, the theoretical phase diagram shows the regions of bimodality in the distribution of the free molecules of mRNA target of a miRNA. In presence of extrinsic noise the bimodality region (given by the colour map) is wider than the case without extrinsic noise and it is present also for low values of interaction strength between miRNA and targets - differently than the case of absence of extrinsic noise.

Our experiments allow to match the parameters of Figure 6.6(a) with empirical ones. Indeed, (i) the extrinsic noise levels are given by the different cell cycle phases; (ii)

the mRNA transcription rate (k_R) corresponds to the eYFP fluorescence intensity and (iii) the miRNA-target interaction strength g is the number of binding sites present on the 3'UTR region of the mCherry gene. Changing the cell cycle phases corresponds to changing the levels of extrinsic noise and choosing different levels of eYFP intensity is equal to change the transcription rate of the mRNA (k_R), as shown by the white arrows in Figure 6.6(a).

In Figure 6.6(b) we present the very first results of such empirical phase diagram. Here the case of 7 binding sites is shown. The cells have been sorted according to their cell cycle phase (level of noise). Since the case G0/G1 is the phase in which cells grow and double their components, we assume it being the case of maximum noise after the unsorted condition in which all the cell cycle phases are considered together. The phase G2 instead is the one in which the cells have doubled their contents and the DNA is already doubled, therefore we expect it being the phase of minimum noise level. The S phase, instead, we assume lying in the middle between the former two. Of course, a proper quantification of such noise levels will be necessary to confirm such assumptions.
Each single cell has its values of intensities of mCherry and eYFP. By choosing 4 bins of eYFP fluorescent intensities (i.e. changing the transcription rate of the target, k_R) the distribution of the mCherry expression of the cells belonging to the given bin and the given cell cycle phase has been obtained. As emphasised by the dashed blue line in Figure 6.6(b), there is a wide region within such line in which the target expression distribution is bimodal. Such region gets wider as the level of noise increases. A systematic quantification of such bimodality is still lacking. We intend to compute it similarly to what have been performed in [8], where the number of maxima of every distribution has been taken into account. By repeating the same kind of analysis with the cases of 4 and 1 binding sites, we expect such region to shrink and move towards higher values of noise, since in this way, we are decreasing the interaction strength between miRNA and its target.

Even though a more systematic and quantitative analysis is still lacking, the presented preliminary results are promising in giving experimental evidences of the stochastic model broadly discussed in [8]. Moreover, they suggest the possibility of controlling such system not only from a theoretical but also from an experimental point of view.

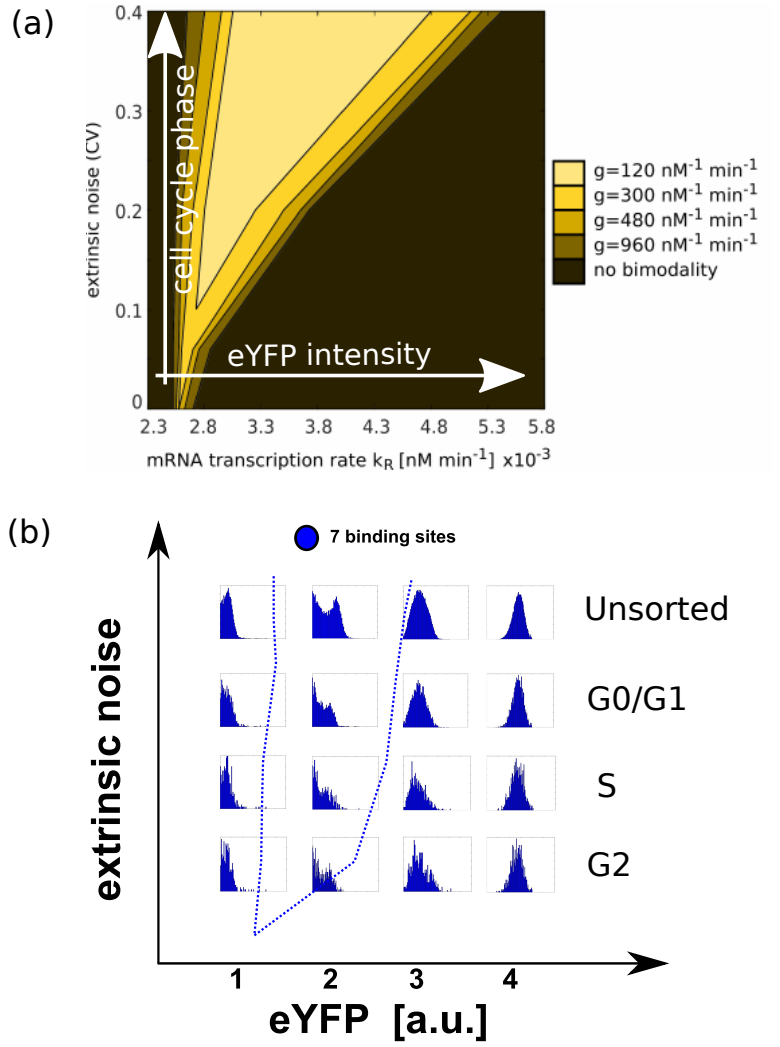

Fig. 6.6 (a) Phase diagram for the bimodality in the distribution of the free molecule of mRNA target of a miRNA. The x-axis represents the transcription rate of the target mRNA (k_R); the y-axis shows levels of the extrinsic noise; and the colours are relative to different values of interaction strengths (parameter g). The white arrows parallel to the x- and y- axis represent the direction of variation of the corresponding experimental parameters, i.e. the cell cycle phase as a proxy of the extrinsic noise and the eYFP intensity for k_R. Adapted from [8]. (b) Similar phase diagram of plot of (a) obtained with the experimental data corresponding to the case of strong interaction strength (7 binding sites for the miRNA). The extrinsic noise is shown as a function of the level of eYFP fluorescence intensity. The distributions are relative to the mCherry expression level for cells sorted according to the cell cycle phase and eYFP levels. From high to low noise (i.e. from top to bottom), the cell cycle phases are: unsorted, G0/G1, S and G2.

6.2.3 Future perspectives

The preliminary experimental results presented in the previous section revealed being in agreement with the theoretical predictions discussed in [8] concerning the role of the extrinsic noise in inducing bimodal gene expression. However, as already mentioned, in order to drive proper conclusions on the compatibility of the empirical and theoretical results, a systematic quantification of the investigated parameters is needed together with the repetition of the very same analysis on the data corresponding to different miRNA-target interaction strengths. In this way it would be possible to experimentally reproduce what have been broadly studied only theoretically up to now, namely the landscape of conditions that allow, through the extrinsic noise, to drive a miRNA-target system to show a bimodal expression of the target.

A further important issue that arises from the discussed results, concerns the distribution of miRNA miR-20a. We showed that the expression of miR-20a changes according to the cell cycle phases and the sorting of the cells with respect to them allows to to control the range of bimodality in the expression of the target. In order to be sure that, as supposed, this is due to a reduction of the noise in the expression of the miR-20a, it should be worth checking whether there is or not a correlation between the distribution of the miRNA and the one of its target. In a positive case, the selection over the cell cycle phase would mean selecting specific expression levels of the miRNA. In such case, the control of the bimodal expression could be due to changes in the mean expression levels of the miRNA rather than to a modulation of the extrinsic noise.

In order to experimentally quantify the miRNA at single cell level, some recently developed assays are available [130, 131]. The quantification is not trivial and such assays are based on a technique called Fluorescence in situ Hybridization (FISH) that allows to label specific sequencies of RNA or DNA with fluorescent probes. By combining FISH with a further technique that amplifies the signal of the DNA, this can then be detected through flow cytometry measurements. In order to exploit such technique for our purposes and reconstruct miR-20a distributions, a check on its compatibility with the already used fluorophores and DNA marker has to be made.

Appendix A

The role of sodium acetate on cell growth

In this appendix we describe the project on monoclonal antibodies carried out during a visiting period at the Centro de Inmunología Molecular (CIM) of Havana in collaboration with Prof. R. Mulet, Dr. J. F. de Cossio Diaz, Prof. A. Lage Castellanos, L. Calzadilla Rosado and B. A. Peréz Fernández.

The aim of the work is the investigation of the impact of sodium acetate on the growth and protein production of a specific human cell line. Since the work is still in progress at the present time, only the preliminary results are discussed here.

The first Section A.1 is dedicated to a brief introduction to the problem and it is followed by the description of methodologies adopted for reaching the goal (Section A.2 and A.3). In Section A.4, we present the first results. Afterwards, in Section A.5 we discuss the criticalities of the adopted methodologies, suggest improvements and test them on new experiments before concluding in Section A.6.

A.1 Bioproduction and aim of the work

Antibodies are proteins produced by plasma cells and used by the immune system to target a specific foreign object (antigen). When they are derived from a single parent cell, they are called monoclonal antibodies (mAbs). mAbs, together with recombinant proteins, are two of the most known examples of the so called protein

therapeutics. This is a large group of bioproducts widely used in research and, as the name suggests, for therapeutical purposes [132]. Its fields of medical applications range from immunology, to organ transplant and to oncology [133]. To give an example of the last field, the use of monoclonal antibodies is widely developed for cancer therapy [134], in particular cancer immunotherapy, where they are used as immune checkpoint inhibitors [135]. Immune checkpoints are pathways critical for the ability of the immune system to control cancer cell growth. Indeed, such pathways are used by many cancers in order to grow by avoiding the immune system. Through the use of mAbs as immune checkpoint inhibitors, the immune system is able to block these pathways and thus stop or slow the cancer growth.

mAbs and recombinant proteins are produced by different systems, from non-mammalian, such as bacteria or plants, to mammalian expression systems, like for example human cell lines. Those produced by mammalian cell lines (both human and animal), have the great advantage to be directly secreted by cells, while those produced by bacteria, for example, must be extracted through cell lysis. Among all mammalian cell lines the most used for producing monoclonal antibodies are Chineese Hamster Ovary (CHO) and mouse myeloma (NS0) [39]. Human cell lines have even more advantages for the applications due to the fact that their produced proteins are similar to those naturally synthesised in humans [132]. Among the products approved as recombinant biotherapeutic products, there are those produced by human fibrosarcoma HT-1080 and embryonic kidney 293 (HEK293) cells [132]. In the work we are presenting in this appendix we focus precisely on this last line and its products. HEK293 cell line in fact is known to rapidly grow and to have an high efficiency in protein production [132]. Normally HEK293 cells live in adhesion, but cell lines modified in order to make them living in suspension exist. In particular, we are considering a HEK293 suspension cell line able to produce extracellular domain protein of epidermal growth factor receptor (Her1-ECD) [136].
The human epidermal grow factor receptor (EGFR or Her1) is a protein consisting of an extracellular ligand binding domain (ECD) that binds to epidermal grow factor (EGF) and plays a role in the regulation of neoplastic processes and development. Indeed, when binding to its specific ligand (EGF) it may lead for instance to a modulation of cell proliferation or to the tissue differentiation in normal tissues or tumours. Indeed, its deregulation and over expression is known to be related to many epithelial tumours like brest, ovarian, prostate, lung and colon tumours [137]. For these reasons cancer therapies that target EGFR have been developed, such as

therapeutic antibodies, immunotoxins and vaccines [138]. On the last topic it is worth mentioning that a clinical trial phase I of a vaccine based on Her1-ECD (released from the Quality Control Department from the Center of Molecular Immunology in Havana) has been recently evaluated on patients affected by prostate cancer [139]. The active ingredient on which such formulation is based is the Her1-ECD protein obtained from the supernatant of HEK293 cell line [139–141].

Since the importance of such therapeutical proteins, researchers have been interested in investigating ways for improving their productivity. Several methods have been developed for increasing the bioproductivity of cells, both for mammalian cells and for bacteria. Concerning bioproductivity in bacteria, we remind for an example the commentary [38] we worked on in collaboration with Dr. F. Ceroni (Imperial College, London). There we describe the work of Rugbjerg and colleagues [142], who developed a bioengineering strategy able to limit the enrichment of non-producing cell populations in cell cultures employed for bioproduction. The strategy is called "product addition" and it is based on placing the genes responsible for the growth under the control of a promoter that is responsive to the bioproduct. In this way the productivity is depending on the product itself. The author tested it in *Escherichia coli* engineered to produce mevalonic acid in long terms cultivations.

Concerning mammalian cells, it has been found that adding chemicals to the growth medium is an efficient way for stimulating the cells to express proteins [41]. Glycerol, dimethyl sulfoxide (DMSO) and sodium butyrate [39–41] are some examples of such chemicals that are able to enhance protein productions in some cell lines. Another interesting example is given by the sodium acetate ($C_2H_3O_2Na$), the sodium salt of the acetic acid that naturally occurs in animal tissues and plants. There are evidences showing that cell viability of human gastric adenocarcinoma (AGS) epithelial cells, for example, is stimulated by sodium acetate in a dose-dependent manner as well as their ability to produce some proteins [143].

The aim of our work fits here. We are interested in studying how the growth and productivity of the above cited HEK293 cell line changes when exposed to different concentrations of sodium acetate. To address such issue in a quantitative way, we took advantage of the cell growth protocol discussed in Chapter 2 in order to perform a systematical monitoring of cell growth. Then, the quantification of the produced proteins at different growth phases allows to relate the bioproductivity of

the examined cell line to the concentration of sodium acetate supplied to the growth medium.

In this appendix we focus on the first part of the project, namely the optimisation of the growth protocol. The details of the adopted techniques are explained in the next sections.

A.2 Summary of the experimental methodology

To the purposes of this project, we adopted the same experimental and data analysis methodologies described in Chapter 2. To briefly summarise it, we set up batch cell growth experiments (see Section A.3 for details) by seeding cells in the growth medium supplied with different concentrations of sodium acetate. The growth was followed day by day by taking micrographs of representative samples of the population and then by counting cells through the use of an image segmentation algorithm (Section A.3.1). A supernatant aliquote was saved each day in order to later quantify the protein production corresponding to the desired growth phase. This last step has been carried out by the researchers at CIM and will not be described here. As preliminary results in Section A.4, we present the growth curves obtained by applying the segmentation algorithm and then their analysis performed by exploiting the same analytical tools described in Section 2.3.3. In the last Section A.5 we discuss the criticalities of such adopted methodology and suggest a new one based on an improved image segmentation algorithm.

A.3 Experiments

The cell line under exam is HEK293, human embrionic kidney cells living in suspension. Like Jurkat cells (see Section 2.3.2), also these tend to form clusters when growing. A macroscopic difference between the two lines is that while the disruption of clusters for Jurkat cells is not critical, it is so for HEK293. Here indeed the disruption of clusters through pipetting is less efficient and may cause a stronger stress to cells which tend to change their shape. To give an example, Figure A.1 (a) and (c) show representative micrographs of samples in which there are many cells with jagged boundaries as well as many big clusters. This is more likely to happen

when the cell concentration is high (Figure A.1(c)). These issues has been taken into account in the image segmentation algorithm in order to properly estimate the cell number (see Section A.3.1).

The experiments were performed with two different initial concentrations: $N_0 = 3 \cdot 10^5$ cell/ml and $N_0 = 3 \cdot 10^6$ cell/ml in order to investigate also the inoculum size influence on the growth.

We set up batch experiments by seeding the fixed initial concentration of cells in flasks containing 10 ml of growth medium. Since we were interested in studying the effect of sodium acetate, we prepared 5 flasks with the growth medium (DMEM) considered as standard for culturing these cells, supplied with an increasing concentration of sodium acetate, from 0 mM up to 20 mM. Cells were then incubated at 37°C with 5% CO_2 in an incubator that was constantly shaking the flasks gently.

Analogously to the procedure adopted in Chapter 2, cell growth was followed by counting cells day by day through the use of a custom-made image segmentation algorithm able to analyse micrographs of samples representative of the population.

Every day, two different aliquotes of the population were taken from each batch and put in tubes of 1.5 ml: one for later quantification of medium components and the other one for counting purposes.

The first aliquote was centrifuged in order to allow cells to form a pellet at the bottom of the tube. Afterwards the surnatant was transferred into another clean tube and freezed. These tubes will be used for the quantification of the bioproducts present in the medium (in progress, not shown).

The second aliquote, instead, was necessary for cell counting. Each tube was gently vortexed or pipetted in order to destroy clusters as much as possible. Then $3x10\,\mu l$ of the mixture composed by cells+medium was injected in mono-use Burker's chambers (see paragraph 2.3.2) and bright field images were taken with an inverted microscope (magnification $10x$).

In order to estimate the cell concentration, we adapted the image segmentation algorithm described in Chapter 2 for the analysis of such new images. In the next section we describe the details of such image processing code.

A.3.1 Image segmentation and growth parameters

For counting cells we adapted the image segmentation algorithm described in Section 2.3.2 to the study of HEK293 cells. It is important to highlight that, since the strong presence of cell clusters in the images, in this new case the estimate of cell concentration have been performed with an indirected method, namely by determining a representative area of single cells and therefore by dividing the total area of detected cells/clusters by the representative one. Due to this, the image segmentation algorithm mainly aimed at segmenting objects and calculate their areas. For this purposes, the first part of the code described in Section 2.3.2 was not considered, since the Watershed procedure tend to change too much the object shapes. Therefore, the second part of the mentioned algorithm - the one concerning sizes - was the one adapted to our new purposes.

We remind to Section 2.3.2 for details on functions and image processing methodologies mentioned in the following description.

After a contrast adjustment of the original image in order to better distinguish alive cells (the brightest objects) from the background and residuals (the darkest objects), a Gaussian filter and a threshold binarization is performed. The threshold is chosen by analysing the intensity profile of the contrasted image and setting a value equal to the mean of the intensity profile plus one standard deviation, since the images were generally very bright. Afterwards a *closing* (dilatation of objects followed by an erosion) and hole filling procedures are applied in order to possibly split big objects and make them being compact and easily recognisable. Then a first segmentation through *bwlabel* function is performed and the sizes of the detected objects are determined by using *regionprops* function. At this point a threshold on the minimum accepted size is set in order to eliminate residuals present in the image. At the end, a new labelling has been performed and the final segmented image is produced: one label is associated to each object. Figure A.1(b) and (d) show two examples of the segmented results: each detected object is labeled with a different colour for an easy visualisation.

Since, as said at the beginning, the cluster disruption process caused suffering of the cells, most of them resulted not round-shaped as they instead should be. For this reason, together with the fact that many visible clusters did not have a regular

shape, the threshold on the circularity present in the algorithm in Section 2.3.2 was not applied. In this way the software is able to recognise even irregular objects.

At the end, the algorithm saves the information on the total number of detected objects, the area of each object (A_i) and their sum that is the total segmented area (A_t).

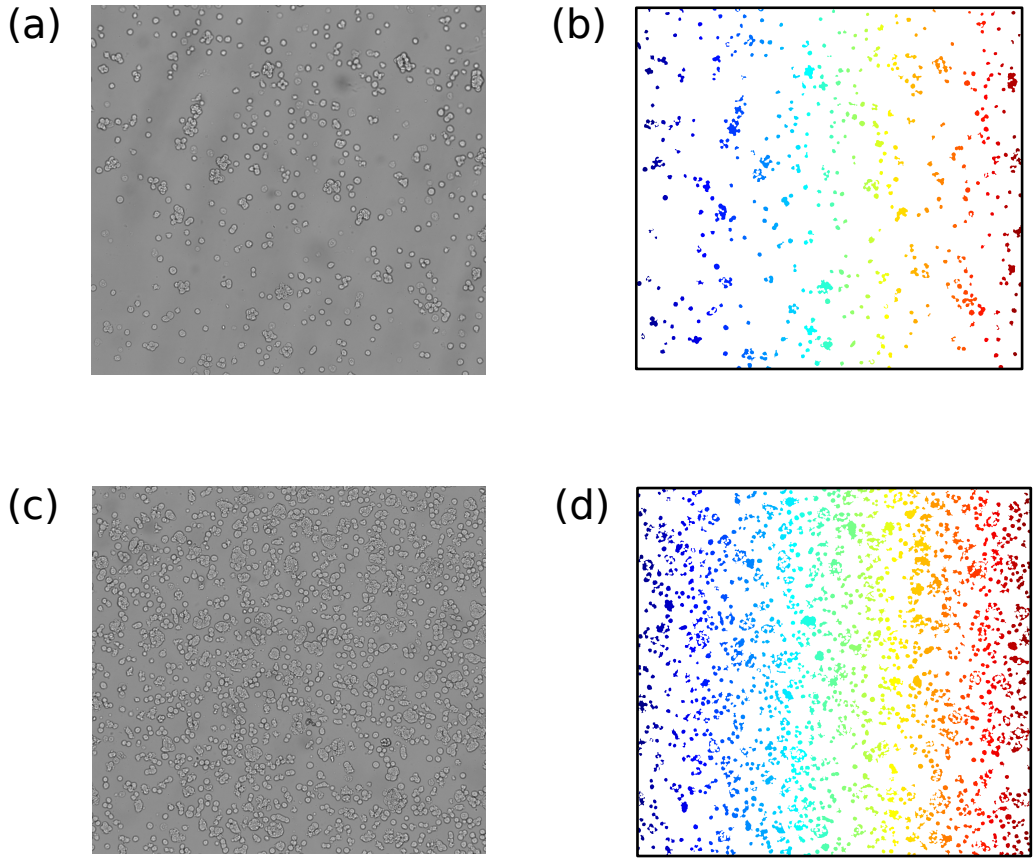

Fig. A.1 Two examples of micrographs (a) and (c) and their respective image segmentation results (b) and (d), where each RGB colour represents a different detected element. (a) and (b) are an example of lower concentration while (c) and (d) set of images refers to a high inoculum experiment.

The concentration of cells was estimated through the analysis of the single areas distribution. We chose the most common value of the distribution (A_{mo}) as the representative area. This choice was driven by the fact that the size distributions were very broad with a defined mode (see Figure A.2 for an example). Therefore, the choice of the mean value (red line) would not have been a good estimate since

it takes into account also the right tail, representative of the clusters. The mode of the distribution (black line) instead, resulted to be located at smaller values than the mode and seemed a more reasonable proxy for the average dimension of single cells.

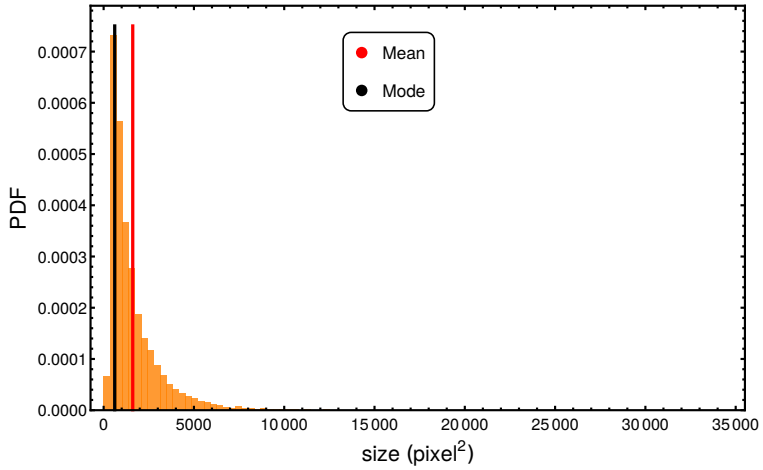

Fig. A.2 Example of size distribution of object detected by the algorithm. The highest value of area is not visible in the plot and corresponds to a size of 34750 px^2. The red line represents the mean of the distribution, the black line the mode. The mode has been chosen as representative of single cell areas.

Thus, by dividing the total area by the mode, we obtain the number of objects with average area A_{mo}. By converting such number into concentration, we finally obtained the growth curves. Figure A.3(a) and (b) in next section show examples of such curves in logarithmic scale for both low and high inoculum experiments at two different sodium acetate concentrations.

In order to quantify the growth parameters, we fit the logarithm of the growth curves, normalized to the initial density ($ln(N/N_0)$) with the sigmoidal shape function described in Section 2.3.3 and here reported for simplicity:

$$ln\frac{N}{N_0} = \frac{A}{1+exp\left[\frac{4\lambda_{max}}{A}(t_{lag}-t)+2\right]}$$

The parameters A, λ_{max} and t_{lag} are the maximum reachable value, the maximum growth rate and the lag time respectively. Thus, after the analytical definition of the three growth phases as described in Section 2.3.3, the growth rate of the exponential phase (λ) can be calculated as the linear fit of data within that phase. This is the

variable that we are interested in relating to the sodium acetate concentration, as described in the following section A.4.

A.4 Toxic activity of sodium acetate

This section is devoted to the first results obtained by following the procedures described in the previous section. The discussion here reported concerns the impact of sodium acetate concentration on the growth of cell populations.
The section is structured as follows: we (i) firstly present and discuss the growth curves in order to comment the performance of the algorithm and the procedure adopted for cell density determination. To this it follows (ii) the analysis on the growth parameters in relation with results of Chapter 2.

A.4.1 Robust growth curves

In this section we discuss the performances of the image segmentation algorithm and the methodology for estimating the concentration, through general considerations on the growth curves.

Figure A.3 (a) and (b) show examples of growth curves relative to low and high inoculum sizes respectively, expressed as the logarithm of the concentration of cells (N) *vs* time. Dots are experimental data, while solid lines represent their trends. The two colours represent two different concentrations of sodium acetate added to the medium, 0 mM in red and 20 mM in blue. In both plots, the experimental data are reported as the average value of the three replicates and the error bars are their dispersion.

First, we observe that, in general, the error bars are within the $10-15\%$. This suggests that (i) there is a good compatibility of the technical replicates among each others; (ii) the image segmentation algorithm gives results consistent among each others independently on the overall cell concentration. Thus, despite all the problems faced with the image analysis, the algorithm seems to show consistent performances when applied to similar conditions.

Secondly, the concentrations of cells estimated with our protocol at time $t = 0$ are higher than the expected initial concentrations ($3 \cdot 10^5$ and $3 \cdot 10^6$ cell/ml). It is worth

noting that some variability among samples is usual during the preparation of the experiments due to practical issues. However, a reason for the higher values that we found can be dig up also in the method we adopted for estimating them. As already discussed, we decided to use as a reference value the mode of the distributions of the size. It is then possible that in some cases, this value results lower than the real single cell sizes, thus giving a higher number of cells.
In order to obtain more reliable results, more refined methods for estimating cell concentrations should be found. Since in our analysis we are not focusing on absolute values of cell densities, but on their trends, we will discuss the results obtained through this method in the rest of the present section. Afterwards, driven by the above observations, in Section A.5 we present a new segmentation algorithm and sampling preparation protocol that we developed in order improve the cell density estimate.

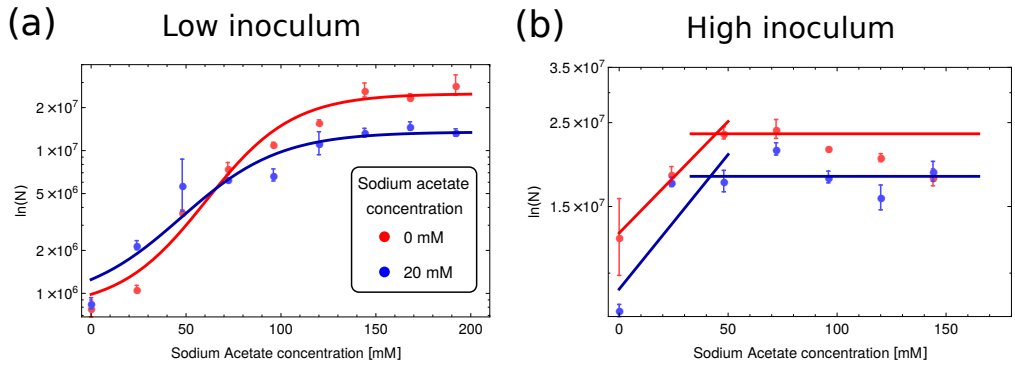

Fig. A.3 Examples of growth curves obtained for two different concentrations of sodium acetate: 0 mM in blue, 20 mM in red. Dots are experimental data expressed as the mean value of 3 replicates. Error bars are given by the replicates semidispersion. (a). Results of low inoculum experiment ($N_0 \sim 3 \cdot 10^5$cell/ml). Solid lines are the sigmoidal shape fit of the experimental data. (b). Growth curves from high inoculum experiment ($N_0 \sim 3 \cdot 10^6$cell/ml). Solid intersected lines represent the trend of the exponential and saturation phases of the experimental data.

A.4.2 The tuning of the parameters of growth

The fitting procedure described in paragraph A.3.1, was applied to each single growth curve.
We refer to Figure A.3 - where the average trend of the replicates is shown - for making some considerations on the way the parameters were estimated in the two

set of experiments.
By focusing on Figure A.3 (a), it is possible to notice how the experimental data could be fit with a sigmoidal shape curve and the three phases of growth - i.e. lag, exponential and saturation phases - can be recognised. Therefore in this case, the growth parameters have been determined for each curve through the procedure described in paragraph A.3.1, namely by fitting the data with a sigmoidal shape curve to define the growth phases and then compute the exponential growth rate.
From Figure A.3 (b), instead, it is evident that the data do not present a lag phase. In order to define the exponential growth rate (λ) and the carrying capacity (k) we performed two linear fits (represented by the solid lines in Figure A.3 (b)). For determining the parameter λ, the first 3 data points were considered; while for k, we computed the average of the data corresponding to a saturation phase.

Despite the methodologies we adopted for inferring growth parameters, the general trend of the cell populations in these two conditions agrees with findings of Chapter 2. There, in fact, we claimed that populations of Jurkat cells seeded at high concentrations showed a smaller lag time duration and tend to quickly saturate. A first observation of HEK293 growth curves under study seems to suggest a similar behaviour.

In order to deep investigate the relation of growth parameters both with the inoculum size and especially with sodium acetate, in the following sections, we focus on the behaviour of the three growth parameters so far considered.

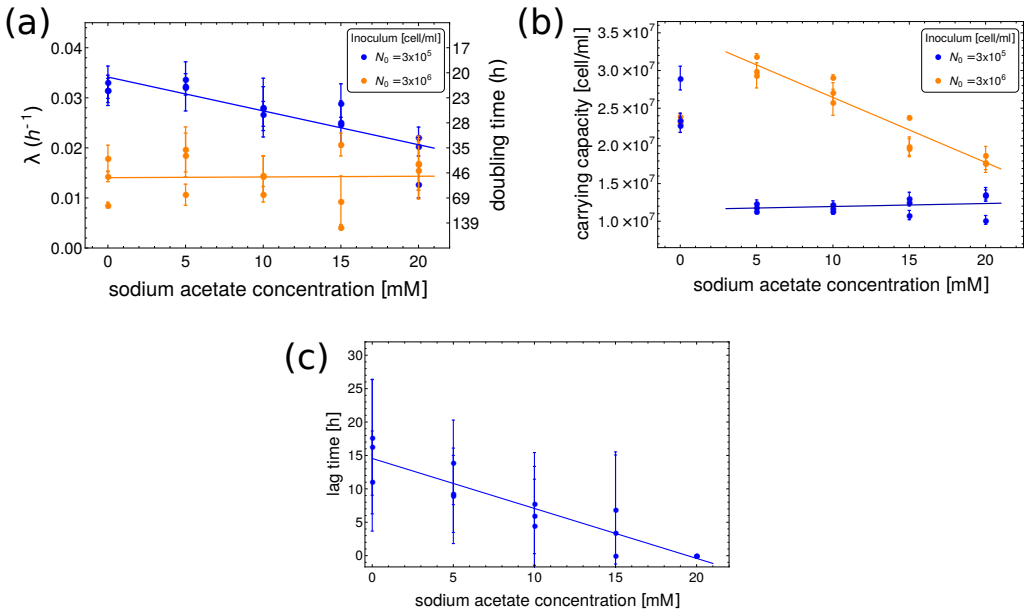

Fig. A.4 Growth parameters analysis. (a) shows the exponential growth rate (λ) as function of sodium acetate concentration. Blue dots are relative to the lower inoculum experiment ($N_0 \sim 3 \cdot 10^5$ cell/ml), orange dots refer to higher inoculum ($N_0 \sim 3 \cdot 10^6$ cell/ml). Each point is a single experiment. Error bars are obtained through the fitting procedures that estimates the parameters. The lines represent the trend of the data. (b) represents the carrying capacity vs sodium acetate concentration. Again here the error bars are given by the fits. In this case, the general trends of the data are relative to sodium acetate concentrations bigger than 0. In (c) the lag time vs sodium acetate concentration only relative to the experiment at low inoculum. Error bars are the errors of the fit to obtain the lag time. The line is the linear fit of the data.

Exponential growth rate λ

Figure A.4 (a), shows the exponential growth rate as function of the concentration of sodium acetate for two different inoculum sizes (low inoculum in blue and high inoculum in orange).

We focus on the difference between the growth rates at 0mM of sodium acetate. Here, again, we find results in agreement with the Jurkat inoculum-dependent growth studied in Chapter 2. Indeed, the growth rate of the population seeded with low inoculum is higher than that of the population with higher N_0, suggesting that such N_0 is close to the carrying capacity.

Focusing now on the growth rates obtained with different concentrations of sodium acetate, we observe a different trend between the two experiments. In case of low inoculum (blue data), the growth rate decreases when increasing the external

metabolite concentration. This suggests a dose-dependent toxic activity of sodium acetate on population growth.
Data for high inoculum show instead a different trend. In this case, by changing concentration of sodium acetate, the growth rate remains constant and takes the same value obtained in absence of the metabolite. It is interesting to notice that such value is closer to that of cells seeded at low concentration and growing in 20mM of sodium acetate. These latter results suggest then that a higher concentration of cells may balance the toxic effect of the metabolite, since even without it, the population does not grow with high rates.

In order to deeper investigate the growth rate dependence on the sodium acetate metabolite, further experiments are necessary. In addition to the repetition of the same experiments here presented for confirming or not these first hypothesis, other possible conditions could be investigated by varying the inoculum, investigating in this way the initial seeding dependence of the system.

Carrying capacity

Another interesting variable representative of the growth is the carrying capacity k, that corresponds to the concentration reached by the population in the saturation phase at time $t = \infty$. For the low inoculum experiments, k is obtained from the parameter A of the fit (see Section 2.3.3) and takes into account the initial concentration N_0 as follows:

$$k = \exp[A] \cdot N_0. \tag{A.1}$$

For the high-inoculum experiments, k has been calculated as the mean value of saturation data. The trends of the carrying capacity as function of the concentration of sodium acetate are shown in Figure A.4 (b).

Analogously to the discussion on the growth rate, we firstly focus on the values of the growth rates in absence of the external metabolite. Interestingly, the curves of the two sets of experiments saturate at similar levels. This again, seems in agreement with the fact that the level of saturation is a property of the growth medium and it does not depend on the history of the population (see Chapter 2 for details on this topic).
Looking at the dependence of k on the concentration of sodium acetate, interesting

results arise. For the low-inoculum data, the carrying capacity does not depend on the concentration of the toxic metabolite. Its value is visibly lower than that in absence of sodium acetate, as it is evident also from the growth curves of Figure A.4.1(a). This result supports the hypothesis of the toxic activity of sodium acetate on HEK293 at low concentrations: it seems to influence the growth rate in a dose-dependent manner while sets the maximum achievable concentration to a fixed value. A completely different behaviour emerges from the high-inoculum experiments. In this case, the saturation level seems to depend negatively on the concentration of sodium acetate and for higher concentrations of the metabolite tends to the value reached by the low-inoculum experiments. Such results suggest then that highly concentrated HEK293 cells seem to grow with a growth rate independent on the concentration of the sodium acetate, but they can saturate at different levels depending on the dose of the metabolite present in the environment. The higher the concentration of the metabolite, the lower the saturation level. Also in this case a toxic effect of the sodium metabolite on the growth of the population is suggested.

However, it has to be said again, that the results obtained by the segmentation algorithm in this case are less reliable than those obtained for low concentration experiments, therefore a thorougher analysis would be necessary.

Lag time

The last variable to be considered for the growth curves is the lag time, namely the time a population needs to adapt to the new environment. As in Section A.4.2, growth curves relative to the experiment at high seeding do not present a visible lag time. For this reason, Figure A.4 (d) shows only data of the low-inoculum experiment.The lacking of lag time (i.e. lag time equal to 0 h) for cell populations seeded at high densities, agrees with the inoculum dependent behaviour of the lag time found in Chapter 2 for Jurkat cells. Even though for the low inoculum case a lag time could be computed, the big error bars shown in the plot suggest a difficulty of the fit to estimate it. Also in such case, thus, it would be necessary to increase the frequency of the measurements in the lag phase, in order to increase the resolution of the data and allow the fitting function to better estimate the value.

Despite of this, Figure A.4 (d) shows a negative trend of the lag time with the increasing of the sodium acetate concentration.
When the sodium acetate is present at high concentrations (such as 20 mM), it

influences the growth making the cells faster to adapt to the environment, but letting them grow with a low growth rate and reaching a level of carrying capacity lower than the case without sodium acetate.

A.5 Criticalities and improvements

The results presented in the previous section seem to be promising towards the direction of a quantification of a toxic activity of the sodium acetate on the growth of population of HEK293 cells living in suspension. However, all the conclusions we drew are strongly dependent on the estimate of the growth curves, from which then the interesting growth parameters are calculated. The bottleneck of such analysis lies in the daily estimate of the size of the growing population, i.e. the concentration of cells. As discussed in Section A.3.1, both the way the cells were prepared for counting (pipetting and vortexing) and the algorithm we adopted to count presented high criticalities and needed to be improved.
For this reason, we suggest here a new protocol for sample preparation and a new counting algorithm able to (i) reduce the number of big clusters in the micrographs, (ii) recognise alive from dead cells and (iii) better estimate the size of single cells. We tested then such protocol by performing a new set of experiments with populations of the same HEK293 cells.

Since the investigated cells have a strong attitude to form clusters and the pipetting method adopted strongly influenced cells vitality, a better way to prepare the samples is to avoid to pipet the cells and only shaking them. This allows the cluster to dissolve better without causing a changing in the shape of the cells. Irregular shapes are more difficult to be automatically detected than regular shapes. Moreover, a dilution (in the growth medium) of the sample to be counted can be performed in order to have less cells in the field of view of the micrographs. Such dilution factor must then be accounted during the conversion to obtain the concentration of cells.
In order to discriminate alive from dead cells, we used the Trypan blue, a blue dye analogous to the methylen blue used in Chapter 2, that enters only the dead cells making them appearing blue and darker than the alive ones. Such distinction is visible in the bright field micrograph shown in Figure A.5(a) where the darkest spots are dead cells. Such image, together with Figure A.5(c) is an example of micrographs taken by following such new sample preparation protocol.

By using as input micrographs like those shown in Figure A.5(a) and (c), we developed a new and more sophisticated image segmentation algorithm able to (i) distinguish single cells and evaluate their size; (ii) recongnize clusters and evaluate their area and (iii) estimate the number of single cells contained in such clusters by using as representative size of the single cells the mode of the distribution of such sizes. In the following we briefly explain how the algorithm works.

A.5.1 New image segmentation algorithm to recognise single cells and clusters

We present here the main steps of the new image segmentation algorithm. (We remind to section 2.3.2 for the definition of the Matlab built-in functions mentioned below).
First of all the new algorithm requires as input a micrograph of a representative sample of the population of cells, previously added with the Trypan blue used to stain the dead cells. In the new set of experiments we performed, (as for the old set) we repeated the counting 15 times for each sample, in other words we took 15 images of the same sample. To do it, we prepared 3 counting chambers with cell-medium mixed with an 1 : 2 dilution of Trypan blue.

Here are the details of the code that works on one micrograph at a time.
First of all, the contrast of the image is adjusted and the mode of the distribution of the pixel intensities is considered. Since the cells in the sample are diluted, the major part of the field of the image will be covered by the background. The mode of the distribution of the pixels gives then a proxy information on the intensity of the background. This will be exploited later on to in order to recognise dead cells, that are darker than the background.
Afterwards, by exploiting the Matlab function *edge*, the edges of all the objects present in the image are recognised. After other two steps to close the open segments and fill the closed areas, all the single objects are then recognised and labeled (*bwlabel*). A threshold on the minimum size of the recognised objects is then applied in order to discard all those elements too small to be considered cells.
Here the dead cells recognition occurs. Indeed, the algorithm computes the mode of the distribution of the original pixels of each recognised object and discards all those with values of the mode lower than the background. Such discarded object could

be either dead cells (if round shaped) or residues of the Trypan blue in the growth medium (amorphous shapes). Now, the area of each remained object is computed. In such image (that we call here Im_T for seek of simplicity) there will be both single cells and clusters and it will be used later-on to discriminate the two kinds of objects. At this point, the circularity of all the objects is calculated and a narrow threshold is set in order to select only the round shaped ones: these will be the single cells. The area of each object is computed.
In this way, the total number of single cell (N_s) present in the input image is obtained together with their areas. By running such algorithm over all the images corresponding to the same sample (15 in our case), it is possible to have enough statistics to properly estimate the representative value of the single cells (let's call such value A_s), defined as the mode of the single cell distribution sizes.

At this point the only missing part concerns the recognition of the clusters and the estimate of the number of single cells of area A_s contained in them. To this purpose, we considered two segmented images: the one with both single cells and clusters undistinguished (Im_T) and the one containing only single cells Im_S. By subtracting the second to the first one, a new image is obtained and it only contains the clusters. Of such clusters all those whose size is smaller than A_s and bigger than $200A_s$ are discarded, being them either single cells that have been wrongly detected as clusters or artefacts. The total area of the remaining clusters is then computed and divided by A_m. In this way, the estimate number (N_c) of cells of area A_s included in the clusters is obtained. In the end, by adding, for each image its corresponding N_s and N_c, the total number of cells is computed. By accounting the dilution factor due by the sample preparation and that due to the adding of the Trypan blue then the concentration of cells is finally obtained.

Figure A.5 shows an example of the input and the final output of the described algorithm. (a) and (c) are two examples of representative micrographs used as the starting point of the algorithm. Figure A.5(b) and (d) are instead the results of the running of the algorithm: in red are the single cells while in blue are the clusters.

Such algorithm revealed being robust and showed good performances across the analysed samples: both for samples with a few and a lot of clusters it managed to discriminate them from single cells and estimate the concentration.

In the next section we present the results of the new experiments we performed by adopting the new sample preparation protocol and counting algorithm.

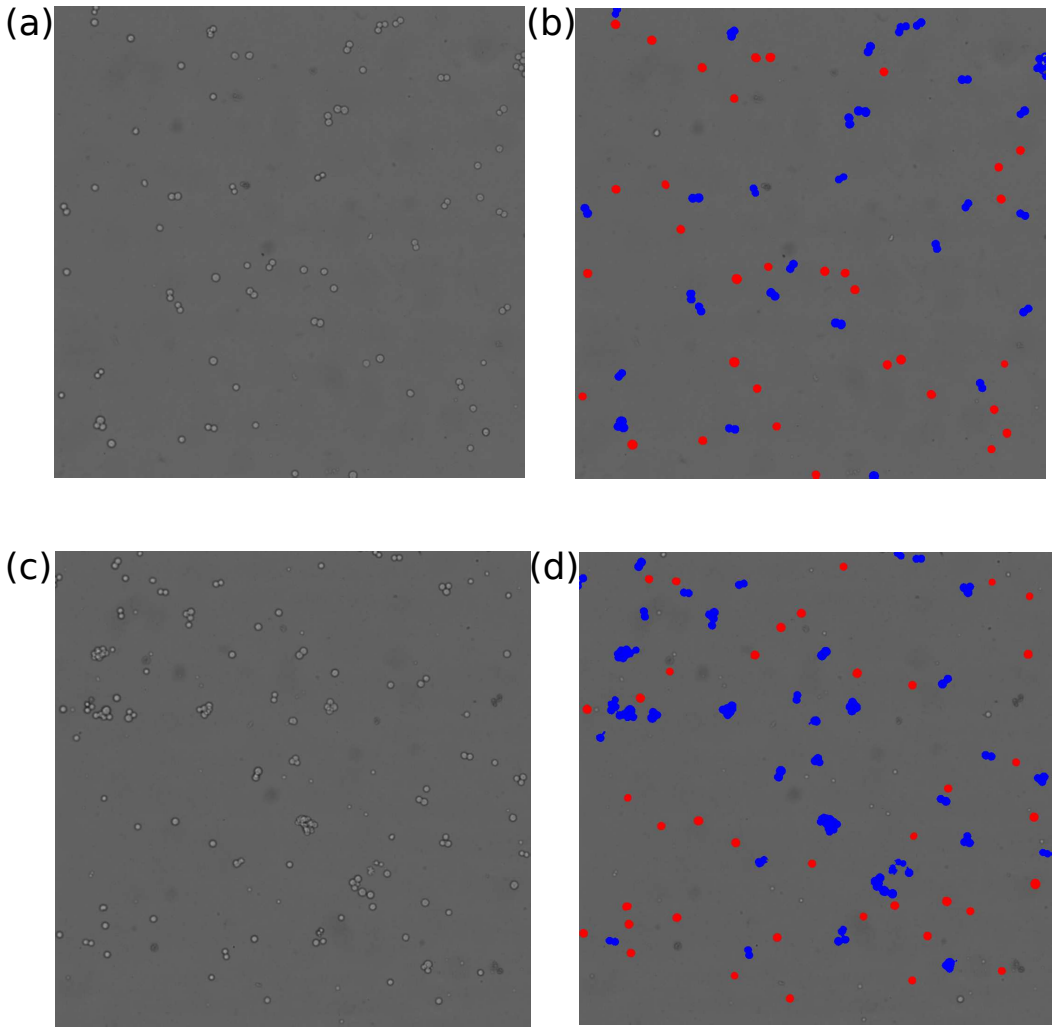

Fig. A.5 Examples of inputs (a,c) and corresponding outputs (b,d) of the new counting algorithm. The darkest spots in (a) and (b) are dead cells. In (b,d) the single cells are coloured in red and clusters in blue.

A.5.2 Testing of the algorithm on a second set of experiments

We repeated the same experiments described in Section A.3. We investigated the same inoculum densities of cells as previous experiments with only three concentrations of sodium acetate (i.e. 0 mM, 10 mM and 20 mM). Cells were cultured in flasks (with a different geometry than the former ones) with 20 ml of growth medium.

By daily monitoring the growth with the discussed protocol and algorithm, we obtained the following new results, shown in Figure A.6.

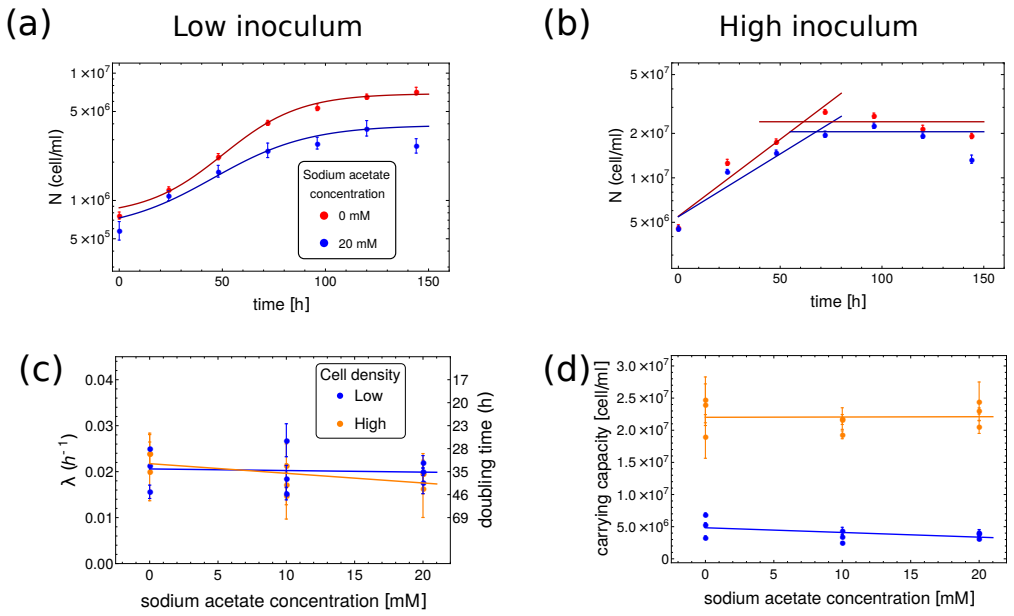

Fig. A.6 Results obtained through the use of the new algorithm and the new sample preparation protocol on the second set of experiments. (a) and (b) are representative growth curves at low and high inoculum density respectively. Red stands for a concentration of 0 mM of sodium acetate while blue is 20mM. Dots are experimental data, the lines show their trends. Each curve corresponds to a single replicate. The error bars are the propagation of the error on the counting. (c) Growth rate trend as a function of the sodium acetate concentration. Each dot corresponds to a single replicate, the error bars are the error of the fit. The lines represent the trend of the data (linear fits). Low density data are shown in blue, while high density ones are in orange. (d) Carrying capacity as a function of the sodium acetate concentration. Dots are experimental data and their error bars are the error of the fit. When not visible is because they have the same size than the dots. The lines represent linear fits of the experimental data. It is valid the same colour legend of plot in (d).

Figure A.6(a) and (b) show two representative growth curves for each set of experiment, low and high inoculum sizes respectively. Each growth curve represents a single replicate and the error bar is the error on the concentration obtained by propagating the standard deviation on the average of the counted number of cells of the 15 micrographs. Such errors are roughly within the 10% of the measurement suggesting thus a good compatibility among the different repeated measurements. The solid lines of A.6 are the fit of the growth curves through the sigmoidal shape function used to determine the growth phases and their representative parameters. Curves

belonging to the experiments at high inoculum sizes, as in the old experiments, show lack of lag time, therefore the phases of growth (exponential and saturation) have been determined as follows. The first one is a linear fit among the first four points of the curve. The slope represents the exponential growth rate λ. The saturation phase, instead has been considered as the average of the maximum value of the curve joined with its previous and following point.
Figure A.6(c) and (d) shows the results concerning the growth rate λ and the carrying capacity, analogously to Figure A.4(a) and (b). According to such data, the results seem opposite to the previous experiments. There, the growth rate for low cell density was decreasing in a sodium acetate dose-dependent manner, while that of high inoculum was constant. Here, the two values seem similar and with a slight decreasing trend for the data belonging to the high inoculum density. However, all the data are compatible within the error bars. A more evident difference is in the carrying capacity. First of all it shows different values between high and low inoculum densities even at 0mM of sodium acetate. This does not agree with the expectation of having a constant carrying capacity with respect to the inoculum size (See Chapter 2). In order to drive proper conclusions on such topic, it is necessary firstly to repeat the experiment in order to increase the statistics. Secondly, it is worth to systematically check with the image segmentation algorithm whether its performances drastically change when the concentration of cells in the flask is very high. In such case, indeed, the overall amount of clusters is increased and this may still affect the measurements causing the non homogeneity of the prepared sample.
Despite the difference between the two inoculum densities, the carrying capacity seems to remain roughly constant with a possible slight decrease for low inoculum densities depending on the dose of sodium acetate. Such trend is the opposite than that presented in Section A.4. What seems to be more evident is that the fluctuations of the carrying capacities become smaller as the sodium acetate concentration increases. However a larger set of experiments would be necessary in order to drive any conclusion.
We do not show any result concerning the lag time, since in all cases the lag phase resulted very short, even for the cases of low inoculum density. To better resolve such phase, a monitoring of the cell concentration more frequently in time would be necessary.

In light of the results obtained so far, it is not possible yet to drive quantitative conclusions on the effect of sodium acetate when supplied to a population of HEK293

cells. The only observation that can be made on this purpose concerns the fact that, in all cases, the increasing of sodium acetate in the growth medium seems to suggest a toxic activity on the growth parameters. None of the found trends were increasing together with the increasing of sodium acetate concentration. In all cases, the trends were either independent or decreasing in a dose-dependent manner with the sodium acetate.
Moreover, we can claim that the new methodology that we adopted to monitor cell growth is more sophisticated than the first one. Thus, it would be necessary to adapt such new counting algorithm for analysing the old data, in order to better estimate cell concentration of the first set of experiments and compare with the second one. It would be not trivial, given the high density of clusters and irregular shaped cells in the old micrographs, but it will certainly give better estimations and thus more reliable results. However, improvements are still necessary concerning high density data.

A.6 Conclusions

In this appendix we presented a quantitative method to systematically investigate the influence of sodium acetate on the growth of populations of HEK293 cells. We performed batch cell growth experiments with different (fixed) sodium acetate concentrations and two different initial population densities. The growth was daily monitored through automatic cell counting. This was performed by exploiting a custom-made image segmentation algorithm that requires as input micrographs of samples representative of the population.

We showed that the bottleneck of such quantification lies in such micrographs. The studied cells indeed tend to form strong clusters when growing. The dissolution of such clusters is necessary in order to have an homogeneous sample and measure its concentration in an automatic way. Such measurement is performed by taking an aliquote of a sample representative of the population under exam. We discussed how the pipetting and vortexing of such samples alters cell vitality, causing a change in the shape of the cells that in this way becomes irregular. Moreover, the higher is the concentration of cells in the sample, the higher the concentration of big clusters in the micrograph. Thus, by joining such effects, the estimate of the effective concentration of cells of the population through an image segmentation algorithm results difficult.

In order to improve the samples preparation, thus we showed firstly how it is better to not pipet the cells but only shake them. In this way cells undergo a smaller stress and do not change their shape, that remains regular and then easier to be detected by an the counting algorithm. Secondly, we suggested to dilute the sample when preparing the counting chambers in order to have a sparser distribution of cells over the slide. Again, this allows better performances of the algorithm.

Concerning the algorithm itself, we presented a first version applied the images with irregular shaped single cells and an high amount of clusters. Such algorithm allows to have a very rough estimate of the cells. Through this, indeed, the total segmented area is considered (i.e. undistinguished clusters and single cells) and divided by the mode of the distribution of the areas of the single segmented objects. We consider such last value as a proxy for the size of the single cells, on average more abundant than clusters. Such method gives a too rough estimate of cell concentration. We presented then an improved version that, starting from micrographs obtained through diluting samples, it is able to (i) recognise separately all the single cells and clusters present in an image, (ii) compute the mode of the distribution of the single cells as a proxy of the representative size of a cell and (iii) estimate the total number of such representative single cells contained in the clusters. In this way, by adding the number of counted real single cells to that of the single cells estimated within the clusters, the final concentration of cells is computed. Such algorithm is more sophisticated than the former one and gives also the possibility to recognise dead cells. In this case, indeed, we stained dead cells with a dye that makes them appearing darker than the alive ones in the images. The algorithm discards all the dark objects, counting thus only bright ones. A further improvement of such algorithm would be to recognise, among the dark objects those that are cells. This can be done by exploiting the roundness of cells that is maintained even when dead.

The described project is still in progress and, as future steps we intend to adapt such second algorithm to the old images in order to firstly have a better estimate of the concentration of cells and thus more reliable growth curves and secondly, to test the efficiency of the presented method to discriminate single cells and clusters in more crowded images.

By exploiting such methodologies we obtained robust growth curves through which we analytically determined the growth parameters useful for the quantification of the influence of sodium acetate on HEK293 cells growth. The results obtained by

exploiting the two described methodologies of sampling preparation and counting, suggest the possibility of a general light dose-dependent toxic activity of the sodium acetate on the growth. Moreover, they underline a possible cross correlation between such effect and the initial density of cell population. Going into details however, we obtained different results between the application of the two methods, thus more detailed quantitative discussions cannot be done yet. Despite of the application of the new algorithm to the old samples in order to better quantify the growth, further experiments will be necessary in order to have a more robust statistics to drive conclusions. In addition to this, the quantification of the bioproducts will allow to quantitative cross correlate the productivity of such cells with the sodium acetate concentration supplied to the medium and thus growth variables.

The work here presented will be continued and it will be part of a publication in preparation.

Appendix B

The role of STIM1 in endothelial cell endosomes trafficking

This project has been performed in collaboration with Prof. G. Serini and Dr. G. Villari from Institute for Cancer Research and Treatment (Candiolo, Italy), and aims at investigating endosomes trafficking.
Our collaborators found that STIM1, a protein that promotes the formation of a molecular motor complex (dynein/dynactin) formation, regulates the traffic of late and early endosomes, vesicles located within cells. Our contribution has been the development of an image processing algorithm able to track fluorescent endosomes in time-lapse videos taken with a confocal microscope. Through the tracking, characteristic quantities related to the morphology and the dynamics of the detected endosomes have been quantified and then used to support the experimental observations.
The work is object of a publication in preparation together with Serini's lab entitled: "STIM1 regulates the dynein/dynactin complex formation contributing to the transport of early and late endosomal traffic on microtubules".

In the following Section B.1, we report a brief introduction to the biological issue. Section B.2 is devoted to methods and to the image segmentation algorithm we developed and finally the last Section B.3 we present the principal results of the project.

B.1 Overview on endosome trafficking and project outline

Endosomes are membrane-delimited compartments found in the cytoplasm of eukaryotic cells that control the recycling or the degradation of membrane components [144]. They are formed by endocytosis, an either receptor- or non-receptor-mediated process through which cells internalise external materials (fluids or macromolecules) by invaginating the plasma membrane and thus forming vesicles and vacuoles. Endosomes can be considered as intracellular transport carriers since they transport the extracellular molecules within the cell. The first endosomes to be formed are called early endosomes (EE). Some material that they contain can return back to the plasma membrane to be recycled, thanks to the so-called recycling endosomes. Other molecules, instead, can be brought towards the inner part of the cell by the so called late endosomes (LE). Once there, they might be degraded by the lysosomes, membrane-enclosed organelles able to break down all biological composites into single units. The passage from early to late endosomes is called endosome maturation [144]. As early endosomes can be considered the first sorting stations of the cell, late endosomes serve as second sorting and trafficking hub of the endosomial system. Evidences suggest that such trafficking is also influenced by endosomes interaction with the endoplasmatic reticulum (ER) which is present in the whole cytoplasm [145]. Endosomial system can thus be considered as a dynamic and organised complex system, where its single components can autonomously move [146]. These movements are driven by molecular motors, namely molecular machines able to consume and convert energy into motion. In particular, they convert chemical energy into work allowing the intracellular cargo transport along cytoskeletal structures, such as microtubules or actin filaments. Two examples of widely studied molecular motors are kinesins and dyneins, usually responsible for the centrifugal (anterograde) and centripetal (retrograde) transport respectively [145].

The aim of the project lies in this framework focusing on the influence of the protein STIM1 on the early and late endosomal trafficking in human endothelial cells.
STIM1 (Stromal interaction molecule 1) is a transmembrane protein known for its ability to control endothelial cell migration [147] and barrier functions [148]. For example, in condition of Ca^{2+} depletion in the endoplasmatic reticulum, STIM1

activates a plasma membrane Ca^{2+} channel (ORAI1) allowing extracellular calcium to enter and refill the storing sites [149]. Another known function of STIM1 is its ability to bind to a protein (+TIP protein EB1) located at the pheripherical end of microtubules by allowing the elongation of endoplasmatic reticulum tubules together with microtubules [149].

Through mass spectometry and biochemical experiments (co-immunoprecipitation experiments) our collaborators identified STIM1 as an interactor of neuropilin 1 (NRP1), an endocytic receptor known to be involved in endosome trafficking [150]. Therefore the project was focused exactly on the investigation of the role of STIM1 in endosome trafficking, in particular on its influence on molecular motors involved in the dynamics of both early and late endosomes. For this purpose, confocal images and time-lapse videos have been performed on fluorescent marked endosomes. A detailed image analysis allowed to quantify information on characteristics and behaviours of endosomes under different conditions. Our contribution to the project fits here, in the development of an algorithm of image segmentation that allows to track endosomes over time and identifies their total number, size and position. In the next Section B.2 we present a description of the algorithm and the analysed quantities. The results of both this analysis and other related experiments, are presented in the last Section B.3.

B.2 Methods for endosome tracking

In order to quantitatively investigate the influence of STIM1 on endosome trafficking in endothelial cells, it has been necessary to quantify variables related to the morphology and the dynamics of the organelles exposed to different conditions. For this purpose, once the environmental conditions, i.e. cell treatment, have been determined, endosomes were marked with a fluorescent dye (red for LE and green for EE) in order to make them visible under a confocal microscope. Thus, both time-lapse videos (2 frames per second) and fixed images were taken. We developed an image segmentation and particle tracking algorithm that implements built-in Matlab functions and is able to recognise and follow endosomes frame by frame. In this way, for each detected endosome we quantified size, distance from the nucleus and the deviation of the motion from a linear trajectory (persistence).

The algorithm takes as input single frames of time lapse videos. The image acquisition has been obtained in RGB colour scale, thus as a first preliminary step, the channel of interest is selected. From this point on, the algorithm consists of five different steps which are described below:

segmentation to firstly recognise objects.

check1 (from frame 2 on) to map new objects into the old ones, as long as the centroid of new ones is located within a certain range from the old ones.

connected components to check that no more than one object of the new frame is mapped in one object of the previous frame. If this happens, the new endosome closest to the old one is considered as the correct one to map. The other ones become new objects.

check2 to look for lost matches in mapping due to the presence of more than one connected components.

variables saving to save interesting variables, i.e. objects sizes and centroid positions.

N.B. For all the details about the cited Matlab functions and about the methods, see Section 2.3.2.

In the first *segmentation* step, an image segmentation is performed by labelling recognised objects (through the function *bwlabel*). The threshold for the labelling was chosen as the sum of the mean value of the intensity distribution of the image plus twice its standard deviation. In this way, eventual differences between frames are taken into account and part of the background is removed. For each detected object then, after a proper removal of all those objects with size lower than a threshold, the position of each centroid is saved. The segmentation of the first frame (frame corresponding to time $t = 0$ s) ends here together with saving the information on the size of each detected endosome. Size and centroid positions are defined through the Matlab function *regionprops*.

From the second frame on, the aim is to track endosomes over time. In order to do it, the code detects elements in the new frame with the same segmentation procedure used in the first one (*segmentation*) and then maps the new labels into the old ones. For this purpose, a check between the object of the actual frame and the

previous one is performed (*check1*). For example: if endosome called B at time t (B_t) is located within a certain range with respect to endosome C at time $t-1$ (C_{t-1}), then the name B_t is changed in C_t, because they represent the same object. If, in the second frame, no match is found then the name of the endosome is saved in an array that is updated frame by frame and that contains (for each frame) the names of all the endosomes that disappeared. At this moment a second check is necessary (*connected components* step) in order to be sure of not having two endosomes mapped in the same endosome of the previous frame, i.e. two or more with the same label. For this purpose, by running on all endosome names, the algorithm checks the number of connected components. If it is bigger than 1, it chooses as the right endosome for the new mapping that endosome which has the minimum spatial distance from the centroid of the endosome with the same name of the previous frame. New names are then given to the other components.

Now a new check *(check2*) with the same procedure used for *check1*, is necessary to assure that those endosomes which were considered as "disappeared" were truly so (and not wrongly labelled). If a match is not found, then the endosome maintains its label.

At this point the code ends by saving in arrays the size and centroid position of the present endosomes (*variables saving*). These information are then used for the calculation and the identification of the interesting variables describing the system. The so described code is repeated for all the frames of the time-lapse.
The output of this part of the code is the total number of detected objects per frame, together with their size and centroid position.

Since the endosomes are free to fluctuate in different planes of the endoplasmatic reticulum, they can eventually be lost frame to frame and eventually re-appear in the same position after a while. From the trajectories of the detected endosomes this information is available and used to split trajectories. Figure B.1 shows an example that we use to clarify this point. In the plot, the size of endosome #174 (name given after the *check2* step) over time is shown. A value equal to 0 means that none endosome can be mapped into this one at that specific frame. Due to the fact that endosomes can appear and disappear, we decided to consider the detected endosome as two different objects when there are at least three frames without any detection of the specific endosome. This is shown by the three different colours and new names (in the legend) of endosome #174 in Figure B.1. Afterwards, for

the analysis of the morphological and dynamical quantities, we consider only those endosomes that can be seen for at minimum 21 frames ($10\,s$), like the green one in the plot.

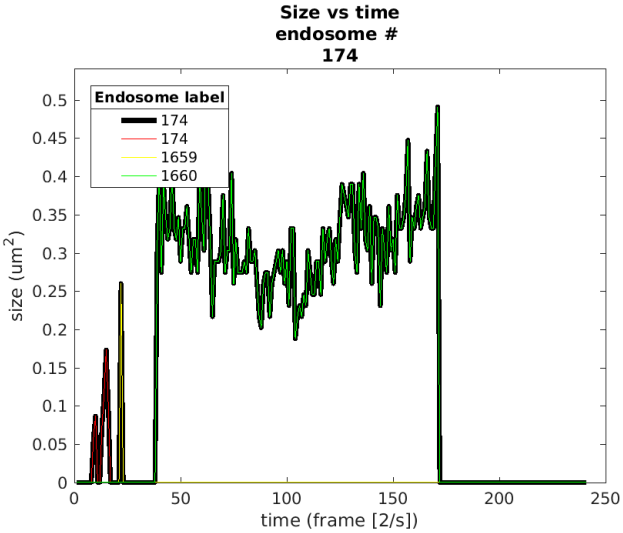

Fig. B.1 Example of the size trend in time of an early endosome detected with the image segmentation and tracking code. The three colours represent the three endosomes in which the first detected one (#174) is split. For further analysis only the green one will be considered, since it appears continuously for more than 21 frames.

As already mentioned, the variables of interest are number, size, distance from the nucleus and deviation from a linear trajectory:

Number is total number of elements detected in each frame. For this purpose, all endosomes were considered, independently of the duration of their trajectories.

Size is the conversion in μm^2 of the size in pixel obtained from the segmentation algorithm.

Distance from the nucleus is the distance between the punctual position of the endosome and the centre of the cell nucleus.

Persistence is a deviation from a linear trajectory and is calculated as the ratio between the minimum path (p_{min}) and the effective one (p_{eff}).

Each of these variable is discussed in Section B.3.

For each experiment we performed a comparison between the results relative to endosomes in cells with silenced STIM1 protein and those from cells of control (wild type) (see Section B.3 for details). The way we adopted for analysing differences in the distributions of the observables relative to the two cases is the Kolmogorov-Smirnov test that we briefly describe in the next paragraph.

Kolmogorov-Smirnov test

Kolmogorov-Smirnov test is usually adopted for defining whether two samples come from the same population or not. The parameter characteristic for this test (D_k), given by the maximum distance between the empirical distribution functions of the two samples under exam $F_1(x)$ and $F_2(x)$ of sizes N_1 and N_2 respectively.

$$D_k = max\{|F_1(x) - F_2(x)|\} \tag{B.1}$$

If the two distributions come from the same population, then the parameter D_k must be lower than a critical value D_{crit} :

$$D_{crit} = c(\alpha)\sqrt{\frac{N_1 + N_2}{N_1 N_2}} \tag{B.2}$$

where α is the significativity chosen for the test. $c(\alpha)$, for n sufficiently large, is equal to the inverse of the Kolmogorov distribution. For significativity $\alpha = 5\%$ and populations bigger than 50 data, then $c(\alpha) = 1.36$.

B.3 Results and discussion

The relation between endosome trafficking and STIM1 has been studied in human umbilical vein endothelial cells (HUVEC).

As mentioned in Section B.1, from mass spectroscopy and co-immunoprecipitation experiments, it has been found that STIM1 is related to the protein NRP1, a receptor involved in endocytosis. Biochemical analysis showed that the silencing of STIM1 (siSTIM1) increases the NRP1 endocytosis, and thus early endosomes formation. This suggested a possible influence of STIM1 on endosomes formation and eventually also on their dynamics. In order to investigate this issue, two different sets of

experiments were set up with the same cells treated in two different ways. A first set used as a control (CTL) was compared to the second one, in which the protein STIM1 has been silenced. Endosomes of these cells were marked with fluorescent dyes in order to make them visible under a microscope. In this way, we have time lapse videos for early endosomes and fixed confocal images for late endosomes.

Figure B.2 compares the features of CTL and siSTIM1 endosomes obtained from the image analysis described in Section B.2. Early endosome plots refer to time lapse videos and therefore the temporal trend of number and size is shown. For the distance from the nucleus and the persistence, we show the trend at $10\,s$ from the beginning of the endosome trajectories. Late endosomes, instead were analysed in single frame images and therefore we show the distribution of static quantities. A schematic representation of these first results are shown in Figure B.3(a).

If the silencing of STIM1 increases early endosomes formation, the number of early endosomes in siSTIM1 should be higher than the control. However we found fewer endosomes than the control (see Figure B.2(a)), but with average bigger sizes (Figure B.2(b)). This suggests that STIM1 silencing may promote the formation of early endosomes which then tend to fuse together, resulting bigger and fewer then expected. Note that the sizes of siSTIM1 EE are much more dispersed than in the control.

By looking at the same variables on late endosomes (Figure B.2(a) and (b), LE), they show, as expected, a behaviour opposite to that of EE. Late endosomes in cells where STIM1 have been silenced are fewer. Concerning the size, from Figure B.2(b), LE siSTIM1 have the same size of the control. However, this can be due to thresholding issues in the segmentation algorithm, necessary to remove undesired objects. Further experiments would be necessary in order to detect whether a difference in sizes is present between silenced and not silenced STIM1 conditions.

By focusing on the distribution of the distances from the nucleus of siSTIM1 early endosomes (Figure B.2(c)), it is worthy to notice that it is narrower than the control, and centred towards lower values. In order to support this observation and thus the existence of a difference between the two distributions, a Kolmogorov-Smirnov test has been performed (see Section B.2). The cumulative distribution functions relative to the test are shown in the second plot of Figure B.2(c). The test gave a result higher than the threshold ($D_k = 0.21$ and $D_{crit} = 0.12$), thus confirming the difference between the two distributions. About late endosomes, we notice an

opposite trend: siSTIM1 LE are located farther from the nucleus with respect to their control.

Figure B.2(d) shows the persistence of early endosomes as function of their minimum path. The persistence gives information on the deviation from a linear trajectory and is the ratio between the minimum and the effective path made by the endosomes in time. The persistence of early endosomes seems to be less dispersed and lower than that of their controls. This result suggests that siSTIM1 EE follow a path which is less linear than that followed by the control, namely when silencing STIM1 the motion seems less guided.

From literature, it is known that the microtubules-dependent motility of late endosomes is driven by a machinery that involves the dynein/dynactin motor complex, responsible for a retrogade (centripetal) motion [145]. Additional biochemical analysis have been performed then in this direction and showed that STIM1 is effectively related to the dynein. Deeper investigations have been carried out by performing new experiments. This time, the dynein activity has been inhibited (siDyn in Figure B.3), and thus the known retrograde motor suppressed. To inhibit dynein activity, cells were treated with Ciliobrevin D, a repressor of dynein ATPase activity. In this condition, LE were located far from the nucleus, as expected, while EE showed the same behaviour obtained with STIM1 silencing: they were bigger than the control and close to the nucleus. See Figure B.3(b) for a scheme of these results. To recapitulate, all these findings allow us to support a theory in which the presence of STIM1 promotes the formation of the complex dynein/dynactin, required for both retrograde and anterograde transport in early and late endosomes respectively.

Since for EE the dynein silencing did not restore the STIM1 control situation (endosomes far from the nucleus), it is supposed the presence of a second retrograde motor in addition to the dynein, responsible for the perinuclear accumulation of early endosomes. As mentioned in Section B.1, kinesins are usually associated to anterograde transports, however some of them exist as retrograd motors. KIFC1 is a retrograd kinesin expressed by the cells we examined. Therefore experiments of KIFC1 silencing have been performed and showed that it causes the perinuclear accumulation of EE, while it was not influent on LE (see schematics of in Figure B.3(c)). This allowed to assume then that KIFC1 was responsible for motility of the early but not of the late endosomes. The association of silencing of both KIFC1

and dynein restored the same early endosomes position obtained in the control of siSTIM experiments, namely EE far from the nucleus and dispersed in the cytoplasm, as represented in Figure B.3(d). A deeper quantification of the dynamics of such new treatments concerning the silencing of dynein and KIFC1 will be performed by exploiting the same image segmentation algorithm used to investigate siSTIM1 protein.

To conclude, a summary of the results obtained with the described experiments are shown in Figure B.3. There, each square represents a portion of a cell where the nucleus is emphazised by a circle, the membrane by a black line. The green and red filled circles represent early and late endosomes respectively. Circles dimensions are representative for endosome size.

We found that protein STIM1 inhibits endocytosis mediated by NRP1 receptor and promotes the assembly of dynein/dynactin complex, which is responsible for the centripetal and centrifugal transport along microtubuli of LE and EE respectively. Furthermore, we identified a kinesin (KIFC1) that acts as an additional retrograde motor only for EE and not for LE. Recently, Herbert and colleagues [151] proved the existence of dyneins, usually involved in retrograd transports, which are necessary for anterograd movements. Similarly we found a kinesin, usually an anterograd motors, involved in retrograd transports.

The results here presented are object of a publication in preparation together with Serini's lab.

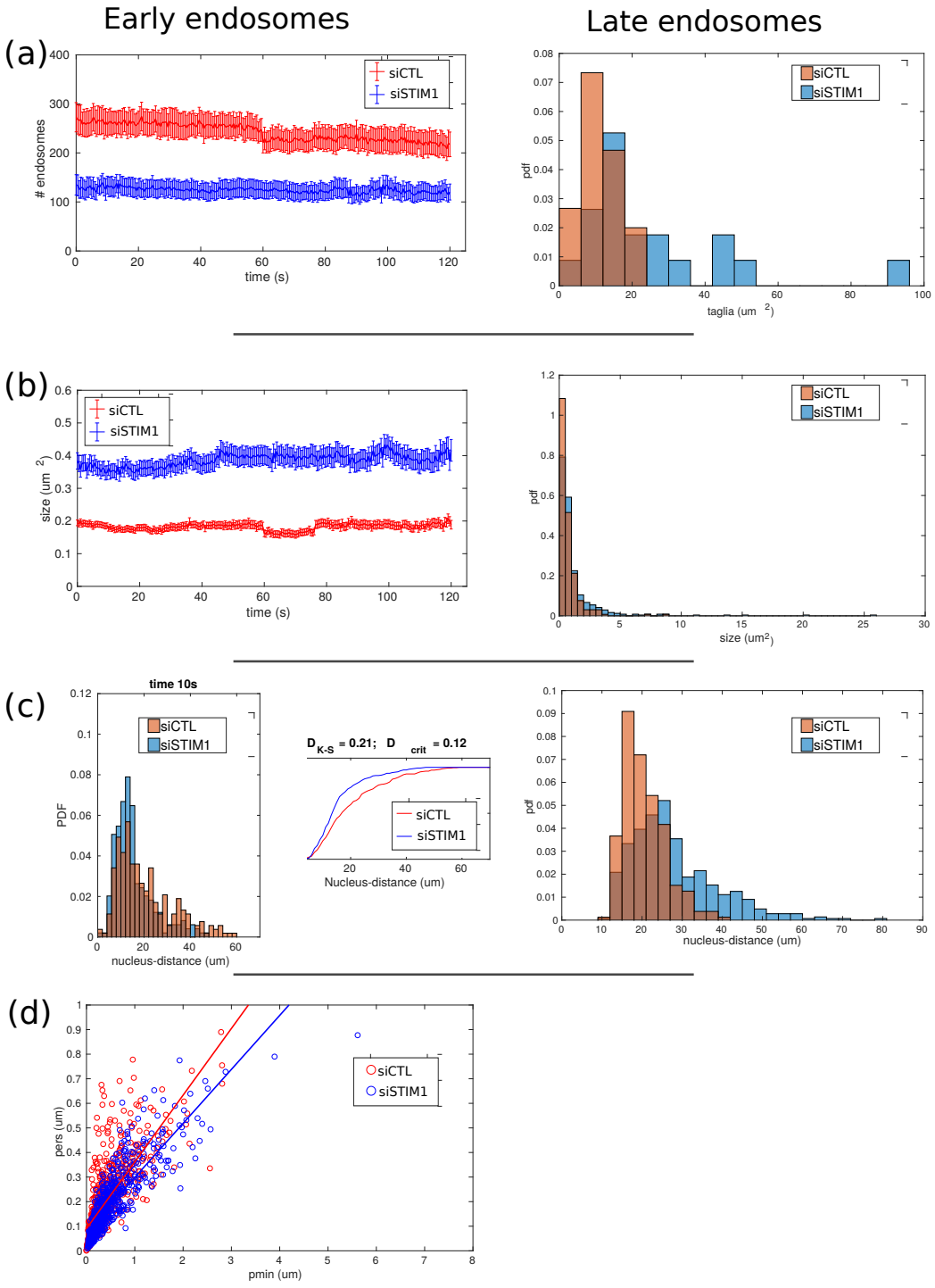

Fig. B.2 Comparisons between features of early (EE, left column) and late endosomes (LE, right). Blue data represent siSTIM1 condition, red/orange ones the control. (a) Shows the average number of endosomes per cell *vs* time for EE and number distribution for LE. In(b) again temporal average trend for EE and distribution at a fixed time for LE of endosome sizes. Line (c) is dedicated to distance to nucleus distributions. For EE the plot of cumulative distribution functions showing the difference between the two populations is shown. Last line (d) represents the persistence versus minimal path of early endosomes, not available for LE. Blue and red lines here represent general trend of correspective data.

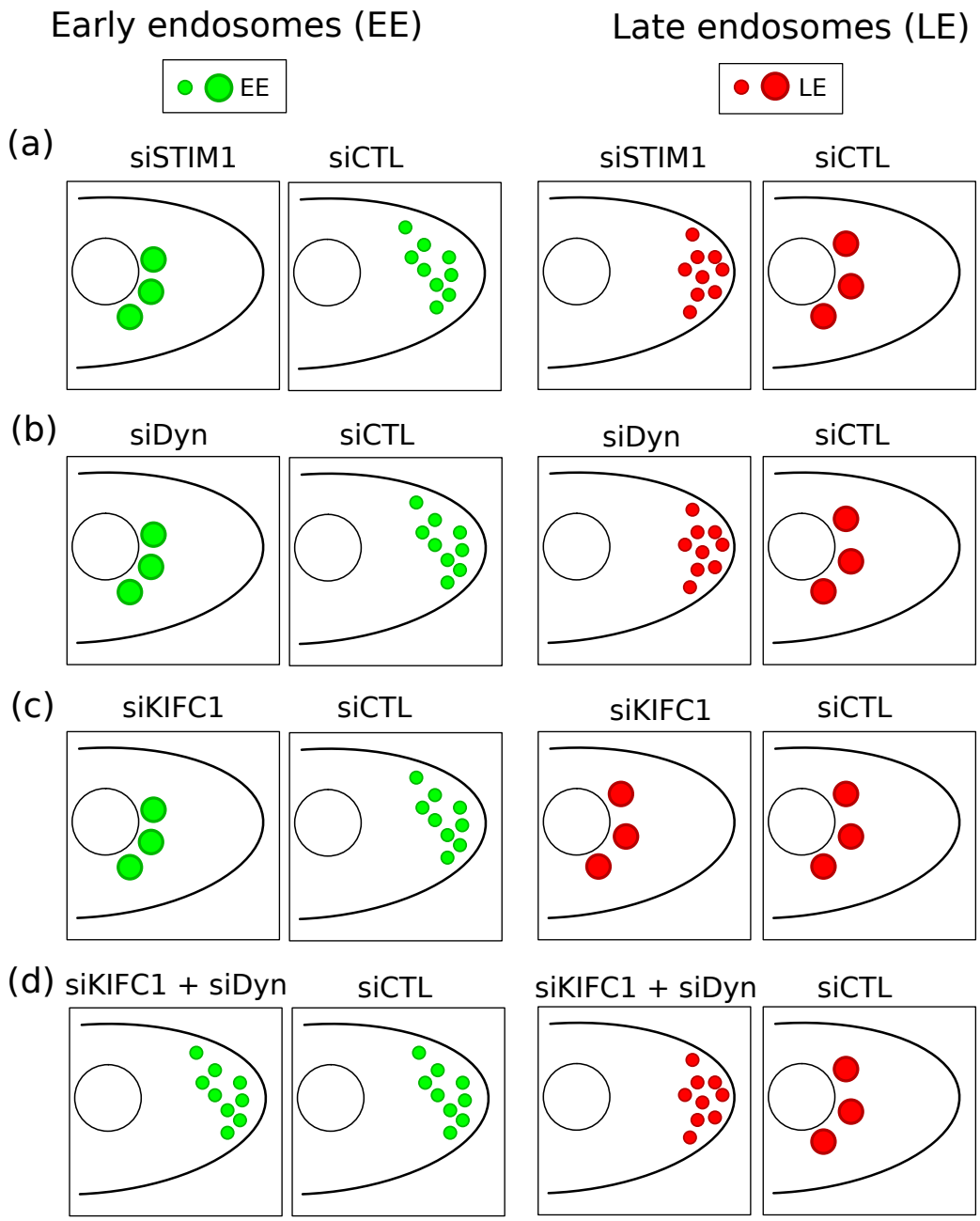

Fig. B.3 Schematic representation of the main results of the project. Each square represents a portion of a cell, where the empty circle is the nucleus and the black line the membrane. Early and late endosomes are represented respectively as green and red filled circles whose sizes give an idea of the differences in endosomes size. From top to bottom the control (siCTL) situation is shown together with the following conditions: (a) STIM1 silencing (siSTIM1); (b) dyneine silencing (siDyn); (c) KIFC1 silencing (siKIFC1) and (d) the combination of silencing of KIFC1 and dynein together.

References

[1] Konstantinos P. Koutsoumanis and Alexandra Lianou. Stochasticity in colonial growth dynamics of individual bacterial cells. *Applied and Environmental Microbiology*, 79(7):2294–2301, 2013.

[2] J Monod. The growth of bacterial cultures. *Annu. Rev. Microbiol.*, 3(1):371–394, October 1949.

[3] Matthew Scott. Quantitative moethods in bacterial physiology. ICTP Spring College on the Physics of Complex Systems Trieste, Italy 2017.

[4] BioRad. Cell counting protocol - counting cells with a hemocytometer and a microscope. Available at: http://www.bio-rad.com/it-it/applications-technologies/cell-counting-methods?ID=LUSOLB470.

[5] Matthew Scott, Carl W. Gunderson, Eduard M. Mateescu, Zhongge Zhang, and Terence Hwa. Interdependence of cell growth and gene expression: Origins and consequences. *Science*, 330(6007):1099–, November 2010.

[6] Matthew Scott and Terence Hwa. Bacterial growth laws and their applications. *Current Opinion in Biotechnology*, 22(4):559–565, 2011.

[7] Francesco Fazi and Clara Nervi. Microrna: basic mechanisms and transcriptional regulatory networks for cell fate determination. *Cardiovascular Research*, 79(4):553–561, 2008.

[8] Marco Del Giudice, Stefano Bo, Silvia Grigolon, and Carla Bosia. On the role of extrinsic noise in microrna-mediated bimodal gene expression. *PLOS Computational Biology*, 14(4):1–26, 04 2018.

[9] Shankar Mukherji, Margaret S Ebert, Grace X Y Zheng, John S Tsang, Phillip A Sharp, and Alexander van Oudenaarden. Micrornas can generate thresholds in target gene expression. *Nature Genetics*, 43:854–, August 2011.

[10] Carla Bosia, Francesco Sgrò, Laura Conti, Carlo Baldassi, Davide Brusa, Federica Cavallo, Ferdinando Di Cunto, Emilia Turco, Andrea Pagnani, and Riccardo Zecchina. Rnas competing for micrornas mutually influence their fluctuations in a highly non-linear microrna-dependent manner in single cells. *Genome Biology*, 18(1):37, Feb 2017.

[11] M Del Giudice. *Molecular sequestration as a way to control variability and induce crosstalkS*. PhD thesis, 2019.

[12] M H Zwietering, I Jongenburger, F M Rombouts, and K van 't Riet. Modeling of the bacterial growth curve. *Applied and Environmental Microbiology*, 56(6):1875–1881, June 1990.

[13] J. D. Watson and F. H. C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171:737–, April 1953.

[14] Athene Donald. Where physics meets biology. *The Guardian*, 2012.

[15] Shashank Shekhar, Lian Zhu, Linas Mazutis, Allyson E Sgro, Thomas G Fai, and Marija Podolski. Quantitative biology: where modern biology meets physical sciences. *Molecular biology of the cell*, 25(22):3482–3485, November 2014.

[16] William Bialek, Andrea Cavagna, Irene Giardina, Thierry Mora, Edmondo Silvestri, Massimiliano Viale, and Aleksandra M. Walczak. Statistical mechanics for natural flocks of birds. *Proc Natl Acad Sci USA*, 109(13):4786–, March 2012.

[17] Ibon Santiago. Nanoscale active matter matters: Challenges and opportunities for self-propelled nanomotors. *Nano Today*, 19:11–15, 2018.

[18] Simon K. Schnyder, John J. Molina, Yuki Tanaka, and Ryoichi Yamamoto. Collective motion of cells crawling on a substrate: roles of cell shape and contact inhibition. *Scientific Reports*, 7(1):5163–, 2017.

[19] Nicholas C Darnton, Linda Turner, Svetlana Rojevsky, and Howard C Berg. Dynamics of bacterial swarming. *Biophysical Journal*, 98(10):2082–2090, January 2010.

[20] Ezio Marinari. Un nuovo incontro: aspetti fondamentali di fisica e biologia, in cent anni di relatività. *Enciclopedia Treccani, Italia www.treccani.it/site/Scuola/Zoom/scuolazoom.htm*, 2005.

[21] Gunnar Jeschke. *Advanced Physical Chemistry Statistical Thermodynamics, Lecture Notes*. 2015.

[22] Sriram Ramaswamy. The mechanics and statistics of active matter. *Annual Review of Condensed Matter Physics*, 1(1):323–345, 2010.

[23] William Bialek. Perspectives on theory at the interface of physics and biology. *Reports on Progress in Physics*, 81(1):012601, 2018.

[24] Edgard B. Carvalho and Wayne R. Curtis. The effect of inoculum size on the growth of cell and root cultures ofhyoscyamus muticus: Implications for reactor inoculation. *Biotechnology and Bioprocess Engineering*, 4(4):287–293, 1999.

[25] Ana C. Gregório, Nuno A. Fonseca, Vera Moura, Manuela Lacerda, Paulo Figueiredo, Sérgio Simões, Sérgio Dias, and João Nuno Moreira. Inoculated cell density as a determinant factor of the growth dynamics and metastatic efficiency of a breast cancer murine model. *PLOS ONE*, 11(11):1–19, 11 2016.

[26] Walter Gulik, Anna Nuutila, Ko L. Vinke, Hens J. G. ten Hoppen, and Sef Heijnen. *Effect of CO2 airflow rate and inoculation density on the batch growth of Catharanthus roseus cell suspensions in stirred fermentors*, volume 10. May 1994.

[27] Kanokwan Kanokwaree and Pauline M. Doran. Effect of inoculum size on growth of atropa belladonna hairy roots in shake flasks. *Journal of Fermentation and Bioengineering*, 84(4):378–381, 1997.

[28] R. C. L. Marteijn, M. M. A. Oude-Elferink, D. E. Martens, C. D. de Gooijer, and J. Tramper. Effect of low inoculation density in the scale-up of insect cell cultures. *Biotechnol Progress*, 16(5):795–799, September 2008.

[29] E N. Rodriguez, M Perez, P Casanova, and L Martinez. *Effect of Seed Cell Density on Specific Growth Rate Using CHO Cells as Model*. January 2001.

[30] Sadettin S. Ozturk and Bernhard Ø. Palsson. Effect of initial cell density on hybridoma growth, metabolism, and monoclonal antibody production. *Journal of Biotechnology*, 16(3):259–278, 1990.

[31] Matthew Scott, Stefan Klumpp, Eduard M Mateescu, and Terence Hwa. Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Molecular Systems Biology*, 10(8), 2014.

[32] Eshel Ben-Jacob, Donald S. Coffey, and Herbert Levine. Bacterial survival strategies suggest rethinking cancer cooperativity. *Trends in Microbiology*, 20(9):403–410, 2012.

[33] Guillaume Lambert, Luis Estévez-Salmeron, Steve Oh, David Liao, Beverly M. Emerson, Thea D. Tlsty, and Robert H. Austin. An analogy between the evolution of drug resistance in bacterial communities and malignant tissues. *Nature Reviews Cancer*, 11:375–, April 2011.

[34] Carla Bosia, Andrea Pagnani, and Riccardo Zecchina. Modelling competing endogenous rna networks. *PLOS ONE*, 8(6):1–13, 06 2013.

[35] Matteo Figliuzzi, Enzo Marinari, and Andrea De Martino. Micrornas as a selective channel of communication between competing rnas: a steady-state theory. *Biophysical journal*, 104(5):1203–1213, March 2013.

[36] Javad Noorbakhsh, Alex H Lang, and Pankaj Mehta. Intrinsic noise of microrna-regulated genes and the cerna hypothesis. *PloS one*, 8(8):e72676–e72676, August 2013.

[37] Peter S. Swain, Michael B. Elowitz, and Eric D. Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences*, 99(20):12795–12800, 2002.

[38] Chiara Enrico Bena, Alice Grob, Mark Isalan, Carla Bosia, and Francesca Ceroni. Commentary: Synthetic addiction extends the productive life time of engineered escherichia coli populations. *Frontiers in Bioengineering and Biotechnology*, 6:77–, May 2018.

[39] Zhou Jiang and Susan T. Sharfstein. Sodium butyrate stimulates monoclonal antibody over-expression in cho cells by improving gene accessibility. *Biotechnol. Bioeng.*, 100(1):189–194, November 2007.

[40] G Stamatoyannopoulos, CA Blau, B Nakamoto, B Josephson, Q Li, E Liakopoulou, B Pace, T Papayannopoulou, SW Brusilow, and G Dover. Fetal hemoglobin induction by acetate, a product of butyrate catabolism [see comments]. *Blood*, 84(9):3198–, November 1994.

[41] J. Rodriguez, M. Spearman, N. Huzel, and M. Butler. Enhanced production of monomeric interferon-ßby cho cells through the control of culture conditions. *Biotechnology Progress*, 21(1):22–30, 2005.

[42] Ping Wang, Lydia Robert, James Pelletier, Wei Lien Dang, Francois Taddei, Andrew Wright, and Suckjoon Jun. Robust growth of escherichia coli. *Current biology : CB*, 20(12):1099–1103, June 2010.

[43] Sattar Taheri-Araghi, Serena Bradde, John T Sauls, Norbert S Hill, Petra A Levin, Johan Paulsson, Massimo Vergassola, and Suckjoon Jun. Cell-size control and homeostasis in bacteria. *Current biology : CB*, 25(3):385–391, February 2015.

[44] M.O. Schaechter, O Maaloe, and N.O. Kjeldgaard. *Dependency of Medium and Temperature on Cell Size and Chemical Composition During Balanced Growth of Salmonella Typhimurium*, volume 19. January 1959.

[45] Sattar Taheri-Araghi, Steven D. Brown, John T. Sauls, Dustin B. McIntosh, and Suckjoon Jun. Single-cell physiology. *Annual Review of Biophysics*, 44(1):123–142, 2015. PMID: 25747591.

[46] Mikihiro Hashimoto, Takashi Nozoe, Hidenori Nakaoka, Reiko Okura, Sayo Akiyoshi, Kunihiko Kaneko, Edo Kussell, and Yuichi Wakamoto. Noise-driven growth rate gain in clonal cellular populations. *Proceedings of the National Academy of Sciences*, 113(12):3251–3256, 2016.

[47] Farshid Jafarpour, Charles S. Wright, Herman Gudjonson, Jedidiah Riebling, Emma Dawson, Klevin Lo, Aretha Fiebig, Sean Crosson, Aaron R. Dinner, and Srividya Iyer-Biswas. Bridging the timescales of single-cell and population dynamics. *Phys. Rev. X*, 8:021007, Apr 2018.

[48] Philipp Thomas. Making sense of snapshot data: ergodic principle for clonal cell populations. *Journal of The Royal Society Interface*, 14(136), 2017.

[49] C. Enrico Bena, M. Del Giudice, T. Gueudré, M. Miotto, E. Turco, A. De Martino, and C. Bosia. Inoculum-density dependent growth reveals inherent cooperative effects and stochasticity in cancer cell cultures. *arXiv:1710.10978*, 2017.

[50] J. Barrett Deris, Minsu Kim, Zhongge Zhang, Hiroyuki Okano, Rutger Hermsen, Alexander Groisman, and Terence Hwa. The innate growth bistability and fitness landscapes of antibiotic-resistant bacteria. *Science*, 342(6162), 2013.

[51] Gene-Wei Li, David Burkhardt, Carol Gross, and Jonathan S. Weissman. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, 157(3):624–635, April 2014.

[52] Antonio A. Alonso, Ignacio Molina, and Constantinos Theodoropoulos. Modeling bacterial population growth from stochastic single-cell dynamics. *Applied and Environmental Microbiology*, 80(17):5241–5253, 2014.

[53] József Baranyi. Stochastic modelling of bacterial lag phase. *International Journal of Food Microbiology*, 73(2):203 – 206, 2002. 3rd International Conference on Predictive Modelling in Foods.

[54] R.L Buchanan, R.C Whiting, and W.C Damert. When is simple good enough: a comparison of the gompertz, baranyi, and three-phase linear models for fitting bacterial growth curves. *Food Microbiology*, 14(4):313 – 326, 1997.

[55] Carmen Pin and József Baranyi. Kinetics of single cells: Observation and modeling of a stochastic process. *Applied and Environmental Microbiology*, 72(3):2163–2169, 2006.

[56] J. Postma, C.H. Hok-A-Hin, and J.H. Oude Voshaar. Influence of the inoculum density on the growth and survival of rhizobium leguminosarum biovar trifolii introduced into sterile and non-sterile loamy sand and silt loam. *FEMS Microbiology Letters*, 73(1):49–57, 1990.

[57] Peter L. Irwin, Ly-Huong T. Nguyen, George C. Paoli, and Chin-Yi Chen. Evidence for a bimodal distribution of escherichia coli doubling times below a threshold initial cell concentration. *BMC Microbiology*, 10(1):207–, 2010.

[58] József Baranyi and Terry A. Roberts. A dynamic approach to predicting bacterial growth in food. *International Journal of Food Microbiology*, 23(3):277–294, 1994.

[59] Daniele De Martino, Fabrizio Capuani, and Andrea De Martino. Quantifying the entropic cost of cellular growth control. *Phys. Rev. E*, 96(1):010401–, July 2017.

[60] Daniel Madar, Erez Dekel, Anat Bren, Anat Zimmer, Ziv Porat, and Uri Alon. Promoter activity dynamics in the lag phase of escherichia coli. *BMC Systems Biology*, 7(1):136–, 2013.

[61] I.A.M. Swinnen, K. Bernaerts, E.J.J. Dens, A.H. Geeraerd, and J.F. Van Impe. Predictive modelling of the microbial lag phase: a review. *International Journal of Food Microbiology*, 94(2):137 – 159, 2004.

[62] József Baranyi. Comparison of stochastic and deterministic concepts of bacterial lag. *Journal of Theoretical Biology*, 192(3):403 – 408, 1998.

[63] N. Bacaër. *Verhulst and the logistic equation (1838). In: A Short History of Mathematical Population Dynamics.* Springer, London, 2011.

[64] Takasu. Lecture 9: Logistic growth models. 2009.

[65] József Baranyi, Susan M. George, and Zoltán Kutalik. Parameter estimation for the distribution of single cell lag times. *Journal of Theoretical Biology*, 259(1):24 – 30, 2009.

[66] Judith Campisi and Fabrizio d'Adda di Fagagna. Cellular senescence: when bad things happen to good cells. *Nature Reviews Molecular Cell Biology*, 8:729–, September 2007.

[67] J A Kennedy and F Barabé. Investigating human leukemogenesis: from cell lines to in vivo models of human leukemia. *Leukemia*, 22:2029–, August 2008.

[68] Rebekah Bartelt, Noemi Cruz-Orcutt, Michaela Collins, and Jon Houtman. *Comparison of T Cell Receptor-Induced Proximal Signaling and Downstream Functions in Immortalized and Primary T Cells*, volume 4. February 2009.

[69] Robert T. Abraham and Arthur Weiss. Jurkat t cells and development of the t-cell receptor signalling paradigm. *Nature Reviews Immunology*, 4:301–, April 2004.

[70] Melanie Werner, D Schmoldt, Frank Hilbrig, Valerie Jerome, Alexander Raup, Kenny Zambrano, Holger Hübner, Rainer Buchholz, and Ruth Freitag. *High Cell Density Cultivation of Human Leukemia T Cells (Jurkat Cells) in Semipermeable Polyelectrolyte Microcapsules*, volume Eng. Life Sci. December 2014.

[71] Graham Pawelec, A. Borowitz, Peter H. Krammer, and Peter Wernet. Constitutive interleukin 2 production by the jurkat human leukemic t cell line. *Eur. J. Immunol.*, 12(5):387–392, 1982.

[72] S Venuta, R Mertelsmann, K Welte, SP Feldman, CY Wang, and MA Moore. Production and regulation of interleukin-2 in human lymphoblastic leukemias studied with t-cell monoclonal antibodies. *Blood*, 61(4):781–, April 1983.

[73] Qiangzhong Ma, Yawen Wang, Agnes Shuk-Yee Lo, Erica M. Gomes, and Richard P. Junghans. Cell density plays a critical role in ex vivo expansion of t cells for adoptive immunotherapy, 2010.

[74] P A Sandstrom and T M Buttke. Autocrine production of extracellular catalase prevents apoptosis of the human cem t-cell line in serum-free medium. *Proc Natl Acad Sci USA*, 90(10):4708–, May 1993.

[75] Woods Gonzalez. *Digital Image Processing (Third Edition)*. Pearson Education, 2008.

[76] Eddins Gonzalez, Woods. *Digital Image Processing Using MATLAB*. Pearson Education.

[77] József Baranyi and Terry A. Roberts. Mathematics of predictive food microbiology. *International Journal of Food Microbiology*, 26(2):199 – 218, 1995.

[78] Gaetano Cannelli. *Metodologie Sperimentali in Fisica, Introduzione al metodo scientifico*. 2010.

[79] M D'Arrigo, G D García de Fernando, R Velasco de Diego, J A Ordóñez, S M George, and C Pin. Indirect measurement of the lag time distribution of single cells of listeria innocua in food. *Applied and environmental microbiology*, 72(4):2533–2538, April 2006.

[80] P. A. Stephens, W. J. Sutherland, and R. P. Freckleton. What is the allee effect? 87(1):185–190, 1999.

[81] W. C. Allee. Studies in animal aggregations: Causes and effects of bunching in land isopods. *Journal of Experimental Zoology*, 45(1):255–277.

[82] Jim Cushing and Jarred T Hudson. Evolutionary dynamics and strong allee effects. *Journal of biological dynamics*, 6:941–58, 03 2012.

[83] Satya N Majumdar and Arnab Pal. *Extreme value statistics of correlated random variables Lectures by Satya N. majumdar, Notes by Arnab Pal*. June 2014.

[84] Srinivas S Somanchi, Kelsey J McCulley, Anitha Somanchi, Leo L Chan, and Dean A Lee. A novel method for assessment of natural killer cell cytotoxicity using image cytometry. *PLoS ONE*, 10(10):e0141074–, October 2015.

[85] M. C. Marchetti, J. F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, Madan Rao, and R. Aditi Simha. Hydrodynamics of soft active matter. *Rev. Mod. Phys.*, 85:1143–1189, Jul 2013.

[86] G. De Magistris and D. Marenduzzo. An introduction to the physics of active matter. *Physica A: Statistical Mechanics and its Applications*, 418:65–77, 2015.

[87] Lina Matera, Miguel Cutufia, Massimo Geuna, Maria Contarini, Stefano Buttiglieri, Shawn Galin, Annamaria Fazzari, and Cristina Cavaliere. Prolactin is an autocrine growth factor for the jurkat human t-leukemic cell line. *Journal of Neuroimmunology*, 79(1):12 – 21, 1997.

[88] Barry Canton, Anna Labno, and Drew Endy. Refinement and standardization of synthetic biological parts and devices. *Nature Biotechnology*, 26:787–, July 2008.

[89] Stefan Klumpp, Zhongge Zhang, and Terence Hwa. *Growth Rate-Dependent Global Effects on Gene Expression in Bacteria*, volume 139. December 2009.

[90] Ha Bremer and Patrick Dennis. *Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate*, volume 2. January 1996.

[91] David Lubertozzi and Jay D. Keasling. Developing aspergillus as a host for heterologous expression. *Biotechnology Advances*, 27(1):53–75, 2009.

[92] Julian N Rosenberg, George A Oyler, Loy Wilkinson, and Michael J Betenbaugh. A green light for engineered algae: redirecting metabolism to fuel a biotechnology revolution. *Current Opinion in Biotechnology*, 19(5):430–436, 2008.

[93] Edoardo M. Airoldi, Curtis Huttenhower, David Gresham, Charles Lu, Amy A. Caudy, Maitreya J. Dunham, James R. Broach, David Botstein, and Olga G. Troyanskaya. Predicting cellular growth from gene expression signatures. *PLOS Computational Biology*, 5(1):e1000257–, January 2009.

[94] Juan I. Castrillo, Leo A. Zeef, David C. Hoyle, Nianshu Zhang, Andrew Hayes, David CJ Gardner, Michael J. Cornell, June Petty, Luke Hakes, Leanne Wardleworth, Bharat Rash, Marie Brown, Warwick B. Dunn, David Broadhurst, Kerry O'Donoghue, Svenja S. Hester, Tom PJ Dunkley, Sarah R. Hart, Neil Swainston, Peter Li, Simon J. Gaskell, Norman W. Paton, Kathryn S. Lilley, Douglas B. Kell, and Stephen G. Oliver. Growth control of the eukaryote cell: a systems biology study in yeast. *Journal of Biology*, 6(2):4–, 2007.

[95] Jonathan Hickson, S Diane Yamada, Jonathan Berger, John Alverdy, James O'Keefe, Bonnie Bassler, and Carrie Rinker-Schaeffer. *Societal interactions in ovarian cancer metastasis: A quorum-sensing hypothesis*, volume 26. June 2008.

[96] Bart De Spiegeleer, Frederick Verbeke, Matthias DHondt, An Hendrix, Christophe Van De Wiele, Christian Burvenich, Kathelijne Peremans, Olivier De Wever, Marc Bracke, and Evelien Wynendaele. The quorum sensing peptides phrg, csp and edf promote angiogenesis and invasion of breast cancer cells in vitro. *PLOS ONE*, 10(3):e0119471–, March 2015.

[97] Davide Ruggero and Pier Paolo Pandolfi. Does the ribosome translate cancer? *Nature Reviews Cancer*, 3:179–, March 2003.

[98] Andrija Finka, Vishal Sood, Manfredo Quadroni, Paolo De Los Rios, and Pierre Goloubinoff. Quantitative proteomics of heat-treated human cells show an across-the-board mild depletion of housekeeping proteins to massively accumulate few hsps. *Cell Stress and Chaperones*, 20(4):605–620, 2015.

[99] I. Watanabe and S. Okada. Effects of temperature on growth rate of cultured mammalian cells (l5178y). *J Cell Biol*, 32(2):309–, February 1967.

[100] Shengming Zhu, Jiangang Wang, Bingkun Xie, Zhiguo Luo, Xiukun Lin, and D. Joshua Liao. Culture at a higher temperature mildly inhibits cancer cell growth but enhances chemotherapeutic effects by inhibiting cell-cell collaboration. *PLOS ONE*, 10(10):e0137042–, October 2015.

[101] Luigi Rovati, Paola Fabbri, Luca Ferrari, and Francesco Pilati. *Plastic Optical Fiber pH Sensor Using a Sol-Gel Sensing Matrix*. February 2012.

[102] Rosalind C. Lee, Rhonda L. Feinbaum, and Victor Ambros. The c. elegans heterochronic gene lin-4 encodes small rnas with antisense complementarity to lin-14. *Cell*, 75(5):843–854, 1993.

[103] David P Bartel. Micrornas: Genomics, biogenesis, mechanism, and function. *Cell*, 116(2):281–297, 2004.

[104] Nicolas E. Buchler and Matthieu Louis. Molecular titration and ultrasensitivity in regulatory networks. *Journal of Molecular Biology*, 384(5):1106–1119, 2008.

[105] Azra Krek, Dominic Grün, Matthew N Poy, Rachel Wolf, Lauren Rosenberg, Eric J Epstein, Philip MacMenamin, Isabelle da Piedade, Kristin C Gunsalus, Markus Stoffel, and Nikolaus Rajewsky. Combinatorial microrna target predictions. *Nature Genetics*, 37:495–, April 2005.

[106] Leonardo Salmena, Laura Poliseno, Yvonne Tay, Lev Kats, and Pier Paolo Pandolfi. A cerna hypothesis: the rosetta stone of a hidden rna language? *Cell*, 146(3):353–358, August 2011.

[107] Ugo Ala, Florian A. Karreth, Carla Bosia, Andrea Pagnani, Riccardo Taulli, Valentine Léopold, Yvonne Tay, Paolo Provero, Riccardo Zecchina, and Pier Paolo Pandolfi. Integrated transcriptional and competitive endogenous rna networks are cross-regulated in permissive molecular environments. *Proc Natl Acad Sci USA*, 110(18):7154–, April 2013.

[108] Mads Kærn, Timothy C. Elston, William J. Blake, and James J. Collins. Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics*, 6:451–, May 2005.

[109] Marina Bessarabova, Eugene Kirillov, Weiwei Shi, Andrej Bugrim, Yuri Nikolsky, and Tatiana Nikolskaya. Bimodal gene expression patterns in breast cancer. *BMC Genomics*, 11(1):S8–, 2010.

[110] David P. Bartel. Micrornas: Target recognition and regulatory functions. *Cell*, 136(2):215–233, 2009.

[111] Allan M. Gurtan and Phillip A. Sharp. The role of mirnas in regulating gene expression networks. *Journal of Molecular Biology*, 425(19):3582–3600, 2013.

[112] David P. Bartel. Metazoan micrornas. *Cell*, 173(1):20–51, 2018.

[113] Davide Cora', Angela Re, Michele Caselle, and Federico Bussolino. Microrna-mediated regulatory circuits: outlook and perspectives. 14(4):045001–, 2017.

[114] Divya Lenkala, Bonnie LaCroix, Eric R. Gamazon, Paul Geeleher, Hae Kyung Im, and R. Stephanie Huang. The impact of microrna expression on cellular proliferation. *Human Genetics*, 133(7):931–938, Jul 2014.

[115] Yong Zhao and Deepak Srivastava. A developmental view of microrna function. *Trends in Biochemical Sciences*, 32(4):189 – 197, 2007.

[116] Zhenyi Su, Zuozhang Yang, Yongqing Xu, Yongbin Chen, and Qiang Yu. Micrornas in apoptosis, autophagy and necroptosis. *Oncotarget*, 6(11):8474–8490, April 2015.

[117] Martin D. Jansson and Anders H. Lund. Microrna and cancer. *Molecular Oncology*, 6(6):590–610.

[118] Mickael Leclercq, Abdoulaye Baniré Diallo, and Mathieu Blanchette. Prediction of human mirna target genes using computationally reconstructed ancestral mammalian sequences. *Nucleic acids research*, 45(2):556–566, January 2017.

[119] Ramiro Garzon, George A. Calin, and Carlo M. Croce. Micrornas in cancer. *Annu. Rev. Med.*, 60(1):167–179, February 2009.

[120] Dmitri Volfson, Jennifer Marciniak, William J. Blake, Natalie Ostroff, Lev S. Tsimring, and Jeff Hasty. Origins of extrinsic variability in eukaryotic gene expression. *Nature*, 439(7078):861–864, 2006.

[121] M T Pickering, B M Stadler, and T F Kowalik. mir-17 and mir-20a temper an e2f1-induced g1 checkpoint to regulate cell cycle progression. *Oncogene*, 28:140–, October 2008.

[122] Zhou Jueyu, Wen-Li Ma, Shuang Liang, Ye Zeng, Rong Shi, Hai-Lang Yu, Wei-Wei Xiao, and Wen-Ling Zheng. *Analysis of microRNA expression profiles during the cell cycle in synchronized HeLa cells*, volume 42. September 2009.

[123] Erel Levine, Zhongge Zhang, Thomas Kuhlman, and Terence Hwa. Quantitative characteristics of gene regulation by small rna. *PLOS Biology*, 5(9):e229–, August 2007.

[124] Laura Poliseno, Leonardo Salmena, Jiangwen Zhang, Brett Carver, William J. Haveman, and Pier Paolo Pandolfi. A coding-independent function of gene and pseudogene mrnas regulates tumour biology. *Nature*, 465:1033–, June 2010.

[125] Araks Martirosyan, Marco Del Giudice, Chiara Enrico Bena, Andrea Pagnani, Carla Bosia, and Andrea De Martino. Kinetic modelling of competition and depletion of shared mirnas by competing endogenous rnas. In *Computational Biology of Non-Coding RNA: Methods and Protocols*, pages 367–409. Springer New York, New York, NY, 2019.

[126] N.G. van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland Personal Library, 1981.

[127] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.

[128] Domitilla Del Vecchio, Alexander J Ninfa, and Eduardo D Sontag. Modular cell biology: retroactivity and insulation. *Mol Syst Biol*, 4(1):–, January 2008.

[129] Lewis J et al. Alberts B, Johnson A. *Molecular Biology of the Cell. 4th edition. - An Overview of the Cell Cycle*. New York: Garland Science;, 2002.

[130] Filippos Porichis, Meghan G. Hart, Morgane Griesbeck, Holly L. Everett, Muska Hassan, Amy E. Baxter, Madelene Lindqvist, Sara M. Miller, Damien Z. Soghoian, Daniel G. Kavanagh, Susan Reynolds, Brett Norris, Scott K. Mordecai, Quan Nguyen, Chunfai Lai, and Daniel E. Kaufmann. High-throughput detection of mirnas and gene-specific mrna at the single-cell level by flow cytometry. *Nature Communications*, 5:5641–, December 2014.

[131] Meiye Wu, Matthew Piccini, and Anup Singh. *MiRNA detection at single-cell resolution using microfluidic LNA flow-FISH*, volume 1211. September 2014.

[132] Jennifer Dumont, Don Euwart, Baisong Mei, Scott Estes, and Rashmi Kshirsagar. Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Critical Reviews in Biotechnology*, 36(6):1110–1122, November 2016.

[133] Michel Chartrain and Lily Chu. Development and production of commercial therapeutic monoclonal antibodies in mammalian cell expression systems: An overview of the current upstream technologies, 2008.

[134] Lorenzo Galluzzi, Erika Vacchelli, Wolf Hervé Fridman, Jerome Galon, Catherine Sautès-Fridman, Eric Tartour, Jessica Zucman-Rossi, Laurence Zitvogel, and Guido Kroemer. Trial watch: Monoclonal antibodies in cancer therapy. *OncoImmunology*, 1(1):28–37, January 2012.

[135] Jennifer Dine, RuthAnn Gordon, Yelena Shames, Mary Kate Kasler, and Margaret Barton-Burke. Immune checkpoint inhibitors: An innovation in immunotherapy for the treatment and management of patients with cancer. *Asia-Pacific Journal of Oncology Nursing*, 4(2):127–135, January 2017.

[136] Belinda Sánchez Ramírez, Eduardo Suárez Pestana, Greta Garrido Hidalgo, Tays Hernández García, Rolando Pérez Rodríguez, Axel Ullrich, and Luis Enrique Férnandez. Active antimetastatic immunotherapy in lewis lung carcinoma with self egfr extracellular domain protein in vssp adjuvant. *Int. J. Cancer*, 119(9):2190–2199, November 2006.

[137] Belinda Sánchez, Yeranddy Aguiar, Diana R Hernández, Greta Garrido, Rolando Pérez, and Luis E Fernádez. HER1 Vaccine: An autologous EGFR vaccine candidate to treat epithelial tumors. *Biotecnología Aplicada*, 26:342 – 344, 12 2009.

[138] Mette K. Nedergaard, Chris J. Hedegaard, and Hans S. Poulsen. Targeting the epidermal growth factor receptor in solid tumor malignancies. *BioDrugs*, 26(2):83–99, 2012.

[139] Iraida Caballero, Lazaro E Aira, Anabel Lavastida, Xitlally Popa, Javier Rivero, Joaquín González, Mónica Mesa, Narjara González, Kelly Coba, Patricia Lorenzo-Luaces, Barbara Wilkinson, Yuliannis Santiesteban, Yanela Santiesteban, Mayelin Troche, Eduardo Suarez, Tania Crombet, Belinda Sánchez, Angel Casacó, Amparo Macías, and Zaima Mazorra. Safety and immunogenicity of a human epidermal growth factor receptor 1 (her1)-based vaccine in prostate castration-resistant carcinoma patients: A dose-escalation phase i study trial. *Frontiers in pharmacology*, 8:263–263, May 2017.

[140] Katia Garcia Duardo, Yadira Prieto Curbelo, Judith Raymond Pous, Estela Yamilet Rabasa Legón, Belinda Sánchez Ramírez, Kathya Rashida de la Luz Hernández, and Adolfo Castillo Vitoch. Assessment of the impact of manufacturing changes on the physicochemical properties and biological activity of her1-ecd vaccine during product development. *Vaccine*, 33(35):4292–4299, 2015.

[141] Dania León, Yadira Prieto, Eutimio G. Fernández, Noemí Pérez, José A. Montero, Julio Palacios, Dubhe Bulté, Kathya R. de la Luz, Vladimir Peña, Williams Ferro, Belinda Sánchez, Rodolfo Valdés, and Adolfo Castillo. Purification process development for her1 extracellular domain as a potential therapeutic vaccine. *Journal of Chromatography B*, 877(27):3105–3110, 2009.

[142] Peter Rugbjerg, Kira Sarup-Lytzen, Mariann Nagy, and Morten Otto Alexander Sommer. Synthetic addiction extends the productive life time of engineered escherichia coli populations. *Proc Natl Acad Sci USA*, 115(10):2347–, March 2018.

[143] Juan Sun, Lifu Bi, Yaojun Chi, Kazuo Aoki, and Junichi Misumi. Effect of sodium acetate on cell proliferation and induction of proinflammatory cytokines: A preliminary evaluation. *Food and Chemical Toxicology*, 43(12):1773–1780, 2005.

[144] Cameron C. Scott, Fabrizio Vacca, and Jean Gruenberg. Endosome maturation, transport and functions. *Seminars in Cell & Developmental Biology*, 31:2–10, 2014.

[145] Juan S Bonifacino and Jacques Neefjes. Moving and positioning the endolysosomal system. *Current Opinion in Cell Biology*, 47:1–8, 2017.

[146] Jacques Neefjes, Marlieke M.L. Jongsma, and Ilana Berlin. Stop or go? endosome positioning in the establishment of compartment architecture, dynamics, and function. *Trends in Cell Biology*, 27(8):580–594, 2017.

[147] Melissa J. Phillips and Gia K. Voeltz. Structure and function of er membrane contact sites with other organelles. *Nature Reviews Molecular Cell Biology*, 17:69–, December 2015.

[148] Arti V Shinde, Rajender K Motiani, Xuexin Zhang, Iskandar F Abdullaev, Alejandro P Adam, José C González-Cobos, Wei Zhang, Khalid Matrougui, Peter A Vincent, and Mohamed Trebak. Stim1 controls endothelial barrier function independently of orai1 and ca2+ entry. *Science signaling*, 6(267):ra18–ra18, March 2013.

[149] Feng-Chiao Tsai, Akiko Seki, Hee Won Yang, Arnold Hayer, Silvia Carrasco, Seth Malmersjö, and Tobias Meyer. A polarized ca(2+), diacylglycerol, and stim1 signaling system regulates directed cell migration. *Nature cell biology*, 16(2):133–144, January 2014.

[150] Hong-Bo Pang, Gary B Braun, Tomas Friman, Pedro Aza-Blanc, Manuel E Ruidiaz, Kazuki N Sugahara, Tambet Teesalu, and Erkki Ruoslahti. An endocytosis pathway initiated through neuropilin-1 and regulated by nutrient availability. *Nature communications*, 5:4904–4904, October 2014.

[151] Amy L. Herbert, Meng-meng Fu, Catherine M. Drerup, Ryan S. Gray, Breanne L. Harty, Sarah D. Ackerman, Thomas O'Reilly-Pol, Stephen L. Johnson, Alex V. Nechiporuk, Ben A. Barres, and Kelly R. Monk. Dynein/dynactin is necessary for anterograde transport of mbp mrna in oligodendrocytes and for myelination in vivo. *Proceedings of the National Academy of Sciences*, 114(43):E9153–E9162, 2017.